Data obfuscation in Data Factory with Delphix Compliance Services

Article • 05/23/2023

The following how-to outlines the use of Delphix Compliance Services (DCS) in Data Factory in Microsoft Fabric dataflows to mask sensitive data prior to delivery.

> **Important**
>
> Microsoft Fabric is currently in PREVIEW. This information relates to a prerelease product that may be substantially modified before it's released. Microsoft makes no warranties, expressed or implied, with respect to the information provided here. Refer to **Azure Data Factory documentation** for the service in Azure.

DCS is a highly scalable masking API service that automatically masks personally identifiable information (PII), supplanting manual processes for delivering compliant data. Its out-of-the-box and configurable algorithms replace sensitive data values with fictitious yet realistic ones, so teams mitigate risk while ensuring end-users can easily consume the right data.

Masked data maintains multicloud referential integrity, is production-like in quality, and remains fully functional for accurate analysis or testing. Note that a DCS account needs to be created prior to use, and you can sign up for a free trial .

What is the challenge?

The cloud is filled with personally identifiable information (PII), fueling privacy and security risk. PII from production apps needs to flow to downstream systems for analytics, exposing organizations to risks or creating data silos. Power Query and DCS automate data compliance and security to unblock data movement.

Breaking down data silos is difficult:

- Data must be manipulated to fit a common format. ETL pipelines must be adapted to each system of record and must scale to support the massive data sets of modern enterprises.

- Compliance with regulations regarding sensitive information must be maintained when data is moved from systems of record. Customer content and other sensitive elements must be obscured without impacting the business value of the data set.

How do DCS and Data Factory solve automating compliant data?

The movement of secure data is a challenge for all organizations. Delphix makes achieving consistent data compliance easy, while Data Factory enables connecting and moving data seamlessly. Together Delphix and Data Factory make the delivery of on-demand, compliant data easy.

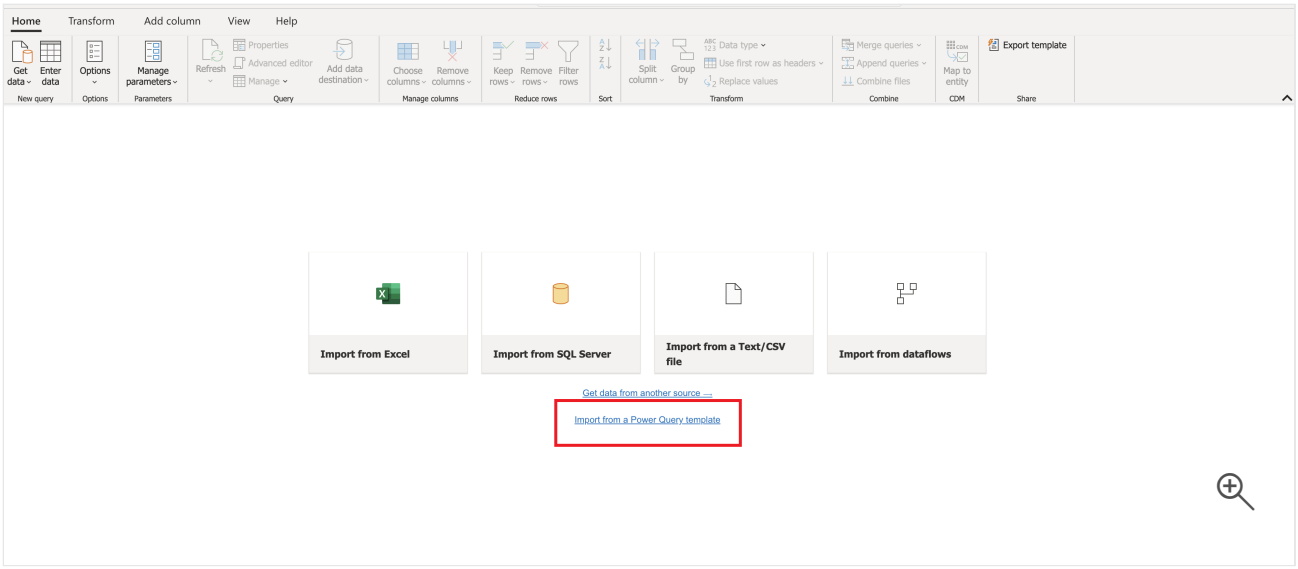
Using Data Factory data flows, you can create a workflow that automates the following steps:

- Read data from the desired source.
- Map sensitive fields to appropriate masking algorithms (and manage as a central configuration table).
- Call DCS masking APIs to replace sensitive data elements with similar but fictitious values.
- Load the compliant data to a desired target.

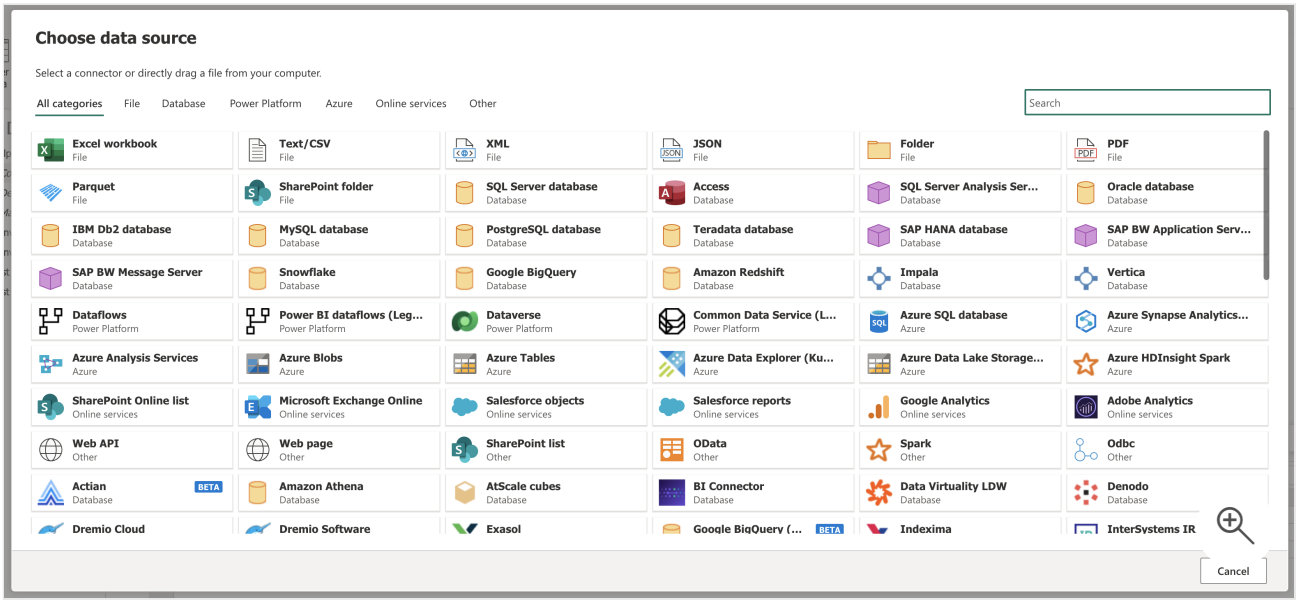
How to get started

Go to the Delphix free preview page to request a free trial of DCS. The Delphix team then contacts you for access and provides the template that is used in the example setup scenario described in this article.

In Power Query, upload the provided template by selecting **Import from a Power Query Template**, and then select the Power Query template file to import. This selection loads a set of queries.

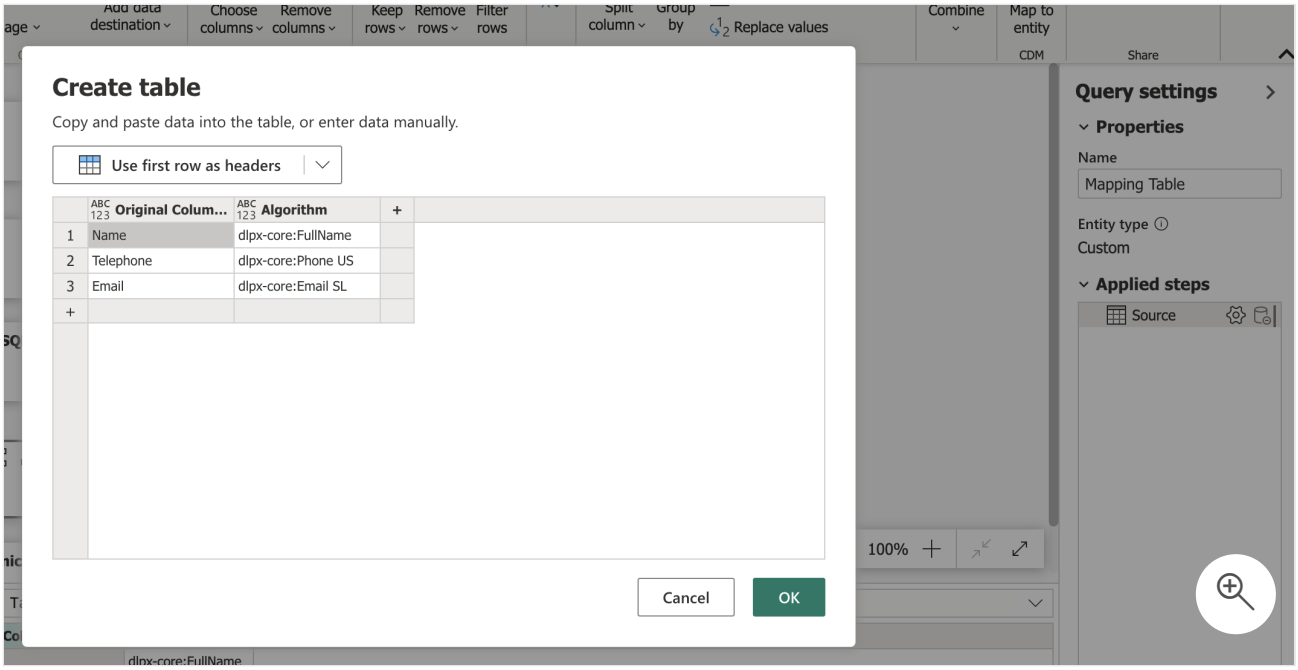


Import the data source that contains sensitive data that you would like masked.

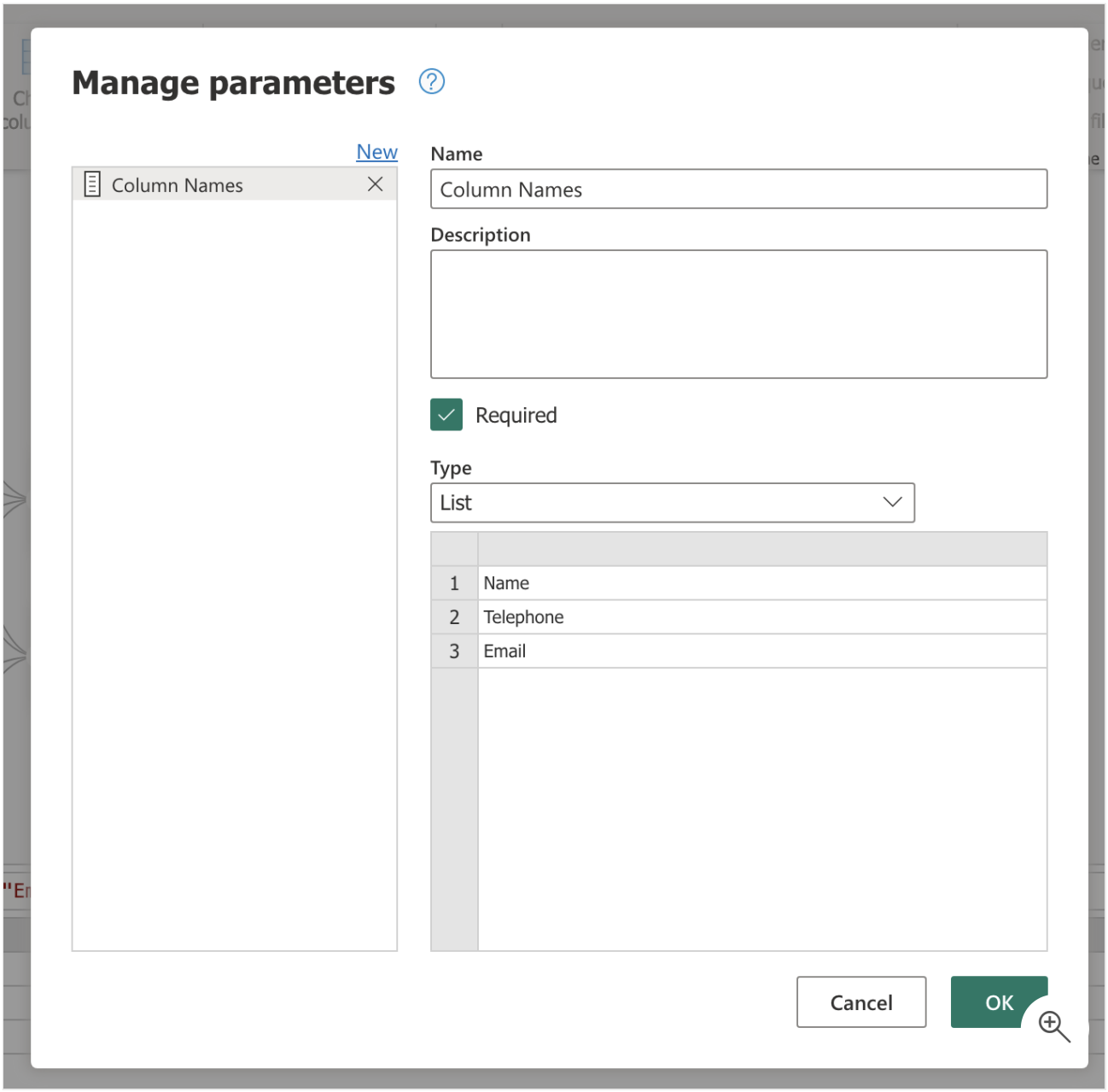


The mapping table is where you configure what fields to mask, and which Delphix masking algorithms to use. In the **Query Settings** pane, right-click on the gear icon. Enter the column names where sensitive data resides in **Original Column**. Enter the corresponding Delphix algorithm in **Algorithm**. Details on available algorithms can be found in the Delphix documentation .

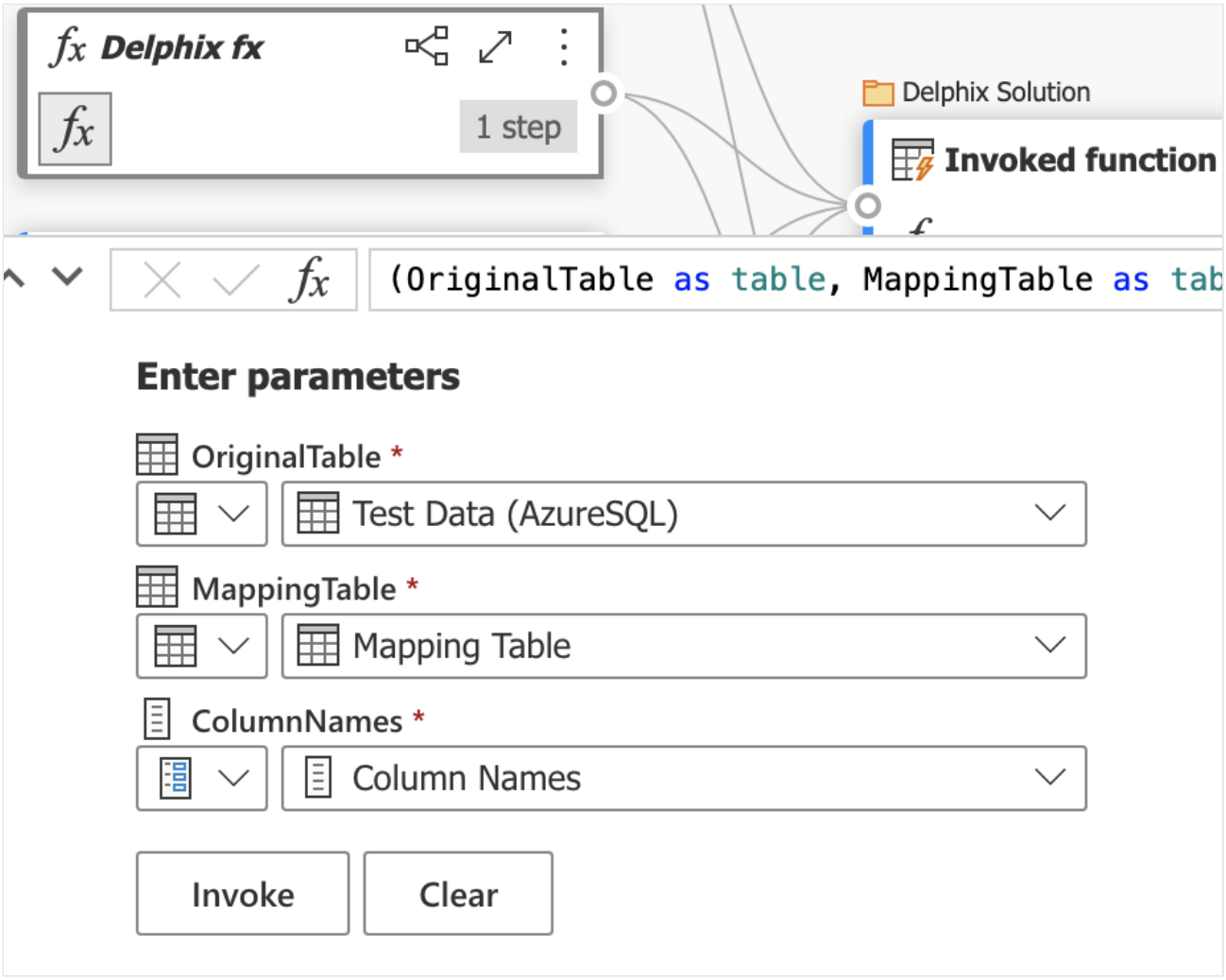
Delphix's out-of-the-box masking algorithms can be customized, or new algorithms can be defined if needed. All Delphix masking algorithms replace sensitive data with fictitious, yet realistic values, and do so consistently across data sets.



This mapping table can be a global configuration across any tables you would like to mask. Should you want to leave any columns unmasked for a given table, the **Column Names** parameter serves as a filter. Copy and paste the list of **Original Columns** (from the mapping table) into the **Column Names** parameter, and delete any column names that you would like to leave unmasked.



You're now ready to mask your data. Select **Delphix fx** and enter parameters as displayed in the following image (with the `OriginalTable` field as the data source that contains sensitive data).



Once this change is complete, select **Invoke** to run the data flow. This selection automatically calls the DCS masking API service to mask data prior to delivery to the destination of your choice.

Your data is now ready to be used safely by end users. The data is masked consistently, ensuring that references remain intact across data sets. As an example, George Smith becomes Edward Robinson regardless of data source or destination, ensuring it's still valuable for integrated analytics scenarios.

Next steps

- Delphix free preview page
- Delphix documentation