

Quickstart with Kyligence Cloud on Azure

This article will show you how to quickly deploy and get started with Kyligence Cloud on the Azure platform. By using the built-in New York Taxi dataset and model, and by crafting a visual chart with Kyligence Insight, a built-in visual data analysis tool. It is recommended that you use the Chrome (64.0.* or higher version) browser for the following steps.

- [What you'll know](#)
- [Basic concepts](#)
- [Prerequisites](#)
- [Deploying Kyligence Cloud](#)
- [Create a workspace](#)
- [Create a project, create a table, synchronize a table](#)
- [Data Analysis](#)
- [Uninstall](#)

What you'll know

In this tutorial, you'll learn about the following:

- Quickly deploy Kyligence Cloud on the Azure platform
- Create workspaces and projects in Kyligence Cloud
- Import built-in sample datasets and models
- Deploy the visual data analysis tool Kyligence Insight and create an analysis chart
- Clean up Kyligence Cloud-related resources

Basic concepts

- Workspaces: Workspaces are the first-level management units under Kyligence Cloud, each using a different cluster, logically isolated from each other, and data is not shared, for example, you can deploy 3 workspaces to meet your needs in accordance with development, testing, and production requirements.
- Projects: Projects are a level-one management unit under each workspace, projects under the same workspace share the same set of cluster resources,

you can create multiple projects in one workspace and serve different business scopes, in one project, you can design multiple models and query and analyze.

- **Data Catalog:** Data Catalog is a metadata management service in Kyligence Cloud that reads files from cloud object storage (Blob, ADLS Gen2, S3, etc.) and defines their table structure. A workspace with a data source type of Object Storage creates and uses a data catalog. Each workspace creates a data catalog, which is shared by all items in the workspace, and the data assets are published and managed in a unified way.
- **Synchronized tables:** Because the data catalog is shared by all items under the workspace, the tables in the data catalog need to be synchronized to the current project before the user can perform data analysis.
- **Model:** Model, which is also the logical semantic layer. A model is a set of tables and the associations (Join Relationships) between them. The model defines fact tables, dimension tables, measures, dimensions, and a set of indexes. The model and its indexes define the estimates to be performed when the data is loaded, and currently support star and snowflake models.
- **Index:** The index is built when the data is loaded and the index is used to speed up queries. Indexes are divided into Aggregate Index and Table Index. Aggregate indexing is essentially a combination of multiple dimensions and measures, suitable for answering aggregate queries, such as total sales for a year, and detail indexes are essentially multiple indexes of large and wide tables, which are suitable for answering detail queries that are scarce to records, such as a user's last 100 transactions.
- **Loading data:** In order to speed up queries, data needs to be loaded into the model from the source table, and an index will be built in the process, which is the estimation process for the data. Each data load produces a Segment. The model after the data is loaded can serve queries, and queries executed on the model are greatly accelerated due to the estimated calculations.
 - **Incremental Data Load:** You can define a partition date or time column on the fact table. Depending on the partition column, you can do incremental loading of very large datasets by time range.
 - **Full Load:** If a partition column is not defined, all the data in the source table is loaded at once.
 - **Rebuild Index:** Users can adjust the definition of models and indexes at any time. For loaded data, the index on it needs to be rebuilt with a new definition. If the user requests that certain queries be accelerated, the system may also optimize the model and index, which in turn triggers the indexing to be rebuilt.
- **Query acceleration:** The ability to speed up queries by automatically optimizing models and indexes. Models and indexes can be automatically optimized based on historical query patterns and dataset characteristics. This saves users a lot of time manually designing models and indexes.

Prerequisites

Kyligence Cloud uses azure app registration to obtain operational authorization to deploy the required resources, make sure that the Azure subscription you are using has permission to create the following services, and if you do not confirm that you have the following permissions, contact your cloud platform administrator:

Resource	Type	Version	Provider
VNet	Network	-	Azure
Network Security Group	Network	-	Azure
Elastic Load Balancer	Network	-	Azure
Network Interface	Network	-	Azure
Azure Database for MySQL	Database	MySQL 5.7	Azure
VM	Compute	OS: Ubuntu 16.04	Azure
Azure Blob Storage	Storage	-	Azure
Disk Storage	Storage	-	Azure

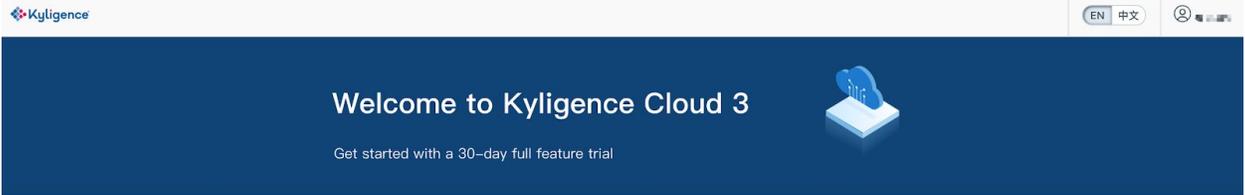
Before you deploy Kyligence Cloud, you need to prepare the following

- Kyligence Cloud trial license
- Directory (tenant) ID
- Application (client)ID
- The application password

Deployment

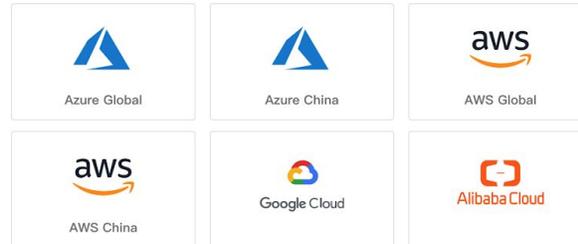
Role: IT Engineer

1. Please visit the Kyligence Cloud [Installer Wizard](#) and then select the platform to deploy Kyligence Cloud services.

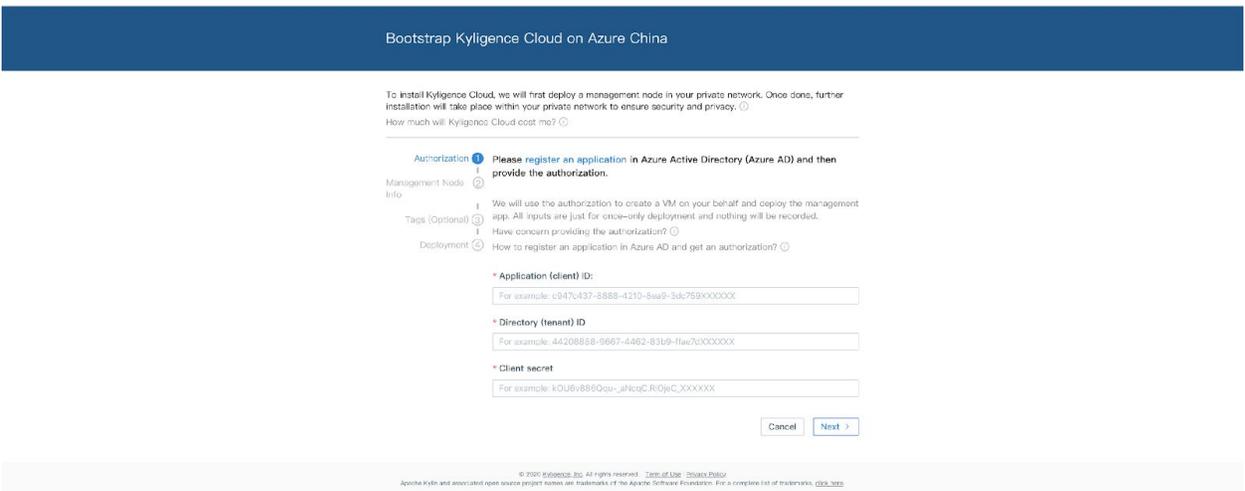


Select a cloud platform

Only 3 steps to provision Kyligence Cloud for you. [Learn more](#)



2. **Tenant ID, Application ID and Application Key** are required during this process. If you do not have these resources available, please refer to the [Prerequisites](#) section or contact your Platform Administrator.



3. On the basic information page, please fill in the following details:

Authorization ⓘ Please provide information about the VM and the management app to be created.

Management Node Info ⓘ

Tags (Optional) ⓘ The resource group where the management node and its dependant resources are going to be created.

Deployment ⓘ

* Resource Groups (Require contributor role) ⓘ

* Identifier
A name prefix to identify all the resources that are going to be created

* VM SSH Username ⓘ
azureuser

* VM SSH Password
The password to login the VM going to be created
 The password should be at least 12 characters in length
 Must contain at least capital letter, one lower case letter, one number and a special character (expect characters " ")

* VM SSH Password Confirm
The password to login the VM going to be created

* VM Type
The type of the VM going to be created

* RDS Type

* Pricing Tier
 General Purpose ⓘ Memory Optimized ⓘ

* Compute Generation
 Gen5 ⓘ Gen4 ⓘ

* vCore
Please select vCore

* Access rules ⓘ
The IP address range in CIDR notation (e.g. 192.168.1.0/24)

< Previous Next >

- Resource Groups: Please select the resource group to deploy Kyligence Cloud services.
- Identifier: Use tags to identify Kyligence Cloud and related resources.
- VM SSH Username: The username to log in to the VM where the Kyligence Cloud server is deployed.
- VM SSH Password: The password to log in to the VM where the Kyligence Cloud server is deployed.
- VM Type: Select the VM type.
- RDS Type: Select the database type and configuration that stores Kyligence Cloud metadata
- Access Rule: Configure the IP range that can access the Kyligence Cloud service. To configure multiple CRIDs, please refer to the [Security group configuration](#) to add the inbound rules for the network security group.

4.(Optional) Click **Next** to create tags for the resources deployed by Kyligence Cloud, click **create tags**.

Bootstrap Kylogence Cloud on Azure China

To install Kylogence Cloud, we will first deploy a management node in your private network. Once done, further installation will take place within your private network to ensure security and privacy. ☺

How much will Kylogence Cloud cost me? ☺

- Authorization ① You can tag resources created by Kylogence Cloud, and later you can categorize resources through tags in the Azure portal.
- Management Node Info ② For more information about tags, please refer to the Azure official documentation.
- Tags (Optional) ③ If you've set a policy for tagging when you deploy resources, please do not add tags with the same name here.
- Deployment ④ If you need to manage the resources you create, please create tags here.

< Previous Deploy

© 2020 Kylogence, Inc. All rights reserved. [Terms of Use](#) | [Privacy Policy](#)
Apache Kylin and associated open source project names are trademarks of the Apache Software Foundation. For a complete list of trademarks, [click here](#).

Bootstrap Kylogence Cloud on Azure China

To install Kylogence Cloud, we will first deploy a management node in your private network. Once done, further installation will take place within your private network to ensure security and privacy. ☺

How much will Kylogence Cloud cost me? ☺

- Authorization ① You can tag resources created by Kylogence Cloud, and later you can categorize resources through tags in the Azure portal.
- Management Node Info ② For more information about tags, please refer to the Azure official documentation.
- Tags (Optional) ③ If you've set a policy for tagging when you deploy resources, please do not add tags with the same name here.
- Deployment ④

Name	Value
<input type="text" value="Please input"/>	<input type="text"/>

< Previous Deploy

© 2020 Kylogence, Inc. All rights reserved. [Terms of Use](#) | [Privacy Policy](#)
Apache Kylin and associated open source project names are trademarks of the Apache Software Foundation. For a complete list of trademarks, [click here](#).

Or skip this step to start the deployment and jump the deployment process page.

Notice: Kylogence Cloud will automatically create the required resources and services in the selected resource group usually it takes 20 mins depends on Network environment.

5.After the deployment is successful, click on the Kylogence Cloud address on the page to launch Kylogence Cloud. The first time you do so, you will be prompted to upload the license file.

If you are using Kyligence Cloud for the first time, you can apply for a trial license on [Kyligence's official website](#) and download it from the **trial welcome email**.

6. Please input the Username and Password to log in. After logging in, we recommend that you go to the Help menu to change this password.

- Username: ADMIN
- Password: KYLIN

Create Workspace

Role: IT Engineer

After Kyligence Cloud is deployed, you need to create a workspace to connect the data source and speed up queries.

Next, please fill in the following information on the Add Workspace page:

The screenshot shows the 'Add Workspace' page in the Kyligence Cloud interface. The page title is 'Add Workspace'. The main content area is titled 'Workspace' and contains the following fields and options:

- Workspace name:** A text input field with the placeholder 'Please Input...'
- Datasource type:** A selection area with four options: 'Azure Data Lake Storage Gen 2' (selected), 'Azure Blob Storage', 'Snowflake', and 'Azure Synapse Analytics'.
- Please input your storage account info:** A section with a radio button selected for 'Please get the related info from "Azure-Storage accounts", please refer [user manual](#) for more details.'
- Storage account name:** A text input field with the placeholder 'Please Input...'
- Storage account key:** A text input field with the placeholder 'Please Input...'
- Query engine SSH:** A text input field with a help icon.

At the bottom right of the form, there are two buttons: 'Cancel' and 'Review + create'.

- Workspace Name: Please enter a workspace name.
- Datasource Type:
 - o You can select **Azure Data Lake Storage Gen 2** and fill in the storage account name and storage account key.
 - o You can select **Azure Blob Storage** and fill in the storage account name and storage account key.
 - o You can select **Snowflake** and fill in the username, password and JDBC connection string, for example:
 - o `jdbc: snowflake: //demo.snowflakecomputing.com/?db=demo&warehouse=demo`

- o You can select **Azure Synapse Analytics** and fill in the username, password and JDBC connection string, for example:

```
o Region is china:  
o jdbc:sqlserver://demo.database.chinacloudapi.cn:1433;database=demo;  
o  
o Region is global:  
o jdbc:sqlserver://demo.database.windows.net:1433;database=demo;
```

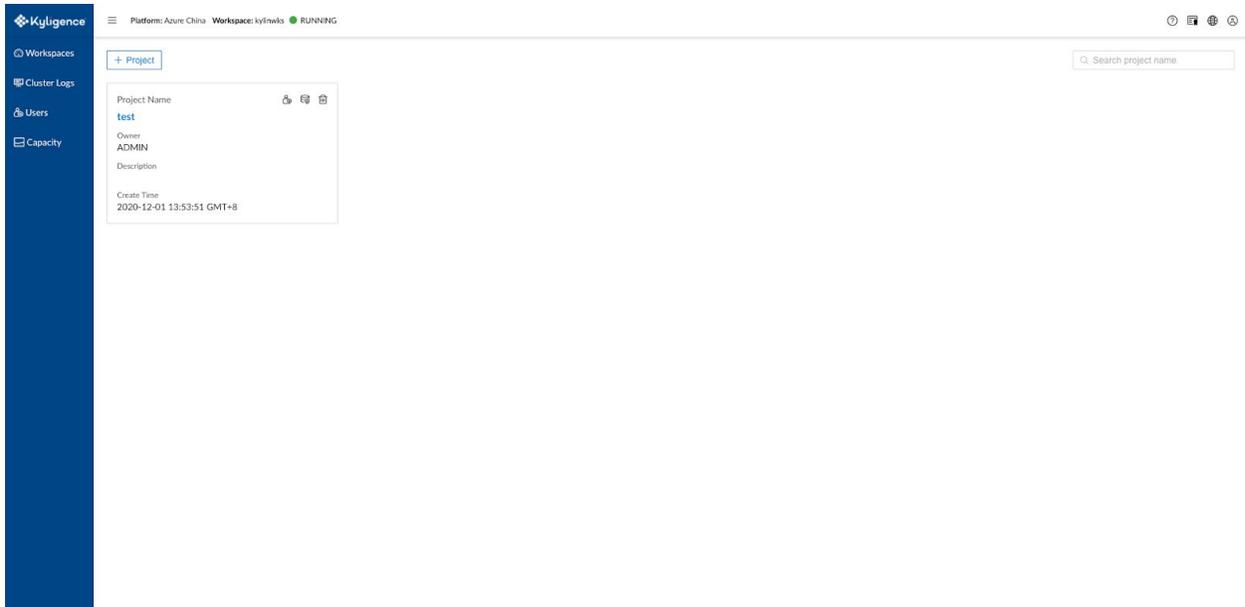
- **SSH Credential:** Please create an SSH key to access your query engine.
- **(Optional) Tags:** You can add tags to the resources created in this workspace, so that your cloud platform administrator can manage the resources on the cloud. The tags added when creating Kyilgence Cloud will be automatically added to the resources in this workspace. You can also add Modify existing tags or add new tags.
- **Cluster Size:** Please enter the volume of data. Kyilgence Cloud will recommend the cluster configuration for you. You can click "Enable Custom Size" to view the cluster details.

After completing the cluster information form, click **Submit and Start**, Kyilgence Cloud will automatically create the clusters. The creation process takes about 5-10 minutes.

Create Project

Role: Data Engineer

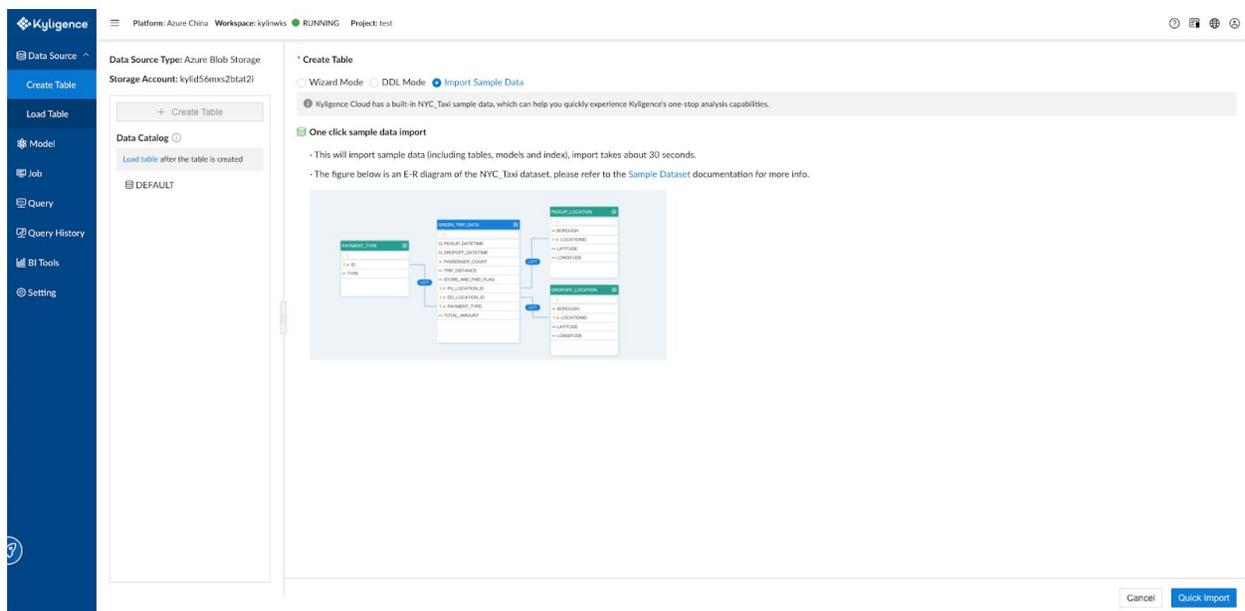
After the workspace is created, you need to create a project in the workspace. You can create multiple projects in one workspace to serve different business areas by sharing one cluster resource.



Import Sample Data

This feature currently only supports Azure Data Lake Storage Gen 2 and Azure Blob Storage on the Azure platform. If you use Azure Synapse Analytics or Snowflake as your data source, please refer to the [Datasource documentation](#) to add the data source.

Kyligence Cloud has a built-in NYC_Taxi dataset that contains travel data for the green taxi in January 2019. We will use this dataset for a demonstration. Visit the **Datasource-Create Table** page, click **Create Table**, select **Import sample data** and click on the **Quick import** to import data. For a description of the sample model and a data dictionary, refer to the [NYC Taxi 数据集](#) documentation.



When the prompt "Import finished" appears, please go to the model page. From there, you can see the imported sample model named **nyc_taxi_green_trip**.

If you need to use the sample table for modeling, please refer to the [model chapter](#) to create the model.

Then you need to load data and build the index for the model before you query. In this example, we choose to load the data in full. Visit the **Model** page, click the small arrow on the left of the model name to open the model details page and click **Aggregate Index-Build Index** below the model to build the data index.

The screenshot shows the Kylogence interface with the 'Model List' page. The selected model is 'nyc_taxi_green_trip'. Below the table, the 'ER Diagram' tab is active, displaying a star schema where 'GREEN_TRIP_DATA' is the central Fact Table, and 'PICKUP_LOCATION', 'DROPOFF_LOCATION', and 'PAYMENT_TYPE' are associated Lookup Tables. The relationships are labeled as 'LEFT'.

Model Name	Index Amount	Storage	Expansion Rate	Usage	Owner	Actions
nyc_taxi_green_trip	31	0 B	0%	0	ADMIN	Build Index

It takes about 5-8 minutes to build the data index. You can query the progress on the **Job** page. Once the build is completed, you can use the dataset for analysis.

The screenshot shows the 'Jobs List' page in Kylogence. A job titled 'Build Index' is shown for the target subject 'nyc_taxi_green_trip'. The job status is 0%, indicating it is still in progress. The start time is 2020-12-01 14:17:16 GMT+8 and the duration is 0.12 mins.

Job Type	Target Subject	Data Range	Job Status	Start Time	Duration	Actions
Build Index	nyc_taxi_green_trip	Full Load	0%	2020-12-01 14:17:16 GMT+8	0.12 mins	Refresh

Visit the **Query** page and query the data using SQL.

The screenshot shows the Kylogence SQL Editor interface. The top navigation bar includes 'Platform: Azure China', 'Workspace: kylinwks', 'RUNNING', and 'Project: test'. The left sidebar contains navigation options like 'Data Source', 'Create Table', 'Load Table', 'Model', 'Job', 'Query', 'Query History', 'BI Tools', and 'Setting'. The main area displays a SQL query: `select * from GREEN_TRIP_DATA`. Below the query editor, the 'Query Information' section shows 'Query ID: 0c642393-1445-4cc2-8d99-63877226a2e9', 'Duration: 3.13s', and 'Answered By: Object Storage'. The 'Query Results' section shows a table with 6 columns: pickup_datetime, dropoff_datetime, passenger_count, trip_distance, store_and_fwd_flag, and pu_location. The table contains 10 rows of data.

pickup_datetime	dropoff_datetime	passenger_count	trip_distance	store_and_fwd_flag	pu_location
2018-12-21 15:17:29	2018-12-21 15:18:57	5	0.0	N	244
2019-01-01 00:10:16	2019-01-01 00:16:32	2	0.86	N	97
2019-01-01 00:27:11	2019-01-01 00:31:38	2	0.66	N	49
2019-01-01 00:46:20	2019-01-01 01:04:54	2	2.68	N	189
2019-01-01 00:19:06	2019-01-01 00:39:43	1	4.53	N	82
2019-01-01 00:12:35	2019-01-01 00:19:09	1	1.05	N	49
2019-01-01 00:47:55	2019-01-01 01:00:01	1	3.77	N	255
2019-01-01 00:12:47	2019-01-01 00:30:50	1	4.1	N	76
2019-01-01 00:16:23	2019-01-01 00:39:46	1	7.75	N	25

Visualization

Role: Analyst

Kylogence Insight is an out-of-the-box visualization tool in Kylogence Cloud. Use this tool for the demonstration.

The screenshot shows the Kylogence BI Tools page. The top navigation bar is the same as in the previous screenshot. The left sidebar is also the same. The main area displays a row of six icons for different visualization tools: Kylogence Insight (Official), Kylogence MDX (Official), Tableau, Power BI, Qlik, and Excel. Each icon has a small 'Install' button below it.

Install and Launch Kylogence Insight

On the **BI Tools** page, click "Install and start" under the built-in Kylogence Insight icon and wait for the installation to enable Kylogence Insight.

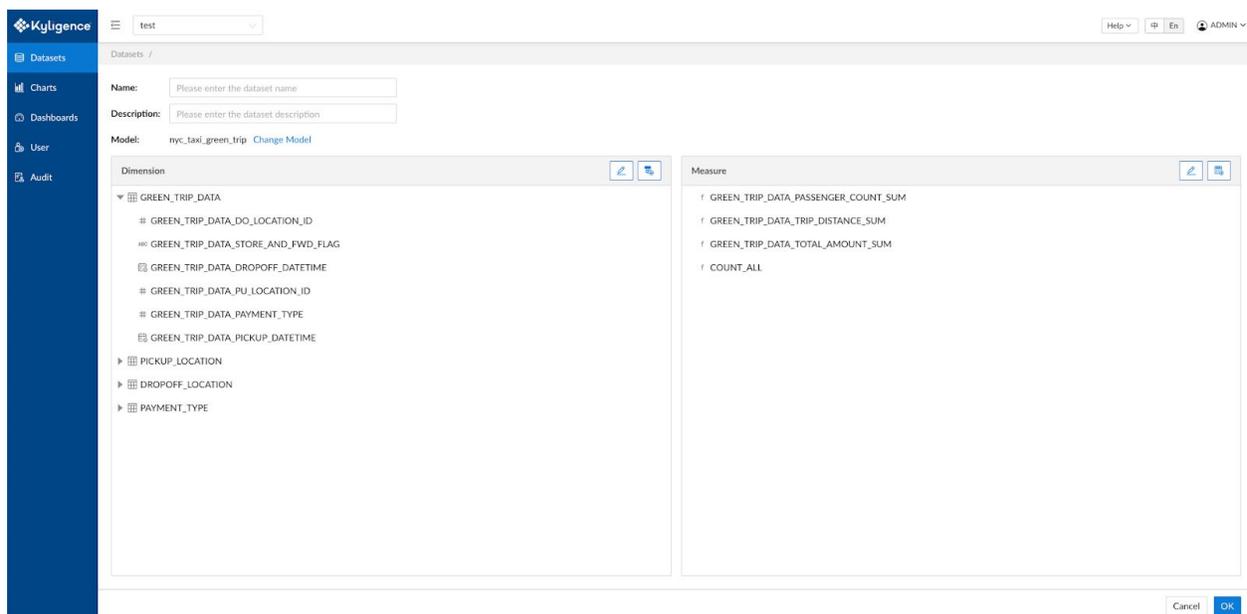
Create Dataset

1. Select the sample project and navigate to the **Dataset** page. Then click the **+Dataset** button in the top left corner and set the **Dataset Usage** to **SQL**.
2. Define Dataset: First, name the dataset *nyc_taxi_green_trip* in **Basic Information** and click Next.

Drag the desired model to the right in **Define Relationships** and click Next.

1. Define Semantics: In **Defining Semantics**, you can define as follows:
2. Click the edit button on the right of the dimension to change the dimension name, for example, rename *PICKUP_DATETIME* in the *GREEN_TRIP_DATA* table to "PICKUP_TIME".
3. Click the edit button on the right of the measure to change the measure name.
4. Click the **+Add Hierarchy** button to create a hierarchy.
5. Click the **+Calculated Measure** button and enter **Calculated Measure Name** and **Expression** to create a calculated measure.

Once all of the definitions have been completed, click the **Submit** button to save the dataset.



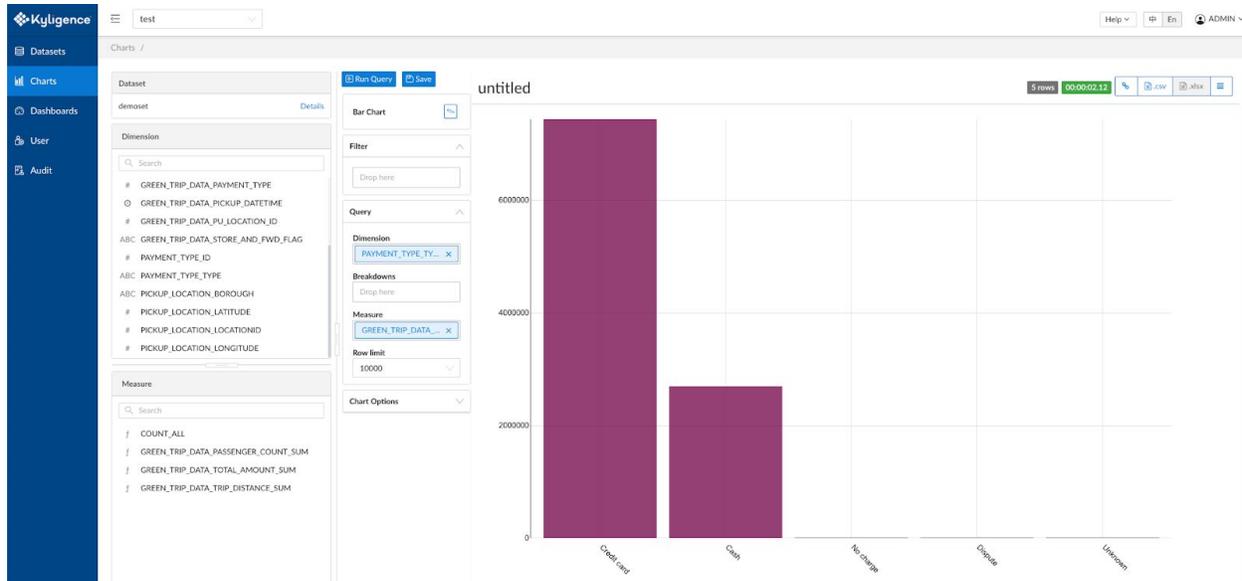
Create Chart

Click **Chart** from the navigation bar, then click the **+Chart** button, select the new *Nyc_Taxi* dataset, and click **Submit** to start visualization.

Drag dimensions and measures to the right panel, then click **Run Query** to execute the query and render the chart.

Take this as an example, first select the **Click to change visualization type** button on the page and select the visualization type as **Bar Chart**. Drag "PAYMENT_TYPE_TYPE" as the dimension, drag "GREEN_TRIP_DATA_TOTAL_AMOUNT_SUM" as the measure, then you can get the total order amount of each payment type.

Once you have the chart, you can save your chart by clicking the **Save** button on the page, or you can download the query result set by clicking the **Export CSV** button.



Uninstall

Role: IT Engineer

You can save money by release resources on your cloud in the following ways, depending on your needs.

- Option 1 - Stop workspace: If you need to continue to use this workspace's data and services later, you can use the workspace list page, stop the workspace after the workspace is stopped, the compute resources deployed in this workspace will be deleted, but the data storage service will remain, your reports, models, indexes in this work will be retained, but the workspace will not be able to provide query and build services in the stop state, you can start this workspace at any time when you need to.
- Option 2 - Delete workspace: If you do not need to use this workspace you can select the Delete workspace, after deletion, the calculation and storage resources for this workspace will be deleted and the data will not be recoverable after deletion.

