Technical preview of masking rules in IBM® Cloud Private for Data

Contents

1. Overview

If you have confidential data in your database tables or data files, you can mask that data so that it still can be analyzed, but the real content is not visible to users.

Data catalog might contain PII (personally identifiable information) data such as credit card number, salary information, age, phone number, email address, and so on. You want to analyze such data, but at the same time you want to protect the privacy of individuals to which this data refers. Therefore, you can mask such data, which means that the real values are replaced with fake values.

Data is masked based on the `if … then` rules, which you create by dragging block elements. If the conditions are met, specified action is taken. For example, if a column Phone has the term Employee Phone Number assigned, and it has label PII assigned, then the values of this column are masked. When data is masked, fake values are used instead of the real values.

Note: This is a technology preview and is not yet supported for use in production environments.

2. Enabling masking rules

Before you can start using the masking feature, you must complete the following steps:

1. Disable the `com.ibm.iis.ug.masking.disabled` flag.
    a. Go to the location of `iisAdmin` tool.
        i. Linux/UNIX: `IS_install_path/ASBServer/bin/iisAdmin.sh`
        ii. Windows: `IS_install_path\ASBServer\bin\iisAdmin.bat`
    b. Run the following command:
        i. Linux/UNIX: `./iisAdmin.sh -unset -key com.ibm.iis.ug.masking.disabled`
        ii. Windows: `iisAdmin.bat -unset -key com.ibm.iis.ug.masking.disabled`
2. Enable the `featureMaskingRules` flag by accessing the following URL: `https://<IP_address>:<port>/ibm/iis/igcui/discover?featureMaskingRules=on`

3. Masking rules structure

Masking rules consist of the following types of elements:

- logic
- conditions
- actions

Logic

The basic logic is `if condition then action`. You can expand this logic by adding more conditions and joining them by using `and`, `or`, and `not` operators. For example, `if condition 1 and not condition 2 or not condition 3 then action` as in "If the asset has term Age assigned and does not have the label Child assigned or does not have the label Adolescent assigned then mask data without preserving format".

Conditions

You can use the following conditions in your data masking rules:

- `asset has the term term_name assigned` - the rule is applied only when the specified term is assigned to an asset that you want to mask. The term must be assigned to a database column or a data file field. You cannot apply masking rules for entire database tables or data files.
- `asset has the label label_name assigned` - the rule is applied only when the specified label is assigned to an asset that you want to mask. The label must be assigned to a database column or a data file field. You cannot apply masking rules for entire database tables or data files.
- `user has the role role_name` - the rule is applied only when the user who previews data has the specified role.
- `user is in the user group group_name` – the rule is applied only when the user who previews data is in the specified user group.

Actions

You can use the following actions in your data masking rules:

- `mask data and preserve the format of data_format` - this action hides the real values of specified assets, but it preserves their data format. For example, when you want to hide an email address, the new value is a random email address. You can still use such masked data for analysis, or as a joining property.
  You can preserve the following data formats:
  - `email` – for example, the original value `john.smith@gmail.com` might become `GqlPKljuIdYa4e4q8t9HENU5@UF`.
  - `SSN` – for example, the original value `111-33-2222` might become `422-67-8164`.
  - `Hashing` – for example, the original value `The first of his name` might become `NPD 9_RHW 4K ROB BL7H`.
- `mask data without preserving the format` - this action hides the real values of specified assets, and the new values are hashed. For example, when you want to hide a phone number, the new value might contain letters and digits.

You can still use such masked data for analysis, or as a joining property. A real value is always masked with the same random value.

- `deny access to data` – this action blocks access to data if the specified conditions are met. You can for example block access for a specific user group. As a result, the values of the affected columns are blurred, and the columns are marked with the access denied icon. You can't use such columns for analysis or as joining properties.

4. Creating masking rules

1. Go to **Organize** > **Data catalog**.
2. From the **Create** menu, select **Masking rule**.
3. Add the name and description for the rule.
4. Select a condition, and connect it to the **if** block. You can select many conditions and join them by using operators from the **Logic** elements.
5. Select the action, and connect it to the **then** block.
   Important: You can add only one action in the rule.
6. Optional: You can complete the following additional tasks on the blocks when you right-click the block:
   - Duplicate block
   - Add comment
   - Select inline inputs, which moves blocks inside another block if applicable
   - Collapse block
   - Disable block
   - Delete block
7. Save the changes.

5. Viewing masked data

After you create masking rules, you can verify whether data is masked. The values are masked only in the data preview of database tables and data files.

Complete the following steps to preview data:

1. Search for the database table or a data file that is affected by a masking rule.
2. From the menu in the asset details page or in the search results list, select **Preview**.

Note: When data masking is enabled, you don’t need to provide credentials for the data source that was used to connect to the asset.

The column with the masked data is marked with an icon, depending on what actions are applied:

- If one of the masking actions is applied, the column is marked with the lock icon, and the replaced data is displayed.
- If the deny access action is applied, the column is marked with the access denied icon, and the values are blurred.

6. Accessing masking rules

If you already have masking rules, you can search for them like any other asset type, either in **Organize** > **Data catalog** in Data Privacy asset group, or from the toolbar.

7. Example scenario

You have a table where you store customer data. This data contains customer email addresses. You want to mask these email addresses. Complete the following steps:

1. Create a term `Customer Email Address` in **Organize** > **Business glossary** > **Terms** > **Create Term**, or in **Organize** > **Data catalog** > **Create** > **Term**.
2. Make sure that you have a label to mark sensitive data. You can add it to the catalog by importing glossary assets. Let's assume that you have two labels to mark sensitive data, `PII` and `GDPR`.
3. You want to mask data in database column `Email`. Search for it from the toolbar, or in **Organize** > **Data catalog**.
4. Open its details page. From the menu, click **Edit**.
    a. In the **Labels** field, find one of your labels and add it to the list.
    b. In the **Assigned to Terms** field, find your newly created term and add it to the list.
    c. Save the changes.
5. Create a masking rule. Go to **Organize** > **Data catalog** > **Create** > **Masking rule**.
6. Add name `Mask email address` and provide description `This rule masks email addresses of customers`.
7. Specify the rule logic.
    a. Your rule must have more than one condition. Click **Logic**, select the **and** element, and drag it to the **if** element.
    b. Click **Conditions**, and select **the asset has the term assigned**. Drag it to the **and** element, and find the `Customer Email Address` term. Save it.

c. Click **Logic**, select the **or** element, and drag it to the **and** element.
d. Click **Conditions**, and select **the asset has the label assigned**. Drag it to the **or** element, and find the `PII` label. Save it.
e. Click **Conditions**, and select **the asset has the label assigned**. Drag it to the **or** element, and find the `GDPR` label. Save it.
f. Click **Actions**, select **mask data and preserve the format of email**, and drag it to the **then** element.
g. Save the rule. This rule masks data for an asset that has the `Customer Email Address` assigned, and the `PII` or `GDPR` label assigned. As a result, the example value `john.smith@ibm.com` might become `Khf9jFR8bnyk9@hsd.com`.

8. Preview the data to check whether it's masked. Search for the database table that contains the `Email` column and preview data for this database table. The column `Email` contains masked values instead of real values, and it's marked with the lock icon.