

DATAGUISE

DgSecure - User Guide

Version 7.2

Copyright © 2020

Dataguise, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, or otherwise—without the written permission of Dataguise, Inc.

Trademarks

Dataguise, the Dataguise logo, DgSecure, DgDiscover, DgMasker, and DgDashboard are registered trademarks of Dataguise, Inc. All other trademarks are property of their respective owners.

Changes

The material in this document is for information only and is subject to change without notice. While reasonable efforts have been made in the preparation of this document to assure its accuracy, Dataguise makes no representation or warranty expressed or implied, as to its completeness, accuracy, or suitability, and assumes no liability resulting from errors or omissions in this document or from the use of the information contained herein. Dataguise reserves the right to make changes in the product design without reservation and without notification to its users.

Dataguise, Inc.
39650 Liberty Street, Suite 400

Fremont, CA 94538
877-632-0522

<http://www.dataguise.com>

Contents

1	Preface	10
1.1	Introduction of DgSecure	10
1.2	Logging In	11
1.2.1	Home Page	11
1.3	Sign Out	18
1.4	Related Source	19
2	Getting Started	20
2.1	Detection	20
2.2	Sampling configuration	20
2.3	Masking	20
2.4	Encryption/Decryption	20
2.5	Privacy	20
2.6	Monitoring	20
3	Dashboard	21
4	Connection Manager	23
4.1	Concept	23
4.2	Create a Connection	23
4.2.1	RDBMS	23
4.2.2	NoSQL	53
4.2.3	AWS	57
4.2.4	Azure	66
4.3	List a Connection	72
4.3.1	RDBMS	72
4.3.2	NoSQL	76
4.3.3	AWS	77
4.3.4	Azure	77
5	Sensitive Type	78
5.1	Default Sensitive Type	78
5.2	New Sensitive Type	80
5.3	Inherit Sensitive Type	86
5.4	Edit Sensitive Type	89
5.5	Confidence Factor	89
6	Policy	91

6.1	Concept	91
6.2	Create a Policy.....	92
6.2.1	DBMS.....	92
6.2.2	Hadoop & Files	94
6.3	List Policies	95
6.3.1	DBMS.....	95
6.3.2	Hadoop & Files	97
6.4	Export Policy.....	99
6.5	Import Policy	100
7	Task	102
7.1	Concept	102
7.1.1	Source System	103
7.1.2	Task Type.....	103
7.2	Create Task.....	104
7.2.1	RDBMS	104
7.2.2	NoSQL.....	132
7.2.3	Hadoop.....	138
7.2.4	Files	158
7.2.5	AWS.....	168
7.2.6	Azure	203
7.2.7	Google Cloud.....	238
7.2.8	SharePoint.....	244
7.3	List a Task	249
7.3.1	RDBMS	249
7.3.2	Hadoop.....	253
7.3.3	NoSQL.....	257
7.3.4	Files	259
7.3.5	Azure	261
7.3.6	AWS.....	263
7.3.7	Google Cloud.....	270
7.3.8	SharePoint.....	272
7.4	Masking Options	275
7.4.1	Static Mask.....	275
7.4.2	Character Mask	276

7.4.3	Format Preservation Mask (FPM)	276
7.4.4	IntelliMask Mask	277
7.4.5	Random Mask	277
7.4.6	NPI Mask	279
7.4.7	Compose Mask.....	280
7.4.8	Compose Math Expression Mask.....	280
7.4.9	Date Synch Mask.....	281
7.4.10	Name Synch Mask.....	281
7.4.11	Email Policy Mask.....	281
7.4.12	Expression Mask.....	282
7.4.13	Full Name Mask.....	282
7.4.14	Regular Expression Mask	283
7.4.15	Custom Lookup Mask.....	284
7.4.16	Custom Mask.....	285
7.4.17	Shuffle Mask.....	288
7.4.18	JSON Mask.....	288
7.4.19	XML Mask.....	288
7.4.20	AES Encryption/Decryption.....	289
7.4.21	FPE encryption/Decryption	289
7.4.22	Partial Field FPM	290
7.4.23	Custom Masking.....	290
7.4.24	Custom Transformation	291
8	Scheduler	292
8.1	Schedule a Task.....	292
8.2	List a Scheduler	293
9	Results.....	296
9.1	RDBMS	296
9.1.1	Detection.....	296
9.1.2	Masking.....	302
9.2	NoSQL.....	304
9.3	Hadoop.....	306
9.3.1	HDFS.....	306
9.3.2	Hive	317
9.3.3	Hbase	327

9.4	Files	328
9.4.1	By Task	329
9.4.2	By Date Range	333
9.4.3	Saved Remediation	335
9.5	AWS.....	336
9.5.1	S3.....	336
9.5.2	RedShift/RDS.....	343
9.6	Azure	351
9.6.1	Azure Blob/Data Lake.....	351
9.6.2	Databases.....	355
9.7	Google Cloud.....	361
9.7.1	By Task	362
9.7.2	By Date Range	366
9.8	SharePoint.....	368
9.8.1	By Task	368
10	Reports.....	371
10.1	Overview	371
10.1.1	Export as PDF	388
10.2	Periodic Reports.....	388
10.3	Tableau Reports	389
10.4	GDPR View	391
10.4.1	Server Details	398
10.5	RoA & RtE.....	399
10.6	DgMonitor.....	403
10.7	Comprehensive Reports.....	404
10.8	Audit Reports	406
10.8.1	By Users.....	408
10.8.2	By Events.....	411
10.9	Hadoop.....	412
10.9.1	Entitlement Report	412
10.9.2	Time Based Report.....	422
10.10	Exported Reports	424
11	Monitor	425
11.1	Overview	425

11.2	Alert Rules	427
11.3	Alerts	429
12	Attributes	431
12.1	Concept	431
12.2	Attribute	431
12.2.1	Create an Attribute	431
12.2.2	Edit an Attribute	433
12.2.3	List an Attribute	436
12.3	Attribute Assignment	439
12.3.1	RDBMS	439
12.3.2	Hadoop	441
13	Structure Management	443
13.1	Concept	443
13.2	Create a Structure	443
13.2.1	RDBMS	443
13.2.2	Hadoop	445
13.2.3	Files	455
13.2.4	AWS	455
13.2.5	Azure	456
13.3	Edit a Structure	456
13.3.1	RDBMS	456
13.3.2	Hadoop	457
13.3.3	Files	467
13.3.4	AWS	467
13.3.5	Azure	467
13.4	List a Structure	467
13.4.1	RDBMS	467
13.4.2	Hadoop	469
13.4.3	Files	474
13.4.4	AWS	474
13.4.5	Azure	474
14	Output Directory	475
14.1	Concept	475
14.2	Create a Source Directory	475

14.2.1	Hadoop.....	475
14.2.2	Files	477
14.2.3	AWS.....	477
14.2.4	Azure	477
14.3	Edit a Source Directory	477
14.3.1	Hadoop.....	477
14.3.2	Files	478
14.3.3	AWS.....	478
14.3.4	Azure	479
14.4	List an Output Directory.....	479
14.4.1	Hadoop.....	479
14.4.2	Files	479
14.4.3	AWS.....	480
14.4.4	Azure	480
15	Access Control.....	481
15.1	ACL Management.....	481
15.2	Role Management.....	482
16	Privacy.....	485
16.1	Concept.....	485
16.2	Components of DSAR and Privacy	486
16.3	DSAR Workflow - Overview.....	487
16.4	DSAR.....	489
16.4.1	Add IDPs.....	490
16.4.2	Add/Edit Identifier	490
16.4.3	Add Data Group	493
16.4.4	Create Connection - DSAR.....	495
16.4.5	Add Data Subject.....	498
16.4.6	Create Task.....	506
16.4.7	Build Metadata and Use API to Populate Quick Scan Columns	507
16.4.8	Schedule RoA/RtE	510
16.4.9	Request RoA/RtE.....	512
16.4.10	View Report.....	515
16.4.11	Search Obligation	519
16.5	Configure Privacy	520

16.5.1	Add IDPs	521
16.5.2	Add/Edit Identifier	521
16.5.3	Add Data Group	524
16.5.4	Add Data Subject.....	525
16.5.5	Create Connection – Privacy	532
16.5.6	Add System	536
16.5.7	Create Task.....	539
16.5.8	Execute Task (As per the Schedule)	544
16.5.9	View Report.....	546
Appendix A: Verifying Hadoop Results		550
Task Results.....		550
Detailed Results		551
Appendix B: EDI Transaction Set.....		553
Appendix C: Error Messages		555
Appendix D: Role Based Access Controls.....		558
Appendix E: Voltage SimpleAPI.....		559
Appendix E: Snappy Files Support.....		561
Appendix F: Lightweight Primitives.....		562
Appendix G: Key-Pair Authentication for Snowflake Connections		564
Appendix H: Detection in Kerberized Clusters.....		568
Appendix I: SL Masking		569
Configure SL Masking for Kerberos User in Oracle		569
Enabling SL Masking in Oracle		570
Enable Random-SL Masking Support for HBase		571
Appendix J: Configure Multiple SQL Server Instances in DgSecure		575
Appendix K: Confidence Factor.....		577

1 Preface

1.1 Introduction of DgSecure

Dataguisse DgSecure is a complete data security solution that enables enterprises to leverage their data to achieve greater business goals while minimizing the risk of exposure and running afoul of data handling regulations such as PII, PCI, HIPAA and GDPR.



DgSecure provides core solution for finding Sensitive data (DETECT), encrypting or masking (PROTECT) it, providing an overall report (AUDIT) and observing the progress over a period of time (MONITOR).

- **DETECT**
DgSecure DETECT capability enables you to discover, count and report on Sensitive data encountered in the database; On-Premises and in the Cloud based. It processes structured, semi-structured and unstructured data formats.
- **PROTECT**
The PROTECT capability ensures that the data is protected. There is an automated, policy based encryption for sensitive data for various file formats. The sensitive data gets encrypted with fictitious content using one of many available data replacement options.
- **AUDIT**
It understands and assess the security risks across the enterprise and in the cloud. The AUDIT report display the details about all the events which were executed by the user, based on the role.
- **MONITOR**
DgSecure MONITOR offers an early warning system focused solely on safeguarding the sensitive data. It tracks how and where Sensitive data is being accessed. It also generates alerts.

1.2 Logging In

To log in to the DgSecure, follow the below steps:

1. Click the DgSecure link: https://<ip_address>/dgadmin/login.html
2. Enter your credentials.

3. Click the **Sign In** button.

1.2.1 Home Page

The Home page displays all task instances the user is authorized to see along with any associated notifications.

Task Instance ID	Task Type	Task Name	Cluster	Status	Start Time	End Time
5	Hadoop	Test_detection	Hadoop Cluster	Paused at 0.00%	Dec-11-2019 04:29:19	Dec-11-2019 04:32:08
7	Hadoop	Test_detection	Hadoop Cluster	Paused at 0.00%	Dec-11-2019 04:42:18	Dec-11-2019 04:44:08
8	Hadoop	Test_detection	Hadoop Cluster	Paused at 0.00%	Dec-11-2019 04:44:31	Dec-11-2019 05:08:54
15	Hadoop	Test_detection	Hadoop Cluster	Paused at 0.00%	Dec-11-2019 05:09:28	Dec-11-2019 05:17:57
16	Hadoop	Test_detection	Hadoop Cluster	Paused at 0.00%	Dec-11-2019 05:18:38	Dec-11-2019 05:21:29
17	Hadoop	Test_detection	Hadoop Cluster	Paused at 0.00%	Dec-11-2019 05:21:36	Dec-11-2019 05:30:21
6	Hadoop	Test_detection	Hadoop Cluster	Failed at 0.00%	Dec-11-2019 04:32:20	Dec-11-2019 04:36:47
18	Hadoop	Test_detection	Hadoop Cluster	Failed at 0.00%	Dec-11-2019 05:32:05	Dec-11-2019 05:39:26
1	Azure Data	Test_detection	Azure Cluster	Failed at 0.00%	Dec-11-2019 02:00:10	Dec-11-2019 02:10:51
2	Azure Data	Detection_2	Azure Cluster	Failed at 0.00%	Dec-11-2019 02:04:21	Dec-11-2019 02:10:51

Notification Description	Data Source	Notification Time	Notification Details	Select All
RDBMS Detection job completion notification	RDBMS	Feb-10-2020 02:11:01	Click here for details	<input type="checkbox"/>
RDBMS start job notification	RDBMS	Feb-10-2020 02:10:41	Click here for details	<input type="checkbox"/>
RDBMS cancel job notification	RDBMS	Feb-10-2020 02:08:59	Click here for details	<input type="checkbox"/>
RDBMS Detection job completion notification	RDBMS	Feb-10-2020 02:04:11	Click here for details	<input type="checkbox"/>

The page is divided into two panes. These are:

1. Task Details:

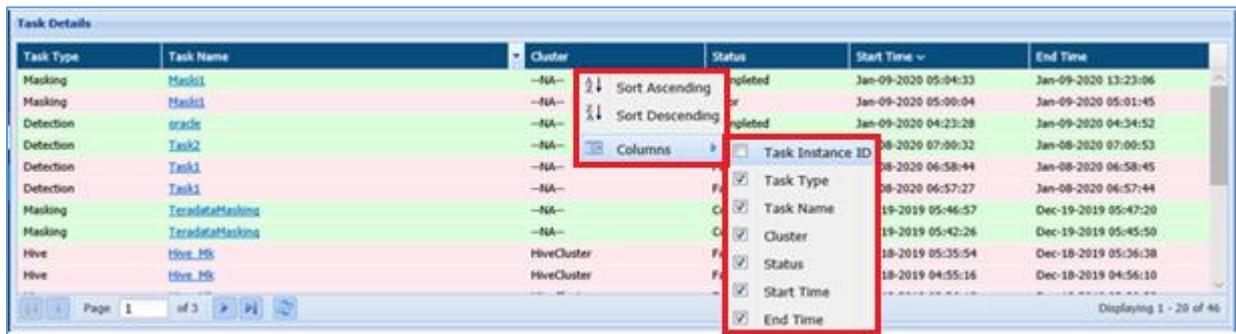
A detailed information about the task will be displayed on this pane. The details about the task includes Task Instance ID, Task Type, Task Name, Cluster, Status, Start Time and End Time.

Additionally, you can sort the values in columns in ascending and descending order and can also view the list of columns headers in the pane.

If you click on column header, following options will appear:

- a. **Sort Ascending** – Click on the Sort Ascending option on column header to sort the column values in ascending order, i.e. A-Z.
- b. **Sort Descending** - Click on the Sort Descending option on column header to sort the column values in descending order, Z-A.
- c. **Columns** – Click the Columns option to include or remove any column from this pane. When you click on Columns option, it will display the list of all Column headers. By default, all the columns are checked.

To remove a column from the pane, check the checkbox next to the column header name.



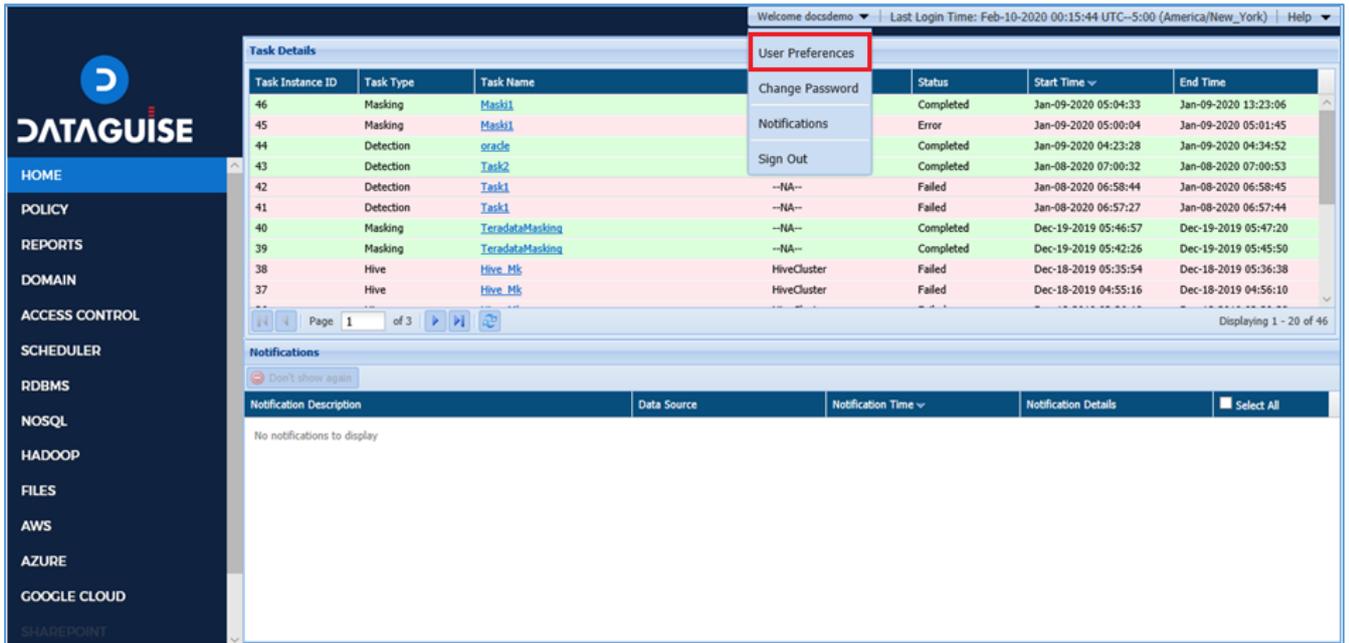
2. Notifications:

This pane shows the event based notifications related to the selected tasks. Notifications are either event based or time based. The details include Notification Description, Data Source, Notification Time, Notification Details.

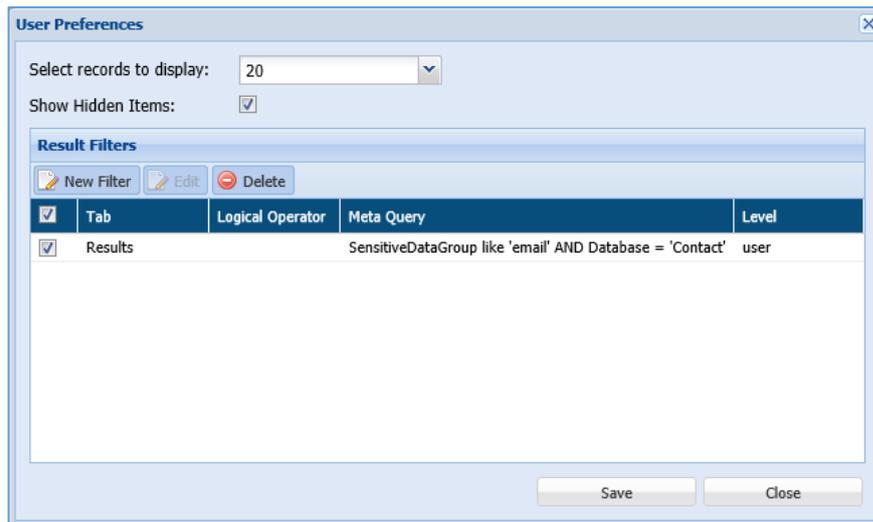
1.2.1.1 User Preferences

User Preferences allows you to create/edit/delete and update the filters. You can also unhide the items by checking the Show Hidden Item checkbox.

To access User Preferences dialog box, Click **Welcome <username>** drop-down in **Home** Page. Click User Preferences.



1. Click the User Preferences option from the drop-down. It will open User Preferences dialog box.



- a) **Select records to display:** Select the number of records to be displayed on the Home page.
- b) **Show Hidden Item:** Check this checkbox to display the hidden tasks on the screen.
- c) **Result Filters:** It will display the list of filters created by you. Using Preferences option you can create, update, delete and edit filters. Follow the below steps to create a filter:

The screenshot shows a 'New Filter' dialog box with the following fields and values:

- Tab: Results
- Logical Operator: No Operator
- Filter Level: user
- Meta Query: SensitiveDataGroup like 'email' AND Database = 'Contact'

Buttons: Save, Cancel

- i. Click the **New Filter** button.
- ii. Enter the details:
 1. **Tab** name where the filter will be applied. For Example, Results, Detailed Results, Skipped Results, etc.
 2. **Logical Operator** is used when new filter is appended with previously created filter. If there is no previous filter, then do not apply logical operator.

For Example: if we have Filter F1 and creating a New Filter F2, then the results for both filters will be combined.

3. **Filter Level:** A user can select the User or System filter in Filter Level drop-down.
4. **MetaQuery:** you can use the defined keywords to create Meta Query. Meta Query will be parsed to the actual query.

To create a Meta Query remove the space and special characters from the header names and use relational operators such as '=' (Equals), '!=' (Not Equals), < (Less Than), > (Greater Than) and Like in between the query.

For Example: if we want to add a filter where Sensitive Type should contain address and table/view should be equal to Base table. Then query will be:

SensitiveType Like 'Address' AND Database = 'Contact'

- iii. **Edit:** Click on **Edit** button. It will allow you to edit a filter.
- iv. **Delete:** Check the checkbox against the filter which you wish to delete. Click the **Delete** button.

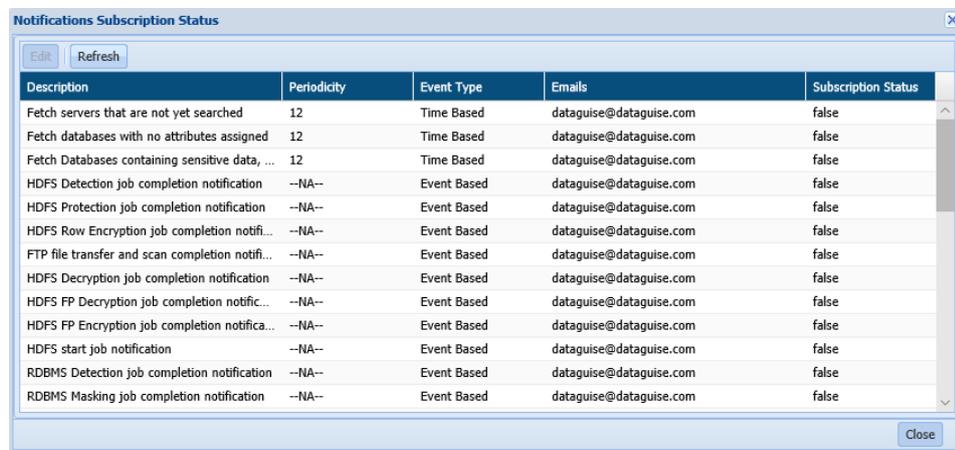
Note: Keywords for Meta Query

Collection	DatabaseUser	FilePath	NullRatio	SampledDate
Column	DataType	FileSize	ObjectPath	SampledRows
ColumnName	DetectionType	FileType	ObjectType	SampleMode
Confidence	DirectoryPath	HitCount	QuickSearch	SensitiveDataGroup
ConnectionName	Error	HostName	ReferentialType	SensitiveDataType
Content Read	FieldName	KeyPath	ReferringTo	SensitiveGroupConfidence
Database	FieldNameMatch	Masked	RowScanned	SensitiveType
DatabaseName	FieldNo	MatchCount	RowsScanned	SkippedReasons
DatabaseType	FileName	NullCount	Safe	Table
TableName	TableSize	TableView	TaskName	TotalData
TotalRows	ValuesScanned			

1.2.1.2 Notifications

In here, you can define the notifications for logged in user in DgAdmin. The Notifications are either event or time based. You can also define the periodicity of the notifications in DgAdmin.

Once you have defined notifications in DgAdmin, you can select to receive it by clicking on Notifications in the drop-down on top right corner of the page.

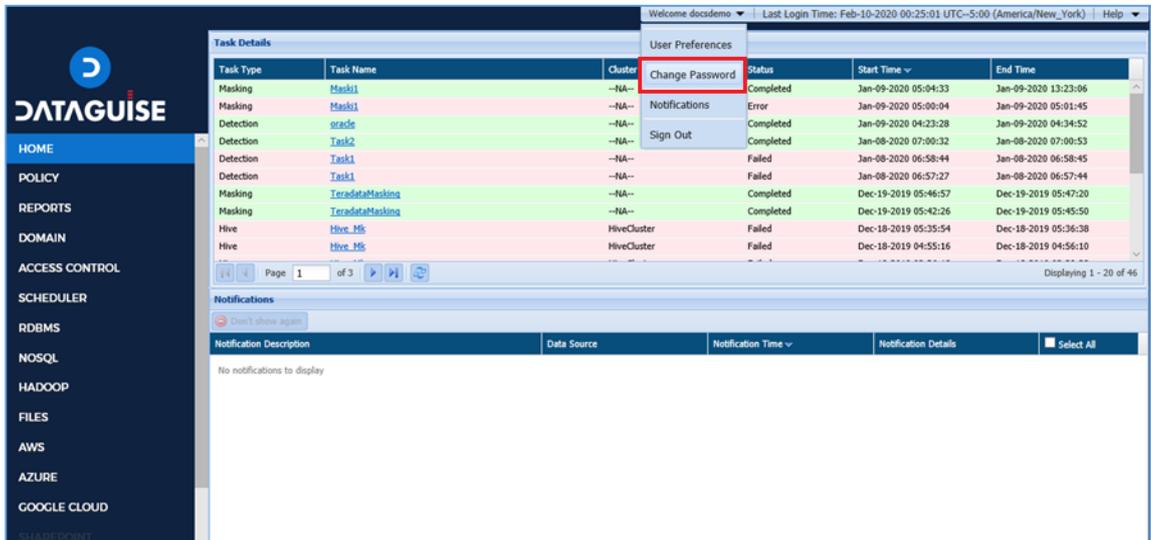


To activate a notification, you need to change the value in the Subscription status to 'True'.

1.2.1.3 Password Change

To change password. Click the **Welcome <username>** drop-down, follow below steps:

1. Go to **Welcome <username>**. Click **Change Password** from the drop-down.



Change Password

Current Password:

New Password:

Confirm Password:

- i. Click the **Change Password**.
- ii. Enter your **Current Password**.
- iii. Enter your **New Password**.

***Note:** Password is case- sensitive.

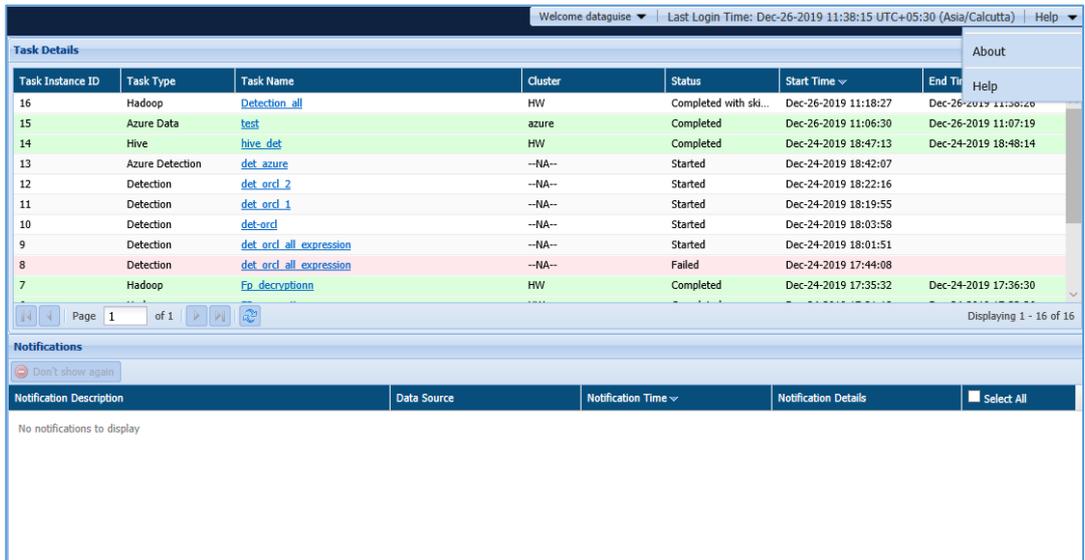
- iv. Confirm your **New Password** by entering it again in **Confirm Password** textbox.
- v. Click **Change Password** button to set the new password as your credentials.
- vi. Click **Cancel** button, if you do not want to save the changes.

1.2.1.4 Help

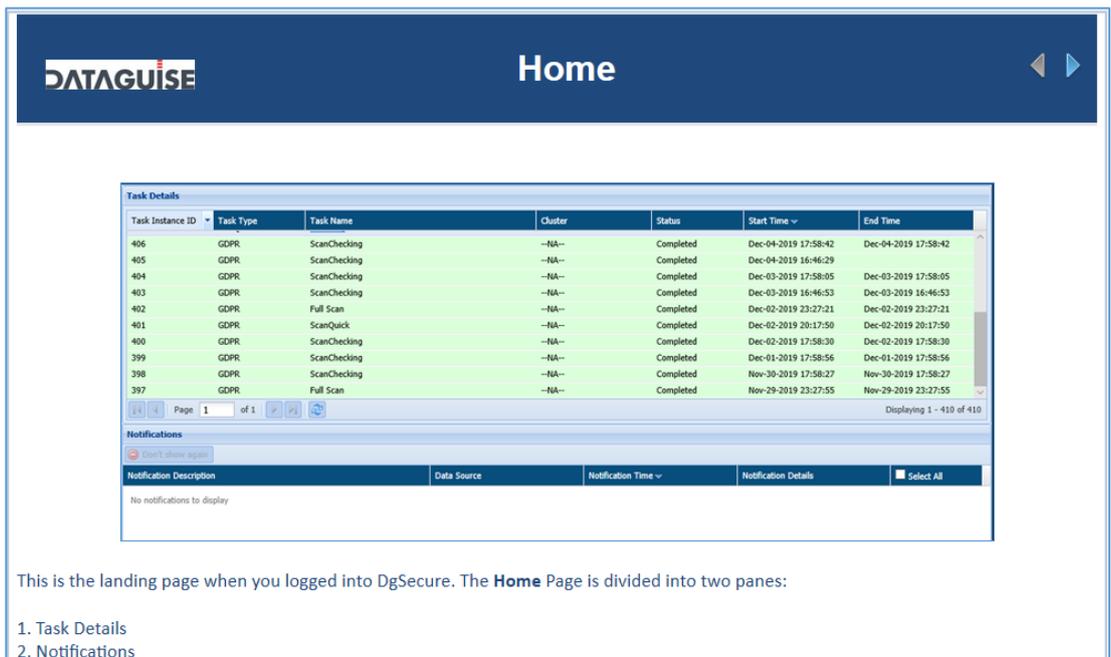
A screen specific discussion of the DgSecure application is found in the Online Help. The Online Help can be accessed from any page in DgSecure via Help button.

To access the Online Help, follow the below steps:

1. Go to the top right corner of the application. Click the Help Button.



2. Click the Help option. This will open a new window in the browser which will display screen specific information.



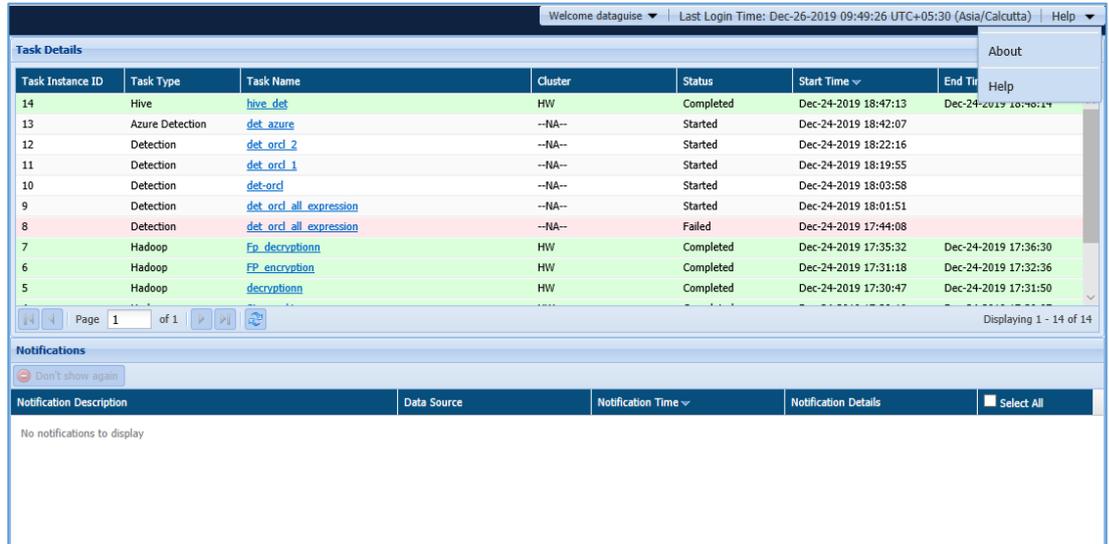
This is the landing page when you logged into DgSecure. The Home Page is divided into two panes:

1. Task Details
2. Notifications

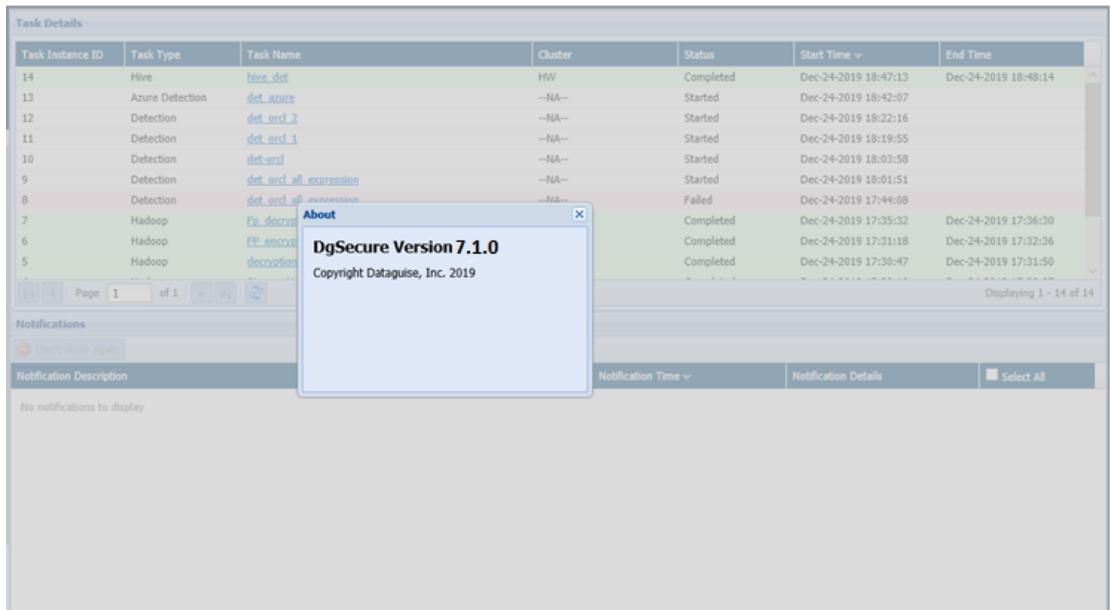
1.2.1.5 About

The About option give the update about the application version. To see the version of the application, follow the below steps.

1. Go to the top right corner of the application. Click the Help drop-down.



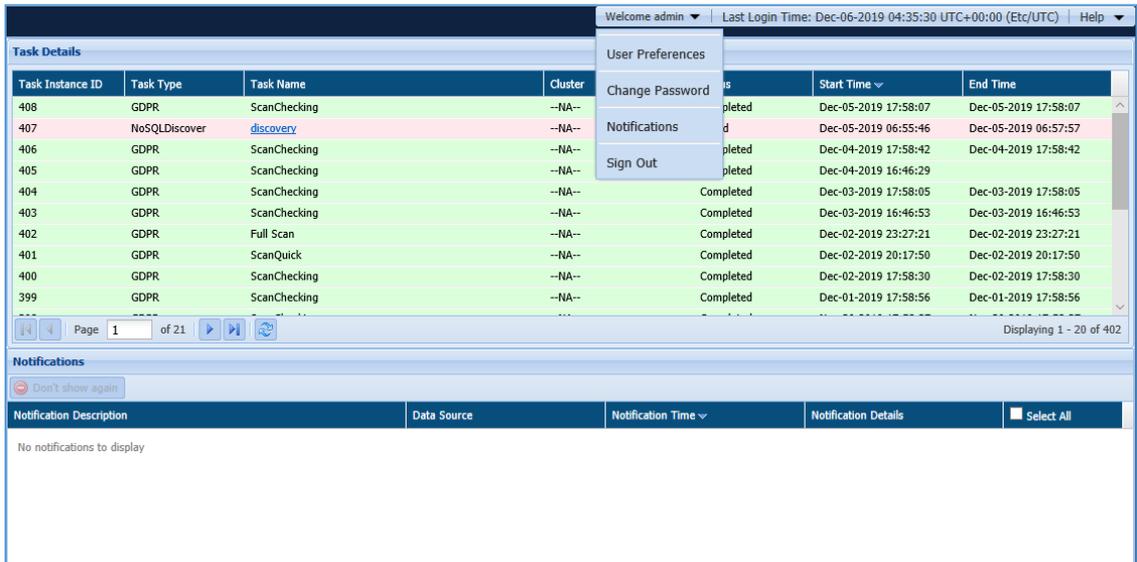
2. Click the About option. It will show the current version of the DgSecure application.



1.3 Sign Out

To sign-out from the DgSecure application, follow the below steps:

1. Go to the top right corner of the application. Click the **Welcome <username>** drop-down.



2. Click the **Sign-Out** option. This option lets your sign-out from the DgSecure application.

1.4 Related Source

This document is intended to support user deployment of DgSecure on the customer site. Other documents are also available:

- **DgSecure Installation and Configuration Guide** – This document walks through how to install and configure DgSecure on both single and multi-node systems. It also documents the DgSecure Admin controller.
- **Online Help** – Accessible from any screen in DgSecure, it shares in-depth information on each screen as well as context specific tips.

2 Getting Started

2.1 Detection

In Detection, you need to locate and identify sensitive data, based on the policy defined within big data, and traditional repositories such as databases and file stores.

2.2 Sampling configuration

In DgSecure, the Sampling Configuration will detect the sensitive information in the database. By Default, the DgSecure provides two options for sampling the data.

2.3 Masking

In DgSecure, you can create, edit or delete a Masking tasks to encrypt the sensitive information which were spotted in the detection task.

2.4 Encryption/Decryption

In DgSecure, you can replace sensitive data with fictitious content using one of many available data replacement options.

2.5 Privacy

Regulations such as the European Union's General Data Protection Regulation (GDPR) and California's Consumer Privacy Act (CCPA) are forcing organizations worldwide, to revisit their Data Privacy policies and practices. Other privacy regulations around the world are likely to have data subject rights as their core principles. While there are minor variations in the specific types of information involved and the conditions under which these rights are to be respected, the fundamental requirements are similar.

2.6 Monitoring

DgSecure Monitor allows users to set alerts around data handling and/or data access in HFDS, S3, MapR-FS, Hive, Oracle, or Teradata. Rules can be set around several conditions including specific users, specific data type groups, specific systems, a specific command, source path, or destination path. Rules can be set using one or all of these conditions.

The alerts are based off the targeted platform's log files. Set alert rules on the **Alert Rules** page. Review triggered alerts on the **Alerts** page. The **Monitor Overview** page provides a comprehensive overview. In order to most effectively utilize DgSecure Monitor, DgSecure detection tasks need to be run in order to identify where sensitive data resides.

The Monitor Overview page provides a centralized location to track the monitoring status of known sensitive data. Monitoring capabilities are broken down according to the percentage of sensitive types covered by alert rules, the number of people who receive alert notifications, source systems, related DgSecure policies, and alerts issued over the past 24 hours.

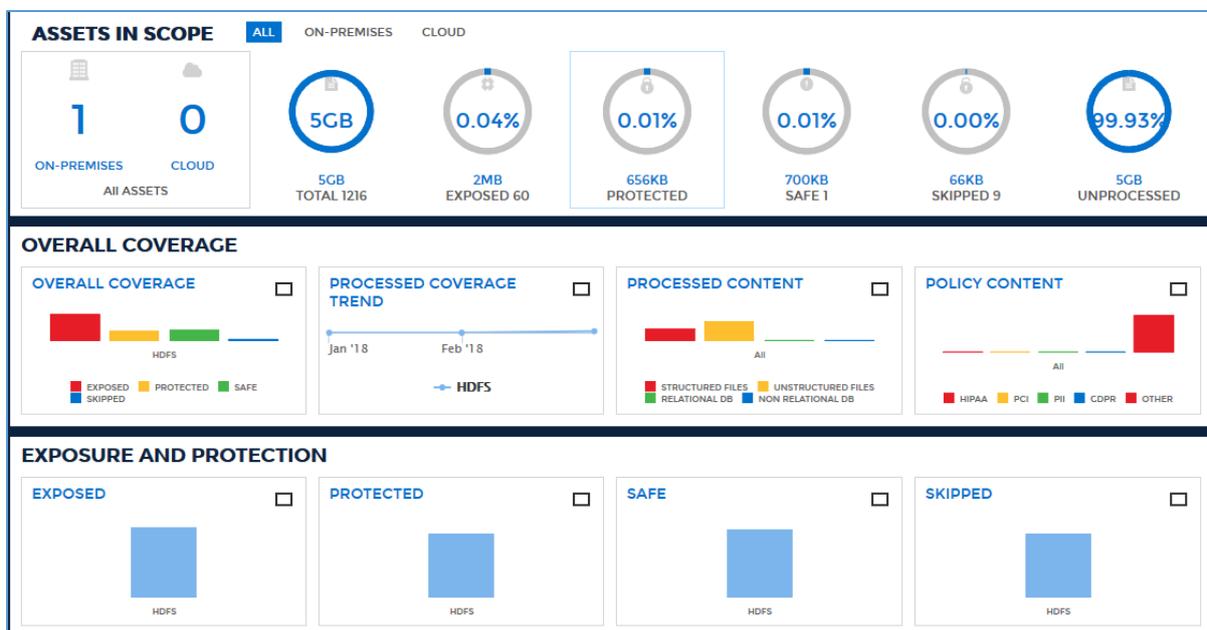
3 Dashboard

The **Overview page** provides a complete, up-to-the minute picture of the security of the company's sensitive data. It provides information in three rows of panels:

1. Assets in Scope
2. Overall Coverage
3. Exposure & Protection.

The graphics in each row provide insight of the sensitive data in the data sources DBMS, HDFS, S3, Hadoop.

Access the **Overview page** from the menu under **Reports > Overview**. It is divided into three separate panels which are described below.

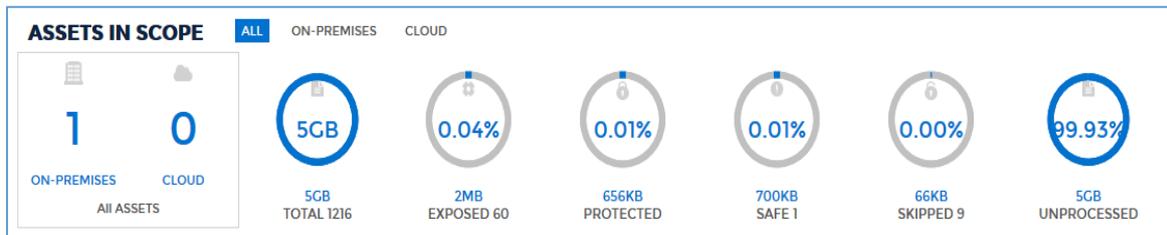


1. Assets in Scope

The Assets in Scope charts provide an overview of your data assets. The first graphic indicates how many assets are on-premises vs in the cloud. Any AWS and Azure assets are considered in the cloud, while all other Hadoop and RDBMS assets are considered to be on-premises or in the cloud according to their designation. Hadoop location is set when setting up the cluster connection. RDBMS location is set when creating a database connection. One asset is equal to one Hadoop cluster or one database. The second chart relates the total number of sensitive objects discovered across all data assets.

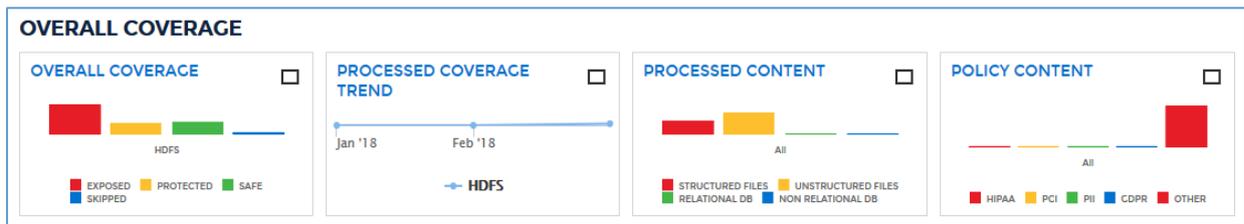
A sensitive object is any table or file that contains at least one sensitive data element. The third chart indicates the percentage of files and tables that contain sensitive data, while the fourth chart indicates the percentage of sensitive objects DgSecure has masked or encrypted. The fifth chart indicates the percentage of safe items within the assets. The sixth

chart indicates the percentage of skipped data. The seventh chart indicates the percentage of unprocessed data.



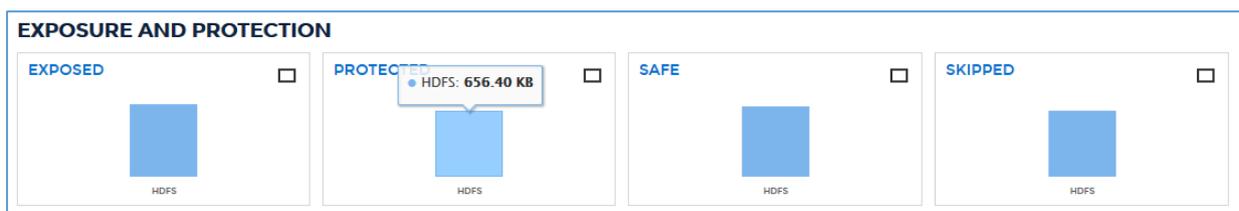
2. Overall Coverage

These Overall Coverage graphs offer insight into the current level of sensitive data coverage across your data assets. The first graph shows the number of tables and files involved in the coverage. The "Safe" bar indicates scanned files found to have no sensitive data. The second graph shows when sensitive data detection and protection occurred. The third graph shows the breakdown of structured vs unstructured data. The fourth graph shows the breakdown of structured vs unstructured data.



3. Exposure and Protection

The Exposure & Protection graphs break down the objects according to whether the sensitive data in them has been protected (masked or encrypted), exposed (detected but unprotected) or skipped. Any files or tables that have yet to be scanned show in the unscanned graph.



4 Connection Manager

4.1 Concept

In order for DgSecure to detect, protect and monitor sensitive information on a database, a connection to the database from DgSecure must first be established. The Connections screen under Connection Manager houses all the connections made to different databases and enables the user to create a new connection or edit the existing connections in DgSecure.

Based on the IDP type i.e., Detection IDP and Masking IDP, all connections can broadly be classified as connections for Detection and connections for masking. A detection connection can be used to create a task for Detection and Metadata discovery and for executing masking, encryption and decryption tasks a masking connection is used.

Further sections detail the steps for creating connections, editing connections and viewing available connections in DgSecure.

***Note:** Before creating a connection, ensure that the required IDP is up and running before creating before creating a connection. If the IDP is inactive or, wrong IP ranges have been assigned to it, a new connection could still get created, however, an error will come up at the time of task execution with this connection and the task will not be processed.

4.2 Create a Connection

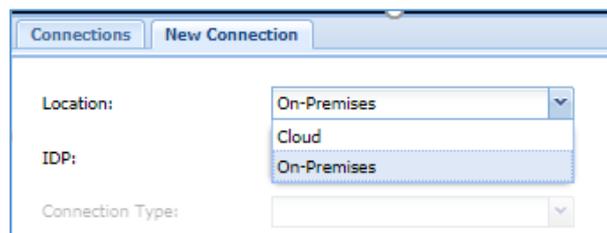
To execute detection and/or masking on databases, the first step is creating a connection to the database from the DgSecure controller. Detection and Masking IDPS are required to process the required tasks.

For more details on creating an IDP, please refer to section **6.2.1. Create IDPs** of the DgSecure Admin Guide. This section outlines the steps to create and/or edit a connection.

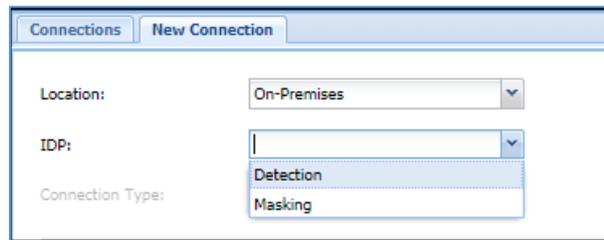
4.2.1 RDBMS

DgSecure supports masking and detection in RDBMS. This section outlines how to create a connection in RDBMS for masking and detection. Perform the following steps:

1. Click **RDBMS > Connection Manager > Connections > New Connection** tab.
2. Select the **Location** either **On premises** or **Cloud**, depending on where your database is located.



3. Select **IDP** type either **Detection** or **Masking**, depending on which of the two tasks has to be executed on the database.

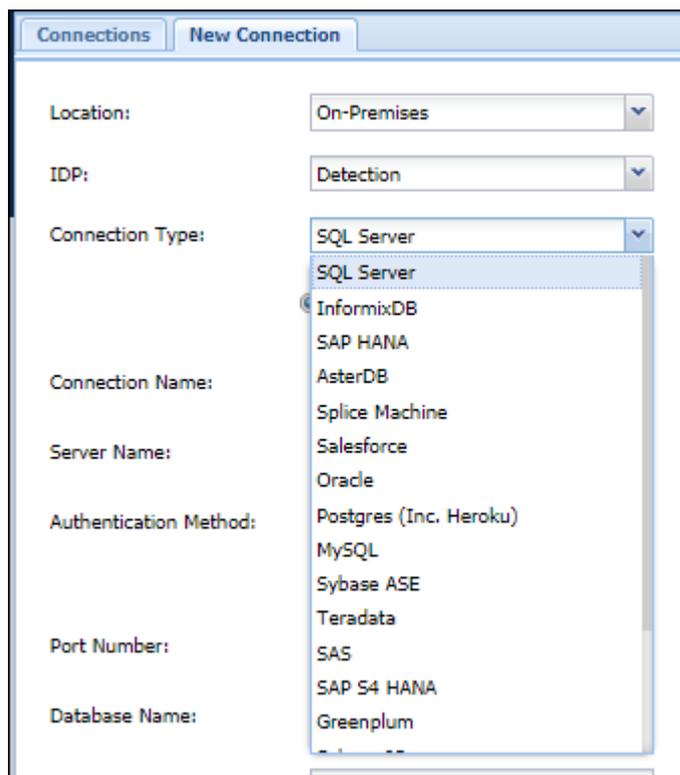


4. Following are the steps to create a connection for [Detection](#) and [Masking](#)

Detection

When creating a detection task, you can filter Databases and fetch metadata to specify databases and schemas for detection. Perform the following steps to create a detection connection for your database:

1. Select the database from the **Connection Type** drop-down.



Following are the different options for different databases:

Following are the options specific to the SQL Server:



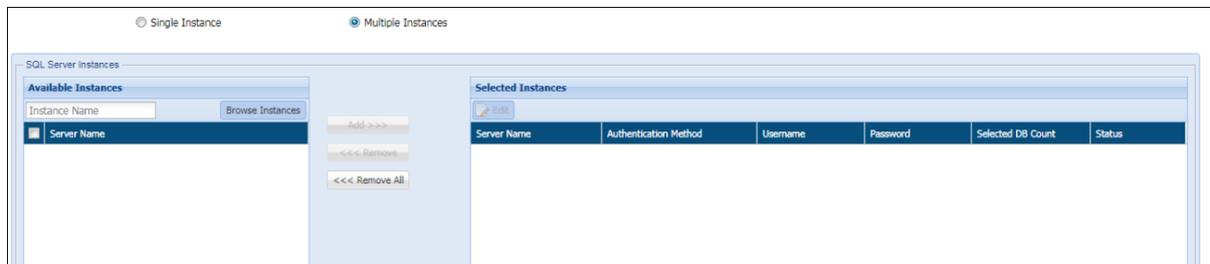
The screenshot shows a configuration form for a SQL Server connection. The 'Connection Type' is set to 'SQL Server'. Under 'Authentication Method', 'SQL Server' is selected. The 'Connection Name' is 'SQL_1', 'Server Name' is 'SQLmac', 'Port Number' is '1433', 'Database Name' is 'master', and 'User Name' is 'SA'. The 'Password' field is masked with four asterisks. The 'SSL' checkbox is checked. The 'IP Address' field contains '100.101.102.103'. The 'Use Connection Strings' checkbox is unchecked. A 'Connection String' field at the bottom contains the template: `jdbc:dsgserver://<hostname>:instance=<instance>`.

- a) **Single Instance:** Provide the following details.
 - i. **Connection Name:** Enter unique connection name. This field accepts letters, numbers, and symbols.
 - ii. **SSL:** For an additional layer of security, check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
 - iii. **Server Name:** Enter the name of the SQL Server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by **Find DBMS**
 - iv. **IP Address:** Enter the IP Address for the connection.
 - v. **Authentication Method:** The following authentication methods are available for SQL Server:
 - **SQL Server:** Authenticates the connection using SQL Server Username and Password.
 - **Windows Impersonation:** Authenticates the connection using the Windows Impersonation Username and Password.
 - **Windows:** Authenticates the connection based on the connection to the local IP address.
 - vi. **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
 - vii. **Database Name:** Enter the name of the database.

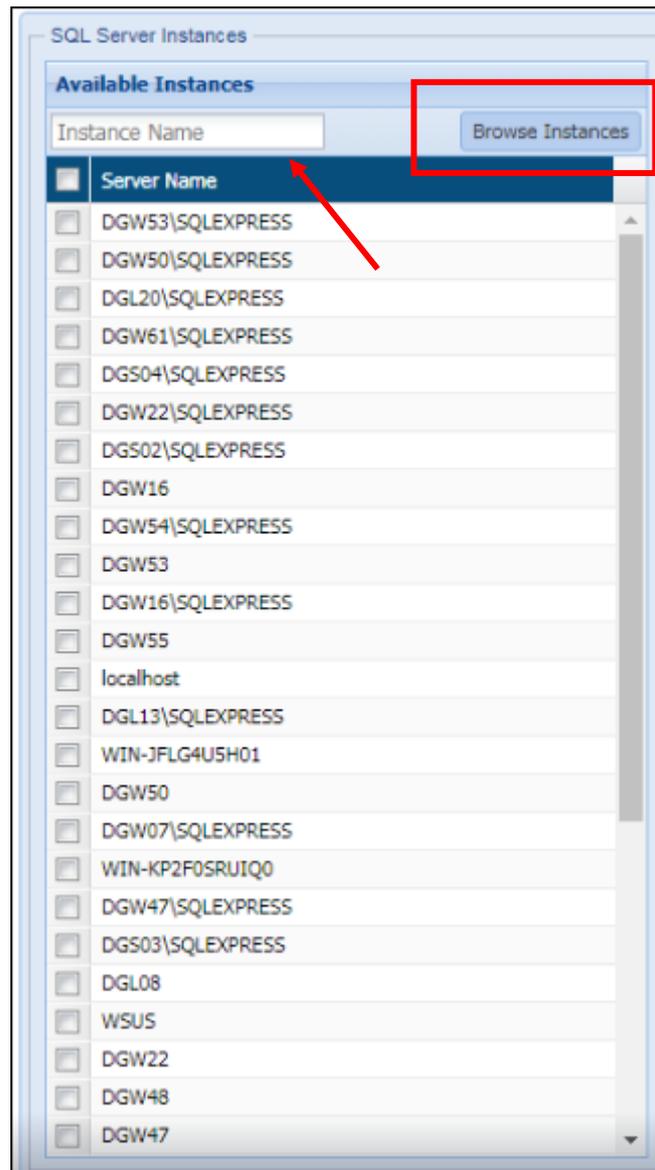
- viii. **User Name:** Enter the database user name.
- ix. **Password:** Enter the database password.
- x. **Use Connection String:** It specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to authenticate the connection using a connection string instead of through username and password.

b) **Multiple Instances:** To connect to multiple instances on an SQL Server, perform the following steps:

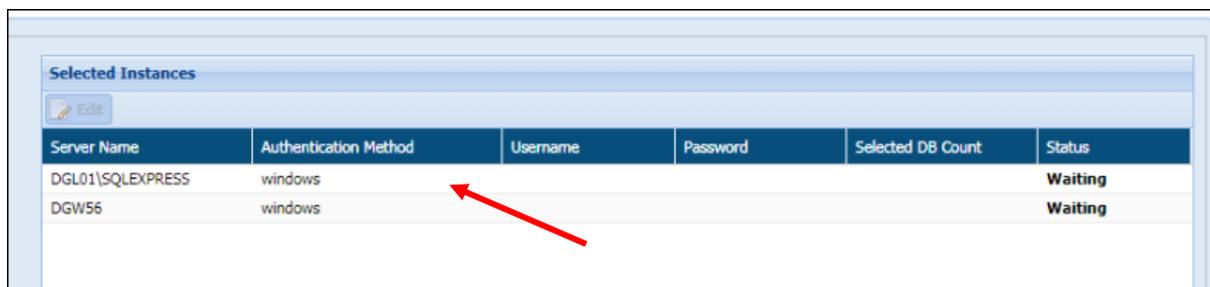
- xi. Select the option Multiple Instances, below the Connection Type. The SQL Server Instances panel will appear:



- xii. Enter the name of the SQL Server instance or click **Browse Instances** to view all the instances available on the SQL Server.



- xiii. Select the required instances and click **Add**. The default authentication method is Windows.



- xiv. To change the authentication method for an instance, select the instance and click **Edit**. The following pop-up will appear:

- xv. Provide the following details in the **Edit Connection Details** popup:
- **Location:** On-Premises or Cloud.
 - **Connection Name:** Enter unique connection name. This field accepts letters, numbers, and symbols.
 - **Server Name:** This field is auto-populated.*
 - **IP Address:** Enter the IP Address for the connection.
 - **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
 - **Authentication Method:** The following authentication methods are available for SQL Server:
 - i. **SQL Server:** Authenticates the connection using SQL Server **Username** and **Password**.
 - ii. **Windows Impersonation:** Authenticates the connection using the Windows Impersonation **Username** and **Password**.
 - iii. **Windows:** Authenticates the connection based on the connection to the local IP address.

***NOTE:** The prerequisites for creating a connection to multiple instances on an SQL Server, are:

- The sqlcmd utility must be installed.
- Sqlcmd-L command must be executed to get the list of the available instances.
- The Server name is auto-populated when the detection agent fetches the name of the server where the command sqlcmd-L is executed.

Refer to [Appendix J: Configure Multiple SQL Server Instances in DgSecure](#) for more details.

InformixDB

Following are the options specific to the InformixDB :

Connection Type:	InformixDB	Instance Name:	Dep
Connection Name:	Info_1	IP Address:	101.102.103.104
Hostname:	Info_db_1	Port Number:	9090
User Name:	Admin	Password:	*****

- Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- Hostname:** Enter the Hostname of the InformixDB or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- IP Address:** Enter the IP Address for the connection.
- Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- User Name:** Enter the database user name.
- Password:** Enter the database password.

SAP HANA

Following are the options specific to the SAP HANA:

Connection Type:	SAP HANA	Hostname:	hana_1
Connection Name:	Sap_hana	Port Number:	39013
IP Address:	100.101.102.102	Database Name:	Dep_db
Username:	Admin	Password:	*****

- Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- Hostname:** Enter the Hostname of the SAP HANA database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- IP Address:** Enter the IP Address for the connection.
- Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.

- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

AsterDB

Following are the options specific to the AsterDB:

Connection Type:	AsterDB	IP Address:	100.101.102.103
Connection Name:	AsterDB1	Database Name:	dep_db
Hostname:	as_db	User Name:	admin
Port Number:	2406	Password:	*****

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the AsterDB or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

Splice Machine

Following are the options specific to the Splice Machine:

Connection Type:	Splice Machine	Group/IP Range:	Default_IPRange_DMA_1
Connection Name:	SM	Database Name:	dep
Hostname:	SP_1	User Name:	admin
Port Number:	1527	Password:	*****
Authentication Method:	<input checked="" type="radio"/> Native <input type="radio"/> Kerberos with Principal		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.

- b) **Hostname:** Enter the Hostname of the Splice Machine database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks.
- c) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **Authentication Method:** There are two authentication methods for Splice Machine databases, these are as follows:
 1. **Native:** Provide the following details to authenticate:
 - User Name: Enter the database user name.
 - Password: Enter the database password.
 2. **Kerberos with Principal:** If Splice machine has been installed with Kerberos, provide the following details of the Kerberos setup to authenticate:
 - Principal Name: Enter the principal name.
 - Keytab: Enter the keytab.

Salesforce

Following are the options specific to the Salesforce

Connection Type:	<input type="text" value="Salesforce"/>	IP Address:	<input type="text" value="100.102.103.104"/>
Connection Name:	<input type="text" value="Salesforce_conn"/>	Security Token:	<input type="text" value="XXXXXXXX"/>
Username:	<input type="text" value="admin"/>	Password:	<input type="password" value="*****"/>

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **IP Address:** Enter the IP Address for the connection.
- c) **Security Token:** Enter the Salesforce security token.
- d) **User Name:** Enter the database user name.
- e) **Password:** Enter the database password.

Oracle

Following are the options specific to the Oracle:

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Connection Type:** There are two connection types in oracle:
 1. **Basic:** Provide the following details for a basic connection:
 - **Hostname:** Enter the Hostname of the Oracle server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
 - **IP Address:** Enter the IP Address for the connection.
 - **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
 - **SID or Service Name:** Enter the Service ID (SID) or Service Name.
 2. **TNS:** Provide the following details for a basic connection:
 - **TNS Name:** Enter the TNS Name of the Oracle server.
 - **IP Address:** Enter the IP Address for the connection.
- d) **Authentication Method:** The following authentication methods are available for Oracle Server:
 1. **Oracle:** Authenticates the connection using Oracle Server Username and Password.
 2. **Kerberos:** Authenticates the connection using the Kerberos.
 3. **Kerberos with Principal:** Authenticates using Kerberos system’s Principal name and password.

***Note:** For using the authentication method, Kerberos, and Kerberos with Principal, the environment should be set up on the IDP. Following are the steps to perform the same:

- i. Check that correct krb5.conf file is available on the IDP machine. If not, copy the krb5.conf file on Windows directory or /etc/ directory.
- ii. Ping KDC server and database server using hostname from IDP to verify network connectivity. Add entries in /etc/hosts file, if required.
- iii. Verify that the clock time should be same on IDP, KDC and DB2, Oracle server machine.
- iv. (For Kerberos) Obtain TGT manually at agent machine.
 - a. Login to OS with the Kerberos user.
 - b. Run “kinit user” command to obtain TGT. Verify using “klist” command.
 - i. kinit is available in Java/jre/bin folder.
 - ii. The default paths are set when client is configured, but if faced with any issues add the following property in “javaOptions” in jetty-embedded.properties present in agent installation. Once this step has been done, restart the agent.

-Doracle.net.kerberos5_cc_name=<path to krb5 cache file> -
Djava.security.krb5.conf=<path to krb5.conf>

- v. Connect through DgSecure by entering the required information.

- e) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string instead of SSL.

Postgres (Inc. Heroku)

Following are the options specific to the Postgres (Inc. Heroku):

Connection Type:	<input type="text" value="Postgres (Inc. Heroku)"/>		
Connection Name:	<input type="text" value="Postgres"/>	SSL:	<input checked="" type="checkbox"/>
Hostname:	<input type="text" value="pos_1"/>	IP Address:	<input type="text" value="100.101.102.103"/>
Port Number:	<input type="text" value="5432"/>	Database Name:	<input type="text" value="dep_db"/>
User Name:	<input type="text" value="admin"/>	Password:	<input type="password" value="****"/>

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.

- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.

***Note:** It is mandatory to check the SSL checkbox if connecting with Heroku database.

- c) **Hostname:** Enter the Hostname of the Postgres server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **User Name:** Enter the database user name.
- h) **Password:** Enter the database password.

MySQL

Following are the options specific to the MySQL Server:

Connection Type:	MySQL	IP Address:	100.101.102.103
Connection Name:	mysql_conn	Password:	*****
Hostname:	mysql_1		
Port Number:	3306		
User Name:	admin		
Use Connection Strings:	<input checked="" type="checkbox"/>		
Connection String:	jdbc:mysql://<ip address>:<port number>		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.

- c) **Hostname:** Enter the Hostname of the MySQL server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.
- h) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to connect using the connection string instead of the hostname and port number.

Sybase ASE

Following are the options specific to the Sybase ASE:

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide
- c) **Hostname:** Enter the Hostname of the Sybase ASE server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

- h) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to connect using the connection string instead of the hostname and port number.

Teradata

Following are the options specific to the Teradata:

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the Teradata server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Authentication Method:** The following authentication methods are available for Teradata Server:
 1. **Teradata:** Authenticates the connection using Teradata Server Username and Password.
 2. **Kerberos:** Authenticates the connection using the Kerberos.
 3. **Kerberos with Principal:** Authenticates using Kerberos system’s Principal name and password.
- e) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string.

SAS

Following are the options specific to the SAS:

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **IP Address:** Enter the IP Address for the connection.

SAP S4 Hana

Following are the options specific to the SAP S4 Hana:

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the SAP S4 Hana Server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

Greenplum

Following are the options specific to the Greenplum:

Connection Type:	<input type="text" value="Greenplum"/>	IP Address:	<input type="text" value="100.101.102.103"/>
Connection Name:	<input type="text" value="green_conn"/>	Database Name:	<input type="text" value="dep"/>
Hostname:	<input type="text" value="greenplum_1"/>	Password:	<input type="password" value="*****"/>
Port Number:	<input type="text" value="5432"/>		
User Name:	<input type="text" value="admin"/>		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the Greenplum database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

Sybase IQ

Following are the options specific to the Sybase IQ:

Connection Type:	<input type="text" value="Sybase IQ"/>	IP Address:	<input type="text" value="100.101.102.103"/>
Connection Name:	<input type="text" value="sybase_conn"/>	Database Name:	<input type="text" value="dep"/>
Hostname:	<input type="text" value="sybase_1"/>	Password:	<input type="password" value="*****"/>
Port Number:	<input type="text" value="2638"/>		
User Name:	<input type="text" value="admin"/>		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the Sybase IQ database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks.
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.

- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

DB2

Following are the options specific to the DB2:

- a) **Connection Name:** Enter unique Connection Name. This field accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the DB2 database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **Authentication Method:** There are two authentication methods for DB2 databases, these are as follows:
 1. **Native:** Provide the following details to authenticate:
 - User Name: Enter the database user name.
 - Password: Enter the database password.
 2. **Kerberos with Principal:** If Splice machine has been installed with Kerberos, provide the following details of the Kerberos setup to authenticate:
 - Principal Name: Enter the principal name.
 - Keytab: Enter the keytab.

- h) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string instead of SSL.

Netezza

Following are the options specific to the Netezza:

Connection Type:	Netezza	IP Address:	100.101.102.103
Connection Name:	netezza_conn	Database Name:	dep
Hostname:	net_1	Password:	*****
Port Number:	5432		
User Name:	admin		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the Netezza database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks.
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

Snowflake

Following are the options specific to the Snowflake:

Connection Type:	Snowflake	URL:	
Authentication Type:	<input checked="" type="radio"/> Username-Password <input type="radio"/> Key-Pair	Database Name:	dep
Connection Name:	snowflae_conn	Password:	*****
Warehouse name:	Snowflake_1		
Group/IP Range:	Default_IPRange_DMA_1		
User Name:	admin		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Warehouse name:** Enter the name of the Snowflake warehouse.
- c) **URL:** Enter the URL to the snowflake setup
- d) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- e) **Database Name:** Enter the name of the database.
- f) **Authentication Method:** There are to authentication methods for DB2 databases, these are as follows:
 - 1. **Username password:** Provide the username and password to authenticate.
 - 2. **Key pair:** Provide the username, keypath and passphrase to authenticate.
- g) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string instead of SSL.

SAS SPD

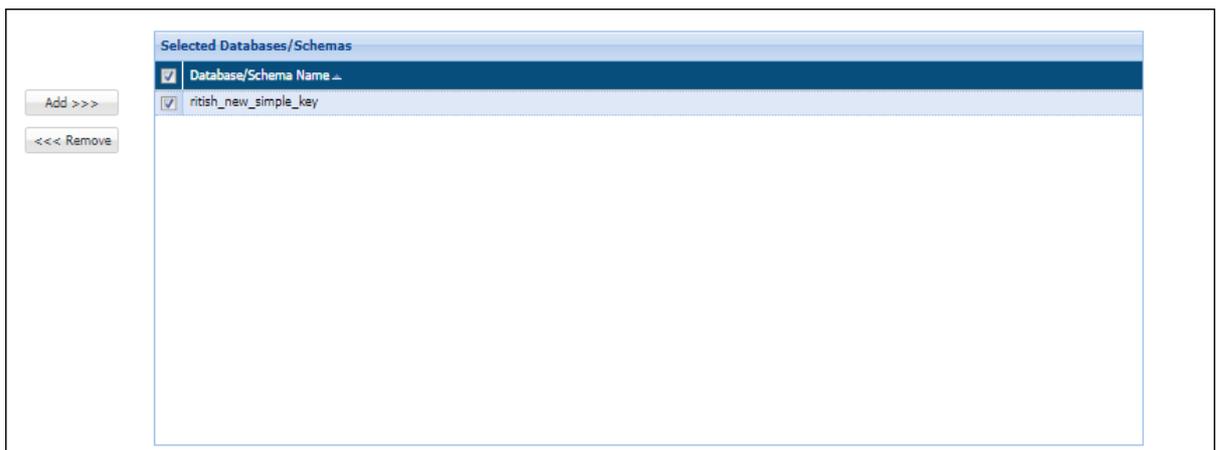
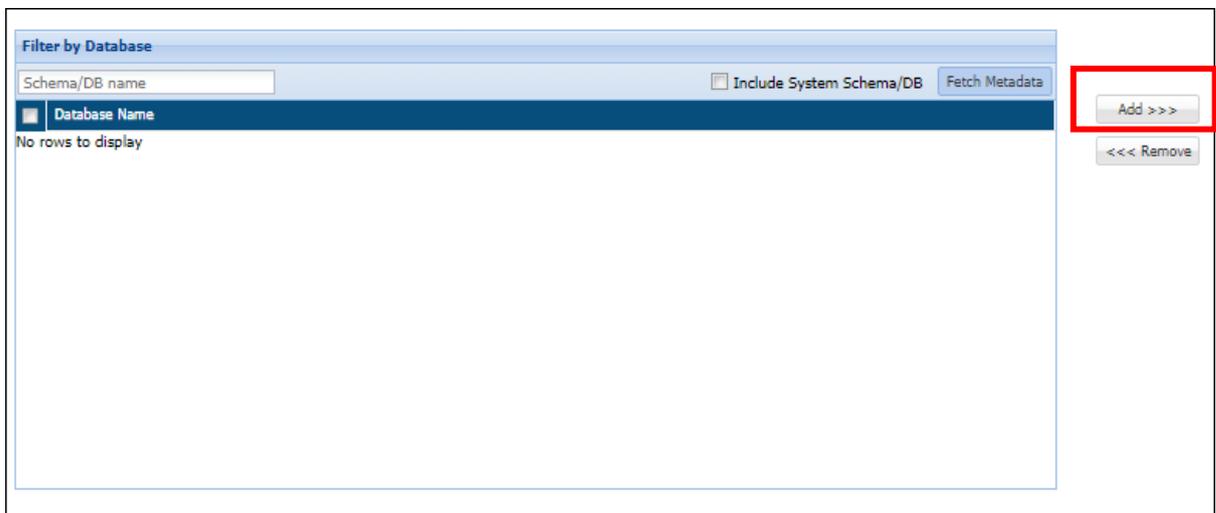
Following are the options specific to the SAS SPD :

Connection Type:	<input type="text" value="SASSPD"/>	Port Number:	<input type="text" value="5432"/>
Connection Name:	<input type="text" value="sasspd_conn"/>	SNET Port Numbers:	<input type="text" value="1000"/>
Hostname:	<input type="text" value="sasspd1"/>	Group/IP Range:	<input type="text" value="Default_IPRange_DMA_1"/>
Libname domain:	<input type="text" value="abd"/>	Password:	<input type="text" value="*****"/>
SNET Hostname:	<input type="text" value="snet1"/>		
User Name:	<input type="text" value="admin"/>		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the SAS SPD database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks.
- c) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- d) **Libname domain:** Enter the Libname domain for the connection.
- e) **SNET Port Number:** Enter the SNET Port Number.

- f) **SNET Hostname:** Enter the SNET Hostname.
- g) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- h) **User Name:** Enter the database user name.
- i) **Password:** Enter the database password.

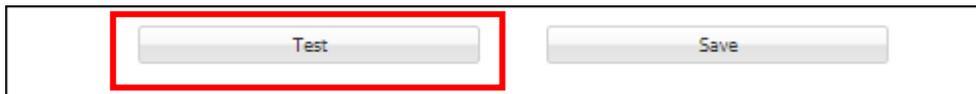
2. After providing the connection details enter the name of the database in the bottom left filter by database panel, select the required schemas and databases and click **Add**. The selected databases will appear in the panel on the bottom right.



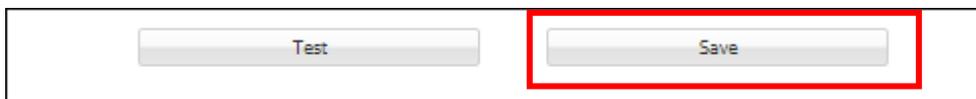
***NOTE:**

- For multiple Instances in SQL Server, select the required instance and click **Edit** to view and select the databases and schemas.
- For SAS and SAS SPD servers there is no **Fetch Metadata** functionality, user has to select the directories for SAS and for SAS SPD, select the directories available on the SNET server at the time of task execution.

3. Click the **Test** button, to test the connection.

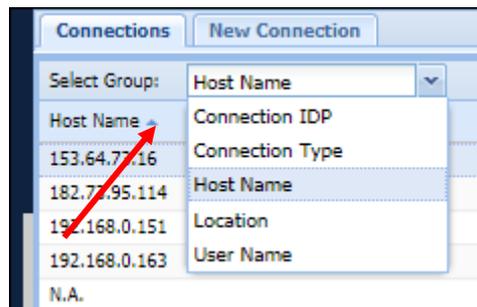


4. Click **Save** button, to save the changes.

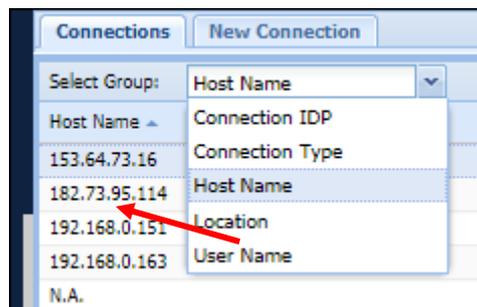


5. To edit a connection, go RDBMS > Connection Manager > Connections.

6. Select the connection classification from the Select Group dropdown.



7. Select the classification type.



8. Select the connection.

ID	Connection Name
7	t1_det
6	t_det

9. Click **Edit**.

ID	Connection Name
7	t1_det
6	t_det

10. Location, IDP and connection type cannot be edited. A connection can be edited using the same steps as task creation.

4.2.1.1 Masking

Perform the following steps to create a masking connection for different databases:

1. After selecting the Location (On-premises or Cloud) and IDP (Masking), select the database from the **Connection Type** drop-down.

Location: Cloud

IDP: Masking

Connection Type: [Open dropdown menu]

- SQL Server
- Oracle
- Postgres (Inc. Heroku)
- MySQL
- Teradata
- DB2
- Snowflake

Connection Name: [Greyed out]

Connection Type: [Greyed out]

IP Address: [Greyed out]

2. Masking is supported on the following RDBMS databases:

SQL Server

Following are the options specific to the SQL Server:



- a) **Connection Name:** Enter unique Connection Name. This field accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Server Name:** Enter the name of the SQL Server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Authentication Method:** The following authentication methods are available for SQL Server:
 - **SQL Server:** Authenticates the connection using SQL Server Username and Password.
 - **Windows Impersonation:** Authenticates the connection using the Windows Impersonation Username and Password.
 - **Windows:** Authenticates the connection based on the connection to the local IP address.
- f) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- g) **Database Name:** Enter the name of the database.
- h) **User Name:** Enter the database user name.
- i) **Password:** Enter the database password.
- j) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox,

and provide the connection string. Use this option if you want to authenticate the connection using a connection string instead of through username and password.

Oracle

Following are the options specific to the Oracle:

The screenshot shows a configuration form for an Oracle connection. Key fields include:

- Connection Type:** Oracle
- Connection Name:** mk_orcl_det_Con
- Connection Type:** Basic (selected), TNS
- IP Address:** 192.168.0.163
- Host Name:** 192.168.0.163
- Port Number:** 1521
- Authentication Method:** Oracle (selected), Kerberos, Kerberos with Principal
- User Name:** fivethousand_qa_performance
- Connection String:** jdbc:oracle:thin:@//<hostname>:<portnumber>/<servicename>

- a) **Connection Name:** Enter unique Connection Name. This field accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Connection Type:** There are two connection types in Oracle:
 1. **Basic:** Provide the following details for a basic connection:
 - **Hostname:** Enter the Hostname of the Oracle server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
 - **IP Address:** Enter the IP Address for the connection.
 - **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
 - **SID or Service Name:** Enter the Service ID (SID) or Service Name.
 2. **TNS:** Provide the following details for a basic connection:
 - **TNS Name:** Enter the TNS Name of the Oracle server.
 - **IP Address:** Enter the IP Address for the connection.
- d) **Authentication Method:** The following authentication methods are available for Oracle Server:
 - **Oracle:** Authenticates the connection using Oracle Server Username and Password.
 - **Kerberos:** Authenticates the connection using the Kerberos.
 - **Kerberos with Principal:** Authenticates using Kerberos system's Principal

name and password.

- e) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the **Use Connection String** checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string instead of SSL.

Postgres (Inc. Heroku)

Following are the options specific to the Postgres (Inc. Heroku):

Connection Type:	Postgres (Inc. Heroku)	SSL:	<input checked="" type="checkbox"/>
Connection Name:	Postgres	IP Address:	100.101.102.103
Hostname:	pos_1	Database Name:	dep_db
Port Number:	5432	Password:	*****
User Name:	admin		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the Postgres server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **User Name:** Enter the database user name.
- h) **Password:** Enter the database password.

MySQL

Following are the options specific to the MySQL Server:

Connection Type: MySQL

Connection Name: mysql_conn

Hostname: mysql1 IP Address: 100.101.102.103

Port Number: 3306

Database Name: dep

User Name: admin Password: *****

Use Connection String:

Connection String: jdbc:mysql://<ip address>:<port number>/<database>

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the MySQL server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **User Name:** Enter the database user name.
- h) **Password:** Enter the database password.
- i) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to connect using the connection string instead of the hostname and port number.

Teradata

Following are the options specific to the Teradata:

Connection Type: Teradata

Connection Name: teradata_conn

Hostname: teradata1 IP Address: 100.101.102.103

Port Number: 1025

Database Name: dep

User Name: admin Password: *****

Use Connection String:

Connection String: jdbc:mysql://<ip address>:<port number>/<database>

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Hostname:** Enter the Hostname of the Teradata server or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- c) **IP Address:** Enter the IP Address for the connection.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.
- h) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string.

DB2

Following are the options specific to the DB2:

Connection Type:	DB2	SSL:	<input checked="" type="checkbox"/>
Connection Name:	db2	IP Address:	100.101.102.103
Hostname:	db2	Database Name:	dep
Port Number:	50000	Keytab:	12345
User Name:	admin		
Use Connection String:	<input type="checkbox"/>		
Connection String:	jdbc:db2://<hostname>:<portnumber>:<database>		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the DB2 database or select from the list of available IPs. You can also search for the hostnames and IP addresses of databases by Find DBMS tasks
- d) **IP Address:** Enter the IP Address for the connection.

- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **User Name:** Enter the database user name.
- h) **Keytab:** Enter the keytab value.
- i) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to secure the connection using connection string instead of SSL.

Snowflake

Following are the options specific to the Snowflake:

Connection Type:	<input type="text" value="Snowflake"/>		
Authentication Type:	<input checked="" type="radio"/> Username-Password	<input type="radio"/> Key-Pair	
Connection Name:	<input type="text" value="snowflae_conn"/>		
Warehouse name:	<input type="text" value="Snowflake_1"/>	URL:	<input type="text"/>
Group/IP Range:	<input type="text" value="Default_IPRange_DMA_1"/>	Database Name:	<input type="text" value="dep"/>
User Name:	<input type="text" value="admin"/>	Password:	<input type="text" value="*****"/>

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **Warehouse name:** Enter the name of the Snowflake warehouse.
- c) **URL:** Enter the URL to the snowflake setup
- d) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- e) **Database Name:** Enter the name of the database.
- f) **Authentication Method:** There are to authentication methods for DB2 databases, these are as follows:
 - **Username password:** Provide the username and password to authenticate.
 - **Key pair:** Provide the username, keypath and passphrase to authenticate.
- g) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox,

and provide the connection string. Use this option if you want to secure the connection using connection string instead of SSL.

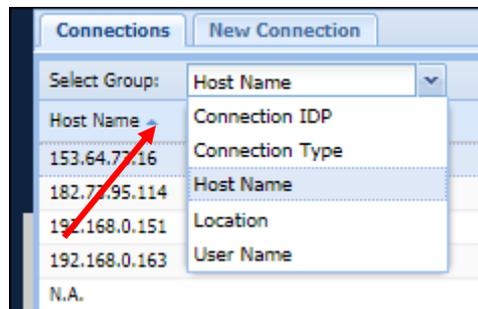
3. Click the **Test** button, to test the connection.



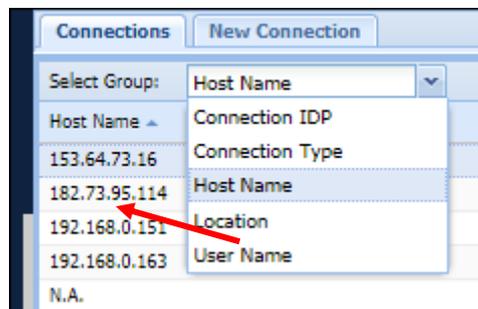
4. Click **Save** button, to save the changes.



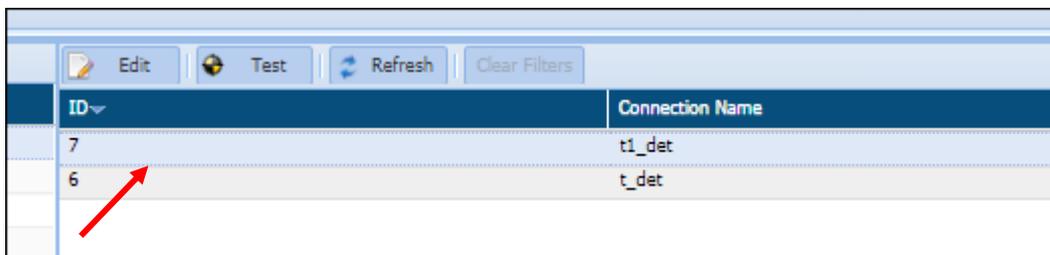
5. To edit a connection, go to RDBMS > Connection Manager > Connections.
6. Select the connection classification from the Select Group dropdown.



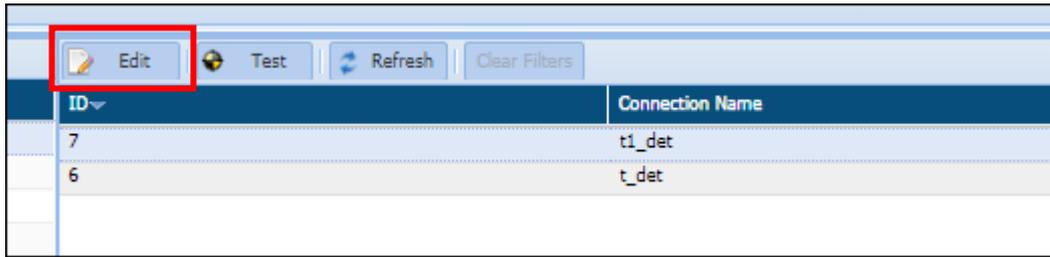
7. Select the classification type.



8. Select the connection.



9. Click **Edit**.

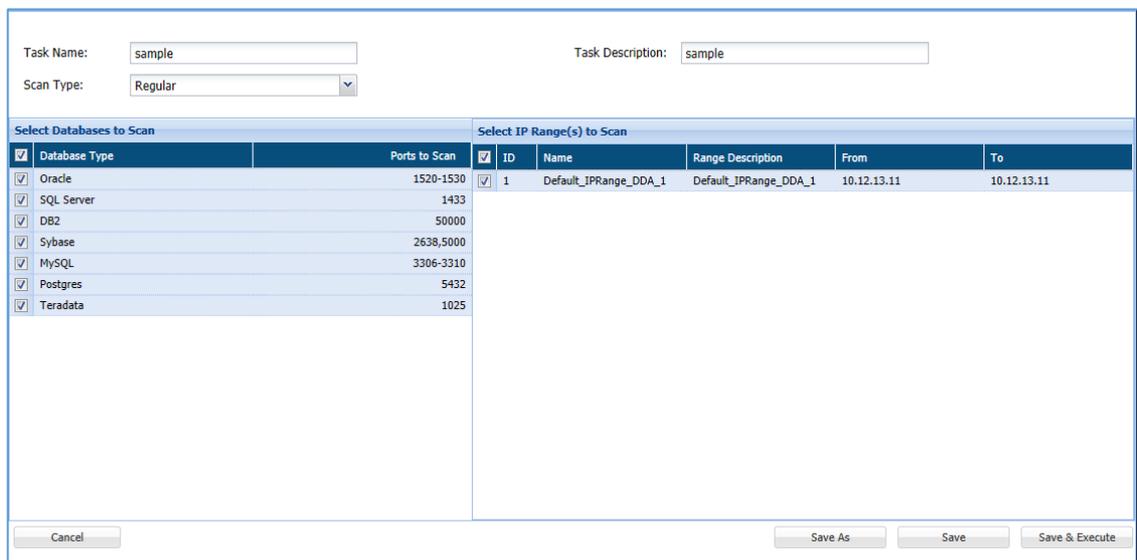


4.2.1.2 Create a Task in Find DBMS

In RDBMS, you can also search for an IP address using the FIND DBMS in DgSecure application.

To create a Connection. Click **RDBMS > Connection Manager > FIND DBMS > Tasks > New Task** tab.

The below image shows the user interface for creating a Task.



- Task Name:** Enter the Task Name. The Task Name text box accepts letters, numbers and symbols. The name should be unique to each individual task name.
- Task Description:** Enter the description for the task name.
- Scan Type:** Select either 'Regular' or 'Intense' from the Scan Type drop-down.
 - Regular:** This option allows you to search for the default port in the network. For Example: Oracle default port is 1521.
 - Intense:** This option search for the custom as well as default port in the network.
- Select Database to Scan:** This pane displays the list of all available database along with their assigned ports. The ports are editable. You can select the available databases from the given list.

Select Databases to Scan	
<input checked="" type="checkbox"/> Database Type	Ports to Scan
<input checked="" type="checkbox"/> Oracle	1520-1530
<input checked="" type="checkbox"/> SQL Server	1433
<input checked="" type="checkbox"/> DB2	50000
<input checked="" type="checkbox"/> Sybase	2638,5000
<input checked="" type="checkbox"/> MySQL	3306-3310
<input checked="" type="checkbox"/> Postgres	5432
<input checked="" type="checkbox"/> Teradata	1025

c) **Select IP Ranges(s) to Scan:** This pane shows the available scan ranges. You can select as many connections for scanning.

***Note:** All the IP ranges which are being displayed in the window are setup in the DgSecure Admin Application.

Select IP Range(s) to Scan					
<input checked="" type="checkbox"/>	ID	Name	Range Description	From	To
<input checked="" type="checkbox"/>	1	Default_IPRange_DDA_1	Default_IPRange_DDA_1	10.12.13.11	10.12.13.11

- d) **Cancel:** Click the **Cancel** button, if you do not want to save the changes.
- e) **Save:** Click the **Save** Button, if you want to save the changes.
- f) **Save & Execute:** Click the **Save & Execute** button, if you wish to save and execute the task at the same time.
- g) **Save As:** Click the **Save As** button, if you have edited the name of the task.

4.2.2 NoSQL

DgSecure Supports Detection for the NoSQL databases: MongoDB, Couchbase and Cassandra. Each of these databases is available under the NoSQL in the menu.

To create a Connection. Click **NoSQL > CONNECTION MANAGER > All CONNECTIONS > ADD NEW CONNECTION** tab.

The below image shows the user interface for creating a connection.

1. **Location:** Select either 'On-Premises' or 'Cloud' option from the Location drop-down.
2. **Connection Type:** Select the Connection Type from the given option. NoSQL support Cassandra, MongoDB and Couchbase connections.

3. **Connection Name:** Enter unique Connection Name. This field accepts letters, numbers, symbols.
4. **Extra Option:** For security, enable Mongo DB provide **AuthSource** and **AuthMechanism** in the **Extra Option** panel. To set the AuthSource and AuthMechanism, perform the steps:

***Note:** The **Extra Option** field is visible only when 'MongoDB' is selected in **Connection Type**.

- i. Click the **View/Modify** button.

NOSQL / CONNECTION MANAGER

ALL CONNECTIONS ADD NEW CONNECTION

Location
On-Premises

Connection Type
MongoDB

Connection Name

Extra option VIEW/MODIFY

User Name

- ii. Enter the values for **AuthSource** and **AuthMechanism** parameters.

EXTRA OPTIONS + ADD OPTIONS

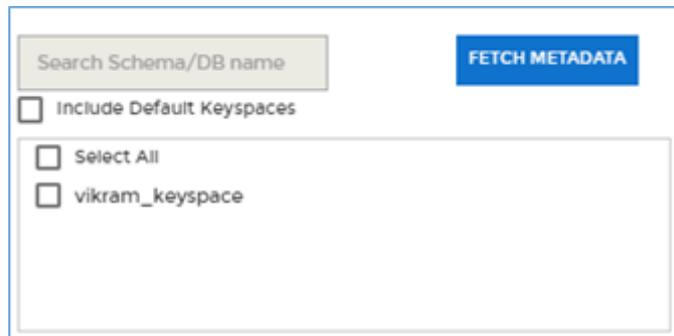
PARAM	VALUE	DELETE
authSource	Sexternal	
authMechanism	PLAIN	

CANCEL SAVE

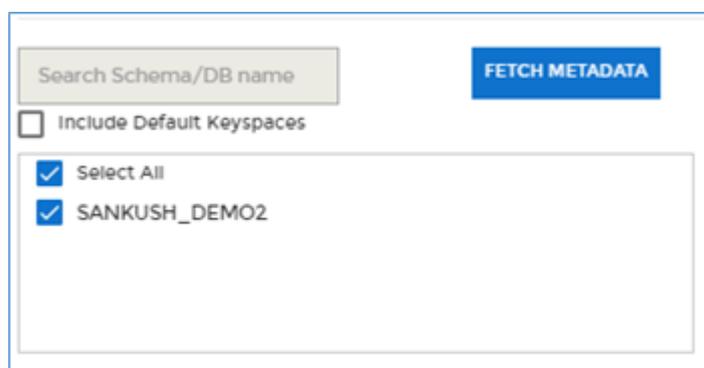
- Click on **+Add Options** to add a new parameter and value textbox.
- Click **Save** to make the changes effective.

5. **Hostname/IP Address:** Enter the Host Name or IP address in the field.
6. **Port Number:** Enter the Port Number for establishing the connection. You can also edit the port number later, if required.
7. **Group/IP Range:** Select the Group/IP Range from the Group/IP range drop-down.
8. **User Name:** Enter the database user name.
9. **Password:** Enter the database password.

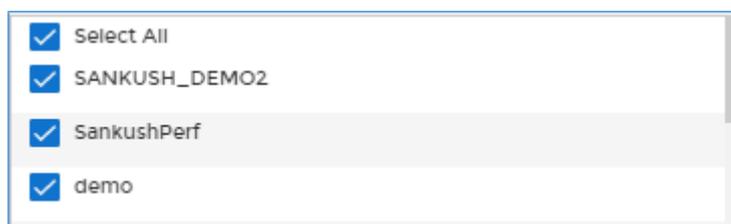
- Filter By Database:** This pane will list down all the database/schema for the entered connection. To populate the bottom window, click Fetch Metadata button and it will list down the available database/schema for the connection.



To select the database/schema name, check the checkbox next to the database or you can search the database by entering the schema name in the textbox.



- Add:** Click the **Add** button, if you want to add the selected database name in **Selected Databases/Schemas** panel.
- Remove:** Click the **Remove** button, if you want to remove the selected database from the **Selected Databases/Schemas** panel.
- Selected Databases/Schemas:** This pane will display the list of all selected schemas or databases name.



- Cancel:** Click the **Cancel** button, if you do not want to save the changes.

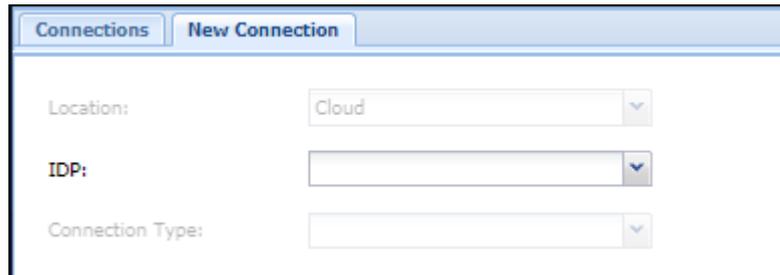
15. **Test:** Click the **Test** button, if you wish to test the connection before using it.

16. **Save:** Click the **Save** button to make the changes effective.

4.2.3 AWS

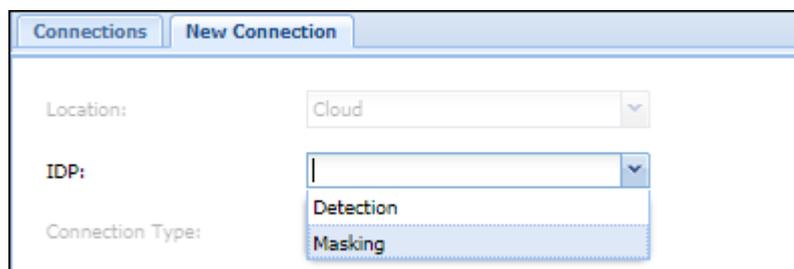
DgSecure supports masking and detection in Redshift on AWS. Perform the following steps to create a connection:

1. Click AWS>RDS/Redshift > Connection Manager > Connections > New Connection tab.



2. The location is **Cloud** by default.

3. Select the IDP type as **Detection** or **Masking**, depending on task that has to be executed on the database.

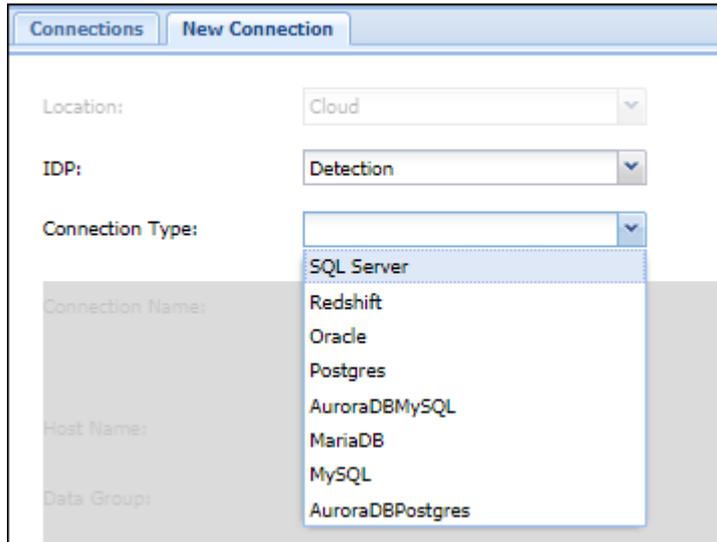


4. Steps to create a connection for Detection and Masking are as follows:

Detection

Perform the following steps to create a detection connection for your database:

a) Select the database from the **Connection Type** drop-down.



5. Detection is supported on the following RDS/Redshift databases:

SQL Server

Refer to section [SQL Server](#) in RDBMS.

Redshift

Following are the options specific to Redshift:



- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the Redshift server or select from the list of available IPs.
- d) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.

- g) **User Name:** Enter the database user name.
- h) **Password:** Enter the database password.
- i) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to connect using the connection string instead of the hostname and port number.

Oracle

Refer to section [Oracle](#) in RDBMS.

Postgres

Refer to section [Postgres \(Inc. Heroku\)](#) in RDBMS.

AuroraDBMySQL

Following are the options specific to AuroraDBMySQL:

Connection Type:	AuroraDBMySQL	SSL:	<input checked="" type="checkbox"/>
Connection Name:	aurora_conn	Group/IP Range:	Default_IPRange_DMA_1
Hostname:	aurora1	Password:	*****
Port Number:	3306		
User Name:	admin		

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the AuroraDBMySQL server or select from the list of available IPs.
- d) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

MariaDB

Following are the options specific to MariaDB:

Connection Type:	MariaDB	SSL:	<input type="checkbox"/>
Connection Name:	maria_conn	Group/IP Range:	Default_IPRange_DMA_1
Hostname:	maria1	Password:	*****
Port Number:	3306		
User Name:	admin		

- Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- Hostname:** Enter the Hostname of the MariaDB server or select from the list of available IPs.
- Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- User Name:** Enter the database user name.
- Password:** Enter the database password.

MySQL

Refer to section [MySQL](#) in RDBMS.

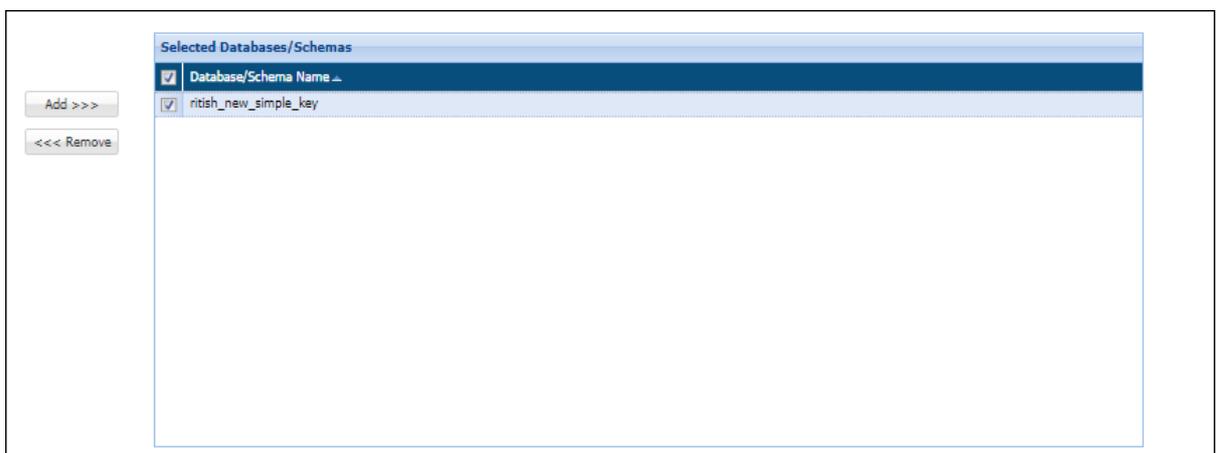
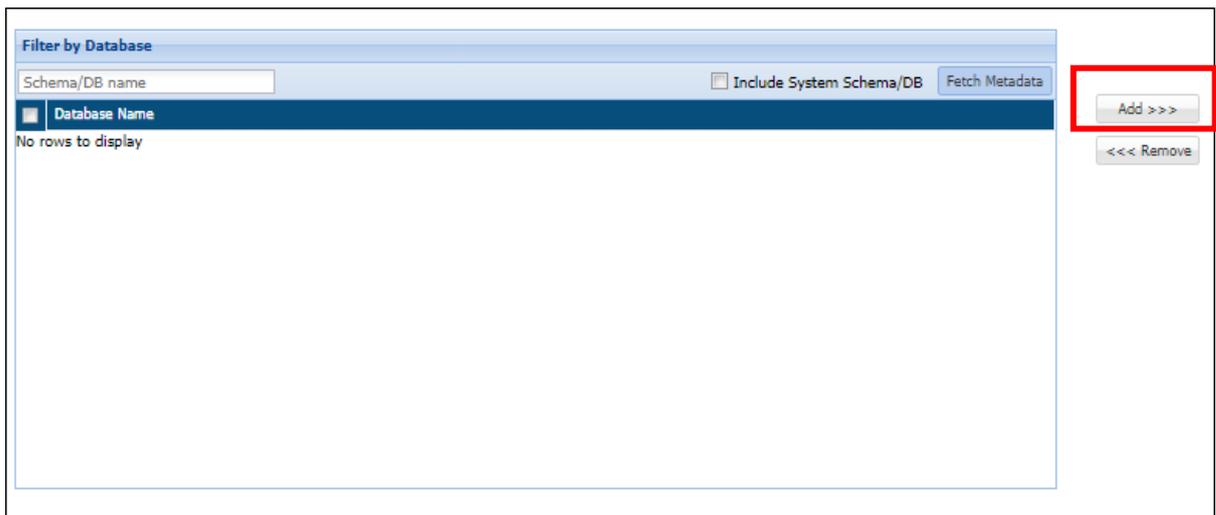
AuroraDBPostgres

Following are the options specific to AuroraDBPostgres:

Connection Type:	AuroraDBPostgres	SSL:	<input type="checkbox"/>
Connection Name:	aurora_conn	Group/IP Range:	Default_IPRange_DMA_1
Hostname:	aurora1	Database Name:	dep
Port Number:	0	Password:	*****
User Name:	admin		

- Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.

- c) **Hostname:** Enter the Hostname of the AuroraDBPostgres server or select from the list of available IPs.
 - d) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
 - e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
 - f) **Database Name:** Enter the name of the database.
 - g) **User Name:** Enter the database user name.
 - h) **Password:** Enter the database password.
6. After providing the connection details enter the name of the database in the bottom left filter by database panel, select the required schemas and databases and click **Add**. The selected databases will appear in the panel on the bottom right.



7. Click the **Test** button, to test the connection.

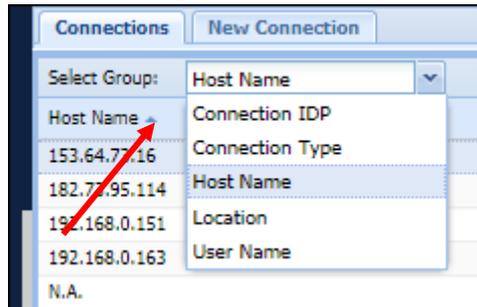


8. Click **Save** button, to save the changes.

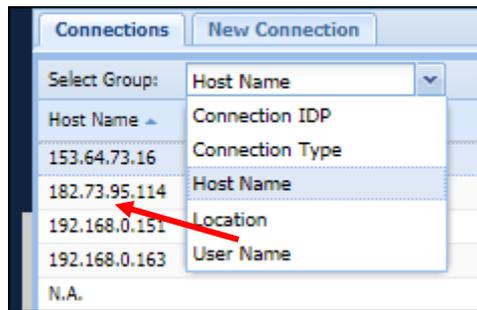


9. To edit a connection, go to AWS>RDS/Redshift > Connection Manager > Connections.

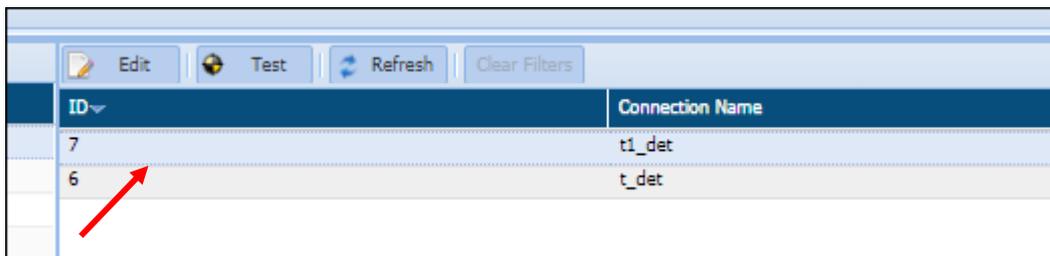
10. Select the connection classification from the Select Group dropdown.



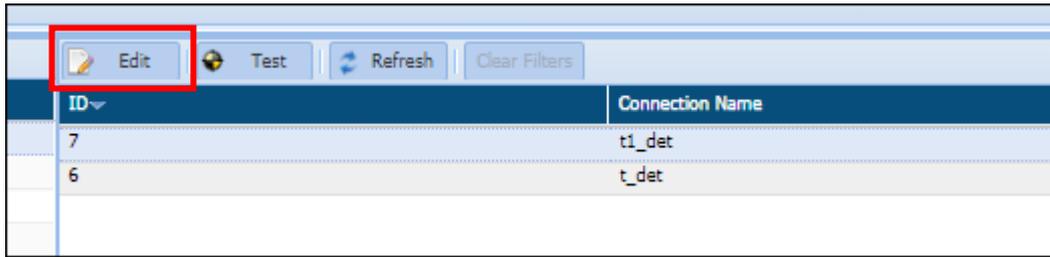
11. Select the classification type.



12. Select the connection.



13. Click **Edit**.

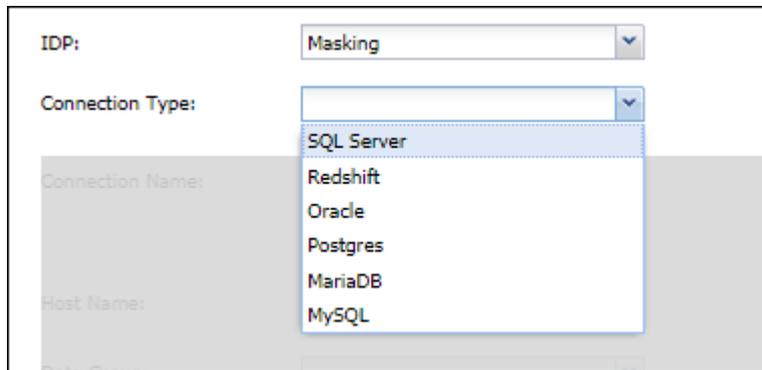


14. IDP and connection type cannot be edited. A connection can be edited using the same steps as task creation.

Masking

Perform the following steps to create a masking connection for different databases:

1. Select the database from the **Connection Type** drop-down.



2. Masking is supported on the following RDS/Redshift databases:

SQL Server

Refer to section [SQL Server](#)

Redshift

Following are the options specific to Redshift:

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.
- b) **SSL:** For an additional layer of security check the SSL (Secure Socket Layer) checkbox. For details on how to setup SSL please refer to DgSecure Admin Guide.
- c) **Hostname:** Enter the Hostname of the Redshift server or select from the list of available IPs.
- d) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **User Name:** Enter the database user name.
- h) **Password:** Enter the database password.
- i) **Use Connection String:** A Connection String specifies the information about the data source and the means of connecting to it. Check the Use Connection String checkbox, and provide the connection string. Use this option if you want to connect using the connection string instead of the hostname and port number.

Oracle

Refer to section [Oracle](#)

Postgres

Refer to section [Postgres \(Inc. Heroku\)](#).

MariaDB

Following are the options specific to MariaDB:

Connection Type:	MariaDB	Hostname:	maria1
Connection Name:	maria_conn	Port Number:	3306
Group/IP Range:	Default_IPRange_DMA_1	Password:	*****
Database Name:	dep		
User Name:	admin		

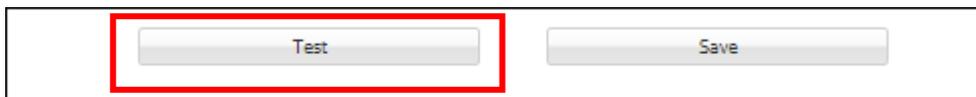
- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.

- b) **Hostname:** Enter the Hostname of the MariaDB server or select from the list of available IPs.
- c) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- d) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- e) **Database Name:** Enter the name of the database.
- f) **User Name:** Enter the database user name.
- g) **Password:** Enter the database password.

MySQL

Refer to section [MySQL](#)

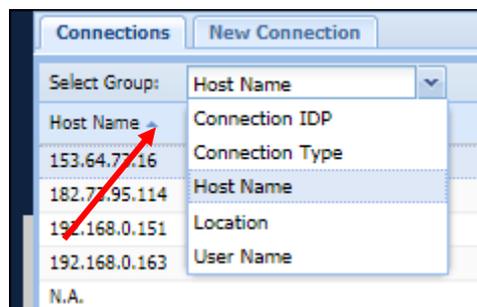
3. Click the **Test** button, to test the connection.



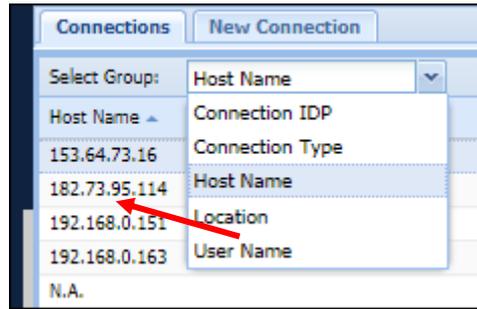
4. Click **Save** button, to save the changes.



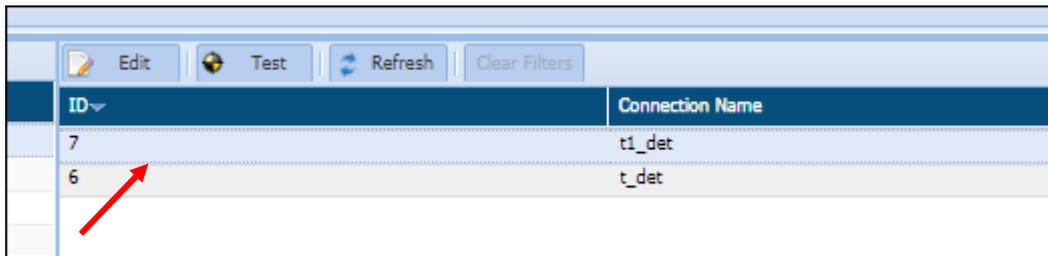
5. To edit a connection, go to AWS>RDS/Redshift > Connection Manager > Connections.
6. Select the connection classification from the **Select Group** dropdown.



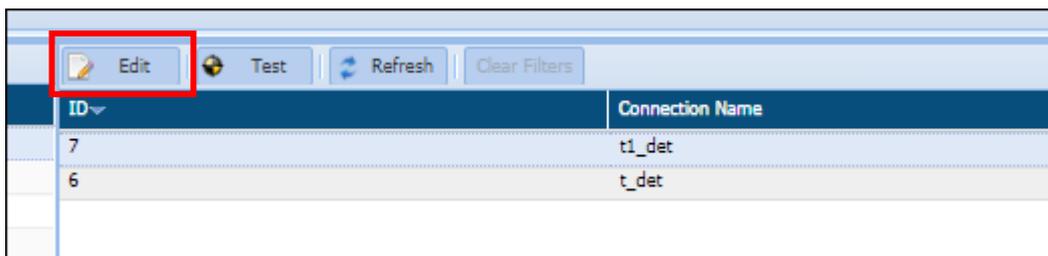
7. Select the classification type.



8. Select the connection.



9. Click **Edit**.

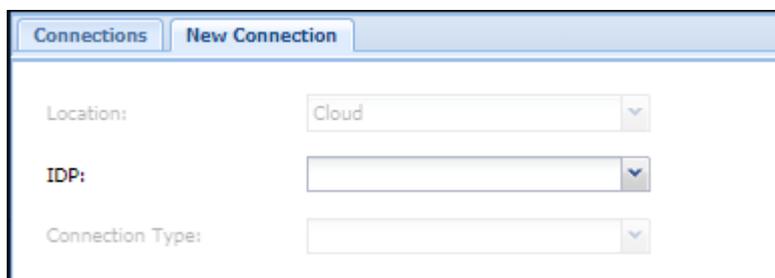


IDP and connection type cannot be edited. A connection can be edited using the same steps as task creation.

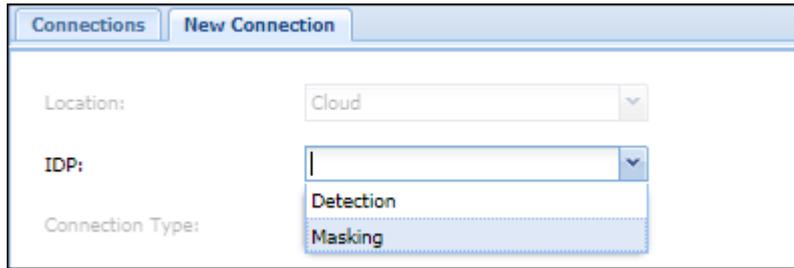
4.2.4 Azure

DgSecure supports masking and detection in Databases on Azure. Perform the following steps to create a connection:

1. Click **Azure>Databases > Connection Manager > Connections > New Connection** tab.



2. The location is **Cloud** by default.
3. Select the IDP type as Detection or Masking, depending on task that has to be executed on the database.

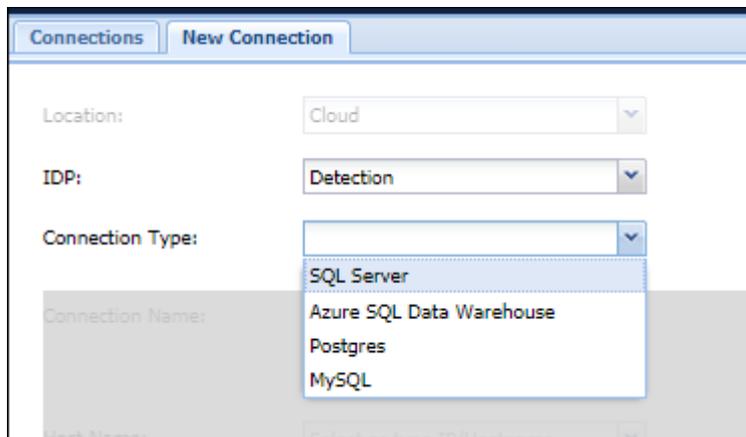


4. Following are the steps to create a connection for Detection and Masking:

Detection

Perform the following steps to create a detection connection for your database:

- a) Select the database from the **Connection Type** drop-down.



- b) Detection is supported on the following RDS/Redshift databases:

SQL Server

Refer to section [SQL Server](#).

Azure SQL Data Warehouse

Following are the options specific to the Azure SQL Data Warehouse:

Connection Type:	Azure SQL Data Warehouse	Group/IP Range:	Default_IPRange_DMA_1
Connection Name:	Azuredata_conn		
Server Name:	azure1		
Authentication Method:	<input checked="" type="radio"/> SQL Server <input type="radio"/> Active Directory Integrated		
Port Number:	1433		
Database Name:	master		
User Name:	admin	Password:	*****

- a) **Connection Name:** Enter unique Connection Name. This fields accepts letters, numbers, and symbols.

- b) **Server Name:** Enter the name of the Azure SQL Data Warehouse Server or select from the list of available IPs.
- c) **Group/IP Range:** Select the Group/IP range from the dropdown. For details on how to create and assign an IP range, please refer to the DgSecure Admin Guide.
- d) **Authentication Method:** The following authentication methods are available for SQL Server:
 - **SQL Server:** Authenticates the connection using SQL Server Username and Password.
 - **Active Directory Integrated:** Authenticates the connection using the details provided at the time of Azure Active Directory setup. For more details refer to DgSecure Admin Guide.
- e) **Port Number:** Enter the Port Number for establishing a connection. You can also edit the port number later, if required.
- f) **Database Name:** Enter the name of the database.
- g) **User Name:** Enter the database user name.
- h) **Password:** Enter the database password.

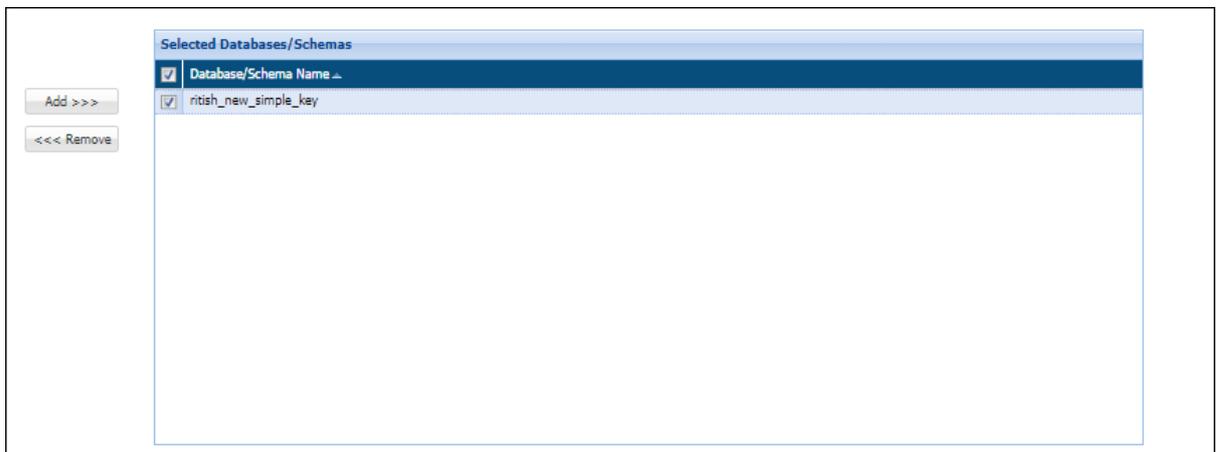
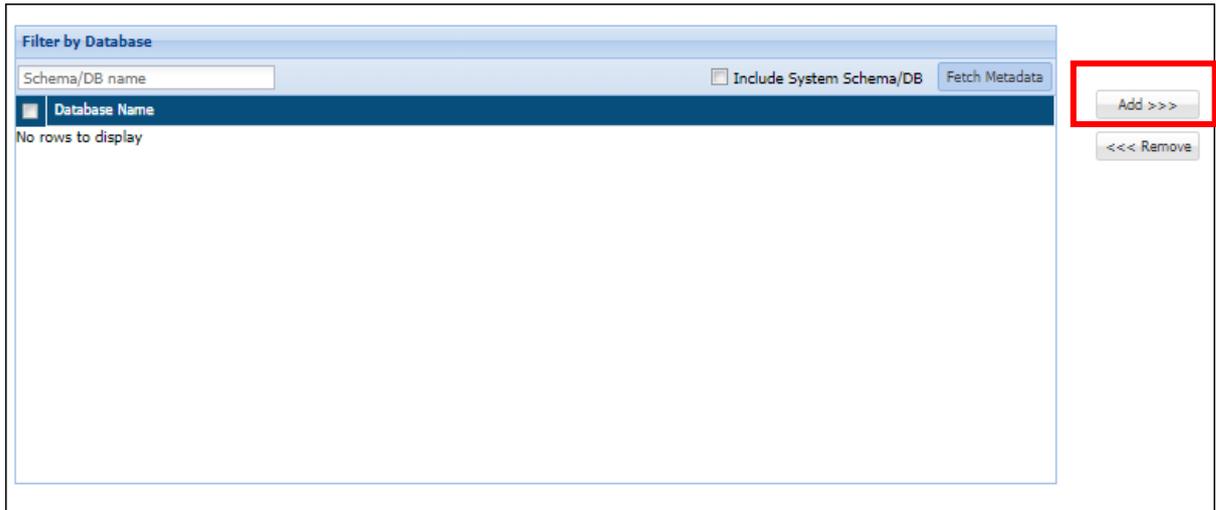
Postgres

Refer to section [Postgres \(Inc. Heroku\)](#).

MySQL

Refer to section [MySQL](#)

5. After providing the connection details enter the name of the database in the bottom left filter by database panel, select the required schemas and databases and click **Add**. The selected databases will appear in the panel on the bottom right.



6. Click the **Test** button, to test the connection.

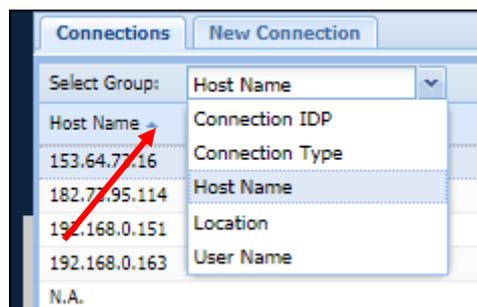


7. Click **Save** button, to save the changes.

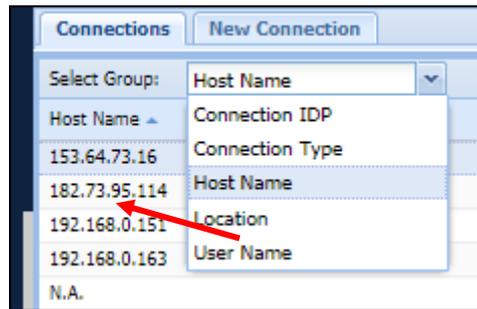


8. To edit a connection, go to AWS>RDS/Redshift > Connection Manager > Connections.

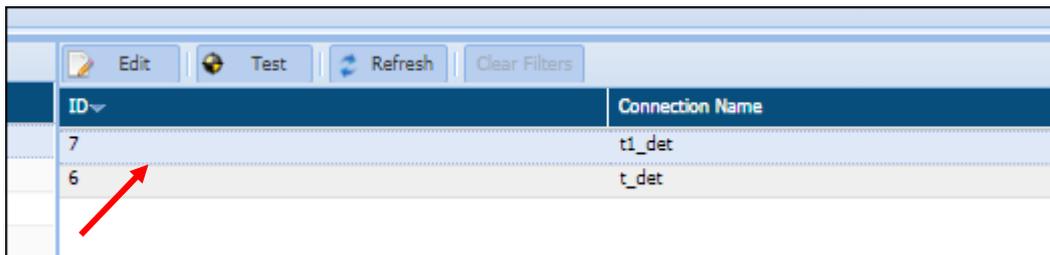
9. Select the connection classification from select group.



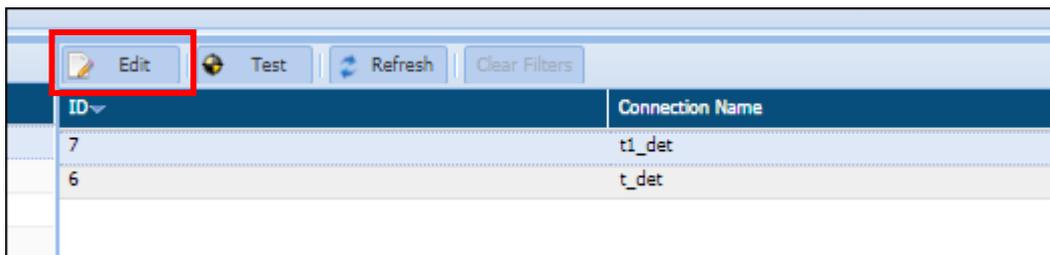
10. Select the classification type.



11. Select the connection.



12. Click edit.

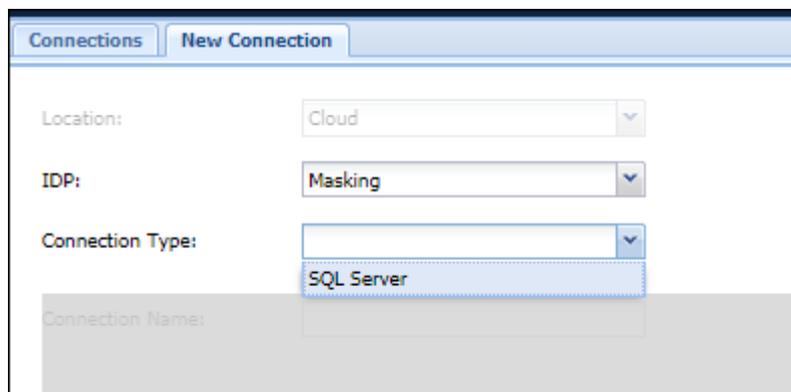


IDP and connection type cannot be edited. A connection can be edited using the same steps as task creation.

4.2.4.1 Masking

Perform the following steps to create a masking connection for the database:

1. Select the database from the **Connection Type** drop-down.



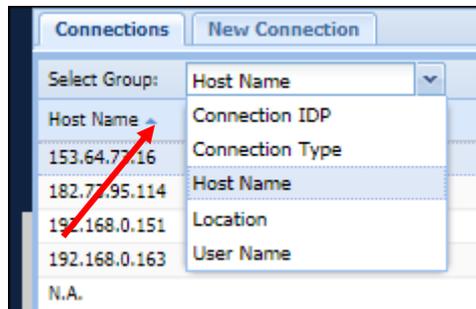
- Masking is supported on SQL Server. Refer to section [SQL Server](#) for details on how to create a connection.
- Click the **Test** button, to test the connection.



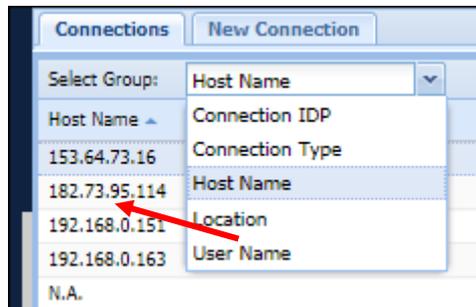
- Click **Save** button, to save the changes.



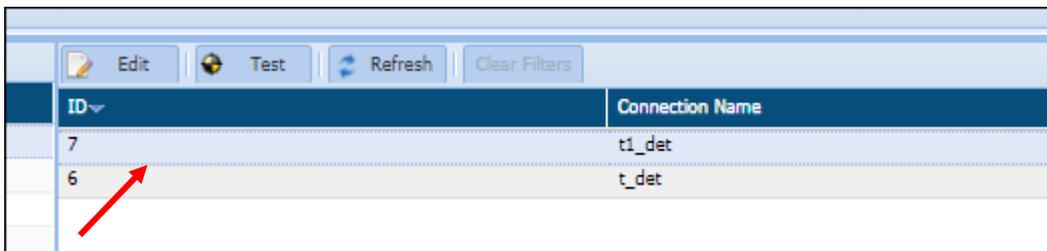
- To edit a connection, go to **AWS>RDS/Redshift > Connection Manager > Connections**.
- Select the connection classification from the **Select Group** dropdown.



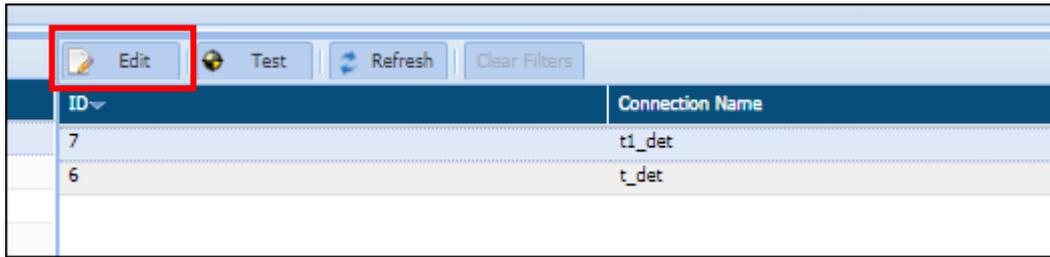
- Select the classification type.



- Select the connection.



- Click **Edit**.



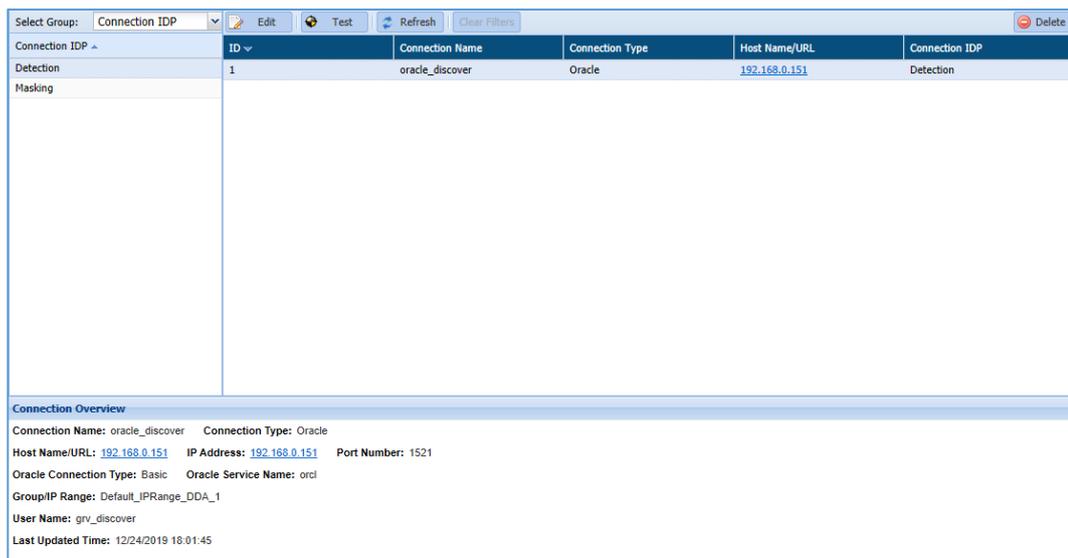
IDP and connection type cannot be edited. A connection can be edited using the same steps as the task creation.

4.3 List a Connection

4.3.1 RDBMS

This section will explain the screen of the Connections tab.

The below screenshot shows the user interface for Connections.



The Connections screen is divided into two panes. These are:

1. Connection List
2. Connection Overview

a) **Connection List:**

This pane list down all the connections groups and number of available connections for each selected connection group. It provides the basic details for the listed connection such as ID, Connection Name, Connection Type, Host Name/URL and Connection IDP.

Connection IDP	ID	Connection Name	Connection Type	Host Name/URL	Connection IDP
Detection	5	DBMS_detection_Oracle_conne...	Oracle	192.168.5.65	Detection
Masking	2	sql_151	SQL Server	192.168.0.151	Detection
	1	SQLServer_Detection_Conn	SQL Server	192.168.0.151	Detection

- ❖ **Select Group:** The Select Group drop-down list the five pre-defined connections groups. These are:
 - Connection IDP
 - Connection Type
 - Host Name
 - Location
 - User Name

- ❖ **Edit:** Click the Edit button, if you want to edit any information for the listed connection.

- ❖ **Test:** Click the Test button to test the listed RDBMS connection.

- ❖ **Refresh:** Click the Refresh button to update the current page with updated information.

- ❖ **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Connection page.

- ❖ **Delete:** Select the Connection which you want to delete. Click the Delete button.

b) Connection Overview

This pane displays the information for the selected connection. The Details include Connection Name, Connection Type, Host Name/URL, IP Address, Port Number, etc.

Connection Overview

Connection Name: oracle_discover Connection Type: Oracle

Host Name/URL: [192.168.0.151](#) IP Address: [192.168.0.151](#) Port Number: 1521

Oracle Connection Type: Basic Oracle Service Name: orcl

Group/IP Range: Default_IPRange_DDA_1

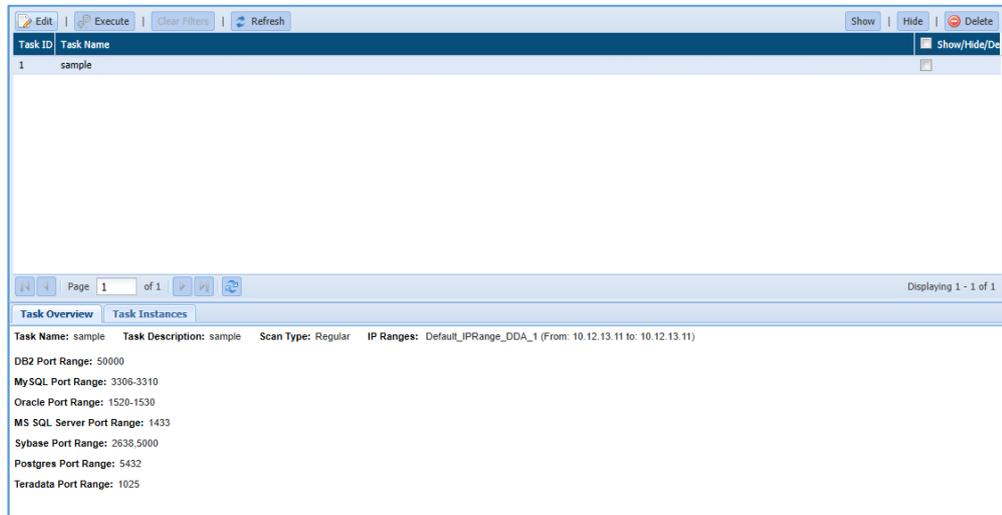
User Name: gnv_discover

Last Updated Time: 12/24/2019 18:01:45

4.3.1.1 Find DBMS

This section will explain the screen of the Task tab in Find DBMS section.

The below screenshot shows the user interface for Task tab.



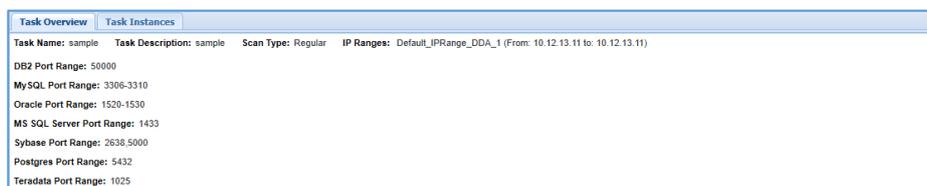
The Task page in Find DBMS is divided into three panes:

1. Task Detail: The Task Detail pane will display the list of all available task in this pane. It will display the information for the task such as Task ID (system generated), Task Name.

You can Edit, Execute, Show, Hide or Delete a task.



2. Task Overview: The Task Overview tab display information for the selected task in the Task Detail pane. The information includes basic details such as Task Name, Task Description, IP Range, Scan Type, DB2 Port Range etc.



3. Task Instance: The Task Instance pane shows information about each instance of the task selected in the Task Detail pane. Information such as ID (system generated), Task Name, Start Time, End Time, Status of the Task.

ID	Task Name	Start Time	End Time	Status
22	sample	Dec-12-2019 12:42:38	Dec-12-2019 12:42:49	Completed
21	sample	Dec-12-2019 11:54:11	Dec-12-2019 11:54:22	Completed

4.3.1.2 Result

The Find DBMS result page displays the information about the databases discovered on an organization’s IP ranges or n network.

To Access the Result page. Click **RDBMS > Connection Manager > Find DBMS > Result**.

The below image shows user interface of the Result Page.

The screenshot shows the 'Task Instances' table with two rows of completed tasks. Below it, the 'Task Instance Details' pane is expanded, showing task parameters and a 'Results Overview' section. The 'Results Overview' section contains a bar chart for MySQL and Oracle, and a pie chart showing a 50/50 split between Oracle and MySQL.

The Result page is divided into three panes. These are:

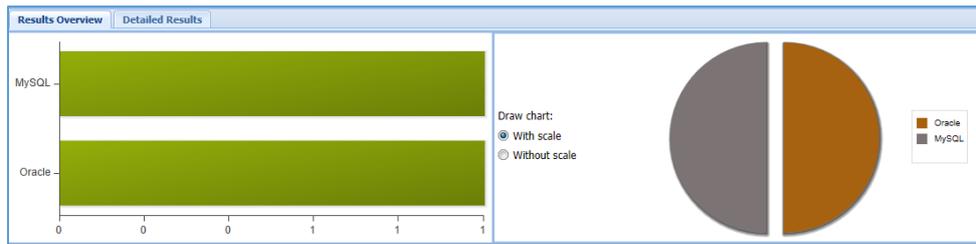
1. **Task Instances:** This pane displays the information for all the tasks. The information includes ID (system generated), Task Name, Status of the task, Start Time, End Time etc.

You can also re-run the task by clicking on **Re-Execute** button.

2. **Task Instance Details:** This pane displays display the parameters and results. Information includes Task Name, Task Description, Scan Type, IP Ranges, DB2 Port Range, etc.

The screenshot shows the 'Task Instance Details' pane with the following information: Task Name: sample, Task Description: sample, Scan Type: Regular, IP Ranges: Default_IPRange_DDA_1(From: 10.12.13.11 to: 10.12.13.11), DB2 Port Range: 50000.

3. **Overview:** The overview pane provides the basic information in Results and Detailed Results tabs. The information is displayed is dependent on the currently selected tab.



- **Results Overview:**

The Results Overview pane gives a graphical summary for the search results. The bar chart shows the number of databases detected. The pie chart on the right shows the same information as the bar chart with each segment representing a different database type.

- **Detailed Results:**

The Detailed Results lists the specific databases that the task instance detected, grouped by type. For each database, it displays the name and IP address of the host machine and the port number on which the database listens for requests.

The screenshot shows the 'Detailed Results' pane with a table containing the following data:

Database Type	IP Address	Host Name	Port Number
MySQL	10.12.13.11	N/A	3306
Oracle	10.12.13.11	N/A	1521

At the top right of the pane are buttons for 'Clear Filters', 'Save Results to File', and 'Save Results to Pdf'. At the bottom, there is a pagination control showing 'Page 1 of 1' and 'Displaying 1 - 2 of 2'.

- **Save Results to File:** Click the Save Results to File button to download the data in Doc format.
- **Save Results to PDF:** Click the Save Results to PDF button to download the data in PDF format.

4.3.2 NoSQL

This section will explain the screen of the Connections tab.

The below screenshot shows the user interface for Connections.

To know more about the connections screen, refer section [RDBMS](#)

4.3.3 AWS

This section will explain the screen of the Connections tab.

The below screenshot shows the user interface for Connections.

To know more about the connections screen, refer section [RDBMS](#)

4.3.4 Azure

This section will explain the screen of the Connections tab.

The below screenshot shows the user interface for Connections.

To know more about the connections screen, refer section [RDBMS](#)

5 Sensitive Type

A **Sensitive Type** is the basis for detection of sensitive information in DgSecure. Different **Sensitive Types** are data elements within databases that indicate or comprise of private and confidential information. This information must be protected from unauthorized access to safeguard the privacy or security of an individual or organization. Detection is performed on a database, to locate this sensitive information within it and the scan is based on the **Sensitive Type** parameter. Various **Sensitive Types** can be used in combinations or as singular entities to detect where sensitive information is stored in a data source. Some examples of such sensitive data are, **Credit Card numbers, Social Security numbers, Addresses** etc. DgSecure houses several Sensitive Types to detect sensitive information suited to various scenarios. DgSecure’s capability to create new **Sensitive Types**, furthers the flexibility and accuracy of detection. Creating a new sensitive type is discussed in detail under further sections.

The function of a **Sensitive Type** is to facilitate and serve as the basis for protection and detection of sensitive information. **Sensitive Types** can be defined as **singleton, dependent or composite** entities. These have been described below:

- **Singleton Sensitive Type**
A standalone **Sensitive Type** that is capable of being uniquely identifiable by its description is defined as a **Singleton Sensitive Type**. Detection of a **Singleton Sensitive Type** is based on the **Sensitive Types’** characteristics, and is independent and unrelated to any other data within the data source.
- **Dependent Sensitive Type**
As the name suggests, this is a **Sensitive Type** that is dependent on another. For example, there is a good chance of finding a credit card number in the same row as a phone number, this can be used to define the dependency between the sensitive information. A simple relationship can be established in such a scenario, by using the **Detect if Found with** option when creating a new **Sensitive Type**.
- **Composite Sensitive Type**
A family of **Sensitive Types** that have compound dependencies on one another can be termed as a **Composite Sensitive Type**. Contact information, for example, consists of many such data points, i.e., addresses, phone numbers, area codes, fax numbers etc., and there a very good possibility of detecting these in the same row. A composite sensitive type in DgSecure can be defined using the feature to **Group** different **Sensitive Types** when creating a new **Sensitive Type**.

5.1 Default Sensitive Type

The **Sensitive Types** tab in **Sensitive Type Manager**, lists all the available **Sensitive Types** in DgSecure. There are more than 90 different available **Sensitive Types** to detect all kinds of sensitive data, be it addresses, emails, phone numbers, bank details, personal information, insurance details or salary. **Sensitive Types** can be used in combinations for further streamlining the search. Multiple **Sensitive Types** can be grouped together under a policy ([Section no.](#)) and selected in isolation at the time of task creation ([Section no.](#))

Sensitive Type	Defined For	Sensitive Group	Confidence Factor Config	Created On	Processing Order	Edit Confidence
Driver License (Nevada)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (New Hampshire)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (New Jersey)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (New York)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (North Carolina)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Ohio)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Pennsylvania)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Rhode Island)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (South Carolina)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (South Dakota)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Texas)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Utah)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Vermont)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Virginia)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Washington)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (West Virginia)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		
Driver License (Wisconsin)	DBMS, Hadoop & Files	Driver License	Modified	November-08-2019 16:13		

Sensitive Type Overview
 Sensitive Type: Bulgaria Addresses Sensitive Type Description: Bulgaria Addresses
 Defined For: DBMS, Hadoop & Files
 Column Data WhiteList: No Column Data BlackList: No
 Detect if found with:
 Sensitive Group Name: European Addresses Sensitive Group Type: Predefined
 Created On: November-08-2019 16:13 Updated On: November-08-2019 16:13

The top panel of the **Sensitive Types** screen lists all the available **Sensitive Types**. When a new **Sensitive Type** is created it will appear on this screen. This panel shows the sensitive type name along with the assigned data type (DBMS or Hadoop & Files), sensitive group, creation date, and processing order.

Processing order is applicable only to the custom **Sensitive Types** and is used to determine which of two similar **Sensitive Types** should be given preference when both are used in a task. Custom **Sensitive Types** at the top of the screen take precedence over those at the bottom. Increase or decrease priority using the arrows in the **Processing Order** column.

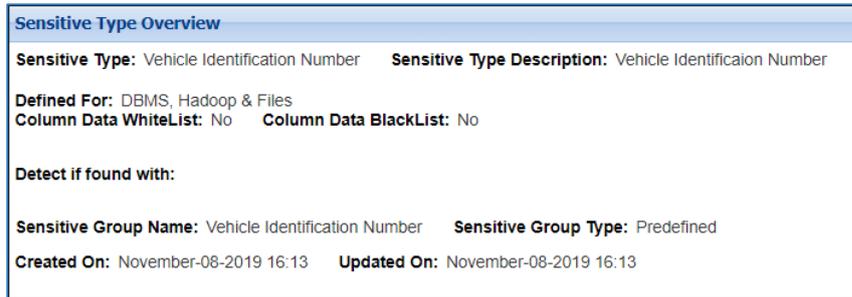
Sensitive Type	Defined For	Sensitive Group	Confidence Factor Config	Created On	Processing Order	Edit Confidence
EmployeeID	DBMS	EmployeeID	Default	October-03-2019 22:58		
HRN	Hadoop & Files	HRN	Default	September-11-2019 22:33		
SalesOrderID	DBMS	SalesOrderID	Modified	August-20-2019 02:31		
BusinessEntityID	DBMS	BusinessEntityID	Modified	August-17-2019 04:29		
AddressID	DBMS	AddressID	Default	August-16-2019 08:58		

Select the **Sensitive Type** to **Edit**, **Delete** or **Add Confidence Parameters** to it ([Confidence factor is discussed under section.](#)) To filter **Sensitive Types** on this screen, use the **Group by** dropdown to select sensitive type groups for viewing.

Sensitive Type	Defined For	Sensitive Group	Confidence Factor Config	Created On	Processing Order	Edit Confidence
Vehicle Identification Number	DBMS, Hadoop & Files	Vehicle Identification Nu...	Modified	November-08-2019 16:13		
Address						
US Address	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
UK Address (Unstructured data only)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
Canada Address (Unstructured data only)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
Address Line (Best suited for structured data)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
Address State (Best suited for structured data)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
Address City (Best suited for structured data)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
Address Zip (Best suited for structured data)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
Address Country (Best suited for structured data)	DBMS, Hadoop & Files	Address	Modified	November-08-2019 16:13		
European Addresses						
Bulgaria Addresses	DBMS, Hadoop & Files	European Addresses	Modified	November-08-2019 16:13		
Sweden Addresses	DBMS, Hadoop & Files	European Addresses	Modified	November-08-2019 16:13		
Hungary Addresses	DBMS, Hadoop & Files	European Addresses	Modified	November-08-2019 16:13		
Finland Addresses	DBMS, Hadoop & Files	European Addresses	Modified	November-08-2019 16:13		

***NOTE: Default sensitive types cannot be edited or deleted, however they can be cloned.**

The bottom panel provides information about the name, description, whether it is applicable to Hadoop & Files or DBMS, the parameters of the sensitive type, the assigned sensitive group, creation date, and update date of the selected **Sensitive Type**.



5.2 New Sensitive Type

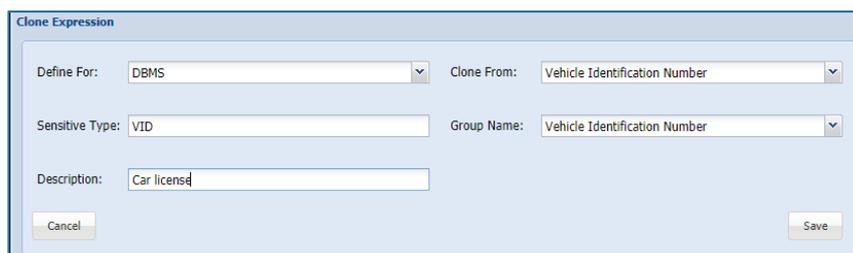
For increased flexibility and efficiency in sensitive data detection suitable to any platform, environment, data source and nature of the data, DgSecure comes complete with the capability to support user defined-custom **Sensitive Types**. A **Sensitive Type** filters data based on a regular expression and narrows down the search using additional inputs such as validation functions, inclusion and exclusion lists etc.

The simplest method to create a **Sensitive Type** in DgSecure is to clone it from an existing **Sensitive Type**. The out-of-the-box **Sensitive Types** in DgSecure are immutable. By cloning from an existing Sensitive Type, DgSecure’s default sensitive types can be edited to conform them to the target data source. Take the following steps to clone a sensitive type:

1. Go to the Sensitive Types tab and click Clone Expression.



2. Provide the required details, select the target data source from the **Define For** dropdown (DBMS or Hadoop and Files), select the **Sensitive Type** you would like to clone from the **Clone From** dropdown, provide a **Sensitive Type** name, select the **Sensitive Type** group from the **Group Name** dropdown and enter a **Description**.



3. **Save the Sensitive Type.** The cloned sensitive type will appear on the **Sensitive Types** screen.

Alternatively, a completely new **Sensitive Type** can also be created. The process for creation of a new **Sensitive Type** differs in DBMS and Hadoop and Files. These have been discussed below.

5.2.1.1 DBMS

Perform the following steps to create a **New Sensitive Type** for **DBMS**:

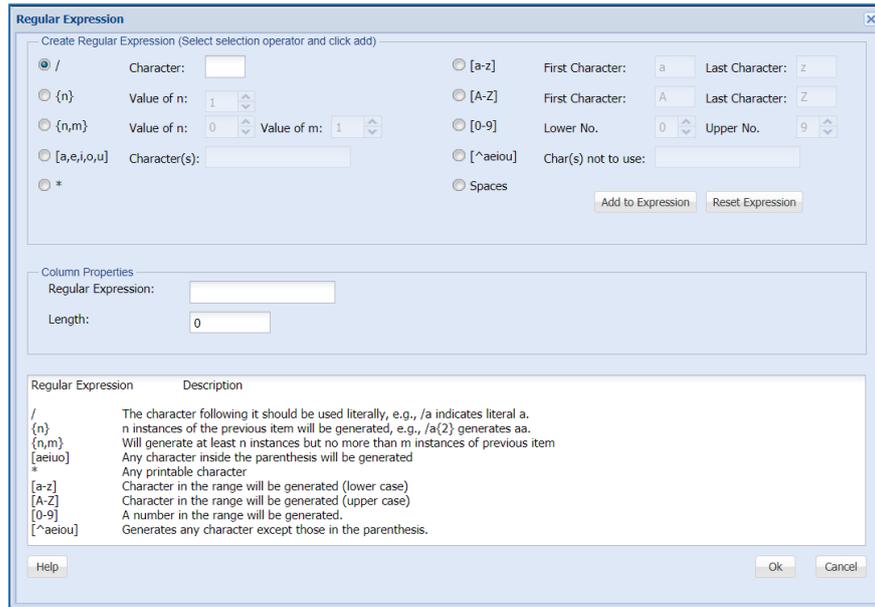
1. Go to the **New Sensitive Type** tab and select **DBMS** from the **Define For** dropdown.

2. Provide a **name, description** and select the **group** of the **Sensitive Type**.

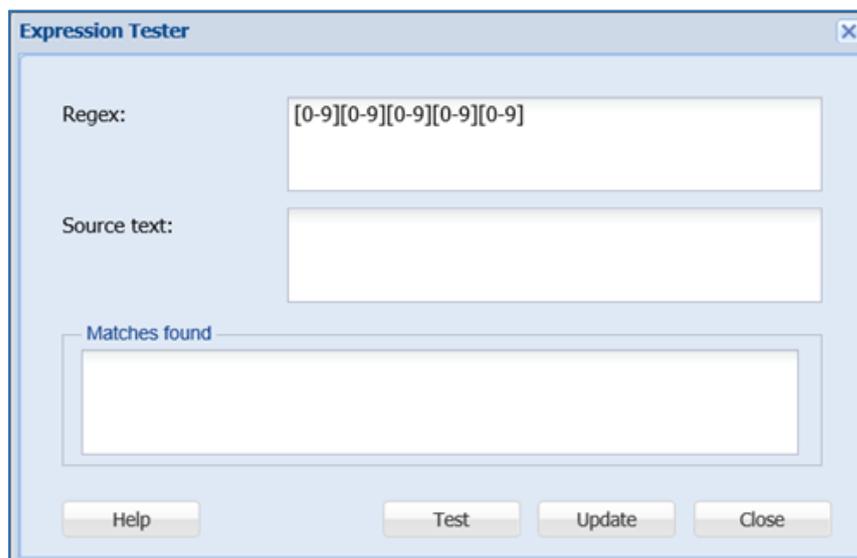
3. Enter a **Data** and **Column Regex**.

Following DgSecure tools have been added to this screen to provide assistance in creating regexes. Alternatively, Regex can also be directly entered into the indicated fields.

- **Expression Builder:** Use expression builder tool to create a Regex.

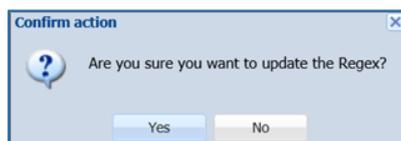


- **Expression Tester:** Use the expression tester tool to test if the created regex is working on a sample data. The **Regex** field is editable.



Perform the following steps to edit a regex:

- Provide a different regular expression in the **Regex** field.
- Click **Update** to save the changes.
- A pop up appears:



Click **Yes** to make the changes effective.

***Note:** It is mandatory to define both Data and Column Regexes to create a sensitive type for DBMS.

- The **Mandatory Field Name Match** checkbox ensures that detection results appear only if Data and Column, both Regexes match.

A screenshot of a form element showing a checkbox labeled "Mandatory Field Name Match" which is checked. Above the checkbox is a button labeled "Expression Builder".

- Select additional **Sensitive Types** under **Detect if Found With** dropdown. This will create dependency between two **Sensitive Types** and report the **Sensitive Type** if found in the same row as the selected **Sensitive Type**.

A screenshot of a form element showing a dropdown menu labeled "Detect If Found With:" with "US Address" selected. Above it is a checked checkbox for "Mandatory Field Name Match".

- Additionally, an **inclusion** or **exclusion list** can also be added to include or exclude certain data and fields from the detection.

A screenshot of a form section titled "Reference Data:" and "Upload Column Name File:". It contains two rows of input fields for "Inclusion List" and "Exclusion List", each with "Browse..", "Remove", and "Download" buttons.

- Upload a **validation function** to verify the detection.

A screenshot of a form element showing a dropdown menu labeled "Validation Function:" with "No Option Selected" selected.

- Click **Add Confidence Factor Config** to setup confidence factor calculations for the **Sensitive Type**. **Confidence factor** is discussed in detail under [Section no.](#)

A screenshot of a form element showing a dropdown menu labeled "Validation Function:" with "No Option Selected" selected, and a button labeled "Add Confidence Factor Config" to its right. Above it are "Inclusion List" and "Exclusion List" fields.

5.2.1.2 Hadoop and Files

Perform the following steps to create a **New Sensitive Type** for **Hadoop and Files**:

1. Go to the **New Sensitive Type** tab and select **Hadoop and Files** from the **Define For** dropdown.

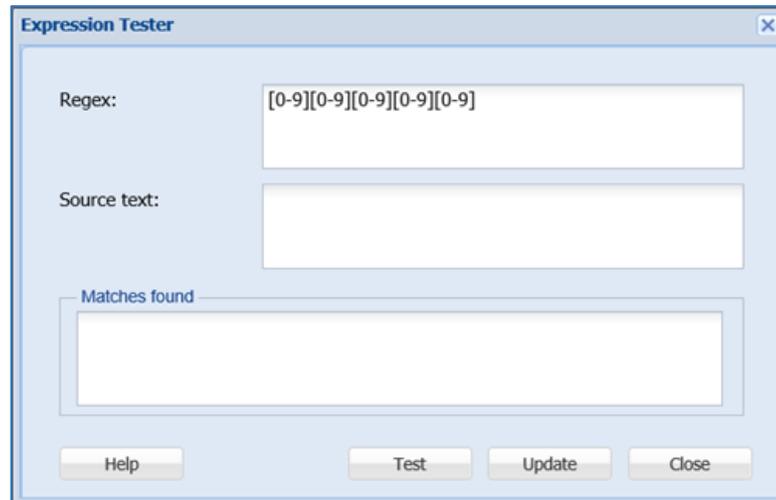
2. Provide a **name**, **description** and select the **group** for the **Sensitive Type**.
3. Enter a Data and Column Regex.

Following DgSecure tools have been added to this screen to provide assistance in creating regexes. Alternatively, Regex can also be directly entered into the indicated fields.

- **Expression Builder:** Use expression builder tool to create a Regex.

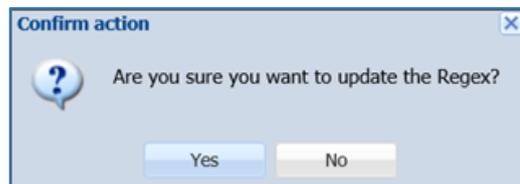
Regular Expression	Description
/	The character following it should be used literally, e.g., /a indicates literal a.
{n}	n instances of the previous item will be generated, e.g., /a(2) generates aa.
{n,m}	Will generate at least n instances but no more than m instances of previous item
[aeiou]	Any character inside the parenthesis will be generated
*	Any printable character
[a-z]	Character in the range will be generated (lower case)
[A-Z]	Character in the range will be generated (upper case)
[0-9]	A number in the range will be generated.
[^aeiou]	Generates any character except those in the parenthesis.

- **Expression Tester:** Use the expression tester tool to test if the created regex is working on a sample data. The **Regex** field is editable.



Perform the following steps to edit a regex:

- Provide a different regular expression in the **Regex** field.
- Click **Update** to save the changes.
- A pop up appears:



Click **Yes** to make the changes effective.

- Check the **Numeric** checkbox if the data is in numeric format.



- The **Mandatory Field Name Match** checkbox ensures that detection results appear only if Data and Column, both Regexes match.



6. Select additional **Sensitive Types** under **Detect if Found With** dropdown. This will create dependency between two **Sensitive Types** and report the **Sensitive Type** if found in the same row as the selected **Sensitive Type**.

Mandatory Field Name Match :

Detect If Found With:

US Address

7. Additionally, a **master list** and an **inclusion** or **exclusion list** can also be added to include or exclude certain data and fields from the detection.

Master List *:

Inclusion List: e.g. C:\dg\sample.txt

Exclusion List: e.g. C:\dg\sample.txt

8. Upload a validation **function** to verify the detection.

Validation Function: No Option Selected

9. Click **Add Confidence Factor Config** to setup confidence factor calculations for the sensitive type. **Confidence Factor** is discussed in detail under [Section 4.5](#).

Inclusion List: e.g. C:\dg\sample.txt Exclusion

Validation Function: No Option Selected

5.3 Inherit Sensitive Type

In addition to the default **Sensitive Types** and custom **Sensitive Types**, properties of a **Sensitive Type** can also be derived from another-parent **Sensitive Type** to create a new **Sensitive Type**. This is termed as inheritance in DgSecure. Inheritance provides the option to pick some or all aspects of previously defined or default **Sensitive Types** and use them to define a new **Sensitive Type**.

The advantage of this feature over **Cloning** of a **Sensitive Type** is that, when the base/ parent **Sensitive Type** is changed, the changes will automatically reflect to the derived **Sensitive Type**. Perform the following steps to create a **Sensitive Type** with the derived properties:

1. Go to New Sensitive Type tab in the Sensitive Type Manager.
2. Select the module from the **Define For** dropdown.

3. Select the **Derive From** option from the **Derivation Type** dropdown to enable the inheritance.

4. Select the required parent **Sensitive Type** from the **Derive From** dropdown.

5. Enter a name for the new derived **Sensitive Type** in the **Sensitive Type** text box. Provide a brief **description** and select the **group** for the **Sensitive Type Description** and **Group Name** fields respectively.

6. There are 3 ways to define the Regex:

- a) To derive the complete Regular Expression (Regex) from the parent **Sensitive Type**, enter (\$base) as column and data Regex. This is the default setting.

Define For: DBMS Derivation Type: Derive From Derive From: Vehicle I

Sensitive Type: Test Sensitive Type Description: A controlled set to test inheritance responses

Group Name: Test

Regex for Data: (\$base) Regex for Column Name: (\$base)

Expression Builder Expression Tester Expression Builder Expression Tester

- b) To edit or make changes to the inherited Regex, enter (\$base) followed by the required addition to the Regex. For example, the Regex (\$base) + 0001, will look for sensitive types as per the definition in the parent **Sensitive Type** and the figure 0001.

Sensitive Types New Sensitive Type

Define For: DBMS Derivation Type: Derive From Derive From: Vehicle Identification Number

Sensitive Type: Test Sensitive Type Description: A controlled set to test inheritance responses

Group Name: Test

Regex for Data: (\$base)+0001 Regex for Column Name: (\$base)

Expression Builder Expression Tester Expression Builder Expression Tester

- c) To create a new Regex instead of a derived Regex, remove (\$base) from the Regex field.

Sensitive Types New Sensitive Type

Define For: DBMS Derivation Type: Derive From Derive From: Vehicle Identification Number

Sensitive Type: Test Sensitive Type Description: A controlled set to test inheritance responses

Group Name: Test

Regex for Data: 0001 Regex for Column Name: (\$base)

Expression Builder Expression Tester Expression Builder Expression Tester

7. For an inherited **Sensitive Type** all additional fields, i.e., **Mandatory Field Name Match**, **Reference Data**, **Upload Column Name** will be derived from the parent **Sensitive Type**. To make any additions, uncheck the **Base** checkbox against the field to provide inputs.

Mandatory Field Name Match : Base:

Reference Data:

Inclusion List: e.g. C:\dg\sample.txt Browse... Remove Download Base Exclusion List: e.g. C:\dg\sample.txt Browse... Remove Download Base

Upload Column Name File:

Inclusion List: e.g. C:\dg\sample.txt Browse... Remove Download Base Exclusion List: e.g. C:\dg\sample.txt Browse... Remove Download Base

Validation Function: No Option Selected Base Add Confidence Factor Config Base

Note: In case of deriving an expression by default properties will be inherited from parent sensitive type.

8. Save the Sensitive Type.

Validation Function: Base Base

Note: In case of deriving a expression by default properties will be inherited from parent sensitive type.

5.4 Edit Sensitive Type

Perform the following steps to **Edit a Sensitive Type**:

1. Select the **Sensitive Type** that needs to be edited from the list on the **Sensitive Types** Screen.



2. Click **Edit**.



3. The following options are greyed out and cannot be edited:
 - a) Define for
 - b) Derivation Type
 - c) Derive From
4. All the other fields can be edited. These have been discussed under the [New Sensitive Type Section](#).

5.5 Confidence Factor

Confidence Factor illustrates the accuracy of Detection. Columns with a **Confidence Factor** less than the threshold value for Detection are not displayed.

6 Policy

6.1 Concept

Policy is a set of pre-defined rules and regulations. In DgSecure, policy comprises a set of guidelines which are created to protect the sensitive information. It contains a set of pre-defined and user-defined Sensitive Types and protection options.

DgSecure supports two types of policies:

1. **Pre-Defined Policies:** These policies are the ones that have been already defined within the DgSecure to improve the ease of usage. GDPR, PII, PCI and HIPAA are some of the Pre-Defined policies in the DgSecure.
2. **Customized Policies:** User can define its own policies and can also customize the Pre-Defined Policies by creating a copy.

The Pre-Defined Policies are:

1. **GDPR (General Data Protection Regulations):** GDPR was officially enacted across the EU on 25th May 2018. It is designed to protect the personal data and privacy of EU (European Union) citizens for every transaction that occurs within EU member states.
2. **HIPAA (Health Insurance Portability and Accountability Act):** HIPAA was enacted in the year 1996. HIPAA is designed to provide the security provisions and data privacy to keep the patients' medical information safe.
3. **PII (Personally Identifiable Information):** PII is any data that could potentially be used to identify a particular individual. PII information includes biometric information, medical information, Personally identifiable information (PIFI) and unique identifiers such as social security numbers. In the U.S., multiple federal laws regulate the protection of PII.
4. **PCI (Payment Card Information/Industry):** PCI was enacted in December 2004. It refers to the payment security standards that ensure all sellers safely and securely accept, store, process and transmit cardholder data during credit card transactions.
5. **CCPA (California Consumer Protection Act):** CCPA will come into effect from January 1st, 2020. This regulation is meant to enhance privacy rights and consumer protection for residents of California, US.

A user can perform various functions in DgSecure. These are:

- i. [Create a policy](#)
- ii. [Edit a policy](#)
- iii. [Export a policy](#)
- iv. [Import a policy](#)

6.2 Create a Policy

A user can create, edit, import and export policies using data sources. There are two type of data sources:

1. **DBMS:** Use DBMS module for creating a policy for structured data. Structured data is stored in well- defined schemas such as database. It is in a tabular form that clearly defines attributes.
2. **Hadoop & Files:** For Unstructured data use Hadoop & Files module. Unstructured data does not have identifiable structure. It lacks in a particular format and sequence.

***Note:** Pre-Defined policies are not editable.

The below section defines the process to create a policy.

6.2.1 DBMS

To create a policy, click **Policy > DBMS > Policy > New Policy** tab.

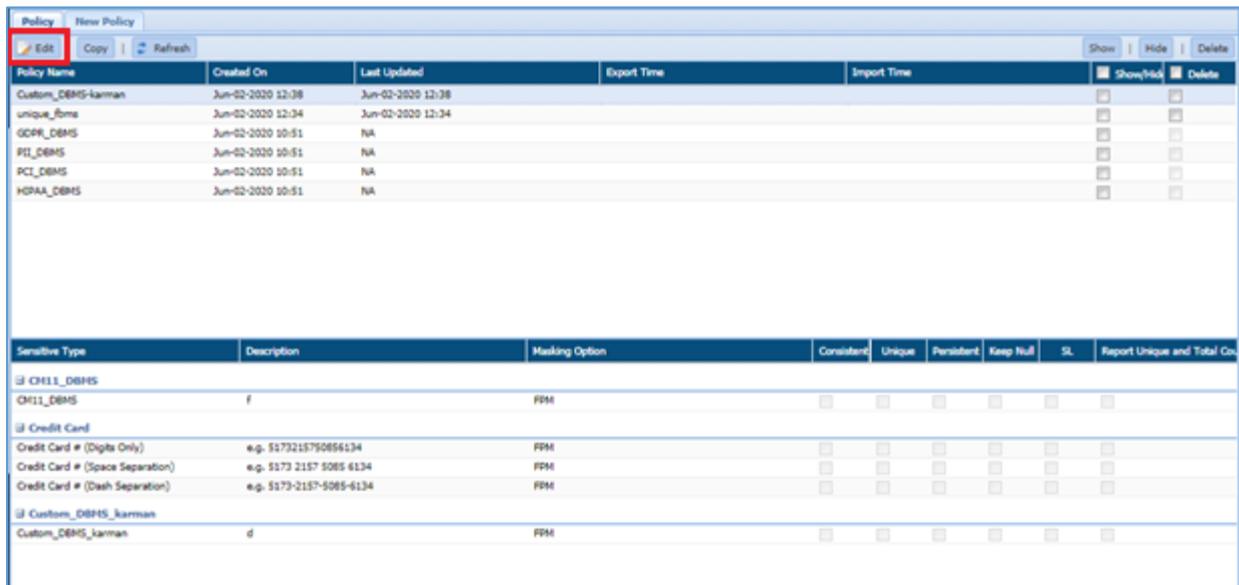
Following screenshot shows the user interface for creating a policy:

Sensitive Type	Description	Masking Option	Consistent	Unique	Persisted	Keep Null	SL	Report Unique
<input type="checkbox"/> Address								
<input type="checkbox"/> UK Address (Unstructured data only)	UK Address	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> US Address	US Address	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Canada Address (Unstructured data only)	Canada Address	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Address State (Best suited for structured data)	Address State	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Address City (Best suited for structured data)	Address City	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Address Country (Best suited for structured da...	Address Country	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> AddressID								
<input type="checkbox"/> AddressID	Address ID	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> BusinessEntityID								
<input type="checkbox"/> BusinessEntityID	Business Entity ID	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Credit Card								
<input type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Select Masking Option	<input type="checkbox"/>					
<input type="checkbox"/> Dates								
<input type="checkbox"/> Credit Card Expiry Date	Credit Card Expiry Date	Select Masking Option	<input type="checkbox"/>					

1. **Policy:** Enter the name of the policy. The name must be unique to for each policy.
2. **Description:** Describe the policy. This fields accepts letters, numbers, symbols and can hold 256 characters.
3. Include **GDPR Dashboard View:** Check this checkbox to display the policy in the **GDPR View** under the **Reports** section.

4. Sensitive **Type**: Select the Sensitive Type to include in the policy. A user can also define its own Sensitive Type in the **Policy > Sensitive Type Manager > New Sensitive Type** tab. User can select ‘n’ number of sensitive types while creating a policy. To select an entire group, check the checkbox next to the Sensitive Group.
5. Masking **Option**: Select the Masking option from the drop – down list. You have to select the CUPS options. There are six parameters for CUPS. These are:
 - i. **Consistent (C)**
 - ii. **Unique (U)**
 - iii. **Persistent (P)**
 - iv. **Keep Null**
 - v. **Stateless (SL)**
 - vi. **Report Unique and Total Count**
6. **Cancel**: Click the **Cancel** button, if the you do not want to save the changes.
7. **Save As**: Click the **Save As** button to save the policy with new name.
8. **Save**: Click the **Save** button to save the policy.
9. To edit a policy, go to **POLICY** tab and click **Edit** button. This functionality will enable you to edit a policy.

Select the policy in Policy panel and click **EDIT**.



6.2.2 Hadoop & Files

To create a policy in Hadoop & Files. Click **Policy > Hadoop & Files > Policy > New Policy** tab. The following screenshot shows the user interface for creating a policy:

Sensitive Type	Description	Protection Option	Consistent	Report Unique Count
Address				
<input type="checkbox"/> Canada Address (Unstructured data only)	Canada Address	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> US Address	US Address	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> UK Address (Unstructured data only)	UK Address	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Address State (Best suited for structured data)	Address State	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Address City (Best suited for structured data)	Address City	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Address Country (Best suited for structured data)	Address Country	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
Credit Card				
<input type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
Dates				
<input type="checkbox"/> Credit Card Expiry Date	Credit Card Expiry Date	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> DOB (Best suited for structured data)	Date Of Birth	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> DOD (Best suited for structured data)	Date Of Death	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> Date (Best suited for structured data)	Date	Select Protection Option	<input type="checkbox"/>	<input type="checkbox"/>

1. **Policy:** Enter the name of the policy. The name must be unique for each policy.
2. **Description:** Describe the policy. This field accepts letters, numbers, symbols and can hold 256 characters.
3. **Include GDPR Dashboard View:** Check this checkbox to display the policy in the **GDPR View** under the **Reports** section.
4. **Sensitive Type:** Select the Sensitive Type to include in the policy. A user can also define its own Sensitive Type in the **Policy > Sensitive Type Manager > New Sensitive Type** Tab. User can select 'n' number of Sensitive Types while creating a policy. To select an entire group, check the checkbox next to the Sensitive Group.
5. **Protection Option:** Select the Protection option from the drop-down list. You also need to select the CUPs Option. There are two parameters for CUPs Options. These are:
 - i. **Consistent (C)**
 - ii. **Report Unique Count**
6. **Cancel:** Click the **Cancel** button, if you do not want to save the changes.
7. **Save As:** Click the **Save As** button to save the policy with new a new name.
8. **Save:** Click the **Save** button to save the policy.

- To edit a policy, go to **POLICY** tab and click **Edit** button. This functionality will allow you to edit the information for the policy.

Select the policy in the Policy panel and click **EDIT**.

Policy Name	Created On	Last Updated	Export Time	Import Time	Show/Hide	Delete
intel	Jun-03-2020 10:32	Jun-03-2020 10:32			<input type="checkbox"/>	<input type="checkbox"/>
char	Jun-03-2020 10:28	Jun-03-2020 10:28			<input type="checkbox"/>	<input type="checkbox"/>
static	Jun-03-2020 10:26	Jun-03-2020 10:26			<input type="checkbox"/>	<input type="checkbox"/>
Random_c	Jun-03-2020 10:24	Jun-03-2020 10:40			<input type="checkbox"/>	<input type="checkbox"/>
ran_cons	Jun-03-2020 10:24	Jun-03-2020 10:24			<input type="checkbox"/>	<input type="checkbox"/>
ran	Jun-03-2020 10:24	Jun-03-2020 10:24			<input type="checkbox"/>	<input type="checkbox"/>
fpm_cons	Jun-02-2020 16:27	Jun-02-2020 16:27			<input type="checkbox"/>	<input type="checkbox"/>
unique	Jun-02-2020 14:50	Jun-02-2020 14:50			<input type="checkbox"/>	<input type="checkbox"/>
FPM_karman	Jun-02-2020 12:35	Jun-02-2020 12:35			<input type="checkbox"/>	<input type="checkbox"/>
FPM	Jun-02-2020 11:43	Jun-02-2020 11:43			<input type="checkbox"/>	<input type="checkbox"/>
GDPR_Hadoop	Jun-02-2020 10:51	NA			<input type="checkbox"/>	<input type="checkbox"/>
PII_Hadoop	Jun-02-2020 10:51	NA			<input type="checkbox"/>	<input type="checkbox"/>
PCI_Hadoop	Jun-02-2020 10:51	NA			<input type="checkbox"/>	<input type="checkbox"/>

Sensitive Type	Description	Protection Option	Consistent	Report Unique Count
Credit Card				
Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
Credit Card # (Digits Only)	e.g. 5173215750856134	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
Dates				
DOB (Best suited for structured data)	Date Of Birth	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
IP Address				
IP Address	IP Address	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
Social Security				
Social Security # (Digits Only)	e.g. 232883211	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
Social Security # (Space Separation)	e.g. 232 88 3211	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>
Social Security # (Dash Separation)	e.g. 232-88-3211	Intellmask	<input type="checkbox"/>	<input type="checkbox"/>

6.3 List Policies

6.3.1 DBMS

This screen displays the list of pre-defined policies and the Sensitive Type associated with each policy.

Policy Name	Created On	Last Updated	Export Time	Import Time	Show/Hide	Delete
test	Nov-19-2019 14:44	Nov-19-2019 14:44			<input type="checkbox"/>	<input type="checkbox"/>
GDPR_DBMS	Nov-08-2019 16:13	NA	Nov-13-2019 15:56		<input type="checkbox"/>	<input type="checkbox"/>
PII_DBMS	Nov-08-2019 16:13	NA			<input type="checkbox"/>	<input type="checkbox"/>
PCI_DBMS	Nov-08-2019 16:13	NA			<input type="checkbox"/>	<input type="checkbox"/>
HIPAA_DBMS	Nov-08-2019 16:13	NA			<input type="checkbox"/>	<input type="checkbox"/>

Sensitive Type	Description	Masking Option	Consistent	Unique	Persistent	Keep Null	SL	Report Unique and Total
Address								
US Address	US Address	FPM	<input type="checkbox"/>					
Address Line (Best suited for structur...	Address Street and Unit	FPM	<input type="checkbox"/>					
Address State (Best suited for structu...	Address State	FPM	<input type="checkbox"/>					
Address City (Best suited for structur...	Address City	FPM	<input type="checkbox"/>					
Address Zip (Best suited for structure...	Address Zip	FPM	<input type="checkbox"/>					
Address Country (Best suited for stru...	Address Country	FPM	<input type="checkbox"/>					
Credit Card								
Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	FPM	<input type="checkbox"/>					
Names								
Part Names(FirstName or LastName)	Part Names(First Name,Last Name, First/LastNa...	FPM	<input type="checkbox"/>					

1. **Edit** – This functionality allows user to edit any customized policy. Select a policy in Policy panel and click **Edit**.

***Note:** Only Custom Policies are editable.

2. **Copy** – User can create a copy of pre-defined and customized policy. To create a copy of the policy, follow the below steps:
 - a) Select the policy that you want to copy.
 - b) Click **Copy**.
3. Enter the details such as **Policy, Description**. Select the **Sensitive Type** and the **CUPS** option.
4. Click **Save**. This will successfully create a copy of the policy.

The screenshot shows the 'Save Policy' dialog box. At the top, there are two input fields: 'Policy:' with the value 'GDPR_copy' and 'Description:' with the value 'GDPR policy_1'. To the right of these fields is a checkbox labeled 'Include GDPR Dashboard View'. Below the input fields is a table with the following columns: Sensitive Type, Description, Masking Option, Consistent, Unique, Persist, Keep Null, SL, and Report Unique. The table is grouped into sections: Address, AddressID, BusinessEntityID, Credit Card, and Dates. The 'Credit Card' section is expanded, showing three rows with checked checkboxes in the 'Sensitive Type' column. At the bottom of the dialog, there are three buttons: 'Cancel', 'Save As', and 'Save'. The 'Save' button is highlighted with a red box.

5. **Refresh** – Click this button to refresh the current page with the updated information.
6. **Show** – Click this button to unhide the policy.
7. **Hide** – To hide any policy, follow the below steps to hide any policy:
 - a) Click the **Show/Hide** checkbox for the selected policy.
 - b) Click the **Hide** button. The policy will get greyed out.

Policy Name	Created On	Last Updated	Export Time	Import Time	Show/Hide	Delete
CCPA_DBMS_1	Nov-13-2019 14:37	Nov-13-2019 14:37			<input checked="" type="checkbox"/>	<input type="checkbox"/>
NewPolicyName	Sep-11-2019 22:13	Sep-11-2019 22:13	Nov-06-2019 11:49		<input type="checkbox"/>	<input checked="" type="checkbox"/>
CCPA_DBMS	Jul-10-2019 4:27	Jul-10-2019 4:27			<input type="checkbox"/>	<input type="checkbox"/>
GDPR_DBMS	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>
PII_DBMS	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>
PCI_DBMS	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>
HIPAA_DBMS	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>

Sensitive Type	Description	Masking Option	Consistent	Unique	Persistent	Keep Null	SL	Report Unique
Address								
Address Line (Best suited for structured...	Address Street and Unit	Random(Address Line 1)	<input type="checkbox"/>					
Address City (Best suited for structured...	Address City	Random(City)	<input type="checkbox"/>					
Address State (Best suited for structure...	Address State	Random(State Name)	<input type="checkbox"/>					
Address Zip (Best suited for structured ...	Address Zip	Random(ZIP)	<input type="checkbox"/>					
Address Country (Best suited for struct...	Address Country	Random(Country Name)	<input type="checkbox"/>					
Email Address								
Email Address	Email Address	Random(Email Addresses)	<input type="checkbox"/>					
Names								
Full Names	Full Names	Random(First Name)	<input type="checkbox"/>					
Social Security								

- Delete** – Check the **Delete** checkbox corresponding to the policy that you want to delete. Click the **Delete** button to delete the selected policies.

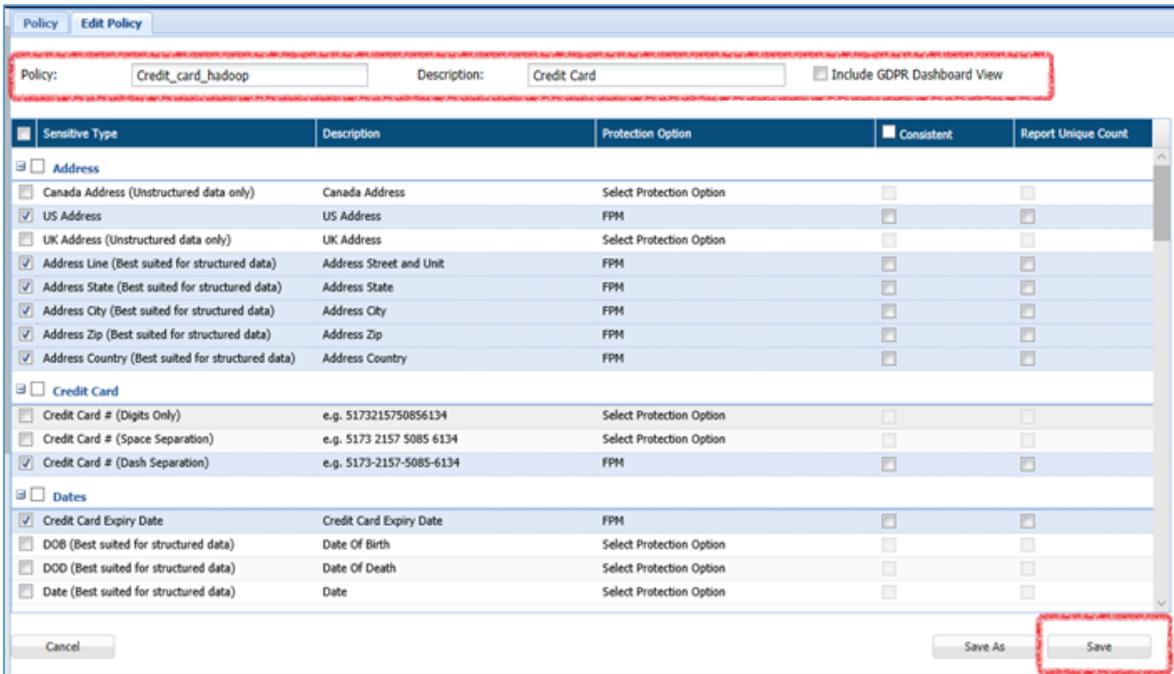
6.3.2 Hadoop & Files

This screen displays the list of pre-defined policies and the Sensitive Type associated with each policy.

Policy Name	Created On	Last Updated	Export Time	Import Time	Show/Hide	Delete
custom_tt	Dec-05-2019 12:29	Dec-05-2019 12:29			<input type="checkbox"/>	<input type="checkbox"/>
custom_t	Dec-04-2019 18:00	Dec-04-2019 18:00			<input type="checkbox"/>	<input type="checkbox"/>
pii_copy	Dec-04-2019 17:01	Dec-04-2019 17:01			<input type="checkbox"/>	<input type="checkbox"/>
policy_with_LUC	May-15-2019 13:56	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
MIX	May-15-2019 13:55	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
Telephone	May-15-2019 13:54	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
SSNO	May-15-2019 13:54	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
NIN	May-15-2019 13:54	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
Names	May-15-2019 13:53	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
EURO	May-15-2019 13:53	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
Dates	May-15-2019 13:53	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>
CCNO	May-15-2019 13:53	Dec-04-2019 17:00	May-15-2019 14:32	Dec-04-2019 17:00	<input type="checkbox"/>	<input type="checkbox"/>

Sensitive Type	Description	Protection Option	Consistent	Report Unique Count
Credit Card				
Credit Card # (Digits Only)	e.g. 5173215750856134	FPM	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Social Security				
Social Security # (Digits Only)	e.g. 232883211	FPM	<input checked="" type="checkbox"/>	<input type="checkbox"/>

1. **Edit** – This functionality allows user to edit any customized policy. Click **Edit** to edit an existing information for the policy.
2. **Copy** – User can create a copy of pre-defined and customized policy. To create a copy of the policy, follow the below steps:
 - a) Select the policy that you want to copy.
 - b) Click **Copy**.
3. Enter the details such as Policy, Description. Select the Sensitive Type and the CUPS option.
4. Click Save. This will successfully create a copy of the policy.



5. **Refresh** – Click on Refresh button to refresh the current page with the updated information.
6. Show – Click the Show button to unhide the policy.
7. Hide – To hide any policy, follow the below steps to hide any policy:
 - c) Click the **Show/Hide** checkbox for the selected policy.
 - d) Click the **Hide** button. The policy will get greyed out.

Policy Name	Created On	Last Updated	Export Time	Import Time	Show/H	Delete
Credit_card_hadoop	Nov-11-2019 16:56	Nov-11-2019 16:56			<input checked="" type="checkbox"/>	<input type="checkbox"/>
VIRs	Jul-27-2019 7:38	Jul-27-2019 7:38			<input type="checkbox"/>	<input type="checkbox"/>
Custom-AES	Jul-22-2019 6:05	Jul-22-2019 6:05			<input type="checkbox"/>	<input type="checkbox"/>
CCPA_Hadoop	Jul-10-2019 4:28	Jul-10-2019 4:28	Jul-11-2019 0:12		<input type="checkbox"/>	<input type="checkbox"/>
GDPR_Hadoop	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>
PII_Hadoop	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>
PCI_Hadoop	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>
HIPAA_Hadoop	Jul-02-2019 8:03	NA			<input type="checkbox"/>	<input type="checkbox"/>

Sensitive Type	Description	Protection Option	Consistent	Report Unique Count
Address				
US Address	US Address	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Address Line (Best suited for structured da...	Address Street and Unit	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Address State (Best suited for structured d...	Address State	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Address City (Best suited for structured data)	Address City	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Address Zip (Best suited for structured data)	Address Zip	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Address Country (Best suited for structured...	Address Country	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Credit Card				
Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	FPM	<input type="checkbox"/>	<input type="checkbox"/>
Dates				
Credit Card Expiry Date	Credit Card Expiry Date	FPM	<input type="checkbox"/>	<input type="checkbox"/>

- Delete** – Check the Delete checkbox corresponding to the policy that you want to delete. Click the **Delete** button to delete the selected policies.

6.4 Export Policy

DgSecure provides the capability to export and import policies. Follow the below steps to export a policy:

Policy Export/Import

Import Options: Error if sensitive type exists Use existing sensitive type Override existing sensitive type

Policy File:

Note: You can import all information related to Policy including reference data in JSON or DGP file format.

DBMS/Hadoop & Files Policies

Policy Name	Type	Created On	Last Updated	Import Time
<input checked="" type="checkbox"/> GDPR_DBMS	DBMS	Nov-08-2019 16:13	NA	
<input type="checkbox"/> GDPR_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PII_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PCI_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> HIPAA_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PII_DBMS	DBMS	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PCI_DBMS	DBMS	Nov-08-2019 16:13	NA	
<input type="checkbox"/> HIPAA_DBMS	DBMS	Nov-08-2019 16:13	NA	

- Click Policy > Export/Import > Policies.

2. Check the checkbox corresponding to the policy that you want to export. Click the **Export Policies** button.
3. Enter the location to save the file in the pop-up. Click the **Save** option and mention the path where file will be downloaded.

***Note:**

1. While exporting a policy, if any custom Sensitive Type is added to the policy which has reference data or Confidence factor config values then these values will also get exported along with the custom Sensitive Type.
2. When a user exports a policy, then Pre-Defined Sensitive Types will not get exported. Only their references will get set in the policy. When policy is getting imported, the Pre-Defined Sensitive Types are immutable and already exists on the system.
3. If any derived sensitive type is part of a policy, then that derived sensitive type with its derivation definition, reference data, confidence factor config will also get exported.
4. If a customer has done some changes in data regex, column name regex or any other property of that Pre-Defined Sensitive Type; DgSecure will not update that Pre-Defined Sensitive Type on the target system while importing a policy on the other system.

In case the customer wants to port the changes performed in the Pre-Defined Sensitive Type from one machine to another, contact support team to get the scripts.

6.5 Import Policy

DgSecure provides the capability to import policies. Follow the below steps to import a policy:

The screenshot shows the 'Policy Export/Import' interface. At the top, there are 'Import Options' with three radio buttons: 'Error if sensitive type exists', 'Use existing sensitive type' (which is selected), and 'Override existing sensitive type'. Below this is a 'Policy File:' field with a text input containing a placeholder path and a 'Browse...' button. A note below the field states: 'Note: You can import all information related to Policy including reference data in JSON or DGP file format.' An 'Import' button is located at the bottom right of this section.

The lower section is titled 'DBMS/Hadoop & Files Policies' and contains a 'Refresh' button and an 'Export Policies' button. Below these is a table with the following data:

Policy Name	Type	Created On	Last Updated	Import Time
<input checked="" type="checkbox"/> GDPR_DBMS	DBMS	Nov-08-2019 16:13	NA	
<input type="checkbox"/> GDPR_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PII_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PCI_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> HIPAA_Hadoop	Hadoop & FILES	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PII_DBMS	DBMS	Nov-08-2019 16:13	NA	
<input type="checkbox"/> PCI_DBMS	DBMS	Nov-08-2019 16:13	NA	
<input type="checkbox"/> HIPAA_DBMS	DBMS	Nov-08-2019 16:13	NA	

1. Click **Policy > Export/Import > Policies**.
2. Check the checkbox corresponding to that policy that you want to import. Click the **Browse** button and select the policy file to import.
3. Select the Import options.

***Note: These import options are applicable only for custom sensitive type.**

Following are the import options:

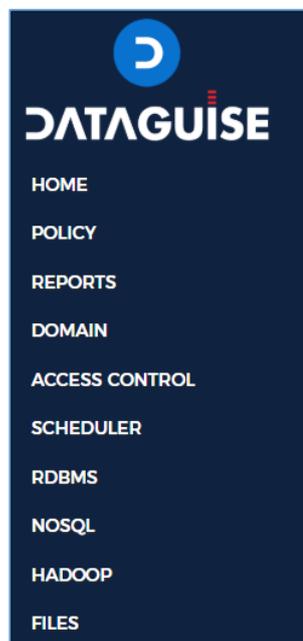
- a) **Error if sensitive type exists:** If sensitive type with the same name already exists, the system will throw an error and will not let you import the policy.
 - b) **Use existing sensitive type:** If the sensitive type with the same name already exists, the system will use the existing sensitive type.
 - c) **Override existing sensitive type:** If the sensitive type with the same name exists, the system will override the properties of the existing sensitive type with the new properties.
4. Click the **Import** button. This will successfully import a policy in the DgSecure.

7 Task

7.1 Concept

DgSecure offers unique solutions to determine and safeguard sensitive information on all types of source systems, relational and non-relational databases, with structured and unstructured data, stored on the premises and even in the cloud. Detection and protection of sensitive data in

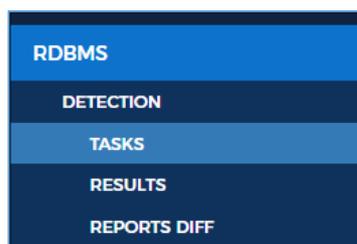
DgSecure is achieved by executing the tasks on the target source systems. This section discusses the steps to create and execute the tasks on each source system. Every licensed source system is available on the main menu of DgSecure, as depicted below:



***Note:** A task can only be executed if the required IDP (Refer to DgSecure Admin Guide section 6.2.1. **Create IDPs**) has been setup and is in “Active” condition, and appropriate [connection](#) has been set to the target source system.

To access the tasks screen for a source system, click **<Source System> -> <Task Type> -> TASKS**

For example, to access detection tasks under RDBMS, click RDBMS->DETECTION->TASKS (depicted in the image below).



7.1.1 Source System

The target data store, e.g. RDBMS, HADOOP, NOSQL etc. are termed as source systems.

7.1.2 Task Type

There are two major types of tasks:

- a) **Detection:** Detection tasks help in obtaining information about where and what type of sensitive data is stored in a source system. To detect sensitive information, it is crucial to define which data entries are to be considered as sensitive elements, e.g. credit card numbers, social security numbers, addresses etc. These sensitive data elements are categorized as **Sensitive Types**.

There are various out-of-the-box sensitive types for the user to select from as well as options to clone or create custom sensitive types. Different sensitive types can be grouped together under a **Policy** to refine the sensitive data detection. Please refer to the sections [Policy](#) and [Sensitive Type](#) for more details.

- b) **Masking:** Masking tasks are used to protect the sensitive data that has been detected using a detection task in the target source system. Users can perform masking and encryption on their data to ensure that sensitive information is not exposed.

DgSecure provides several masking options for securing sensitive data on an organization’s data repositories. The table below shows the masking options available for different data types.

Masking Option Name	Characters (Char)	Variable Characters (Varchar)	Number	Clob
Static mask	X	X	X	X
Character mask	X	X		
Format Preservation Mask (FPM)	X	X	X	
IntelliMask mask	X	X		
Random mask	X	X	X	
NPI mask	X	X	X	
Compose mask	X	X		
Compose Math Expression mask			X	
Date Synch mask		X		
Name Synch mask		X		
Email Policy mask	X	X		
Expression mask			X	
Full Name mask		X		
Regular Expression mask	X	X		
Custom Lookup mask	X	X	X	
Custom mask	X	X	X	

Shuffle mask	X	X	X	
JSON mask		X		X
XML mask		X		X
AES Encryption/Decryption	X	X		
FPE Encryption/Decryption	X	X	X	

For detailed information on each masking option, refer to [7.4 Masking Options](#).

***Note:** Options for task types differ based on the type of source system.

7.2 Create Task

This section will explain the process of creating and editing a task in different source systems.

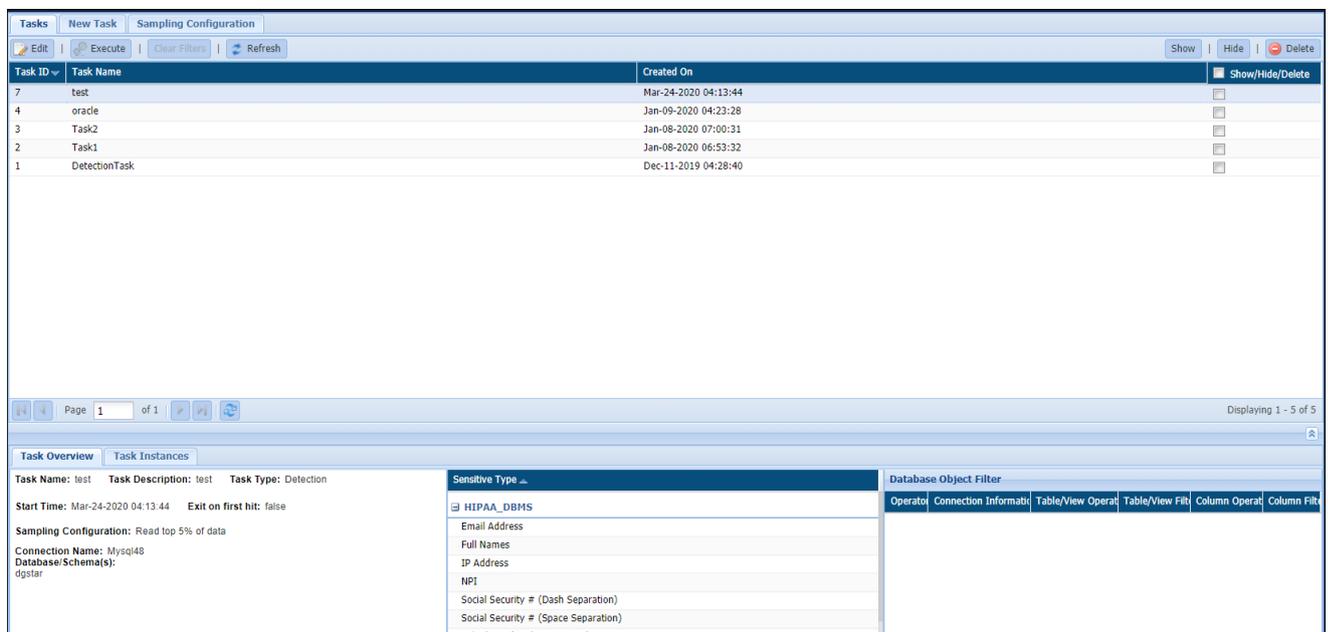
7.2.1 RDBMS

DgSecure supports Detection and Masking in RDBMS databases. Following sections outline the process of creating these tasks.

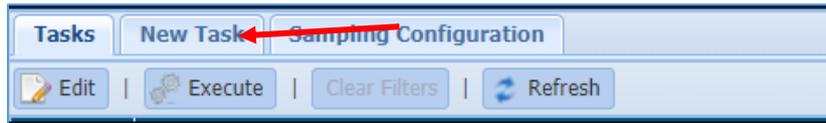
7.2.1.1 Detection Task

Perform the following steps to create a detection task.

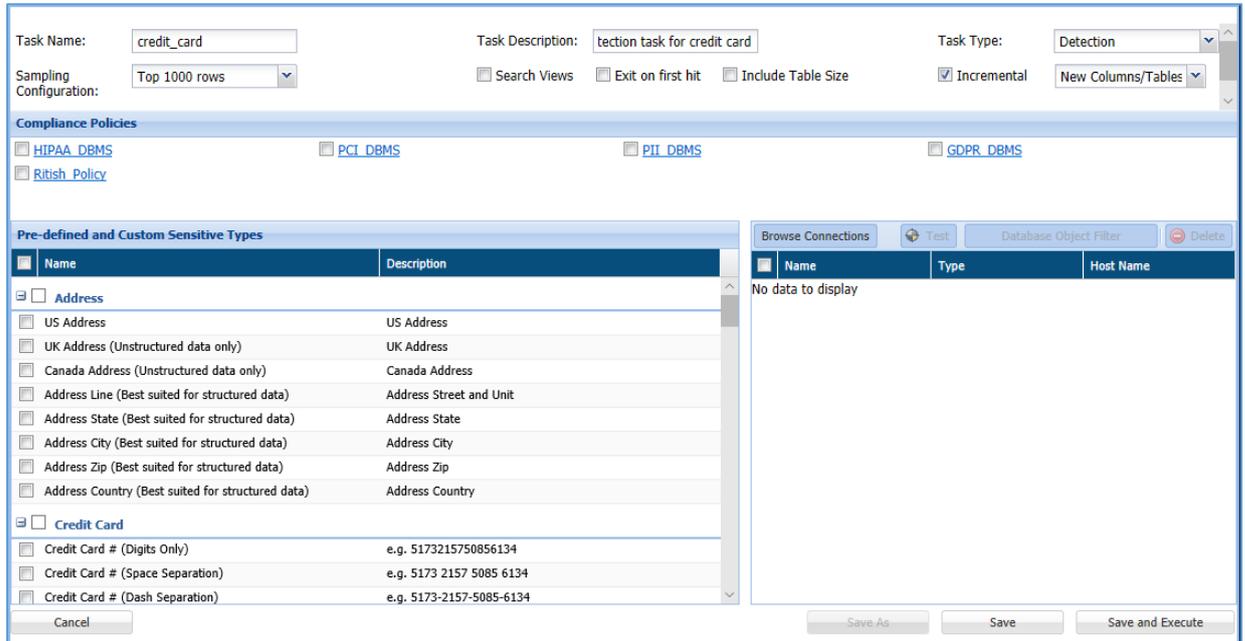
Go to **RDBMS > DETECTION > TASKS**



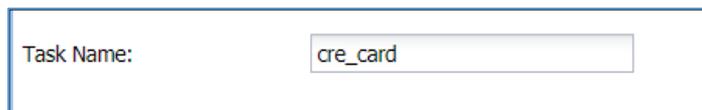
1. To create a new task, click on the **New Task** tab.



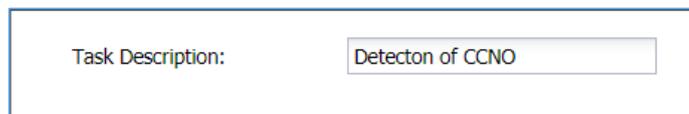
The following image shows the user interface for creating a task.



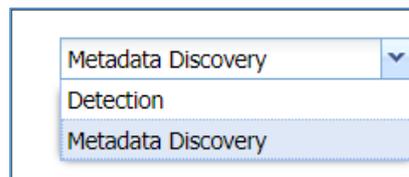
2. Enter a unique **Task Name**. This field supports numeric and character values.



3. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.



4. Select the Task Type: **Detection** or **Metadata Discovery**.

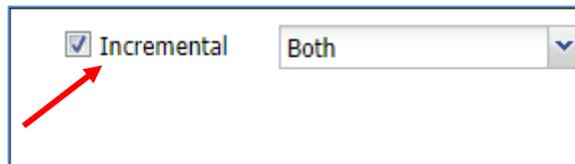


Metadata Discovery scans your database to provide information about the type of data available on the database.

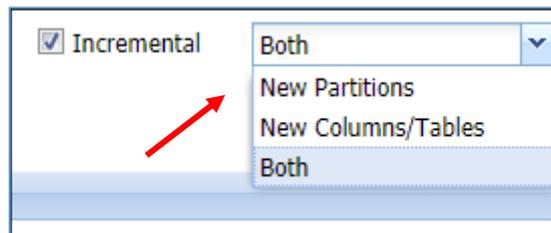
***Note:** Support for Metadata Discovery is now available for Oracle.

5. Check the **Search Views** option to detect the Views tables within the database as well as the tables linked with them. This option is available for the **Detection, Task Type**.
6. Check **Exit on First Hit** option to stop the scanning, when the first Sensitive Type is detected during the scan. This option is available for the **Detection, Task Type**.
7. The **Incremental** option is used to execute sensitive data detection only on the new entries to the database. This option significantly decreases the time taken to scan the database. This option is available for the **Detection, Task Type**. To setup Incremental detection, perform the following steps:

- a) Check the Incremental checkbox.



- b) Select the type of increment to the database that has to be considered for scanning, i.e., addition of new partitions or columns/tables to the database.



8. DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:
 - Top 1000 Rows
 - Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. Check the **Advanced** checkbox to create a new sampling configuration.

You can also configure sampling through **RDBMS > DETECTION > TASKS > SAMPLING CONFIGURATION** tab.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Table Row Count Range: * To: Type: *

By: * Value: *

Table Row Count Range	Sample Type	Sample Value	Sample By	Actions
Default	Top	1000	Rows	
1 to 200	Bottom	100	Rows	

Note:

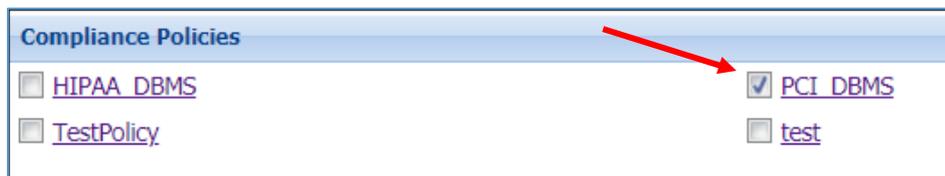
1. Top sampling option is not supported for Teradata, Aster DB and Sybase ASE connections.
2. Bottom sampling option is not supported for Teradata, Aster DB, Snowflake, Informix DB, Splice Machine, Sybase IQ and Sybase ASE connections.
3. Random sampling option is not supported for Snowflake, Informix DB, MySQL and Splice Machine connections. Random sampling is also not supported for Views in Sybase IQ.
4. For performance reasons, except for tables in DB2 schemas, random and bottom sampling is not recommended, where the tables can contain large volume of data.
5. If some unsupported sampling option is chosen, then best applicable sampling option will be applied. For more details regarding chosen sampling option, IDP logs can be checked.

- i. Enter the name of the **Sampling Configuration**.
- ii. Enter the description for sampling.
- iii. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.
- iv. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling. Below are the options for advanced settings:
 - **Table row count range:** Enter numeric value. This value states the starting range of the table from which the records will be sampled.
 - **To:** Enter the numeric value. The value in this field states the ending range of the table till which the records will be sampled.
 - **Type:** Select the sampling configuration type from the **Type** option. There are four options for sampling configuration:
 - i. **Top:** If you select the option **Top**, the sample data for the scan will be selected from the entries at the top of the table, based on the specified range.
 - ii. **Bottom:** If you select the option **Bottom**, the sample data for the scan will be selected from the entries at the bottom of the table. This does not mean the entries will be selected bottom up, instead

depending on the range the last entries in the table will be taken to create a sample data for detection.

- iii. **Random:** Entries from the table that correspond to the specified range, will be selected at random to create a sample set of data for detection.
 - iv. **Complete:** If all the entries in the selected tables of the database have to be scanned for sensitive types, select this option.
 - **By:** To Specify how to pick data for sampling from the table, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
 - **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.
 - v. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.
 - vi. Click the **Save** button to save the changes.
9. Select the required policy under the **Compliance Policies** section. The Compliance Policy panel displays all the Pre-Defined and Customized Policies. Users can select any number of policies while creating or editing a task.

Sensitive types associated with the selected policy can be viewed in the panel below this panel **Pre-Defined and Custom Sensitive Types**. Selecting a policy is not a mandatory step, users can also proceed to select individual sensitive types. For more information, refer to section [Policy](#).



10. Select the required sensitive types for the scan from the **Pre-defined and Custom Sensitive Types** section. Refer to the following screenshot:

Pre-defined and Custom Sensitive Types	
Name	Description
<input type="checkbox"/> ABA_test	
<input type="checkbox"/> ABA_test	ABA
<input type="checkbox"/> Address	
<input type="checkbox"/> US Address	US Address
<input type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
<input type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input type="checkbox"/> Address State (Best suited for structured data)	Address State
<input type="checkbox"/> Address City (Best suited for structured data)	Address City
<input type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input type="checkbox"/> Address Country (Best suited for structured data)	Address Country
<input checked="" type="checkbox"/> Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
<input type="checkbox"/> Dates	

The **Pre-Defined and Custom Sensitive Types** panel lists down all the Sensitive Types. The Sensitive Type associated with the policy gets selected in the Pre-Defined and Custom Sensitive panel and cannot be removed from the scan, however any number of sensitive types can be added to the scan.

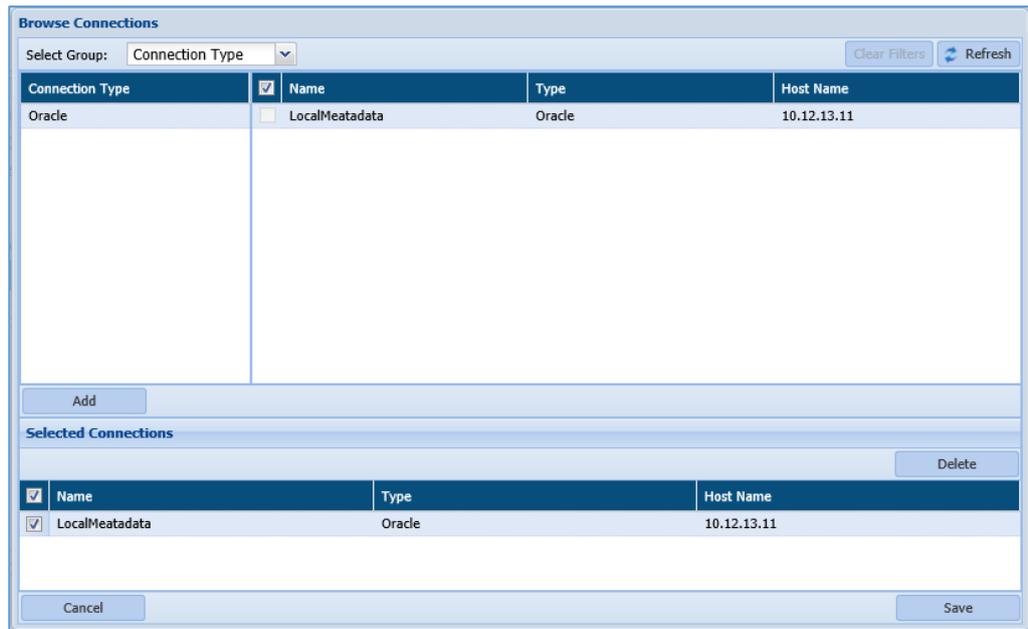
***NOTE:** If a policy is selected, user can still add more sensitive types to the scan but the sensitive types under the selected policies cannot be excluded from the scan.

11. The **Database Connection** panel lists down all the available RDBMS connections. Any number of connections can be selected for a task. This panel list down all the available connections. For details on how to create and manage connections refer [Connection Manager](#). Perform the following steps to choose a connection:

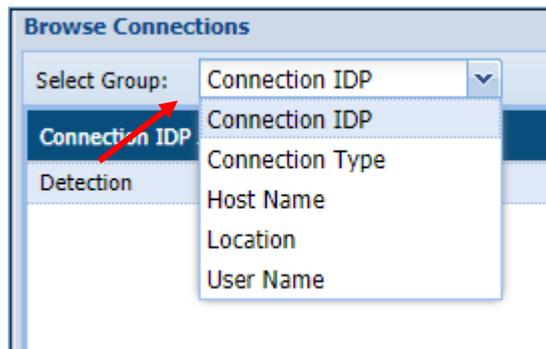
vii. Click Browse Connections.

Browse Connections		
Name	Type	Host Name

viii. The **Browse Connections** dialog box will be displayed. This screen categorizes connections based on user preferences.



- ix. Click on the **Select Group** dropdown and select the required option from the sub groups displayed on the left panel to sort the available connections.

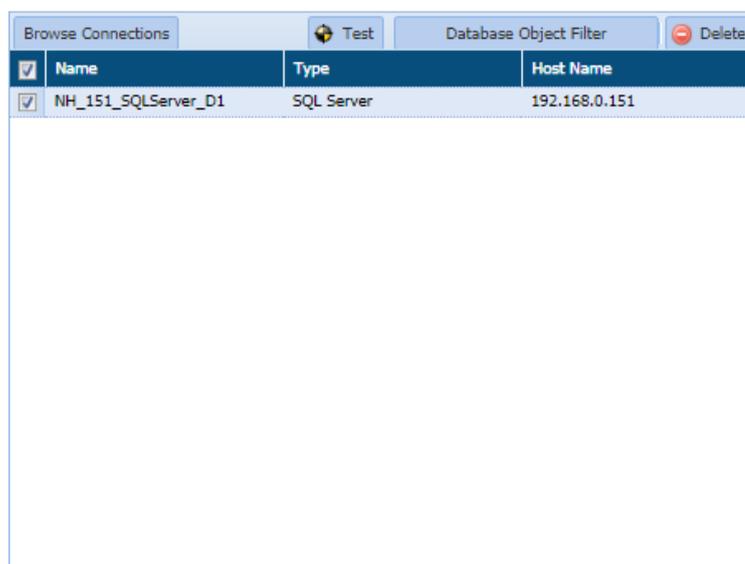


The **Select Group** drop-down has five options:

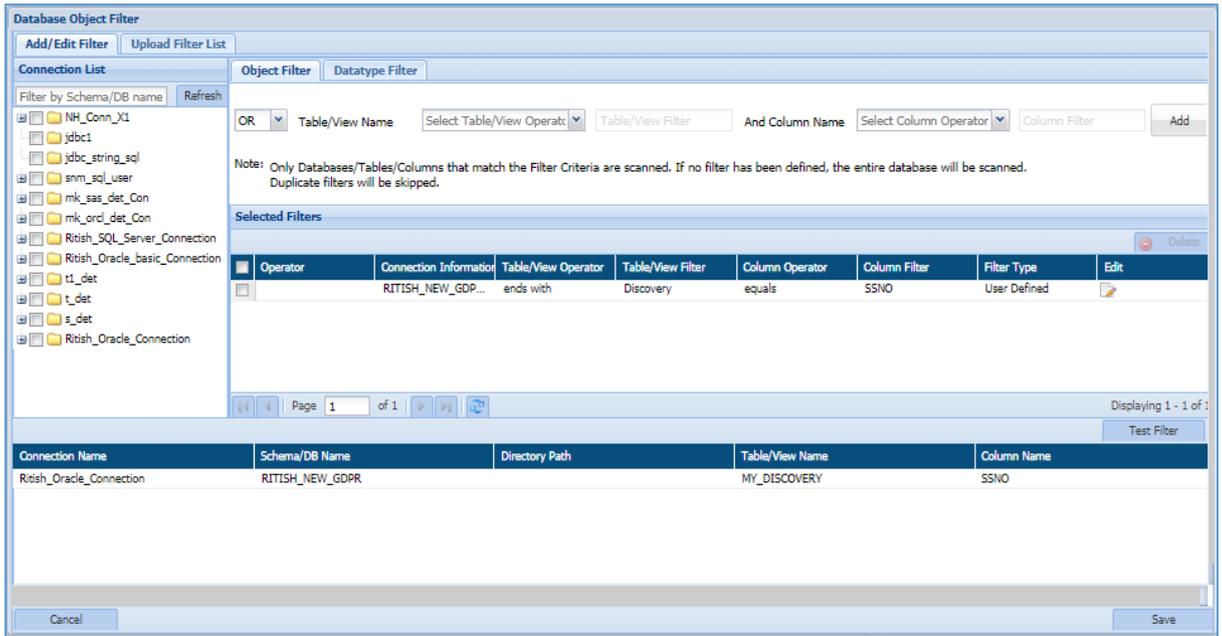
1. **Connection IDP:** Categorizes the available connections based on the types of IDPs available, i.e., Detection and Masking.
2. **Connection Type:** Categorizes the available connections based on the type of server connected to, i.e., Oracle, Teradata, SQL server etc.
3. **Host Name:** Categorizes the list of available connections based on Host Names.
4. **Location:** Categorizes the available connections based on the location of the target source system server, i.e., On-Premises and Cloud.

5. **User Name:** Categorizes the list of available connections based on the Usernames.
- x. Click **Add** to include the selected database connection in the **Selected Connection** panel.
- xi. Check the Selected connection. Click the **Save** button to include connections.
- xii. Click the **Test** button to test the listed RDBMS connection.
- xiii. Click the **Database Object Filter** button to filter tables and/or columns. Once filters are defined, then only those databases/tables/columns that match the filter are scanned.

Check the checkbox next to the connection to enable the Database Object Filter.



You can **Add/Edit Filter** in two ways i.e. either by specifying in manually it manually or by uploading the filter list in the **Upload Filter List** tab.



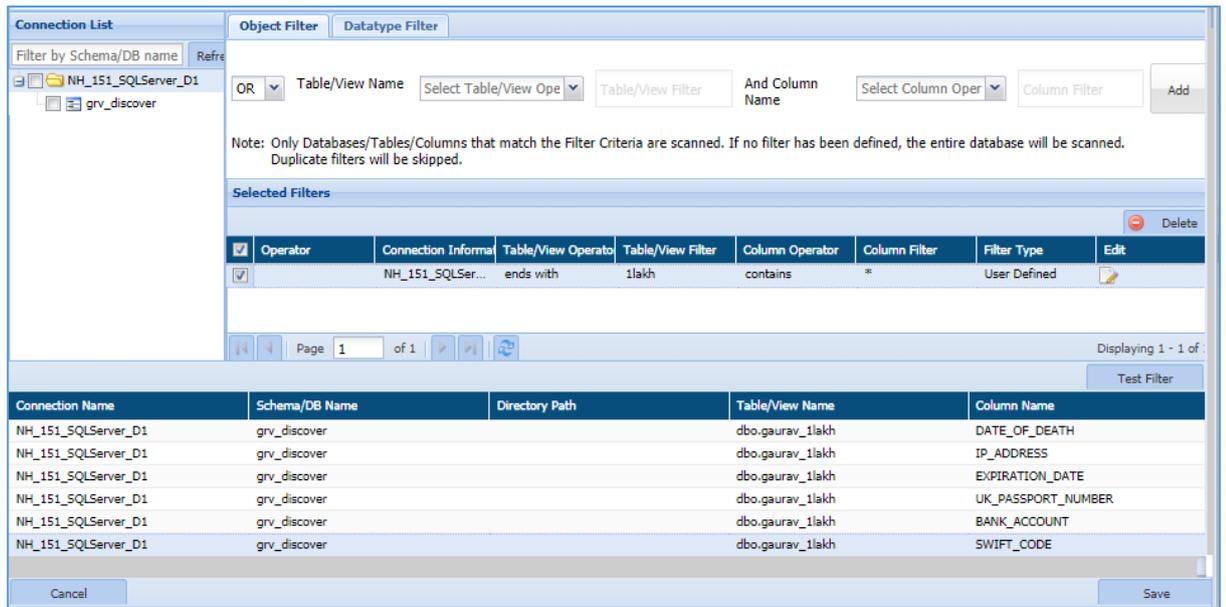
- **Add/Edit Filter:** This tab allows you to apply the filter for the selected connection. Once the filter has been applied then only those databases/tables/columns and datatypes that matched the criteria will be scanned.

There are two types of filter which can be applied.

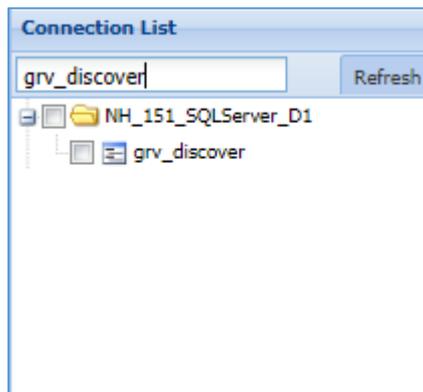
1. Object Filter
2. Datatype Filter

1. Object Filter

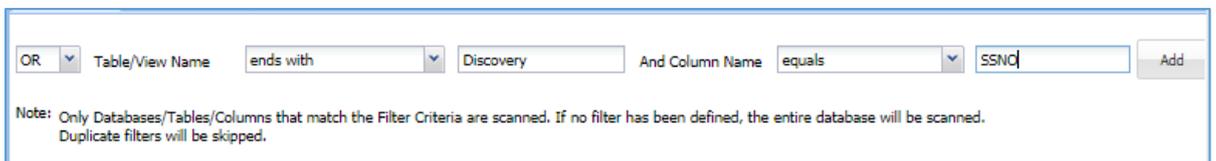
The Object Filter allows you to scan the database/tables/columns based on the defined filter. To apply an Object Filter, perform the following steps:



- a) Select the connection from the Connection List panel or enter the DB/Schema name in the Filter by Schema/DB Name textbox.



- b) Apply the **Object Filter** in the top panel by specifying the **Operator, Table/View** and **Column** name.



For example, in the above image the object filter specifies the table name should end with 'Discovery' and the selected table should contain SSNO column.

There are eight types of Operators based on which you can select the Table and Column name.

1. **Equals:** This operator will check whether the given table/column name exist in the selected database/table. It will return the matched records if the condition is fulfilled.
 2. **Not Equal to:** This operator will return all the records except the given table/column name.
 3. **Contains:** This operator will return only those tables/columns which matched the given criteria.
 4. **Does not contain:** The functionality of this operator is similar to the **Not Equal to** operator, since it returns all the records except the given table/column name.
 5. **Starts with:** This operator will return all the tables/column whose name starts with the given criteria.
 6. **Does not start with:** The functionality of this operator is similar to the **Does not contain** and **Not Equal to**, since it will return all the tables/column name except the one which has been entered.
 7. **Ends with:** This operator will return all the tables/column name whose name ends with the given input.
 8. **Does not ends with:** The functionality of this operator is similar to the **Does not contain** and **Not Equal to**, since it will return all the tables/column name except the one which has been entered.
- c) Click the **Add** button to add the filter in the **Selected Filters** panel.
- d) The **Selected Filters** panel will list down all the user defined filters. It displays information about the filters such as Connection Name, Table/View Operator, Table/View Filter, Column name, etc.

Operator	Connection Information	Table/View Operator	Table/View Filter	Column Operator	Column Filter	Filter Type	Edit
<input type="checkbox"/>	RITISH_NEW_GDP...	ends with	Discovery	equals	SSNO	User Defined	

- To edit a filter, click the button in the **Edit** column. This functionality allows you to re-select the Operator, Table/View and Column name.

- To delete a filter, check the checkbox for that filter and click the **Delete** button. This button will get enabled when you check the checkbox for any filter.

9. Click **Test** button to test the filter. It lists down the result matching the filter criteria.

Connection Name	Schema/DB Name	Directory Path	Table/View Name	Column Name
Ritish_Oracle_Connection	RITISH_NEW_GDPR		MY_DISCOVERY	SSNO

10. Click **Save** button to make the changes effective.

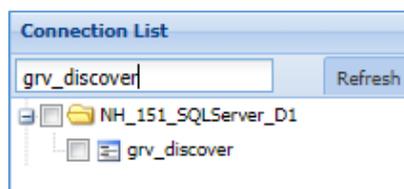
11. Click **Cancel** button, if you do not want to save the changes.

1. Datatype Filter

This tab allows you to specify the datatypes to be included or excluded while scanning based on the selection.

To apply Datatype Filter, perform the following steps:

- Select the connection from the **Connection List** panel or enter the Database/Schema name in the **Filter by Schema/DB Name** textbox.



- b) Select the **Filter Type** from the given option. By default, **Include** filter type is selected.
- **Include:** This option allows to select the datatype that need to be included while scanning.
 - **Exclude:** This option allows to select the datatype that need to be excluded while scanning.
 - **Additional Datatype:** This field allows you to add a datatype if it's not present in the DataType panel. Once that datatype is entered, click **Add** button to specify it in the DataType panel.

- c) Check the checkbox for the datatype which you want to include or exclude based on the selection of **Filter Type** in above panel.

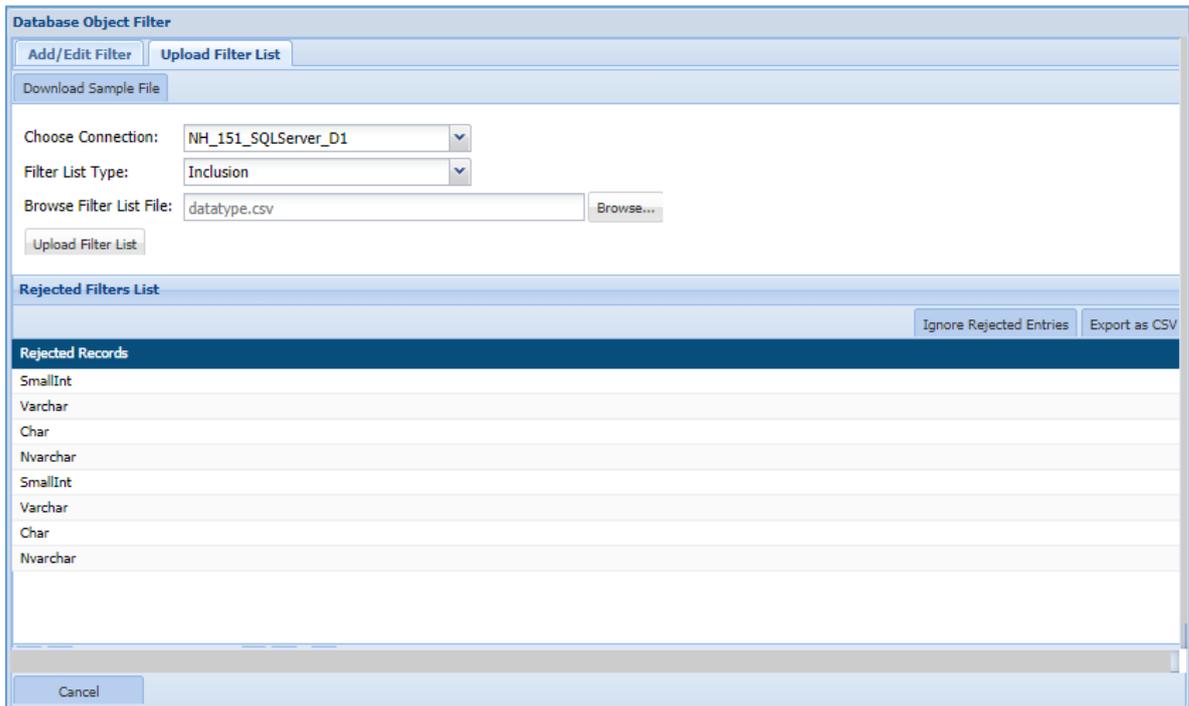
For example, in the above image the CHAR and VARCHAR datatypes are included in the scanning process since Filter Type specify the option as 'Include'. If 'Exclude' is selected as Filter Type then selected CHAR and VARCHAR datatype will be excluded from the scanning process.

- d) Click **Test** button to test the whether any column in the database contains the selected datatype. This functionality will list down all the columns that contain the selected datatype.

Connection Name	Schema/DB Name	Directory Path	Table/View Name	Column Name	Test Filter
Ritish_Oracle_Connection	RITISH_NEW_GDPR		MY_DISCOVERY	SSNO	

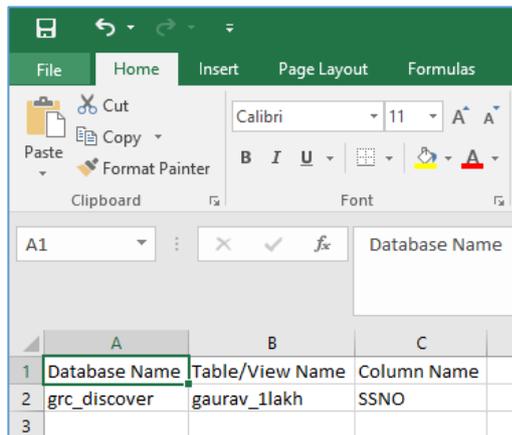
- e) Click **Save** button to make the changes effective.
 f) Click **Cancel** button if you do not want to save the changes.

- **Upload Filter List:** This tab allows you to upload a file containing the list of all columns.



To upload the filter list, perform the following steps.

- Select the connection from the **Choose Connection** drop-down.
- The **Download Sample File** button will be enabled once a connection has been chosen. Enter the Database, Table/View and Column Name in the sample file.



- Select the **Filter List Type** from either 'Inclusion' or 'Exclusion'. This functionality allows you to specify whether to include or exclude the Database, Table/View and Column name.

Filter List Type: Exclusion

Inclusion

Exclusion

- d) Click the **Browse** button to search and upload the saved sample file containing the list of Database, Table/View and Column name which will be either excluded or included.

Browse Filter List File: FiltersSampleFileForSQLServer (2).csv Browse...

- e) Click **Upload Filter List** button to add the defined filters in the **Selected Filters** panel under **Add/Edit Filter** tab.

Selected Filters							Delete
Operator	Connection Informa	Table/View Operato	Table/View Filter	Column Operator	Column Filter	Filter Type	Edit
<input type="checkbox"/>	NH_151_SQLSer...	equals	gaurav_ilakh	equals	SSNO	Include	

Displaying 1 - 1 of

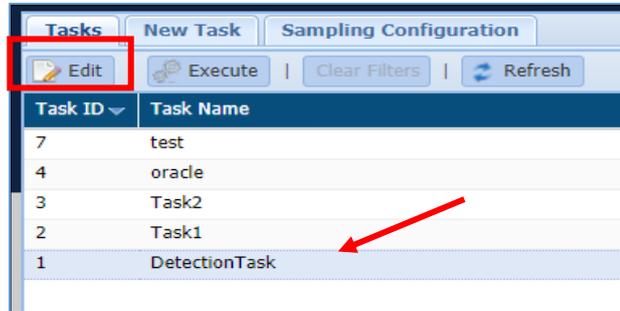
- f) The **Rejected Filters List** will display all the entries from the uploaded list which are not in a proper format as specified in the sample file.

Rejected Filters List		Ignore Rejected Entries	Export as CSV
Rejected Records			
grc_discover,,SSNO			
srch_ccno.emp_d,			
discover_city,Emp_detail,City			

- To download the list of rejected entries, click **Export as CSV** button. A downloaded file will contain the list of rejected entries which were not formatted as per specified format.
 - To remove the rejected entries from the **Rejected Filters List** panel, click the **Ignore Rejected Entries**. This functionality will remove all the rejected entries from the panel.
- g) Click **Save** button to make the changes effective.
- h) Click **Cancel** button if you do not want to save the changes.

9. Click **Save**, if you want to execute the task later else click **Save and Execute**. The results of the task and its status can be viewed under **RDBMS>DETECTION>RESULTS** (Refer to [Result](#) section).

10. To edit an existing task, select the required task from the list of tasks on the Tasks screen and click edit. A task can be edited using the same steps for task creation.



***Note:** Task defining features such as Incremental, Exit on First Hit, Search View and the Task Type cannot be edited. Some of the originally selected options can be modified but new options cannot be added.

7.2.1.2 Masking Task

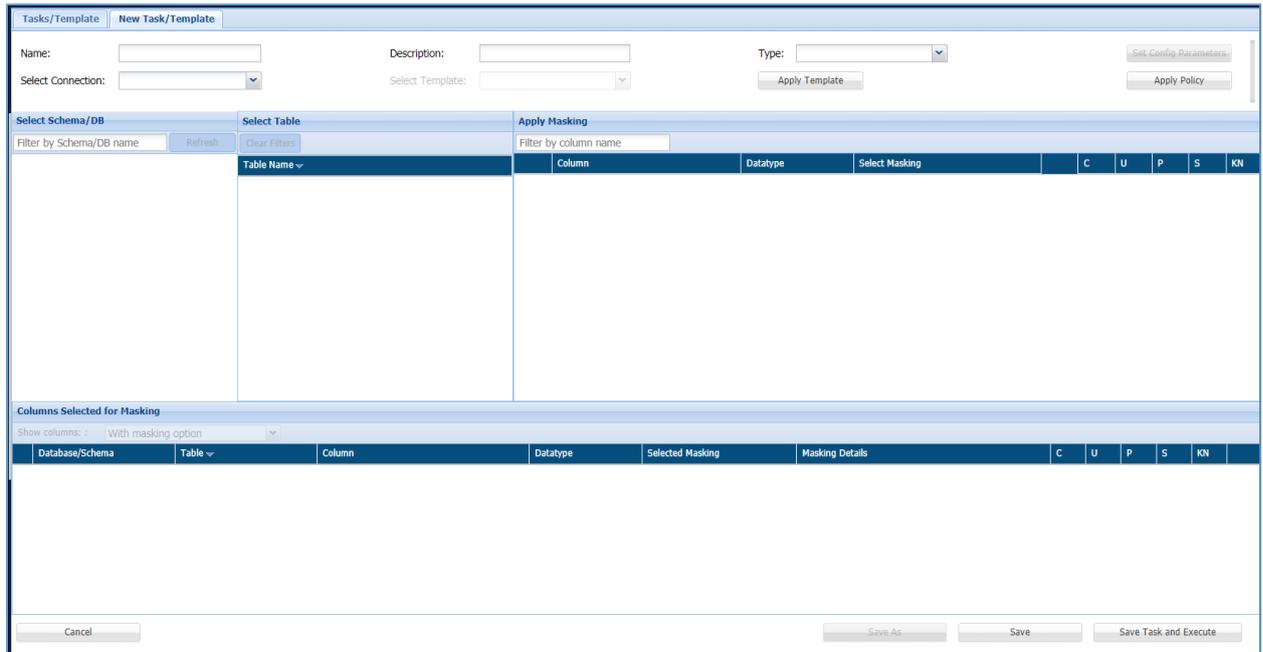
Perform the following steps to create a masking task:

Access the TASKS/TEMPLATE screen, click **RDBMS > MASKING > TASKS/TEMPLATES**

1. To create a new task or template for masking tasks, click on the **New Task/Template** tab.



The following image shows the user interface for creating a task.



2. Enter a unique **Task Name**. This field supports numeric and character values.

Name:

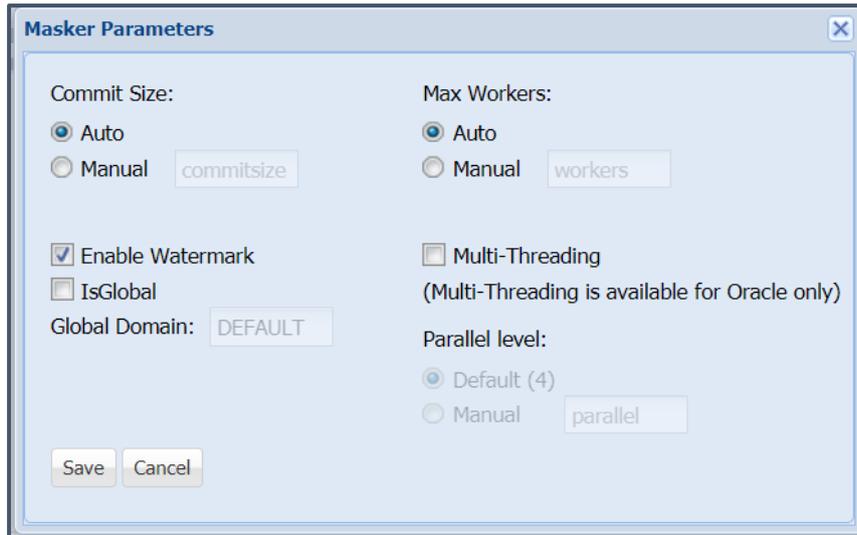
3. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.

Description:

4. Click on the **Type** dropdown. Select **Task** to create a masking task or **Template** to create a template for masking tasks.

Type:

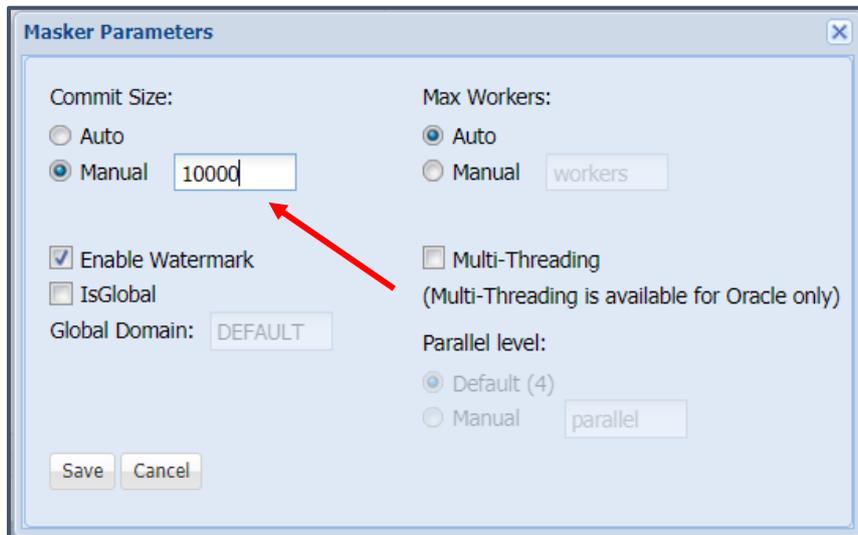
5. To setup additional settings to improve masking abilities click the **Set Config Parameters** button. The following popup, **Masker Parameters** will appear.



The following environment settings for the masking operations can be configured:

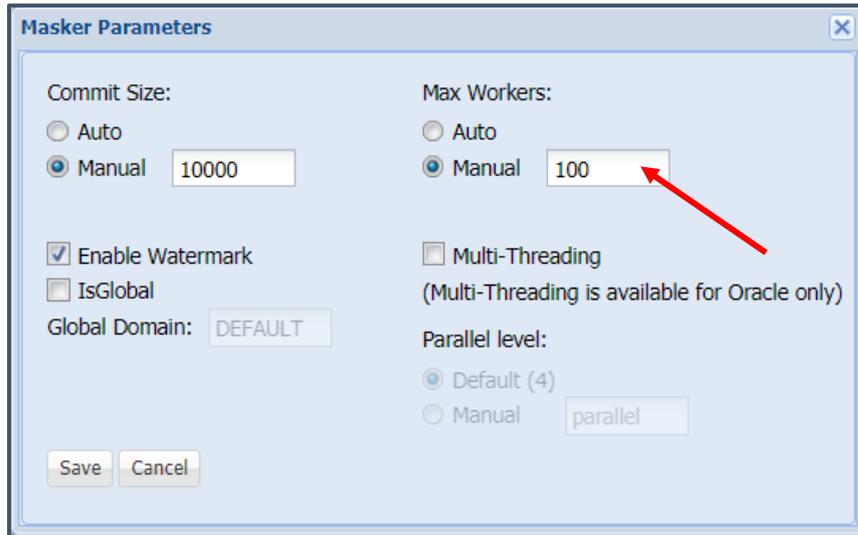
7.2.1.2.1 Commit Size

DgSecure executes masking on large files in batches for better performance and quicker turnaround. Select the option **Manual**, under **Commit Size** and enter the number of rows to be masked within each batch, or select **Auto** to continue with default batch size. For example, in the following screenshot DgSecure has been configured to mask the data in batches of 10000 rows each.



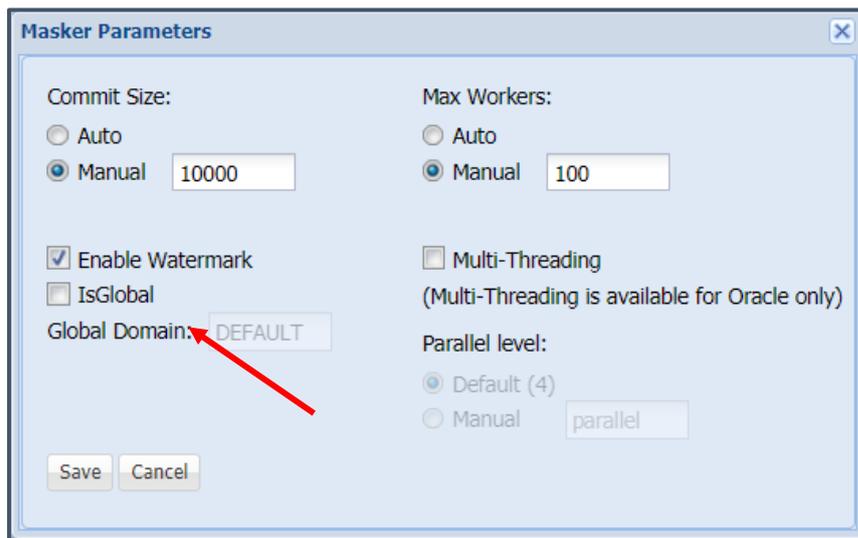
7.2.1.2.2 Max Workers

Select the option **Manual** under the heading **Max Workers**, and enter the number of tables to be masked simultaneously. To continue with DgSecure’s default setting for defining the maximum tables to be masked at one go, select the option **Auto**. For example, in the following screenshot DgSecure has been configured to mask the data in batches of 100 tables each. This feature is particularly useful when there are multiple small tables on the database.



7.2.1.2.3 Enable watermark

The feature to enable watermark for masked values provides a report of masked tables after detection. To setup this feature, simply check the **Enable Watermark** option and **Save** the configuration.



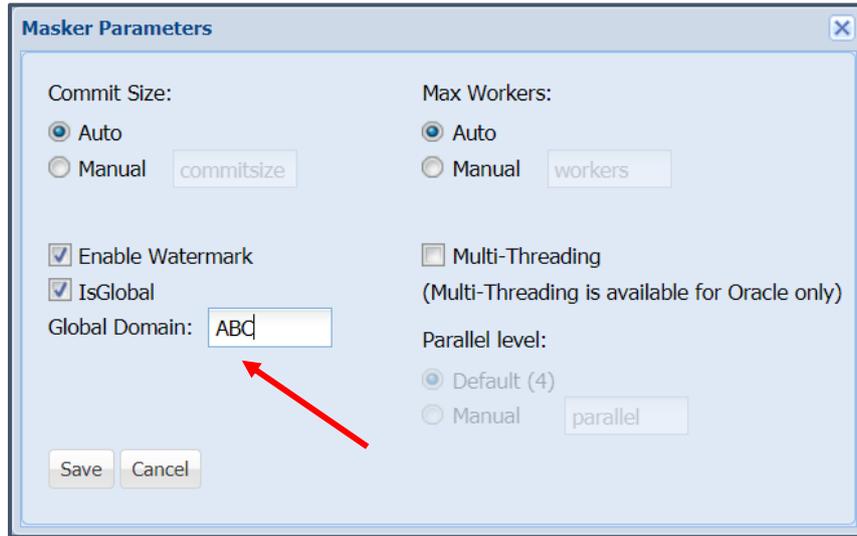
Executing a masking task with the Enable Watermark option checked, will ensure that all the masked tables be marked with **Y** under the column **Masked** in the **Detailed Results** tab in **Detection Results**.

Results																			
Detailed Results																			
Skipped Columns																			
Logs																			
Data Scanned																			
Non-Sensitive Scanned Tables																			
Clear Filters Save Filter																			
Remediate Upload Safelist Edit Safelist Save Results																			
Sensiti	Hostna	Databa	Directo	Table	Column	Data T	Data L	Nullabl	Key Pa	Databa	Confid	Row S	Values	Hit Cou	Null Co	Quick S	Masked	Safe	
Add...	192...	DA...	NA	AN...	LAS...	VA...	50	Y	NA	mk	36	26	26	11	0	N	Y	N	
Add...	192...	DA...	NA	AN...	FIR...	VA...	50	Y	NA	mk	36	26	26	11	0	N	Y	N	
Add...	192...	DA...	NA	AN...	FIR...	VA...	50	Y	NA	mk	21	12	12	3	0	N	Y	N	
Add...	192...	DA...	NA	AN...	LAS...	VA...	50	Y	NA	mk	35	12	12	5	0	N	Y	N	
Add...	192...	DA...	NA	AN...	FIR...	VA...	50	Y	NA	mk	20	17	17	4	0	N	Y	N	
Add...	192...	DA...	NA	AN...	LAS...	VA...	50	Y	NA	mk	25	17	17	5	0	N	Y	N	

7.2.1.2.4 IsGlobal

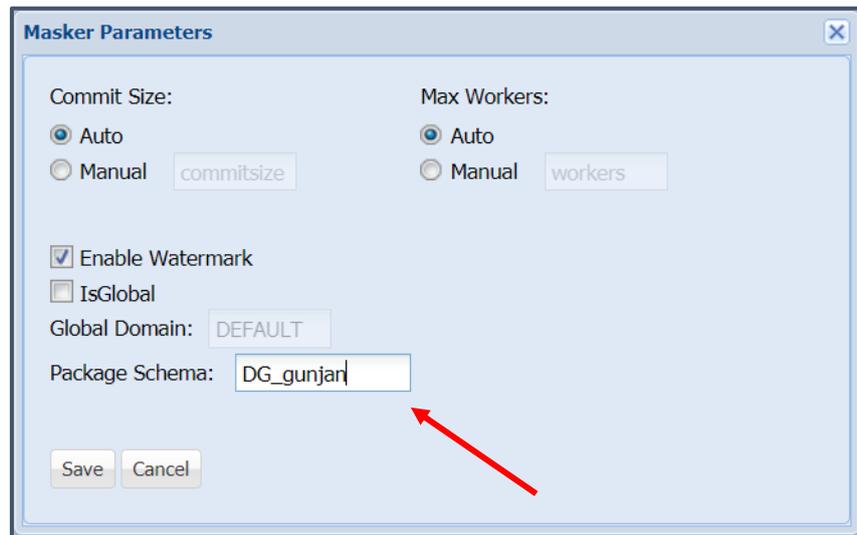
The feature **IsGlobal** is used to consistently mask data across different databases using **SL masking**. Check the **IsGlobal** option, enter the key value to be used to mask the data and **Save**. All masking tasks where this unique key is set will mask the data consistently when SL masking is executed. SL masking is an algorithm based masking, thus, using a unique key for masking across different databases ensures consistent results.

This feature is not supported in PostgreSQL and MySQL.

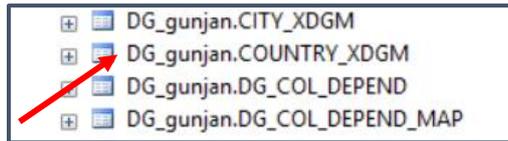


7.2.1.2.5 Package Schema

The option Package Schema is used to name the DgSecure masking objects created by the Makser IDP on the target database. Enter the value for naming the schemas in the **Package Schema** field and **Save** the configuration.



For example, the following screenshot displays how all the new schemas have been named using the provided **Package Schema** string.

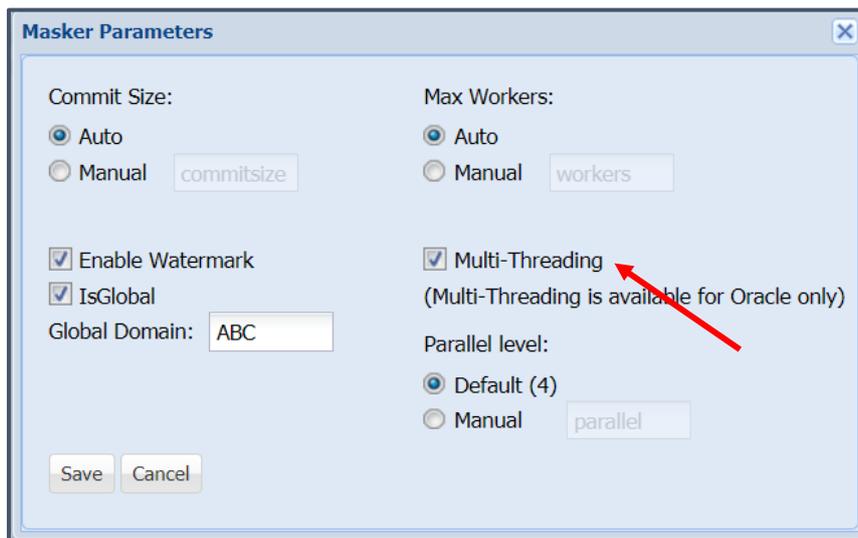


This feature is not supported in Oracle.

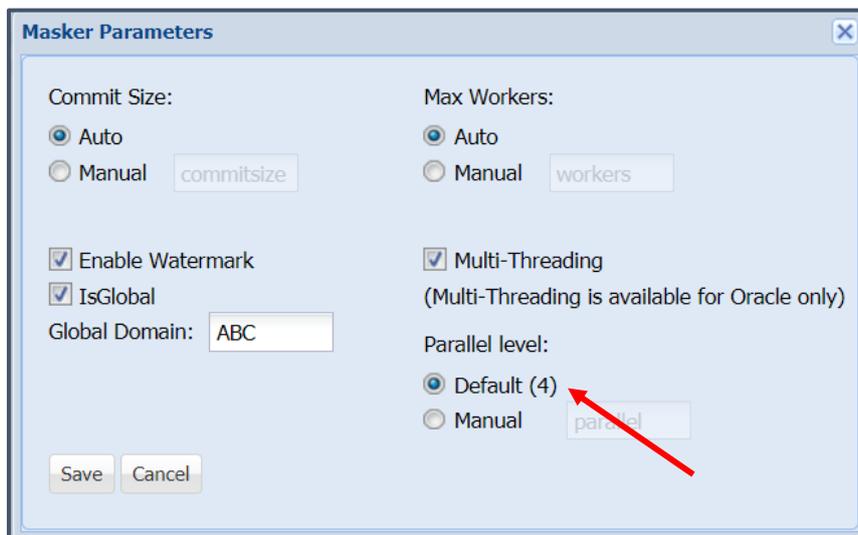
7.2.1.2.6 Multithreading

Multithreading can be applied in Oracle to reduce the time taken to execute masking tasks and improve performance. Perform the following steps to setup multithreading.

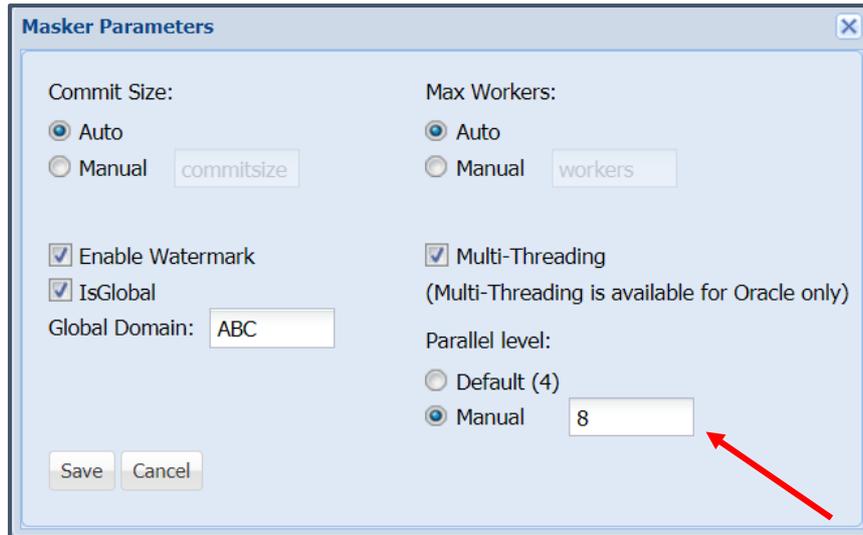
- a) Check the option multithreading.



- b) Provide the number of parallel levels of the table to be masked simultaneously. The default Parallel Level for multithreading is 4.



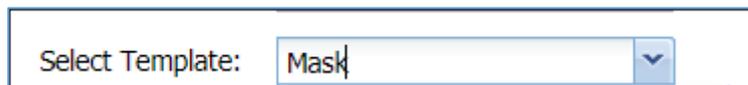
- c) Depending upon the controller machine configurations this number can be adjusted under the option Manual.



NOTE: Ensure that the necessary grants have been provided to the Masker User on oracle to use Multithreading. The script to provide these grants is available at the following location:

[Installed
Directory] \Dataguisse\DgSecure\Agents\DgMaskerAgent\expandedArchive\
WEB-INF\classes\PrerequisiteFiles\Scripts\Oracle\create_user

6. Click **Select Connection** to view the list of the available database connections for masking. For details on how to create a connection for masking refer to [Connection Manager](#).
7. The **Select Template** option provides a list of user created templates for masking. This option will be greyed out if no templates have been created. To create a new template select the Type as Template in step 6 and follow the next steps for creating a task and save the template. Once created, the new template will appear in the select template dropdown on the new task screen.

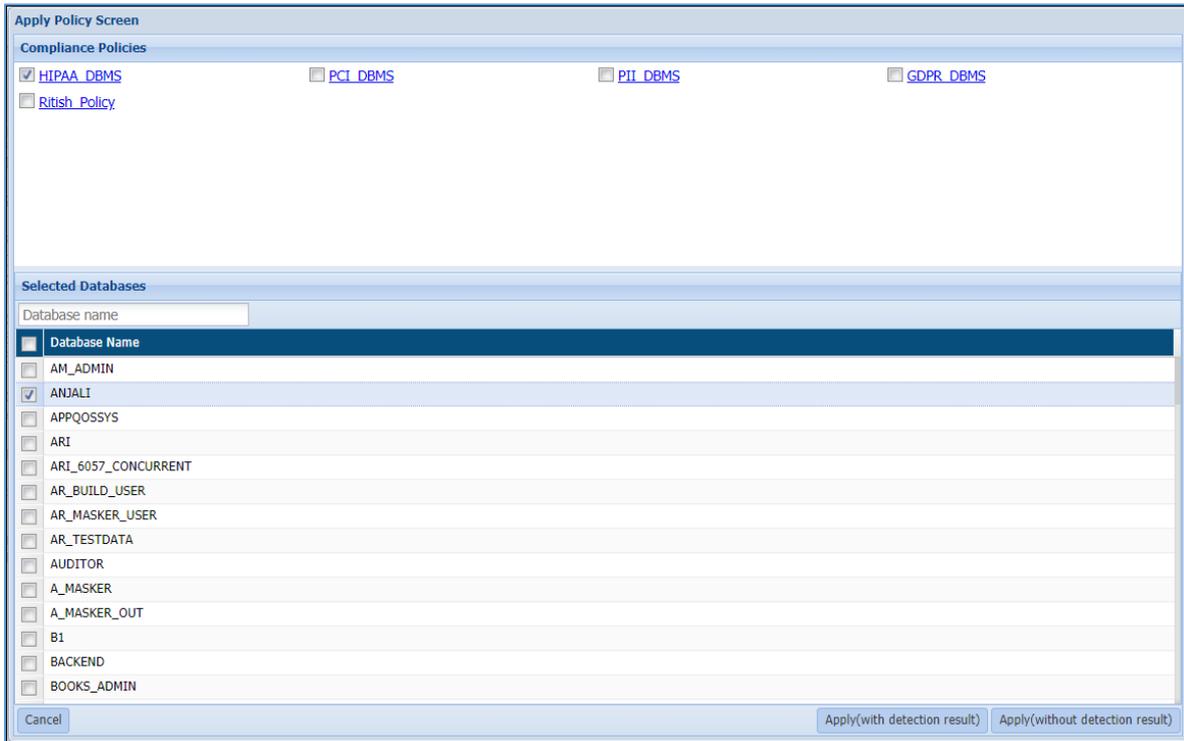


***Note:** Templates are connections specific and only one template can be selected at a time.

8. Click the **Apply Template** button to apply the selected template in **Select Template** dropdown.



9. Click the **Apply Policy** button to choose the available Compliance Policies.

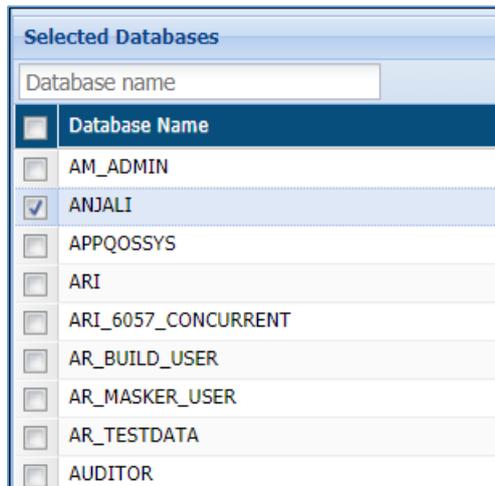


Perform the following steps to apply a policy to the task or template:

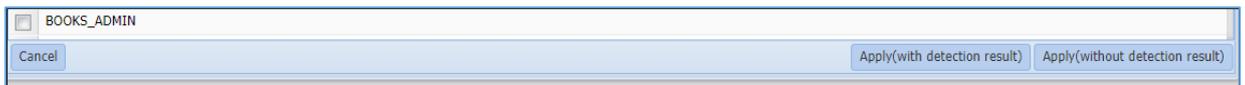
- a) Select the required policy/policies under the **Compliance Policies** panel.



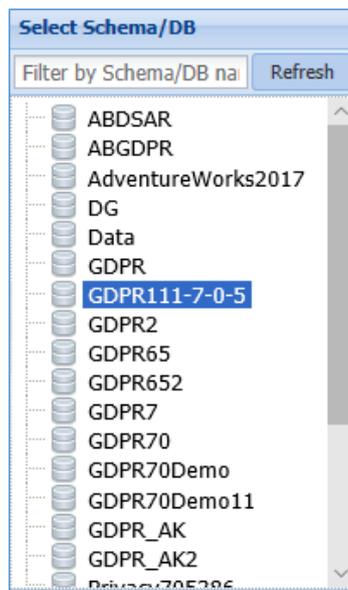
- b) Select the databases you need to mask by checking the checkbox next to the Database Name.



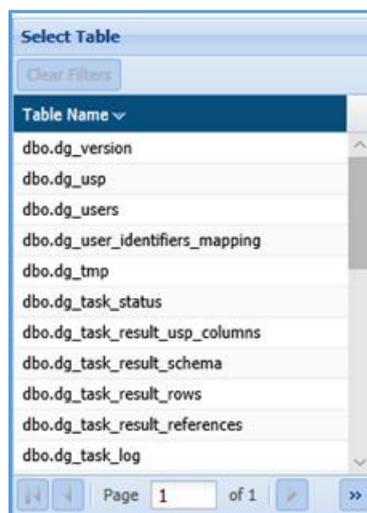
- c) Click the **Apply** (with detection result) button to apply the selected policy with detection results or click the Apply (without detection result) button to apply the selected policy without results. Click Cancel to redo your selection.



- 10. The **Select Schema/DB** panel will display the list of databases or schemas for a selected connection in **Select Connection** drop-down.

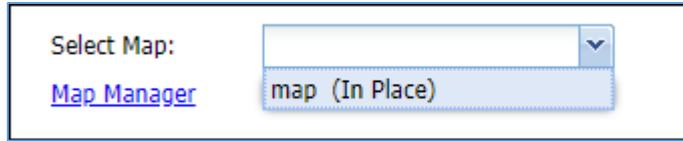


- 11. Select the table from the **Select Table** panel. This panel lists all the tables for the selected database or schema.

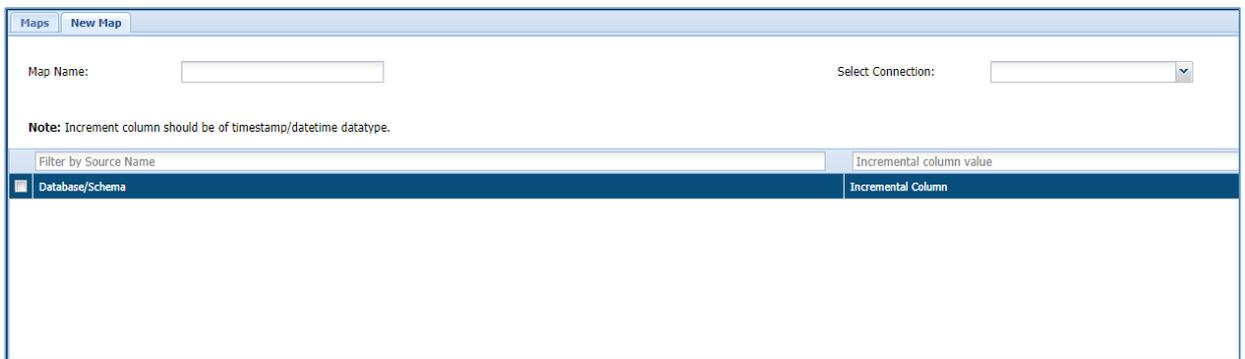


- 12. Check the **Incremental** checkbox to apply incremental masking to the database. This feature is useful to mask new values added in a database after masking has been executed on it. Only the new entries will be masked, thus, the time taken for masking would be reduced. Perform the following steps to make your masking task incremental:

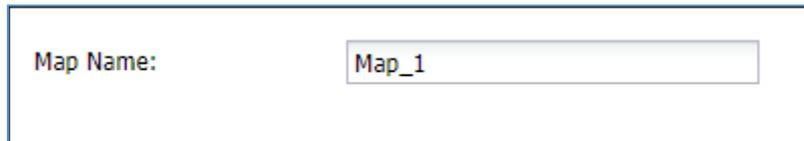
- i. Check **Incremental** checkbox.
- ii. The **Select Map** dropdown will appear. Select the required map. Maps define incremental columns within a database which are considered for indexing the data in order to mask the new rows added to the database.



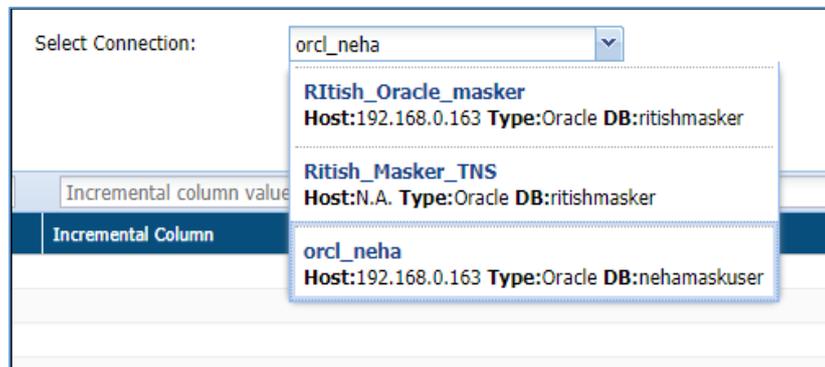
- iii. To add a map, click on the **Map Manager** link.
- iv. The **Map Manager** screen can also be accessed from **RDBMS->Masking->MAP MANAGER** menu bar. The following image displays the **Map Manager** interface:



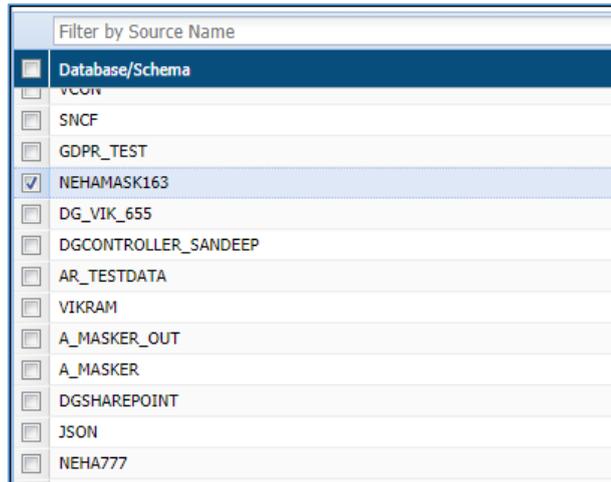
- v. Enter a **Map Name**.



- vi. Select a connection that contains the database on which incremental masking has to be applied.



- vii. Select the required Database/Schema.



- viii. Enter the name of the incremental Column to the corresponding database and ensure that it is a timestamp or date-time datatype.



- ix. Save the map. It will be available in the **Select Map** dropdown on the new task/template screen.

NOTE: To apply incremental masking the following criteria must be met:

- Create a map to execute incremental masking and apply it to the task.
- Relational tables cannot be included.
- Column holding date, time stamp or a numeric incremental value should be added as incremental column
- The incremental column cannot be masked.

13. The **Apply Masking** panel display the list of all the columns for the selected table in Select Table pane. For detailed information on all the available masking options in DgSecure refer section [7.4 Masking Options](#). Perform the following steps to apply masking option to the columns in the database:

- a) Select the masking option from the drop-down against the column entry. You can apply the masking to the selected column by checking on the checkbox corresponding to the column name.

Column	Datatype	Select Masking		C	U	P	S	KN	SL
checkSum	bigint	No Masking	▼	Edit	<input type="checkbox"/>				
connectionid	int	No Masking	▼	Edit	<input type="checkbox"/>				
constraintCols	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
dbName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
groupId	int	No masking available		Edit	<input type="checkbox"/>				
keyType	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
ResultTableId	int	No masking available		Edit	<input type="checkbox"/>				
schemaName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
tableName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				

To further enhance the results of masking following options can be combined with the masking options:

- i. **Consistent:** Masks the data of the selected table with consistent values. E.g. If the name John is masked as FVGB, throughout the table it will be masked the same way.
- ii. **Unique:** Masks each entry with a unique value.
- iii. **Persistent:** Similar to Consistent, however in this case the same values will be masked consistently across all the tables of the database.
- iv. **Sync:** This option allows tracking of masked entries in different tables if any of the same entries are found in other tables they will also be masked.
- v. **Keep Null:** The cells containing null values are kept null even after masking.
- vi. **Stateless:** Persistently masks the data without creating any metadata. No extra space is required to create masking tables.

Column	Datatype	Select Masking		C	U	P	S	KN	SL
checkSum	bigint	No Masking	▼	Edit	<input type="checkbox"/>				
connectionid	int	No Masking	▼	Edit	<input type="checkbox"/>				
constraintCols	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
dbName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
groupId	int	No masking available		Edit	<input type="checkbox"/>				
keyType	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
ResultTableId	int	No masking available		Edit	<input type="checkbox"/>				
schemaName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
tableName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				

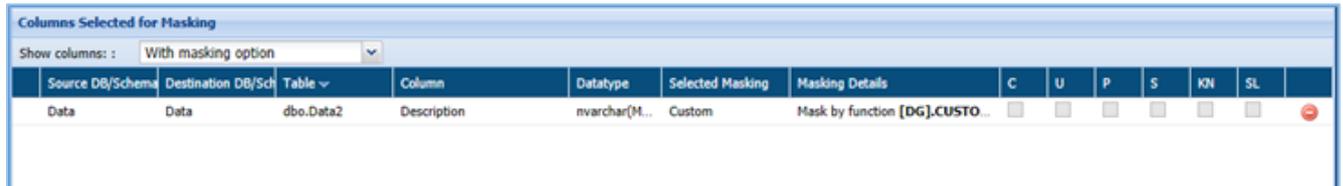
***NOTE:**

Stateless Masking cannot be performed on the following databases:

- Postgres
- MySQL
- Snowflake
- Maria DB

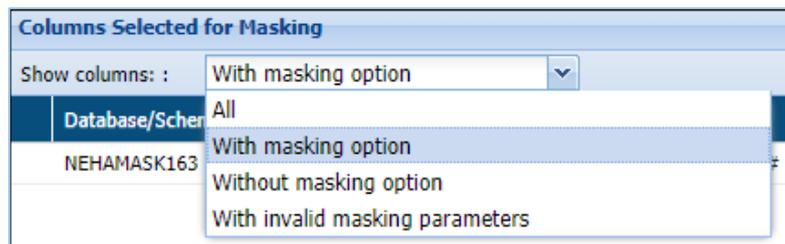
Stateless support on redshift is limited to FPM and Names Masking.

14. View the selection under the **Columns Selected for Masking** panel.



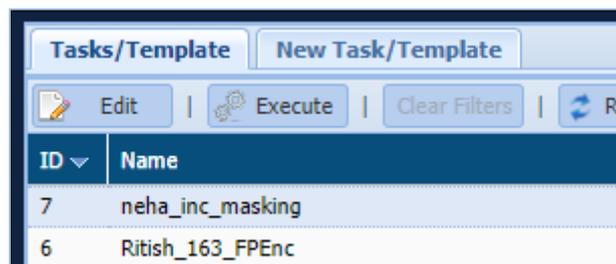
15. Select the option from the **Show Columns** drop-down. There are four options:

- b) All
- c) With masking option
- d) Without masking option
- e) With invalid masking parameters



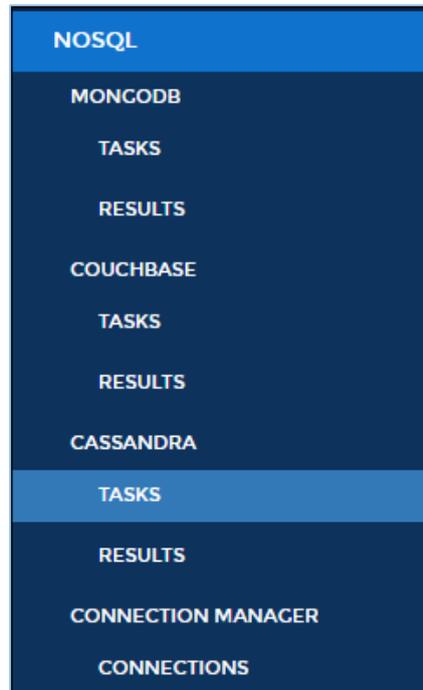
16. Click **Save** to save the task or template or save and execute to execute the masking task.

17. To edit a task, select the task from the **Tasks/Template** tab and click edit. A task can be edited using the same steps for creating a new task.



7.2.2 NoSQL

DgSecure Supports Detection on the NoSQL databases: MongoDB, Couchbase and Cassandra. Each of these databases is available under the NoSQL in the menu.

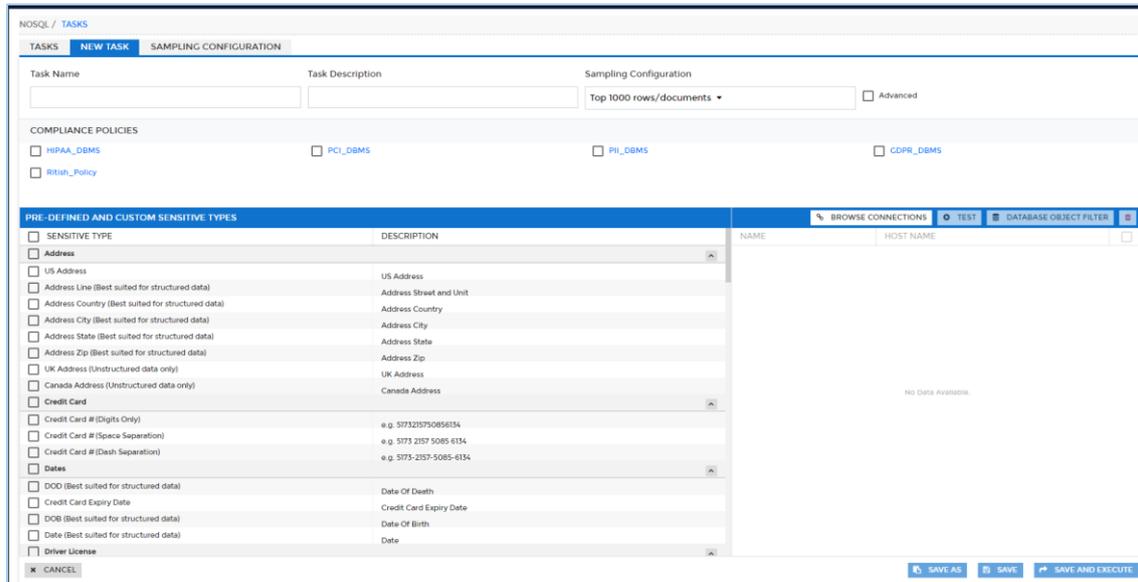


To create task for a NoSQL database, go to

NOSQL > <NOSQL Database > TASKS

The steps for creating tasks for detection on NoSQL databases perform the following steps, these steps are common for all the databases:

1. Go to **NOSQL > <NOSQL Database > TASKS > New Task** tab.



2. Enter a unique **Task Name**. This field supports numeric and character values.

Task Name

cre_card

3. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.

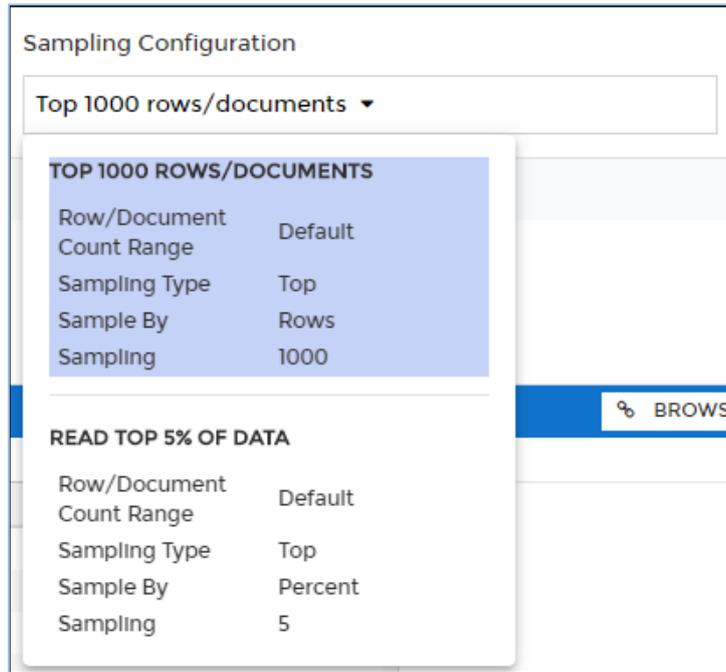
Task Description

detect_ccnq

4. DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

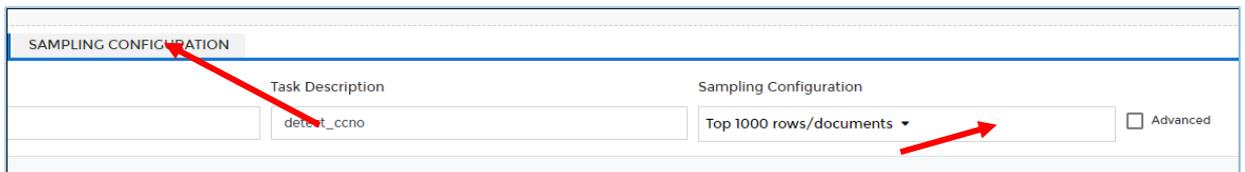
- Top 1000 Rows
- Read top 5% of data

Select the Sampling Configuration from the dropdown.



Perform the following steps to create a new sampling configuration:

- i. Go to **NOSQL -> TASKS -> SAMPLING CONFIGURATION** tab or check the Advanced checkbox.



- ii. Enter the name of the Sampling Configuration.

Name

- iii. Enter the description.

Description

- iv. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.

Set Sampling Config As Default

Show Advance Sampling Details

- v. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling.

Set Sampling Config As Default

Show Advance Sampling Details

Below are the options for advanced settings:

Row/Document Count Range	To	By *	Value *	Type
1	100	Rows/Documents	100	Top
				+ ADD

- **Table row count range:** Enter starting range for the database.
- **To:** Enter the ending range for the database.
- **Type:** Select the sampling configuration type from the **Type** option. Only sampling from the top of the database is available for NoSQL. The option **Top**, creates a sample data for the scan from the entries at the top of the database, based on the specified range.
- **By:** To Specify how to pick data for sampling from the database, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
- **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.

- vi. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.
- vii. Click the **Save** button to save the changes.

5. The **Compliance Policy** panel displays all the Pre-Defined and Customized Policies. Users can select any number of policies while creating or editing a task. Sensitive types associated with the selected policy can be viewed in the panel below this panel **Pre-Defined and Custom Sensitive Types**. Selecting a policy is not a mandatory step, users can also proceed to select individual sensitive types. For more information, refer to [Policy](#) section.

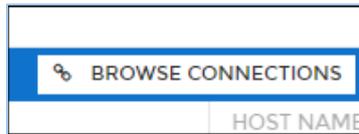


6. The **Pre-Defined and Custom Sensitive Types** panel lists down all the Sensitive Types. The Sensitive Type associated with the policy gets selected in the Pre-Defined and Custom Sensitive panel and cannot be removed from the scan, however any number of sensitive types can be added to the scan.

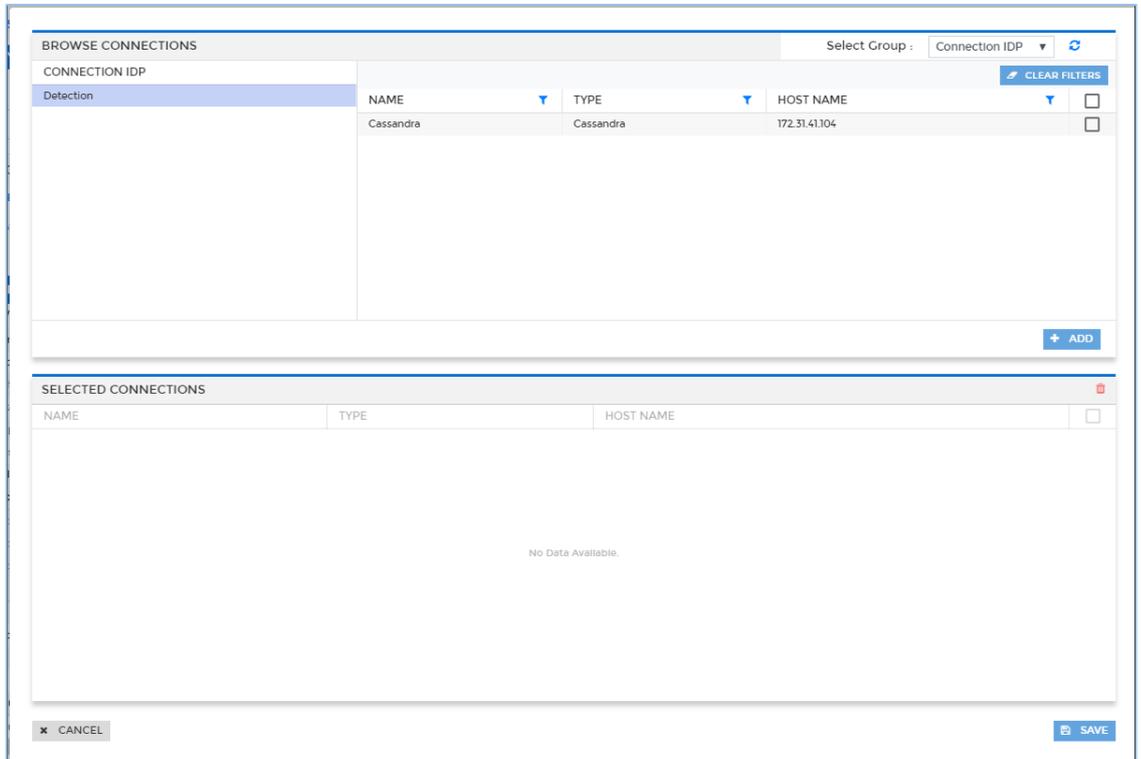
PRE-DEFINED AND CUSTOM SENSITIVE TYPES	
SENSITIVE TYPE	DESCRIPTION
<input type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input type="checkbox"/> Address Country (Best suited for structured data)	Address Country
<input type="checkbox"/> Address City (Best suited for structured data)	Address City
<input type="checkbox"/> Address State (Best suited for structured data)	Address State
<input type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
<input checked="" type="checkbox"/> Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
<input type="checkbox"/> Dates	
<input type="checkbox"/> DOD (Best suited for structured data)	Date Of Death
<input type="checkbox"/> Credit Card Expiry Date	Credit Card Expiry Date
<input type="checkbox"/> DOB (Best suited for structured data)	Date Of Birth
<input type="checkbox"/> Date (Best suited for structured data)	Date
<input type="checkbox"/> Driver License	
<input type="checkbox"/> Driver License (Miscellaneous)	

7. The right hand panel lists down all the available NoSQL connections. Any number of connections can be selected for a task. This panel list down all the available connections. For details on how to create and manage connections refer Connection Manager. Perform the following steps to choose a connection:

- a) Click **Browse Connections**.



- b) The **Browse Connections** dialog box will be displayed. This screen categorizes connections based on user preferences.



- i. Click on the Select Group dropdown and select the required option from the sub groups displayed on the left panel to sort the available connections.

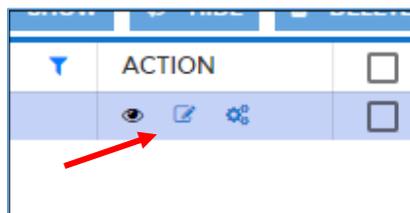


The **Select Group** drop-down has five options:

1. **Connection IDP:** Categorizes the available connections based on the types of IDPs available, i.e., Detection.
2. **Connection Type:** Categorizes the available connections based on the type of server connected to, i.e., MongoDB, Couchbase and Cassandra.
3. **Host Name:** Categorizes the list of available connections based on Host Names.

4. **Location:** Categorizes the available connections based on the location of the target source system server, i.e., On-Premises and Cloud.
5. **User Name:** Categorizes the list of available connections based on the Usernames.
 - ii. Check the checkbox next to the connection name and click **Add** to include the selected database connection in the **Selected Connection** panel.
 - iii. Click the **Save** button to include connections in **Connection** panel.
8. Click the **Test** button to test the listed NoSQL connection.
9. Click the **Database Object Filter** button to filter tables and/or columns. Once filters are defined, then only those databases/tables/columns that match the filter are scanned.
10. Click **Save** to save the task or to execute the task click save and execute.
11. To edit an existing task, select the required task from the list of tasks on the **Tasks** screen and click the edit icon . A task can be edited using the same steps for task creation.

NOSQL / TASKS						
TASKS	NEW TASK	SAMPLING CONFIGURATION	CLEAR FILTER	SHOW	HIDE	DELETE
TASKID ↓	TASK NAME	CREATED ON	ACTION			<input type="checkbox"/>
2	All_inone_karman	Jun-02-2020 13:02:14	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	Task_T_karman	Jun-02-2020 12:51:45	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



7.2.3 Hadoop

DgSecure supports Detection and Protection on different Hadoop distribution systems like HDFS, Hive and HBASE. Various protection options are available for these distributions based on the type of data-structure or unstructured.

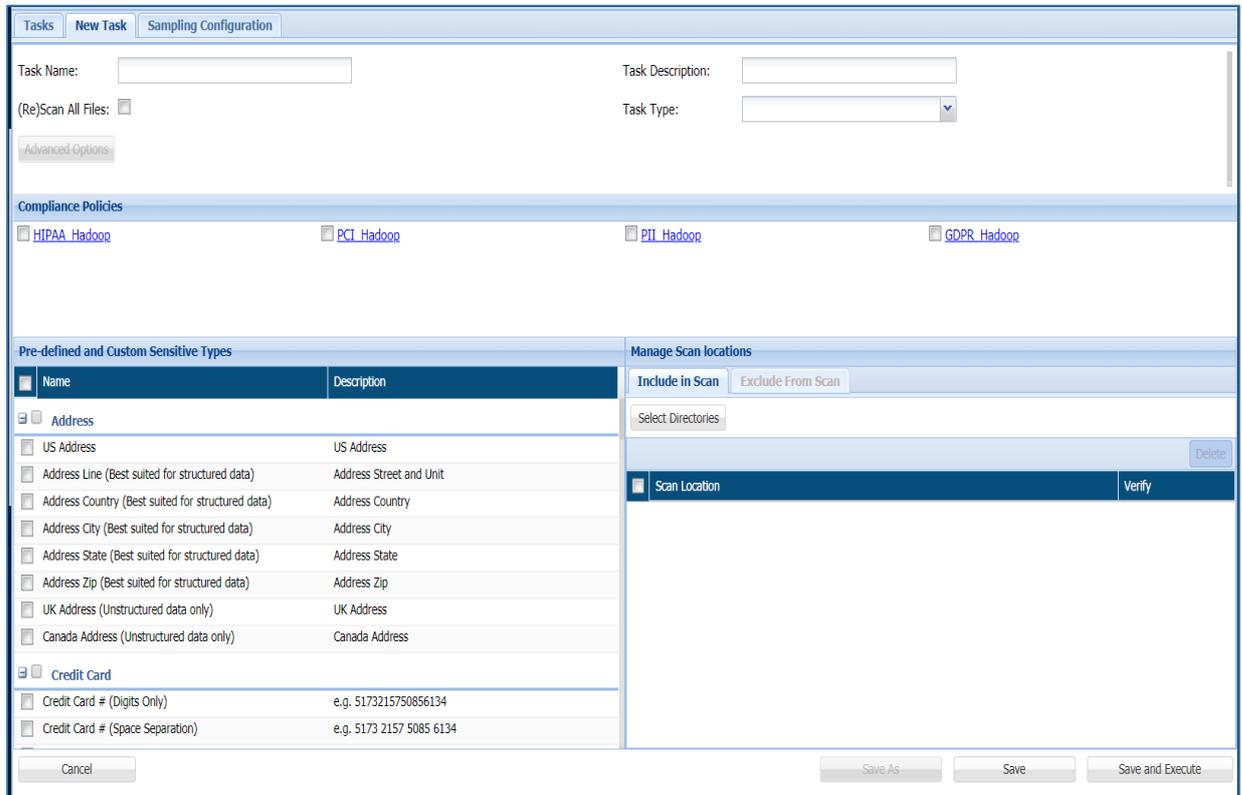
7.2.3.1 HDFS

Detection, Metadata discovery, Masking, Encryption and Decryption operations can be performed on HDFS systems. To create a task to for these operations perform the following

steps

7.2.3.1.1 Create an HDFS task

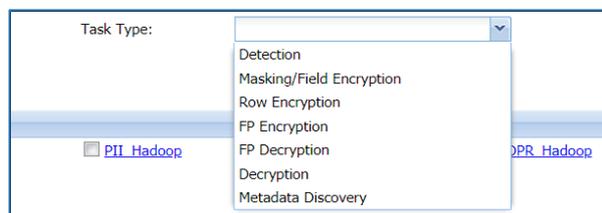
1. Go to **HADOOP > HDFS > TASKS**. Select the **New Task** tab.



2. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.



3. Choose a Task Type.



The list of available task types and the options specific to that task type are explained below:

Detection

Detection tasks scan the target source system for sensitive data elements. The following options are specific to the detection tasks:

- a) **Scan All Files:** Check this checkbox to scan all the available files for a given connection between the dates specified in **Files Modified After** and **Files Modified Before** drop down.

(Re)Scan All Files:
 Dump Metadata:
 Files Modified After: 05/02/2020 05:00 Files Modified Before: 05/04/2020 05:00

- b) **Dump Metadata:** Check **Dump Metadata** option to remove the metadata files after scanning.
- c) **Advanced Options:** Define a batch size for scanning the source system in batches. Click on the **Advanced Options** button to define the batch size of the data.

Detection Advanced Options

Auto Batch Size

Batch Size(Files): 0 Batch min size(MB): 512

Cancel Ok

Define the number of files per batch in the **Batch Size(Files)** option or check the **Auto Batch Size** option to enter the minimum batch size in **Batch min size(MB)** option.

- d) **Read Files:** Choose to read the entire file or a part of the file at random.

Read Files: Entire Files Part of Files

- e) **Exit on first hit:** Check this checkbox to report the table or database as sensitive, at the first event of detection of a sensitive type.

Exit on first hit:

- f) Sampling Configuration

Sampling Configuration: Advanced

DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

- i. Top 1000 Rows
- ii. Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. To create a new sampling configuration perform the following steps:

- i. Go to **HDFS -> TASKS -> SAMPLING CONFIGURATION** tab or check the Advanced checkbox.

File Size Range (Bytes)	Sample Value	Sample By	Actions
Default	1000	Rows	

- ii. Enter the name of the Sampling Configuration.

Name: *

- iii. Enter the description.

Description:

- iv. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.



- v. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling.

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

Below are the options for advanced settings:

- **File Size Range:** Enter the range for the sample in Bytes.
 - **To:** Enter the ending range for the sample.
 - **By:** To Specify how to pick data for sampling from the source system, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
 - **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.
- vi. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.
 - vii. Click the **Save** button to save the changes.

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

Masking/Field Encryption

Masking or Field Encryption hides sensitive data in the target source system by replacing it with a system generated value. DgSecure provides various masking options to ensure the usability of data after it has been protected. Masking/Field Encryption can be applied on structured as well as unstructured data. For more details refer to section [7.4 Masking Options](#). Following options are available to customize a Masking/Field Encryption task:

- a) **(RE)Scan All Files:** Scan all available files for a given connection.

2. **Structured:** Check the structured option if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#)

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

Row Encryption

Row encryption is ideal for unstructured data. Following additional option is available:

- a) **Scan All Files:** Scan all available files for a given connection.

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

FP Encryption

FP or Format Preserving encryption can only be executed on structured files. Following additional options are available:

- a) **Scan All Files:** Scan all available files for a given connection.
- b) **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

FP Decryption

FP or Format Preserving Decryption is used to decrypt the data encrypted by an FP Encryption task. Following additional options are available

- a) **Scan All Files:** Scan all available files for a given connection.
3. **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

Decryption

Decryption tasks are used to decrypt data on which Encryption has been performed.

- a) **Scan All Files:** Scan all available files for a given connection.

4. **Structured:** Check the structured if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

Metadata Discovery

Metadata Discovery scans the source system to provide information about the type of data available on the source system. Following additional option can be set for Metadata Discovery:

- a) **Recursion Levels:** Select number of recursions levels to be scanned. This option is available only for Metadata Discovery.

Recursion Levels: All Custom

To proceed for remaining steps, go to step 4 [Compliance Policy](#).

5. **Compliance Policy** can be set with all the task types in HDFS except Metadata Discovery. For more details about compliance policies, refer to section [Policy](#). After selecting the required options, perform the following steps:

- a) Select the required policies.

Compliance Policies

<input checked="" type="checkbox"/> HIPAA_Hadoop	<input checked="" type="checkbox"/> PCI_Hadoop	<input checked="" type="checkbox"/> PII_Hadoop	<input type="checkbox"/> GDPR_Hadoop
--	--	--	--

6. **Pre-defined and Custom Sensitive Types** are available for all task types in HDFS except Metadata Discovery. Select the required sensitive types.

Pre-defined and Custom Sensitive Types	
Name	Description
<input checked="" type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input checked="" type="checkbox"/> Address Country (Best suited for structured data)	Address Country
<input checked="" type="checkbox"/> Address City (Best suited for structured data)	Address City
<input checked="" type="checkbox"/> Address State (Best suited for structured data)	Address State
<input checked="" type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input checked="" type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input checked="" type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
Dates	
<input type="checkbox"/> DOB (Best suited for structured data)	Date Of Birth

Cancel

***Note:**

1. Row Encryption uses default row encryption configuration for masking. This will mask all the entries of the row and is best suited to unstructured datatypes such as text files.
2. FP Encryption uses default encryption configuration to protect the original data format. This option is best suited to structured datatypes.
3. FP Decryption can only be executed on data that has been encrypted using FP Encryption.
4. Decryption can be executed on data that has been encrypted using FP Encryption or Field Encryption.

7. If you select Masking/Encryption as the Task Type, Protection Option and Consistent fields are also available. Select the required Protection Option for the selected sensitive types. For details about all the masking options available in HDFS refer to 7.4 Masking Options.

Pre-defined and Custom Sensitive Types				
<input type="checkbox"/>	Name	Description	Protection Option	Consistent
<input type="checkbox"/> Address				
<input type="checkbox"/>	US Address	US Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Line (Best suited for structur...	Address Street and Unit	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Country (Best suited for stru...	Address Country	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Address City (Best suited for structur...	Address City	AES Encryption	<input type="checkbox"/>
<input type="checkbox"/>	Address State (Best suited for structu...	Address State	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Zip (Best suited for structure...	Address Zip	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	UK Address (Unstructured data only)	UK Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Canada Address (Unstructured data o...	Canada Address	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/> Credit Card				
<input checked="" type="checkbox"/>	Credit Card # (Digits Only)	e.g. 5173215750856134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Random (Credit Card Numbers)	<input type="checkbox"/>

- If you select **Metadata Discovery** as the task type, the panel **Metadata File Types** will be displayed. Check the required file types for the scan by checking the checkbox against the Name of the required file type.

Metadata File Types		
<input type="checkbox"/>	Name	Type
<input type="checkbox"/>	AVRO	avro
<input checked="" type="checkbox"/>	SEQ	seq
<input checked="" type="checkbox"/>	RC	rc
<input type="checkbox"/>	ORC	orc
<input type="checkbox"/>	PARQUET	parquet
<input checked="" type="checkbox"/>	XML	xml
<input type="checkbox"/>	TEXT	txt
<input type="checkbox"/>	EDI	edi
<input type="checkbox"/>	CSV	csv
<input type="checkbox"/>	PDF	pdf
<input type="checkbox"/>	JSON	json
<input type="checkbox"/>	MSEXCEL	msexcel

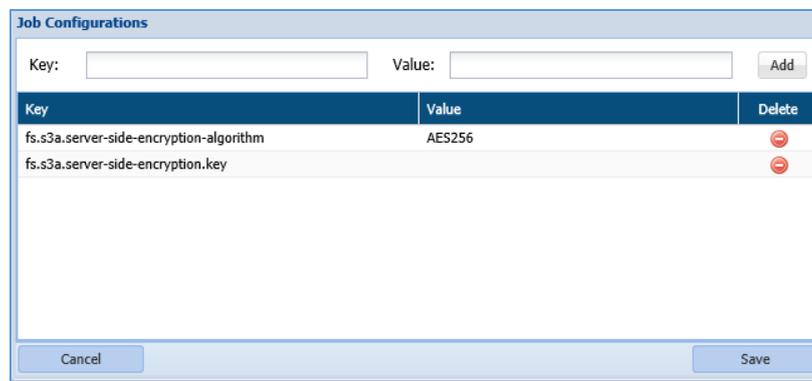
- Manage Scan Locations:** Specify which directories to scan.
 - Include in Scan:** Click **Select Directories** to choose the directory to perform the task.

Manage Scan locations

Include Files that failed previously:

Scan Location	Verify
<input type="checkbox"/> /munish	<input type="checkbox"/>

- viii. **Include Files that Failed Previously:** check this option to include all those files which were previously not scanned due to some exception. This option will appear when **Detection** is selected in **Task Type** field.
- ix. **Delete Input Files on Job Completion:** check this option to delete all the original input files, which were included, post masking. This option will appear when **Masking/Field Encryption, Row Encryption, FP Encryption** is selected in **Task Type** field.
- x. **Job Configuration:** check the checkbox to setup the parameters list. The value for job configuration will contain the pre-defined key and the value. If you have not specified any **Job Configuration**, then default parameter list will be executed.

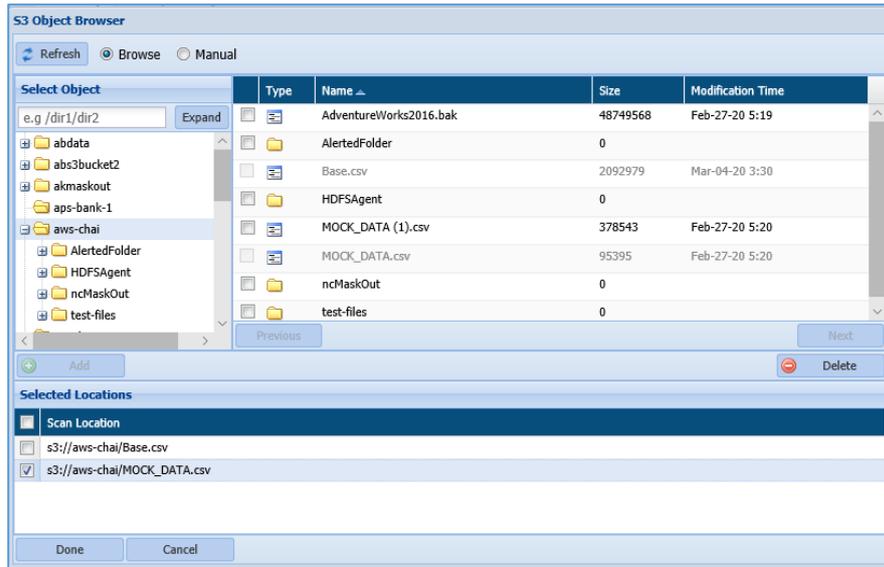


There are two ways in which buckets can be selected in **S3 Object Browser**:

- Browse
- Manual

To include buckets for scanning, click **Select Bucket** button, perform the following steps:

Browse: This option lets you browse the objects from **Select Object** panel. To select the objects, perform the below steps:



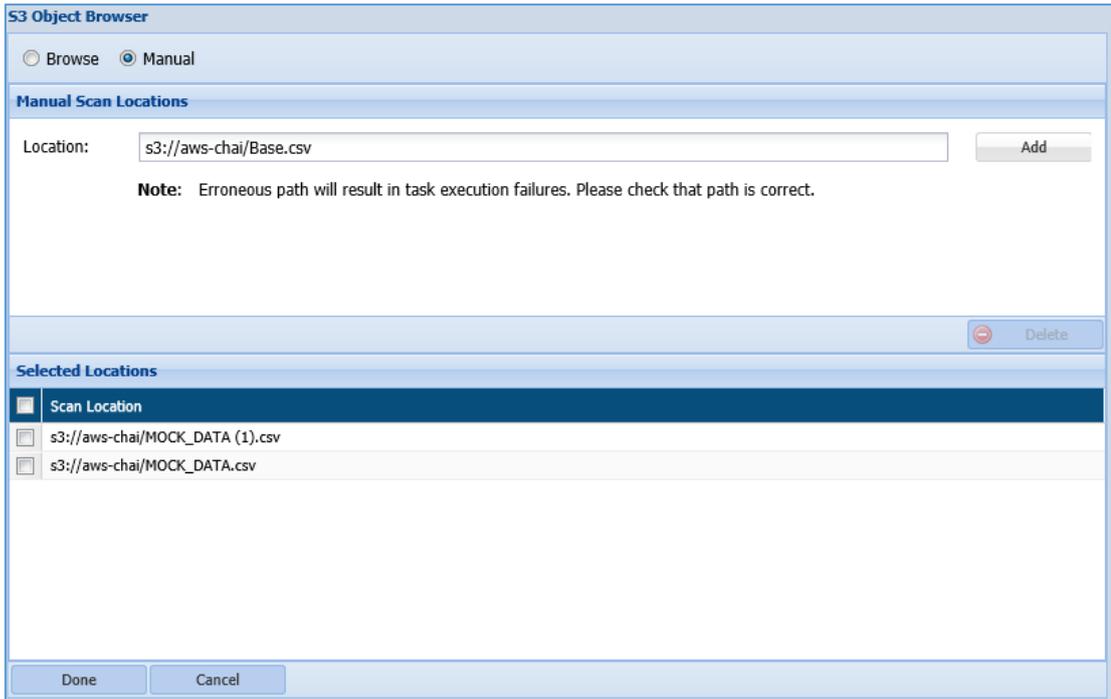
- i. Select the folder from the directory to view the objects for that folder. All the objects will be displayed in the right panel.
- ii. Select the object in the right panel by checking the checkbox and click **Add** button. This functionality will include the objects in the **Selected Locations** panel.



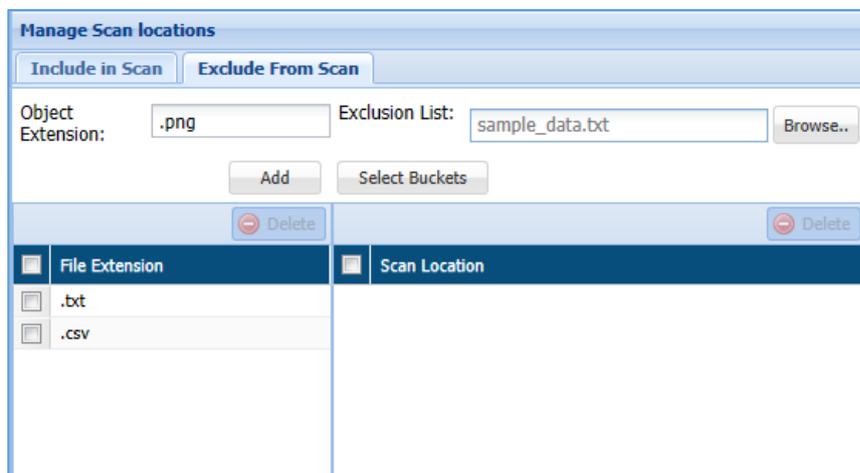
To delete the objects from the **Selected Location** panel, check the checkbox next to the **Scan Location** name and click **Delete** button. The **Delete** button will be enabled, once the **Scan Location** checkbox is checked.

- iii. Click **Done** button to include the objects in **Manage Scan Location**.

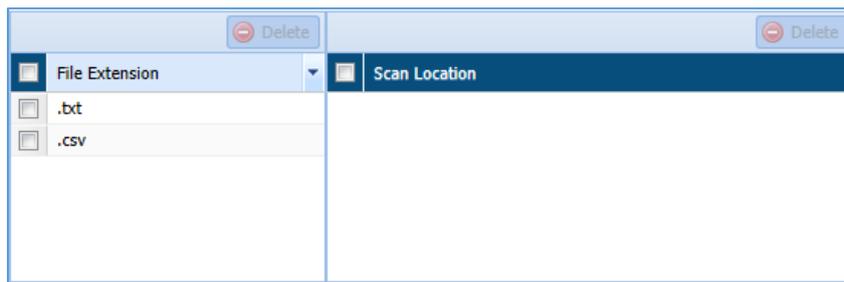
Manual: This option allows you to select the object from the bucket manually. Perform the below steps:



- i. Enter the path for the scan location that you want to include.
 - ii. Click **Add** button to include the path for the object in Selected Location panel.
 - iii. Click **Done** to include the selected objects in the Manage Scan Location.
 - iv. To delete the object from the **Selected Location** panel, check the checkbox next to the scan location name and click **Delete**.
10. **Exclude from Scan:** Select objects to be excluded at the time of task execution or browse the path to an exclusion list. The **Exclude From Scan** tab will be enabled when **Detection** and **Masking/Field Encryption** are selected in Task Type field.



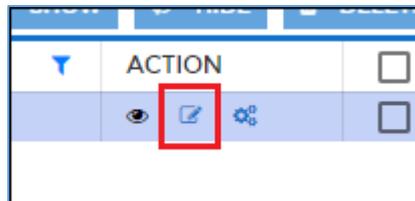
- i. The **Object Extension** field allow you to specify the objects extension (.txt, .csv, etc) that need to be excluded.
- ii. The **Extension List** field allow you either to **Browse** the path for the object or to select the object using the **Select Buckets** button. This functionality lets you specify the exclusion list of all objects that need not be included in the scanning.
- iii. The below panel list down information for both excluded **File Extension** and the **Scan Location** selected above.



***Note:** To decrypt tasks, ensure that appropriate roles have been assigned before executing the encryption task. For more details, refer to section [Role Management](#).

- 11. After creating the required task, i.e., detection, masking, encryption, decryption or metadata discovery, click save to save the task to schedule later, or save and execute to execute it right away.

To edit an existing task, select the required task from the list of tasks on the **HBASE Tasks** screen and click **Edit**. A task can be edited using the same steps for task creation.



7.2.3.2 HBase

DgSecure supports Detection and Protection on HBASE systems. To create a task to for these operations perform the following steps:

7.2.3.2.1 Create a task

1. Go to HADOOP>HBASE>TASKS. Select the New Task tab.

PRE-DEFINED AND CUSTOM SENSITIVE TYPES		COLUMNS SELECTED			
SENSITIVE TYPE	DESCRIPTION	NAMESPACE	TABLE	COLUMN FAMILY	COLUMN
Address		ABC	table_abc	CCNO	ccno_dash
US Address	US Address	ABC	table_abc	CCNO	ccno_digit
Address Line (Best suited for structured data)	Address Street and Unit	ABC	table_abc	CCNO	ccno_space
Address Country (Best suited for structured data)	Address Country				
Address City (Best suited for structured data)	Address City				
Address State (Best suited for structured data)	Address State				
Address Zip (Best suited for structured data)	Address Zip				
UK Address (Unstructured data only)	UK Address				
Canada Address (Unstructured data only)	Canada Address				
ccno_digit					
ccno_digit	ccno				
Credit Card					
Credit Card # (Digits Only)	e.g. 5173215750856134				

12. Select cluster from the **Select Cluster** drop-down.

13. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.

14. Choose a **Task Type**.

15. The available task types and the options specific to that task type are explained below:

Detection

Detection tasks scan the target source system for sensitive data elements. The following

options are specific to the detection tasks:

a) **Full Scan:** Select this option to scan all the records in the table.

16. **Sampling by Number of Rows/Region:** Upon choosing this option, only specified number of rows (from Top) would be scanned from each region of the table storage.

For e.g., if value entered is 1000 and the table is spread across 5 regions, then top 1000 rows in lexicographical order from each region would be scanned, making the task to scan total of 5000 rows from table of detection.

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

Protection

Protection hides sensitive data in the target source system by replacing it with a system generated value.

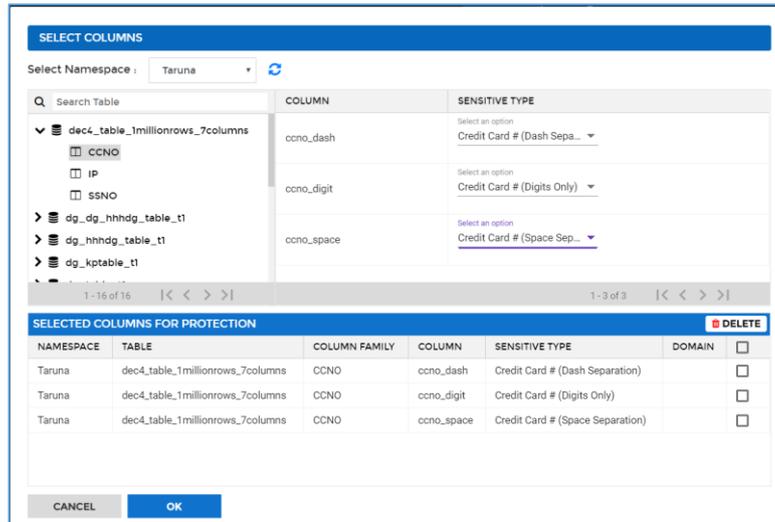
a) **Cell Versions:** Click **Specify Version** to select a specific version of cells to mask and enter the version number. Alternatively, to mask all versions, select the option **All**.

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

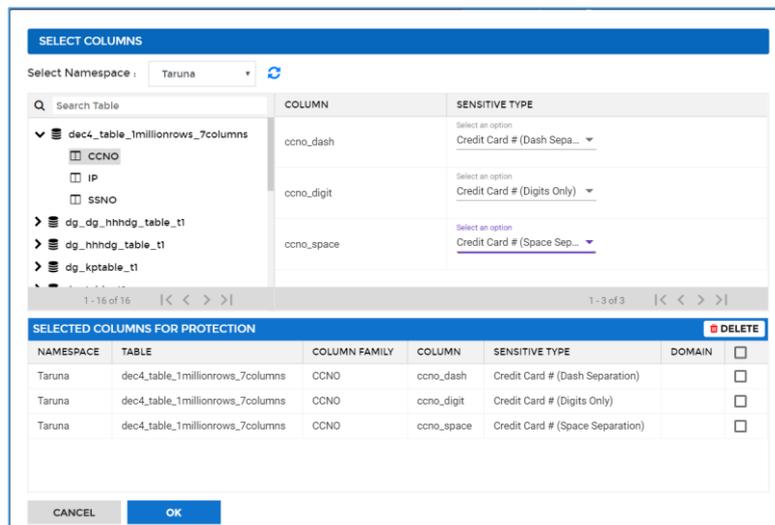
17. **Compliance Policy** can be set with all the task types. For more details about compliance policies, refer to section [Policy](#). After selecting the required options, perform the following steps:

a) Select the required policies.

18. Pre-defined and Custom Sensitive Types are available for all task types. Select the required sensitive types.



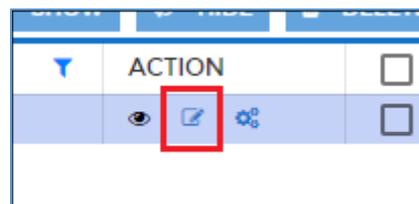
21. Select the required **Sensitive Type** for the selected columns.



22. Click **OK** to save the selection.

23. After creating the required task, click **Save** to save the task to schedule later, or **Save and Execute** to execute it right away.

To edit an existing task, select the required task from the list of tasks on the **HBASE Tasks** screen and click **Edit**. A task can be edited using the same steps for task creation.



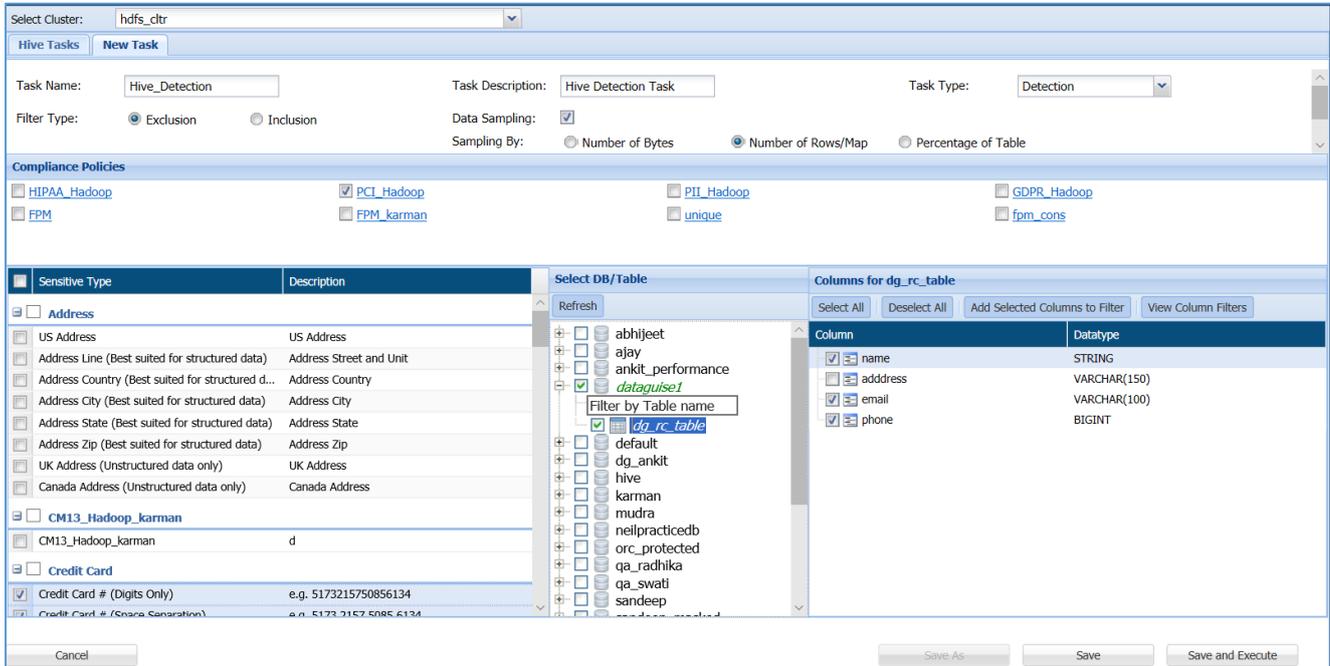
7.2.3.3 Hive

DgSecure supports Detection and Protection on Hive systems. To create a task to for these

operations perform the following steps:

7.2.3.3.1 Create a task

1. Go to **HADOOP > HIVE > TASKS**. Select the New Task tab.

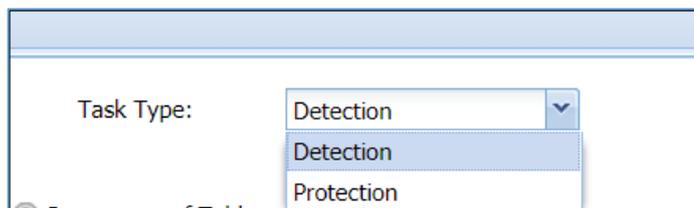


2. Select cluster from the **Select Cluster** drop-down.

24. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.



25. Choose a **Task Type**.



26. The available task types and the options specific to that task type are explained below:

Detection

Detection tasks scan the target source system for sensitive data elements.

Filter Type:	<input checked="" type="radio"/> Exclusion	<input type="radio"/> Inclusion	Data Sampling:	<input checked="" type="checkbox"/>
			Sampling By:	<input type="radio"/> Number of Bytes <input checked="" type="radio"/> Number of Rows/Map <input type="radio"/> Percentage of Table
			Number of Rows/Map:	1000

The following options are specific to the detection tasks:

- a) **Filter Type:** Select **Exclusion** to exclude the selected tables while performing the detection task or **Inclusion** to include the selected tables while performing the detection task.

- 27. **Sampling By:** DgSecure is equipped with the data sampling to limit the area of scan which helps in reducing the time taken for detection. Check this checkbox to run detection on part of the data. Following are the available options:
 - i. **Number of Bytes** - Select this option to define the sample size in bytes/KB/MB/GB. Enter the size in the **Number of Bytes** text box.
 - ii. **Number of Rows/Map** - Select this option to define the sample size in number of rows or map. Enter the size in the **Number of Rows/Map** text box.
 - iii. **Percentage of Table** – Select this option to define the sample size in percentage of total tables. Enter the percentage in the **Percentage of Table** text box.

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

Protection

Protection hides sensitive data in the target source system by replacing it with a system generated value.

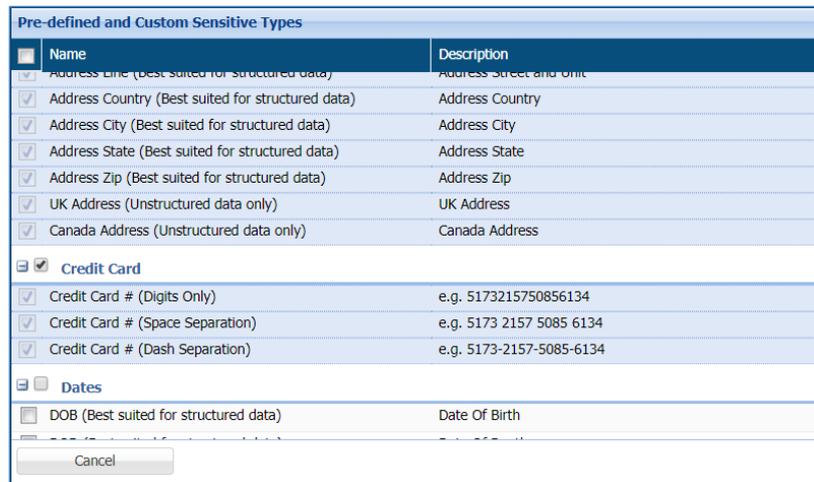
To proceed for remaining steps, go to step 6 [Compliance Policy](#).

- 28. **Compliance Policy** can be set with all the task types. For more details about compliance policies, refer to section [Policy](#). After selecting the required options, perform the following steps:

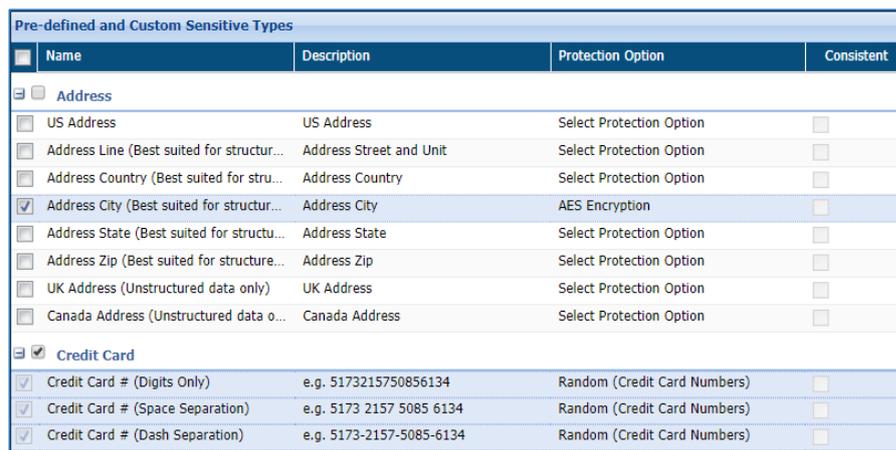
- a) Select the required policies.

Compliance Policies			
<input checked="" type="checkbox"/> HIPAA_Hadoop	<input checked="" type="checkbox"/> PCI_Hadoop	<input checked="" type="checkbox"/> PII_Hadoop	<input type="checkbox"/> GDPR_Hadoop

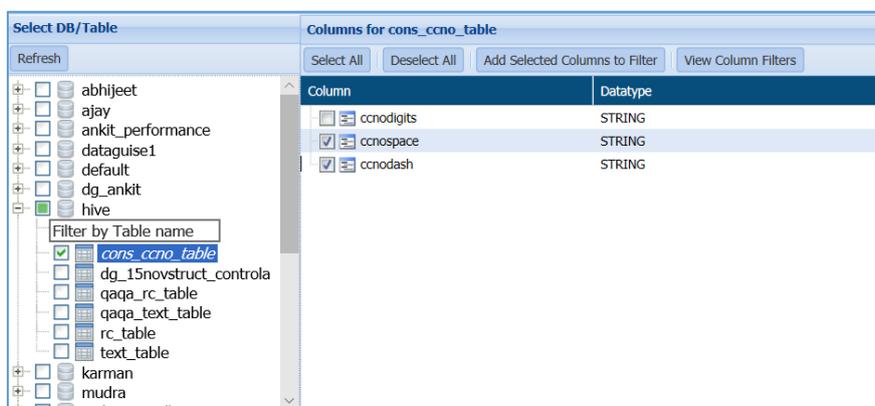
29. Pre-defined and Custom Sensitive Types are available for all task types. Select the required sensitive types.



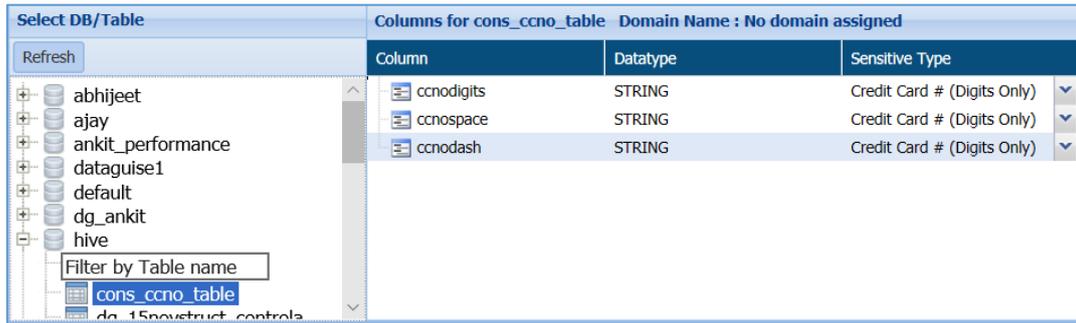
If you select **Protection** as the **Task Type**, **Protection Option** and **Consistent** fields are also available. Select the required **Protection Option** for the selected sensitive types. For details about all the masking options available refer to [Masking Options](#).



30. Select the database or table in the **Select DB/Table** panel.



- a) Select the table and columns to scan. If you select **Protection** as the **Task Type**, there are some extra fields:



- 31. Select the table and columns to scan. Select the sensitive type of the columns.
- 32. If you select **Protection** as the task type, the panel **Columns Selected For Masking** will be displayed.

Database	Table	Column	Data Type	Sensitive Type	Domain Name
hive	cons_ccno_table	ccnodigits	STRING	Credit Card # (Digits Only)	
hive	cons_ccno_table	ccnospace	STRING	Credit Card # (Digits Only)	
hive	cons_ccno_table	ccnodash	STRING	Credit Card # (Digits Only)	

- 33. After creating the required task, click **Save** to save the task to schedule later, or **Save and Execute** to execute it right away.

To edit an existing task, select the required task from the list of tasks on the **Hive Tasks** screen and click **Edit**. A task can be edited using the same steps for task creation.

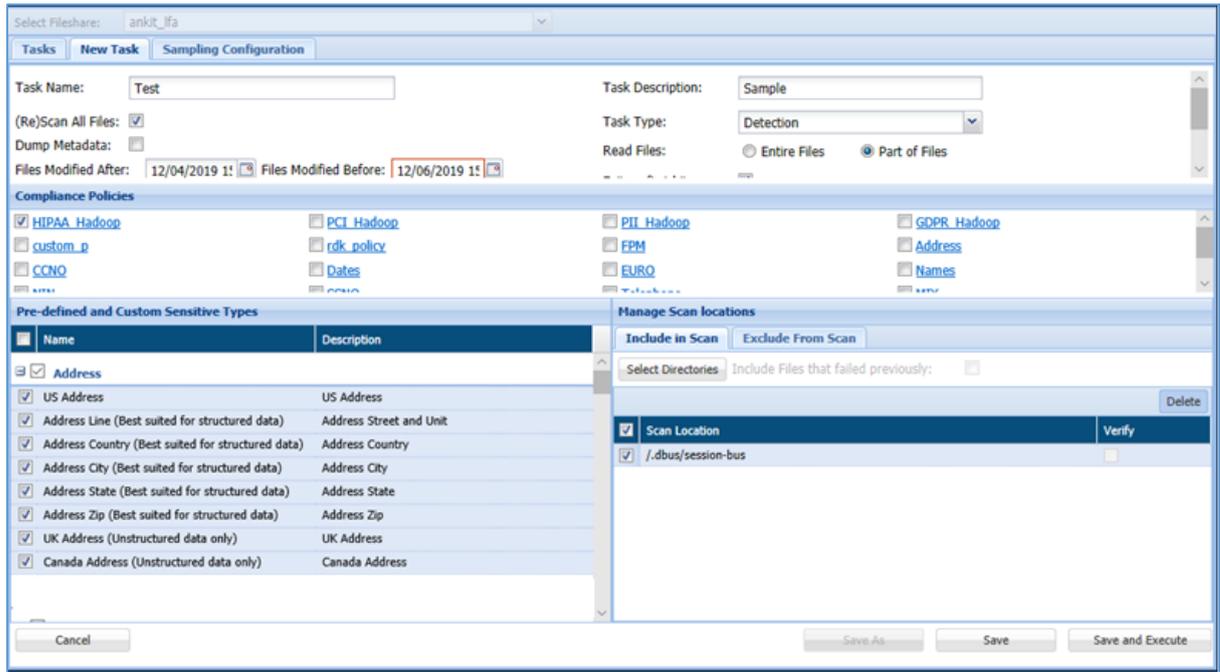


7.2.4 Files

DgSecure supports Detection and Protection on different files. Detection, Masking, Encryption and Decryption operations can be performed on Files systems. To create a task to for these operations perform the following steps:

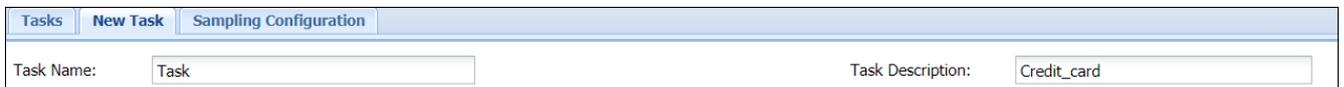
7.2.4.1.1 Create a task

- 1. Go to **FILES>TASKS**. Select the **New Task** tab.



34. Select files from the **Select Fileshare** drop-down.

35. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.



36. Choose a **Task Type**.



37. The list of available task types and the options specific to that task type are explained below:

Detection

Detection tasks scan the target source system for sensitive data elements. The following options are specific to the detection tasks:

- a) **Scan All Files:** Check this checkbox to scan all the available files for a given connection between the dates specified in **Files Modified After** and **Files Modified Before** drop down.

(Re)Scan All Files:
 Dump Metadata:
 Files Modified After: 05/02/2020 05:00 Files Modified Before: 05/04/2020 05:00

38. **Dump Metadata:** Check **Dump Metadata** option to remove the metadata files after scanning.

39. **Read Files:** Choose to read the entire file or a part of the file at random.

Read Files: Entire Files Part of Files

40. **Exit on first hit:** Check this checkbox to report the table or database as sensitive, at the first event of detection of a sensitive type.

Exit on first hit:

41. **Sampling Configuration:**

Sampling Configuration: Top 1000 rows
 Advanced

DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

- Top 1000 Rows
- Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. To create a new sampling configuration, perform the following steps:

- Go to **HDFS > TASKS > SAMPLING CONFIGURATION** tab or check the Advanced checkbox.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

File Size Range (Bytes)	Sample Value	Sample By	Actions
Default	1000	Rows	

ii. Enter the name of the Sampling Configuration.

Name: *

iii. Enter the description.

Description:

iv. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.

Set Sampling Config as Default:

v. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling.

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

Below are the options for advanced settings:

- **File Size Range:** Enter the range for the sample in Bytes.
- **To:** Enter the ending range for the sample.
- **By:** To Specify how to pick data for sampling from the source system, there are two ways:
 - i. **By Rows:** Select '**Rows**' from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select '**Percent**' from the drop-down, to sample a percentage of the data
- **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.
- vi. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.
- vii. Click the **Save** button to save the changes.

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

Masking/Field Encryption:

Masking or Field Encryption hides sensitive data in the target source system by replacing it with a system generated value. DgSecure provides various masking options to ensure the usability of data after it has been protected. Masking/Field Encryption can be applied on structured as well as unstructured data. For more details, refer to [Masking Option](#). Following options are available to customize a Masking/Field Encryption task:

- a) **(Re)Scan All Files:** Scan all available files for a given connection.
42. **Structured:** Check the structured option if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#).

Row Encryption

Row encryption is ideal for unstructured data. Following additional option is available:

- a) **(Re)Scan All Files:** Scan all available files for a given connection.

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

FP Encryption

FP or Format Preserving encryption can only be executed on structured files. Following additional options are available:

a) **(Re)Scan All Files:** Scan all available files for a given connection.

43. **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

FP Decryption

FP or Format Preserving Decryption is used to decrypt the data encrypted by an FP Encryption task. Following additional options are available

a) **(Re)Scan All Files:** Scan all available files for a given connection.

44. **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

Decryption

Decryption tasks are used to decrypt data on which Encryption has been performed.

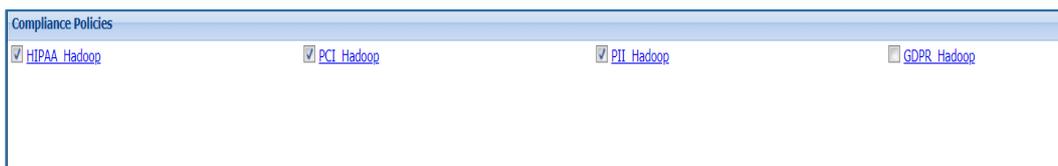
a) **(Re)Scan All Files:** Scan all available files for a given connection.

45. **Structured:** Check the structured if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step 6 [Compliance Policy](#).

46. **Compliance Policy** can be set with all the task types. For more details about compliance policies, refer to section [Policy](#). After selecting the required options, perform the following steps:

a) Select the required policies.



47. **Pre-defined and Custom Sensitive Types** are available for all task types. Select the required sensitive types.

Pre-defined and Custom Sensitive Types		
<input type="checkbox"/>	Name	Description
<input checked="" type="checkbox"/>	Address Line (Best suited for structured data)	Address Street and Unit
<input checked="" type="checkbox"/>	Address Country (Best suited for structured data)	Address Country
<input checked="" type="checkbox"/>	Address City (Best suited for structured data)	Address City
<input checked="" type="checkbox"/>	Address State (Best suited for structured data)	Address State
<input checked="" type="checkbox"/>	Address Zip (Best suited for structured data)	Address Zip
<input checked="" type="checkbox"/>	UK Address (Unstructured data only)	UK Address
<input checked="" type="checkbox"/>	Canada Address (Unstructured data only)	Canada Address
<input checked="" type="checkbox"/> Credit Card		
<input checked="" type="checkbox"/>	Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/>	Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/>	Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
<input type="checkbox"/> Dates		
<input type="checkbox"/>	DOB (Best suited for structured data)	Date Of Birth
Cancel		

***NOTE:**

- Row Encryption uses default row encryption configuration for masking. This will mask all the entries of the row and is best suited to unstructured datatypes such as text files.
- FP Encryption uses default encryption configuration to protect the original data format. This option is best suited to structured datatypes.
- FP Decryption can only be executed on data that has been encrypted using FP Encryption.
- Decryption can be executed on data that has been encrypted using FP Encryption or Field Encryption.

48. If you select **Masking/Encryption** as the **Task Type**, **Protection Option** and **Consistent** fields are also available. Select the required **Protection Option** for the selected sensitive types. For details about all the masking options available in Files refer to [Masking Options](#).

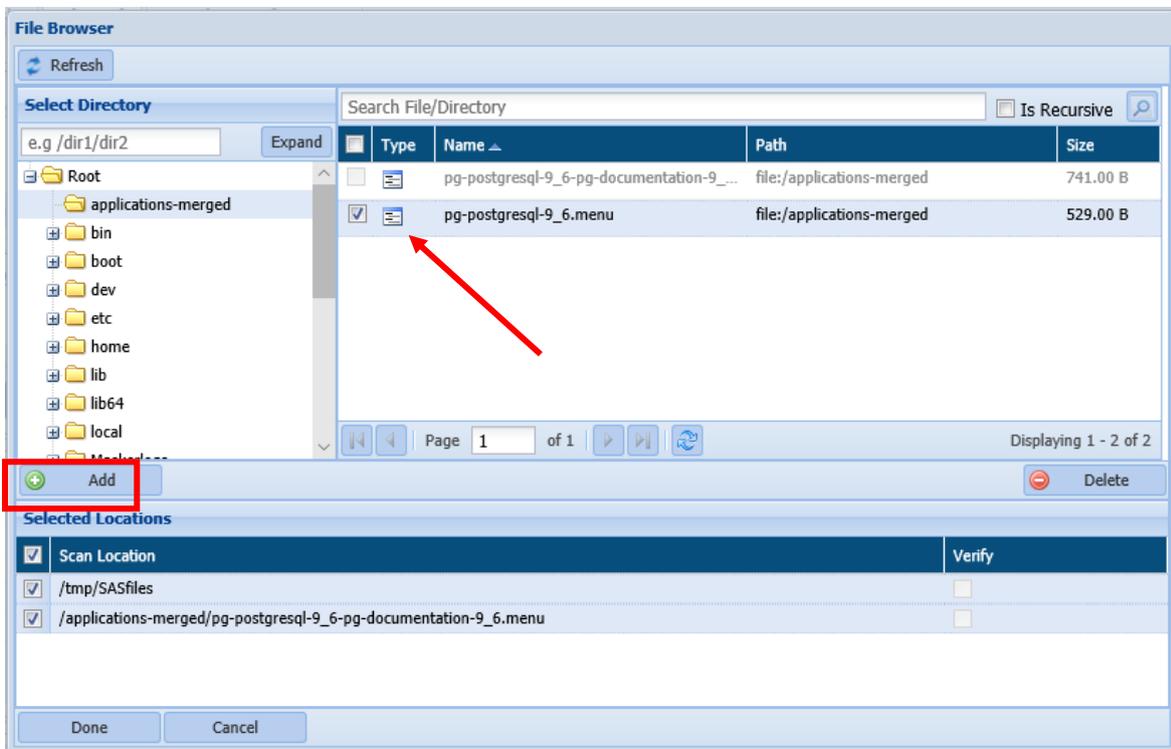
Pre-defined and Custom Sensitive Types				
<input type="checkbox"/>	Name	Description	Protection Option	Consistent
<input type="checkbox"/> Address				
<input type="checkbox"/>	US Address	US Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Line (Best suited for structur...	Address Street and Unit	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Country (Best suited for stru...	Address Country	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Address City (Best suited for structur...	Address City	AES Encryption	<input type="checkbox"/>
<input type="checkbox"/>	Address State (Best suited for structu...	Address State	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Zip (Best suited for structure...	Address Zip	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	UK Address (Unstructured data only)	UK Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Canada Address (Unstructured data o...	Canada Address	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/> Credit Card				
<input checked="" type="checkbox"/>	Credit Card # (Digits Only)	e.g. 5173215750856134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Random (Credit Card Numbers)	<input type="checkbox"/>

49. **Manage Scan Locations:** Specify which directories to scan.

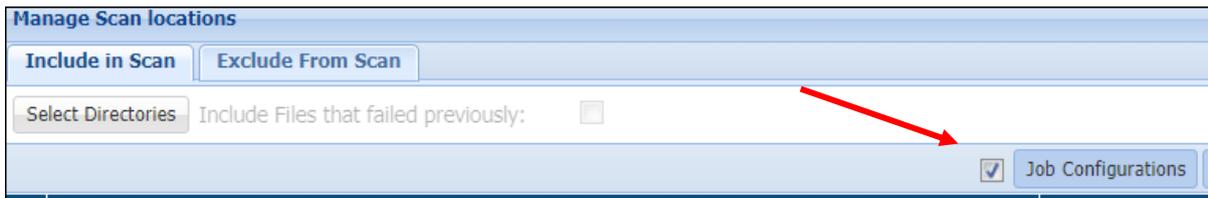
- **Include in Scan:** Perform the following steps to select directories to include in the scan.
 - i. Click **Select Directories** to choose the directory to perform the task.



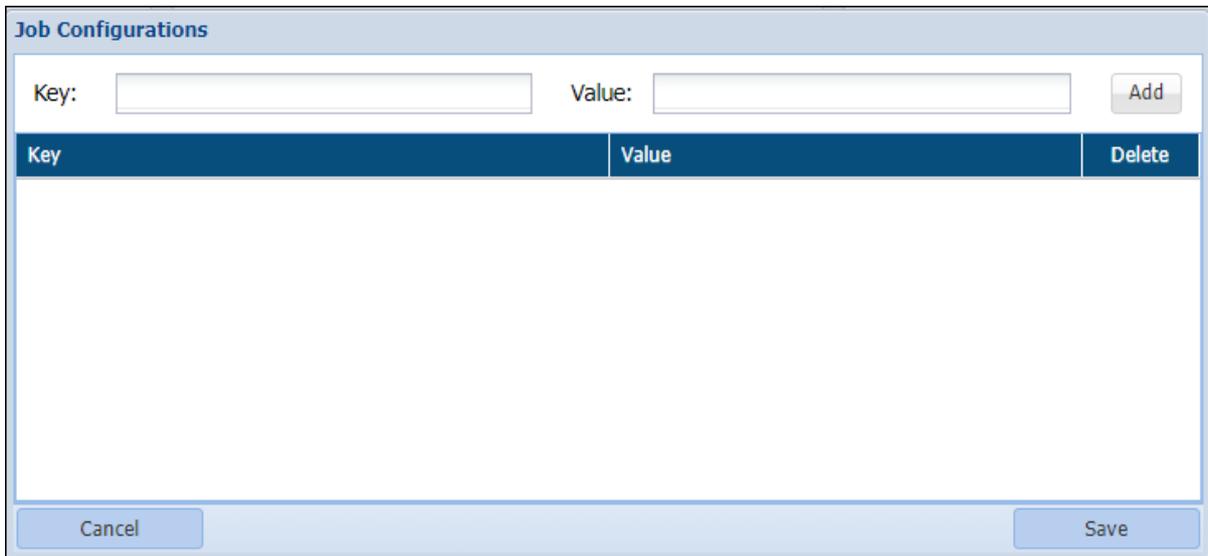
- ii. Select the required directories and click **Add**.



- iii. Check the **Is Recursive** checkbox for recurring data.
- iv. Click the **Add** button. It will add the file in **Selected Locations** Pane.
- v. Click **Done** to save your selection.
- vi. To provide keys for the search check the checkbox against **Job Configuration**.

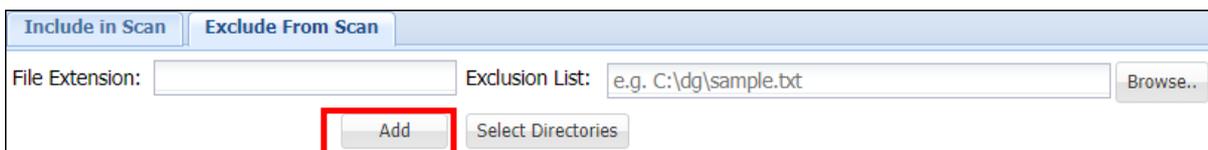


vii. Add the key and value and save.



- viii. **Include Files that Failed Previously** check box will be greyed out for a fresh scan. Check this option if some elements of a previously executed task got skipped or the task was completed with errors. This option is available only for detection tasks.
- ix. **Delete Input Files on Job Completion** option is available for masking and encryption tasks. Check this check box to delete the input data after task execution.
- x. To delete a scan location from the list select it and click **Delete**.

- **Exclude from Scan:** Select directories to be excluded at the time of task execution or browse the path to an exclusion list. Perform the following steps:
 - i. Enter the file extension and browse location to the exclusion list if you have prepared an exclusion list. Click add.



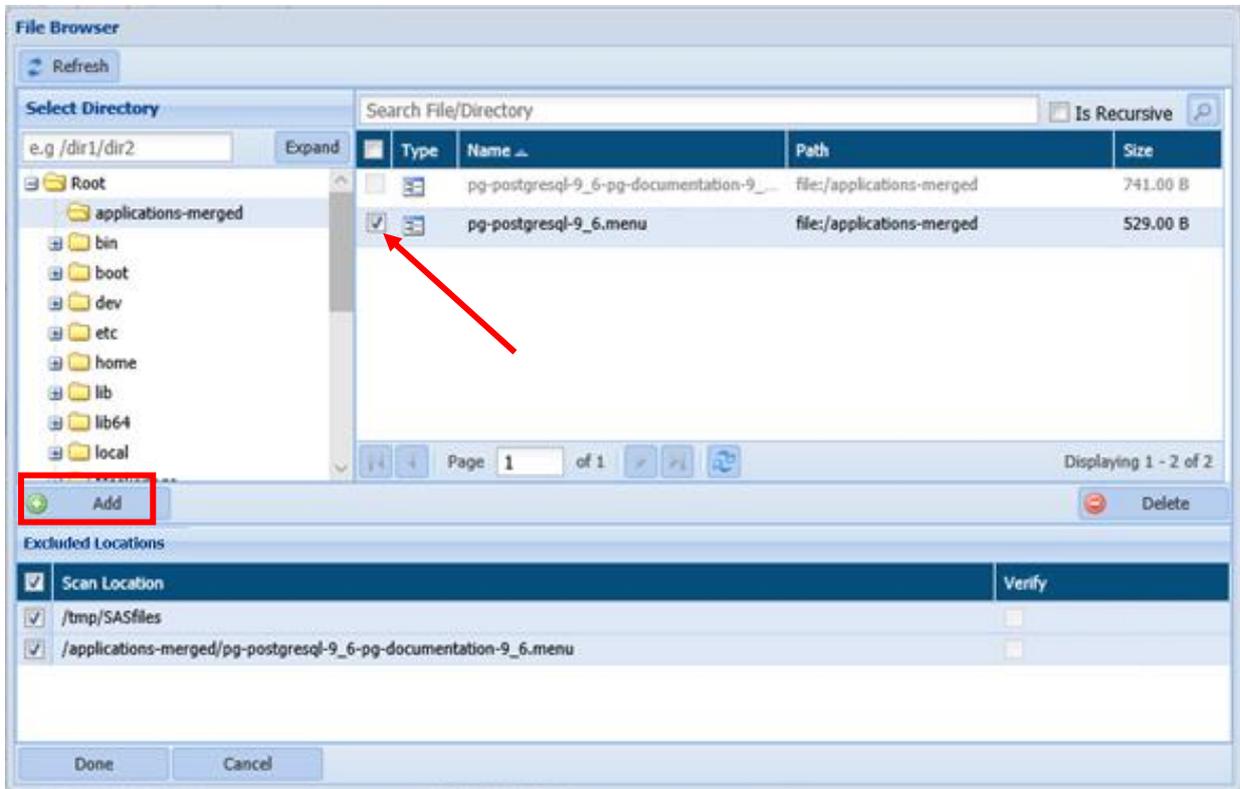
- ii. Click **Select Directories** to manually select the directories to be excluded from the task.

Include in Scan | Exclude From Scan

File Extension: Exclusion List: Browse..

Add | **Select Directories**

iii. Select the required directories and click **Add**.

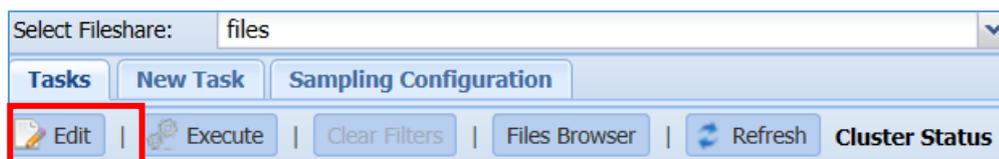


- iv. Check the **Is Recursive** checkbox for recurring data.
- v. Click **Done** to save your selection.

***NOTE:** To decrypt tasks, ensure that appropriate roles have been assigned before executing the encryption task. For more details refer to section [Role Management](#).

50. After creating the required task, i.e., detection, masking, encryption, decryption or metadata discovery, click save to save the task to schedule later, or save and execute to execute it right away.

To edit an existing task, select the required task from the list of tasks on the Tasks screen and click **Edit**. A task can be edited using the same steps for task creation.



7.2.5 AWS

DgSecure supports Detection and Protection on different Hadoop distribution systems like S3, RDS/RedShift. Various protection options are available for these distributions based on the type of data-structure or unstructured.

7.2.5.1 S3

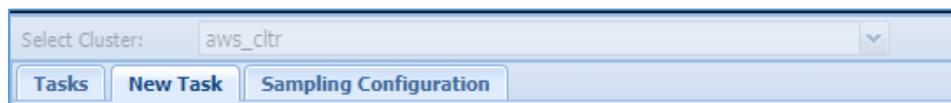
Detection, Metadata discovery, Masking, Encryption and Decryption operations can be performed on S3 systems. To create a task for these operations, perform the following steps.

7.2.5.1.1 Create an S3 task

1. Go to **AWS > S3 > TASKS** and select the cluster from the **Select Cluster** drop-down.



2. To create a new task, click on the **New Task** tab.



The following image shows the user interface for creating a task.

The screenshot shows the task configuration interface with the following details:

- Task Name:** CCard
- Task Description:** Detection task for Credit Card
- (Re)Scan All Objects:**
- Dump Metadata:**
- Task Type:** Detection
- Read Objects:** Entire Objects Part of Objects
- Exit on first hit:**
- Compliance Policies:**
 - HIPAA_Hadoop
 - PCI_Hadoop
 - PII_Hadoop
 - GDPR_Hadoop
 - custom_un
 - fom
 - aes
- Pre-defined and Custom Sensitive Types:**

Name	Description
<input checked="" type="checkbox"/> Address	
<input checked="" type="checkbox"/> US Address	US Address
<input checked="" type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input checked="" type="checkbox"/> Address Country (Best suited for structured data)	Address Country
<input checked="" type="checkbox"/> Address City (Best suited for structured data)	Address City
<input checked="" type="checkbox"/> Address State (Best suited for structured data)	Address State
<input checked="" type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input checked="" type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input checked="" type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
<input type="checkbox"/> Credit Card	
<input type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
- Manage Scan locations:**
 - Include in Scan:** **Exclude From Scan:**
 - Select Buckets:** **Include Objects that failed previously:**
 - Scan Location:** **Verify:**

Buttons at the bottom: Cancel, Save As, Save, Save and Execute.

51. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.

Task Name: Task Description:

52. Choose a **Task Type** from given options.

Task Type:

The list of available task types and the options specific to that task type are explained below:

Detection

Detection tasks scan the target source system for sensitive data elements. The following options are specific to the detection tasks:

(Re)Scan All Objects:
 Dump Metadata:
 Objects Modified After: Objects Modified Before:
 Advanced Options
 Task Type:
 Read Objects: Entire Objects Part of Objects
 Exit on first hit:
 Sampling Configuration:
 Advanced

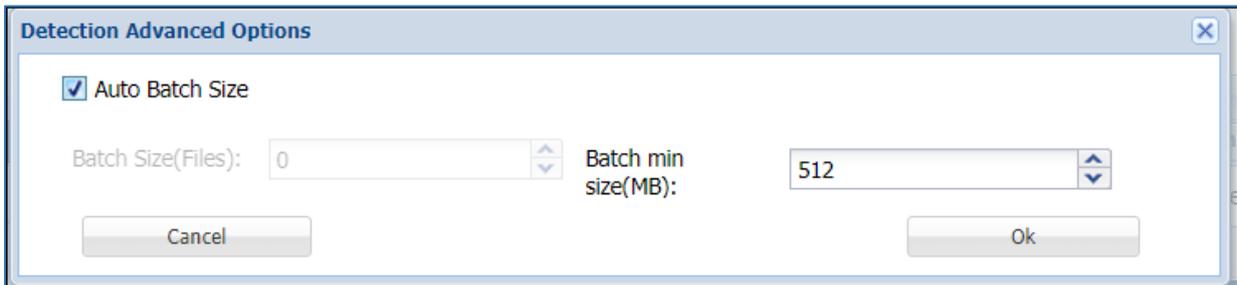
- a) **(Re)Scan All Files**: Check this checkbox to scan all the available objects for a given connection between the dates specified in **Objects Modified After** and **Objects Modified Before** drop down.

The **Objects Modified After** and **Objects Modified Before** drop down will be visible when **(Re)Scan All Objects** checkbox is checked.

(Re)Scan All Objects:
 Dump Metadata:
 Objects Modified After: Objects Modified Before:
 Advanced Options

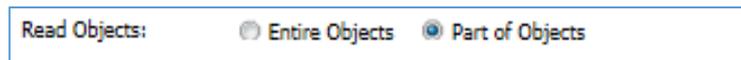
53. **Dump Metadata:** Check **Dump Metadata** option to remove the metadata files after scanning.

54. **Advanced Options:** Define a batch size for scanning the source system in batches. Click on the **Advanced Options** button to define the batch size of the data.



Define the number of files per batch in the **Batch Size(Files)** option or check the **Auto Batch Size** option to enter the minimum batch size in **Batch min size(MB)** option.

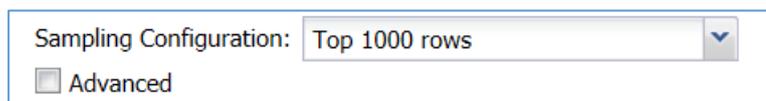
55. **Read Objects:** Choose to read the **Entire Objects** or a **Part of Objects** at random.



56. **Exit on first hit:** Check this checkbox to report the table or database as sensitive, at the first event of detection of a sensitive type.



57. **Sampling Configuration**



DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

- Top 1000 Rows
- Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. To create a new sampling configuration perform the following steps:

- i. Go to **AWS > S3 > TASKS > SAMPLING CONFIGURATION** tab or check the Advanced checkbox.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

File Size Range (Bytes)	Sample Value	Sample By	Actions
Default	1000	Rows	
1 to 100	50	Percent	

- ii. Enter the name of the Sampling Configuration.

Name: *

- iii. Enter the description for the sample.

Description:

- iv. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.

Set Sampling Config as Default:

- v. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling.

Below are the options for advanced settings:

- **File Size Range (Bytes):** Enter the starting range for the sample in bytes.
 - **To:** Enter the ending range for the sample.
 - **By:** To Specify how to pick data for sampling from the source system, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
 - **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling **By-Rows** is selected and denotes the percentage of sampling **By-Percent** is selected.
- vi. After setting up the required details for configuration, click **Add** to add the user-defined sampling configuration to the list.
 - vii. Click the **Save** button to save the changes.

To proceed for remaining steps, go to step [Compliance Policy](#).

Masking/Field Encryption

Masking or Field Encryption hides sensitive data in the target source system by replacing it with a system generated value. DgSecure provides various masking options to ensure the usability of data after it has been protected. Masking/Field Encryption can be applied on structured as well as unstructured data. For more details refer to [Structure Management](#). Following options are available to customize a Masking/Field Encryption task:

- a) **(Re)Scan All Files:** Scan all available files for a given connection.

58. **Structured:** Check the structured option if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step [Compliance Policy](#).

Row Encryption

Row encryption is ideal for unstructured data. Following additional option is available:

- a) **(Re)Scan All Files:** Scan all available files for a given connection.

To proceed for remaining steps, go to step [Compliance Policy](#).

FP Encryption

FP or Format Preserving encryption can only be executed on structured files. Following additional options are available:

- a) **(Re)Scan All Files:** Scan all available files for a given connection.
- b) **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step [Compliance Policy](#).

FP Decryption

FP or Format Preserving Decryption is used to decrypt the data encrypted by an FP Encryption task. Following additional options are available

- a) **(Re)Scan All Files:** Scan all available files for a given connection.
- b) **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step [Compliance Policy](#).

Decryption

Decryption tasks are used to decrypt data on which Encryption has been performed.

- a) **(Re)Scan All Files:** Scan all available files for a given connection.

- 59. **Structured:** Check the structured if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step [Compliance Policy](#).

Metadata Discovery

Metadata Discovery scans the source system to provide information about the type of data available on the source system. Following additional option can be set for Metadata Discovery:

- a) **Recursion Levels:** Select number of recursions levels to be scanned. This option is available only for Metadata Discovery.

Recursion Levels: All Custom

To proceed for remaining steps, go to step [Compliance Policy](#).

60. **Compliance Policy** can be set with all the task types in HDFS except Metadata Discovery. For more details about compliance policies, refer to section [Policy](#) . After selecting the required options, perform the following steps:

- a) Select the required policies.

Compliance Policies

[HIPAA_Hadoop](#)
 [PCI_Hadoop](#)
 [PII_Hadoop](#)
 [GDPR_Hadoop](#)

61. **Pre-defined and Custom Sensitive Types** are available for all task types in HDFS except Metadata Discovery. Select the required sensitive types.

Pre-defined and Custom Sensitive Types	
<input type="checkbox"/> Name	Description
<input checked="" type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input checked="" type="checkbox"/> Address Country (Best suited for structured data)	Address Country
<input checked="" type="checkbox"/> Address City (Best suited for structured data)	Address City
<input checked="" type="checkbox"/> Address State (Best suited for structured data)	Address State
<input checked="" type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input checked="" type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input checked="" type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
<input checked="" type="checkbox"/> Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
<input type="checkbox"/> Dates	
<input type="checkbox"/> DOB (Best suited for structured data)	Date Of Birth

Cancel

***Note:**

1. Row Encryption uses default row encryption configuration for masking. This will mask all the entries of the row and is best suited to unstructured datatypes such as text files.
2. FP Encryption uses default encryption configuration to protect the original data format. This option is best suited to structured datatypes.
3. FP Decryption can only be executed on data that has been encrypted using FP Encryption.
4. Decryption can be executed on data that has been encrypted using FP Encryption or Field Encryption.

62. If you select Masking/Encryption as the Task Type, Protection Option and Consistent fields are also available. Select the required Protection Option for the selected sensitive types. For details about all the masking options available in HDFS refer to [Masking Options](#).

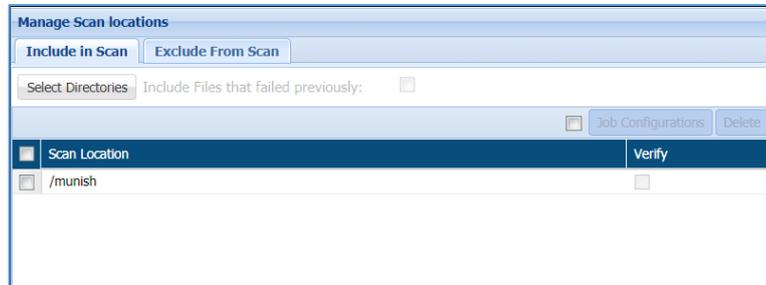
Pre-defined and Custom Sensitive Types				
<input type="checkbox"/>	Name	Description	Protection Option	Consistent
<input type="checkbox"/> Address				
<input type="checkbox"/>	US Address	US Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Line (Best suited for structur...	Address Street and Unit	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Country (Best suited for stru...	Address Country	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Address City (Best suited for structur...	Address City	AES Encryption	<input type="checkbox"/>
<input type="checkbox"/>	Address State (Best suited for structu...	Address State	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Zip (Best suited for structure...	Address Zip	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	UK Address (Unstructured data only)	UK Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Canada Address (Unstructured data o...	Canada Address	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/> Credit Card				
<input checked="" type="checkbox"/>	Credit Card # (Digits Only)	e.g. 5173215750856134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Random (Credit Card Numbers)	<input type="checkbox"/>

63. If you select **Metadata Discovery** as the task type, the panel **Metadata File Types** will be displayed. Check the required file types for the scan by checking the checkbox against the Name of the required file type.

Metadata File Types		
<input type="checkbox"/>	Name	Type
<input type="checkbox"/>	AVRO	avro
<input checked="" type="checkbox"/>	SEQ	seq
<input checked="" type="checkbox"/>	RC	rc
<input type="checkbox"/>	ORC	orc
<input type="checkbox"/>	PARQUET	parquet
<input checked="" type="checkbox"/>	XML	xml
<input type="checkbox"/>	TEXT	txt
<input type="checkbox"/>	EDI	edi
<input type="checkbox"/>	CSV	csv
<input type="checkbox"/>	PDF	pdf
<input type="checkbox"/>	JSON	json
<input type="checkbox"/>	MSEXCEL	msexcel

64. **Manage Scan Locations:** This panel specify which directories to scan.

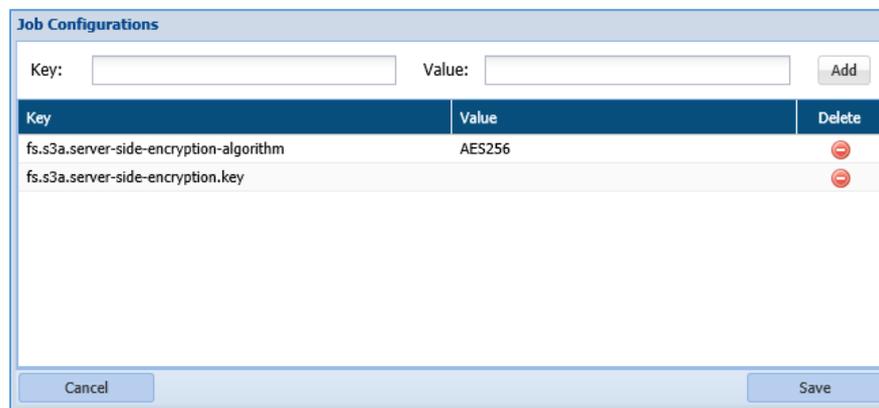
a) **Include in Scan:** Click **Select Directories** to choose the directory to perform the task.



viii. **Include Files that Failed Previously:** check this option to include all those files which were previously not scanned due to some exception. This option will appear when **Detection** is selected in **Task Type** field.

ix. **Delete Input Files on Job Completion:** check this option to delete all the original input files, which were included, post masking. This option will appear when **Masking/Field Encryption, Row Encryption, FP Encryption** is selected in **Task Type** field.

x. **Job Configuration:** check the checkbox to setup the parameters list. The value for job configuration will contain the pre-defined key and the value. If you have not specified any **Job Configuration**, then default parameter list will be executed.

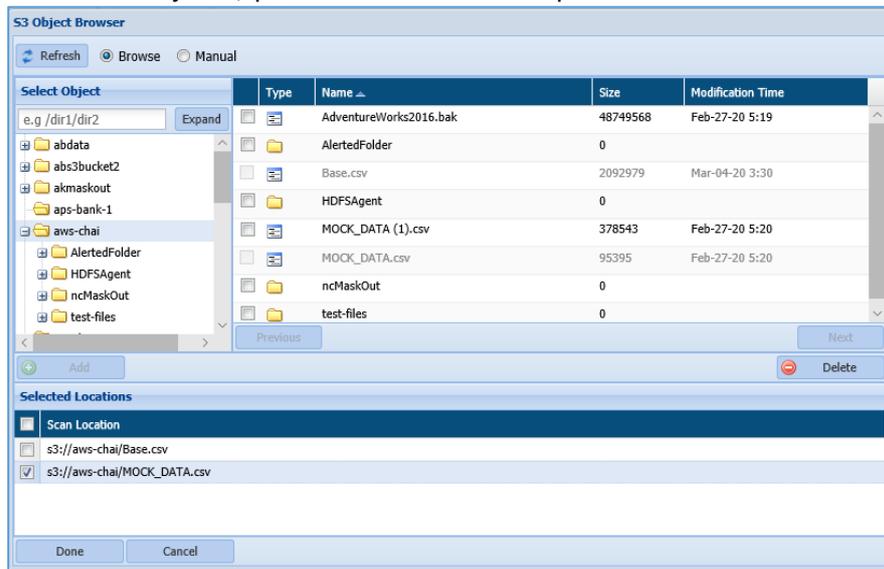


There are two ways in which buckets can be selected in **S3 Object Browser**:

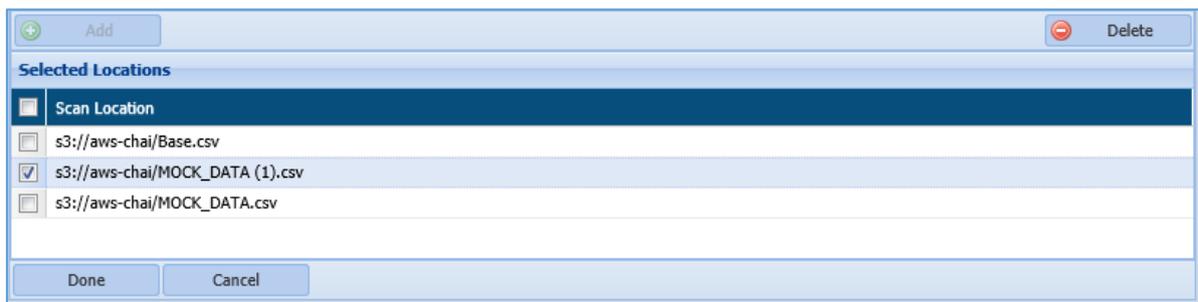
- Browse
- Manual

To include buckets for scanning, click **Select Bucket** button, perform the following steps:

Browse: This option lets you browse the objects from **Select Object** panel. To select the objects, perform the below steps:



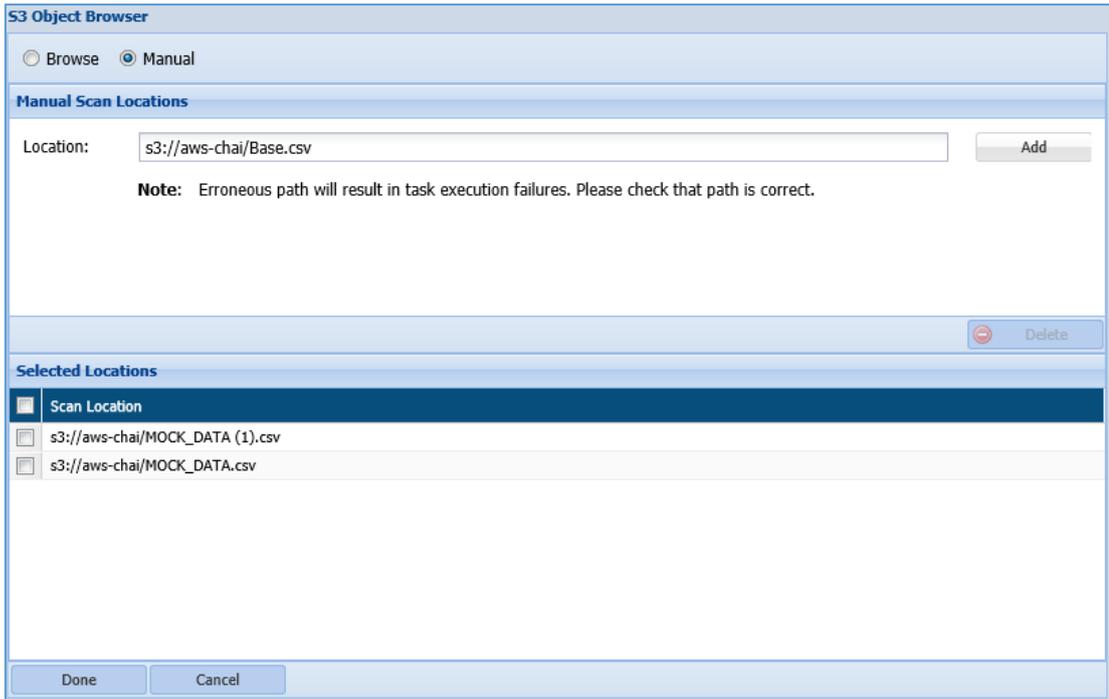
- i. Select the folder from the directory to view the objects for that folder. All the objects will be displayed in the right panel.
- ii. Select the object in the right panel by checking the checkbox and click **Add** button. This functionality will include the objects in the **Selected Locations** panel.



To delete the objects from the **Selected Location** panel, check the checkbox next to the **Scan Location** name and click **Delete** button. The **Delete** button will be enabled, once the **Scan Location** checkbox is checked.

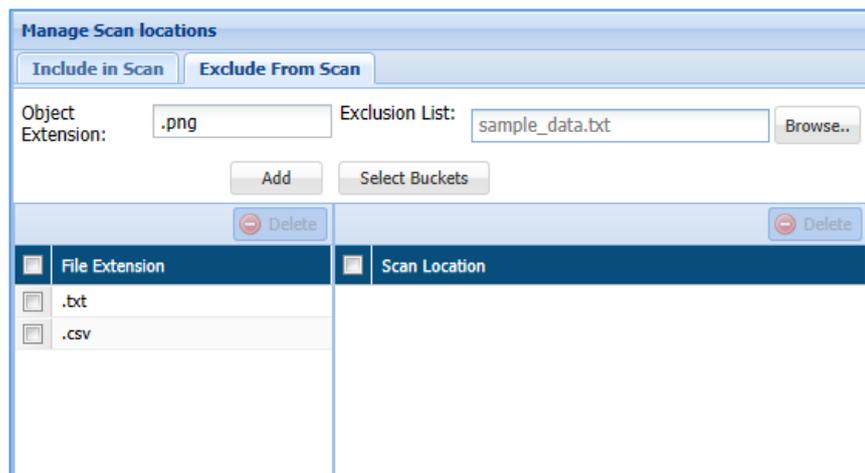
- iii. Click **Done** button to include the objects in **Manage Scan Location**.

Manual: This option allows you to select the object from the bucket manually. Perform the below steps:



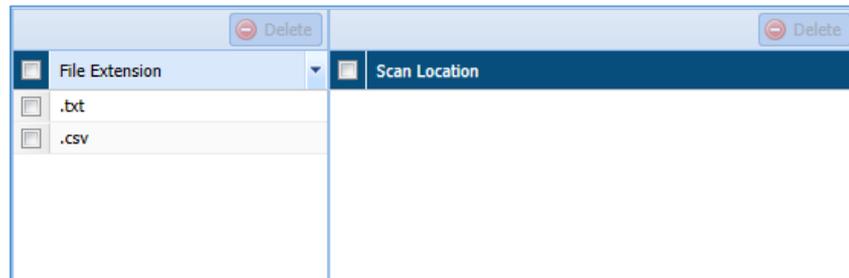
- i. Enter the path for the scan location that you want to include.
- ii. Click **Add** button to include the path for the object in Selected Location panel.
- iii. Click **Done** to include the selected objects in the Manage Scan Location
- iv. To delete the object from the **Selected Location** panel, check the checkbox next to the scan location name and click **Delete**.

65. **Exclude from Scan:** Select objects to be excluded at the time of task execution or browse the path to an exclusion list. The **Exclude From Scan** tab will be enabled when **Detection** and **Masking/Field Encryption** are selected in Task Type field.



- i. The **Object Extension** field allow you to specify the objects extension (.txt, .csv, etc) that need to be excluded.

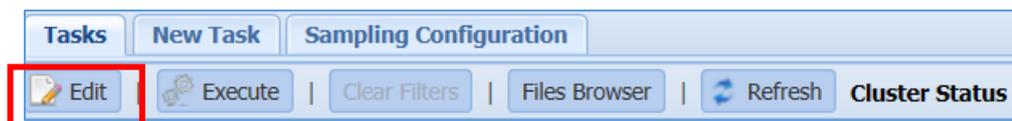
- ii. The **Extension List** field allow you either to **Browse** the path for the object or to select the object using the **Select Buckets** button. This functionality lets you specify the exclusion list of all objects that need not be included in the scanning.
- iii. The below panel list down information for both excluded **File Extension** and the **Scan Location** selected above.



***Note:** To decrypt tasks, ensure that appropriate roles have been assigned before executing the encryption task. For more details, refer to [Role Management](#)

- 66. After creating the required task, i.e., detection, masking, encryption, decryption or metadata discovery, click save to save the task to schedule later, or save and execute to execute it right away.

To edit an existing task, select the required task from the list of tasks on the Tasks screen and click **Edit**. A task can be edited using the same steps for task creation.



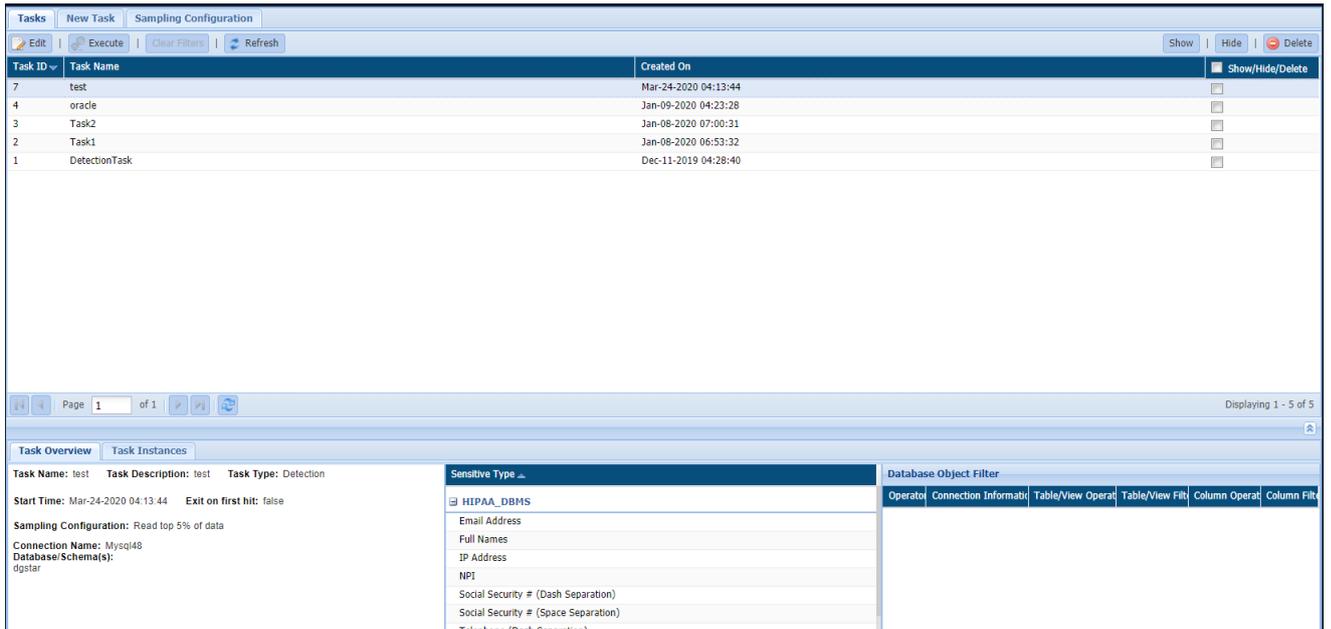
7.2.5.2 RDS/RedShift

DgSecure supports Detection and Masking in RDS/RedShift databases. Following sections outline the process of creating these tasks.

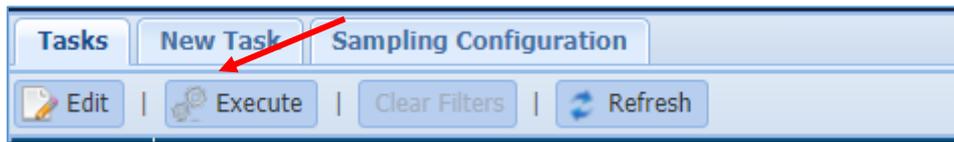
Detection

Perform the following steps to create a detection task.

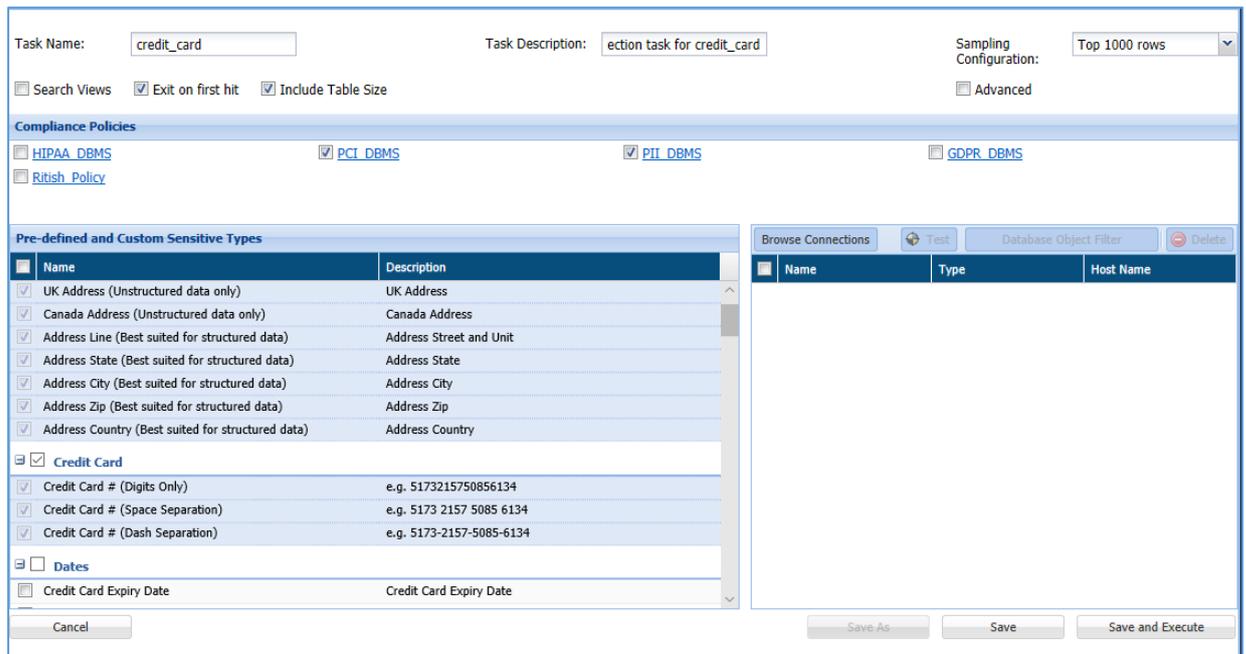
Go to **AWS > RDS/REDSHIFT > DETECTION > TASKS**.



1. To create a new task, click on the **New Task** tab.



The following image shows the user interface for creating a task.



2. Enter a unique **Task Name**. This field supports numeric and character values.

Task Name:	<input type="text" value="cre_card"/>
------------	---------------------------------------

3. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.

Task Description:	<input type="text" value="Detecton of CCNO"/>
-------------------	---

4. Check the **Search Views** option to detect the Views tables within the database as well as the tables linked with them. This option is available for the **Detection, Task Type**.
5. Check **Exit on First Hit** option to stop the scanning, when the first Sensitive Type is detected during the scan. This option is available for the **Detection, Task Type**.

67. DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

- i. Top 1000 Rows
- ii. Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. Check the **Advanced** checkbox to create a new sampling configuration.

You can also configure sampling through **AWS > RDS/REDSHIFT > DETECTION > TASKS > SAMPLING CONFIGURATION** tab.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Table Row Count Range: * To: Type: *

By: * Value: *

Table Row Count Range	Sample Type	Sample Value	Sample By	Actions
Default	Top	1000	Rows	 
1 to 200	Bottom	100	Rows	 

Note:

1. Top sampling option is not supported for Teradata, Aster DB and Sybase ASE connections.
2. Bottom sampling option is not supported for Teradata, Aster DB, Snowflake, Informix DB, Splice Machine, Sybase IQ and Sybase ASE connections.
3. Random sampling option is not supported for Snowflake, Informix DB, MySQL and Splice Machine connections. Random sampling is also not supported for Views in Sybase IQ.
4. For performance reasons, except for tables in DB2 schemas, random and bottom sampling is not recommended, where the tables can contain large volume of data.
5. If some unsupported sampling option is chosen, then best applicable sampling option will be applied. For more details regarding chosen sampling option, IDP logs can be checked.

a) Enter the name of the Sampling Configuration.

68. Enter the description for sampling.

69. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.

70. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling. Below are the options for advanced settings:

- **Table row count range:** Enter numeric value. This value states the starting range of the table from which the records will be sampled.
- **To:** Enter the numeric value. The value in this field states the ending range of the table till which the records will be sampled.
- **Type:** Select the sampling configuration type from the **Type** option. There are four options for sampling configuration:
 - i. **Top:** If you select the option **Top**, the sample data for the scan will be selected from the entries at the top of the table, based on the specified range.

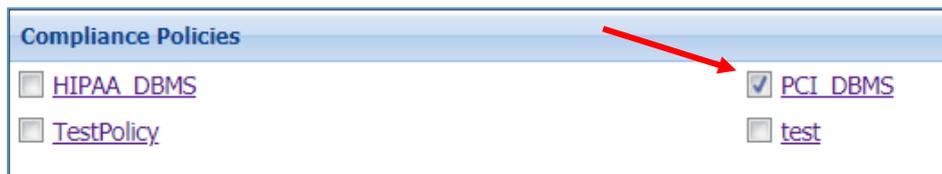
- ii. **Bottom:** If you select the option **Bottom**, the sample data for the scan will be selected from the entries at the bottom of the table. This does not mean the entries will be selected bottom up, instead depending on the range the last entries in the table will be taken to create a sample data for detection.
 - iii. **Random:** Entries from the table that correspond to the specified range, will be selected at random to create a sample set of data for detection.
 - iv. **Complete:** If all the entries in the selected tables of the database have to be scanned for sensitive types, select this option.
- **By:** To Specify how to pick data for sampling from the table, there are two ways:
 - i. **By Rows:** Select '**Rows**' from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select '**Percent**' from the drop-down, to sample a percentage of the data
 - **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.

71. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.

72. Click the **Save** button to save the changes.

73. Select the required policy under the **Compliance Policies** section. The Compliance Policy panel displays all the Pre-Defined and Customized Policies. Users can select any number of policies while creating or editing a task. Sensitive types associated with the selected policy can be viewed in the panel below this panel **Pre-Defined and Custom Sensitive Types**.

Selecting a policy is not a mandatory step, users can also proceed to select individual sensitive types. For more information, refer to section [Policy](#).



74. Select the required sensitive types for the scan from the **Pre-defined and Custom Sensitive Types** section. Refer to the following screenshot:

Pre-defined and Custom Sensitive Types	
Name	Description
ABA_test	
<input type="checkbox"/> ABA_test	ABA
Address	
<input type="checkbox"/> US Address	US Address
<input type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
<input type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input type="checkbox"/> Address State (Best suited for structured data)	Address State
<input type="checkbox"/> Address City (Best suited for structured data)	Address City
<input type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input type="checkbox"/> Address Country (Best suited for structured data)	Address Country
Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
Dates	

The **Pre-Defined and Custom Sensitive Types** panel lists down all the Sensitive Types. The Sensitive Type associated with the policy gets selected in the Pre-Defined and Custom Sensitive panel and cannot be removed from the scan, however any number of sensitive types can be added to the scan.

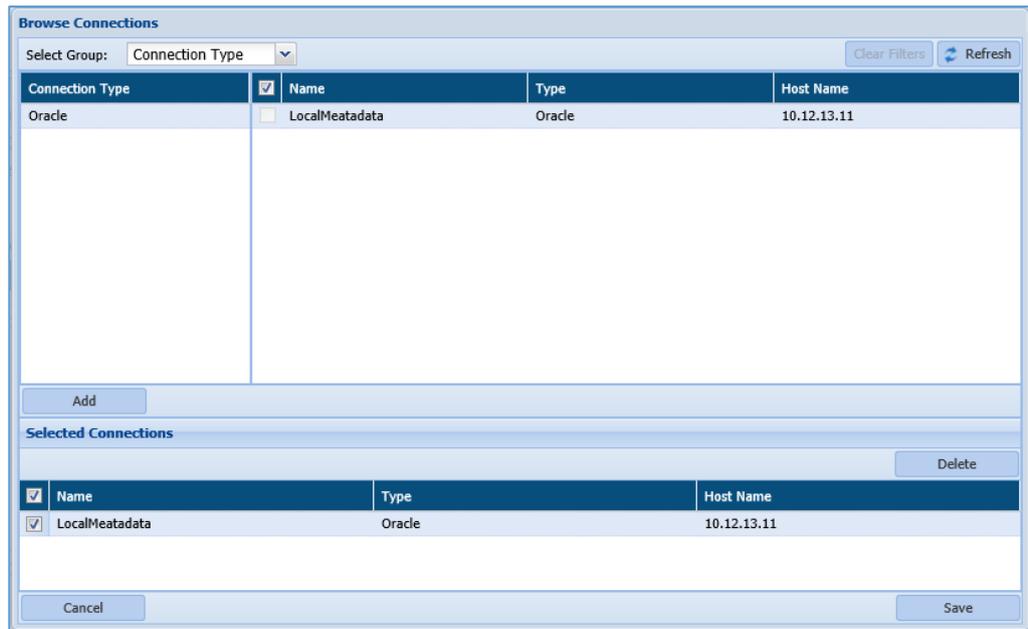
***NOTE:** If a policy is selected, user can still add more sensitive types to the scan but the sensitive types under the selected policies cannot be excluded from the scan.

75. The Database Connection panel lists down all the available RDBMS connections. Any number of connections can be selected for a task. This panel list down all the available connections. For details on how to create and manage connections refer [Connection Manager](#). Perform the following steps to choose a connection:

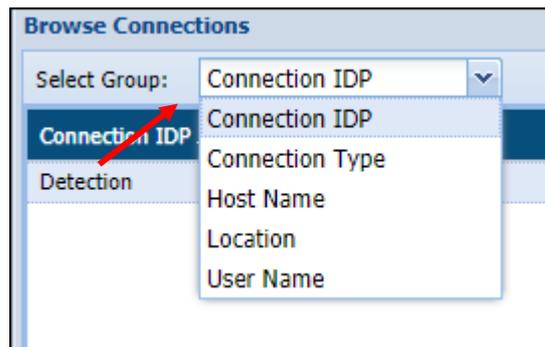
- a) Click Browse connections.

Browse Connections		
Name	Type	Host Name

76. The **Browse Connections** dialog box will be displayed. This screen categorizes connections based on user preferences.



- iii. Click on the Select Group dropdown and select the required option from the sub groups displayed on the left panel to sort the available connections.



The **Select Group** drop-down has five options:

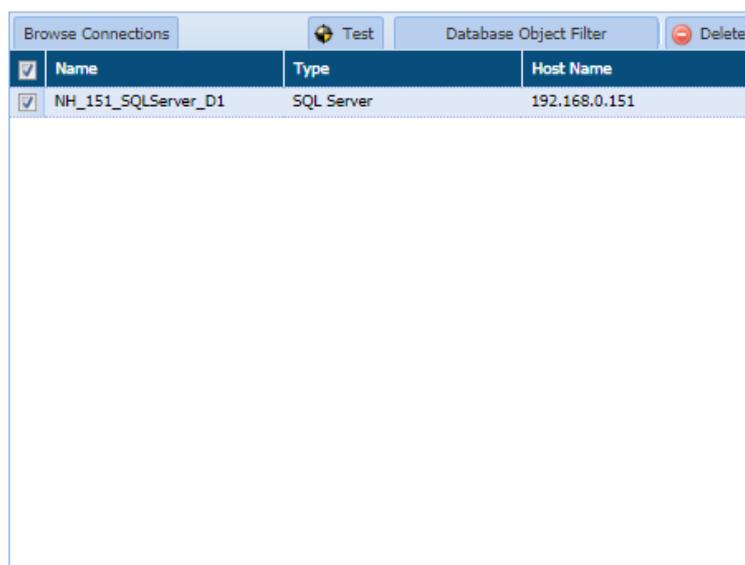
1. **Connection IDP:** Categorizes the available connections based on the types of IDPs available, i.e., Detection and Masking.
2. **Connection Type:** Categorizes the available connections based on the type of server connected to, i.e., Oracle, Teradata, SQL server etc.
3. **Host Name:** Categorizes the list of available connections based on Host Names.
4. **Location:** Categorizes the available connections based on the location of the target source system server, i.e., On-Premises and Cloud.
5. **User Name:** Categorizes the list of available connections based on the Usernames.

- iv. Click **Add** to include the selected database connection in the **Selected Connection** panel.
- v. Check the Selected connection. Click the **Save** button to include connections.

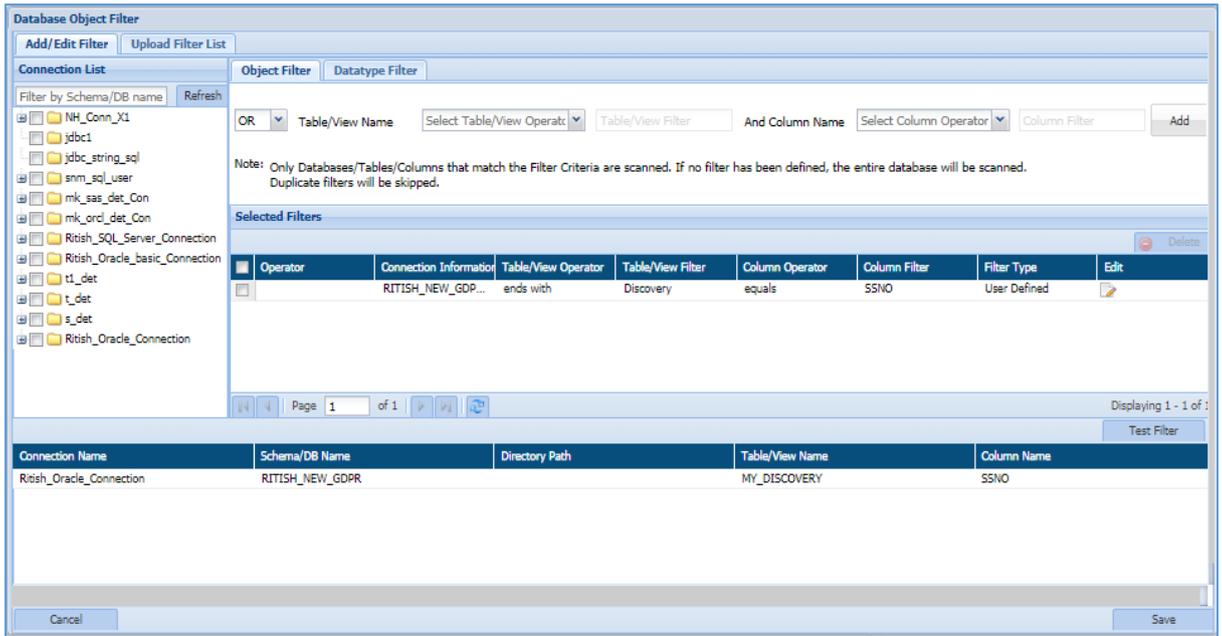
77. Click the **Test** button to test the listed RDBMS connection.

78. Click the **Database Object Filter** button to filter tables and/or columns. Once filters are defined, then only those databases/tables/columns that match the filter are scanned.

Check the checkbox next to the connection to enable the Database Object Filter.



You can **Add/Edit Filter** in two ways i.e. either by specifying in manually it manually or by uploading the filter list in the **Upload Filter List** tab.



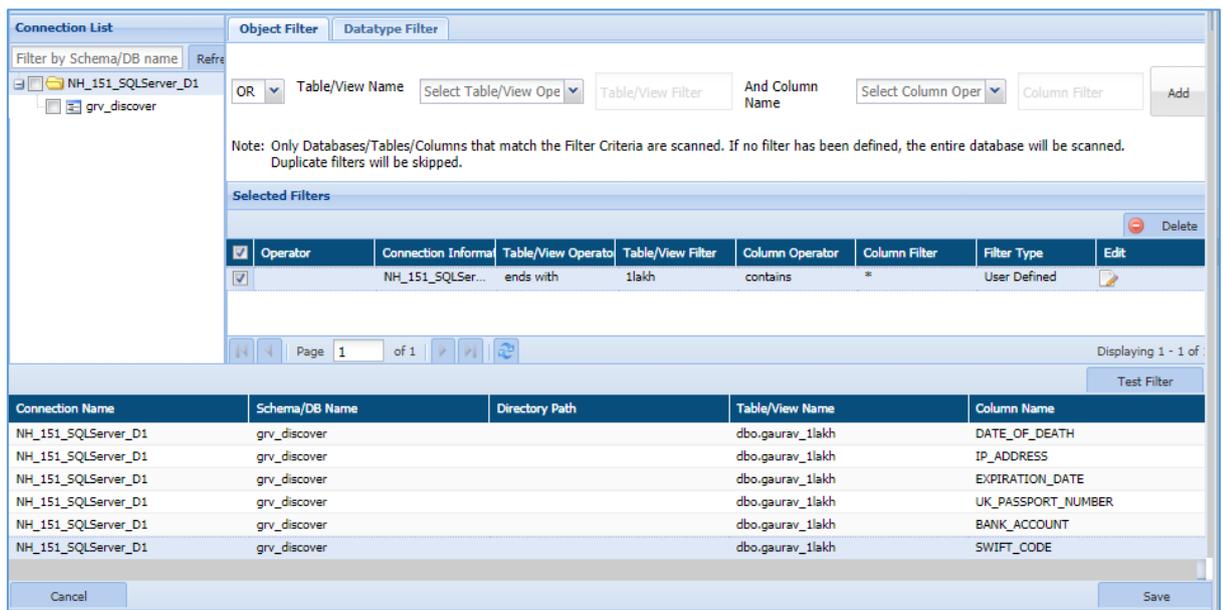
- **Add/Edit Filter:** This tab allows you to apply the filter for the selected connection. Once the filter has been applied then only those databases/tables/columns and datatypes that matched the criteria will be scanned.

There are two types of filter which can be applied.

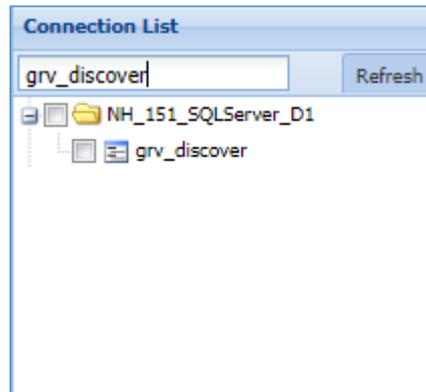
1. Object Filter
2. Datatype Filter

Object Filter

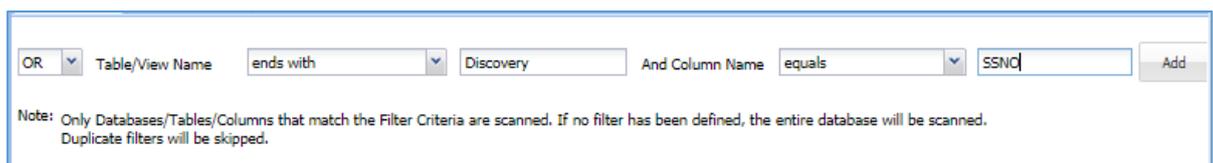
The Object Filter allows you to scan the database/tables/columns based on the defined filter. To apply an Object Filter, perform the following steps:



- a) Select the connection from the Connection List panel or enter the DB/Schema name in the Filter by Schema/DB Name textbox.



- b) Apply the **Object Filter** in the top panel by specifying the **Operator, Table/View** and **Column** name.



For example, in the above image the object filter specifies the table name should end with 'Discovery' and the selected table should contain SSNO column.

There are eight types of Operators based on which you can select the Table and Column name.

1. **Equals:** This operator will check whether the given table/column name exist in the selected database/table. It will return the matched records if the condition is fulfilled.
2. **Not Equal to:** This operator will return all the records except the given table/column name.
3. **Contains:** This operator will return only those tables/columns which matched the given criteria.
4. **Does not contain:** The functionality of this operator is similar to the **Not Equal to** operator, since it returns all the records except the given table/column name.
5. **Starts with:** This operator will return all the tables/column whose name starts with the given criteria.

6. **Does not start with:** The functionality of this operator is similar to the **Does not contain** and **Not Equal to**, since it will return all the tables/column name except the one which has been entered.
 7. **Ends with:** This operator will return all the tables/column name whose name ends with the given input.
 8. **Does not ends with:** The functionality of this operator is similar to the Does not contain and Not Equal to, since it will return all the tables/column name except the one which has been entered.
- c) Click the **Add** button to add the filter in the **Selected Filters** panel.
9. The **Selected Filters** panel will list down all the user defined filters. It displays information about the filters such as Connection Name, Table/View Operator, Table/View Filter, Column name, etc.

Operator	Connection Information	Table/View Operator	Table/View Filter	Column Operator	Column Filter	Filter Type	Edit
<input type="checkbox"/>	RITISH_NEW_GDP...	ends with	Discovery	equals	SSNO	User Defined	

Page 1 of 1 | Displaying 1 - 1 of 1

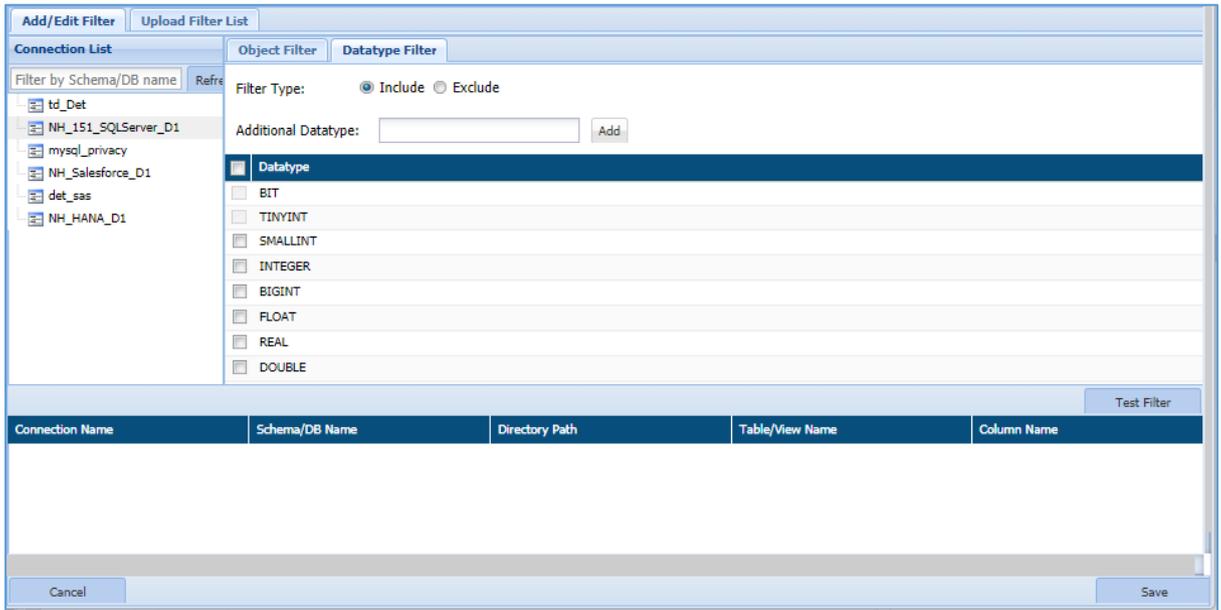
- To edit a filter, click the button in the **Edit** column. This functionality allows you to re-select the Operator, Table/View and Column name.
 - To delete a filter, check the checkbox for that filter and click the **Delete** button. This button will get enabled when you check the checkbox for any filter.
10. Click **Test** button to test the filter. It lists down the result matching the filter criteria.

Connection Name	Schema/DB Name	Directory Path	Table/View Name	Column Name
Rish_Oracle_Connection	RITISH_NEW_GDPR		MY_DISCOVERY	SSNO

11. Click **Save** button to make the changes effective.
12. Click **Cancel** button, if you do not want to save the changes.

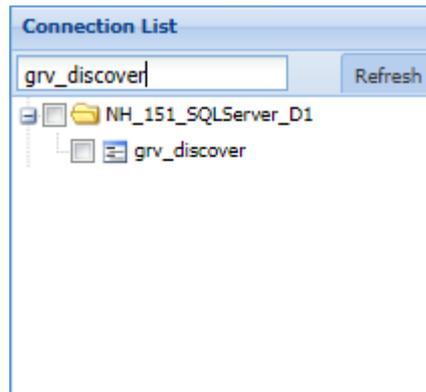
Datatype Filter

This tab allows you to specify the datatypes to be included or excluded while scanning based on the selection.



To apply Datatype Filter, perform the following steps:

- a) Select the connection from the **Connection List** panel or enter the Database/Schema name in the **Filter by Schema/DB Name** textbox.



- b) Select the **Filter Type** from the given option. By default, **Include** filter type is selected.
 - **Include:** This option allows to select the datatype that need to be included while scanning.
 - **Exclude:** This option allows to select the datatype that need to be excluded while scanning.
 - **Additional Datatype:** This field allows you to add a datatype if it's not present in the Data Type panel. Once that datatype is entered, click **Add**

button to specify it in the DataType panel.

Filter Type: Include Exclude

Additional Datatype:

13. Check the checkbox for the datatype which you want to include or exclude based on the selection of **Filter Type** in above panel.

Filter Type: Include Exclude

Additional Datatype:

Datatype	
<input type="checkbox"/>	NUMERIC
<input type="checkbox"/>	DECIMAL
<input checked="" type="checkbox"/>	CHAR
<input checked="" type="checkbox"/>	VARCHAR
<input type="checkbox"/>	LONGVARCHAR
<input type="checkbox"/>	DATE
<input type="checkbox"/>	TIME
<input type="checkbox"/>	TIMESTAMP
<input type="checkbox"/>	BINARY

For example, in the above image the CHAR and VARCHAR datatypes are included in the scanning process since Filter Type specify the option as 'Include'. If 'Exclude' is selected as Filter Type then selected CHAR and VARCHAR datatype will be excluded from the scanning process.

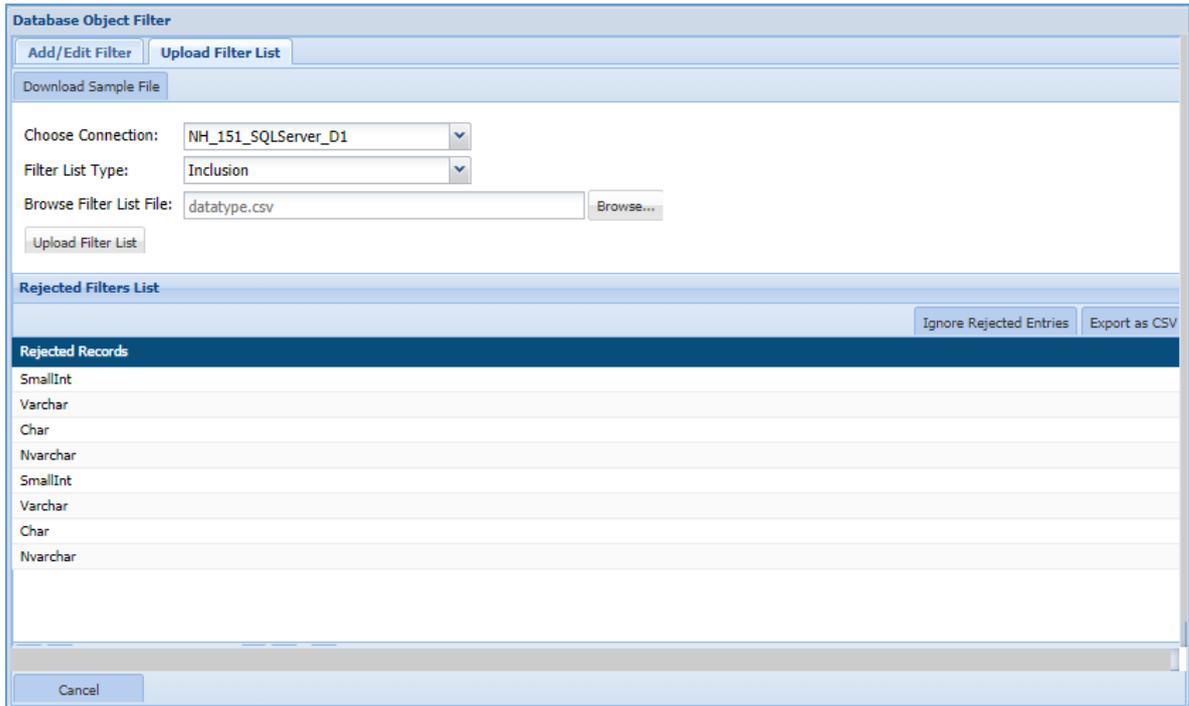
14. Click **Test** button to test the whether any column in the database contains the selected datatype. This functionality will list down all the columns that contain the selected datatype.

Connection Name	Schema/DB Name	Directory Path	Table/View Name	Column Name
Ritish_Oracle_Connection	RITISH_NEW_GDPR		MY_DISCOVERY	SSNO

15. Click **Save** button to make the changes effective.

16. Click **Cancel** button if you do not want to save the changes.

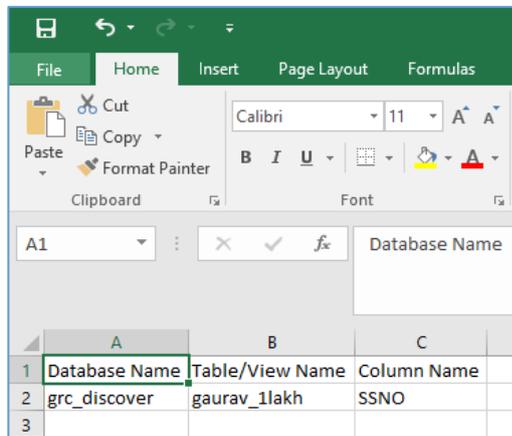
- **Upload Filter List:** This tab allows you to upload a file containing the list of all columns.



To upload the filter list, perform the following steps.

- a) Select the connection from the **Choose Connection** drop-down.

17. The **Download Sample File** button will be enabled once a connection has been chosen. Enter the Database, Table/View and Column Name in the sample file.



18. Select the **Filter List Type** from either 'Inclusion' or 'Exclusion'. This functionality allows you to specify whether to include or exclude the Database, Table/View and Column name.

Filter List Type: Exclusion

Inclusion

Exclusion

19. Click the **Browse** button to search and upload the saved sample file containing the list of Database, Table/View and Column name which will be either excluded or included.

Browse Filter List File: FiltersSampleFileForSQLServer (2).csv Browse...

20. Click **Upload Filter List** button to add the defined filters in the **Selected Filters** panel under **Add/Edit Filter** tab.

Selected Filters							Delete
Operator	Connection Informa	Table/View Operato	Table/View Filter	Column Operator	Column Filter	Filter Type	Edit
<input type="checkbox"/>	NH_151_SQLSer...	equals	gaurav_ilakh	equals	SSNO	Include	

Displaying 1 - 1 of

21. The **Rejected Filters List** will display all the entries from the uploaded list which are not in a proper format as specified in the sample file.

Rejected Filters List		Ignore Rejected Entries	Export as CSV
Rejected Records			
grc_discover,,SSNO			
srch_ccno.emp_d,			
discover_city,Emp_detail,City			

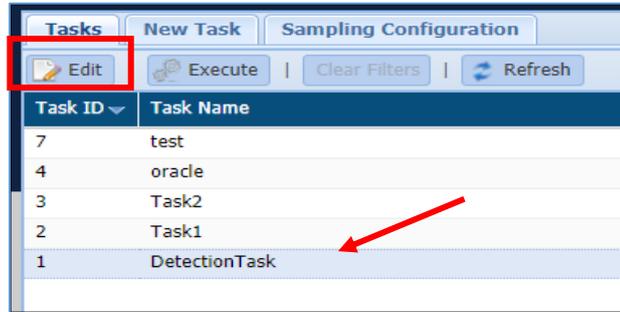
- To download the list of rejected entries, click **Export as CSV** button. A downloaded file will contain the list of rejected entries which were not formatted as per specified format.
- To remove the rejected entries from the **Rejected Filters List** panel, click the **Ignore Rejected Entries**. This functionality will remove all the rejected entries from the panel.

22. Click **Save** button to make the changes effective.

23. Click **Cancel** button if you do not want to save the changes.

24. Click **Save**, if you want to execute the task later else click **Save and Execute**. The results of the task and its status can be viewed under **RDBMS>DETECTION>RESULTS** (Refer to Section [Results](#)).

To edit an existing task, select the required task from the list of tasks on the Tasks screen and click edit. A task can be edited using the same steps for task creation.



***Note:** Task defining features such as Incremental, Exit on First Hit, Search View and the Task Type cannot be edited. Some of the originally selected options can be modified but new options cannot be added.

Masking Task

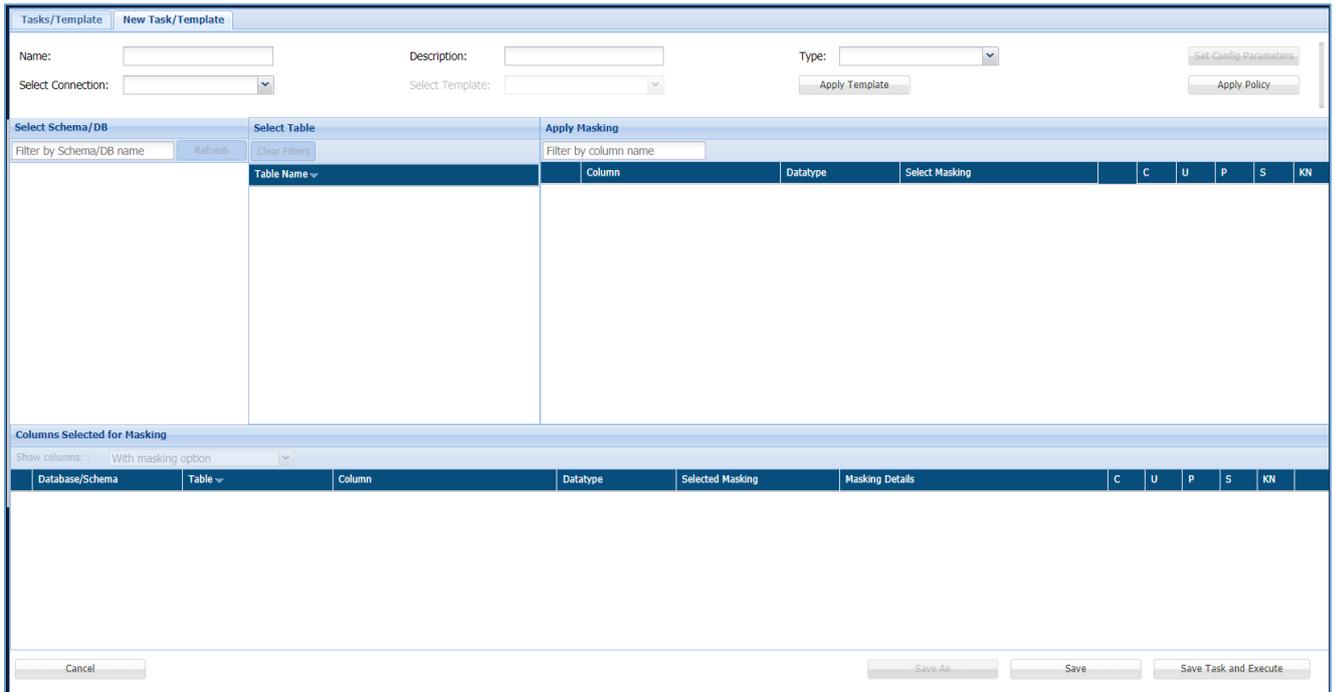
Perform the following steps to create a masking task:

Access the NEW TASKS/TEMPLATE screen, click **AWS > RDS/REDSHIFT > MASKING > TASKS/TEMPLATES > NEW Task/Template** tab.

1. To create a new task or template for masking tasks, click on the **New Task/Template** tab.



25. The following image shows the user interface for creating a task.



26. Enter a unique **Task Name**. This field supports numeric and character values.

Name:

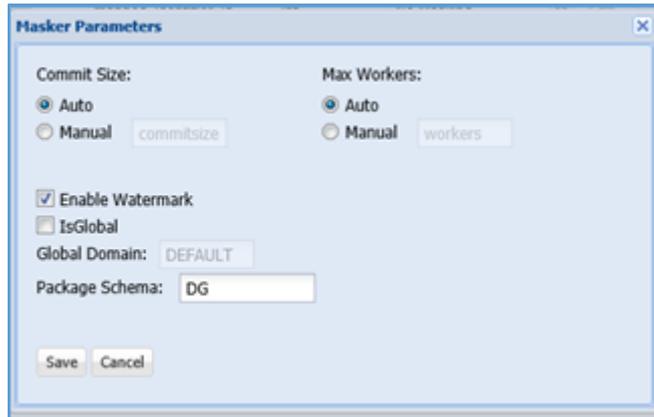
27. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.

Description:

28. Click on the **Type** dropdown. Select **Task** to create a masking task or **Template** to create a template for masking tasks.

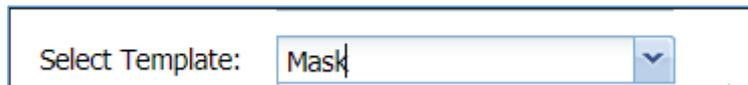
Type:

29. Click the **Set Config Parameters** button. It will display the environment settings for the masking operations such as **Commit Size**, **Max Workers**, **Package Schema**, **IsGlobal** and **Enable Watermark**.



30. Click **Select Connection** to view the list of the available database connections for masking. For details on how to create a connection for masking refer to [Connection Manager](#)

31. The **Select Template** option provides a list of user created templates for masking. This option will be greyed out if no templates have been created. To create a new template select the Type as Template in step 6 and follow the next steps for creating a task and save the template. Once created, the new template will appear in the select template dropdown on the new task screen.

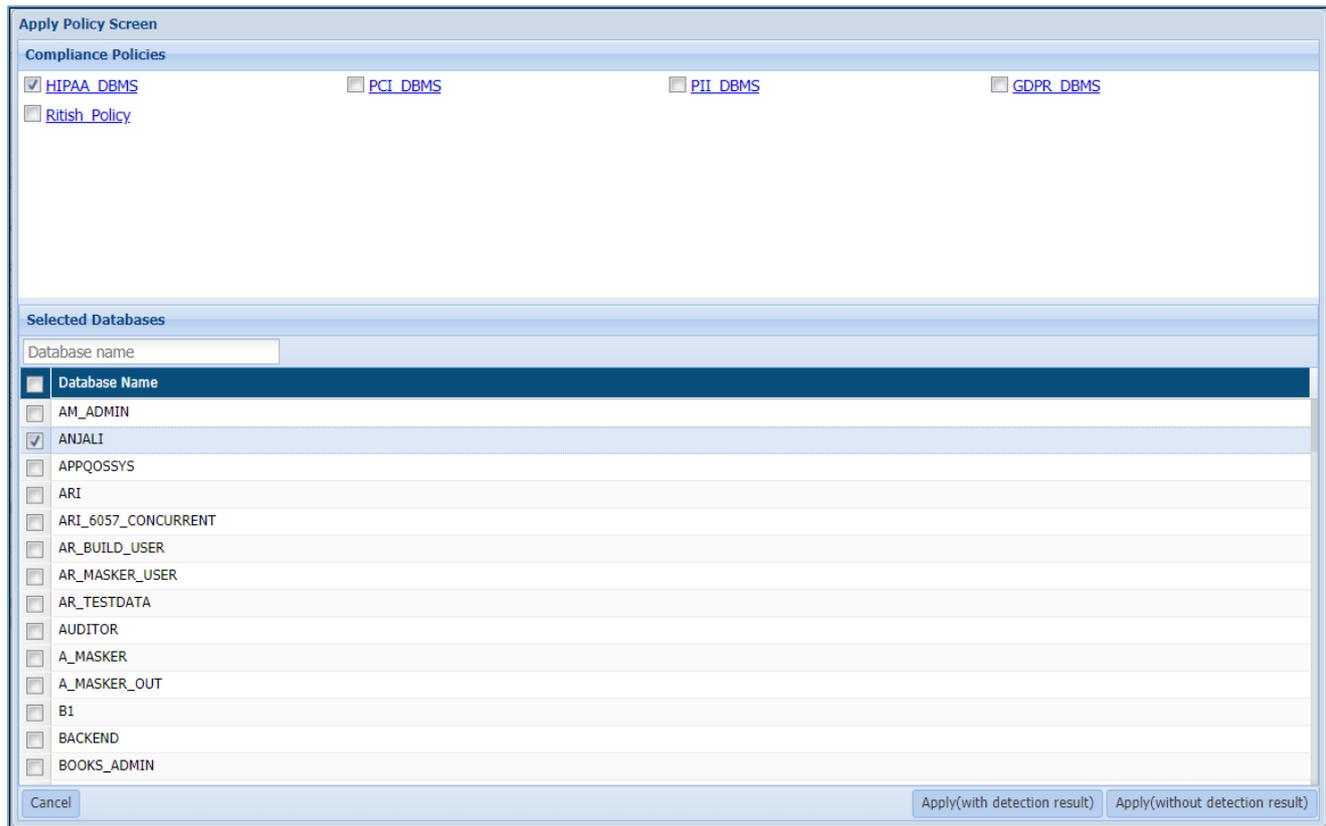


***Note:** Templates are connections specific and only one template can be selected at a time.

32. Click the **Apply Template** button to apply the selected template in **Select Template** drop-down.

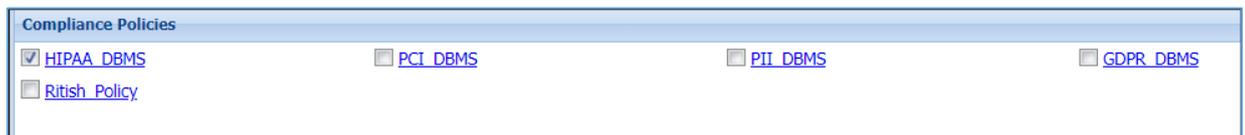


33. Click the **Apply Policy** button to choose the available Compliance Policies.

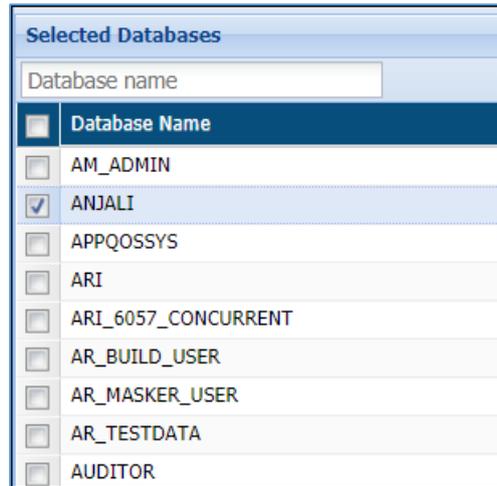


Perform the following steps to apply a policy to the task or template:

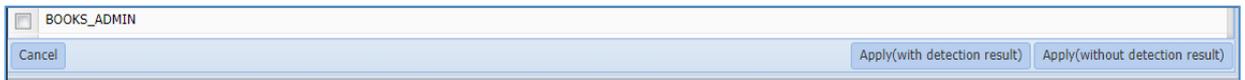
- a) Select the required policy/policies under the **Compliance Policies** panel.



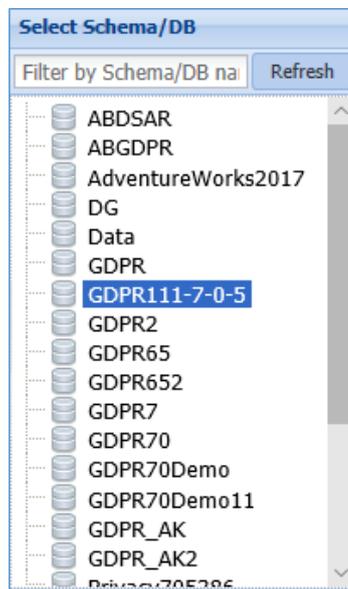
34. Select the databases you need to mask by checking the checkbox next to the Database Name.



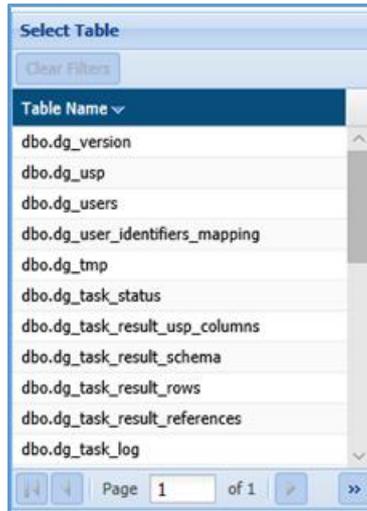
35. Click the **Apply** (with detection result) button to apply the selected policy with detection results or click the Apply (without detection result) button to apply the selected policy without results. Click Cancel to redo your selection.



36. The Select Schema/DB pane will display the list of databases or schemas for a selected connection in **Select Connection** drop-down.

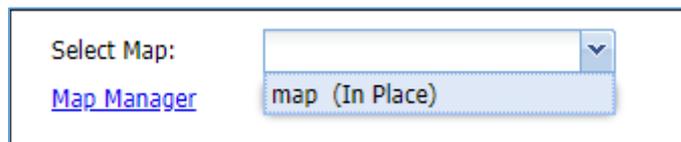


37. Select the table from the Select Table pane. This pane lists all the tables for the selected database or schema.

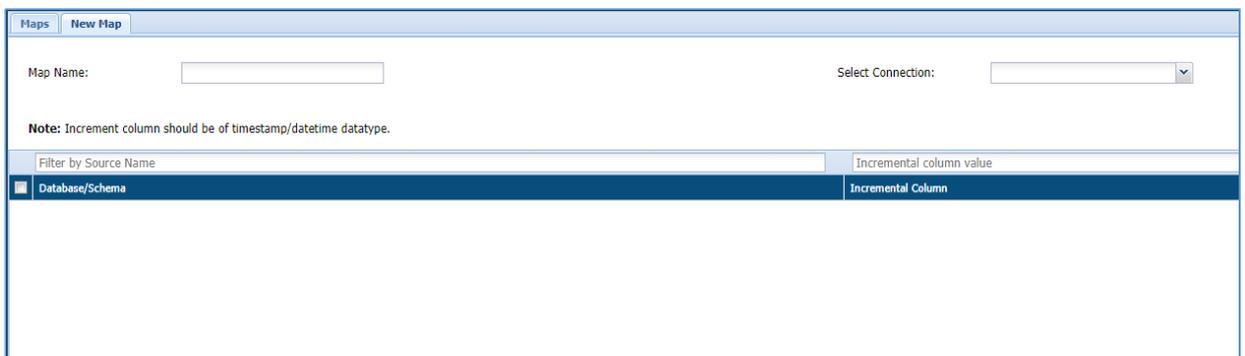


38. Check the Incremental checkbox to apply incremental masking to the database. This feature is useful to mask new values added in a database after masking has been executed on it. Only the new entries will be masked, thus, the time taken for masking would be reduced. Perform the following steps to make your masking task incremental:

- i. Check incremental checkbox.
- ii. The Select map Dropdown will appear. Select the required map. Maps define incremental columns within a database which are considered for indexing the data in order to mask the new rows added to the database.



- iii. To add a map click on the Map Manager link
- iv. The Map Manager Screen can also be accessed from the menu bar. The following image displays the Map Manager interface:



- v. Enter a Map Name.

Map Name:

- vi. Select a connection that contains the database on which incremental masking has to be applied.

Select Connection:

- Ritish_Oracle_masker**
Host:192.168.0.163 Type:Oracle DB:ritishmasker
- Ritish_Masker_TNS**
Host:N.A. Type:Oracle DB:ritishmasker
- orcl_neha**
Host:192.168.0.163 Type:Oracle DB:nehamaskuser

Incremental column value

Incremental Column

- vii. Select the required Database/Schema.

Filter by Source Name

<input type="checkbox"/>	Database/Schema
<input type="checkbox"/>	VCON
<input type="checkbox"/>	SNCF
<input type="checkbox"/>	GDPR_TEST
<input checked="" type="checkbox"/>	NEHAMASK163
<input type="checkbox"/>	DG_VIK_655
<input type="checkbox"/>	DGCONTROLLER_SANDEEP
<input type="checkbox"/>	AR_TESTDATA
<input type="checkbox"/>	VIKRAM
<input type="checkbox"/>	A_MASKER_OUT
<input type="checkbox"/>	A_MASKER
<input type="checkbox"/>	DGSHAREPOINT
<input type="checkbox"/>	JSON
<input type="checkbox"/>	NEHA777

- viii. Enter the name of the incremental Column to the corresponding database and ensure that it is a timestamp or date-time datatype.

Database/Schema	Incremental Column
<input type="checkbox"/> JSON_XML_SCRIPT	
<input type="checkbox"/> M_MASKING	
<input type="checkbox"/> DG_USER	
<input type="checkbox"/> NEHA	
<input type="checkbox"/> GDPR_72	
<input type="checkbox"/> BUILD_N	
<input type="checkbox"/> SNCF_MASKER	
<input type="checkbox"/> VCON	
<input type="checkbox"/> SNCF	
<input type="checkbox"/> GDPR_TEST	
<input checked="" type="checkbox"/> NEHAMASK163	DATE_OF_BIRTH
<input type="checkbox"/> DG_VIK_655	

- ix. Save the map. It will be available in the Select Map dropdown on the new task/template screen.

NOTE: To apply incremental masking the following criteria must be met:

1. Create a map to execute incremental masking and apply it to the task.
2. Relational tables cannot be included.
3. Column holding date, time stamp or a numeric incremental value should be added as incremental column
4. The incremental column cannot be masked.

39. The Apply Masking pane display the list of all the columns for the selected table in Select Table pane. For detailed information on all the available masking options in DgSecure refer to [Masking Options](#). Perform the following steps to apply masking option to the columns in the database:

- a) Select the masking option from the drop-down against the column entry. You can apply the masking to the selected column by checking on the checkbox corresponding to the column name.

Apply Masking										
Filter by column name										
Column	Datatype	Select Masking		C	U	P	S	KN	SL	
checksum	bigint	No Masking	▼	Edit	<input type="checkbox"/>					
connectionid	int	No Masking	▼	Edit	<input type="checkbox"/>					
constraintCols	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
dbName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
🔑 groupId	int	No masking available		Edit	<input type="checkbox"/>					
keyType	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
🔑 ResultTableId	int	No masking available		Edit	<input type="checkbox"/>					
schemaName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
tableName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					

40. To further enhance the results of masking following options can be combines with the masking options:
- i. **Consistent:** Masks the data of the selected table with consistent values. E.g. If the name John is masked as FVGB, throughout the table it will be masked the same way.
 - ii. **Unique:** Masks each entry with a unique value.
 - iii. **Persistent:** Similar to Consistent, however in this case the same values will be masked consistently across all the tables of the database.
 - iv. **Sync:** This options allows tracking of masked entries in different tables if any of the same entries are found in other tables they will also be masked.
 - v. **Keep Null:** The cells containing null values are kept null even after masking.
 - vi. **Stateless:** Persistently masks the data without creating any metadata. No extra space is required to create masking tables.

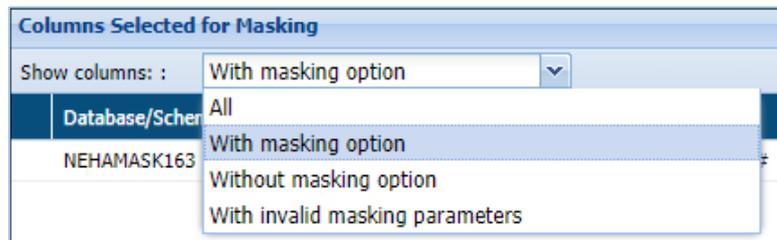
Column	Datatype	Select Masking		C	U	P	S	KN	SL
checkSum	bigint	No Masking	▼	Edit	<input type="checkbox"/>				
connectionid	int	No Masking	▼	Edit	<input type="checkbox"/>				
constraintCols	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
dbName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
groupId	int	No masking available		Edit	<input type="checkbox"/>				
keyType	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
ResultTableId	int	No masking available		Edit	<input type="checkbox"/>				
schemaName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				
tableName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>				

41. View the selection under the **Columns Selected for Masking** panel.

Source DB/Schema	Destination DB/Sch	Table	Column	Datatype	Selected Masking	Masking Details	C	U	P	S	KN	SL
Data	Data	dbo.Data2	Description	nvarchar(M...	Custom	Mask by function [DG].CUSTO...	<input type="checkbox"/>					

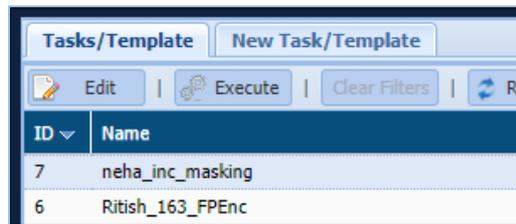
42. Select the option from the Show Column drop-down. There are four options:

- a) All
- b) With masking option
- c) Without masking option
- d) With invalid masking parameters



43. Click save to save the task or template or save and execute to execute the masking task.

44. To edit a task, select the task from the Tasks/Template tab and click edit. A task can be edited using the same steps for creating a new task.



7.2.6 Azure

7.2.6.1 Azure Blob/Data Lake

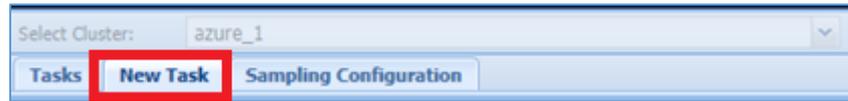
Detection, Metadata discovery, Masking, Encryption and Decryption operations can be performed on Azure Blob/Data Lake systems. To create a task to for these operations, perform the following steps.

7.2.6.1.1 Create an Azure Blob/Data Lake task

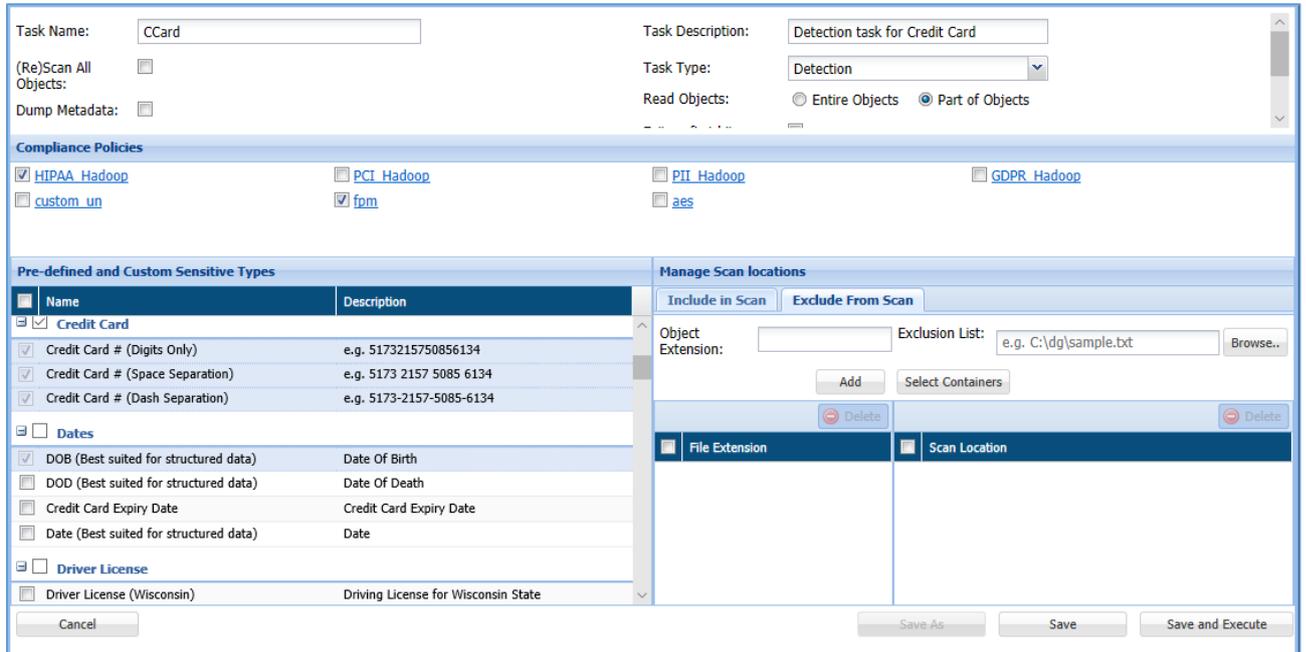
1. Go to **AZURE > AZURE BLOB/DATA LAKE > TASKS**. Select the cluster.



2. To create a new task, click on the **New Task** tab.



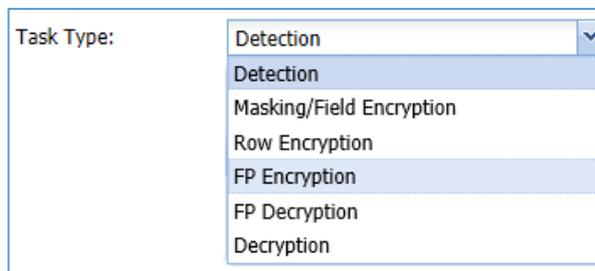
The following image shows the user interface for creating a task.



45. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.



46. Choose a **Task Type** from given options.



The list of available task types and the options specific to that task type are explained below:

Detection

Detection tasks scan the target source system for sensitive data elements. The following options are specific to the detection tasks:

- a) **Scan All Files:** Check this checkbox to scan all the available objects for a given connection between the dates specified in **Objects Modified After** and **Objects Modified Before** drop down.

The **Objects Modified After** and **Objects Modified Before** drop down will be visible when **(Re)Scan All Objects** checkbox is checked.

- 47. **Dump Metadata:** Check **Dump Metadata** option to remove the metadata files after scanning.

- 48. **Advanced Options:** Define a batch size for scanning the source system in batches. Click on the **Advanced Options** button to define the batch size of the data.

Define the number of files per batch in the **Batch Size(Files)** option or check the **Auto Batch Size** option to enter the minimum batch size in **Batch min size(MB)** option.

- 49. **Read Objects:** Choose to read the Entire Objects or a Part of Objects at random.

50. **Exit on first hit:** Check this checkbox to report the table or database as sensitive, at the first event of detection of a sensitive type.

Exit on first hit:

51. Sampling Configuration

Sampling Configuration: Top 1000 rows Advanced

DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

- Top 1000 Rows
- Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. To create a new sampling configuration perform the following steps:

i. Go to **AZURE > AZURE BLOB/DATA LAKE > TASKS > SAMPLING CONFIGURATION** tab or check the Advanced checkbox.

File Size Range (Bytes)	Sample Value	Sample By	Actions
Default	1000	Rows	
1 to 100	50	Percent	

ii. Enter the name of the Sampling Configuration.

Name: *

iii. Enter the description for the sample.

Description:

iv. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.

Set Sampling Config as Default:

v. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling.

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

Below are the options for advanced settings:

- **File Size Range (Bytes):** Enter the starting range for the sample in bytes.
 - **To:** Enter the ending range for the sample.
 - **By:** To Specify how to pick data for sampling from the source system, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
 - **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling **By-Rows** is selected and denotes the percentage of sampling **By-Percent** is selected.
- vi. After setting up the required details for configuration, click **Add** to add the user-defined sampling configuration to the list.
- vii. Click the **Save** button to save the changes.

To proceed for remaining steps, go to step [Compliance Policy](#).

Masking/Field Encryption

Masking or Field Encryption hides sensitive data in the target source system by replacing it with a system generated value. DgSecure provides various masking options to ensure the usability of data after it has been protected. Masking/Field Encryption can be applied on structured as well as unstructured data. For more details refer to [Masking Options](#) . Following options are available to customize a Masking/Field Encryption task:

- a) **(RE)Scan All Files:** Scan all available files for a given connection.
- b) **Structured:** Check the structured option if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#)

To proceed for remaining steps, go to step [Compliance Policy](#).

Row Encryption

Row encryption is ideal for unstructured data. Following additional option is available:

- a) **Scan All Files:** Scan all available files for a given connection.

To proceed for remaining steps, go to step [Compliance Policy](#).

FP Encryption

FP or Format Preserving encryption can only be executed on structured files. Following additional options are available:

- a) **Scan All Files:** Scan all available files for a given connection.
- b) **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step [Compliance Policy](#).

FP Decryption

FP or Format Preserving Decryption is used to decrypt the data encrypted by an FP Encryption task. Following additional options are available

- a) **Scan All Files:** Scan all available files for a given connection.

52. **Structured:** This option is checked by default. For details about how to create a structure refer to the section [Structure Management](#).

To proceed for remaining steps, go to step [Compliance Policy](#).

Decryption

Decryption tasks are used to decrypt data on which Encryption has been performed.

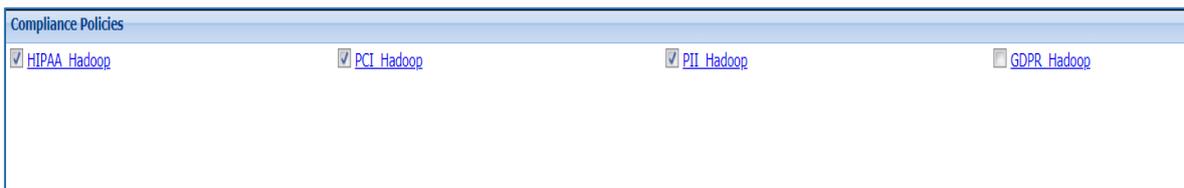
- a) **Scan All Files:** Scan all available files for a given connection.

- 53. **Structured:** Check the structured if the data that needs to be masked is structured. For details about how to create a structure refer to the section [Structure Management](#)

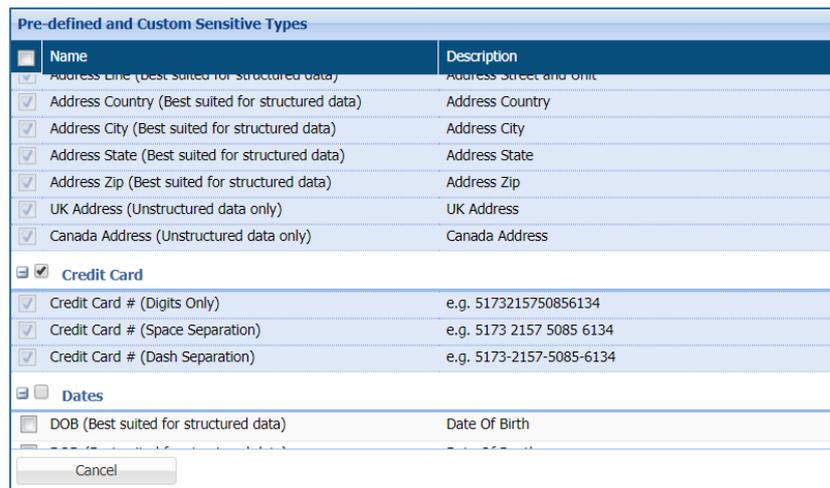
To proceed for remaining steps, go to step [Compliance Policy](#).

- 54. **Compliance Policy** can be set with all the task types in HDFS except Metadata Discovery. For more details about compliance policies, refer to section [Policy](#) . After selecting the required options, perform the following steps:

- a) Select the required policies.



- 55. **Pre-defined and Custom Sensitive Types** are available for all task types in HDFS except Metadata Discovery. Select the required sensitive types.



***NOTE:**

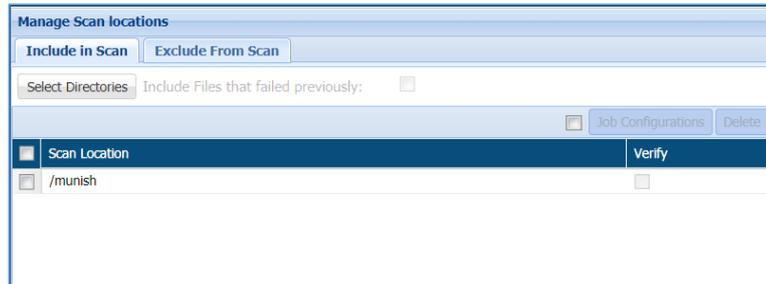
- Row Encryption uses default row encryption configuration for masking. This will mask all the entries of the row and is best suited to unstructured datatypes such as text files.
- FP Encryption uses default encryption configuration to protect the original data format. This option is best suited to structured datatypes.
- FP Decryption can only be executed on data that has been encrypted using FP Encryption.
- Decryption can be executed on data that has been encrypted using FP Encryption or Field Encryption.

56. If you select Masking/Encryption as the Task Type, Protection Option and Consistent fields are also available. Select the required Protection Option for the selected sensitive types. For details about all the masking options available in HDFS refer to [Masking Options](#)

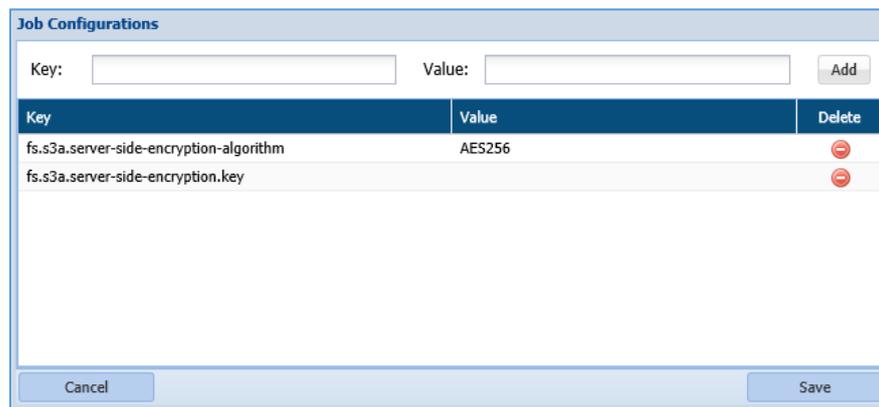
Pre-defined and Custom Sensitive Types				
<input type="checkbox"/>	Name	Description	Protection Option	Consistent
<input type="checkbox"/>	Address			
<input type="checkbox"/>	US Address	US Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Line (Best suited for structur...	Address Street and Unit	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Country (Best suited for stru...	Address Country	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Address City (Best suited for structur...	Address City	AES Encryption	<input type="checkbox"/>
<input type="checkbox"/>	Address State (Best suited for structu...	Address State	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Address Zip (Best suited for structure...	Address Zip	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	UK Address (Unstructured data only)	UK Address	Select Protection Option	<input type="checkbox"/>
<input type="checkbox"/>	Canada Address (Unstructured data o...	Canada Address	Select Protection Option	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card			
<input checked="" type="checkbox"/>	Credit Card # (Digits Only)	e.g. 5173215750856134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134	Random (Credit Card Numbers)	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134	Random (Credit Card Numbers)	<input type="checkbox"/>

57. **Manage Scan Locations:** Specify which directories to scan.

- a) **Include in Scan:** Click **Select Directories** to choose the directory to perform the task.



- viii. **Include Files that Failed Previously:** check this option to include all those files which were previously not scanned due to some exception. This option will appear when **Detection** is selected in **Task Type** field.
- ix. **Delete Input Files on Job Completion:** check this option to delete all the original input files, which were included, post masking. This option will appear when **Masking/Field Encryption, Row Encryption, FP Encryption** is selected in **Task Type** field.
- x. **Job Configuration:** check the checkbox to setup the parameters list. The value for job configuration will contain the pre-defined key and the value. If you have not specified any **Job Configuration**, then default parameter list will be executed.

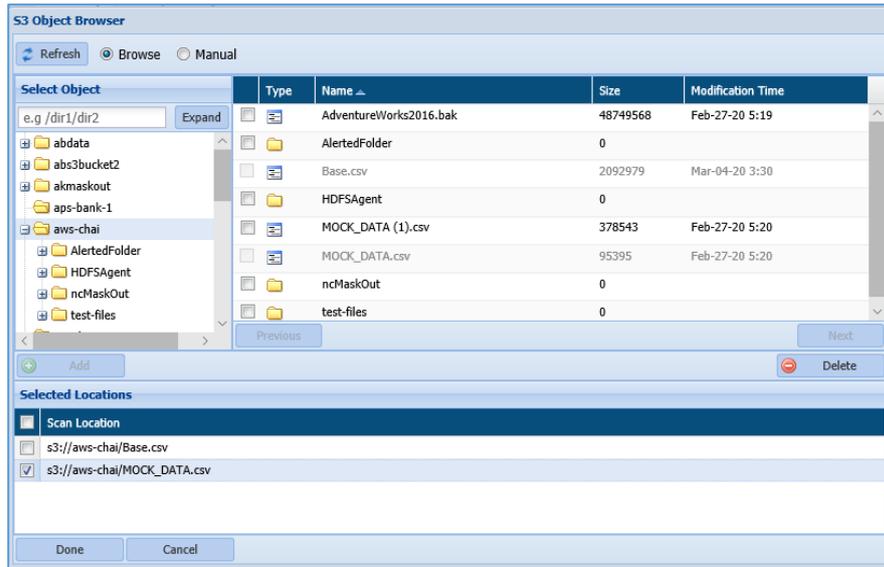


There are two ways in which buckets can be selected in **S3 Object Browser**:

- Browse
- Manual

To include buckets for scanning, click **Select Bucket** button, perform the following steps:

Browse: This option lets you browse the objects from **Select Object** panel. To select the objects, perform the below steps:



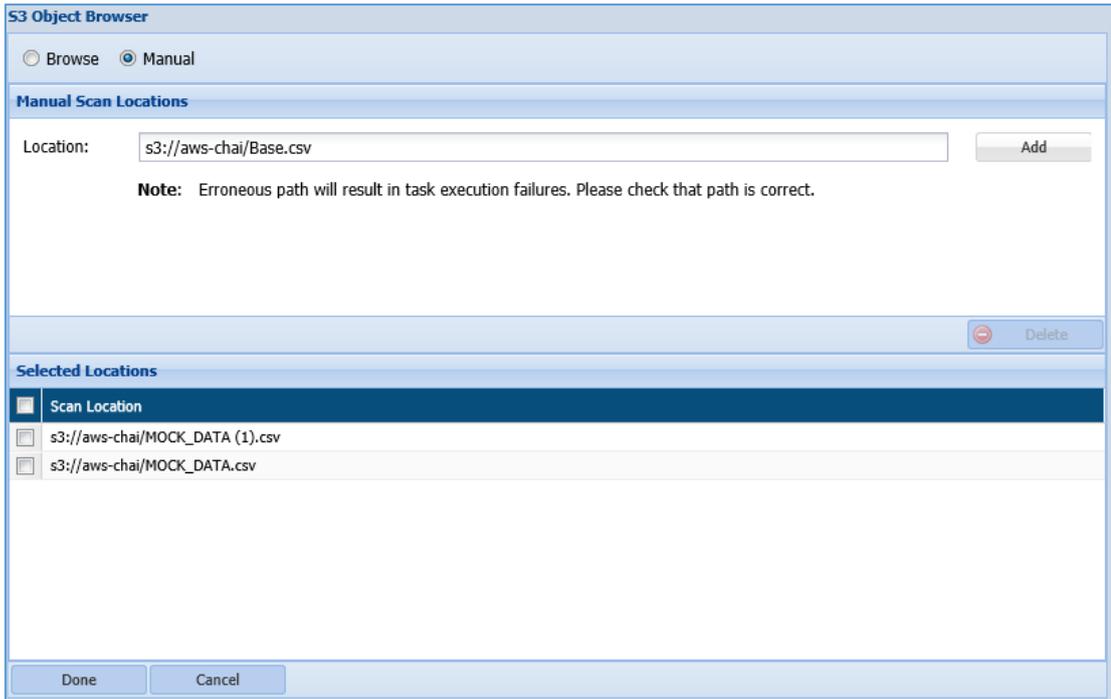
- i. Select the folder from the directory to view the objects for that folder. All the objects will be displayed in the right panel.
- ii. Select the object in the right panel by checking the checkbox and click **Add** button. This functionality will include the objects in the **Selected Locations** panel.



To delete the objects from the **Selected Location** panel, check the checkbox next to the **Scan Location** name and click **Delete** button. The **Delete** button will be enabled, once the **Scan Location** checkbox is checked.

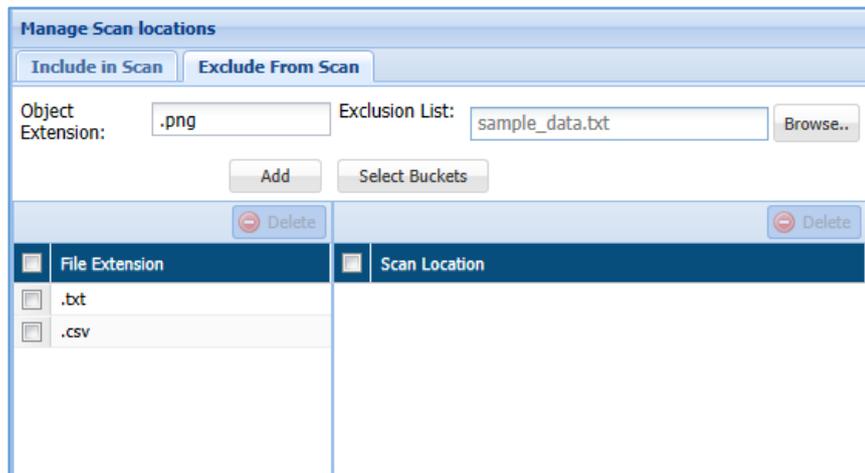
- iii. Click **Done** button to include the objects in **Manage Scan Location**.

Manual: This option allows you to select the object from the bucket manually. Perform the below steps:



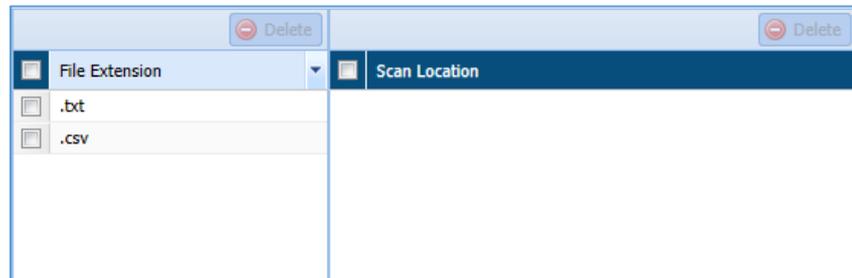
- i. Enter the path for the scan location that you want to include.
- ii. Click **Add** button to include the path for the object in Selected Location panel.
- iii. Click **Done** to include the selected objects in the Manage Scan Location
- iv. To delete the object from the **Selected Location** panel, check the checkbox next to the scan location name and click **Delete**.

58. **Exclude from Scan:** Select objects to be excluded at the time of task execution or browse the path to an exclusion list. The **Exclude From Scan** tab will be enabled when **Detection** and **Masking/Field Encryption** are selected in Task Type field.



- i. The **Object Extension** field allow you to specify the objects extension (.txt, .csv, etc) that need to be excluded.

- ii. The **Extension List** field allow you either to **Browse** the path for the object or to select the object using the **Select Buckets** button. This functionality lets you specify the exclusion list of all objects that need not be included in the scanning.
- iii. The below panel list down information for both excluded **File Extension** and the **Scan Location** selected above.



***Note:** To decrypt tasks, ensure that appropriate roles have been assigned before executing the encryption task. For more details, refer to section [Role Management](#)

- 59. After creating the required task, i.e., detection, masking, encryption, decryption or metadata discovery, click save to save the task to schedule later, or save and execute to execute it right away.

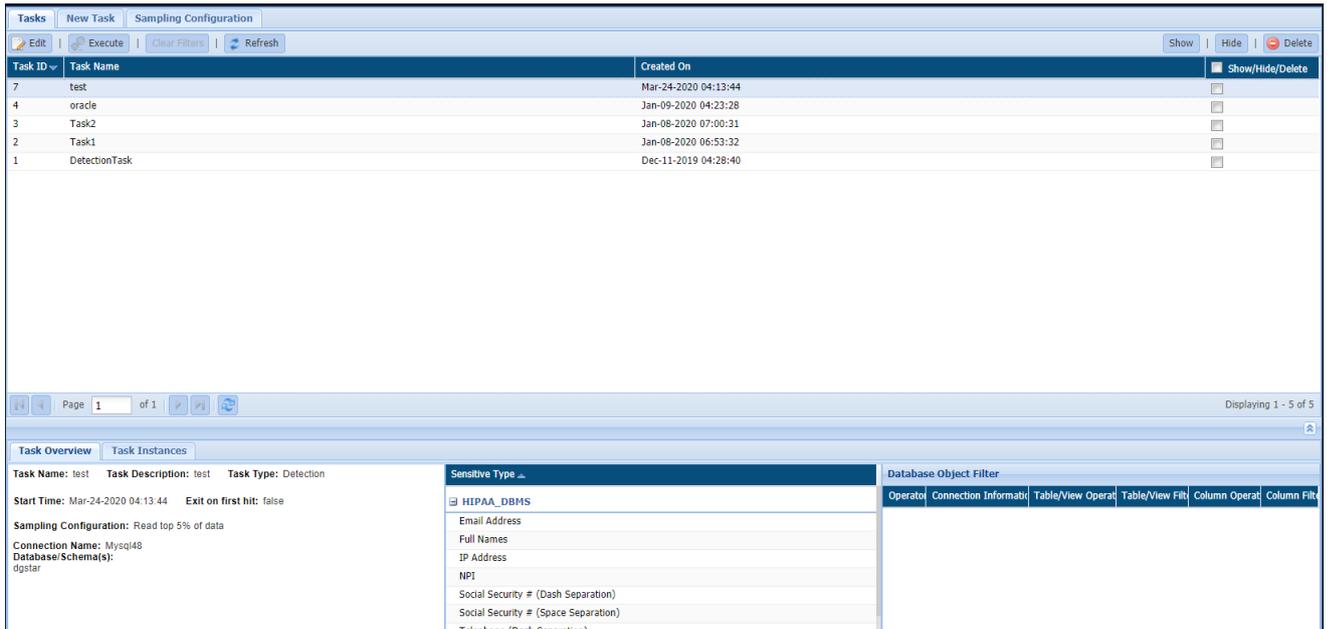
7.2.6.2 Databases

DgSecure supports Detection and Masking in RDS/RedShift databases. Following sections outline the process of creating these tasks.

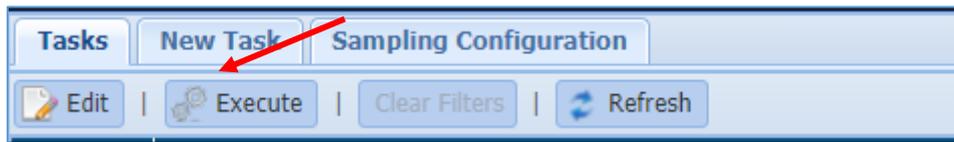
7.2.6.2.1 Detection

Perform the following steps to create a detection task.

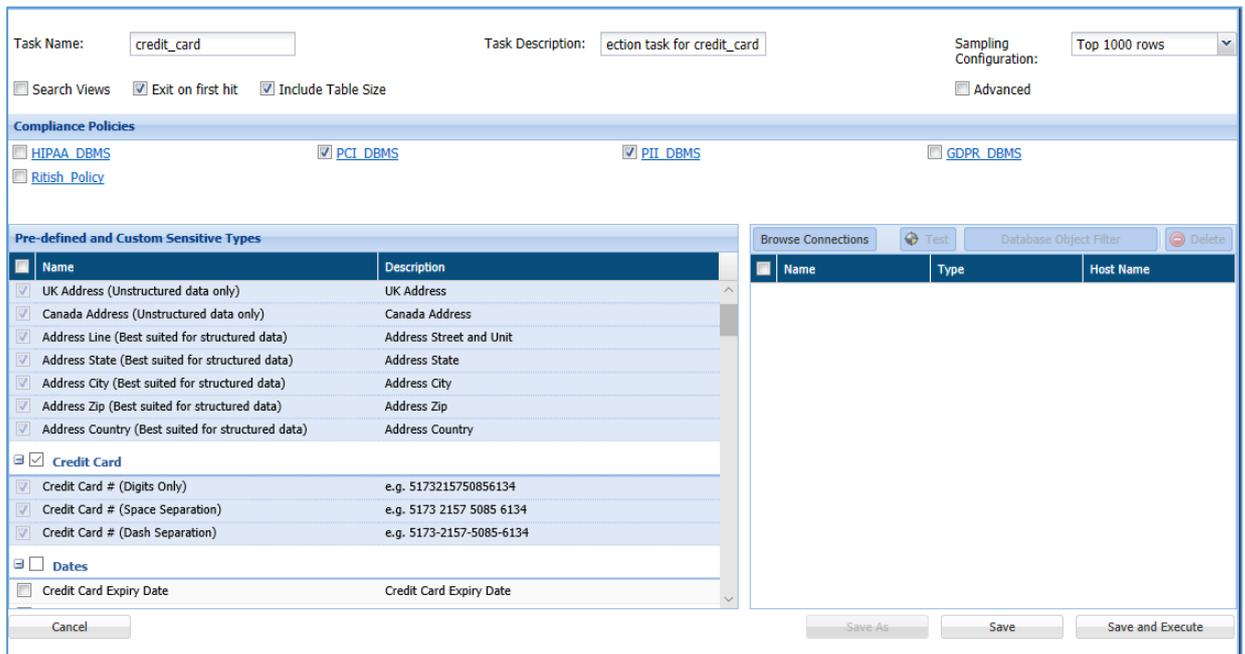
Go to **AZURE > DATABASES > DETECTION > TASKS**.



1. To create a new task, click on the **New Task** tab.



The following image shows the user interface for creating a task.



2. Enter a unique **Task Name**. This field supports numeric and character values.

Task Name:	<input type="text" value="cre_card"/>
------------	---------------------------------------

3. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.

Task Description:	<input type="text" value="Detector of CCNO"/>
-------------------	---

4. Check the **Search Views** option to detect the Views tables within the database as well as the tables linked with them. This option is available for the **Detection, Task Type**.
5. Check **Exit on First Hit** option to stop the scanning, when the first Sensitive Type is detected during the scan. This option is available for the **Detection, Task Type**.

60. DgSecure is equipped with data sampling to limit the area of scan which helps in reducing the time taken for detection. Default options to scan sample data from the database are:

- Top 1000 Rows
- Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. Check the **Advanced** checkbox to create a new sampling configuration.

You can also configure sampling through **AWS > RDS/REDSHIFT > DETECTION > TASKS > SAMPLING CONFIGURATION** tab.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Table Row Count Range: * To: Type: *

By: * Value: *

Table Row Count Range	Sample Type	Sample Value	Sample By	Actions
Default	Top	1000	Rows	 
1 to 200	Bottom	100	Rows	 

Note:

1. Top sampling option is not supported for Teradata, Aster DB and Sybase ASE connections.
2. Bottom sampling option is not supported for Teradata, Aster DB, Snowflake, Informix DB, Splice Machine, Sybase IQ and Sybase ASE connections.
3. Random sampling option is not supported for Snowflake, Informix DB, MySQL and Splice Machine connections. Random sampling is also not supported for Views in Sybase IQ.
4. For performance reasons, except for tables in DB2 schemas, random and bottom sampling is not recommended, where the tables can contain large volume of data.
5. If some unsupported sampling option is chosen, then best applicable sampling option will be applied. For more details regarding chosen sampling option, IDP logs can be checked.

- Enter the name of the Sampling Configuration.
- Enter the description for sampling.
- Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.
- Check the option **Show Advance Sampling Details** to set the advanced settings for sampling. Below are the options for advanced settings:
 - **Table row count range:** Enter numeric value. This value states the starting range of the table from which the records will be sampled.
 - **To:** Enter the numeric value. The value in this field states the ending range of the table till which the records will be sampled.
 - **Type:** Select the sampling configuration type from the **Type** option. There are four options for sampling configuration:
 - Top:** If you select the option **Top**, the sample data for the scan will be selected from the entries at the top of the table, based on the specified range.
 - Bottom:** If you select the option **Bottom**, the sample data for the scan will be selected from the entries at the bottom of the table. This does not mean the entries will be selected bottom up, instead

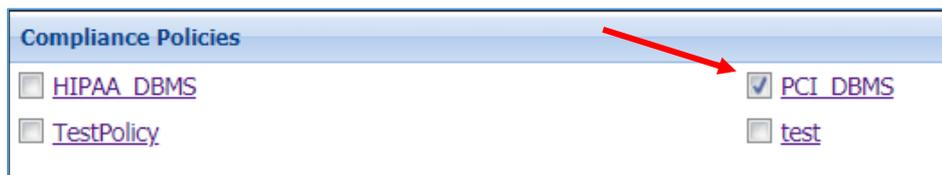
depending on the range the last entries in the table will be taken to create a sample data for detection.

- iii. **Random:** Entries from the table that correspond to the specified range, will be selected at random to create a sample set of data for detection.
- iv. **Complete:** If all the entries in the selected tables of the database have to be scanned for sensitive types, select this option.
- **By:** To Specify how to pick data for sampling from the table, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
- **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.

61. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.

62. Click the **Save** button to save the changes.

63. Select the required policy under the **Compliance Policies** section. The Compliance Policy panel displays all the Pre-Defined and Customized Policies. Users can select any number of policies while creating or editing a task. Sensitive types associated with the selected policy can be viewed in the panel below this panel **Pre-Defined and Custom Sensitive Types**. Selecting a policy is not a mandatory step, users can also proceed to select individual sensitive types. For more information, refer to section [Policy](#) .



64. Select the required sensitive types for the scan from the **Pre-defined and Custom Sensitive Types** section. Refer to the following screenshot:

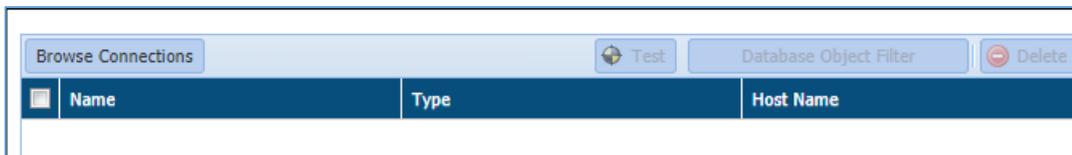
Pre-defined and Custom Sensitive Types	
Name	Description
<input type="checkbox"/> ABA_test	
<input type="checkbox"/> ABA_test	ABA
<input type="checkbox"/> Address	
<input type="checkbox"/> US Address	US Address
<input type="checkbox"/> UK Address (Unstructured data only)	UK Address
<input type="checkbox"/> Canada Address (Unstructured data only)	Canada Address
<input type="checkbox"/> Address Line (Best suited for structured data)	Address Street and Unit
<input type="checkbox"/> Address State (Best suited for structured data)	Address State
<input type="checkbox"/> Address City (Best suited for structured data)	Address City
<input type="checkbox"/> Address Zip (Best suited for structured data)	Address Zip
<input type="checkbox"/> Address Country (Best suited for structured data)	Address Country
<input checked="" type="checkbox"/> Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
<input type="checkbox"/> Dates	

The **Pre-Defined and Custom Sensitive Types** panel lists down all the Sensitive Types. The Sensitive Type associated with the policy gets selected in the Pre-Defined and Custom Sensitive panel and cannot be removed from the scan, however any number of sensitive types can be added to the scan.

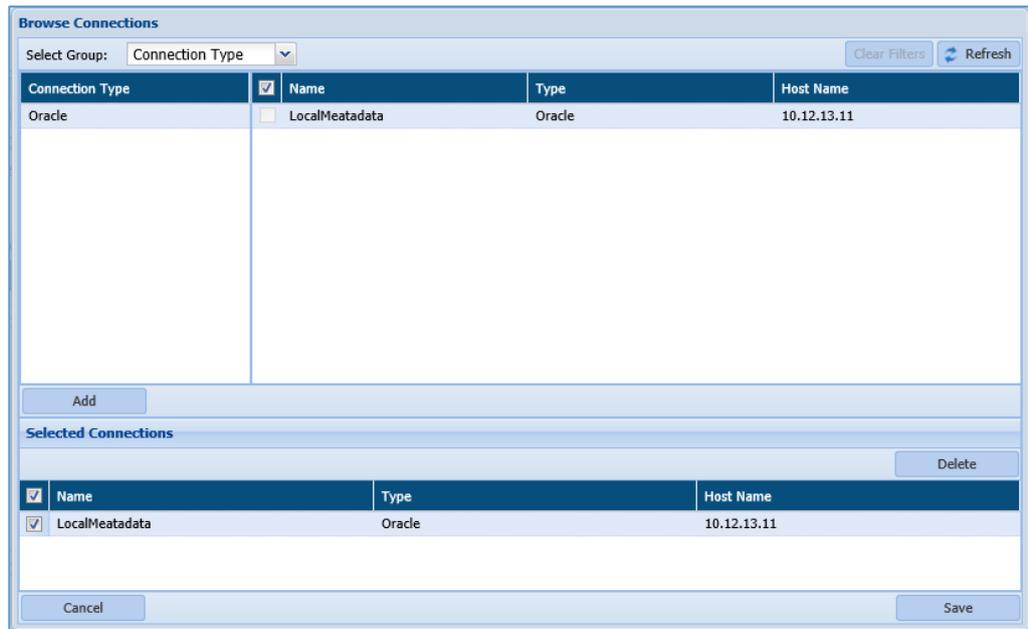
***NOTE:** If a policy is selected, user can still add more sensitive types to the scan but the sensitive types under the selected policies cannot be excluded from the scan.

65. The Database Connection panel lists down all the available RDBMS connections. Any number of connections can be selected for a task. This panel list down all the available connections. For details on how to create and manage connections refer [Connection Manager](#) Perform the following steps to choose a connection:

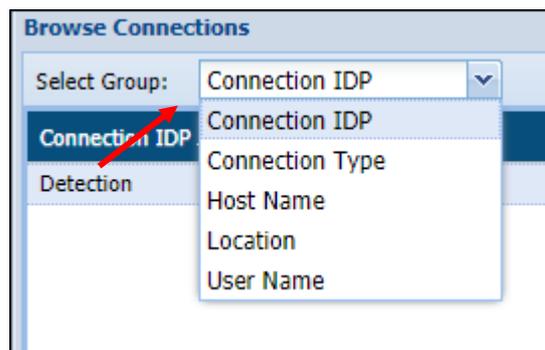
- a) Click Browse connections.



66. The **Browse Connections** dialog box will be displayed. This screen categorizes connections based on user preferences.



- i. Click on the Select Group dropdown and select the required option from the sub groups displayed on the left panel to sort the available connections.



The **Select Group** drop-down has five options:

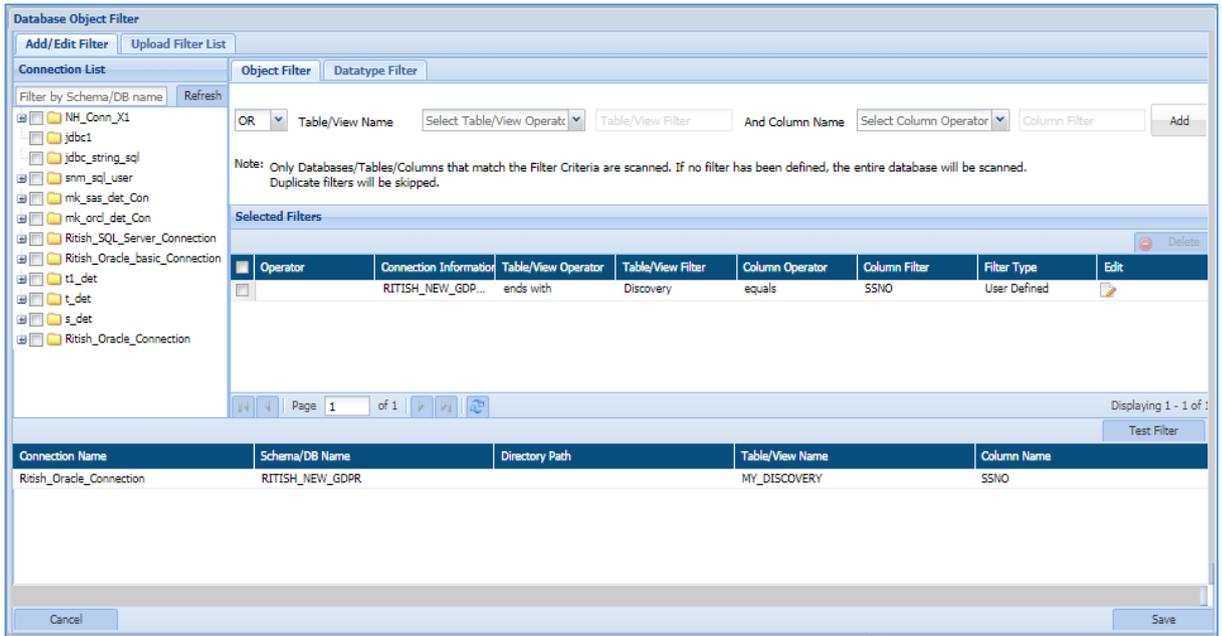
1. **Connection IDP:** Categorizes the available connections based on the types of IDPs available, i.e., Detection and Masking.
2. **Connection Type:** Categorizes the available connections based on the type of server connected to, i.e., Oracle, Teradata, SQL server etc.
3. **Host Name:** Categorizes the list of available connections based on Host Names.
4. **Location:** Categorizes the available connections based on the location of the target source system server, i.e., On-Premises and Cloud.

5. **User Name:** Categorizes the list of available connections based on the Usernames.
- ii. Click **Add** to include the selected database connection in the **Selected Connection** panel.
- iii. Check the Selected connection. Click the **Save** button to include connections.
- iv. Click the **Test** button to test the listed RDBMS connection.
- v. Click the **Database Object Filter** button to filter tables and/or columns. Once filters are defined, then only those databases/tables/columns that match the filter are scanned.

Check the checkbox next to the connection to enable the Database Object Filter.

<input checked="" type="checkbox"/>	Name	Type	Host Name
<input checked="" type="checkbox"/>	NH_151_SQLServer_D1	SQL Server	192.168.0.151

You can **Add/Edit Filter** in two ways i.e. either by specifying in manually it manually or by uploading the filter list in the **Upload Filter List** tab.



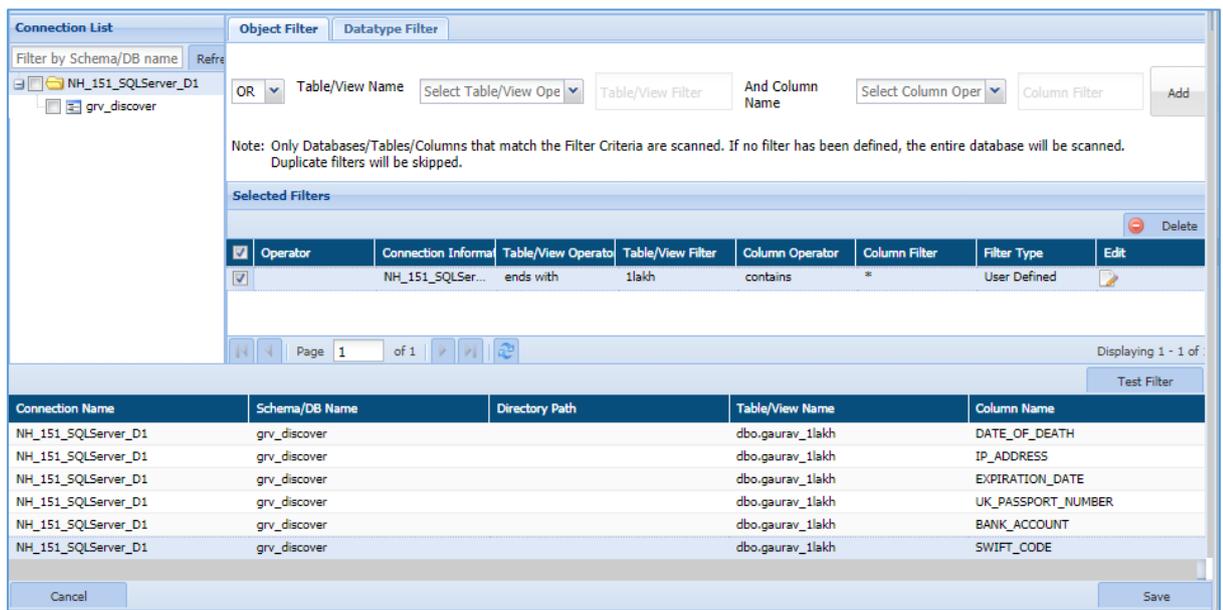
- **Add/Edit Filter:** This tab allows you to apply the filter for the selected connection. Once the filter has been applied then only those databases/tables/columns and datatypes that matched the criteria will be scanned.

There are two types of filter which can be applied.

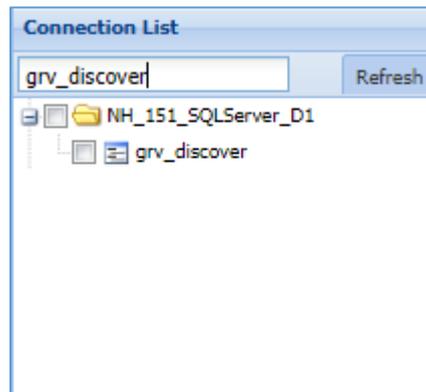
1. Object Filter
2. Datatype Filter

Object Filter

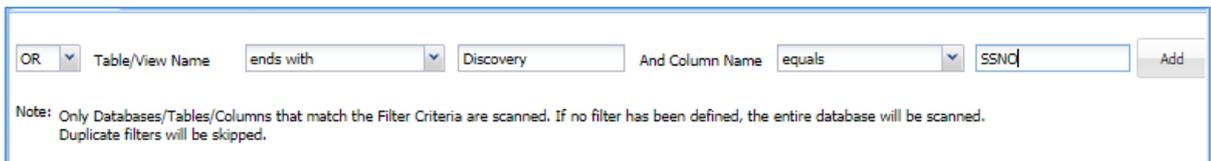
The Object Filter allows you to scan the database/tables/columns based on the defined filter. To apply an Object Filter, perform the following steps:



- a) Select the connection from the Connection List panel or enter the DB/Schema name in the Filter by Schema/DB Name textbox.



- b) Apply the **Object Filter** in the top panel by specifying the **Operator, Table/View** and **Column** name.



For example, in the above image the object filter specifies the table name should end with 'Discovery' and the selected table should contain SSNO column.

There are eight types of Operators based on which you can select the Table and Column name.

1. **Equals:** This operator will check whether the given table/column name exist in the selected database/table. It will return the matched records if the condition is fulfilled.
2. **Not Equal to:** This operator will return all the records except the given table/column name.
3. **Contains:** This operator will return only those tables/columns which matched the given criteria.
4. **Does not contain:** The functionality of this operator is similar to the **Not Equal to** operator, since it returns all the records except the given table/column name.
5. **Starts with:** This operator will return all the tables/column whose name starts with the given criteria.

6. **Does not start with:** The functionality of this operator is similar to the **Does not contain** and **Not Equal to**, since it will return all the tables/column name except the one which has been entered.
 7. **Ends with:** This operator will return all the tables/column name whose name ends with the given input.
 8. **Does not ends with:** The functionality of this operator is similar to the Does not contain and Not Equal to, since it will return all the tables/column name except the one which has been entered.
- c) Click the **Add** button to add the filter in the **Selected Filters** panel.
9. The **Selected Filters** panel will list down all the user defined filters. It displays information about the filters such as Connection Name, Table/View Operator, Table/View Filter, Column name, etc.

Operator	Connection Information	Table/View Operator	Table/View Filter	Column Operator	Column Filter	Filter Type	Edit
<input type="checkbox"/>	RITISH_NEW_GDP...	ends with	Discovery	equals	SSNO	User Defined	

- To edit a filter, click the  button in the **Edit** column. This functionality allows you to re-select the Operator, Table/View and Column name.
- To delete a filter, check the checkbox for that filter and click the **Delete** button. This button will get enabled when you check the checkbox for any filter.

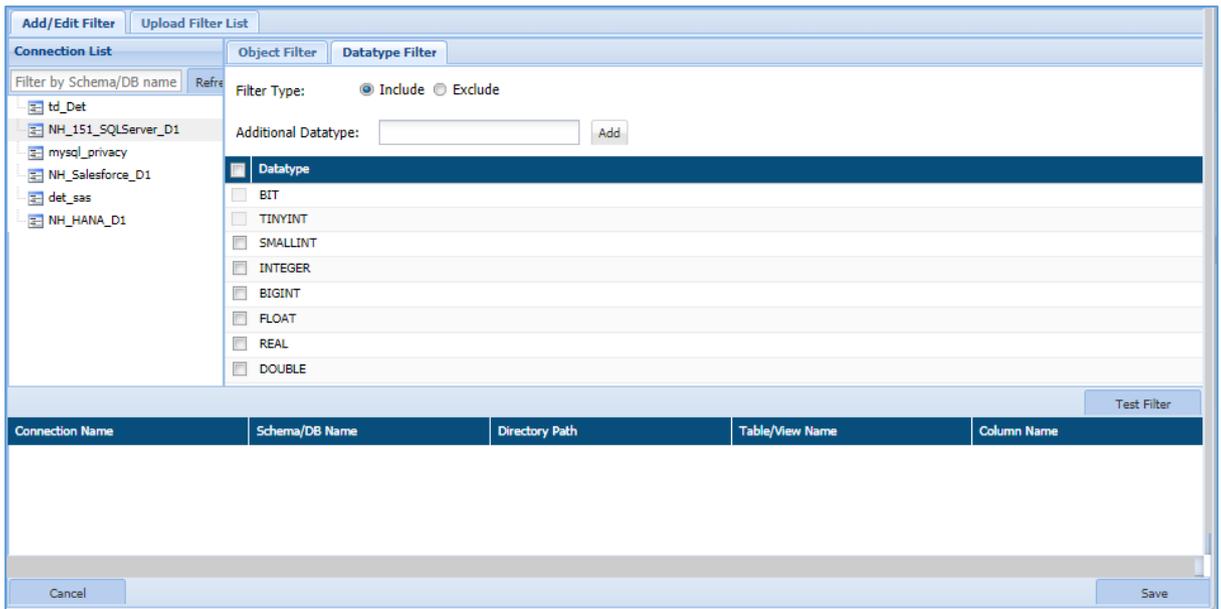
10. Click **Test** button to test the filter. It lists down the result matching the filter criteria.

Connection Name	Schema/DB Name	Directory Path	Table/View Name	Column Name
Ritish_Oracle_Connection	RITISH_NEW_GDPR		MY_DISCOVERY	SSNO

11. Click **Save** button to make the changes effective.
12. Click **Cancel** button, if you do not want to save the changes.

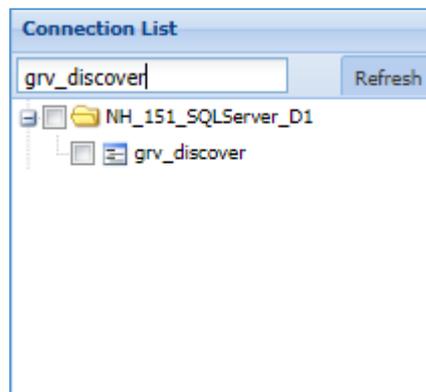
Datatype Filter

This tab allows you to specify the datatypes to be included or excluded while scanning based on the selection.



To apply Datatype Filter, perform the following steps:

- a) Select the connection from the **Connection List** panel or enter the Database/Schema name in the **Filter by Schema/DB Name** textbox.



- b) Select the **Filter Type** from the given option. By default, **Include** filter type is selected.
 - **Include:** This option allows to select the datatype that need to be included while scanning.
 - **Exclude:** This option allows to select the datatype that need to be excluded while scanning.
 - **Additional Datatype:** This field allows you to add a datatype if it's not present in the DataType panel. Once that datatype is entered, click **Add** button to specify it in the DataType panel.

Filter Type: Include Exclude

Additional Datatype:

13. Check the checkbox for the datatype which you want to include or exclude based on the selection of **Filter Type** in above panel.

Filter Type: Include Exclude

Additional Datatype:

Datatype	
<input type="checkbox"/>	NUMERIC
<input type="checkbox"/>	DECIMAL
<input checked="" type="checkbox"/>	CHAR
<input checked="" type="checkbox"/>	VARCHAR
<input type="checkbox"/>	LONGVARCHAR
<input type="checkbox"/>	DATE
<input type="checkbox"/>	TIME
<input type="checkbox"/>	TIMESTAMP
<input type="checkbox"/>	BINARY

For example, in the above image the CHAR and VARCHAR datatypes are included in the scanning process since Filter Type specify the option as 'Include'. If 'Exclude' is selected as Filter Type then selected CHAR and VARCHAR datatype will be excluded from the scanning process.

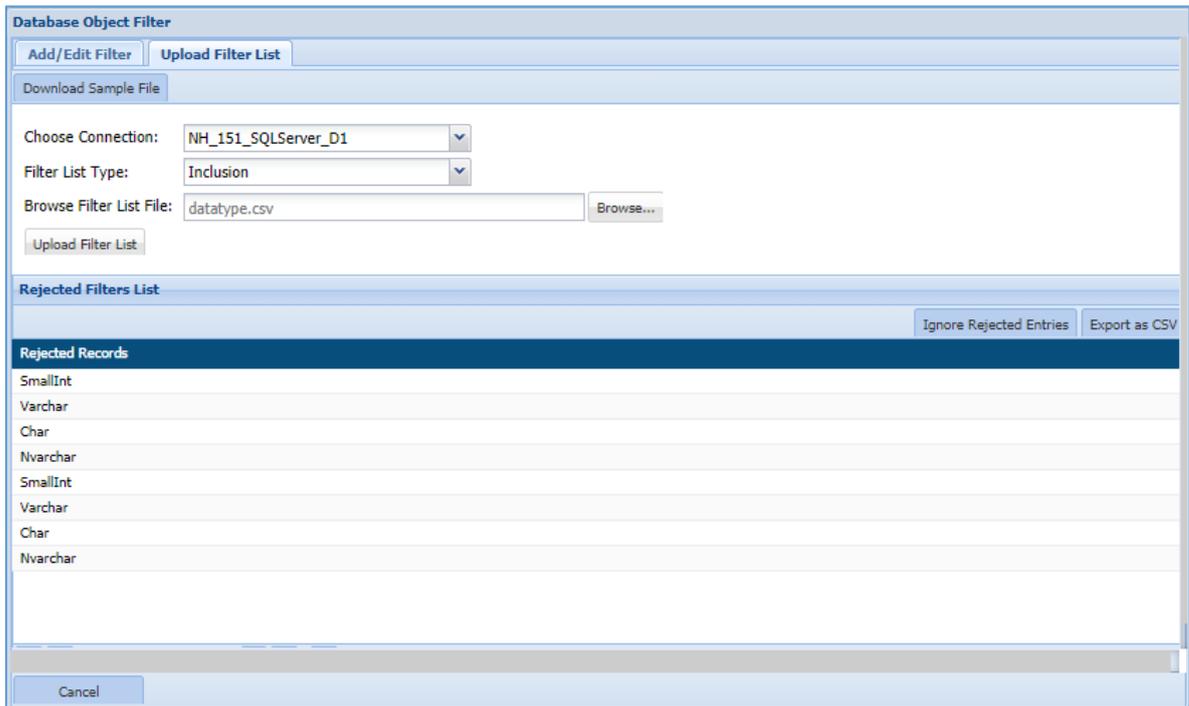
14. Click **Test** button to test the whether any column in the database contains the selected datatype. This functionality will list down all the columns that contain the selected datatype.

Connection Name	Schema/DB Name	Directory Path	Table/View Name	Column Name
Ritish_Oracle_Connection	RITISH_NEW_GDPR		MY_DISCOVERY	SSNO

15. Click **Save** button to make the changes effective.

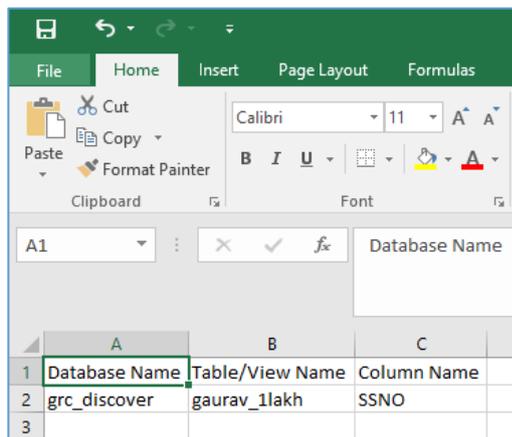
16. Click **Cancel** button if you do not want to save the changes.

- **Upload Filter List:** This tab allows you to upload a file containing the list of all columns.



To upload the filter list, perform the following steps.

- a) Select the connection from the **Choose Connection** drop-down.
17. The **Download Sample File** button will be enabled once a connection has been chosen. Enter the Database, Table/View and Column Name in the sample file.



18. Select the **Filter List Type** from either 'Inclusion' or 'Exclusion'. This functionality allows you to specify whether to include or exclude the Database, Table/View and Column name.

Filter List Type: Exclusion

- Inclusion
- Exclusion

19. Click the **Browse** button to search and upload the saved sample file containing the list of Database, Table/View and Column name which will be either excluded or included.

Browse Filter List File: FiltersSampleFileForSQLServer (2).csv Browse...

20. Click **Upload Filter List** button to add the defined filters in the **Selected Filters** panel under **Add/Edit Filter** tab.

Selected Filters							Delete
Operator	Connection Informa	Table/View Operato	Table/View Filter	Column Operator	Column Filter	Filter Type	Edit
<input type="checkbox"/>	NH_151_SQLSer...	equals	gaurav_ilakh	equals	SSNO	Include	

Displaying 1 - 1 of

21. The **Rejected Filters List** will display all the entries from the uploaded list which are not in a proper format as specified in the sample file.

Rejected Filters List		Ignore Rejected Entries	Export as CSV
Rejected Records			
grc_discover,,SSNO	srch_ccno.emp_d,	discover_city,Emp_detail,City	

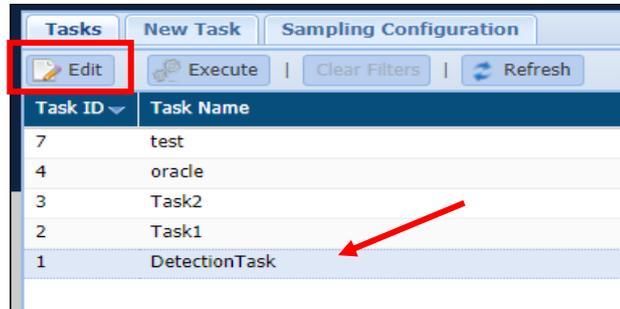
- To download the list of rejected entries, click **Export as CSV** button. A downloaded file will contain the list of rejected entries which were not formatted as per specified format.
- To remove the rejected entries from the **Rejected Filters List** panel, click the **Ignore Rejected Entries**. This functionality will remove all the rejected entries from the panel.

22. Click **Save** button to make the changes effective.

23. Click **Cancel** button if you do not want to save the changes.

24. Click **Save**, if you want to execute the task later else click **Save and Execute**. The results of the task and its status can be viewed under **RDBMS>DETECTION>RESULTS** (Refer to Section [Results](#)).

To edit an existing task, select the required task from the list of tasks on the Tasks screen and click edit. A task can be edited using the same steps for task creation.



***Note:** Task defining features such as Incremental, Exit on First Hit, Search View and the Task Type cannot be edited. Some of the originally selected options can be modified but new options cannot be added.

7.2.6.2.2 Masking Task

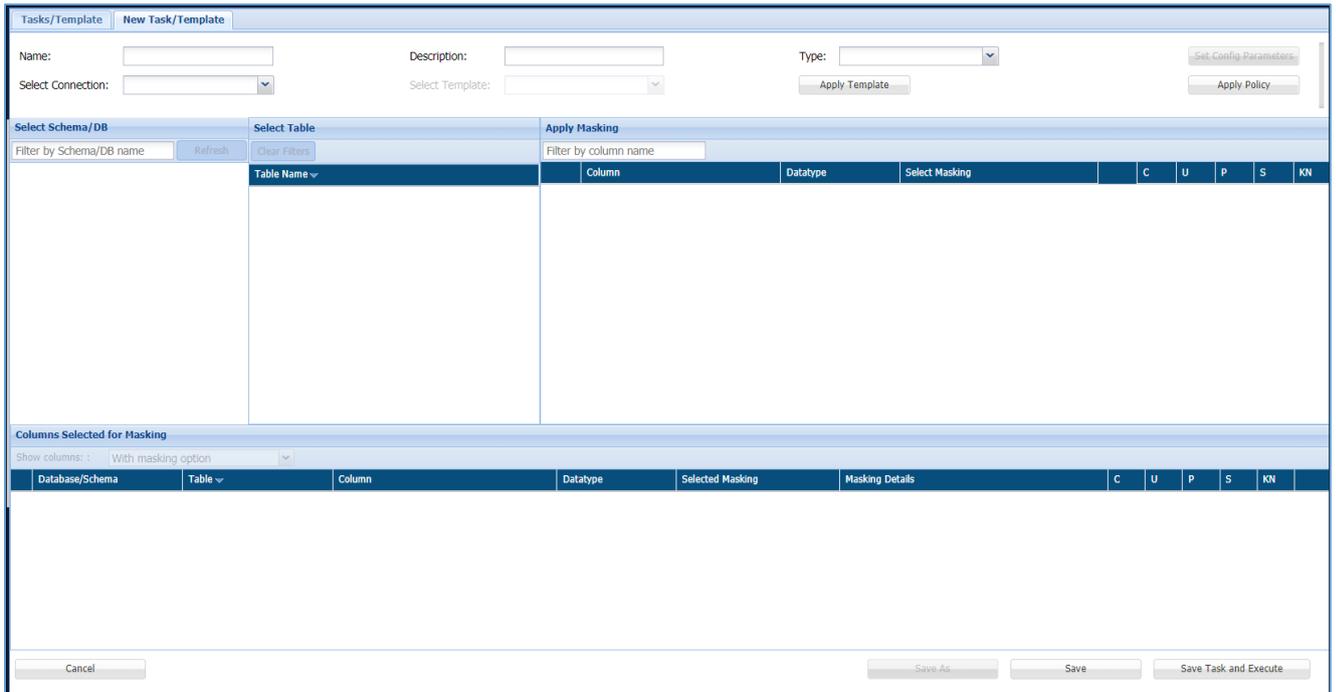
Perform the following steps to create a masking task:

Access the **NEW TASKS/TEMPLATE** screen, click **AZURE > DATABASES > MASKING > TASKS/TEMPLATES > NEW Task/Template** tab.

1. To create a new task or template for masking tasks, click on the **New Task/Template** tab.



The following image shows the user interface for creating a task.



25. Enter a unique **Task Name**. This field supports numeric and character values.

Name:

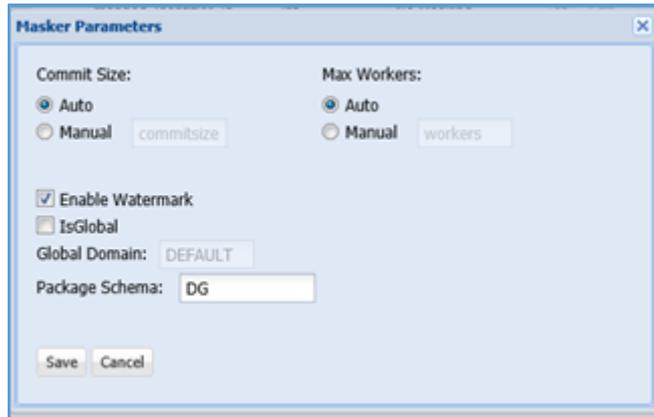
26. Enter a **Task Description** of maximum 254 characters. This field supports numeric and character values.

Description:

27. Click on the **Type** dropdown. Select **Task** to create a masking task or **Template** to create a template for masking tasks.

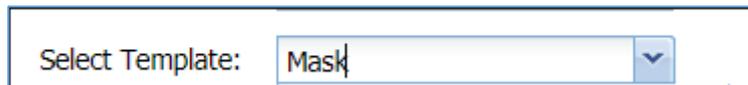
Type:

28. Click the **Set Config Parameters** button. It will display the environment settings for the masking operations such as **Commit Size**, **Max Workers**, **Package Schema**, **IsGlobal** and **Enable Watermark**.



29. Click **Select Connection** to view the list of the available database connections for masking. For details on how to create a connection for masking refer to [Connection Manager](#)

30. The **Select Template** option provides a list of user created templates for masking. This option will be greyed out if no templates have been created. To create a new template select the Type as Template in step 6 and follow the next steps for creating a task and save the template. Once created, the new template will appear in the select template dropdown on the new task screen.

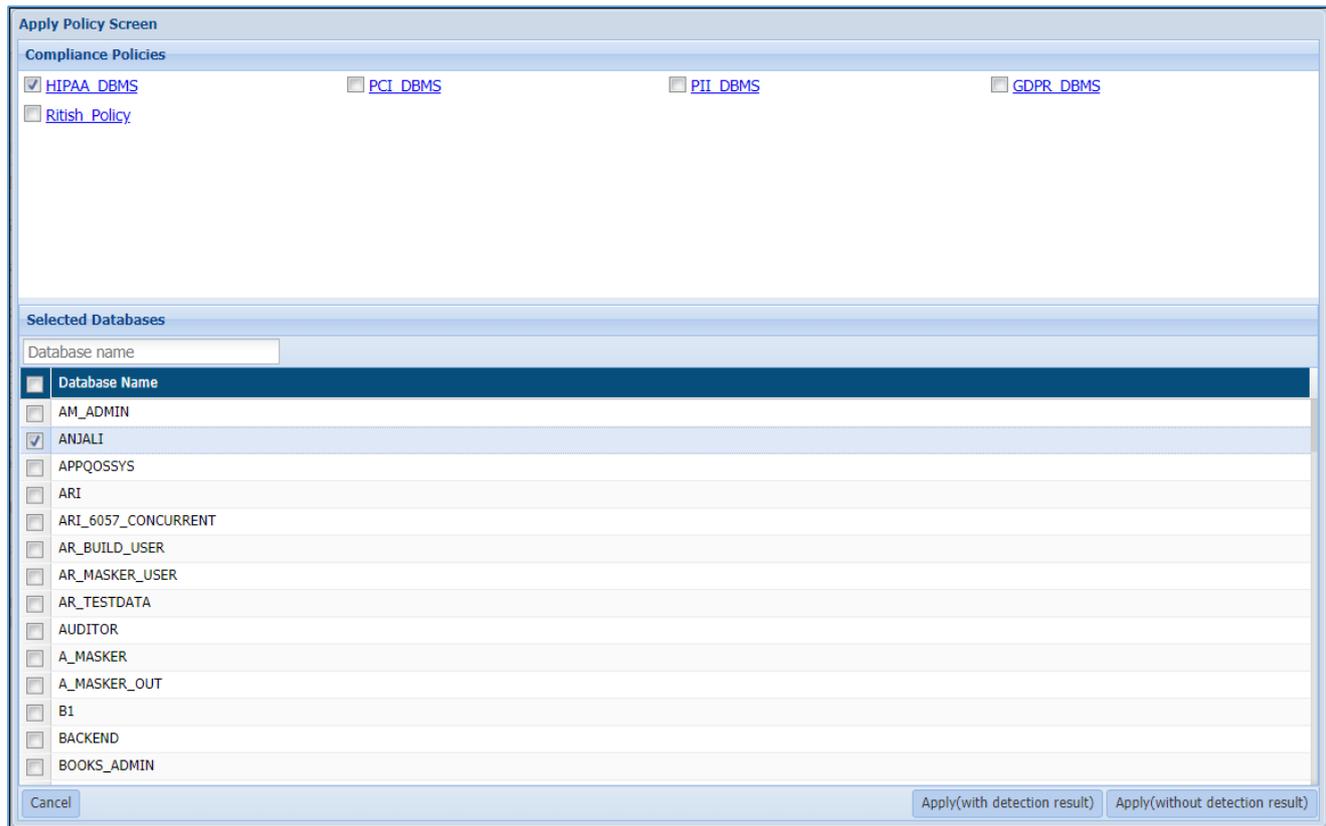


***Note:** Templates are connections specific and only one template can be selected at a time.

31. Click the **Apply Template** button to apply the selected template in **Select Template** drop-down.

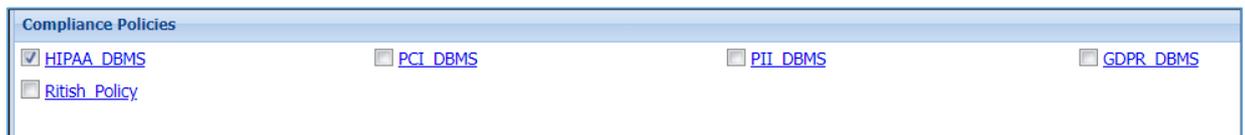


- a) Click the **Apply Policy** button to choose the available Compliance Policies.

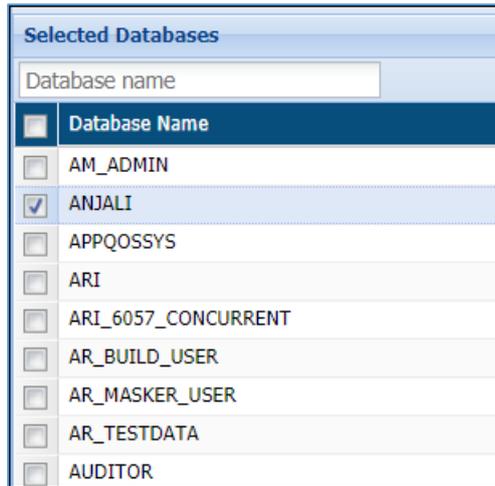


Perform the following steps to apply a policy to the task or template:

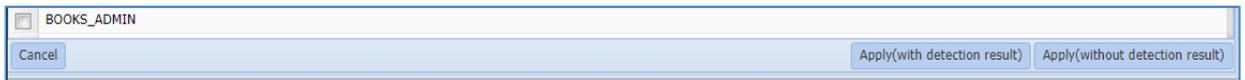
- i. Select the required policy/policies under the **Compliance Policies** panel.



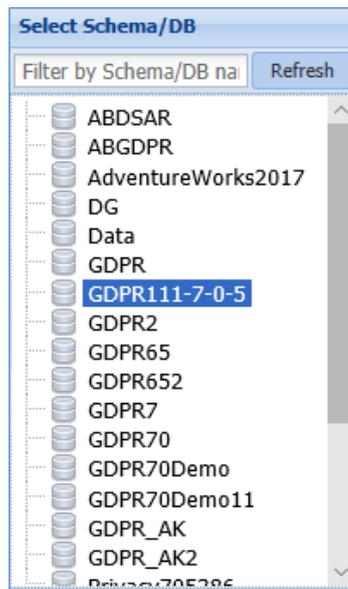
- ii. Select the databases you need to mask by checking the checkbox next to the Database Name.



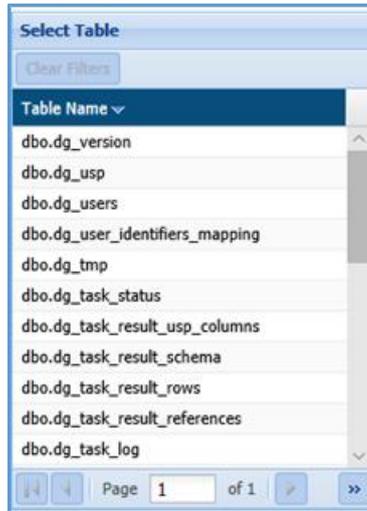
- iii. Click the **Apply** (with detection result) button to apply the selected policy with detection results or click the Apply (without detection result) button to apply the selected policy without results. Click Cancel to redo your selection.



- 32. The Select Schema/DB pane will display the list of databases or schemas for a selected connection in **Select Connection** drop-down.



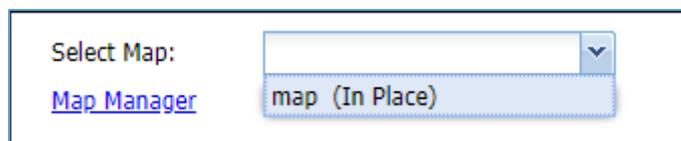
- 33. Select the table from the Select Table pane. This pane lists all the tables for the selected database or schema.



34. Check the Incremental checkbox to apply incremental masking to the database. This feature is useful to mask new values added in a database after masking has been executed on it. Only the new entries will be masked, thus, the time taken for masking would be reduced. Perform the following steps to make your masking task incremental:

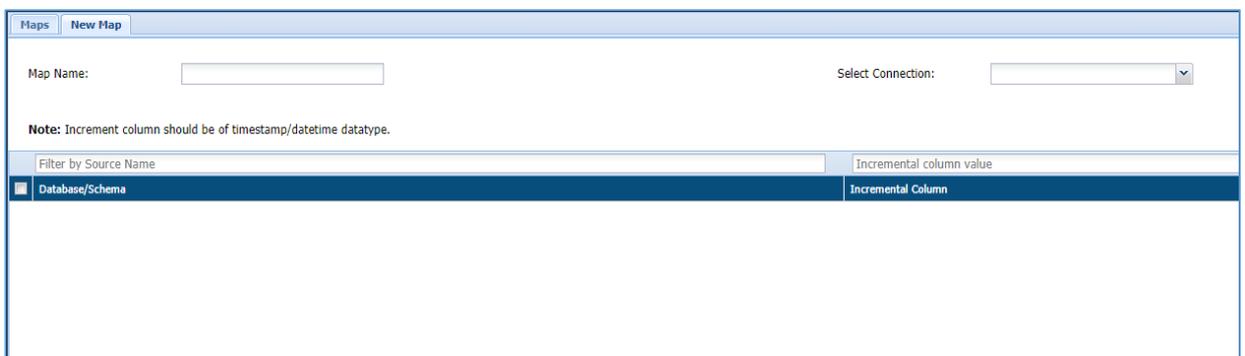
a) Check incremental checkbox.

35. The Select map Dropdown will appear. Select the required map. Maps define incremental columns within a database which are considered for indexing the data in order to mask the new rows added to the database.



36. To add a map click on the Map Manager link

37. The Map Manager Screen can also be accessed from the menu bar. The following image displays the Map Manager interface:



38. Enter a Map Name.

Map Name:

39. Select a connection that contains the database on which incremental masking has to be applied.

Select Connection:

- Ritish_Oracle_masker**
Host:192.168.0.163 Type:Oracle DB:ritishmasker
- Ritish_Masker_TNS**
Host:N.A. Type:Oracle DB:ritishmasker
- orcl_neha**
Host:192.168.0.163 Type:Oracle DB:nehamaskuser

Incremental column value

Incremental Column

40. Select the required Database/Schema.

Filter by Source Name

<input type="checkbox"/>	Database/Schema
<input type="checkbox"/>	VCON
<input type="checkbox"/>	SNCF
<input type="checkbox"/>	GDPR_TEST
<input checked="" type="checkbox"/>	NEHAMASK163
<input type="checkbox"/>	DG_VIK_655
<input type="checkbox"/>	DGCONTROLLER_SANDEEP
<input type="checkbox"/>	AR_TESTDATA
<input type="checkbox"/>	VIKRAM
<input type="checkbox"/>	A_MASKER_OUT
<input type="checkbox"/>	A_MASKER
<input type="checkbox"/>	DGSHAREPOINT
<input type="checkbox"/>	JSON
<input type="checkbox"/>	NEHA777

41. Enter the name of the incremental Column to the corresponding database and ensure that it is a timestamp or date-time datatype.

Database/Schema	Incremental Column
<input type="checkbox"/> JSON_XML_SCRIPT	
<input type="checkbox"/> M_MASKING	
<input type="checkbox"/> DG_USER	
<input type="checkbox"/> NEHA	
<input type="checkbox"/> GDPR_72	
<input type="checkbox"/> BUILD_N	
<input type="checkbox"/> SNCF_MASKER	
<input type="checkbox"/> VCON	
<input type="checkbox"/> SNCF	
<input type="checkbox"/> GDPR_TEST	
<input checked="" type="checkbox"/> NEHAMASK163	DATE_OF_BIRTH
<input type="checkbox"/> DG_VIK_655	

42. Save the map. It will be available in the Select Map dropdown on the new task/template screen.

NOTE: To apply incremental masking the following criteria must be met:

1. Create a map to execute incremental masking and apply it to the task.
2. Relational tables cannot be included.
3. Column holding date, time stamp or a numeric incremental value should be added as incremental column.
4. The incremental column cannot be masked.

43. The Apply Masking pane display the list of all the columns for the selected table in Select Table pane. For detailed information on all the available masking options in DgSecure refer to [Masking Options](#) . Perform the following steps to apply masking option to the columns in the database:

- a) Select the masking option from the drop-down against the column entry. You can apply the masking to the selected column by checking on the checkbox corresponding to the column name.

Apply Masking										
Filter by column name										
Column	Datatype	Select Masking		C	U	P	S	KN	SL	
checksum	bigint	No Masking	▼	Edit	<input type="checkbox"/>					
connectionid	int	No Masking	▼	Edit	<input type="checkbox"/>					
constraintCols	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
dbName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
groupId	int	No masking available		Edit	<input type="checkbox"/>					
keyType	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
ResultTableId	int	No masking available		Edit	<input type="checkbox"/>					
schemaName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					
tableName	nvarchar	No Masking	▼	Edit	<input type="checkbox"/>					

To further enhance the results of masking following options can be combined with the masking options:

- iv. **Consistent:** Masks the data of the selected table with consistent values. E.g. If the name John is masked as FVGB, throughout the table it will be masked the same way.
- v. **Unique:** Masks each entry with a unique value.
- vi. **Persistent:** Similar to Consistent, however in this case the same values will be masked consistently across all the tables of the database.
- vii. **Sync:** This options allows tracking of masked entries in different tables if any of the same entries are found in other tables they will also be masked.
- viii. **Keep Null:** The cells containing null values are kept null even after masking.
- ix. **Stateless:** Persistently masks the data without creating any metadata. No extra space is required to create masking tables.

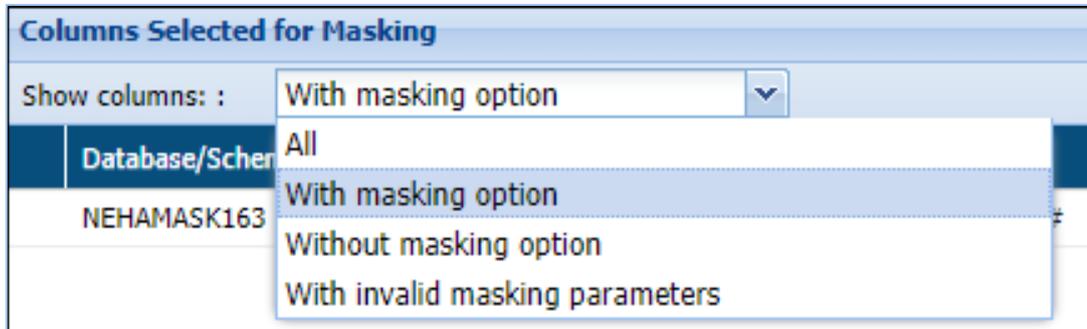
Column	Datatype	Select Masking	Edit	C	U	P	S	KN	SL
checksum	bigint	No Masking	Edit	<input type="checkbox"/>					
connectionid	int	No Masking	Edit	<input type="checkbox"/>					
constraintCols	nvarchar	No Masking	Edit	<input type="checkbox"/>					
dbName	nvarchar	No Masking	Edit	<input type="checkbox"/>					
groupId	int	No masking available	Edit	<input type="checkbox"/>					
keyType	nvarchar	No Masking	Edit	<input type="checkbox"/>					
ResultTableId	int	No masking available	Edit	<input type="checkbox"/>					
schemaName	nvarchar	No Masking	Edit	<input type="checkbox"/>					
tableName	nvarchar	No Masking	Edit	<input type="checkbox"/>					

44. View the selection under the **Columns Selected for Masking** panel.

Source DB/Schema	Destination DB/Sch	Table	Column	Datatype	Selected Masking	Masking Details	C	U	P	S	KN	SL
Data	Data	dbo.Data2	Description	nvarchar(M..	Custom	Mask by function [DG].CUSTO...	<input checked="" type="checkbox"/>	<input type="checkbox"/>				

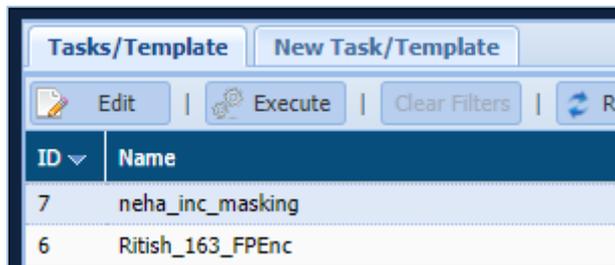
45. Select the option from the Show Column drop-down. There are four options:

- a) All
- b) With masking option
- c) Without masking option
- d) With invalid masking parameters



46. Click save to save the task or template or save and execute to execute the masking task.

47. To edit a task, select the task from the Tasks/Template tab and click edit. A task can be edited using the same steps for creating a new task.



7.2.7 Google Cloud

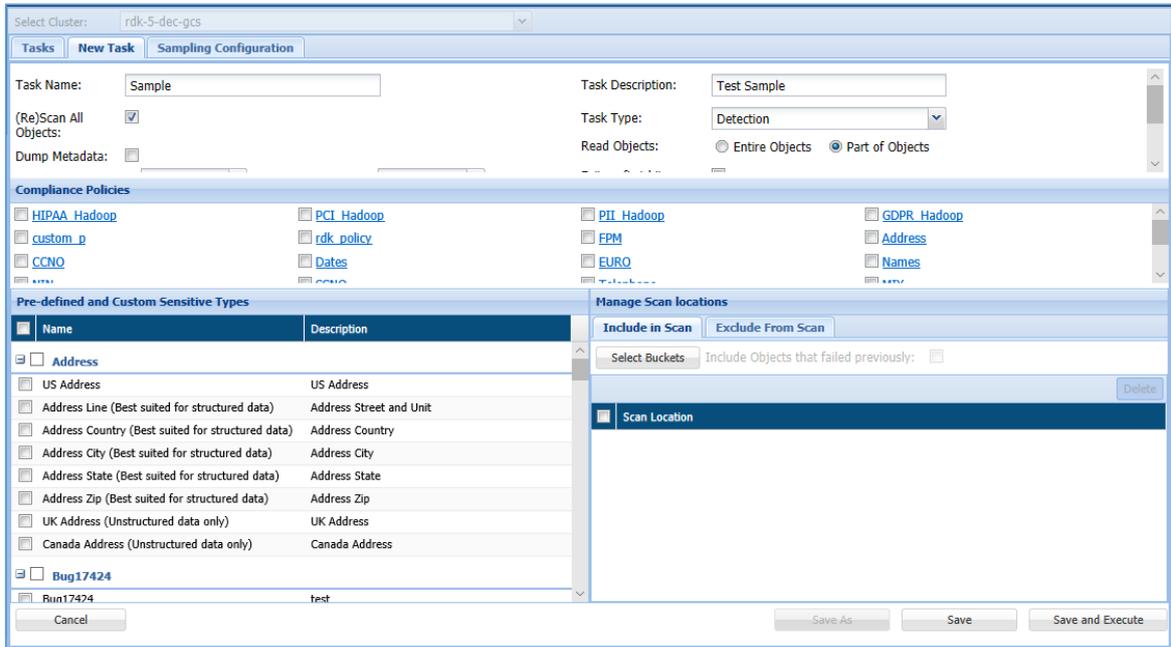
7.2.7.1 GCS

In Google Cloud, you can create and manage the GCS.

These tasks search for sensitive data hosted in Google Cloud Storage (GCS). They are designed to run on dataproc clusters spun up using the DgSecure Cloud IDP. The Cloud IDP is responsible for interacting with Google Cloud platform in order to spin up the cluster, manage the cluster's life cycle and installing the GCS IDP on the cluster.

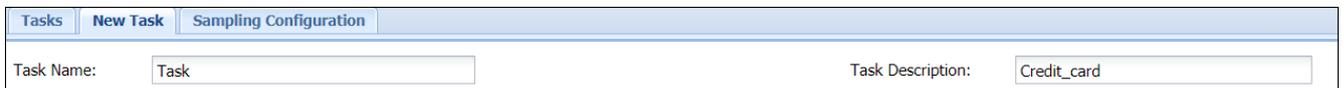
7.2.7.1.1 Create a task

1. Go to Google Cloud > GCS > Tasks. Select the New Task tab.



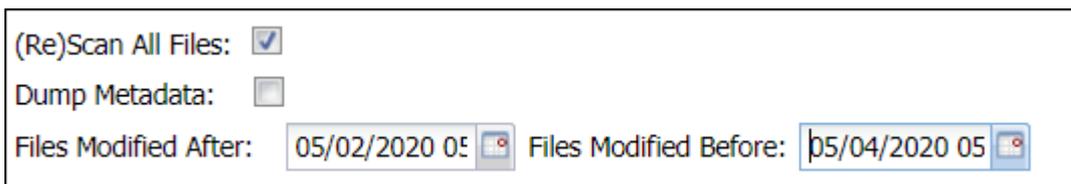
2. Select cluster from the **Select Cluster** drop-down.

48. Enter a meaningful **Task Name** and **Task Description**. The task name must be unique to the task. It can be up to 256 characters and consist of letters, numbers, certain symbols (! @ # \$ _), without any spaces. The definition can be as long as the task name and contain any combination of numbers, letters, and symbols.



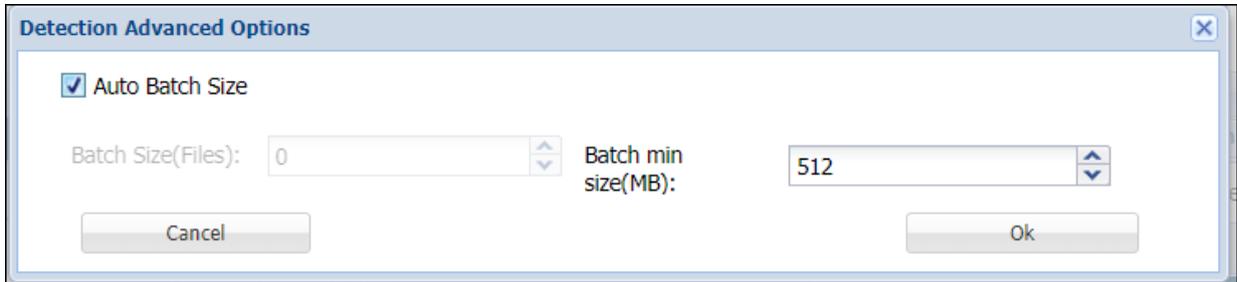
49. Choose **Detection** as the **Task Type**. Following are the options, specific to the **Detection**

50. Check **Scan All Files** checkbox to scan all the available files for a given connection between the dates specified in **Files Modified After** and **Files Modified Before** drop down.



51. Check **Dump Metadata** option to remove the metadata files after scanning.

52. Click on the **Advanced Options** button to define the batch size of the data.



Define the number of files per batch in the **Batch Size (Files)** option or check the **Auto Batch Size** option to enter the minimum batch size in **Batch min size (MB)** option.

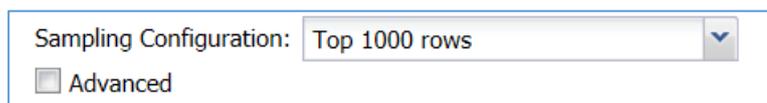
53. Choose to read the entire file or a part of the file at random in the **Read Files** options.



54. Check **Exit on first hit** checkbox to report the table or database as sensitive, at the first event of detection of a sensitive type.



DgSecure is equipped with **Sampling Configuration** to limit the area of scan which helps in reducing the time taken for detection.



Default options to scan sample data from the database are:

- Top 1000 Rows
- Read top 5% of data

The sampling configuration field is set to sample Top 1000 Rows by default. To create a new sampling configuration perform the following steps:

- Go to **GCS -> TASKS -> SAMPLING CONFIGURATION** tab or check the Advanced checkbox.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

File Size Range (Bytes)	Sample Value	Sample By	Actions
Default	1000	Rows	 

55. Enter the name of the Sampling Configuration.

Name: *

56. Enter the description.

Description:

57. Check the option **Set Sampling Config as Default** to set the Sampling Configuration as the default configuration for all your tasks.

Set Sampling Config as Default:

58. Check the option **Show Advance Sampling Details** to set the advanced settings for sampling.

Sampling Criteria Per Map

File Size Range (Bytes): * To:

By: * Value: *

Below are the options for advanced settings:

- **File Size Range:** Enter the range for the sample in Bytes.
- **To:** Enter the ending range for the sample.
- **By:** To Specify how to pick data for sampling from the source system, there are two ways:
 - i. **By Rows:** Select **Rows** from the drop-down, to sample data based on the number of rows.
 - ii. **By Percent:** Select **Percent** from the drop-down, to sample a percentage of the data
- **Value:** Enter the numeric value. It will specify the total number of records to be processed if sampling By-Rows is selected and denotes the percentage of sampling By-Percent is selected.

59. After setting up the required configuration, click **Add** to add the user-defined sampling configuration to the list.

60. Click the **Save** button to save the changes.

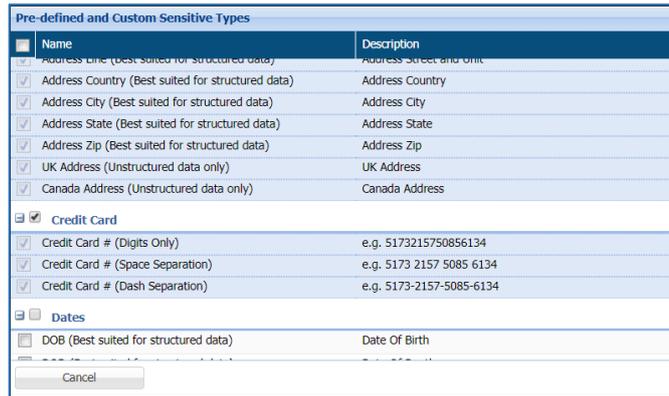
61. Compliance Policy can be set with all the task types in HDFS except Metadata Discovery. For more details about compliance policies, refer to section [Policy](#) . After selecting the required options, perform the following steps:

a) Select the required policies.

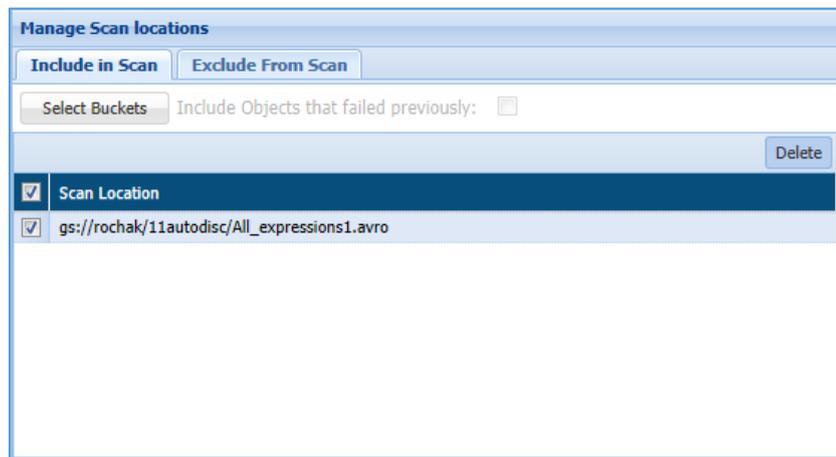
Compliance Policies

HIPAA_Hadoop PCI_Hadoop PII_Hadoop GDPR_Hadoop

62. Select the required sensitive types in the **Pre-defined and Custom Sensitive Types** panel.

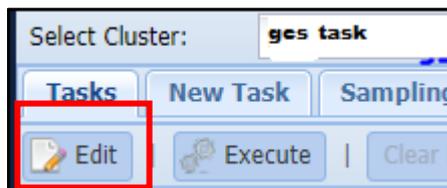


63. Specify location of the files to scan in the **Manage Scan Location** panel. It displays either task target directories or excluded directories, depending on which tab is selected. You can add more objects by clicking the **Select Buckets** button.



- a) **Delete:** Check the **Delete** checkbox corresponding to the scan location that you want to delete. Click the **Delete** button to delete the selected scan location.
- b) **Select Buckets:** Click the **Select Buckets** button, if you want to include the object, follow the below steps for including an object:
 - i. Click **Select Buckets** button. It will open **GCS File Browser** window.
 - ii. Select objects from the **Select Object** Pane. Check the File in the right pane of the **GCS Object Browser** window.
 - iii. Click the **Add** button. It will add the file in **Selected Locations** pane.
 - iv. Check the file in the Selected Locations pane.
 - v. Click Done.

64. **Include Objects that failed previously:** Check this option, if you want to include the objects which skipped in the previous run. Such objects will get included when you re-run the task.
65. **Exclude From Scan:** You can specify an object or a folder that need to be excluded in the targeted scan path. There are two ways in which you can specify an object is excluded. These are:
- **Object Extension:** Enter the type of **Object Extension** which need to be excluded.
For Example: .txt, .docx, .CSV, etc.
 - **Add:** Click Add button to include the **Object Extension** type in the **File Extension** pane.
 - **Exclusion List:** You can upload object **Exclusion List** either from your local machine using **Browse** button or **GCS Object Browser**.
 - **Select Buckets:** Click **Select Buckets** to include object Extension list from **GCS File Browser**.
 - **Delete:** Check the **Delete** checkbox corresponding to the **File Extension** or **Scan Location** that you want to delete. Click the **Delete** button to delete it.
66. After creating the required task, i.e., detection, masking, encryption, decryption or metadata discovery, click save to save the task to schedule later, or save and execute to execute it right away.
67. To edit an existing task, select the required task from the list of tasks on the Tasks screen and click **Edit**. A task can be edited using the same steps for task creation.



7.2.7.2 *BigQuery*

7.2.8 **SharePoint**

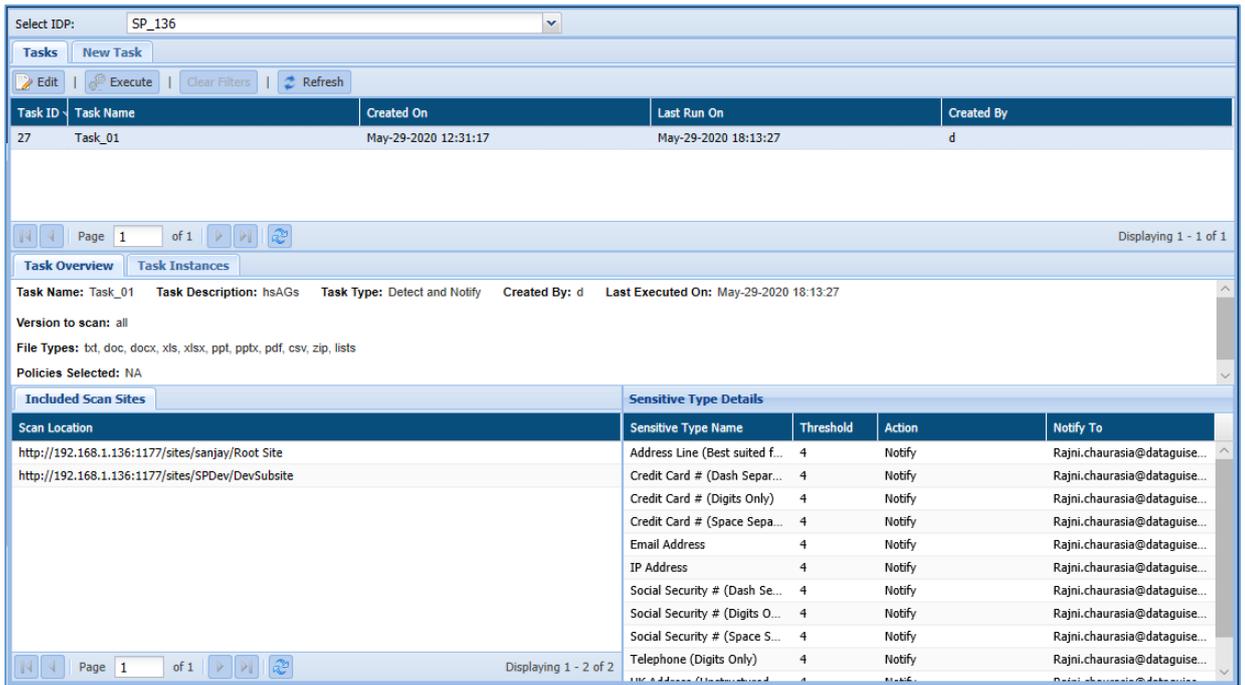
It supports for the detection of sensitive data in Microsoft Sharepoint. **SHAREPOINT** sub-menu

provides the facility to create and manage tasks to identify the data that is subtle to the data handling regulations.

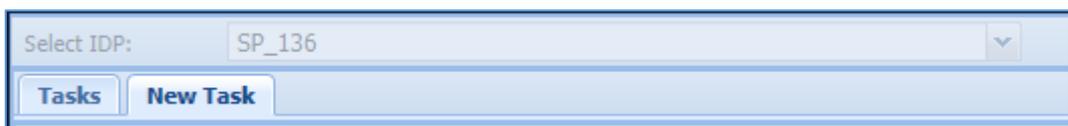
Sharepoint IDP supports detection on various file types: text, excel, pdf, csv, ppt, doc and zip. Detection task locates and identifies sensitive data as per the applied policy.

The below steps provides the facility to create a new SharePoint task. A task consists of one or more policies, an action such as detection, notify etc., and a target scan path.

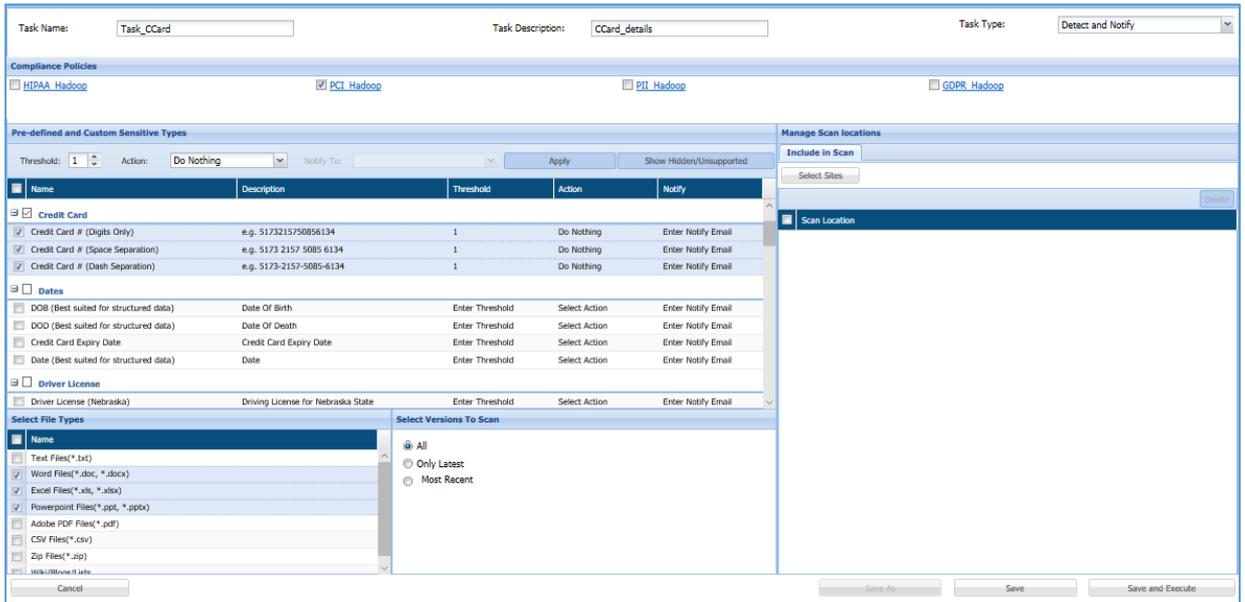
Go to **SHAREPOINT > TASKS**.



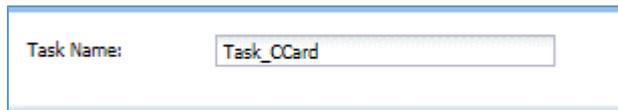
1. Select an IDP from **Select IDP** drop –down and click **NEW TASK** tab.



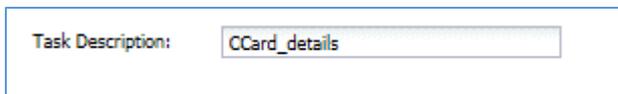
2. The following image shows the user interface for creating a task.



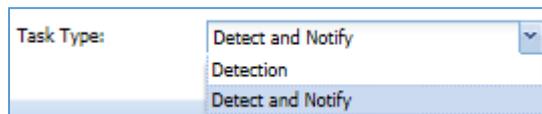
3. **Task Name:** Enter the name of the task.



68. **Task Description:** Enter the description of the task.



69. **Task Type:** Select the type of task. Following are the options:

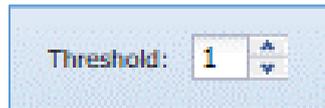


- a) **Detection** - Detects the sensitive data.
- b) **Detection and Notify** – Detects the sensitive data and sends the notification emails to the email IDs mentioned in the **Notify To** field.

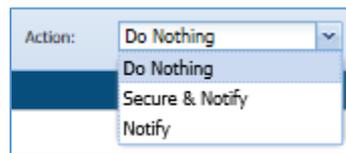
70. **Compliance Policies:** Select the policy based on which, the sensitive data you want to identify.



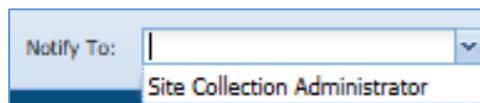
71. **Threshold:** This field will appear only for the Task Type: Detection and Notify. Enter the number of records. The system will perform the action defined in the **Action** field when the number of sensitive records exceeds the number defined in this field.



72. **Action:** This field will appear only for the Task Type: Detection and Notify. Select the action that you want to perform when the number of sensitive records exceeds the number defined in the **Threshold** field. Following are the options:



- a) **Secure & Notify:** Removes access rights of all the users except administrator and sends the notification emails to the email IDs mentioned in the **Notify To** field.
73. **Notify:** Sends the notification emails to the email IDs mentioned in the **Notify To** field.
74. **Notify To:** This field will be enabled when value for **Action:** Secure & Notify and Notify. Select the email IDs to send the notifications. Emails IDs of sharepoint administrator and site owner are already configured. You can also enter the email IDs manually.



75. Click the **Apply** button to apply all the changes made for Threshold, Action and Notify field.
76. Click **Show Hidden/Unsupported** button to view the data types that are not supported or hidden.

77. **Pre-defined and Custom Sensitive Type:** This grid displays all the pre-defined data types. It allows you to define Threshold, Action and Notify field specific to a data type. You can also select the data types to be included for the sensitive data scan.

Pre-defined and Custom Sensitive Types	
Show Hidden/Unsupported	
Name	Description
<input checked="" type="checkbox"/> Credit Card	
<input checked="" type="checkbox"/> Credit Card # (Digits Only)	e.g. 5173215750856134
<input checked="" type="checkbox"/> Credit Card # (Space Separation)	e.g. 5173 2157 5085 6134
<input checked="" type="checkbox"/> Credit Card # (Dash Separation)	e.g. 5173-2157-5085-6134
<input type="checkbox"/> Dates	
<input type="checkbox"/> DOB (Best suited for structured data)	Date Of Birth
<input type="checkbox"/> DOD (Best suited for structured data)	Date Of Death
<input type="checkbox"/> Credit Card Expiry Date	Credit Card Expiry Date
<input type="checkbox"/> Date (Best suited for structured data)	Date
<input type="checkbox"/> Driver License	
<input type="checkbox"/> Driver License (Nebraska)	Driving License for Nebraska State

78. **Select File Types:** Select all the file types that you want to scan. The supported file types are text, excel, powerpoint, pdf, csv, zip, word files.

Name
<input type="checkbox"/> Text Files(*.txt)
<input checked="" type="checkbox"/> Word Files(*.doc, *.docx)
<input checked="" type="checkbox"/> Excel Files(*.xls, *.xlsx)
<input checked="" type="checkbox"/> Powerpoint Files(*.ppt, *.pptx)
<input type="checkbox"/> Adobe PDF Files(*.pdf)

79. **Select Versions To Scan:** Select the data file version to scan. The available options are:

Select Versions To Scan

All

Only Latest

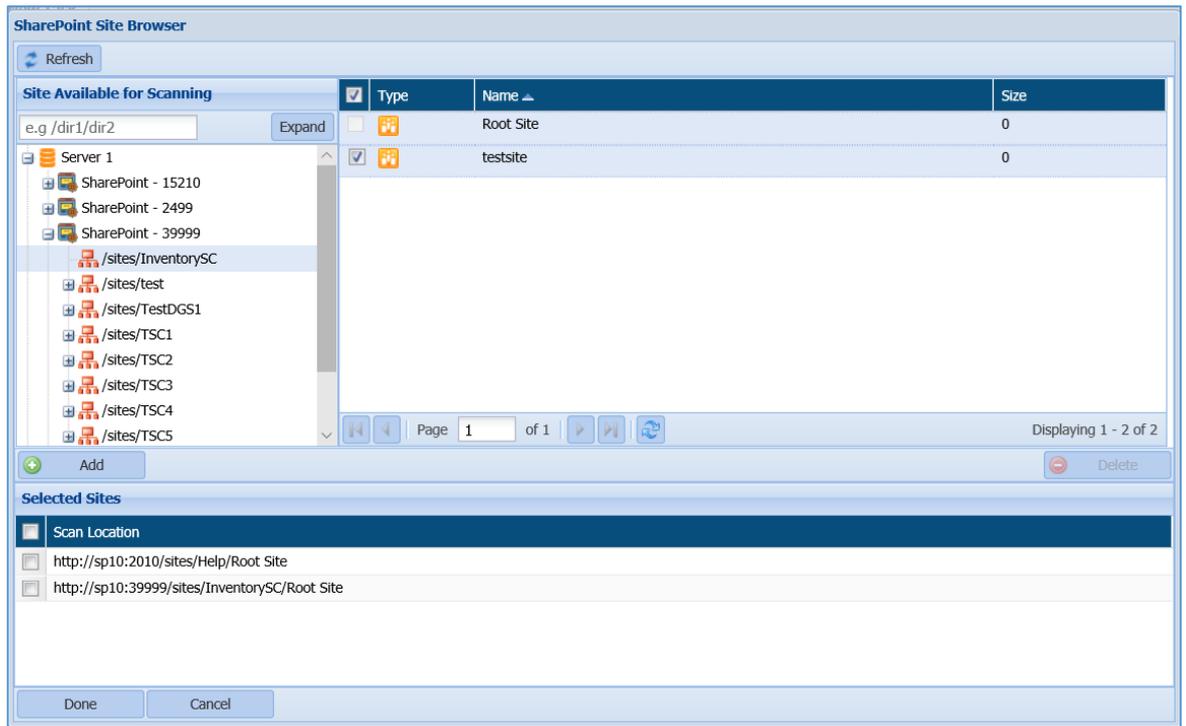
Most Recent

a) **All** - Select this option, if you want to scan the complete dataset.

80. **Only Latest** - Select this option, if you want to scan only the latest version of the dataset.

81. **Most Recent**-Specify the number of recent versions of the dataset that you want to check.

82. Click **Select Sites** button to include all the sites that you want to scan.



- i. Browse the site level from the **Site Available for Scanning** panel.
- ii. To select the sites, check the checkbox corresponding to the site names.
- iii. Click **Add** to include the selected sites in **Selected Sites** panel.
- iv. Select the locations by checking the checkbox next to the scan location name.
- v. Click **Done** button and the selected sites will be available in the **Select Sites** pane.
- vi. If you want to delete any site from the list, select the site and click **Delete**.

83. Click **Save** to save the filters. Click **Save and Execute** to save the filters and start the scan on the dataset. The saved task will be available on the **Task** page.

7.3 List a Task

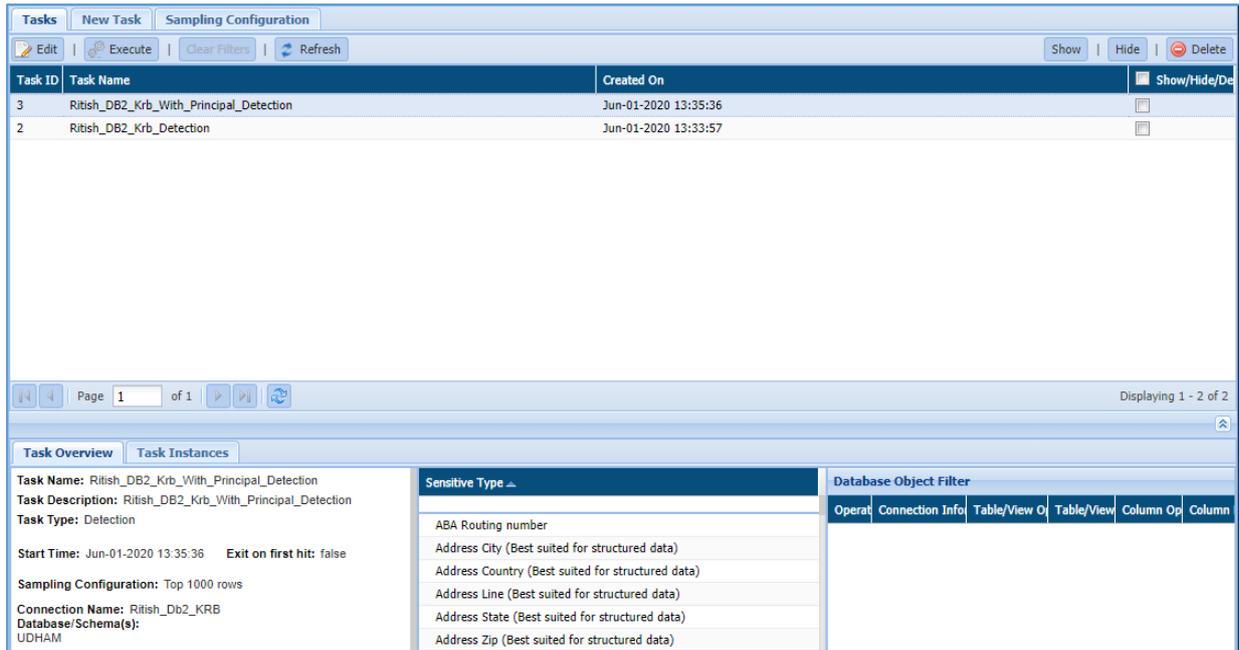
This section will explain the screen which appears when clicked on Task.

7.3.1 RDBMS

In RDBMS, a default screen appears when click on **Task** sub-menu. This page offers in-depth details about the detection tasks that the user has permission to view.

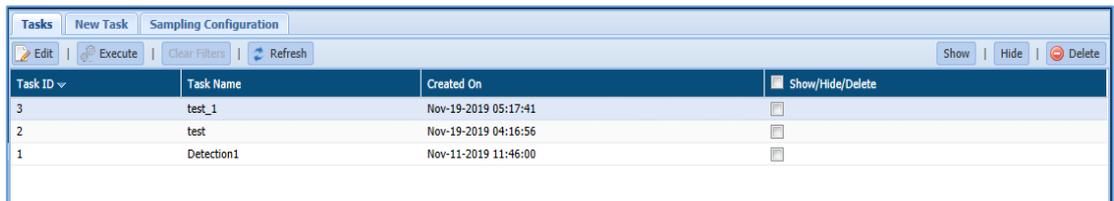
Detection

The below screenshot shows the user interface of the default screen of the Detection Task page.

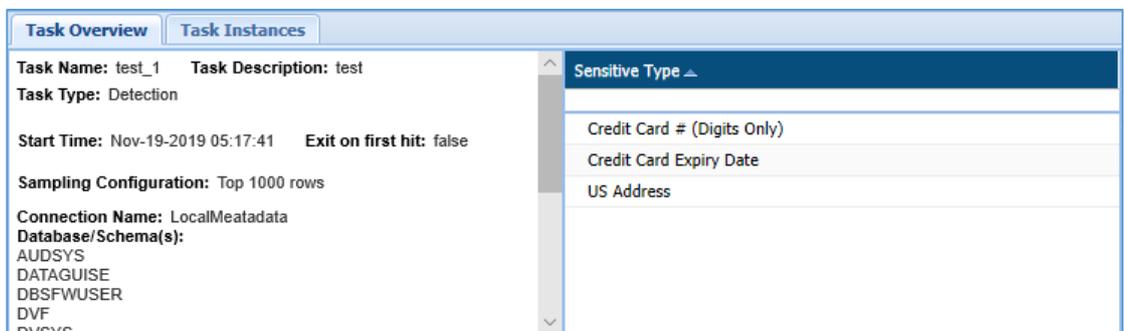


To access the **Task** page from the menu. Click **RDBMS > Detection > Tasks**.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task creation date, system-generated task ID.



84. **Task Overview tab:** It displays the basic information for the selected task. It displays the information for the selected task such as the Task Name, Task Description, Task Type, Sampling Configuration, Connection Name, Database schema, list of Sensitive Type, etc.



- a) **Database Object Filter:** The database Object Filter pane displays only those databases/tables/columns that match the filter applied. The columns display information such as the applied operator, connection information, Table/View operator, Table/View Filter, column operator and column filter.

Database Object Filter					
Operator	Connection In	Table/View	Table/View	Column Oper	Column Fil
	SQLSrvDis...	not equ...	Person...	not equal to	Demog...
AND	SQLSrvDis...	not equ...	dbo.Da...	not equal to	XmlEvent

- 85. **Task Instances tab:** it displays the related information about each instance of the task selected in the detection task panel. Information such as Task Name, Task ID (system generated), Status of the task, Start time and End time.

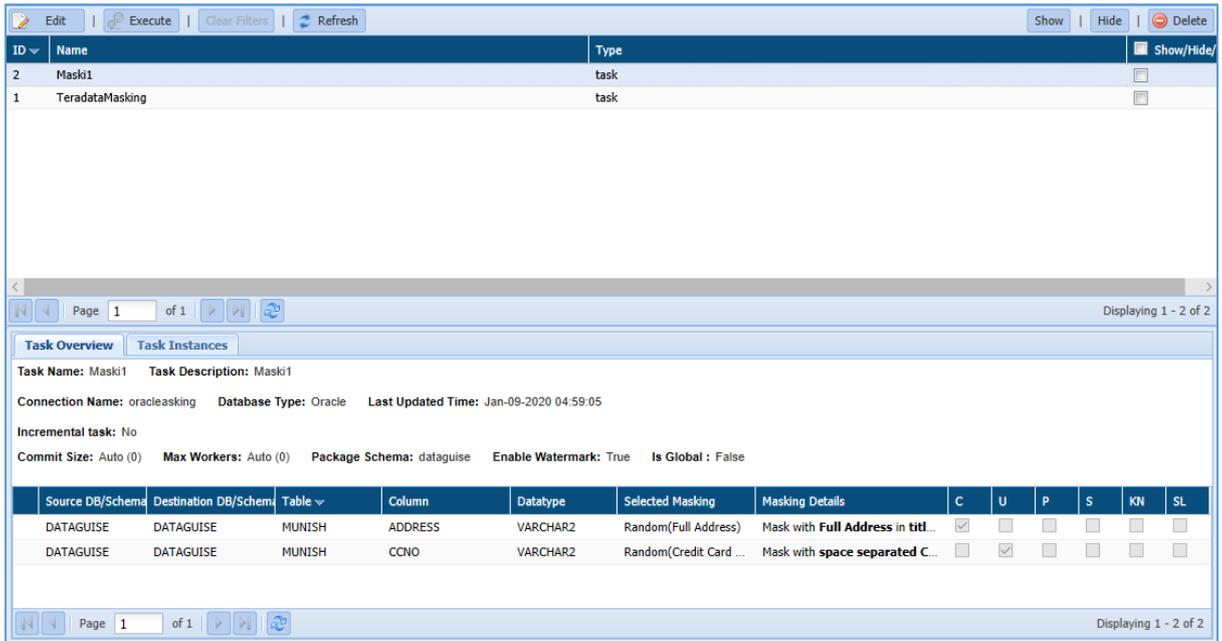
Task Overview		Task Instances			
ID	Task Name	Status	Start Time	End Time	
20	test_1	Completed	Nov-26-2019 23:46:21	Nov-26-2019 23:46:41	
19	test_1	Completed	Nov-26-2019 23:45:58	Nov-26-2019 23:46:19	
18	test_1	Cancelled	Nov-26-2019 03:49:07	Nov-26-2019 03:49:28	

The status of the task can be in any of the listed states:

- i. Initializing
- ii. Completed
- iii. Running
- iv. Failed
- v. Stopped

Masking

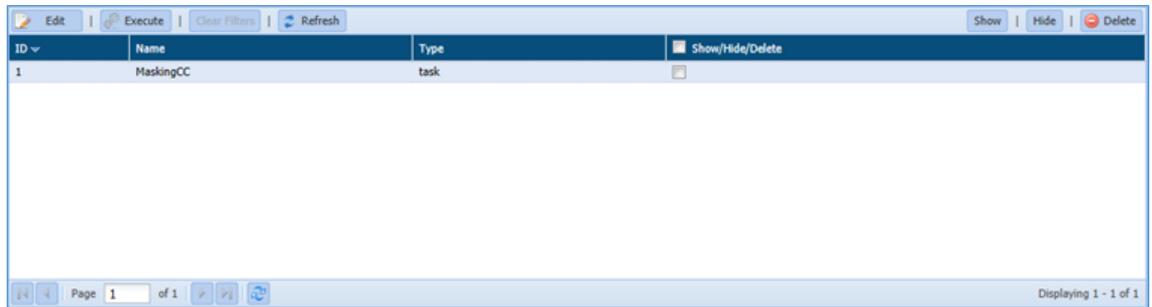
The below screenshot shows the user interface of the default screen of the Masking Task page.



The Tasks/Template screen is divided into three panels. These are:

1. Masking Tasks/Template:

The Masking Tasks/Template panel shows the tasks and the properties associated with each masking task. The task properties include details such as Task ID (system generated), Task Name, Type, etc.



86. Task Overview

The Task Overview pane displays the details for the selected task. It includes information such as Task Name, Task Description, Database Type, Connection Name, etc. It also displays information about the masking details for the Sensitive Type such as SourceDB/Schema, Table, Column, Datatype, Selected Masking, Masking Details, CUPS option, etc.

Source DB/Schema	Destination DB/Schema	Table	Column	Datatype	Selected Masking	Masking Details	C	U	P	S	KN	SL
Data	Data	dbo.Data2	Description	nvarchar(MAX)	Custom	Mask by function [DG].CUSTO...	<input type="checkbox"/>					

87. Task Instances

The Task Instances pane displays the information about each instance of the task selected in the Detection Tasks panel. Information includes such as Task ID (system generated), Task Name, Status, Start Time and, End Time.

ID	Task Name	Start Time	End Time	Status
377	MaskingCC	Nov-18-2019 23:12:46	Nov-18-2019 23:12:57	Completed
376	MaskingCC	Nov-18-2019 23:11:44	Nov-18-2019 23:11:56	Completed
375	MaskingCC	Nov-18-2019 23:10:59	Nov-18-2019 23:11:10	Completed
308	MaskingCC	Oct-07-2019 19:16:59	Oct-07-2019 19:17:30	Completed

7.3.2 Hadoop

7.3.2.1 HDFS

1. Task Details

The top panel of the HDFS Tasks page lists all the tasks that you have permission to view. When you select a task in this list, information about the task is displayed in the tabbed panel below. Click the HDFS Browser button to view the directories and files in the targeted cluster.

Select Cluster: Hadoop Cluster

Tasks | New Task | Sampling Configuration

Edit | Execute | Clear Filters | HDFS Browser | Refresh | Cluster Status: Active | Show | Hide | Delete

Task ID	Task Name	Created On	Last Run On	Created By	Show/Hide/De
1	Test_detection	Dec-11-2019 05:44:44	Dec-16-2019 11:36:21	DocsDemo	

Page 1 of 1 | Displaying 1 - 1 of 1

Task Overview | Task Instances

Task Name: Test_detection | Task Description: Detection | Created By: DocsDemo | Last Executed On: Dec-16-2019 11:36:21 | Scan Type: Full | Dump Metadata: false

Task Type: Detection | Read Files: Part of Files | Sampling Configuration: Top 1000 rows | Include Files that failed previously: False | Files Modified After: NA | Files Modified Before: NA

Batch Size/Files: 30 | Min Batch Size(MB): NA

Included Scan Locations	Sensitive Type Details
Scan Location	Sensitive Type Name
/munish	URL
	Social Security # (Space Separation)
	Social Security # (Dash Separation)
	Telephone (Space Separation)
	Telephone (Dash Separation)
	Telephone (Standard)
	Telephone (Standard without spaces)

Page 1 of 1 | Displaying 1 - 1 of 1

88. Task Overview

For this tab, the panel displays the date and time that the task was last executed on, and details the task's composition. The displayed parameters include:

- a) The types of sensitive data that are searched for (e.g., social security and credit card numbers).

89. The type of task that is performed — discovery, masking, encryption by row, encryption by cell, or decryption.

90. The directory path details.

91. Task Instances.

Task Overview | Task Instances

Task Name: Test_detection | Task Description: Detection | Created By: DocsDemo | Last Executed On: Dec-16-2019 11:36:21 | Scan Type: Full | Dump Metadata: false

Task Type: Detection | Read Files: Part of Files | Sampling Configuration: Top 1000 rows | Include Files that failed previously: False | Files Modified After: NA | Files Modified Before: NA

Batch Size/Files: 30 | Min Batch Size(MB): NA

Included Scan Locations	Sensitive Type Details
Scan Location	Sensitive Type Name
/munish	URL
	Social Security # (Space Separation)
	Social Security # (Dash Separation)
	Telephone (Space Separation)
	Telephone (Dash Separation)
	Telephone (Standard)
	Telephone (Standard without spaces)

Page 1 of 1 | Displaying 1 - 1 of 1

92. Task Instance

When a task is executed, the resulting scan, with its unique start time and results, is called a task instance. For this tab, the panel tracks these instances ordered by the date of execution. For each instance, the tab displays:

- a) The task name.
- 93. The start and finish times.
- 94. The status (Started, Running, Completed, or Canceled).

Task Overview		Task Instances									
ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Hostname	Instance IP Address	Show
23	Test_detection	Completed	Dec-16-2019 11:3...	Dec-16-2019 11:3...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
20	Test_detection	Completed	Dec-11-2019 05:4...	Dec-11-2019 05:4...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
19	Test_detection	Completed	Dec-11-2019 05:4...	Dec-11-2019 05:4...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
18	Test_detection	Failed at 0.00%	Dec-11-2019 05:3...	Dec-11-2019 05:3...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
17	Test_detection	Paused at 0.00%	Dec-11-2019 05:2...	Dec-11-2019 05:3...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
16	Test_detection	Paused at 0.00%	Dec-11-2019 05:1...	Dec-11-2019 05:2...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
15	Test_detection	Paused at 0.00%	Dec-11-2019 05:0...	Dec-11-2019 05:1...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
8	Test_detection	Paused at 0.00%	Dec-11-2019 04:4...	Dec-11-2019 05:0...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
7	Test_detection	Paused at 0.00%	Dec-11-2019 04:4...	Dec-11-2019 04:4...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	
6	Test_detection	Failed at 0.00%	Dec-11-2019 04:3...	Dec-11-2019 04:3...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.dg...	10.12.13.11	

7.3.2.2 Hive

1. Task Details

The top panel of the Hive Tasks page lists all the tasks that you have permission to view. When you select a task in this list, information about the task is displayed in the panels below.

Select Cluster: hdfsCluster

Hive Tasks New Task

Edit | Execute | Clear Filters | Refresh | Cluster Status Active Show | Hide | Delete

Task ID	Task Name	Created On	Last Run On	Created By	Show/Hide/De
8	pPection_M	Dec-18-2019 15:49:44	Dec-18-2019 15:49:53	dataguisse	<input type="checkbox"/>
6	pProtection_M	Dec-18-2019 15:49:19		dataguisse	<input type="checkbox"/>
5	hiveRestAPI/task_percentage1	Dec-18-2019 15:46:39	Dec-18-2019 15:47:05	dataguisse	<input type="checkbox"/>
3	Hive_P	Dec-17-2019 16:41:38	Dec-17-2019 17:58:38	dataguisse	<input type="checkbox"/>
1	Hive_test	Dec-17-2019 16:39:48	Dec-17-2019 17:57:48	dataguisse	<input type="checkbox"/>

Page 1 of 1 Displaying 1 - 5 of 5

Task Details

Task Name: Hive_test **Task Description:** Hive_test **Created By:** dataguisse **Last Executed On:** Dec-17-2019 17:57:48 **Task Type:** Detection

Policies Selected: HIPAA_Hadoop,PCI_Hadoop
Sampling By: Number of Rows/Map **Sampling Value:** 1000

Table Selected	Sensitive Types Detail
Table Name	Sensitive Types
external_text_table	Telephone (Dot Separation) URL Social Security # (Space Separation) Social Security # (Dash Separation) Telephone (Space Separation) Telephone (Dash Separation) Telephone (Standard) Telephone (Standard without spaces) Full Names

95. Task Overview

For this tab, the panel displays the date and time that the task was last executed on, and details the task's composition. The displayed parameters include:

- The types of sensitive data that are searched for (e.g., social security and credit card numbers).
- The type of task that is performed — discovery, masking, encryption by row, encryption by cell, or decryption.
- The directory path details.
- Task Instances.

Task Details

Task Name: Hive_test **Task Description:** Hive_test **Created By:** dataguisse **Last Executed On:** Dec-17-2019 17:57:48 **Task Type:** Detection

Policies Selected: HIPAA_Hadoop,PCI_Hadoop
Sampling By: Number of Rows/Map **Sampling Value:** 1000

96. Table Selected

Lists the tables selected for the scan under the selected task and provides details of the selected table name, column name, Sensitive Type, data type and Domain name.

Table Selected				
Table Name	Column Name	Sensitive Type	Data Type	Domain Name
external_text_table	ccno_digits	Credit Card # (Digi...	STRING	HDFS

97. Sensitive Type Detail

Lists all the sensitive types selected in the selected task.

Sensitive Types Detail
Sensitive Types
Credit Card # (Digits Only)
Credit Card # (Space Separation)
Credit Card # (Dash Separation)

7.3.2.3 Hbase

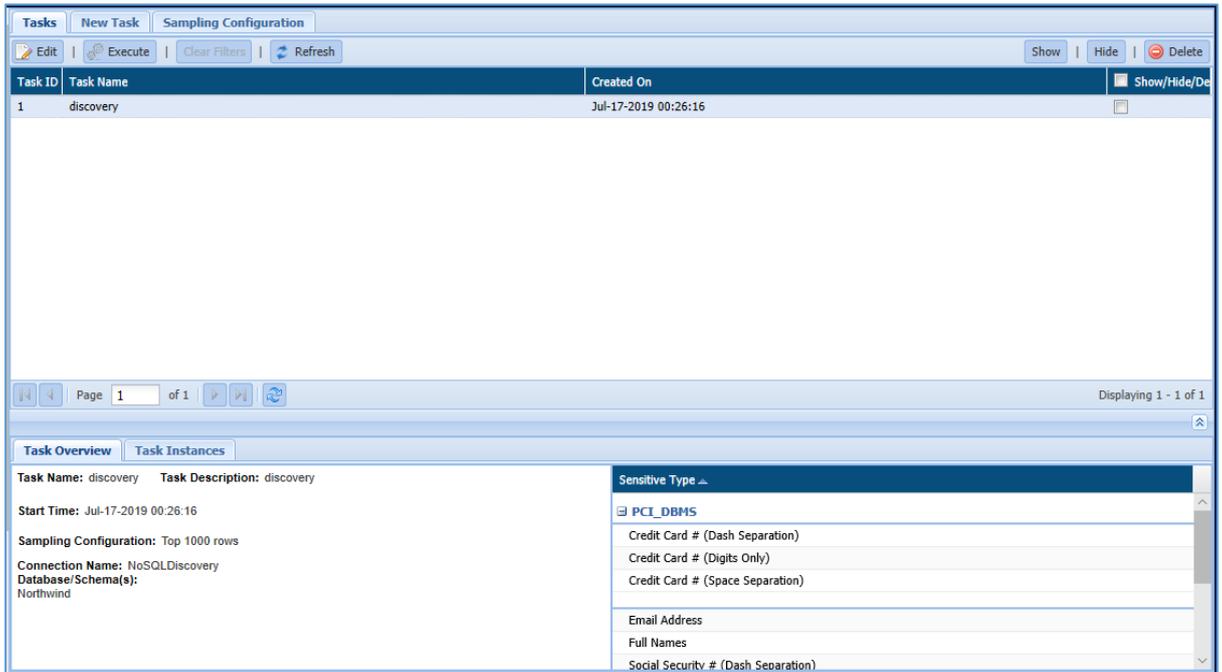
To view all the existing tasks, go to **HBASE> HBASE Tasks**.

Hbase Tasks		New Task						SHOW	HIDE
TASKID ↓	TASK NAME	CREATED ON	LAST RUN ON	CREATED BY	ACTION			SHOW/HIDE	
5	task_1_demo	Dec-06-2019 10:51:17	Dec-06-2019 10:51:26	d	👁	📄	⚙	<input type="checkbox"/>	

7.3.3 NoSQL

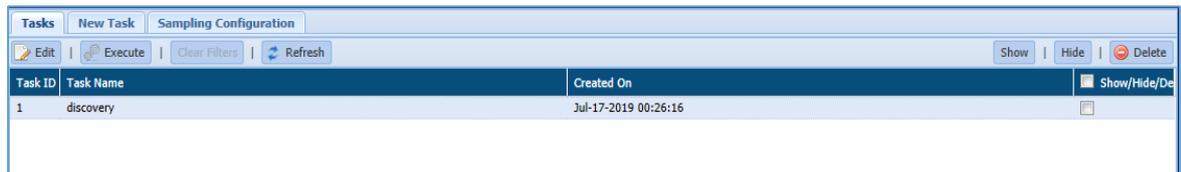
In NoSQL, screen appears when click on **Tasks** sub-menu. This page offers in-depth details about the tasks that the user has permission to view.

The below image shows the user interface of the screen of the **Tasks** page.

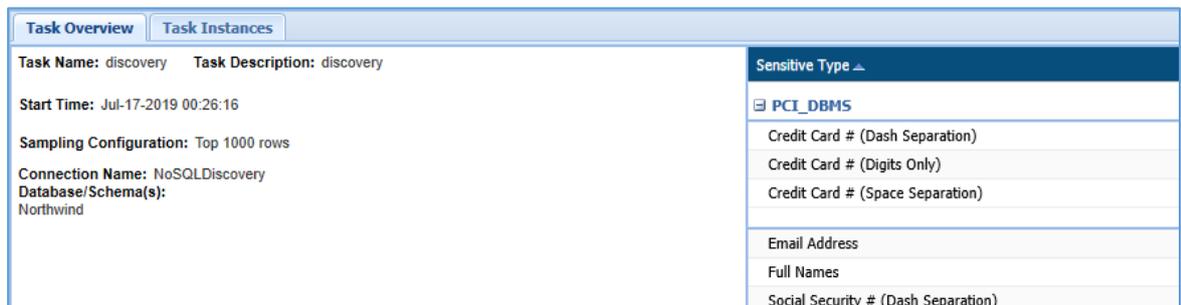


To access the **Task** page. Click **NoSQL > Detection > Tasks**.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task creation date, system-generated task ID.



98. **Task Overview tab:** It displays the basic information for the selected task. It displays information such as the Task Name, Task Description, Sampling Configuration, Connection Name, Database schema, list of Sensitive Type, etc.



99. **Task Instances tab:** it displays the information about each instance of the task selected in the Tasks panel. Information such as Task Name, Task ID (system generated), Status of the task, Start time and End time.

ID	Task Name	Status	Start Time	End Time
407	discovery	Started	Dec-05-2019 06:55:46	
158	discovery	Completed	Aug-05-2019 18:55:54	Aug-05-2019 18:56:28
157	discovery	Failed	Aug-05-2019 18:45:34	Aug-05-2019 00:00:00
127	discovery	Completed	Jul-20-2019 18:29:47	Jul-20-2019 18:30:11
126	discovery	Completed	Jul-20-2019 18:24:55	Jul-20-2019 18:25:19
125	discovery	Completed	Jul-20-2019 18:22:35	Jul-20-2019 18:22:59
124	discovery	Failed	Jul-20-2019 18:12:51	Jul-20-2019 00:00:00

The status of the task can be in any of the listed states:

- i. Completed
- ii. Running
- iii. Failed
- iv. Stopped

7.3.4 Files

In Files, screen appears when click on **Tasks**. This page offers in-depth details about the tasks that the user has permission to view.

The below image shows the user interface of the screen.

Task ID	Task Name	Created On	Last Run On	Created By	Show/Hide/
41	f_bug	Dec-05-2019 12:56:13	Dec-05-2019 12:59:24	d	<input type="checkbox"/>
40	files_150mb_entire	Dec-05-2019 12:50:52	Dec-05-2019 12:50:55	d	<input type="checkbox"/>
39	files_150mb	Dec-05-2019 12:50:36	Dec-05-2019 12:50:37	d	<input type="checkbox"/>
28	det_all_files_limit	Dec-04-2019 15:09:38	Dec-04-2019 15:09:43	d	<input type="checkbox"/>
24	lfa_dec	Dec-04-2019 14:27:55	Dec-04-2019 14:27:56	d	<input type="checkbox"/>
23	lfa_row	Dec-04-2019 14:22:05	Dec-04-2019 14:22:09	d	<input type="checkbox"/>
22	lfa_maks	Dec-04-2019 14:20:29	Dec-04-2019 14:20:30	d	<input type="checkbox"/>
13	det_all_files	Dec-04-2019 13:23:06	Dec-04-2019 13:23:08	d	<input type="checkbox"/>
9	t_non	Dec-04-2019 13:05:03	Dec-04-2019 13:05:04	d	<input type="checkbox"/>
8	test1	Dec-04-2019 13:01:29	Dec-04-2019 13:01:31	d	<input type="checkbox"/>
4	test	Dec-04-2019 12:50:20	Dec-04-2019 12:50:24	d	<input type="checkbox"/>

Task Name: f_bug	Task Description: test	Created By: d	Last Executed On: Dec-05-2019 12:59:24	Scan Type: Full	Dump Metadata: false
Task Type: Detection	Read Files: Part of Files	Sampling Configuration: Top 1000 rows	Include Files that failed previously: False	Files Modified After: NA	Files Modified Before: NA

Scan Location	Sensitive Type Name
/tmp/rochak/cc	Bug17424

To access the **Task** page. Click **Files > Tasks > Tasks** tab.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task such as Task ID, Task Name, creation date, Created by, etc.

Task ID	Task Name	Created On	Last Run On	Created By	Show/Hide
41	f_bug	Dec-05-2019 12:56:13	Dec-05-2019 12:59:24	d	<input type="checkbox"/>
40	files_150mb_entire	Dec-05-2019 12:50:52	Dec-05-2019 12:50:55	d	<input type="checkbox"/>
39	files_150mb	Dec-05-2019 12:50:36	Dec-05-2019 12:50:37	d	<input type="checkbox"/>
28	det_all_files_limit	Dec-04-2019 15:09:38	Dec-04-2019 15:09:43	d	<input type="checkbox"/>
24	ifa_dec	Dec-04-2019 14:27:55	Dec-04-2019 14:27:56	d	<input type="checkbox"/>
23	ifa_row	Dec-04-2019 14:22:05	Dec-04-2019 14:22:09	d	<input type="checkbox"/>

100. **Task Overview:** It displays the basic information for the selected task. It displays information such as Task Name, Task Description, Created by, Last executed on, Sampling Configuration, etc.

101. **Task Overview** tab also includes Included Scan Location and Exclude From Scan tab, Sensitive Type Details panel.

- a) **Included Scan Location:** This window will display list of directory location from where the files are scanned.
- b) **Exclude from Scan:** This tab will display File Extension and Scan Location pane.
- c) **Sensitive Type Details:** This tab displays the list of Sensitive Type for the selected task.

Task Overview		Task Instances	
Task Name: f_bug	Task Description: test	Created By: d	Last Executed On: Dec-05-2019 12:59:24
Task Type: Detection	Read Files: Part of Files	Sampling Configuration: Top 1000 rows	Include Files that failed previously: False
Files Modified After: NA	Files Modified Before: NA	Dump Metadata: false	
Included Scan Locations		Sensitive Type Details	
Exclude From Scan		Sensitive Type Name	
Scan Location		Sensitive Type Name	
/tmp/rochak/cc		Bug17424	
Page 1 of 1		Displaying 1 - 1 of 1	

102. **Task Instances:** it displays the information about each instance of the task selected in the Tasks panel. Information such as Task ID, Task Name, Status, Start time and End time, Executed By, IDP Hostname, IDP IP Address, etc.

Task Overview		Task Instances									
ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Host	Instance IP Ad	Show/Hide
85	f_bug	Completed	Dec-05-2019 1...	Dec-05-2019 1...	d	root	rack160-hdp26-qa	192.168.0...	rack160-hd...	192.168.0...	<input type="checkbox"/>
84	f_bug	Completed	Dec-05-2019 1...	Dec-05-2019 1...	d	root	rack160-hdp26-qa	192.168.0...	rack160-hd...	192.168.0...	<input type="checkbox"/>
83	f_bug	No Conforming Files	Dec-05-2019 1...		d	root	rack160-hdp26-qa	192.168.0...	rack160-hd...	192.168.0...	<input type="checkbox"/>
82	f_bug	No Conforming Files	Dec-05-2019 1...		d	root	rack160-hdp26-qa	192.168.0...	rack160-hd...	192.168.0...	<input type="checkbox"/>

Page 1 of 1 Displaying 1 - 4 of 4

The status of the task can be in any of the listed states:

- i. Completed
- ii. Running
- iii. Failed
- iv. Stopped

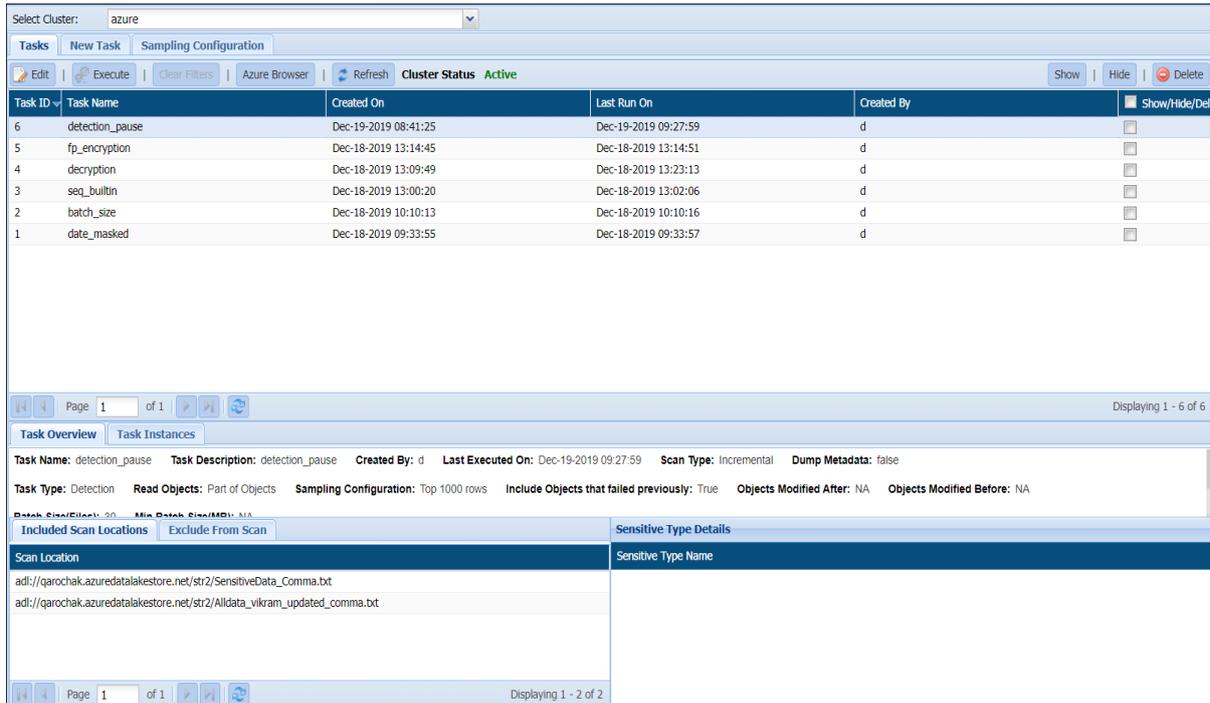
7.3.5 Azure

1. Task

Use Azure tasks to define and launch tasks that search for, and optionally mask or encrypt sensitive information within an HDFS cluster. Create tasks with the information to search for (such as credit card numbers and names), the locations to search in, and the masking option to apply.

103. Task Details

The top panel of the Azure Tasks page lists all the tasks that you have permission to view. When you select a task in this list, information about the task is displayed in the tabbed panel below.

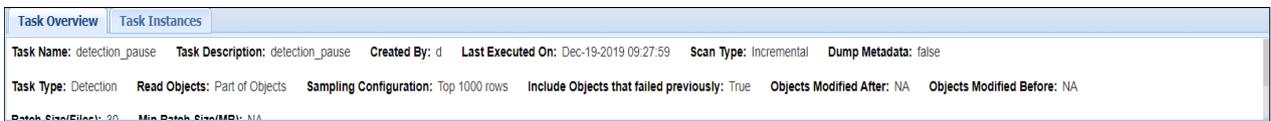


104. Task Overview

For this tab, the panel displays the date and time that the task was last executed on, and details the task's composition. The displayed parameters include:

- a) The types of sensitive data that are searched for (e.g., social security and credit card numbers).

- 105. The type of task that is performed — discovery, masking, encryption by row, encryption by cell, or decryption.
- 106. The directory path details.
- 107. Task Instances.



Following tabs get populated on the screen when the Task Overview tab is selected:

- a) **Include Scan Locations and Exclude From Scan**
The locations added to the Include list. When the Included Scan Location tab is selected, the instance's target directory, the structure assigned to the directory, and the domain to which it is assigned are shown. Not all scan locations are associated with a structure or domain.

108. The locations added to the Exclude list. When the Exclude from Scan tab is selected, the panel displays any excluded file extensions and file paths.

Included Scan Locations			Exclude From Scan		
Scan Location	Structure	Domain			
adl://qarochak.azuredatalakestore.net/swati3/builtinFile_longw...	seq1	sw			

109. **Sensitive Type Details**

The Sensitive Type Details panel shows the task instance's sensitive types.

Sensitive Type Details	
Sensitive Type Name	
URL	
Social Security # (Space Separation)	
Social Security # (Dash Separation)	
Telephone (Space Separation)	
Telephone (Dash Separation)	

110. **Task Instance**

When a task is executed, the resulting scan, with its unique start time and results, is called a task instance. For this tab, the panel tracks these instances ordered by the date of execution. For each instance, the tab displays:

- a) The task name.
- 111. The start and finish times.
- 112. The status (Started, Running, Completed, or Canceled).

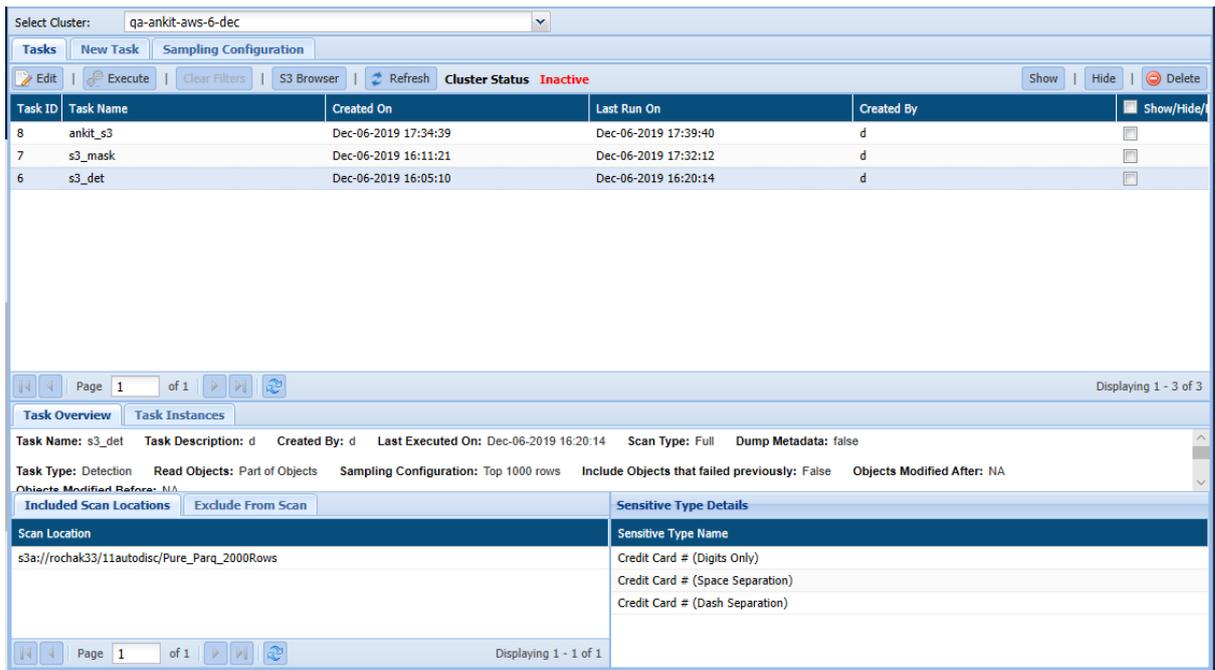
Task Overview		Task Instances									
ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Hostname	Instance IP Address	Show/Hide
6	fp_encryption	Failed	Dec-18-2019 13:1...	Dec-18-2019 13:1...	d	hdfs	hn0-qa-tea	10.0.0.26	maninderlinux	10.0.0.7	<input type="checkbox"/>

7.3.6 AWS

7.3.6.1.1 S3

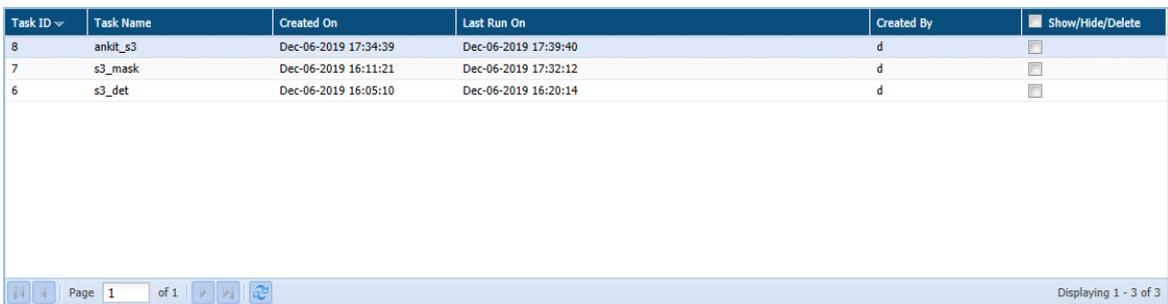
In S3, screen appears when clicked on **Tasks**. This page offers in-depth details about the tasks that the user has permission to view.

The below image shows the user interface of the screen.

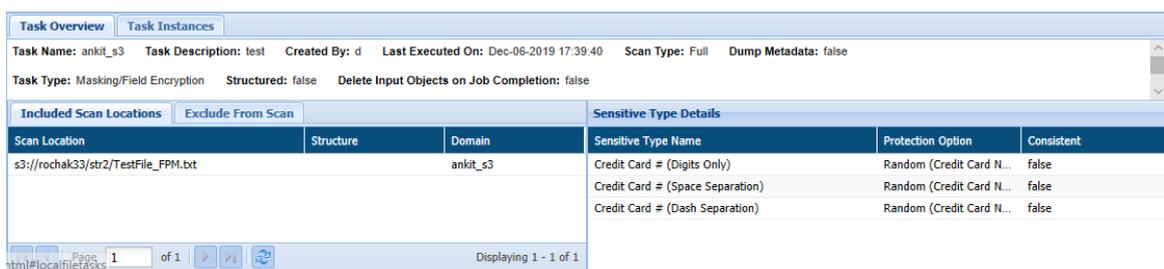


To access the **Task** page. Click **AWS > S3 > Tasks** tab.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task such as Task ID (system generated), Task Name, Created On, Last Run On, Created by, etc.



113. **Task Overview:** It displays the details for the selected task. It includes information such as Task Name, Task Description, Created by, Last Executed On, Sampling Configuration, Scan Type, etc.



Task Overview tab also includes **Included Scan Location** tab, **Exclude From Scan** tab, and **Sensitive Type Details** pane.

- a) **Included Scan Location:** This pane displays the list of included files for the selected scan locations. Information includes Scan Location, Structure, Domain.

Included Scan Locations		Exclude From Scan
Scan Location	Structure	Domain
s3://rochak33/str2/TestFile_FPM.txt		ankit_s3

Page 1 of 1 Displaying 1 - 1 of 1

- 114. **Exclude from Scan:** This pane displays the list of excluded file extensions and scan locations of an excluded object.

Included Scan Locations		Exclude From Scan
Object Extension	Scan Location	

- 115. **Sensitive Type Details:** This tab displays the list of Sensitive Type for the selected task.

Sensitive Type Details		
Sensitive Type Name	Protection Option	Consistent
Credit Card # (Digits Only)	Random (Credit Card N...	false
Credit Card # (Space Separation)	Random (Credit Card N...	false
Credit Card # (Dash Separation)	Random (Credit Card N...	false

- 116. **Task Instances:** it displays the information about each instance of the task selected in the Detection Tasks panel. Information such as Task ID (system generated), Task Name, Status, Start Time and End Time, Executed By, User, IDP Hostname, IDP IP Address, etc.

ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Host	Instance IP Ad	Show/H
269	s3_mask	Completed	Dec-06-2019 1...	Dec-06-2019 1...	d	hdfs	ip-172-31-31-8	172.31.31.8	rack160-hd...	192.168.0...	<input type="checkbox"/>
268	s3_mask	Completed	Dec-06-2019 1...	Dec-06-2019 1...	d	hdfs	ip-172-31-31-8	172.31.31.8	rack160-hd...	192.168.0...	<input type="checkbox"/>
233	s3_mask	Completed	Dec-06-2019 1...	Dec-06-2019 1...	d	hdfs	ip-172-31-31-8	172.31.31.8	rack160-hd...	192.168.0...	<input type="checkbox"/>

Page 1 of 1 Displaying 1 - 3 of 3

The status of the task can be in any of the listed states:

- i. Initializing
- ii. Completed
- iii. Running
- iv. Failed
- v. Stopped

7.3.6.1.2 RDS/RedShift

In RDS/RedShift, you can check tasks for both Detection and Masking module.

Detection

In RDS/RedShift, screen appears when clicked on **Tasks**. This page offers in-depth details about the tasks that the user has permission to view.

The below image shows the user interface of the screen.

The screenshot displays the 'Tasks' page in the Dataguise interface. At the top, there are tabs for 'Tasks', 'New Task', and 'Sampling Configuration'. Below the tabs are action buttons: 'Edit', 'Execute', 'Clear Filters', 'Refresh', 'Show', 'Hide', and 'Delete'. The main table lists tasks with columns for Task ID, Task Name, and Created On. Two tasks are visible: 'rds_schedule_det' (ID 11) and 'detect_allPrimaryDBs' (ID 9). Below the table, there is a detailed view for the selected task 'rds_schedule_det'. This view includes fields for Task Name, Task Description, Task Type (Detection), Start Time, Exit on first hit, Sampling Configuration, and Connection Name. A 'Sensitive Type' dropdown menu is open, showing options like 'Credit Card # (Dash Separation)', 'Credit Card # (Digits Only)', and 'Credit Card # (Space Separation)'. There is also a 'Database Object Filter' section with columns for Operat, Connection Info, Table/View O, Table/View, Column Op, and Column.

To access the **Task** page. Click **AWS > RDS/RedShift > Detection > Tasks** tab.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task such as Task ID (system generated), Task Name, Created On, Last Run On, Created by, etc.

Task ID	Task Name	Created On	Show/Hide/De
11	rds_schedule_det	Dec-06-2019 10:35:01	<input type="checkbox"/>
9	detect_allPrimaryDBs	Dec-05-2019 13:27:18	<input type="checkbox"/>

Page 1 of 1

Displaying 1 - 2 of 2

117. **Task Overview:** It displays the details for the selected task. It includes information such as Task Name, Task Description, Task Type, Exit on first hit, Start Time, Sampling Configuration, etc.

Task Overview | Task Instances

Task Name: rds_schedule_det | Task Description: d

Task Type: Detection

Start Time: Dec-06-2019 10:35:01 | Exit on first hit: false

Sampling Configuration: Top 1000 rows

Connection Name: sql_rds_d
Database/Schema(s): Sankush

Connection Name: pg_rds_d
Database/Schema(s):

Sensitive Type

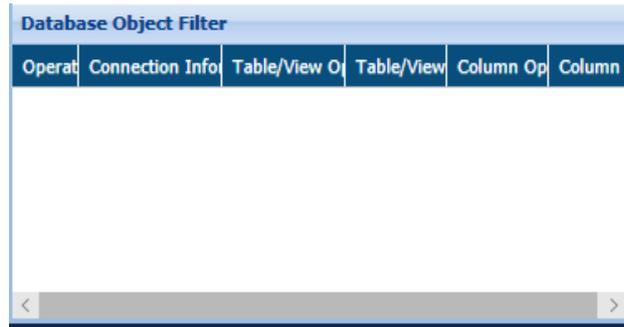
- PCI_DBMS
 - Credit Card # (Dash Separation)
 - Credit Card # (Digits Only)
 - Credit Card # (Space Separation)

Task Overview tab also includes **Sensitive Type Details** and **Database Object Filter** pane.

- a) **Sensitive Type Details:** This tab displays the list of Sensitive Type for the selected task.

Sensitive Type Details		
Sensitive Type Name	Protection Option	Consistent
Credit Card # (Digits Only)	Random (Credit Card N...	false
Credit Card # (Space Separation)	Random (Credit Card N...	false
Credit Card # (Dash Separation)	Random (Credit Card N...	false

118. **Database Object Filter:** The database Object Filter pane displays only those databases/tables/columns that match the filter applied. The columns display information such as the applied operator, connection information, Table/View operator, Table/View Filter, column operator and column filter.



119. **Task Instances:** it displays the information about each instance of the task selected in the Detection Tasks panel. Information includes such as Task ID (system generated), Task Name, Status, Start Time and, End Time and Show/Hide checkbox.

ID	Task Name	Status	Start Time	End Time	Show/Hide
935	rds_schedule_det	Failed	Dec-08-2019 10:37:09	Dec-08-2019 10:42:43	<input type="checkbox"/>
541	rds_schedule_det	Failed	Dec-07-2019 10:37:43	Dec-07-2019 10:43:17	<input type="checkbox"/>
162	rds_schedule_det	Completed	Dec-06-2019 10:40:56	Dec-06-2019 10:44:09	<input type="checkbox"/>

Page 1 of 1 Displaying 1 - 3 of 3

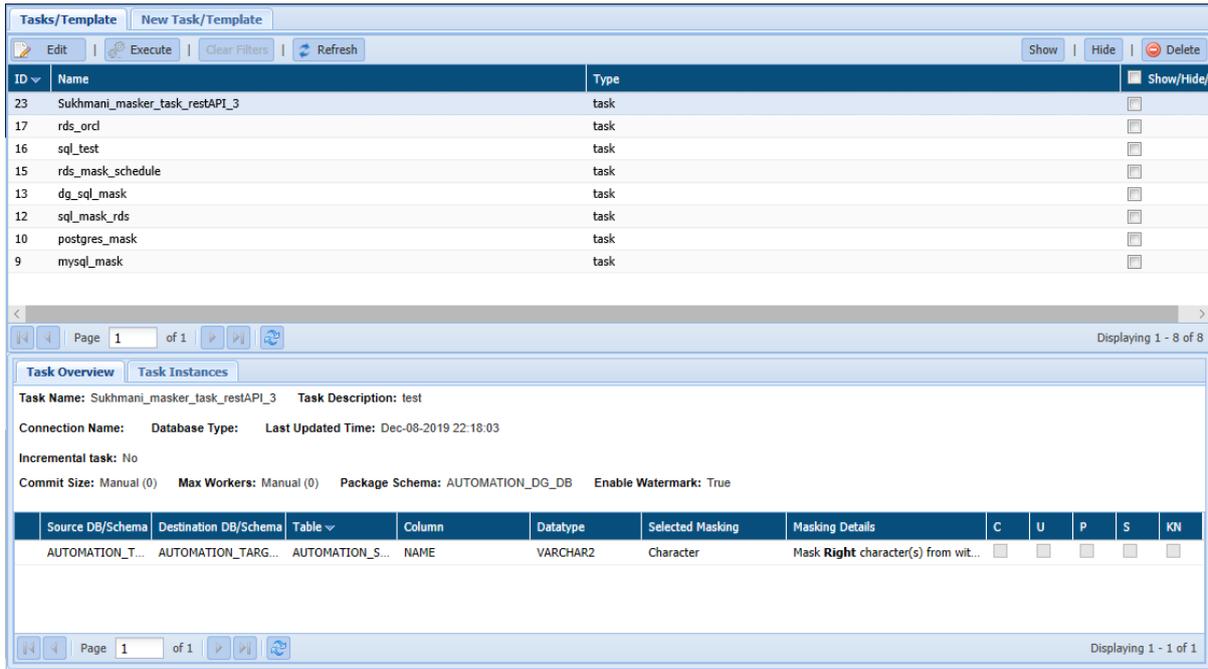
The status of the task can be in any of the listed states:

- i. Initializing
- ii. Completed
- iii. Running
- iv. Failed
- v. Stopped

Masking

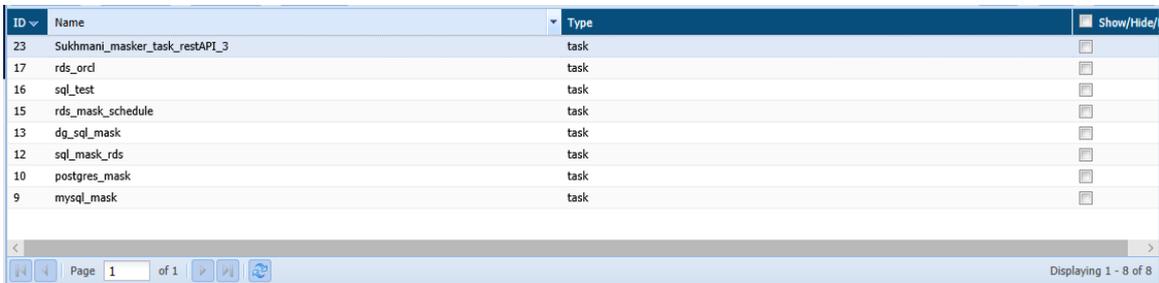
In RDS/RedShift, screen appears when clicked on **Tasks**. This page offers in-depth details about the tasks that the user has permission to view.

The below image shows the user interface of the screen.



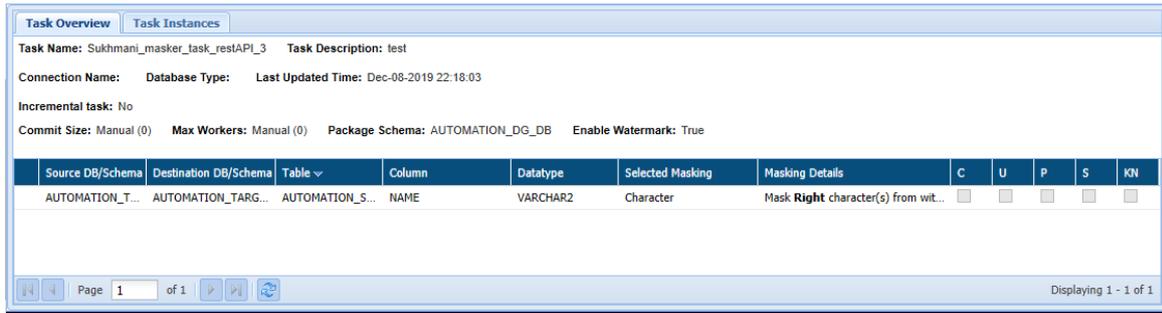
To access the **Task** page. Click **AWS > RDS/RedShift > Detection > Tasks** tab.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task such as Task ID (system generated), Task Name, Type, etc.

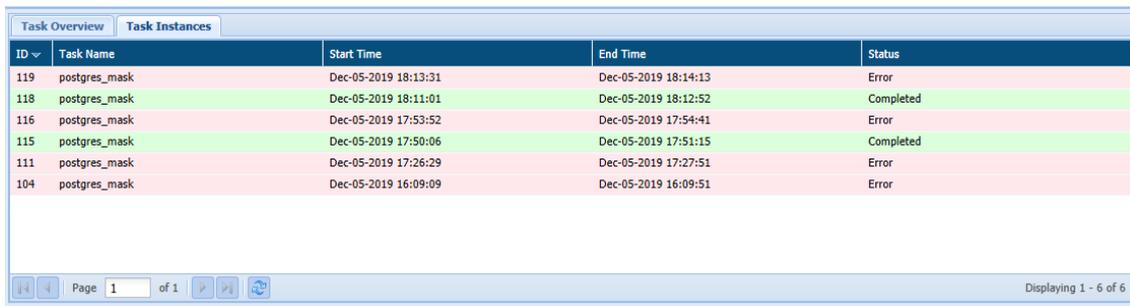


120. **Task Overview:** It displays the details for the selected task. It includes information such as Task Name, Task Description, Database Type, etc.

It also displays information about the database such as SourceDB/Schema, Table, Column, Datatype, Masking Details, etc.



121. **Task Instances:** it displays the information about each instance of the task selected in the Detection Tasks panel. Information includes such as Task ID (system generated), Task Name, Status, Start Time and, End Time.



The status of the task can be in any of the listed states:

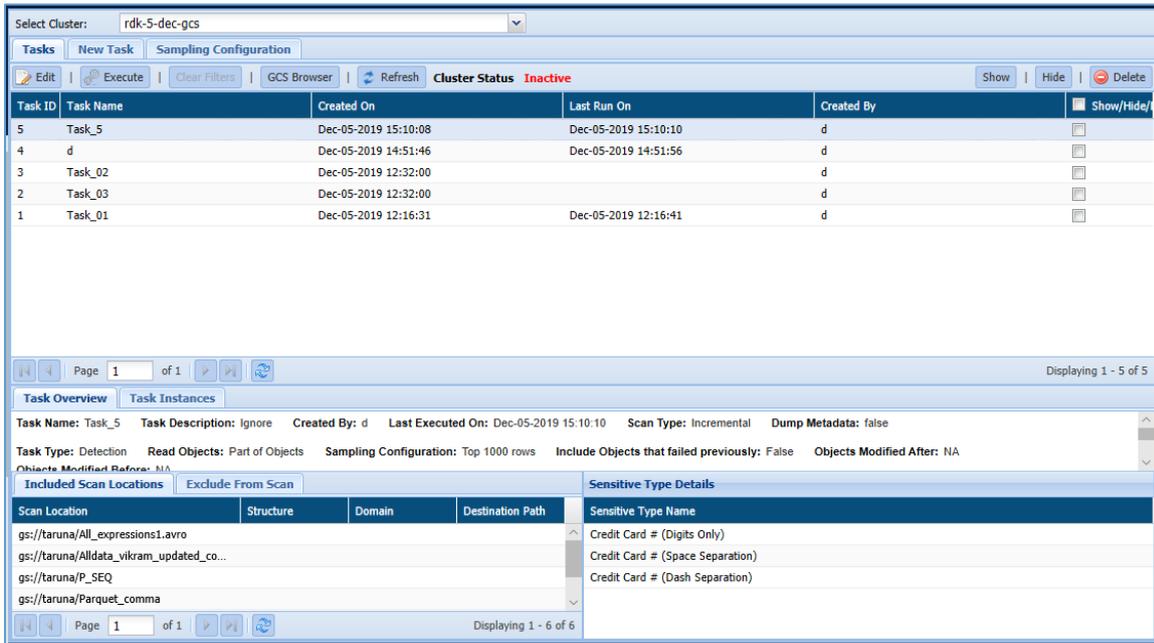
- i. Initializing
- ii. Completed
- iii. Running
- iv. Failed
- v. Stopped

7.3.7 Google Cloud

7.3.7.1 GCS

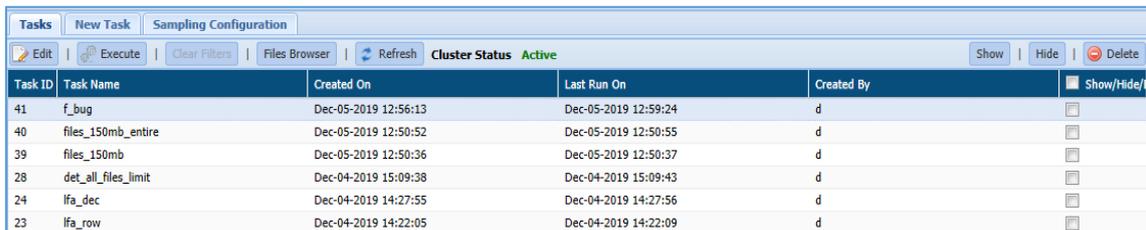
In GCS, screen appears when clicked on **Tasks**. This page offers in-depth details about the tasks that the user has permission to view.

The below image shows the user interface of the screen.



To access the **Task** page. Click **Google Cloud > GCS > Tasks** tab.

1. **Detection Task Panel:** The detection task panel shows the tasks and the properties associated with each task. Properties include details about the task such as Task ID, Task Name, creation date, Created by, etc.



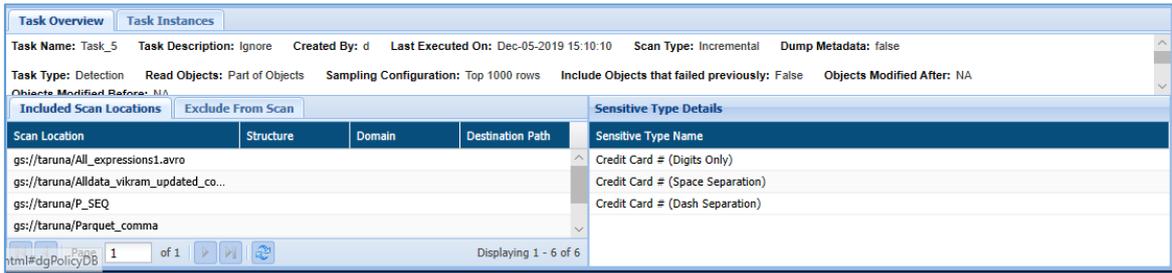
122. **Task Overview:** It displays the details for the selected task. It displays information such as Task Name, Task Description, Created by, Last executed on, Sampling Configuration, Scan Type, etc.

Task Overview tab also includes **Included Scan Location** tab, **Exclude From Scan** tab, and **Sensitive Type Details** pane.

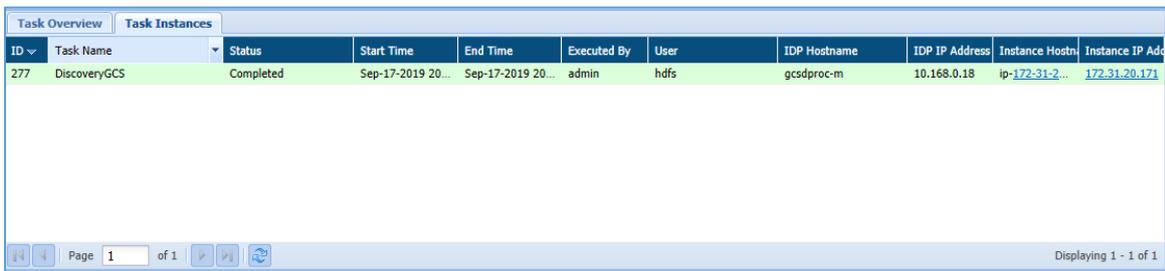
- a) **Included Scan Location:** This pane displays the list of Included Objects for the selected scan locations.

123. **Exclude from Scan:** This pane displays the list of excluded **Object Extension** and **Scan Location** of an object.

124. **Sensitive Type Details:** This tab displays the list of Sensitive Type for the selected task.



125. **Task Instances:** it displays the information about each instance of the task selected in the Tasks panel. Information such as Task ID, Task Name, Status, Start time and End time, Executed By, IDP Hostname, IDP IP Address, etc.



The status of the task can be in any of the listed states:

- i. Completed
- ii. Initializing
- iii. Running
- iv. Failed
- v. Stopped

7.3.8 SharePoint

The below screenshot shows the user interface of the default screen of the **Tasks** tab.

The screenshot shows the Dataguise interface with the 'Tasks' tab selected. At the top, there is a 'Select IDP' dropdown set to 'SP_136'. Below it are buttons for 'Edit', 'Execute', 'Clear Filters', and 'Refresh'. A table lists tasks with columns: Task ID, Task Name, Created On, Last Run On, and Created By. One task is visible: Task ID 27, Task Name Task_01, Created On May-29-2020 12:31:17, Last Run On May-29-2020 18:13:27, Created By d. Below the table is a pagination bar showing 'Page 1 of 1' and 'Displaying 1 - 1 of 1'. The 'Task Overview' tab is active, showing details for Task_01: Task Description: hsAGs, Task Type: Detect and Notify, Created By: d, Last Executed On: May-29-2020 18:13:27. It also shows 'Version to scan: all' and 'File Types: txt, doc, docx, xls, xlsx, ppt, pptx, pdf, csv, zip, lists'. Below this is the 'Included Scan Sites' table with two entries: http://192.168.1.136:1177/sites/sanjay/Root Site and http://192.168.1.136:1177/sites/SPDev/DevSubsite. To the right is the 'Sensitive Type Details' table with columns: Sensitive Type Name, Threshold, Action, and Notify To. It lists various sensitive types like Address Line, Credit Card #, Email Address, IP Address, Social Security #, and Telephone (Digits Only), all with a threshold of 4 and a 'Notify' action.

To access the **Task** tab from the menu. Click **SHAREPOINT > Tasks** tab.

1. Task Panel:

This panel shows the tasks and the properties associated with each task. Properties include details about the task such as Task ID (system generated), Task Name, Task Creation date.

This screenshot shows a portion of the Dataguise interface, specifically the task list table. It includes the 'Edit', 'Execute', 'Clear Filters', and 'Refresh' buttons at the top. The table has columns: Task ID, Task Name, Created On, Last Run On, and Created By. The single entry is: Task ID 27, Task Name Task_01, Created On May-29-2020 12:31:17, Last Run On May-29-2020 18:13:27, Created By d. The pagination bar at the bottom shows 'Page 1 of 1' and 'Displaying 1 - 1 of 1'.

126. Task Overview tab:

This panel displays the basic information for the selected task. It displays the information such as the Task Name, Task Description, Task Type, Sampling Configuration, Connection Name, Database schema, list of Sensitive Type, etc.

It also displays information for the Included Sites for scanning and Sensitive Type Details.

- **Included Scan Sites:** This panel will display the list of all the sites which were selected for scanning
- **Sensitive Type Details:** This panel will display the list of Sensitive Type discovered while scanning.

Included Scan Sites		Sensitive Type Details			
Scan Location	Sensitive Type Name	Threshold	Action	Notify To	
http://192.168.1.136:1177/sites/sanjay/Root Site	Address Line (Best suited f...	4	Notify	Rajni.chaurasia@datagui...	
http://192.168.1.136:1177/sites/SPDev/DevSubsite	Credit Card # (Dash Separ...	4	Notify	Rajni.chaurasia@datagui...	
	Credit Card # (Digits Only)	4	Notify	Rajni.chaurasia@datagui...	
	Credit Card # (Space Sepa...	4	Notify	Rajni.chaurasia@datagui...	
	Email Address	4	Notify	Rajni.chaurasia@datagui...	
	IP Address	4	Notify	Rajni.chaurasia@datagui...	
	Social Security # (Dash Se...	4	Notify	Rajni.chaurasia@datagui...	
	Social Security # (Digits O...	4	Notify	Rajni.chaurasia@datagui...	
	Social Security # (Space S...	4	Notify	Rajni.chaurasia@datagui...	
	Telephone (Digits Only)	4	Notify	Rajni.chaurasia@datagui...	

127. Task Instances

This panel will display the related information about each instance of the task selected in the detection task panel. Information such as Task Name, Task ID (system generated), Status of the task, Start time and End time.

ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Hostn	Instance IP Ad
114	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
113	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
112	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
111	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
110	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
109	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
108	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41
107	Task_01	Completed	May-29-2020 1...	May-29-2020 1...	d	d	192.168.1.136	192.168.1.136	DG-D-W001	10.12.15.41

The status of the task can be in any of the listed states:

- i. Initializing
- ii. Completed
- iii. Running
- iv. Fail
- v. Stopped

7.4 Masking Options

DgSecure provides several masking options for securing sensitive data on an organization's data repositories. The table below shows the masking options available for different data types.

Masking Option Name	Characters (Char)	Variable Characters (Varchar)	Number	Clob
Static mask	X	X	X	X
Character mask	X	X		
Format Preservation Mask (FPM)	X	X	X	
IntelliMask mask	X	X		
Random mask	X	X	X	
NPI mask	X	X	X	
Compose mask	X	X		
Compose Math Expression mask			X	
Date Synch mask		X		
Name Synch mask		X		
Email Policy mask	X	X		
Expression mask			X	
Full Name mask		X		
Regular Expression mask	X	X		
Custom Lookup mask	X	X	X	
Custom mask	X	X	X	
Shuffle mask	X	X	X	
JSON mask		X		X
XML mask		X		X
AES Encryption/Decryption	X	X		
FPE Encryption/Decryption	X	X	X	

7.4.1 Static Mask

Use the static mask to insert the same value into every field in the column. You can enter either a static value (such as AAA or 123) or a null value.

The Static masking option is suitable for columns that meet the following conditions:

- Do NOT have uniqueness constraints.
- Do not require a variety of values for downstream testing or analysis purposes.

7.4.2 Character Mask

Use the Character mask to replace one or more characters at the beginning or end of a string with a meaningless character. The Character mask uses these characters:

- # pound sign
- \$ dollar sign
- @ at-the-rate-of sign
- % percent sign
- . period

The Character mask is suitable for structured fields with an invariable format, like account or credit card numbers, where:

- Only a specific portion of the field is sensitive
- Non-sensitive portions contain useful information you want to retain.
- The sensitive part of the data is at the beginning or end of the field. Valid masks include #23456, ###456, #####6, 12345#, 123###, 1#####, but not 12##56.

If you want to mask a segment in the middle of a structured field, consider Intellimask instead.

7.4.3 Format Preservation Mask (FPM)

Use Format Preservation Mask (FPM) to insert a random value that preserves the format of the original data:

***Note:**

Don't use Format Preservation Mask (FPM) on columns of unique values unless the possible number of potential random values is large. For example, if a column contains only unique 2-digit numbers, the pool of potential values is only 99. With such a small possible number of replacement values, the FPM mask will generate duplicate values.

When you apply FPM to a column with unique values, DgMasking monitors the number of "collisions" (duplicate values) produced. Once 20 duplicate values are created, DgMasking stops attempting to apply the mask and skip the column entirely.

The Format Preservation Mask is suitable when:

- The number of characters or the length of the field is not changed.
- Capital letters are masked with random capital letters.
- Small letters are masked with random small letters.
- Digits are masked with random digits.
- Special characters are left as is.

The FPM mask is suitable for structured data with two exceptions:

- If business rules constrain the value or some part of the value. For example, if business rules require a telephone number to have a valid area code and exchange, then FPM is not a good masking choice. Instead, consider using the Random masking option, which

generates random but valid data for common sensitive types such as telephone, credit card, and Social Security numbers. For non-standard sensitive types, consider using the Regular Expression mask, which allows you to apply any necessary constraints.

- If the data in the masked column will be analysed. Random input would render any analysis meaningless. Instead, consider using the Shuffle mask.

FPM supports masking in a variety of European Languages: French, German, Italian, Spanish, Czech, Slovak, Danish, Finnish, Estonian, Latvian, and Polish.

Additionally, Format Preservation Masking (FPM) should not be used on data requiring unique values unless the possible number of potential random values is large. For example, if a column contains only unique 2-digit numbers, the pool of potential values is only 99. With such a small possible number of replacement values, the FPM mask will generate duplicate values.

In order to use Format Preserving Masking, an FPM key must first be generated. To do this, navigate to http://<hostname_or_IP_address>:8111/HDFSIDP/FPMKeygen.html. Enter the domain to which the key should be assigned, a passphrase and passphrase for key. Click **Generate**.

7.4.4 IntelliMask Mask

Use Intellimask to replace a subset of characters with random values based on a regular expression. The rest of the field is left intact.

The Intellimask is suitable for structured fields with an invariable format that meet the following conditions:

- Only a specific portion of the information is sensitive.
- The non-sensitive portion contains useful information you want to leave intact.
- The sensitive portion has format and/or value constraints that are enforced at the database or application level, making it impossible to mask with a simple Character mask.

7.4.5 Random Mask

Use the random mask to insert a random, fictitious value that retains the properties of the original data. The random mask generates random numeric, string, and date values. It can also generate realistic, fictitious values for these common sensitive types:

- Credit card number, Social security number.
- Street address, City, State, and Country
- Email address, Telephone number
- First name, last name, and full name

Choose a random value that matches the type of data you want to mask.

Option	Parameters	Parameter Description
Address Line 1	Case Options	Choose UPPERCASE, lowercase, or Title Case (capitalizes the first letter of each word)
Credit Card Numbers	Select Providers Manually	Credit card number structure varies by provider. Place a check mark next to the variations you want to include in the randomly-generated data. The choices are: American Express, Discover, Master Card, Visa, and User-Defined. If you choose User-Defined, enter a regular expression that describes the custom format.
	Random Number	Generates credit card numbers that are realistic but not necessarily valid. (In other words, they do not contain functioning check digits.) Use this option only if your application does not validate credit card numbers.
	Auto Generation	Generates credit card numbers that match the format of the values being masked. Visa numbers are masked with Visa numbers; MasterCard numbers are masked with MasterCard numbers, and so on.
	Digit/ Space/ Dash Separation	Indicate how to format the credit card numbers: Using digits only With spaces or dashes separating the components of the number.
Date	Minimum/ Maximum	Define a range of output values by selecting a minimum and maximum date.
	Increment By	Select a range of days, months, and years to use to increment the original date.
	Decrement By	Select a range of days, months, and years to use to decrement the original date.
Timestamp	Minimum/ Maximum date	Define a range of output values by selecting a minimum and maximum date and time.
	Format	Select a date format: ddMMyyyy HH:MM:SS, MMddyyyy HH:MM:SS, or yyyyMMdd HH:MM:SS
Email Address	No parameters	
First Name	Case Options	Choose UPPERCASE, lowercase, or Title Case.
First and Last Name	Case Options	Choose UPPERCASE, lowercase, or Title Case.

Option	Parameters	Parameter Description
Last Name	Case Options	Choose UPPERCASE, lowercase, or Title Case.
Random String	(Various output options)	Choose from five options: Alphabetic characters (letters)—choose mixed case, upper case, or lower case Alphanumeric characters (letters and digits) Printable characters (letters, digits, and symbols)
Social Security Number	Digit/ Space/ Dash Separation	Indicate how to format the Social Security numbers: Using digits only. With spaces or dashes separating the components of the number.
	Valid numbers	Generate numbers that could pass validation checks (i.e., numbers that follow the rules of allocation used by the Social Security Administration).
	Random numbers	Generate random numbers that follow the correct format but would not necessarily pass a validation check.
Telephone Number	Determine Generation Type	Choose the type of separation (if any) to use between the components of the phone number (area code, exchange, and subscriber number): Digits Only Space Separation Dash Separation Standard. Standard format is: (123) 456-7890.
Number	Minimum/ Maximum	Define a range of output values by selecting a Minimum and Maximum number.
	Precision	Indicate the number of decimal places (if any) to include in the number.
	Padding	Pads the value with leading zeros, if necessary, to achieve the specified field length.

7.4.6 NPI Mask

Use the NPI mask to insert a number that follows the formatting and rules of the National Provider Identifier (NPI). DgSecure applies the Luhn algorithm to generate valid 10-digit combinations in sequence from lowest to highest. You further randomize this sequence by providing a starting value (called an NPI "seed") and an increment.

Note: Use this masking option only on columns with a numeric data type of at least 10 digits

7.4.7 Compose Mask

Use the Compose mask to generate a value by combining full or partial values from other fields in the same row.

The Compose option is suitable:

- When the field to be masked is a dependent of information obtained from other source fields in the same row.
- One or more of those source fields will be separately masked.
- You want to make the value in the dependent field reflects the new (masked) values in the source fields.

To make sure the dependent field matches the source field, DgSecure masks the source fields before the composed field.

7.4.8 Compose Math Expression Mask

Use this Compose Math Expression mask to generate a value by performing a calculation on two or more fields in the row. The components of the math expression can include any numeric column in the table and any of the following operators:

- + (for addition)
- - (for subtraction)
- * (for multiplication)
- / (for division)

The Compose Math Expression mask is suitable where:

- When the field to be masked is a numeric and derives its value from other source fields in the same row.
- One or more of those source fields will be separately masked.
- You want to make the value in the dependent field reflects the new (masked) values in the source fields.

To make sure the dependent field matches the source field, DgSecure masks the source fields before the composed field.

7.4.9 Date Synch Mask

Use the Date Synch mask to replace the date in the target field with a date from another field in the same row. You can adjust the format of the date as necessary.

Note: This masking option is not available on SQL Server.

This specialized mask is useful with de-normalized tables in which the same value (in this case a date) appears two or more times in a single row. If you mask one of those instances, the remaining instances should be updated with the new value. To accomplish that, you apply a Date Synch mask to those instances.

To make sure the dates are synchronized, DgSecure masks the primary date field before the dependent field(s).

7.4.10 Name Synch Mask

Use the Name Synch mask to replace the name in the target field with a name from another field in the same row, automatically adjusting the format as necessary.

Note: This masking option is not available on SQL Server.

This specialized mask is suitable for de-normalized tables in which the same value (in this case a name) appears two or more times in a single row. If you mask one of those instances, the remaining instances should be updated with the new value. To accomplish that, you apply a Name Synch mask to those instances.

The Name Synch masking option automatically detects and duplicates the name format used in the target field. However, there are instances where Name Synch cannot determine the format of a name field. You need to supply a default format for these occasions, for example, when the field is null.

To make sure the dates are synchronized, DgSecure masks the primary date field before the dependent field(s).

7.4.11 Email Policy Mask

Use the Email Policy mask to generate an email address by combining values from two other fields in the same row (along with a domain and extension that you specify).

The Email Policy option is best used in situations where business rules require:

- A specific, invariable format for email addresses.
- Synchronization between elements of the email address and information contained in other fields (typically the first and last name fields).

To make sure the email fields are synchronized, DgSecure masks the source field before the composed field(s).

7.4.12 Expression Mask

Use the Expression mask to modify a numeric value in any of three ways:

- Incrementing or decrementing it by a set percentage or number.
- Replacing it with a random value within a specified range.
- Replacing it with a random value from a specified sequence.

The expression mask is suitable for:

- Numeric fields that will not be statistically analysed.
- Preserving statistical information such as a column’s minimum value, maximum value, sum, or average value, consider using the Shuffle option instead.
- Generating random values from a specified sequence is a good way to mask numeric primary keys such as employee or patient IDs. This option, called a "scrambled sequence", does not produce duplicate values, and so it does not violate the primary key constraint.

Used alone, incrementing or decrementing is probably not sufficient protection for sensitive data such as salary, because if even one original value is known, the pattern is easy to decipher. It is best to use this option in conjunction with masking on other fields such as names or other identifiers.

If you use a Scrambled Sequence to mask a primary key, make sure to select the Synchronized CUPS option as well, to ensure that the new primary key values are written to the corresponding foreign keys as well.

7.4.13 Full Name Mask

Use the Full Name mask to replace a full name with a fictitious full name of the same format. This mask is suitable for full name columns that do *not* need to be synched with First Name and Last Name columns.

Note: If you need to keep a full name in synch with first and last names that appear in the same row, use the Compose masking option instead.

Use the Regular Expression mask to insert a value based on a regular expression you define.

This mask is suitable for:

- Non-standard sensitive data for which there are no pre-built masking formulas.
- Data that is constrained to a specific format, specific values, or both.

Verify that the result of the expression matches the data type of the column you are masking.

Component	Description	Expression
DC	A two-letter department code.	[A-Z]{2}
-	Literal hyphen.	/-

Component	Description	Expression
Yy	The last two digits of the fiscal year.	[0-9]{2}
-	Literal hyphen.	/-
T	A one-letter code indicating the contract type. Possible values are: a, b, c, and d. Letter must be lower case.	[a-d]
-	Literal hyphen.	/-
Nnnn	A three or four-digit serial number.	[0-9]{3,4}

Here is an example of a regular expression:

[A-Z]{2}/-[0-9]{2}/-[a-d]/-[0-9]{3,4}

Sample input:	Sample output:
FN-10-c-349	EX-91-a-5234

7.4.14 Regular Expression Mask

Use the Regular Expression mask to insert a value based on a regular expression you define.

This mask is suitable for:

- Non-standard sensitive data for which there are no pre-built masking formulas.
- Data that is constrained to a specific format, specific values, or both.

Verify that the result of the expression matches the data type of the column you are masking.

Component	Description	Expression
DC	A two-letter department code.	[A-Z]{2}
-	Literal hyphen.	/-
Yy	The last two digits of the fiscal year.	[0-9]{2}
-	Literal hyphen.	/-
T	A one-letter code indicating the contract type. Possible values are: a, b, c, and d. Letter must be lower case.	[a-d]
-	Literal hyphen.	/-
Nnnn	A three or four-digit serial number.	[0-9]{3,4}

Here is an example of a regular expression:

[A-Z]{2}/-[0-9]{2}/-[a-d]/-[0-9]{3,4}

Sample input:	Sample output:
FN-10-c-349	EX-91-a-5234

7.4.15 Custom Lookup Mask

This mask retrieves and inserts a substitute value using a lookup table. The lookup table can be in any schema (or database, in the case of an SQL Server) that DgSecure is able to connect to.

This masking option is designed for use with existing masked data sets. Use it when you have rich sets of realistic but fictitious data that you have organized in relational tables for the express purpose of masking data. Tables like this are commonly set up to mask sensitive types such as names, addresses, and telephone numbers.

To configure a Custom Lookup mask, you must:

- Specify the two columns that will form the join: The Base Reference column (in the table of actual data) and the Lookup Reference column (in the table of fictitious data). These columns need not have the same name, but they must share values in common. In this example, the Emp ID column and the EmployeeID column form the join.
- Specify the column that contains the fictitious data that will replace the actual data. This is called the Lookup Data column, and in this example the column is named TelephoneNo.

During the actual masking operation, DgMasking reads each value in the Base Reference column and searches for a matching value in the Lookup Reference column. When it finds a match, it retrieves and inserts the associated lookup data. In the example below, DgMasking has found a match for the value 102 and will now replace the original telephone number (415-456-1234) with the telephone number in the Lookup Data column (617-234-4567).

Base Reference Column	Column to be masked		Lookup Reference Column	Lookup Data Column
Emp ID	Telephone		EmployeeID	TelephoneNo
104	510-834-3643		100	905-123-4567
100	510-567-2347		101	510-567-2378
102	415-456-1234	↔	102	617-234-4567
101	905-234-6789		103	415-456-1237
103	510-345-3456		104	510-456-2345

Note: The lookup table and the table containing the column to be masked must have at least one column in common on which to base a join. Make sure that the data type of the lookup data matches the data type of the column you are masking.

7.4.16 Custom Mask

7.4.16.1 Overview of the Feature

Custom masking option allows users to create their own customized functions on databases for data masking. The dropdown list of Custom function screen displays the custom masking functions within the database for which a connection has been created from DGSecure Connection Manager. To apply custom masking, user will choose a function from the functions list. A parameter list will be populated with text boxes where user can assign a table column, a static value to the parameter or column properties. User can also assign the column properties to the parameter as '\$masked column name\$', '\$column length\$', '\$column data type\$', '\$columns scale\$', '\$column precision \$' by selecting it from the drop down list. These properties are added so that user can have access to column properties in custom function.

A value or table column can be assigned to the field as per the parameter datatype and parameter length. User can apply a filter on the function names to get the desired list of functions.

Policy change –New masking option added – 'custom'. User can apply the custom function masking from the policy screen by choosing the custom option. User will select the desired database connection on policy screen and can select the custom functions to apply masking on detected columns using policy

Use the Custom mask for special situations that require complex data masking. The Custom masking option lets you prepare a set of instructions that perform the desired data transformation

For example, a column in your data named National ID contains different types of values, such as Social Security Numbers (SSN), Individual Taxpayer Identification Numbers (ITIN), and Social Insurance Numbers (SIN). You can create a custom mask (or script) that does the following:

- Checks the format of the current value.
- If the value has the format of a Social Security Number, applies the Random/SSN masking option to generate a fictitious but valid replacement.
- If the value has some other format, applies the Format Preservation masking option (FPM) to generate a random value of the identical format.

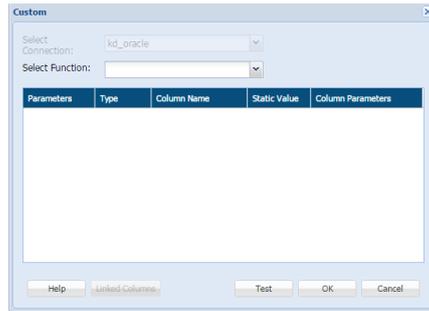
Note: The concept of custom masking has been changed. Custom masking function is created at the database. The 'Custom Functions' tab under masking head has been removed from the DgSecure UI screen.

7.4.16.2 Externals of the Feature:

a. User Interface:

On UI, custom masking option will appear on 2 screens - On Policy and on Masking screens:

- Screen appears as shown below:



b. **DGCL Impact:**

DGCL will be responsible for following things:

- Fetching the List of Functions
- Getting the function parameters
- Assigning static values or table columns to the parameters provided, the length and datatypes match.
- Assigning column properties to the parameters.

7.4.16.3 Detailed Design:

Flow of Custom Masking:

1. Masking through Policy

- Create a Policy and select 'custom' option to mask the detected column.
- Drop down in custom function screen will load all functions from back end database with which connection is created in connection manager. User can type the name of the function to load or search the function.
- After selecting the function, the list of parameters will be displayed.
- Assign static value or column properties against the parameter
- Apply detection through the created policy.
- Create masking task by selecting the above created policy.
- Execute the task to mask the data.

Note: The option to test custom function does not appear on the policy screen. This option is only available for custom masking from the masking screen.

2. Custom Masking

- Create a masking task.
- Select masking type as - 'custom'
- Drop down in custom function screen will load all the functions from the back end database with which connection is created in connection manager. User can search a function by its name.

- After selecting function the parameters list is displayed.
- Assign column name, static value or column properties against parameter. User can enter the value in single Column only.
- User can test the function. Results will be displayed on UI.

Impacts, if any, of the design on:

- i. High Availability in the DgSecure system: No
- ii. Web Services with Other Components: Earlier there was only one web service which was able to create and test custom function. Now there are 2 web services:

Yes – A new web service has been added to the masker IDP which can be called from DgController. This web service will return a list of all functions with which connection has been created from the Connection manager in DGSecure UI. Prefix of the function can be used to filter functions. Test Custom Function web service is required to test the function call by database. This web service will return the function result.

- iii. Encryption in the Controller Repository, Results Database: No
- iv. Audit Reports in DgSecure: No.
- v. DgSecure RBAC – Changes needed for RBAC with this feature: No.
- vi. Controller Snapshots – Feature requires an addition to the fields saved in the Task Instance snapshot in the Controller: YES. (following tables are required to increase the size of the column 'param2'
 1. dg_masking_temp_columns
 2. dg_masking_columns
 3. dg_policy_details
 4. dg_policy_details_hist
 5. dg_masking_column_hist
 6. dg_mask_cols_snapshot
- vii. DDL Changes – Captures DDL changes: Yes.
Alter the table 'dg_columns' –
Set the size of columns param2 and param3 to 2000. Save the json of parameters as param2 and save the function call as param3 column.
- viii. Logging – especially overly verbose logging and leakage of sensitive data in logs:
No.

7.4.16.4 Performance and Scalability Considerations

No.

7.4.16.5 Security Considerations

These changes have been made to improve security. Earlier any user logged on to DGSecure could create any type of custom masking, which could also be used for creating malicious queries or content. With the new security features only users that have rights to databases can create custom functions.

7.4.16.6 Other impacts

If user creates multiple connections with multiple databases, user will have to create a custom masking function on all those databases for which connections have been created.

DgSecure requires a specific script format for each supported database. For more information, contact Dataguisse Technical Support.

7.4.17 Shuffle Mask

Use the Shuffle mask to change the order of values within a table column. In other words, the values in the column are "shuffled" so that each value ends up in a randomly-assigned row that is different than the one it started out in.

If there are fields that need to remain associated, you can shuffle them together as a unit so that they maintain their relationship. You can shuffle up to seven fields in tandem.

This mask is suitable for:

- Preserving statistical information, such as a column's minimum, maximum, sum, or average value.
- Ensuring that the masked data will not violate business rules at the database or application level. (Since the data is repositioned rather than replaced, the values are guaranteed to meet all

7.4.18 JSON Mask

Use the Json names to present the CLOB (Character Large Object) field in Oracle table. It will encrypt or mask the fields within the CLOB.

- CUPS are disabled for Json masking.
- JSON masking is supported for NVARCHAR_MAX, VARCHAR, VARCHAR_MAX and TEXT data types.

Note: For JSON/XML masking support minimum required version is 1.6

7.4.19 XML Mask

Use XML tags to present the CLOB field in Oracle table.

- CUPS are disabled for XML masking.

- XML, XML_TYPE, VARCHAR, VARCHAR_MAX, NVARCHAR_MAX, TEXT and CLOB are the only data types for which masking is available.

Note: For JSON/XML masking support minimum required version is 1.6

7.4.20 AES Encryption/Decryption

Use AES (Advanced Encryption Standard) encryption/decryption to protect classified information. AES uses symmetric block cipher to encrypt the sensitive data.

This masking is suitable for:

- AES encryption/decryption supports CHAR and VARCHAR datatypes.
- Maximum length of the column should be less than or equal to length of encrypted string.

Formulae for calculating length of AES encrypted string using from plain text.

$$\text{Enc_text_length} = ((4 * ((\text{length}(\text{plain_text}) / 16 + 1) * 16) / 3) + 3) \& '3$$

Plain text Min Length	Plain text Max Length	Encrypted String Length
1	15	24
16	31	44
32	47	64

For Example:

19 bytes is encrypted using AES results in 32 binary bytes. The 32 bytes represented as a printable string in Base64 is 44 characters. In summary, encrypting 19 bytes with AES will always result in 44 character Base64 bytes.

7.4.21 FPE encryption/Decryption

Use the FPE encryption/decryption to preserve the format of the data. It encrypts the data in such a way that output is in the same format as the input (the plaintext).

This masking if suitable for:

- FPE encryption/decryption supports CHAR, VARCHAR and NUMBER data types.

7.4.22 Partial Field FPM

This option masks a specified portion of the sensitive data. It only applies to numeric characters. It is only applicable to HDFS tasks. This is a good protection option for numbers that need to be protected but still retain their format. This masking option requires a FPM Key. This key can be generated through the `.../HDFSIDP/FPMKeygen.html` page.

7.4.23 Custom Masking

Use the custom masking option to create a user-defined protection solution. It is only applicable to HDFS tasks.

To leverage this solution,

- 1) place the CustomProtectionDefinition.json file at
`{Installation_path}/webapps/dgcontroller/`

***Note:** This JSON file is configurable. It contains the mapping of custom masking plugins with the sensitive types, controller will pass this mapping to GUI and then GUI will populate the custom masking options for the correspondingly mapped sensitive types on tomcat restart

Example JSON:

```
[
  {
    "sensitiveType": "Credit Card (Digits Only)",
    "customProtectionDetails": [{
      "protectionName": "Retain Credit Card Type",
      "paramsList": ["numberOfCharsToRetain", "delimiter"]
    }],
    {
      "protectionName": "Address Line 1",
      "paramsList": ["Country", "Case"]
    }
  ]
```

- 2) And two jar files (CustomProtection.jar and Retain Card Type.jar) at the path specified by the `"local.custom.masking.jar.path"` property in `HDFSIDPConfig.properties`. This property is configurable. By default, the path is `{Installation Path}/HDFSIDP/expandedArchive/WEB-INF/plugins/custom_protection/`
- 3) Restart the HDFSIDP.
- 4) The custom masking solution is ready to use. Apply the masking solution on the **Hadoop Policy** page or **New/Edit HDFS Task** page.

7.4.24 Custom Transformation

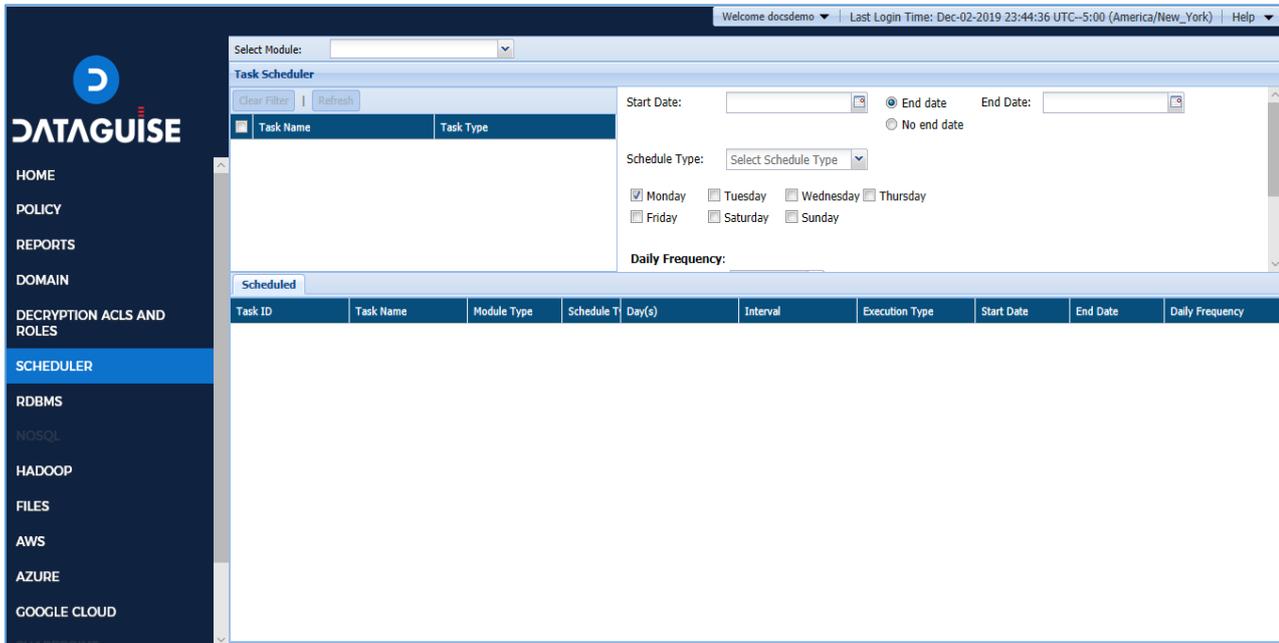
This masking option calls a web service to allow external masking/encryption. It is only available for credit card numbers. Also, this masking option is only available when the property in the `projparams.properties` file is turned on. For more information on how to configure this property, please refer to the *Installation and Configuration Guide*.

8 Scheduler

Scheduler provides an ability to schedule a task at a pre-defined time or at a specified time interval.

To access the scheduler page. Click **Scheduler**.

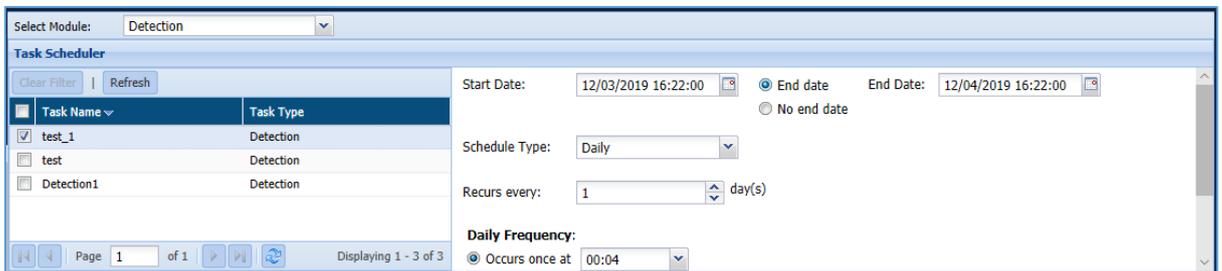
The below screenshot shows the user interface of a Scheduler.



8.1 Schedule a Task

This section will explain the process of scheduling a task.

Following are the steps for scheduling a task:



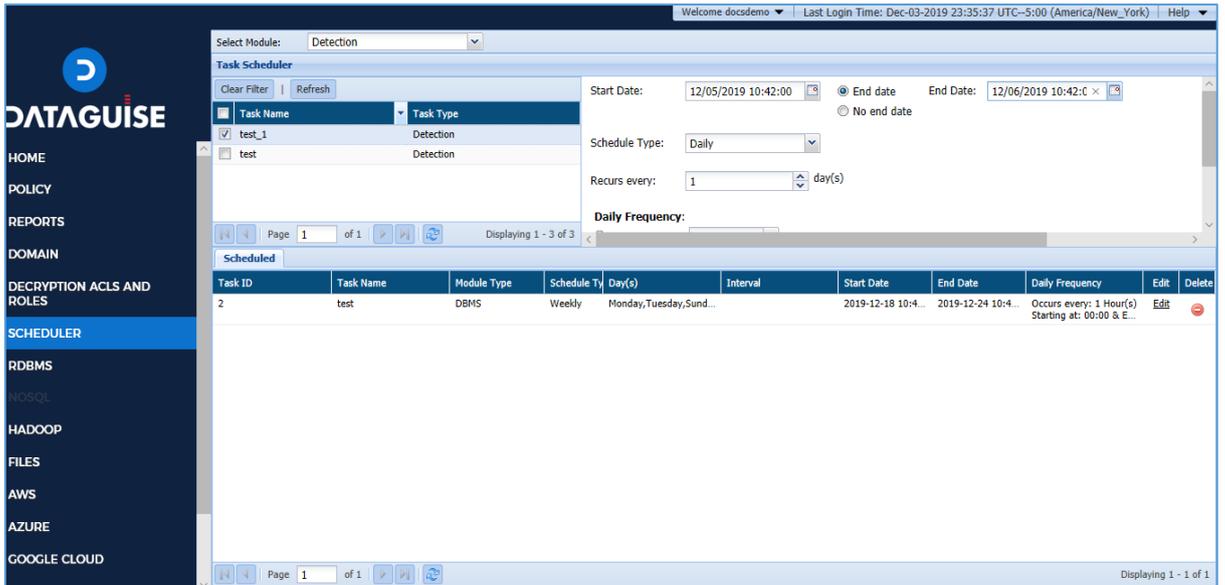
1. Click the **Select Module** drop-down to select the module.
2. Check the required task in the **Task Scheduler** panel.
3. Select the task schedule **Start Date**.

4. Select the task schedule **End date** or **No end date**, if task scheduled is forever.
5. Select **Schedule Type** from the drop-down.
 - i. **Daily:** Select this option to schedule the task daily.
 - Select the frequency of the task in the **Recurs**.
 - Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - Select **Occurs every** and mention the hours, **Starting at** and **Ending at**, if you want to perform the task in every few hours.
 - ii. **Hourly:** Select this option to schedule the task in every few hours.
 - Select the **Interval Time**.
 - iii. **Monthly:** Select this option to schedule the task on the selected days of the month.
 - Select **Day** of one or more **Months** to perform the task.
 - Select a specific **Weekday** of one or more **Months** to perform the task.
 - Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - Select **Occurs every** and mention the Hours, **Starting at** and **Ending at**, if you want to perform the task in every few hours.
 - iv. **Once:** Select this option to schedule the task only once.
 - v. **Weekly:** Select this option to schedule the task on the selected days of the week.
 - Check the **Days** of the week to schedule the task.
 - Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - Select **Occurs every** and mention the hours, **Starting at** and **Ending at**, if you want to perform the task in every few hours.
6. Click the **Schedule** button to schedule tasks.
7. Click the **Reset** button to re-scheduled tasks, if required.

8.2 List a Scheduler

This section will explain the screen of the **Scheduler** section.

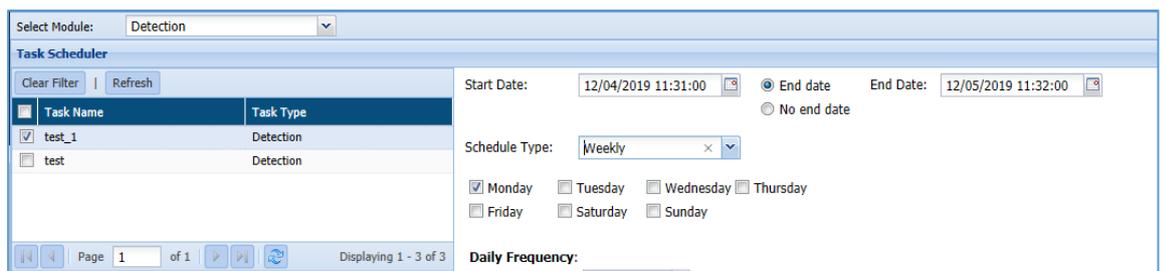
The below screenshot shows the user interface of the **Scheduler**.



- **Task Scheduler Panel:** This panel lists down all the tasks for the selected module. It displays information about the task such as Task Name, Task Type. A user can also schedule a task by providing inputs such as the Start Date, End Date, Schedule Type, etc.

You can also **Refresh** and **Clear Filters** from the Task Scheduler Panel.

- Clear Filters:** Click the **Clear Filters** button to remove any applied filters from the column.
- Refresh:** Click the **Refresh** button to update the current screen.



- **Scheduled:** It displays the list of all scheduled tasks and the details related to it. The screen displays details for the scheduled task such as **Task ID** (system generated), **Task Name**, **Module Type**, **Schedule Type**, **Day(s)**, **Interval**, **Start Date**, **End Date**, **Daily Frequency** etc.

You can also **Edit** and **Delete** a scheduled task from the Scheduled panel.

- Edit:** A user can also edit a scheduled task by clicking the Edit Button. This will allow user to edit or update the information.

ii. **Delete:** Click the **Delete** button to delete the scheduled task.

Scheduled										
Task ID	Task Name	Module Type	Schedule Type	Day(s)	Interval	Start Date	End Date	Daily Frequency	Edit	Delete
2	test	DBMS	Weekly	Monday,Tuesday,Sund...		2019-12-18 10:4...	2019-12-24 10:4...	Occurs every: 1 Hour(s) Starting at: 00:00 & E...	Edit	

9 Results

9.1 RDBMS

9.1.1 Detection

9.1.1.1 Results

The **Result** screen displays information about the Sensitive data detected in the existing database. This screen is divided into three panels.

1. Task Instances
2. Task Instances Detail
3. Overview

The screenshot displays the Dataguise interface. The top section, 'Task Instances', contains a table with the following data:

ID	Task Name	Status	Start Time	End Time	Remediated	Re-Execute	Show/Hide
44	oracle	Completed	Jan-09-2020 04:23:28	Jan-09-2020 04:34:52	No	▶▶	☐
43	Task2	Completed	Jan-08-2020 07:00:32	Jan-08-2020 07:00:53	No	▶▶	☐
42	Task1	Failed	Jan-08-2020 06:58:44	Jan-08-2020 06:58:45	No	▶▶	☐

The middle section, 'Task Instance Details', shows information for Task Instance ID: 44, Task Name: oracle, Task Type: Detection, Start Time: Jan-09-2020 04:23:28, and Sampling Configuration: Top 1000 rows. A 'Sensitive Type' dropdown is set to 'HIPAA_DBMS', and a 'Database Object Filter' table is empty.

The bottom section, 'Sensitive Data Group', displays a table of sensitive data:

Sensitive Data Group	Database	Directory Path	Table	Column	Key Path	Sensitive Group Config	Referential Type	Referring To
Credit Card	DATAGUISE	NA	MUNISH	CCNO	NA	100		
Credit Card	DATAGUISE	NA	SUPPORTTEST	CREDITCARD	NA	30		
Email Address	DGCONTROLLER_...	NA	DG_DATABASE_ATTRIB...	CONTACT_EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_HADOOP_ATTRIBUT...	CONTACT_EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATIONS_US...	EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATIONS_US...	EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATION_SEN...	EMAILCONTENT	NA	16		

- **Task Instances**
The Task Instance pane display information for the masking tasks. This pane will display information such as ID (system generated), Task Name, Status, Start Time, End Time, etc.

This screenshot shows the 'Task Instances' panel with the same data as the previous screenshot:

ID	Task Name	Status	Start Time	End Time	Remediated	Re-Execute	Show/Hide
44	oracle	Completed	Jan-09-2020 04:23:28	Jan-09-2020 04:34:52	No	▶▶	☐
43	Task2	Completed	Jan-08-2020 07:00:32	Jan-08-2020 07:00:53	No	▶▶	☐
42	Task1	Failed	Jan-08-2020 06:58:44	Jan-08-2020 06:58:45	No	▶▶	☐

1. **Refresh:** Click the **Refresh** button. It will update the current page with the updated information

2. **Clear Filters:** Click the **Clear Filters** button. It will remove any applied filters on the Tasks page.
3. **Resume:** Click the **Resume** button to re-start the task where it was initially stopped.
4. **Show:** Click the **Show** button to unhide a task.
5. **Hide:** To hide any task, follow the below steps:
 - i. Check **Show/Hide** checkbox for the policy.
 - ii. Click the **Hide** button. The policy will get greyed out.
6. **With Results:** Check the **With Results** checkbox, if you want to see the results for the selected masking task.
7. **Queue For Remediation:** Click this button to queue the task for remediation workflow. This option allows you to queue the task for remediation manually.

Once the task has been queued for remediation, the value for the remediation can be seen in the **Task Instances** panel under the column name **REMEDIATED**. It specifies the value as **YES** or **NO**.

ID	Task Name	Status	Start Time	End Time	Remediated	Re-Execute	Show/Hide
183	email_privacy	Completed	May-19-2020 17:35:08	May-19-2020 17:35:29	Yes		<input type="checkbox"/>
153	Ritish_DGWalker_Oracle	Completed	May-18-2020 10:18:17	May-18-2020 10:18:21	No		<input type="checkbox"/>
152	Ritish_DGWalker_Oracle_copy_by_parteek...	Completed	May-18-2020 10:17:13	May-18-2020 10:17:27	No		<input type="checkbox"/>

To queue the task for remediation workflow automatically, set the value of **AUTO QUEUE FOR REMEDIATION** to **YES** in DgAdmin application under **SETTINGS**.

8. **Remediated Results:** check the **Remediated Results** checkbox, if you want to view the final results for the selected fields which are marked as **Remediated**. The remediated result will be displayed in the **Detailed Result** tab under the **Overview** panel.

- **Task Instance Details**

The Task Instance Details pane will display the basic information for the task selected in the Task Instance pane. The includes details such as Task Name, Task Description, Start Time, End Time, Task Type, etc.

It also provides additional information for the Sensitive Type and the Database Object Filter.

Task Instance Details		Sensitive Type	Database Object Filter					
Task Name:	oracle	<ul style="list-style-type: none"> HIPAA_DBMS Email Address Full Names IP Address NPI 	Operat	Connection Info	Table/View O	Table/View	Column Op	Column
Task Instance ID:	44		No rows to display.					
Task Type:	Detection							
Start Time:	Jan-09-2020 04:23:28							
Sampling Configuration:	Top 1000 rows							

- **Overview**

The bottom panel will show the detailed information for the selected task. The information displayed is dependent on the currently selected tab.

Sensitive Data Group	Database	Directory Path	Table	Column	Key Path	Sensitive Group Conf	Referential Type	Referring To
Credit Card	DATAGUISE	NA	MUNISH	CCNO	NA	100		
Credit Card	DATAGUISE	NA	SUPPORTTEST	CREDITCARD	NA	30		
Email Address	DGCONTROLLER_...	NA	DG_DATABASE_ATTRIB...	CONTACT_EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_HADOOP_ATTRIBUT...	CONTACT_EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATIONS_US...	EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATIONS_US...	EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATION_SEN...	EMAILCONTENT	NA	16		

- Results:** The Result tab provide information for the Sensitive Type encountered in the database along with the additional information such as Column Name, Table Name, Sensitive Group Confidence.
- Detailed Results:** This pane list the Sensitive table columns that are discovered, along with their location, data type, Table Name, Column Name, Data Type, etc.
- Skipped Columns:** This pane list down any skipped columns along with the data type identified in that column.
- Data Scanned:** The Data Scanned pane list down the sampled and the total data targeted with the scan.
- Non Sensitive Scanned Tables:** This pane list down all the Database Name along with their Host Name, Directory Path, Table Name and Table Size which were found non – sensitive during the scan process.
- Logs:** This pane list any errors which occurred during the task execution. It displays the log information for the errors such as Database Name, Schema Name, Column Name, Error Description, and Error.

***Note:** The **Metadata Results** and **Metadata Relation Results** tab are available in Results screen only when Task Type = ‘Metadata Discovery’ is selected for Oracle and Teradata connections.

- Metadata Results:** This panel will display the list of all the columns names along with the other information such as column size, column type, constraints (primary key, foreign key, surrogate key etc). This tab displays the association of a column name with a table and database name.

The highlighted records in the tab are the results of Metadata Discovery.

Database Name	Table Name	Column Name	Column Size	Column Type	Sensitive Type	Table Size	Constraints
ritish_new_simple_key	dbo.view1	ABA_ROUTING_NUMB...	100	varchar		16.0 K	
ritish_new_simple_key	dbo.MY_DISCOVERY	ABA_ROUTING_NUMB...	100	varchar		16.0 K	
ritish_new_simple_key	dbo.RITISH_DISCOVERY4	ADDRESS	100	varchar	Canada Address (Unstructured data only)	16.0 K	
ritish_new_simple_key	dbo.view1	ADDRESS	100	varchar	Canada Address (Unstructured data only)	16.0 K	
ritish_new_simple_key	dbo.MY_DISCOVERY	ADDRESS	100	varchar	Canada Address (Unstructured data only)	16.0 K	
ritish_new_simple_key	dbo.RITISH_TABLE	ADDRESS	100	varchar	Canada Address (Unstructured data only)	0.0	
ritish_new_simple_key	dbo.RITISH_DISCOVERY10	ADDRESS	100	varchar	Canada Address (Unstructured data only)	16.0 K	

8. **Metadata Relation Results:** This panel will display the relation between the parent and child tables for any database. It will display other information based on which a relation can be inferred between the child and parent tables. The information includes column names as Child Column, Parent Column and Relation.

Child Column	Child Table	Parent Column	Database Name	Parent Table	Relation
ID	dbo.RITISH_DISCOVERY4	ID	ritish_new_simple_key	dbo.RITISH_DISCOVERY9	ID_DISC4_FK
ID	dbo.RITISH_DISCOVERY9	ID	ritish_new_simple_key	dbo.RITISH_DISCOVERY8	ID_DISC9_FK
ID	dbo.RITISH_DISCOVERY8	ID	ritish_new_simple_key	dbo.RITISH_DISCOVERY7	ID_DISC8_FK
ID	dbo.RITISH_DISCOVERY7	ID	ritish_new_simple_key	dbo.RITISH_DISCOVERY6	ID_DISC7_FK
ID	dbo.RITISH_DISCOVERY6	ID	ritish_new_simple_key	dbo.RITISH_DISCOVERY5	ID_DISC6_FK
ID	dbo.RITISH_DISCOVERY5	ID	ritish_new_simple_key	dbo.RITISH_DISCOVERY4	ID_DISC5_FK
NPI	dbo.RITISH_DISCOVERY2	NPI	ritish_new_simple_key	dbo.RITISH_DISCOVERY3	ID_DISC2_FK

9. **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
10. **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
11. **Remediate:** Click this button to remediate any result in **DETAILED RESULTS** tab. To remediate the selected result, perform the following steps:
 - a) Select the record in the **Detailed Result** tab and click the **Remediate** button.

Sensitive Typ	Hostnam	Databas	Director	Table	Column	Data T	Data Len	Nullabl	Key Pri	Databas	Confid	Row Sc	Values	HR Col	Null Co	Quick	Masked	Safe	Table	Database	Referent	Referring	Remed S
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	82	10	10	8	0	Y	N	N	TA...	SQL S...			Unre...
Email Add	192...	grv...	NA	dbo.GA...	EMAI...	var...	100	Y	NA	sonam	78	1,000	1,000	752	50	Y	N	N	TA...	SQL S...			Unre...
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	77	2,000	2,000	1,484	101	Y	N	N	TA...	SQL S...			Unre...
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	76	2,000	2,000	1,477	102	Y	N	N	TA...	SQL S...			Unre...
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	76	2,000	2,000	1,468	102	Y	N	N	TA...	SQL S...			Unre...
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	77	2,000	2,000	1,485	103	Y	N	N	TA...	SQL S...			Unre...

- b) Enter the details for **Remediation Action** and **Remediation Value (%)** field in the popup.

Remediation

Scope

Host Name: 192.168.0.151 DB Name: grv_discover

Remediation Action: Select remediation action Remediation Value(%): 1

Table Name: dbo.gaurav_10lakh_1 Field Name: EMAIL_ADDRESS

Sensitive Type: Email Address Enabled

Cancel Save

- i. Following are the options in **Remediation Action** field.

- **Mark as Non Sensitive:** Select this option to mark the record as Non Sensitive. The system will ignore this record in future re-run.
- **Mark as Correct:** Select this option to mark the record as Sensitive. The system will not ignore this record in future re-run.
- **Decrease the Confidence Factor:** select this option to decrease the confidence factor in Remediation Value (%) field.
- **Increase the Confidence Factor:** Select this option increase the confidence factor in Remediation Value (%) field.

ii. Enter the remediation value in percentage. This field will be enabled when you select **Decrease** or **Increase the Confidence Factor** in **Remediation Action** field.

Once the remediation value has been set and saved, the same can be seen under Saved Remediation tab.

HostName	Scope Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Act	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

9.1.1.2 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

HostName	Scope Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Act	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

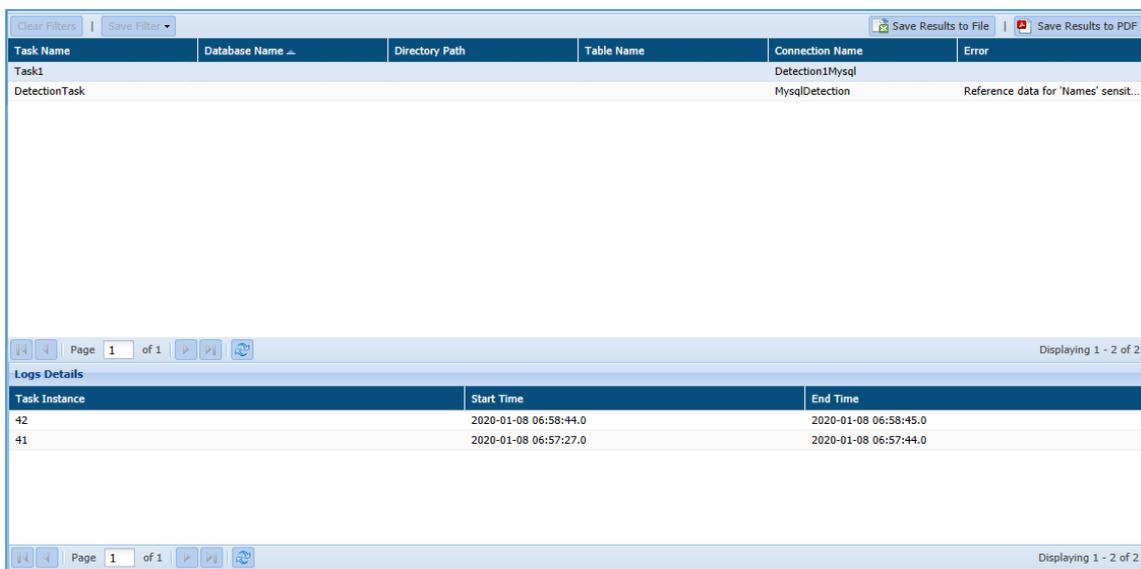
To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.1.1.3 Logs

In Logs tab, you can view the list of all the Task Name along with the errors occurred during the task execution. The Logs tab is divided into two panes:

1. Logs List
2. Logs Details



- **Logs List:** This pane will display the list of all the Tasks Name along with the errors occurred during the task execution. It will also display the information such as Database Name, Directory Path, Table Name and Connection Name.
 1. **Save Results to File:** Click the Save Results to File button to save the Logs List information in .csv format.
 2. **Save Results to PDF:** Click the Save Results to PDF button to save the Logs List information in PDF format.

Task Name	Database Name	Directory Path	Table Name	Connection Name	Error
Task1				Detection1Mysql	
DetectionTask				MysqlDetection	Reference data for 'Names' sensit...

- **Logs Details:** The Logs Details pane will display the information for the selected task in the Logs List. The information includes Task Instance, Start Time and End Time.

Task Instance	Start Time	End Time
42	2020-01-08 06:58:44.0	2020-01-08 06:58:45.0
41	2020-01-08 06:57:27.0	2020-01-08 06:57:44.0

9.1.2 Masking

The Result page displays information about the data masked with each masked task instance. The Result page is divided into three panes:

- i. Task Instances
- ii. Task Instance Details
- iii. Overview

ID	Task Name	Status	Start Time	End Time	Re-Execute	Re-Try	Show
46	Mask1	Completed	Jan-09-2020 05:04:33	Jan-09-2020 13:23:06	▶▶		<input type="checkbox"/>
45	Mask1	Error	Jan-09-2020 05:00:04	Jan-09-2020 05:01:45	▶▶	▶▶	<input type="checkbox"/>
40	TeradataMasking	Completed	Dec-19-2019 05:46:57	Dec-19-2019 05:47:20	▶▶		<input type="checkbox"/>
39	TeradataMasking	Completed	Dec-19-2019 05:42:26	Dec-19-2019 05:45:50	▶▶		<input type="checkbox"/>

Task Instance Details
 Task Name: Mask1 | Start Time: Jan-09-2020 05:04:33 | End Time: Jan-09-2020 13:23:06 | Incremental task: No | Rows to Mask: 1000000 | Rows Masked: 1000000
 Rows with Masking Errors: 0

Connection Name	Source DB	Destination DB	Table	Column	CUPS Options	Masking Action	# of Rows	Errors
oracleasing	DATAGUISE	DATAGUISE	MUNISH	ADDRESS	C	Random(Full Address)	1,000,000	0
oracleasing	DATAGUISE	DATAGUISE	MUNISH	CCNO	U	Random(Credit Card ...)	1,000,000	0

- **Task Instances:**

The Task Instance pane displays information for the masking tasks. This pane will display information such as ID (system generated), Task Name, Status, Start Time, End Time, etc.

ID	Task Name	Status	Start Time	End Time	Re-Execute	Re-Try	Show/Hide
46	Mask1	Completed	Jan-09-2020 05:04:33	Jan-09-2020 13:23:06	▶▶		<input type="checkbox"/>
45	Mask1	Error	Jan-09-2020 05:00:04	Jan-09-2020 05:01:45	▶▶	▶▶	<input type="checkbox"/>
40	TeradataMasking	Completed	Dec-19-2019 05:46:57	Dec-19-2019 05:47:20	▶▶		<input type="checkbox"/>
39	TeradataMasking	Completed	Dec-19-2019 05:42:26	Dec-19-2019 05:45:50	▶▶		<input type="checkbox"/>

1. **Refresh:** Click the Refresh button. It will update the current page with the updated information
2. **Clear Filters:** Clear Filters: Click the Clear Filters button. It will remove any applied filters on the Tasks page.
3. **Show:** Click the Show button to unhide a task.
4. **Hide:** To hide any task, follow the below steps:
 - i. Check Show/Hide checkbox for the policy.
 - ii. Click the Hide button. The policy will get greyed out.
5. **With Results:** Check the With Results checkbox, if you want to see the results for the selected masking task.

- **Task Instance Details:**

The task Instance Details pane will display the basic information for the masking task selected in the Task Instance pane. The information includes details related to tasks such as Task Name, Start Time, End Time, Incremental Task, etc.

Task Name: Mask1	Start Time: Jan-09-2020 05:04:33	End Time: Jan-09-2020 13:23:06	Incremental task: No	Rows to Mask: 1000000	Rows Masked: 1000000
Rows with Masking Errors: 0					

- **Overview:**

The bottom pane shows the detailed information for the selected task. The information displayed is dependent on the current selected tab.

Connection Name	Source DB	Destination DB	Table	Column	CLUPS Options	Masking Action	# of Rows	Errors
oracleasking	DATAGUISE	DATAGUISE	MUNISH	ADDRESS	C	Random(Full Address)	1,000,000	0
oracleasking	DATAGUISE	DATAGUISE	MUNISH	CCNO	U	Random(Credit Card ...	1,000,000	0

1. **Detailed Results:** The Detailed Results panel lists each masked column along with details such as the Connection Name, Source DB, Destination DB, Table, Column, CUPS Option, etc.
2. **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
3. **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
4. **Logs:** This pane list any errors which occurred during the task execution. It display the log information for the errors such as Database Name, Schema Name, Column Name, Error Description, Error.

Database Name	Schema Name	Table Name	Column Name	Error Description	Error
No rows to display					

9.2 NoSQL

The Results tab displays the information about the sensitive information found in the database. The tab is divided into three panes. These are:

The screenshot displays the Dataguise interface with three main panes:

- Task Instances:** A table listing task instances with columns for ID, Task Name, Status, Start Time, End Time, Re-Execute, and Show/Hide. It shows three instances, one completed and two failed.
- Task Instance Details:** A pane for a specific task instance (ID: 150) showing details like Start Time, Sampling Configuration, and Connection Name. It also includes a 'Sensitive Type' dropdown menu with options like PCI_DBMS and Credit Card #.
- Results:** A table showing sensitive data groups with columns for Sensitive Data Group, Database, Collection, and Field Name. It lists various fields like Description, Address, City, etc., from the Northwind database.

1. Task Instances
 2. Task Instances Detail
 3. Overview
- **Task Instances:** This window will display the information for the task such as ID (system generated), Task Name, Status, Start Time, and End Time.

ID	Task Name	Status	Start Time	End Time	Re-Execute	Show/Hide
407	discovery	Failed	Dec-05-2019 06:55:46	Dec-05-2019 06:57:57	▶▶	<input type="checkbox"/>
158	discovery	Completed	Aug-05-2019 18:55:54	Aug-05-2019 18:56:28	▶▶	<input type="checkbox"/>
157	discovery	Failed	Aug-05-2019 18:45:34	Aug-05-2019 00:00:00	▶▶	<input type="checkbox"/>

Click the **Re-Execute** button to re-run the task.

- **Task Instance Details:**

The Task Instance Details pane will display the task properties such as Task Name, Task Instance, Task Type, Start Time, list of Sensitive Type, etc.

1. **Sensitive Type:** This pane displays the list of Sensitive Type.

- **Overview:**

The bottom panel will show the detailed information for the selected task. The information displayed is dependent on the currently selected tab.

Sensitive Data Group	Database	Collection	Field Name
Names	Northwind	customers	Description
Names	Northwind	customers	Address
Names	Northwind	customers	City
Names	Northwind	customers	CompanyName
Names	Northwind	customers	ContactName
Names	Northwind	customers	ContactTitle
Names	Northwind	customers	Region

1. **Results:** The result tab will display the high level information of the Sensitive Types for the selected task in the detection task pane. The information includes Sensitive Data Group, Database, Collection and Field Name.

2. **Detailed Results:** The Detailed Results tab will provide you the detailed information for the Sensitive Type encountered in the database. The information will include Sensitive Type, Hostname, Database, Collection, Field Name, Confidence, etc.

Sensitive Type	Hostname	Database	Collection	Field Name	Confidence	Hit Count	Quick Search
Full Names	13.56.82.12	Northwind	products	ProductName	4	7	N
Full Names	13.56.82.12	Northwind	categories	Description	6	1	N
Full Names	13.56.82.12	Northwind	suppliers	ContactName	70	16	Y
Full Names	13.56.82.12	Northwind	suppliers	Address	5	3	N
Full Names	13.56.82.12	Northwind	suppliers	HomePage	3	2	N
Full Names	13.56.82.12	Northwind	suppliers	ContactTitle	5	3	N
Full Names	13.56.82.12	Northwind	suppliers	City	6	4	N

- 3. **Data Scanned:** List the total number of rows in the scanned tables and how many of those rows DgSecure actually scanned.

Database Name	Collection	Total Rows	Total Size (Bytes)	Sampled Data (Bytes)	Sampled Documents
Northwind	categories	8	3209	3208	8
Northwind	customers	91	27410	27391	91
Northwind	employee-territories	49	2695	2695	49
Northwind	northwind	3000	558626	186000	1000
Northwind	order-details	2155	212387	98000	1000
Northwind	orders	1660	641346	386000	1000
Northwind	products	77	17814	17787	77

- **Save Results to File:** Click the **Save Results to File** to download the data in Doc format.

9.3 Hadoop

9.3.1 HDFS

9.3.1.1 Results-By Task

The **By Task** tab under HDFS Results, provides the details of the selected HDFS tasks. The Task Instances panel, at the top of the page, contains a list of all the HDFS tasks that you have permission to view.

1. Select a task instance to display its parameters and results in the panels below.
2. Click Refresh to update the list with any recently-launched task instances. Click Resume to restart canceled tasks.

The screenshot displays the Dataguise interface. At the top, there's a 'Task Instances' table with columns for ID, Task Name, Status, Start Time, End Time, Executed By, User, IDP Hostname, IDP IP Address, Instance Hostname, Instance IP, Remediated, Re-Executed, and Show. Below this is the 'Task Instance Overview' panel, which includes fields for Task Name, Task Description, Created By, Last Executed On, Scan Type, Dump Metadata, Task Type, Read Files, Sampling Configuration, Include Files that failed previously, Files Modified After, and Files Modified Before. To the right of the overview is a table for 'Included Scan Locations' with columns for Scan Location, Structure, and Domain. Below the overview is the 'Results Overview' panel, which contains several charts: a bar chart for 'Structured & UnStructured' files, a bar chart for 'Files Found' (Sensitive vs Non-Sensitive), a pie chart for 'Draw chart: With scale' (Address, Credit Card, Email Address, IP Address, NPI, Social Security, Telephone, URL), and a pie chart for 'Draw chart: Without scale' (CSV).

The middle panel shows the Task Instance Overview, Included and Excluded scanning locations and Sensitive Type Details. While the scan is running, the panel displays a progress bar that shows you the total number of tables to be scanned and the number that have been scanned so far. The various tabs on this tab are discussed below:

- **Task Instance Overview:** Provides identifying information about the selected task instance and describes the work that it performed.

This screenshot shows a detailed view of the 'Task Instance Overview' panel. It contains the following information:

- Task Name:** Test_detection
- Task Description:** Detection
- Created By:** DocsDemo
- Last Executed On:** Dec-18-2019 01:25:35
- Scan Type:** Full
- Dump Metadata:** false
- Task Type:** Detection
- Read Files:** Part of Files
- Sampling Configuration:** Top 1000 rows
- Include Files that failed previously:** False
- Files Modified After:** NA
- Files Modified Before:** NA
- Sampled Data:** 2.90 KB
- Total Data:** 2.90 KB

- **Included Scan Locations:** The locations added to the Include list. When the Included Scan Location tab is selected, the instance's target directory, the structure assigned to the directory, and the domain to which it is assigned are shown. Not all scan locations are associated with a structure or domain.

This screenshot shows the 'Included Scan Locations' panel. It features a table with the following data:

Scan Location	Structure	Domain
/munish		

 Below the table, there are navigation controls including 'Page 1 of 1' and 'Displaying 1 - 1 of 1'.

- **Excluded From Scan:** The locations added to the Exclude list. When the Exclude from Scan tab is selected, the panel displays any excluded file extensions and file paths.

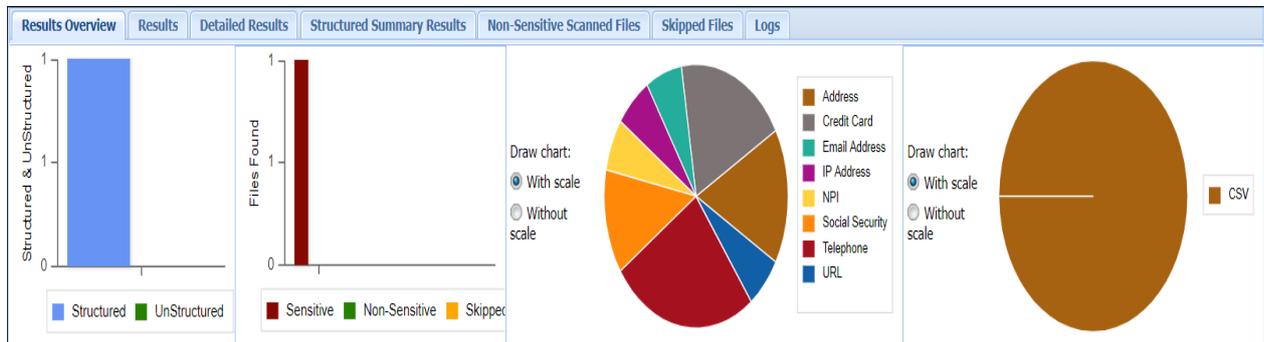
Included Scan Locations		Exclude From Scan	
File Extension	Scan Location	File Extension	Scan Location

- **Sensitive Type Details:** The Sensitive Type Details panel shows the task instance's sensitive types.

Sensitive Type Name
URL
Social Security # (Space Separation)
Social Security # (Dash Separation)
Telephone (Space Separation)
Telephone (Dash Separation)

The tabs shown in the bottom panel are as follows:

- **Results Overview:** For the Results Overview tab, the panel gives a graphical summary of the results for the selected task instance. Hovering over data points in any of the graphs shows the number of hits and hit percentage for a file or sensitive type.



- **Results:** For the Results tab, the panel displays sensitive data by group, hit count, scan setting, task name, whether the scan was incremental, and whether the data was masked/encrypted.

Task Name	Sensitive Data Group	File Path	File Type	File Owner	Detection Type	Content Read	Hit Count	Modification Date
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	26	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	30	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	Social Security	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	20	2019-11-25 02:37:00
Test_detection	Telephone	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	40	2019-11-25 02:37:00

- Detailed Results:** The Detailed Results tab displays the sensitive data by type. When the Detailed Results tab is selected, two buttons appear in the menu bar of the bottom panel. You can, Save results to file or Save Results to PDF.

Task Name	Sensitive Data Type	File Path	File Type	File Owner	Detection Type	Content Read	Hit Count	Modification Date
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	2	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Credit Card # (Dark Connection)	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00

The columns in the Results and Detailed Results tables are described below:

Column	Description
Task Name	The name of the task that was executed to create this instance.
Sensitive Data Type	A specific sensitive type within a sensitive type group. The sensitive type and sensitive type group may be synonymous.
Sensitive Data Group	The type of sensitive information that has been discovered.

File Path	The location of the file with sensitive data. Hit Count. The number of values in in the listed file path that match the search criteria.
File Type	The type of file scanned.
File Owner	The user for task execution.
Detection Type	The type of detection, structured or simulated.
Content Read	The sample type of the data scanned.
Hit Count	The number of sensitive types detected.
AES/Masked	Indicates whether or not the sensitive type was masked. This column displays "False" if the sensitive data was not masked and "True" if the sensitive data was masked.

FP/Row Encrypted	Indicates whether or not the sensitive type was encrypted. This column displays "False" if the sensitive data was not encrypted and "True" if the sensitive data was encrypted.
FP/Row Decrypted	Displays "True" if the scan was incremental and "False" if it was a full scan.
Modification Date	The date on which the task was last modified.

- **Structured Summary Results:** This tab only appears for Detection results. It indicates the number of sensitive data items - categorized by sensitive data type - the task instance discovered and which column contains the sensitive elements.

Sensitive Type	File Path	Field No.	Confidence	Match Count	Rows Scanned	Sample Mode	Null Count	Null Ratio	Field Name Match
Address City (Best sui...	/munish/AllData.csv	Column 11	14	2	12	Sample: 0,r,1000;	1	8	false
Telephone (Standard)	/munish/AllData.csv	Column 11	87	10	12	Sample: 0,r,1000;	1	8	true
Telephone (Standard ...	/munish/AllData.csv	Column 12	87	10	12	Sample: 0,r,1000;	1	8	true
Address City (Best sui...	/munish/AllData.csv	Column 12	7	1	12	Sample: 0,r,1000;	1	8	false
IP Address	/munish/AllData.csv	Column 14	87	10	12	Sample: 0,r,1000;	1	8	true
NPI	/munish/AllData.csv	Column 15	87	10	12	Sample: 0,r,1000;	1	8	true
Address City (Best sui...	/munish/AllData.csv	Column 16	70	10	12	Sample: 0,r,1000;	1	8	false

- **Non Sensitive Scanned Files:** Lists out all the files where no sensitive content was detected.

File Path	File Type	File Size	File Owner	Modification Date
No non-sensitive scanned files.				

- **Skipped Files:** The files that were skipped during scan are listed here.

File Path	Skipped Reason	File Type	File Owner	Modification Date
No file skipped.				

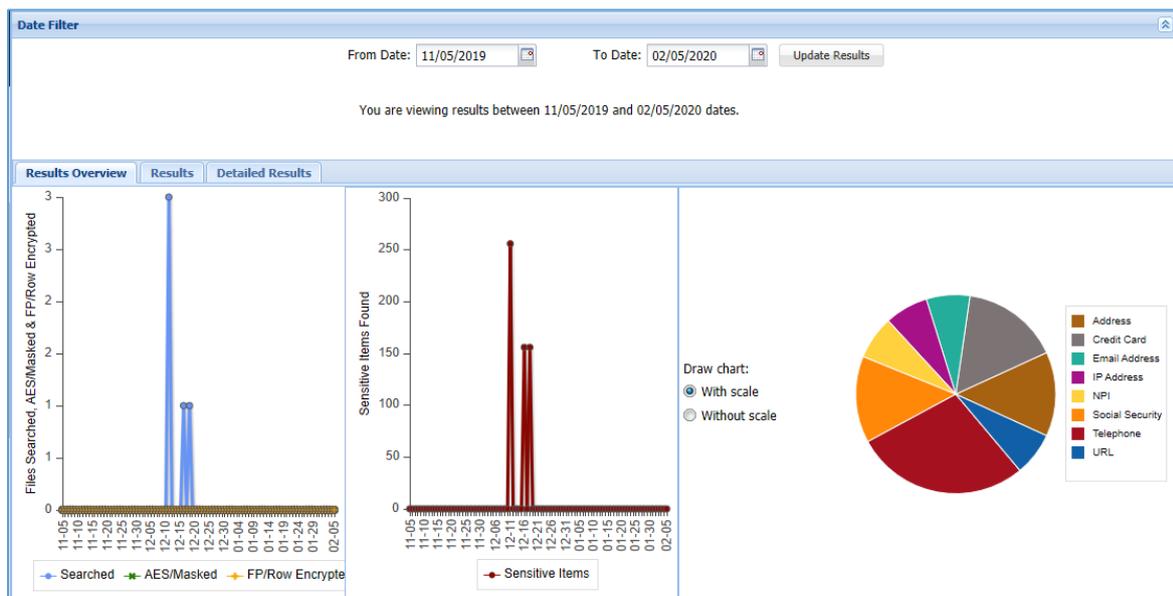
- **Logs:** For the Logs tab, the bottom panel lists any errors that occurred while the task was running (for example, if a target machine was offline and therefore unsearchable).

Error	File Name	Directory Path
No log to show.		

9.3.1.2 Results-By Date Range

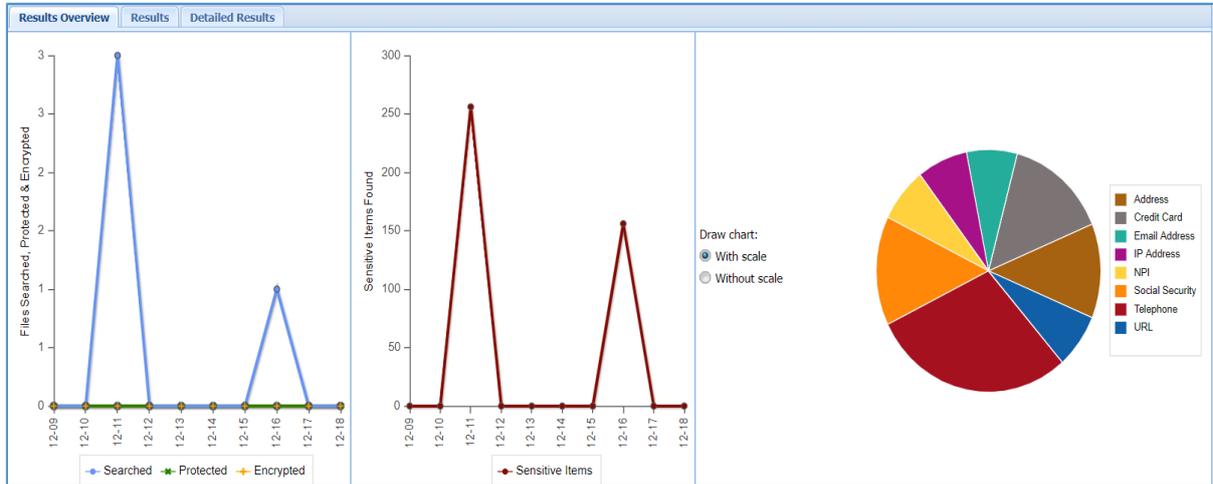
The By Date Range tab under HDFS Results, provides the details of HDFS tasks executed within a period of time. To generate time based results, take the following steps:

1. Enter From Date and To Date, which you require results between.
2. Click Update Results.
3. The three tabs on the bottom panel will get populated with the details of task, data scanned, data protected, data encrypted, sensitive items found and the breakup of different sensitive items detected on your systems within the specified period.



The bottom panel of this screen contains the following three tabs:

- **Results Overview:** For the Results Overview tab, the panel gives a graphical summary of the results for the selected task instance. Hovering over data points in any of the graphs shows the number of hits and hit percentage for a file or sensitive type. Only sensitive types uncovered by the selected task instance are displayed.



- **Results:** For the Results tab, the panel displays sensitive data by group, hit count, scan setting, task name, whether the scan was incremental, and whether the data was masked/encrypted.

Task Name	Sensitive Data Group	File Path	File Type	File Owner	Detection Type	Content Rel	Hit Count	AES/Maske	FP/Row Encrypt	FP/Row Decrypt	Modification Date
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	16	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	16	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	16	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	30	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	30	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	30	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00

- **Detailed Results:** The Detailed Results tab displays the sensitive data by type.

Task Name	Sensitive Data Type	File Path	File Type	File Owner	Detection Type	Content Re	Hit Count	AES/Maske	FP/Row Encrypt	FP/Row Decrypt	Modification Date
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Standard...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	URL	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Standard)	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Dash Se...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Da...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Sp...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Space Se...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Standard...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card # (Digits...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Da...	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Samplin...	10	false	false	false	2019-11-25 02:37:00

The Results and Detailed Results can be downloaded as a pdf and csv. Details about the following fields are shown under the Results and Detailed Results tabs:

Column	Description
Task Name	The name of the task that was executed to create this instance.
Sensitive Data Type	A specific sensitive type within a sensitive type group. The sensitive type and sensitive type group may be synonymous.
Sensitive Data Group	The type of sensitive information that has been discovered.

File Path	The location of the file with sensitive data. Hit Count. The number of values in in the listed file path that match the search criteria.
File Type	The type of file scanned.
File Owner	The user for task execution.
Detection Type	The type of detection, structured or simulated.
Content Read	The sample type of the data scanned.
Hit Count	The number of sensitive types detected.
AES/Masked	Indicates whether or not the sensitive type was masked. This column displays "False" if the sensitive data was not masked and "True" if the sensitive data was masked.

FP/Row Encrypted	Indicates whether or not the sensitive type was encrypted. This column displays "False" if the sensitive data was not encrypted and "True" if the sensitive data was encrypted.
FP/Row Decrypted	Displays "True" if the scan was incremental and "False" if it was a full scan.
Modification Date	The date on which the task was last modified.

9.3.1.3 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

Remediation							
HostName	Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Acti	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.3.2 Hive

9.3.2.1 Results-By Task

The **By Task** tab under Hive Results, provides the details of the selected Hive tasks.

The Task Instances panel, at the top of the page, contains a list of all the Hive tasks that you have permission to view.

1. Select a task instance to display its parameters and results in the panels below.
2. Click Refresh to update the list with any recently-launched task instances. Click Resume to restart canceled tasks.

Select Cluster: Hadoop Cluster

By Task | By Date Range | Saved Remediation

Task Instances

Pause | Refresh | Clear Filter | Resume | Queue For Remediation Remediated Results | With Results | Show | Hide

ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Host	Instance IP Ad	Remediated	Re-Execute	Show/H
25	Test_detection	Completed	Dec-18-2019 0...	Dec-18-2019 0...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.d...	10.12.13.11	No		
23	Test_detection	Completed	Dec-16-2019 1...	Dec-16-2019 1...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.d...	10.12.13.11	No		
20	Test_detection	Completed	Dec-11-2019 0...	Dec-11-2019 0...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.d...	10.12.13.11	No		
19	Test_detection	Completed	Dec-11-2019 0...	Dec-11-2019 0...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.d...	10.12.13.11	No		
18	Test_detection	Failed at 0.00%	Dec-11-2019 0...	Dec-11-2019 0...	DocsDemo	hdfs	document.dg.local	10.12.13.11	document.d...	10.12.13.11	No		

Page 1 of 1 Displaying 1 - 13 of 13

Task Instance Overview

Task Name: Test_detection Task Description: Detection Created By: DocsDemo
 Last Executed On: Dec-18-2019 01:25:35 Scan Type: Full Dump Metadata: false

Task Type: Detection Read Files: Part of Files Sampling Configuration: Top 1000 rows
 Include Files that failed previously: False Files Modified After: NA Files Modified Before: NA

Sampled Data: 2.90 KB Total Data: 2.90 KB

Scan Location	Structure	Domain	Sensitive Type Name
/munish			URL
			Social Security # (Space Separation)
			Social Security # (Dash Separation)
			Telephone (Space Separation)
			Telephone (Dash Separation)

Page 1 of 1 Displaying 1 - 1 of 1

Results Overview | Results | Detailed Results | Structured Summary Results | Non-Sensitive Scanned Files | Skipped Files | Logs

The middle panel shows the Task Instance Overview, Included and Excluded scanning locations and Sensitive Type Details. While the scan is running, the panel displays a progress bar that shows you the total number of tables to be scanned and the number that have been scanned so far. The various tabs on this tab are discussed below:

- **Task Instance Overview:** Provides identifying information about the selected task instance and describes the work that it performed.

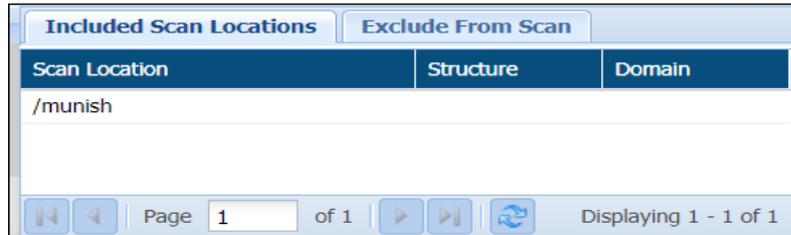
Task Instance Overview

Task Name: Test_detection Task Description: Detection Created By: DocsDemo
 Last Executed On: Dec-18-2019 01:25:35 Scan Type: Full Dump Metadata: false

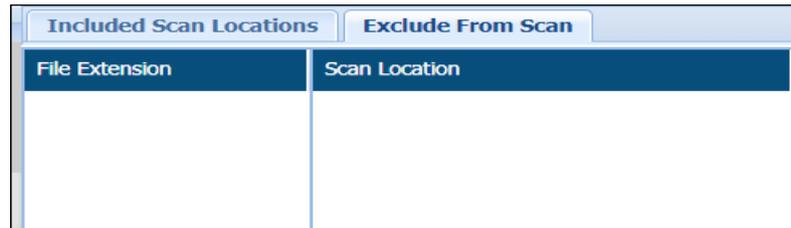
Task Type: Detection Read Files: Part of Files Sampling Configuration: Top 1000 rows
 Include Files that failed previously: False Files Modified After: NA Files Modified Before: NA

Sampled Data: 2.90 KB Total Data: 2.90 KB

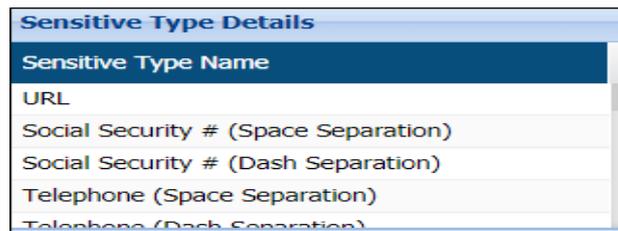
- **Included Scan Locations:** The locations added to the Include list. When the Included Scan Location tab is selected, the instance's target directory, the structure assigned to the directory, and the domain to which it is assigned are shown. Not all scan locations are associated with a structure or domain.



- **Excluded From Scan:** The locations added to the Exclude list. When the Exclude from Scan tab is selected, the panel displays any excluded file extensions and file paths.

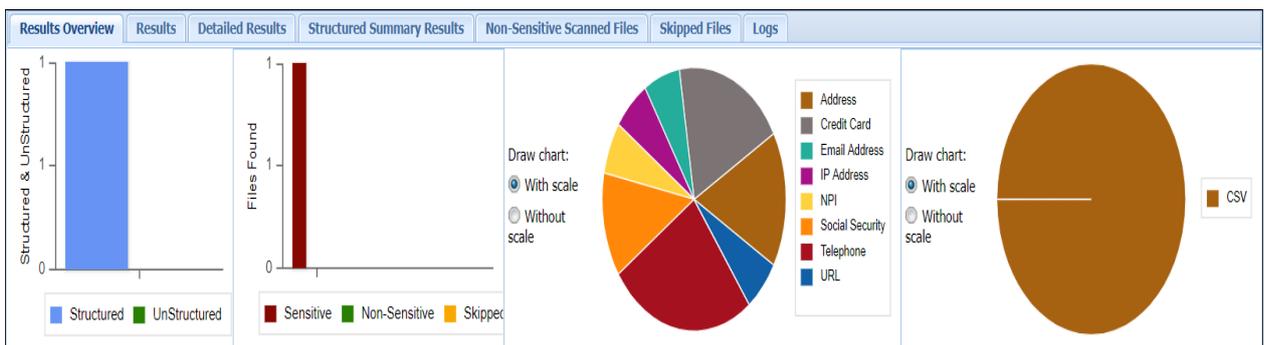


- **Sensitive Type Details:** The Sensitive Type Details panel shows the task instance's sensitive types.



The tabs shown in the bottom panel are as follows:

- **Results Overview:** For the Results Overview tab, the panel gives a graphical summary of the results for the selected task instance. Hovering over data points in any of the graphs shows the number of hits and hit percentage for a file or sensitive type.



- **Results:** For the Results tab, the panel displays sensitive data by group, hit count, scan setting, task name, whether the scan was incremental, and whether the data was masked/encrypted.

Task Name	Sensitive Data Group	File Path	File Type	File Owner	Detection Type	Content Read	Hit Count	Modification Date
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	26	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	30	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	Social Security	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	20	2019-11-25 02:37:00

- Detailed Results:** The Detailed Results tab displays the sensitive data by type. When the Detailed Results tab is selected, two buttons appear in the menu bar of the bottom panel. You can save results to file or save results to PDF.

Task Name	Sensitive Data Type	File Path	File Type	File Owner	Detection Type	Content Read	Hit Count	Modification Date
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	2	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Address City (Best suited for stru...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	1	2019-11-25 02:37:00
Test_detection	Credit Card # (Dark Generation)	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(100...	10	2019-11-25 02:37:00

The columns in the Results and Detailed Results tables are described below:

Column	Description
Task Name	The name of the task that was executed to create this instance.
Sensitive Data Type	A specific sensitive type within a sensitive type group. The sensitive type and sensitive type group may be synonymous.

Sensitive Data Group	The type of sensitive information that has been discovered.
File Path	The location of the file with sensitive data. Hit Count. The number of values in in the listed file path that match the search criteria.
File Type	The type of file scanned.
File Owner	The user for task execution.
Detection Type	The type of detection, structured or simulated.
Content Read	The sample type of the data scanned.
Hit Count	The number of sensitive types detected.
AES/Masked	Indicates whether or not the sensitive type was masked. This column displays "False" if the sensitive data was not masked and "True" if the sensitive data was masked.
FP/Row Encrypted	Indicates whether or not the sensitive type was encrypted. This column displays "False" if the sensitive data was not encrypted and "True" if the sensitive data was encrypted.
FP/Row Decrypted	Displays "True" if the scan was incremental and "False" if it was a full scan.

Modification Date	The date on which the task was last modified.
-------------------	---

- **Structured Summary Results:** This tab only appears for Detection results. It indicates the number of sensitive data items - categorized by sensitive data type - the task instance discovered and which column contains the sensitive elements.

Sensitive Type	File Path	Field No.	Confidence	Match Count	Rows Scanned	Sample Mode	Null Count	Null Ratio	Field Name Match
Address City (Best sui...	/munish/AllData.csv	Column 11	14	2	12	Sample: 0,r,1000;	1	8	false
Telephone (Standard	/munish/AllData.csv	Column 11	87	10	12	Sample: 0,r,1000;	1	8	true
Telephone (Standard ...	/munish/AllData.csv	Column 12	87	10	12	Sample: 0,r,1000;	1	8	true
Address City (Best sui...	/munish/AllData.csv	Column 12	7	1	12	Sample: 0,r,1000;	1	8	false
IP Address	/munish/AllData.csv	Column 14	87	10	12	Sample: 0,r,1000;	1	8	true
NPI	/munish/AllData.csv	Column 15	87	10	12	Sample: 0,r,1000;	1	8	true
Address City (Best sui...	/munish/AllData.csv	Column 16	70	10	12	Sample: 0,r,1000;	1	8	false

- **Non Sensitive Scanned Files:** Lists out all the fills where no sensitive content was detected.

File Path	File Type	File Size	File Owner	Modification Date
No non-sensitive scanned files.				

- **Skipped Files:** The files that were skipped during scan are listed here.

File Path	Skipped Reason	File Type	File Owner	Modification Date
No file skipped.				

- **Logs:** For the Logs tab, the bottom panel lists any errors that occurred while the task was running (for example, if a target machine was offline and therefore unsearchable).

Error	File Name	Directory Path
No log to show.		

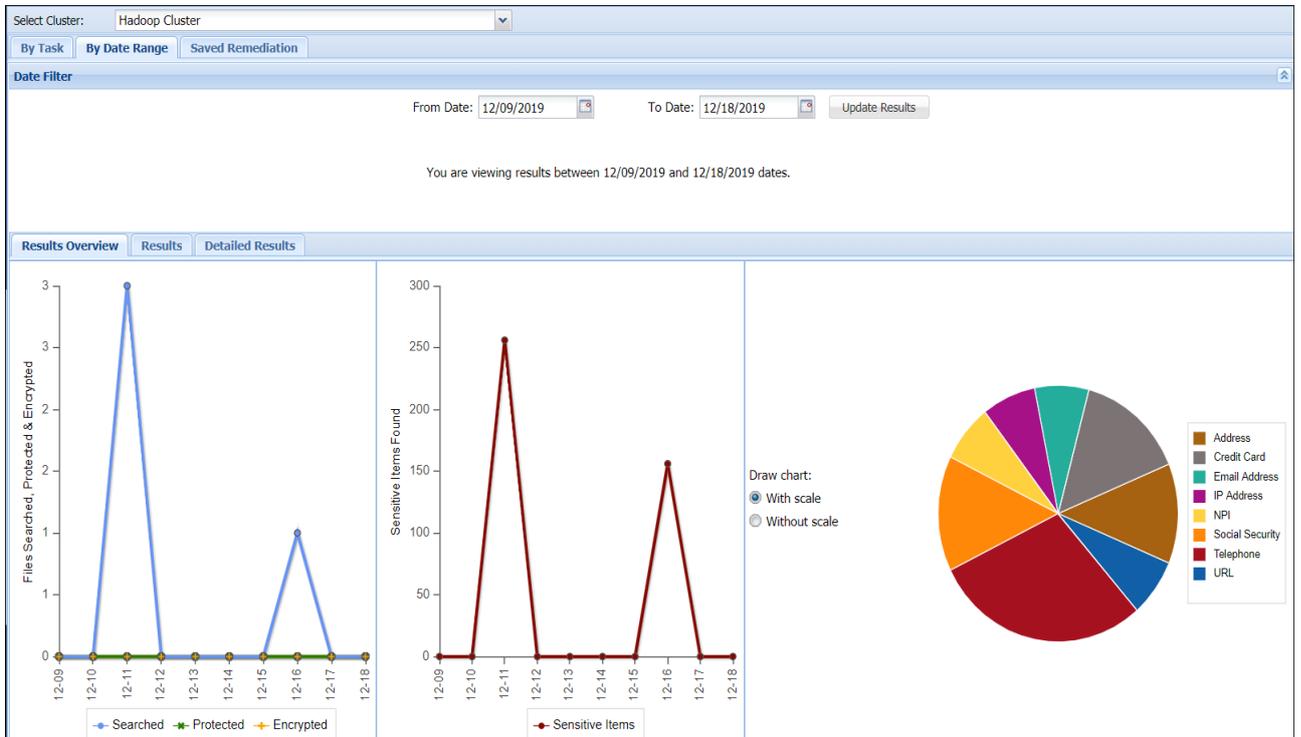
9.3.2.2 Results-By Date Range

The By Date Range tab under Hive Results, provides the details of HDFS tasks executed within a period of time.

To generate time based results, take the following steps:

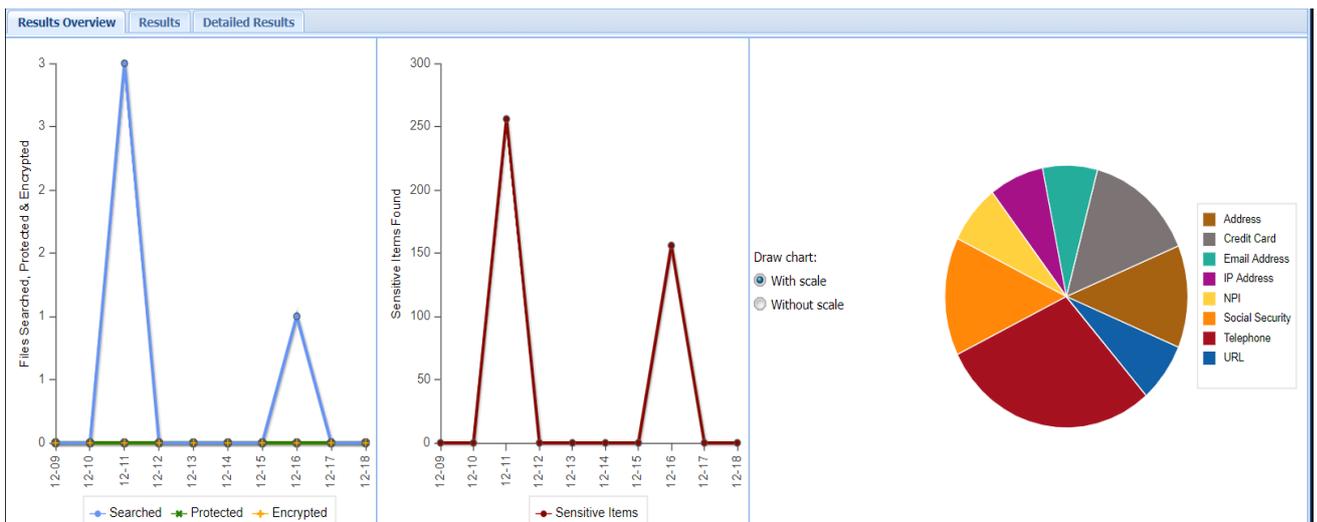
1. Enter From Date and To Date, which you require results between.
2. Click Update Results.

3. The three tabs on the bottom panel will get populated with the details of task, data scanned, data protected, data encrypted, sensitive items found and the breakup of different sensitive items detected on your systems within the specified period.



The bottom panel of this screen contains the following three tabs:

- **Results Overview:** For the Results Overview tab, the panel gives a graphical summary of the results for the selected task instance. Hovering over data points in any of the graphs shows the number of hits and hit percentage for a file or sensitive type. Only sensitive types uncovered by the selected task instance are displayed.



- **Results:** For the Results tab, the panel displays sensitive data by group, hit count, scan setting, task name, whether the scan was incremental, and whether the data was masked/encrypted.

Results Overview Results Detailed Results											
Clear Filters Save Filter											
Save Results to File Save Results to Pdf											
Task Name	Sensitive Data Group	File Path	File Type	File Owner	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted	Modification Date
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	16	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	16	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	30	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	30	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	20	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	20	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	20	false	false	false	2019-11-25 02:37:00

Page 1 of 2 | Displaying 1 - 20 of 24

- **Detailed Results:** The Detailed Results tab displays the sensitive data by type.

Results Overview Results Detailed Results											
Clear Filters Save Filter											
Save Results to File Save Results to Pdf											
Task Name	Sensitive Data Type	File Path	File Type	File Owner	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted	Modification Date
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Standard wit...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	URL	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Standard)	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Dash Separat...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Dash S...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Space ...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Space Separ...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Email Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Telephone (Standard wit...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Credit Card # (Digits Only)	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Dash S...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	NPI	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	Social Security # (Space ...	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	IP Address	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00
Test_detection	URL	/munish/AllData.csv	CSV	hdfs	Structured	Sampling(...)	10	false	false	false	2019-11-25 02:37:00

Page 1 of 2 | Displaying 1 - 20 of 40

The Results and Detailed Results can be downloaded as a pdf and csv. Details about the following fields are shown under the Results and Detailed Results tabs:

Column	Description
--------	-------------

Task Name	The name of the task that was executed to create this instance.
Sensitive Data Type	A specific sensitive type within a sensitive type group. The sensitive type and sensitive type group may be synonymous.
Sensitive Data Group	The type of sensitive information that has been discovered.
File Path	The location of the file with sensitive data. Hit Count. The number of values in in the listed file path that match the search criteria.
File Type	The type of file scanned.
File Owner	The user for task execution.
Detection Type	The type of detection, structured or simulated.

Content Read	The sample type of the data scanned.
Hit Count	The number of sensitive types detected.
AES/Masked	Indicates whether or not the sensitive type was masked. This column displays "False" if the sensitive data was not masked and "True" if the sensitive data was masked.
FP/Row Encrypted	Indicates whether or not the sensitive type was encrypted. This column displays "False" if the sensitive data was not encrypted and "True" if the sensitive data was encrypted.
FP/Row Decrypted	Displays "True" if the scan was incremental and "False" if it was a full scan.

Modification Date	The date on which the task was last modified.
-------------------	---

9.3.3 Hbase

To view results of executed tasks go to results screen under HBASE. The top panel shows all the task instances on the selected cluster.

Task Instances

ID	TASK NAME	STATUS	START TIME	END TIME	ACTION	SHOW / HIDE
34	Version_3	Completed	Dec-10-2019 16:38:53	Dec-10-2019 16:39:51		<input type="checkbox"/>
29	Version_3	Completed	Dec-10-2019 15:46:02	Dec-10-2019 15:47:13		<input type="checkbox"/>
28	Version_3	Completed	Dec-10-2019 15:15:02	Dec-10-2019 15:17:14		<input type="checkbox"/>
27	PII_1lakh	Completed	Dec-10-2019 15:04:59	Dec-10-2019 15:06:26		<input type="checkbox"/>
24	task_1million_pci	Completed	Dec-10-2019 14:02:02	Dec-10-2019 14:07:48		<input type="checkbox"/>
23	PCI_1lakh	Completed	Dec-10-2019 13:58:48	Dec-10-2019 14:00:06		<input type="checkbox"/>

Detailed Results

SENSITIVE TYPE	NAMESPACE	TABLE	COLUMN FAMILY	COLUMN	PROTECTION TYPE
Social Security # (Space Separation)	karman	table1_uncom_ver_2	Social Security Number	S_Space	FPM
Social Security # (Digits Only)	karman	table1_uncom_ver_2	Social Security Number	S_Digits	FPM
Social Security # (Dash Separation)	karman	table1_uncom_ver_2	Social Security Number	S_Dash	FPM
Credit Card # (Space Separation)	karman	table1_uncom_ver_2	CREDIT CARD NUMBER	Space	FPM
Credit Card # (Digits Only)	karman	table1_uncom_ver_2	CREDIT CARD NUMBER	Digits	FPM
Credit Card # (Dash Separation)	karman	table1_uncom_ver_2	CREDIT CARD NUMBER	Dash	FPM

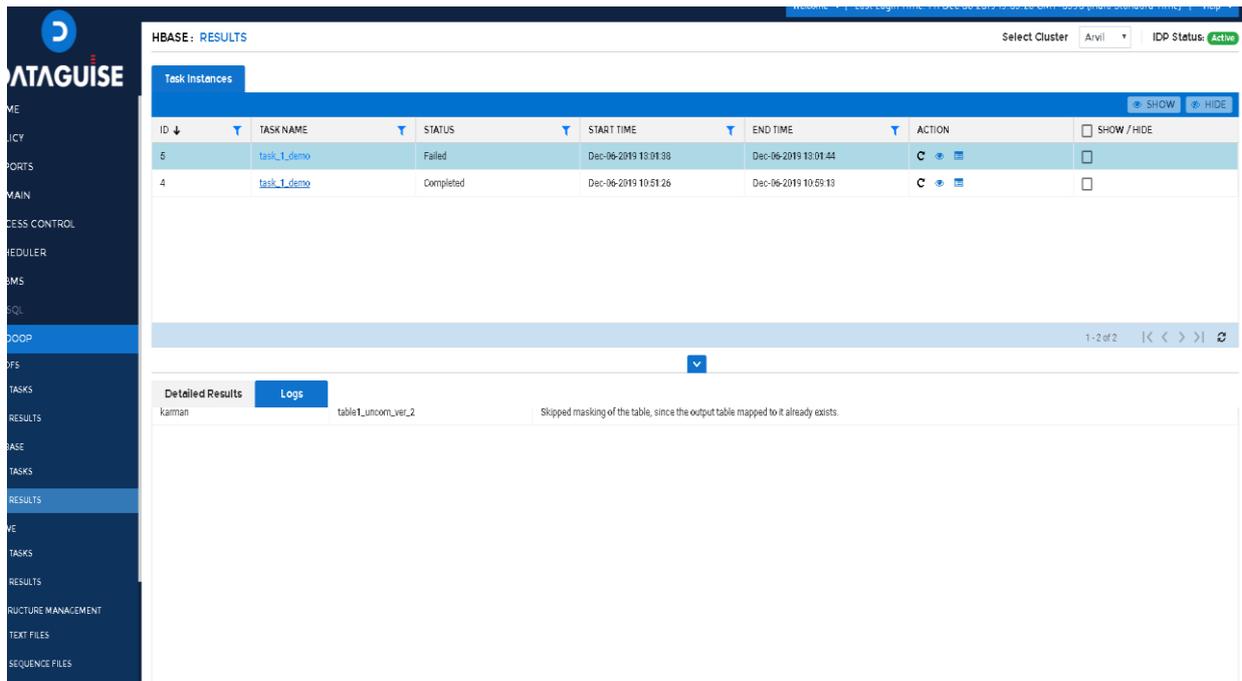
The bottom panel has two tabs:

1. Detailed Results: Provides complete description of the selected task instance.

Detailed Results

SENSITIVE TYPE	NAMESPACE	TABLE	COLUMN FAMILY	COLUMN	PROTECTION TYPE
Credit Card # (Space Separation)	karman	table1_uncom_ver_2	CREDIT CARD NUMBER	Space	FPM
Credit Card # (Digits Only)	karman	table1_uncom_ver_2	CREDIT CARD NUMBER	Digits	FPM
Credit Card # (Dash Separation)	karman	table1_uncom_ver_2	CREDIT CARD NUMBER	Dash	FPM

2. Logs: Generates results of errors of any file could not be masked due to an issue at the HBASE level, i.e., was skipped for some reason.

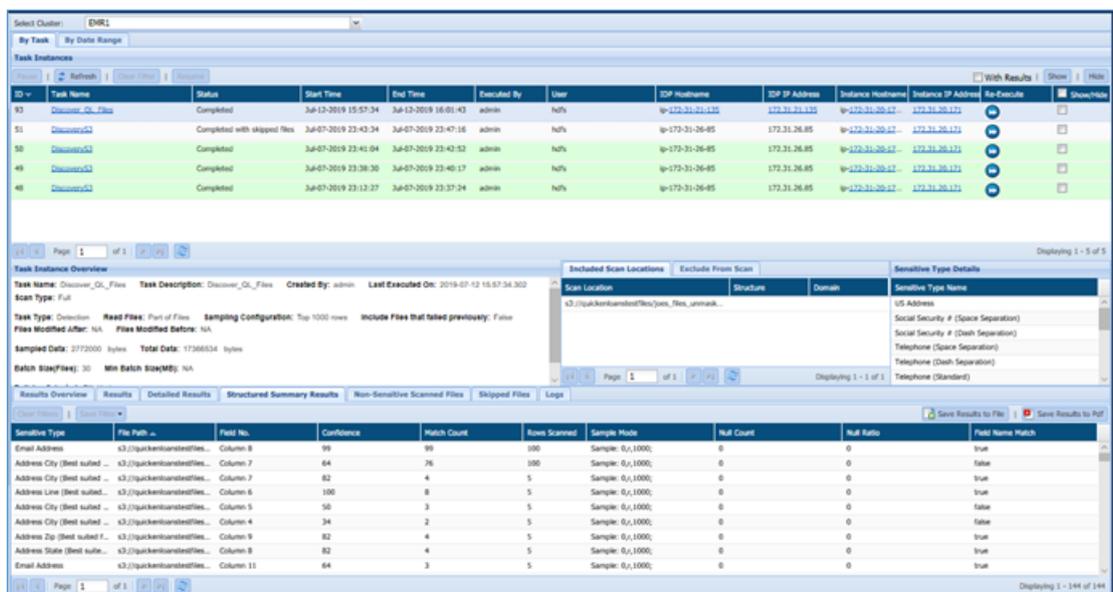


9.4 Files

The Result page displays the status and the results of the Files tasks. The information can be tracked:

1. By Task
2. By Date Range
3. Saved Remediation

The below image shows the user interface of “By Task” tab when clicked on Result page:



9.4.1 By Task

The 'By Task' pane is divided into three panes. These are:

1. Task Instances
2. Task Instance Overview
3. Overview

- **Task Instances**

The Task Instances pane display all the task instances along with their ID (system generated), Task Name, Status, Start Time, End Time, Executed By, User, IDP Hostname, IDP IP Address, Instance Hostname, Instance IP Address, Re-Execute, etc.

Click the Re-Execute button, if you wish to re-run the task again.

ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Hostname	Instance IP Address	Re-Execute	Show/Hide
199	DiscoveryFiles	Completed with s...	Aug-21-2019 ...	Aug-21-2019 ...	admin	root	ip-172-31-20-171...	172.31.20...	ip-172-31-...	172.31.20...	⏮	☐
198	DiscoveryFiles	No Conforming Files	Aug-21-2019 ...		admin	root	ip-172-31-20-171...	172.31.20...	ip-172-31-...	172.31.20...	⏮	☐
195	DiscoveryFiles	Completed with s...	Aug-21-2019 ...	Aug-21-2019 ...	admin	root	ip-172-31-20-171...	172.31.20...	ip-172-31-...	172.31.20...	⏮	☐

1. **Pause:** Click the Pause button to pause the running task for a while.
2. **Refresh:** Click the Refresh button. It will update the current page with the updated information.
3. **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Tasks page.
4. **Resume:** Click the Resume button to re-start the process from the point where it was paused.
5. **Show:** Click the Show button to unhide a task.
6. **With Results:** Check the **With Results** checkbox, if you want to see the results for the selected masking task.
7. **Queue For Remediation:** Click this button to queue the task for remediation workflow. This option allows you to queue the task for remediation manually.

Once the task has been queued for remediation, the value for the remediation can be seen in the **Task Instances** panel under the column name **REMEDIATED**. It specifies the value as **YES** or **NO**.

ID	Task Name	Status	Start Time	End Time	Remediated	Re-Execute	Show/Hide
183	email_privacy	Completed	May-19-2020 17:35:08	May-19-2020 17:35:29	Yes	⏮	☐
153	Ritish DGWalker Oracle	Completed	May-18-2020 10:18:17	May-18-2020 10:18:21	No	⏮	☐
152	Ritish DGWalker Oracle copy by parteek...	Completed	May-18-2020 10:17:13	May-18-2020 10:17:27	No	⏮	☐

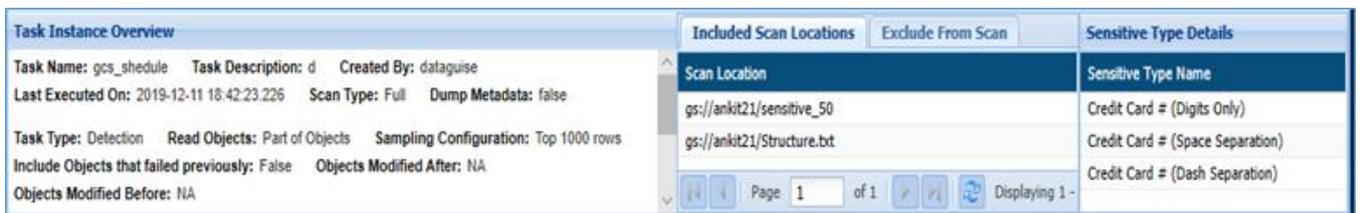
To queue the task for remediation workflow automatically, set the value of **AUTO QUEUE FOR REMEDIATION** to **YES** in DgAdmin application under **SETTINGS**.

8. **Remediated Results:** check the **Remediated Results** checkbox, if you want to view the final results for the selected fields which are marked as **Remediated**. The remediated result will be displayed in the **Detailed Result** tab under the **Overview** panel.
9. **Hide:** To hide any task, follow the below steps:
 - i. Check Show/Hide checkbox for the policy.
 - ii. Click the Hide button. The policy will get greyed out.

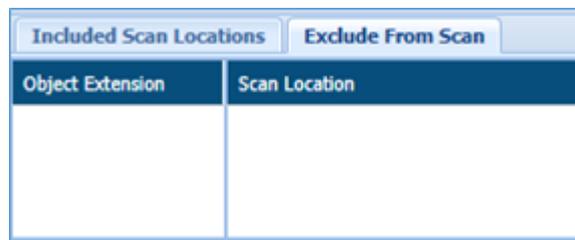
• **Task Instance Overview**

The Task Instance Overview pane will display the basic information for the Files task selected in the Task Instance pane. This includes details such as Task Name, Task Description, Created By, Last Executed On, etc.

The Task Instance Overview pane will also display information for the Scanned Locations in Included Scan Locations and Exclude From Scan tab, Sensitive Type Details.



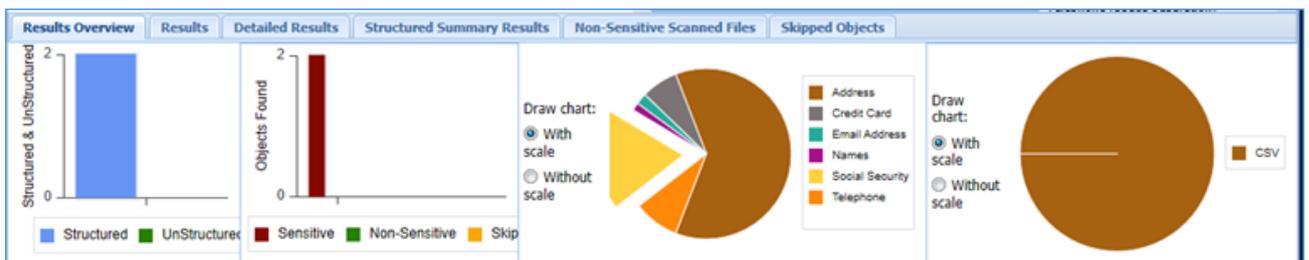
1. **Included Scan Locations:** It displays the information about the task target directories.
2. **Exclude From Scan:** It will display the list of excluded object extension and the file location.



3. **Sensitive Type Details:** It display the list of all Sensitive Types discovered in the task.

• **Overview**

The bottom panel will show the detailed information for the selected task. The information displayed is dependent on the currently selected tab.



1. **Result Overview:** The Result Overview tab gives a graphical summary of the results for the selected task instance. Hovering over the data points in any of the graphs shows the number of hits and hit percentage for a file or an expression.
2. **Results:** The Results tab displays sensitive data groups, hit count, task name, Object Path, Object Type, Detection Type and Content Read.

Task Name	Sensitive Data Group	Object Path	Object Type	Detection Type	Content Read	Hit Count
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	ORC	Structured	Sampling(1000 ...	76
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	AVRO	Structured	Sampling(1000 ...	27
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	5
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	5
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	23
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	22

3. **Detailed Results:** The Detailed Results tab will provide you the detailed information for the Sensitive Type discovered in the database. The information will include Sensitive Data Type, Task Name, Object Path, Object Type, Detection Type, Content Read and Hit Count.

Task Name	Sensitive Data Type	Object Path	Object Type	Detection Type	Content Read	Hit Count
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	AVRO	Structured	Sampling(1000 ...	41
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	146
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	5
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	2
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	ORC	Structured	Sampling(1000 ...	9
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	6

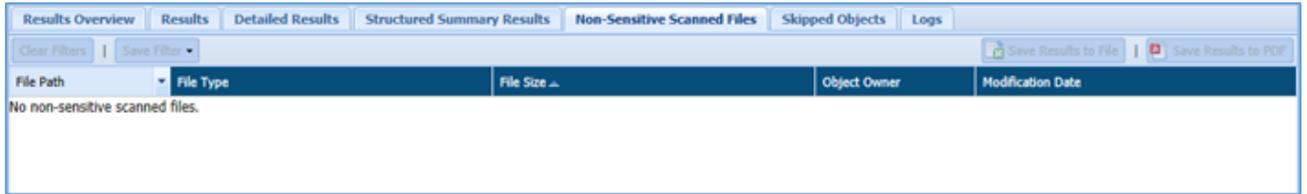
4. Structured Summary Results:

Sensitive Type	File Path	Field No.	Confidence	Match Count	Rows Scanned	Sample Mode	Null Count	Null Ratio	Field Name Match
Email Address	s3://quickenloan...	Column 8	99	99	100	Sample: 0,r,1000;	0	0	true
Address City (Be...	s3://quickenloan...	Column 7	64	76	100	Sample: 0,r,1000;	0	0	false
Address City (Be...	s3://quickenloan...	Column 7	82	4	5	Sample: 0,r,1000;	0	0	true
Address Line (Be...	s3://quickenloan...	Column 6	100	8	5	Sample: 0,r,1000;	0	0	true
Address City (Be...	s3://quickenloan...	Column 5	50	3	5	Sample: 0,r,1000;	0	0	false
Address City (Be...	s3://quickenloan...	Column 4	74	2	5	Sample: 0,r,1000;	0	0	false

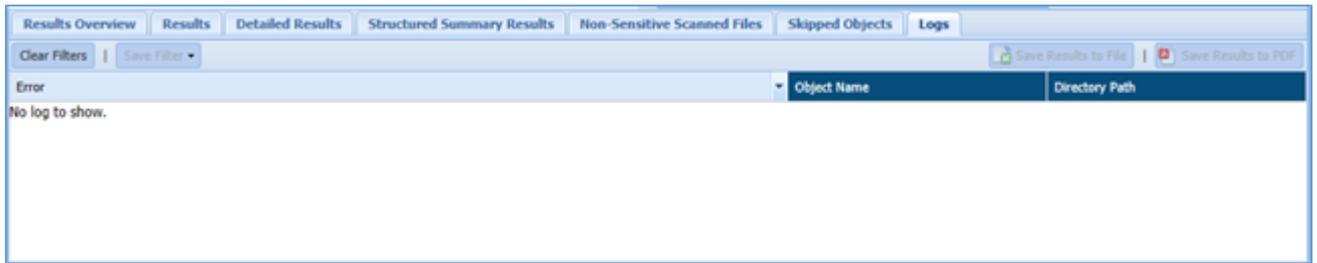
5. **Skipped Objects:** This tab will provide you with the list of objects that were skipped during the execution. It displays information about the Object Path, Skipped Reason, Object Type, Object Owner and Notification Date.

Object Path	Skipped Reason	Object Type	Object Owner	Modification Date
No file skipped.				

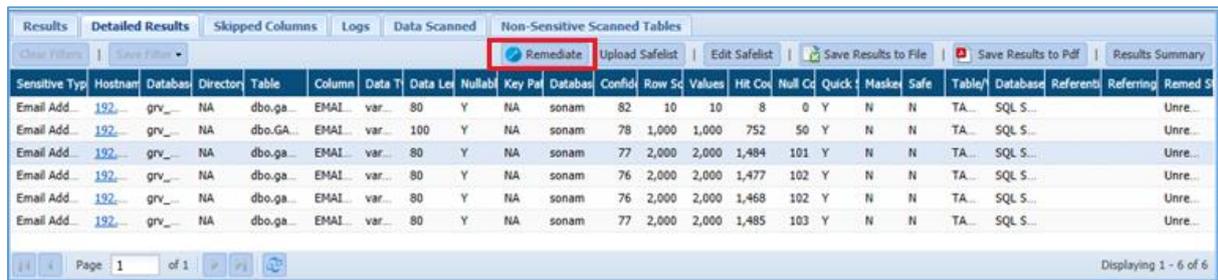
6. Non-Sensitive Scanned Files:



7. **Logs:** It will list errors that occurred during task execution.



- **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
 - **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
12. **Remediate:** Click this button to remediate any result in **STRUCTURED SUMMARY** tab. To remediate the selected result, perform the following steps:
- c) Select the record in the **Detailed Result** tab and click the **Remediate** button.



d) Enter the details for **Remediation Action** and **Remediation Value (%)** field in the popup.

Remediation

Scope

Host Name: DB Name:

Remediation Action: Remediation Value(%):

Table Name: Field Name:

Sensitive Type: Enabled

- iii. Following are the options in **Remediation Action** field.
 - **Mark as Non Sensitive:** Select this option to mark the record as Non Sensitive. The system will ignore this record in future re-run.
 - **Mark as Correct:** Select this option to mark the record as Sensitive. The system will not ignore this record in future re-run.
 - **Decrease the Confidence Factor:** select this option to decrease the confidence factor in Remediation Value (%) field.
 - **Increase the Confidence Factor:** Select this option increase the confidence factor in Remediation Value (%) field.
- iv. Enter the remediation value in percentage. This field will be enabled when you select **Decrease** or **Increase the Confidence Factor** in **Remediation Action** field.

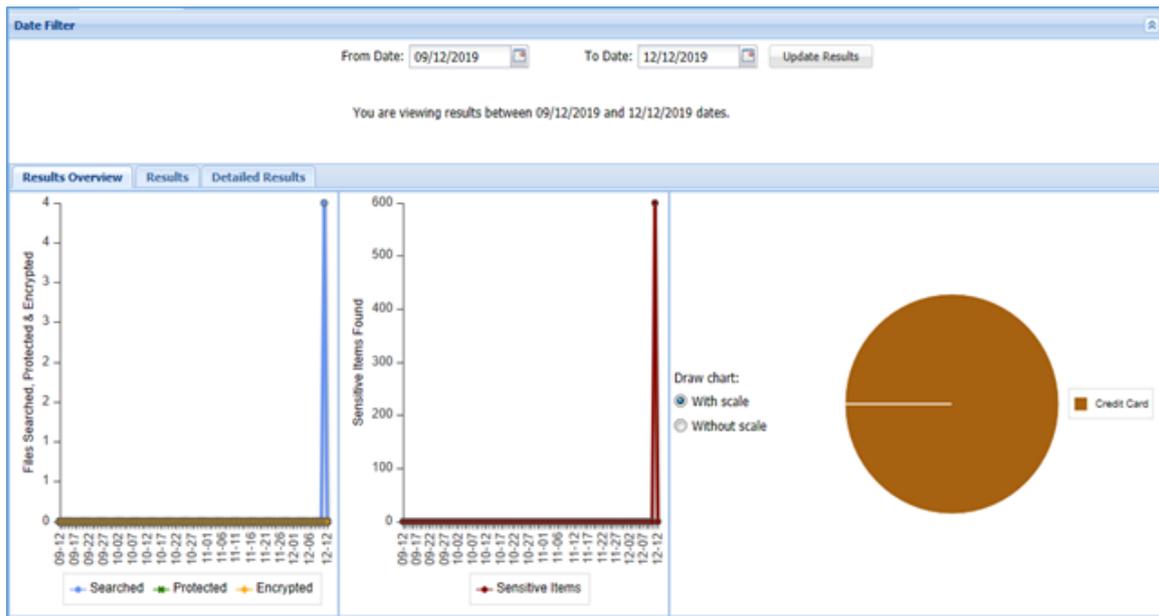
Once the remediation value has been set and saved, the same can be seen under Saved Remediation tab.

Scope		Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Acti	Enabled
HostName	Database Name						
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

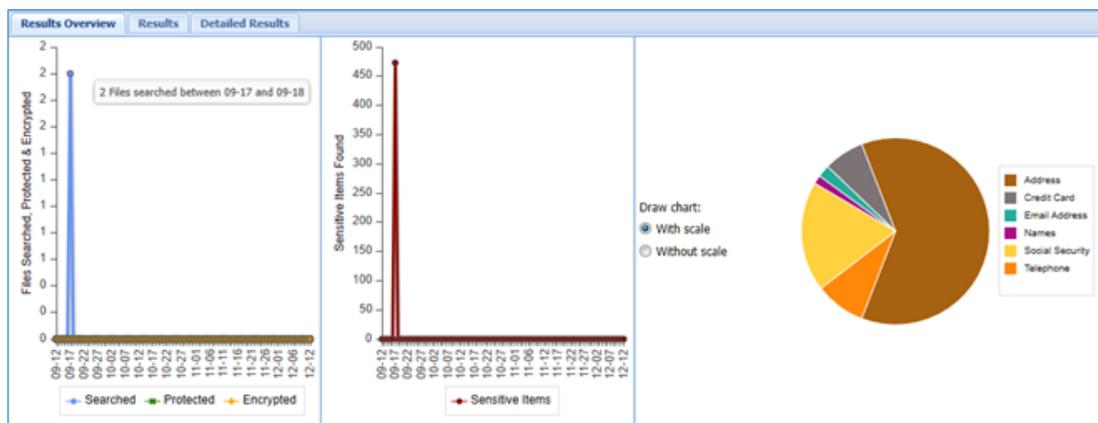
9.4.2 By Date Range

The 'By Date Range' tab is divided into two panes. These are:

1. Date Filter
2. Overview



- **Date Filter:** In Date Filter pane, you can specify the date range. Using this option, you can filter out results based on the date range.
- **Overview**
The Overview pane display the graphical representation of the information in Results Overview tab and detailed information in Results and Detailed Results tab.



1. **Results Overview:** The Results Overview tab display information in form of graphical representation. It displays two bar chart and a pie chart.
 - The first bar chart shows the number of files searched, encrypted and masked.
 - The second bar chart shows the total number of sensitive files found.
 - The pie chart shows the Sensitive Type found in each segment.
2. **Results:** The Results tab lists down the Sensitive Data Group detected in the specified time range. It includes information such as Task Name, Sensitive Data Group, Object Path, Object Type, Detection Type, Content Read, Hit Count, AES/Masked, FP/Row Encrypted, FP/Row Decrypted.

Task Name	Sensitive Data Group	Object Path	Object Type	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted
DiscoveryGCS	Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	26	false	false	false
DiscoveryGCS	Address	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	266	false	false	false
DiscoveryGCS	Credit Card	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	30	false	false	false
DiscoveryGCS	Credit Card	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	3	false	false	false
DiscoveryGCS	Email Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Names	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	7	false	false	false
DiscoveryGCS	Social Security	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	20	false	false	false
DiscoveryGCS	Social Security	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	70	false	false	false
DiscoveryGCS	Telephone	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	40	false	false	false
DiscoveryGCS	Telephone	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	1	false	false	false

- Detailed Results:** The Detailed Results tab lists down the Sensitive Types detected in the specified time range, grouped by type and location. The tab also displays information such as Task Name, Sensitive Data Group, Object Path, Object Type, Detection Type, Content Read, Hit Count, and whether the data was masked.

Task Name	Sensitive Data Type	Object Path	Object Type	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted
DiscoveryGCS	Address Line (Best suited ...)	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	154	false	false	false
DiscoveryGCS	US Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Address City (Best suited ...)	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	16	false	false	false
DiscoveryGCS	Address City (Best suited ...)	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	71	false	false	false
DiscoveryGCS	Address Zip (Best suited f...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	41	false	false	false
DiscoveryGCS	Credit Card # (Space Sep...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	3	false	false	false
DiscoveryGCS	Credit Card # (Digits Only)	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Credit Card # (Space Sep...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Credit Card # (Dash Sepa...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Email Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Full Names	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	7	false	false	false
DiscoveryGCS	Social Security # (Space ...)	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Social Security # (Space ...)	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	70	false	false	false
DiscoveryGCS	Social Security # (Dash S...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Telephone (Space Separa...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false

- **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
- **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
- **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Detection task page.

9.4.3 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

Scope		Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Act	Enabled
HostName	Database Name						
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.5 AWS

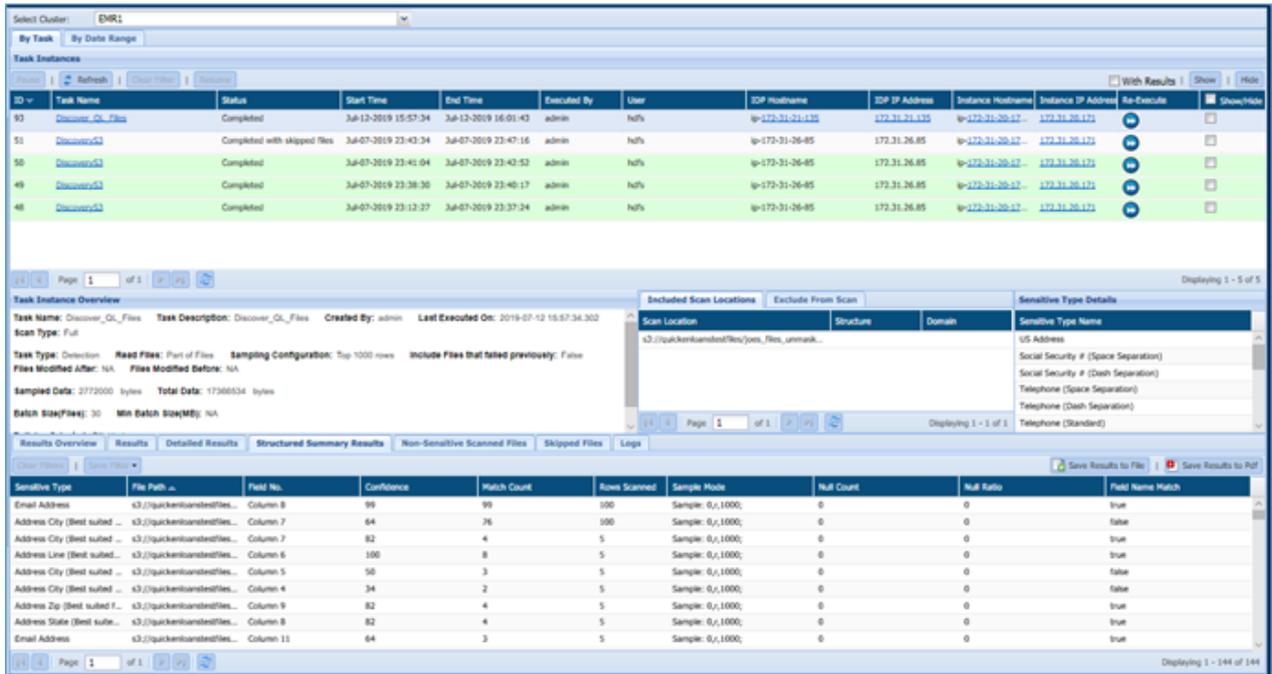
In AWS, you can track the information for the tasks executed in S3 and RedShift/RDS.

9.5.1 S3

The Result page displays the status and the results for the S3 tasks. The information can be tracked:

1. By Task
2. By Date Range
3. Saved Remediation

The below image shows the user interface of “By Task” tab when clicked on Result page:



9.5.1.1 By Task

The 'By Task' tab is divided into three panes. These are:

1. Task Instances
2. Task Instance Overview
3. Overview

- **Task Instances:**

The Task Instances pane display all the task instances along with their ID (system generated), Task Name, Status, Start Time, End Time, Executed By, User, IDP Hostname, IDP IP Address, Instance Hostname, etc. Click the Re-Execute button, if you wish to re-run the task again.

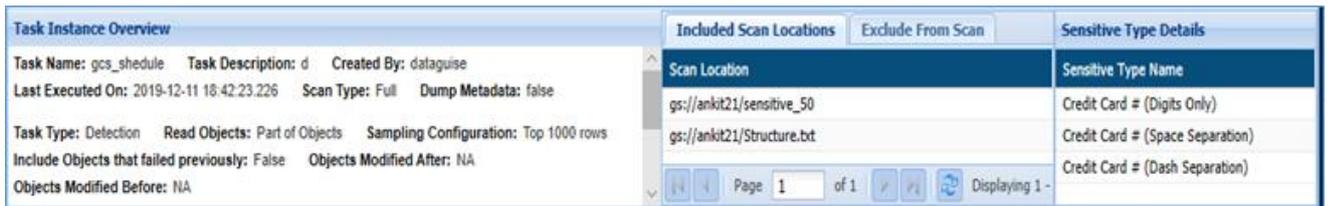
1. Pause: Click the Pause button to pause the running task for a while.
2. Refresh: Click the Refresh button. It will update the current page with the updated information.
3. Clear Filters: Click the Clear Filters button. It will remove any applied filters on the Tasks page.
4. Resume: Click the Resume button to re-start the process from the point where it was paused.
5. Queue For Remediation
6. With Results
7. Show: Click the Show button to unhide a task.
8. Hide: To hide any task, follow the below steps:
 - i. Check Show/Hide checkbox for the policy.

ii. Click the Hide button. The policy will get greyed out.

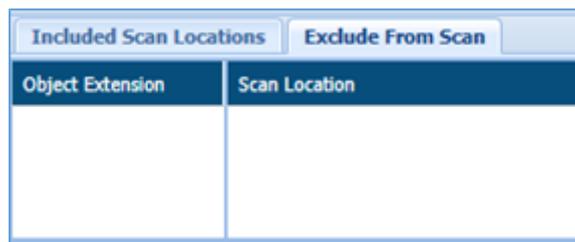
• **Task Instance Overview:**

The Task Instance Overview pane will display the basic information for the GCS task selected in the Task Instance pane. The includes details such as Task Name, Task Description, Created By, Last executed On, etc.

The Task Instance Overview pane will also display information for the Scanned Locations in included Scan Locations and Exclude From Scan tab, Sensitive Type Details.



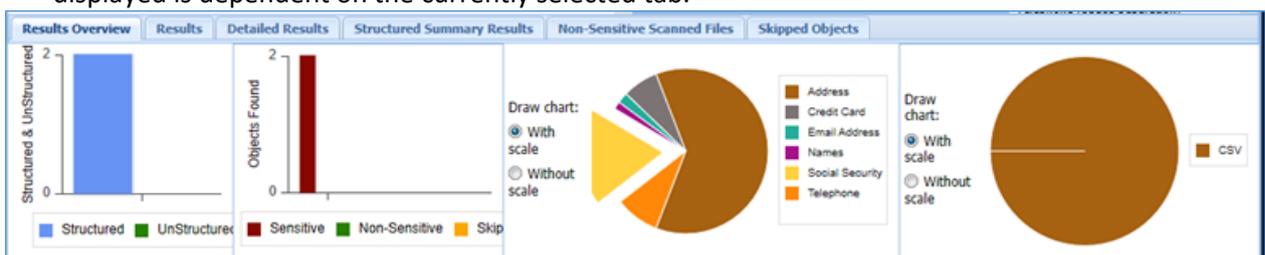
1. **Included Scan Locations:** It displays the information about the task target directories.
2. **Exclude From Scan:** It will display the list of excluded object extension and the file location.



3. **Sensitive Type Details:** It display the list of all Sensitive Types discovered in the task.

▪ **Overview:**

The bottom panel will show the detailed information for the selected task. The information displayed is dependent on the currently selected tab.



1. **Result Overview:** The Result Overview tab gives a graphical summary of the results for the selected task instance. Hovering over the data points in any of the graphs shows the number of hits and hit percentage for a file or an expression.
2. **Results:** The Results tab displays sensitive data groups, hit count, task name, Object Path, Object Type, Detection Type and Content Read.

Task Name	Sensitive Data Group	Object Path	Object Type	Detection Type	Content Read	Hit Count
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	ORC	Structured	Sampling(1000 ...	76
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	AVRO	Structured	Sampling(1000 ...	27
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	5
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	5
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	23
Discover_QL_Files	Address	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	22

3. **Detailed Results:** The Detailed Results tab will provide you the detailed information for the Sensitive Type discovered in the database. The information will include Sensitive Data Type, Task Name, Object Path, Object Type, Detection Type, Content Read and Hit Count.

Task Name	Sensitive Data Type	Object Path	Object Type	Detection Type	Content Read	Hit Count
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	AVRO	Structured	Sampling(1000 ...	41
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	146
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	5
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	2
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	ORC	Structured	Sampling(1000 ...	9
Discover_QL_Files	Address City (Best suited for structu...	s3://quickenloanstestfiles/joes_files...	TEXT	Unstructured	Sampling(1000 ...	6

4. **Structured Summary Results:**

Sensitive Type	File Path	Field No.	Confidence	Match Count	Rows Scanned	Sample Mode	Null Count	Null Ratio	Field Name Match
Email Address	s3://quickenloan...	Column 8	99	99	100	Sample: 0,r,1000;	0	0	true
Address City (Be...	s3://quickenloan...	Column 7	64	76	100	Sample: 0,r,1000;	0	0	false
Address City (Be...	s3://quickenloan...	Column 7	82	4	5	Sample: 0,r,1000;	0	0	true
Address Line (Be...	s3://quickenloan...	Column 6	100	8	5	Sample: 0,r,1000;	0	0	true
Address City (Be...	s3://quickenloan...	Column 5	50	3	5	Sample: 0,r,1000;	0	0	false
Address City (Re...	s3://quickenloan...	Column 4	74	2	5	Sample: 0,r,1000;	0	0	false

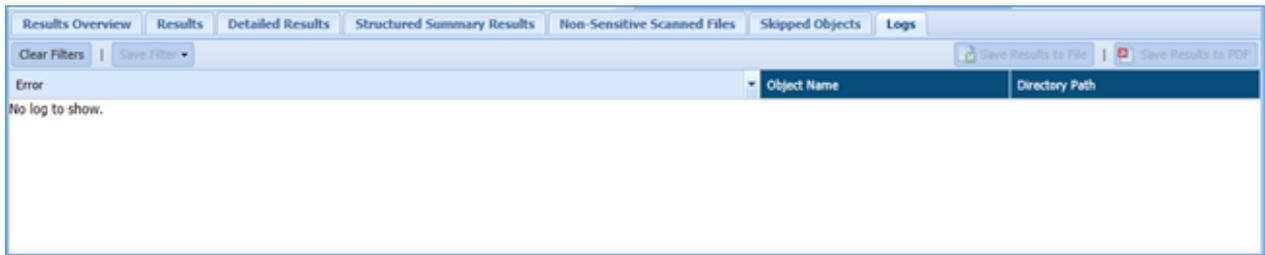
5. **Skipped Objects:** This tab will provide you with the list of objects that were skipped during the execution. It displays information about the Object Path, Skipped Reason, Object Type, Object Owner and Notification Date.

Object Path	Skipped Reason	Object Type	Object Owner	Modification Date
No file skipped.				

6. **Non-Sensitive Scanned Files:**

File Path	File Type	File Size	Object Owner	Modification Date
No non-sensitive scanned files.				

7. **Logs:** It will list errors that occurred during task execution.

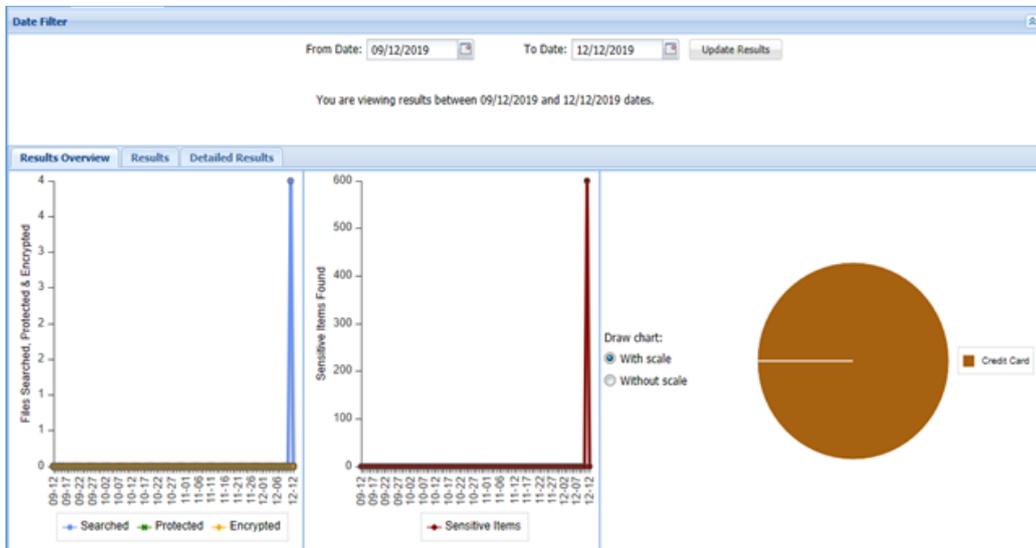


- **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
- **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.

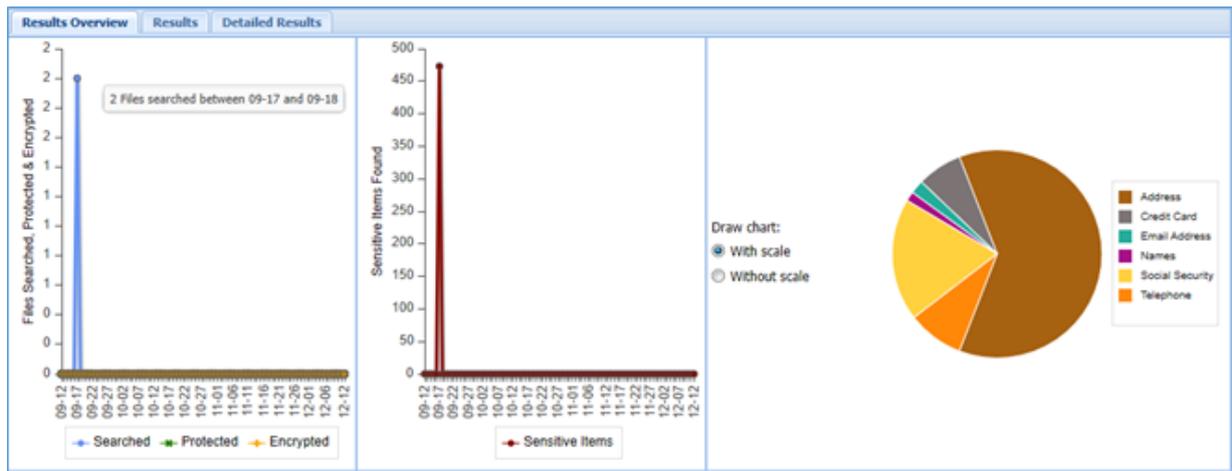
9.5.1.2 *By Date Range*

The 'By Date Range' tab is divided into two panes. These are:

1. Date Filter
2. Overview



- **Date Filter:** In Date Filter pane, you can specify the date range. Using this option, you can filter out result based on the date range.
- **Overview:** The Overview pane display the graphical representation of the information in Results Overview tab and detailed information in Results and Detailed Results tab.



- Results Overview:** The Results Overview tab displays information in graphical form. It includes:
 - The first bar chart shows the number of files searched, encrypted, and masked.
 - The second bar chart shows the total number of sensitive files found.
 - The pie chart shows the sensitive types found in each segment.
- Results:** The Results tab lists the Sensitive Data Groups detected in the specified time range. It includes information such as Task Name, Sensitive Data Group, Object Path, Object Type, Detection Type, Content Read, Hit Count, AES/Masked, FP/Row Encrypted, and FP/Row Decrypted.

Task Name	Sensitive Data Group	Object Path	Object Type	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted
DiscoveryGCS	Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	26	false	false	false
DiscoveryGCS	Address	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	266	false	false	false
DiscoveryGCS	Credit Card	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	30	false	false	false
DiscoveryGCS	Credit Card	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	3	false	false	false
DiscoveryGCS	Email Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Names	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	7	false	false	false
DiscoveryGCS	Social Security	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	20	false	false	false
DiscoveryGCS	Social Security	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	70	false	false	false
DiscoveryGCS	Telephone	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	40	false	false	false
DiscoveryGCS	Telephone	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	1	false	false	false

- Detailed Results:** The Detailed Results tab lists the Sensitive Types detected in the specified time range, grouped by type and location. It also displays information such as Task Name, Sensitive Data Group, Object Path, Object Type, Detection Type, Content Read, Hit Count, and whether the data was masked.

Task Name	Sensitive Data Type	Object Path	Object Type	Detection Type	Content Read	HR Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted
DiscoveryGCS	Address Line (Best suited ...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	154	false	false	false
DiscoveryGCS	US Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Address City (Best suited ...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	16	false	false	false
DiscoveryGCS	Address City (Best suited ...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	71	false	false	false
DiscoveryGCS	Address Zip (Best suited f...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	41	false	false	false
DiscoveryGCS	Credit Card # (Space Sep...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	3	false	false	false
DiscoveryGCS	Credit Card # (Digits Only)	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Credit Card # (Space Sep...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Credit Card # (Dash Sepa...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Email Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Full Names	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	7	false	false	false
DiscoveryGCS	Social Security # (Space ...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Social Security # (Space ...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	70	false	false	false
DiscoveryGCS	Social Security # (Dash S...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Telephone (Space Separa...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false

- **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
- **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
- **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Detection task page.

9.5.1.3 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

HostName	Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Acti	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.5.2 RedShift/RDS

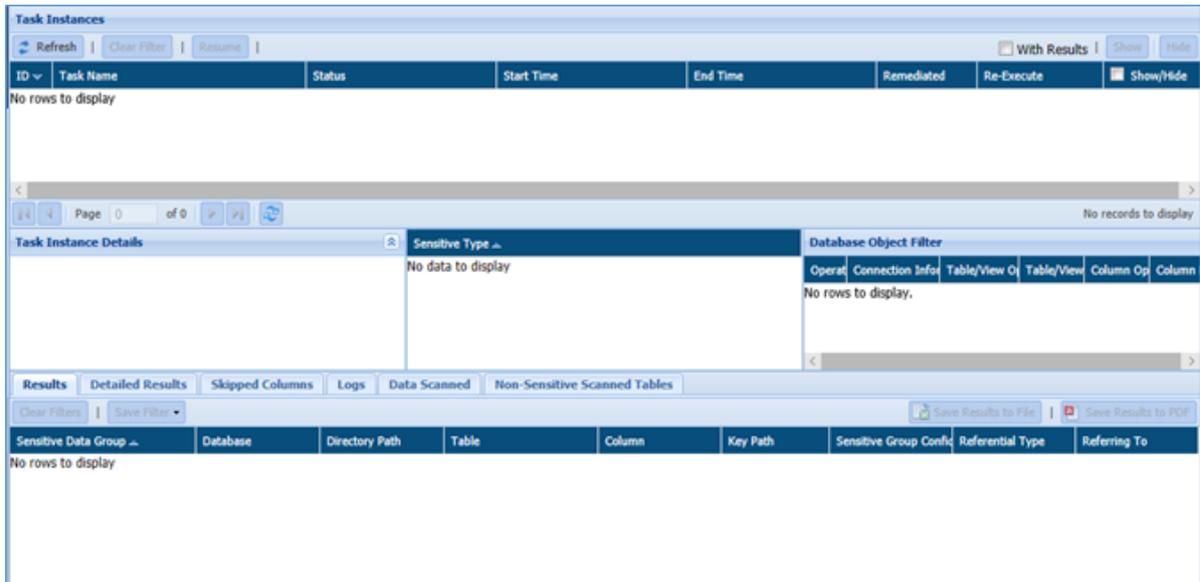
In RedShift/RDS, the Result page displays the status and results of the Detection and Masking tasks.

9.5.2.1 Detection

9.5.2.1.1 Results

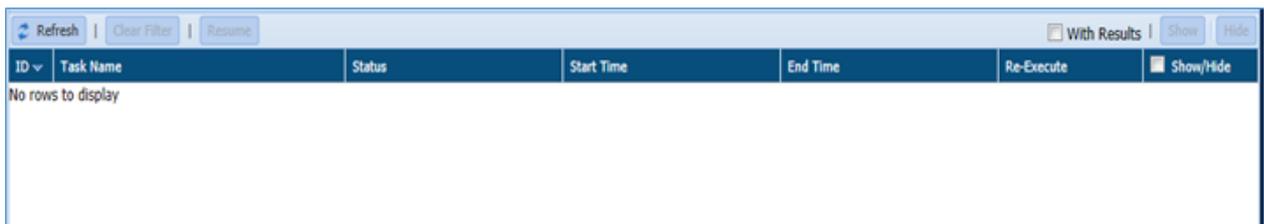
The Results tab is divided in three panes. These are:

1. Task Instances
2. Task Instance Details
3. Overview



- **Task Instances**

This pane will display the information for the task such as ID (system generated), Task Name, Status, Start Time, End Time, Show/Hide, etc. Click the Re-Execute button next to the selected task to re-run it. Selecting a task instance displays its parameters and results in the panels below.



1. **Refresh:** Click the Refresh button. It will update the current page with the updated information.
2. **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Tasks page.
3. **Resume:** Click the Resume button to re-start the process from the point where it was paused.
4. **Show:** Click the Show button to unhide a task.
5. **Hide:** To hide any task, follow the below steps:
 - i. Check Show/Hide checkbox for the policy.
 - ii. Click the Hide button. The policy will get greyed out.
6. **Queue For Remediation:** Click this button to queue the task for remediation workflow. This option allows you to queue the task for remediation manually.

Once the task has been queued for remediation, the value for the remediation can be seen in the **Task Instances** panel under the column name **REMEDIATED**. It specifies the value as **YES** or **NO**.

ID	Task Name	Status	Start Time	End Time	Remediated	Re-Execute	Show/Hide
183	email_privacy	Completed	May-19-2020 17:35:08	May-19-2020 17:35:29	Yes		<input type="checkbox"/>
153	Ritish_DGWalker_Oracle	Completed	May-18-2020 10:18:17	May-18-2020 10:18:21	No		<input type="checkbox"/>
152	Ritish_DGWalker_Oracle_copy_by_parteek...	Completed	May-18-2020 10:17:13	May-18-2020 10:17:27	No		<input type="checkbox"/>

To queue the task for remediation workflow automatically, set the value of **AUTO QUEUE FOR REMEDIATION** to **YES** in DgAdmin application under **SETTINGS**.

7. **Remediated Results:** check the **Remediated Results** checkbox, if you want to view the final results for the selected fields which are marked as **Remediated**. The remediated result will be displayed in the **Detailed Result** tab under the **Overview** panel.
 8. **With Results:** Check the **With Results** checkbox, if you want to see the results for the selected masking task.
- **Task Instance Details:**
The Task Instance Detail pane will display the basic information for the RDS/RedShift detection task selected in the Task Instance pane. It includes details such as Task Name, Task Description, Created By, Last Executed On, etc.

The Task Instance Details pane will also display the list of Sensitive Types and Database Object Filter used.

Task Instance Details		Sensitive Type	Database Object Filter
Task Name: mysql_rds	Task Instance ID: 94	PCI_DBMS	Operat, Connection Infor, Table/View O, Table/View, Column Op, Column
Task Type: Detection		Credit Card # (Dash Separation)	No rows to display.
Start Time: Dec-13-2019 14:00:05		Credit Card # (Digits Only)	
Sampling Configuration: Top 1000 rows		Credit Card # (Space Separation)	

1. **Sensitive Type:** The Sensitive Type pane will display the list of Sensitive Type for the selected task.

2. **Database Object Filter:** The database Object Filter pane displays only those databases/tables/columns that match the filter applied. The columns display information such as the applied operator, connection information, Table/View operator, Table/View Filter, column operator and column filter.

- **Overview:**

The Overview pane will show the overview of basic information for the selected task. The information displayed is dependent on the currently selected tab.

Sensitive Data Group	Database	Directory Path	Table	Column	Key Path	Sensitive Group Conf	Referential Type	Referring To
Address	shubi_detection	NA	Address_zip	Add_zip	NA	100		
Address	shubi_detection	NA	all_mismatch	ccno	NA	84		
Address	shubi_detection	NA	guri_new	address	NA	54		
Address	shubi_detection	NA	guri_new	Address_city	NA	100		
Address	shubi_detection	NA	guri_new	address_country	NA	68		
Address	shubi_detection	NA	guri_new	Address_state	NA	78		
Address	shubi_detection	NA	guri_new	Address_zip	NA	92		

1. **Results:** The result tab will display the high level information of the Sensitive Types for the selected task in the detection task pane. The information includes Sensitive Data Group, Database, Table, Column, Sensitive Group Confidence, etc.

2. **Detailed Results:** The Detailed Results tab will provide you the detailed information for the Sensitive Type encountered in the database. The information will include Sensitive Type, Hostname, Database, Table, Column, Database User, etc.

Sensitive Type	Hostname	Database	Directory	Table	Column	Data Ty	Data Len	Nullabl	Key Pat	Database	Confid	Row Sc	Values	Hit Co	Null C	Quick	Maske	Safe	Table/	Database	Referenti	Referring
Address Cit...	10.12...	dgstar	NA	tmp_stmt	stmt	VAR...	4000	Y	NA	root	7	99	99	8	0	N	N	N	BA...	MySQL		
Address Cit...	10.12...	sys	NA	sys_config	variable	VAR...	128	N	NA	root	14	6	6	1	0	N	N	N	BA...	MySQL	Primar...	
Address St...	10.12...	dgstar	NA	dg_dbm...	location	VAR...	64	Y	NA	root	84	1	1	1	0	N	N	N	BA...	MySQL		
Address St...	10.12...	dgstar	NA	tmp_stmt	stmt	VAR...	4000	Y	NA	root	10	99	99	12	0	N	N	N	BA...	MySQL		
Address Zi...	10.12...	sys	NA	sys_config	value	VAR...	128	Y	NA	root	14	6	6	1	1	N	N	N	BA...	MySQL		
Date (Best ...	10.12...	dgstar	NA	dg_clou...	updat...	DAT...	19	Y	NA	root	40	0	0	0	0	Y	N	N	BA...	MySQL		
Date (Best ...	10.12...	dgstar	NA	dg_dbm...	updat...	DAT...	19	Y	NA	root	40	0	0	0	0	Y	N	N	BA...	MySQL		

- Remediate
- Upload Safelist
- Edit Safelist
- Results Summary

3. **Skipped Columns:** This tab will provide you with the list of columns that were skipped during the execution.

Connection Name	Database Name	Directory Path	Table Name	Column Name	Data Type
No rows to display					

- Logs:** It will list errors that occurred during task execution.

Database Name	Directory Path	Table Name	Connection Name	Error
No rows to display				

- Data Scanned:** List the total number of rows in the scanned tables and how many of those rows DgSecure actually scanned.

Host Name	Database Name	Directory Path	Table Name	Total Rows	Rows Scanned (Max)	Total Data	Sampled Data
No rows to display.							

- Non Sensitive Scanned Tables:** It will list down the tables which contain non sensitive data. This tab includes information such as Host Name, Database Name, Table Name, Table Size(bytes).

- Remediate:** Click this button to remediate any result in **DETAILED RESULTS** tab. To remediate the selected result, perform the following steps:
 - Select the record in the **Detailed Result** tab and click the **Remediate** button.

Sensitive Type		Hostnam	Databas	Director	Table	Column	Data T	Data Len	Nullabl	Key Pal	Databas	Confid	Row Sc	Values	HR Co	Null Co	Quick	Masked	Safe	Table	Database	Referent	Referring	Remed S
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	82	10	10	8	0	Y	N	N	TA...	SQL S...			Unre...	
Email Add	192...	grv...	NA	dbo.GA...	EMAI...	var...	100	Y	NA	sonam	78	1,000	1,000	752	50	Y	N	N	TA...	SQL S...			Unre...	
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	77	2,000	2,000	1,484	101	Y	N	N	TA...	SQL S...			Unre...	
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	76	2,000	2,000	1,477	102	Y	N	N	TA...	SQL S...			Unre...	
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	76	2,000	2,000	1,468	102	Y	N	N	TA...	SQL S...			Unre...	
Email Add	192...	grv...	NA	dbo.ga...	EMAI...	var...	80	Y	NA	sonam	77	2,000	2,000	1,485	103	Y	N	N	TA...	SQL S...			Unre...	

f) Enter the details for **Remediation Action** and **Remediation Value (%)** field in the popup.

Remediation

Scope

Host Name: 192.168.0.151 DB Name: grv_discover

Remediation Action: Select remediation action Remediation Value(%): 1

Table Name: dbo.gaurav_10lakh_1 Field Name: EMAIL_ADDRESS

Sensitive Type: Email Address Enabled

Buttons: Cancel, Save

v. Following are the options in **Remediation Action** field.

- **Mark as Non Sensitive:** Select this option to mark the record as Non Sensitive. The system will ignore this record in future re-run.
- **Mark as Correct:** Select this option to mark the record as Sensitive. The system will not ignore this record in future re-run.
- **Decrease the Confidence Factor:** select this option to decrease the confidence factor in Remediation Value (%) field.
- **Increase the Confidence Factor:** Select this option increase the confidence factor in Remediation Value (%) field.

vi. Enter the remediation value in percentage. This field will be enabled when you select **Decrease** or **Increase the Confidence Factor** in **Remediation Action** field.

Once the remediation value has been set and saved, the same can be seen under Saved Remediation tab.

Remediation							
HostName	Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Acti	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

9.5.2.1.2 Logs

In Logs tab, you can view the list of all the Task Name along with the errors occurred during the task execution. The Logs tab is divided into two panes:

1. Logs List
2. Logs Details

Task Name	Database Name	Directory Path	Table Name	Connection Name	Error
Task1				Detection1Mysql	
DetectionTask				MysqlDetection	Reference data for 'Names' sensit...

Task Instance	Start Time	End Time
42	2020-01-08 06:58:44.0	2020-01-08 06:58:45.0
41	2020-01-08 06:57:27.0	2020-01-08 06:57:44.0

- **Logs List:** This pane will display the list of all the Tasks Name along with the errors occurred during the task execution. It will also display the information such as Database Name, Directory Path, Table Name and Connection Name.

1. **Save Results to File:** Click the Save Results to File button to save the Logs List information in .csv format.
2. **Save Results to PDF:** Click the Save Results to PDF button to save the Logs List information in PDF format.

Task Name	Database Name	Directory Path	Table Name	Connection Name	Error
Task1				Detection1Mysql	
DetectionTask				MysqlDetection	Reference data for 'Names' sensit...

- **Logs Details:** The Logs Details pane will display the information for the selected task in the Logs List. The information includes Task Instance, Start Time and End Time.

Task Instance	Start Time	End Time
42	2020-01-08 06:58:44.0	2020-01-08 06:58:45.0
41	2020-01-08 06:57:27.0	2020-01-08 06:57:44.0

9.5.2.1.3 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

HostName	Scope Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Acti	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.5.2.2 Masking

In Logs tab, you can view the list of all the Task Name along with the errors occurred during the task execution. The Logs tab is divided into two panes:

1. Logs List
2. Logs Details

Task Name	Database Name	Directory Path	Table Name	Connection Name	Error
Task1				Detection1Mysql	
DetectionTask				MysqlDetection	Reference data for 'Names' sensit...

Task Instance	Start Time	End Time
42	2020-01-08 06:58:44.0	2020-01-08 06:58:45.0
41	2020-01-08 06:57:27.0	2020-01-08 06:57:44.0

- **Logs List:** This pane will display the list of all the Tasks Name along with the errors occurred during the task execution. It will also display the information such as Database Name, Directory Path, Table Name and Connection Name.
 1. **Save Results to File:** Click the Save Results to File button to save the Logs List information in .csv format.
 2. **Save Results to PDF:** Click the Save Results to PDF button to save the Logs List information in PDF format.

Task Name	Database Name	Directory Path	Table Name	Connection Name	Error
Task1				Detection1Mysql	
DetectionTask				MysqlDetection	Reference data for 'Names' sensit...

- Logs Details:** The Logs Details pane will display the information for the selected task in the Logs List. The information includes Task Instance, Start Time and End Time.

Task Instance	Start Time	End Time
42	2020-01-08 06:58:44.0	2020-01-08 06:58:45.0
41	2020-01-08 06:57:27.0	2020-01-08 06:57:44.0

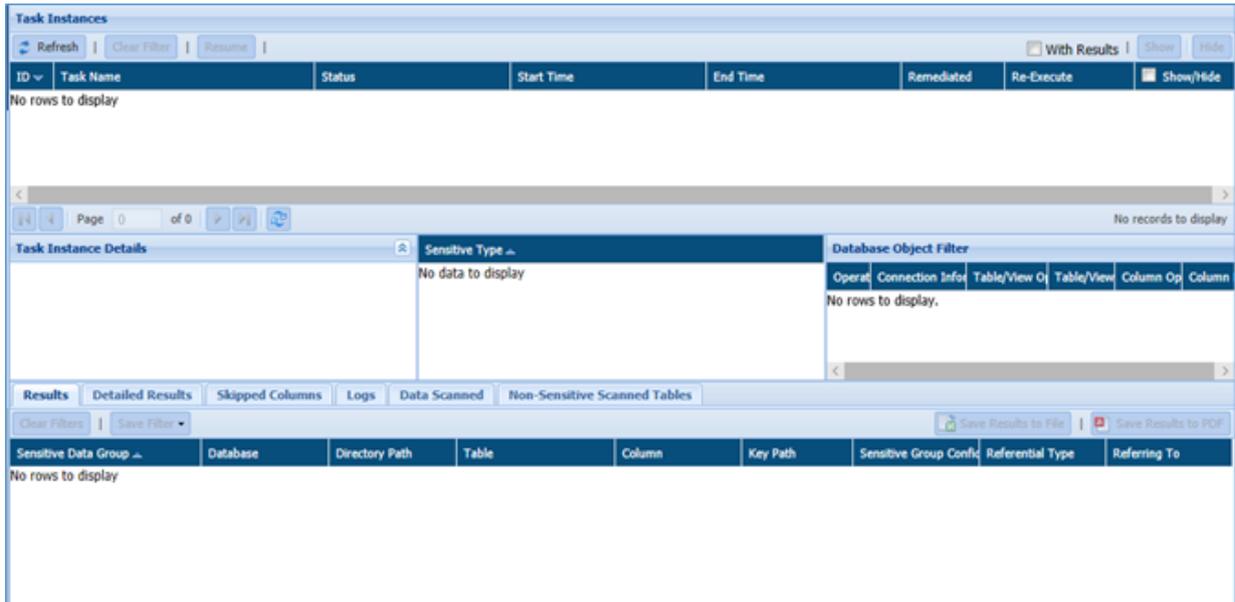
9.6 Azure

9.6.1 Azure Blob/Data Lake

9.6.1.1 By Task

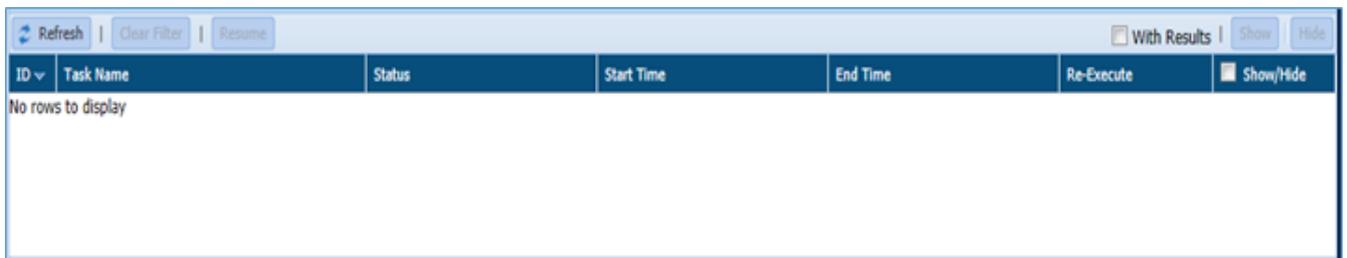
The Results tab is divided in three panes. These are:

1. Task Instances
2. Task Instance Details
3. Overview



4. Task Instances

This pane will display the information for the task such as ID (system generated), Task Name, Status, Start Time, End Time, Show/Hide, etc. Click the Re-Execute button next to the selected task to re-run it. Selecting a task instance displays its parameters and results in the panels below.



Refresh: Click the **Refresh** button. It will update the current page with the updated information.

Clear Filters: Click the **Clear Filters** button. It will remove any applied filters on the Tasks page.

Resume: Click the Resume button to re-start the process from the point where it was paused.

Show: Click the **Show** button to unhide a task.

Hide: To hide any task, follow the below steps:

- i. Check **Show/Hide** checkbox for the policy.
- ii. Click the **Hide** button. The policy will get greyed out.

5. Task Instance Details:

The Task Instance Detail pane will display the basic information for the RDS/RedShift detection task selected in the Task Instance pane. It includes details such as Task Name, Task Description, Created By, Last Executed On, etc.

The Task Instance Details pane will also display the list of Sensitive Types and Database Object Filter used.



6. Sensitive Type: The Sensitive Type pane will displays the list of Sensitive Type for the selected task.

7. Database Object Filter: The database Object Filter pane displays only those databases/tables/columns that match the filter applied. The columns display information such as the applied operator, connection information, Table/View operator, Table/View Filter, column operator and column filter.

8. Overview:

The Overview pane will show the overview of basic information for the selected task. The information displayed is dependent on the currently selected tab.



- **Results:** The result tab will display the high level information of the Sensitive Types for the selected task in the detection task pane. The information includes Sensitive Data Group, Database, Table, Column, Sensitive Group Confidence, etc.
- **Detailed Results:** The Detailed Results tab will provide you the detailed information for the Sensitive Type encountered in the database. The information will include Sensitive Type, Hostname, Database, Table, Column, Database User, etc.

Sensitive Type	Hostnam	Databas	Directory	Table	Column	Data Ty	Data Len	Nullabl	Key Pat	Database	Confid	Row Sc	Values	Hit Co	Null Co	Quick	Maske	Safe	Table	Database	Referenti	Referring
Address Cit...	10.12...	dgstar	NA	tmp_stmt	stmt	VAR...	4000	Y	NA	root	7	99	99	8	0	N	N	N	BA...	MySQL		
Address Cit...	10.12...	sys	NA	sys_config	variable	VAR...	128	N	NA	root	14	6	6	1	0	N	N	N	BA...	MySQL	Primar...	
Address St...	10.12...	dgstar	NA	dg_dbm...	location	VAR...	64	Y	NA	root	84	1	1	1	0	N	N	N	BA...	MySQL		
Address St...	10.12...	dgstar	NA	tmp_stmt	stmt	VAR...	4000	Y	NA	root	10	99	99	12	0	N	N	N	BA...	MySQL		
Address Zi...	10.12...	sys	NA	sys_config	value	VAR...	128	Y	NA	root	14	6	6	1	1	N	N	N	BA...	MySQL		
Date (Best ...	10.12...	dgstar	NA	dg_clou...	updat...	DAT...	19	Y	NA	root	40	0	0	0	0	Y	N	N	BA...	MySQL		
Date (Best ...	10.12...	dgstar	NA	dg_dbm...	updat...	DAT...	19	Y	NA	root	40	0	0	0	0	Y	N	N	BA...	MySQL		

- **Skipped Columns:** This tab will provide you with the list of columns that were skipped during the execution.

Connection Name	Database Name	Directory Path	Table Name	Column Name	Data Type
No rows to display					

- **Logs:** It will list errors that occurred during task execution.

Database Name	Directory Path	Table Name	Connection Name	Error
No rows to display				

- **Data Scanned:** List the total number of rows in the scanned tables and how many of those rows DgSecure actually scanned.

Host Name	Database Name	Directory Path	Table Name	Total Rows	Rows Scanned (Max)	Total Data	Sampled Data
No rows to display.							

- **Non Sensitive Scanned Tables:** It will list down the tables which contain non sensitive data. This tab includes information such as Host Name, Database Name, Table Name, Table Size (bytes).



9.6.1.2 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

Remediation							
Scope		Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Acti	Enabled
HostName	Database Name						
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.6.2 Databases

9.6.2.1 Results

The **Result** screen displays information about the Sensitive data detected in the existing database. This screen is divided into three panels.

4. Task Instances
5. Task Instances Detail
6. Overview

The screenshot shows the 'Task Instances' interface. At the top, there are buttons for 'Refresh', 'Clear Filter', 'Resume', and 'Queue For Remediation'. A table lists three tasks: 'oracle' (ID 44, Completed), 'Task2' (ID 43, Completed), and 'Task1' (ID 42, Failed). Below the table, the 'Task Instance Details' pane for ID 44 is expanded, showing 'Task Name: oracle', 'Task Type: Detection', and 'Start Time: Jan-09-2020 04:23:28'. A 'Sensitive Type' dropdown is set to 'HIPAA_DBMS', and a 'Database Object Filter' pane is empty. At the bottom, a table shows 'Sensitive Data Group' details for various databases and tables, including 'Credit Card' and 'Email Address'.

- **Task Instances**
The Task Instance pane display information for the masking tasks. This pane will display information such as ID (system generated), Task Name, Status, Start Time, End Time, etc.

This screenshot is identical to the one above, showing the 'Task Instances' overview with the task list and the 'Task Instance Details' pane for task ID 44.

9. **Refresh:** Click the **Refresh** button. It will update the current page with the updated information
10. **Clear Filters:** Click the **Clear Filters** button. It will remove any applied filters on the Tasks page.
11. **Resume:** Click the **Resume** button to re-start the task where it was initially stopped.
12. **Show:** Click the **Show** button to unhide a task.
13. **Hide:** To hide any task, follow the below steps:
 - i. Check **Show/Hide** checkbox for the policy.
 - ii. Click the **Hide** button. The policy will get greyed out.
14. **With Results:** Check the **With Results** checkbox, if you want to see the results for the selected masking task.

15. **Queue For Remediation:** Click this button to queue the task for remediation workflow. This option allows you to queue the task for remediation manually.

Once the task has been queued for remediation, the value for the remediation can be seen in the **Task Instances** panel under the column name **REMEDIATED**. It specifies the value as **YES** or **NO**.

ID	Task Name	Status	Start Time	End Time	Remediated	Re-Execute	Show/Hide
183	email_privacy	Completed	May-19-2020 17:35:08	May-19-2020 17:35:29	Yes		
153	Ritish_DGWalker_Oracle	Completed	May-18-2020 10:18:17	May-18-2020 10:18:21	No		
152	Ritish_DGWalker_Oracle_copy_by_parteek...	Completed	May-18-2020 10:17:13	May-18-2020 10:17:27	No		

To queue the task for remediation workflow automatically, set the value of **AUTO QUEUE FOR REMEDIATION** to **YES** in DgAdmin application under **SETTINGS**.

16. **Remediated Results:** check the **Remediated Results** checkbox, if you want to view the final results for the selected fields which are marked as **Remediated**. The remediated result will be displayed in the **Detailed Result** tab under the **Overview** panel.

- **Task Instance Details**

The Task Instance Details pane will display the basic information for the task selected in the Task Instance pane. The includes details such as Task Name, Task Description, Start Time, End Time, Task Type, etc.

It also provides additional information for the Sensitive Type and the Database Object Filter.

- **Overview**

The bottom panel will show the detailed information for the selected task. The information displayed is dependent on the currently selected tab.

Sensitive Data Group	Database	Directory Path	Table	Column	Key Path	Sensitive Group Conf	Referential Type	Referring To
Credit Card	DATAGUISE	NA	MUNISH	CCNO	NA	100		
Credit Card	DATAGUISE	NA	SUPPORTTEST	CREDITCARD	NA	30		
Email Address	DGCONTROLLER_...	NA	DG_DATABASE_ATTRIB...	CONTACT_EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_HADOOP_ATTRIBUT...	CONTACT_EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATIONS_US...	EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATIONS_US...	EMAIL	NA	16		
Email Address	DGCONTROLLER_...	NA	DG_NOTIFICATION_SEN...	EMAILCONTENT	NA	16		

13. **Results:** The Result tab provide information for the Sensitive Type encountered in the database along with the additional information such as Column Name, Table Name, Sensitive Group Confidence.
14. **Detailed Results:** This pane list the Sensitive table columns that are discovered, along with their location, data type, Table Name, Column Name, Data Type, etc.
15. **Skipped Columns:** This pane list down any skipped columns along with the data type identified in that column.
16. **Data Scanned:** The Data Scanned pane list down the sampled and the total data targeted with the scan.
17. **Non Sensitive Scanned Tables:** This pane list down all the Database Name along with their Host Name, Directory Path, Table Name and Table Size which were found non – sensitive during the scan process.
18. **Logs:** This pane list any errors which occurred during the task execution. It displays the log information for the errors such as Database Name, Schema Name, Column Name, Error Description, and Error.
19. **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
20. **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
21. **Remediate:** Click this button to remediate any result in **DETAILED RESULTS** tab. To remediate the selected result, perform the following steps:
 - g) Select the record in the **Detailed Result** tab and click the **Remediate** button.

Sensitive Typ	Hostnam	Databas	Director	Table	Column	Data T	Data Len	Nullabl	Key Pal	Databas	Confid	Row Sc	Values	Hit Co	Null Co	Quick	Masked	Safe	Table	Database	Referenti	Referring	Remed S
Email Add	192	grv	NA	dbo.ga	EMAI	var	80	Y	NA	sonam	82	10	10	8	0	Y	N	N	TA	SQL S			Unre
Email Add	192	grv	NA	dbo.GA	EMAI	var	100	Y	NA	sonam	78	1,000	1,000	752	50	Y	N	N	TA	SQL S			Unre
Email Add	192	grv	NA	dbo.ga	EMAI	var	80	Y	NA	sonam	77	2,000	2,000	1,484	101	Y	N	N	TA	SQL S			Unre
Email Add	192	grv	NA	dbo.ga	EMAI	var	80	Y	NA	sonam	76	2,000	2,000	1,477	102	Y	N	N	TA	SQL S			Unre
Email Add	192	grv	NA	dbo.ga	EMAI	var	80	Y	NA	sonam	76	2,000	2,000	1,468	102	Y	N	N	TA	SQL S			Unre
Email Add	192	grv	NA	dbo.ga	EMAI	var	80	Y	NA	sonam	77	2,000	2,000	1,485	103	Y	N	N	TA	SQL S			Unre

- h) Enter the details for **Remediation Action** and **Remediation Value (%)** field in the popup.

vii. Following are the options in **Remediation Action** field.

- **Mark as Non Sensitive:** Select this option to mark the record as Non Sensitive. The system will ignore this record in future re-run.
- **Mark as Correct:** Select this option to mark the record as Sensitive. The system will not ignore this record in future re-run.
- **Decrease the Confidence Factor:** select this option to decrease the confidence factor in Remediation Value (%) field.
- **Increase the Confidence Factor:** Select this option increase the confidence factor in Remediation Value (%) field.

viii. Enter the remediation value in percentage. This field will be enabled when you select **Decrease** or **Increase the Confidence Factor** in **Remediation Action** field.

Once the remediation value has been set and saved, the same can be seen under Saved Remediation tab.

HostName	Scope Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Act	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

9.6.2.2 Saved Remediation

The Saved Remediation tab allows you to view all the records that have been selected as remediated results in **DETAILED RESULTS** tab under **OVERVIEW** panel.

***Note:** To know more about how to enable and queue the task for remediation workflow, refer section **Settings** under DgAdmin.

Once the remediation value has been set and saved in **DETAILED RESULTS** tab, the same can be seen under **Saved Remediation** tab. The **Saved Remediation** tab will display the information such as Host Name, Database Name under Scope, Table Name, Field Name, Sensitive Type, Remediation Action, etc.

HostName	Scope Database Name	Table Name	Field Name	Sensitive Type	Remediation Action Name	Remediation Act	Enabled
192.168.1.56	RITISH_NEW_GDPR	RITISH_MASK	VARCHAR_COL2	NPI	Decrease Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	EMAIL_ADDRESS	Email Address	Increase Confidence Per...	10	true
153.64.73.16	mandeep_test5	SMOKE_20	ABA_ROUTING_...	ABA Routing number	Mark As Correct		true
153.64.73.16	mandeep_test5	SMOKE_20	URL	URL	Mark As Non Sensitive		true

To edit the remediation details, click **EDIT**. This functionality allows you to update the information for remediated results, if required.

To delete the remediated record, select the record and click **Delete**. This will delete the selected remediated record from the screen.

Click **Refresh** to update the screen.

9.6.2.3 Logs

In Logs tab, you can view the list of all the Task Name along with the errors occurred during the task execution. The Logs tab is divided into two panes:

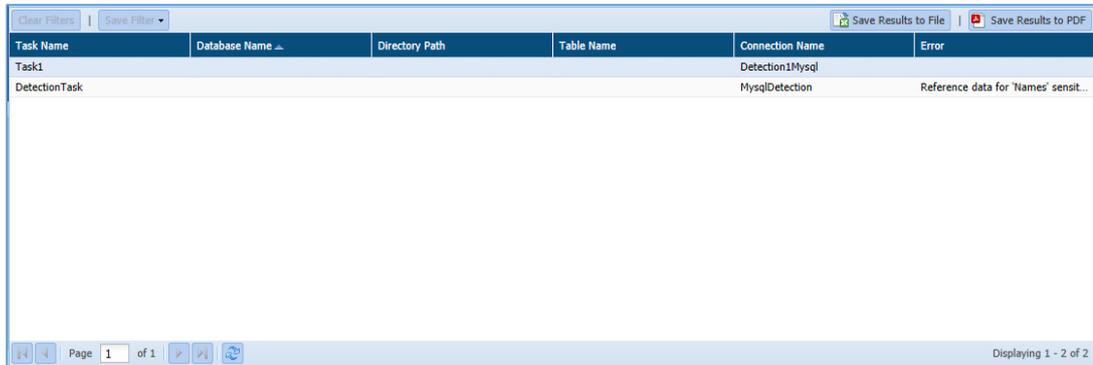
1. Logs List
2. Logs Details

Task Name	Database Name	Directory Path	Table Name	Connection Name	Error
Task1				Detection1Mysql	
DetectionTask				MysqlDetection	Reference data for 'Names' sensit...

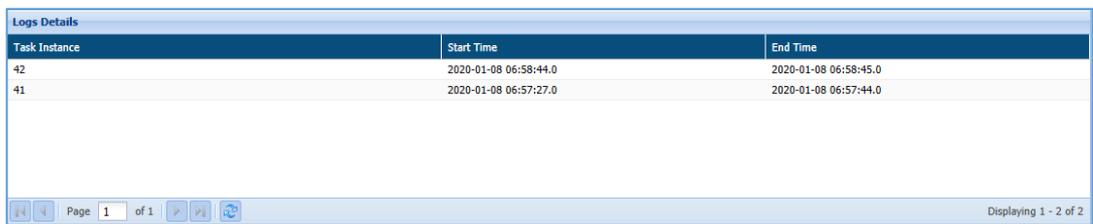
Task Instance	Start Time	End Time
42	2020-01-08 06:58:44.0	2020-01-08 06:58:45.0
41	2020-01-08 06:57:27.0	2020-01-08 06:57:44.0

- **Logs List:** This pane will display the list of all the Tasks Name along with the errors occurred during the task execution. It will also display the information such as Database Name, Directory Path, Table Name and Connection Name.

3. **Save Results to File:** Click the Save Results to File button to save the Logs List information in .csv format.
4. **Save Results to PDF:** Click the Save Results to PDF button to save the Logs List information in PDF format.



- **Logs Details:** The Logs Details pane will display the information for the selected task in the Logs List. The information includes Task Instance, Start Time and End Time.

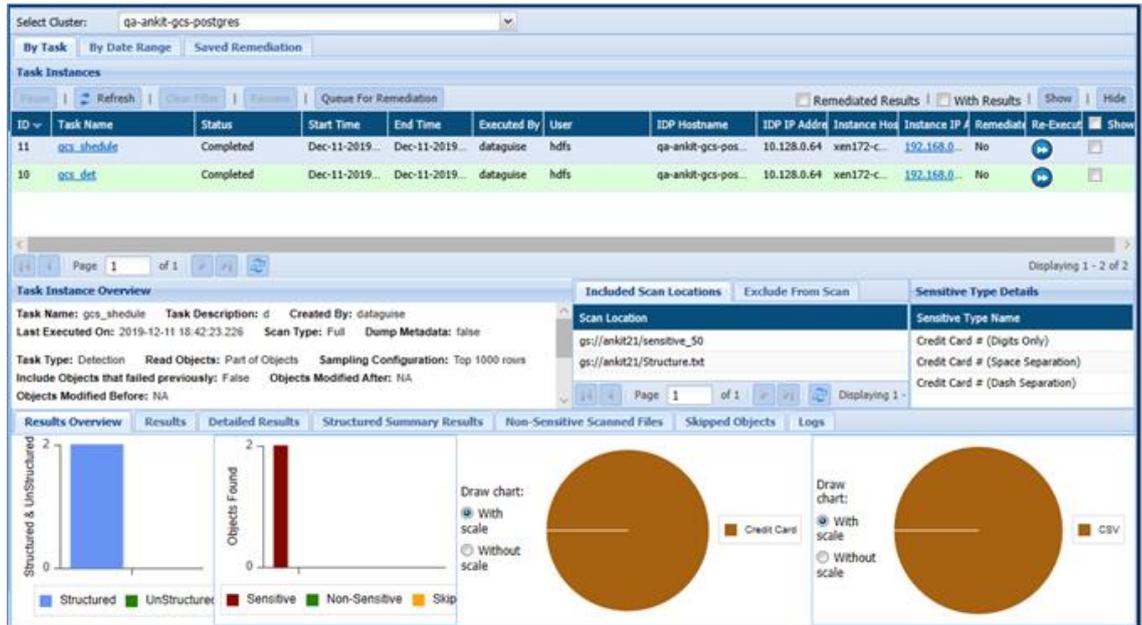


9.7 Google Cloud

The Result page displays the status and the results of the GCS tasks. The information can be tracked:

1. By Task
2. By Date Range
3. Saved Remediation

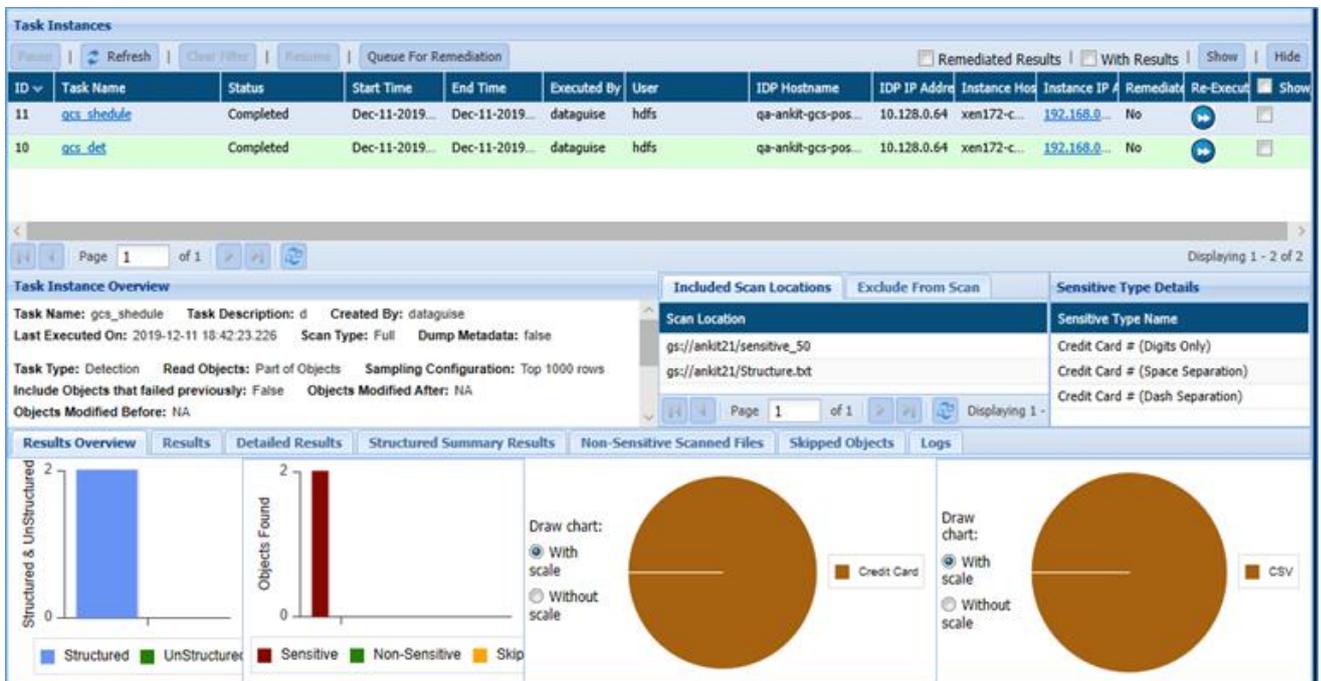
The below image shows the user interface of “By Task” tab when clicked on Result page:



9.7.1 By Task

The 'By Task' tab is divided into three panes. These are:

1. Task Instances
2. Task Instance Overview
3. Overview



- **Task Instances:**

The Task Instances pane display all the task instances along with their ID (system generated), Task Name, Status, Start Time, End Time, Executed By, User, IDP Hostname, IDP IP Address, Instance Hostname, etc. Click the Re-Execute button, if you wish to re-run the task again.

ID	Task Name	Status	Start Time	End Time	Executed By	User	IDP Hostname	IDP IP Address	Instance Hostname	Instance IP Address	Remediate	Re-Execute	Show
11	gcs_schedule	Completed	Dec-11-2019...	Dec-11-2019...	dataguise	hdfs	qa-ankit-gcs-pos...	10.128.0.64	xen172-c...	192.168.0...	No	▶▶	<input type="checkbox"/>
10	gcs_det	Completed	Dec-11-2019...	Dec-11-2019...	dataguise	hdfs	qa-ankit-gcs-pos...	10.128.0.64	xen172-c...	192.168.0...	No	▶▶	<input type="checkbox"/>

1. **Pause:** Click the Pause button to pause the running task for a while.
2. **Refresh:** Click the Refresh button. It will update the current page with the updated information.
3. **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Tasks page.
4. **Resume:** Click the Resume button to re-start the process from the point where it was paused.
5. **Queue For Remediation**
6. **With Results**
7. **Show:** Click the Show button to unhide a task.
8. **Hide:** To hide any task, follow the below steps:
 - i. Check Show/Hide checkbox for the policy.
 - ii. Click the Hide button. The policy will get greyed out.

- **Task Instance Overview:**

The Task Instance Overview pane will display the basic information for the GCS task selected in the Task Instance pane. The includes details such as Task Name, Task Description, Created By, Last Executed On, etc.

The Task Instance Overview pane will also display information for the Scanned Locations in Included Scan Locations and Exclude from Scan tab, Sensitive Type Details.

Scan Location	Sensitive Type Name
gs://ankit21/sensitive_50	Credit Card # (Digits Only)
gs://ankit21/Structure.bt	Credit Card # (Space Separation)
	Credit Card # (Dash Separation)

1. **Included Scan Locations:** It displays the information about the task target directories.

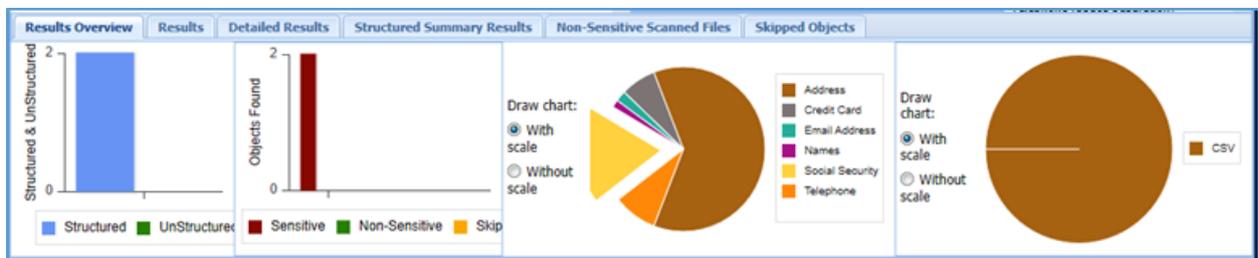
2. **Exclude From Scan:** It will display the list of excluded object extension and the file location.

Included Scan Locations		Exclude From Scan	
Object Extension	Scan Location	Object Extension	Scan Location

3. **Sensitive Type Details:** It display the list of all Sensitive Types discovered in the task.

- **Overview:**

The bottom panel will show the detailed information for the selected task. The information displayed is dependent on the currently selected tab.



4. **Result Overview:** The Result Overview tab gives a graphical summary of the results for the selected task instance. Hovering over the data points in any of the graphs shows the number of hits and hit percentage for a file or an expression.
5. **Results:** The Results tab displays sensitive data groups, hit count, task name, Object Path, Object Type, Detection Type and Content Read.

Task Name	Sensitive Data Group	Object Path	Object Type	Detection Type	Content Read	Hit Count
DiscoveryGCS	Address	gr://lg-data/ADData.csv	CSV	Structured	Scan(100%)	26
DiscoveryGCS	Address	gr://lg-data/normaSchema.csv	CSV	Structured	Scan(100%)	266
DiscoveryGCS	Credit Card	gr://lg-data/ADData.csv	CSV	Structured	Scan(100%)	30
DiscoveryGCS	Credit Card	gr://lg-data/normaSchema.csv	CSV	Structured	Scan(100%)	3
DiscoveryGCS	Email Address	gr://lg-data/ADData.csv	CSV	Structured	Scan(100%)	10
DiscoveryGCS	Names	gr://lg-data/normaSchema.csv	CSV	Structured	Scan(100%)	7
DiscoveryGCS	Social Security	gr://lg-data/ADData.csv	CSV	Structured	Scan(100%)	20
DiscoveryGCS	Social Security	gr://lg-data/normaSchema.csv	CSV	Structured	Scan(100%)	70
DiscoveryGCS	Telephone	gr://lg-data/ADData.csv	CSV	Structured	Scan(100%)	40

6. **Detailed Results:** The Detailed Results tab will provide you the detailed information for the Sensitive Type discovered in the database. The information will include Sensitive Data Type, Task Name, Object Path, Object Type, Detection Type, Content Read and Hit Count.

Task Name	Sensitive Data Type	Object Path	Object Type	Detection Type	Content Read	Hit Count
DiscoveryGCS	Address City (Best suited for structu...	gs://dg-data/normalSchema.csv	CSV	Structured	Scan(100%)	71
DiscoveryGCS	Address City (Best suited for structu...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	16
DiscoveryGCS	Address Line (Best suited for structu...	gs://dg-data/normalSchema.csv	CSV	Structured	Scan(100%)	154
DiscoveryGCS	Address Zip (Best suited for structur...	gs://dg-data/normalSchema.csv	CSV	Structured	Scan(100%)	41
DiscoveryGCS	Credit Card # (Dash Separation)	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10
DiscoveryGCS	Credit Card # (Dinits Only)	nc://do-data/AllData.csv	CSV	Structured	Scan(100%)	10

7. Structured Summary Results:

Sensitive Type	File Path	Field No.	Confidence	Match Count	Rows Scanned	Sample Mode	Null Count	Null Ratio	Field Name Match
Credit Card # (Di...	gs://ankit21/Struc...	Column 2	84	50	50	Sample: 0,r,1000;	0	0	false
Credit Card # (Sp...	gs://ankit21/Struc...	Column 3	84	50	50	Sample: 0,r,1000;	0	0	false
Credit Card # (Da...	gs://ankit21/Struc...	Column 4	84	50	50	Sample: 0,r,1000;	0	0	false
Credit Card # (Di...	gs://ankit21/sensi...	Column 2	57	50	74	Sample: 0,r,1000;	24	32	false
Credit Card # (Sp...	gs://ankit21/sensi...	Column 3	57	50	74	Sample: 0,r,1000;	24	32	false
Credit Card # (Da...	nc://ankit21/sensi...	Column 4	57	50	74	Sample: 0,r,1000;	24	32	false

- Remediate

8. **Skipped Objects:** This tab will provide you with the list of objects that were skipped during the execution. It displays information about the Object Path, Skipped Reason, Object Type, Object Owner and Notification Date.

Object Path	Skipped Reason	Object Type	Object Owner	Modification Date
No file skipped.				

9. Non-Sensitive Scanned Files:

File Path	File Type	File Size	Object Owner	Modification Date
No non-sensitive scanned files.				

10. **Logs:** It will list errors that occurred during task execution.

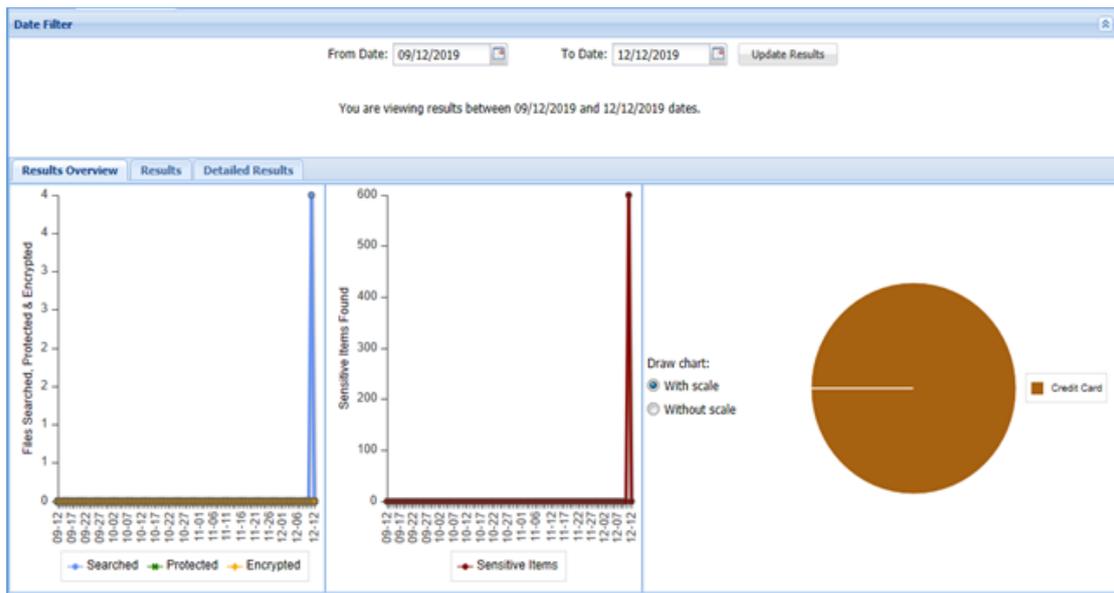
Error	Object Name	Directory Path
No log to show.		

- **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
- **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.

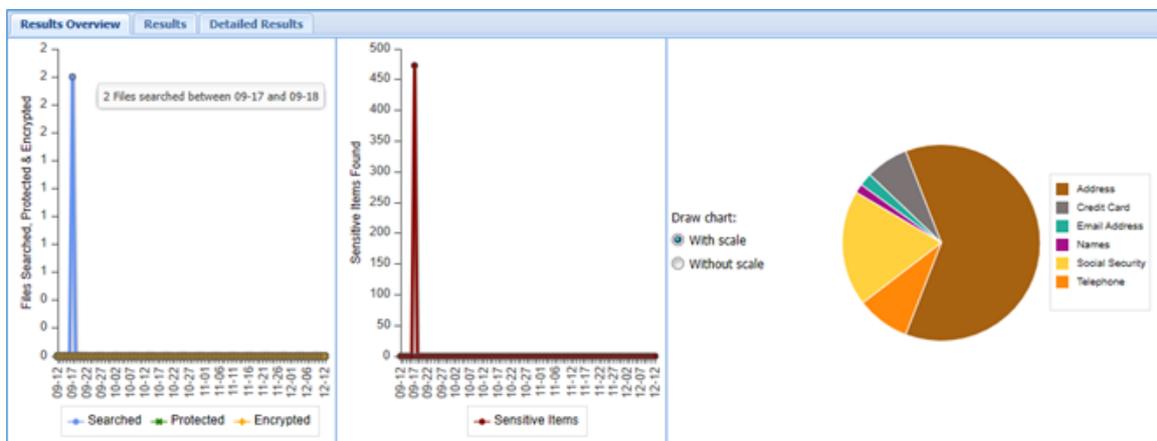
9.7.2 By Date Range

The 'By Date Range' tab is divided into two panes. These are:

1. Date Filter
2. Overview



- **Date Filter:** In Date Filter pane, you can specify the date range.
- **Overview:** The Overview pane display the graphical representation of the information in Results Overview tab and detailed information in Results and Detailed Results tab.



3. **Results Overview:** The Results Overview tab display information in form of graphical representation. It displays two bar chart and a pie chart.

- The first bar chart shows the number of files searched, encrypted and masked.
- The second bar chart shows the total number of sensitive files found.
- The pie chart shows the Sensitive Type found in each segment.

4. **Results:** The Results tab lists down the Sensitive Data Group detected in the specified time range. It includes information such as Task Name, Sensitive Data Group, Object Path, Object Type, Detection Type, Content Read, Hit Count, AES/Masked, FP/Row Encrypted, FP/Row Decrypted.

Task Name	Sensitive Data Group	Object Path	Object Type	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted
DiscoveryGCS	Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	26	false	false	false
DiscoveryGCS	Address	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	266	false	false	false
DiscoveryGCS	Credit Card	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	30	false	false	false
DiscoveryGCS	Credit Card	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	3	false	false	false
DiscoveryGCS	Email Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Names	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	7	false	false	false
DiscoveryGCS	Social Security	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	20	false	false	false
DiscoveryGCS	Social Security	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	70	false	false	false
DiscoveryGCS	Telephone	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	40	false	false	false
DiscoveryGCS	Telephone	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	1	false	false	false

5. **Detailed Results:** The Detailed Results tab lists down the Sensitive Types detected in the specified time range, grouped by type and location. The tab also displays information such as Task Name, Sensitive Data Type, Object Path, Object Type, Detection Type, Content Read, Hit Count, and whether the data was masked.

Task Name	Sensitive Data Type	Object Path	Object Type	Detection Type	Content Read	Hit Count	AES/Masked	FP/Row Encrypted	FP/Row Decrypted
DiscoveryGCS	Address Line (Best suited ...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	154	false	false	false
DiscoveryGCS	US Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Address City (Best suited ...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	16	false	false	false
DiscoveryGCS	Address City (Best suited ...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	71	false	false	false
DiscoveryGCS	Address Zip (Best suited f...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	41	false	false	false
DiscoveryGCS	Credit Card # (Space Sep...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	3	false	false	false
DiscoveryGCS	Credit Card # (Digits Only)	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Credit Card # (Space Sep...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Credit Card # (Dash Sepa...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Email Address	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Full Names	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	7	false	false	false
DiscoveryGCS	Social Security # (Space ...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Social Security # (Space ...	gs://dg-data/normalSche...	CSV	Structured	Scan(100%)	70	false	false	false
DiscoveryGCS	Social Security # (Dash S...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false
DiscoveryGCS	Telephone (Space Separa...	gs://dg-data/AllData.csv	CSV	Structured	Scan(100%)	10	false	false	false

- **Save Results to File:** Click the Save Results to File, if you wish to download the data in text format.
- **Save Results to PDF:** Click the Save Results to PDF, if you wish to download the data in PDF format.
- **Clear Filters:** Click the Clear Filters button. It will remove any applied filters on the Detection task page.

9.8 SharePoint

9.8.1 By Task

The **Results** page displays information about the databases discovered with the sensitive data.

Access the **Results** page from the menu under **SHAREPOINT > Results**.

The screen is divided into six panes that are described below.

1. Task Instance:

This panel displays the list of tasks created for the IDP along with the status, start and end time, executor name, user, IP address etc. You can pause, resume and cancel the task.

To re-execute a task, click button.

2. Task Instance Details:

This panel displays all the information about the executed task. To view task instance details, select the task from the **Task Instance** panel.

Task Instance Details

Task Name: Task_01 Task Description: hsAGs
 Task Type: Detect and Notify Created By: d
 Last Executed On: May-29-2020 17:53:21 Job Id: 2029175321

Version to scan: all

File Types: txt,doc,docx,xls,xlsx,ppt,pptx,pdf,esv,zip,lists

6. Included Scan Sites:

This panel displays the locations that are scanned.

Included Scan Sites

Scan Location
http://192.168.1.136:1177/sites/sanjay/Root Site
http://192.168.1.136:1177/sites/SPDev/DevSubsite
http://192.168.1.136:1177/sites/SPDev/Root Site

Page 1 of 1 Displ

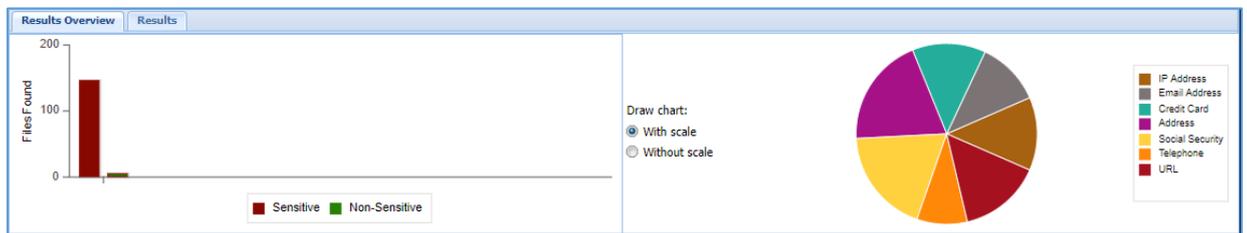
7. Sensitive Type Details:

This panel displays the details about the sensitive data type, discovered.

Sensitive Type Name	Threshold	Action	Notify To
Address Line (Best suited ...	4	Notify	Rajni.chaurasia@dataguis...
Credit Card # (Dash Sepa...	4	Notify	Rajni.chaurasia@dataguis...
Credit Card # (Digits Only)	4	Notify	Rajni.chaurasia@dataguis...
Credit Card # (Space Sep...	4	Notify	Rajni.chaurasia@dataguis...

8. Results Overview:

This panel displays the graphical summary of the scan result for the selected task.



- a) The bar chart on the left shows the number of different types of record. The x-axis represents the type of data. The y-axis represents the number of files with different data types.

9. The pie chart displays the ratio of different sensitive record types for that database.

10. Results:

This panel displays the details of files with sensitive data and the data type. Click **Save Results to File** or **Save Results to PDF** in the top right corner to export the details.

To download the reports in CSV format, click **Save Results to File**. To download the reports in PDF format, click **Save Results to PDF**.

The screenshot shows the 'Results' tab in the Dataguise interface. It features a table with columns for Task Name, File Path, File Type, Sensitive Data Type, File Version, and Hit Counts. The table contains five rows of data. In the top right corner, there are buttons for 'Clear Filters', 'Save Results to File', and 'Save Results to PDF'. At the bottom, there is a pagination control showing 'Page 1 of 69' and a status bar indicating 'Displaying 1 - 20 of 1368'.

Task Name	File Path	File Type	Sensitive Data Type	File Version	Hit Counts
Task_01	http://192.168.1.136:1177/sites/SPDev/IncRoot/New.txt	TXT	Email Address	1	2
Task_01	http://192.168.1.136:1177/sites/SPDev/Shared Documents/SpDev (4).txt	TXT	URL	1	2
Task_01	http://192.168.1.136:1177/sites/SPDev/IncRoot/New.txt	TXT	IP Address	1	2
Task_01	http://192.168.1.136:1177/sites/SPDev/Shared Documents/SpDev (4).txt	TXT	Email Address	1	2
Task_01	http://192.168.1.136:1177/sites/SPDev/Shared Documents/SpDev (3).docx	DOCK	Telephone (Digits Only)	1	11

10 Reports

10.1 Overview

The **Overview page** provides a complete picture of the organization's sensitive data. It displays detailed information about the sensitive data which is **Exposed**, **Protected** (encrypted or masked), **Safe** or **Skipped**. It also shows the information in graphical representation stating the total data processed based on the type of file i.e. structured or unstructured. It also gives a clear picture of the total size of the sensitive data processed, based on the data source type in each policy.

1. **Exposed:** The sensitive data which is not protected during the processing is marked as Exposed.
2. **Protected:** The sensitive data which are encrypted or masked using protection option such as AES Encryption, FPE Masking, etc., while processing is categorized as **Protected**.
3. **Safe:** The sensitive data which is marked safe in Safe-list or the data is already protected is marked as Safe.
4. **Skipped:** The data records which were skipped during the processing is marked as Skipped.

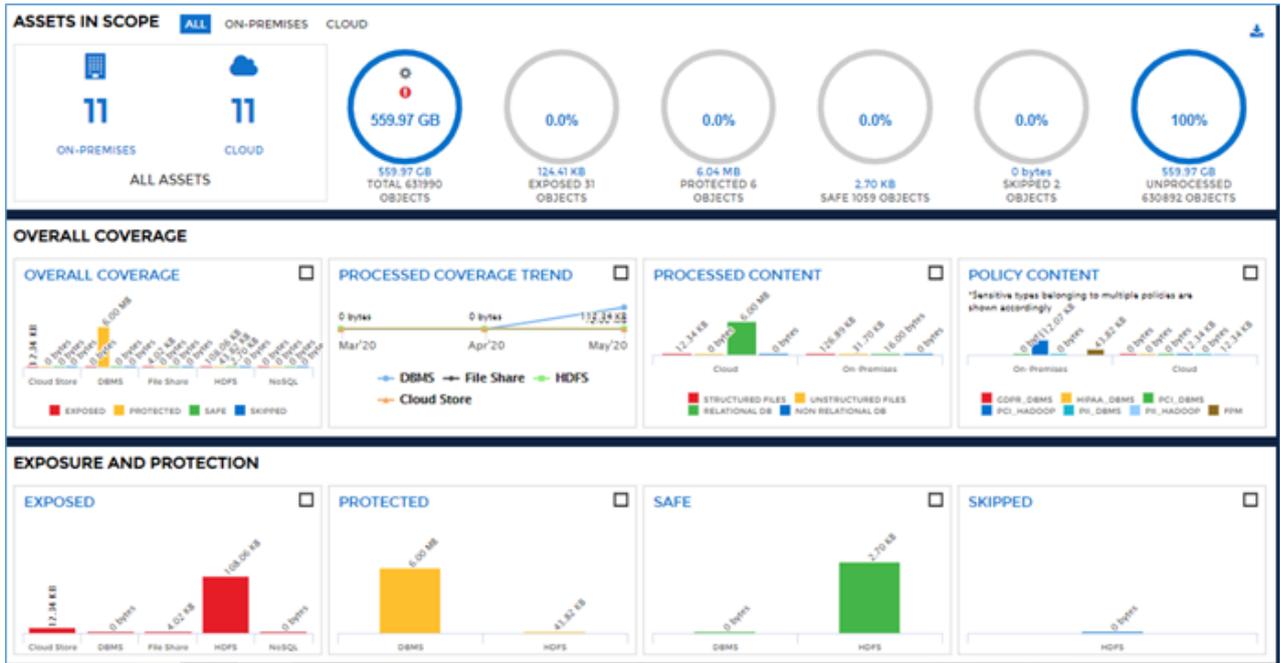
***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more visit section **6.1.5. Manage Roles and Permissions**

To access **Overview** page, click **Reports > Overview**.

The page is divided into three panels.

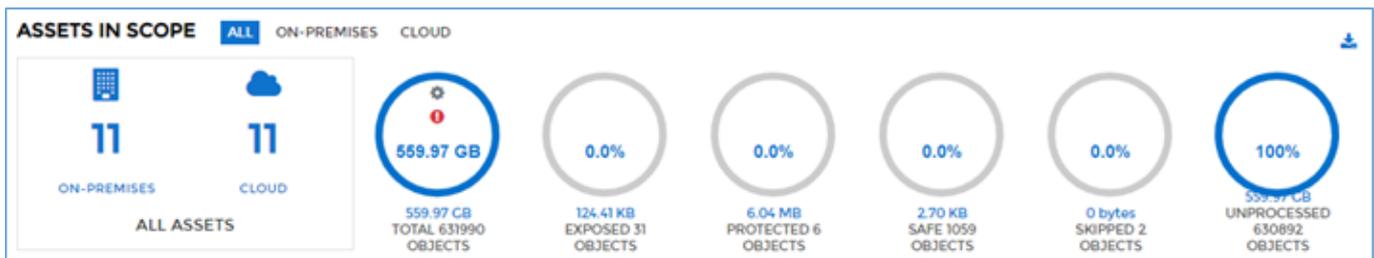
1. ASSETS IN SCOPE
 128. OVERALL COVERAGE
 129. EXPOSTURE AND PROTECTION



The graphics in each row, provide insight of the total objects in all data sources e.g. DBMS, HDFS, S3, Hadoop, etc. You can segregate the result for 'All', 'On-Premises' or 'Cloud' data in the charts, based on the selection made at the top of the screen. Click the  button to download the **Overview** screen in the PDF format.

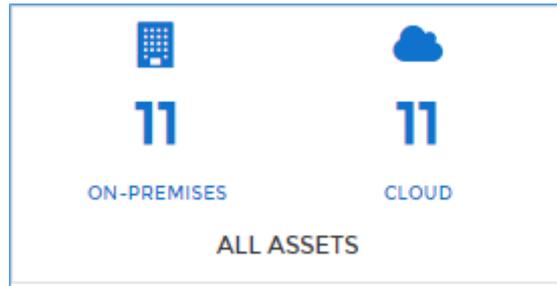
Data assets are termed as data sources or data stores and Data objects are referred as tables/files.

Assets in Scope

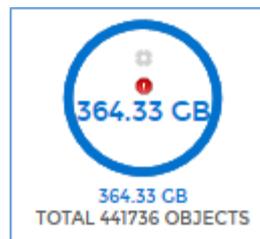


This section provides a complete overview of the data assets. There are various charts that provide information for the sensitive data stored in different tables/files. Click on each chart to view the total size and count of tables/files stored in each data asset.

1. The first chart **All Assets** indicates the total number of data assets present on the **On-Premises** and on the **Cloud** environment.



- The second chart provides the total size and count of all the objects discovered across all the data sources.



To view the size and count of the total data objects in each data source, click the chart and it will display the total size and count of the tables/files for all the data source.

***Note** – If any object has  in the last column, it means that due to some unexpected reasons such as revoking the permission from the user for accessing the database or unexpected shutdown of the system, etc., you are not able to access it.

TOTAL OBJECTS			
NAME	SIZE	COUNT	
Cloud Store	0	0	
DBMS	164.54 CB	102248	
File Share	122.94 CB	473444	
HDFS	272.50 CB	56298	

[CLOSE](#)

To view the in-depth details for each data source, click the data source name and the **Server/Cluster Details** screen will appear.

OVERVIEW: TOTAL > DBMS

SERVERS

REFRESH SHOW DETAILS

NAME	Size	Count
<input checked="" type="checkbox"/> AuroraDBMySQL-QA-SANKUSH-AURORA-MYSQL-INSTANCE-1.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM	9.01 MB	198
<input checked="" type="checkbox"/> AuroraDBPostgres-QASANKUSH-ARORAPOSTGRES-INSTANCE-1.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM		
<input type="checkbox"/> DB2-192.168.0.163		
<input type="checkbox"/> MariaDB-QA-NIHAR-MARIADB.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM		
<input type="checkbox"/> MySQL-SANKUSH-MYSQLMYSQLDATABASE.AZURE.COM		
<input type="checkbox"/> Oracle-192.168.0.163		
<input type="checkbox"/> Oracle-QA-NIHAR-ORACLE12C.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM		
<input type="checkbox"/> Redshift-MANINDERDETECTION.CA0HM05RTETE.US-EAST-1.REDSHIFT.AMAZONAWS.COM		
<input type="checkbox"/> Salesforce-N.A.		
<input type="checkbox"/> SAP S4 HANA-18.185.106.68		
<input type="checkbox"/> Snowflake-https://dataguisse_partner.east-us-2.azure.snowflakecomputing.com		
<input type="checkbox"/> SQL Server-192.168.0.151		

SERVER DETAILS

EXPORT AS PDF

Size 9.01 MB Count 198

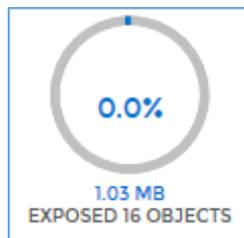
To view the the total size and count of Sensitive Type discovered for the selected server/cluster name, check the checkbox next to the server/cluster name. Click **Show Details** button. This will display the size and count of sensitive type in the **Server/Cluster Details** panel. To update the **Servers/Clusters** panel with all the connections name, click **Refresh** button.

To view reports for DBMS data source, click the **Comprehensive Reports** button. This will display information in form of graphs for various dimensions such as Sensitive Types, Policies, DB Types, etc.

To know more information for **Comprehensive Reports**, visit [Comprehensive Reports](#).



130. The third chart indicates the percentage of the exposed objects that contain sensitive type group. It also displays the total size and the count of the exposed objects.



To view the total size and count of the objects for each data source, click the chart and it will display the total size and count of objects containing the exposed data for each data source.

EXPOSED		
NAME ↑	SIZE	COUNT
Cloud Store	12.34 KB	2
DBMS	0	21
File Share	4.02 KB	1
HDFS	108.06 KB	7
NoSQL	0	0

CLOSE

To view the in-depth details for each data source, click the data source name and the

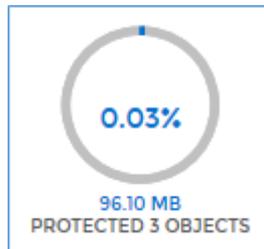
Details screen will appear.

OVERVIEW EXPOSED> HDFS			
CLUSTERS		CLUSTER DETAILS	
<input checked="" type="checkbox"/>	NAME		
<input checked="" type="checkbox"/>	hdfs_citr		
	Name	Exposed 7 file(s)	Sensitive Type Group
	hdfs_citr	7 file(s)	Size 108.06 KB
	/karman	7 file(s)	fpm_PCI_Hadoop 108.06 KB
	/karman/avro	2 file(s)	PCI_Hadoop 8.07 KB
	/karman/avro/All_expressions1.avro	Y	Credit Card 4.02 KB
	/karman/avro/hsp_all_expression.avro	Y	Credit Card 4.05 KB
	/karman/Data_Files_100	5 file(s)	PCI_Hadoop 99.99 KB

This screen displays the information for the exposed objects along with the sensitive type group in each object for the selected data source.

To download the report in PDF format, visit [Export as PDF](#).

- The fourth chart indicates the percentage of the protected objects. These objects contain any sensitive type group and is marked as protected using any of the protection options such as FPE Masking, AES Encryption, Random Masking, etc. This chart also displays the total size and the count of objects which are marked as protected.



To view the size and count of protected objects for each data source, click the chart. This screen will display the size and count of objects for each data source.

PROTECTED		
NAME	SIZE	COUNT
DBMS	6.00 MB	4
HDFS	43.82 KB	2

CLOSE

To view the in-depth details for each data source, click the data source name and the **Details** screen will appear.

OVERVIEW: PROTECTED > DBMS

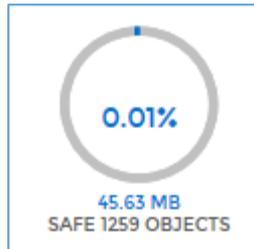
SERVERS		COMPREHENSIVE REPORTS	SERVER DETAILS	
<input checked="" type="checkbox"/>	Oracle-QA-NIHAR-ORACLE12C.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM		Oracle-QA-NIHAR-ORACLE12C.CQGRVY7ZDRME...	Protected 2 table(s) 1 table(s)
<input checked="" type="checkbox"/>	Redshift-MANINDERDETECTION.CA0HM05RTE7E.US-EAST-1.REDSHIFT.AMAZONAWS.COM		Redshift-MANINDERDETECTION.CA0HM05RTE7E...	Protected 2 table(s) 1 table(s)
<input type="checkbox"/>	SQL Server-192.168.0.151		nh_db7238	Protected 2 table(s) 1 table(s)
<input type="checkbox"/>	SQL Server-SANKUS-SQL.DATABASE.WINDOWS.NET		schema_a.table_a1	Protected 2 table(s) Y

Size 6.00 MB

This screen displays the information for the protected objects marked as 'Y' along with the size of each object for the selected data source.

To download the report in PDF format, visit [Export as PDF](#).

132. The fifth chart indicates the percentage of the safe objects within the data sources. It also displays the count and size of all the objects that are marked as safe.



To view the size and count of the safe objects for each data source, click the chart and the screen will display the total size and count for all objects in each data source in which sensitive type group is marked safe.

SAFE		
NAME	SIZE	COUNT
DBMS	0	1058
HDFS	2.70 KB	1

CLOSE

To view the complete details for each data source, click on the data source name and the **Details** screen will appear.

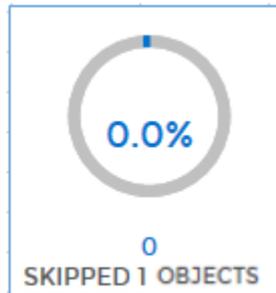
OVERVIEW SAFE> DBMS

SERVERS		COMPREHENSIVE REPORTS	SERVER DETAILS	
<input checked="" type="checkbox"/>	MariaDB-QA-NIHAR-MARIADB.CQCRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM		MariaDB-QA-NIHAR-MARIADB.CQCRVY7ZDRME.U...	2 table(s)
<input checked="" type="checkbox"/>	Oracle-192.168.0.163		nh_7242_user1	2 tables(s)
<input type="checkbox"/>	Oracle-QA-NIHAR-ORACLE12C.CQCRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM		DC_RECREATE_INDEX	Y
<input type="checkbox"/>	Redshift-MANINDERDETECTION.CA0HM05RTE7E.US-EAST-1.REDSHIFT.AMAZONAWS.COM		DC_RECREATE_TRIGGER	Y
<input type="checkbox"/>	Salesforce-N.A.		Oracle-192.168.0.163	4 table(s)
<input type="checkbox"/>	SAP S4 HANA-18.185.106.68		ANJALI	4 tables(s)
<input type="checkbox"/>	Snowflake-https://dataguisse_partner.east-us-2.azure.snowflakecomputing.com		DCS_1000	Y
<input type="checkbox"/>	SQL Server-SANKUS-SQL.DATABASE.WINDOWS.NET		DC_COLUMN_LOCKS	Y
			DC_TABLE_LOCKS	Y
			TEST	Y

This screen displays the information for the objects which are marked as Safe as 'Y' in Safe column along with the size of each table.

To download the report in PDF format, visit [Export as PDF](#).

- The sixth chart indicates the percentage of the skipped objects for each data source. This chart also displays the total size and count of the objects which are skipped.



Click the chart to view the size and count of objects skipped in each data source. The screen will display the size and count for each data source in which objects were skipped.

SKIPPED		
NAME	SIZE	COUNT
HDFS	0	2

CLOSE

To view the complete details for each data source, click the data source name and **Details** screen will appear.

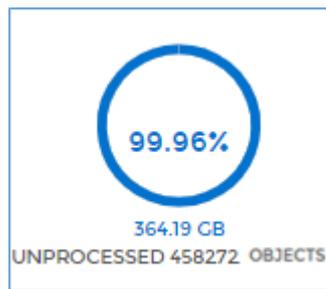
OVERVIEW: SKIPPED > HDFS

CLUSTERS		CLUSTER DETAILS	
<input checked="" type="checkbox"/>	NAME	Name	Skipped 2 file(s)
<input checked="" type="checkbox"/>	hdfs_citr	hdfs_citr	2 file(s)
		/karman	2 file(s)
		/karman/avro	2 file(s)
		/karman/avro/empty_example.avro	Y
		/karman/avro/empty_example2.avro	Y
			Size 0

The **Details** screen will display information of the all skipped objects along with the size for the selected data source.

To download the report in PDF format, visit [Export as PDF](#).

- 134. The seventh chart indicates the percentage of the unprocessed objects. This chart also displays the total size and the count of the objects which are marked as unprocessed.



Click the chart to view the details for each data source. This screen will display the total size and count of unprocessed objects for each data source.

UNPROCESSED OBJECTS		
NAME ↑	SIZE	COUNT
DBMS	176.40 GB	101331
File Share	122.94 GB	468480
HDFS	272.49 GB	56719

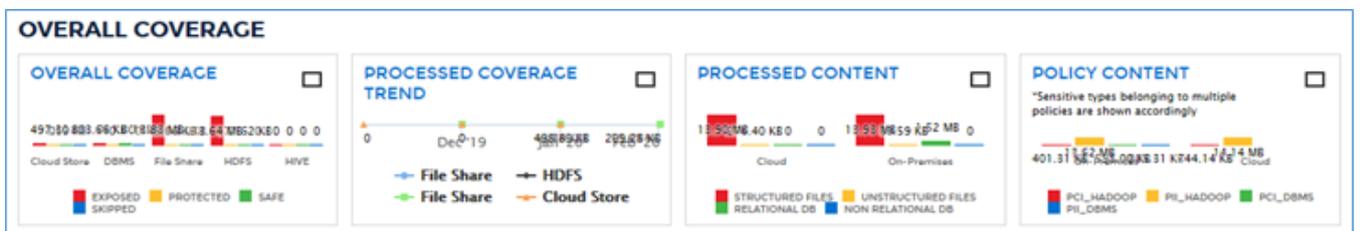
CLOSE

To view the in-depth details for each data source, click the data source name and the **Server Details** screen will appear.

OVERVIEW: UNPROCESSED> HDFS		
CLUSTERS	CLUSTER DETAILS	
<input type="checkbox"/> REFRESH <input type="checkbox"/> SHOW DETAILS <input type="button" value="EXPORT AS PDF"/>		
<input checked="" type="checkbox"/>	NAME	Size 272.49 CB
<input checked="" type="checkbox"/>	hdfs_ctr	Count 56719

To view the the total size and count of Sensitive Type discovered for the selected server/cluster name, check the checkbox next to the server/cluster name. Click **Show Details** button. This will display the size and count of sensitive type in the **Server/Cluster Details** panel.

Overall Coverage

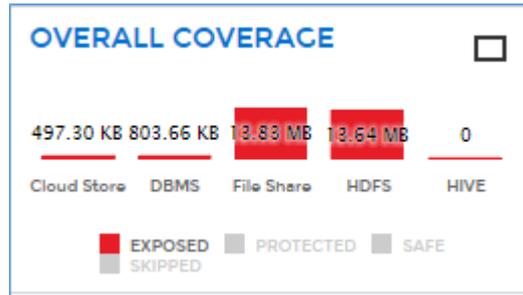


The **Overall Coverage** section displays the information in four different charts. This section offers an insight of the sensitive data based on the different dimension such as Content Type, Policy, Coverage Trend based on the data sources and Overall Coverage graph showing whether data in the objects were Exposed, Protected, Safe or Skipped.

Click on the box provided on the top right corner of each graph, to view it in the full screen as well as to download the graph in PDF format.

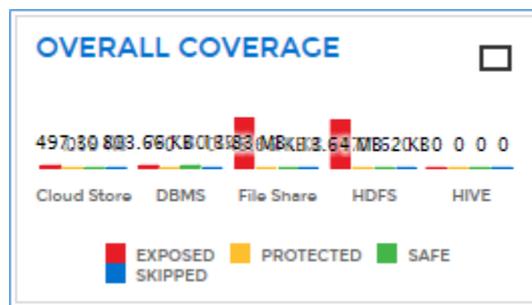


Click on the chart legend mentioned on the bottom of the chart, to view data for specific dimension.



For example, in the above screenshot the you can view data for Exposed chart legend. Since Protected, Safe and Skipped has been hidden.

1. The **Overall Coverage** graph display size of the objects in each data source which are Exposed, Protected, Safe and Skipped.



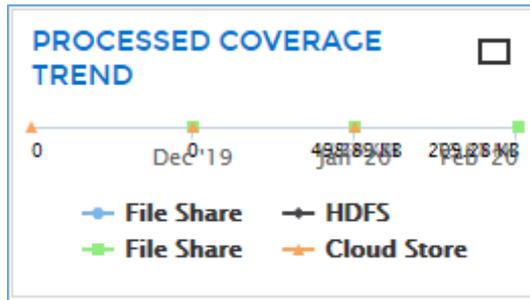
To view the complete details of all the tables/files that are marked as Safe, Protected, Exposed or Skipped, click the chart and the **Details** screen will appear.

SERVERS		SERVER DETAILS			
NAME	Name	Exposed	Protected	Safe	Size
AuroraDBMySQL-QA-SANKUSH-AURORA-MYSQL-INSTANCE-1.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM	AuroraDBMySQL-QA-SANKUSH-AURORA-MYSQL-1...	4 table(s)	0 table(s)	0 table(s)	0
AuroraDBPostgres-QASANKUSH-ARORAPOSTGRES-INSTANCE-1.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM	nh_data	4 tables(s)	0 tables(s)	0 tables(s)	0
MariaDB-QA-NIHAR-MARIA DB.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM	nh_detect1	Y	N	N	0
MySQL-SANKUSH-MYSQL.MYSQL.DATABASE.AZURE.COM	ABA_ROUTING_NUMBER				
Oracle-192.168.0.163	ADDRESS				
Oracle-QA-NIHAR-ORACLE12C.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM	ADDRESS_COUNTRY				
Redshift-MANINDERDETECTION_CA0HM05RTE7E.US-EAST-1.REDSHIFT.AMAZONAWS.COM	ADDRESS_COUNTRY				
Salesforce-N.A.	ADDRESS_COUNTRY				
SAP S4 HANA-18.185.106.68	address_un				
Snowflake-https://datagui.se_partner.east-us-2.azure.snowflakecomputing.com	address_un				
SQL Server-192.168.0.151	ADDRESS_ZIP				
SQL Server-SANKUS-SQL.DATABASE.WINDOWS.NET	CARDHOLDER_NAME				
	CARDHOLDER_NAME				
	CCNO				
	CCNO				
	CITY				
	CITY				
	CUSTOM				

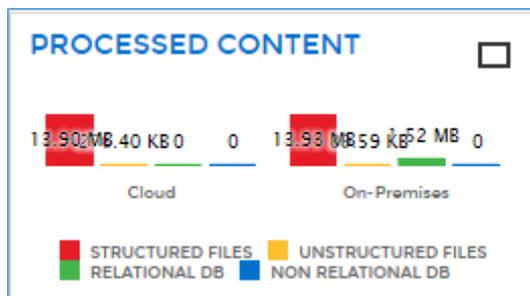
This screen will display the name of all tables/files, size of table/files and the information in form of 'Y' (yes) or 'N' (No) whether the file was marked as Safe, Exposed, Protected or Skipped.

To download the report in the PDF format, visit [Export as PDF](#).

- 135. The Processed Coverage Trend graph displays the information about the data processed in the specific time i.e. month wise. This chart displays the total size of the tables or files and the month on which the processing was done.



- 136. The Processed Content graph shows the breakdown of the data processed based on the schema. This chart categorizes the data based on the schema i.e. structured, unstructured files and relational, non-relational databases.



To view the size and count of the processed data based on the schema, click the chart. It will display the information for the processed content based on the selection.

PROCESSED CONTENT		
NAME	SIZE	COUNT
DBMS	16.00 bytes	1037

[CLOSE](#)

E.g. The above screenshot displays the information of the processed content for structured files in Cloud Store and File Share when clicked on the structured files bar chart.

To view the complete details for each data source, click the data source name. It will display

the **Details** screen.

OVERVIEW: PROCESSED CONTENT > DBMS

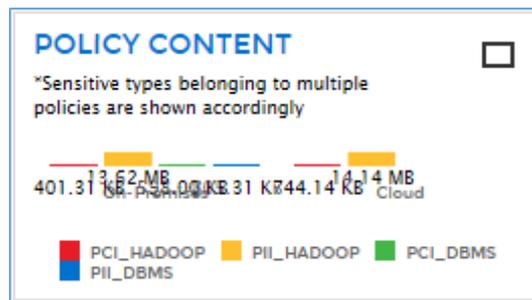
SERVERS		COMPREHENSIVE REPORTS	SERVER DETAILS	
<input checked="" type="checkbox"/>	NAME	<input type="checkbox"/>		
<input checked="" type="checkbox"/>	SQL Server-192.168.0.151	<input type="checkbox"/>	SQL Server-192.168.0.151	relational db 16.00 bytes
<input checked="" type="checkbox"/>	Oracle-192.168.0.163	<input type="checkbox"/>	ccno_data	relational db 16.00 bytes
<input checked="" type="checkbox"/>	SAP S4 HANA-18.185.106.68	<input type="checkbox"/>	dbo.ccno	relational db 16.00 bytes
<input checked="" type="checkbox"/>	Salesforce-N.A.	<input type="checkbox"/>	Oracle-192.168.0.163	relational db 0
			ANJALI	relational db 0
			DCS_1000	relational db 0
			DC_COLUMN_LOCKS	relational db 0
			DC_TABLE_LOCKS	relational db 0
			TEST	relational db 0
			SAP S4 HANA-18.185.106.68	relational db 0
			Salesforce-N.A.	relational db 0

This screen will display information for the selected data source along with the Content Type and Size of the file/table. The **Comprehensive Reports** are available only for DBMS data source. To access **Comprehensive Reports** screen, click the **Comprehensive Reports** button in the left side panel.

To know more information for **Comprehensive Reports**, visit [Comprehensive Reports](#).

To download the report in the PDF format, visit [Export as PDF](#).

- The Policy Content graph shows the breakdown of the content based on the different policies. It displays the size of the sensitive types belonging to multiple policies and data sources.



To view the size and count of the Policy Content on the basis of different data sources, click the chart.

POLICY CONTENT		
NAME ↑	SIZE	COUNT
File Share	4.02 KB	1
HDFS	108.06 KB	7

CLOSE

The screen displays the total count and size of the tables/files across each data sources name. E.g. in the above screenshot displays the information of the policy content for PII Hadoop in HDFS and File Share.

To view the complete details for each data source, click the data source name. It will display the **Details** screen.

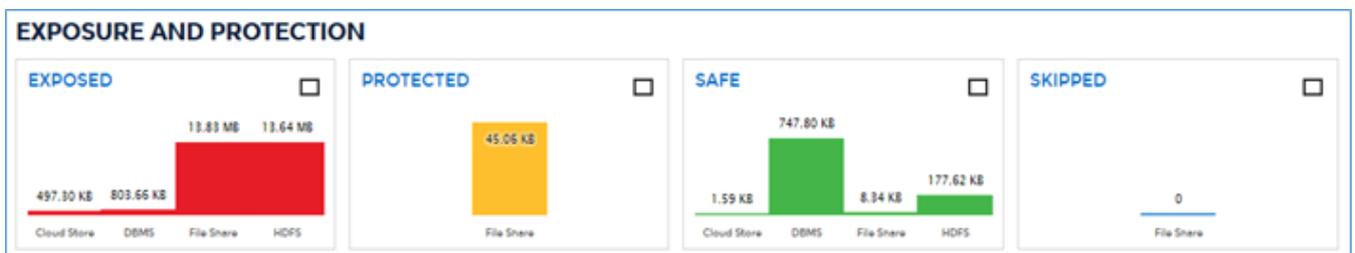
OVERVIEW: POLICY CONTENT > HDFS

CLUSTERS		CLUSTER DETAILS		
REFRESH	SHOW DETAILS	EXPORT AS PDF		
NAME	Name	Sensitive Type Group	Size	Count
hdfs_citr	hdfs_citr	PCI_Hadoop	108.06 KB	7
	<ul style="list-style-type: none"> karman <ul style="list-style-type: none"> avro <ul style="list-style-type: none"> avro/All_expressions1.avro avro/hsp_all_expression.avro Data_Files_100 <ul style="list-style-type: none"> Data_Files_100/builtinFile Data_Files_100/orcFile Data_Files_100/parqFile Data_Files_100/rcFile Data_Files_100/testFile.txt 	<ul style="list-style-type: none"> PCI_Hadoop PCI_Hadoop PCI_Hadoop Credit Card Credit Card PCI_Hadoop Credit Card Credit Card Credit Card Credit Card Credit Card 	<ul style="list-style-type: none"> 108.06 KB 108.06 KB 8.07 KB 4.02 KB 4.05 KB 99.99 KB 28.89 KB 4.50 KB 9.67 KB 27.94 KB 29.00 KB 	<ul style="list-style-type: none"> 7 file(s) 7 file(s) 2 file(s) 1 file(s) 1 file(s) 5 file(s) 1 file(s) 1 file(s) 1 file(s) 1 file(s) 1 file(s)

This screen will display information for the selected data source along with the content type, size of the file or table and the Sensitive Type group it holds.

To download the report in the PDF format, visit [Export as PDF](#).

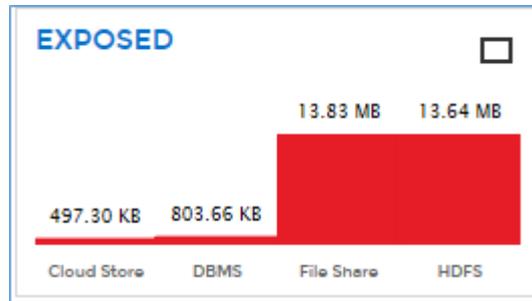
EXPOSURE AND PROTECTION



The **Exposure And Protection** graphs break down the objects based on the sensitive data in them. It provides information whether the sensitive data has been **Protected** (masked or encrypted), **Exposed** (detected but unprotected), **Skipped** or marked **Safe** in the objects.

Click on the **Legend** for each chart to hide or unhide the selected data in the graph.

1. The Exposed graph displays the size of the exposed objects. This graph depicts that during the scanning of the data source how much sensitive data has been detected and is exposed.



E.g., in the above screenshot 497.30 KB of Sensitive data in Cloud store is marked as unsafe. Similarly, the data for other three data sources have been displayed in the graph.

To view the complete details for selected data source, click the data source name. It will display the **Details** screen.

OVERVIEW: EXPOSED > HDFS

CLUSTERS		CLUSTER DETAILS			
NAME		Name	Exposed 7 file(s)	Sensitive Type Group	Size 108.06 KB
<input checked="" type="checkbox"/>	hdfs_citr	hdfs_citr	7 file(s)		108.06 KB
		/karman	7 file(s)	fpm,PCI_Hadoop	108.06 KB
		/karman/avro	2 file(s)	PCI_Hadoop	8.07 KB
		/karman/avro/All_expressions1.avro	Y	Credit Card	4.02 KB
		/karman/avro/hsp_all_expression.avro	Y	Credit Card	4.05 KB
		/karman/Data_Files_100	5 file(s)	PCI_Hadoop	99.99 KB

This screen will display the information for the selected data source along with the Exposed File column specifying the number of tables or files exposed and the name of the sensitive type group which is marked as 'Y' for being exposed or unprotected.

To download the report in the PDF format, visit [Export as PDF](#).

138. The Protected graph will display the total size of the objects which are marked as protected.



E.g., in the above screenshot the total of 45.06 KB of objects is Protected i.e. either masked or encrypted.

To view the complete details for selected data source, click the data source name. It will display the **Details** screen.

OVERVIEW: PROTECTED > DBMS

SERVICES **COMPREHENSIVE REPORTS** **SERVER DETAILS**

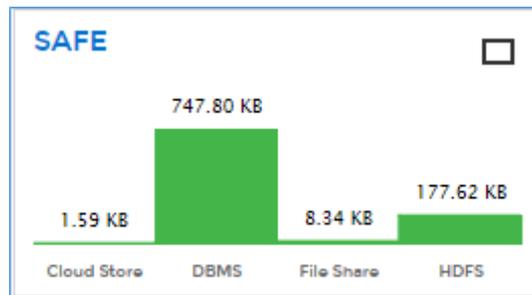
REFRESH SHOW DETAILS EXPORT AS PDF

NAME	Name	Protected	Size
Oracle-QA-NIHAR-ORACLE12C.CQGRVY7ZDRME.US-EAST-1.RDS.AMAZONAWS.COM	Oracle-QA-NIHAR-ORACLE12C.CQGRVY7ZDRME....	1 table(s)	0
Redshift-MANINDERDETECTION.CA0HM05RTE7E.US-EAST-1.REDSHIFT.AMAZONAWS.COM	Redshift-MANINDERDETECTION.CA0HM05RTE7E... nh_db7238 schema_a.table_a1	1 table(s) 1 table(s) Y	6.00 MB 6.00 MB
SQL Server-192.168.0.151	SQL Server-SANKUS-SQL.DATABASE.WINDOWS.... Masker dbo.mask_20	1 table(s) 1 table(s) Y	72.00 bytes 72.00 bytes
SQL Server-SANKUS-SQL.DATABASE.WINDOWS.NET	SQL Server-192.168.0.151 ccno_data	1 table(s) 1 table(s)	16.00 bytes 16.00 bytes

This screen will display the information for the selected data source along with the Protected File column specifying the count of tables/files and the size of each file marked as 'Y' (yes) which depicts that the file is protected.

To download the report in the PDF format, visit [Export as PDF](#).

139. The Safe graph will display the total size of the objects which are marked as Safe.



E.g., in the above screenshot the total of 747.80 KB of objects in DBMS is marked as Safe. Similarly, the objects for other three data sources can be seen in the graph such as 1.59 KB for Cloud store, 8.34. KB for File Share and 177.62 KB for HDFS are marked as Safe.

To view the complete details for selected data source, click the data source name. It will display the **Details** screen.

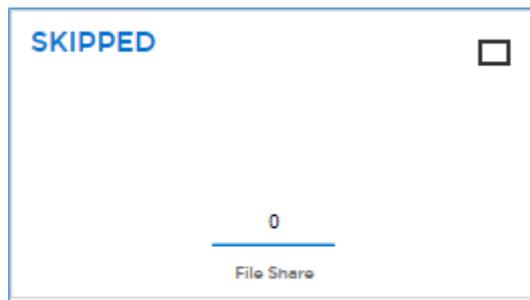
OVERVIEW: SAFE> HDFS

CLUSTERS		CLUSTER DETAILS		
<input checked="" type="checkbox"/>	NAME	Name	Safe 1 file(s)	Size 2.70 KB
<input checked="" type="checkbox"/>	hdfs_cltr	<ul style="list-style-type: none"> hdfs_cltr <ul style="list-style-type: none"> /karman <ul style="list-style-type: none"> /karman/avro <ul style="list-style-type: none"> /karman/avro/All_expressions1.avsc 	1 file(s)	2.70 KB

This screen will display the information for the selected data source along with the Safe File column containing the total count of tables/files and the size of each file marked as 'Y' (yes) which depicts that the file is Safe.

To download the report in the PDF format, visit [Export as PDF](#).

- 140. The Skipped graph display the total size of the objects for each data source that has been skipped. Any files/tables that have not been scanned is shown in the Skipped graph.



To view the complete details for selected data source, click the data source name. It will display the **Details** screen.

OVERVIEW: SKIPPED> HDFS

CLUSTERS		CLUSTER DETAILS		
<input checked="" type="checkbox"/>	NAME	Name	Skipped 2 file(s)	Size 0
<input checked="" type="checkbox"/>	hdfs_cltr	<ul style="list-style-type: none"> hdfs_cltr <ul style="list-style-type: none"> /karman <ul style="list-style-type: none"> /karman/avro <ul style="list-style-type: none"> /karman/avro/empty_example.avro /karman/avro/empty_example2.avro 	2 file(s)	0

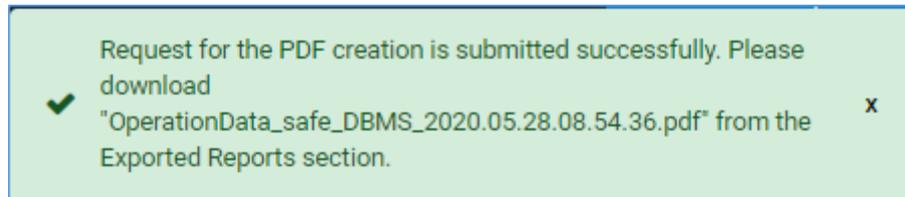
This screen will display the information for the selected data source along with the Skipped File column containing the total count of objects and the size of each file marked as 'Y' (yes) which depicts that the file is not scanned or have been skipped.

To download the report in the PDF format, visit [Export as PDF](#).

10.1.1 Export as PDF

To download the report in PDF format, perform the below steps:

1. Click the **Export as PDF** button. A pop-up with the PDF file name will appear on the screen.



141. To download the generated PDF report, visit [Exported Reports](#).

10.2 Periodic Reports

The **Periodic Reports** provide details about the scanned data sources that how much sensitive data has been exposed, protected, and safe for a specific duration. These reports display the size and count of the sensitive data detected for different source types over a defined period of time.

***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more, visit section **6.1.5. Manage Roles and Permissions**.

To access **Periodic Reports**, click **Reports > Periodic Reports**. Perform the following steps to generate a Periodic Report:

1. Enter the **Start Date** and the **End Date**. By default, the **End Date** will be the current date and the **Start Date** will be 7 days prior to it. You can modify both the dates.

142. Click Update Results .

143. The reports for the specified period of time will display:

Start Date	12/11/2019	End Date	12/18/2019	UPDATE RESULTS
SOURCE TYPE	EXPOSED	PROTECTED	SAFE	
DBMS	288.00 KB (7 Table(s))	0 (0 Table(s))	2.42 MB (57 Table(s))	
HDFS	8.69 KB (3 File(s))	0 (0 File(s))	0 (0 File(s))	

To view the sensitive groups that has been detected, perform the steps:

1. Click on the link under the **Exposed** field to view the sensitive group details for the corresponding source type. A popup appears:

HDFS DETAILS	
SENSITIVE GROUP NAME	PROCESSED COUNT
Address	3
Credit Card	3
Email Address	3
IP Address	3
NPI	3
Social Security	3
Telephone	3
URL	3

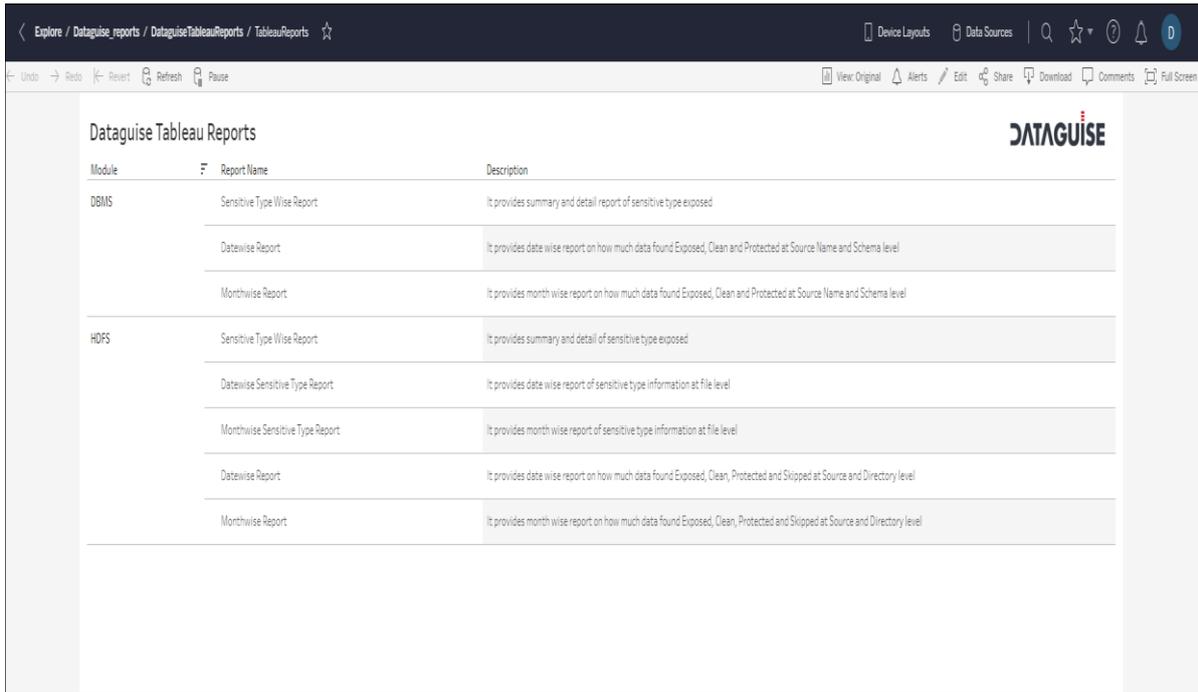
[OK](#)

This dialog box lists all the names of the Sensitive Groups and the total count of the sensitive data detected within each group for the selected Source Type.

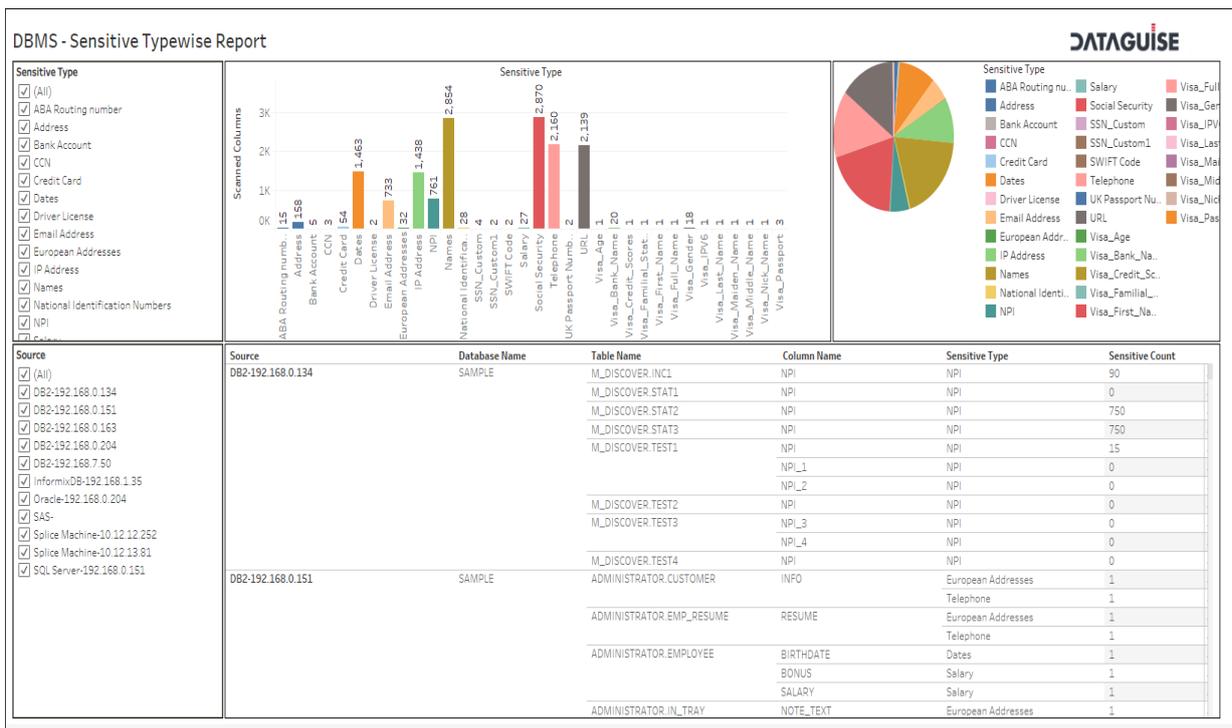
2. Click **OK** to close the pop up.

10.3 Tableau Reports

Go to DgSecure> Reports> Tableau Reports, to view detailed Tableau Reports. You will be logged into Tableau and see the screen below:



Click on the Report Name to view it. For instance, a Sensitive Type Wise Report for DBMS would appear as follows:



A Tableau report offers more flexibility in terms of selecting any number of Sensitive Type or Source System to view detailed graphs in an easily understandable format.

10.4 GDPR View

GDPR View offers insight into the details of the data sources, RoA and RtE request status, and the information related to the alerts and the alert rules.

***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more, visit section **6.1.5. Manage Roles and Permissions**.

Access the **GDPR View** screen from the menu under **Reports > GDPR View**.

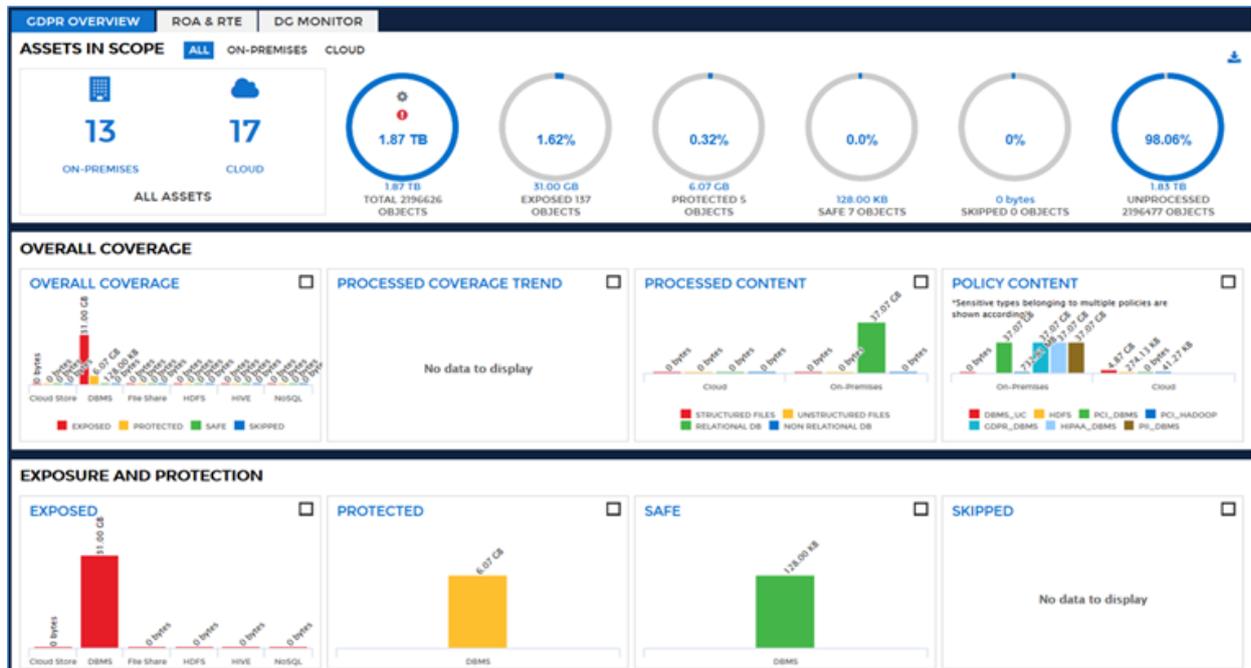
There are three tabs under GDPR View:

1. GDPR Overview
2. RoA (Right of Access) and RtE (Right to Erasure)
3. Dg Monitor

***Note:** If a user checks the **Include GDPR Dashboard View** checkbox on the **Policy > New Policy** tab, only then the data for that policy will be displayed under the **GDPR View** screen.

GDPR Overview

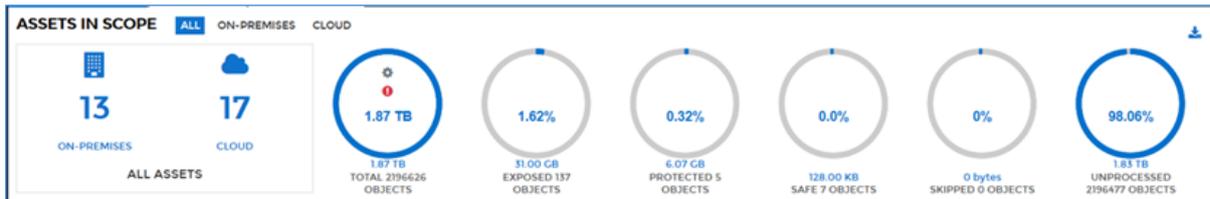
The **GDPR Overview tab** provides the complete details about the data sources.



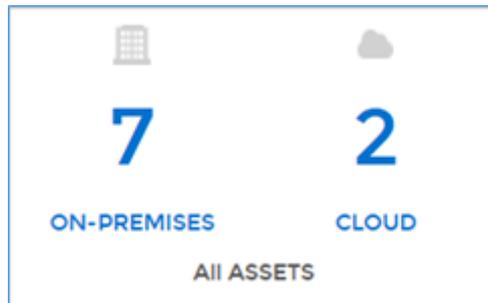
The charts in each row provide insight into the sensitive data status of different data sources. You can segregate the results for **All**, **On-Premises**, or **Cloud** data in the charts, based on the selection made at the top of the screen. Click  to download the **GDPR View** screen.

1. Assets in Scope

This panel provides an overview of the complete data assets.



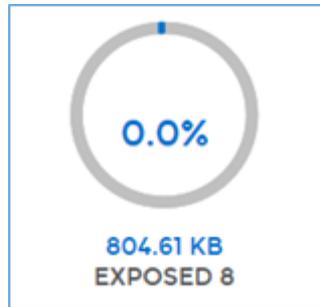
- a) The **All Assets** chart indicates that how many data assets are **On-Premises** and on the **Cloud**. Any AWS and Azure assets are considered in the cloud, while all other Hadoop and RDBMS assets are considered to be on-premises or in the cloud according to their designation. Hadoop location is set when setting up the cluster connection. RDBMS location is set when creating a database connection. One asset is equal to one Hadoop cluster or one database.



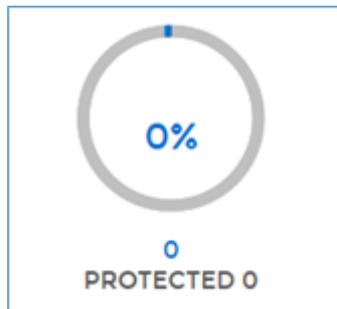
- 144. The **Total** chart indicates the total number of the sensitive data discovered across all the data assets. To see what will happen on clicking the chart, view [Server Details](#).



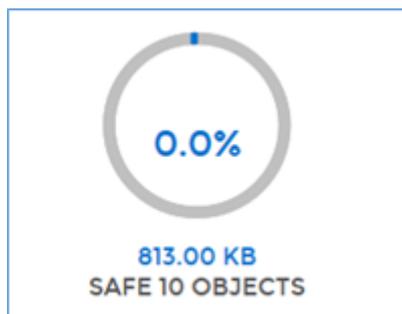
145. The **Exposed** chart indicates the percentage of the exposed files and tables that contain sensitive data. To see what will happen on clicking the chart, view [Server Details](#).



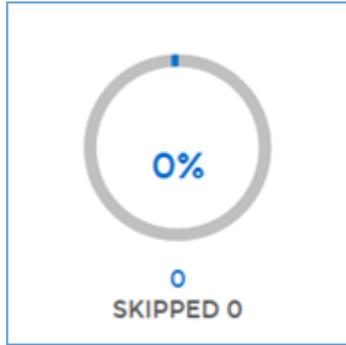
146. The **Protected** chart indicates the percentage of protected sensitive data sources that DgSecure has masked or encrypted. To see what will happen on clicking the chart, view [Server Details](#).



147. The **Safe** chart indicates the percentage of the safe objects within the assets. To see what will happen on clicking the chart, view [Server Details](#).

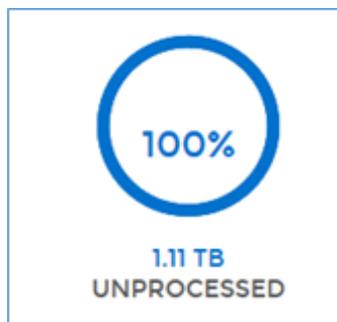


148. The **Skipped** chart indicates the percentage of the skipped objects. To see what will happen on clicking the chart, view [Server Details](#).



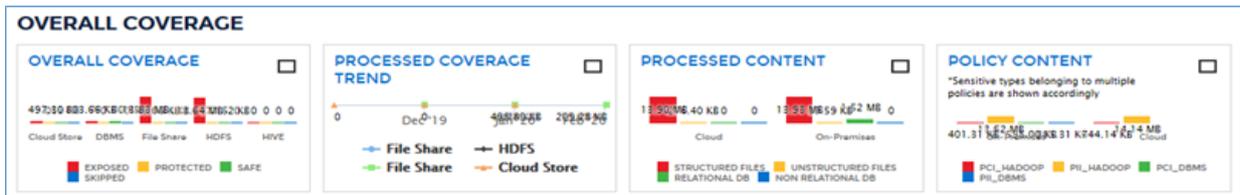
149. The **Unprocessed** chart indicates the percentage of **Unprocessed** objects.

To see what will happen on clicking the chart, view [Server Details](#).

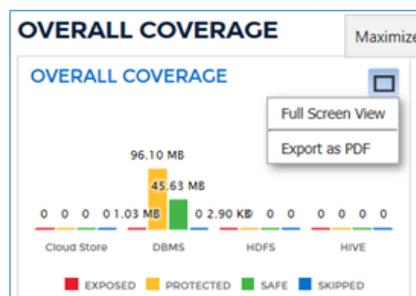


150. **Overall Coverage**

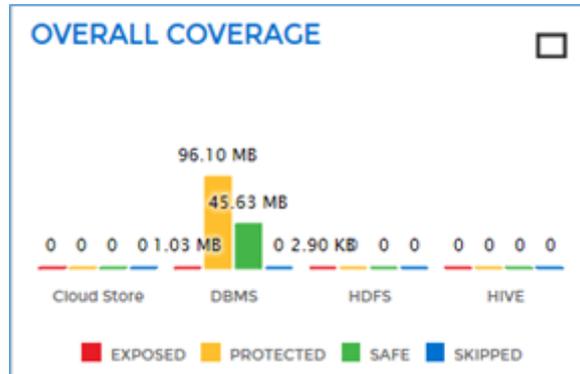
This panel offers insight into the sensitive data coverage across all the data assets.



Click on the rectangular icon shown in the top right corner of each of the graph to view it in the full screen as well as to download it in the PDF format.



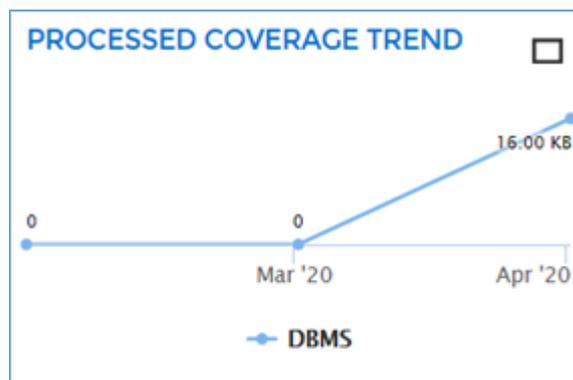
- a) The **Overall Coverage** graph shows the number of tables and files involved in the coverage. The Sensitive data can be in any of the state i.e. **Exposed**, **Protected**, **Safe** and **Skipped**.



- **Exposed:** The Sensitive data which is not protected during the processing, is marked as Exposed.
- **Protected:** The Sensitive data which is encrypted or masked while processing, is categorized as Protected.
- **Safe:** The Sensitive data which is marked as safe in the Safelist and the data which is already protected, is marked as Safe.
- **Skipped:** The data which were skipped during the processing, is marked as Skipped.

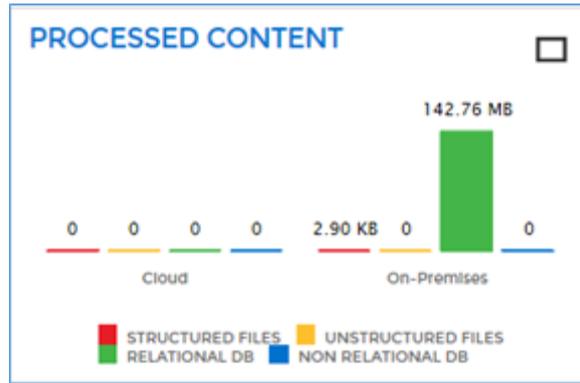
To see what will happen on clicking the graph, view [Server Details](#).

151. The **Processed Coverage** graph shows the coverage trend of the Sensitive Data for different data sources.



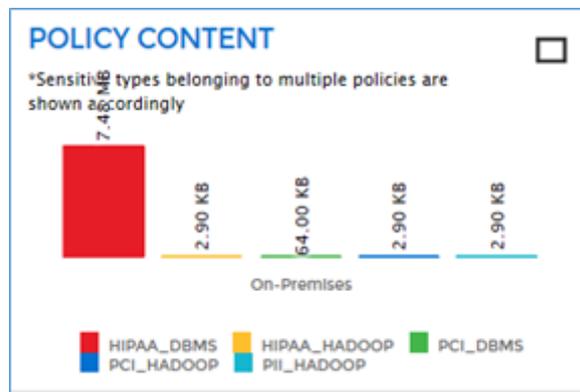
To see what will happen on clicking the graph, view [Server Details](#).

152. The **Processed Content** graph shows the breakdown of the data based on the structured vs unstructured data and relational vs non-relational data sources.



To see what will happen on clicking the graph, view [Server Details](#).

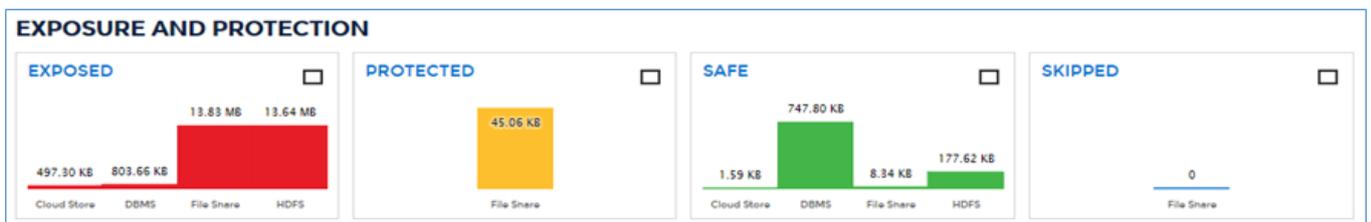
153. The **Policy Content** graph shows the breakdown of the data based on the different policies.



To see what will happen on clicking the graph, view [Server Details](#).

154. **Exposure and Protection**

The **Exposure & Protection** graphs break down the data sources based on the Sensitive data in them. It provides information whether the Sensitive data has been **Protected** (masked or encrypted), **Exposed** (detected but unprotected), **Skipped** or **Safe**.



- a) The **Exposed** graph displays information for the exposed data, data source-wise. This graph basically depicts that during the processing the sensitive data was detected which falls under the unprotected category i.e. Exposed.



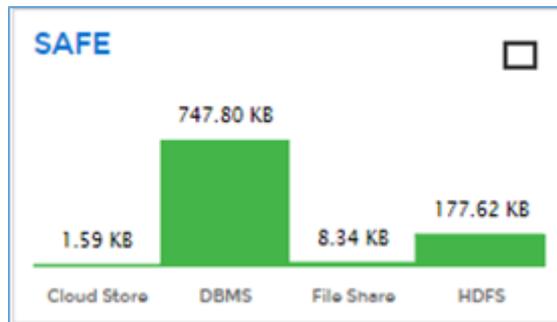
To see what will happen on clicking the graph, view [Server Details](#).

155. The **Protected** graph displays information for the protected data which is either masked or encrypted, data source-wise.



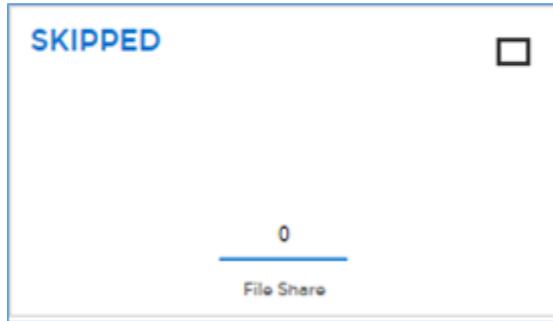
To see what will happen on clicking the graph, view [Server Details](#).

156. The **Safe** graph displays information for the Sensitive data which are marked as Safe, data source-wise.



To see what will happen on clicking the graph, view [Server Details](#).

157. The **Skipped** graph displays information for the Skipped data, data source-wise. Any files or tables that have not been scanned is shown in the Skipped graph.



To see what will happen on clicking the graph, view [Server Details](#).

10.4.1 Server Details

To view the **Server/Cluster Details** and to download the report, perform the following steps:

1. Click on the chart. A pop up will appear that displays total data details in a tabular form, data source-wise.

EXPOSED		
NAME ↑	SIZE	COUNT
Cloud Store	12.34 KB	2
DBMS	0	21
File Share	4.02 KB	1
HDFS	108.06 KB	7
NoSQL	0	0

CLOSE

2. Click on the name of the data source to view the **Server Details** for the selected data source.

GDPR VIEW: EXPOSED > DBMS

SERVERS		COMPREHENSIVE REPORTS	SERVER DETAILS
<input type="checkbox"/>	NAME		Size 35.92 GB
<input type="checkbox"/>	Oracle-192.168.0.163		Count 252

Buttons: REFRESH, SHOW DETAILS, EXPORT AS PDF

3. Check the checkboxes in the left panel listing the servers or clusters name to view the in-depth details, and click **Show Details**. The **Server Details** screen will display the server details as well as the table details.

OVERVIEW EXPOSED> HDFS

CLUSTERS		CLUSTER DETAILS			
<input type="checkbox"/>	NAME	Name	Exposed 7 file(s)	Sensitive Type Group	Size 108.06 KB
<input checked="" type="checkbox"/>	hdfs_citr	<ul style="list-style-type: none"> hdfs_citr <ul style="list-style-type: none"> /karman <ul style="list-style-type: none"> /karman/avro <ul style="list-style-type: none"> /karman/avro/All_expressions1.avro <ul style="list-style-type: none"> Y <ul style="list-style-type: none"> Credit Card <ul style="list-style-type: none"> 4.02 KB /karman/avro/hsp_all_expression.avro <ul style="list-style-type: none"> Y <ul style="list-style-type: none"> Credit Card <ul style="list-style-type: none"> 4.05 KB /karman/Data_Files_100 <ul style="list-style-type: none"> 5 file(s) <ul style="list-style-type: none"> PCI_Hadoop <ul style="list-style-type: none"> 99.99 KB 	7 file(s)		108.06 KB

- Click **Export As PDF** export the report.
- You can access the **Exported Reports** from the menu under **Reports > Exported Reports**.

EXPORTED REPORTS

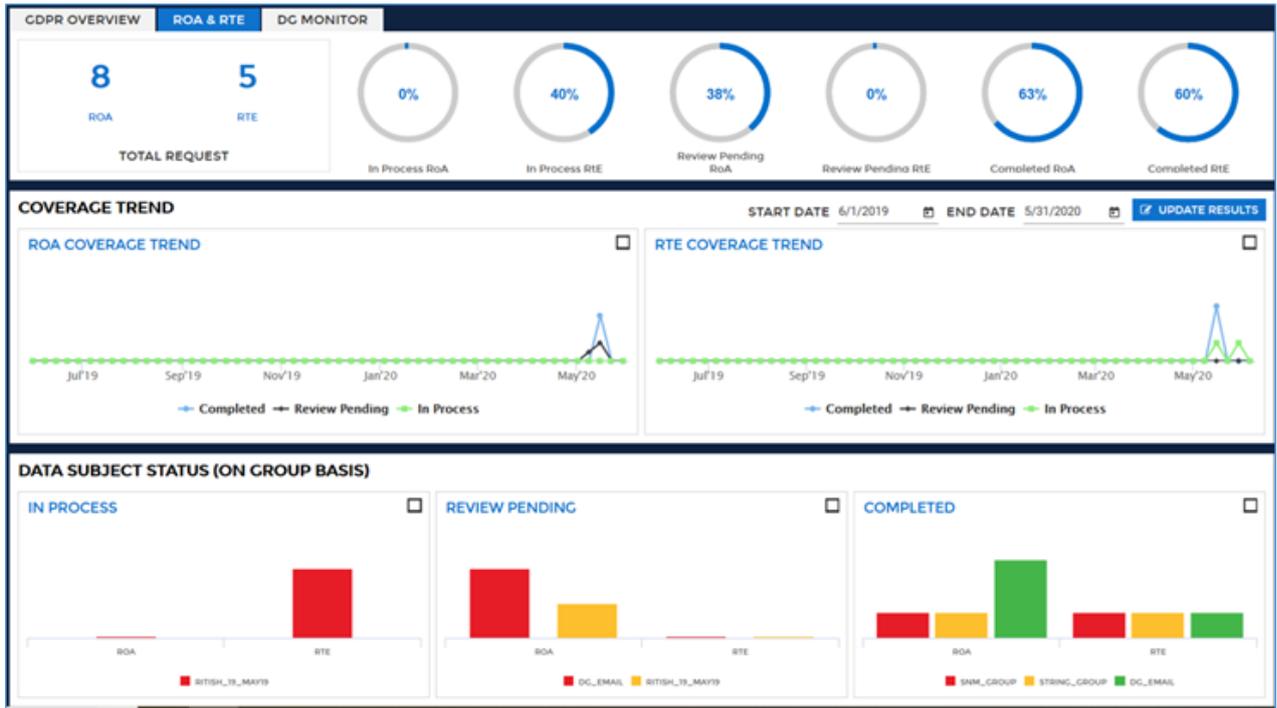
PDF NAME	DATA SOURCE TYPE	CATEGORY	EXPORTED TIME	PROGRESS	DOWNLOAD
OperationData_exposed_DBMS_2020.05.27.22.29.28.pdf	DBMS	exposed	2020-05-27T22:32:47.216+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.14.pdf	DBMS	safe	2020-05-27T22:33:36.866+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.18.pdf	DBMS	safe	2020-05-27T22:34:10.598+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.03.pdf	DBMS	safe	2020-05-27T22:34:12.22+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.31.56.pdf	DBMS	safe	2020-05-27T22:35:17.733+05:30	100%	
OperationData_exposed_DBMS_2020.05.27.22.29.14.pdf	DBMS	exposed	2020-05-27T22:32:07.251+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.22.pdf	DBMS	safe	2020-05-27T22:33:54.315+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.32.44.pdf	DBMS	safe	2020-05-27T22:35:36.585+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.33.25.pdf	DBMS	safe	2020-05-27T22:36:06.924+05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.33.54.pdf	DBMS	safe	2020-05-27T22:36:26.383+05:30	100%	

- Click in the **Download** column to download the report.

10.5 RoA & RtE

RoA and RtE screen provides insight into RoA and RtE request status. The screen is further divided into three panels:

- Overview
 - Coverage Trend
 - Data Subject Status (On Group Basis)



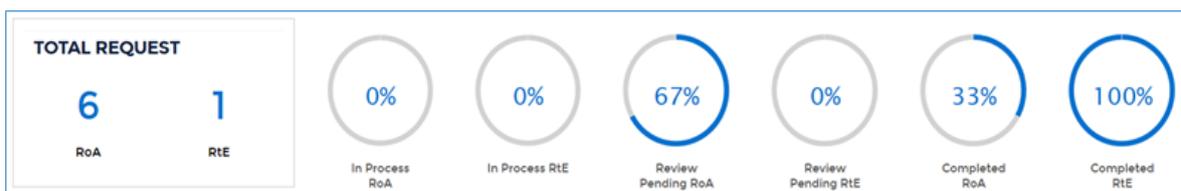
***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more, visit section **6.1.5. Manage Roles and Permissions**.

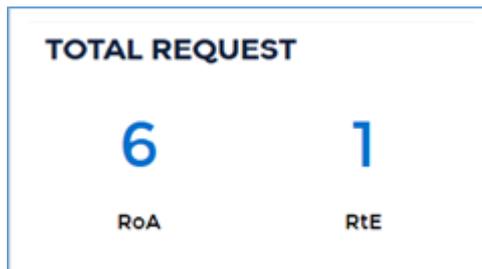
Click the menu under **Reports > GDPR View > RoA and RtE**.

1. Overview

The charts in this panel provide an overview of RoA and RtE requests.



- a) The first chart indicates that how many data subjects have been submitted for RoA/RtE scan. The Right of Access (RoA) allows the data subject to retrieve their personal data, which the controller might have used or processed. The Right to Erasure (RoE) is also known as the Right to be Forgotten under GDPR. It allows the data subject to request the erasure of all their personal data being processed or stored, by a controller.



- b) The second and third charts display the total percentage of data subjects are under Incoming status. Whenever a data subject is created and submitted for RoA/RtE scan, the data subject falls under **In Process** status.



- c) The fourth and fifth charts display the total percentage of data which are awaiting for manual review.



- d) The sixth and seventh charts display the total percentage of data subjects under **Completed** status i.e. fulfilled the request.

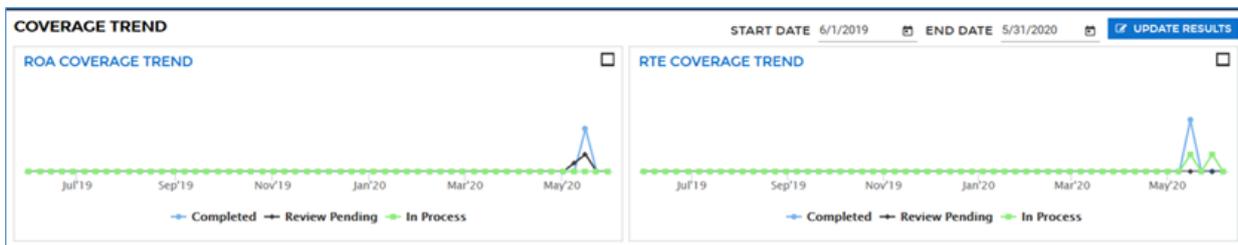


160. Coverage

Trend

The **Coverage Trend** panel offers insight about the number of requests having In Process, Review Pending, and Completed status. The information is displayed on the basis of **Start Date** and **End Date**. By default, the information is displayed for one year. You can modify both the dates. The difference between the start date and end date will remain one year. Click **Update Results** to get the updated result.

The first and second graph shows that how many requests were In Process, Review Pending, and Completed for RoA and RtE respectively for the dates that you have specified.



You can also maximize the graph by clicking on the rectangular icon displayed on the extreme right corner of both the graphs.

161. Data Subject Status (On Group Basis)

A data subject belongs to a data group. The rule specified in the group to which a data subject belongs decides the operation that has to be performed.

The graphs in the **Data Subject Status (On Group Basis)** panel show that how many data subjects for a specific data group are, In Process, Review Pending, and Completed status for RoA and RtE.

On hovering over any bar icon, you can see the name of the data group followed by the number of data subjects present in that data group for a specific status. All the data groups for which information is getting displayed, are listed at the bottom of each graph.



You can also maximize the graph by clicking on the rectangular icon displayed on the

extreme right corner of each graph.

10.6 DgMonitor

DgSecure detection tasks need to be run in order to identify where sensitive data resides. After executing the detection tasks, sensitive data can be located and monitored.

***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more, visit section **6.1.5. Manage Roles and Permissions**.

The **Dg Monitor** screen offers a birds-eye view of the current monitoring statuses of known sensitive data. The screen is divided into three panels:

1. Alert Rules and Alerts

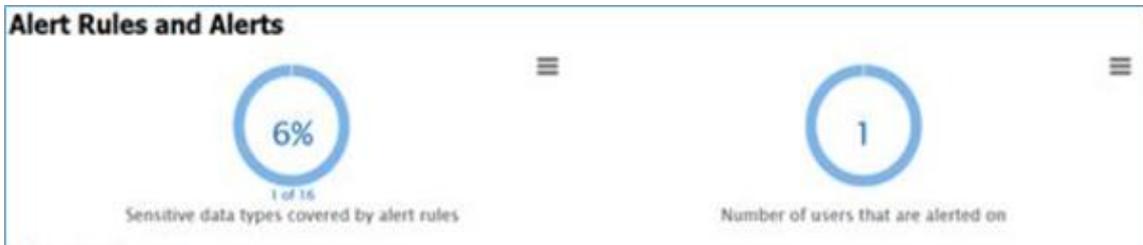
- 162. Alert Rules
- 163. Alert Issued/24HRS



Access the Dg Monitor screen from the menu under **Reports > GDPR View > Dg Monitor**.

1. Alert Rules and Alerts

This panel has two graphs. The first graph shows the percentage of sensitive data covered by the alert rules. Whereas, the second graph shows the number of people who trigger alert notifications.



2. Alert Rules

This panel contains two graphs. The first graph shows how many alert rules are defined for a specific source system. Whereas, the second graph shows how many alert rules are defined for a specific compliance policy.



3. Alert Issued/24HRS

The panel has three graphs, which shows the number of alerts issued on compliance policies, source systems, and users over the past 24 hours respectively.



10.7 Comprehensive Reports

Comprehensive Reports provide information for various dimensions such as Sensitive Type, Servers, and Policies. An information is displayed in graphical representation. This screen

compares the following aspects of functions on DgSecure:

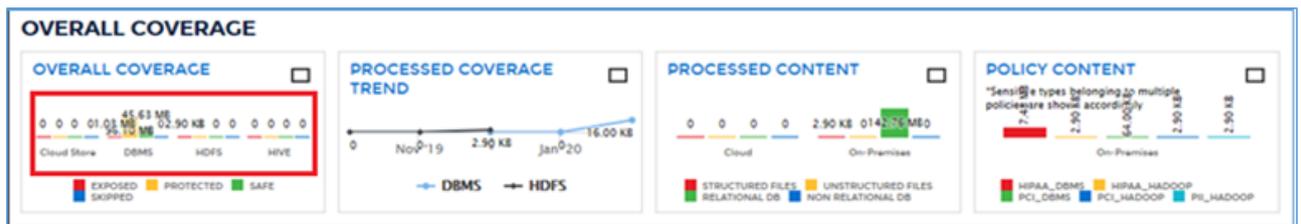
1. Total Count of Servers and Tables Scanned by DB Type
 164. Percentage of Tables Exposed by DB Type
 165. Percentage of Servers Scanned by DB Type
 166. Percentage of Tables Exposed by Sensitive Type
 167. Percentage of Tables Exposed by DB Type by Policy Type

***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more, visit section **6.1.5. Manage Roles and Permissions**.

To access **Comprehensive Reports**, click **Reports > Overview**. Perform the following steps:

1. Click on the bar chart of **Overall Coverage** graph displayed on the **Overview** screen.



2. A screen containing information for server will be displayed. Click the **Comprehensive Report** button on top left corner of the screen.

***Note:** The name of the screen will vary depending on the selected data source. For example, Server Details name will be displayed if DBMS chart is selected. Similarly, the name will change to Cluster Details, if Hadoop chart is selected.

OVERVIEW: TOTAL > DBMS			
SERVERS		SERVER DETAILS	
<input checked="" type="checkbox"/> COMPREHENSIVE REPORTS		<input type="checkbox"/> SERVER DETAILS	
<input type="button" value="REFRESH"/>		<input type="button" value="SHOW DETAILS"/>	
		<input type="button" value="EXPORT AS PDF"/>	
<input checked="" type="checkbox"/>	NAME	TYPE	Size 873.19 GB
<input checked="" type="checkbox"/>	Oracle-N.A.	Oracle	Count 116283
<input checked="" type="checkbox"/>	Sybase IQ-182.73.95.113	Sybase IQ	
<input checked="" type="checkbox"/>	Teradata-153.64.73.16	Teradata	

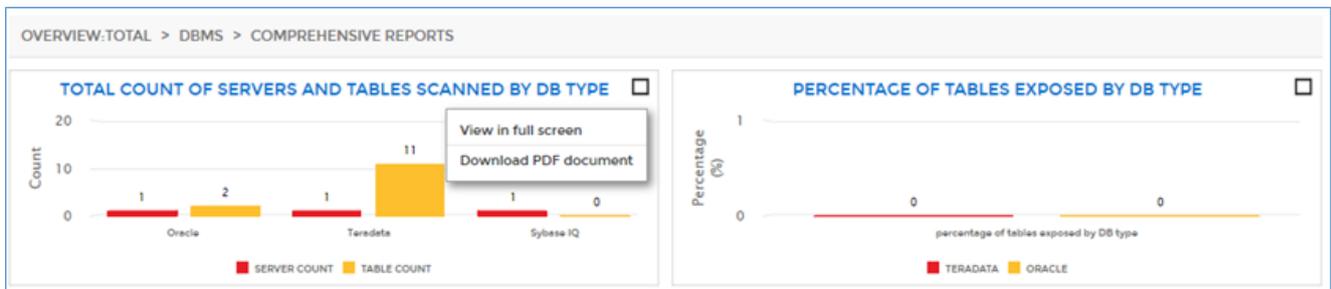
3. Click the **Comprehensive Reports** button and the **Comprehensive Reports** screen will appear.



4. To view a report in full screen or to download a report in PDF format, click Maximize option at the top right corner of each graph.

a) Click **Full Screen View** to maximize the selected graph.

168. Click **Export to PDF** if you want to download the report.



10.8 Audit Reports

The Dashboard provides extensive reports of DBMS and Hadoop. The **Audit Reports Dashboard** allows you to view additional details based on the operations performed by each user such as who logged into the DgSecure application, which task was executed, when did user log off from the application, etc.

***Note:** By default, only Admin can access the **Reports** section. You can provide rights to other users for accessing the **Reports** section through, **User Management > Roles > Edit Product Access Permissions** option in DgAdmin.

To know more, visit section **6.1.5. Manage Roles and Permissions**.

The report provides extensive information for DBMS and Hadoop data source. The **Audit Reports** maintain a logs of all the accesses to DgSecure and the information based on the operations performed by each user.

Within one screen, you can view information either for DBMS or for the Hadoop based on the selection made in **Select Module** drop-down.

REPORTS / AUDIT REPORTS

Select Module: Hadoop

BY USERS | BY EVENTS

USERS Filter by Roles Filter by Username CLEAR FILTER

USER	ROLE	NAME	EMAIL
dataguide	SUPER_ADMIN		dg@dg.in
Others	DEFAULT_USER		

EVENTS From: 2/28/2020 To: 5/28/2020 UPDATE RESULT EXPORT TO PDF

S.NO.	EVENT	EVENT DETAILS	TIMESTAMP
1	Log In	Log In 'dataguide' / 'SUPER_ADMIN'	May-28-2020 02:45:47
2	Log Off	Log Off 'dataguide' / 'SUPER_ADMIN'	May-27-2020 10:31:35
3	User session timeout/expired	User session timeout/expired 'dataguide' / 'SUPER_ADMIN'	May-27-2020 10:31:31
4	User session timeout/expired	User session timeout/expired 'dataguide' / 'SUPER_ADMIN'	May-27-2020 10:17:31
5	Log Off	Log Off 'dataguide' / 'SUPER_ADMIN'	May-27-2020 10:08:48
6	User session timeout/expired	User session timeout/expired 'dataguide' / 'SUPER_ADMIN'	May-27-2020 10:05:31
7	Log In	Log In 'dataguide' / 'SUPER_ADMIN'	May-27-2020 09:36:00

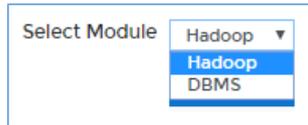
1 - 20 of 320

To access **Audit Reports**, click **Reports > Audit Reports** screen.

There are two types of **Audit Reports** based on which you can view information:

- By Users:** If you want to view the report for a specific User, click the **Reports > Audit Report > By Users** tab.
 - 169. **By Events:** if you want to view report for any event within a specific date range, click the **Reports > Audit Report > By Events** tab.

Within one screen, you can view information either for DBMS or Hadoop based on the selection made in **Select Module** drop-down.



The screens are discussed below:

10.8.1 By Users

The **Audit Reports – By Users** screen is divided into two panels.

1. **Users** - This panel will display the list of all available user.

170. **Events** - This panel displays the list of all operations performed by a user selected in **Users** panel.

The screenshot shows the 'REPORTS / AUDIT REPORTS' interface. At the top, there is a 'Select Module' dropdown set to 'Hadoop'. Below this are two tabs: 'BY USERS' (active) and 'BY EVENTS'. The 'BY USERS' panel contains a table with columns 'USER', 'ROLE', 'NAME', and 'EMAIL'. It lists 'dataguise' with role 'SUPER_ADMIN' and email 'dg@dg.in', and 'Others' with role 'DEFAULT_USER'. There are filters for 'Filter by Roles' and 'Filter by Username', and a 'CLEAR FILTER' button. The 'BY EVENTS' panel shows a table with columns 'S.NO.', 'EVENT', 'EVENT DETAILS', and 'TIMESTAMP'. It displays a list of events such as 'Log In', 'Log Off', and 'User session timeout/expired' for the 'dataguise' user. The date range is set from '2/28/2020' to '5/28/2020'. There are buttons for 'UPDATE RESULT' and 'EXPORT TO PDF'.

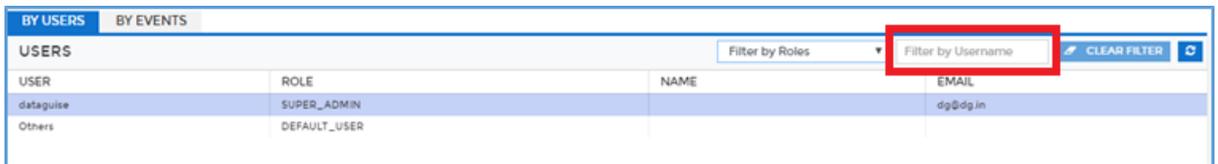
1. Users

The **Users** panel allows you to filter the Events panel data based on the selected the user. This panel displays the information for selected user such as Name, Email and Role assigned to user, etc. The **Filter By Roles** drop-down allows you to select users based on which you can filter the information.



Similarly, you can also filter the **Events** panel by providing the User Name in **Filter by UserName** searchbox. To remove the applied filters, click **Clear Filter** button.

To refresh the **Users** panel with updated records, click .



Perform following steps to view specific users or usernames:

- a) When you select the **By Users** tab. The **Users** panel will display the list of users and their role in **DgSecure**.

- 171. To view report for a specific role or a specific username in the **Users** panel. Select the role in **Filter by Roles** drop-down or enter the username in **Filter by Username** textbox.

The **Filter by Role** will display the result in **User** panel for the selected role. For **Filter by Username**, the **User** panel will display the data for the specific user entered in **Filter by Username**.



- 172. Select the user in the **Users** panel and the **Events** panel will display events for the selected User.

- 173. You can also view **Events** for the selected user based on the date range defined in the **Date Filter**. Click  to update the results.

REPORTS / AUDIT REPORTS

Select Module Hadoop

BY USERS BY EVENTS

USERS SUPER_ADMIN Filter by Username CLEAR FILTER

USER	ROLE	NAME	EMAIL
datagui	SUPER_ADMIN		dg@dg.in

EVENTS From: 2/28/2020 To: 5/28/2020 UPDATE RESULT EXPORT TO PDF

S.NO.	EVENT	EVENT DETAILS	TIMESTAMP
1	Log In	Log In 'datagui' / 'SUPER_ADMIN'	May-28-2020 03:55:20
2	Log In	Log In 'datagui' / 'SUPER_ADMIN'	May-28-2020 03:54:42
3	User session timeout/expired	User session timeout/expired 'datagui' / 'SUPER_ADMIN'	May-28-2020 03:42:01
4	Log In	Log In 'datagui' / 'SUPER_ADMIN'	May-28-2020 03:34:01
5	Log In	Log In 'datagui' / 'SUPER_ADMIN'	May-28-2020 02:45:47
6	Log Off	Log Off 'datagui' / 'SUPER_ADMIN'	May-27-2020 10:31:35
7	User session timeout/expired	User session timeout/expired 'datagui' / 'SUPER_ADMIN'	May-27-2020 10:31:31

1 - 20 of 324

E.g., in the above image when SUPER_ADMIN role is selected in the **Filter by Role** drop-down. The **Events** panel is populated with all the operations performed by the SUPER_ADMIN.

- 174. Click **Export to PDF** button to generate a PDF format of the report.
- 175. Click **Clear Filters** to clear the results.

176. Events

The **Events** panel allows you to filter the data for a defined time period. The **Events** panel will display all the operations performed by the user along with the **Timestamp** and **Event Details**.

Perform following steps to view specific event within a date range:

EVENTS From: 2/28/2020 To: 5/28/2020 UPDATE RESULT EXPORT TO PDF

S.NO.	EVENT	EVENT DETAILS	TIMESTAMP
34	Log In	Log In 'datagui' / 'SUPER_ADMIN'	May-27-2020 03:29:23
35	Log In	Log In 'datagui' / 'SUPER_ADMIN'	May-27-2020 03:28:37
36	Testing Hdfs Conn	Testing Hdfs Conn Hadoop Agent Details	May-27-2020 02:34:49
37	Get Cluster Details	Get Cluster Details Hadoop Agent Details Cluster Type: 'HDI-3.5', Agent Id: '29'	May-27-2020 02:23:22
38	Testing Hdfs Conn	Testing Hdfs Conn Hadoop Agent Details	May-27-2020 02:22:15
39	Testing Hdfs Conn	Testing Hdfs Conn Hadoop Agent Details	May-27-2020 02:22:11
40	Create	Create Infrastructure node 'azure_v2' Type: 'DHA'	May-27-2020 02:22:11

21 - 40 of 324

- a) To modify the **Event** panel for any specific date range. Select starting date in **From** and end date in **To** drop-down.

177. Click the **Update Results** button to apply filters.

S.NO.	EVENT	EVENT DETAILS	TIMESTAMP
34	Log In	Log In 'dataguise' / 'SUPER_ADMIN'	May-27-2020 03:29:23
35	Log In	Log In 'dataguise' / 'SUPER_ADMIN'	May-27-2020 03:28:37
36	Testing Hdfs Conn	Testing Hdfs Conn Hadoop Agent Details	May-27-2020 02:34:49
37	Get Cluster Details	Get Cluster Details Hadoop Agent Details Cluster Type: 'HDI-3.5', Agent Id: '29'	May-27-2020 02:23:22
38	Testing Hdfs Conn	Testing Hdfs Conn Hadoop Agent Details	May-27-2020 02:22:15
39	Testing Hdfs Conn	Testing Hdfs Conn Hadoop Agent Details	May-27-2020 02:22:11
40	Create	Create Infrastructure node 'azure_v2' Type: 'DHA'	May-27-2020 02:22:11

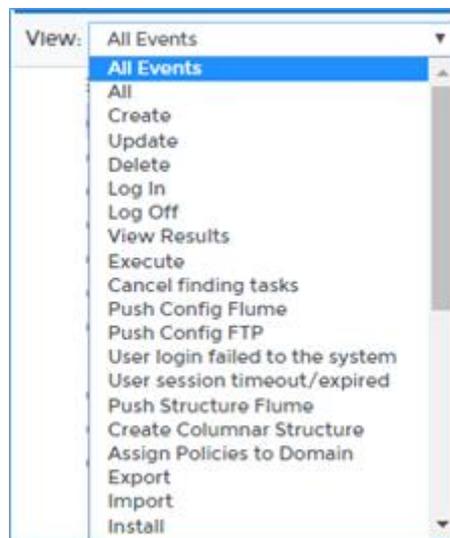
E.g., in the above image the **Events** panel displays the filter results for a given time period.

178. Click  button to update the Events panel with updated data.

179. Click **Export To PDF** button if you want to download the data.

10.8.2 By Events

The **Audit Reports – By Events** screen allows you to filter data for a defined time period. The **Events** screen will display all the operations performed by the user along with the **Timestamp** and **Events Details**. You can also filter the data by selecting a specific event in **View** drop-down. There are around 35 pre-defined events for which you can filter out the data.



To view reports in **By Events** tab, perform the following steps:

- a) Select the event in **View** drop-down.

180. Select the starting date range in **From** and end date in **To** drop-down or events in **View** drop-down. You can select either both the option or just one option from View and Date Range.



181. Click the **Update Result** button to apply the filters.
182. The **Events** panel will be updated for the defined time period or for selected events. It will display all the occurrences of operations performed for a defined time period or for selected event specified in the top panel.

REPORTS / AUDIT REPORTS

Select Module: Hadoop

BY USERS | **BY EVENTS**

View: Execute | From: 2/28/2020 | To: 5/28/2020 | UPDATE RESULT | EXPORT TO PDF

S.NO.	EVENT	USER & ROLE	EVENT DETAILS	TIMESTAMP
1	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-27-2020 03:56:10
2	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-27-2020 03:52:35
3	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 08:08:52
4	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 08:05:12
5	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 06:05:15
6	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 04:31:09
7	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 04:25:21
8	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task.	May-26-2020 03:58:29
9	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task.	May-26-2020 02:47:02
10	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 01:53:59
11	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-26-2020 00:53:47
12	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task: fa_det_schl	May-26-2020 00:28:46
13	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-25-2020 23:53:44
14	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task: fa_det_schl	May-25-2020 23:28:42
15	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-25-2020 22:53:40
16	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task: fa_det_schl	May-25-2020 22:28:35
17	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-25-2020 21:53:30
18	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task: fa_det_schl	May-25-2020 21:28:23
19	Execute	dataguide/SUPER_ADMIN	Execute Task.	May-25-2020 20:53:22
20	Execute	dataguide/SUPER_ADMIN	Execute Hadoop Task: fa_det_schl	May-25-2020 20:28:21

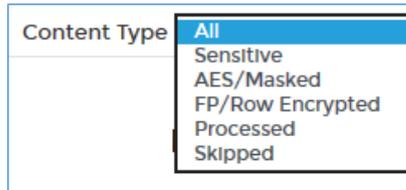
1 - 20 of 73

183. Click  button to updated the **By Events** tab with latest records.
184. Click the **Export To PDF** button to download the report in the PDF format.

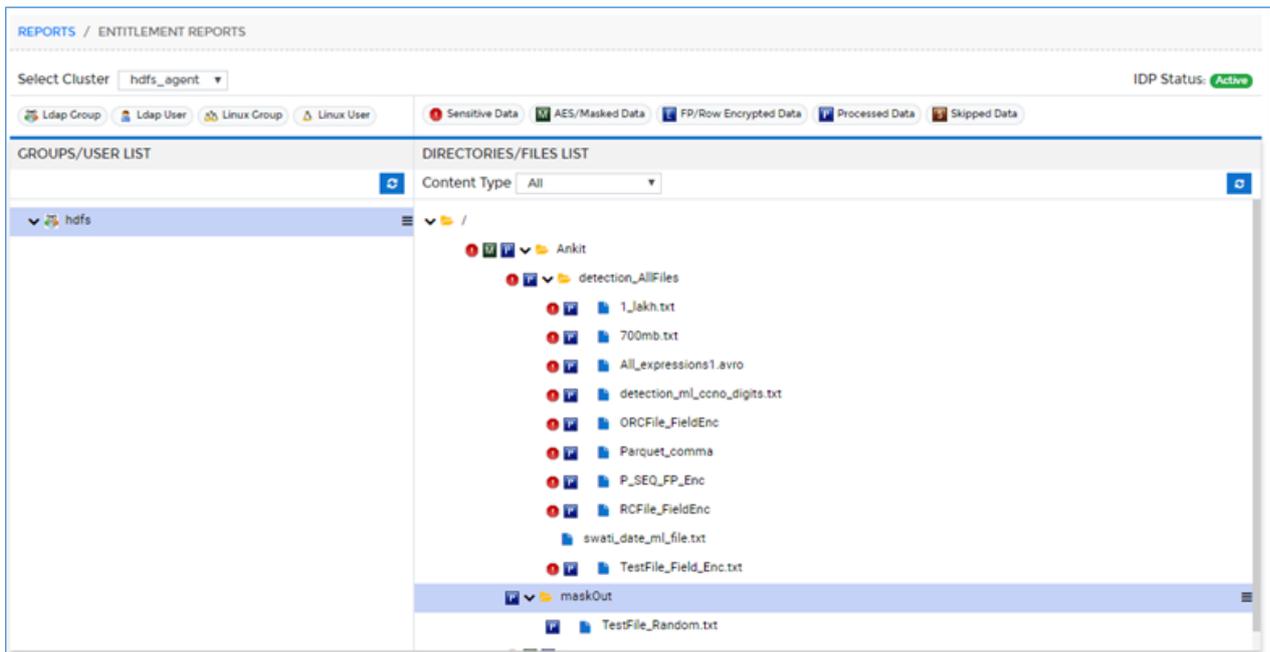
10.9 Hadoop

10.9.1 Entitlement Report

The **Entitlement Report** display the access rights for both specific user / group or directory / file. It also specifies the content type for a data stored in directory/file. A content type for a data stored in a file/directory can be Sensitive, AES/Masked, FP/Row Encrypted, Processed and Skipped.

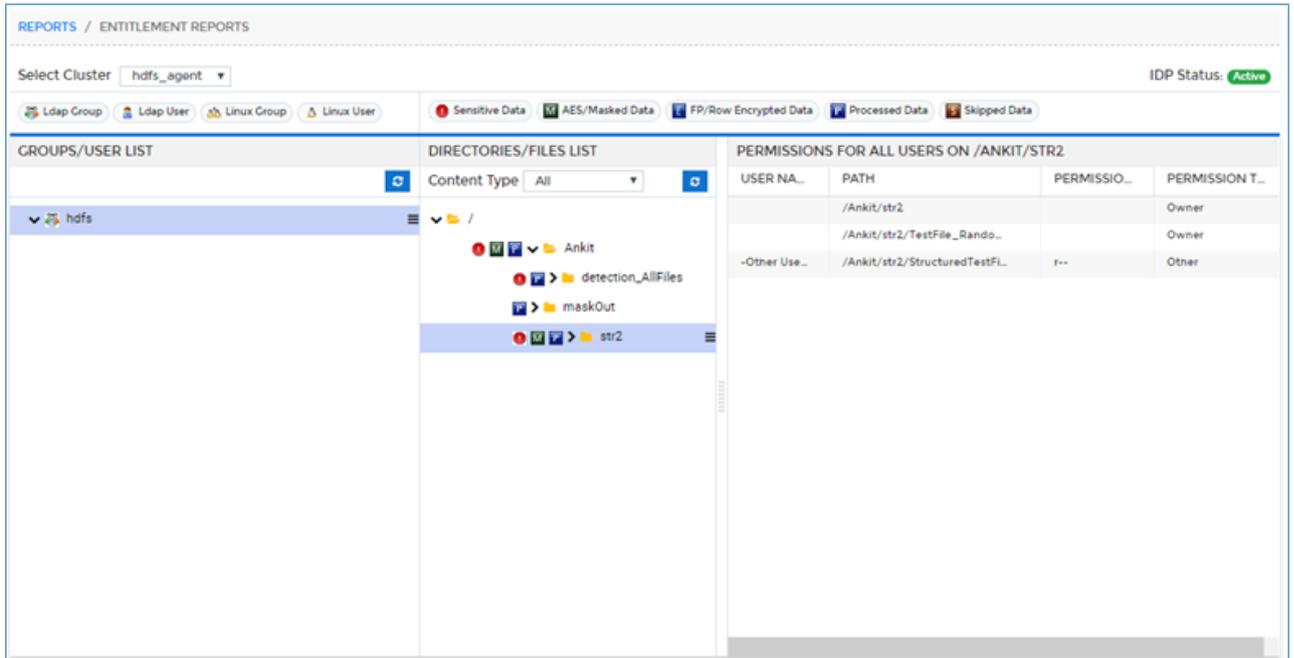


When **Entitlement Report** is processed for a specific user/group, this indicates what their access rights are for a specific file or directory. If it is processed on a directory/file, then it indicates the access rights for a specific user/group.



The Entitlement Report screen is divided into four panels

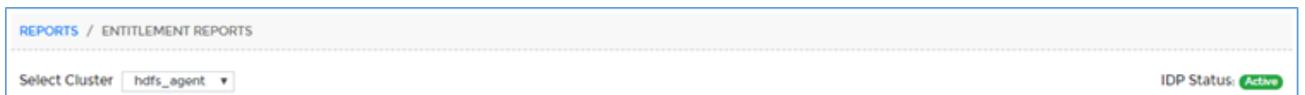
1. Top View
2. Groups/Users List
3. Directories/File List
4. Permission List



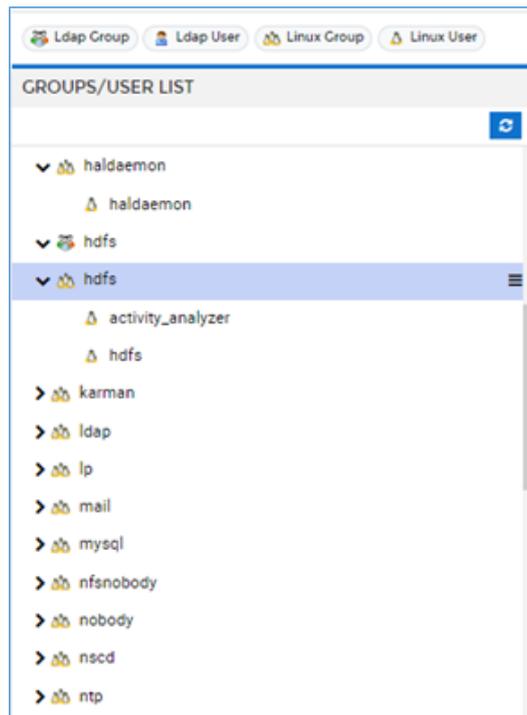
There is one **Entitlement Report** per cluster. To access the **Entitlement Report** screen, click **Reports > Hadoop > Entitlement Report**.

1. Top View

This panel allows you to select the cluster from the Select Cluster drop-down. It also display the status of an IDP i.e. whether it is Active or Inactive.

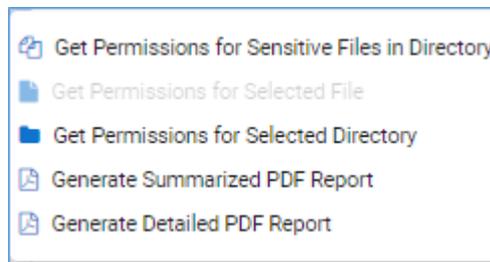


- 2. Groups/User List:** This panel display all the **LDAP/Linux Groups** to which **LDAP/Linux User** is associated for a given cluster. The icon next to the name of group specifies whether the group is LDAP group or Linux group. You can also view the icons detail on top of the **Groups/User List** panel.



To update the Groups/User List panel, click  button.

To get permission or to generate a report, click  next to the group name. This will display the list of permissions associated with a group or a user.



- a) The **Get Permissions for Sensitive Files in Directory** will list down all the files name with absolute path. It also displays other information such as Permissions (Read, Write, Execute), Permission Type, FP/Row Encryption, Sensitive Data, etc.

***To view other columns information scroll the bar to the right.**

PERMISSIONS FOR GROUP: MAIL ON /ANKIT			
	PERMISSIO...	PERMISSION T...	FP
		Other	false
'TestFile_Random.txt		Other	false
'StructuredTestFile.txt	r--	Other	false
<tion_AllFiles		Other	false
<tion_AllFiles/All_expressions1.avro	r--	Other	false
<tion_AllFiles/1_lakh.txt	r--	Other	false
<tion_AllFiles/700mb.txt	r--	Other	false
<tion_AllFiles/TestFile_Field_Enc.txt	r--	Other	false
<tion_AllFiles/detection_ml_ccno_di...	r--	Other	false
<tion_AllFiles/P_SEQ_FP_Enc	r--	Other	false
<tion_AllFiles/Parquet_comma	r--	Other	false
<tion_AllFiles/RCFile_FieldEnc	r--	Other	false
<tion_AllFiles/ORCFile_FieldEnc	r--	Other	false
		Other	false
'TestFile_Random.txt		Other	false
'StructuredTestFile.txt	r--	Other	false

185. The **Get Permission For Selected File** option will list down all the file names with absolute path for which a user has user permission to read, write and execute. The information includes the path of file, permissions and permission type.

PERMISSIONS FOR USER: ADM ON /KARMAN/AVRO/EMPTY_EXAMPLE.AVRO		
PATH	PERMISSIONS	PERMISSION TYPE
/karman/avro/empty_example.avro		Other

186. The **Generate Summarized PDF Report** option allows you to download the report. Upon selecting the option, a **Filter Report** popup appears.

FILTER REPORT

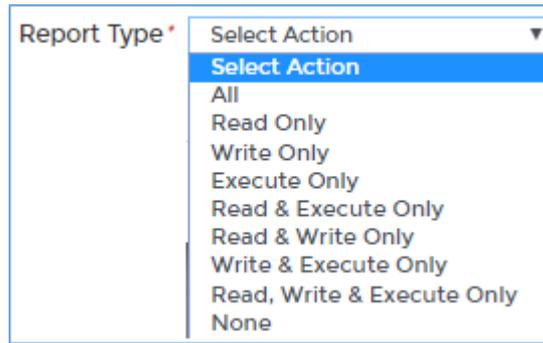
Report Type * Select Action ▼

Include Sub-Directories

✕ CANCEL
➔ GENERATE

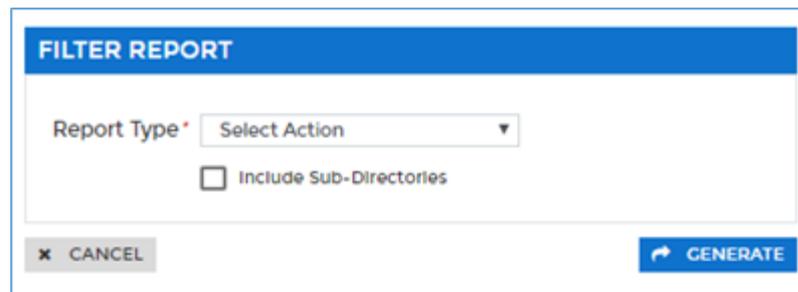
Select the option from the **Report Type** drop-down. To include the sub-directories

in the downloaded report, check the **Include Sub-Directories** checkbox.

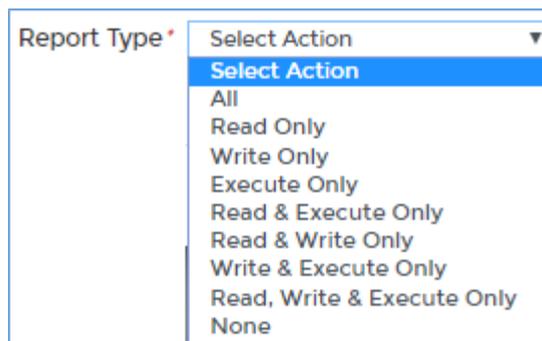


Click **Generate** button. This will download the report based on specified filters.

187. The **Generate Detailed PDF Report** option allows you to download the detailed version of report. Upon selecting the option, a **Filter Report** popup appears.



Select the option from the **Report Type** drop-down. To include the sub-directories in the downloaded report, check the **Include Sub-Directories** checkbox.



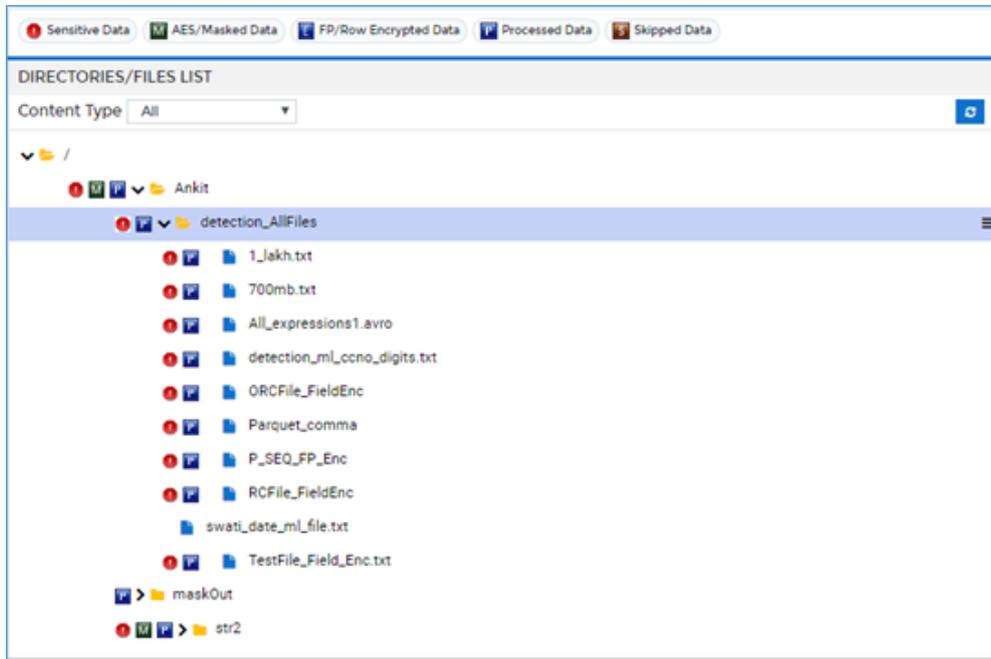
Click **Generate** button. This will download the report based on specified filters.

188. **Directories/Files List**

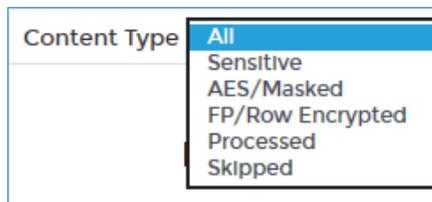
This panel will display the list of all files stored in a directory. The icons next to each file or

folder specify that the folder contains sensitive data which is processed and has been protected by some masking option.

You can view the icons detail on top of the **Directories/Files List** panel.

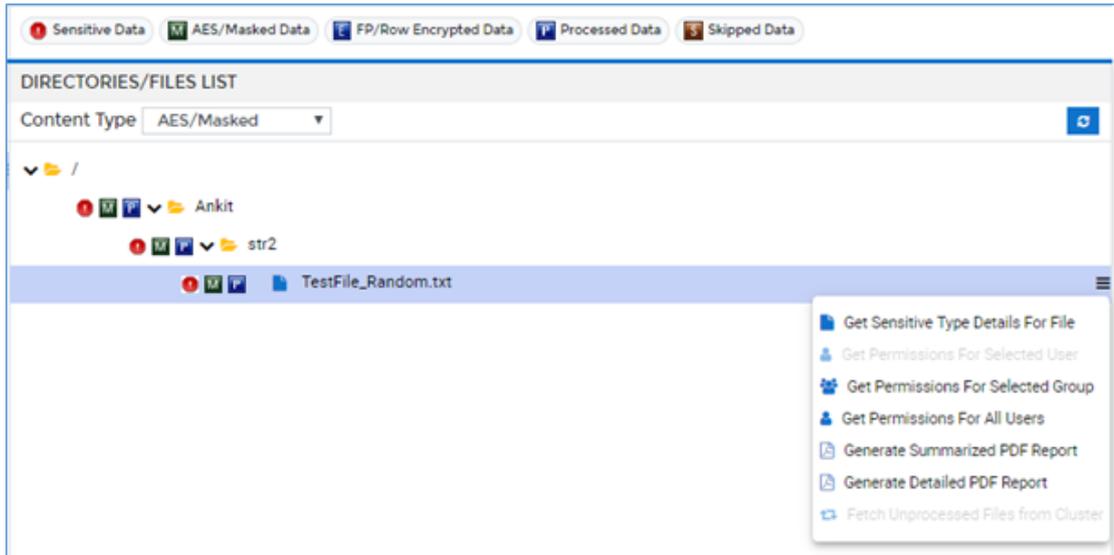


You can also filter out the Directory/File list by specifying the content from Content Type drop-down. A content type for a data stored in a file/directory can be Sensitive, AES/Masked, FP/Row Encrypted, Processed and Skipped.



To update the **Directories/Files List** panel, click  button.

To get permission or to generate a report, click  next to the directory/file name. This will display the list of permissions for the selected directory/file name.



- a) The **Get Sensitive Type Details For File** option will display list of sensitive data associated with the selected file. Upon selecting the option, the Sensitive Data popup appears.

SENSITIVE DATA				
SENSITIVE TYPE	SENSITIVE TYPE COUNT	FP/ROW ENCRYPTED	DETECTED	AES/MASKED
Credit Card # (Dash Separation)	100	false	true	true
Credit Card # (Digits Only)	100	false	true	true
Credit Card # (Space Separation)	100	false	true	true

CANCEL

- 189. The **Get Permission for the Selected User** option will display the list of all files that user can access. The Permission panel displays information such as Path, Permissions and Permission Type for that file.

PERMISSIONS FOR USER: KARMAN ON /KARMAN/AVRO/EMPTY_EXAMPLE.AVRO		
PATH	PERMISSIONS	PERMISSION TYPE
/karman/avro/empty_example.avro		Other

190. The **Get Permission for Selected Group** option will display a list of all files that can be accessed by a group of users. The Permission panel displays information for Path, type of permission for the file and the permission type (Read, Write, Execute).

PERMISSIONS FOR GROUP: HDFS ON /ANKIT/STR2		
PATH	PERMISSIO...	PERMISSION T...
/Ankit/detection_AllFiles/TestFile_Field_...		Owner

191. The **Get Permission for All Users** option will display the list of all files that all users can access. The **Permission** panel displays information such as User Name, absolute path for all listed files in a directory, Permissions, Permission Type.

PERMISSIONS FOR ALL USERS ON /KARMAN/DATA_FILES_100/ORCFILE			
USER NA...	PATH	PERMISSIO...	PERMISSION T...
	/karman/Data_Files_100/or...		Owner
-Other Use...	/karman/Data_Files_100/or...		Other

192. The **Generate Summarized PDF Report** option allows you to download the report. Upon selecting the option, a **Filter Report** popup appears.

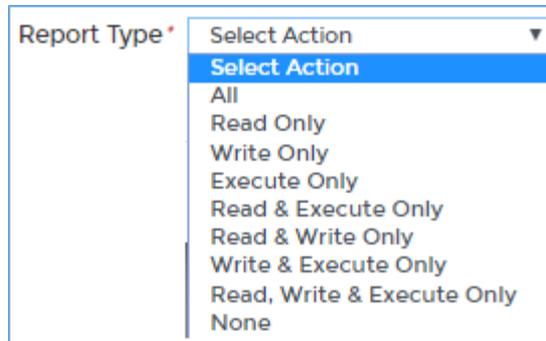
FILTER REPORT

Report Type* Select Action ▼

Include Sub-Directories

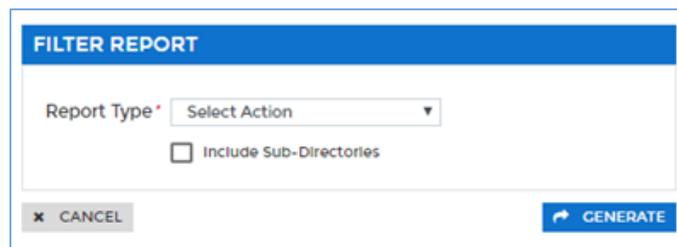
✕ CANCEL
↻ GENERATE

Select the option from the **Report Type** drop-down. To include the sub-directories in the downloaded report, check the **Include Sub-Directories** checkbox.

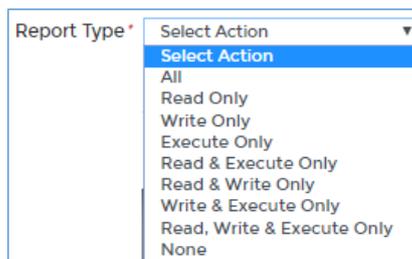


Click **Generate** button. This will download the report based on specified filters.

193. The **Generate Detailed PDF Report** option allows you to download the detailed version of report. Upon selecting the option, a **Filter Report** popup appears.

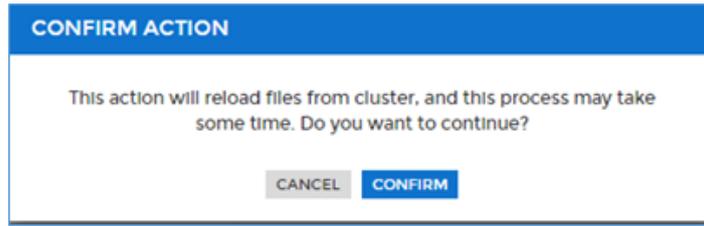


Select the option from the **Report Type** drop-down. To include the sub-directories in the downloaded report, check the **Include Sub-Directories** checkbox.



Click **Generate** button. This will download the report based on specified filters.

194. The **Fetch Unprocessed Files from Cluster** option allows you to re-load the files from cluster. Upon selecting this option, Confirm Action popup appears.

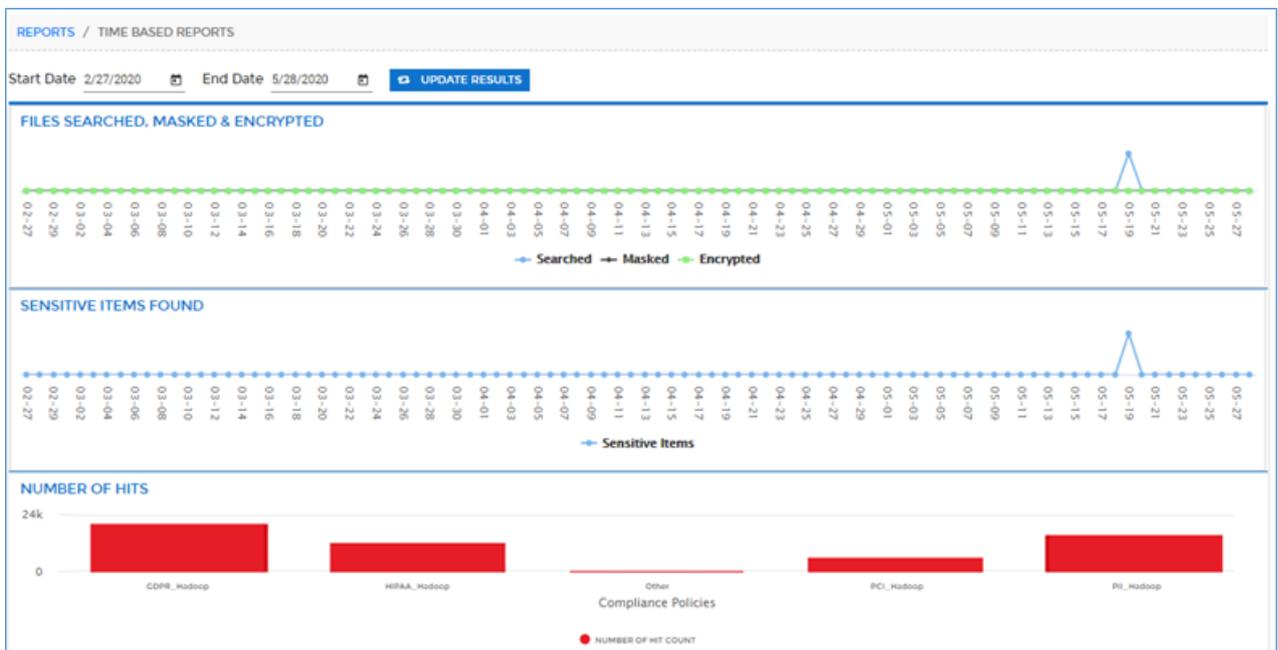


Click **Confirm** to reload the files from cluster.

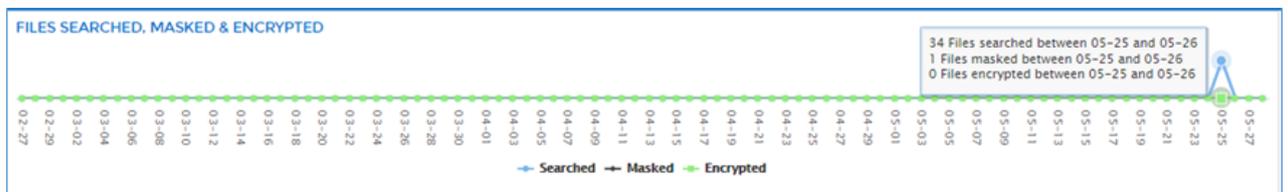
10.9.2 Time Based Report

The **Time Based Report** display results in the form of graphs based on the provided date range. This report display results for the Sensitive Data in form of graphs.

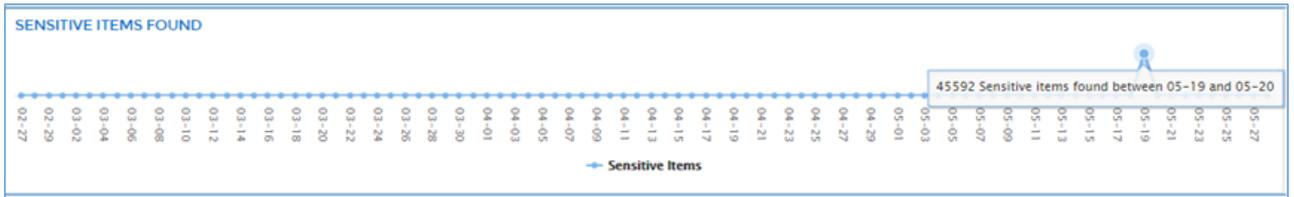
To access **Time Based Report**, click **Reports > Hadoop > Time Based Reports**.



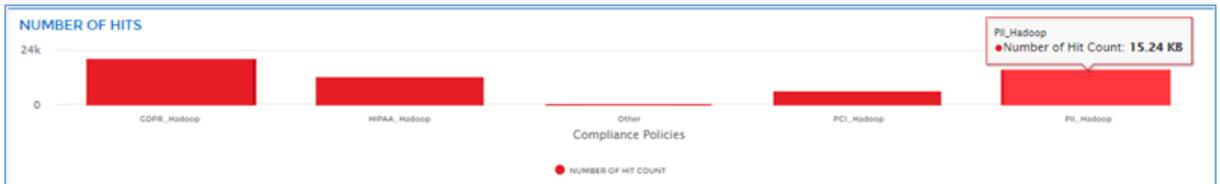
1. The first line chart displays information for the Number of **Files Searched, Masked & Encrypted** within a specified date range.



2. The second line chart display information for the **Number of Sensitive Items** found for searched date range.



3. The third bar chart display information for the **Number of Hits** found for each **Compliance Group** in Hadoop data source.



To access the detailed **Time Based Report**, click on the data point of first and second graph to generate the report. Perform the below steps to generate the tabular **Time Based Report**:

a) Select the **Start** and **End Date** in date filter.

Start Date 2/27/2020 End Date 5/28/2020 [UPDATE RESULTS](#)

b) Click the **Update Results** button. This will update the charts based for the given date range.



c) Click on the data points of first and second graph to generate detailed **Time Based Report**.

TIME BASED REPORTS: DETAILS									
TASK NAME	START TIME	FILES SEARCHED	SENSITIVE FILES	SENSITIVE ITEMS	SIZE OF DATA SEARCHED	FP/ROW ENCRYPTED	AES/MASKED	INCREMENTAL	STATUS
det_ifa	2020-05-25 07:09:16.48	8	8	49	0	false	false	false	COMPLETED
ifa_mask	2020-05-25 07:08:32.266	1	1	300	0	false	true	false	COMPLETED
ifa_det	2020-05-25 06:13:50.38	7	7	33888	0	false	false	false	COMPLETED

10.10 Exported Reports

The **Exported Reports** screen allows you to download the pdf version of the reports generated in the **Overview** and **GDPR** section of the Dashboard.

EXPORTED REPORTS					
PDF NAME	DATA SOURCE TYPE	CATEGORY	EXPORTED TIME	PROGRESS	DOWNLOAD
OperationData_exposed_DBMS_2020.05.27.22.29.28.pdf	DBMS	exposed	2020-05-27T22:32:47.216-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.14.pdf	DBMS	safe	2020-05-27T22:33:36.866-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.18.pdf	DBMS	safe	2020-05-27T22:34:10.598-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.03.pdf	DBMS	safe	2020-05-27T22:34:12.22-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.31.56.pdf	DBMS	safe	2020-05-27T22:35:17.733-05:30	100%	
OperationData_exposed_DBMS_2020.05.27.22.29.14.pdf	DBMS	exposed	2020-05-27T22:32:07.251-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.22.pdf	DBMS	safe	2020-05-27T22:33:54.315-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.32.44.pdf	DBMS	safe	2020-05-27T22:35:36.585-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.33.25.pdf	DBMS	safe	2020-05-27T22:36:06.924-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.33.54.pdf	DBMS	safe	2020-05-27T22:36:26.383-05:30	100%	

This screen will display the information for the reports such as PDF Name, Data Source Type, Category, Export Time, Progress of the report, and button in the **Download** column. To refresh the **Exported Reports** screen, click .

To download the reports, perform the below steps:

1. To access the Exported Reports section, click Reports > Exported Reports.

EXPORTED REPORTS					
PDF NAME	DATA SOURCE TYPE	CATEGORY	EXPORTED TIME	PROGRESS	DOWNLOAD
OperationData_exposed_DBMS_2020.05.27.22.29.28.pdf	DBMS	exposed	2020-05-27T22:32:47.216-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.14.pdf	DBMS	safe	2020-05-27T22:33:36.866-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.18.pdf	DBMS	safe	2020-05-27T22:34:10.598-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.03.pdf	DBMS	safe	2020-05-27T22:34:12.22-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.31.56.pdf	DBMS	safe	2020-05-27T22:35:17.733-05:30	100%	
OperationData_exposed_DBMS_2020.05.27.22.29.14.pdf	DBMS	exposed	2020-05-27T22:32:07.251-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.30.22.pdf	DBMS	safe	2020-05-27T22:33:54.315-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.32.44.pdf	DBMS	safe	2020-05-27T22:35:36.585-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.33.25.pdf	DBMS	safe	2020-05-27T22:36:06.924-05:30	100%	
OperationData_safe_DBMS_2020.05.27.22.33.54.pdf	DBMS	safe	2020-05-27T22:36:26.383-05:30	100%	

2. Click in **Download** column to download the report in PDF format.

11 Monitor

DgSecure Monitor allows users to set alerts around data handling and/or data access in HFDS, S3, MapR-FS, Hive, Oracle, or Teradata. Rules can be set around several conditions including specific users, specific data type groups, specific systems, a specific command, source path, or destination path. Rules can be set using one or all of these conditions.

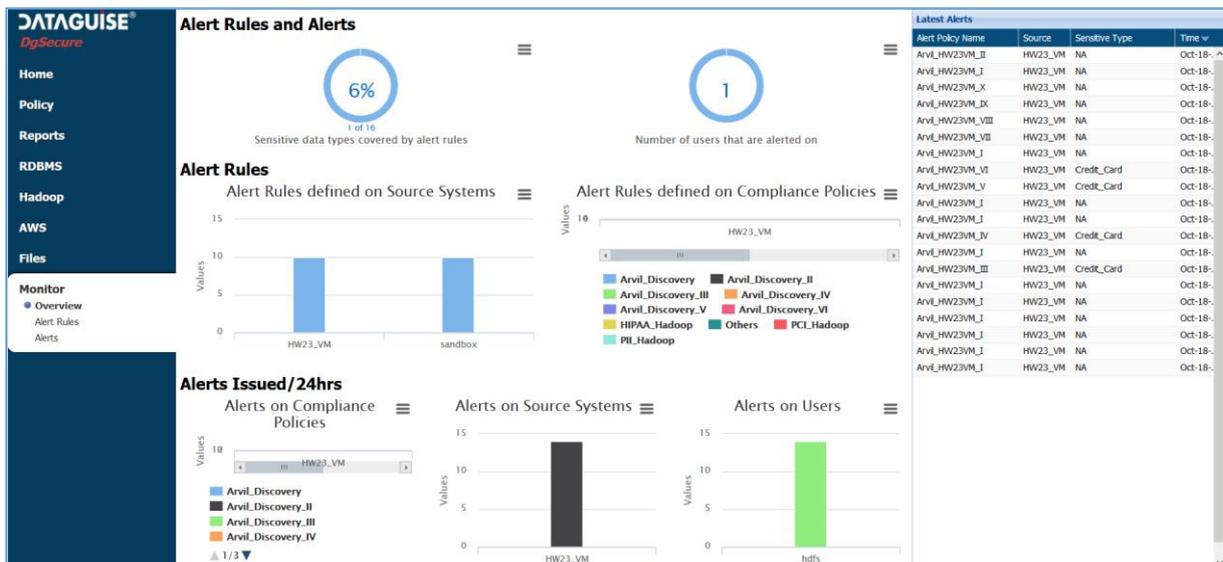
The alerts are based off the targeted platform’s log files. Set alert rules on the **Alert Rules** page. Review triggered alerts on the **Alerts** page. The **Monitor Overview** page provides a comprehensive overview. In order to most effectively utilize DgSecure Monitor, DgSecure detection tasks need to be run in order to identify where sensitive data resides.

The Monitor Overview page provides a centralized location to track the monitoring status of known sensitive data. Monitoring capabilities are broken down according to the percentage of sensitive types covered by alert rules, the number of people who receive alert notifications, source systems, related DgSecure policies, and alerts issued over the past 24 hours.

11.1 Overview

The **Overview** page offers a birds-eye view of the current monitoring statuses of known sensitive data.

Access the **Overview** page from the menu under **Monitor > Overview**. The **Overview** page is divided into two panels which are described below.



The Monitor Overview page provides a centralized location to track the monitoring status of known sensitive data. The majority of the page is a series of graphs showing the percentage of sensitive types covered by alert rules, the number of people who trigger alert notifications, source systems, and related DgSecure policies.



A panel on the right side of the page shows the most recently triggered alerts.

Alert Policy Name	Source	Sensitive Type	Time
Arvil_HW23VM_II	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_X	HW23_VM	NA	Oct-18-
Arvil_HW23VM_IX	HW23_VM	NA	Oct-18-
Arvil_HW23VM_VIII	HW23_VM	NA	Oct-18-
Arvil_HW23VM_VII	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_VI	HW23_VM	Credit_Card	Oct-18-
Arvil_HW23VM_V	HW23_VM	Credit_Card	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_IV	HW23_VM	Credit_Card	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_III	HW23_VM	Credit_Card	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-
Arvil_HW23VM_I	HW23_VM	NA	Oct-18-

11.2 Alert Rules

The **Alert Rules** page offers in-depth details about existing alert rules that the user has permission to see.

Access the **Alert Rules** page from the menu under **Monitor > Alert Rules**. The **Alert Rules** page is divided into two panels which are described below.

Alert Rules Summary

Alert Rule Name	Source System	Authorization Status	Status	Created On
ArviAlertRule_2	HW23_VM	true	<input checked="" type="radio"/>	Aug-09-2016 17:07:58
Dgm_Arvi_For_Monitor	sandbox	false	<input type="radio"/>	Aug-09-2016 11:57:30
DGM_CREDIT_CARD	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_SOCIAL_SECURITY	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_NAMES	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_ADDRESS	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_EMAIL_ADDRESS	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_IP_ADDRESS	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_TELEPHONE	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_URL	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_NPI	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51
DGM_ABA_ROUTING_NUMBER	sandbox	false	<input type="radio"/>	Aug-09-2016 08:55:51

The top panel shows four number sets: alerts triggered, sensitive type groups actively monitored, source systems monitored, and active alerts.

Alert Rules Summary

Alert Rules	Sensitive Type Groups	Source Systems	Alert Rules
2 Not Triggerged, 7 Triggerged	0 Monitored, 8 Not Monitored	1 Monitored, 0 Not Monitored	1 Active, 8 Inactive

The bottom panel shows either all alert rules, default alert rules, or custom alert rules. Whichever table displays, the details of the selected alert rule type (custom, default, or all) are shown. To create a new alert rule, click the **New Alert Rule** button. To edit an existing alert rule, select the desired rule and click **Edit Alert Rule**. The status column shows either a white circle, black circle, or red circle with an exclamation point. A black circle indicates the alert rule is on. A white circle indicates the alert rule is off. A red circle with exclamation point indicates the alert rule is on and has been triggered.

Alert Rules								
New Alert Rule Edit Alert Rule								
All Rules Default Rules Custom Rules								
Alert Rule Name	Source System	Sensitive Type Group	Command	Allowed	Resource	User	Status	Created On
DGM_CREDIT_CARD	sandbox	Credit Card		false			<input type="radio"/>	Jul-06-2016 00:06:49
DGM_NAMES	sandbox	Names		false			<input type="radio"/>	Jul-06-2016 00:06:49
DGM_ADDRESS	sandbox	Address		false			<input type="radio"/>	Jul-06-2016 00:06:49
DGM_EMAIL_ADDRESS	sandbox	Email Address		false			<input type="radio"/>	Jul-06-2016 00:06:49
DGM_IP_ADDRESS	sandbox	IP Address		false			<input type="radio"/>	Jul-06-2016 14:46:54
DGM_TELEPHONE	sandbox	Telephone		false			<input type="radio"/>	Jul-06-2016 00:06:49
DGM_URL	sandbox	URL		false			<input type="radio"/>	Jul-06-2016 00:06:49
DGM_NPI	sandbox	NPI		false			<input type="radio"/>	Jul-06-2016 00:06:49
Arvi_Command_AI	sandbox		rename		/ArviRa/D33, /ArviRa/...	hdfs	<input checked="" type="radio"/>	Jul-06-2016 15:05:32

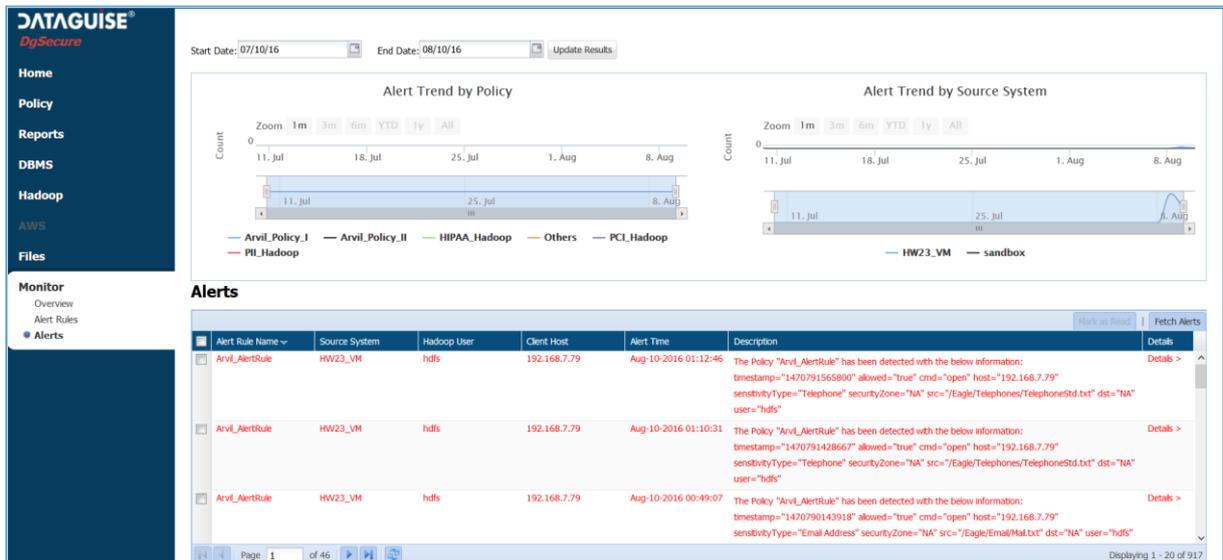
When the **New Alert Rule** button is clicked, a new page displays. Fill out the information as needed. Only fields marked with a red asterisk are mandatory. Multiple values per condition type can be set. Related DgSecure policies are automatically calculated based on the selected sensitive types. Selecting the “Notification De-Duplication” checkbox displays another field that allows the user to select a timeframe in minutes. When selected, only one rule is generated if an alert condition is met multiple times within the timeframe.

***Note:** When using “regex” as an operator for an alert rule, the regex pattern must follow standard regex patterns.

11.3 Alerts

The **Alerts** page offers in-depth details about triggered alerts that the user has permission to see.

Access the **Alerts** page from the menu under **Monitor > Alerts**. The **Alerts** page is divided into two panels which are described below.



The top panel shows two graphs. The first graph displays the number of triggered alerts by policy. Hover the cursor over a policy plot point to see how many times an alert associated with the policy has been triggered. The second graph displays the number of alerts by source system. Hover the cursor over a plot point to see how many times alerts for a specific system have been triggered.



The bottom panel shows the details of triggered alerts including source system, Hadoop user, time, and description of the violated rule.

Alerts Mark as Read | Fetch Alerts

Alert Rule Name	Source System	Hadoop User	Client Host	Alert Time	Description	Details
<input type="checkbox"/> Arvi_AlertRule	HW23_VM	hdfs	192.168.7.79	Aug-10-2016 01:12:46	The Policy "Arvi_AlertRule" has been detected with the below information: timestamp="1470791565800" allowed="true" cmd="open" host="192.168.7.79" sensitivityType="Telephone" securityZone="NA" src="/Eagle/Telephones/TelephoneStd.txt" dst="NA" user="hdfs"	Details >
<input type="checkbox"/> Arvi_AlertRule	HW23_VM	hdfs	192.168.7.79	Aug-10-2016 01:10:31	The Policy "Arvi_AlertRule" has been detected with the below information: timestamp="1470791428667" allowed="true" cmd="open" host="192.168.7.79" sensitivityType="Telephone" securityZone="NA" src="/Eagle/Telephones/TelephoneStd.txt" dst="NA" user="hdfs"	Details >
<input type="checkbox"/> Arvi_AlertRule	HW23_VM	hdfs	192.168.7.79	Aug-10-2016 00:49:07	The Policy "Arvi_AlertRule" has been detected with the below information: timestamp="1470790143918" allowed="true" cmd="open" host="192.168.7.79" sensitivityType="Email Address" securityZone="NA" src="/Eagle/Email/Mal.txt" dst="NA" user="hdfs"	Details >

Page 1 of 46 Displaying 1 - 20 of 917

12 Attributes

12.1 Concept

Attributes classify databases on a network allowing for quicker identification and enhanced insights in the DBMS and Hadoop & Files Dashboard. The Attribute page displays all attribute groups and the attributes that compose them. Once attribute is created, you can assign the attributes using Attribute Assignment section in the DgSecure.

The below sections explains the process of creating, editing an attribute in the Attribute section and assigning attribute in the Attribute Assignment section.

12.2 Attribute

This section will explain the process of creating, editing and deleting an Attribute in DBMS and Hadoop & Files section.

12.2.1 Create an Attribute

12.2.1.1 DBMS

In DBMS, attribute classify databases on a network allowing for a quicker identification and enhanced insights in the DBMS dashboard.

To access the Attribute page. Click **Policy > DBMS > Attribute**.

12.2.1.1.1 Create an Attribute Group

To create an Attribute, you need to first create an Attribute Group.

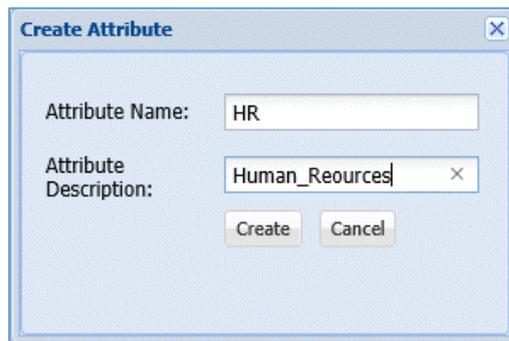
The below image shows the user interface for creating an Attribute Group.

- **Add Group:** Click the Add Group button to create DB group.
- **Group Name:** Enter the unique name for the group.
- **Group Description:** Enter the description for the group.
- **Create:** Click the Create button, if you want to create the DB group.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.1.1.2 Create an Attribute

To create an Attribute, follow the below steps.

The below image shows the user interface for creating an Attribute.



- **Add Attribute:** Click the Add Attribute button to define the Attribute Name and Attribute Description.
- **Attribute Name:** Enter the unique name for the attribute.
- **Attribute Description:** Enter the description for the attribute.
- **Create:** Click the Create button, if you want to save the changes.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.1.2 Hadoop & Files

In Hadoop & Files, attribute classify databases on a network allowing for a quicker identification and enhanced insights in the Hadoop dashboard.

To access the Attribute page. Click **Policy > Hadoop & Files > Attribute**.

12.2.1.2.1 Create an Attribute Group

To create an Attribute, you need to first create an Attribute Group.

The below image shows the user interface for creating an Attribute.



- **Add Group:** Click the Add Group button to create DB group.
- **Group Name:** Enter the unique name for the group.
- **Group Description:** Enter the description for the group.
- **Create:** Click the Create button, if you want to create the DB group.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.1.2.2 Create an Attribute

To create an Attribute, follow the below steps.

The below image shows the user interface for creating an Attribute.

- **Add Attribute:** Click the Add Attribute button to define the Attribute Name and Attribute Description.
- **Attribute Name:** Enter the unique name for the attribute.
- **Attribute Description:** Enter the description for the attribute.
- **Create:** Click the Create button, if you want to save the changes.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.2 Edit an Attribute

12.2.2.1 DBMS

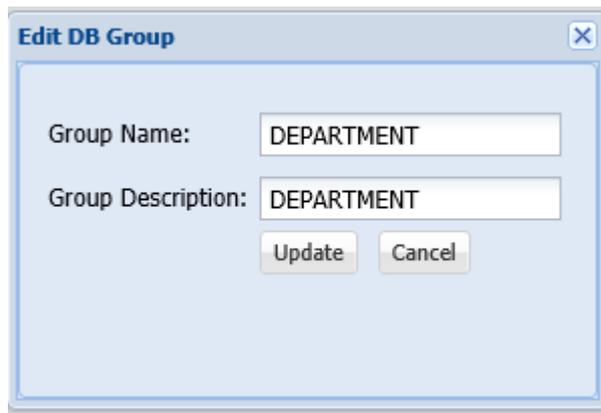
To edit an Attribute. Select the attribute which you wish to edit.

Click POLICY > **DBMS** > **Attribute**> **Edit Group**.

12.2.2.1.1 Edit an Attribute Group

To edit an Attribute Group, follow the below steps.

The below image shows the user interface for editing the information.

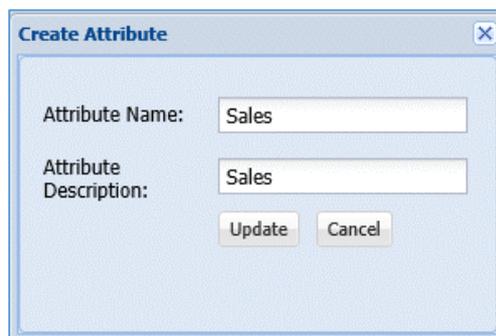


- **Edit Group:** Click the Edit Group button, if you want to edit the group name and description.
- **Group Name:** Edit the name of the group.
- **Group Description:** Edit the description for the Group Name.
- **Update:** Click the Update button, if you want to save the changes with new Group Name.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.2.1.2 Edit an Attribute

To edit an Attribute, follow the below steps.

The below image shows the user interface for editing an attribute.



- **Edit Attribute:** Click the Edit Attribute button to edit the Attribute Name and Attribute Description.
- **Attribute Name:** Edit the name of the attribute.
- **Attribute Description:** Edit the description for the attribute.
- **Update:** Click the Update button, if you want to save the changes.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.2.2 Hadoop & Files

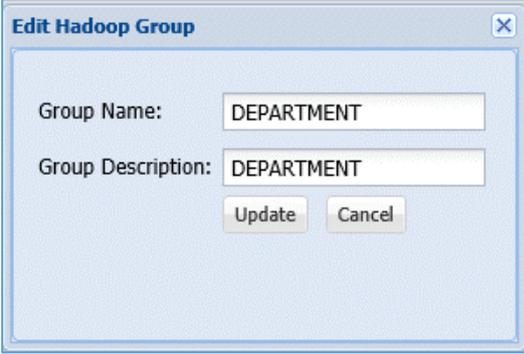
To edit an Attribute. Select the attribute which you wish to edit.

Click POLICY > **DBMS** > **Attribute**> **Edit Group**.

12.2.2.2.1 Edit an Attribute Group

To edit an Attribute Group, follow the below steps.

The below image shows the user interface for editing the information.

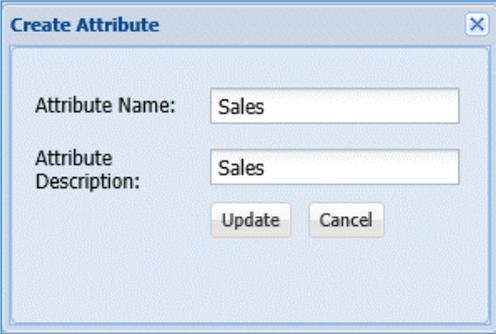


- **Edit Group:** Click the Edit Group button, if you want to edit the group name and description.
- **Group Name:** Edit the name of the group.
- **Group Description:** Edit the description for the Group Name.
- **Update:** Click the Update button, if you want to save the changes with new Group Name.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.2.2.2 Edit an Attribute

To edit an Attribute, follow the below steps.

The below image shows the user interface for editing an attribute.



- **Edit Attribute:** Click the Edit Attribute button to edit the Attribute Name and Attribute Description.
- **Attribute Name:** Edit the name of the attribute.
- **Attribute Description:** Edit the description for the attribute.
- **Update:** Click the Update button, if you want to save the changes.
- **Cancel:** Click the Cancel button, if you do not want to save the changes.

12.2.3 List an Attribute

12.2.3.1 DBMS

In DBMS, this page displays all attribute groups and the attributes associated with the groups.

The below image shows the user interface of the Attribute page.

The screenshot shows the 'Manage DB Attributes' interface. It features two main sections, each with a table and control buttons.

Attribute Groups Table:

Group Name	Group Description	Created At
LOB	Line of Business	20 Dec 2019 07:43:55 GMT
APPLICATION	APPLICATION	20 Dec 2019 07:43:55 GMT
REGION	REGION	20 Dec 2019 07:43:55 GMT
DEPARTMENT	DEPARTMENT	20 Dec 2019 07:43:55 GMT

Attributes Table:

Name	Attribute Description	Attribute Parent
Marketing	Market	
Sales	Sales_dpt	
HR	Human_resource	

- Manage DB Attribute:** This pane displays all the existing attribute groups. These groups organize the attributes but cannot themselves be used as a filter. It displays the Group Name, Description and the Creation Date.

You can also view the list of all attributes associated to the Group.

This screenshot shows the 'Manage DB Attributes' interface, focusing on the attribute groups table. The table is identical to the one in the previous screenshot.

Group Name	Group Description	Created At
LOB	Line of Business	20 Dec 2019 07:43:55 GMT
APPLICATION	APPLICATION	20 Dec 2019 07:43:55 GMT
REGION	REGION	20 Dec 2019 07:43:55 GMT
DEPARTMENT	DEPARTMENT	20 Dec 2019 07:43:55 GMT

- Add Group:** Click the Add Group button to define the Group Name and Description.
- Edit Group:** Click the Edit Group button to edit the information.

- ❖ **Delete Group:** Select the Attribute Group which you want to delete. Click the Delete Group button.

- **Attribute List:** The Attribute list is the bottom pane in the Attribute Page. It will list down all attributes which are associated with the selected Attribute Group. It displays the information such as Name of the Attribute, Attribute Description, Attribute Parent.

Name	Attribute Description	Attribute Parent
Marketing	Marketing	
Sales	Sales	
HR	Human_Resource	

- ❖ **Add Attribute:** Click the Add Attribute button to define the Attribute Name and Attribute Description.
- ❖ **Edit Attribute:** Click the Edit Attribute button to edit the information.
- ❖ **Delete Attribute:** Select the Attribute which you want to delete. Click the Delete Attribute button.

12.2.3.2 Hadoop & Files

In DBMS, this page displays all attribute groups and the attributes associated with the groups.

The below image shows the user interface of the Attribute page.

The screenshot shows the 'Manage DB Attributes' interface. It consists of two main sections. The top section is a table of attribute groups with columns for Group Name, Group Description, and Created At. The bottom section is a table of attributes with columns for Name, Attribute Description, and Attribute Parent.

Group Name	Group Description	Created At
LOB	Line of Business	20 Dec 2019 07:43:55 GMT
APPLICATION	APPLICATION	20 Dec 2019 07:43:55 GMT
REGION	REGION	20 Dec 2019 07:43:55 GMT
DEPARTMENT	DEPARTMENT	20 Dec 2019 07:43:55 GMT

Name	Attribute Description	Attribute Parent
Marketing	Market	
Sales	Sales_dpt	
HR	Human_resource	

- Manage DB Attribute:** This pane displays all the existing attribute groups. These groups organize the attributes but cannot themselves be used as a filter. It displays the Group Name, Description and the Creation Date.

You can also view the list of all attributes associated to the Group.

This screenshot shows the top portion of the 'Manage DB Attributes' interface, specifically the table of attribute groups. It includes the same table as shown in the previous screenshot.

Group Name	Group Description	Created At
LOB	Line of Business	20 Dec 2019 07:43:55 GMT
APPLICATION	APPLICATION	20 Dec 2019 07:43:55 GMT
REGION	REGION	20 Dec 2019 07:43:55 GMT
DEPARTMENT	DEPARTMENT	20 Dec 2019 07:43:55 GMT

- Add Group:** Click the Add Group button to define the Group Name and Description.
- Edit Group:** Click the Edit Group button to edit the information.
- Delete Group:** Select the Attribute Group which you want to delete. Click the Delete Group button.

- Attribute List:** The Attribute list is the bottom pane in the Attribute Page. It will list down all attributes which are associated with the selected Attribute Group. It displays

the information such as Name of the Attribute, Attribute Description, Attribute Parent.

Name	Attribute Description	Attribute Parent
Marketing	Marketing	
Sales	Sales	
HR	Human_Resource	

- ❖ **Add Attribute:** Click the Add Attribute button to define the Attribute Name and Attribute Description.
- ❖ **Edit Attribute:** Click the Edit Attribute button to edit the information.
- ❖ **Delete Attribute:** Select the Attribute which you want to delete. Click the Delete Attribute button.

12.3 Attribute Assignment

Attributes group databases on a network to allow for quicker identification and enhanced reporting. The Attribute Assignment page is where the user can apply attributes to the desired databases.

12.3.1 RDBMS

To assign an attribute to Database/Schema, follow the below steps.

The below image shows the user interface of the Apply Attribute page.

Database/Schema	IP Address	Port No	Vendor	Contact Name	Contact Email	Production	Attributes Applied	New Attributes
vb	192.168.0...	1433	SQL Ser...	Admin	admin@datagui.se.com	<input type="checkbox"/>	Marketing X	
vb	192.168.0...	0	SQL Ser...			<input type="checkbox"/>	Terr X	DEPARTMENT
USER_65SDATA	192.168.0...	0	Oracle			<input type="checkbox"/>	Sales X	Marketing
USER_655	192.168.0...	0	Oracle			<input type="checkbox"/>		Sales
TEST_USER	192.168.0...	0	Oracle			<input checked="" type="checkbox"/>		HR
TESTUSER	192.168.0...	0	Oracle			<input type="checkbox"/>		REGION
TAJMASK2	192.168.0...	0	Oracle			<input type="checkbox"/>		Terr
TAJMASK	192.168.0...	0	Oracle			<input type="checkbox"/>		ZIP
TAJJINDER	192.168.0...	0	Oracle			<input type="checkbox"/>		Click here to add attributes
SUKH	192.168.0...	0	Oracle			<input type="checkbox"/>		Click here to add attributes
SS_GDPR	192.168.0...	1433	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
SS_GDPR	192.168.0...	0	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
sssss	192.168.0...	1433	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
sssss	192.168.0...	0	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
SP_DGHDFSNF...	192.168.0...	0	Oracle			<input type="checkbox"/>		Click here to add attributes
SP_DGDASHBOA...	192.168.0...	0	Oracle			<input type="checkbox"/>		Click here to add attributes
SP_DGCONTROL...	192.168.0...	0	Oracle			<input type="checkbox"/>		Click here to add attributes
SP_DGCONTROL...	192.168.0...	0	Oracle			<input type="checkbox"/>		Click here to add attributes
sonam_json	192.168.0...	1433	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
sonam_json	192.168.0...	0	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
sonam_GDPR	192.168.0...	1433	SQL Ser...			<input type="checkbox"/>		Click here to add attributes
sonam GDPR	192.168.0...	0	SQL Ser...			<input type="checkbox"/>		Click here to add attributes

- **New Attributes:** select the attribute name from the New Attribute drop-down. This functionality allows you to add tags in the Attributes Applied column.

To remove tag, click the 'x' next to the tag name.

To create an Attribute Group and Attribute, click section [1.2 Attribute](#)

- **Contact Name:** Enter the name of the person next to the database under the Contact Name column.

Database/Schema	IP Address	Port No	Vendor	Contact Name	Contact Email	Production	Attributes Applied	New Attributes
vb	192.168.0.151	1433	SQL Server	Admin		<input type="checkbox"/>		Click here to add attributes
vb	192.168.0.151	0	SQL Server	ABC		<input type="checkbox"/>		Click here to add attributes
USER_655DATA	192.168.0.155	0	Oracle	Enter Name		<input type="checkbox"/>		Click here to add attributes
USER_655	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TEST_USER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TESTUSER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TAJMASK2	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TAJMASK	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TAJINDER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes

- **Contact Email:** Enter the email address next to the database under the Contact Email Column.

Database/Schema	IP Address	Port No	Vendor	Contact Name	Contact Email	Production	Attributes Applied	New Attributes
vb	192.168.0.151	1433	SQL Ser...	Admin	admin@dataguisse.com	<input type="checkbox"/>		Click here to add attributes
vb	192.168.0.151	0	SQL Ser...	ABC	abc@xyc.com	<input type="checkbox"/>		Click here to add attributes
USER_655DATA	192.168.0.155	0	Oracle	Tajinder	Tajinder@dataguisse.com	<input checked="" type="checkbox"/>		Click here to add attributes
USER_655	192.168.0.155	0	Oracle		Enter Email	<input type="checkbox"/>		Click here to add attributes
TEST_USER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TESTUSER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TAJMASK2	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes

- **Save Attributes:** Click the Save Attribute button, if you want to save the changes.
- **Production:** Check the Production checkbox to specify that the selected database/schema is the production database.

Database/Schema	IP Address	Port No	Vendor	Contact Name	Contact Email	Production	Attributes Applied	New Attributes
vb	192.168.0.151	1433	SQL Ser...	Admin	admin@dataguisse.com	<input type="checkbox"/>		Click here to add attributes
vb	192.168.0.151	0	SQL Ser...	ABC	abc@xyc.com	<input type="checkbox"/>		Click here to add attributes
USER_655DATA	192.168.0.155	0	Oracle	Tajinder	Tajinder@dataguisse.com	<input checked="" type="checkbox"/>		Click here to add attributes
USER_655	192.168.0.155	0	Oracle		Enter Email	<input type="checkbox"/>		Click here to add attributes
TEST_USER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TESTUSER	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes
TAJMASK2	192.168.0.155	0	Oracle			<input type="checkbox"/>		Click here to add attributes

- **Apply Attributes:** Use the Apply Attribute pane to assign tags or descriptors to database/schema. You can also use these tags to organize DBMS scan reports. This pane will display basic information such as Database/Schema name, IP

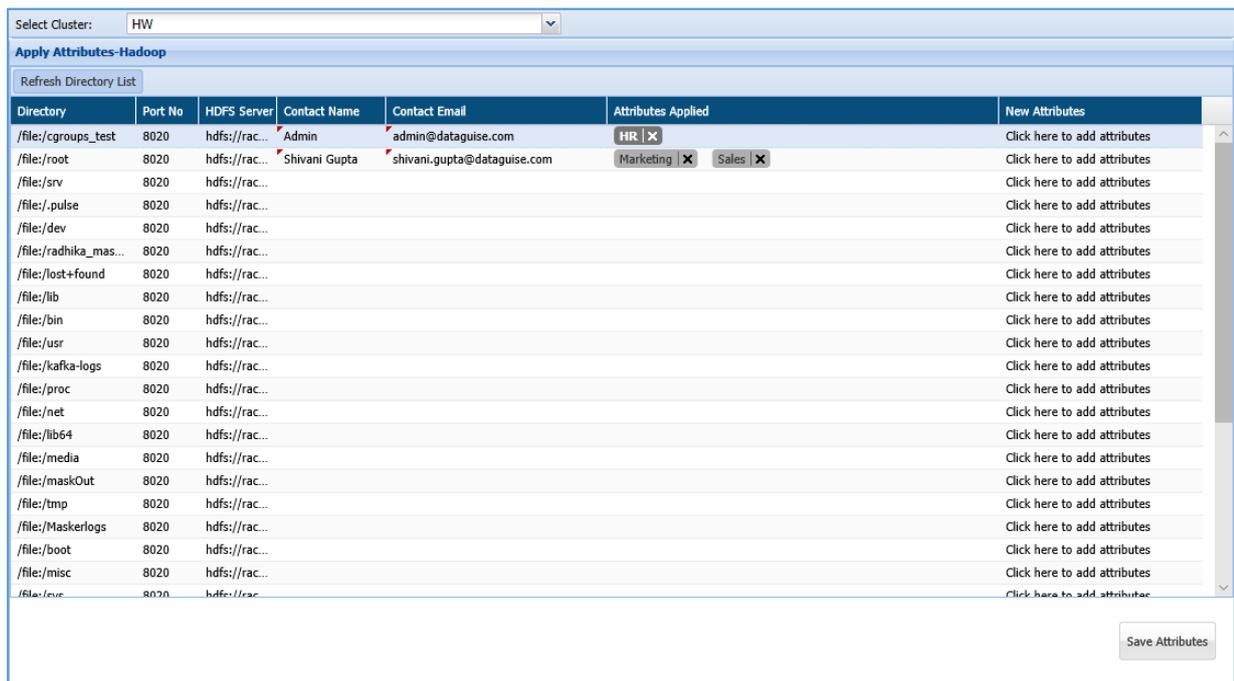
Address, Port No, Vendor, Contact Name, Contact Email, Production, Attributes Applied, New Attributes.

- **Refresh Database List:** Click the Refresh Database List button. It will update the current page with new information.

12.3.2 Hadoop

To assign an attribute to Database/Schema, follow the below steps.

The below image shows the user interface of the Apply Attribute page.



- **New Attributes:** select the attribute name from the New Attribute drop-down. This functionality allows you to add tags in the Attributes Applied column.

To remove tag, click the 'x' next to the tag name.

To create an Attribute Group and Attribute, click section [1.2 Attribute](#)

- **Contact Name:** Enter the name of the person next to the database under the Contact Name column.

Select Cluster: HW

Apply Attributes-Hadoop

Refresh Directory List

Directory	Port No	HDFS Server	Contact Name	Contact Email	Attributes Applied	New Attributes
/file/cgroups_test	8020	hdfs://rack160-hdp26...	Admin	admin@dataguis...	HR X	Click here to add attributes
/file/root	8020	hdfs://rack160-hdp26...	Shivani Gupta	shivani.gupta@d...	Marketing X Sales X	Click here to add attributes
/file/srv	8020	hdfs://rack160-hdp26...	<input type="text" value="Enter Name"/>			Click here to add attributes
/file:/pulse	8020	hdfs://rack160-hdp26...				Click here to add attributes
/file:/dev	8020	hdfs://rack160-hdp26...				Click here to add attributes
/file:/radhika_mask...	8020	hdfs://rack160-hdp26...				Click here to add attributes
/file:/lost+found	8020	hdfs://rack160-hdp26...				Click here to add attributes
/file:/lib	8020	hdfs://rack160-hdp26...				Click here to add attributes

- **Contact Email:** Enter the email address next to the database under the Contact Email Column.

Apply Attributes-Hadoop

Refresh Directory List

Directory	Port No	HDFS Server	Contact Name	Contact Email	Attributes Applied	New Attributes
/file/cgroups_test	8020	hdfs://rack160-hdp26...	Admin	admin@dataguis...	HR X	Click here to add attributes
/file/root	8020	hdfs://rack160-hdp26...	Shivani Gupta	shivani.gupta@d...	Marketing X Sales X	Click here to add attributes
/file/srv	8020	hdfs://rack160-hdp26...				Click here to add attributes
/file:/pulse	8020	hdfs://rack160-hdp26...				Click here to add attributes
/file:/dev	8020	hdfs://rack160-hdp26...				Click here to add attributes

- **Save Attributes:** Click the Save Attribute button, if you want to save the changes.
- **Apply Attributes - Hadoop:** Use the Apply Attribute pane to assign tags or descriptors to database/schema. You can also use these tags to organize DBMS scan reports. This pane will display basic information such as Database/Schema name, IP Address, Port No, Vendor, Contact Name, Contact Email, Production, Attributes Applied, New Attributes.
- **Refresh Database List:** Click the Refresh Database List button. It will update the current page with new information.

13 Structure Management

13.1 Concept

This feature enables user to apply masking on a database without employing detection to find sensitive data. If you know which specific column within a database contain sensitive data, then you can define a structure based on which masking is deployed.

Currently, DgSecure supports structure definitions for Text, Sequence, Avro, RC/ORC and Parquet Files. For RC/ORC files, DgSecure also supports masking/encryption within unstructured columns.

13.2 Create a Structure

This section will explain the process of creating a structure.

13.2.1 RDBMS

In RDBMS, first you need to create a structure and then you can add the column for a masking. You can also add new databases, edit or delete the existing databases.

13.2.1.1 Create a Structure

To create a structure. Click **RDBMS > Structure Management > Add Structure**.

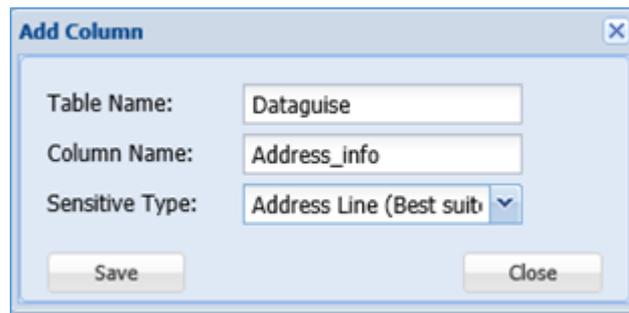
The below image shows the user interface for creating a structure.

- **Add Structure:** Click the Add Structure button to define the Structure Name and Description.
- **Structure Name:** Enter the name of the structure.
- **Description:** Enter the description for the structure name.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

13.2.1.2 Add Column info for a Structure

To add a column for masking. Click **RDBMS > Structure Management > Add Column**.

The below image shows the user interface for adding a column for masking.

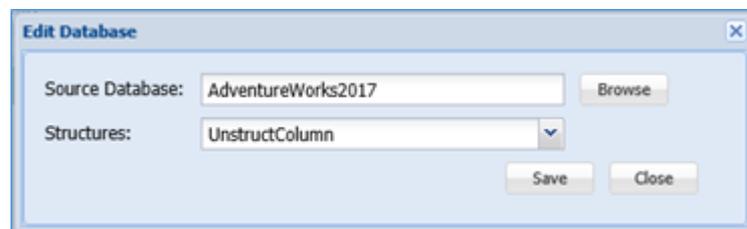


- **Add Column:** Click the Add Column button to specify the Table Name, Column Name and Sensitive Type for masking.
- **Table Name:** Enter the name of the table.
- **Column Name:** Enter the name of the column which need to be masked.
- **Sensitive Type:** Select the Sensitive Type from the drop-down.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

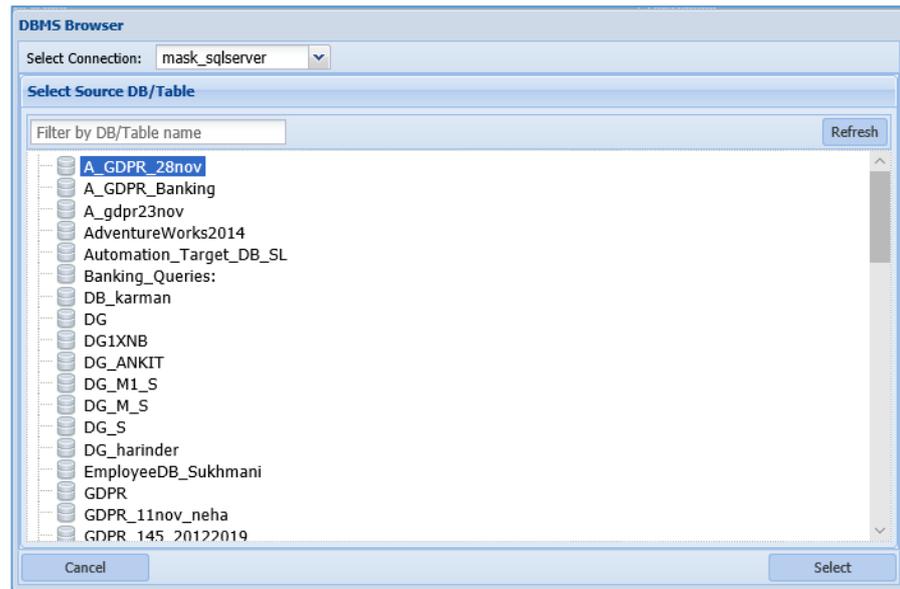
13.2.1.3 Create a Databases

To create a database. Click **RDBMS > Structure Management > Add Database**.

The below image shows the user interface of the creating a database.



- **Add Database:** Click the Add Database button to specify the Source Database Name and the Structure.
- **Source Database:** Enter the name of the database or you can click the Browse button to search for database.
- **Structures:** Select the structure from the Structure drop-down.
- **Browse:** Click the Browse button to search for the database or you can manually enter the name of the database.



1. **Select Connection:** Select the connection from the Select Connection drop-down.
2. Select the DB/Table from the Select Source DB/Table pane.
3. **Refresh:** Click the Refresh button to update the current page with updated information.
4. **Cancel:** Click the Cancel button, if you do not want to save the changes.
5. **Select:** Click the Select button, if you want to save the changes.

- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

13.2.2 Hadoop

Hadoop currently runs against text files (and compressed text files), Sequence files, RC/ORC and Avro files

13.2.2.1 Text Structure

[Create a Structure](#)

To create a structure. Click **Hadoop > Structure Management > Text Files > Add Structure**.

The below image shows the user interface for creating a structure.

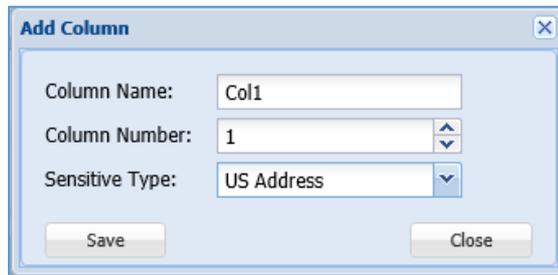
- **Structure Name:** Enter the name of the structure. The structure name accepts letters, numbers and symbols. The name must be unique to each individual structure.
- **Description:** Enter the description for the structure name.
- **Number of Header Rows:** Select the numeric value from the list box.
- **Text Structure Type:** You can select the **Text Structure Type** from the drop-down. The available options are: **Default**, **HiveArray**, **HiveStruct**, and **HiveMap**.
- **Use Delimiter:** This field will appear, when you select **Default** as the **Text Structure Type**. Check this checkbox to identify different columns from the text file, on the basis of the provided delimiter.
- **Column Delimiter:** This field will appear, when you select **Default** or **HiveStruct** as the **Text Structure Type**. Enter the Column Delimiter. The Column Delimiter field accepts several character types such as: Comma, Semicolon, Pipe, Space etc.
- **Position Counter:** This field will appear, when you select **Default** as the **Text Structure Type**. This field will appear only when you will uncheck the **Use Delimiter** checkbox. Select the value of the position counter from the drop-down. Enter the position values at the time of adding a column.
- **Array Delimiter:** This field will appear, when you select **HiveArray** as the **Text Structure Type**. Enter the delimiter for the Array.
- **Select Sensitive Type:** This field will appear, when you select **HiveArray** as the **Text Structure Type**. Select the Sensitive Type from the drop-down.
- **Key-Value Delimiter:** This field will appear, when you select **HiveMap** as the **Text Structure Type**. Enter the key value delimiter.

- **Element Delimiter:** This field will appear, when you select **HiveMap** as the **Text Structure Type**. Enter the element delimiter.
- **Select Sensitive Type for Key:** This field will appear, when you select **HiveMap** as the **Text Structure Type**. Select the Sensitive Type for the key.
- **Select Sensitive Type for Value:** This field will appear, when you select **HiveMap** as the **Text Structure Type**. Select the Sensitive Type for the value.
- **Add Keys:** Check this checkbox to add a new key.
 -  : Click this button to add a new key.
 -  : Click this button to delete a key.
 - **Select Sensitive Type for Key:** Select the Sensitive Type for the key.
 - **Select Sensitive Type for Value:** Select the Sensitive Type for the value.
- **File Pattern:** Enter the File Pattern in the text box. The File Pattern accepts all supported file types, for example, .txt, .xml, .csv

Add column info for a Structure

To add a column for masking. Click **Hadoop > Structure Management > Add Column**.

The below image shows the user interface for adding a column for masking.



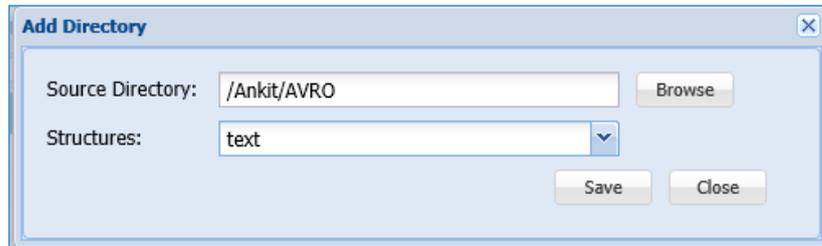
- **Column Name:** Enter the name of the column which need to be masked. The column name accepts letters, numbers and symbols.
- **Column Number:** Select the numeric value for the column from Column Number list box.
- **Offset:** This option will appear when you select **Position Counter** value in the [Create Structure](#) pop-up. Enter the starting position of the cell.
- **Length:** This option will appear when you select **Position Counter** value in the [Create Structure](#) pop-up. Enter the length of the cell.
- **Sensitive Type:** Select the Sensitive Type from the drop-down.
- **Save:** Click the Save button, if you want to save the changes.

- **Close:** Click the Close button, if you do not want to save the changes.

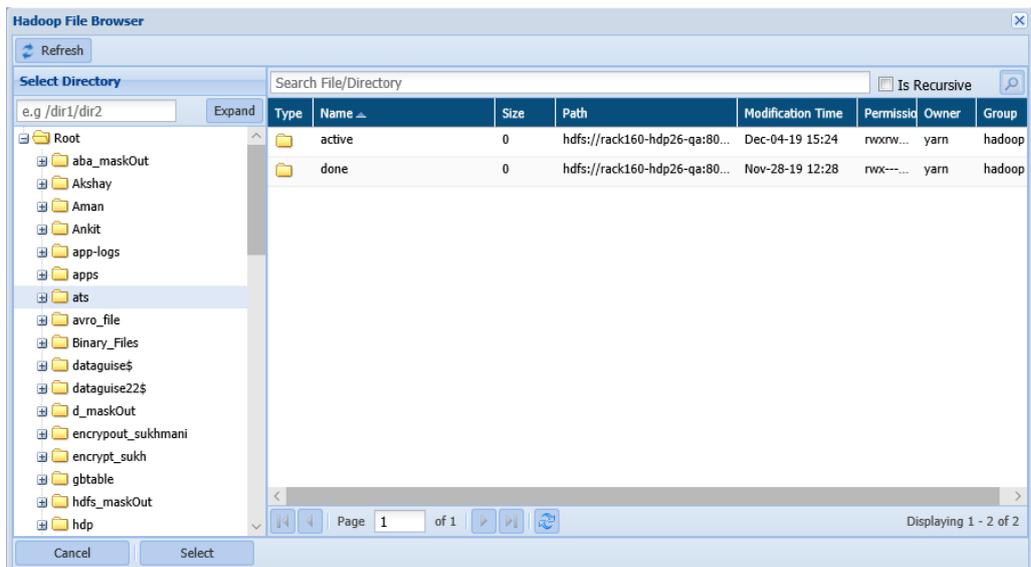
Create a Directory

To create a database. Click **RDBMS > Structure Management > Add Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Enter the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
2. Select the File/Directory from the right pane.
3. **Select:** Click the Select button, if you want to include the file or directory.
4. **Refresh:** Click the Refresh button to update the current page with updated information.
5. **Cancel:** Click the Cancel button, if you do not want to save the changes.

- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

13.2.2.2 Sequence Files

Create a Structure

To create a structure. Click **Hadoop > Structure Management > Sequence Files > Add Structure**. The below image shows the user interface for creating a structure.

The screenshot shows a dialog box titled "Add Sequence File Structure". It contains the following fields and options:

- Structure Name: seq
- Description: seq
- Key Class: org.apache.hadoop.io.Text
- Value Type: BuiltIn
- Value Class: Text
- Select One:
 - Sensitive Type
 - Text Structure
- Text Structure: seq

At the bottom of the dialog, there is a table with two columns: "Field Name" and "Sensitive Type". The table is currently empty. Below the table are "Save" and "Close" buttons.

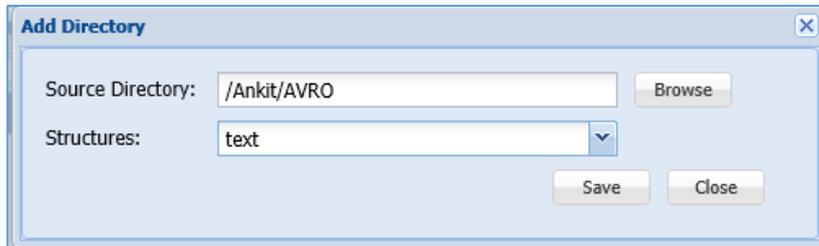
- **Structure Name:** Enter the name of the structure. This field structure accepts numbers, symbols and letters.
- **Description:** Enter the description for the structure name.
- **Key Class:** Enter the Key Class.
- **Value Type:** Select the Value Type from the drop-down. You can select either 'Built-In' or 'SelfDefined'.
 - ❖ **Value Class:** Choose appropriate Value Class from the drop-down.
 - ❖ **Select One:** You can choose either 'Sensitive Type' or 'Text Structure'.
- **Save:** Click the Save button, if you want to save the changes.

- **Close:** Click the Close button, if you do not want to save the changes.

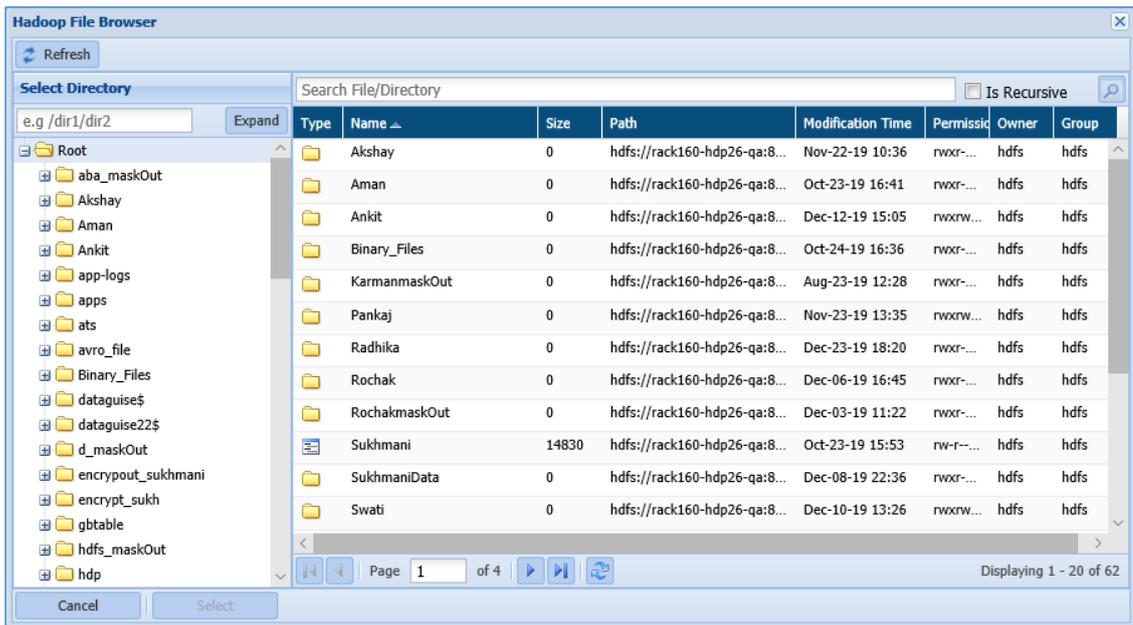
[Create a Directories](#)

To create a database. Click **RDBMS > Structure Management > Add Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Enter the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
2. Select the File/Directory from the right pane.
3. **Select:** Click the Select button, if you want to include the file or directory.

4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.2.2.3 AVRO Files

Create a Structure

To create a structure. Click **Hadoop > Structure Management > AVRO Files > Add Structure**. The below image shows the user interface for creating a structure.

The screenshot shows a dialog box titled "Add Structure". It contains the following elements:

- Structure Name:** A text input field containing "HA".
- Description:** A text area containing "Hadoop".
- File Type:** Two radio buttons, "Local" (selected) and "Hadoop".
- Browse File:** A text input field containing "e.g. C:\dg\sample.avsc" and a "Browse..." button.
- Read File Structure:** A button located below the file path.
- Fields List:** A table with two columns: "Field Name" and "Sensitive Type". The table is currently empty.
- Save and Close:** Two buttons at the bottom right of the dialog.

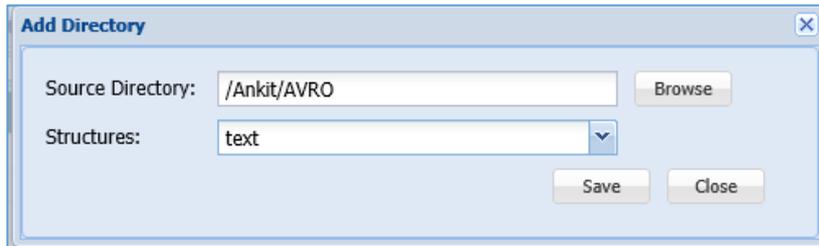
- **Structure Name:** Enter the name of the structure. This field accepts symbols, letters.
- **Description:** Enter the description for the structure name.
- **File Type:** Select either 'Local' or 'Azure Cloud' from the File Type options.
- **Browse File:** Click the 'Browse' button to include the file from your local machine if 'Local' is selected in the File Type option.
- **Read File Structure:** To read the structure of the file, click the Read File Structure.

- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

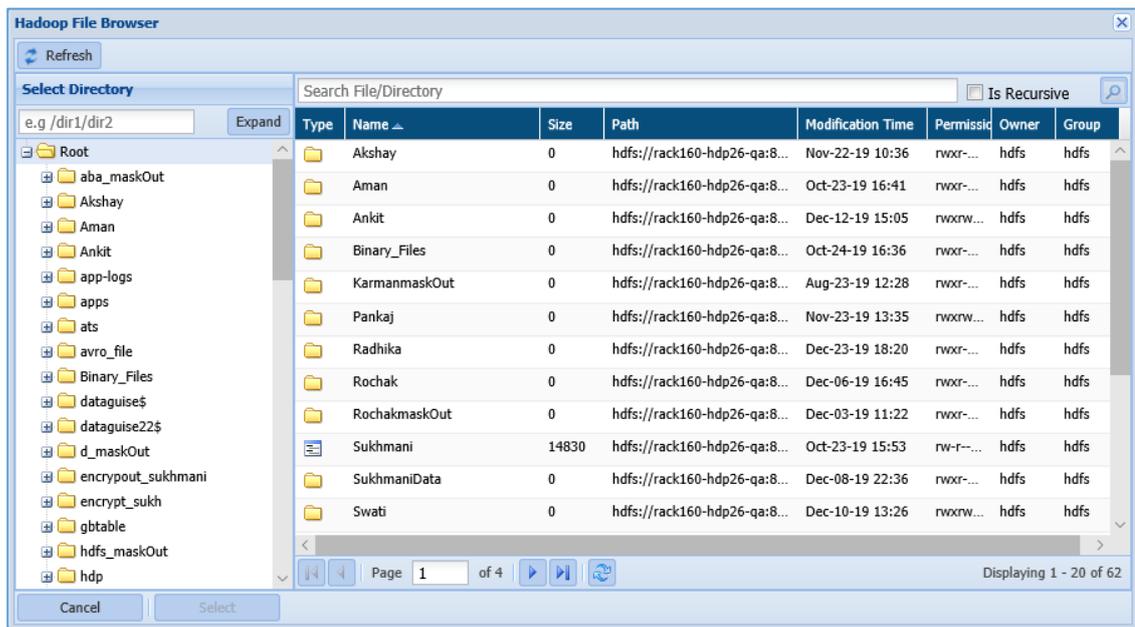
Create a Directories

To create a database. Click **RDBMS > Structure Management > Add Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Enter the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
2. Select the File/Directory from the right pane.
3. **Select:** Click the Select button, if you want to include the file or directory.

4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.2.2.4 RC/ORC Files

Create a Structure

To create a structure. Click **Hadoop > Structure Management > RC/ORC Files > Add Structure**.

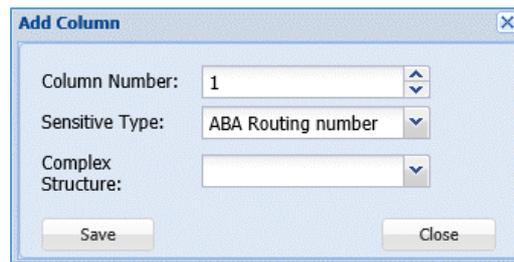
The below image shows the user interface for creating a structure.

- **Structure Name:** Enter the name of the structure. This field accepts letters, symbols and numbers. The structure name should be unique.
- **Description:** Enter the description for the structure name.
- **Number of Header Rows:** Select the numeric value from the list box.
- **Structure Type:** Select the structure type from the given options.
- **File Pattern:** Enter the File Pattern in the text box.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

Add Column info for a Structure

To add a column for masking. Click **Hadoop > Structure Management > RC/ORC Files > Add Column**.

The below image shows the user interface for adding a column for masking.

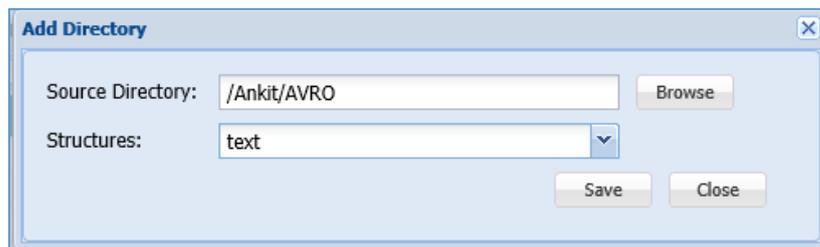


- **Column Number:** Select the column number from the Column Number list box.
- **Sensitive Type:** Select the Sensitive Type from the drop-down.
- **Complex Structure:** Select the option from the Complex Structure drop-down.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

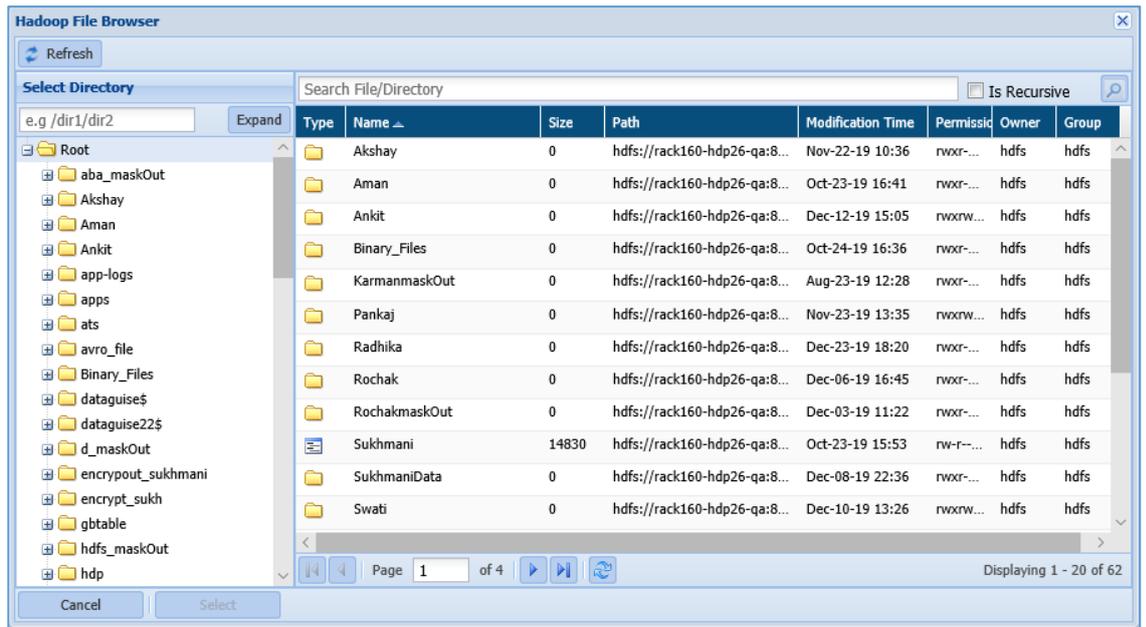
[Create a Directories](#)

To create a database. Click **Hadoop > Structure Management > RC/ORC Files > Add Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Enter the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
 2. Select the File/Directory from the right pane.
 3. **Select:** Click the Select button, if you want to include the file or directory.
 4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.2.3 Files

To access Structure Management in Files. Click **Files > Structure Management**.

To know the process of creating a structure for Text, Sequence, AVRO, RC/ORC and Parquet Files, refer to section [Hadoop](#).

13.2.4 AWS

To access Structure Management in Files. Click **AWS > Structure Management**.

To know the process of creating a structure for Text, Sequence, AVRO, RC/ORC and Parquet Files, refer to section [Hadoop](#).

13.2.5 Azure

To access Structure Management in Files. Click **Azure > Structure Management**.

To know the process of creating a structure for Text, Sequence, AVRO, RC/ORC and Parquet Files, refer to section [Hadoop](#).

13.3 Edit a Structure

13.3.1 RDBMS

13.3.1.1 Edit a Structure

To edit a structure. Click **RDBMS > Structure Management > Edit Structure**.

The below image shows the user interface for creating a structure.

- **Structure Name:** Edit the name of the structure.
- **Description:** Edit the description for the structure name.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

13.3.1.2 Edit Column info for a Structure

To edit a column for masking. Click **RDBMS > Structure Management > Edit Column**.

The below image shows the user interface for adding a column for masking.

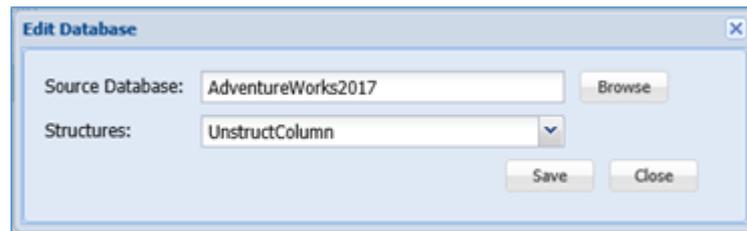
- **Table Name:** Edit the name of the table.

- **Column Name:** Edit the name of the column which need to be masked.
- **Sensitive Type:** Select the Sensitive Type from the drop-down.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

13.3.1.3 Edit a Databases

To edit a database. Click **RDBMS > Structure Management > Edit Database**.

The below image shows the user interface of the creating a database.



- **Source Database:** Edit the name of the database or you can click the Browse button to search for database.
- **Structures:** Select the structure from the Structure drop-down.
- **Browse:** Click the Browse button to search for the database or you can manually enter the name of the database.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

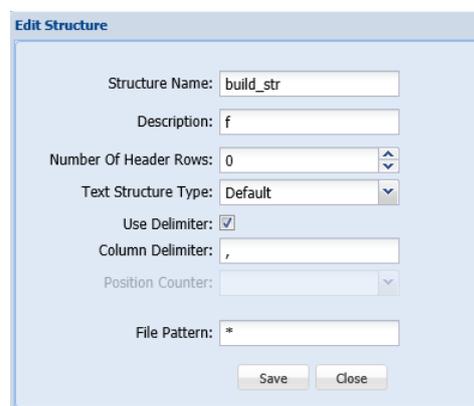
13.3.2 Hadoop

13.3.2.1 Text Structure

[Edit a Structure](#)

To Edit a structure. Click **Hadoop > Structure Management > Text Files > Edit Structure**.

The below image shows the user interface for creating a structure.



- **Structure Name:** Edit the name of the structure. The structure name accepts letters, numbers and symbols. The name must be unique to each individual structure.
- **Description:** Edit the description for the structure name.
- **Number of Header Rows:** Select the numeric value from the list box.
- **Text Structure Type:** You can select the Text Structure Type from the drop-down.
- **Use Delimiter:** Check the Delimiter checkbox, if you want to use this option.
- **Column Delimiter:** Edit the Column Delimiter. The Column Delimiter field accepts several character types such as:
 - Comma
 - Semicolon
 - Pipe
- **Element Delimiter:** Edit the Element Delimiter.
- **Select Sensitive Type for Key:** Select the Sensitive Type from the Select Sensitive Type for Key drop-down.
- **Select Sensitive Type for Value:** Select the given option from the drop-down.
- **Add Keys**
- **File Pattern:** Edit the File Pattern in the text box. The File Pattern accepts all supported file types, for example, .txt, .xml, .csv

Note:

- The Element Delimiter, Select Sensitive Type for Key, Select Sensitive Type for Value, Add Keys options are available when HiveMap is selected in Text Structure Type drop-down.
- Array Delimiter option will be available when HiveArray is selected in the Text Structure Type drop-down.
- Position Counter and Use Delimiter option will be available when Default is selected in the Text Structure Type drop-down.

[Edit column info for a Structure](#)

To edit a column for masking. Click **RDBMS > Structure Management > Edit Column**.

The below image shows the user interface for adding a column for masking.

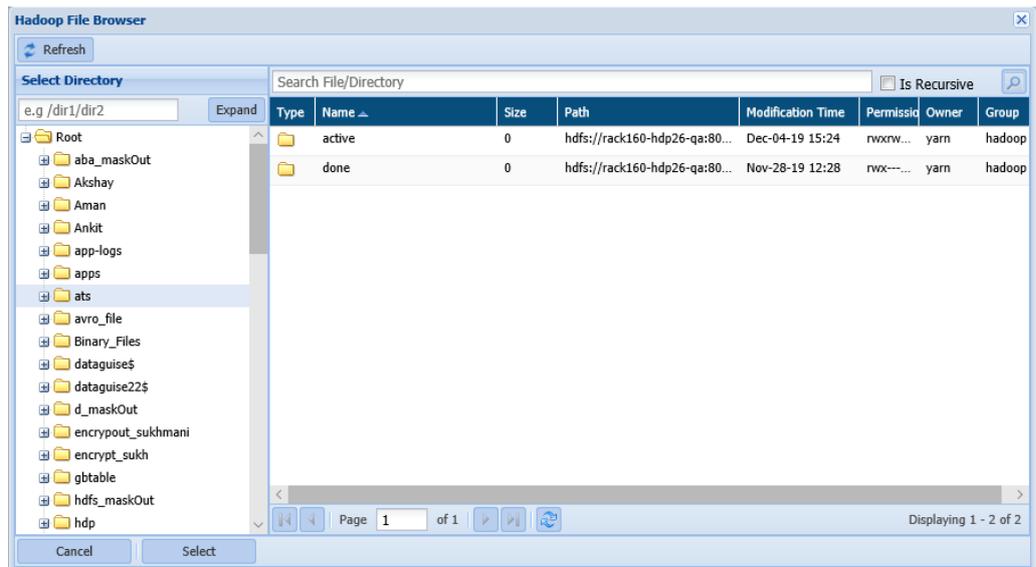
- **Column Name:** Edit the name of the column which need to be masked. The column name accepts letters, numbers and symbols.
- **Column Number:** Select the numeric value for the column from Column Number list box.
- **Sensitive Type:** Select the Sensitive Type from the drop-down.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

[Edit a Directory](#)

To edit a database. Click **RDBMS > Structure Management > Edit Directory**.

The below image shows the user interface of the creating a database.

- **Source Database:** Edit the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
 2. Select the File/Directory from the right pane.
 3. **Select:** Click the Select button, if you want to include the file or directory.
 4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.3.2.2 Sequence Files

[Edit a Structure](#)

To edit a structure. Click **Hadoop > Structure Management > Sequence Files > Edit Structure.**

The below image shows the user interface for creating a structure.

The screenshot shows a dialog box titled "Add Sequence File Structure". It contains the following fields and options:

- Structure Name: seq
- Description: seq
- Key Class: org.apache.hadoop.io.Text
- Value Type: BuiltIn
- Value Class: Text
- Select One:
 - Sensitive Type
 - Text Structure
- Text Structure: seq

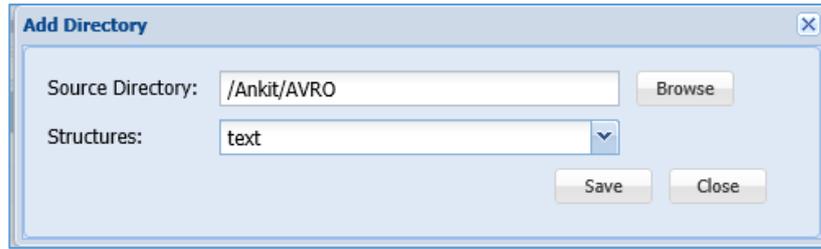
At the bottom of the dialog, there are two buttons: "Save" and "Close".

- **Structure Name:** Edit the name of the structure. This field structure accepts numbers, symbols and letters.
- **Description:** Edit the description for the structure name.
- **Key Class:** Edit the Key Class.
- **Value Type:** Select the Value Type from the drop-down. You can select either 'Built-In' or 'SelfDefined'.
 - ❖ **Value Class:** Choose appropriate Value Class from the drop-down.
 - ❖ **Select One:** You can choose either 'Sensitive Type' or 'Text Structure'.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

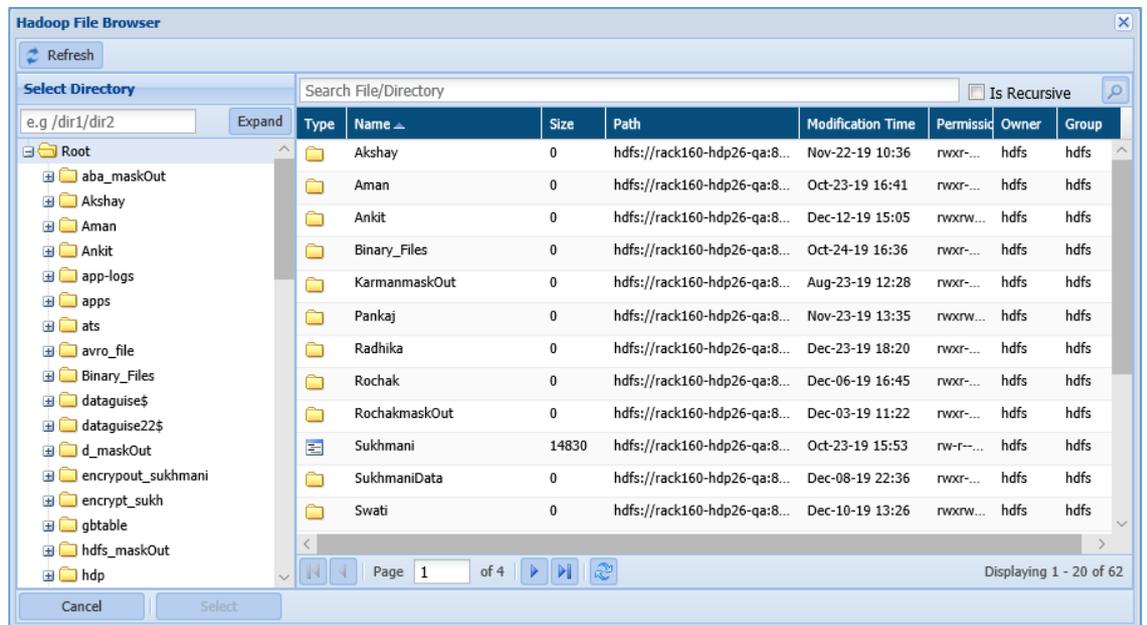
[Edit a Directories](#)

To edit a database. Click **RDBMS > Structure Management > Edit Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Edit the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
 2. Select the File/Directory from the right pane.
 3. **Select:** Click the Select button, if you want to include the file or directory.
 4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.3.2.3 AVRO Files

Edit a Structure

To edit a structure. Click **Hadoop > Structure Management > AVRO Files > Edit Structure**. The below image shows the user interface for creating a structure.

The screenshot shows a dialog box titled "Add Structure". It contains the following elements:

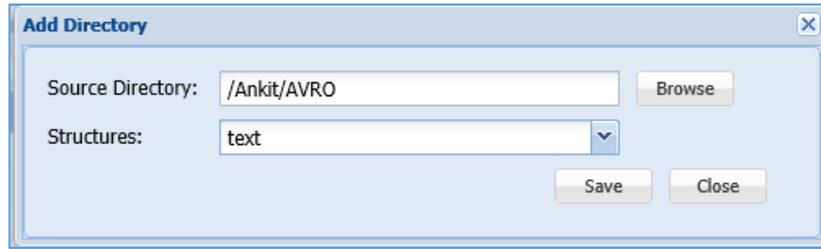
- Structure Name:** A text input field containing "HA".
- Description:** A text area containing "Hadoop".
- File Type:** Two radio buttons, "Local" (selected) and "Hadoop".
- Browse File:** A text input field containing "e.g. C:\dg\sample.avsc" and a "Browse.." button.
- Read File Structure:** A button located below the file path.
- Fields List:** A table with two columns: "Field Name" and "Sensitive Type". The table is currently empty.
- Save and Close:** Two buttons at the bottom right of the dialog.

- **Structure Name:** Edit the name of the structure. This field accepts symbols, letters.
- **Description:** Edit the description for the structure name.
- **File Type:** Select either 'Local' or 'Azure Cloud' from the File Type options.
- **Browse File:** Click the 'Browse' button to include the file from your local machine if 'Local' is selected in the File Type option.
- **Read File Structure:** To read the structure of the file, click the Read File Structure.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

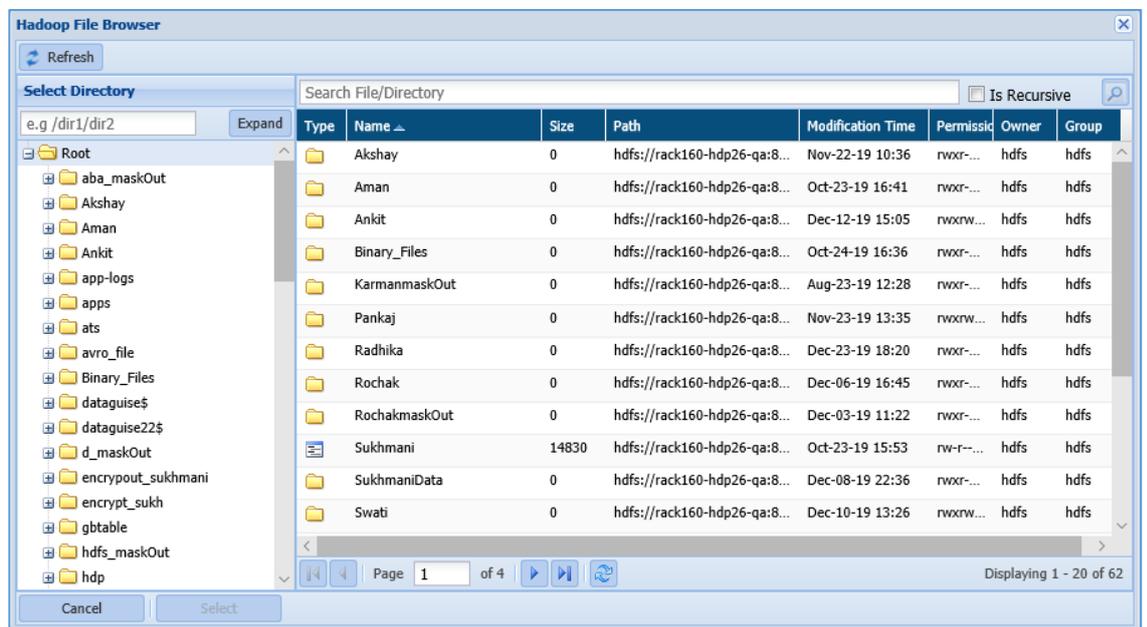
Edit a Directories

To edit a database. Click **Hadoop > Structure Management > AVRO Files > Add Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Edit the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
 2. Select the File/Directory from the right pane.
 3. **Select:** Click the Select button, if you want to include the file or directory.
 4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.3.2.4 RC/ORC Files

Edit a Structure

To edit a structure. Click **Hadoop > Structure Management > RC/ORC Files > Edit Structure**. The below image shows the user interface for creating a structure.

The 'Add Structure' dialog box has the following fields and controls:

- Structure Name: rc
- Description: rc
- Number Of Header Rows: 0 (with up/down arrows)
- Structure Type: RC (with a dropdown arrow)
- File Pattern: * (with a clear 'x' button)
- Buttons: Save, Close

- **Structure Name:** Edit the name of the structure. This field accepts letters, symbols and numbers. The structure name should be unique.
- **Description:** Edit the description for the structure name.
- **Number of Header Rows:** Select the numeric value from the list box.
- **Structure Type:** Select the structure type from the given options.
- **File Pattern:** Edit the File Pattern in the text box.
- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

Edit Column info for a Structure

To edit a column for masking.

Click **Hadoop > Structure Management > RC/ORC Files > Edit Column**.

The below image shows the user interface for adding a column for masking.

The 'Add Column' dialog box has the following fields and controls:

- Column Number: 1 (with up/down arrows)
- Sensitive Type: ABA Routing number (with a dropdown arrow)
- Complex Structure: (with a dropdown arrow)
- Buttons: Save, Close

- **Column Number:** Select the column number from the Column Number list box.
- **Sensitive Type:** Select the Sensitive Type from the drop-down.
- **Complex Structure:** Select the option from the Complex Structure drop-down.

- **Save:** Click the Save button, if you want to save the changes.
- **Close:** Click the Close button, if you do not want to save the changes.

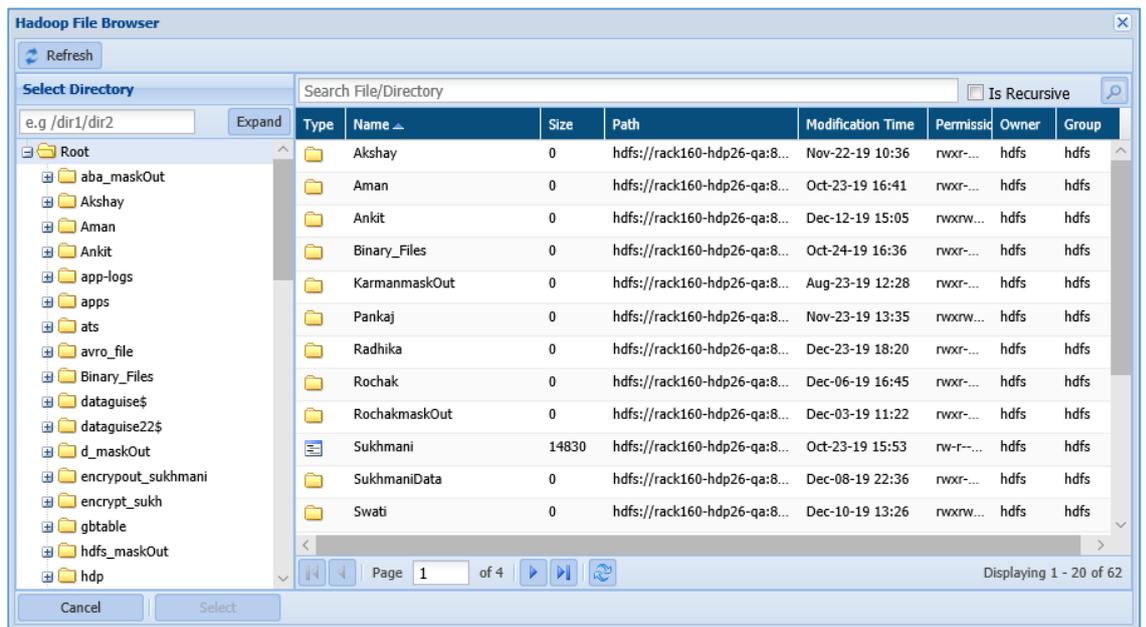
[Edit a Directories](#)

To edit a database. Click **Hadoop > Structure Management > RC/ORC Files > Edit Directory**.

The below image shows the user interface of the creating a database.



- **Source Database:** Edit the name of the directory or you can click the Browse button to search for directory.
- **Structures:** Select the structure from the Structures drop-down.
- **Browse:** Click the Browse button to search for the directory or you can manually enter the name of the database.



1. **Select Directory:** Select the directory from the Select directory pane. The right pane next to Select Directory pane will display the list of all Files and Folders.
2. Select the File/Directory from the right pane.
3. **Select:** Click the Select button, if you want to include the file or directory.

4. **Refresh:** Click the Refresh button to update the current page with updated information.
 5. **Cancel:** Click the Cancel button, if you do not want to save the changes.
- **Save:** Click the Save button, if you want to save the changes.
 - **Close:** Click the Close button, if you do not want to save the changes.

13.3.3 Files

To access Structure Management in Files. Click **Files > Structure Management**.

To know the process of editing a structure for Text, Sequence, AVRO, RC/ORC and Parquet Files, click [Hadoop](#) .

13.3.4 AWS

To access Structure Management in Files. Click **AWS > Structure Management**.

To know the process of editing a structure for Text, Sequence, AVRO, RC/ORC and Parquet Files, click [Hadoop](#).

13.3.5 Azure

To access Structure Management in Files. Click **Azure > Structure Management**.

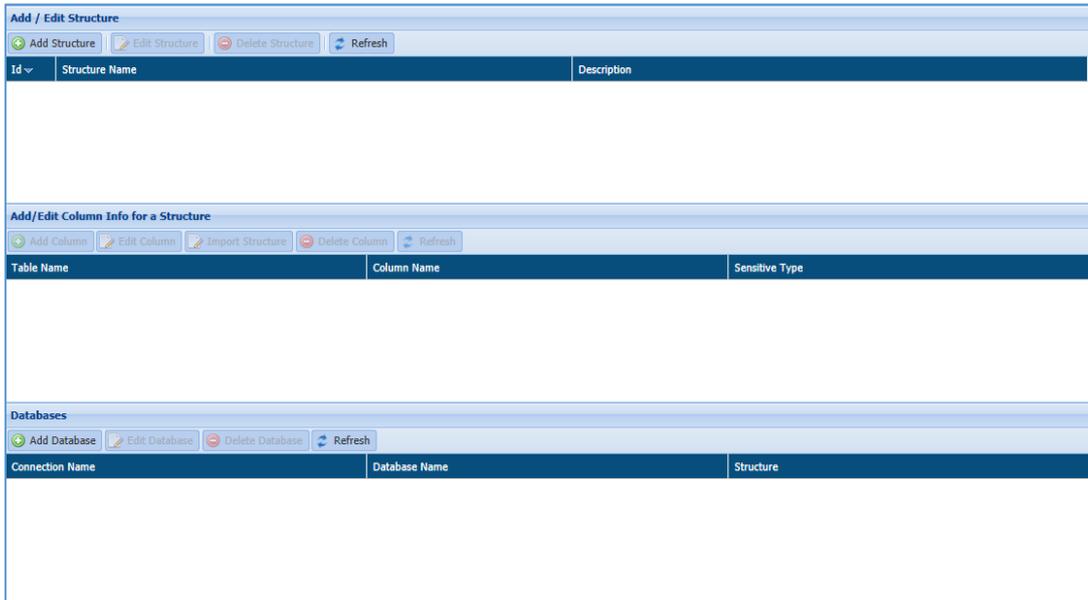
To know the process of editing a structure for Text, Sequence, AVRO, RC/ORC and Parquet Files, click [Hadoop](#).

13.4 List a Structure

13.4.1 RDBMS

This section will explain the screen of the Structure Management.

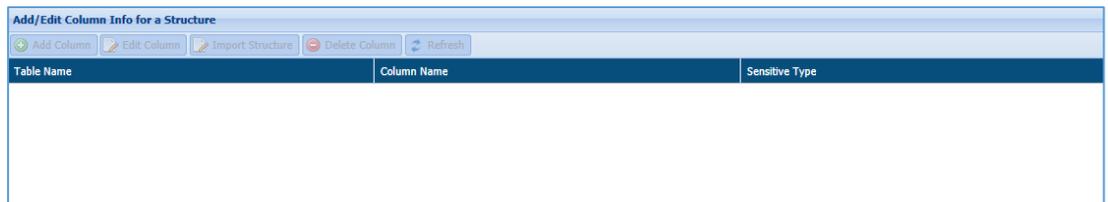
The below screenshot shows the user interface for Structure screen.



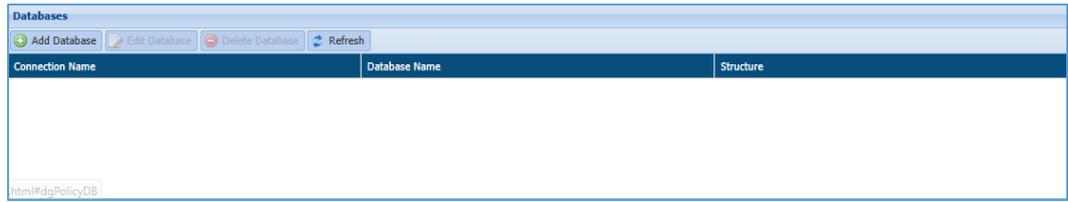
- Add/Edit Structure:**
 The Add/Edit Structure will display the list of all structure that you have created. It includes information about structure like ID (system generated), Structure Name, Description of the structure.



- Add/Edit Column Info for a Column:**
 The Add/Edit Column Info for a Column will display the list masked column list. It will display the information such as Task Name, Column Name and Sensitive Type.



- Databases:**
 The Database pane will display the list all databases that you have added. The details include Connection Name, Database Name and Structure.

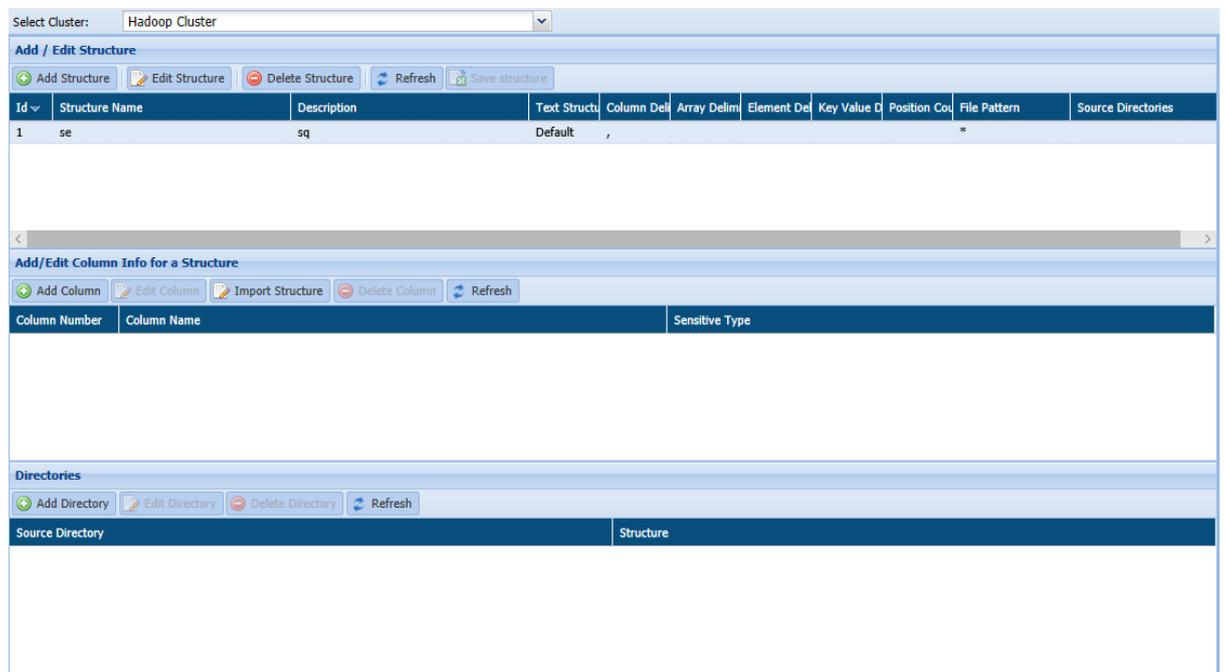


13.4.2 Hadoop

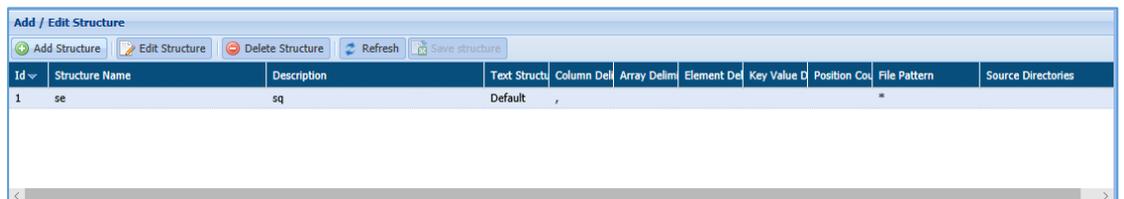
This section will explain the screen of the Structure Management for Text, Sequence, Avro, RC/ORC and Parquet Files.

13.4.2.1 Text Files

The below image shows the user interface of the Text File section.

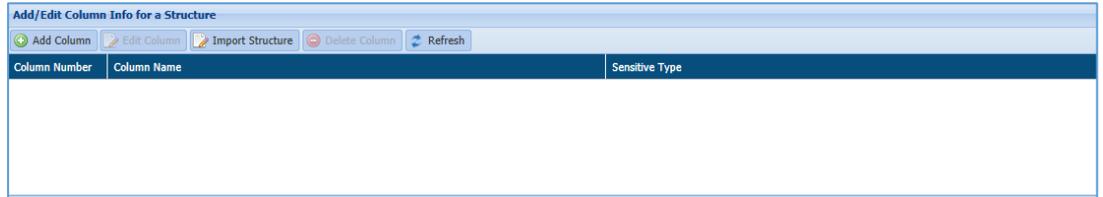


- Add/Edit Structure:**
 The Add/Edit Structure will display the list of all structure that you have created. It includes information about structure like ID (system generated), Structure Name, Description of the structure, Text Structure, Column Delimiter, Array Delimiter, etc.



- **Add/Edit Column Info for a Column:**

The Add/Edit Column Info for a Column will display the list masked column list. It will display the information such as Column Number, Column Name and Sensitive Type.



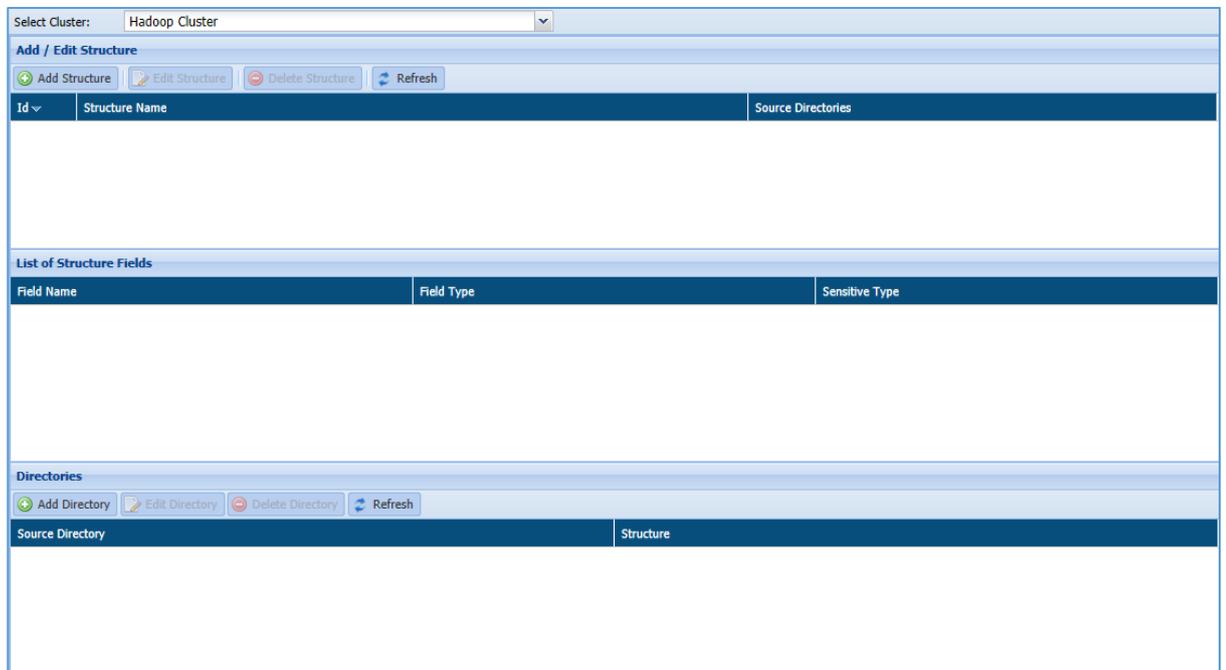
- **Directories:**

The Directories pane will display the list all databases that you have added. The details include Source Directory and Structure.



13.4.2.2 Sequence Files

The below image shows the user interface of the Sequence File section.



- **Add/Edit Structure:**

The Add/Edit Structure will display the list of all structure that you have created. It includes information about structure like ID (system generated), Structure Name,

Source Directory.

Add / Edit Structure		
Add Structure Edit Structure Delete Structure Refresh		
Id	Structure Name	Source Directories

- **List of Structure Fields:**

The List of Structure Fields will display all list of fields name which you have defined while creating a structure. The panel displays information such as Field Name, Field Type and Sensitive Type.

List of Structure Fields		
Field Name	Field Type	Sensitive Type

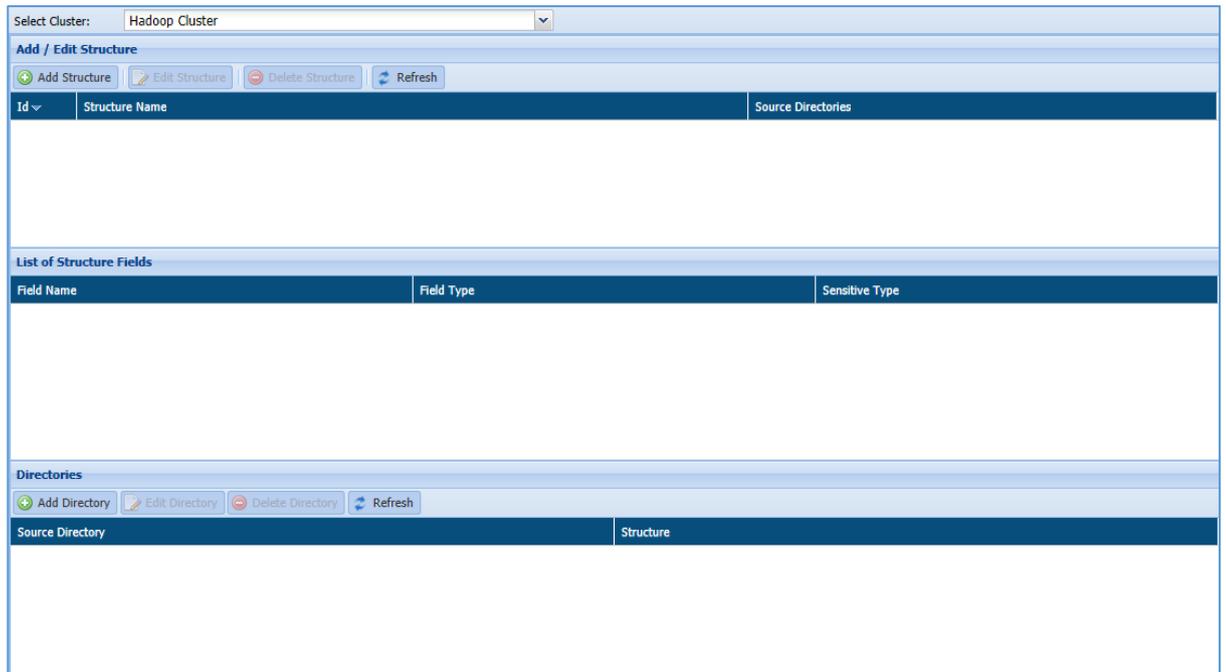
- **Directories:**

The Directories pane will display the list all databases that you have added. The details include Source Directory and Structure.

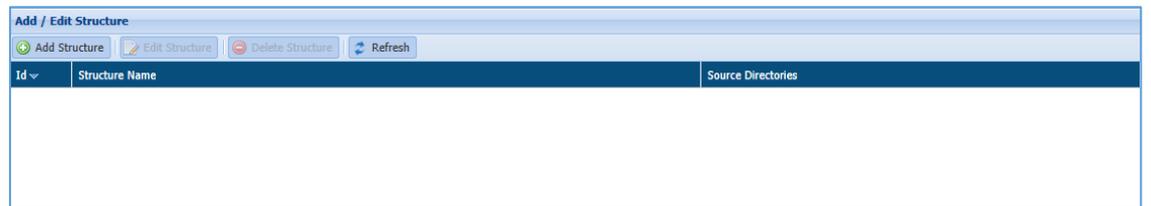
Directories	
Add Directory Edit Directory Delete Directory Refresh	
Source Directory	Structure

13.4.2.3 Avro Files

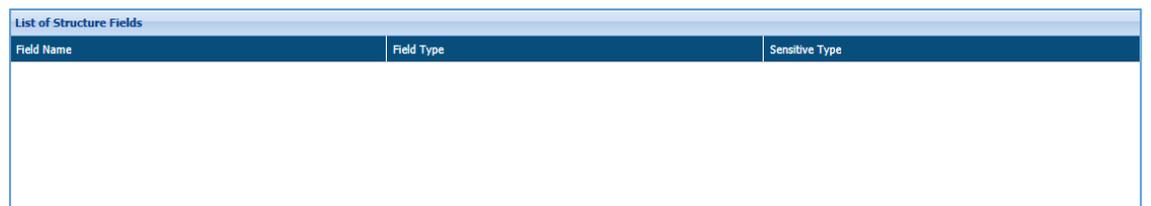
The below image shows the user interface of the Sequence File section.



- Add/Edit Structure:**
 The Add/Edit Structure will display the list of all structure that you have created. It includes information about structure like ID (system generated), Structure Name, Source Directory.



- List of Structure Fields:**
 The List of Structure Fields will display all list of fields name which you have defined while creating a structure. This pane displays information such as Field Name, Field Type and Sensitive Type.



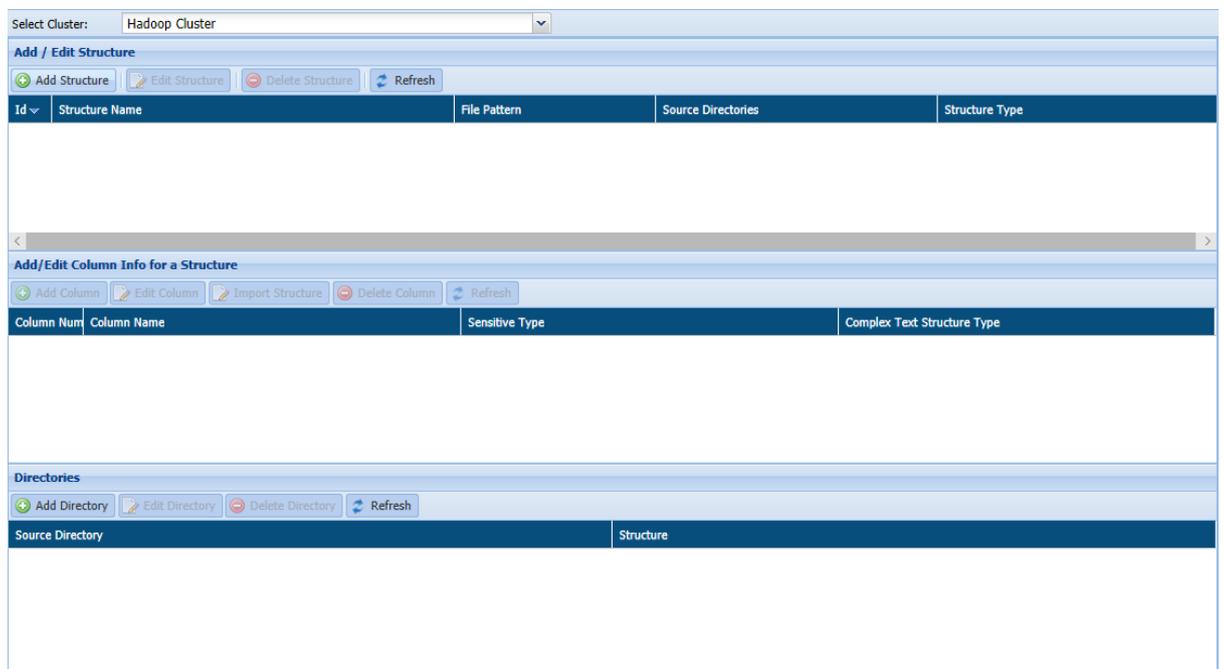
- **Directories:**

The Directories pane will display the list all databases that you have added. The details include Source Directory and Structure.



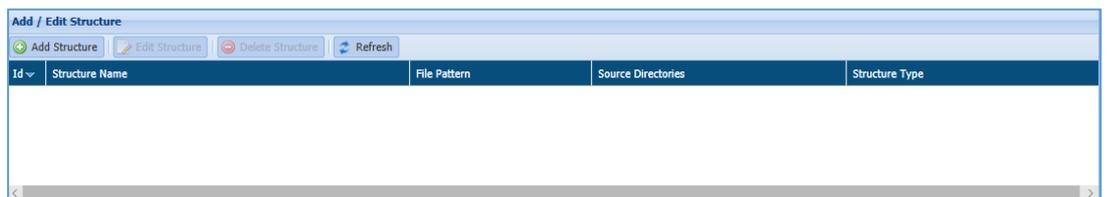
13.4.2.4 RC/ORC Files

The below image shows the user interface of the RC/ORC screen.



- **Add/Edit Structure:**

The Add/Edit Structure will display the list of all structure that you have created. It includes information about structure like ID (system generated), Structure Name, File Pattern, Source Directories and Structure Type



- **Add/Edit Column Info for a Column:**

The Add/Edit Column Info for a Column will display the list masked column list. It will display the information such as Column Number, Column Name, Sensitive Type and Complex Text Structure Type.

Column Num	Column Name	Sensitive Type	Complex Text Structure Type

- **Directories:**

The Directories pane will display the list all databases that you have added. The details include Source Directory and Structure.

Source Directory	Structure

13.4.3 Files

This section will explain the screen of the Structure Management for Text, Sequence, Avro, RC/ORC and Parquet Files.

To know more about this screen, refer to section [Hadoop](#).

13.4.4 AWS

This section will explain the screen of the Structure Management for Text, Sequence, Avro, RC/ORC and Parquet Files.

To know more about this screen, refer to section [Hadoop](#).

13.4.5 Azure

This section will explain the screen of the Structure Management for Text, Sequence, Avro, RC/ORC and Parquet Files.

To know more about this screen, refer to section [Hadoop](#).

14 Output Directory

14.1 Concept

The Output Directory page is used to define the output directory for masking results. This page shows the source and destination directories to which each have been mapped.

The user can designate a new destination directory or edit an existing one by clicking the appropriate button on the screen.

If no destination directory is defined, then DgSecure will create a sub-directory for the masked values.

14.2 Create a Source Directory

This section will explain the process of creating, editing and deleting a source directory.

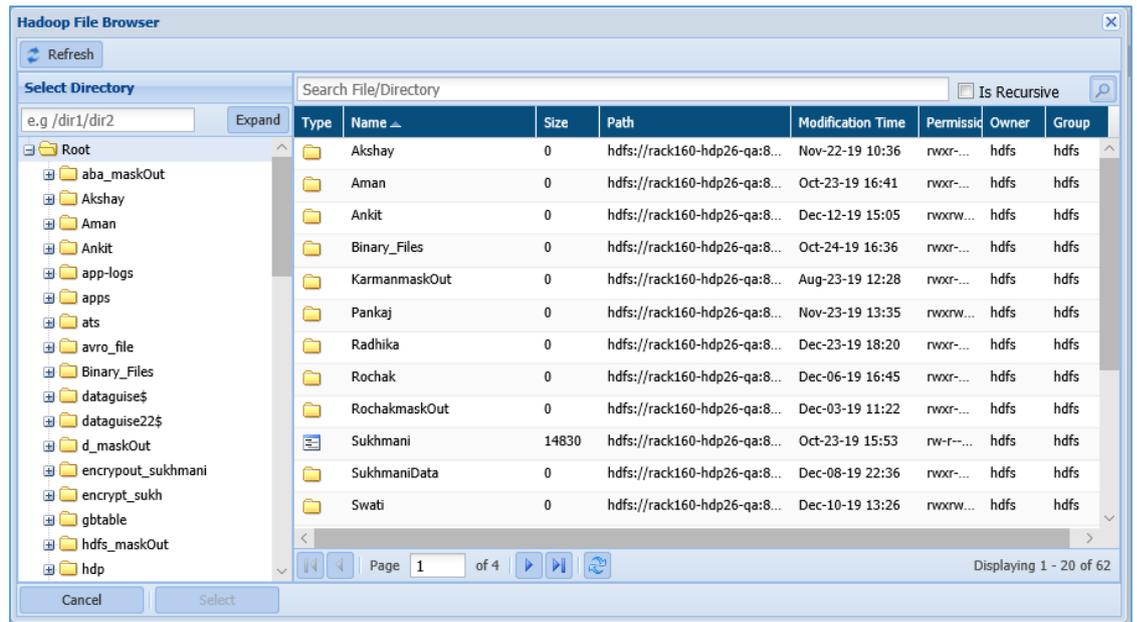
14.2.1 Hadoop

To create a source directory. Click **Hadoop > Output Directory > Add Source Directory**.

The below image shows the user interface for creating a source directory.

The screenshot shows a dialog box titled "Source Directory for Masking". It has three text input fields: "Source Directory" containing "/dev", "Branch Point" containing "/dev", and "Destination Directory" containing "/ncMaskOut". A "Browse" button is positioned to the right of the "Source Directory" field. At the bottom right of the dialog, there are "Save" and "Close" buttons.

- **Add Source Directory:** Click the Add Source Directory button to select the directory from the Hadoop File Browser.
- **Browse:** Click the Browse button to select the directory for masking.



- ❖ **Refresh:** Click the Refresh button to update the Hadoop File Browse dialog box with updated information.
 - ❖ **Select Directory:** The Select Directory pane will list down all the available directories. The right pane will display list of all Files and Folder for the selected directory.
 - ❖ **Cancel:** Click the Cancel button, if you do not want to save the changes.
 - ❖ **Select:** Click the Select button, if you want to include the selected directory for masking.
- **Branch Point:** The Branch point will display the selected directory name.
 - **Destination Directory:** The Destination Directory will display the name of the destination folder name.
 - **Save:** Click the Save button, if you want to save the changes.
 - **Cancel:** Click the Cancel button, if you do not want to save the changes.

14.2.2 Files

To create a source directory in Files. Click **Files > Output Directory > Add Source Directory**.

To know the process of creating a source directory, refer to section [Hadoop](#).

14.2.3 AWS

To create a source directory in Files. Click **AWS > Output Directory > Add Source Directory**.

To know the process of creating a source directory, refer to section [Hadoop](#).

14.2.4 Azure

To create a source directory in Files. Click **AWS > Output Directory > Add Source Directory**.

To know the process of creating a source directory, refer to section [Hadoop](#).

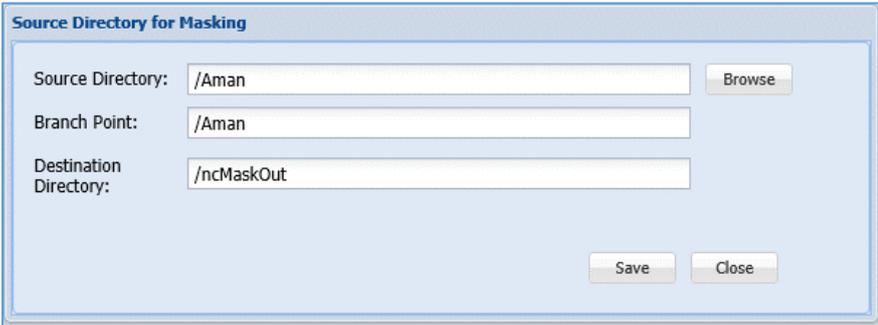
14.3 Edit a Source Directory

14.3.1 Hadoop

To edit a Source Directory. Select the Source Directory which you wish to edit.

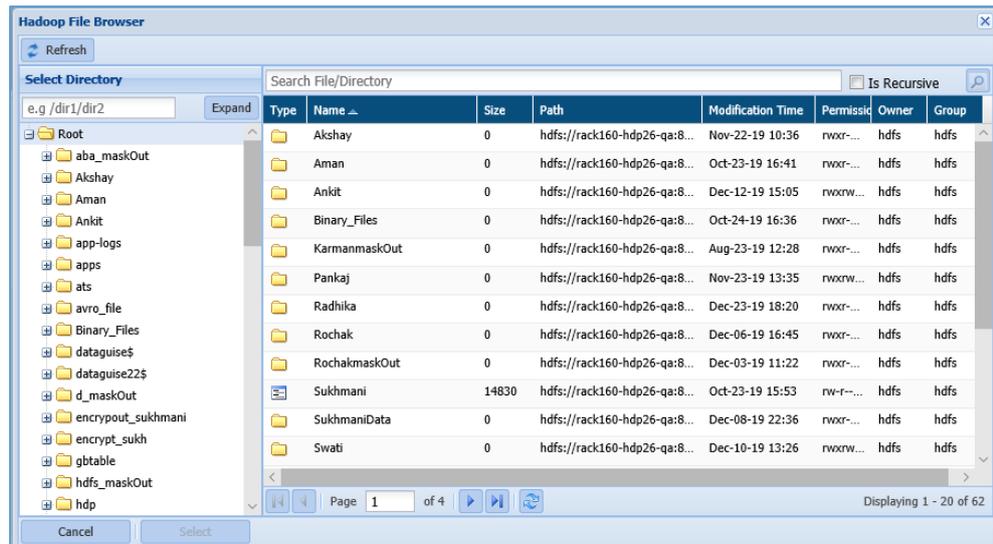
Click Hadoop > **Output Directory > Edit Source Directory**.

The below image shows the user interface for editing a Source directory.



The screenshot shows a dialog box titled "Source Directory for Masking". It has three input fields: "Source Directory" containing "/Aman" with a "Browse" button to its right, "Branch Point" containing "/Aman", and "Destination Directory" containing "/ncMaskOut". At the bottom right, there are "Save" and "Close" buttons.

- **Source Directory:** To Edit the Source Directory, click Browse button to select a new source directory.
- **Browse:** Click the Browse button to select the directory from the Hadoop File Browser.



❖ **Select:** Click the Select button, if you want to save the changes.

❖ **Cancel:** Click the Cancel button, if you do not want to save the changes.

- **Branch Point:** The Branch Point text box will be updated with the updated selected directory name.
- **Destination Directory:** The Destination directory name will be updated accordingly to the selected directory name.

14.3.2 Files

To edit a Source Directory. Select the Source Directory which you wish to edit.

Click **Files > Output Directory > Edit Source Directory**.

To know the process of editing a source directory, refer to section [Hadoop](#).

14.3.3 AWS

To edit a Source Directory. Select the Source Directory which you wish to edit.

Click **AWS > Output Directory > Edit Source Directory**.

To know the process of editing a source directory, refer to section [Hadoop](#).

14.3.4 Azure

To edit a Source Directory. Select the Source Directory which you wish to edit.

Click **Azure > Output Directory > Edit Source Directory**.

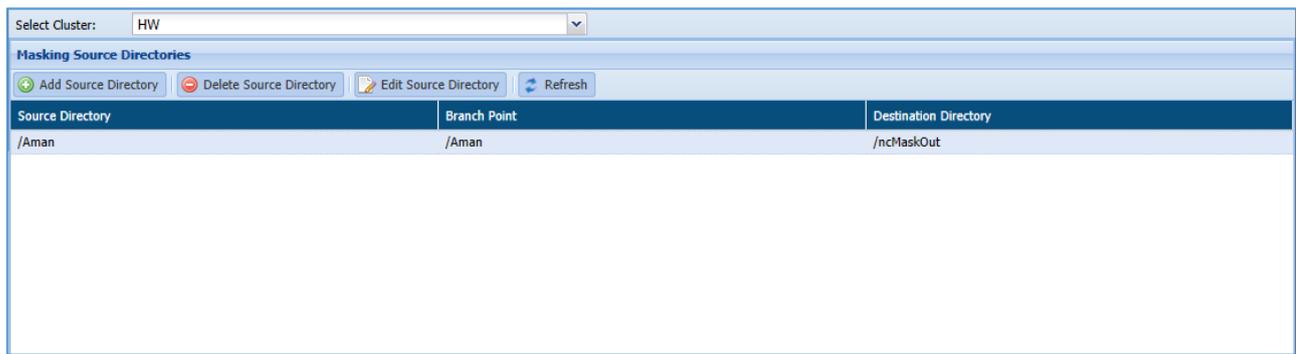
To know the process of editing a source directory, refer to section [Hadoop](#).

14.4 List an Output Directory

14.4.1 Hadoop

This section will explain the screen of the Output Directory.

The below screenshot shows the user interface for Output Directory.



- **Masking Source Directories:** This pane displays the list of all the source directory and the basic details such as the Branch Point and Destination Directory.

You can add, delete, edit the source directory by clicking an appropriate button.

- ❖ **Add Source Directory:** Click the Add Source Directory button to select the source directory from the Browse button.
- ❖ **Edit Source Directory:** Click the Edit Source Directory button to edit the information of Source Directory by updating the selected directory from browser window.
- ❖ **Delete Source Directory:** Select the Source Directory which you want to delete. Click the Delete Source Directory button.

14.4.2 Files

This section will explain the screen of the Output Directory for Files section.

To know the process of editing a source directory, refer to section [Hadoop](#).

14.4.3AWS

This section will explain the screen of the Output Directory for AWS section.

To know the process of editing a source directory, refer to section [Hadoop](#).

14.4.4Azure

This section will explain the screen of the Output Directory for AWS section.

To know the process of editing a source directory, refer to section [Hadoop](#).

15 Access Control

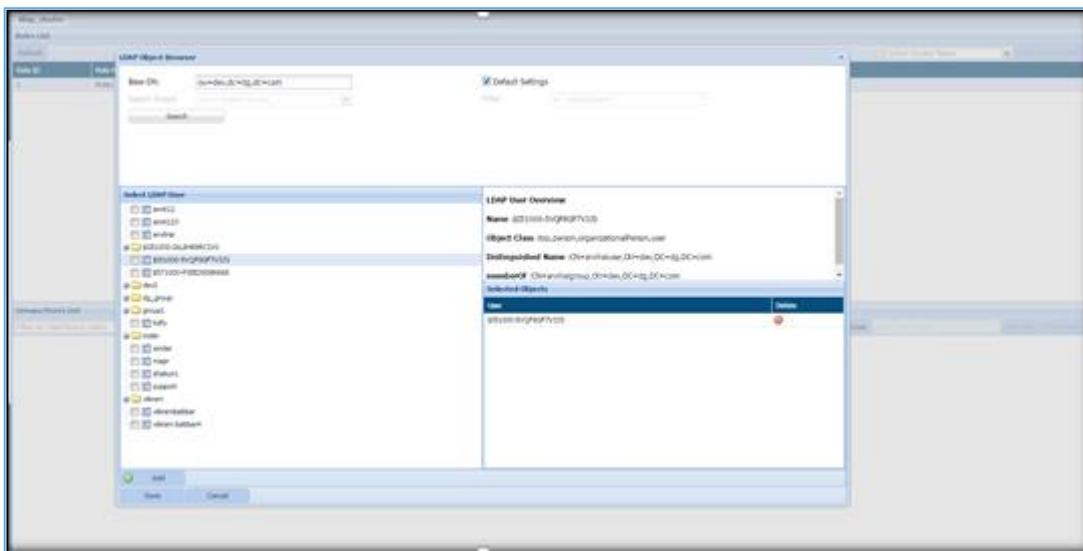
15.1 ACL Management

An access control list (ACL) manages user access to encrypted data. DgSecure’s ACL Management function controls which users can run decryption tasks. The **ACL Management** page displays the available roles and users. The groups/users can be entered manually, chosen with the LDAP browser, or it can come from the Hadoop Control IDP.

Roles are created and managed on the **Role Management** page.

If using Hive Decryption, the contents of the file will be stored in the Hive DB table after decryption.

Access the **ACL Management** page from the menu under **Access Control > ACL Management**. The **ACL Management** page is described below.



The top panel displays roles that have been created as well as their ID description. Select a role in order to view or edit which users are assigned to it. To copy the current ACL to another cluster, select the desired cluster from the dropdown on the left side of the screen. Once the cluster is selected, click the **Copy ACL to Cluster** button. In order for the operation to be successful, the clusters must have the same Groups/Users List.

Role ID	Role Name	Role Description
1	Business Analyst	BA Technologies

The bottom panel lists the users that compose the access control list. There are three different ways, this list is populated.

LDAP/LDAPS: When the cluster is configured with LDAP or LDAPS on the Hadoop Clusters page, the users list is generated by selecting groups/users from the LDAP Object Viewer. To access the LDAP Object Viewer, click the LDAP Object Viewer button. The LDAP Object Viewer uses object classes defined in DgAdmin on the LDAP Object Class Management page to internally search the tree. Object viewer behaviour is determined in DgAdmin on the Settings page. When both LDAP/LDAPS and the Hadoop Control IDP are configured on a cluster, priority is given to LDAP/LDAPS users.

Hadoop Control IDP: When DgSecure's Hadoop Control IDP is installed on the cluster, the ACL user list shows all Linux users with permissions on the cluster.

Manual Entry: Users can be added manually in addition to - or independent of - LDAP and the Hadoop Control IDP. A user that is added manually must already exist on the cluster.



15.2 Role Management

A role defines a user's ability to run both dynamic and bulk decryption tasks in DgSecure. A user with a role missing appropriate permissions that attempts to decrypt a HDFS file cannot successfully decrypt the data in that file. Once a role is created, users can be assigned to that role on the ACL Management page. A user can have multiple roles.

There are two parameters that can be used to restrict a role's permissions: sensitive type, and day/time.

Access the **Role Management** page from the menu under **Access Control > Role Management**. The **Role Management** page is described below.

Manage Roles

Buttons: Add Role, Refresh, Delete Role, Update ACL

Role ID	Role Name	Role Description
1	analyst	Marketing Research

Edit Sensitive Types Access Permissions

Buttons: Save, Refresh

Name	Access
Full Access for RegexGroups	<input type="checkbox"/>
ABA Routing number	<input checked="" type="checkbox"/>
Address	<input type="checkbox"/>
Bank Number	<input checked="" type="checkbox"/>
Card Swipe Data	<input checked="" type="checkbox"/>
CM11	<input type="checkbox"/>

Edit Days Permissions

Buttons: Set Time, Refresh

Day Name	Time Intervals
Sun	
Mon	
Tue	
Wed	
Thu	
Fri	

The top panel is where roles are created or deleted. Once a role is created, it is listed in the top panel along with its ID and description. Select a role to view or edit its details in the panels below.

Buttons: Add Role, Refresh, Delete Role

Role Id	Role name	Role Description
3	qa	qa
4	qa1	qa
5	hdfs	hdfs
6	dataguide	dataguide
7	root	roo

Use the **Edit Sensitive Type Access Permissions** panel to identify which sensitive types a user has permissions to decrypt.

Edit Sensitive Types Access Permissions

Buttons: Save, Refresh

Name	Access
Full Access for RegexGroups	<input type="checkbox"/>
ABA Routing number	<input type="checkbox"/>
Address	<input type="checkbox"/>
Bank Number	<input type="checkbox"/>
Card Swipe Data	<input type="checkbox"/>
CM11	<input type="checkbox"/>

Use the **Edit Days Permissions** panel to identify the days and times when a role is allowed to decrypt. When no time restrictions are set, a role is able to decrypt at any time.

Edit Days Permissions

Set Time Refresh

Day Name	Time Intervals
Fri	06:00:00-23:45:00
Sun	06:00:00-12:00:00,14:00:00-18:30:00
Mon	
Sat	
Thu	
Tue	

16 Privacy

16.1 Concept

Regulations such as the European Union’s General Data Protection Regulation (GDPR) and California’s Consumer Privacy Act (CCPA) are forcing organizations worldwide, to revisit their Data Privacy policies and practices. Other privacy regulations around the world are likely to have data subject rights as their core principles. While there are minor variations in the specific types of information involved and the conditions under which these rights are to be respected, the fundamental requirements are similar.

Dataguide has the following features within the Privacy module of the DgSecure product, to scan, process and retrieve information related to Data Subjects from source systems and to generate useful DSAR reports.

Privacy – The Privacy Screens combine information about Data Subjects with information about systems and users to show the various situations including exposure to third party user risk, and cross border transfer risk. There is also a “Subject Search” screen where an individual Data Subject’s data can be retrieved. This is an interactive, ad-hoc retrieval of Data Subject information which can be used as a manual DSAR data retrieval mechanism. However, in this document, we describe the automated DSAR capability in DgSecure. The Privacy Screens are described in more detail in the Privacy section of the DgSecure User Guide.

Automated DSAR Capability – The European Union’s General Data Protection Regulations (GDPR), give an individual the right to request access to his/her personal information that an organization is handling, and the purposes for which this information is being used. This right is called the Right of Access (RoA). Similarly, an individual can also request the erasure (RtE) of his/her personal information being stored in the organization’s repositories, subject to applicable laws.

- i. **RoA (Right of Access): Right of Access** allows the Data Subject to retrieve their personal data which the organization has obtained or processed. The organization must provide information about the Data Subject that it has processed, collected or that has been transmitted to a third party. In RoA, the organization must provide a copy of the data being stored about the Data Subject in their repositories.
- ii. **RtE (Right to Erasure): The Right to Erasure** is also known as the Right to be Forgotten under GDPR. It allows the Data Subject to request the erasure of all their personal data which is being processed or stored by the organization. The scope of the erasure will be subject to other applicable laws.

The Automated DSAR capability in DgSecure enables DSAR requests to be processed in an automated fashion at pre-determined schedules, while still allowing for a manual inspection and approval/rejection by a Data Protection Officer (DPO).

16.2 Components of DSAR and Privacy

i. **Data Group**

DgSecure uses the notion of Data Groups or Data Source Groups to provide isolation between different sub-organizations within an organization. It is a form of multi-tenancy, just for the DSAR capability. **Data Group, Data Source Group** and **Groups** are synonymous under the Privacy module.

For example, an organization might be a conglomerate which has two businesses, one selling Toys and another selling Medicines. The Data Subjects pertaining to these two businesses are different, and so are the systems that contain the Data Subject information. In this case, the administrator would create two Data Source Groups, one for Toys and the other for Medicines. Systems containing Data Subject information for the Toys business will be added to the Toys Data Source Group. The list of Data Subjects for the Toys business will also be associated with the Toys Data Source Group. Identifiers will be associated with the Data Subjects in this Data Source Group and rules for searching for these identifiers will also be associated with this Data Source Group. A similar set of associations will be made for the Medicines business. Note that any overlaps in systems or Data Subjects can be handled by associating them with both Data Source Groups.

Most organizations will just need one Data Source Group to be defined, and all the systems, Data Subjects and rules will be assigned to that single Data Source Group.

ii. **Data Subjects**

A Data Subject is an entity that has identifying information such as name, address, email ID, SSNO, and CCNO. These pieces of identifying information are called Identifiers. The name and datatype of an identifier are specified while creating an the identifier in DgSecure. The values of identifiers for a given Data Subject are entered while adding that Data Subject to the system. The rules defined in Data Group helps in identifying the Data Subject’s entries in the source systems while retrieving the data in response to a DSAR request.

For example, in the below table, Jane Doe, John Doe and John Douglas are the Data Subjects. These Data Subjects have identifying information such as the Name, SSNO, Region in which they reside, and Email IDs.

Name	SSNO	Region	email_id
Jane Doe	09867537	California	jane.doe@gmail.com
John Doe	03567812	New Jersey	john.doe@gmail.com
John Douglas	05648975	New Jersey	john.doe.90@gmail.com

In DgSecure, there are three ways to add the Data Subjects:

- a. Manual Entry
- b. By uploading a file containing data subject details, with known format
- c. By uploading a file containing data subject details, with unknown format

iii. **Identifiers**

To establish an identity for a Data Subject, we need to know its identifying attributes or features. Name, Phone, SSNO, CCNO, Email, Phone Number, etc. are examples for identifying features of a Data Subject. Such attributes are known as Identifiers. The values for identifiers are provided at the time of adding the Data Subject. Identifiers serve as the basic building block in finding Data Subjects. There are 21 predefined Identifiers.

There are three types of Identifiers within DgSecure.

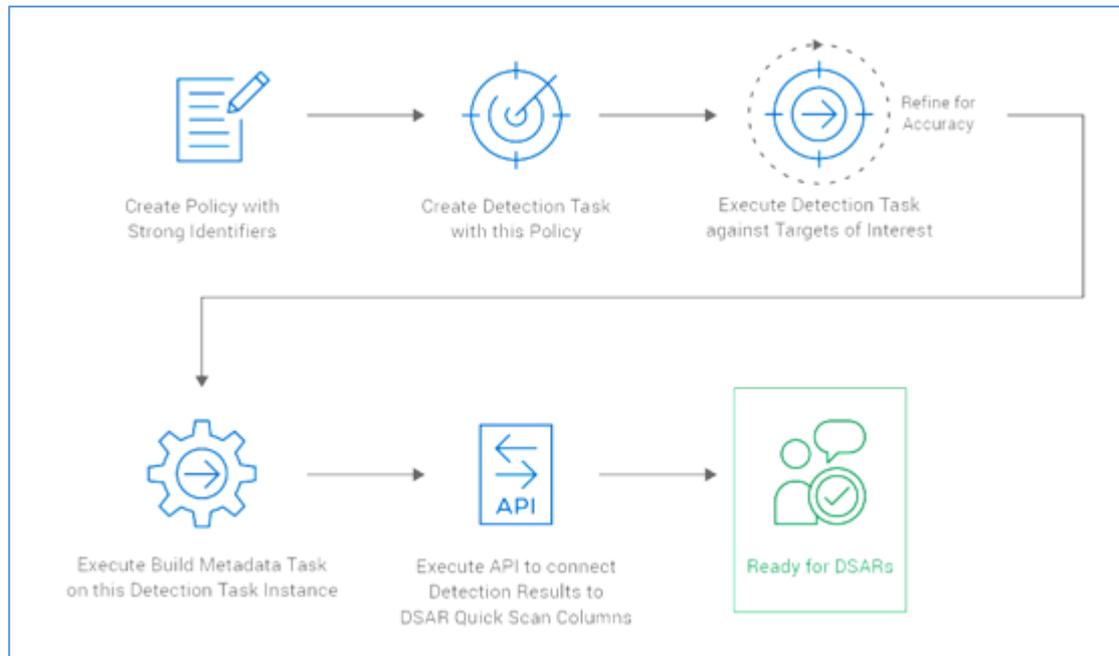
1. **Strong Identifiers:** These identifiers help in identifying a Data Subject uniquely. These strong Identifiers are also the ones required to retrieve information related to the Data Subject multiple sets of systems. Examples of strong identifiers are Credit Card Number, Social Security, and Email Address.
2. **Weak Identifiers:** These identifiers help in further defining a Data Subject but are not by themselves enough to uniquely identify a Data Subject. Examples are Region, Zip Code, and Name.
3. **Mandatory Identifiers:** Mandatory identifiers are mandatory only for the Privacy Screens and not for the Automated DSAR flow. They are used for slicing and dicing the Data Subject information in the Privacy Screens. Mandatory identifiers can be either Strong or Weak. Examples are Region, Country, Zip Code, Department, and Organization.

Name	SSNO	Region	email_id
Jane Doe	09867537	California	jane.doe@gmail.com
John Doe	03567812	New Jersey	john.doe@gmail.com
John Douglas	05648975	New Jersey	john.doe.90@gmail.com

E.g., in the above table Jane Doe, John Doe and John Douglas are the Data Subjects. The attributes of the Data Subjects i.e. their Name, SSNO, Region and email_id are the Identifiers. The email_id and SSNO are the Strong Identifiers since they uniquely identify each Data Subject. The Region is a Weak as well as Mandatory Identifier.

16.3 DSAR Workflow - Overview

The DSAR Workflow diagram provides an overall view of the process to be followed while generating DSAR reports using the Automated DSAR flow.



When a DSAR request comes in to DgSecure, the system has two ways to scan the Data Subject’s data:

1. **Quick Scan:** With this option, the user performs a standard detection with a policy consisting of strong identifiers (as the diagram depicts above). When a DSAR request is processed, only those columns in the tables where strong identifiers are found, will be searched. This is far more efficient and accurate than the alternative, namely, Full Scan, described below.
2. **Full Scan:** Here, based on the strong identifiers specified, DgSecure will scan every column in all relevant Data Sources (see below for a discussion of Data Groups) purely based on the values passed in, and return with the information.

***Note:** The Full Scan option will be deprecated in version 7.2 and will be discontinued in future releases of the DgSecure.

The DSAR workflow diagram above summarizes the process which need to be followed for generating a DSAR report, using Quick Scan. The following are the pre-requisites to configure DSAR:

- Add Identifiers
- Add Connections
- Add Data Groups

These pre-requisite steps are covered in detail in sections 4.2 to 4.4 below.

Based on the above diagram the methodology is as follows:

4. **Create Policy with Strong Identifiers:** The initial step is to create a policy with the Strong Identifiers used for Data Subjects in the Data Group. The

Strong Identifiers are added by including the corresponding Sensitive Types while creating a Policy.

5. **Create Detection Task with the Policy:** Once policy has been defined, the next step is to create a detection task with this policy.
6. **Execute Detection Task:** The columns containing the Strong Identifiers will be detected when we execute the detection task. The results can be refined by tweaking confidence factor parameters, using the remediation workflow to eliminate common false-positive situations, and setting up the threshold for the confidence factor. These topics are covered in the Detection Task section of the DgSecure User Guide.
7. **Execute Build Metadata Task:** After the detection results for the target of interest have been generated, we will schedule and execute the “Build Metadata” task on the detection task instance that was just executed. The “Build Metadata” task copies the list of columns containing the strong identifiers to the database associated with the Privacy IDP.
8. **Execute API:** The last step is to connect the detection results with the Quick Scan columns. This is done by executing a REST API call on the Privacy IDP.

Now, the system is ready to accept DSAR requests and process them efficiently. The above steps are the overall flow for executing DSARs. In the following sections we will go into one-time setup of the Privacy IDP, Data Source Group, Identifiers, and Rules, which are necessary to process DSARs.

16.4 DSAR

Perform the following steps to configure DSAR in DgSecure.

***Note** – Steps 1 to 7 are one-time configuration steps, last 2 steps, Request RoA/RtE and View Report are the ones that will be performed repeatedly as DSAR requests come in.

1. [Add IDPs](#)
2. [Add/Edit Identifier](#)
3. [Add Data Group](#)
4. [Create Connection - DSAR](#)
5. [Add Data Subject](#)
6. [Create Task](#)
7. [Build Metadata and Use API to Populate Quick Scan Columns](#)
8. [Schedule RoA/RtE](#)
9. [Request RoA/RtE](#)
10. [View Report](#)

16.4.1 Add IDPs

To add an IDP, login into the **DgAdmin**. To Access the IDP screen, click **IDP Management > IDPs > Add IDP**. The **IDP Management** panel will appear.

ID	Name	Hostname / IP	Port	IDP Type	Status	SSL Type	Start Time	Decommission	View Log
1	DBMS Detection	192.168.1.65	8889	DBMS Detection IDP	Active	None	02/18/2020 05:00:07	<input type="checkbox"/>	
2	DBMS Masking	localhost	8888	DBMS Masking IDP	Active	None	02/18/2020 04:59:55	<input type="checkbox"/>	
22	Hadoop Detection	localhost	8111	Hadoop Data IDP	Active	None	02/18/2020 04:59:54	<input type="checkbox"/>	
24	Azure Data IDP	52.167.126.253	8111	Azure Data IDP	Inactive	None		<input type="checkbox"/>	
25	Azure Cloud IDP	13.68.17.227	8081	Azure Cloud IDP	Inactive	None		<input type="checkbox"/>	
26	Hive	localhost	9980	Hive IDP	Active	None	02/18/2020 04:59:54	<input type="checkbox"/>	
27	d2	localhost	8889	DBMS Detection IDP	Active	None	02/18/2020 04:59:57	<input type="checkbox"/>	
28	Files	192.168.5.65	8222	File IDP	Inactive	None		<input type="checkbox"/>	

Follow the below steps to add an IDP:

1. Click the **Add IDP** button. The **Add / Edit IDP** dialog box will appear.

The dialog box contains the following fields:

- Name:** Privacy_IDP
- IDP Type:** Privacy IDP (selected from a dropdown menu)
- Hostname / IP:** 192.165.1.35
- Port:** 8884 (selected from a dropdown menu)

Buttons: Close, Save

- I. Enter the descriptive name for the IDP in the **Name** textbox.
- II. Enter the **hostname or IP** address of the IDP’s host machine in the **Hostname/IP** textbox.
- III. Select the **Privacy IDP** as the IDP Type from the drop-down.
- IV. Enter the **port number** which the IDP will use to send and receive information.

2. Click **Save**.

16.4.2 Add/Edit Identifier

To access the Identifiers screen, log into the **DgSecure** Application. Click **Privacy > Configuration > Identifiers > Add Identifier** button. The **Add/Edit Identifier** panel will appear.

PRIVACY MANAGEMENT | CONFIGURATION > ADD/EDIT IDENTIFIERS

IDENTIFIER NAME: BikeStoreDSAREmail

IDENTIFIER DESCRIPTION: Email

IS MANDATORY STRONG IDENTIFIER

SELECT SENSITIVE TYPE: [Dropdown]

SELECT DATATYPE: Select Type [Dropdown]

SELECTED OPERATOR: [Dropdown]

[ADD]

SELECTED IDENTIFIER		
DATATYPE	OPERATOR	ACTIONS
STRING	EQUALS	[Edit] [Delete]

[CANCEL] [UPDATE]

There are 21 pre-defined Identifiers. Below are their names:

First_Name	Driver_License_Number	Home_Address_Street
Middle_Name	Passport_Number	Home_Address_City
Last_Name	Cell_Number	Purchase_Date
NationalID	Home_Phone_Number	Home_Address_State
Username	Product_Purchased	Home_Address_Zip_Code
Other_Info	Country	CustomerID
Email	Credit_Card_Number	Home_Address_County/Province

To add a new identifier, follow the below steps.

1. Enter the name of the identifier in the **Identifier Name** text box.
2. Enter the description of the identifier in the **Identifier Description** text box.
3. If the identifier is mandatory, check the **Is Mandatory** checkbox. At the time of creating a Data Subject, value for these Identifiers must be provided.
4. If the identifier is a strong identifier, check the **Strong Identifier** checkbox. **Strong identifier** is the identifier that value must be present at the time of scanning.

When the **Strong Identifier** checkbox is selected, the **Select Sensitive Type**, **Select DataType**, **Selected Operator** fields and **Selected Identifier** pane will be visible.

The screenshot shows a configuration window with the following elements:

- SELECT SENSITIVE TYPE:** A dropdown menu with "Email Address" selected.
- SELECT DATATYPE:** A dropdown menu with "string" selected.
- SELECTED OPERATOR:** A dropdown menu with "contains" selected, accompanied by an "ADD" button.
- SELECTED IDENTIFIER:** A table with the following structure:

DATATYPE	OPERATOR	ACTIONS
STRING	EQUALS	[edit] [delete]
- Buttons:** "CANCEL" and "SAVE" buttons at the bottom right.

- The **Sensitive Type** field will appear only if you have checked the **Strong Identifier** checkbox. Select the sensitive type in the **Select Sensitive Type** drop-down.

The screenshot shows a list of sensitive types with checkboxes:

- ABA Routing number
- Address
- Bank Account
- Credit Card
- CustomFName
- Dates

- The **Data Type** field will appear only if you have checked the **Strong Identifier** checkbox. Select the data type in the **Select Datatype** drop-down.

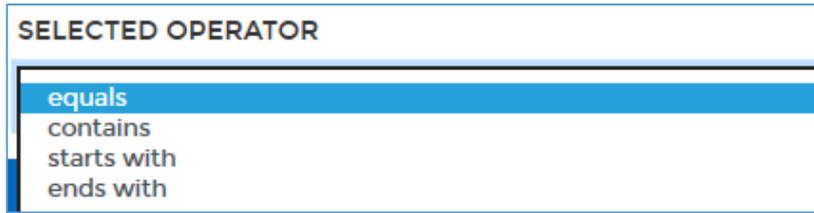
The available data types are: **String, Numeric** and **Date**.

The screenshot shows the "SELECT DATATYPE" dropdown menu with the following options:

- Select Type
- string
- numeric
- date

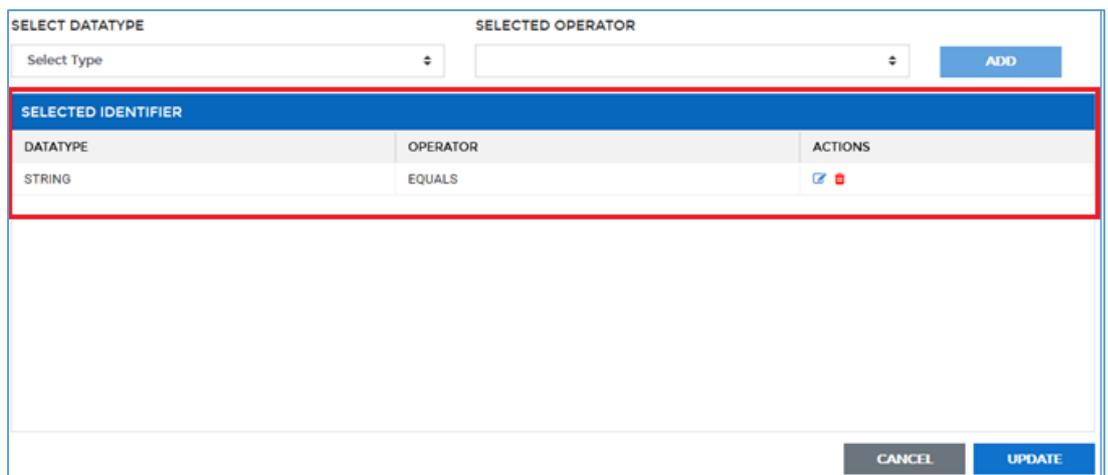
- The **Selected Operator** field will appear only if you have checked the **Strong Identifier** checkbox. Select the operator in the **Selected Operator** drop-down.

The list will display the available operators based on the selected **Data Type**.



In the above screenshot, the listed operators are displayed when **Data Type = 'String'** is selected.

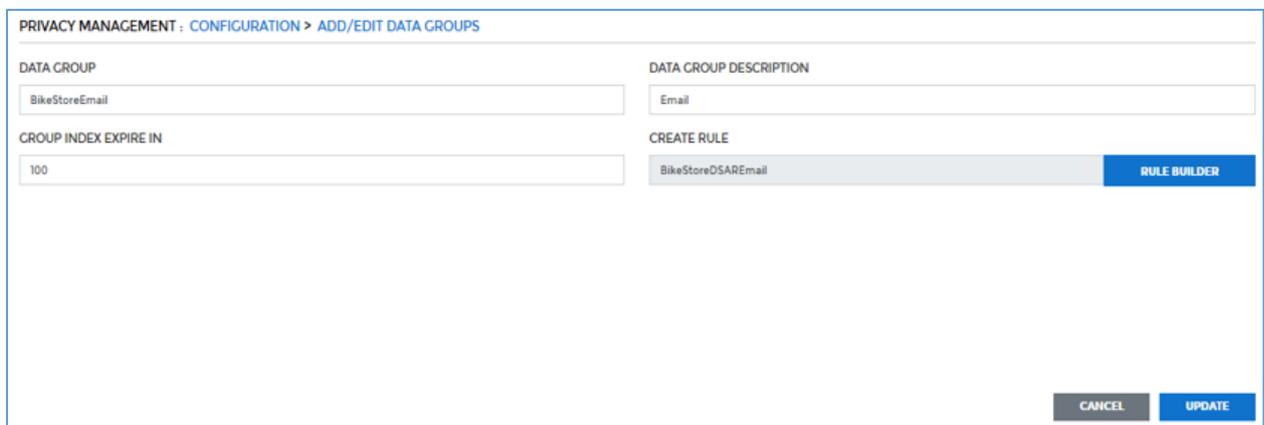
8. Click the **Add** button to add the datatype. The added datatypes will be available in the **Selected Identifier** panel.
9. Click **Save**.



To edit a Datatype or Operator, click . To delete a DataType or Operator, click .

16.4.3 Add Data Group

To access the **Data Groups** screen, click **Privacy > Configuration > Groups > Add Group** button. The **Add/Edit Data Group** screen will appear.



To add a new **Data Group**, perform the following steps:

1. Enter the name of the Data Group in the **Data Group** text box.
2. Enter description of the Data Group in the **Data Group Description** text box.
3. Enter the expiry days of the Data Group in the **Group Index Expire In** text box. The **Group Index Expire In** is the expiration time (in days) of this group’s Data Subject-related information (including transactional data), stored in the Privacy IDP’s database.

If a Group’s data expires, then existing scan results (which have not yet been approved by the DPO) for Data Subjects in that Group become invalid and the status of these Data Subjects changes to **Incoming** state.

4. To create the rule for the Data Group in the **Create Rule** text box. Click **Rule Builder** button.

- i. Select at least one strong identifier in the **Select Identifier** drop-down.
- ii. The query will be displayed in the **Rule Query** text box.
- iii. Click **Save**.

5. Click the **Save** button. This will save the Data Group and it will be available on the **Data Group** screen.

PRIVACY MANAGEMENT : CONFIGURATION > DATA GROUPS ADD GROUP

ID	GROUP NAME	GROUP DESCRIPTION	GROUP EXPIRY	RULE	ACTION
4	BikeStoreEmail	Email	100	BikeStoreDSAREmail	
3	FNLNPrivacy	FirstName LName for Privacy	100	(FName_Privacy AND LName_Privacy)	
2	Email	Email	100	Email_DSAR	
1	FNameLName	FName and LName	100	(FName_DSAR AND LName_DSAR)	

To view the details of a Data Group, click . To edit a Data Group, click . To delete a Data Group, click .

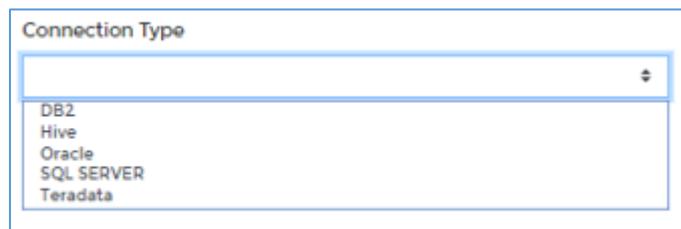
The Rules are used when a search for identities is being executed on the target data stores, in response to a DSAR request.

16.4.4 Create Connection - DSAR

For each **Group** created as described in the above sections, create **Connections** to all databases within which the **RoA** and/or **RtE** scans have to be executed.

To add a connection, click **Privacy > Configuration > Connection Manager > Add New Connection** tab.

1. Select the connection type from the **Connection Type** drop-down. It displays all the supported database types: Oracle, DB2, Hive, Oracle, Teradata and SQL.



2. Enter the **Connection Name**. This field accepts letters, numbers, symbols, and holds up to 256 characters. The name must be unique for each individual connection.

***Note** – The **Connection Type** and **Connection Name** cannot be edited once created.

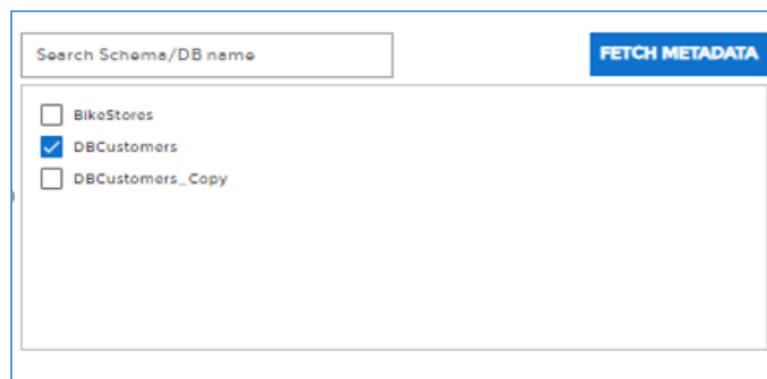
3. Enter the **Hostnames** and **IP address** of the database.
4. Enter the **Port Number**.
5. Select the **Data Group**.
6. Enter the database **User Name**.
7. Enter the database **Password**.
8. Enter the **IP address**.
9. Select the **Authentication Method** according to the Connection Type.

***Note** – The **SQL Server**, **Windows Impersonation** and **Windows** authentication method are available when **SQL server** is selected as **Connection Type**.

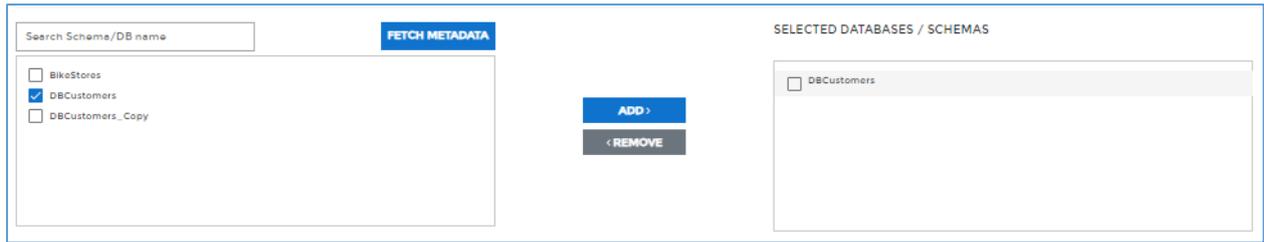
Following are the available authentication methods:

- **SQL Server**
- **Windows Impersonation**
- **Windows**

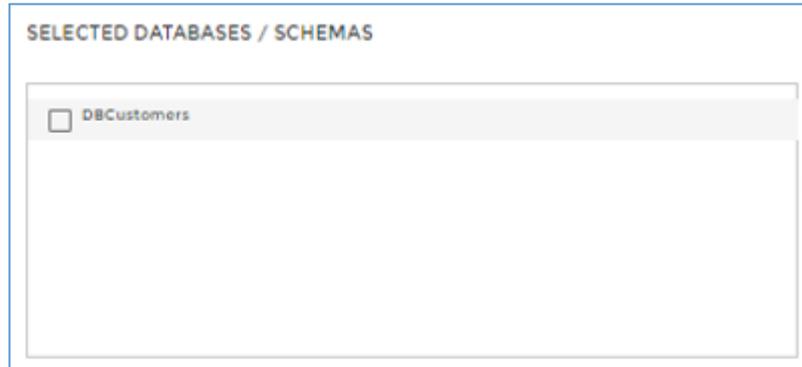
10. Enter the **URL**.
11. Select the **Type** either 'Basic' or 'TNS'.
12. Enter the **Service ID**.
13. Enter the **Service Name**.
14. After completing all entries, click **Fetch Metadata**. A list of all databases on the server will appear under the **Filter by Database** pane.



15. Check the checkbox next to the database name to select it. Click the **Add** button to add the databases in **Selected Database/Schemas panel**.



16. The selected databases will be added to the **Selected Databases/Schemas** list.



17. Click **Remove** if you wish to remove any selected Databases from the **Selected Database/Schemas** panel.

18. Click the **Save** button to save the changes.

19. The **Connection** is now ready. You can view the list of all connections in **Connection Manager** screen.

PRIVACY MANAGEMENT : CONFIGURATION > CONNECTION MANAGER

ALL CONNECTIONS		ADD NEW CONNECTION			
ID	CONNECTION NAME	CONNECTION TYPE	HOST NAME/URL	CONNECTION IDP	ACTIONS
9	CustomerDB	SQL Server	34.223.113.94	GDPR	
8	BikeStoreDSAR	SQL Server	34.223.113.94	GDPR	
7	SQLServerQuickscan	SQL Server	35.239.136.84	GDPR	
6	TeradataEmail	Teradata	153.64.73.15	GDPR	
5	SQLServerEmail	SQL Server	35.239.136.84	GDPR	
3	SQLServer	SQL Server	35.239.136.84	GDPR	

To view the Connection details, click . To edit a Connection, click . To delete a Connection, click .

16.4.5 Add Data Subject

To access **Data Subjects** screen, click **Privacy > Configuration > Data Subjects > New Data Subject** tab. The **New Data Subject** panel will appear.

PRIVACY MANAGEMENT / CONFIGURATION / DATA SUBJECTS

DATA SUBJECTS NEW DATA SUBJECT

Manual File with known format File with unknown format

DATA SUBJECT NAME * OTHER INFO DATA GROUP *

STRONG IDENTIFIERS

IDENTIFIER	DATATYPE	OPERATOR	VALUE
Email *	STRING	EQUALS	<input type="text" value="danette_williams@illinois.com"/>

IDENTIFIER ATTRIBUTES

NAME	DESCRIPTION	VALUE
First_Name *	First Name	<input type="text" value="Danette"/>
Last_Name *	Last Name	<input type="text" value="Williams"/>
Country *	Country	<input type="text" value="UK"/>

OBLIGATIONS

OBLIGATION NAME	VALUE
GDPR-Compliant	<input type="text" value="True"/>
CCPA-Compliant	<input type="text" value="False"/>
esdsadd	<input type="text" value="False"/>

There are three ways in which you can add new **Data Subject**. These are:

1. Manual

2. Upload File with Known Format – Files with Known Format are files that conform to the standard (or sample) format for Data Subjects supported by DgSecure. Format here refers to the sequence of identifiers in the file. XLS, XLSX, and CSV files of Known Format are supported.

3. Upload File with Unknown Format – Files with Unknown Format may have the identifiers in any sequence. They do not conform to the standard (or sample) format supported by DgSecure. The association between the headers in the file and identifiers is made after the upload. Only CSV files of Unknown format are supported.

To add a new Data Subject, follow the below steps:

- 1. Manual:** Select the **Manual** option to provide all the required information manually. Perform the following steps to create a Data subject manually:

PRIVACY MANAGEMENT / CONFIGURATION / DATA SUBJECTS

DATA SUBJECTS **NEW DATA SUBJECT**

Manual
 File with known format
 File with unknown format

DATA SUBJECT NAME *
 OTHER INFO
 DATA GROUP *

STRONG IDENTIFIERS

IDENTIFIER	DATATYPE	OPERATOR	VALUE
Email *	STRING	EQUALS	<input type="text" value="danette_williams@ilois.com"/>

IDENTIFIER ATTRIBUTES

NAME	DESCRIPTION	VALUE
First_Name *	First Name	<input type="text" value="Danette"/>
Last_Name *	Last Name	<input type="text" value="Williams"/>
Country *	Country	<input type="text" value="UK"/>

OBLIGATIONS

OBLIGATION NAME	VALUE
GDPR-Compliant	<input type="text" value="True"/>
CCPA-Compliant	<input type="text" value="False"/>
esdsedd	<input type="text" value="False"/>

- I. Enter the name of the Data Subject in the **Data Subject Name** text box.
- II. Enter extra information of the Data Subject in the **Other Info** text box, if any.
- III. Select the Data Group from the **Data Group** drop-down. This will display the list of **Strong Identifier** and **Identifier Attributes** associated with the selected Data Group.
- IV. Enter the values of the **Strong Identifiers**. This field will appear only if you are creating the Data Subject manually.

STRONG IDENTIFIERS

IDENTIFIER	DATATYPE	OPERATOR	VALUE
FName_DSAR *			
	STRING	EQUALS	<input type="text" value="Shivani"/>
LName_DSAR *			
	STRING	EQUALS	<input type="text" value="Gupta"/>

- V. Enter the value of the mandatory Identifiers in the **Identifier Attribute** list.

IDENTIFIER ATTRIBUTES		
NAME	DESCRIPTION	VALUE
First_Name *	First Name	<input type="text" value="Shivani"/>
Last_Name *	Last Name	<input type="text" value="Gupta"/>
Email *	Personal Email Address	<input type="text" value="gupta.shivani1402@gmail.com"/>
Country *	Country	<input type="text" value="India"/>

VI. Select the value for Obligation from either **TRUE** or **FALSE**.

OBLIGATIONS	
OBLIGATION NAME	VALUE
GDPR-Compliant	<input type="text" value="True"/>
CCPA-Compliant	<input type="text" value="False"/>

VII. Click **Save**.

2. **Upload File with Known Format:** Select the **File with Known Format** option to upload the information from the file.

Perform the below steps to create the Data Subject by uploading a file.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS NEW DATA SUBJECT

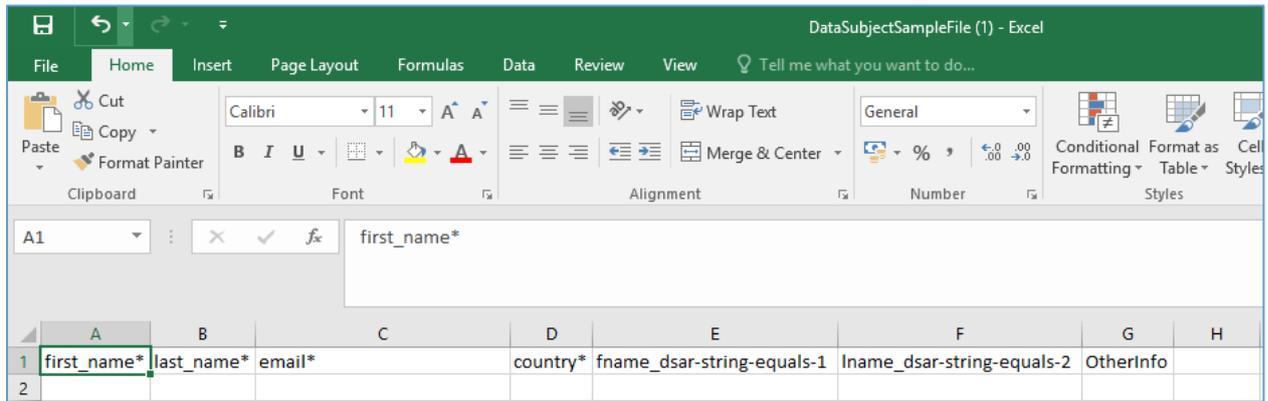
Manual
 File with known format
 File with unknown format

DATA GROUP:
 DATA SUBJECT FILE:
BROWSE UPLOAD

CANCEL
DOWNLOAD SAMPLE FILE
DOWNLOAD LOG FILE

- I. Select the Data Group from the **Data Group** drop-down.
- II. To download a sample file, click the **Download Sample File** button in the bottom of the screen. This file contains the column headers for the selected Data Group, based on which you will enter the data.

***Note – The Download Sample File button will be enabled only when you select a Data Group from the Data Group drop-down.**

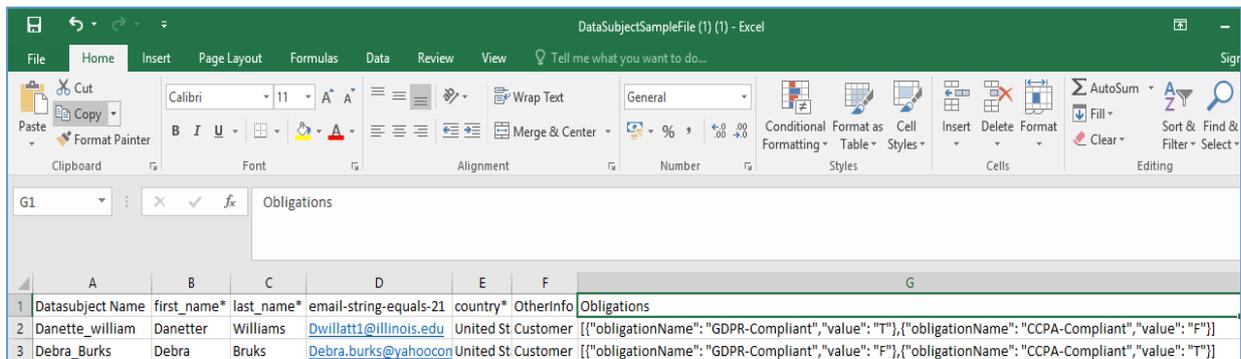


- III. Fill the data in the downloaded sample file as per the specified format and save it on your local machine.

While uploading the Data Subject information through file, specify the Obligation format as defined below:

```
[{"obligationName": "GDPR-Compliant", "value": "F"},
{"obligationName": "CCPA-Compliant", "value": "T"}]
```

Where 'T' stands for True and 'F' for false



- IV. Click the **Browse** button to upload the saved file from your local machine.

***Note – Only xlsx, xls and csv file are allowed.**

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS **NEW DATA SUBJECT**

Manual File with known format File with unknown format

DATA GROUP: BikeStoreEmail DATA SUBJECT FILE: DataSubjectSampleFile.csv **BROWSE** **UPLOAD**

Note : Last data subject saved successfully. please download log file for more details.

- V. Click the **Upload** button. The data for **Data Subject** will be available on the **Data Subjects** tab.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS					NEW DATA SUBJECT	DELETE
	FIRST NAME	GROUP NAME	OTHER INFO	PDF STATUS	ACTIONS	
<input type="checkbox"/>	Danette_Willatt	Email	Customer	No Data		
<input type="checkbox"/>	Debra_Burks	BikeStoreEmail	Customer	Generate Report		
<input type="checkbox"/>	Sandeep_Singh	Email	Telemarket	Generate Report		
<input type="checkbox"/>	Roberto_Tamburello	Email	Finance	Erased		
<input type="checkbox"/>	Ken_Sanchez	FNameLName	Admin	Generate Report		
<input type="checkbox"/>	Terri_Duffy	FNLNPrivacy	Analyst	No Data		

- 3. **Upload File with Unknown Format:** Select the **File with Unknown Format** option to upload information from an unknown file format.

Perform the below steps to create Data Subject by uploading a file format.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

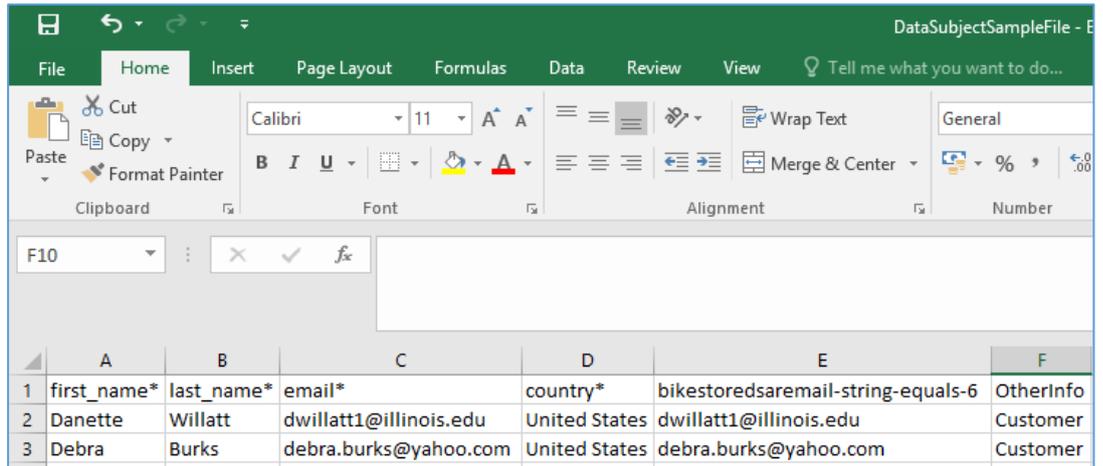
DATA SUBJECTS **NEW DATA SUBJECT**

Manual File with known format File with unknown format

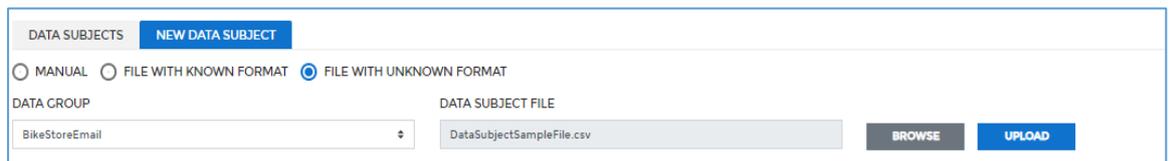
DATA GROUP: FNameLName DATA SUBJECT FILE: DataSubjectSampleFile.csv **BROWSE** **UPLOAD**

- I. Select the Data Group from the **Data Group** drop-down.
- II. Create a csv file on your local machine which will contain the details of **Data Subject**.

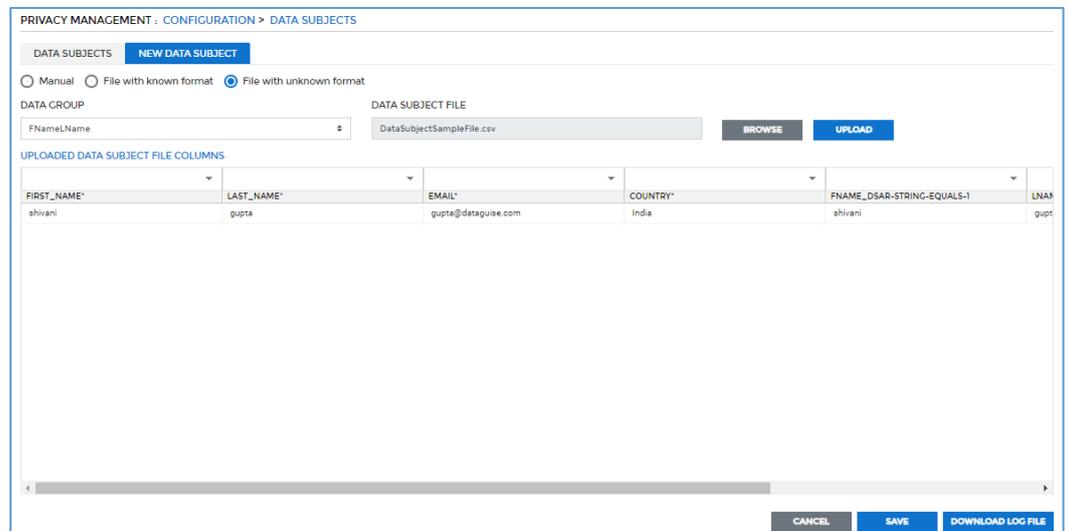
***Note – Only csv file is allowed.**



III. Click the **Browse** button to search the file from your local machine.



IV. Click the **Upload** button to the csv file. The uploaded data will appear in the **Uploaded Data Subject File Columns** panel.



V. Once the file is uploaded then tag the file columns headers with the **Strong Identifiers**.

***Note** - Identifiers in '**bold**' are strong Identifiers. Identifiers with '*' are mandatory Identifiers. Identifiers in '**bold**' with '*' are both strong and mandatory Identifiers.

To tag file columns with the **Strong Identifiers**.

- a. Click on the drop-down above of the column headers.

UPLOADED DATA SUBJECT FILE COLUMNS

FIRST_NAME*	LAST_NAME*	EMAIL*	COUNTRY*	BIKESTORESAREMAIL-STRING-EQUAL-6
Shivani	Gupta	shivani.gupta@gmail.com	India	shivani.gupta@bike.com
Tania	Thakur	tania.thakur@gmail.com	India	tania.thakur@bike.com
Gunjan	Bhatnagar	gunjan.bhatnagar@gmail.com	India	gunjan.bhatnagar@bike.com

- b. Select the **Strong Identifier** from the given list.

UPLOADED DATA SUBJECT FILE COLUMNS

FIRST_NAME*	LAST_NAME*	No Selection	COUNTRY*	BIKESTORESAREMAIL-STRING-EQUAL-6
Shivani	Gupta	First_Name *	India	shivani.gupta@bike.com
Tania	Thakur	Middle_Name	India	tania.thakur@bike.com
Gunjan	Bhatnagar	Last_Name *	India	gunjan.bhatnagar@bike.com

- VI. Click the **Save** button to save the changes. The saved Data Subjects will be displayed on the **Data Subjects** tab.

PRIVACY MANAGEMENT > CONFIGURATION > DATA SUBJECTS

	FIRST NAME	GROUP NAME	OTHER INFO	PDF STATUS	ACTIONS
<input type="checkbox"/>	Shivani	BikeStoreEmail	Customer	No Data	✖ GF
<input type="checkbox"/>	Gunjan	BikeStoreEmail	Customer	No Data	✖ GF
<input type="checkbox"/>	Tania	BikeStoreEmail	Customer	No Data	✖ GF
<input type="checkbox"/>	Danette_William	Email	Customer	No Data	✖ GF
<input type="checkbox"/>	Debra_Burks	BikeStoreEmail	Customer	Generate Report	✖ GF
<input type="checkbox"/>	Sandeep_Singh	Email	Telemarket	Generate Report	✖ GF
<input type="checkbox"/>	Roberto_Tamburello	Email	Finance	Erased	✖ GF
<input type="checkbox"/>	Ken_Sanchez	PhoneNumber	Admin	Generate Report	✖ GF
<input type="checkbox"/>	Terri_Duffy	PIINPrivacy	Analyst	No Data	✖ GF

16.4.5.1 Add Obligation

To access **Add New Obligation** screen, click **Privacy > Configuration > Obligations > Add New Obligation** tab. The **Add New Obligation** panel will appear.

PRIVACY MANAGEMENT / CONFIGURATION / OBLIGATION

OBLIGATION ADD NEW OBLIGATION DELETE

Obligation Name: Email

Description: Obligation for sending emails

Expire In (Years): 1

Other Information

CANCEL SAVE

To add Obligation, follow the below steps:

1. Enter a unique name for Obligation. This field supports both the numeric and character values.

195. Enter the description for the Obligation.

196. Enter the numeric value for **Expire In (years)**. Once Obligation expires then you will not be able to send any notification to the user.

197. To enter **Other Information**, click the  next to the heading on right side of the screen. This will open the Other Information panel.

PRIVACY MANAGEMENT / CONFIGURATION / OBLIGATION

OBLIGATION ADD NEW OBLIGATION DELETE

Obligation Name: Email

Description: obligation for sending email

Expire In (years): 1

Other Information

Enter the information for the **Name** and **Value** field under **Other Information** panel.

- i. To delete Name and Value field, click  button.
- ii. To add more fields for Name and Value, click **Add** button.

- 198. Click **Save** to make the changes effective.
- 199. Click **Cancel**, if you do not want to save the changes.

16.4.6 Create Task

The next step is to create a task for scanning the data.

To access the **Task** screen, click **Privacy > Configuration > Task > Add Task** button.

1. Enter the **Task Name** and **Task Description**. The Task Name accepts both numeric and character value.
2. Select the **Scan Type**. There are two ways in which scanning can be done.
 1. **Quick Scan**: This option results in only the columns in the Quick Scan column list being scanned. The Quick Scan columns are populated from

the discovered columns during either a Detection Task execution (described in the beginning of this document) or during a **Full Scan** execution. You can also manually add columns to the list of Quick Scan columns. This feature reduces the scan time.

***Note:** To create the Quick Scan columns from the Detection Task results, visit Section 18.4.7, Build Metadata and Use API to Populate Quick Scan Columns.

2. **Full Scan:** This option scans the entire database irrespective of the user selection.
3. Select the **Data Group** from the drop-down. The list of all the Data Groups will appear in the drop down.
4. Connections for the selected Data Group will be displayed in the **Connection** panel. Select the required connections.
5. **Save** the Task. The target data sources will be the ones that are included via the connections in this task.
6. The task will now be listed in the list of tasks under the **Tasks** screen.

PRIVACY MANAGEMENT : [CONFIGURATION](#) > [TASKS](#)

ID	TASK NAME	TASK DESCRIPTION	ACTION
5	BikeStoreDSARFull	Full Scan	
4	SQLServerFNLNQuick	SQL Server - GROUP: FNameLName_Privacy	
3	Teradata_Email	Teradata Email	
2	SQLServerFullEmail	SQL Server Email Full Scan	
1	SQLServerFull	SQLServer Full	

[ADD TASK](#)

To view the Task Details, click . To edit details of a Task, click .

16.4.7 Build Metadata and Use API to Populate Quick Scan Columns

This step populates the IDP’s repository with detected results metadata using the Build Metadata Task, and then subsequently uses a REST API call to connect the metadata to Quick Scan columns

Before executing this API, perform the following steps:

1. [Create Detection Connection](#)
2. [Create and execute the Detection Task](#)

To populate the Quick Scan columns, perform the below steps:

16.4.7.1 Add System (Optional Step)

System Information adds metadata about each target system. This an optional step for Automated DSARs, whereas it is mandatory for the Privacy Screens. This is addressed

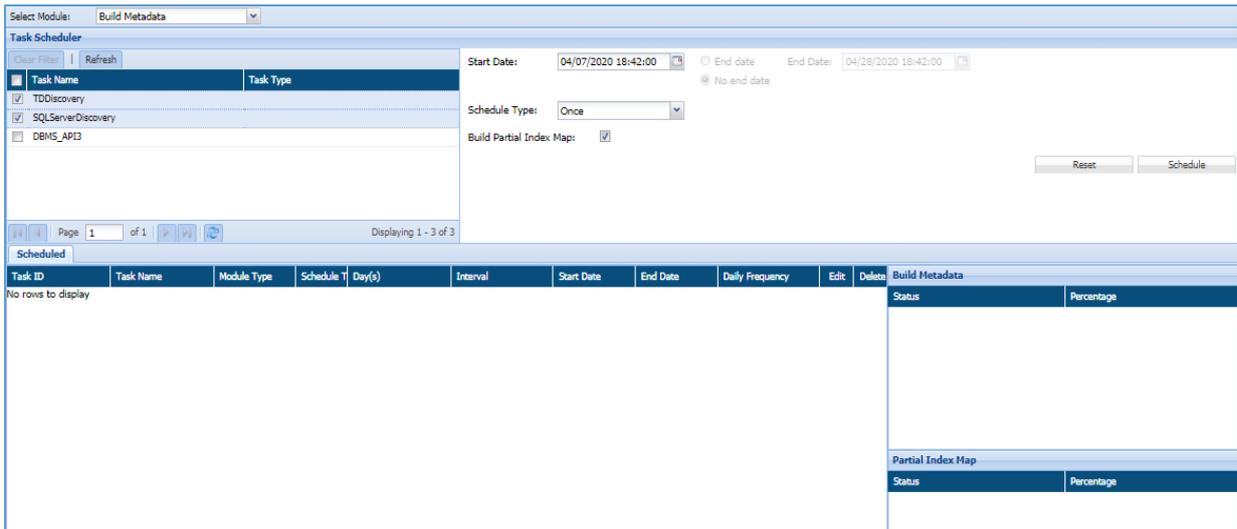
in detail in the Privacy section of the DgSecure User Manual.

16.4.7.2 Create and Execute the Build Metadata Task

The Build Metadata task is used to push the DBMS discovery results from the controller to the IDP’s repository.

To create a **Metadata**:

1. Go to the **Scheduler** and select the module **Build Metadata** from the **Select Module** drop-down.



2. Select the required task.
3. Select the task schedule start date.
4. Select the **end date** for a task. If the task is scheduled forever, select **no end date**.
5. Select the schedule type. The available options are:
 3. **Weekly**: Select this option to perform the task on the selected days of the week.
 - I. Select the days of the week to perform the task.
 - II. Select the daily frequency of the task. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - III. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
 4. **Hourly**: Select this option to perform the task in every few hours.
 1. Select the time interval.
 2. **Monthly**: Select this option to perform the task on the selected days of the month.
 3. Select days of one or more months to perform the task.

4. Select a specific weekday of one or more months to perform the task.
 5. Select the frequency of the task. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 6. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
5. **Daily:** Select this option to perform the task daily.
7. Select the frequency of the task in the **Recurs.**
- IV. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - V. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
6. **Once:** Select this option to perform the task only once.
6. Click the **Schedule** button to complete the action.
 7. Scheduled tasks will appear in the bottom panel. You can reschedule a task by editing them from this screen. You can delete a task also.

16.4.7.3 *Populate the Quick Scan columns*

Use the below API to populate Quick Scan columns from the detection results which are copied by Build Metadata. This API will create the Quick Scan columns internally.

URL:

http://hostname:port/GDPRAgent/services/GDPRAgent/buildDsarQuickScanColumns?controllerid=controllerId

The “controllerId” is generated while installing DgSecure.

Sample Call:

<http://localhost:8080/GDPRAgent/services/GDPRAgent/buildDsarQuickScanColumns?controllerid=dataguise>

***Note:** Host and port should be of Privacy IDP.

Request Body:

```
[{
    "discoverConnId":3,
    "privacyConnId":12,
    "privacyDataGroupName":"Customer_1"
},
```

```
{
    "discoverConnId":16,
    "privacyConnId":12,
    "privacyDataGroupName":"Human_Resource"
}]
```

CURL Command:

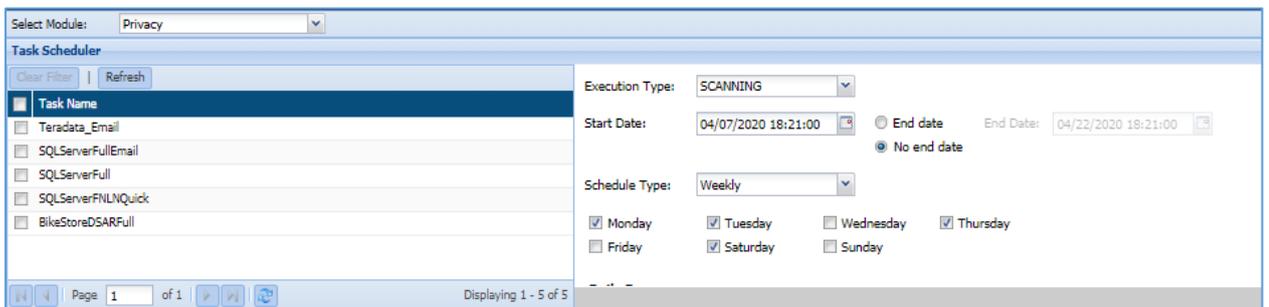
```
[root@azuredataguise-VM dataguise]# curl --location --request POST
'http://localhost:8884/GDPRAgent/services/GDPRAgent/buildDsarQuickScan
Columns?controllerid=dataguise' --header 'Content-Type: application/json'
--data
' [{"discoverConnId":7,"privacyConnId":8,"privacyDataGroupName":"grp1"} ]'
{"message":"Quick Sacn Columns updated
sucessfully","response":null,"result":"SUCCESS","resultCode":null}
```

16.4.8 Schedule RoA/RtE

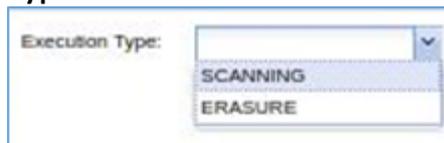
Once the **task** and the **request** are ready the next step is to schedule the task.

To schedule a task in **DSAR**, access the **Scheduler** screen. Click **Scheduler** from the side menu.

1. Go to the **Scheduler** Screen.
2. Select **Privacy** from the **Select Module** drop-down.



3. Select the required **Task**.
4. Select the **Execution Type** either **Scanning** or **Erasure**. For **RoA**, select **Scanning** as the **Execution Type**. For **RtE**, select **Erasure** as **Execution Type**.



5. Select the **Start Date** and the **End Date**.

6. Select the **Schedule Type** from the drop-down. The available options are:
 7. **Weekly**: Select this option to perform the task on the selected days of the week.
 - i. Select the days of the week to perform the task on specific days.
 - ii. Select the daily frequency of the task. Select **Occurs Once at** and mention the time, if you want to perform the task once in a day.
 - iii. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
 8. **Hourly**: Select this option to perform the task in every few hours.
 - i. Select the time interval.
 9. **Monthly**: Select this option to perform the task on the selected days of the month.
 - i. Select days of one or more months to perform the task.
 - ii. Select a specific weekday of one or more months to perform the task.
 - iii. Select the frequency of the task. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - iv. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
 10. **Daily**: Select this option to perform the task daily.
 - i. Select the frequency of the task in the **Recurs**.
 - ii. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - iii. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
 11. **Once**: Select this option to perform the task only once.
7. Click on the **Schedule** button to complete the action.
8. The list of scheduled tasks will appear in the **Scheduler** tab in the bottom panel.

Task ID	Task Name	Module Type	Schedule T	Day(s)	Interval	Execution Type	Start Date	End Date	Daily Frequency
5	BikeStoreDSARFull	GDPR	Weekly	Thursday		SCANNING	2020-02-20 15:...		Occurs once at: 00:00
4	SQLServerFNLNQuick	GDPR	Weekly	Wednesday		SCANNING	2020-02-05 10:...		Occurs once at: 00:00
3	Teradata_Email	GDPR	Weekly	Monday		SCANNING	2020-01-20 11:...		Occurs once at: 00:00
2	SQLServerFullEmail	GDPR	Weekly	Monday		ERASURE	2020-01-20 09:...		Occurs once at: 00:00
1	SQLServerFull	GDPR	Weekly	Monday,Saturday		SCANNING	2020-01-20 09:...		Occurs once at: 00:00

Page 1 of 1 Displaying 1 - 5 of 5

At this point, the system is ready to receive DSAR requests.

16.4.9 Request RoA/RtE

RoA and RtE Requests can be either manually entered or can be generated via the API, which in turn will be called by a front-end intake UI. The front-end intake UI will typically be designed by the organization implementing the automated DSAR flow, and will be embedded in an authenticated customer- or employee-facing portal.

To generate a request for RoA/RtE, perform the below steps.

16.4.9.1 Submit the Request Manually

To request RoA/RtE manually, click **Privacy > DSAR Workflow > DSAR Applications** screen.

1. Click the **DSAR Application** screen.

PRIVACY MANAGEMENT : DSAR APPLICATION								RESET	SUBMIT
DATA SUBJECT NAME	DATA GROUP	OTHER INFO	LAST ROA DATE	LAST RTE DATE	PRE-FETCH DATE	<input type="checkbox"/> ROA	<input type="checkbox"/> RTE	<input type="checkbox"/> PRE-FETCH DATA	ACTIONS
Danette_Willatt	Email	Customer	2020-02-21 02:20:56			<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Debra_Burka	BikeStoreEmail	Customer	2020-02-21 23:34:31		2020-02-20 14:25:58	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
Sandeep_Singh	Email	Telemarket	2020-02-20 14:25:28			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Roberto_Tamburello	Email	Finance	2020-01-20 22:36:10	2020-01-20 22:57:01		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Ken_Sanchez	FNameLName	Admin	2020-02-21 02:20:56			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Terri_Duffy	FNLNPrivacy	Analyst	2020-02-05 23:58:15			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

2. Select the **RoA** or **RtE** checkbox corresponding to the Data Subjects for which you want to perform **RoA/RtE**. Click the **Save** button.
3. On saving the request, the **Status** of the Data Subjects will be in **Incoming** state in **RoA/RtE** screen under **DSAR Workflow**. This screen is typically visible to the DPO.

4. The **Status** of Data Subject in **RoA/RtE** are:

- i. **No Data:** If the status is **No Data**, it means that no data was found in the target database when scanning was completed.

200. **Erased:** If the status is **Erased**, it means that the data of selected Data Subject has been deleted in the target data stores.

201. **Review Pending:** If status is **Review Pending**, it means DPO or Admin have to review the data after report has been generated. Once reviewed, they can approve it.

202. **Incoming:** If status is **Incoming**, it means the Data Subject is scheduled for processing. When the scheduled task gets triggered, the status of Data Subject will change to Incoming state.

PRIVACY MANAGEMENT : DSAR WORKFLOW > ROA							DISAPPROVE	APPROVE	REJECT	FREEZE	RESUBMIT
	DATA SUBJECT NAME	OTHER INFO	STATUS	DATA GROUP	ROA DATE	DATA	ACTIONS				
<input type="checkbox"/>	Danette_Willatt	Customer	No Data	Email	2020-02-21 02:20:56	No Data					
<input type="checkbox"/>	Debra_Burks	Customer	Review Pending	BikeStoreEmail	2020-02-21 23:34:31	Generate Report					
<input type="checkbox"/>	Sandeep_Singh	Telemarket	Review Pending	Email	2020-02-20 14:25:28	Generate Report					
<input type="checkbox"/>	Roberto_Tamburello	Finance	Erased	Email	2020-01-20 22:36:10	Erased					
<input type="checkbox"/>	Ken_Sanchez	Admin	Review Pending	FNameLName	2020-02-21 02:20:56	Generate Report					
<input type="checkbox"/>	Terri_Duffy	Analyst	Review Pending	FNLNPrivacy	2020-02-05 23:58:15	Generate Report					

PRIVACY MANAGEMENT : DSAR WORKFLOW > RTE							DISAPPROVE	APPROVE	REJECT	FREEZE	RESUBMIT
	DATA SUBJECT NAME	OTHER INFO	STATUS	DATA GROUP	LAST RTE DATE	DATA	ACTIONS				
<input type="checkbox"/>	Roberto_Tamburello	Finance	Erased	Email	2020-01-20 22:57:01	Erased					

5. The DPO has rights to approve, disapprove, reject, freeze, or resubmit a frozen request, by clicking the buttons provided on the top of the screen.

1. **Disapprove:** Click the **Disapprove** button, if RoA/RtE report is approved mistakenly. A DPO can disapprove a report. The Data Subject’s status will change to **Review Pending**, once it has been disapproved.
2. **Approve:** Click the **Approve** button, if the RoA report is valid. The DPO approves the scanned data. For **RtE**, approve means ready to be erased. The step after this is that i) in the case of an RoA, a report (which could be custom modified) will be sent back to the source (typically the same portal where the web intake page was

located) or ii) in the case of an RtE, the data found is deleted from the target data stores.

***Note on final report format served to end user:**

If the incoming requests are generated via the intake API which is integrated with a web intake portal, custom coding and interfacing with the web intake portal will be required for step i) above, because the final format of the report is typically different for each organization.

3. **Reject:** Click the **Reject** button to reject the RoA/RtE request. Data Subject Requests with status as **Review Pending** can be rejected.
4. **Freeze:** Click the **Freeze** button, if the RoA/RtE report contains false positive results. The DPO can freeze the request. Downstream processing will not take place in this case.
5. **Resubmit:** Click the **Resubmit** button, if the Data Subject's report status is **Frozen** or **In-Error**.
 - i. In **RoA**, when the **Resubmit** button is clicked the status of the Data Subject will change to **Incoming** state i.e. ready to be scanned, if it was in **Frozen** or **In-Error** state.
 - ii. In **RtE**, when the **Resubmit** button is clicked the status of the Data Subject will change to **Review Pending** i.e. ready to be approved in order to be erased, if it was in **In-Error** state.

16.4.9.2 Submit the Request via API (integration with Intake Web Page)

A request for RoA/RtE using can be submitted using the API described below. This API will submit a request for RoA or RtE. This API is typically called by an Intake button on an authenticated web page from where a logged in Data Subject can make an RoA or RtE request.

Endpoint URL:

http://host:port /dgcontroller/services/v1/products/Privacy/request/

Sample Call:

<http://localhost:8088/dgcontroller/services/v1/products/Privacy/request/>

Header Parameters:

sessionId

Request Body:

{

```

"requestType":"ROA",
"identifierValues":["rob0@adventure-works.com"]
}

```

Here “requestType” can be either RoA or RtE.
 “identifierValues” are the values of Strong Identifiers.

Response Body:

It will provide a request ID which user can use for generating and downloading the report.

16.4.10 View Report

The last step is to view the report. The report can be generated or downloaded either manually or via API.

16.4.10.1 View the Report Manually

You can view the processed report in:

1. Data Subject
2. RoA (Request of Access)

1. Data Subject:

To view the report in the Data Subject screen, click **Privacy > Configurations > Data Subjects** tab.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS						
DATA SUBJECTS					NEW DATA SUBJECT	DELETE
<input type="checkbox"/>	FIRST NAME	GROUP NAME	OTHER INFO	PDF STATUS	ACTIONS	
<input type="checkbox"/>	Danette_Willatt	Email	Customer	No Data		
<input type="checkbox"/>	Debra_Burks	BikeStoreEmail	Customer	Generate Report		
<input type="checkbox"/>	Sandeep_Singh	Email	Telemarket	Generate Report		
<input type="checkbox"/>	Roberto_Tamburello	Email	Finance	Erased		
<input type="checkbox"/>	Ken_Sanchez	FNameLName	Admin	Generate Report		
<input type="checkbox"/>	Terri_Duffy	FNLNPrivacy	Analyst	Generate Report		

Items per page: 20 1 - 6 of 6 << >> <>

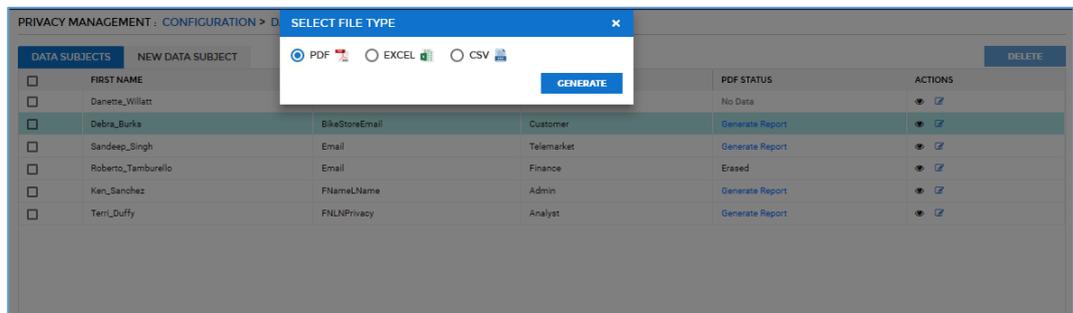
This screen will display the basic details of the Data Subject such as **First Name, Group Name, Other Information of the Data Subject, PDF Status**.

To view the details of a Data Subject, click . To edit a Data Subject, click .

You can also generate report for each of the Data Subject by clicking on the **Generate Report** in the PDF Status column. Perform the following steps to generate a report for the Data Subject.

DATA SUBJECTS					NEW DATA SUBJECT	DELETE
<input type="checkbox"/>	FIRST NAME	GROUP NAME	OTHER INFO	PDF STATUS	ACTIONS	
<input type="checkbox"/>	Danette_Willatt	Email	Customer	No Data		
<input type="checkbox"/>	Debra_Burks	BikeStoreEmail	Customer	Generate Report		
<input type="checkbox"/>	Sandeep_Singh	Email	Telemarket	Generate Report		
<input type="checkbox"/>	Roberto_Tamburello	Email	Finance	Erased		
<input type="checkbox"/>	Ken_Sanchez	FNameLName	Admin	Generate Report		
<input type="checkbox"/>	Terri_Duffy	FNLNPrivacy	Analyst	Generate Report		

- i. Go to the **Data Subjects** screen. Click on the **Generate Report** hyperlink in **PDF Status** Column.
- ii. Select the **File Type** from the given option. Click **Generate** button.
 - a. PDF
 - b. Excel
 - c. CSV



- iii. The status in the **PDF Status** column will change from **Generate Report** to **Download Report**.
- iv. Click on the **Download Report** and specify a location where the report need to be saved.
- v. The Final report will provide the scanned results.

Scanned Data Results
Generated By : admin On : 2020-03-07 05:55:49:493
Data Subject : Debra_Burks

Connection Type: SQL Server Connection Name: BikeStoreDSAR Host name (or) IP Address: 34.223.113.94
Database: BikeStore Schema: sales Table: customers
Table Hierarchy: BikeStore.sales>customers

customer_id	first_name	last_name	phone	email	street	city	state	zip_code
1	Debra	Burks	null	debra.burks@bikestore.com	4773 Thorne Ave	Oakland Park	NY	14177
1445	Debra	Burks	null	debra.burks@bikestore.com	4773 Thorne Ave	Oakland Park	NY	14177
1780	Debra	Burks	null	debra.burks@bikestore.com	4773 Thorne Ave	Oakland Park	NY	14177

Connection Type: SQL Server Connection Name: BikeStoreDSAR Host name (or) IP Address: 34.223.113.94
Database: BikeStore Schema: sales Table: orders
Table Hierarchy: BikeStore.sales>customers -> BikeStore.sales>orders

order_id	customer_id	order_status	order_date	required_date	shipped_date	store_id	staff_id
500	1	1	2016-12-09	2016-12-10	2016-12-12	2	6
1555	1	2	2018-04-18	2018-04-18	null	2	7
1610	1	3	2018-11-18	2018-11-18	null	2	6

Connection Type: SQL Server Connection Name: BikeStoreDSAR Host name (or) IP Address: 34.223.113.94
Database: BikeStore Schema: sales Table: order_items
Table Hierarchy: BikeStore.sales>customers -> BikeStore.sales>orders -> BikeStore.sales>order_items

order_id	item_id	product_id	quantity	unit_price	discount
500	1	19	2	7000.00	0.00
500	2	22	2	245.00	0.00
500	3	23	1	700.00	0.00
500	4	20	2	1349.00	0.00
1555	1	24	2	145.00	0.00
1555	2	156	1	6499.00	0.00
1555	3	130	1	800.00	0.00
1555	4	120	1	1000.00	0.00
1555	5	174	1	1199.00	0.00
1610	1	153	1	8999.00	0.00
1610	2	283	2	118.00	0.00

2. RoA

To view report in the **Data Subject** screen, click **Privacy > DSAR Workflow > RoA**.

PRIVACY MANAGEMENT : DSAR WORKFLOW > ROA

				DISAPPROVE	APPROVE	REJECT	FREEZE	RESUBMIT
	DATA SUBJECT NAME	OTHER INFO	STATUS	DATA GROUP	ROA DATE	DATA		ACTIONS
<input type="checkbox"/>	Danette_Willatt	Customer	No Data	Email	2020-02-21 02:20:56	No Data		
<input type="checkbox"/>	Debra_Burks	Customer	Review Pending	BikeStoreEmail	2020-02-21 23:34:31	Generate Report		
<input type="checkbox"/>	Sandeep_Singh	Telemarket	Review Pending	Email	2020-02-20 14:25:28	Generate Report		
<input type="checkbox"/>	Roberto_Tamburello	Finance	Erased	Email	2020-01-20 22:36:10	Erased		
<input type="checkbox"/>	Ken_Sanchez	Admin	Review Pending	FNameLName	2020-02-21 02:20:56	Generate Report		
<input type="checkbox"/>	Terri_Duffy	Analyst	Review Pending	FNLNPrivacy	2020-02-05 23:58:15	Generate Report		

You can also generate a report for each of the Data Subject. Perform the following steps to generate the report for Data Subject.

- i. Click on the **Generate Report** hyperlink in **Data** column.
- ii. Select the **File Type** from the given option. Click **Generate** button.
 1. PDF
 2. Excel
 3. CSV

PRIVACY MANAGEMENT : DSAR WORKFLOW > **SELECT FILE TYPE**

PDF EXCEL CSV

GENERATE

	DATA SUBJECT NAME	OTHER INFO	STATUS	DATA GROUP	ROA DATE	DATA	ACTIONS
<input type="checkbox"/>	Danette_Willatt				2020-02-21 02:20:56	No Data	
<input checked="" type="checkbox"/>	Debra_Burks	Customer	Review Pending	BikeStoreEmail	2020-02-21 23:34:31	Generate Report	
<input type="checkbox"/>	Sandeep_Singh	Telemarket	Review Pending	Email	2020-02-20 14:25:28	Generate Report	
<input type="checkbox"/>	Roberto_Tamburello	Finance	Erased	Email	2020-01-20 22:36:10	Erased	
<input type="checkbox"/>	Ken_Sanchez	Admin	Review Pending	FNameLName	2020-02-21 02:20:56	Generate Report	
<input type="checkbox"/>	Terri_Duffy	Analyst	Review Pending	FNLNPrivacy	2020-02-05 23:58:15	Generate Report	

- iii. The status of the **Data Subject Name** in the **Data** column will change from **Generate Report** to **Download Report**.
- iv. Click on **Download Report** and specify the location where the report needs to be saved.
- v. The Final report will provide the scanned results.

Scanned Data Results
Generated By : admin On : 2020-03-07 05:55:49.493
Data Subject : Debra_Burks

Connection Type: SQL Server Connection Name: BkafStoreDSAR Host name (or) IP Address: 34.223.113.94
Database: BkafStore Schema: sales Table: customers
Table Hierarchy: BkafStore.sales.customers

customer_id	first_name	last_name	phone	email	street	city	state	zip_code
1	Debra	Burks	mail	debra.burks@yahoo.com	9773 Thorne Ave	Orland Park	IL	14177
1446	Debra	Burks	mail	debra.burks@yahoo.com	9773 Thorne Ave	Orland Park	IL	14177
7891	Debra	Burks	mail	debra.burks@yahoo.com	9773 Thorne Ave	Orland Park	IL	14177

Connection Type: SQL Server Connection Name: BkafStoreDSAR Host name (or) IP Address: 34.223.113.94
Database: BkafStore Schema: sales Table: orders
Table Hierarchy: BkafStore.sales.customers -> BkafStore.sales.orders

order_id	customer_id	order_status	order_date	required_date	shipped_date	store_id	staff_id
1009	1	1	2019-12-09	2019-12-10	2019-12-11	1	1
1553	1	1	2019-08-18	2019-08-18	mail	1	1
1613	1	1	2019-11-18	2019-11-18	mail	1	1

Connection Type: SQL Server Connection Name: BkafStoreDSAR Host name (or) IP Address: 34.223.113.94
Database: BkafStore Schema: sales Table: order_items
Table Hierarchy: BkafStore.sales.customers -> BkafStore.sales.orders -> BkafStore.sales.order_items

order_id	item_id	product_id	quantity	list_price	discount
1009	1	2	2	7999.99	0.05
1009	2	22	2	7499.99	0.20
1009	3	23	1	7999.99	0.07
1009	6	10	2	1549.00	0.10
1553	1	24	2	549.99	0.10
1553	2	106	1	6499.99	0.10
1553	3	105	1	409.99	0.10
1553	4	109	1	1999.99	0.05
1553	6	114	1	1199.99	0.20
1613	1	153	1	4999.99	0.07
1613	2	283	2	119.99	0.05

16.4.10.2 Accessing the Reports via API

The reports can also be generated and downloaded via the APIs.

1. Generate a report

This API will submit a request to generate report. A report can be generated in three file formats i.e. PDF, Excel and CSV.

URL:

<http://host:port/dgcontroller/services/v1/products/Privacy/report/generae>

Sample Call:

<http://localhost:8088/dgcontroller/services/v1/products/Privacy/report/generae>

Header Parameters:

sessionId

Request Body:

```
{
    "requestId":5,
    "format":"PDF"
```

```
}
```

The “requestId” is the ID generated using the first API: [http://host:port /dgcontroller/services/v1/products/Privacy/request/](http://host:port/dgcontroller/services/v1/products/Privacy/request/).

Response Body:

It will display a success message on the screen.

2. Download the Report

This API will post a request to download a report which is generated using the above API. This API will download the report in any of specified file format i.e. PDF, Excel and CSV.

URL:

<http://host:port/dgcontroller/services/v1/products/Privacy/report/download>

Sample Call:

<http://localhost:8088/dgcontroller/services/v1/products/Privacy/report/download>

Header Parameters

sessionId

Request Body:

```
{  
    "requestId":5,  
}
```

The “requestId” is the ID generated using API:

<http://host:port /dgcontroller/services/v1/products/Privacy/request/>.

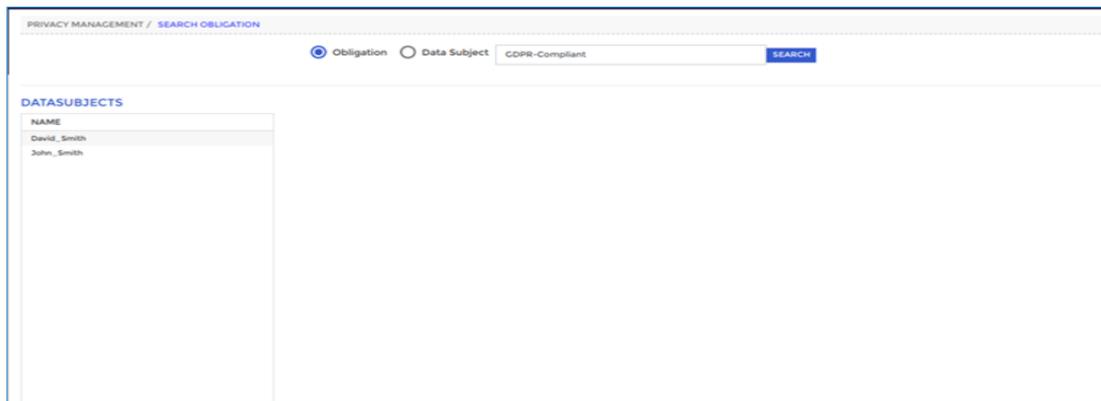
Response Body:

In case of success, file will be downloaded on the host machine.

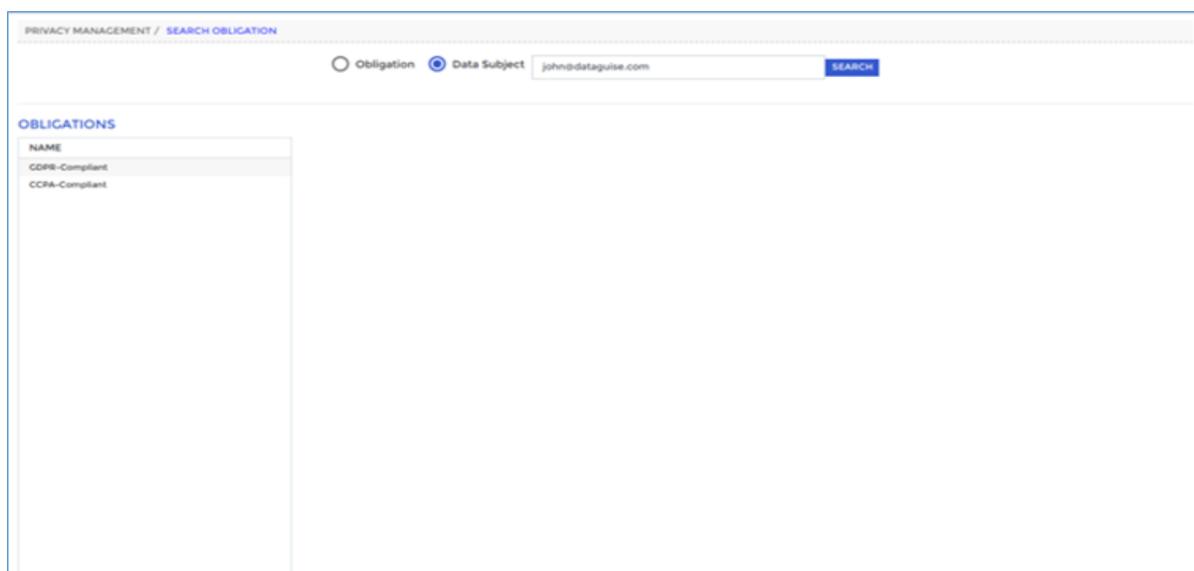
16.4.11 Search Obligation

This is the main screen where the end user of DgSecure can select obligations and get lists of data subjects falling within its scope, can select a data subject and get all obligations affecting that data subject.

On entering the obligation name in the search box, we can get the list of datasubjects associated with that obligation.



For getting the list of obligations associated with the datasubject, we have to provide the strong identifier value of the datasubject in the searchbox.



16.5 Configure Privacy

Perform the following steps to configure Privacy in DgSecure:

***Note** – Steps 1 to 7 are one-time configuration steps, and the last 2 steps, Execute Task and View Report, are the ones that need to be performed repeatedly.

1. [Add IDPs](#)
2. [Add/Edit Identifier](#)
3. [Add Data Group](#)
4. [Add Data Subject](#)
5. [Add System](#)
6. [Create Connection - Privacy](#)
7. [Create Task](#)
8. [Execute Task \(As per the Schedule\)](#)

9. [View Report](#)

16.5.1 Add IDPs

To add an IDP, login into the **DgAdmin**. To Access the IDP screen, click **IDP Management > IDPs > Add IDP**. The **IDP Management** panel will appear.

ID	Name	Hostname / IP	Port	IDP Type	Status	SSL Type	Start Time	Decommission	View Log
1	DBMS Detection	192.168.1.65	8889	DBMS Detection IDP	Active	None	02/18/2020 05:00:07	<input type="checkbox"/>	
2	DBMS Masking	localhost	8888	DBMS Masking IDP	Active	None	02/18/2020 04:59:55	<input type="checkbox"/>	
22	Hadoop Detection	localhost	8111	Hadoop Data IDP	Active	None	02/18/2020 04:59:54	<input type="checkbox"/>	
24	Azure Data IDP	52.167.126.253	8111	Azure Data IDP	Inactive	None		<input type="checkbox"/>	
25	Azure Cloud IDP	13.68.17.227	8081	Azure Cloud IDP	Inactive	None		<input type="checkbox"/>	
26	Hive	localhost	9980	Hive IDP	Active	None	02/18/2020 04:59:54	<input type="checkbox"/>	
27	d2	localhost	8889	DBMS Detection IDP	Active	None	02/18/2020 04:59:57	<input type="checkbox"/>	
28	Files	192.168.5.65	8222	File IDP	Inactive	None		<input type="checkbox"/>	

Follow the below steps to add an IDP:

3. Click the **Add IDP** button. The **Add / Edit IDP** dialog box will appear.

Add/Edit IDP

Name: IDP Type:

Hostname / IP: Port:

1. Enter the descriptive name for the IDP in the **Name** textbox.
2. Enter the **hostname or IP** address of the IDP’s host machine in the **Hostname/IP** textbox.
3. Select the **Privacy IDP** as the IDP Type from the drop-down.
4. Enter the **port number** which the IDP will use to send and receive information.

4. Click **Save**.

5.

16.5.2 Add/Edit Identifier

To access the Identifiers screen, log into the **DgSecure** Application.

Click **Privacy > Configuration > Identifiers > Add Identifier** button. The **Add/Edit Identifier** panel will appear.

PRIVACY MANAGEMENT | CONFIGURATION > ADD/EDIT IDENTIFIERS

IDENTIFIER NAME: BikeStoreDSAREmail

IDENTIFIER DESCRIPTION: Email

IS MANDATORY STRONG IDENTIFIER

SELECT SENSITIVE TYPE: [Dropdown]

SELECT DATATYPE: Select Type [Dropdown]

SELECTED OPERATOR: [Dropdown] **ADD**

SELECTED IDENTIFIER		
DATATYPE	OPERATOR	ACTIONS
STRING	EQUALS	[Edit] [Delete]

CANCEL **UPDATE**

6.

7.

To add a new identifier, follow the below steps.

8. Enter the name of the identifier in the **Identifier Name** text box.
9. Enter the description of the identifier in the **Identifier Description** text box.
10. If the identifier is mandatory, check the **Is Mandatory** checkbox. At the time of creating a Data Subject, value for these Identifiers must be provided.
11. If the identifier is a strong identifier, check the **Strong Identifier** checkbox. **Strong identifier** is the identifier that value must be present at the time of scanning.

When the **Strong Identifier** checkbox is selected, the **Select Sensitive Type**, **Select DataType**, **Selected Operator** fields and **Selected Identifier** pane will be visible.

SELECT SENSITIVE TYPE: Email Address [Dropdown]

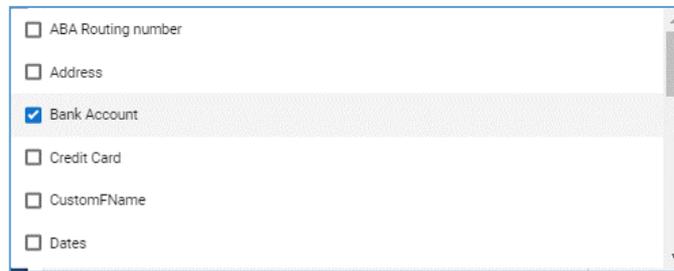
SELECT DATATYPE: string [Dropdown]

SELECTED OPERATOR: contains [Dropdown] **ADD**

SELECTED IDENTIFIER		
DATATYPE	OPERATOR	ACTIONS
STRING	EQUALS	[Edit] [Delete]

CANCEL **SAVE**

12. The **Sensitive Type** field will appear only if you have checked the **Strong Identifier** checkbox. Select the sensitive type in the **Select Sensitive Type** drop-down.



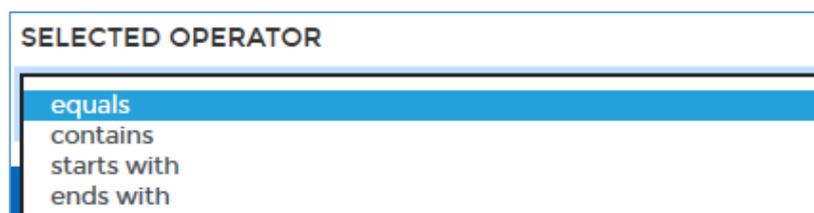
13. The **Data Type** field will appear only if you have checked the **Strong Identifier** checkbox. Select the data type in the **Select Datatype** drop-down.

The available data types are: **String**, **Numeric** and **Date**.



14. The **Selected Operator** field will appear only if you have checked the **Strong Identifier** checkbox. Select the operator in the **Selected Operator** drop-down.

The list will display the available operators based on the selected **Data Type**.



In the above screenshot, the listed operators are displayed when **Data Type = 'String'** is selected.

15. Click the **Add** button to add the datatype. The added datatypes will be available in the **Selected Identifier** panel.

SELECT DATATYPE		SELECTED OPERATOR
Select Type		
ADD		
SELECTED IDENTIFIER		
DATATYPE	OPERATOR	ACTIONS
STRING	EQUALS	
CANCEL UPDATE		

To edit a Datatype or Operator, click . To delete a Datatype or Operator, click .

16. Click **Save**.

16.5.3 Add Data Group

To access the **Data Groups** screen, click **Privacy > Configuration > Groups > Add Group** button. The **Add/Edit Data Group** screen will appear.

PRIVACY MANAGEMENT : CONFIGURATION > ADD/EDIT IDENTIFIERS

IDENTIFIER NAME: BikeStoreDSAREmail

IDENTIFIER DESCRIPTION: Email

IS MANDATORY STRONG IDENTIFIER

SELECT SENSITIVE TYPE: [Dropdown]

SELECT DATATYPE: [Dropdown]

SELECTED OPERATOR: [Dropdown] **ADD**

SELECTED IDENTIFIER		
DATATYPE	OPERATOR	ACTIONS
STRING	EQUALS	
CANCEL UPDATE		

To add a new **Data Group**, perform the following steps:

- 17.
18. Enter the name of the Data Group in the **Data Group** text box.
19. Enter description of the Data Group in the **Data Group Description** text box.
20. Enter the expiry days of the Data Group in the **Group Index Expire In** text box. The **Group Index Expire In** is the expiration time of the group in days.
21. To create the rule for the Data Group in the **Create Rule** text box. Click **Rule Builder** button.

() AND OR

Select Identifier :

Select an option

Rule Query :

(Email_DSAR) AND (FName_DSAR)

RESET

CANCEL SAVE

- vi. Select at least one strong identifier in the **Select Identifier** drop-down.
- vii. The query will be displayed in the **Rule Query** text box.
- viii. Click **Save**.

22. Click the **Save** button. This will save the Data Group and it will be available on the **Data Group** screen.

PRIVACY MANAGEMENT : CONFIGURATION > DATA GROUPS

[ADD GROUP](#)

ID	GROUP NAME	GROUP DESCRIPTION	GROUP EXPIRY	RULE	ACTION
4	BikeStoreEmail	Email	100	BikeStoreDSAREmail	
3	FNLNPrivacy	FirstName LName for Privacy	100	(FName_Privacy AND LName_Privacy)	
2	Email	Email	100	Email_DSAR	
1	FNameLName	FName and LName	100	(FName_DSAR AND LName_DSAR)	

To view the details of a Data Group, click . To edit a Data Group, click . To delete a Data Group, click .

16.5.4 Add Data Subject

To access **Data Subjects** screen, click **Privacy > Configuration > Data Subjects > New Data Subject** tab. The **New Data Subject** panel will appear.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS **NEW DATA SUBJECT**

MANUAL
 FILE WITH KNOWN FORMAT
 FILE WITH UNKNOWN FORMAT

DATA SUBJECT NAME *
 OTHER INFO
 DATA GROUP *

STRONG IDENTIFIERS

IDENTIFIER	DATATYPE	OPERATOR	VALUE
Email_DSAR *	STRING	EQUALS	<input type="text" value="dwillatt1@illinois.edu"/>

IDENTIFIER ATTRIBUTES

NAME	DESCRIPTION	VALUE
First_Name *	First Name	<input type="text" value="Danette"/>
Last_Name *	Last Name	<input type="text" value="Willatt"/>
Email *	Personal Email Address	<input type="text" value="dwillatt1@illinois.edu"/>
Country *	Country	<input type="text" value="United States"/>

There are three ways in which you can create new **Data Subject**. These are:

1. Manual
2. File with Known Format
3. File with Unknown Format

To add a new Data Subject, follow the below steps:

23. **Manual:** Select the **Manual** option to provide all the required information manually. Perform the following steps to create a Data subject manually:
- 24.

25. **PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS**

DATA SUBJECTS **NEW DATA SUBJECT**

MANUAL FILE WITH KNOWN FORMAT FILE WITH UNKNOWN FORMAT

DATA SUBJECT NAME * OTHER INFO DATA GROUP *

STRONG IDENTIFIERS

IDENTIFIER	DATATYPE	OPERATOR	VALUE
Email_DSAR *	STRING	EQUALS	<input type="text" value="dwillatt1@illinois.edu"/>

IDENTIFIER ATTRIBUTES

NAME	DESCRIPTION	VALUE
First_Name *	First Name	<input type="text" value="Danette"/>
Last_Name *	Last Name	<input type="text" value="Willatt"/>
Email *	Personal Email Address	<input type="text" value="dwillatt1@illinois.edu"/>
Country *	Country	<input type="text" value="United States"/>

CANCEL SAVE

1. Enter the name of the Data Subject in the **Data Subject Name** text box.
2. Enter extra information of the Data Subject in the **Other Info** text box, if any.
3. Select the Data Group from the **Data Group** drop-down. This will display the list of **Strong Identifier** and **Identifier Attributes** associated with the selected Data Group.
4. Enter the values of the **Strong Identifiers**. This field will appear only if you are creating the Data Subject manually.

STRONG IDENTIFIERS

IDENTIFIER	DATATYPE	OPERATOR	VALUE
FName_DSAR *	STRING	EQUALS	<input type="text" value="Shivani"/>
LName_DSAR *	STRING	EQUALS	<input type="text" value="Gupta"/>

5. Enter the value of the mandatory Identifiers in the **Identifier Attribute** list.

IDENTIFIER ATTRIBUTES		
NAME	DESCRIPTION	VALUE
First_Name *	First Name	<input type="text" value="Shivani"/>
Last_Name *	Last Name	<input type="text" value="Gupta"/>
Email *	Personal Email Address	<input type="text" value="gupta.shivani1402@gmail.com"/>
Country *	Country	<input type="text" value="India"/>

6. Click **Save**.

- 26. **File with Known Format:** Select the **File with Known Format** option to upload the information from the file.
- 27.
- 28. Perform the below steps to create the Data Subject by uploading a file.
- 29.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS NEW DATA SUBJECT

Manual
 File with known format
 File with unknown format

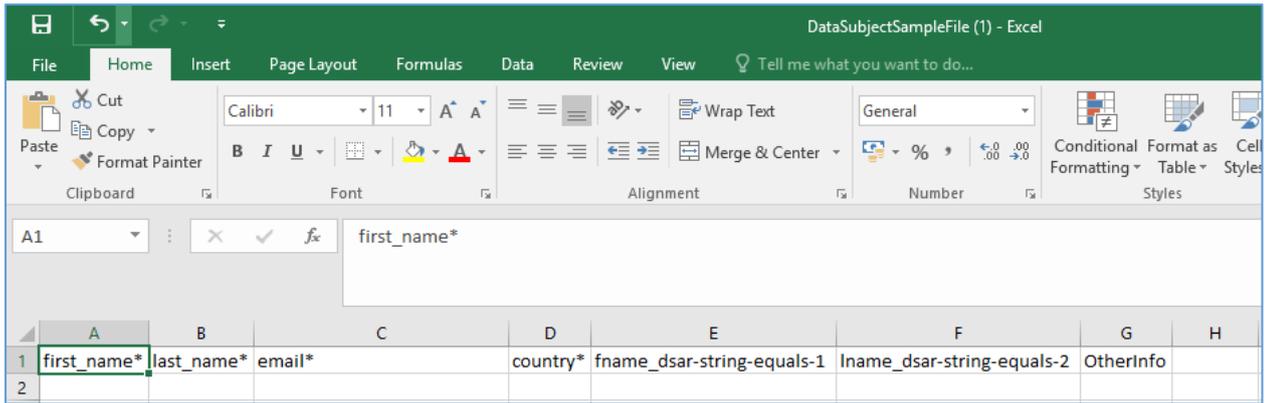
DATA GROUP:

DATA SUBJECT FILE: BROWSE

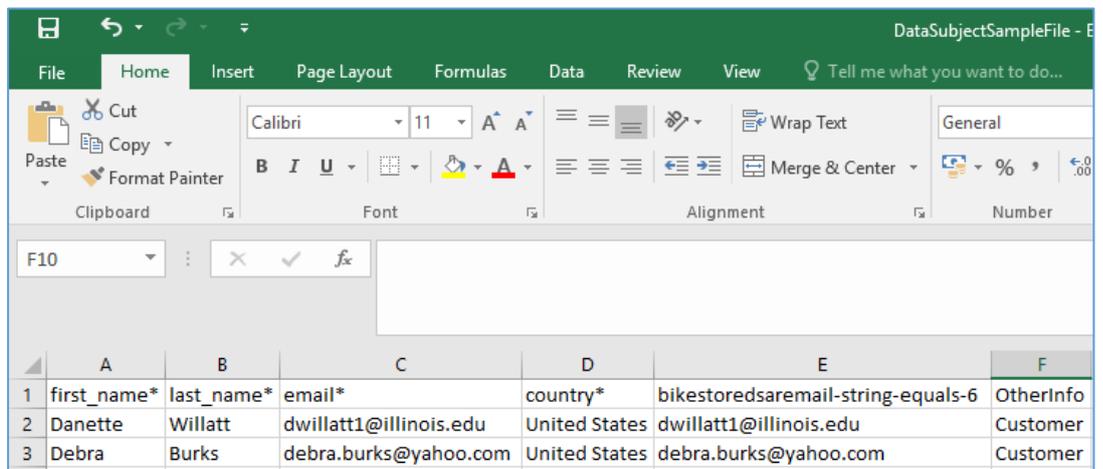
30.
31.

1. Select the Data Group from the **Data Group** drop-down.
2. To download a sample file, click the **Download Sample File** button in the bottom of the screen. This file contains the column headers for the selected Data Group, based on which you will enter the data.

Note – The **Download Sample File** button will be enabled only when you select a Data Group from the **Data Group** drop-down.

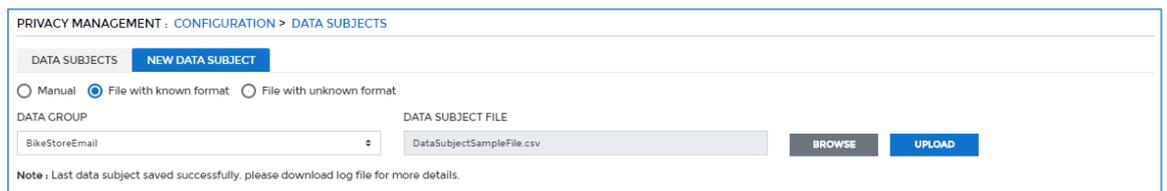


3. Fill the data in the downloaded sample file as per the specified format and save it on your local machine.



4. Click the **Browse** button to upload the saved file from your local machine.

Note – Only **xlsx**, **xls** and **csv** file are allowed.



5. Click the **Upload** button. The data for **Data Subject** will be available on the **Data Subjects** tab.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS		NEW DATA SUBJECT				DELETE
<input type="checkbox"/>	FIRST NAME	GROUP NAME	OTHER INFO	PDF STATUS	ACTIONS	
<input type="checkbox"/>	Danette_Willatt	Email	Customer	No Data		
<input type="checkbox"/>	Debra_Burks	BikeStoreEmail	Customer	Generate Report		
<input type="checkbox"/>	Sandeep_Singh	Email	Telemarket	Generate Report		
<input type="checkbox"/>	Roberto_Tamburello	Email	Finance	Erased		
<input type="checkbox"/>	Kan_Sanchez	FName,Name	Admin	Generate Report		
<input type="checkbox"/>	Terr_Duffy	FNLNPrivacy	Analyst	No Data		

32. **File with Unknown Format:** Select the **File with Unknown Format** option to upload information from the unknown file format.

33.

34. Perform the below steps to create Data Subject by uploading a file format.

35.

PRIVACY MANAGEMENT : CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS **NEW DATA SUBJECT**

Manual
 File with known format
 File with unknown format

DATA GROUP: FName,Name DATA SUBJECT FILE: DataSubjectSampleFile.csv **BROWSE** **UPLOAD**

36.

1. Select the Data Group from the **Data Group** drop-down.
2. Create a csv file on your local machine which will contain the details of **Data Subject**.

Note – Only csv file is allowed.

	A	B	C	D	E	F
1	first_name*	last_name*	email*	country*	bikestoresaremail-string-equals-6	OtherInfo
2	Danette	Willatt	dwillatt1@illinois.edu	United States	dwillatt1@illinois.edu	Customer
3	Debra	Burks	debra.burks@yahoo.com	United States	debra.burks@yahoo.com	Customer

3. Click the **Browse** button to search the file from your local machine.

- Click the **Upload** button to the csv file. The uploaded data will appear in the **Uploaded Data Subject File Columns** panel.

FIRST_NAME*	LAST_NAME*	EMAIL*	COUNTRY*	FNAME_DSAR-STRING-EQUALS-1	LNAR*
shivani	gupta	gupta@dataguise.com	India	shivani	gupt

- Once the file is uploaded then tag the file columns headers with the **Strong** and **Mandatory Identifiers**.

Note - Identifiers in '**bold**' are Strong Identifiers. Identifiers with '*' are Mandatory Identifiers. Identifiers in '**bold**' with '*' are both Strong and Mandatory Identifiers.

To tag file columns with the **Strong** and **Mandatory Identifiers**.

- Click on the drop-down above of the column headers.

FIRST_NAME*	LAST_NAME*	EMAIL*	COUNTRY*	BIKESTORESAREMAIL-STRING-EQUAL-6
Shivani	Gupta	shivani.gupta@gamil.com	India	shivani.gupta@bike.com
Tania	Thakur	tania.thakur@gmail.com	India	tania.thakur@bike.com
Gunjan	Bhatnagar	gunjan.bhatnagar@gmail.com	India	gunjan.bhatnagar@bike.com

- Select the **Strong** and **Mandatory Identifier** from the given list.

UPLOADED DATA SUBJECT FILE COLUMNS

FIRST_NAME*	LAST_NAME*		COUNTRY*	BIKESTORESAREMAIL-STRING-EQUAL-6
Shivani	Gupta	No Selection	India	shivani.gupta@bike.com
Tania	Thakur	First_Name *	India	tania.thakur@bike.com
Gunjan	Bhatnagar	Middle_Name	India	gunjan.bhatnagar@bike.com
		Last_Name *		
		NationalID		
		Driver_License_Number		
		Passport_Number		

6. Click the **Save** button to save the changes. The saved Data Subjects will be displayed on the **Data Subjects** tab.

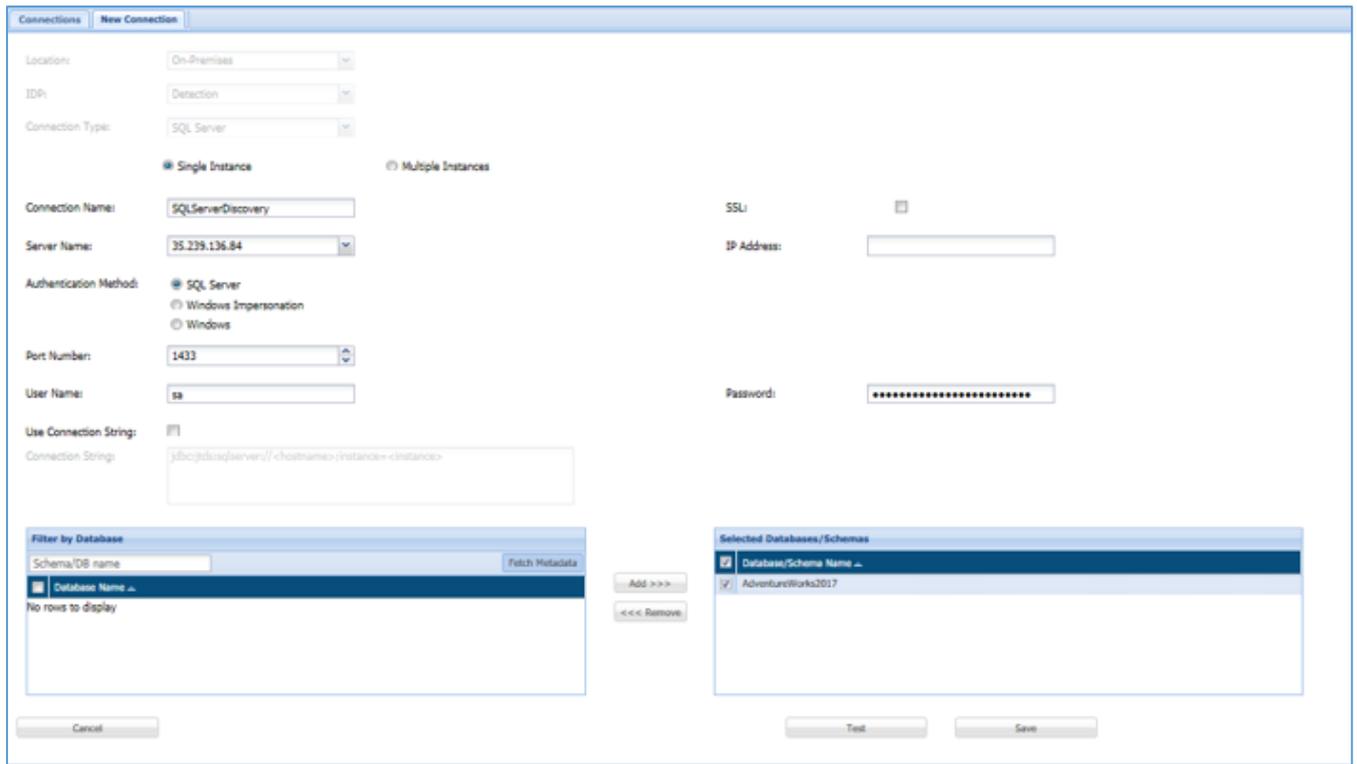
PRIVACY MANAGEMENT > CONFIGURATION > DATA SUBJECTS

DATA SUBJECTS	NEW DATA SUBJECT					DELETE
	FIRST NAME	GROUP NAME	OTHER INFO	PDF STATUS	ACTIONS	
<input type="checkbox"/>	Shivani	BikeStoreEmail	Customer	No Data		
<input type="checkbox"/>	Gunjan	BikeStoreEmail	Customer	No Data		
<input type="checkbox"/>	Tania	BikeStoreEmail	Customer	No Data		
<input type="checkbox"/>	Danette_Willatt	Email	Customer	No Data		
<input type="checkbox"/>	Debra_Burks	BikeStoreEmail	Customer	Generate Report		
<input type="checkbox"/>	Sandeep_Singh	Email	Telemarket	Generate Report		
<input type="checkbox"/>	Roberto_Tamburello	Email	Finance	Erased		
<input type="checkbox"/>	Ken_Sanchez	FNameLName	Admin	Generate Report		
<input type="checkbox"/>	Terri_Duffy	FNLNPrivacy	Analyst	No Data		

16.5.5 Create Connection – Privacy

For each **Data Group** created in the above sections, create connection with the database where you want to scan the sensitive data.

To create a connection, access the **Connection Manager**. Click **RDBMS > Connection Manager > Connections > New Connection** tab.



Follow the below steps to create a connection.

37. Select either **On-Premises** or **Cloud** option from the **Location** drop-down.
38. Select **Detection** from the **IDP** drop-down. A connection designated for the Detection IDP only applies to Detection tasks.
39. Select the connection type from the **Connection Type** drop-down. It lists down all the supported databases types.

Note – The **Location**, **IDP** and **Connection Type** cannot be edited, once the connection has been created.

40. Enter unique **Connection Name**. This fields accepts letters, numbers, and symbols.
41. Check the **SSL** checkbox if you want to make your connection secure.
42. Select the **Connection Type** either **Basic** or **TNS**.

1. **TNS Name:** Enter unique TNS name.

Note: The **TNS Name** field will appear only when you select the Connection Type as **TNS**.

- 43.

- 44. Enter the **Host Name** or select the **IP address** from the **Host Name** drop-down. You can also search for the hostnames and IP addresses of databases by **Find DBMS** tasks.
- 45. Enter the **IP Address** for the connection.
- 46. Enter the **Port Number** for establishing a connection. You can also edit the port number later, if required.
- 47. Enter the **Service ID (SID)** or **Service Name**.
- 48. Enter the database **user name**.
- 49. Enter the database **password**.
- 50. Check the **Use Connection String** checkbox, if you want to enter the Connection string.
- 51. The **Connection String** text box will appear only when you have checked the **Use Connection String** checkbox.

The **Connection String** specify the information about the data source and the means of connecting to it.

Use Connection String:

Connection String:

- 52. Select the **Authentication Method** according to the Connection Type.
- 53.
- 54. **Note** – The **SQL Server**, **Windows Impersonation** and **Windows** authentication method are available when **Connection Type** SQL server is selected.
- 55.
- 56. Following are the available Authentication Methods:

1. **SQL Server**
2. **Windows Impersonation**
3. **Windows**

- 57. Click **Fetch Metadata**. A list of all databases on the server will appear under the **Filter by Database** pane.

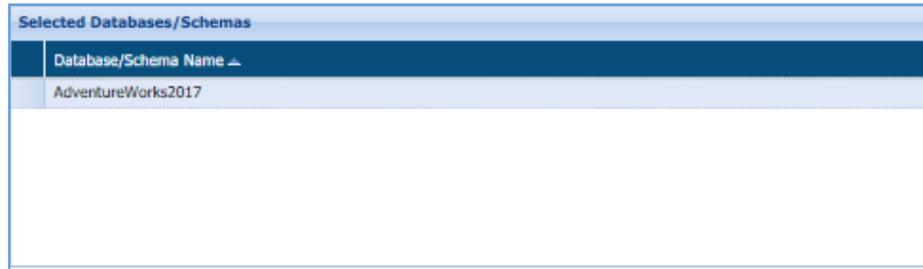
Filter by Database

Schema/DB name

<input type="checkbox"/> Database Name
<input type="checkbox"/> AKPrivacy
<input checked="" type="checkbox"/> AdventureWorks2017BAK
<input type="checkbox"/> DGMask

58. To select a database, check the checkbox next to the listed database. Click the **Add** button.

59. The selected databases will be added to the **Selected Databases/Schemas** list.



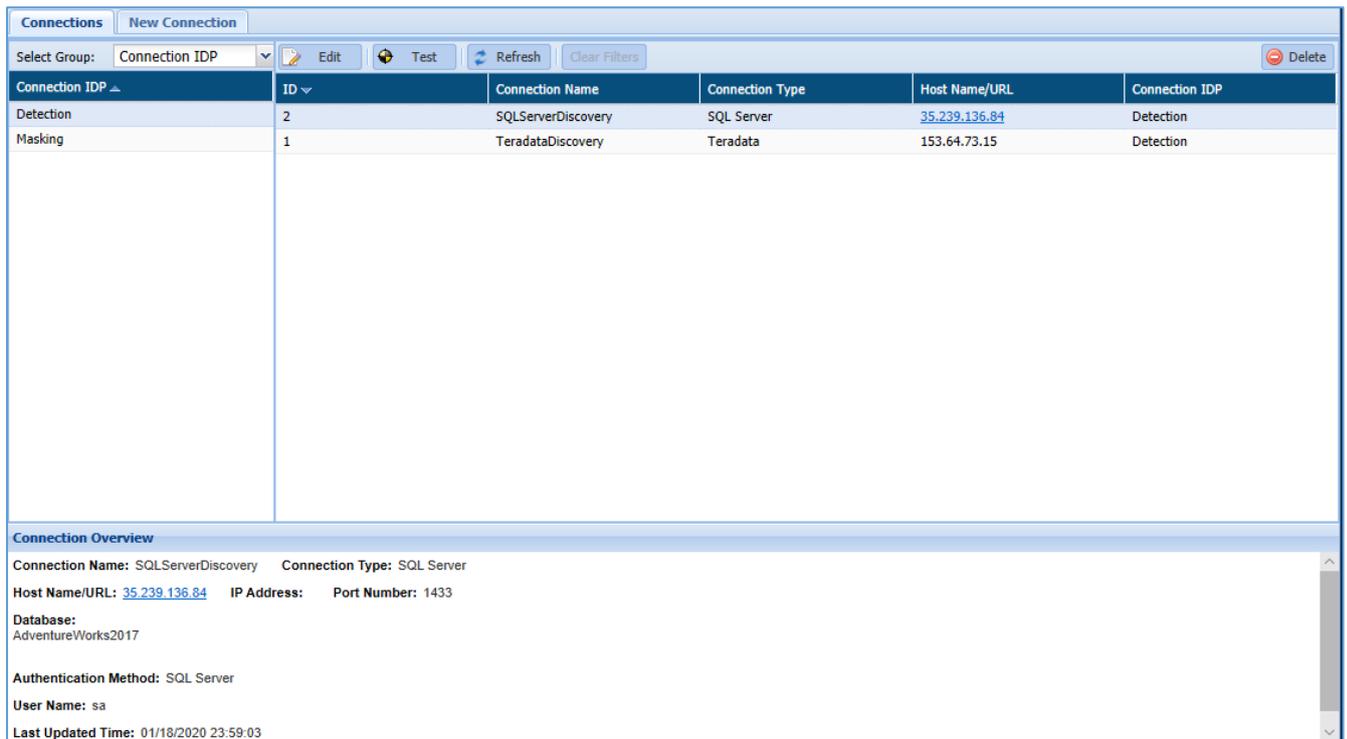
60. To remove a database from the **Selected Database/Schema** list, click **Remove** button.

61. Click the **Save** button, if you want to save the connection.

62. Click the **Cancel** button, if you do not want to save the connection.

63. Click the **Test** button, if you want to test the connection before using it to ensure that the connection has been established.

64. Once the **Connection** is ready. It will be available on **Connection Manager** screen.



16.5.6 Add System

To access **Systems** screen, click **Privacy > Configuration > Systems > New System** tab. The **New System** screen will appear.

There are two ways in which you can add a new **System**.

1. Manual
2. File Upload

65. **Manual**: Select the **Manual** option to provide all the required information of the System manually. Perform the following step to add a system manually:

66. Enter the details for creating a System such as System Name, System Type, Classification, Owner of the system, Controller, Processing Purpose, Country, Retention Period and Host Name.

- i. System Name:** Enter the name of the system. A System is a host on which there can be single or multiple database servers.
- ii. System Type:** Enter the category of the system.

- iii. **Classification:** Enter the classification of the system.
- iv. **System Owner:** Enter the owner of the system.
- v. **Controller:** Enter the controller name of the system.
- vi. **Processing Purpose:** Enter the **Processing Purpose** for which system is created. E.g., HR data, Employee data, etc.
- vii. **Country:** Select the country name from the drop-down. This field specifies the name of the country in which system is located.
- viii. **Retention Period:** Enter the numeric value. This field defines the period, when retention needs to be done on the system.
- ix. **Hostname/TNS Name:** Enter the **Hostname** or **TNS Name**. This field should contain unique hostname.

67. Click **Save**.

68.

69. The System will be created and the basic details of the System will be displayed in the **System** screen.

PRIVACY MANAGEMENT : CONFIGURATION > SYSTEMS

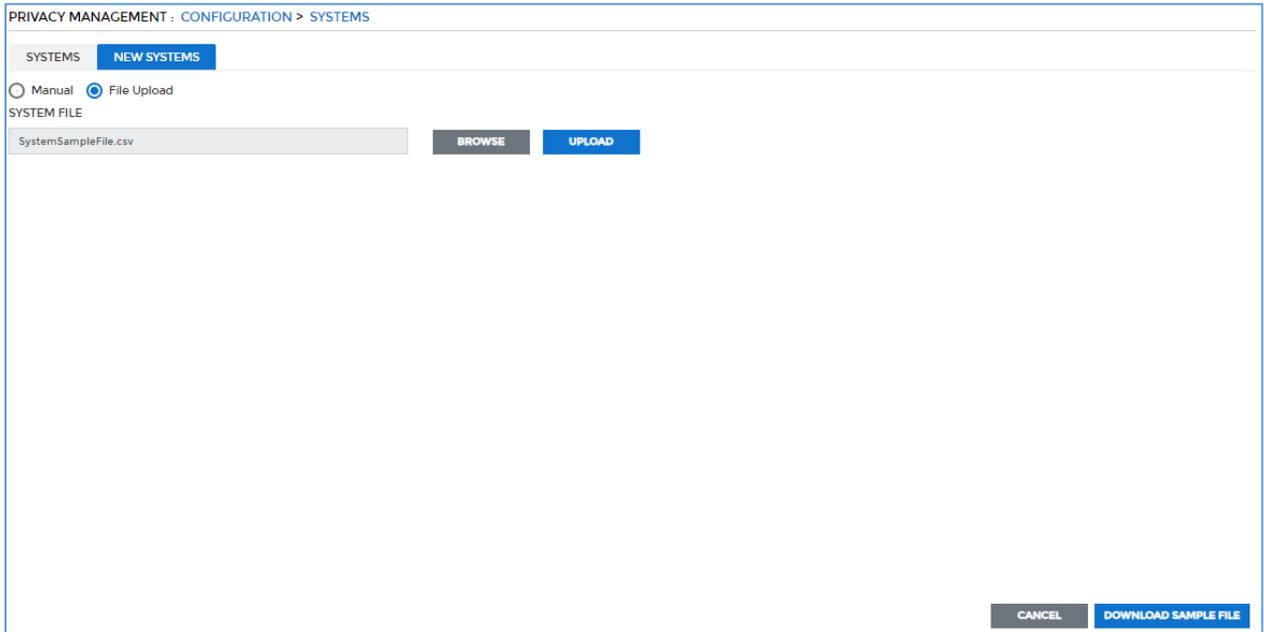
SYSTEMS		NEW SYSTEMS			
SYSTEM NAME	SYSTEM TYPE	HOST NAME	CLASSIFICATION	SYSTEM OWNER	ACTIONS
Y-Solowarm	Marketing	Prevacid	Public	Marketing	  
Alpha	HR	DELFLEX	Confidential	Services	  
Cardguard	Engineering	DYNA-HEX 4	Confidential	Services	  
Voyatouch	HR	FORTAZ	Confidential	Business Development	  
Namfix	Sales	ZOFTRAN	Secret	Training	  
Voltsillam	Sales	Thyroid Assist	Public	Services	  
Viva	Marketing	Lisinopril and Hydrochlorothiazide	Secret	Training	  
Sonsing	Sales	Lachesis Murtus Kit Refill	Secret	Human Resources	  
Zamit	HR	Cataracts	Confidential	Engineering	  
HRPrivate	windows	35.239.136.84	Private	Preeti	  
BikeStoreDB	SQL Server	34.223.113.94	Confidential	Services	  

Items per page: 20 1 - 11 of 11   

To view the System details, click . To edit the System details, click . To delete the System, click .

70. **File Upload:** Select **File Upload** option to create a System by uploading information through a file.

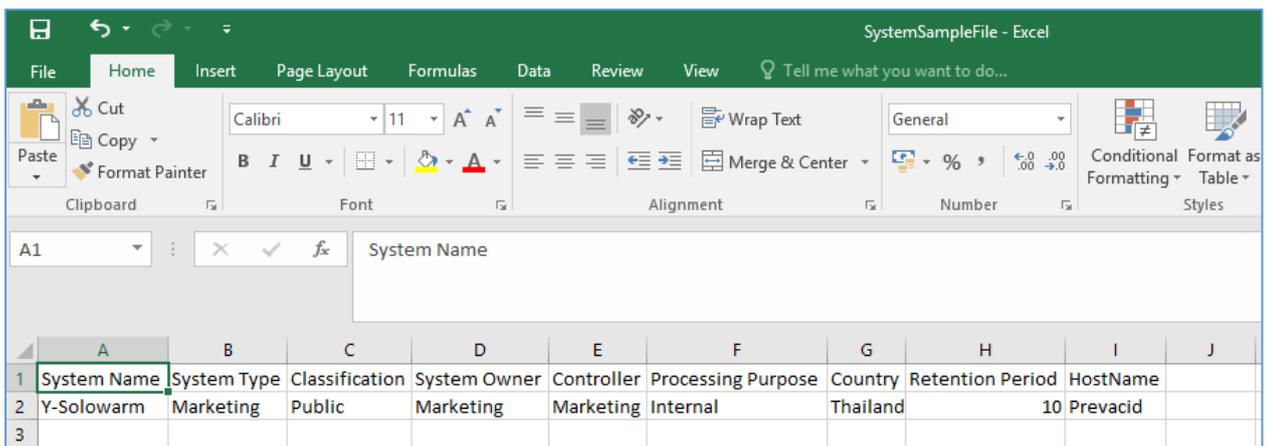
71.



72.

Perform the following steps to add a System:

1. To download the system sample file, click the **Download Sample File** button in the bottom of the screen.
2. Fill in the data for the System in the downloaded sample file as per the specified format and save it on your local machine.



3. Click the **Browse** button to search the location of the saved file on your local machine.

PRIVACY MANAGEMENT : CONFIGURATION > SYSTEMS

SYSTEMS **NEW SYSTEMS**

Manual File Upload

SYSTEM FILE

SystemSampleFile.csv **BROWSE** **UPLOAD**

- Click the **Upload** button to upload the csv file. This file contains the information of the System. Once the file is uploaded, the system details will be added on the **Systems** screen.

PRIVACY MANAGEMENT : CONFIGURATION > SYSTEMS

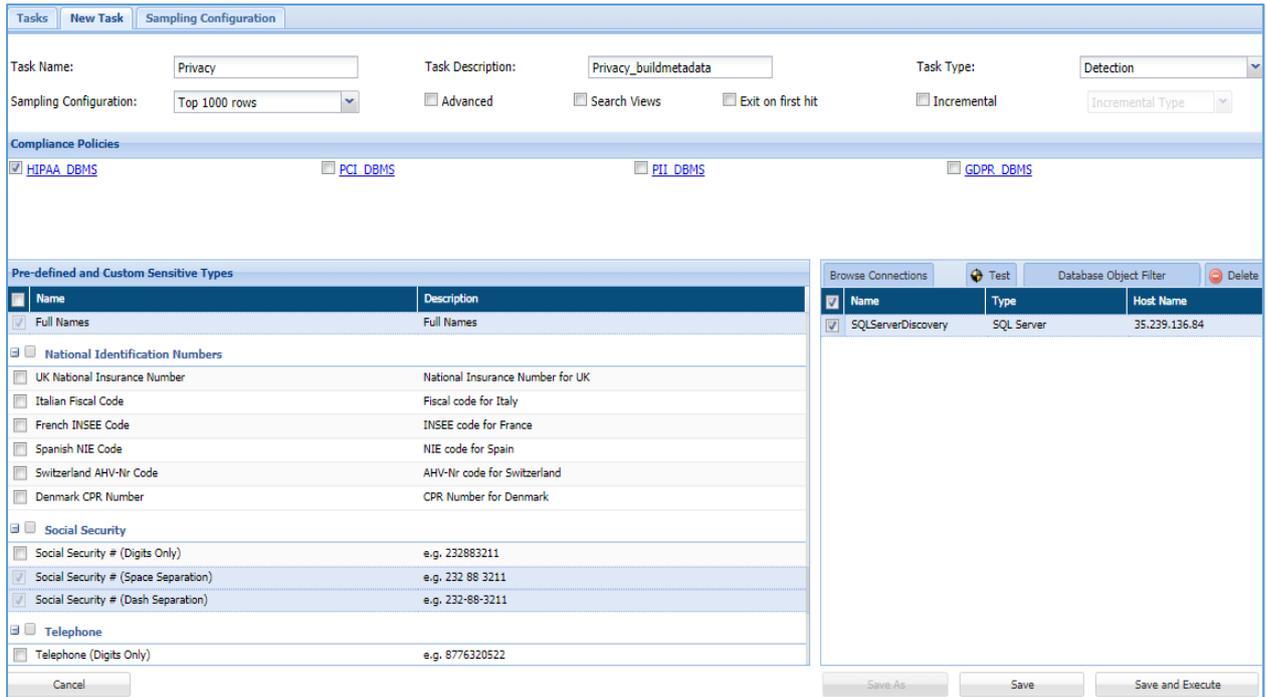
SYSTEM NAME	SYSTEM TYPE	HOST NAME	CLASSIFICATION	SYSTEM OWNER	ACTIONS
Y-Solowarm	Marketing	Prevacid	Public	Marketing	
Alpha	HR	DELFLEX	Confidential	Services	
Cardguard	Engineering	DYNA-HEX 4	Confidential	Services	
Voyatouch	HR	FORTAZ	Confidential	Business Development	
Namfix	Sales	ZOFRAN	Secret	Training	
Voltsillam	Sales	Thyroid Assist	Public	Services	
Viva	Marketing	Lisinopril and Hydrochlorothiazide	Secret	Training	
Sonsing	Sales	Lachesis Mutus Kit Refill	Secret	Human Resources	
Zamit	HR	Cataracts	Confidential	Engineering	
HRPrivate	windows	35.239.136.84	Private	Preeti	
BikeStoreDB	SQL Server	34.223.113.94	Confidential	Services	

Items per page: 20 1 - 11 of 11

16.5.7 Create Task

Detection task identifies sensitive data in known databases. If the database IP address and port numbers are already known, a **Detection** task can be run immediately. If this information is not known, it can be found using a **Find DBMS** task.

To access the **New Detection Task** screen, click **RDBMS > Detection > Tasks > New Task** tab. The **New Task** panel will be displayed.



The **New / Edit Detection Task** screen is divided into four panels which are described below.

1. Task List
2. Compliance
3. Pre-defined and Custom Sensitive Type
4. Connections

1. **Task List**

This panel displays the **Task Name** and **Task Description**. Both fields accept letters, numbers, and symbols, and hold up to 256 characters. The name must be unique for each individual task. The task name cannot be edited once a task is created. Altering the Task Name and clicking **Save As** will create a new task.

Select the sampling configuration type from the **Sampling Configuration** drop-down. By default, there are three options.

5. **Top 1000 rows**: Select this option to sample the top 1000 rows for sampling.
6. **Read Top 5% of data**: Select this option to sample the top 5% data for the sampling.
7. **Full**: Select this option to do the sampling of the entire database.

Check the **Advance** checkbox to define the sampling configuration.

Sampling Configuration

Name: * Description: Set Sampling Config as Default:

Show Advance Sampling Details

Table Row Count Range: * To: Type: *

By: * Value: *

Table Row Count Range	Sample Type	Sample Value	Sample By	Actions
Default	Top	1000	Rows	

Note:

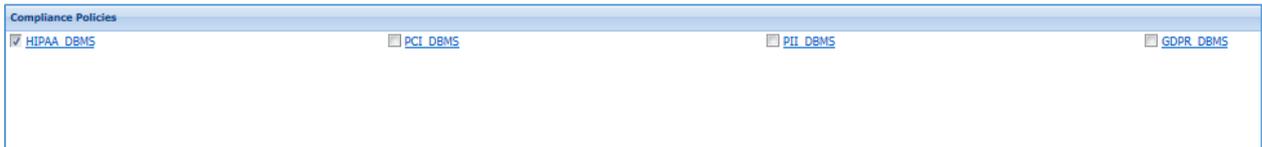
1. Top and Bottom sampling options are not supported for Teradata, Aster DB, Sybase and Sybase IQ connections.
2. Random and Bottom sampling options are not supported for Snowflake, Informix DB and Splice Machine connections.
3. Random sampling option is not supported for MySQL connection.

- i. Enter the **Name** and **Description** of the Sampling Configuration.
- ii. **Set Sampling Config as Default:** Check the checkbox if you want to set the sampling configuration as default.
- iii. **Show Advance Sampling Details:** Check the checkbox to set the advance setting of sampling configuration.
- iv. Enter table row count range start value in **Table Row Count Range** box. This value specifies the starting range for sampling.
- v. Enter table row count range end value in **To** box.
- vi. Select the **Type** of the sampling from the given options. There are four options. These are:
 - a. **Top:** Select the **Top** option in the drop-down. It will process the records from the top of the database.
 - b. **Bottom:** Select the **Bottom** option in the drop-down. It will process the records from the bottom of the database.
 - c. **Random:** Select the **Random** option in the drop-down. It will process the random records in the database irrespective of their position.
 - d. **Complete:** Select the **Complete** option in the drop-down. It will process entire database.

- vii. Select the type by which data is to be scanned in the **By** drop down. The options are:
 - 8. **Rows:** Select **Rows** from the drop-down, if you want to see the number of records sampled.
 - 9. **Percent** Select **Percent** from the drop-down, if you want to see the percent of records sampled.
- viii. Select the value of rows or percent in the **Value**. Click **Save** button to save the configuration.

2. **Compliance Policies**

The **Compliance policies** panel is where all DBMS policies are listed. You can select any number of the policies while creating a **New Task**. When a policy is selected, the sensitive types associated with that policy are automatically selected in the **Pre-defined and Custom Sensitive types** panel.



3. **Pre-defined and Custom Sensitive Type**

When a policy is selected, the sensitive types associated with that policy are automatically selected in the Pre-defined and Custom sensitive types panel. It is also possible to add sensitive types to a task independent of a policy. Clicking the checkbox in the upper left corner of the panel selects all sensitive types. The sensitive types are categorized by sensitive group; each group expands and collapses by clicking on the +/- button next to the group name or by clicking the group name itself.

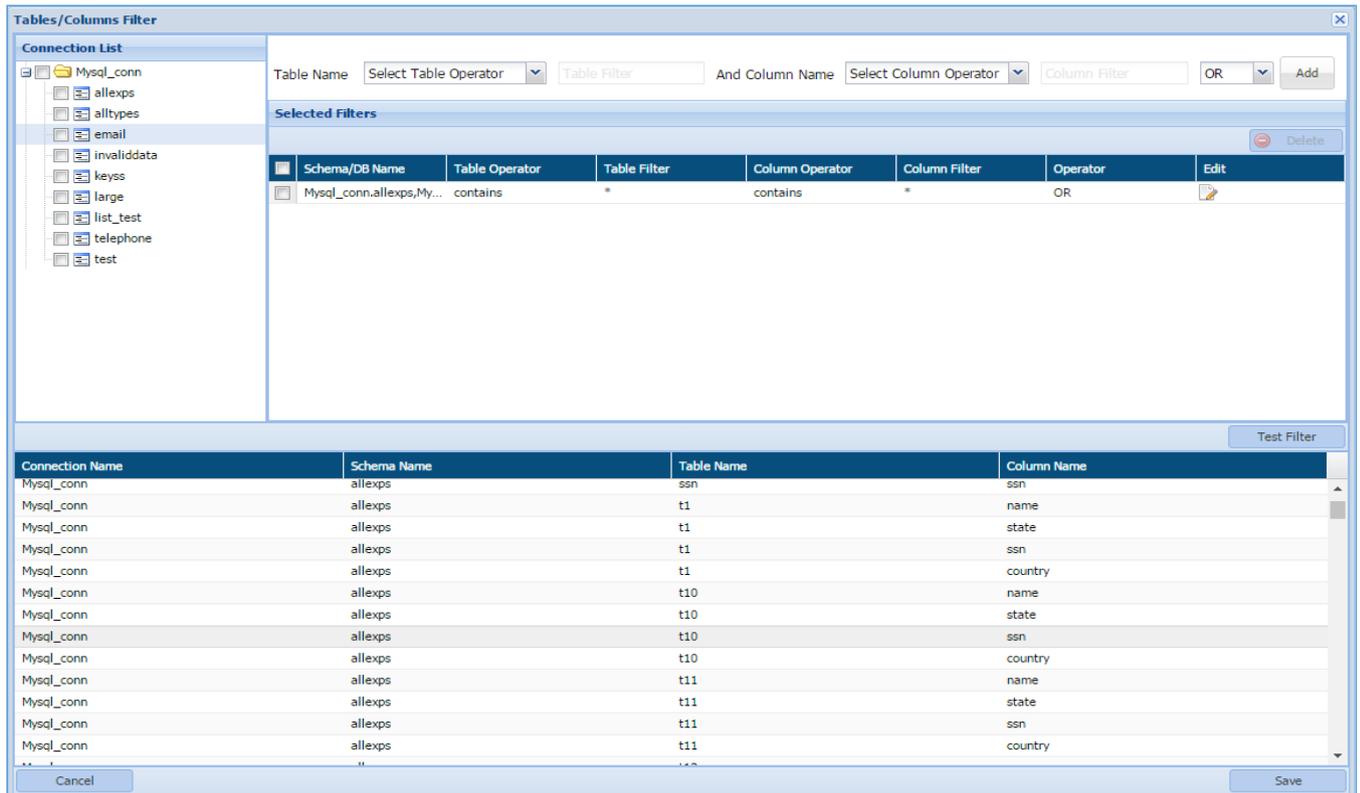
Pre-defined and Custom Sensitive Types	
Name	Description
<input checked="" type="checkbox"/> Full Names	Full Names
<input type="checkbox"/> National Identification Numbers	
<input type="checkbox"/> UK National Insurance Number	National Insurance Number for UK
<input type="checkbox"/> Italian Fiscal Code	Fiscal code for Italy
<input type="checkbox"/> French INSEE Code	INSEE code for France
<input type="checkbox"/> Spanish NIE Code	NIE code for Spain
<input type="checkbox"/> Switzerland AHV-Nr Code	AHV-Nr code for Switzerland
<input type="checkbox"/> Denmark CPR Number	CPR Number for Denmark
<input type="checkbox"/> Social Security	
<input type="checkbox"/> Social Security # (Digits Only)	e.g. 232883211
<input checked="" type="checkbox"/> Social Security # (Space Separation)	e.g. 232 88 3211
<input checked="" type="checkbox"/> Social Security # (Dash Separation)	e.g. 232-88-3211
<input type="checkbox"/> Telephone	
<input type="checkbox"/> Telephone (Digits Only)	e.g. 8776320522
<input checked="" type="checkbox"/> Telephone (Space Separation)	e.g. 877 632 0522
<input checked="" type="checkbox"/> Telephone (Dash Separation)	e.g. 877-632-0522
<input checked="" type="checkbox"/> Telephone (Dot Separation)	e.g. 877.632.0522
<input checked="" type="checkbox"/> Telephone (Standard)	e.g. (877) 632-0522
<input checked="" type="checkbox"/> Telephone (Standard without spaces)	e.g. (877)632-0522

4. **Connections**

The **Select Connections** panel is where the user selects the RDBMS connection(s). There is no limit to the number or type of the connections that can be selected. The connections that appear in this panel are those which have been created on the **Connection Manager** screen.

Browse Connections		Test	Database Object Filter	Delete
Name	Type	Host Name		
<input checked="" type="checkbox"/> SQLServerDiscovery	SQL Server	35.239.136.84		

Click the **Database Object Filter** button to filter certain tables and/or columns out of the search. When filters are defined, only those Databases/Tables/Columns that match the filter are scanned.



Once you will create a task, schedule it.

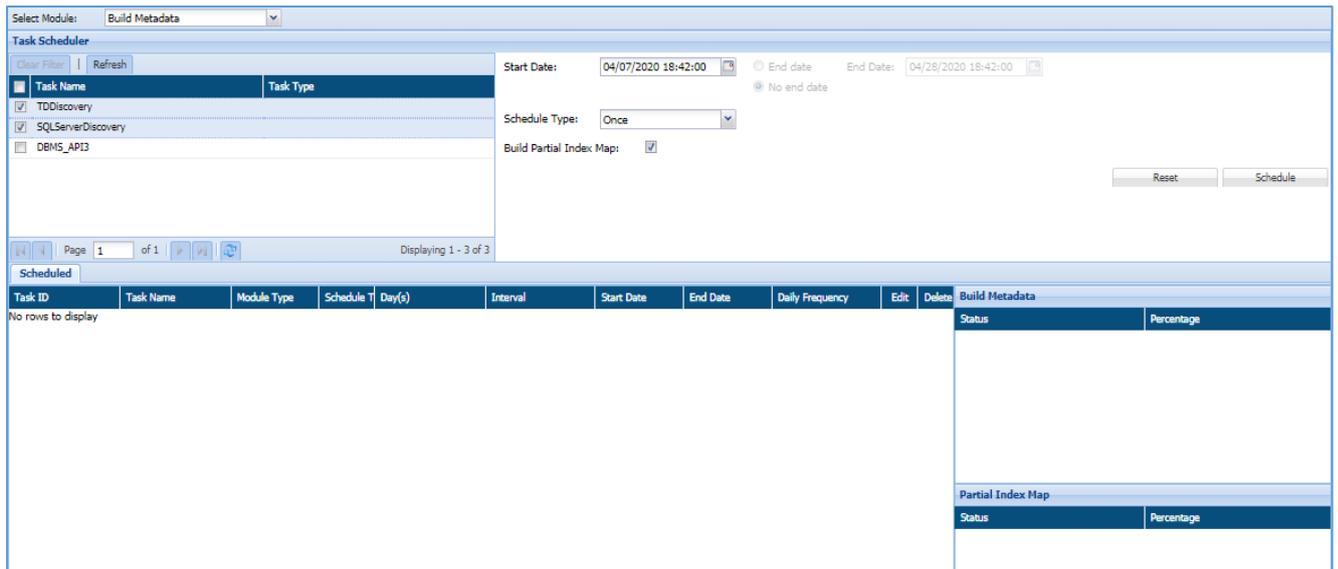
16.5.8 Execute Task (As per the Schedule)

The next step is to schedule, and execute the detection task as well as build the metadata.

The Metadata summarizes the basic information about the data. It makes finding and working with particular instances of data easier. The Build Metadata task is used to push the DBMS discovery results from the controller agent to the Privacy agent repository. It also saves the system, database, schema, column and regex information into agent repository.

Once the task has been scheduled, the results will get generated automatically and It will push those results on the Privacy IDPs.

To schedule a task and to create **Metadata**, access the **Scheduler** screen. Click **Scheduler** from the side bar. The Scheduler screen will be displayed.



To schedule a Task, follow the below steps.

73. Go to the **Scheduler** and select the module **Build Metadata** from the **Select Module** drop-down
74. Select the required task.
75. Select the task schedule start date.
76. Select the **end date** for a task. If the task is scheduled forever, select **no end date**.
77. Select the schedule type. The available options are:
 - i. **Weekly**: Select this option to perform the task on the selected days of the week.
 1. Select the days of the week to perform the task.
 2. Select the daily frequency of the task. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 3. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
 - ii. **Hourly**: Select this option to perform the task in every few hours.
 4. Select the time interval.
 5. **Monthly**: Select this option to perform the task on the selected days of the month.
 6. Select days of one or more months to perform the task.
 7. Select a specific weekday of one or more months to perform the task.
 8. Select the frequency of the task. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 9. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.

- iii. **Daily:** Select this option to perform the task daily.
 - 10. Select the frequency of the task in the **Recurs.**
 - 11. Select **Occurs once at** and mention the time, if you want to perform the task once in a day.
 - 12. Select **Occurs every** and mention the hours, start time and end time, if you want to perform the task in every few hours.
 - 13. **Once:** Select this option to perform the task only once.

78. Click the **Schedule** button to complete the action.

79. Scheduled tasks will appear in the bottom panel. You can reschedule a task by editing them from this screen. You can delete a task also.

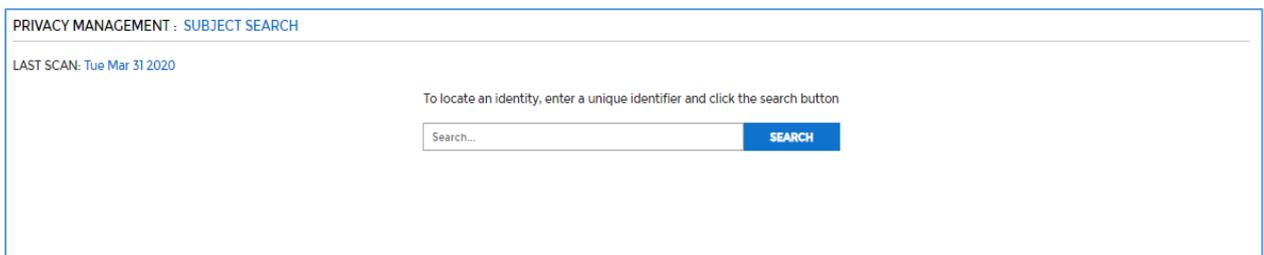
80.

16.5.9 View Report

Using the Subject Search screen, you can view the report for any Data Subject as per the specified Strong Identifier in the Search text box. It is an interactive way to access the report for any Data Subject.

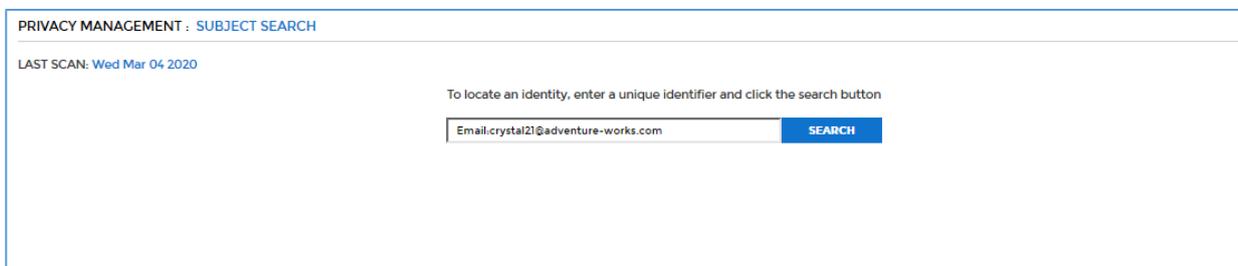
To access the **Subject Search** screen, click **Privacy > Subject Search**.

The **Subject Search** screen allows you to view the report based on the given **Strong Identifier**. It displays the list of all the databases in which the Data Subject is detected.



81. Enter the strong identifier and click **Search**. The **Results** pop-up will appear.

82.



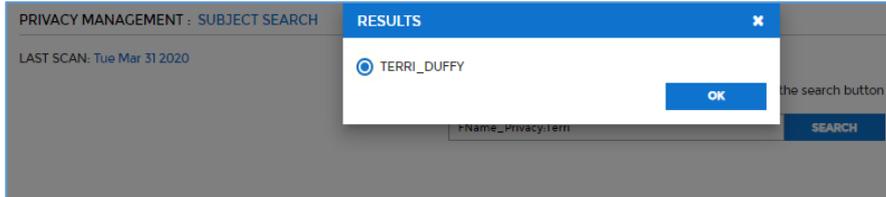
83.

84.

85.

86. Choose the identity in the **RESULTS** pop-up. Click **Ok**.

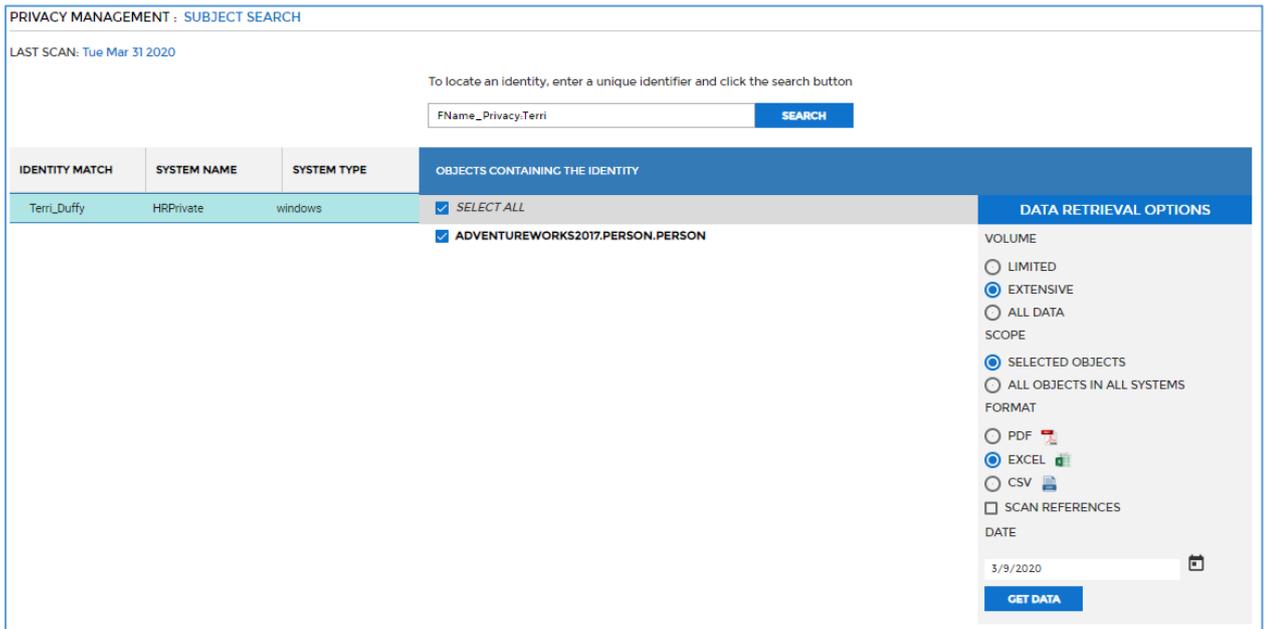
87.



88.

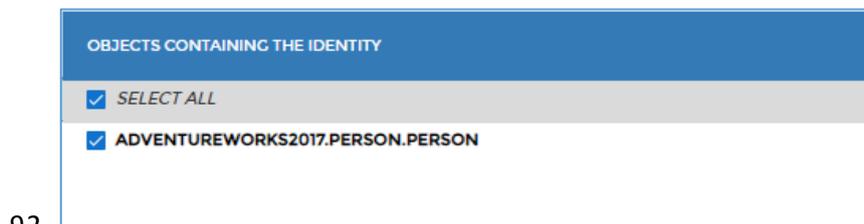
89.

90. The detailed data will be displayed showing the **Objects** name and **Data Retrieval Options**.



The left pane will display the matched **Identity** along with the **System Name** and **System Type** to which the identity is related. The middle pane will display the list of all the **Objects** i.e. tables in which the information for matched identity exists. The right pane will display the **Data Retrieval Options**.

91. Select the **Objects** by checking the check box next to each object name. The listed objects contain the data for matched identity.



92.

93.

94. Select the **Data Retrieval Options** and click **Get Data**. It will appear the **Report** tab.

95.

96.

97. The options available in **Data Retrieval Options** pane are:

98.

- i. **Volume:** Select **Limited**, **Extensive** or **All Data** option in volume.
 - 1. **Limited:** Select **Limited** option, if you want to scan only the first two records of identifier in the selected object.
 - 2. **Extensive:** Select **Extensive** option, if you want to scan the first 10-15 records in the selected object.
 - 3. **All Data:** Select **All Data** option, if you want to scan the entire list of selected objects.

- ii. **Scope:** Using the **Scope** option you can specify the Objects i.e. tables that need to be selected for generating a report. There are two options.
 - 1. **Selected Objects:** Select **Selected Objects** if you want to select few objects.
 - 2. **All Objects in All Systems:** Select **All Objects in All Systems** if you want to select all the listed objects.

- iii. **Format:** You can download the report in any of the specified format. There are three options available for format:
 - 1. PDF
 - 2. Excel

3. CSV

- iv. **Scan References:** Check the **Scan References** checkbox, if you want to look for the foreign key references in the other objects. Using this option, the foreign key which are associated in other objects are also selected for the scanning.
- v. **Date:** Select the date for which you have received a request from the customer. DPO or Admin who operates on DSAR Application needs to put this date manually before clicking **GET DATA**.

99. To download the Report, click **Privacy > Reports > SAR**.

PRIVACY MANAGEMENT : [REPORTS](#) > [SAR](#)

IDENTITY	REQUEST RECEIVED DATE	DATA RETRIEVAL START DATE	DATA RETRIEVAL COMPLETION DATE	VOLUME	SCOPE	FORMAT	REPORT
Terri_Duffy	2020-03-16 11:49:33	2020-03-16 06:22:25	2020-03-16 06:22:33	limited	selectedObject	CSV	DOWNLOAD_REPORT
Terri_Duffy	2020-02-07 15:38:32	2020-02-07 23:41:05	2020-02-07 23:41:10	allVolume	allObjectsInAllSystem	PDF	DOWNLOAD_REPORT
Terri_Duffy	2020-02-07 15:35:46	2020-02-07 23:37:43	2020-02-07 23:37:48	allVolume	allObjectsInAllSystem	CSV	DOWNLOAD_REPORT
Terri_Duffy	2020-02-05 09:42:55	2020-02-05 17:44:53	2020-02-05 17:45:02	allVolume	allObjectsInAllSystem	PDF	DOWNLOAD_REPORT

100.

101.

1. Select the report you wish to download and click **Download_Report** hyperlink in the **Report** column.

Appendix A: Verifying Hadoop Results

Verify that Hadoop tasks are finding targeted sensitive data by comparing the Task Results file or Detailed Results file with the file on which the task was run.

Task Results files can be found at `hadoop fs -cat /dataguise$/results/(taskname)/(task_instance_ID)`

Detailed Results files can be found at `hadoop fs -cat /dataguise$/results/(taskname)/(task_instance_ID)/summary_results_structured/`

The specific contents of the task results file differ slightly according to file type, and whether the file is treated as structured or unstructured.

Task Results files show the type and location of the discovered sensitive element. However, the discovered sensitive data element's actual value can be suppressed by setting the `results.suppressRealValues` property in the `HDFSIDPConfig.properties` file. For instance, if DgSecure discovers the email address dataguise@dataguise.com when this property is turned on, dataguise@dataguise.com will not appear in the results file. Please see the *Installation & Configuration Guide, section 3.1.1 HDFS IDP* for more information.

Task Results

Task results are available for tasks run against both structured and unstructured file types. The information displayed in the file varies according to the type of file that was scanned.

The format of a task results file is the same for structured text, structured Sequence, Avro, RC/ORC, and Parquet files.

When reading the files, please note that all offsets are from the beginning of the file.

Unstructured Text Files:

Begin offset of Line, End offset of Line, Begin offset of sensitive item, End offset of sensitive item,

***Unstructured Sequence Files:**

row # (within the map), col # (always 1), 0, 0,

***Note:** A Sequence file where the value field is unstructured text is considered "unstructured".

***Structured Files of All Types:**

row # (within the map), col #, sensitive item count, and whether the discovered item is a dependent or not (a numeric string representing the sensitive type ID = yes, a 0=no)

***Note:** A separate results file is created for each map:

Here is an example task results file for the scan of an ORC file:

```
hdfs://centos.cdh503:8020/user/hive/warehouse/orc_userdatatest/000000
_0 ----- (PATH OF ORIGINAL FILE)
  SSN-D 1      2      1      0      410312381
  Name  1      6      1      0      BRIAN
  EmAdr 1      7      1      0      BRIANxxx@gmail.com
  ABA   2      2      1      0      312765553
  Name  2      6      1      0      ERIK
  EmAdr 2      7      1      0      ERIKxxx@gmail.com
  ABA   3      2      1      0      343583325
  Name  3      6      1      0      EDWARD
  EmAdr 3      7      1      0      EDWARDxxx@gmail.com
  Name  4      6      1      0      DAVID
  EmAdr 4      7      1      0      DAVIDxxx@gmail.com
  ABA   5      2      1      0      686222131
  Name  5      6      1      0      TOMMY
  EmAdr 5      7      1      0      TOMMYxxx@gmail.com
  ABA   6      2      1      0      211333981
  Name  6      6      1      0      SEAN
  EmAdr 6      7      1      0      SEANxxx@gmail.com
```

Detailed Results

Field Results are only available for Detection tasks run against structured files. This file shows each sensitive type discovered, the type of scan (full vs incremental), the ratio of matched to total rows, and the number of columns in which sensitive data was found.

When the sensitive data was discovered using a Dependent expression, the Field Results file treats the Dependent expression as one expression. In the field column, multiple column numbers would be displayed. Also, each sensitive comprising expression would be listed in the SensitiveType column.

Filed(s):	Match/Total Rows		SensitiveType		Sample Mode
Column 2:	156/13474		NPI		Full scan.
Column 3:	13318/13474		Telephone (Digits Only)		Full scan.
Column 1:	2973/13474		Social Security # (Digits Only)		Full scan.
Column 12:	10495/13474		ABA Routing number		Incremental scan.
Column 0:	13474/13474		Credit Card # (Digits Only)		Full scan.

NPI found in 1 column(s).

Social Security # (Digits Only) found in 1 column(s).

Telephone (Digits Only) found in 1 column(s).

ABA Routing number found in 1 column(s).

Credit Card # (Digits Only) found in 1 column(s).

Appendix B: EDI Transaction Set

Currently, Dataguise uses EDI transaction set 837 "Healthcare claim". In order to alter the transaction set being used, please contact Dataguise Professional Services.

The transaction set is used to parse data in the EDI file format. The parsing starts from the tag of each EDI line and is based on the value of each column or the position of each column. Once parsing is done for an EDI line, the sensitive data in that line is detected. Dataguise sensitive type ID is also called regex ID under some documentations.

The EDI transaction set is combined with menu.txt and multiple dictionary files. Menu.txt is the catalog for all dictionary files. Menu.txt can contain multiple lines and each line defines one column of a tag.

One directory file can contain multiple lines and each line defines one value for the related column.

menu.txt format:

```
[EDI tag]*[EDI column number (start from 0)]*[dictionary file name]
```

dictionary file format:

```
[EDI value for the column defined in menu.txt](+[offset from such column]):[Dataguise sensitive type ID list (split by ',')]#[comments]
```

Example 1:

menu.txt:

...

```
NM1*8*Identification Code Qualifier.txt
```

...

Above line means: if an EDI line starts with tag "NM1", we look up dictionary file "Identification Code Qualifier.txt" for how to process its column 8.

Identification Code Qualifier.txt:

...

```
34(+1):4,5,6#Social Security Number
```

...

Above line means: if column 8 has value "34", its next column (+1) can contain Security Security Number (whose sensitive type ID is 4, 5, or 6)

Here is an EDI line which will match above dictionary entry:

```
NM1*IL*1*Duo*John*X*Mr**34*365118888*12*05
```

Example 2:

menu.txt:

...

```
N4*4*country.txt
```

...

Above line means: if an EDI line starts with tag "N4", we look up dictionary file "country.txt" for how to process it column 4.

country.txt:

```
~(+0):77#Country
```

Above line means: regardless what value is on column 4 (this is meaning of ~), its current column (+0) can be country name.

Here is an EDI line which will match above dictionary entry:

```
N4*Fremont*CA*94538*USA
```

Appendix C: Error Messages

Error Text	Error Code
Task does not exist	100
Cluster Name does not exist	101
ScanLocation does not exist	102
Cluster name should not be null	103
Cluster name should not be blank	104
Cluster does not exist	105
Policy does not exist	106
Specified path doesn't exist	107
Unable to find the task	108
Specified task type does not exist	109
No results found!	110
Unsupported JSON	111
Input JSON is malformed	112
Syntax Error	113
Special characters are not allowed except underscores	114
The system cannot find the path specified	115
No cluster has been added to session	116
Access denied	117
Task Status	118
Unable to get details of instance id	119
Please remove the trailing slash(es) in scanlocation	120
No such Hadoop policy found	121
No policy specified	122
Both excludeListFilePath and excludedScanPathList cannot be specified	123
Path specified to Exclude File list does not exist	124
Please select only .txt file for Exclude File list	125
Unsupported tasktype	126
Special characters are not allowed except underscore	127
Search Parameters can only be specified for Auto Discovery tasks	128
File Extensions in excludedFileExtensions should start with .	129
Either deleteExclusionList should be set to false or excludeListFilePath should not be specified	130
Structured flag cannot be set to true for {Row Encryption, Discovery, Detection} tasks	131
OutputColumnForm can only be specified for {columnar masking} task	132
DeleteInputFiles cannot be specified	133
Search Parameters must be specified	134
discoveryCriteria must be specified	135
fullFileScan must be specified	136
fileModifiedBefore can never be greater than today's date	137
File Modified Date Range is invalid.	138
Scan location not associated with any structure therefore verifyStructure cannot be set	139

to true	
Scan Location are not part of any domain;	140
Validation Error	141
Text Structure Validation Error	142
Sequence Structure Validation Error	143
structure does not exist	144
Column Number can't be repeated within the structure	145
Field Name does not exist	146
columnNumber cannot be less than 1	147
Either columnType or complexStructureName must be specified in column definition	148
Only one of columnType or complexStructureName could be specified in column definition	149
encKeyDesc does not exist	150
Server not reachable	151
domain does not exist	152
directory path does not exist	153
Only directories are allowed to be specified	154
Cannot update scanpath which is not yet associated anywhere	155
Entity and Key Value not exist	156
Structure verification failed for given Task Type	157
outputDir cannot be specified with structure definition	158
outputDir must be specified	159
outputDir cannot be blank	160
branchPoint cannot be specified with structure definition	161
branchPoint must be specified	162
Database does not exist	163
Please provide database	164
Table does not exist	165
Please provide table name	166
samplingType is not set	167
samplingValue is not set	168
At least one Policy must be specified	169
Task name in command is different from the task name defined in JSON file	171
Table does not exist in database	172
Table does not exist in given database	173
column does not match in Table	174
Table does not exist in tableName List	175
encKeyName does not exist	176
fpEncKeyName does not exist	177
At least one Policy/Expression must be specified	178
Invalid computecluster name	179
At least one Connection to Search must be specified	180
sampleDataStart cannot have value - {Complete} with useSafeList set to true	181
useSafeList can only be set for {discover with count} tasks	182

Table Filter should only contain letters, numbers and _	183
Column Filter should only contain letters, numbers and _	184
No connection name found	185
No policy name with name	186
Invalid Tasktype or Connection IDP value	187
Group Name does not exist	188
connection already exists	189
Database Name does not exist	190
Schema does not exist	191
Invalid Command, cannot specify cluster name	192
Unable to get the task information	254
Session is no longer valid	255
Invalid input	256

Appendix D: Role Based Access Controls

Role-based access control (RBAC) is a method of regulating access to computer and/or network resources based on the roles of individual users within an enterprise. In this context, access is the ability of an individual user to perform a specific task, such as view, create, or modify a file.

Users belonging to a specific role create a policy and members of other roles will not have access to these policies. The user creating the policy will have access to set permissions on the custom policies – Create, Update, Read, D. These users will have privileges to create and manage structures, policies, and sensitive types. Admins will be able to define global policies (open to all users) and if a private policy needs to be opened to the public, an Admin will need to be contacted. Tasks will not be executed nor be able to be modified if they contain a policy that the user does not have access to.

Users with “DEFAULT” role have “Product” and “Owner” access by default on Policies, Sensitive Types, Domain, Structures, etc. This means they are able to see the objects created by themselves in addition to the predefined ones. They can also be given access to objects created by others.

Under “Owner Access” permissions, every role type user will be able to access their own created items.

Under “Full Access” permissions, every role type will be able to access all items – those created by themselves and others.

To provide access on individual items (task, policy, connections, structures, etc) to specific role types, Super_Admin can set the “R,U,D,E” permissions as required.

Appendix E: Voltage SimpleAPI

In order to use the Voltage SimpleAPI to obfuscate your sensitive data with DgSecure, please follow the instructions below.

1. Create **CustomProtectionDefinition.json** file with following parameters for the sensitive types on which we need to enable **VolProtection**

```
{
"sensitiveType": "Credit Card # (Digits Only)",
"customProtectionDetails": [{
"protectionName": "VolProtection",
"paramsList": ["TaskType","ProtectionType","Format","Identity","SharedSecret"]
}]
},
{
"sensitiveType": "Credit Card # (Space Separation)",
"customProtectionDetails": [{
"protectionName": "VolProtection",
"paramsList": ["TaskType","ProtectionType","Format","Identity","SharedSecret"]
}]
}]
}
```

- A. Put the **CustomProtectionDefinition.json** file under directory `.../webapps/dgcontroller/`
- B. Update `CustomProtectionConfig.properties` file at `.../HDFSIDP/expandedArchive/WEB-INF/classes/com/dataguise/Hadoop/util/`

Setting up the configuration properties:

a. **policyURL**=<https://voltage-pp-0000.dataprotection.voltage.com/policy/clientPolicy.xml>

This could be different on production environment.

b. **trustStore** with root certificates must be available on all nodes of cluster at the same path with resetting the permissions to be world-readable.

c. Update the `trustStore` property with the `trustStore` path.

Ex: **trustStore=/opt/voltage/trustStore**

d. Provide the cache path for caching the cryptographic keys and other centrally managed information in file.

cache =/opt/voltage/cache

- Other properties are optional.
- C. Put **VolProtection.jar** file under .../HDFSIDP/expandedArchive/WEB-INF/plugins/custom_protection
 - D. Restart tomcat.
 - E. Restart HDFSIDP service.

 - F. Create Masking/Encryption policy for VolProtection task by setting the protection parameters.

Acceptable values for VolProtection parameters are:

TaskType: It could be one of the following-

- a. *Protect*
- b. *Access*
- c. *Digest*

ProtectionType: Acceptable values for ProtectionType parameter:-

For TaskType **Protect/Access**:-

- a. *FPE*
- b. *AES*

For TaskType **Digest**:-

- a. *DigestMD5*
- b. *DigestSHA1*
- c. *DigestSHA224*
- d. *DigestSHA256*
- e. *DigestSHA384*
- f. *DigestSHA512*

Format: It is the FormatName which is defined for FPE on voltage appliance.

Example: **AlphaNumeric**

Identity: authIdentity is defined for particular FormatMapping on voltage appliance.

Example: [test@test.int](#) or

[test@voltage.com](#)

SharedSecret: Shared secret to be used for Key Server authentication.

Example: "**voltage123**"

- G. Create a domain and associate above policy to it along with the input directory.
- H. Create a structure for an input file.
- I. Create and execute structured Protection(Masking/Encryption) task on input file with VolProtectionPolicy.

Appendix E: Snappy Files Support

In order to run masking/encryption on Snappy Compressed file type, following these steps.

1. Copy the following files from your cluster:

- a. libsnappy.so
- b. libsnappy.so.1
- c. libsnappy.so.1.1.4

2. Copy the above files to a location of your choice on the machine where DgSecure is installed:

Ex. /usr/java/jdk1.7.0_67/lib/amd64

3. Update the jetty-embedded.properties file as follows

(/.../Dataguide/DgSecure/IDPs/HDFSIDP):

JavaOptions=-Dhdp.version=<HadoopVersion> -Djava.library.path=<path selected in step 2>

Ex. JavaOptions=-Dhdp.version=2.3.4.0-3485 -

Djava.library.path=/usr/java/jdk1.7.0_67/lib/amd64

4. Restart the HDFS IDP before running your next task.

Appendix F: Lightweight Primitives

These are created to ease the task creation process for users that keep the task type, options, and policies remain the same, whereas the directory paths and their corresponding structure definitions are the things that change per task.

The lightweight primitives take advantage of this information and we can create a task template with the repeated details: task type, task options, and policies, and stores these details into the database.

Lightweight Primitives

Task Template: The key new primitive is the *Task Template*. The Task Template defines the following:

1. Task type
2. Task options
3. Policies

The Task Template is stored in the database.

Structure Definition: With the lightweight primitives, the `create structure` statement is not needed. The JSON file description of the structure is still created as a file, and passed on with scan paths to execute a task template. Structures are *not* stored in the database, as the `create structure` statement is never executed.

There is no “Task Definition” in the lightweight primitives set.

Task Instance: Gets created for each execution of a task template.

In an abstract sense, the task template gets executed as follows:

```
execute task_template (ListOf<scan path, structure for scan path>)
```

Initially, Task Templates will only be supported for Hadoop, and only from DGCL. Later, support will be extended to S3, GCS, DBMS etc., and also support from the UI will be available.

Usage of the Lightweight Primitives

1. Create/copy a JSON task template file, containing the task type, options, and policies, in the local file system where DGCL executes
2. Create/copy a JSON structure definition file in the local file system where DGCL executes.

From DGCL:

```
create hadoop task_template from <task_template_file>
```

```
run hadoop task_template (  
    {"scanpath1", "structureFile1"},  
    {"scanpath2", "structureFile2"},  
    {"scanpath3", "structureFile3"}  
)
```

Here, structureFile1..3 refer to files that describe the structure in JSON format. Additionally, we will explore the ability to plug in the structure JSON in the `run task_template` statement, so that the structure file is not needed.

From a database storage perspective, only the task template and details for each task execution (i.e. task instance) are stored. There is no storage of task definitions or structure definitions. The task template definition (which will be a JSON description), will contain additional options to control what will be saved in the database, further reducing the storage requirements for the executing task.

Appendix G: Key-Pair Authentication for Snowflake Connections

To configure the public/private key pair:

From the command line in a terminal window, generate a private key:

```
$ openssl genrsa 2048 | openssl pkcs8 -topk8 -inform PEM -out rsa_key.p8
```

OpenSSL prompts for a passphrase used to encrypt the private key file. Record this passphrase. You will input it when connecting to Snowflake.

From the command line, generate the public key by referencing the private key:

```
$ openssl rsa -in rsa_key.p8 -pubout -out rsa_key.pub
```

Copy the public and private key files to a local directory for storage. Record the path to the files.

***Note** that the private key is stored using the PKCS#8 (Public Key Cryptography Standards) format and is encrypted using the passphrase you specified in the previous step; however, the file should still be protected from unauthorized access using the file permission mechanism provided by your operating system. It is your responsibility to secure the file when it is not being used.

DgSecure will then apply **AES encryption** layer above this encrypted private key and store it on the Agent side:

```
DBMS Detection Agent (\ dgDiscoverAgent \ WEB-
```

```
INF \ classes \ com \ dataguisse \ discoverAgent \ keystore \ <username>.txt);
```

```
DBMS Masking Agent (\ dgAgent \ WEB-
```

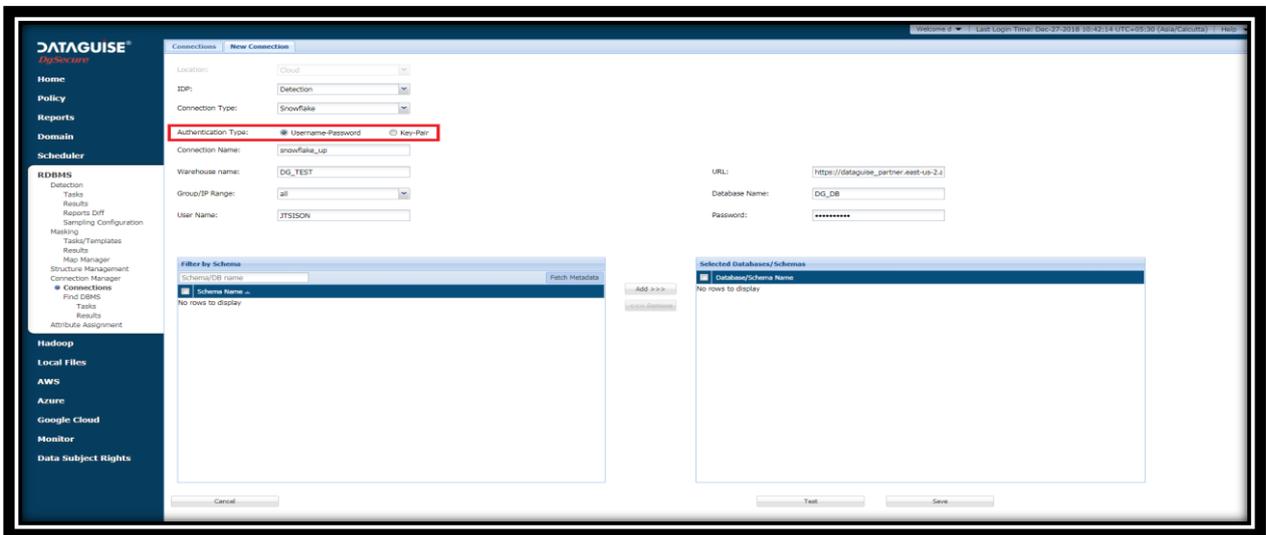
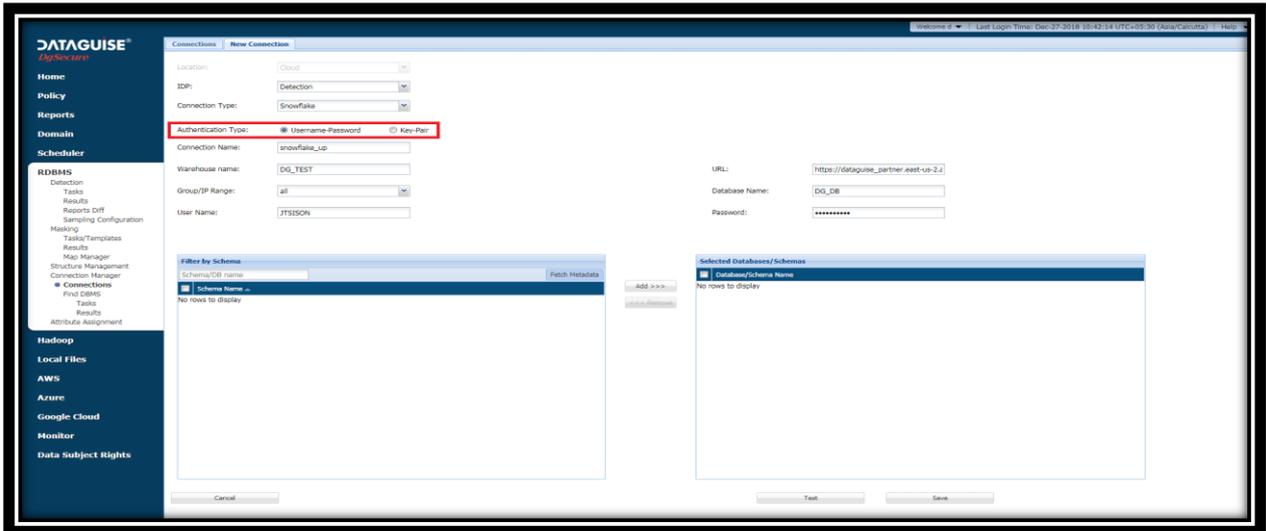
```
INF \ classes \ com \ dataguisse \ Worker \ Masker \ PrerequisiteFiles \ keystore \ <username>.txt)
```

Assign the public key to the Snowflake user using [ALTER USER](#). For example:

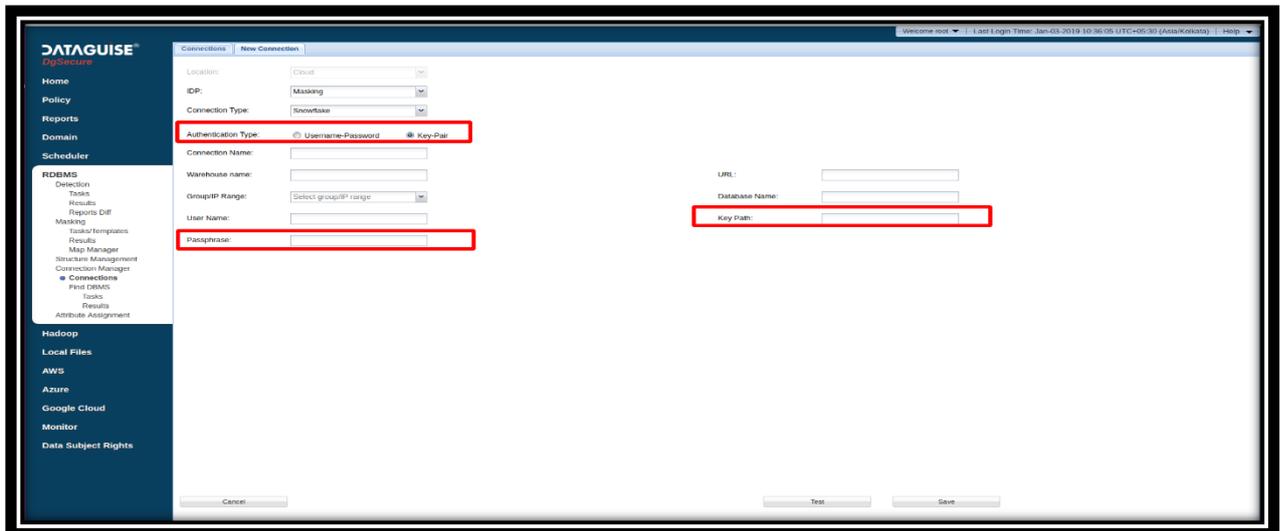
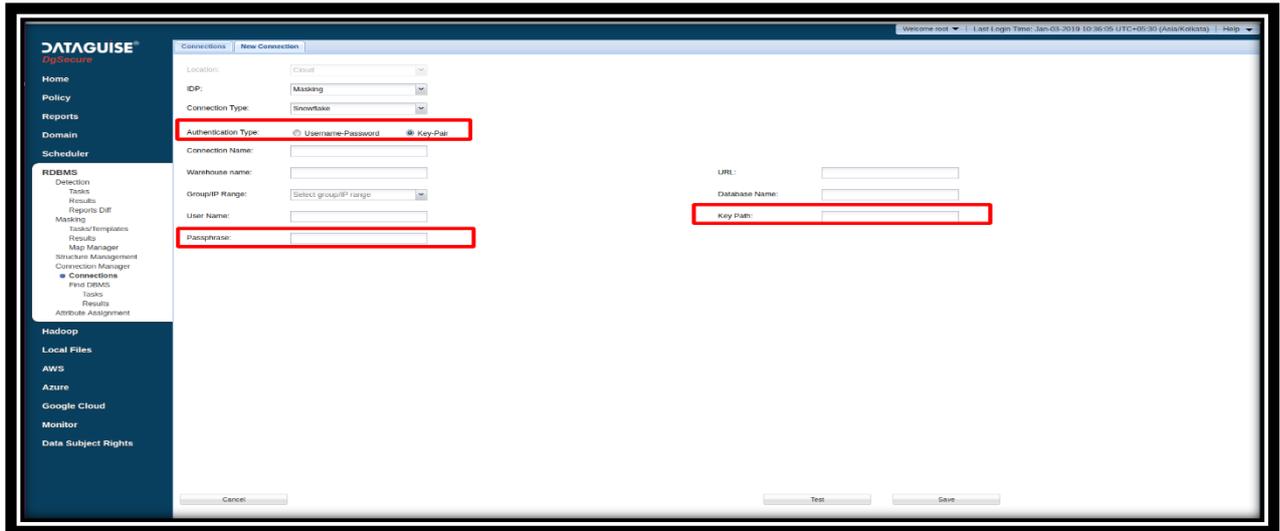
```
alter user jsmith set rsa_public_key='MIIBIjANBgkqh...';
```

DgSecure:

Authentication Type: Username-Password



Authentication Type: Key-Pair



Private Key path and passphrase are required to authenticate.

NOTE: Private Key should be placed on Agent side.

Case a: Private Key path contains the private key

Agent looks into the specified location for the private key. If found, Agent uses it and stores it in AES encrypted format (location above-mentioned).

Note supreme priority is given to the 'Key Path' such that Agent always stores the updated private key as provided by the user.

Case b: Private Key path does not hold the private key

Agent is unable to fetch the private key from the specified location. Agent looks for the private key on its side (<username>.txt) and tries to authenticate using it.

'Test Connection' and running of (Detection or Masking) task demands establishing a

connection with the specified database using the mentioned credentials; hence this forces to store the encrypted private key in AES encrypted format on Agent side.

Appendix H: Detection in Kerberized Clusters

Following are the config properties required for YARN-specific Kerberos Configurations, which we need to configure in the spark.properties file available at the location:

.../HDFSAGENT/expandedArchive/WEB-INF/classes/spark.properties

On IDP deployment machine.

- **spark.yarn.principal**
Principal to be used to login to KDC, while running on secure clusters.
- **spark.yarn.keytab**
The full path to the file that contains the keytab for the principal specified above. This keytab will be copied to the node running the YARN Application Master via the YARN Distributed Cache, and will be used for renewing the login tickets and the delegation tokens periodically.
- **spark.yarn.access.hadoopFileSystems**
A comma-separated list of secure Hadoop filesystems your Spark application is going to access.

For example,

```
spark.yarn.access.hadoopFileSystems=hdfs://nn1.com:8032,hdfs://nn2.com:8032,  
webhdfs://nn3.com:50070.
```

The Spark application must have access to the filesystems listed and Kerberos must be properly configured to be able to access them (either in the same realm or in a trusted realm). Spark acquires security tokens for each of the filesystems so that the Spark application can access those remote Hadoop filesystems.

- **spark.yarn.kerberos.relogin.period**
How often to check whether the kerberos TGT should be renewed. This should be set to a value that is shorter than the TGT renewal period (or the TGT lifetime if TGT renewal is not enabled). The default value should be enough for most deployments.

Appendix I: SL Masking

Configure SL Masking for Kerberos User in Oracle

To perform SL masking, we need to create a database user to load the library into; this user is identified externally by the kerberos user.

To create a user, follow the below steps before masking is run:

1. Create a new user which is identified externally by Kerberos user. The username must be same and should be without domain name. The names are case sensitive. Following is the command:

```
create user <KRBUSER_NAME> identified externally as '<KRBUSER@EXAMPLE.COM>'
```

For example,

```
create user ORAKRB identified externally as 'ORAKRB@DGAD.COM'.
```

2. Give the grants required for masking to <KRBUSER>; e.g. ORAKRB. The grants can be found under prerequisites folder for oracle in DgMaskerAgent installation.

Steps 3-4 are specific to SL masking

3. Load the jar file in <KRBUSER> using below command.
4. `loadjava -user sys/sys -schema ORAKRB <location for DgFPELib.jar>`
where ORKRB is the <KRBUSER> as specified in the step 1 example.
5. Give grants mentioned in GrantsToAccessJavaFromRDBMS.sql to <KRBUSER>.

Run the masking task.

***Note:** During SL masking task execution, if an error related to run time permission is encountered then provide the grant specified below and re-execute the task.

```
Call dbms_java.grant_permission( 'ORAKRB', 'SYS:java.lang.RuntimePermission',  
'accessClassInPackage.sun.misc',")
```

Where 'ORAKRB' is the Kerberos user i.e. <KRBUSER>

Enabling SL Masking in Oracle

Some pre-requisites are required before executing SL masking and Encryption/Decryption in Oracle for 'On-Premises'. Follow the below steps to executing SL masking in Oracle:

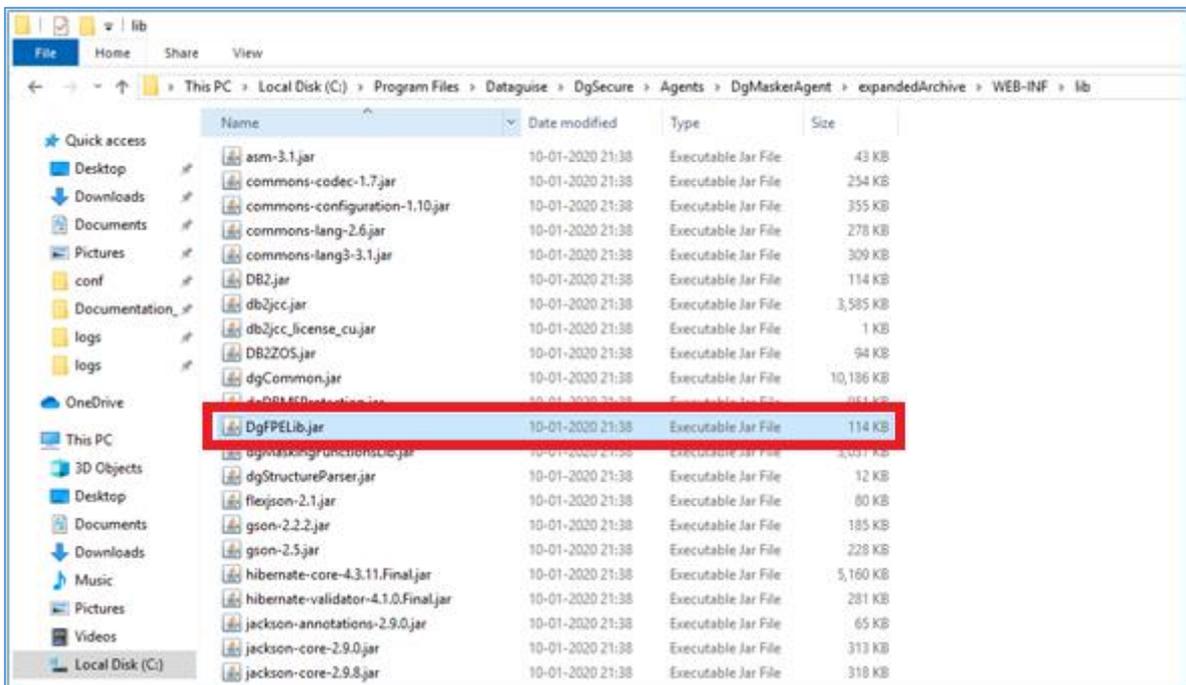
1. Open **GrantsToAccessJavaFromRDBMS.sql** file from location: ".../x.x.x/x.x.x.0.0.0_x_DD-M-YYYY/PrerequisiteScripts/Oracle/".

This file has grant scripts that are needed to access java function from RDBMS. Follow the steps written in this file to provide the grants.

2. Copy **DgFPELib.jar** file to Oracle server location.

Windows location:

C:/Program Files/Dataguisse/DgSecure/Agents/DgMaskerAgent/expandedArchive/WEB-INF/lib



Linux location:

/opt/Dataguisse/DgSecure/Agents/DgMaskerAgent/expandedArchive/WEB-INF/lib

Name	Size	Changed
..		2/21/2020
asm-3.1.jar	43 KB	2/7/2020
commons-codec-1.7.jar	254 KB	2/7/2020
commons-configuration-1.10.jar	355 KB	2/7/2020
commons-lang-2.6.jar	278 KB	2/7/2020
commons-lang3-3.1.jar	309 KB	2/7/2020
DB2.jar	114 KB	2/7/2020
db2jcc.jar	3,585 KB	2/7/2020
db2jcc_license_cu.jar	1 KB	2/7/2020
DB2ZOS.jar	94 KB	2/7/2020
dgCommon.jar	10,188 KB	2/7/2020
dgDBMSPProtection.jar	951 KB	2/7/2020
DgFPELib.jar	114 KB	2/7/2020
dgMaskingFunctionsLib.jar	3,037 KB	2/7/2020
dgStructureParser.jar	12 KB	2/7/2020
flexjson-2.1.jar	80 KB	2/7/2020
gson-2.2.2.jar	185 KB	2/7/2020

- To load jar file, execute below mentioned command.

Windows: execute the below command from command prompt, using the same username and password with which connection is created in Connection Manager in DgSecure.

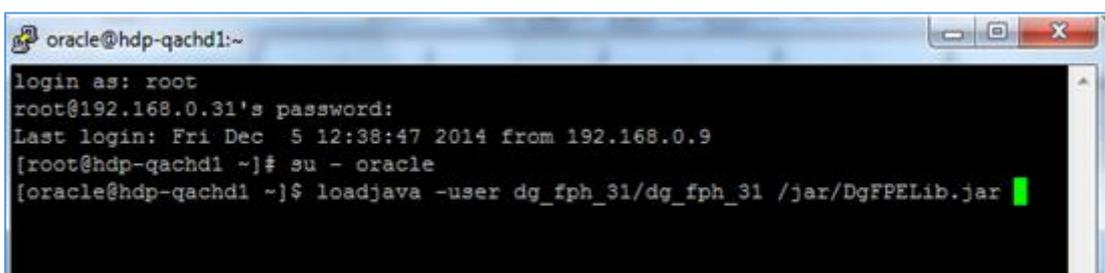
`loadjava -user <username>/<password> <location of DgFPELib.jar file>`

For e.g. `loadjava -user hr/hr D:\DgFPELib.jar`

Linux:

`loadjava(space)-user(space)Dg-user(dgmasker)/Dg-user(password)(space)location of the JAR file e.g./opt/FPE/DgFPElib.jar.`

For e.g. `loadjava -user dg_fhp_31/dg_fhp_31 /jar/DgFPELib.jar`



In the above screenshot, the `dg_fhp_31` is the masking user. This user is used to create connection for masking.

Enable Random-SL Masking Support for HBase

The system has been enhanced to provide Random SL Masking support to HBase. This maintains the

consistency of the repetitive values while masking. The repeated sensitive information spotted during the detection task in the database will be masked with the same value throughout to maintain consistency.

To support this functionality, the following property has been added:

```
# Restrict overwriting of any table, which is also selected as a destination table for masking.
# By default, this is set to Y to restrict the overwriting. Change to N, for forceful overwrite.
# If the destination table selected in masking task already exists, masking of the source table would be skipped, if set to Y.
dg.restrict.destination.table.overwrite = Y

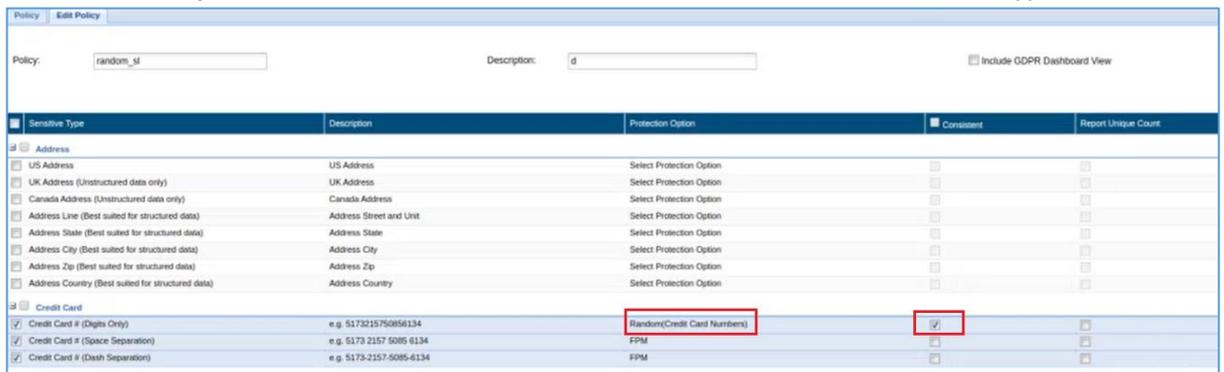
# Consistent masking engine type, default value is SL
dg.cengine = SL
```

The file in which the above mentioned property exists can be located under:

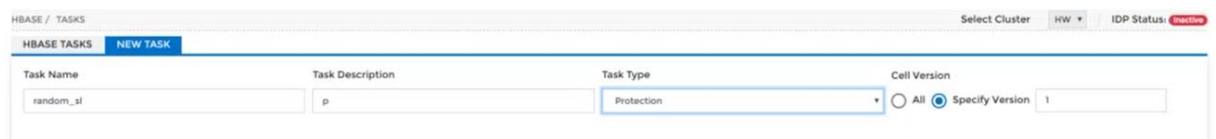
<Installation_Directory>/Datagui/DgSecure/Agents/HBaseAgent/expandedArchive/Web-INF/classes/HBaseIDPConfig.properties

Perform the following steps to provide Random SL Masking support in HBase:

1. Create a new policy for HBase. While selecting the sensitive type, select the **Random** option for **Protection option** and check the **Consistent** checkbox for the selected sensitive type.

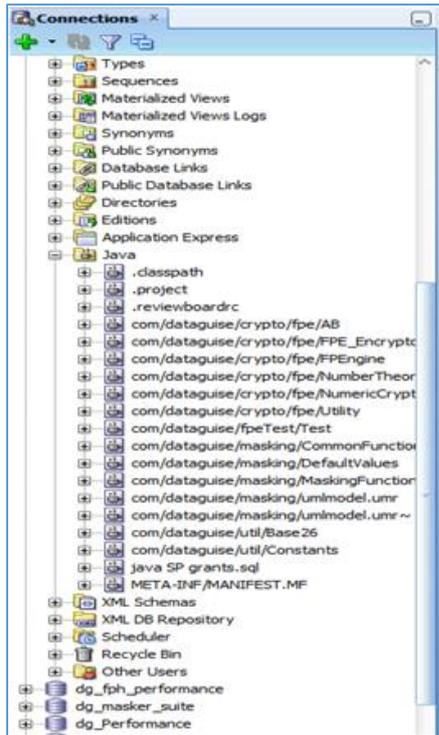


2. Create an HBase task. Go to **HBase > Tasks > New Tasks**.



3. **Save and Execute** the task.

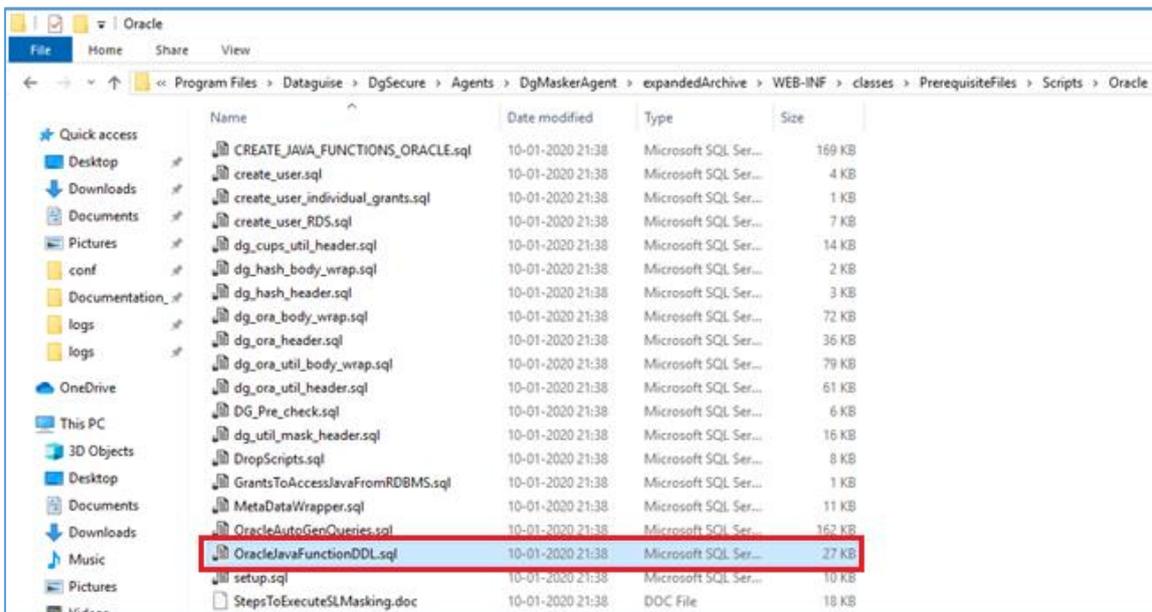
4. Verify JAR file is uploaded or not, check JAVA folder in DG User (dgMasker).



- Run **OracleJavaFunctionDDL** file in Dg User (DgMasker).

Windows location: The script is available under following location

‘C:\Program Files\Dataguisse\DgSecure\Agents\DgMaskerAgent\expandedArchive\WEB-INF\classes\PrerequisiteFiles\Scripts\Oracle’



Linux location:

'/opt/Dataguisse/DgSecure/Agents/DgMaskerAgent/expandedArchive/WEB-INF/classes/PrerequisiteFiles/Scripts/Oracle'

6. Set "**JAVA_JIT_ENABLED**" to true.

Once the permission has been granted, below mentioned masking options will be available for SL masking:

- a) Regular Expression
- b) FPM
- c) IntelliMask
- d) Random (All available masking options)
- e) NPI

Upon selecting SL option in DgSecure, the Consistent (C) and Unique (U) options gets selected automatically. However, for some of the masking options only Consistent (C) is selected and Unique (U) is disabled; and they cannot be selected manually. These masking options are:

- a) Regular Expression
- b) Intellimask
- c) Random (Address Line 1, Address Line 2, First Name, Last Name, First and Last Name, Email Addresses, City, State Country and Random String)

If you want to generate Consistent Unique data for Regular Expression, Intellimask and Random (Address Line 1, Address Line 2, First Name, Last Name, First and Last Name, Email, Addresses, City, State, Country and Random String). You can apply FPM or Shuffle masking with SL option.

There are some other important points related to SL:

1. SL option does not allow to select persistent dataset from sync window.
2. SL option is not supported with Custom masking.

Appendix J: Configure Multiple SQL Server Instances in DgSecure

DgSecure supports detection on multiple SQL Server instances. To connect with multiple instances, the Detection IDP must have access to the SQL Server to get a list of all the available instances. To achieve this an sqlcmd utility must be installed on the machine where the Detection IDP is installed. To download sqlcmd utility, click on the following link:

<https://www.microsoft.com/en-us/download/details.aspx?id=53591>

There are two ways to obtain the list of SQL Server Instances:

1. Execute the following .net function, after executing the function call this function from the java:

```

using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Data.Sql;
using System.Data;

namespace ConsoleApplication2
{
    class Program
    {
        static void Main()
        {
            using (DataTable sqlSources =
                SqlDataSourceEnumerator.Instance.GetDataSources())
            {
                foreach (DataRow source in sqlSources.Rows)
                {
                    string servername;
                    string instanceName =
                        source["InstanceName"].ToString();

                    if (!string.IsNullOrEmpty(instanceName))
                    {
                        servername = source["InstanceName"] +
                            "\\\" + source["ServerName"];
                    }
                    else
                    {
                        servername =
                            source["ServerName"].ToString();
                    }
                    Console.WriteLine(" Server Name:{0}",
                        servername);
                    Console.WriteLine("      Version:{0}",
                        source["Version"]);
                    Console.WriteLine();
                }
                Console.ReadKey();
            }
        }
    }
}

```

***NOTE:** Following are limitations to using the above mentioned .net function:

- SQL Server Browser must be running at the time of execution of this function else incorrect information may be displayed.
 - Only the named SQL Server instances will be listed.
-

2. Execute the sqlcmd-L command to get the list of available instances. The code to execute this command through java is as follows:

```
-----  
-----  
  
        Runtime rt = Runtime.getRuntime();  
  
        Process pr = rt.exec("sqlcmd -L");  
  
        BufferedReader input = new BufferedReader(new  
InputStreamReader(pr.getInputStream()));  
  
        String line=null;  
  
        while((line=input.readLine()) != null) {  
            System.out.println(line);  
        }  
        int exitVal = pr.waitFor();  
  
        System.out.println("Exited with error code  
"+exitVal);
```

***NOTE:** The command 'sqlcmd-L' can only be executed on SQL Server machine containing MS Tools. To install MS Tools click on the following link:

<https://docs.microsoft.com/en-us/sql/linux/sql-server-linux-setup-tools>

Appendix K: Confidence Factor

MISSCOUNT = TOTALCOUNT - (HITCOUNT + NULLCOUNT);

CONFIDENCE = 100 * (HITCOUNT - MISSCOUNT) / (TOTALCOUNT - NULLCOUNT)

IF (ISHEADERMATCH)

CONFIDENCE += HEADERWEIGHTAGE - REDUCTIONFACTOR * (TOTALCOUNT - NULLCOUNT) / TOTALCOUNT;

Here, TOTALCOUNT refers to the row scanned and HITCOUNT refers to the match count.