Collibra

Collibra Data Intelligence Cloud

Collibra Protect

Version 2024.01

Collibra

Collibra Data Intelligence CloudCollibra Data Governance Center - Collibra Protect

Release date: Tuesday, January 2, 2024

Revision date: January 02, 2024

You can find the most up-to-date technical documentation on our Documentation Center at

https://productresources.collibra.com/docs/collibra/latest/#cshid=protect

Contents

Scenarios for using Protect

This topic describes how Collibra Protect helps you to:

- Use the metamodel graph to establish and enforce protection policies on Business Processes, Data Categories, and Data Sets.
- Apply a range of protection mechanisms to data sources using classifications.
- Support privacy preferences, such as consent management, data subject access requests, and the right to be forgotten, via row-filtering mechanisms.
- Conduct an audit of relevant protection at data sources and use reporting to demonstrate compliance in data storage and consumption.

Discover and classify personal information

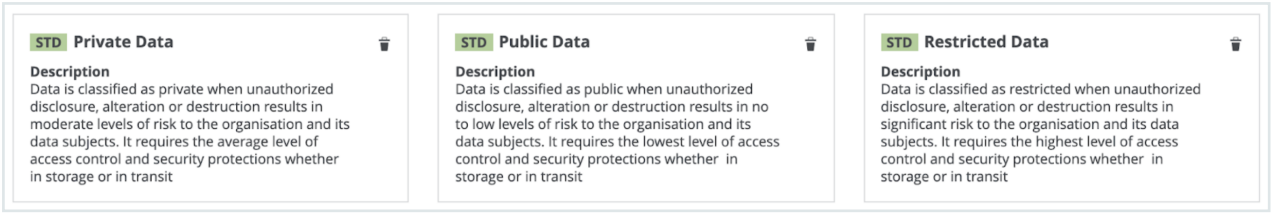
Suppose that you want to help your organization find personal information.

To achieve this, typically, your Privacy team sets up the Data Classification Policy, where they classify the data used in the organization based on the sensitivity or the business criticality of the data. This determines the required levels of security for the applications that store that data or the applications that are used for the transit of the data.

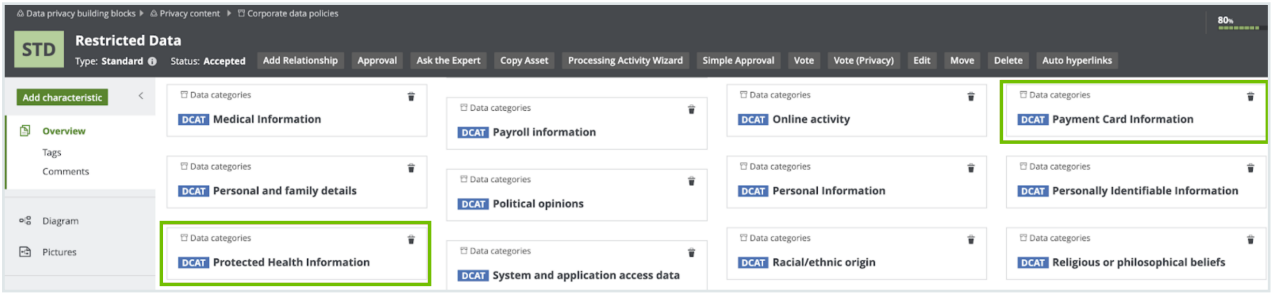
Consider the following three classifications for sensitivity:

- Public data, which is least sensitive.
- Private data, which is slightly more sensitive than the public data.
- Restricted data, which is the most sensitive data and therefore requires the highest level of access controls and security protection.

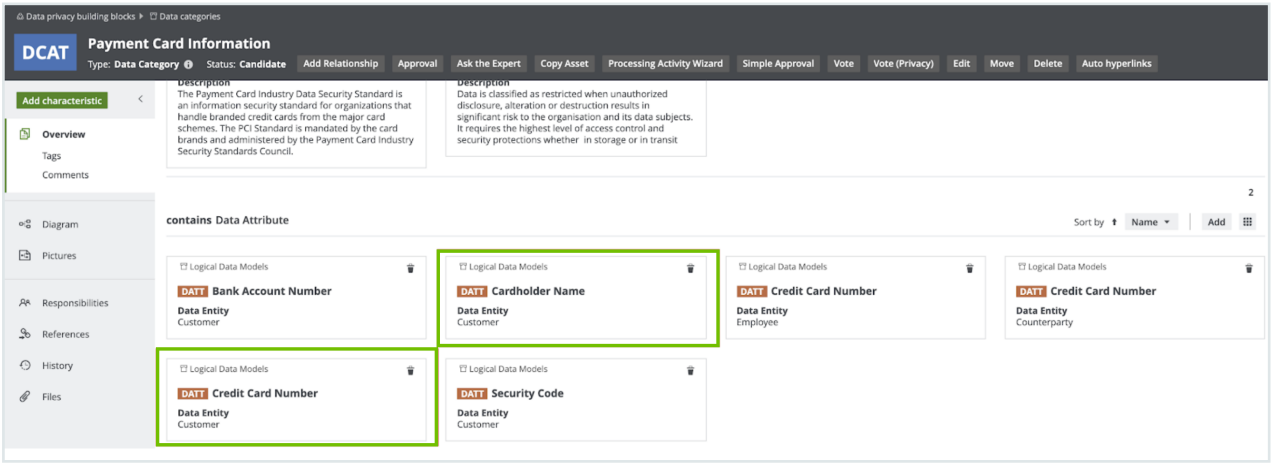
The following image shows the standard subassets of the Data Classification policy.



The Privacy team determines the data categories to which these subassets apply. For example, they can determine that Restricted Data applies to the following data categories: Gender, Social Security Number, Payment Card Information.

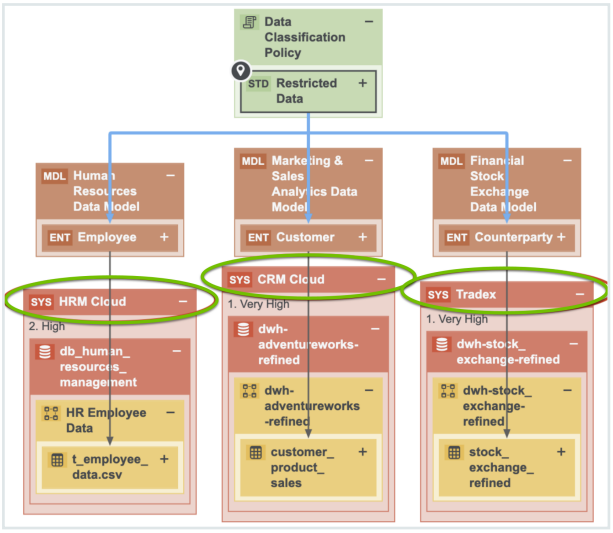


The Privacy team determines the sensitivity and the required security at the data category level as opposed to the column level. At the data category level, the Privacy team then determines what data elements belong to the identified data categories. For example, the Payment Card Information data category groups the Cardholder Name and the Credit Card Number, among other information.



In this model, Data Attributes are grouped under the Data Category. This is how the Privacy layer is linked to the logical data model. This promotes collaboration between the Privacy team and the Governance team. In addition, this allows the automated data classification of the organization's personal information, which makes views such as the

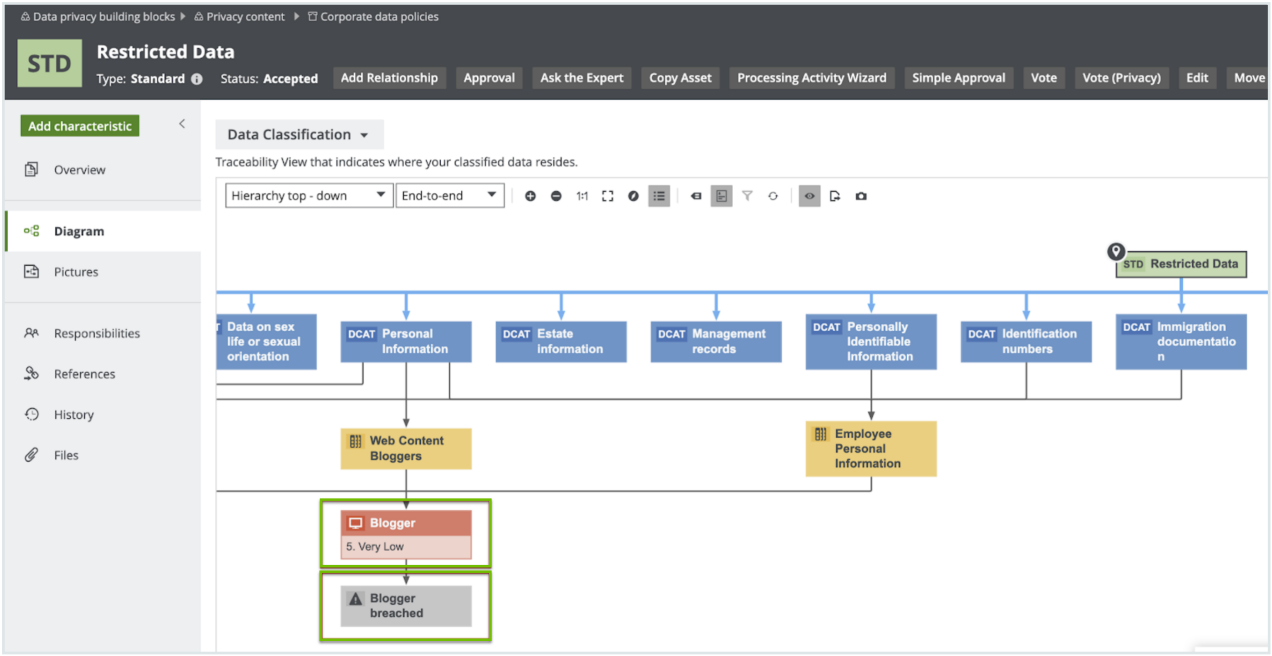
Restricted Data Overview diagram, available at the most sensitive data category, Standard Restricted Data.



In the above image, the applications in which the restricted data resides are highlighted.

The Privacy team determines the policies and standards that determine which data categories are sensitive to the organization and what the required levels of protection are. The Data Governance team maps those data categories to the applications where that data resides. The Security team determines what the security levels on those applications are. Thus, the view captured in the above image requires collaboration among teams.

Consider the traceability diagram called Data Classification under the Restricted Data standard. This standard contains the most sensitive information and thus requires the highest level of security controls; however, it resides on an application that has very low security. Because of this, the Information Security team needs to take the necessary remediation actions and improve the security levels on Blogger. As shown in the following image, an investigation is already ongoing on the potential data breach on Blogger.

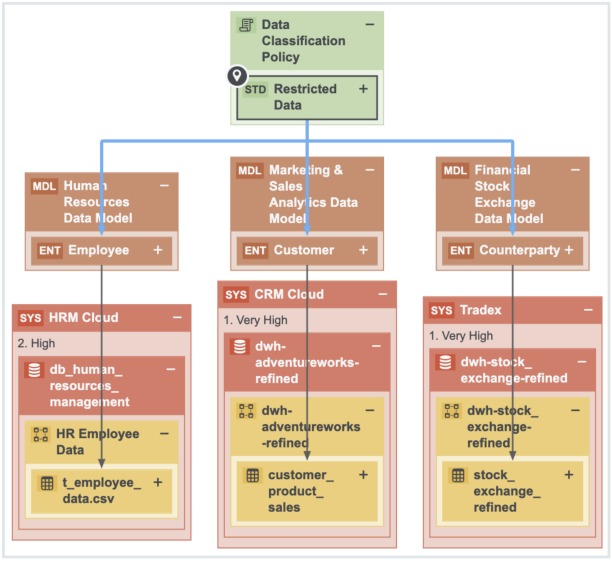


Data classification capabilities and guided stewardship

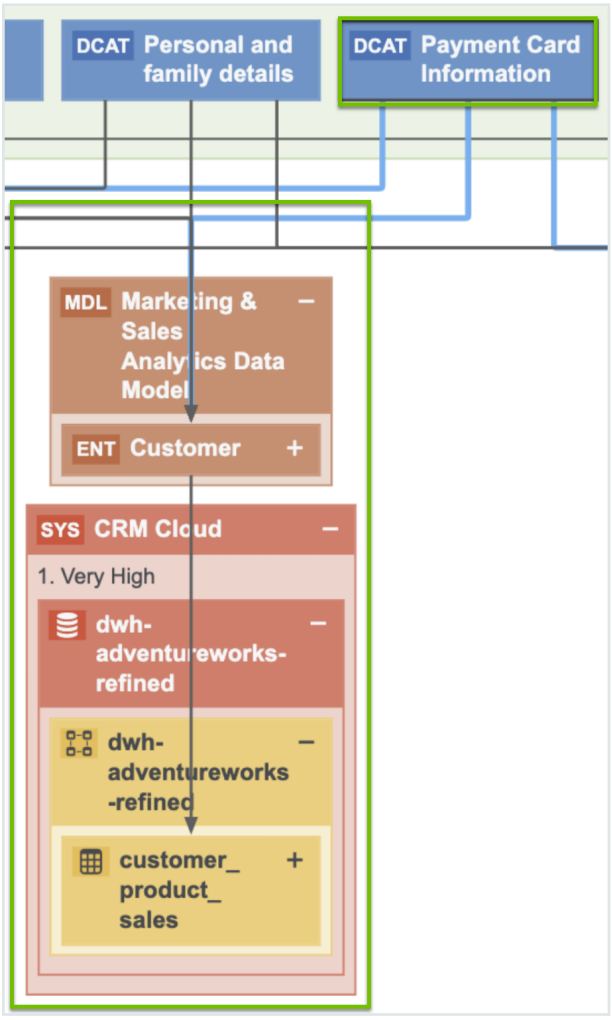
This section describes how Collibra Data Privacy leverages the data classification capabilities in Catalog. Thus far, we learned that the Restricted Data standard groups data categories, which group data attributes. In the example, the Payment Card Information data category contains the Credit Card Number data attribute.

Guided stewardship is a semi-automated process of mapping columns and tables to logical data attributes. It enables content tables to be mapped to data attributes. After scanning a table and then applying guided stewardship in which the Steward selects attributes from the suggestions coming from the automated mapping, the column is mapped to the Credit Card Number. Moreover, when a column is mapped to a data attribute, the column is also mapped to a data category because of the relation between the data category and the data attribute.

The result of classifying one application with the Catalog's Data Classification is shown in the following image.

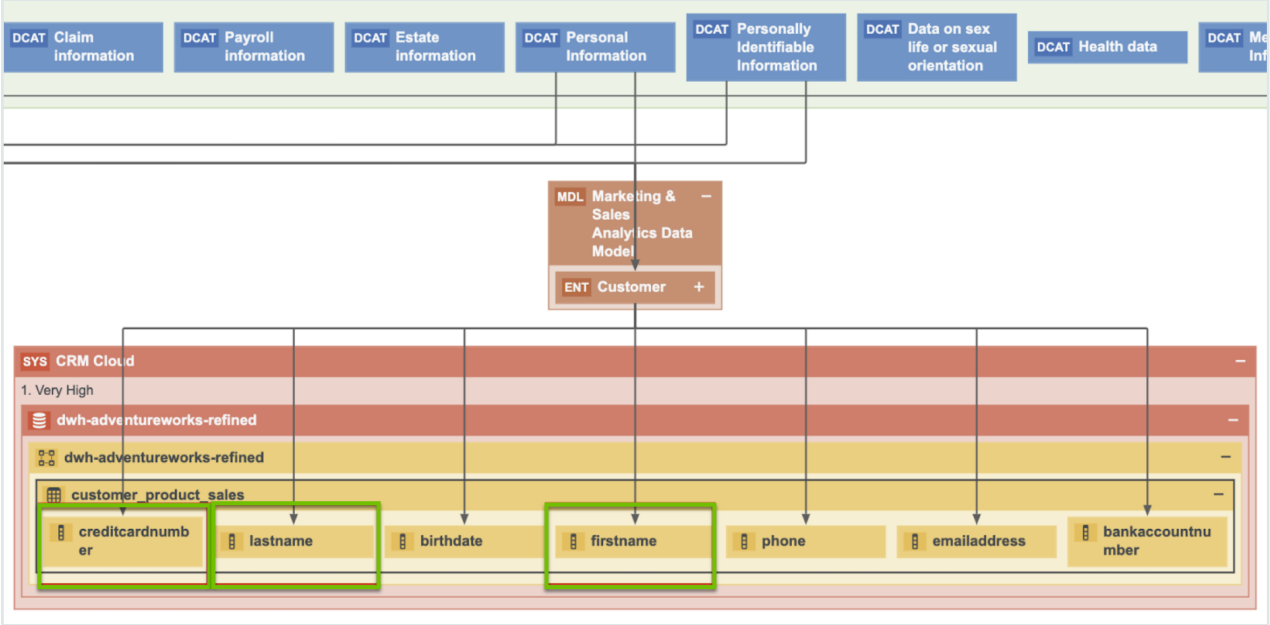


Restricted Data groups multiple data categories. The following image shows the data attributes that the Payment Card Information data category groups.



By applying guided stewardship and data classification, the data attributes are mapped to the columns. Thus, by using Catalog's data classification capabilities, the Data Governance team can find personal information and sensitive personal information.

It is important to know the context to determine which information is considered personal information. For example, Name can be the name of a customer or an employee, in which case Name is considered personal information. Name can also be the name of another organization. This context can be provided only by a Steward. Therefore, data classification and guided stewardship will help the Steward mapping customer's names to the Name column. Because the Privacy team has mapped names and family details, you can safely assume that this is Personal Information. Similarly, Credit Card Number can be the credit card number of another organization, but it is the Steward who has mapped the number to the Credit Card Number data attribute belonging to the Customer data entity, and as a result, we know that the payment card information is very restricted data.



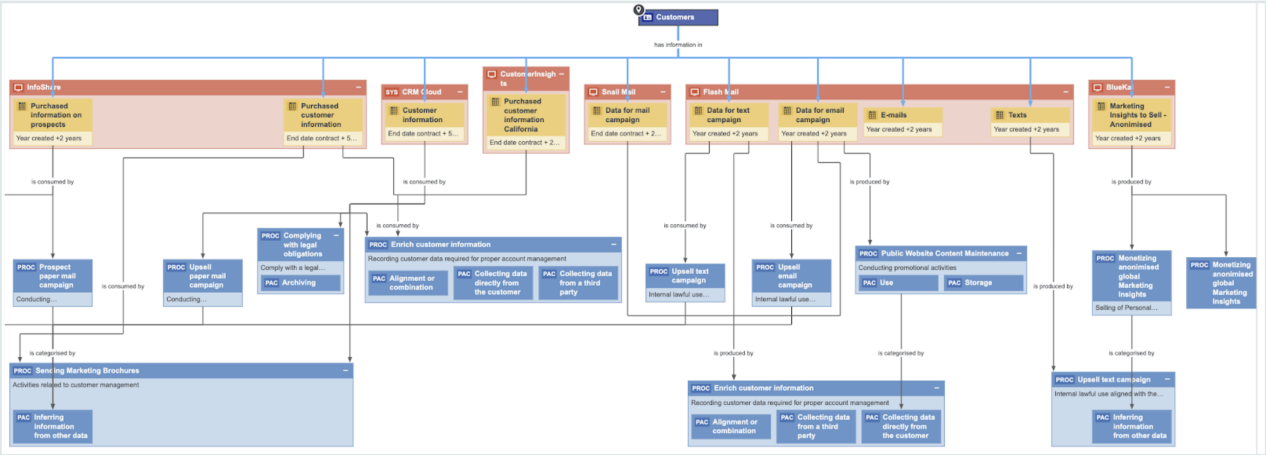
This is an example of how guided stewardship, Catalog's data classification combined with guided stewardship and Data Privacy, gives you a vertical view on where Personal Information resides.

Customer requests and consent management

The previous sections described how we help customers find their Personal Information across applications. This section describes how we help customers manage data subject requests and consent. Collibra has the relevant metadata that is necessary for a partner application that fulfils the data subject requests or manages consent to operate. These applications need the metadata about where the data resides, where you store customer information, how you use the information, why you use the information, and what your legal basis is, so that they can determine for which applications you need consent and for which processes you need instance for a consent. Collibra has and governs the required metadata. In addition, through APIs, Collibra can integrate with those applications to feed them with the metadata that they need to function.

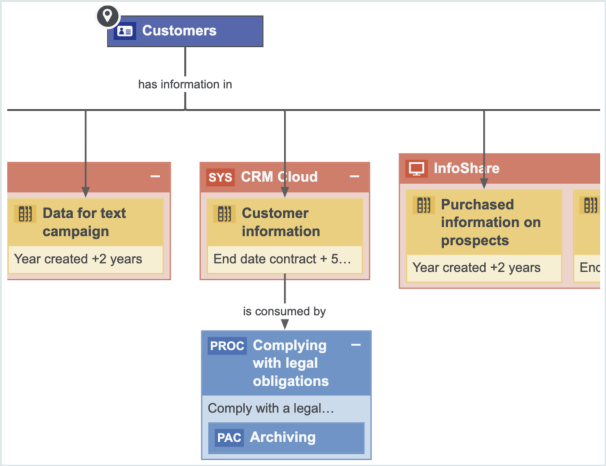
Consider the customer data. Collibra knows where this data resides and how it is being used. This is an outcome of obtaining input from the business users during the onboarding of the Business Processes where users are asked what data they use, which applications they use, and for what purpose they use the data. When further onboarding of those business processes by the Stewards takes place, one of these steps is mapping the Business Processes to the data, and then also helping those Business Stewards with the mapping through the data classification capabilities in Catalog.

The following image shows a traceability view, which is a result of collaboration with the Business team, Data Governance team, and other teams.



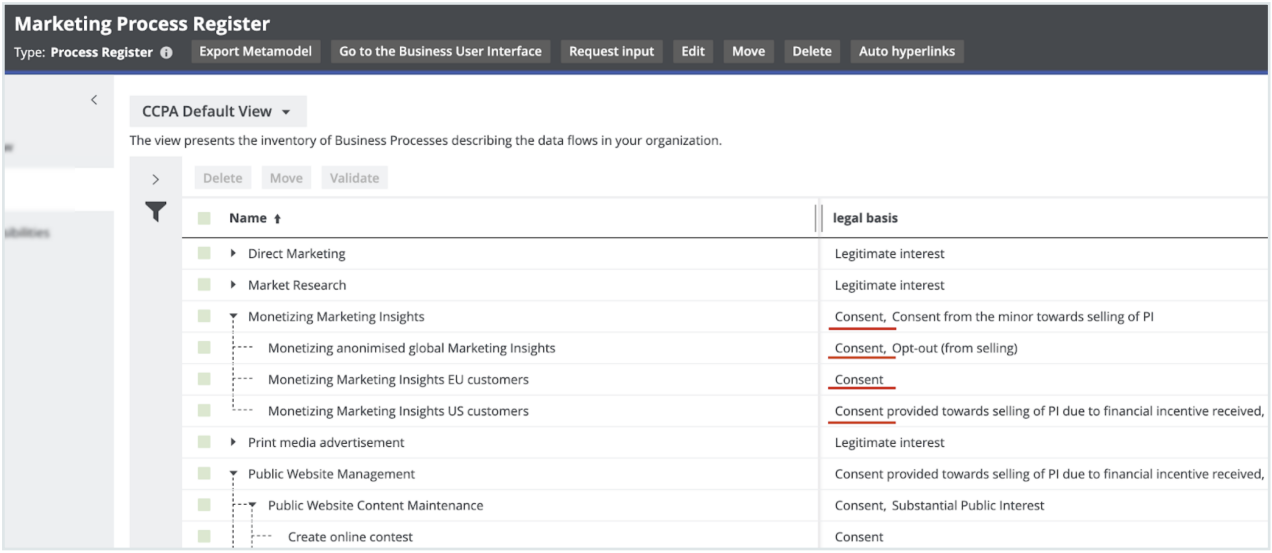
The above image shows where data resides and why it is used. It shows all the applications that contain customer data, and also the related retention periods, which can

be imported when a customer wants to exercise their right to be forgotten. Collibra knows in which applications the data resides and the business processes that use that data. Thus, we know why and how we are using our customer data. This determines how to respond to the right to be forgotten because there are often Business Processes where you have the real legitimate reason to retain the customer's personal information.



When a customer wants to exercise their right to be forgotten, we can remove the information in these applications; however, we need to store the customer information in the above table in order to comply with the legal obligation. Therefore, it is not only important to know where your personal information resides, but also why you are using it. Such information is important information for applications that process data subject requests (DSRs). You can integrate with the application that does the DSRs and create a workflow to process DSRs. Based on the input of the information and metadata that you will find in Collibra, you can validate the request. When the request is approved, you can point the applications to the Stewards and send them a task to perform the action that appears in the data subject request, such as, removing the data or extracting the data and sending it to a customer.

The same approach can be applied to the integrated consent management applications. These applications need to know the processes for reaching the consent, and such applications reside in the Records of Processing Activities (called Process Register in Collibra), so that you can see all the processes that rely on the consent and the data categories for which you need consent.

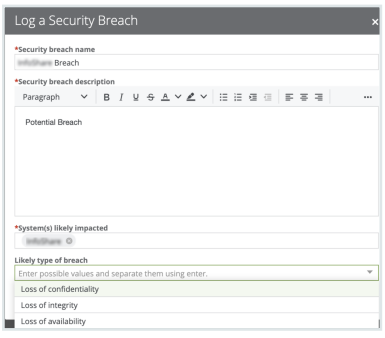


These are stored in the data sets that can also contain granular information, such as the individual data elements for which you want to obtain consent—this combines the information about which business processes require consent and the data categories for which you need consent to process all information in Collibra. The information governed in Collibra can be then sent to the consent management application that is used to manage consent.

Potential data breach workflow

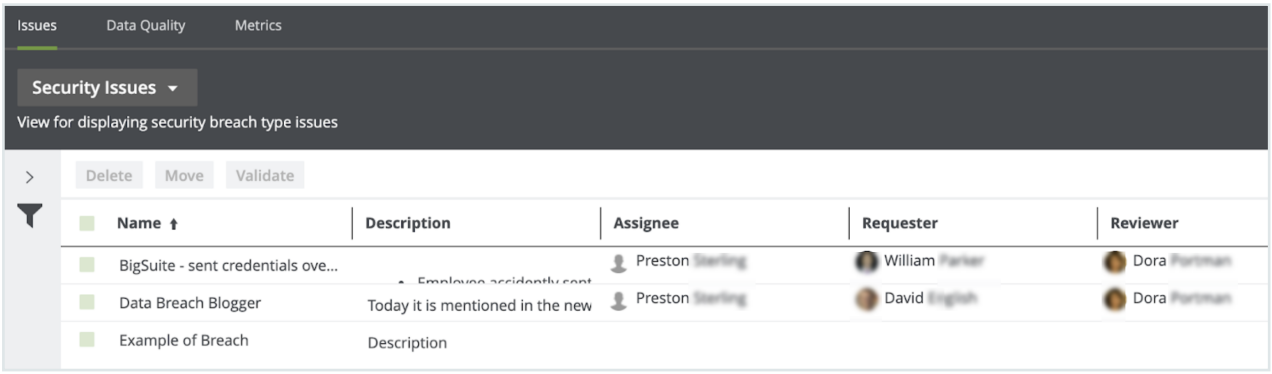
This section describes how Collibra helps when a data breach occurs.

With Collibra Data Privacy, Collibra for Desktop, or Collibra for Mobile, you can report any suspicious behavior by logging a potential data breach.



If your organization has suffered a potential data breach, you can determine the application that needs to be investigated and the type of breach that may have occurred,

and then log a potential data breach. The related workflow will require the Community Manager on the data governance counsel to assign an Issue Manager who will investigate the breach. The Issue Manager will then investigate the issue, assess the potential impact of the breach, determine the reporting requirements (for example, to whom the incident must be reported), and plan the remediation actions to address the risks. The reporting evidence needs to be stored. If you go to Data Helpdesk, you can find an overview of all the breaches that are being investigated.



Collibra can help with investigating the impact of the breach because of the knowledge of which data resides in the applications and the processes that use those applications. Such a holistic view on where the data resides, which applications are involved, and the processes that rely on these applications can help in assessing the impact on customers following a data breach. Collibra can not only help an organization log and investigate a data breach but also help analyze the impact of the breaches because Collibra knows where the data resides and how it is being used. In addition, it contains a history of all the breaches (including potential ones) that would have been logged.

How do we get there?

This section describes the Records of Processing Activities (called Process Register in Collibra), Business Process discovery capabilities, data categorization and classification, and different prescriptive paths for reaching from the logical data layer envisioned in the metamodel graph and connected data sets to a physical data layer present in columns located directly at the data source.

Create and maintain Process Register (RoPA)

Process Register is an essential part of privacy compliance, foreseen directly by GDPR article 30 as a Record of Processing Activities (RoPA) and derived from CCPA requirements for performing data mapping in the organization. Process Register enables to store assets of the Business Process type that describes processes in the organization that involve personal data. In Collibra, Business Processes reflect the requirements stated by Processing Activity in GDPR.

Business Process onboarding

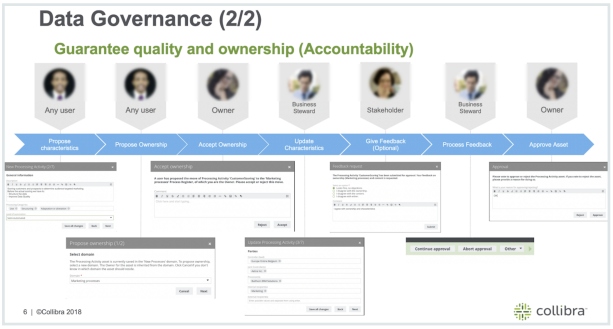
Business Processes may be onboarded by business users as well as privacy stewards through dedicated workflow implementing guided stewardship principle in Collibra Data Privacy. During onboarding, multiple roles collaborate in providing content to the onboarded Business Process. Because of the dedicated tasks and required approval and feedback, assets are onboarded in a governed way.

In the scenario on the Personal Information (PI) Discovery, it was described how Collibra helps with discovering Personal Information. But equally important to knowing where you are storing personal information is knowing why you are using personal information. That is, what the legal context of using that PI is. This context is created within Process Registers, throughout the usage of Business Processes that describe the processes conducted by organization relating to the usage of personal information.

Typically, that information does not reside with one person that can help you document that knowledge. That information is stored within multiple areas across the organization and it may not be easy to centralize this information and ensure that the information is up to date. To help you with this task, CollibraData Privacy comes with the Business Process discovery capabilities.

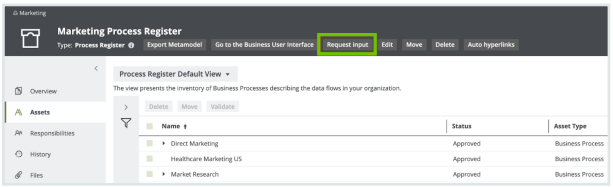
Consider a high-level overview of Data Privacy Business Process discovery capabilities. It commences with the Business Users describing the Business Processes in their terms. They will describe the data being used, applications being used, and any third parties with which they share information. After describing the Business Process, the owner of the Business Process will accept the ownership of that particular Business Process. When the ownership is accepted, the experts or the stewards will further onboard the proposed Business Process. This means that they will ensure that the Business Process is accurate

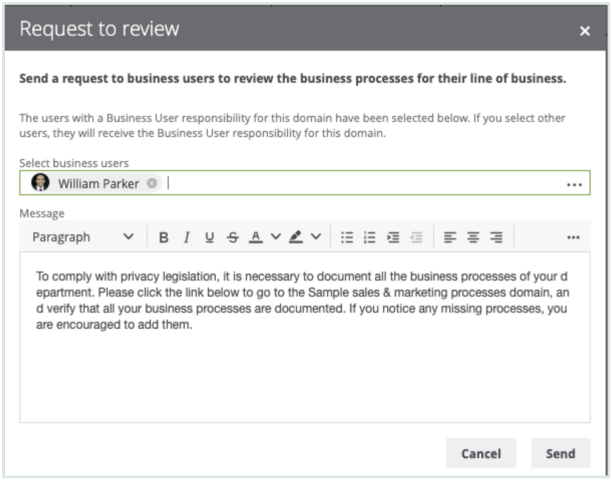
and actionable because that Business Process provides business context on how we use personal information and we must ensure that the description is accurate. Therefore, in principle, you will have the Business Steward, Privacy Steward, and Data Steward, each adding business metadata, adding privacy metadata, and performing data mapping, respectively. After the stewards have updated the characteristics, you can optionally obtain feedback from the stakeholders. The following sections describe each step involved in the process.



Requesting business users' input

The information related to Business Processes may be requested from the Business User directly from Data Privacy Process Register. Typically, this will be done by those who work on the Privacy program. With the **Request input** button, an email will be generated for the selected business users, which can provide relevant information on the business side of the process. You can have a guiding text that explains the purpose of your request. If you click **Send**, an email is sent to the business user with an invitation to contribute to the Process Register.



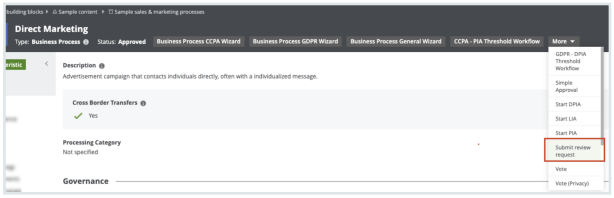


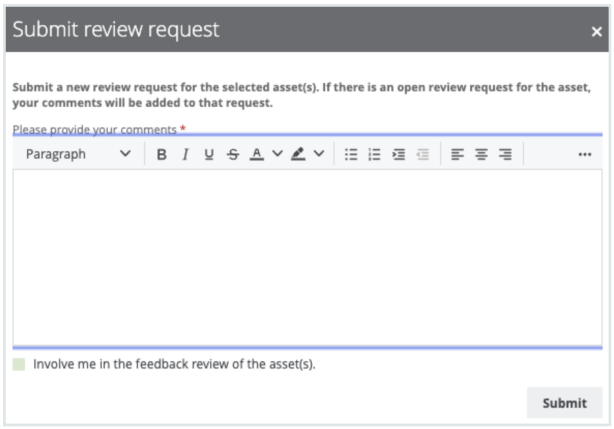
Maintain RoPA (Process Register) over time with review requests

While the successful result of the asset onboarding process is a new asset with the status Approved, asset change management is the standardized procedure for making changes to such approved assets.

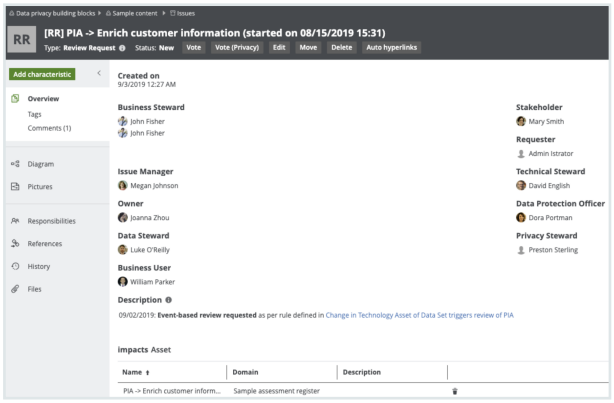
You may have many reasons to review an approved asset. Data Privacy groups such reasons into three categories and offers three corresponding means to trigger a review request:

- Manual: A trigger that is manually actioned by a user if, for example, the user wants to request a review of a Business Process asset considered to be incomplete or inaccurate. Any user can manually request a review of an approved asset.



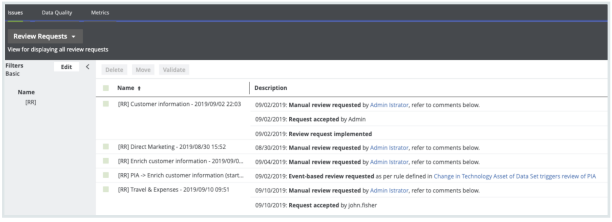


- Time-based: A trigger that is automatically actioned at a specified frequency. This is useful for assessment assets for which you might be required to review periodically to comply with a regulation.



- Event-based: A trigger that is automatically actioned by the fact of changes made to specified characteristics of the related asset.

All of the review requests are available in Data Helpdesk.



Perform assessments

Conduct PIA and DPIA

If a business process is likely to introduce a level of risk to the rights and freedom of natural persons, the Business Steward or the Data Protection Officer must perform the following:

- Privacy Impact Assessment (PIA), if complying with CCPA
- Data Privacy Impact Assessment (DPIA), if complying with GDPR

To determine whether or not you need to perform such an assessment for a Business Process asset, you must run a Threshold workflow.

The potential for business processes to expose the rights and freedom of natural persons to risk is significant. Privacy Impact Assessments (PIA) and Data Privacy Impact Assessments (DPIA) assess the risks to the rights and freedom of data subjects, born of a specific business process.

After onboarding a Business Process asset, the relevant Threshold workflow helps you determine whether or not a PIA or DPIA is needed. If it is determined that an assessment is necessary, the Owner or the Business Steward for the Business Process asset must complete the relevant workflow:

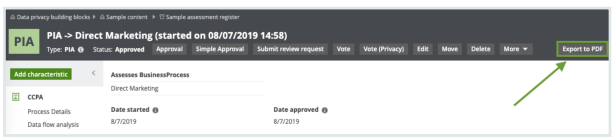
- PIA, if complying with CCPA
- DPIA, if complying with GDPR

Print assessment results

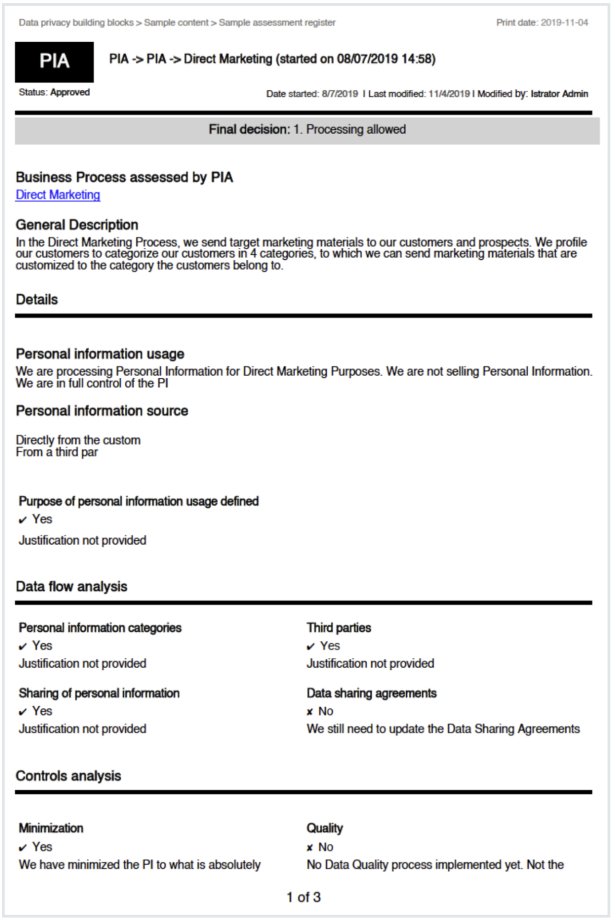
Assessments are a way for an organization to demonstrate compliance. You can export and print the PIA results in a unified way. You can also download a PIA asset page as a printable PDF, regardless of the status of the PIA asset.

Steps

1. Go to the relevant PIA asset page.



2. Click **Export to PDF**.
   » The PDF is downloaded to your computer.



Data privacy building blocks > Sample content > Sample assessment register Print date: 2019-11-04

PIA **PIA -> PIA -> Direct Marketing (started on 08/07/2019 14:58)**

Status: **Approved**

Date started: 8/7/2019 I Last modified: 11/4/2019 I Modified by: **Istrator Admin**

Final decision: 1. Processing allowed

Business Process assessed by PIA

Direct Marketing

General Description

In the Direct Marketing Process, we send target marketing materials to our customers and prospects. We profile our customers to categorize our customers in 4 categories, to which we can send marketing materials that are customized to the category the customers belong to.

Details

Personal information usage

We are processing Personal Information for Direct Marketing Purposes. We are not selling Personal Information. We are in full control of the PI

Personal information source

Directly from the custom
From a third par

Purpose of personal information usage defined

✔ Yes

Justification not provided

Data flow analysis

Personal information categories
✔ Yes
Justification not provided

Third parties
✔ Yes
Justification not provided

Sharing of personal information
✔ Yes
Justification not provided

Data sharing agreements
✘ No
We still need to update the Data Sharing Agreements

Controls analysis

Minimization
✔ Yes
We have minimized the PI to what is absolutely

Quality
✘ No
No Data Quality process implemented yet. Not the

1 of 3

Chapter 3

Essentials

This section contains information that can help you use Collibra Protect to the best of its ability.

Data protection types

This topic describes the types of protection that you can apply to your data via Protect.

> Tip *Data* refers to the tables and columns in a database.

Access-based protection

Access-based protection is the most basic type of protection that you can apply to your data. It involves providing the right users or groups access to the data based on the Collibra assets.

> Note Access-based protection is available only in data access rules.

Column-based protection

Column-based protection allows you to mask the data in specific columns so that the original data is not shown; for example, masking a column that contains credit card numbers.

You can mask the columns that are a part of a data category or a data classification. When granting access to a certain asset, you can apply the masking on only a subset of the asset if the subset is also a part of the data category or the data classification.

The following masking options are available:

- **Default masking:** Shows the data as **0**.
- **Hashing:** Shows the data as a set of different letters, numbers, and symbols.
- **Show last:** Shows the last few characters of the data. You can choose to show the last 1 through 20 characters of the data. The most common choice is 4.
- **No masking:** Shows the original data.

Note

- Column-based protection is available in both data protection standards and data access rules.
- For information about custom masking, go to Custom masking.

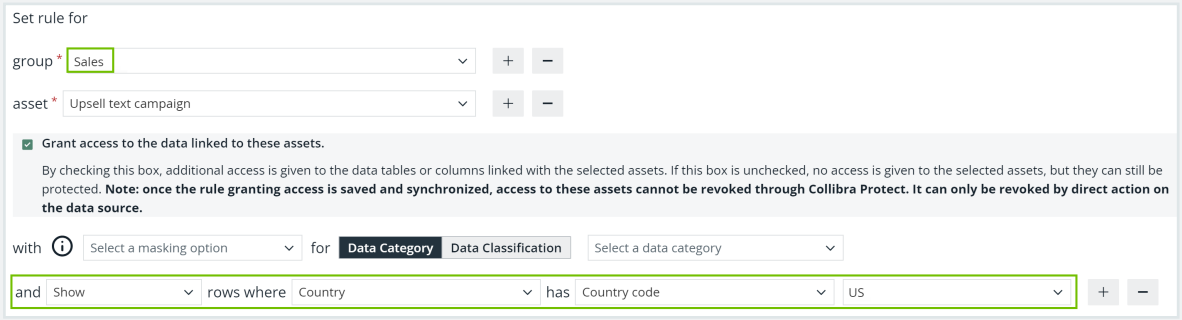
Example Suppose that you want the Human Resources (HR) group to be able to access a data set of US-based customers. Suppose that certain parts of the data set need to be hidden from the HR group because they contain restricted data, such as personally identifiable information (PII). Then, you can further protect the data by applying column-based protection or row-based protection.

Row-based protection

Row-based protection uses row filters to allow you to show or hide specific rows of a table. It is based on the values stored in the cell of a table.

Note Row-based protection is available only in data access rules.

Example Suppose that you want the Sales group to be able to access the data set of US-based customers. Then, you can create a data access rule and use the row-filtering option in the rule to show only those rows in the table that contain US in a column.



Custom masking

Custom masking is a feature that extends the data protection capabilities of Protect. Protect offers a set of out-of-the-box masking options. Custom masking allows you to define your own data protection methods.

You can manage custom masking via API. For more information, go to the Collibra Protect API documentation.

> Note
> - Custom masking functions are available only in Databricks and Snowflake. If you try to apply custom masking to a column in AWS Lake Formation or BigQuery, the out-of-the-box default masking is automatically applied to the column instead.
> - You cannot delete a custom masking function that is used in a data protection standard or a data access rule.

Example

The following is an example of a POST request for custom masking in Snowflake.

```
{
    "name": "My custom masking",
    "mappings": [
      {
        "provider": "Snowflake",
        "mappings": [
          {
            "dataType": "string",
            "functionName": "hash_my_string"
          },
          {
            "dataType": "number",
            "functionName": "hash_my_number"
          }
        ]
      }
    ]
}
```

If you apply **My custom masking** to a Snowflake column containing the value **Collibra**, the value is replaced by the result of the following Snowflake function: `hash_my_string (Collibra)`. However, if you apply this custom masking to a date column, the default

masking is automatically applied instead. This is because the POST request does not include any mapping for the date data type.

> Important The functionName specified in the mapping cannot contain spaces and cannot exceed 255 characters. Ensure that the masking functions exist on your data source provider. If a function does not exist, synchronization fails.

Masking functions

The following is an example of the syntax for a custom masking function in Databricks.

```
create or replace function mydb.myschema.mystring_function(value
STRING)
  RETURNS STRING
  RETURN concat("---", sha2(value, 0) , "+++");
```

The following is an example of the syntax for a custom masking function in Snowflake.

```
create or replace function mydb.myschema.mystring_function(value
VARCHAR)
    RETURNS VARCHAR
    AS
    $$
       concat('---', sha2(value) , '+++')
    $$;
```

Compatibility between Protect and Edge capability

Protect and Edge capabilities use different delivery mechanisms, which can result in compatibility differences. For example, you might have a version of Protect that supports custom masking, and a version of the Edge capability does not support it. If you use custom masking in a standard or rule, and your installed Edge capability does not support custom masking, synchronization is not triggered.

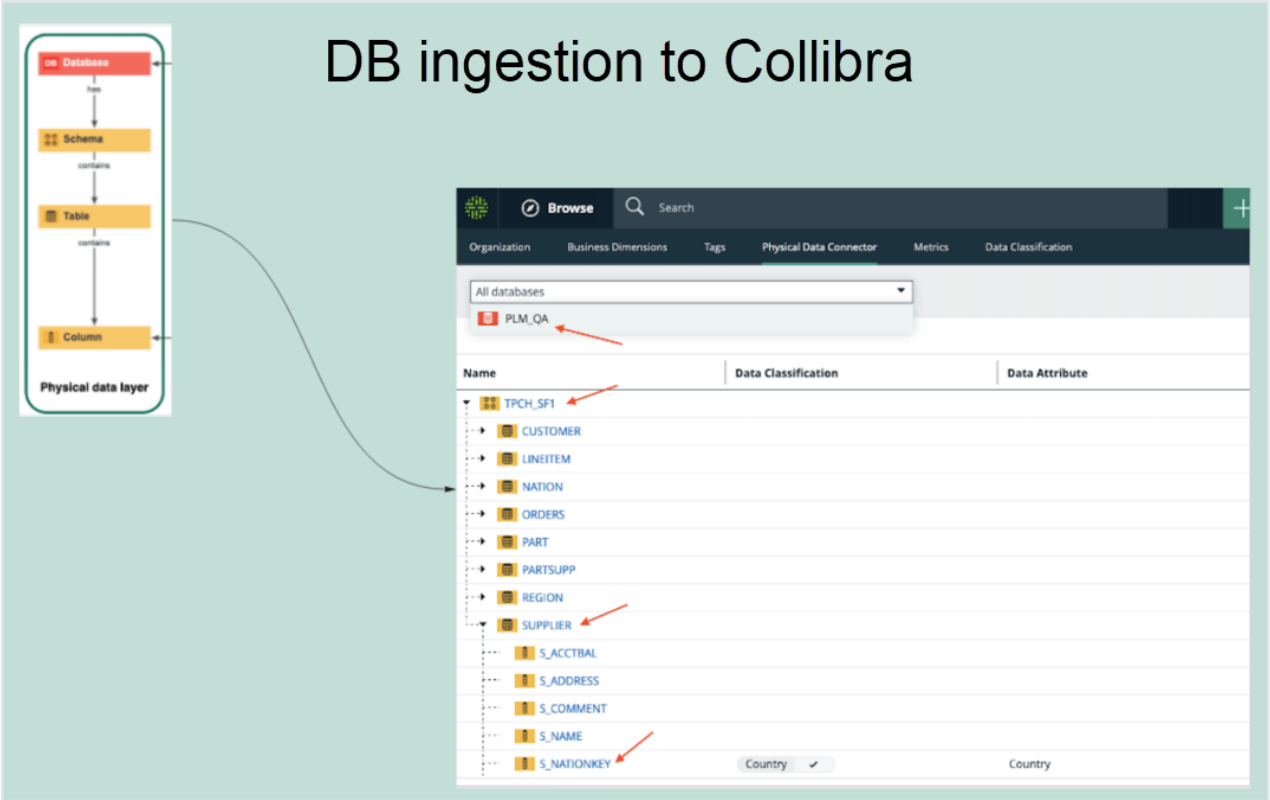
Technical background

This topic explains the connection of the data in a database (DB) with the physical layer (equivalent assets in Collibra Data Intelligence Cloud) and the logical layer (the packaged

model).

Consider the following DB.

When ingesting this DB to Collibra Data Intelligence Cloud, the physical layer is created, in addition to an asset for each of the schemas, tables, and columns, as depicted in the following image.



After the physical layer is created in Collibra, the logical layer can be created on top of the physical layer, as follows:

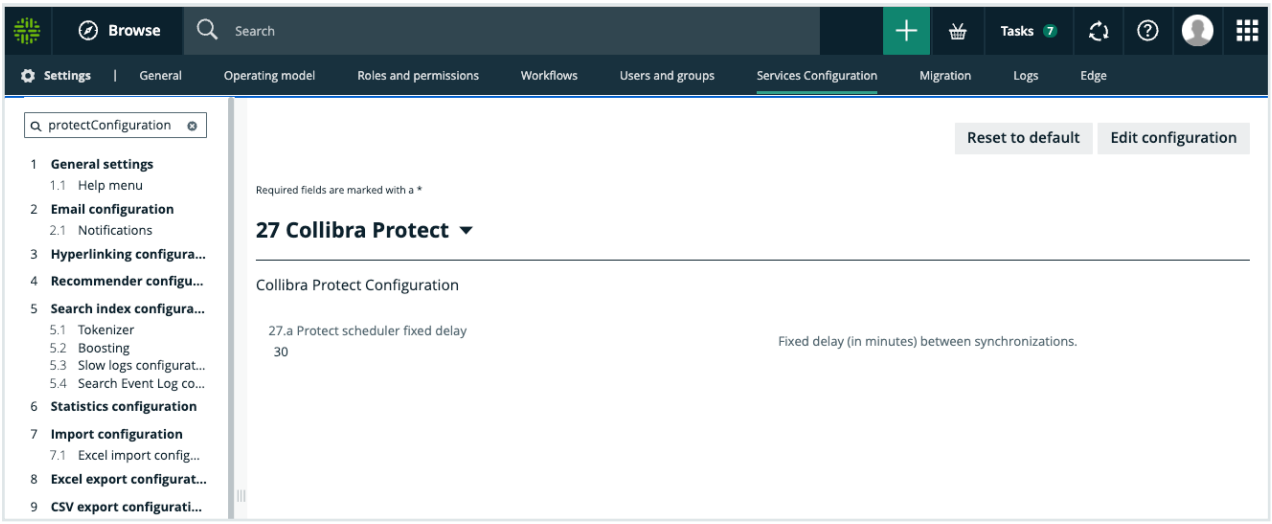
- Select any column and classify it as any available data classification. Alternatively, you can allow Collibra to classify the column for you.
- Assign the column to a data attribute.
- Create additional assets or use the existing assets of different types (Business Process, Data Category, or Data Set) to establish a relation with the columns.

> Note Protect supports only those columns that are linked to Table assets. It does not support Database View assets.

Synchronization

Collibra Protect automatically synchronizes data protection standards and data access rules with the databases of the data source providers such as BigQuery and Snowflake at regular intervals. This synchronization runs in the background on a configured frequency. By default, the frequency is every 60 minutes. You can, however, change the frequency

through the **Protect scheduler fixed delay** field on the Services Configuration tab in Collibra.



> Important If the **Services Configuration** tab is not shown to you, create a support ticket asking the following JVM Parameter be added to your Collibra Infrastructure Configuration: **-DPROTECT_SYNC_SCHEDULER_DELAY=PT60M**. After the parameter is added, restart Collibra to synchronize the Protect policies with the data source providers.

The synchronization includes the following processes:

- Aggregation of all data protection standards and data access rules with a computation of the following:
    - Which columns need to be masked for which groups
    - Which tables need to have a row filter
    - Which tables and columns need to be granted access
- On the databases of the data source providers such as Snowflake:
    - Creation and application of masking
    - Creation and application of row filters
    - Granting of access to groups on tables and columns (depending on the underlying database)

Data protection standards and data access rules

Collibra Protect protects your data through data protection standards and data access rules.

Data protection standards create a primary layer of protection for similar types of data by masking the data wherever it is stored, whereas data access rules create an additional layer of protection by managing access and enhancing protection for specific usages.

This topic explains when to create a data protection standard over a data access rule and vice versa, and what to consider when creating them.

When to create a standard over a rule and vice versa

- Suppose that columns containing the first and last names are a part of the Personally Identifiable Information (PII) data category. Then, regardless of the databases, tables, and schemas to which those columns belong, you can create a data protection standard that targets all of those columns by selecting the PII data category in the standard and masking it.
  Then, you can create a data access rule that grants access to a specific group, for a specific data set, while knowing that all PII within this data set will be masked by the data protection standard.
- Suppose that a data protection standard is created to mask a column that is classified as Personally Identifiable Information (PII) for everyone. You, however, want to unmask that PII column for a specific group. You can do so by creating for the same group a data access rule to unmask the classified column, because data access rules take priority over data protection standards.
- Suppose that you want to grant access to a group, but the protection from the data protection standard is not enough because there might be other sensitive data within a supported asset. Then, you can create a data access rule to add additional layers of protection over the ones that were set by the data protection standard. You can further protect the data by applying additional masking on the data or by filtering the data using the row-filtering option in the rule.

What to consider when creating standards or rules

When creating data protection standards or data access rules for assets, consider how the assets are grouped. Suppose that you have a Business Process asset, BP, which contains the following Data Set assets: DS1, DS2, and DS3. Instead of creating a data protection standard or data access rule for each of the three Data Set assets (DS1, DS2, and DS3), consider creating a standard or rule that targets the Business Process asset (BP), to save your time.

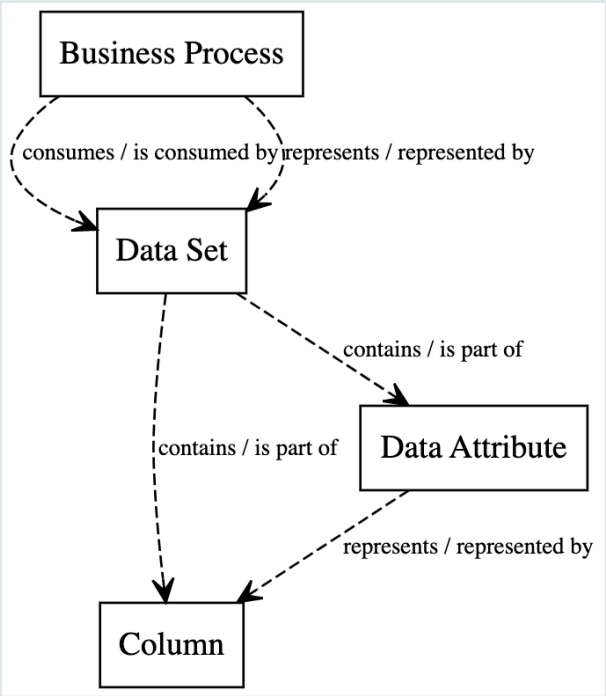
Prescriptive paths

You can use Collibra Protect to secure the data in the assets of the packaged asset types, such as Business Process, Data Category, and Data Set, in addition to the assets of any new or modified asset types.

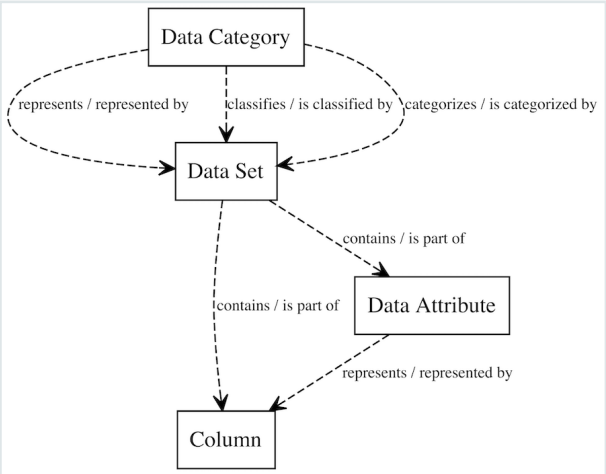
The asset that you select when creating a data protection standard or a data access rule is related to the physical data layer, such as tables and columns, through a set of relations and intermediate assets. These relations are paths that Protect uses to traverse from the selected asset (business or logical layer) to a column (physical data layer) in order to find the column that needs protection. Such traversal follows a set of prescriptive paths. Each asset type has a set of prescriptive paths for traversing to the Column asset, as depicted in the following sections.

> Note Depending on your permission, you can customize the prescriptive paths. For more information, go to Customization of prescriptive paths.

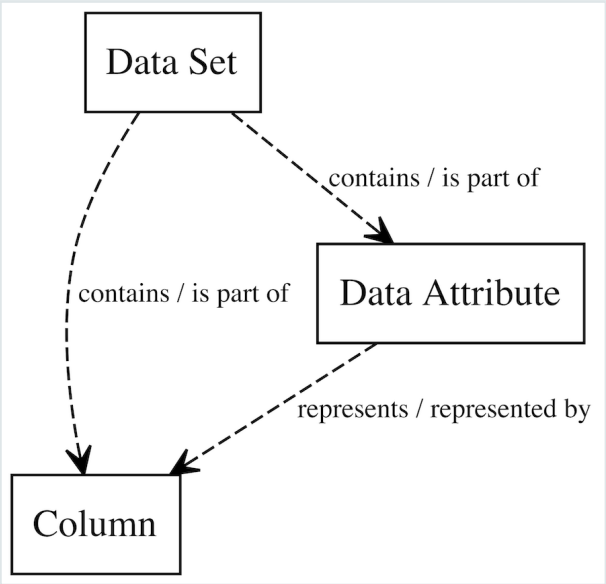
Business Process



Data Category



Data Set



Customization of prescriptive paths

Collibra Protect supports the following asset types:

- **Packaged asset types:** Business Process, Data Category, and Data Set
- **Custom asset types:** These are the packaged asset types that you have modified or the asset types that you have created. If you modify the attributes and relations of a packaged asset type, then the packaged asset type becomes a custom asset type.

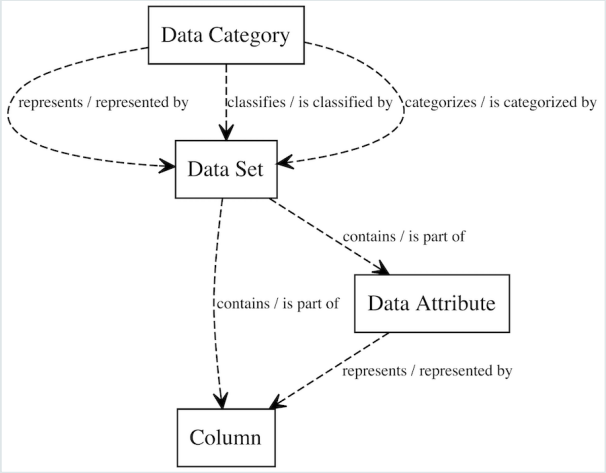
If you have the **Protect** > **Administration** global permission, you can customize the prescriptive paths for the asset types through APIs. The customization may include creating, modifying, or deleting the prescriptive paths: for example, adding or modifying the prescriptive paths for packaged and custom asset types, defining how the asset types relate to columns, and removing any obsolete prescriptive paths.

The customized prescriptive paths are applied to data protection standards and data access rules.

> Note You cannot remove a customized prescriptive path if an asset type linked to the prescriptive path is used in a standard or rule.

Protect supports a maximum of 10 asset types. Each asset type can have a maximum of 6 relations and a maximum depth of 3. However, when customizing the prescriptive path for an asset type, we recommend that you provide only one relation for the asset type. Prescriptive paths must always end in a Column asset type (that is, 00000000-0000-0000-0000-000000031008).

The following image is an example of a prescriptive path that has 6 relations and a depth of 3.



Restore the default asset types

If you want to restore the default asset types defined by Collibra, a PATCH operation must be performed on each asset type. The list of asset types and their specifications are as follows.

If Data Privacy is not installed

Data Set (00000000-0000-0000-0001-000400000001)

```
                    {
    "description": "Prescriptive path from Data Set to Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
```

```
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-000000031008"
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-
000000031005",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007094",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      }
    ],
    "assetTypeId": "00000000-0000-0000-0001-000400000001"
  }
```

Data Category (00000000-0000-0000-0000-000000031109)

```
                  {
    "description": "Prescriptive path from Data Category to
Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
```

```
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
```

```
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      ],
      "assetTypeId": "00000000-0000-0000-0000-000000031109"
    }
```


Business Process (00000000-0000-0000-0000-000000031103)


```
                    {
      "description": "Prescriptive path from Data Set to Column",
      "relations": [
        {
          "relationTypeId": "00000000-0000-0000-0000-
000000007062",
          "relationTypeDirection": "SOURCE",
          "assetType": {
            "assetTypeId": "00000000-0000-0000-0000-000000031008"
          }
        },
        {
          "relationTypeId": "00000000-0000-0000-0000-
000000007062",
          "relationTypeDirection": "SOURCE",
          "assetType": {
            "assetTypeId": "00000000-0000-0000-0000-
```

```
000000031005",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007094",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      }
    ],
    "assetTypeId": "00000000-0000-0000-0001-000400000001"
  }
```

If Data Privacy is installed

Data Set (00000000-0000-0000-0001-000400000001)

```
                    {
    "description": "Prescriptive path from Data Set to Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-000000031008"
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-
000000031005",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007094",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
```

```
            }
          }
        }
      ],
      "assetTypeId": "00000000-0000-0000-0001-000400000001"
    }
```

Data Category (00000000-0000-0000-0000-000000031109)

```
                    {
    "description": "Prescriptive path from Data Category to
Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
```

```
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
```

```
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007315",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "c0e00000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007315",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "c0e00000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      }
```

```
    ],
    "assetTypeId": "00000000-0000-0000-0000-000000031109"
  }
```

Business Process (00000000-0000-0000-0000-000000031103)

```
                  {
    "description": "Prescriptive path from Business Process to
Column",
    "relations": [
      {
        "relationTypeId": "c0e00000-0000-0000-0000-
000000007314",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "c0e00000-0000-0000-0000-
000000007314",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "c0e00000-0000-0000-0000-
000000007314",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
```

```
000000031008"
                }
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
```

```
            }
          }
        ],
        "assetTypeId": "00000000-0000-0000-0000-000000031103"
      }
```

Set up Protect

This topic describes how to set up Protect and establish a connection between your data source and Protect.

> Tip
> The information in this topic varies depending on the data source that you select.
>
> Data source
>
> 

Before you begin

AWS Lake Formation

1. Download the JDBC driver for Amazon Athena.
2. Create a JDBC connection from your Edge site to Amazon Athena.

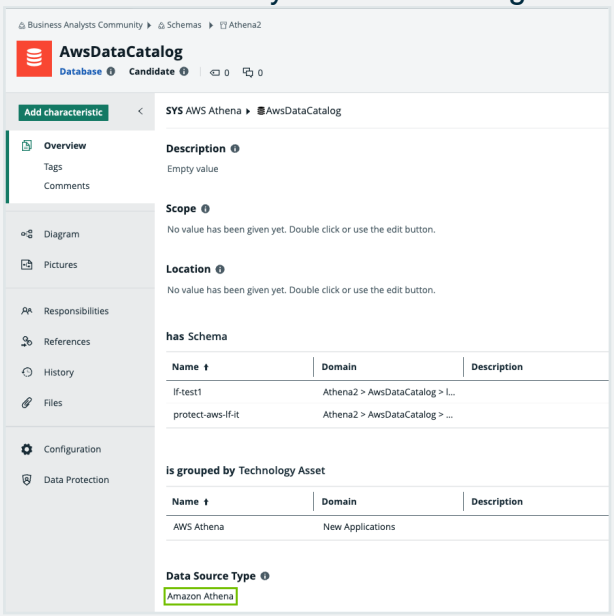
   > Tip When creating the connection, in the **Connection provider** field, select **Generic JDBC connection**.

3. Add the Catalog JDBC ingestion capability to the Edge site.

   > Tip
   > When adding the capability:
   >
   > - In the **Capability template** field, select **Catalog JDBC Ingestion**.
   > - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

Tip The following image shows an ingested AWS Lake Formation database. The **Data Source Type** attribute containing the value **Amazon Athena** is added to the database asset only after the Catalog JDBC ingestion process is complete.



BigQuery

1. Download the JDBC driver for Google BigQuery.
2. Create a JDBC connection from your Edge site to Google BigQuery.

   Tip When creating the connection, in the **Connection provider** field, select **Generic JDBC connection**. In the **Connection properties** section, set the value of the **Other** connection property to **SupportNativeDataType=True**.

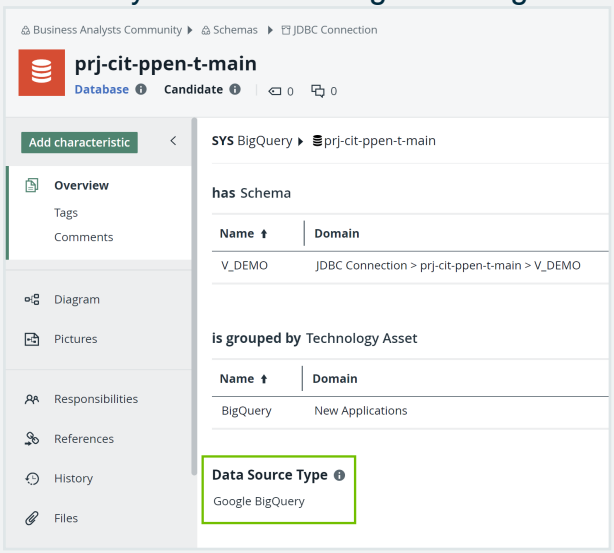
3. Add the Catalog JDBC ingestion capability to the Edge site.

   Tip
   When adding the capability:

   - In the **Capability template** field, select **Catalog JDBC Ingestion**.
   - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

Tip The following image shows an ingested BigQuery database. The **Data Source Type** attribute containing the value **Google BigQuery** is added to the database asset only after the Catalog JDBC ingestion process is complete.



Watch a video

Databricks

1. Download the JDBC driver for Databricks.
2. Create a JDBC connection from your Edge site to Databricks.

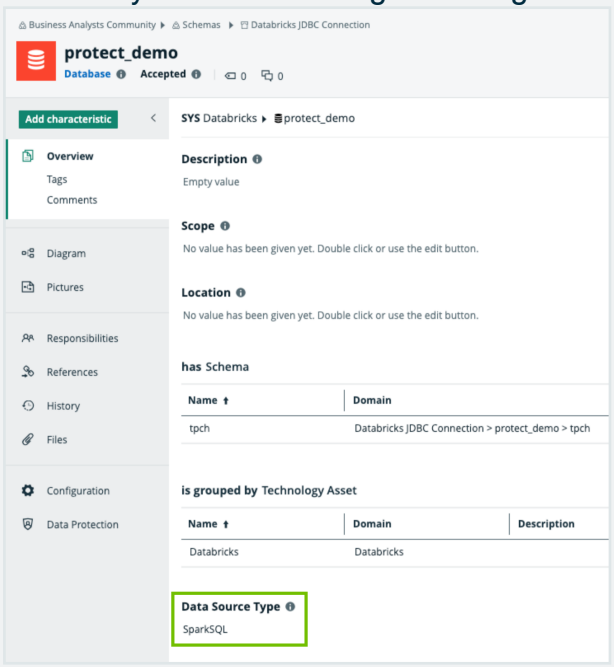
> Tip When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.

3. Add the Catalog JDBC ingestion capability to the Edge site.

> Tip
> When adding the capability:
>
> - In the **Capability template** field, select **Catalog JDBC Ingestion**.
> - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

> Tip The following image shows an ingested Databricks database. The **Data Source Type** attribute containing the value **SparkSQL** is added to the database asset only after the Catalog JDBC ingestion process is complete.



Snowflake

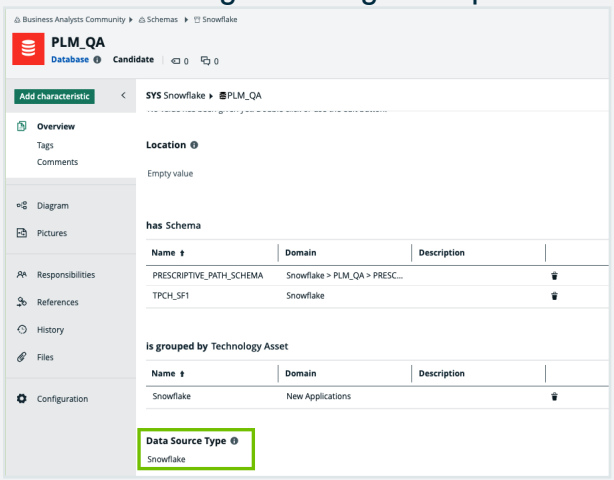
1. Download the JDBC driver for Snowflake.
2. Create a JDBC connection from your Edge site to Snowflake.

> Tip When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.

3. Add the Catalog JDBC ingestion capability to the Edge site.

> Tip
> When adding the capability:
>
> - In the **Capability template** field, select **Catalog JDBC Ingestion**.
> - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

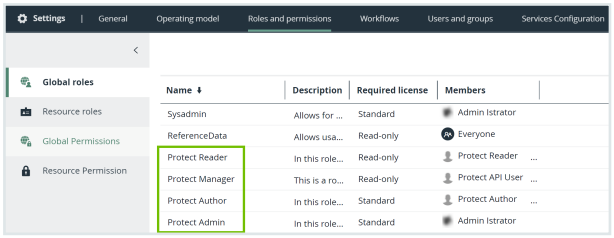
> Tip The following image shows an ingested Snowflake database. The **Data Source Type** attribute containing the value **Snowflake** is added to the database asset only after the Catalog JDBC ingestion process is complete.
>
> 
>

Steps

AWS Lake Formation

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.

2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create an AWS connection from the Edge site to Amazon Athena.

> Tip
> - When creating the connection, in the **Connection provider** field, select **AWS connection**.
> - Ensure that the user associated with the Access Key ID used in the connection has the required permissions.

4. Add the Protect for AWS Lake Formation capability to the Edge site.

> Tip
> - When adding the capability:
>   - In the **Capability template** field, select **Collibra Protect for AWS Lake Formation**.
>   - In the **Connection** field, select the connection that you created in Step 3.
> - Do not add more than one Protect for AWS Lake Formation capability to the Edge site.

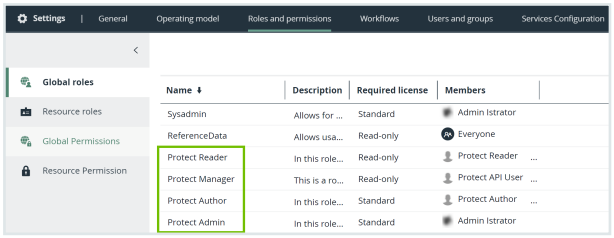
» Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

BigQuery

> Note Apart from the JDBC connection created for the Catalog ingestion, Protect for BigQuery requires an extra connection, which is the GCP connection. The GCP connection is necessary because Protect requires access to certain GCP APIs that cannot be reached through the JDBC connection alone. The GCP connection ensures that data protection is enforced.

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.

2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create a GCP connection from the Edge site to Google BigQuery.

> Tip
> - When creating the connection, in the **Connection provider** field, select **GCP connection**.
> - Ensure that the user associated with the GCP Service Account used in the connection has the required permissions.

4. Add the Protect for BigQuery capability to the Edge site.

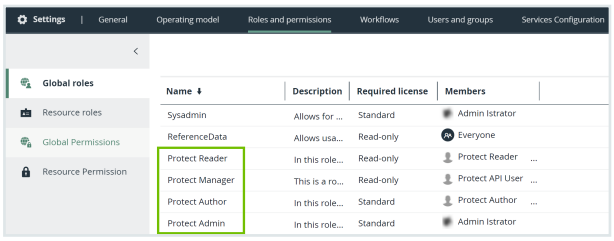
> Tip
> - When adding the capability:
>   - In the **Capability template** field, select **Collibra Protect for Google BigQuery**.
>   - In the **Connection** field, select the connection that you created in Step 3.
> - Do not add more than one Protect for BigQuery capability to the Edge site.
> - If the version of the capability is **1.97.1**, then ensure that the JSON content in the **GCP Service Account** field in the GCP connection you created is Base64 encoded. You can find the version of the capability in the **Version** column on the **Capabilities** tab.

» Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

Watch a video

Databricks

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.
2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create a Username/Password JDBC connection from the Edge site to Databricks.

   Tip
   - When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.
   - Ensure that the user associated with the Databricks role used in the connection has the required privileges.

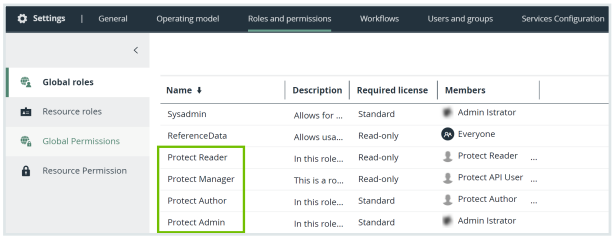
4. Add the Protect for Databricks capability to the Edge site.

   Tip
   - When adding the capability:
     - In the **Capability template** field, select **Collibra Protect for Databricks**.
     - In the **Connection** field, select the connection that you created in Step 3.
   - Do not add more than one Protect for Databricks capability to the Edge site.

   » Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

Snowflake

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.
2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create a Username/Password JDBC connection from the Edge site to Snowflake.

   Tip
   - When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.
   - Ensure that the user associated with the Snowflake role used in the connection has the required privileges.

4. Add the Protect for Snowflake capability to the Edge site.

   Tip
   - When adding the capability:
     - In the **Capability template** field, select **Collibra Protect for Snowflake**.
     - In the **Connection** field, select the connection that you created in Step 3.
   - Do not add more than one Protect for Snowflake capability to the Edge site.

   » Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

Protect global roles and permissions

The following tables describe the global roles and global permissions that are specific to Collibra Protect.

Global role	Description
Protect Reader	A user who can view Protect with read-only access to data protection standards and data access rules.
Protect Author	A user who can: • Create standards and rules. • Modify or delete only the standards and rules that they created. • View imported policies. • View groups. • Generate audit logs as an individual contributor.
Protect Admin	A user who has the same permissions as a Protect Author. In addition, this user can modify or delete the standards and rules created by others, and access additional APIs.

> Note The **Protect Manager** global role is intended only for the Protect system user.

Global permission	Description
Product Rights > Protect	The **Read-only** license is required for this permission. The permission allows a user to access Collibra Protect. All Protect global roles and the **Edge site** global role have this permission.

Global permission	Description
Protect > Edit	The **Standard** license is required for this permission. The permission allows a user to: • Create standards and rules. • Modify only the standards and rules that they created. • Delete only the standards and rules that they created.
Protect > Administration	The **Standard** license is required for this permission. The permission allows a user to: • Create standards and rules. • Modify all standards and rules. • Delete all standards and rules.

Open Protect

This topic describes how to open Collibra Protect and what is shown on the **Protect** landing page.

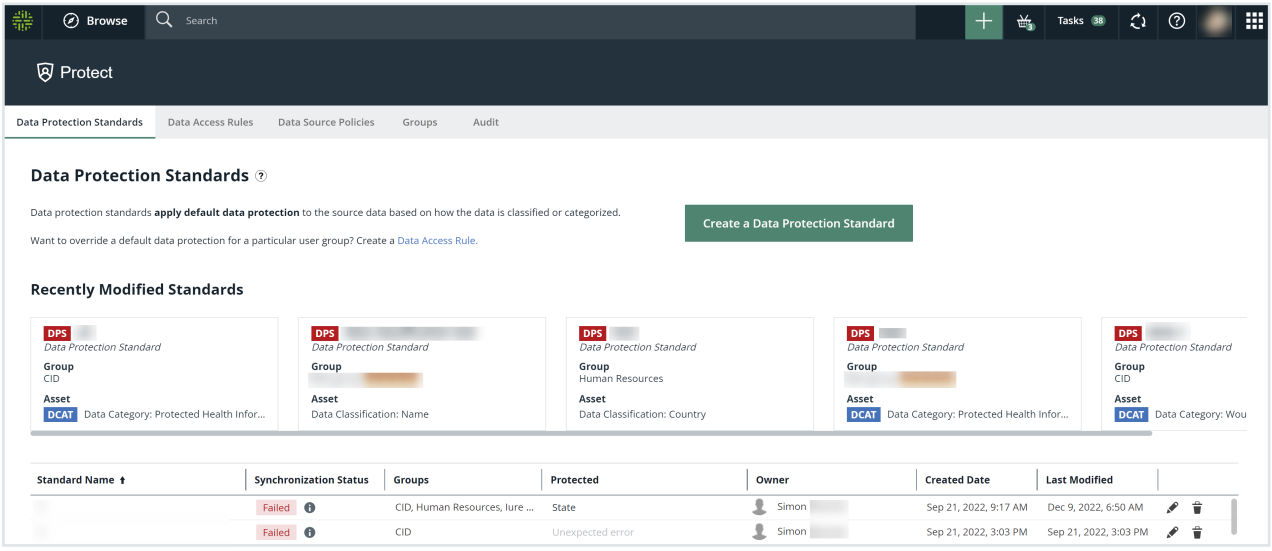
Requirements and permissions

You have a global role that has the **Protect** global permission.

Steps

On the main menu, click , and then click **Protect**.

» The **Protect** landing page opens.



Protect landing page

The following table describes the tabs that are shown on the **Protect** landing page depending on your role.

Tab	Description
Data Protection Standards	Data protection standards to define data source access to data types based on data categories, data attributes, or data classifications.
Data Access Rules	Data access rules to grant specific groups different accesses to the same data in business processes, data categories, or data sets. Note Data access rules take priority over data protection standards.
Data Source Policies	Policies that are active in the data source tables.
Groups	Groups that are mapped to the roles in data sources for use in data protection standards and data access rules.
Audit	Option to generate an audit log of the ingested data from the data sources.

Protect groups

You must create at least one Protect group before creating a data protection standard or a data access rule. Each Protect group is associated with a role in the data source provider.

> Note In BigQuery, *roles* are referred to as *principals*.

The **Groups** tab in Protect contains an overview of the Protect groups that are created for standards and rules. The table on the **Groups** tab contains the Protect groups that are active in the data source.

This topic describes how to create a Protect group and what is shown on the **Groups** tab in Protect.

Create a Protect group

Requirements and permissions

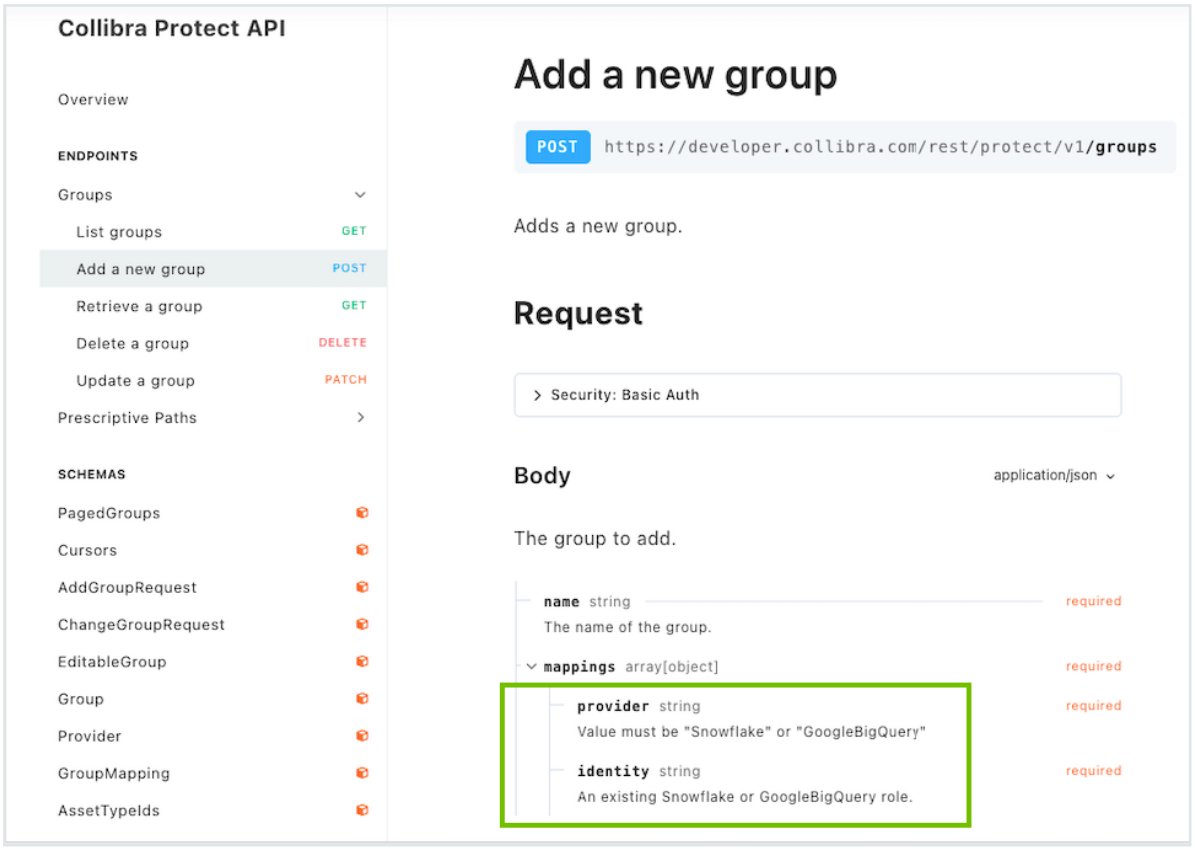
You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

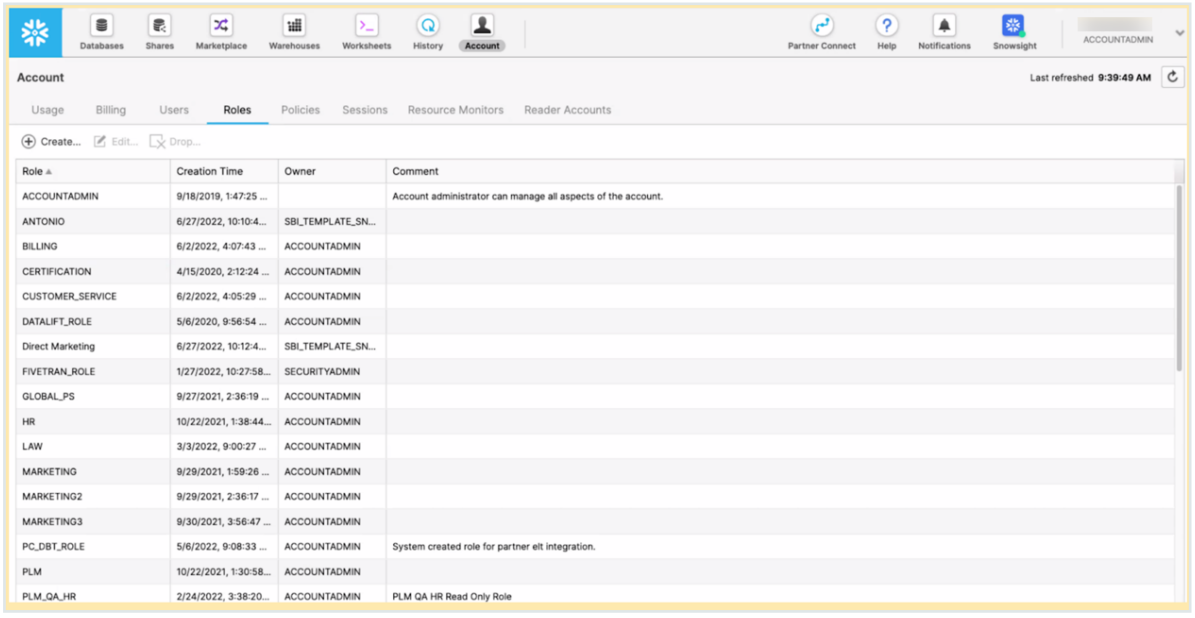
1. Open Protect.
2. On the **Groups** tab, click **Collibra Protect Group API**.
3. For the next steps, go to Add a new group.

Details

- When creating a Protect group, you are prompted to specify the data source provider (**AWSLakeFormation**, **Databricks**, **GoogleBigQuery**, or **Snowflake**) and the existing role from the provider to map the role to the group.

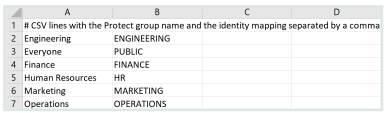


- The following image shows the roles in Snowflake.



- The following images show a CSV file (named **protect_groups.csv**) that contains Protect groups to be added to Collibra, and a bash script that adds those Protect

groups to Collibra for Snowflake.



	A	B	C	D
1	# CSV lines with the Protect group name and the identity mapping separated by a comma			
2	Engineering	ENGINEERING		
3	Everyone	PUBLIC		
4	Finance	FINANCE		
5	Human Resources	HR		
6	Marketing	MARKETING		
7	Operations	OPERATIONS		

```
#!/usr/bin/env bash

# COLLIBRA_URL should point to your Collibra deployment
COLLIBRA_URL="https://my_company.collibra.com"

# COLLIBRA_AUTH should contain the Collibra user and password separated by a colon
# This user should be able to create Protect groups (ie have the global role Protect Author and/or Admin)
COLLIBRA_AUTH="user:password"

# Which provider does the identity map to? Value must be "Snowflake", "GoogleBigQuery", etc
PROTECT_GROUP_PROVIDER="Snowflake"

if [[ -z "${COLLIBRA_URL}" ]]; then
    echo "Environment variable COLLIBRA_URL has not been defined"
    exit 1
fi
if [[ -z "${COLLIBRA_AUTH}" ]]; then
    echo "Environment variable COLLIBRA_AUTH has not been defined"
    exit 1
fi

{
    read # Ignore first line in csv file
    while IFS=, read -r field1 field2
    do
        echo "Add group $field1 for $PROTECT_GROUP_PROVIDER with identity $field2"
        curl -u "${COLLIBRA_AUTH}" -X POST "${COLLIBRA_URL}/rest/protect/v1/groups" -H "accept: application/json" -H "Content-Type: application/json" -d @- << EOF
        {
            "name": "$field1",
            "mappings":
            [
                {
                    "provider": "${PROTECT_GROUP_PROVIDER}",
                    "identity": "$field2"
                }
            ]
        }
EOF
    done
} < protect_groups.csv
```

Groups tab

The following table describes the columns that are shown in the table on the **Groups** tab.

Column	Description
Group Name	The name of the group.
System Reference	References to identify the data source provider and the native identifier associated with the group.
Created By	The name of the user who created the group.
Created Date	The date when the group was created.

> Note
> - Multiple Protect groups can be mapped to the same data source identity.
> - Within a single Protect group, only one mapping per data source is supported.

Data protection standards

Data protection standards in Collibra Protect protect your data by masking similar types of data wherever it is stored, through column-based protection.

Create a data protection standard

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.
- You have the **Catalog** global role. This role is required to view data classifications for selection in a data protection standard.

Before you begin

Ensure that Protect groups have been created.

Steps

1. Open Protect.
2. Click the **Data Protection Standards** tab.
3. Click **Create a Data Protection Standard**.
   » The **Create a Data Protection Standard** dialog box appears.
4. Enter the required information.

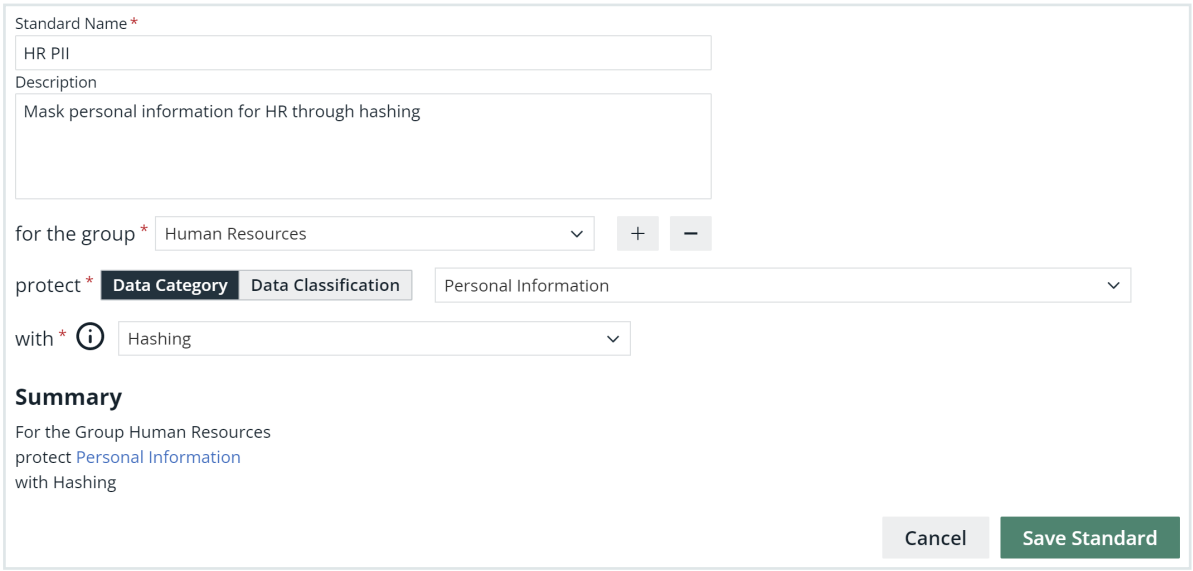
Details

Field	Description
Standard Name	Enter a name for the data protection standard.
Optional: Description	Enter a description for the data protection standard.

Field	Description
Group	Select the group for the data protection standard. Tip You can add more groups by using the plus icon.
Protect (Data Category/Data Classification)	Click **Data Category** or **Data Classification**, and then select the data category or data classification that you want to protect. Note If the association between the data classification and a column is not yet accepted yet, the standard ignores the column.

Field	Description
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. Tip This field contains the following options: ◦ **Default masking** ◦ **Hashing** ◦ **Show last**

Tip
- The **Summary** section shows a summary of the standard.



5. Click **Save Standard**.
   » A message appears stating that the standard is sent to source, and the standard is shown in the table on the **Data Protection Standards** tab.

Modify a data protection standard

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

  > Note If you have the the **Protect** > **Edit** global permission, you can modify only the data protection standard that you created. If you have the **Protect** > **Administration** global permission, you can modify any data protection standard.

- You have the **Catalog** global role. This role is required to view data classifications for selection in a data protection standard.
- You have the permissions to view the assets that are associated with the data protection standard. Otherwise, the **Unauthorized Asset** value is shown to you when you modify the standard.

Steps

1. Open Protect.
2. In the table, in the row containing the standard that you want to modify, click ✎.
   » The **Edit a Data Protection Standard** dialog box appears.
3. Modify the required information.

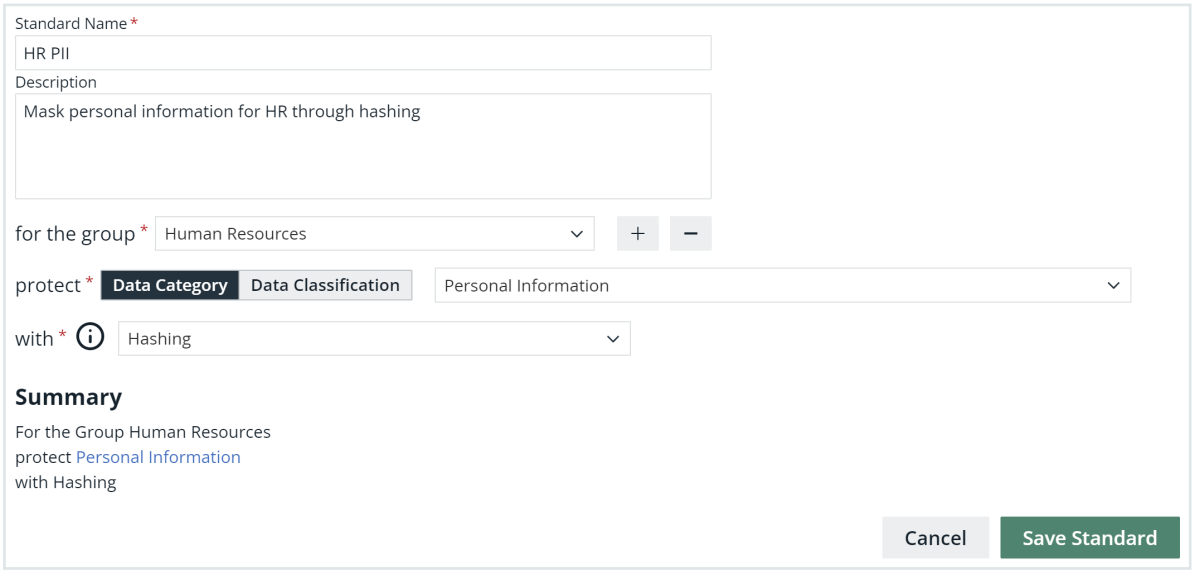
Details

Field	Description
Standard Name	Enter a name for the data protection standard.

Field	Description
Optional: Description	Enter a description for the data protection standard.
Group	Select the group for the data protection standard. Tip You can add more groups by using the plus icon.
Protect (Data Category/Data Classification)	Click **Data Category** or **Data Classification**, and then select the data category or data classification that you want to protect. Note If the association between the data classification and a column is not yet accepted yet, the standard ignores the column.

Field	Description
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. Tip This field contains the following options: ◦ **Default masking** ◦ **Hashing** ◦ **Show last**

Tip
- The **Summary** section shows a summary of the standard.



4. Click **Save Standard**.
   » A message appears stating that the standard is sent to source, and the standard is shown in the table on the **Data Protection Standards** tab.

Delete a data protection standard

Requirements and permissions

You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

1. Open Protect.
2. Click the **Data Protection Standards** tab.
3. In the table, in the row containing the standard that you want to delete, click 🗑.
   » The **Delete Data Protection Standard** dialog box appears.
4. Click **Delete**.
   » A message appears stating that the request to delete the standard is received.

> Tip You can check the status of the standard in the **Synchronization Status** column in the table on the **Data Protection Standards** tab.

Data Protection Standards tab

The **Data Protection Standards** tab in Protect contains an overview of data protection standards. The **Recently Modified Standards** section on the tab shows the 5 last modified data protection standards.

The following table describes the columns that are shown in the table on the **Data Protection Standards** tab.

Column	Description
Standard Name	The name of the standard.
Synchronization Status	The status of synchronization between the standard in Protect and that in the data source.
Groups	The groups for which the standard is created.
Protected	The assets that the standard protects. Tip Depending on your role, you can view the details of an asset by clicking the asset link in this column.
Owner	The name of the user who created the standard.
Created Date	The date and time when the standard was created.
Last Modified	The date and time when the standard was last modified.

Synchronization status

The following table describes the statuses that may be shown in the **Synchronization status** column on the **Data Protection Standard** tab.

> Tip To view the status of the data protection standard in the data source, go to the database of the data source provider.

Synchronization Status	Description
Active	The standard is enforced in the data source.
Pending	The standard is created or modified and is pending synchronization.
Failed	The synchronization of the standard has failed. Tip For more information about the error, click next to the status.
Delete Pending	The standard will be deleted during the next synchronization.
Not Deleted	The standard could not be deleted. Tip For more information about the error, click next to the status.

> Note Protect periodically synchronizes with your data source providers to update the status of the data protection standards in Collibra, except if the status is **Failed**. For more information, go to Synchronization.

Data access rules

Data access rules in Collibra Protect protect your data by managing access and enhancing protection for specific usages. They protect your data by:

- Managing access to the data (access-based protection)
- Masking the data (column-based protection)
- Filtering the data (row-based protection)

Create a data access rule

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.
- You have the **Catalog** global role. This role is required to view data classifications for selection in a data access rule.

Before you begin

Ensure that Protect groups have been created.

Steps

1. Open Protect.
2. Click the **Data Access Rules** tab.
3. Click **Create a Data Access Rule**.
   » The **Create a Data Access Rule** dialog box appears.
4. Enter the required information.

Details

Field	Description
Rule Name	Enter a name for the data access rule.
Optional: Description	Enter a description for the data access rule.

Field	Description
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.
Asset	Select the data asset that the rule is protecting. Tip ◦ This field contains Business Process, Data Category, and Data Set assets, in addition to assets of custom asset types. ◦ For more information, go to Technical background and Prescriptive paths. ◦ You can add more groups by using the plus icon.

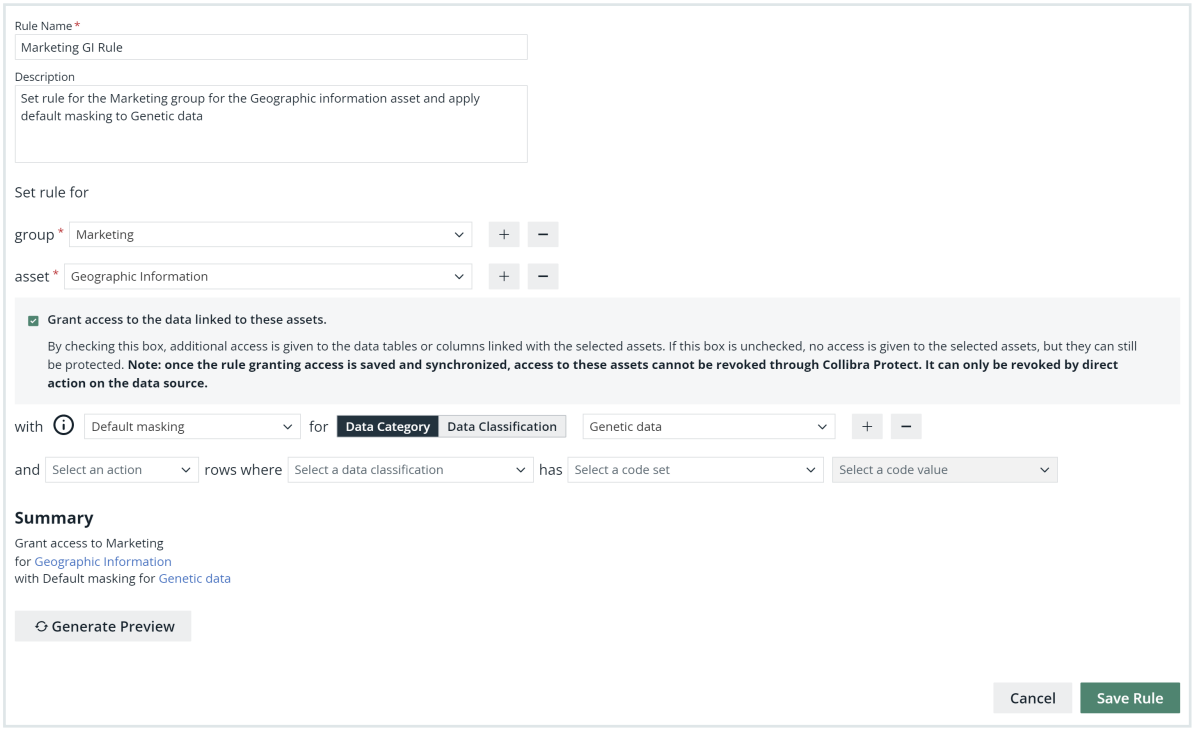
Field	Description
Optional: With (masking option)	◦ Select the type of masking that you want to apply to a data category or data classification. Tip This field contains the following options: ■ **Default masking** ■ **Hashing** ■ **Show last** ■ **No masking** ◦ Click **Data Category** or **Data Classification**, and then select the data category or data classification for the selected masking option. Note If the association between the data classification and a column is not accepted yet, the rule ignores the column.

Field	Description
	Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	a. Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. Tip This field contains the following options: ◦ **Show** ◦ **Hide** b. In the **rows where** field, select the data classification that you want to show or hide. c. In the **has** field, select the code set for the selected data classification. d. In the next field, select the code value for the selected code set. Tip You can add more data classifications for row-filtering by using the plus icon.

Tip

- The **Grant access to the data linked to these assets** checkbox, which is selected by default, is applicable to only Databricks and Snowflake. A selected checkbox indicates that you are allowing the selected groups to access those tables and columns in the database that are linked to the selected assets. If you do not want the selected groups to have this level of access, clear the checkbox.
- The **Summary** section shows a summary of the rule.



5. To preview the rule, click **Generate Preview**.

   Tip The preview shows only the first 1,000 affected columns. The drop-down list box below the **Generate Preview** button is used to switch between the assets that you selected in the rule. Each asset has its own preview table.

6. Click **Save Rule**.

   » A message appears stating that the rule is sent to source, and the rule is shown in the table on the **Data Access Rules** tab.

Modify a data access rule

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

  > Note If you have the the **Protect** > **Edit** global permission, you can modify only the data access rule that you created. If you have the **Protect** > **Administration** global permission, you can modify any data access rule.

- You have the **Catalog** global role. This role is required to view data classifications for selection in a data access rule.
- You have the permissions to view the assets that are associated with the data access rule. Otherwise, the **Unauthorized Asset** value is shown to you when you modify the rule.

Steps

1. Open Protect.
2. In the table, in the row containing the rule that you want to modify, click ✎.
   » The **Edit a Data Access Rule** dialog box appears.
3. Modify the required information.

   Details

Field	Description
Rule Name	Enter a name for the data access rule.
Optional: Description	Enter a description for the data access rule.

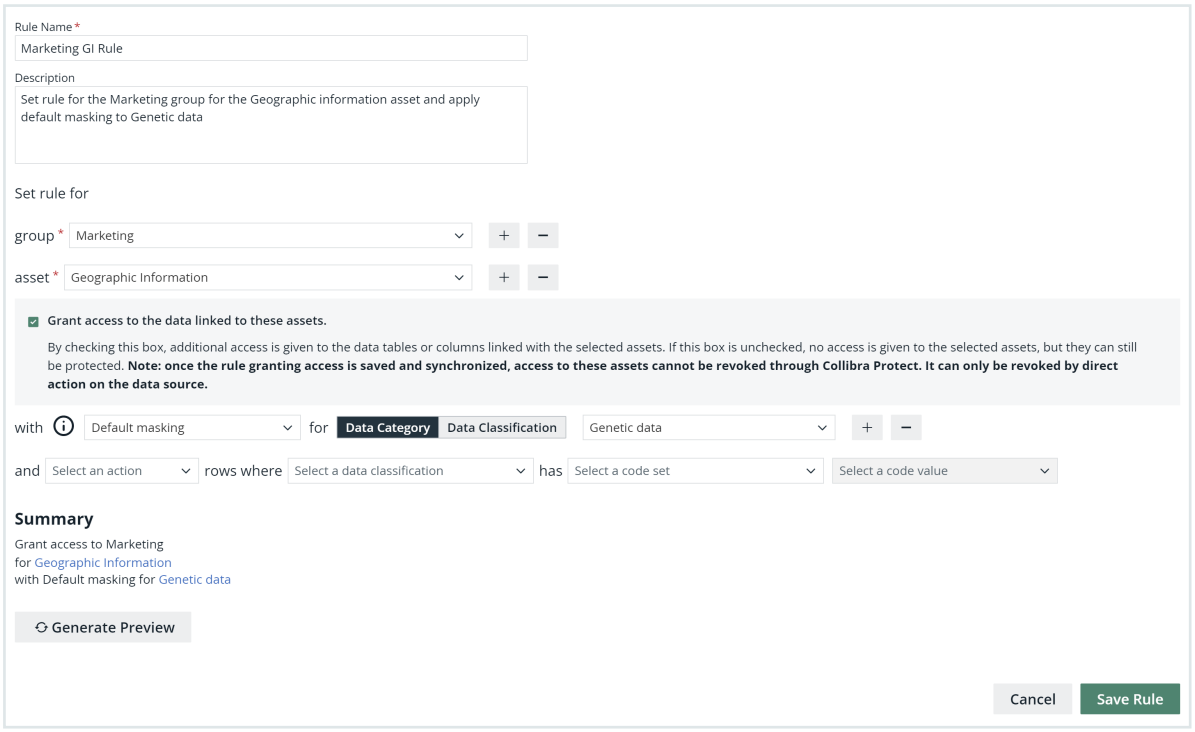
Field	Description
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.
Asset	Select the data asset that the rule is protecting. Tip ◦ This field contains Business Process, Data Category, and Data Set assets, in addition to assets of custom asset types. ◦ For more information, go to Technical background and Prescriptive paths. ◦ You can add more groups by using the plus icon.

Field	Description
Optional: With (masking option)	◦ Select the type of masking that you want to apply to a data category or data classification. Tip This field contains the following options: ■ **Default masking** ■ **Hashing** ■ **Show last** ■ **No masking** ◦ Click **Data Category** or **Data Classification**, and then select the data category or data classification for the selected masking option. Note If the association between the data classification and a column is not accepted yet, the rule ignores the column.

Field	Description
	Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	a. Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. Tip This field contains the following options: ◦ **Show** ◦ **Hide** b. In the **rows where** field, select the data classification that you want to show or hide. c. In the **has** field, select the code set for the selected data classification. d. In the next field, select the code value for the selected code set. Tip You can add more data classifications for row-filtering by using the plus icon.

> Tip
> - A selected checkbox indicates that you are allowing the selected groups to access those tables and columns in the database that are linked to the selected assets. If you do not want the selected groups to have this level of access, clear the checkbox.
> - The **Summary** section shows a summary of the rule.



4. To preview the rule, click **Generate Preview**.

> Tip The preview shows only the first 1,000 affected columns. The drop-down list box below the **Generate Preview** button is used to switch between the assets that you selected in the rule. Each asset has its own preview table.

5. Click **Save Rule**.
   » A message appears stating that the rule is sent to source, and the rule is shown in the table on the **Data Access Rules** tab.

Delete a data access rule

Requirements and permissions

You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

1. Open Protect.
2. Click the **Data Access Rules** tab.
3. In the table, in the row containing the rule that you want to delete, click .
   » The **Delete Data Access Rule** dialog box appears.
4. Click **Delete**.
   » A message appears stating that the request to delete the rule is received.

> Tip You can check the status of the rule in the **Synchronization Status** column in the table on the **Data Access Rules** tab.

Data Access Rules tab

The **Data Access Rules** tab in Protect contains an overview of data access rules. The **Recently Modified Rules** section on the tab shows the 5 last modified data access rules.

The following table describes the columns that are shown in the table on the **Data Access Rules** tab.

Column	Description
Rule Name	The name of the rule.
Synchronization Status	The status of synchronization between the rule in Protect and that in the data source.
Groups	The groups for which the rule is created.
Affected Assets	The assets that the rule protects. Tip Depending on your role, you can view the details of an asset by clicking the asset link in this column.
Owner	The name of the user who created the rule.
Created Date	The date and time when the rule was created.
Last Modified	The date and time when the rule was last modified.

Synchronization status

The following table describes the statuses that may be shown in the **Synchronization status** column on the **Data Access Rule** tab.

> Tip To view the status of the data access rule in the data source, go to the database of the data source provider.

Synchronization Status	Description
Active	The rule is enforced in the data source.
Pending	The rule is created or modified and is pending synchronization.
Failed	The synchronization of the rule has failed. Tip For more information about the error, click next to the status.
Delete Pending	The rule will be deleted during the next synchronization.
Not Deleted	The rule could not be deleted. Tip For more information about the error, click next to the status.

> Note Protect periodically synchronizes with your data source providers to update the status of the data access rules in Collibra, except if the status is **Failed**. For more information, go to Synchronization.

Data source policies (beta)

Data source policies are the policies that are native to a data source, for example, AWS Lake Formation data filters, BigQuery policy tags, and Snowflake masking policies. Data protection standards and data access rules created in Protect result in policies in the data sources. Protect enforces its standards and rules by creating and applying the data source policies on the physical data layer (tables and columns).

Import data source policies

Requirements and permissions

- You have the **Protect Author** or **Protect Admin** global role.
- The **Manage all resources** global permission is assigned to the **Edge site** global role.

Steps

You can import policies from your data source to Protect by using the Collibra Protect Data Source Policies API. The following is a template of a cURL command that you can use.

```
curl --location --request POST 'https://<collibra-environment-
url>/rest/protect/v1/policies/import' --header 'Authorization:
Basic <user:password encoded in base64>' --header 'Content-
Type: application/json' -d '{"databaseId": "<database-asset-
ID>"}' -v
```

> Note
> In the template:
>
> - Replace the placeholders indicated by "<>" with the actual values for your Collibra environment.
> - *database-asset-ID* refers to the ID of the database asset in Collibra that maps to the database in your data source.

Data Source Policies tab

The **Data Source Policies** tab contains an overview of the native data source policies. The table on the tab contains the policies that are active in the data source. These include both the policies that already exist in your data source and the policies that are automatically created by Protect in your data source.

The following table describes the columns that are shown in the table on the **Data Source Policies** tab.

Column	Description
Policy Name	The name of the policy in the data source.
Policy Logic	The logic that the data source uses to enforce the policy. For example, Snowflake runs an SQL script when you try to access protected data.
Tags	The names of the tags associated with the policy.
Data Source	The data source provider.

Data source providers

Collibra Protect periodically synchronizes with an aggregation of all data protection standards and data access rules. These standards and rules form a data source-agnostic representation containing all databases, schemas, tables, and columns, as well as their protections and accesses. The synchronization process then triggers the Edge capabilities, such as **Collibra Protect for Snowflake**. These Edge capabilities are responsible for translating the representation to actions toward the data source provider using their technology. This process might involve JDBC and REST calls to perform low-level operations to guarantee that the protections and accesses are applied.

Protect for AWS Lake Formation

To protect your AWS Lake Formation data, Protect uses AWS Lake Formation's permissions and data filters. The name of the data category or data classification selected in a data protection standard becomes an AWS Lake Formation tag (LF-tag) with the same name. The tag is then applied to all affected columns.

AWS Lake Formation policies

AWS Lake Formation protects your data by either granting access to or revoking access from one or more columns via permissions and data filters.

> Note AWS Lake Formation does not support data masking.

When you create a data protection standard or data access rule, one or more permissions and data filters are created in AWS Lake Formation. Each permission includes a data filter

to control access to data. Additionally, for a data protection standard, AWS Lake Formation tags (LF-tags) are created and assigned to columns.

> Note In this topic, the term *policies* refers to AWS Lake Formation permissions and data filters.

Data filters

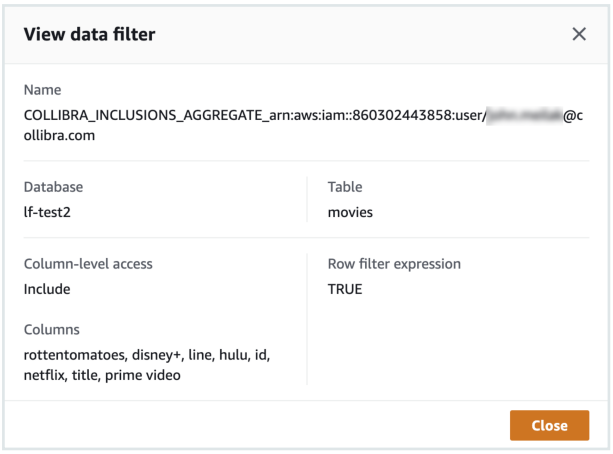
The following table contains the equivalent AWS Lake Formation data filter for a given Protect masking type.

Protect masking type	Equivalent AWS Lake Formation data filter
Default masking	Exclude
Hashing	Exclude
Show last	Exclude
No masking	Include

Each data filter belongs to a specific table in your AWS Data Catalog.

A data filter includes the following information:

- **Name:** The name of the data filter.
- **Table:** The name of the table whose columns are included or excluded.
- **Database:** The name of the database that contains the table.
- **Columns:** A list of columns to include or exclude in query results.
- **Column-level access:** The type of access—either include or exclude—for the columns.
- **Row filter expression:** An expression that specifies the rows to include in query results. The value **TRUE** indicates that all the rows in the table are shown.



> Note Protect safeguards your data in AWS Lake Formation by aggregating all the data protection standards and rules so that a single data filter is created in AWS Lake Formation per table per group. If multiple standards or rules exist for excluding columns, a single data filter with all the columns excluded is created. If a rule is then created for including columns, a data filter with all the columns included is created and the previously excluded columns are no longer considered.

Revoking existing policies for an effective data protection

To effectively protect your AWS Lake Formation data using Protect, you must first revoke any existing AWS Lake Formation policies. Data protection standards and access rules control access to tables and columns for IAM users by creating policies in AWS Lake Formation. To ensure that these policies work as intended, any previous policies granted to those users must be revoked.

> Example Suppose that Joe has full access to the **customers** table. If a data protection standard that hides PII is created and synchronized with AWS Lake Formation, policies are created for Joe. Those policies allow Joe only limited access to the **customers** table by excluding the PII columns. However, the policies will not work if Joe's existing full access to the **customers** table is not first revoked.

AWS Lake Formation group mapping

The Protect group mapping for AWS Lake Formation must follow the syntax for IAM identifiers.

Suppose that you want to create a Protect group named **Sales** that maps to the AWS IAM user **arn:aws:iam::000000000000:user/sales@example.com**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "AWSLakeFormation",
     "identity":
"arn:aws:iam::000000000000:user/sales@example.com"
    }
  ]
}
```

AWS Lake Formation permissions

To perform actions in AWS Lake Formation, Protect uses an AWS connection. This AWS connection must be configured with an AWS IAM user that has the following permissions on all the specified services.

```
{
    "Version": "2012-10-17",
    "Statement":
    [
        {
            "Effect": "Allow",
            "Action":
            [
                "athena:ListDataCatalogs",
                "athena:GetQueryExecution",
                "athena:StartQueryExecution",
                "cloudtrail:LookupEvents",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetTableVersions",
                "glue:GetTables",
                "lakeformation:AddLFTagsToResource",
                "lakeformation:CreateDataCellsFilter",
                "lakeformation:CreateLFTag",
                "lakeformation:DeleteDataCellsFilter",
                "lakeformation:DeleteLFTag",
                "lakeformation:GetLFTag",
                "lakeformation:GetResourceLFTags",
                "lakeformation:GrantPermissions",
                "lakeformation:ListDataCellsFilter",
                "lakeformation:ListLFTags",
                "lakeformation:ListPermissions",
                "lakeformation:RemoveLFTagsFromResource",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:GetBucketAcl"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Deny",
            "Action":
            [
                "lakeformation:PutDataLakeSettings"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS APIs

The following table explains the functions of the AWS APIs that are used by Protect for AWS Lake Formation.

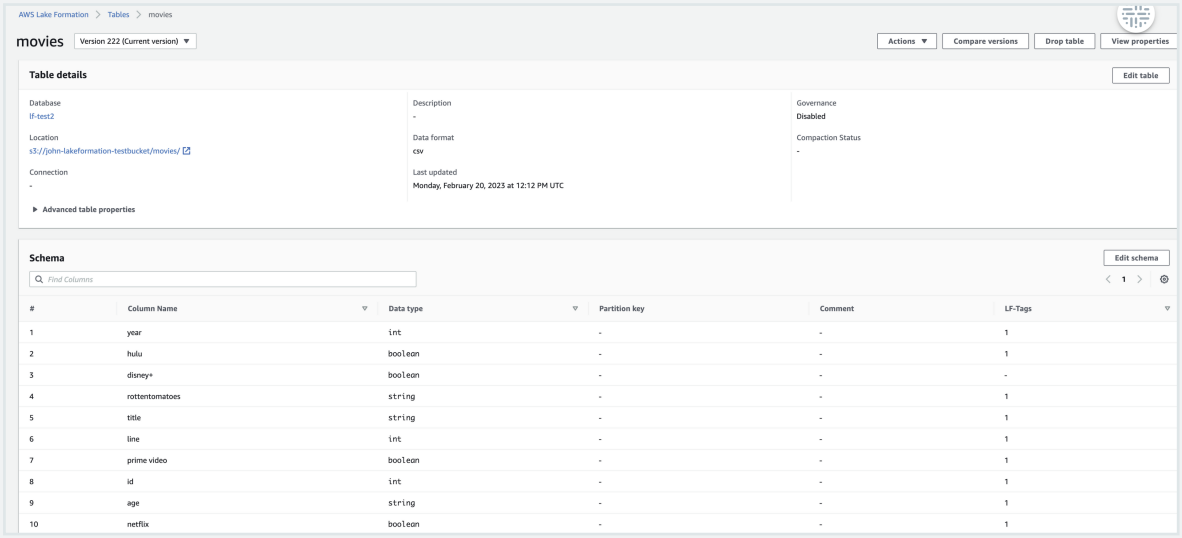
AWS API	Function
athena	Gets information from the AWS Glue Data Catalog. Note Catalog ingestion for AWS databases is performed by using the Amazon Athena service. However, not all the databases ingested from Athena are AWS Lake Formation databases. Hence, Protect needs to identify if a database ingested from Athena is also recognized by AWS Lake Formation. This can be achieved by making an API call to Athena's ListDataCatalogs.
cloudtrail	Shows the audit log in Protect.
glue	Gets a list of tables for a database.
lakeformation	• Creates, deletes, and lists an AWS Lake Formation tag (LF-tag). • Adds and removes an LF-Tag from a resource (column). • Creates, deletes, and lists data filters. • Adds and removes permissions from a resource (table).

AWS Lake Formation examples

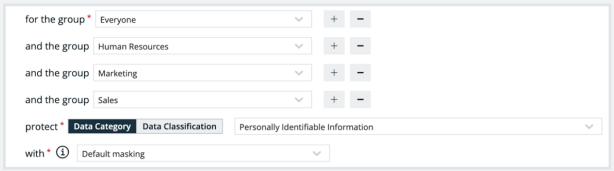
This topic contains examples of how AWS Lake Formation behaves with respect to certain data protection standards and data access rules.

Example

Suppose that a table named **movies** exists in AWS Lake Formation. This table contains Personally Identifiable Information (PII). The PII data category contains all the columns from **movies**, except for **age** and **year**.

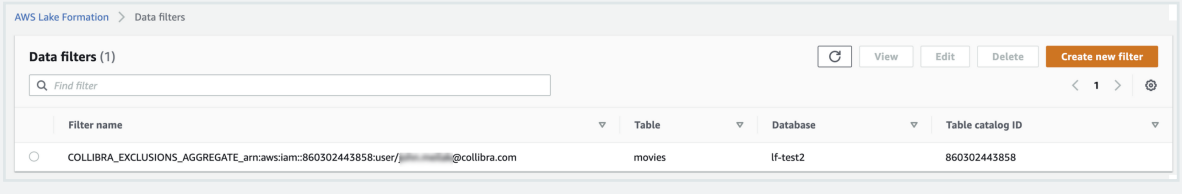


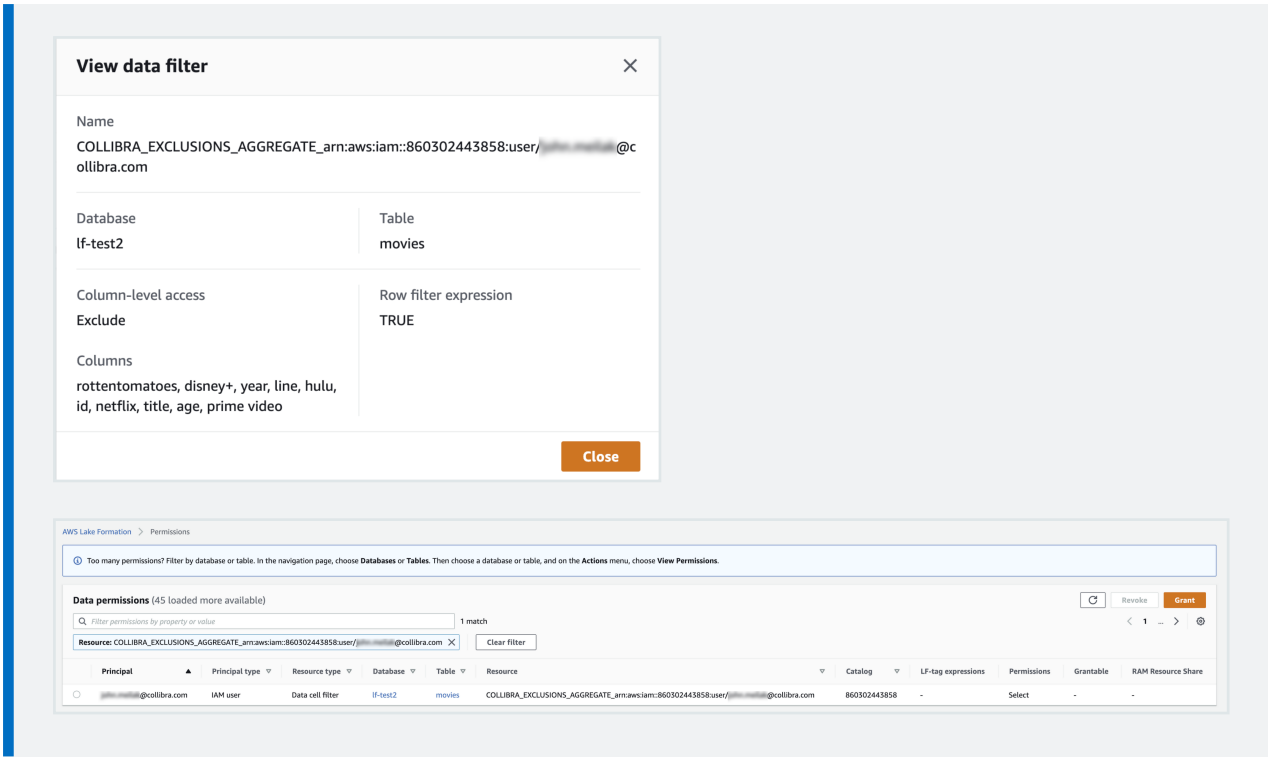
A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



Behavior

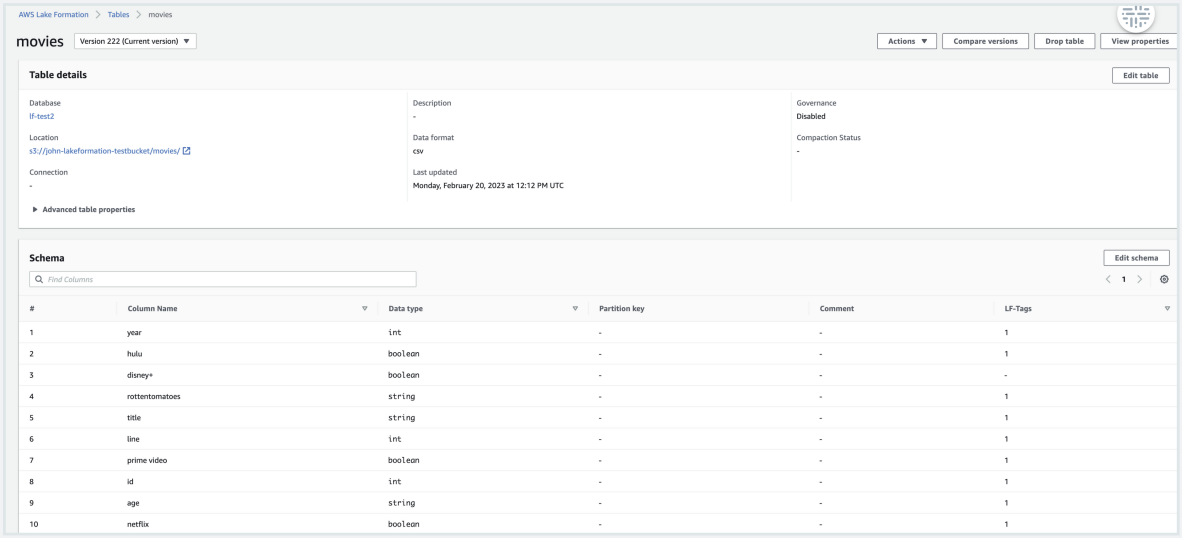
When the standard is synchronized and active, an exclusion data filter is created in AWS Lake Formation. This exclusion data filter hides all the PII columns from the specified groups. The exclusion data filter is named COLLIBRA_EXCLUSIONS_AGGREGATE_/<arn>.



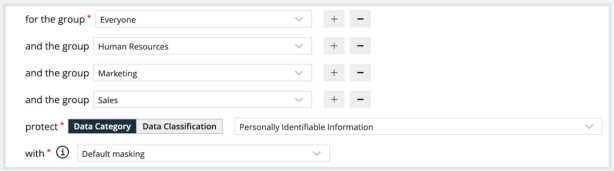

View data filter
Name
COLLIBRA_EXCLUSIONS_AGGREGATE_arn:aws:iam::860302443858:user/[redacted]@collibra.com
Database
lf-test2
Table
movies
Column-level access
Exclude
Row filter expression
TRUE
Columns
rottentomatoes, disney+, year, line, hulu, id, netflix, title, age, prime video
Close
AWS Lake Formation
Permissions
Too many permissions? Filter by database or table. In the navigation page, choose Databases or Tables. Then choose a database or table, and on the Actions menu, choose View Permissions.
Data permissions (45 loaded more available)
Revoke
Grant
Filter permissions by property or value
1 match
Resource: COLLIBRA_EXCLUSIONS_AGGREGATE_arn:aws:iam::860302443858:user/[redacted]@collibra.com
Clear filter
Principal
Principal type
Resource type
Database
Table
Resource
Catalog
LF-tag expressions
Permissions
Grantable
RAM Resource Share
[redacted]@collibra.com
IAM user
Data cell filter
lf-test2
movies
COLLIBRA_EXCLUSIONS_AGGREGATE_arn:aws:iam::860302443858:user/[redacted]@collibra.com
860302443858
-
Select
-
-

Example

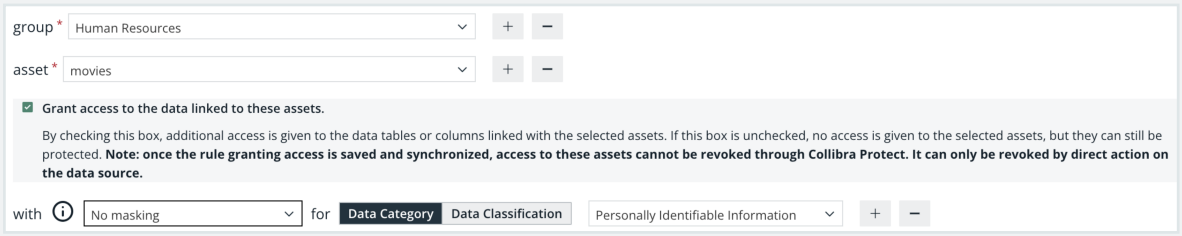
Suppose that a table named **movies** exists in AWS Lake Formation. This table contains Personally Identifiable Information (PII). The PII data category contains all the columns from **movies**, except for **age** and **year**.



A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.

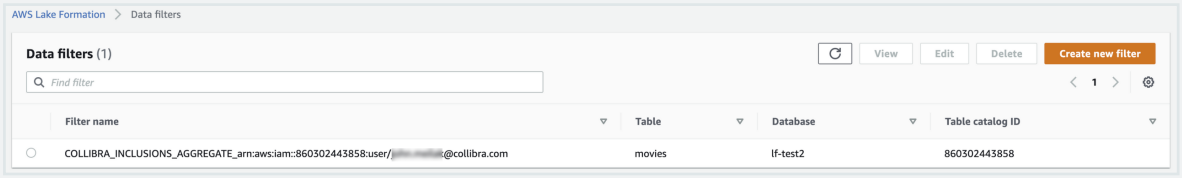


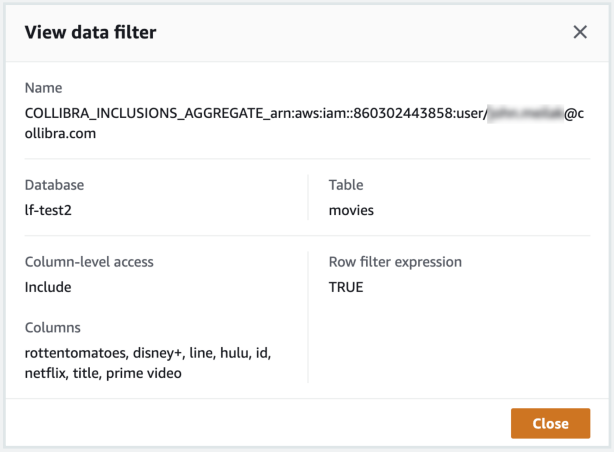
However, a rule that applies to the **Human Resources** group has been created. This rule does not require any masking for the PII columns in **movies**.



Behavior

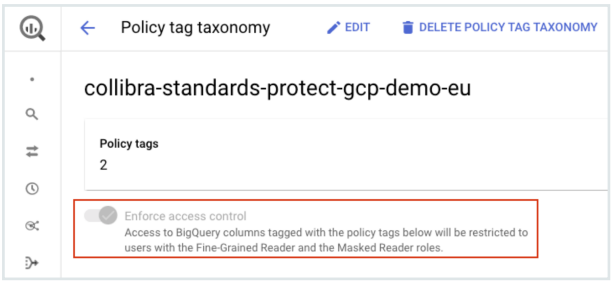
Because the rule takes priority over the standard, when the standard and the rule are synchronized and active, an inclusion data filter resulting from the rule is created in AWS Lake Formation, instead of an exclusion data filter resulting from the standard. This inclusion data filter shows all the PII columns in the **movies** table to the **Human Resources** group. The inclusion data filter is named COLLIBRA_INCLUSIONS_AGGREGATE_/<arn>.





Protect for BigQuery

To protect your BigQuery data, Protect uses Google's policy tags to create tags and assign the tags to the BigQuery columns. These tags control who can access the tagged data. Only the Protect groups specified in your data protection standards and data access rules can access the tagged BigQuery columns.



BigQuery masking rules

Each Protect masking type has an equivalent counterpart in BigQuery called a masking rule. As such, masking rules in BigQuery correspond to masking types in Protect.

> Note The BigQuery masking rules are not the same as the Protect data access rules.

The following table contains the equivalent BigQuery masking rule for a given Protect masking type.

Protect masking type	Equivalent BigQuery masking rule
Default masking	Default masking value
Hashing	Hash (SHA256) > Note BigQuery supports the **Hash (SHA256)** masking rule only for certain columns depending on their data types. If **Hash (SHA256)** cannot be applied to a certain column due to the data type of the column, the following masking rule is applied instead: **Default masking value**.
Show last	Default masking value > Note BigQuery does not support the **Show last** masking type. The **Show last** masking type is supported only on Snowflake.

Protect masking type	Equivalent BigQuery masking rule
No masking	Fine-Grained Reader Note Each Protect group to which you assign standards has an equivalent counterpart in BigQuery called a GCP principal. BigQuery grants the **Fine-Grained Reader** role to the assigned GCP principal to allow the GCP principal to view the data to which no masking is applied in Protect.

BigQuery data types

The following table contains the BigQuery masking rule that Protect supports for a given BigQuery data type.

Summary

- Protect supports the BigQuery **Default masking value** rule for all types of columns.
- Protect does not support the BigQuery **Nullify** rule for any type of column.
- Protect supports the BigQuery **Hash (SHA256)** rule only for the following types of columns: BYTES, STRING.

BigQuery data type	BigQuery masking rule supported by Protect
ARRAY	Default masking value
BIGNUMERIC	Default masking value
BOOL	Default masking value
BYTES	• Default masking value • Hash (SHA256)
DATE	Default masking value
DATETIME	Default masking value
FLOAT64	Default masking value

BigQuery data type	BigQuery masking rule supported by Protect
GEOGRAPHY	Default masking value
INT64	Default masking value
INTERVAL	Default masking value
JSON	Default masking value
NUMERIC	Default masking value
STRING	• Default masking value • Hash (SHA256)
STRUCT	Default masking value
TIME	Default masking value
TIMESTAMP	Default masking value

BigQuery group mapping

The Protect group mapping for BigQuery must follow the syntax for principal identifiers.

Suppose that you want to create a Protect group named **Sales** that maps to the BigQuery group email address **sales@example.com**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "GoogleBigQuery",
     "identity": "group:sales@example.com"
    }
  ]
}
```

BigQuery permissions

To perform actions in BigQuery, Protect uses a GCP connection. This GCP connection must be configured with a service account that has the following permissions.

- `bigquery.dataPolicies.create`
- `bigquery.dataPolicies.delete`
- `bigquery.dataPolicies.get`
- `bigquery.dataPolicies.getIamPolicy`
- `bigquery.dataPolicies.list`
- `bigquery.dataPolicies.setIamPolicy`
- `bigquery.dataPolicies.update`
- `bigquery.datasets.get`
- `bigquery.datasets.getIamPolicy`
- `bigquery.jobs.create`
- `bigquery.rowAccessPolicies.create`
- `bigquery.rowAccessPolicies.delete`
- `bigquery.rowAccessPolicies.list`
- `bigquery.rowAccessPolicies.setIamPolicy`
- `bigquery.rowAccessPolicies.update`
- `bigquery.tables.get`
- `bigquery.tables.getData`
- `bigquery.tables.list`
- `bigquery.tables.setCategory`
- `bigquery.tables.update`
- `datacatalog.categories.getIamPolicy`
- `datacatalog.categories.setIamPolicy`
- `datacatalog.taxonomies.create`
- `datacatalog.taxonomies.get`
- `datacatalog.taxonomies.list`
- `datacatalog.taxonomies.update`
- `logging.logEntries.list`
- `resourcemanager.projects.get`

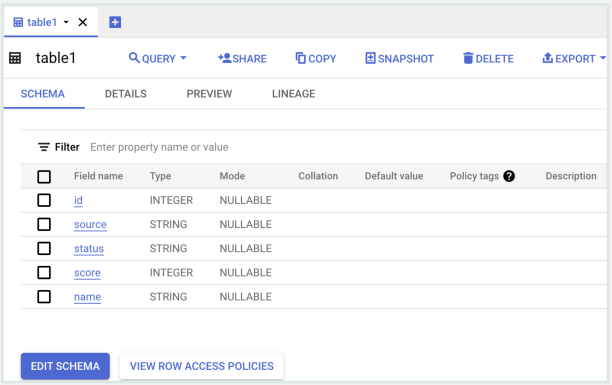
In addition, ensure that the following APIs are enabled for the GCP projects used by Protect:

- BigQuery API
- BigQuery Data Policy API
- Google Cloud Data Catalog API
- Cloud Logging API

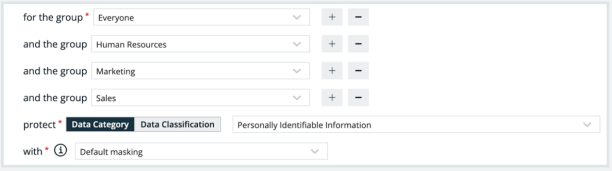
BigQuery examples

This topic contains examples of how BigQuery behaves with respect to certain data protection standards and data access rules.

Example
Suppose that a table named **table1** exists in BigQuery. This table contains Personally Identifiable Information (PII). The PII data category contains all the columns from **table1**.

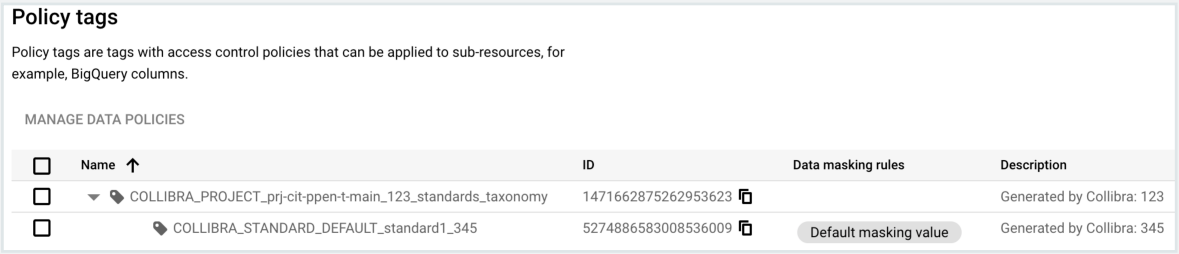


A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.

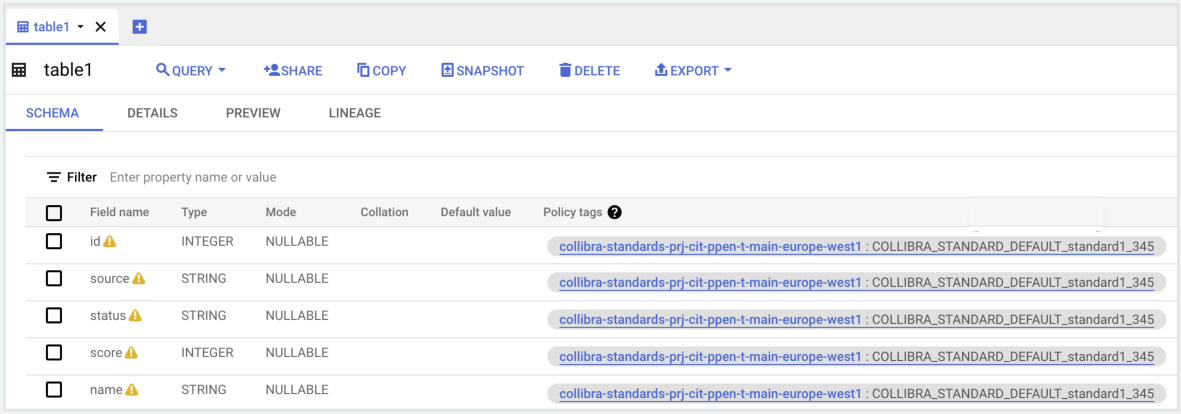


Behavior

When the standard is synchronized and active, a standard policy tag is created in BigQuery's taxonomy. The standard policy tag is named COLLIBRA_STANDARD_DEFAULT_<data protection standard name><data protection standard ID>.

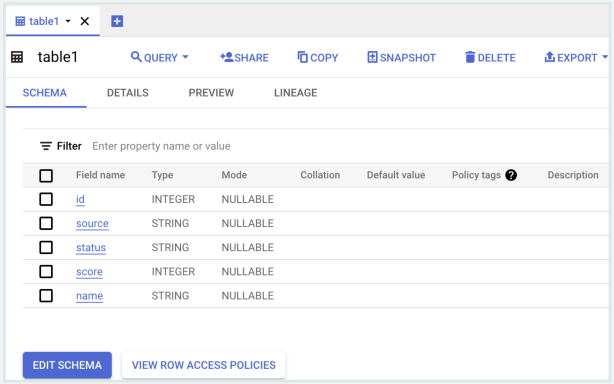


The following image shows how the policy tags are applied to the columns in **table1**.

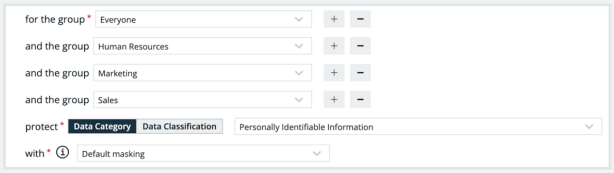


All the columns are assigned the same standard policy tag and are protected by default masking because they belong to the PII data category (selected in the standard).

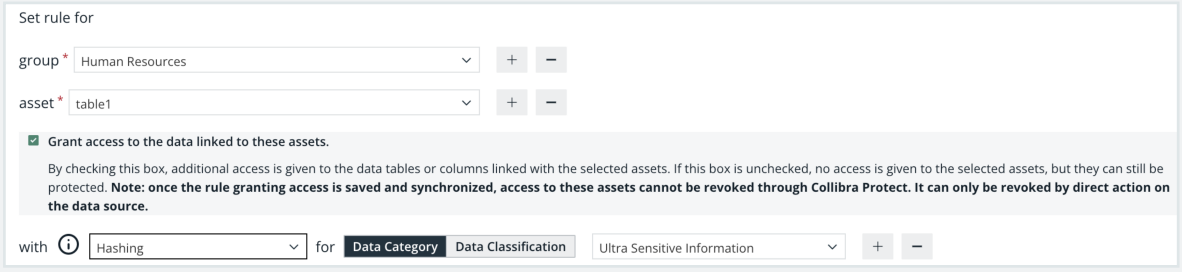
Example
Suppose that a table named **table1** exists in BigQuery. This table contains Personally Identifiable Information (PII) and Ultra Sensitive Information (USI). The PII data category contains all the columns from **table1**, except for **id** and **source**. The USI data category contains only the **status** column.



A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



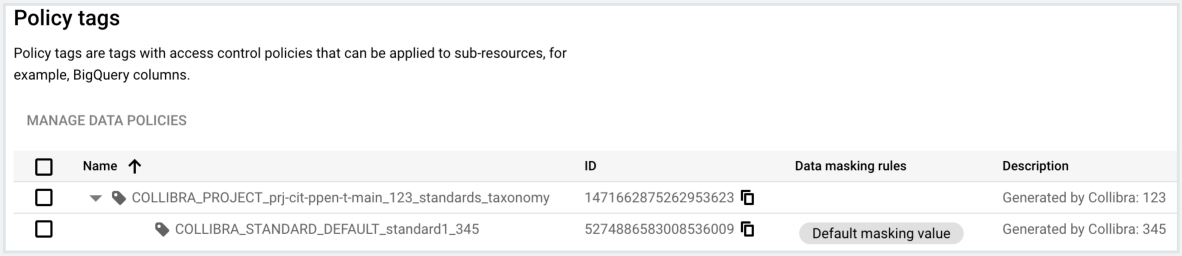
However, a rule that applies to the **Human Resources** group has been created. This rule requires hashing for the USI columns in **table1**.

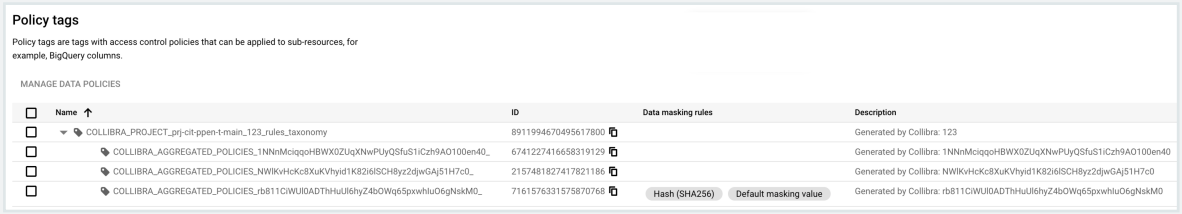


Behavior

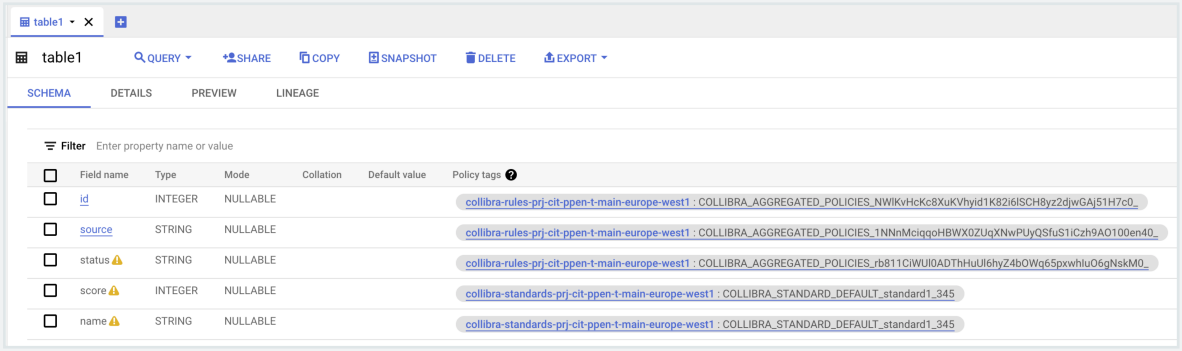
When the standard and rule are synchronized and active, policy tags are created in BigQuery's taxonomy. The standard policy tag is named COLLIBRA_STANDARD_DEFAULT_<data protection standard name><data protection standard ID>. The

rule policy tag is named COLLIBRA_AGGREGATED_POLICIES_ <rulesaccesshash>.





The following image shows how the policy tags are applied to the columns in **table1**.



- The **id** and **source** columns do not belong to the PII data category (selected in the standard) or the USI data category (selected in the rule). Therefore, they are not protected by either the standard or the rule. However, they are still assigned a rule policy tag with the Fine-Grained Reader access to allow users to view the original data.
- The **name** and **score** columns belong to the PII data category (selected in the standard). They are assigned the same standard policy tag and are protected by default masking.
- The **status** column belongs to both the PII data category (selected in the standard) and the USI data category (selected in the rule). Because the rule takes priority over the standard, the **status** column is assigned only the rule policy tag and is protected by hashing.

Protect for Databricks

To protect your Databricks data, Protect uses Databricks's column-based masking functions. These masking functions are applied to columns to enforce data protection.

> Note The Databricks functions are in the public preview stage. Protect for Databricks will be generally available (GA) whenever the Databricks functions reach the GA stage.

Databricks policies

Databricks has the following types of policies:

- Column-based
- Row-based

Each of these policy types can be created either in Protect or on Databricks.

Data access standards created in Protect result in column-based policies on Databricks. Column-based policies are applied directly to the columns on Databricks.

Row filters in data access rules result in row-based policies on Databricks. Row-based policies are applied to the tables on Databricks.

Databricks data types

Databricks provides several functions to transform the data. This topic describes how Databricks transforms the data for a given Protect masking type.

- **Default masking:** Databricks does not support this masking type. Protect, however, uses the default masking type to apply protection to a wide range of data types. A default masking value is applied to each column according to the data type of the column.
  Default masking values for data types

Column data type	Databricks data type	Default masking value
NUMERIC	BIGINT	bigint('0')
BIGNUMERIC	BIGINT	bigint('0')
BYTEINT	BIGINT	bigint('0')
BIGINT	BIGINT	bigint('0')
BINARY	BINARY	binary('00')
VARBINARY	BINARY	binary('00')
BYTES	BINARY	binary('00')
BOOLEAN	BOOLEAN	false
DATE	DATE	1970-01-01
DATETIME	DATE	1970-01-01
DECIMAL	DECIMAL(p,s)	decimal('0.0')
DOUBLE	DOUBLE	double('0.0')
DOUBLE PRECISION	DOUBLE	double('0.0')
REAL	DOUBLE	double('0.0')
FLOAT	FLOAT	float('0.0')
FLOAT4	FLOAT	float('0.0')
FLOAT8	FLOAT	float('0.0')
INT	INT	int('0')
NUMBER	NUMBER	int('0')

Column data type	Databricks data type	Default masking value
BIT	INT	int('0')
INTEGER	INT	int('0')
SMALLINT	SMALLINT	smallint('0')
STRING	STRING	mask('S','*')
CHAR	STRING	mask('S','*')
CHARACTER	STRING	mask('S','*')
VARCHAR	VARCHAR	mask('S','*')
TEXT	STRING	mask('S','*')
TIMESTAMP	TIMESTAMP	1970-01-01 00:00:00.000
TIME	TIMESTAMP	1970-01-01 00:00:00.000
TIMESTAMP_NTZ	TIMESTAMP	1970-01-01 00:00:00.000
TIMESTAMP_LTZ	TIMESTAMP	1970-01-01 00:00:00.000
TIMESTAMP_TZ	TIMESTAMP	1970-01-01 00:00:00.000
TINYINT	TINYINT	tinyint('0')
ARRAY	ARRAY <elementType >	array()
MAP	MAP < keyType,valueType >	map()

Column data type	Databricks data type	Default masking value
STRUCT	STRUCT < [fieldName : fieldType [NOT NULL] [COMMENT str][, ...]] >	struct(0) or struct(0,0) Tip The dynamic value depends on how many fields are defined for the STRUCT datatype.

- **Hashing:** Uses the following Databricks functions:
  - `SHA2` (for strings)
  - `HASH` (for numbers)
  - `right(hash(value), (precision - scale))` (for decimals)
- **Show last:** Uses the following expressions:
  - `right(value,n)` (for strings)
  - `mod(value, cast(power(10,n) AS INT))` (for integers)
  - `regexp_replace(substr(string(value), length(value) - (n-1), n), '^$', '0')` (for floating-point numbers and decimals)

    Tip In the expressions, *value* indicates the content and *n* indicates the number of characters to be shown.
- **No masking:** Returns the raw content.

Note
- You can apply the **Hashing** and **Show last** masking types to only the following Databricks data types: BIGINT, DECIMAL, DOUBLE, FLOAT, INT, SMALLINT, STRING, and TINYINT.
- If a selected masking type cannot be applied to a certain data type–for example, when you attempt to apply the **Hashing** masking type to the DATE data type–the **Default masking** type is applied to the data type to guarantee protection.

Databricks group mapping

The Protect group mapping for Databricks must follow the syntax for principals.

Suppose that you want to create a Protect group named **Sales** that maps to the Databricks group **SALES**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "Databricks",
     "identity": "SALES"
    }
  ]
}
```

Databricks privileges

To perform actions in Databricks, Protect uses an Edge connection. This Edge connection must be configured with a role that is the owner of the catalog or schema in Databricks.

Databricks examples

This topic contains examples of how Databricks behaves with respect to certain data protection standards and data access rules.

Example
Suppose that:

- The **Personally Identifiable Information** (PII) and **Personal Information** (PI) data categories exist in Databricks. These two data categories contain a column named **DOB**.
- A standard that applies to the **HR** group has been created. This standard requires hashing for the PII data category.
- A standard that applies to the **Marketing** group has been created. This standard requires default masking for the PI data category.

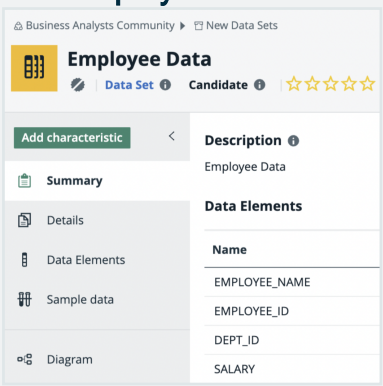
Behavior

When the standards are synchronized and active, a function is created in Databricks for each standard and linked to the **DOB** column. A single column masking policy that combines the two policies is then created and applied to the **DOB** column. This column masking policy includes the protection defined in each standard.

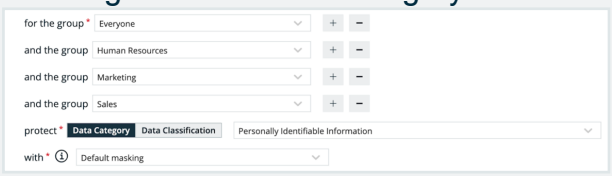
```
CASE
  WHEN (
    current_user() == 'HR'
    or is_account_group_member('HR')
  ) THEN hash(val)

WHEN (
    current_user() == 'Marketing'
    or is_account_group_member('Marketing')
  ) THEN 0
  ELSE val
END
```

Example
Suppose that:

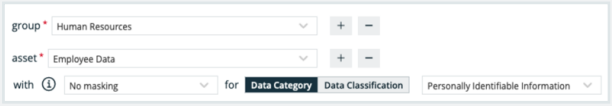
- The **Personally Identifiable Information** (PII) data category exists in Databricks.
- The **Employee Data** data set exists in Databricks. This data set contains PII.



- A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



- A rule that applies to the **Human Resources** group has been created. This rule does not require any masking for the PII columns in the **Employee Data** asset.



Behavior

When the standard is synchronized and active, masking policies are created in Databricks—one policy for each column. The masking functions are named collibra_masking_policy_<asset ID>.

Column	Type	Comment	Tags	Mask
EMPLOYEE_NAME	string			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_9d293821_f1fa_4564_bc20_6fb33331256c
EMPLOYEE_ID	int			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_f0cc1791_5238_4404_9314_ab13f226c605
DEPT_ID	int			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_14cbeca4_0c58_42a2_944b_88838469d140
SALARY	decimal(10,0)			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_cfd4ff71_6940_4736_b2d4_8cbc50e51a5b

The following image shows a masking policy for the STRING data type. The data that is shown in the policy depends on the masking type selected in the standard and rule. In the policy, `val` indicates the value as it is stored in the table.

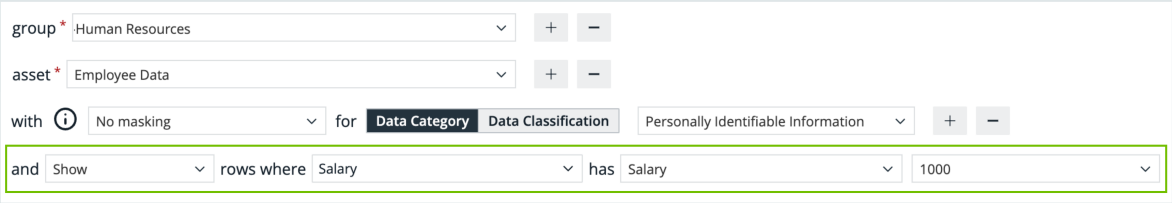
```
  WHEN (
    current_user() == 'HR'
    or is_account_group_member('HR')
  ) THEN val
  WHEN (
    current_user() == 'Everyone'
    or is_account_group_member('Everyone')
  ) THEN mask('S', '*')
  WHEN (
    current_user() == 'Marketing'
    or is_account_group_member('Marketing')
  ) THEN mask('S', '*')
  WHEN (
    current_user() == 'Sales'
    or is_account_group_member('Sales')
  ) THEN mask('S', '*')
  ELSE val
```

According to the standard, the **Everyone**, **Human Resources**, **Marketing**, and **Sales** groups have masked access to the data. However, according to the rule, the **Human Resources** group has unmasked access to the data. As a result:

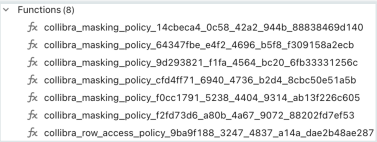
- The column is not masked for the **Human Resources** group.
- The column is masked for the **Everyone**, **Marketing**, and **Sales** groups.

Example

Consider the above example with the row filter added, as shown in the following image.



Behavior




```
CREATE
OR REPLACE FUNCTION protect_dev_catalog.tpch_dev.COLLIBRA_
ROW_ACCESS_POLICY_9ba9f188_3247_4837_a14a_dae2b48ae287
(SALARY decimal(10, 0)) RETURN IF(
  (
    (
      current_user() == 'HR'
      or is_account_group_member('HR')
    )
    and SALARY IN (1000)
  ),
  true,
  false
)
```

The row access functions are named collibra_row_access_policy_<asset ID>. The masking and row access policy functions are created at the schema level in Databricks.

Note Protect for Databricks supports Databricks external tables.

Protect for Snowflake

To protect your Snowflake data, Protect uses Snowflake's tag-based masking policies. The name of the data category or data classification selected in a data protection standard becomes a tag with the same name. The tag is then applied to all affected columns to enforce data protection.

Snowflake policies

Snowflake has the following types of policies:

- Column-based
- Row-based
- Tag-based

Each of these policy types can be created either in Protect or on Snowflake.

Data access rules created in Protect result in column-based policies on Snowflake. Column-based policies are applied directly to the columns on Snowflake.

Row filters in data access rules result in row-based policies on Snowflake. Row-based policies are applied to the tables on Snowflake.

Data protection standards created in Protect result in tag-based policies on Snowflake. The tags are subsequently applied to the columns on Snowflake.

Snowflake data types

Snowflake provides several functions to transform the data. This topic describes how Snowflake transforms the data for a given Protect masking type.

- **Default masking:** Snowflake does not support this masking type. Protect, however, uses the default masking type to apply protection to a wide range of data types. A default masking value is applied to each column according to the data type of the column.

Default masking values for data types

Column data type	Snowflake data type	Default masking value
NUMBER	NUMBER	0
DECIMAL	NUMBER	0
NUMERIC	NUMBER	0
INT	NUMBER	0
INTEGER	NUMBER	0
BIGINT	NUMBER	0
SMALLINT	NUMBER	0
TINYINT	NUMBER	0
BYTEINT	FLOAT	0
FLOAT	FLOAT	0
FLOAT4	FLOAT	0
FLOAT8	FLOAT	0
DOUBLE	FLOAT	0

Column data type	Snowflake data type	Default masking value
DOUBLE PRECISION	FLOAT	0
REAL	FLOAT	0
VARCHAR	VARCHAR	*
CHAR	VARCHAR	*
CHARACTER	VARCHAR	*
STRING	VARCHAR	*
TEXT	VARCHAR	*
BINARY	BINARY	00
VARBINARY	BINARY	00
BOOLEAN	BOOLEAN	false
DATE	DATE	1970-01-01
DATETIME	TIMESTAMP_NTZ	1970-01-01 00:00:00.000
TIME	TIME	00:00:00
TIMESTAMP	TIMESTAMP_NTZ	1970-01-01 00:00:00.000
TIMESTAMP_LTZ	TIMESTAMP_LTZ	1969-12-31 16:00:00.000-0800 Note This may change depending on the time zone.

Column data type	Snowflake data type	Default masking value
TIMESTAMP_NTZ	TIMESTAMP_NTZ	1970-01-01 00:00:00.000
TIMESTAMP_TZ	TIMESTAMP_TZ	1969-12-31 16:00:00.000-0800 Note This may change depending on the time zone.
VARIANT	VARIANT	0
OBJECT	OBJECT	{}
ARRAY	ARRAY	[]
GEOGRAPHY	GEOGRAPHY	{"coordinates": [0,0],"type": "Point"} (aka point(0, 0) and visualization can change based on user preferences)

- **Hashing:** Uses the following Snowflake functions:
  - `SHA2` (for strings)
  - `HASH` (for numbers)
- **Show last:** Uses the following expressions:
  - `substr(to_varchar(value), length(value) - n, n)` (for strings)
  - `mod(value, power(10,n))` (for numbers)
  - Tip In the expressions, *value* indicates the content and *n* indicates the number of characters to be shown.
- **No masking:** Returns the raw content.

> Note
> - You can apply the **Hashing** and **Show last** masking types to only the following Snowflake data types: FLOAT, NUMBER, and STRING.
> - If a selected masking type cannot be applied to a certain data type–for example, when you attempt to apply the **Hashing** masking type to the DATE data type–the **Default masking** type is applied to the data type to guarantee protection.

Snowflake group mapping

The Protect group mapping for Snowflake must follow the syntax for identifiers.

Suppose that you want to create a Protect group named **Sales** that maps to the Snowflake role **SALES**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "Snowflake",
     "identity": "SALES"
    }
  ]
}
```

Snowflake privileges

To perform actions in Snowflake, Protect uses an Edge connection. This Edge connection must be configured with a role that has the following privileges in Snowflake.

Snowflake privilege	Description
`[APPLY MASKING POLICY]`	To apply masking policies. Required for the role performing the actions.

Snowflake privilege	Description
`[APPLY ROW ACCESS POLICY]`	To apply row access policies. Required for the role performing the actions.
`[APPLY TAG]`	To apply tags. Required for the role performing the actions.
`[IMPORTED PRIVILEGES]`	To import privileges. Required for the role performing the actions.
`[MANAGE GRANTS]`	To manage access privileges. Required for the role performing the actions.
`[USAGE]`	To manage usage access on databases and schemas involved in the protection. Required on each database and schema where policies are applied to the role performing the actions.
`[CREATE MASKING POLICY]`	To create masking policies. Required on each schema where policies are applied to the role performing the actions.
`[CREATE ROW ACCESS POLICY]`	To create row access policies. Required on each schema where policies are applied to the role performing the actions.
`[CREATE TAG]`	To create tags. Required on each schema where policies are applied to the role performing the actions.

Example Suppose that a role named **PROTECT** exists in Snowflake and this role is responsible for managing access privileges on all schemas within a database named **DEMO**. To enable the Snowflake **PROTECT** role to perform an action in Snowflake, the following statements can be used.

```
GRANT APPLY MASKING POLICY ON ACCOUNT TO ROLE PROTECT;
GRANT APPLY ROW ACCESS POLICY ON ACCOUNT TO ROLE PROTECT;
GRANT APPLY TAG ON ACCOUNT TO ROLE PROTECT;
GRANT MANAGE GRANTS ON ACCOUNT TO ROLE PROTECT;
GRANT IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE TO ROLE
PROTECT;
GRANT USAGE ON DATABASE DEMO TO ROLE PROTECT;
GRANT USAGE ON ALL SCHEMAS IN DATABASE DEMO TO ROLE
PROTECT;
GRANT CREATE MASKING POLICY ON ALL SCHEMAS IN DATABASE DEMO
TO ROLE PROTECT;
GRANT CREATE ROW ACCESS POLICY ON ALL SCHEMAS IN DATABASE
DEMO TO ROLE PROTECT;
GRANT CREATE TAG ON ALL SCHEMAS IN DATABASE DEMO TO ROLE
PROTECT
```

Snowflake examples

This topic contains examples of how Snowflake behaves with respect to certain data protection standards and data access rules.

Example
Suppose that:

- The **Personally Identifiable Information** (PII) and **Personal Information** (PI) data categories exist in Snowflake. These two data categories contain a column named **DOB**.
- A standard that applies to the **HR** group has been created. This standard requires hashing for the PII data category.
- A standard that applies to the **Marketing** group has been created. This standard requires default masking for the PI data category.

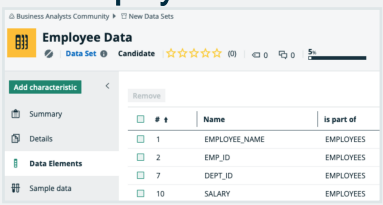
Behavior

When the standards are synchronized and active, a tag policy is created in Snowflake for each standard and linked to the **DOB** column. A single column masking policy that combines the two tag policies is then created and applied to the **DOB** column. This column masking policy includes the protection defined in each standard.

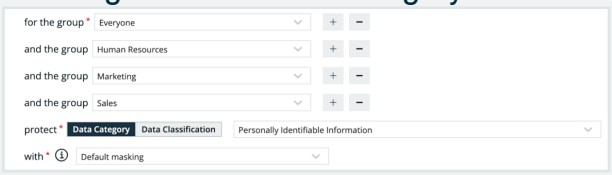
```
CASE
        WHEN CURRENT_ROLE() = 'HR' THEN hash(val)::NUMBER
        WHEN CURRENT_ROLE() = 'MARKETING' THEN 0
        ELSE val
    END
```

Example
Suppose that:

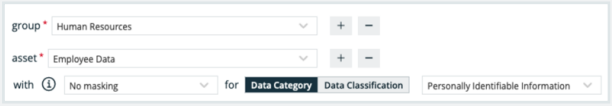
- The **Personally Identifiable Information** (PII) data category exists in Snowflake.
- The **Employee Data** data set exists in Snowflake. This data set contains PII.



- A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



- A rule that applies to the **Human Resources** group has been created. This rule does not require any masking for the PII columns in the **Employee Data** asset.



Behavior

Standard

When the standard is synchronized and active, 14 masking policies are created in Snowflake–one policy for each Snowflake data type. These masking policies are associated with the **Personally Identifiable Information** tag and are created at the schema level. The tag is assigned to those columns that need to be protected. The masking policies are named COLLIBRA/MASKING_POLICY/<asset ID>/<Snowflake type>.

Results Data Preview

Query ID SQL 84ms 18 rows

Filter result... Copy

Row	created_on	name	database_name	schema_name	kind	owner
1	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/ARRAY	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
2	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/BINARY	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
3	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/BOOLEAN	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
4	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/DATE	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
5	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/FLOAT	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
6	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/GEOGRAPHY	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
7	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/NUMBER	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
8	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/OBJECT	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
9	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/STRING	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
10	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIME	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
11	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIMESTAMP	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
12	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIMESTAMP_LTZ	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
13	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIMESTAMP_TZ	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
14	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/VARIANT	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN

At runtime, Snowflake fetches the right masking policy based on the column data type.

```
SHOW TAGS;
```

Results Data Preview

Query ID SQL 48ms 2 rows

Filter result... Copy

Row	created_on	name	database_name	schema_name	owner	comment
1	2022-09-06 03:46:10.054...	Personally Identifiable Information	PROTECT_QA	DEMO	ACCOUNTADMIN	Generated by Collibra: 28d226cc-0ab0-4d23-b912-985312fb36b1

The following image shows a masking policy for the STRING data type. The data that is shown in the policy depends on the masking type selected in the standard. In the policy, `val` indicates the value as it is stored in the table.

Details

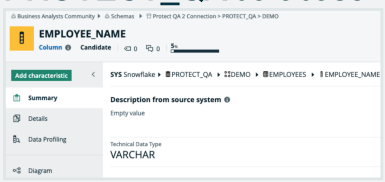
```
CASE
        WHEN CURRENT_ROLE() = 'PUBLIC' THEN '*'
        WHEN CURRENT_ROLE() = 'HR' THEN '*'
        WHEN CURRENT_ROLE() = 'MARKETING' THEN '*'
        WHEN CURRENT_ROLE() = 'SALES' THEN '*'
        ELSE val
    END
```

Rule

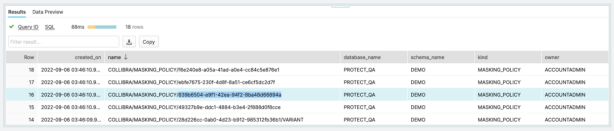
A rule results in a combination of grant instructions, dynamic masking, and row access policies.

The rule grants access of the **Employee Data** data set to the **Human Resources** group, as indicated by the selected **Grant access...** checkbox in the rule. Then, the corresponding Snowflake role for the group can access each database, schema, and table in the data set. In addition, the column masking policy is applied to those columns that need to be protected.

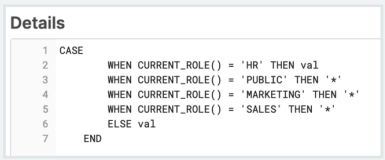
Consider the **EMPLOYEE_NAME** column in the **Employee Data** data set. This column belongs to the **EMPLOYEES** table within the **DEMO** schema in the **PROTECT_QA** database.



In Snowflake, each column that is categorized as PII within the **Employee Data** dataset inherits the masking policy that is applied to the column in Protect. The masking policies created at the schema level are named COLLIBRA/MASKING_POLICY/<asset ID>.



The following image shows the masking policy created for the **EMPLOYEE_NAME** column.



Summary

According to the standard, the **Everyone**, **Human Resources**, **Marketing**, and **Sales** groups have masked access to the data. However, according to the rule, the **Human Resources** group has unmasked access to the data. As a result, the **EMPLOYEE_NAME** column has both a policy tag and a column masking policy applied to it via the standard and the rule, respectively.

In Snowflake, if both a policy tag and a column masking policy exist for the same column, the column masking policy takes priority and the policy tag is not assigned to the column. To ensure that the protection defined in the standard is not ignored, the column masking policy also considers the conditions defined in the standard (policy tag).

Thus, when a standard is created for the **Human Resources**, **Marketing**, and **Sales** groups to mask a column, and when a rule is created for the **Human Resources** group to not mask the same column, the result is as follows:

- The column is not masked for the **Human Resources** group.
- The column is masked for the **Marketing** and **Sales** groups.

Protect audit (beta)

An audit log in Protect contains information about the queries that were run to access the data and the data that was accessed.

This topic describes how to generate an audit log for Protect and what is shown in an audit log.

> Tip
> The information in this topic varies depending on the data source that you select.
>
> Data source
>
> 

Generate an audit log

You can generate an audit log of access records from the data source on the **Audit** page.

> Note The time that it takes for the actions performed in a data source to appear in an audit log in Protect varies from several minutes to hours, depending on the data source.

Requirements and permissions

You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

1. Open Protect.
2. Click the **Audit** tab.

3. Click **BigQueryDatabricksLake FormationSnowflake**.
4. In the **AWS Region** field, select the hosting region for your Amazon Web Services.
5. Click one of the following buttons: **Today**, **Yesterday**, **A week ago**, **30 days ago**.

   Tip The start date corresponding to the button that you clicked is shown in the **Start Date** field. Alternatively, you can enter or select a date in the **Start Date** field.

6. Click **Generate Log**.
   » The audit log is generated.

Important
- The generation of an audit log may take up to a minute. After clicking **Generate Log**, do not navigate away from the **Audit** page because doing so cancels the audit log generation.
- The audit log contains the first 1,000 records from the selected start date. If you want to view the remaining records, contact your data source administrator.

Audit log data

The following table describes the columns that are shown in an audit log.

AWS Lake FormationDatabricksBigQuerySnowflake

Column	Description
Query ID	The ID of the query in Snowflake.
Query Start Time	The date and time of the query in Snowflake.
Source User Name	The name of the user in Snowflake who ran the query to access the data.
Direct Objects Accessed	The database object (a table or a view) that was used to access the data.
Base Objects Accessed	The database object that was accessed.

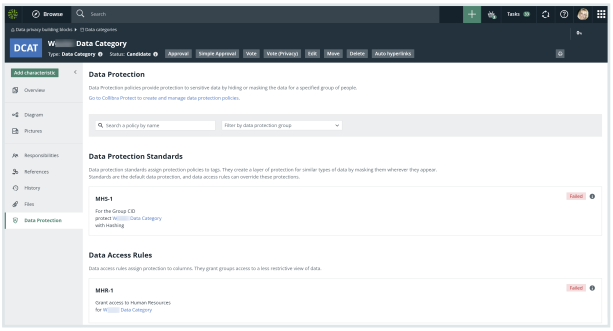
Column	Description
Event Name	The name of the event in AWS Lake Formation.
Date	The date and time of the event in AWS Lake Formation.
Source User Name	The name of the user in AWS Lake Formation who ran the event to access the data.
Event Source	The source of the event, for example, AWS Athena.
Resources	The resources that were accessed.
Method Name	The name of the method in BigQuery.
Date	The date and time of the method in BigQuery.
Principal	The name of the user in BigQuery who ran the method to access the data.
Resource Name	The resource that was accessed.
Action Name	The name of the action in Databricks.
Objects Accessed	The objects that were used to access the data.
Email	The email address of the user in Databricks who ran the action to access the data.
Query Start Time	The date and time of the action in Databricks.

Asset data protection

The asset pages for the following asset types contain the **Data Protection** tab to allow you to view, filter, create, and manage data protection standards and data access rules:

- Business Process
- Data Category
- Data Set
- Custom asset types such as Column, Database, Schema, and Table, derived from the aforementioned asset types via prescriptive paths

> Note Data protection standards support only Data Category assets and data classifications.



View or filter standards and rules

Requirements and permissions

You have the **Protect Reader** global role.

Steps

On the asset page (for the one of the aforementioned asset types), click the **Data Protection** tab.

» Data protection standards and data access rules that are linked to the asset are shown.

Tip
- To filter the standards and rules by name, in the **Search a policy by name** field, enter the name of the standard or rule that you want to view.
- To filter the standards and rules by group, in the **Filter by data protection group** field, select the group for which you want to view the standard or rule.

Create or manage standards and rules

Requirements and permissions

You have the **Protect Author** and **Protect Admin** global roles.

Steps

1. On the asset page (for the one of the aforementioned asset types), click the **Data Protection** tab.
2. Click the following link: **Go to Collibra Protect to create and manage data protection policies.**

Tip For information about how to create and manage data protection standards and data access rules, go to Data Protection Standards tab and Data Access Rules tab.

Chapter 13

Why certain standards and rules fail

Certain data protection standards or data access rules may fail due to logical errors. This section describes some of the common scenarios that cause them to fail.

Different types of masking affecting the same column

This topic contains examples to describe how data protection standards and data access rules behave when different types of masking affect the same column.

> Note In the topic, the term *agent* refers to a data category or a data classification.

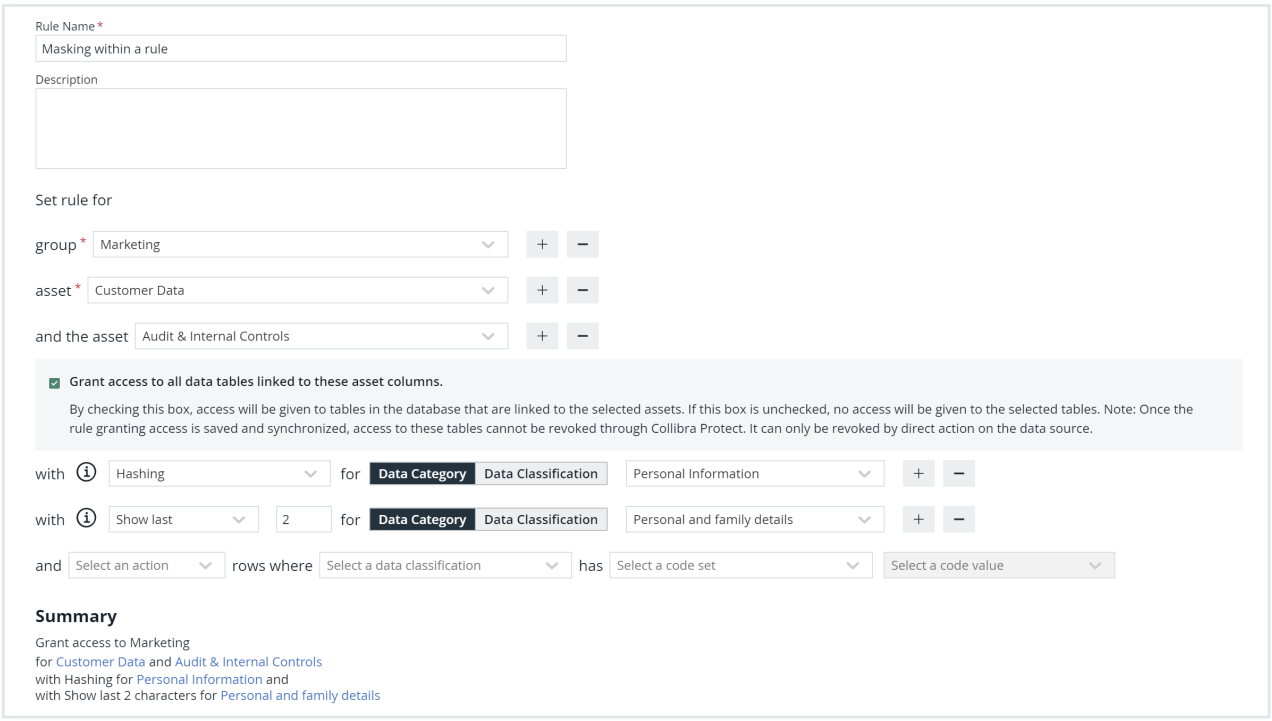
Masking within a rule

Scenario

A rule that is set for a group masks multiple agents using different types of masking, and the agents share the same column. This scenario is applicable regardless of whether the rule applies to a single asset or multiple assets.

Example

Consider a rule that is set for the **Marketing** group. The rule masks the **Personal Information** data category by hashing and masks the **Personal and family details** data category by showing only the last two digits. Suppose that both these data categories share the same column. Then, the rule will fail because the same column cannot be masked using two different masking types for a given group.

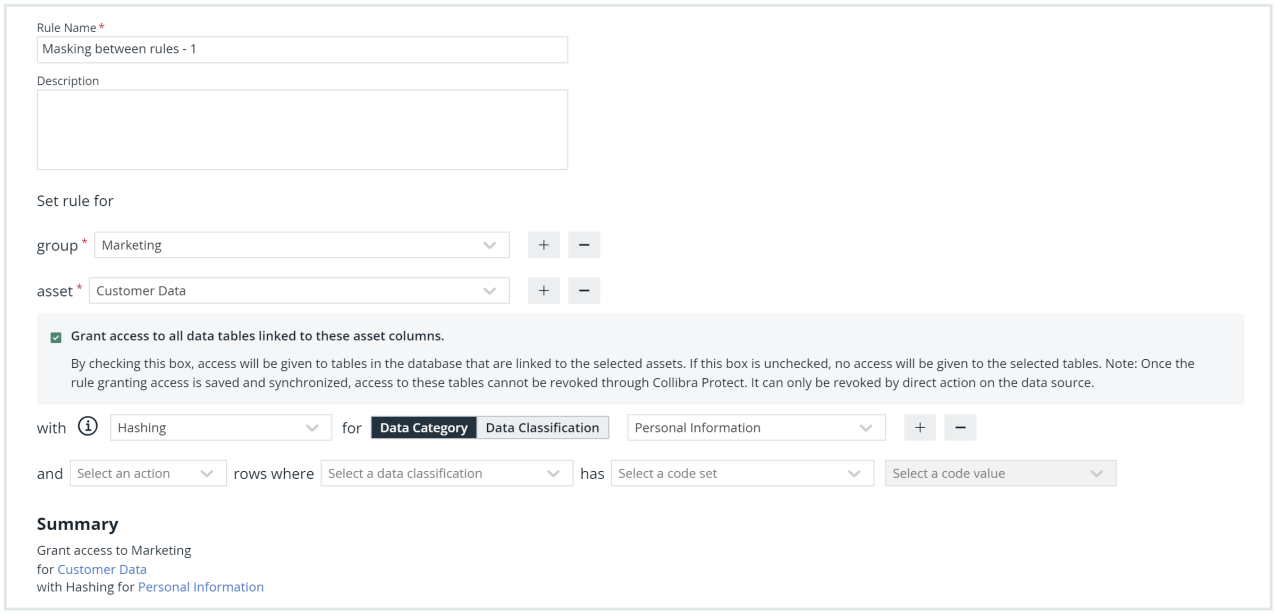


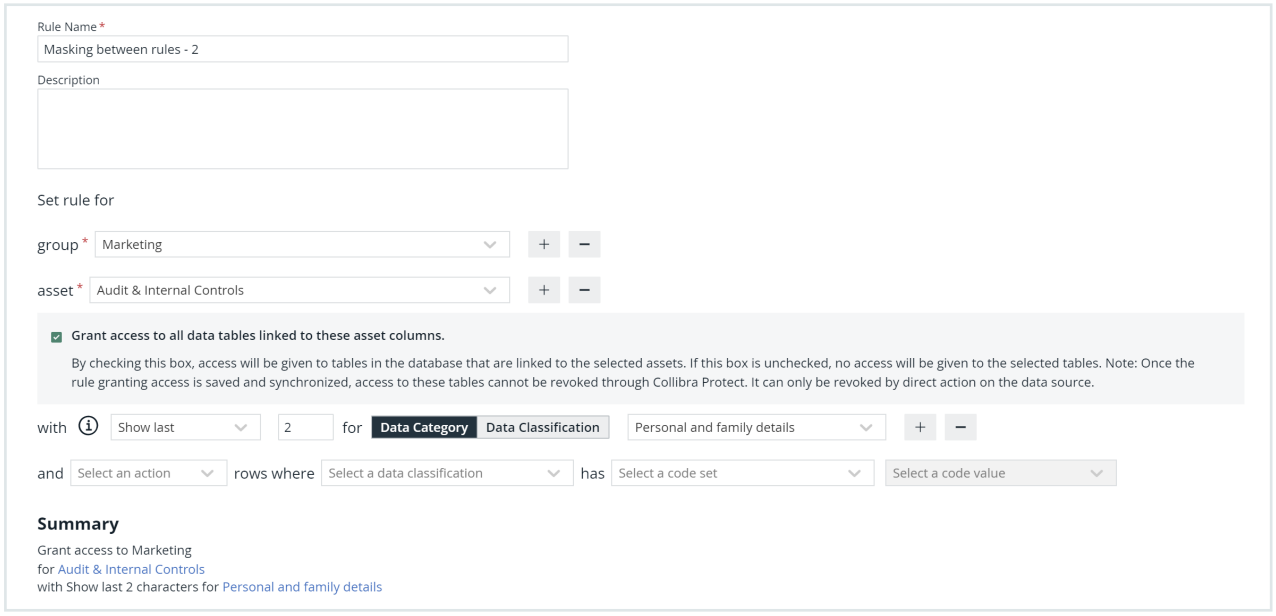
Masking between rules

This scenario is similar to the previous scenario except that this scenario considers two rules, instead of one, that are set for the same group. The masking types for the agents in the two rules are different, and both the agents share the same column. Then, a conflict occurs because the same column cannot be masked using two different masking types for a given group.

When two rules conflict with each other, if the synchronization status of only one of them is **Active**, then the other rule fails. If, however, the synchronization status of both the rules is **Active** or **Pending**, then both of them fail.

This scenario is applicable regardless of whether the agents are the same or different, and regardless of whether the rule applies to a single asset or multiple assets.





Conflicting filters affecting the same column

This topic contains examples to describe how data protection standards and data access rules behave when conflicting filters affect the same column.

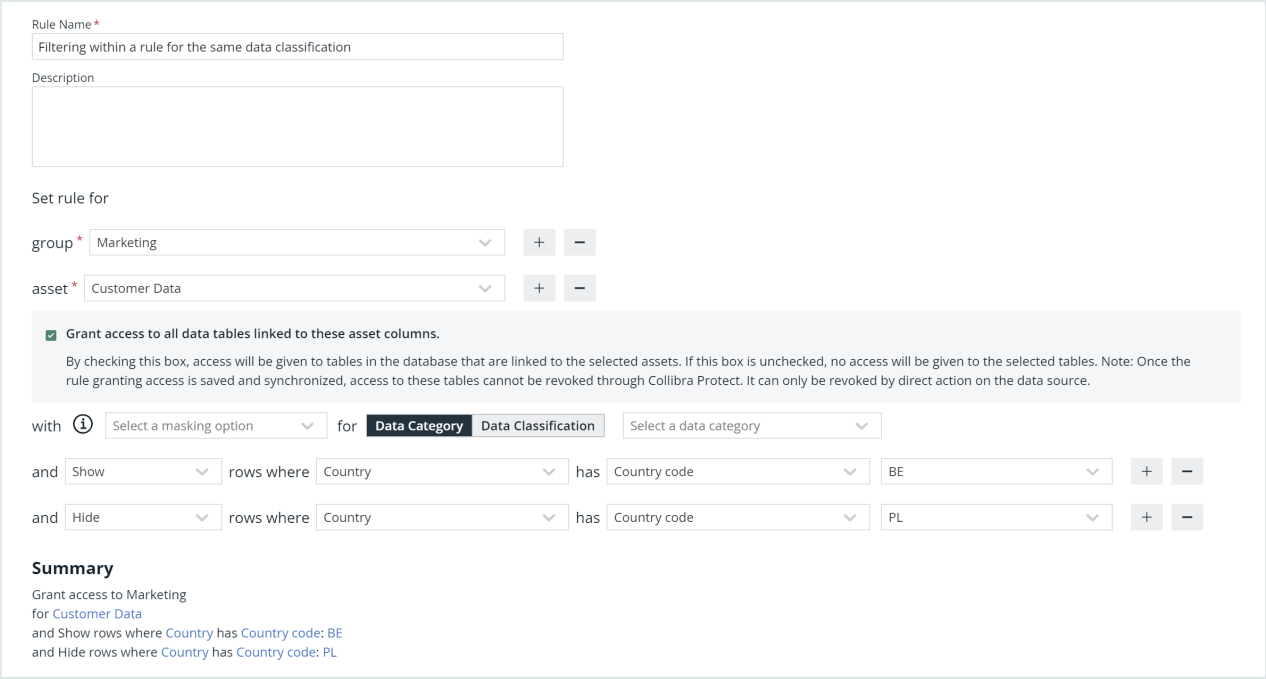
Filtering within a rule for the same data classification

Scenario

A rule that is set for a group contains conflicting filters for the same data classification.This scenario is applicable regardless of whether the rule applies to a single asset or multiple assets.

Example

Consider a rule that is set for the **Marketing** group and the **Customer Data** asset. The rule contains two filters for the **Country** data classification.



If any of the tables in the asset contain a column that is classified as **Country**:

- The first filter shows the rows that contain **BE** in that column.
- The second filter hides the rows that contain **PL** in that column.

Then, this rule will fail because two conflicting filters affect the same column.
When applying a filter for a specific data classification, you must select only one type of action. That is, you can choose to either show rows based on one or more values or hide rows based on one or more values. You must not use the show and hide filter actions together for the same data classification.

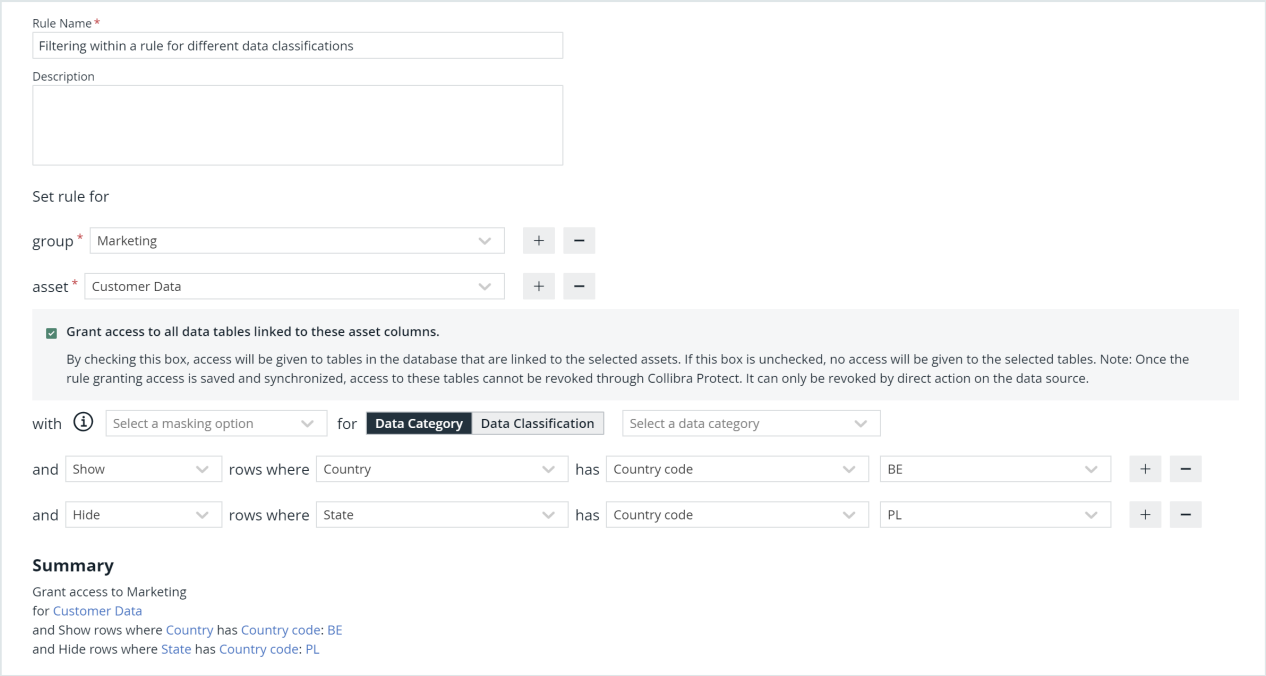
Filtering within a rule for different data classifications

Scenario

A rule that is set for a group contains conflicting filters for different data classifications that share the same column. This scenario is applicable regardless of whether the rule applies to a single asset or multiple assets.

Example

Consider a rule that is set for the **Marketing** group and the **Customer Data** asset. The rule contains two filters: one for the **Country** data classification, and another for the **State** data classification.



If any of the tables in the asset contain columns that are classified as **Country**, the first filter shows only the rows that contain **BE** in those columns.

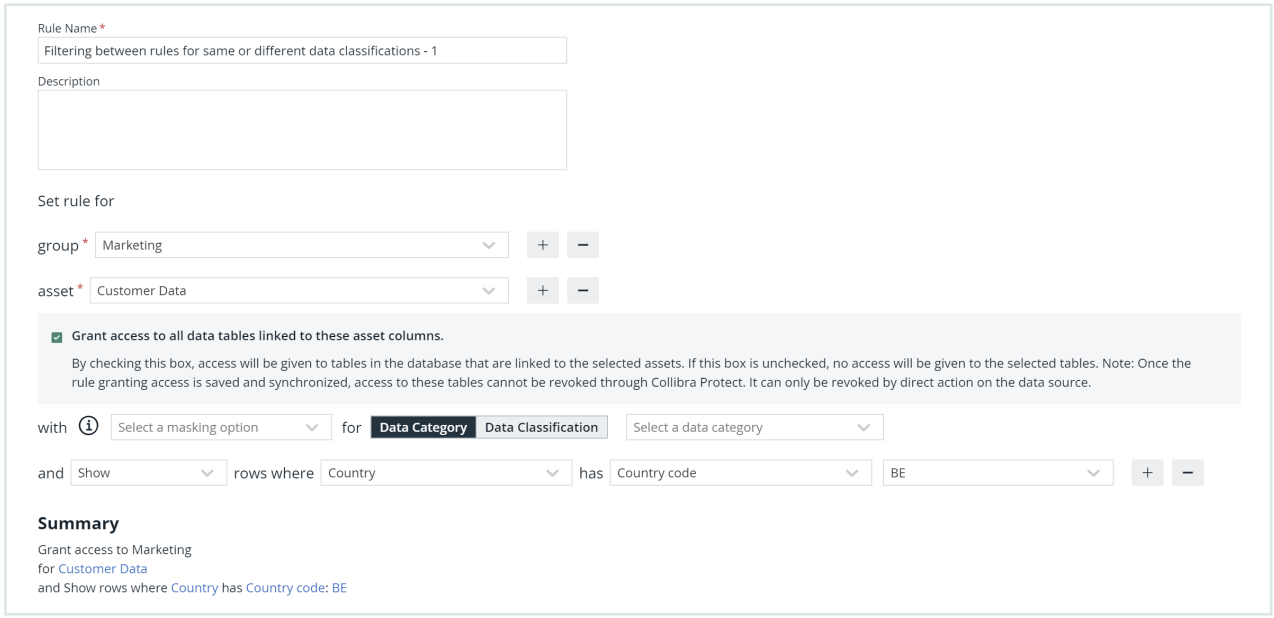
If any of the tables in the asset contain columns that are classified as **State**, the second filter hides only the rows that contain **PL** in those columns.

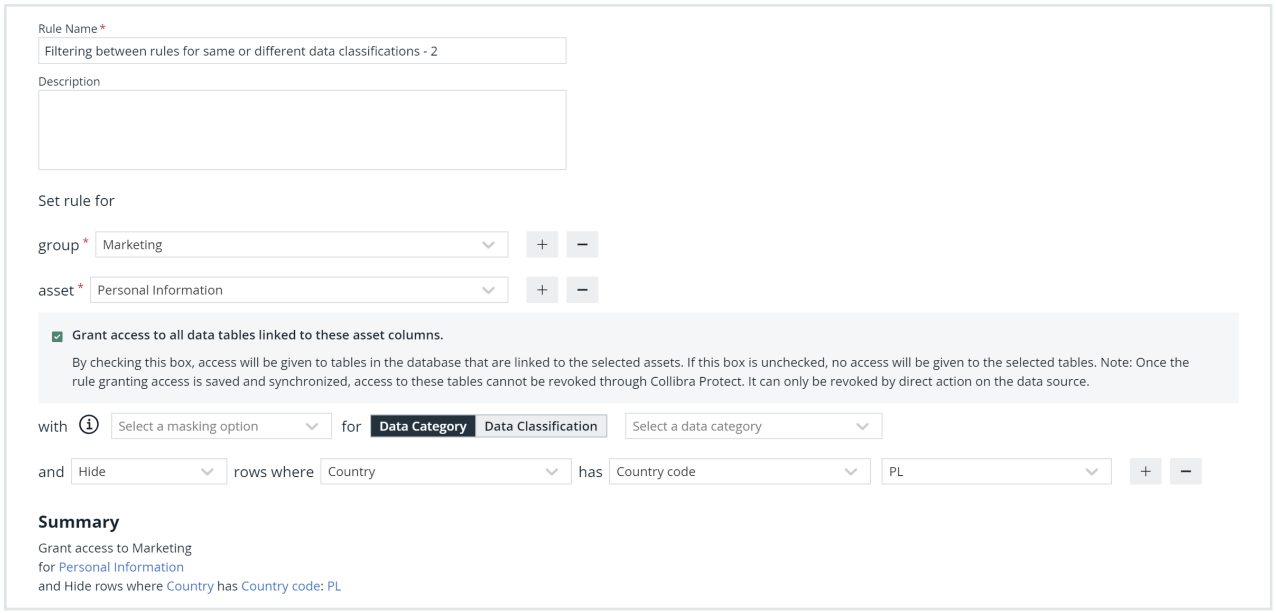
Suppose that a column is classified as both **Country** and **State**. That is, data classifications **Country** and **State** share the same column. Then, this rule will fail because two conflicting filters affect the same column.

Filtering between rules for same or different data classifications

This scenario is similar to the previous scenarios except that this scenario considers two rules, instead of one, that are set for the same group. The filter in one rule is different from the filter in the other rule, and both the filters affect the same column. Then, a conflict occurs because two conflicting filters affect the same column.

When two rules conflict with each other, if the synchronization status of only one of them is **Active**, then the other rule fails. If, however, the synchronization status of both the rules is **Active** or **Pending**, then both of them fail.




Rule Name *
Filtering between rules for same or different data classifications - 2
Description
Set rule for
group * Marketing
asset * Personal Information
Grant access to all data tables linked to these asset columns.
By checking this box, access will be given to tables in the database that are linked to the selected assets. If this box is unchecked, no access will be given to the selected tables. Note: Once the rule granting access is saved and synchronized, access to these tables cannot be revoked through Collibra Protect. It can only be revoked by direct action on the data source.
with
Select a masking option
for
Data Category
Data Classification
Select a data category
and
Hide
rows where
Country
has
Country code
PL
Summary
Grant access to Marketing
for Personal Information
and Hide rows where Country has Country code: PL

Protect

Protect is a capability of Collibra to protect sensitive data and grant varying levels of access to the data to specific groups of people through policies that do not require you to code. It enables you to build and manage policies in a single place and enforce those policies across various cloud platforms to ensure that sensitive data is always protected. You can enforce data protection at the source database level directly from the Protect interface, and apply advanced data protection through masking, redacting, and hashing.

Protect simplifies access governance and eliminates the need for repetitive actions and approvals. By providing permission to view information to only those who need it, Protect minimizes risk and promotes a safe data culture in your organization.

You can also use Protect to provide differential access, for example, to give everyone access to a data set but allow certain type of access to only certain groups of people based on data categories.

> Note Protect supports the following data source providers:
>
> - AWS Lake Formation
> - BigQuery
> - Databricks
> - Snowflake

Scenarios for using Protect

This topic describes how Collibra Protect helps you to:

- Use the metamodel graph to establish and enforce protection policies on Business Processes, Data Categories, and Data Sets.
- Apply a range of protection mechanisms to data sources using classifications.
- Support privacy preferences, such as consent management, data subject access requests, and the right to be forgotten, via row-filtering mechanisms.
- Conduct an audit of relevant protection at data sources and use reporting to demonstrate compliance in data storage and consumption.

Discover and classify personal information

Suppose that you want to help your organization find personal information.

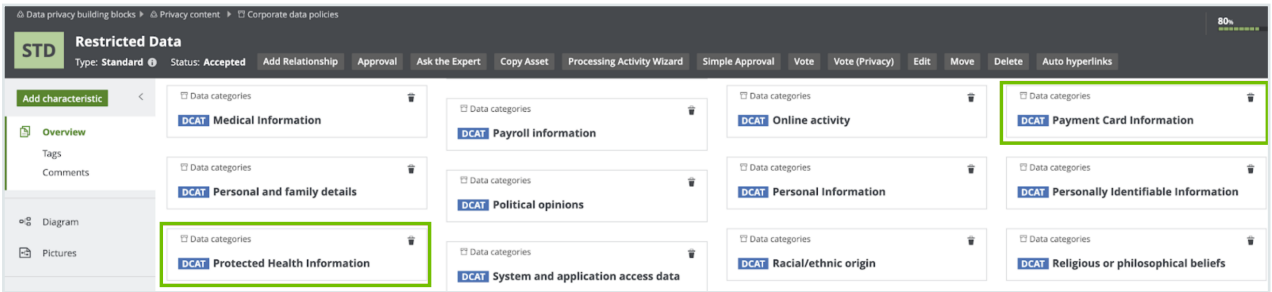
To achieve this, typically, your Privacy team sets up the Data Classification Policy, where they classify the data used in the organization based on the sensitivity or the business criticality of the data. This determines the required levels of security for the applications that store that data or the applications that are used for the transit of the data.

Consider the following three classifications for sensitivity:

- Public data, which is least sensitive.
- Private data, which is slightly more sensitive than the public data.
- Restricted data, which is the most sensitive data and therefore requires the highest level of access controls and security protection.

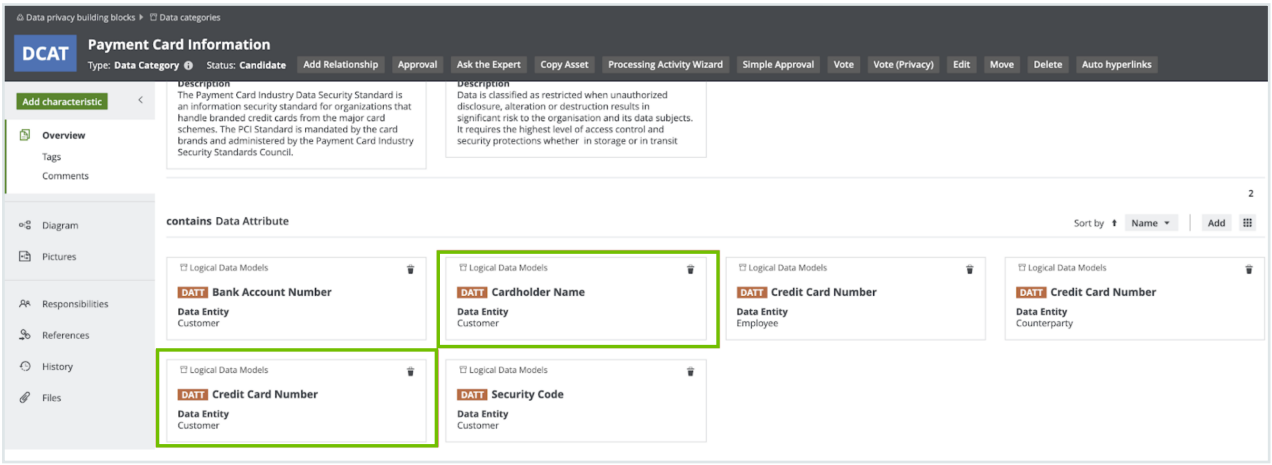
The following image shows the standard subassets of the Data Classification policy.

The Privacy team determines the data categories to which these subassets apply. For example, they can determine that Restricted Data applies to the following data categories: Gender, Social Security Number, Payment Card Information.

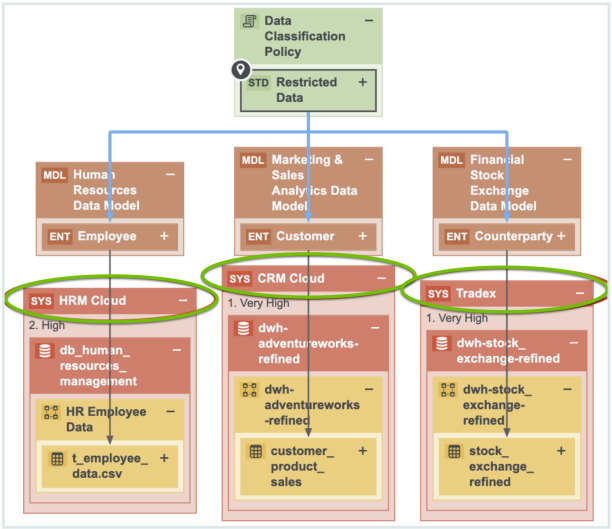


The Privacy team determines the sensitivity and the required security at the data category level as opposed to the column level. At the data category level, the Privacy team then determines what data elements belong to the identified data categories. For example, the

Payment Card Information data category groups the Cardholder Name and the Credit Card Number, among other information.



In this model, Data Attributes are grouped under the Data Category. This is how the Privacy layer is linked to the logical data model. This promotes collaboration between the Privacy team and the Governance team. In addition, this allows the automated data classification of the organization's personal information, which makes views such as the Restricted Data Overview diagram, available at the most sensitive data category, Standard Restricted Data.

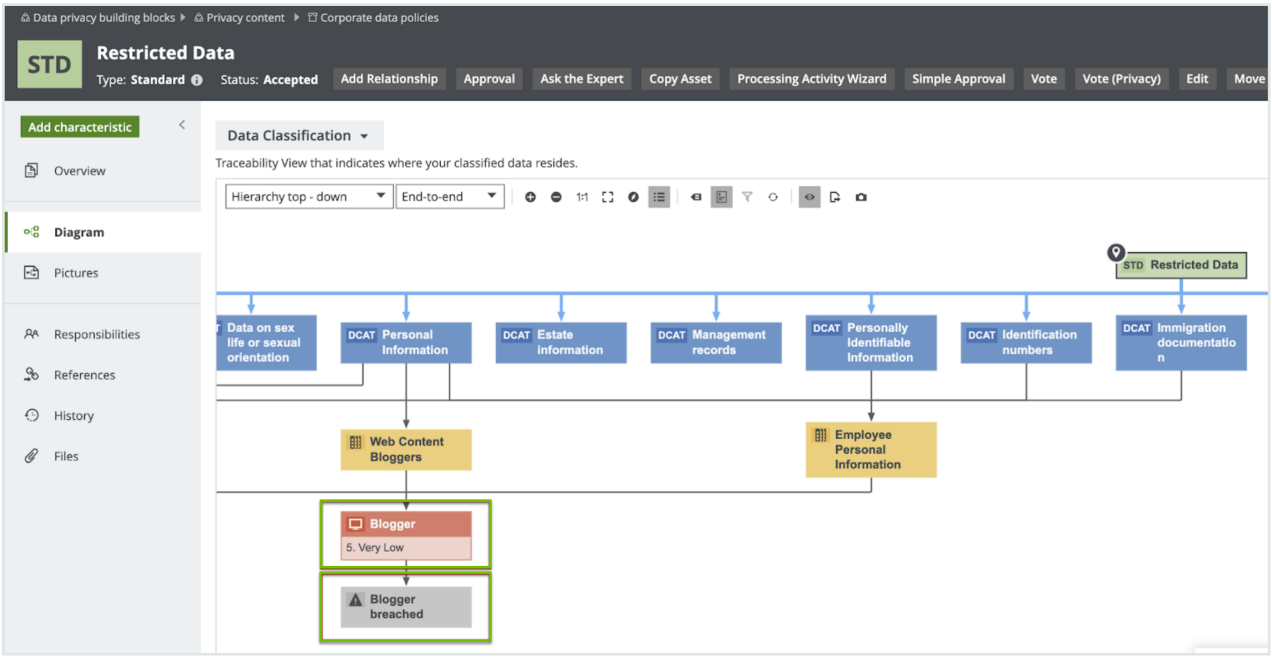


In the above image, the applications in which the restricted data resides are highlighted.

The Privacy team determines the policies and standards that determine which data categories are sensitive to the organization and what the required levels of protection are. The Data Governance team maps those data categories to the applications where that

data resides. The Security team determines what the security levels on those applications are. Thus, the view captured in the above image requires collaboration among teams.

Consider the traceability diagram called Data Classification under the Restricted Data standard. This standard contains the most sensitive information and thus requires the highest level of security controls; however, it resides on an application that has very low security. Because of this, the Information Security team needs to take the necessary remediation actions and improve the security levels on Blogger. As shown in the following image, an investigation is already ongoing on the potential data breach on Blogger.



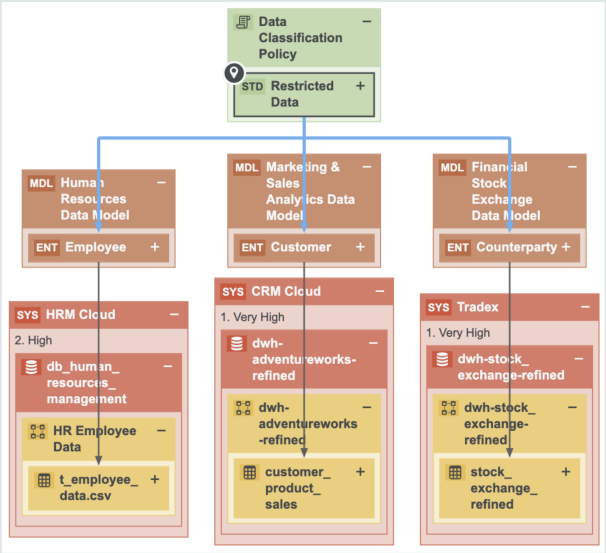
Data classification capabilities and guided stewardship

This section describes how Collibra Data Privacy leverages the data classification capabilities in Catalog. Thus far, we learned that the Restricted Data standard groups data categories, which group data attributes. In the example, the Payment Card Information data category contains the Credit Card Number data attribute.

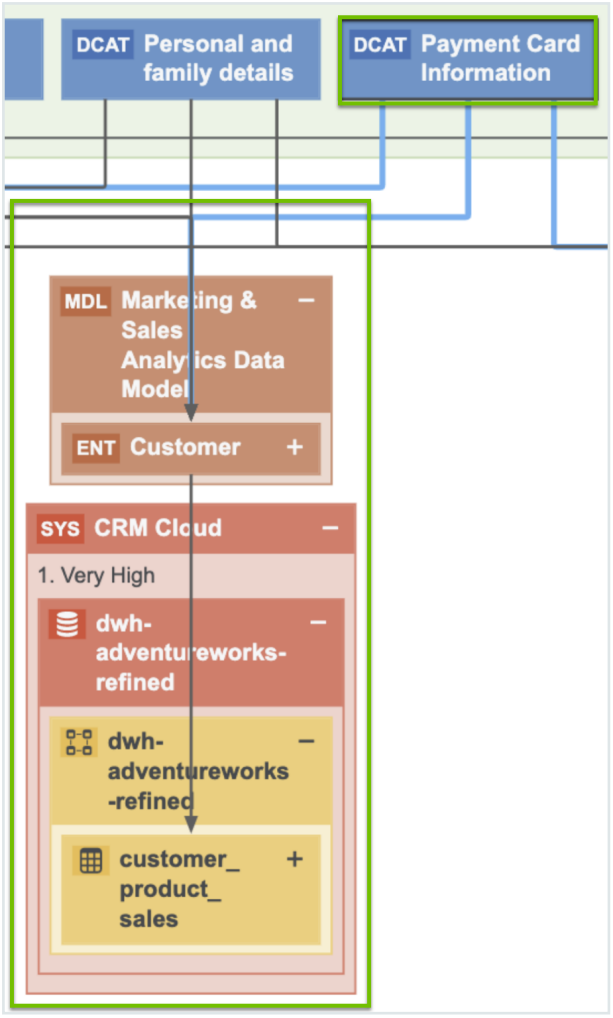
Guided stewardship is a semi-automated process of mapping columns and tables to logical data attributes. It enables content tables to be mapped to data attributes. After scanning a table and then applying guided stewardship in which the Steward selects

attributes from the suggestions coming from the automated mapping, the column is mapped to the Credit Card Number. Moreover, when a column is mapped to a data attribute, the column is also mapped to a data category because of the relation between the data category and the data attribute.

The result of classifying one application with the Catalog's Data Classification is shown in the following image.

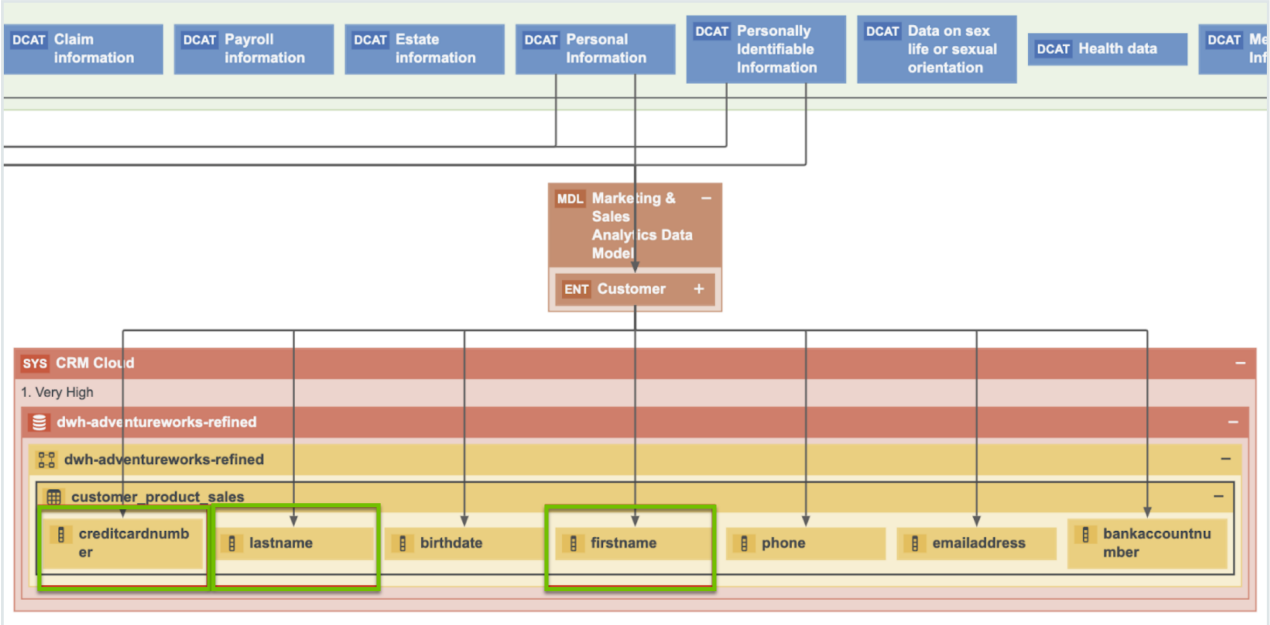


Restricted Data groups multiple data categories. The following image shows the data attributes that the Payment Card Information data category groups.



By applying guided stewardship and data classification, the data attributes are mapped to the columns. Thus, by using Catalog's data classification capabilities, the Data Governance team can find personal information and sensitive personal information.

It is important to know the context to determine which information is considered personal information. For example, Name can be the name of a customer or an employee, in which case Name is considered personal information. Name can also be the name of another organization. This context can be provided only by a Steward. Therefore, data classification and guided stewardship will help the Steward mapping customer's names to the Name column. Because the Privacy team has mapped names and family details, you can safely assume that this is Personal Information. Similarly, Credit Card Number can be the credit card number of another organization, but it is the Steward who has mapped the number to the Credit Card Number data attribute belonging to the Customer data entity, and as a result, we know that the payment card information is very restricted data.



This is an example of how guided stewardship, Catalog's data classification combined with guided stewardship and Data Privacy, gives you a vertical view on where Personal Information resides.

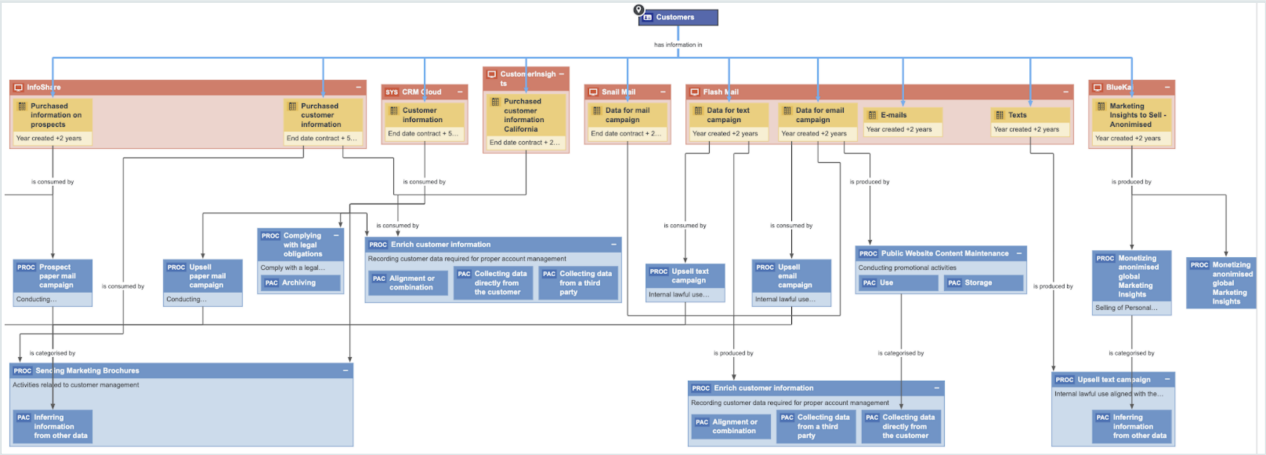
Customer requests and consent management

The previous sections described how we help customers find their Personal Information across applications. This section describes how we help customers manage data subject requests and consent. Collibra has the relevant metadata that is necessary for a partner application that fulfils the data subject requests or manages consent to operate. These applications need the metadata about where the data resides, where you store customer information, how you use the information, why you use the information, and what your legal basis is, so that they can determine for which applications you need consent and for which processes you need instance for a consent. Collibra has and governs the required metadata. In addition, through APIs, Collibra can integrate with those applications to feed them with the metadata that they need to function.

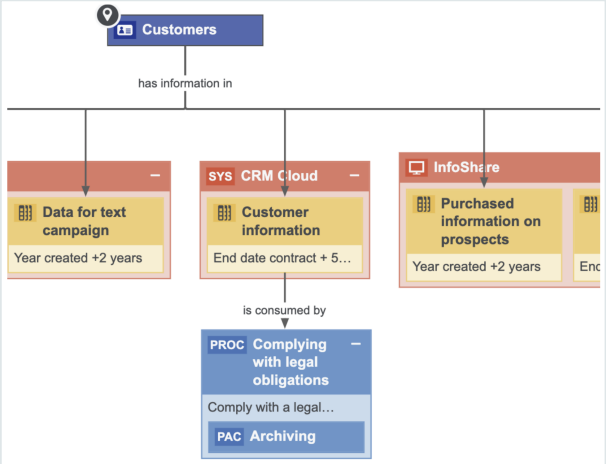
Consider the customer data. Collibra knows where this data resides and how it is being used. This is an outcome of obtaining input from the business users during the onboarding of the Business Processes where users are asked what data they use, which applications they use, and for what purpose they use the data. When further onboarding of those business processes by the Stewards takes place, one of these steps is mapping the

Business Processes to the data, and then also helping those Business Stewards with the mapping through the data classification capabilities in Catalog.

The following image shows a traceability view, which is a result of collaboration with the Business team, Data Governance team, and other teams.



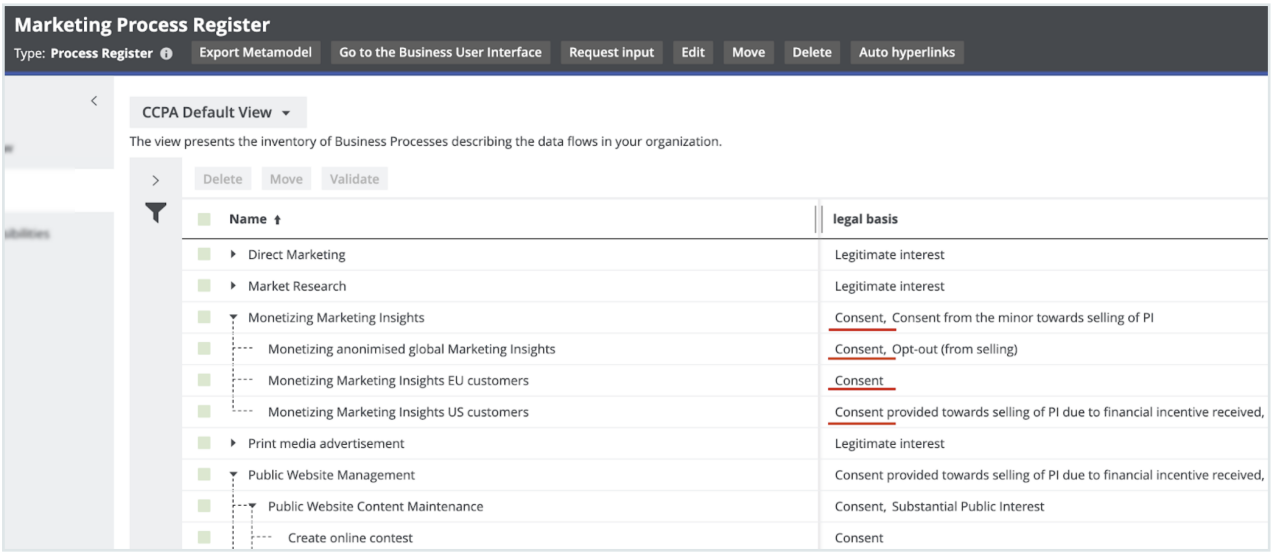
The above image shows where data resides and why it is used. It shows all the applications that contain customer data, and also the related retention periods, which can be imported when a customer wants to exercise their right to be forgotten. Collibra knows in which applications the data resides and the business processes that use that data. Thus, we know why and how we are using our customer data. This determines how to respond to the right to be forgotten because there are often Business Processes where you have the real legitimate reason to retain the customer's personal information.



When a customer wants to exercise their right to be forgotten, we can remove the information in these applications; however, we need to store the customer information in

the above table in order to comply with the legal obligation. Therefore, it is not only important to know where your personal information resides, but also why you are using it. Such information is important information for applications that process data subject requests (DSRs). You can integrate with the application that does the DSRs and create a workflow to process DSRs. Based on the input of the information and metadata that you will find in Collibra, you can validate the request. When the request is approved, you can point the applications to the Stewards and send them a task to perform the action that appears in the data subject request, such as, removing the data or extracting the data and sending it to a customer.

The same approach can be applied to the integrated consent management applications. These applications need to know the processes for reaching the consent, and such applications reside in the Records of Processing Activities (called Process Register in Collibra), so that you can see all the processes that rely on the consent and the data categories for which you need consent.



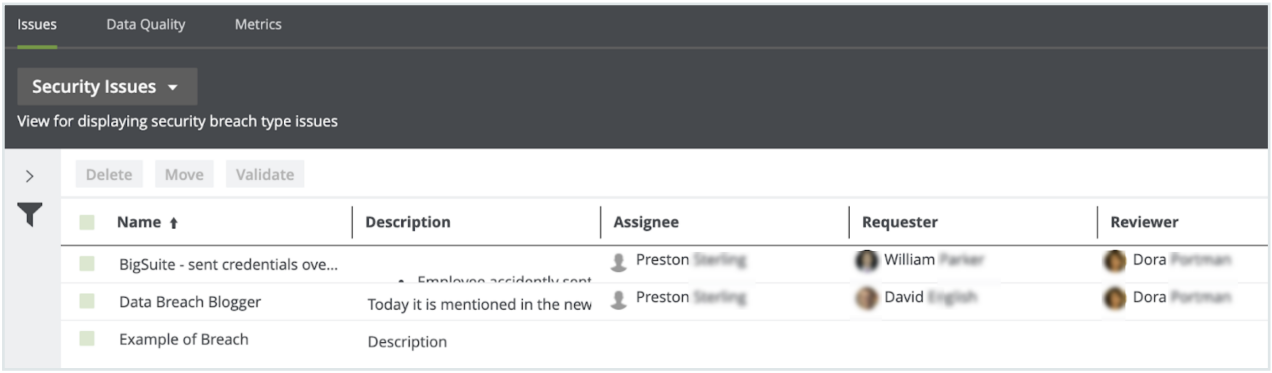
These are stored in the data sets that can also contain granular information, such as the individual data elements for which you want to obtain consent–this combines the information about which business processes require consent and the data categories for which you need consent to process all information in Collibra. The information governed in Collibra can be then sent to the consent management application that is used to manage consent.

Potential data breach workflow

This section describes how Collibra helps when a data breach occurs.

With Collibra Data Privacy, Collibra for Desktop, or Collibra for Mobile, you can report any suspicious behavior by logging a potential data breach.

If your organization has suffered a potential data breach, you can determine the application that needs to be investigated and the type of breach that may have occurred, and then log a potential data breach. The related workflow will require the Community Manager on the data governance counsel to assign an Issue Manager who will investigate the breach. The Issue Manager will then investigate the issue, assess the potential impact of the breach, determine the reporting requirements (for example, to whom the incident must be reported), and plan the remediation actions to address the risks. The reporting evidence needs to be stored. If you go to Data Helpdesk, you can find an overview of all the breaches that are being investigated.



Collibra can help with investigating the impact of the breach because of the knowledge of which data resides in the applications and the processes that use those applications. Such a holistic view on where the data resides, which applications are involved, and the

processes that rely on these applications can help in assessing the impact on customers following a data breach. Collibra can not only help an organization log and investigate a data breach but also help analyze the impact of the breaches because Collibra knows where the data resides and how it is being used. In addition, it contains a history of all the breaches (including potential ones) that would have been logged.

How do we get there?

This section describes the Records of Processing Activities (called Process Register in Collibra), Business Process discovery capabilities, data categorization and classification, and different prescriptive paths for reaching from the logical data layer envisioned in the metamodel graph and connected data sets to a physical data layer present in columns located directly at the data source.

Create and maintain Process Register (RoPA)

Process Register is an essential part of privacy compliance, foreseen directly by GDPR article 30 as a Record of Processing Activities (RoPA) and derived from CCPA requirements for performing data mapping in the organization. Process Register enables to store assets of the Business Process type that describes processes in the organization that involve personal data. In Collibra, Business Processes reflect the requirements stated by Processing Activity in GDPR.

Business Process onboarding

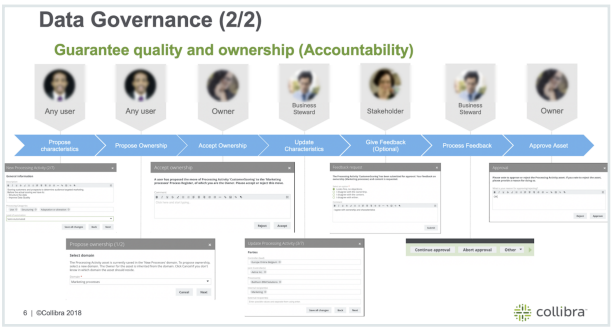
Business Processes may be onboarded by business users as well as privacy stewards through dedicated workflow implementing guided stewardship principle in Collibra Data Privacy. During onboarding, multiple roles collaborate in providing content to the onboarded Business Process. Because of the dedicated tasks and required approval and feedback, assets are onboarded in a governed way.

In the scenario on the Personal Information (PI) Discovery, it was described how Collibra helps with discovering Personal Information. But equally important to knowing where you are storing personal information is knowing why you are using personal information. That is, what the legal context of using that PI is. This context is created within Process

Registers, throughout the usage of Business Processes that describe the processes conducted by organization relating to the usage of personal information.

Typically, that information does not reside with one person that can help you document that knowledge. That information is stored within multiple areas across the organization and it may not be easy to centralize this information and ensure that the information is up to date. To help you with this task, CollibraData Privacy comes with the Business Process discovery capabilities.

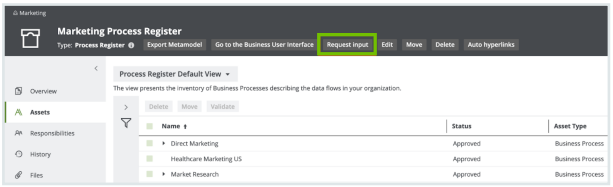
Consider a high-level overview of Data Privacy Business Process discovery capabilities. It commences with the Business Users describing the Business Processes in their terms. They will describe the data being used, applications being used, and any third parties with which they share information. After describing the Business Process, the owner of the Business Process will accept the ownership of that particular Business Process. When the ownership is accepted, the experts or the stewards will further onboard the proposed Business Process. This means that they will ensure that the Business Process is accurate and actionable because that Business Process provides business context on how we use personal information and we must ensure that the description is accurate. Therefore, in principle, you will have the Business Steward, Privacy Steward, and Data Steward, each adding business metadata, adding privacy metadata, and performing data mapping, respectively. After the stewards have updated the characteristics, you can optionally obtain feedback from the stakeholders. The following sections describe each step involved in the process.

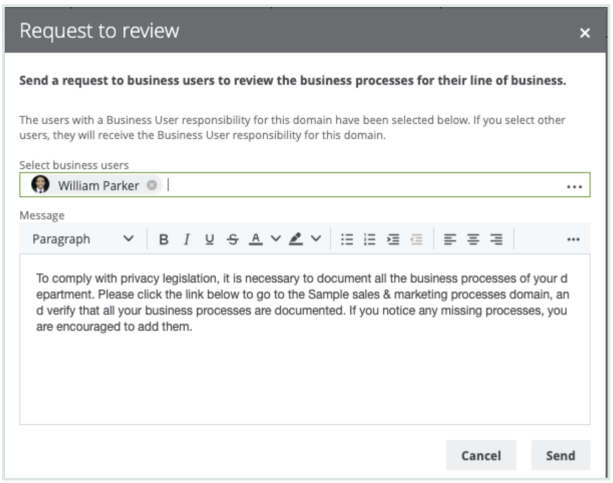


Requesting business users' input

The information related to Business Processes may be requested from the Business User directly from Data Privacy Process Register. Typically, this will be done by those who work on the Privacy program. With the **Request input** button, an email will be generated for the

selected business users, which can provide relevant information on the business side of the process. You can have a guiding text that explains the purpose of your request. If you click **Send**, an email is sent to the business user with an invitation to contribute to the Process Register.



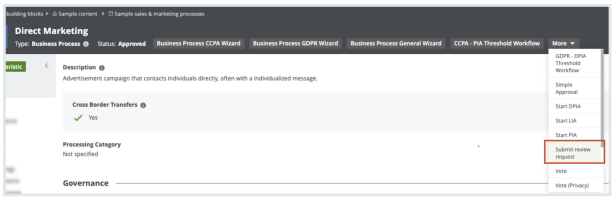


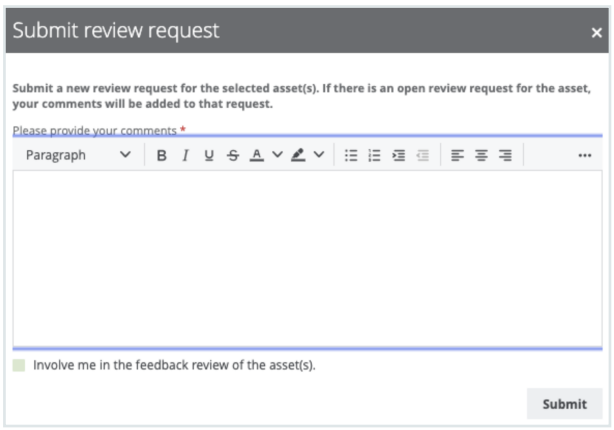
Maintain RoPA (Process Register) over time with review requests

While the successful result of the asset onboarding process is a new asset with the status Approved, asset change management is the standardized procedure for making changes to such approved assets.

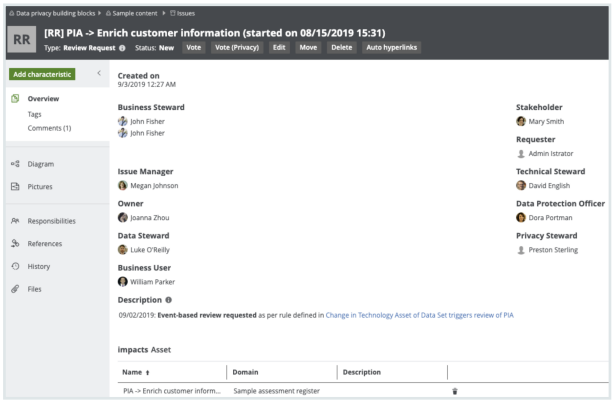
You may have many reasons to review an approved asset. Data Privacy groups such reasons into three categories and offers three corresponding means to trigger a review request:

- Manual: A trigger that is manually actioned by a user if, for example, the user wants to request a review of a Business Process asset considered to be incomplete or inaccurate. Any user can manually request a review of an approved asset.



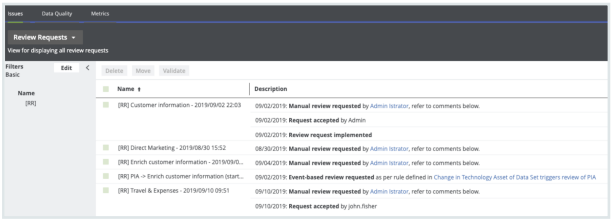


- Time-based: A trigger that is automatically actioned at a specified frequency. This is useful for assessment assets for which you might be required to review periodically to comply with a regulation.



- Event-based: A trigger that is automatically actioned by the fact of changes made to specified characteristics of the related asset.

All of the review requests are available in Data Helpdesk.



Perform assessments

Conduct PIA and DPIA

If a business process is likely to introduce a level of risk to the rights and freedom of natural persons, the Business Steward or the Data Protection Officer must perform the following:

- Privacy Impact Assessment (PIA), if complying with CCPA
- Data Privacy Impact Assessment (DPIA), if complying with GDPR

To determine whether or not you need to perform such an assessment for a Business Process asset, you must run a Threshold workflow.

The potential for business processes to expose the rights and freedom of natural persons to risk is significant. Privacy Impact Assessments (PIA) and Data Privacy Impact Assessments (DPIA) assess the risks to the rights and freedom of data subjects, born of a specific business process.

After onboarding a Business Process asset, the relevant Threshold workflow helps you determine whether or not a PIA or DPIA is needed. If it is determined that an assessment is necessary, the Owner or the Business Steward for the Business Process asset must complete the relevant workflow:

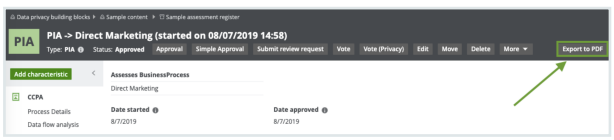
- PIA, if complying with CCPA
- DPIA, if complying with GDPR

Print assessment results

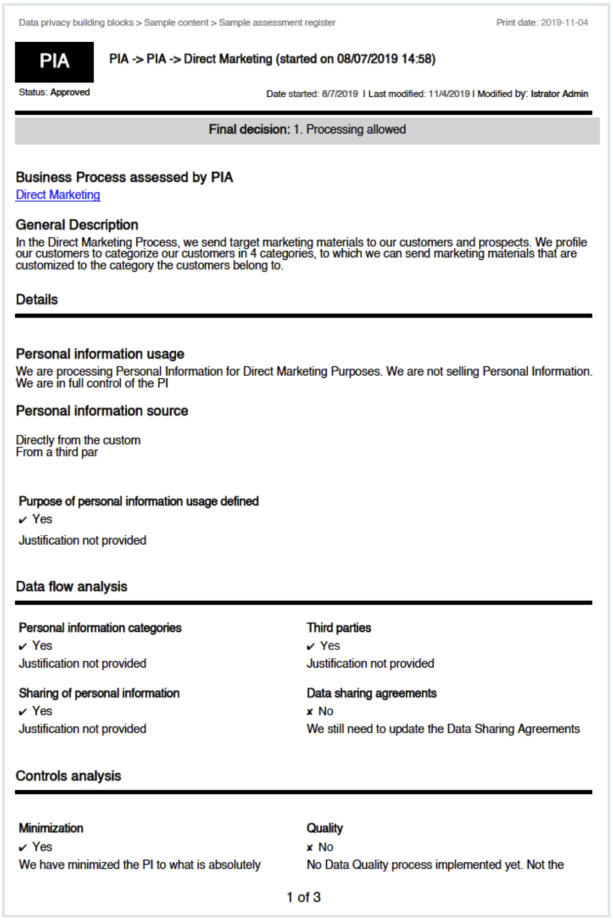
Assessments are a way for an organization to demonstrate compliance. You can export and print the PIA results in a unified way. You can also download a PIA asset page as a printable PDF, regardless of the status of the PIA asset.

Steps

1. Go to the relevant PIA asset page.



2. Click **Export to PDF**.
   » The PDF is downloaded to your computer.



Data privacy building blocks > Sample content > Sample assessment register Print date: 2019-11-04

PIA **PIA -> PIA -> Direct Marketing (started on 08/07/2019 14:58)**

Status: **Approved**

Date started: 8/7/2019 I Last modified: 11/4/2019 I Modified by: **Istrator Admin**

Final decision: 1. Processing allowed

Business Process assessed by PIA

Direct Marketing

General Description

In the Direct Marketing Process, we send target marketing materials to our customers and prospects. We profile our customers to categorize our customers in 4 categories, to which we can send marketing materials that are customized to the category the customers belong to.

Details

Personal information usage

We are processing Personal Information for Direct Marketing Purposes. We are not selling Personal Information. We are in full control of the PI

Personal information source

Directly from the custom
From a third par

Purpose of personal information usage defined

✔ Yes

Justification not provided

Data flow analysis

Personal information categories
✔ Yes
Justification not provided

Third parties
✔ Yes
Justification not provided

Sharing of personal information
✔ Yes
Justification not provided

Data sharing agreements
✘ No
We still need to update the Data Sharing Agreements

Controls analysis

Minimization
✔ Yes
We have minimized the PI to what is absolutely

Quality
✘ No
No Data Quality process implemented yet. Not the

1 of 3

Chapter 14

Essentials

This section contains information that can help you use Collibra Protect to the best of its ability.

Data protection types

This topic describes the types of protection that you can apply to your data via Protect.

> Tip *Data* refers to the tables and columns in a database.

Access-based protection

Access-based protection is the most basic type of protection that you can apply to your data. It involves providing the right users or groups access to the data based on the Collibra assets.

> Note Access-based protection is available only in data access rules.

Column-based protection

Column-based protection allows you to mask the data in specific columns so that the original data is not shown; for example, masking a column that contains credit card numbers.

You can mask the columns that are a part of a data category or a data classification. When granting access to a certain asset, you can apply the masking on only a subset of the asset if the subset is also a part of the data category or the data classification.

The following masking options are available:

- **Default masking:** Shows the data as **0**.
- **Hashing:** Shows the data as a set of different letters, numbers, and symbols.
- **Show last:** Shows the last few characters of the data. You can choose to show the last 1 through 20 characters of the data. The most common choice is 4.
- **No masking:** Shows the original data.

Note

- Column-based protection is available in both data protection standards and data access rules.
- For information about custom masking, go to Custom masking.

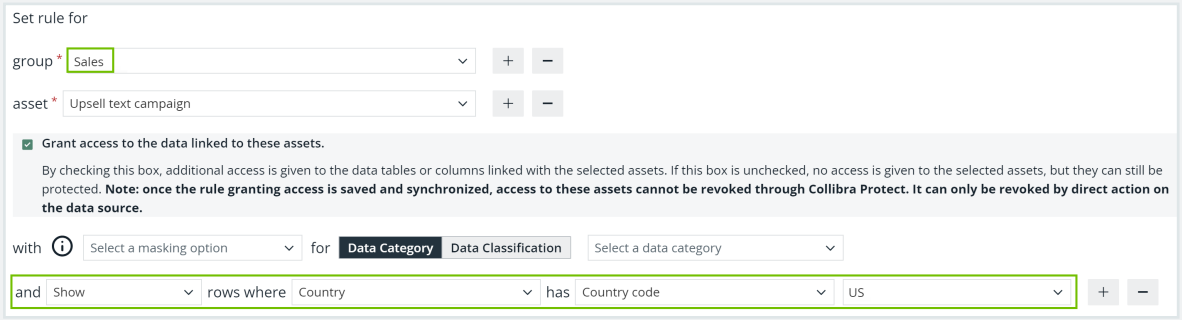
Example Suppose that you want the Human Resources (HR) group to be able to access a data set of US-based customers. Suppose that certain parts of the data set need to be hidden from the HR group because they contain restricted data, such as personally identifiable information (PII). Then, you can further protect the data by applying column-based protection or row-based protection.

Row-based protection

Row-based protection uses row filters to allow you to show or hide specific rows of a table. It is based on the values stored in the cell of a table.

Note Row-based protection is available only in data access rules.

Example Suppose that you want the Sales group to be able to access the data set of US-based customers. Then, you can create a data access rule and use the row-filtering option in the rule to show only those rows in the table that contain US in a column.



Custom masking

Custom masking is a feature that extends the data protection capabilities of Protect. Protect offers a set of out-of-the-box masking options. Custom masking allows you to define your own data protection methods.

You can manage custom masking via API. For more information, go to the Collibra Protect API documentation.

> Note
> - Custom masking functions are available only in Databricks and Snowflake. If you try to apply custom masking to a column in AWS Lake Formation or BigQuery, the out-of-the-box default masking is automatically applied to the column instead.
> - You cannot delete a custom masking function that is used in a data protection standard or a data access rule.

Example

The following is an example of a POST request for custom masking in Snowflake.

```
{
    "name": "My custom masking",
    "mappings": [
      {
        "provider": "Snowflake",
        "mappings": [
          {
            "dataType": "string",
            "functionName": "hash_my_string"
          },
          {
            "dataType": "number",
            "functionName": "hash_my_number"
          }
        ]
      }
    ]
}
```

If you apply **My custom masking** to a Snowflake column containing the value **Collibra**, the value is replaced by the result of the following Snowflake function: `hash_my_string (Collibra)`. However, if you apply this custom masking to a date column, the default

masking is automatically applied instead. This is because the POST request does not include any mapping for the date data type.

> Important The functionName specified in the mapping cannot contain spaces and cannot exceed 255 characters. Ensure that the masking functions exist on your data source provider. If a function does not exist, synchronization fails.

Masking functions

The following is an example of the syntax for a custom masking function in Databricks.

```
create or replace function mydb.myschema.mystring_function(value
STRING)
  RETURNS STRING
  RETURN concat("---", sha2(value, 0) , "+++");
```

The following is an example of the syntax for a custom masking function in Snowflake.

```
create or replace function mydb.myschema.mystring_function(value
VARCHAR)
    RETURNS VARCHAR
    AS
    $$
       concat('---', sha2(value) , '+++')
    $$;
```

Compatibility between Protect and Edge capability

Protect and Edge capabilities use different delivery mechanisms, which can result in compatibility differences. For example, you might have a version of Protect that supports custom masking, and a version of the Edge capability does not support it. If you use custom masking in a standard or rule, and your installed Edge capability does not support custom masking, synchronization is not triggered.

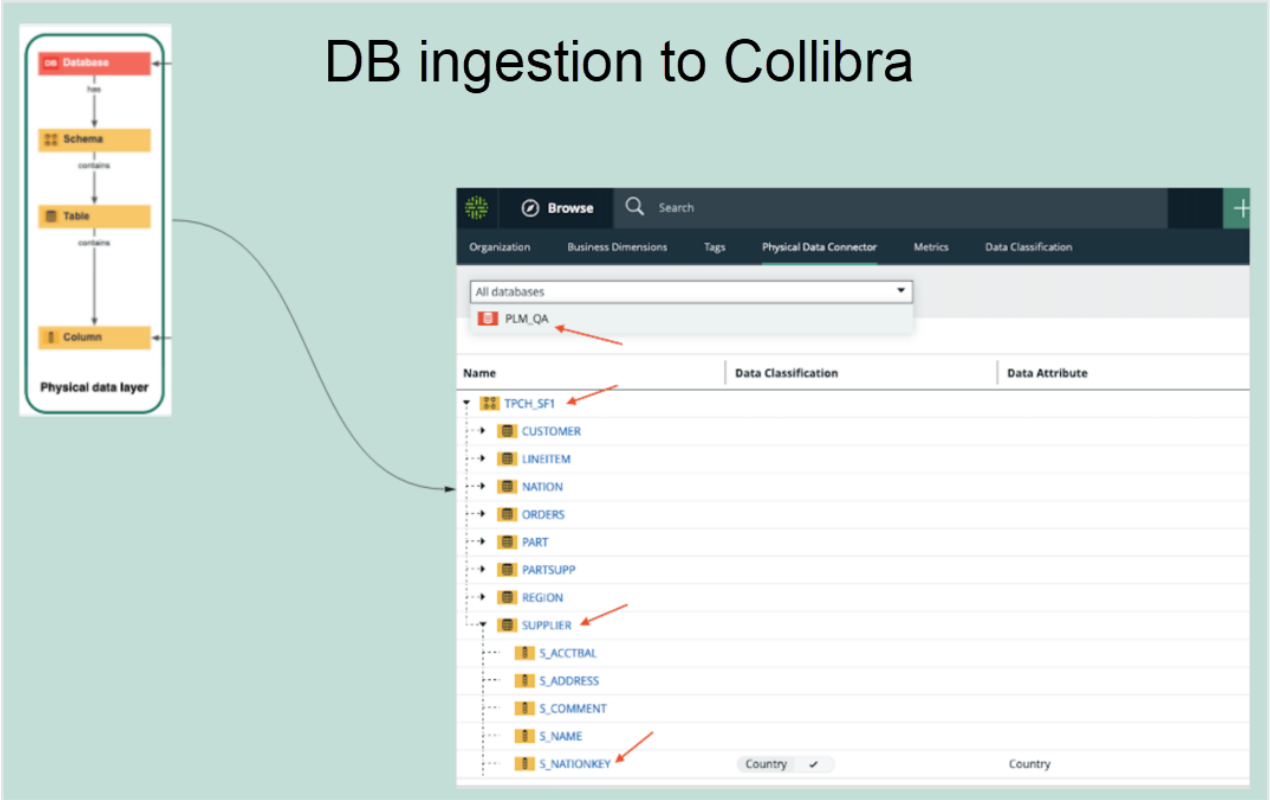
Technical background

This topic explains the connection of the data in a database (DB) with the physical layer (equivalent assets in Collibra Data Intelligence Cloud) and the logical layer (the packaged

model).

Consider the following DB.

When ingesting this DB to Collibra Data Intelligence Cloud, the physical layer is created, in addition to an asset for each of the schemas, tables, and columns, as depicted in the following image.



After the physical layer is created in Collibra, the logical layer can be created on top of the physical layer, as follows:

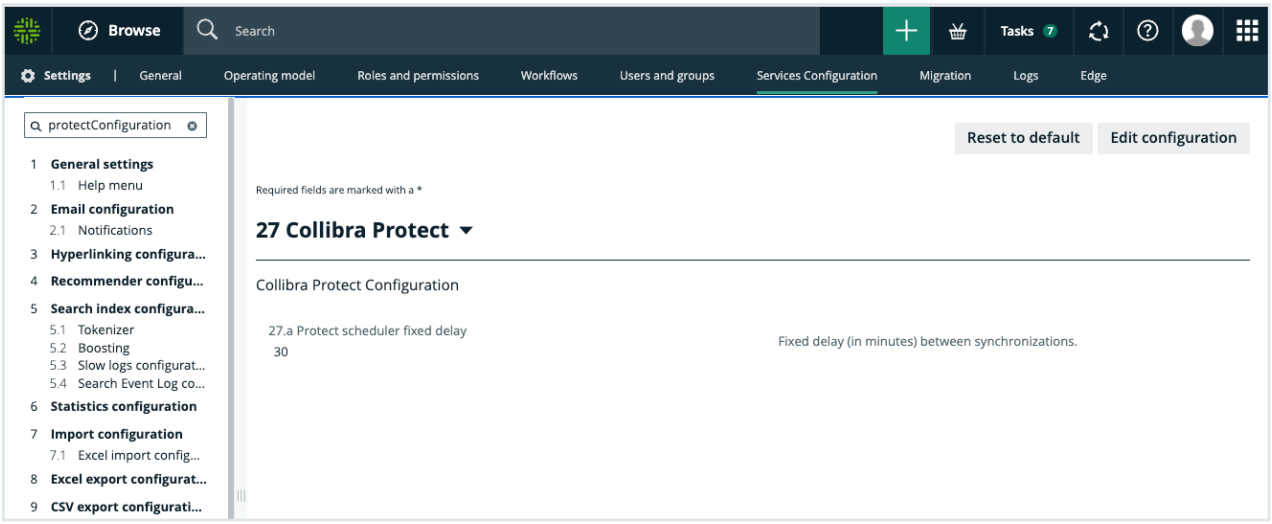
- Select any column and classify it as any available data classification. Alternatively, you can allow Collibra to classify the column for you.
- Assign the column to a data attribute.
- Create additional assets or use the existing assets of different types (Business Process, Data Category, or Data Set) to establish a relation with the columns.

> Note Protect supports only those columns that are linked to Table assets. It does not support Database View assets.

Synchronization

Collibra Protect automatically synchronizes data protection standards and data access rules with the databases of the data source providers such as BigQuery and Snowflake at regular intervals. This synchronization runs in the background on a configured frequency. By default, the frequency is every 60 minutes. You can, however, change the frequency

through the **Protect scheduler fixed delay** field on the Services Configuration tab in Collibra.



> Important If the **Services Configuration** tab is not shown to you, create a support ticket asking the following JVM Parameter be added to your Collibra Infrastructure Configuration: **-DPROTECT_SYNC_SCHEDULER_DELAY=PT60M**. After the parameter is added, restart Collibra to synchronize the Protect policies with the data source providers.

The synchronization includes the following processes:

- Aggregation of all data protection standards and data access rules with a computation of the following:
    - Which columns need to be masked for which groups
    - Which tables need to have a row filter
    - Which tables and columns need to be granted access
- On the databases of the data source providers such as Snowflake:
    - Creation and application of masking
    - Creation and application of row filters
    - Granting of access to groups on tables and columns (depending on the underlying database)

Data protection standards and data access rules

Collibra Protect protects your data through data protection standards and data access rules.

Data protection standards create a primary layer of protection for similar types of data by masking the data wherever it is stored, whereas data access rules create an additional layer of protection by managing access and enhancing protection for specific usages.

This topic explains when to create a data protection standard over a data access rule and vice versa, and what to consider when creating them.

When to create a standard over a rule and vice versa

- Suppose that columns containing the first and last names are a part of the Personally Identifiable Information (PII) data category. Then, regardless of the databases, tables, and schemas to which those columns belong, you can create a data protection standard that targets all of those columns by selecting the PII data category in the standard and masking it.
  Then, you can create a data access rule that grants access to a specific group, for a specific data set, while knowing that all PII within this data set will be masked by the data protection standard.
- Suppose that a data protection standard is created to mask a column that is classified as Personally Identifiable Information (PII) for everyone. You, however, want to unmask that PII column for a specific group. You can do so by creating for the same group a data access rule to unmask the classified column, because data access rules take priority over data protection standards.
- Suppose that you want to grant access to a group, but the protection from the data protection standard is not enough because there might be other sensitive data within a supported asset. Then, you can create a data access rule to add additional layers of protection over the ones that were set by the data protection standard. You can further protect the data by applying additional masking on the data or by filtering the data using the row-filtering option in the rule.

What to consider when creating standards or rules

When creating data protection standards or data access rules for assets, consider how the assets are grouped. Suppose that you have a Business Process asset, BP, which contains the following Data Set assets: DS1, DS2, and DS3. Instead of creating a data protection standard or data access rule for each of the three Data Set assets (DS1, DS2, and DS3), consider creating a standard or rule that targets the Business Process asset (BP), to save your time.

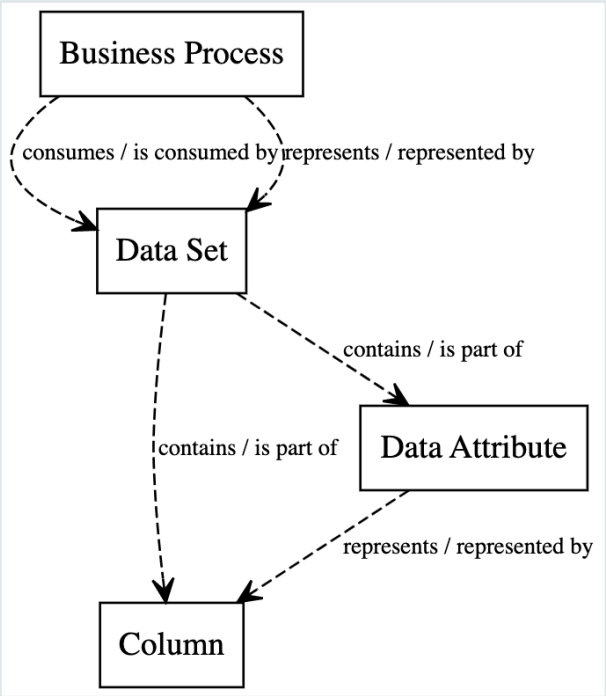
Prescriptive paths

You can use Collibra Protect to secure the data in the assets of the packaged asset types, such as Business Process, Data Category, and Data Set, in addition to the assets of any new or modified asset types.

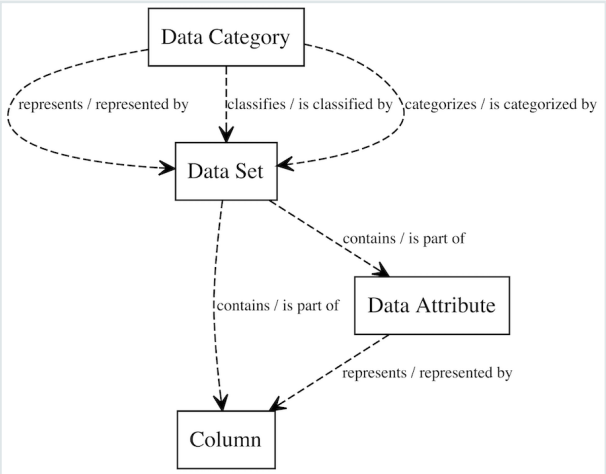
The asset that you select when creating a data protection standard or a data access rule is related to the physical data layer, such as tables and columns, through a set of relations and intermediate assets. These relations are paths that Protect uses to traverse from the selected asset (business or logical layer) to a column (physical data layer) in order to find the column that needs protection. Such traversal follows a set of prescriptive paths. Each asset type has a set of prescriptive paths for traversing to the Column asset, as depicted in the following sections.

> Note Depending on your permission, you can customize the prescriptive paths. For more information, go to Customization of prescriptive paths.

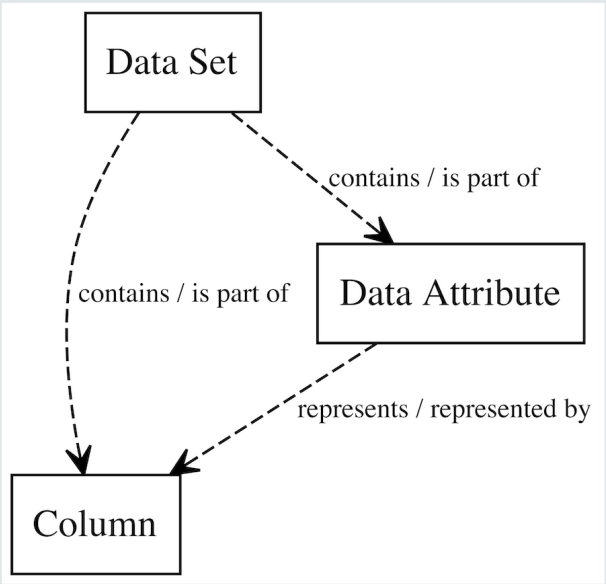
Business Process



Data Category



Data Set



Customization of prescriptive paths

Collibra Protect supports the following asset types:

- **Packaged asset types:** Business Process, Data Category, and Data Set
- **Custom asset types:** These are the packaged asset types that you have modified or the asset types that you have created. If you modify the attributes and relations of a packaged asset type, then the packaged asset type becomes a custom asset type.

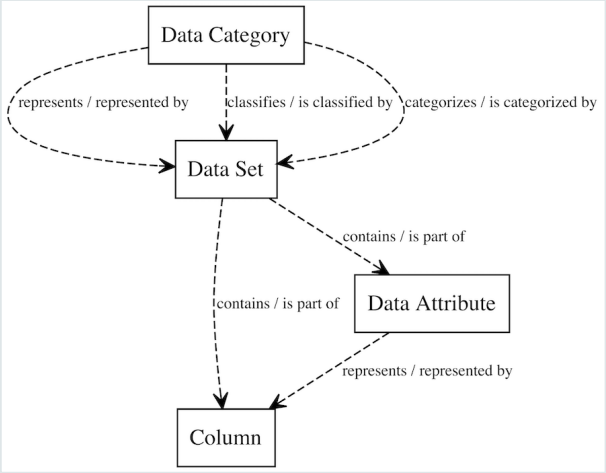
If you have the **Protect** > **Administration** global permission, you can customize the prescriptive paths for the asset types through APIs. The customization may include creating, modifying, or deleting the prescriptive paths: for example, adding or modifying the prescriptive paths for packaged and custom asset types, defining how the asset types relate to columns, and removing any obsolete prescriptive paths.

The customized prescriptive paths are applied to data protection standards and data access rules.

> Note You cannot remove a customized prescriptive path if an asset type linked to the prescriptive path is used in a standard or rule.

Protect supports a maximum of 10 asset types. Each asset type can have a maximum of 6 relations and a maximum depth of 3. However, when customizing the prescriptive path for an asset type, we recommend that you provide only one relation for the asset type. Prescriptive paths must always end in a Column asset type (that is, 00000000-0000-0000-0000-000000031008).

The following image is an example of a prescriptive path that has 6 relations and a depth of 3.



Restore the default asset types

If you want to restore the default asset types defined by Collibra, a PATCH operation must be performed on each asset type. The list of asset types and their specifications are as follows.

If Data Privacy is not installed

Data Set (00000000-0000-0000-0001-000400000001)

```
                    {
    "description": "Prescriptive path from Data Set to Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
```

```
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-000000031008"
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-
000000031005",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007094",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      }
    ],
    "assetTypeId": "00000000-0000-0000-0001-000400000001"
  }
```

Data Category (00000000-0000-0000-0000-000000031109)

```
                  {
    "description": "Prescriptive path from Data Category to
Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
```

```
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
```

```
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      }
    ],
    "assetTypeId": "00000000-0000-0000-0000-000000031109"
  }
```

Business Process (00000000-0000-0000-0000-000000031103)

```
                {
    "description": "Prescriptive path from Data Set to Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-000000031008"
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-
```

```
000000031005",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007094",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      }
    ],
    "assetTypeId": "00000000-0000-0000-0001-000400000001"
  }
```

If Data Privacy is installed

Data Set (00000000-0000-0000-0001-000400000001)

```
                    {
    "description": "Prescriptive path from Data Set to Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-000000031008"
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007062",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0000-
000000031005",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007094",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
```

```
            }
          }
        }
    ],
    "assetTypeId": "00000000-0000-0000-0001-000400000001"
  }
```

Data Category (00000000-0000-0000-0000-000000031109)

```
                  {
    "description": "Prescriptive path from Data Category to
Column",
    "relations": [
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
```

```
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007007",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
```

```
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007315",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "c0e00000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007315",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "c0e00000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
        }
      }
```

```
    ],
    "assetTypeId": "00000000-0000-0000-0000-000000031109"
  }
```

Business Process (00000000-0000-0000-0000-000000031103)

```
                  {
    "description": "Prescriptive path from Business Process to
Column",
    "relations": [
      {
        "relationTypeId": "c0e00000-0000-0000-0000-
000000007314",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "c0e00000-0000-0000-0000-
000000007314",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "c0e00000-0000-0000-0000-
000000007314",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
```

```
000000031008"
                }
              }
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031008"
            }
          }
        }
      },
      {
        "relationTypeId": "00000000-0000-0000-0000-
000000007038",
        "relationTypeDirection": "SOURCE",
        "assetType": {
          "assetTypeId": "00000000-0000-0000-0001-
000400000001",
          "relation": {
            "relationTypeId": "00000000-0000-0000-0000-
000000007062",
            "relationTypeDirection": "SOURCE",
            "assetType": {
              "assetTypeId": "00000000-0000-0000-0000-
000000031005",
              "relation": {
                "relationTypeId": "00000000-0000-0000-0000-
000000007094",
                "relationTypeDirection": "SOURCE",
                "assetType": {
                  "assetTypeId": "00000000-0000-0000-0000-
000000031008"
                }
              }
            }
          }
```

```
            }
          }
        ],
        "assetTypeId": "00000000-0000-0000-0000-000000031103"
    }
```

Set up Protect

This topic describes how to set up Protect and establish a connection between your data source and Protect.

> Tip
> The information in this topic varies depending on the data source that you select.
>
> Data source
>
> 

Before you begin

AWS Lake Formation

1. Download the JDBC driver for Amazon Athena.
2. Create a JDBC connection from your Edge site to Amazon Athena.

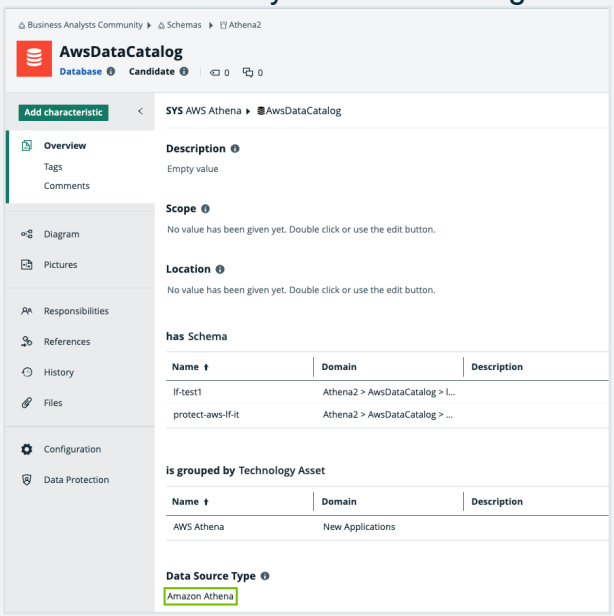
   > Tip When creating the connection, in the **Connection provider** field, select **Generic JDBC connection**.

3. Add the Catalog JDBC ingestion capability to the Edge site.

   > Tip
   > When adding the capability:
   >
   > - In the **Capability template** field, select **Catalog JDBC Ingestion**.
   > - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

Tip The following image shows an ingested AWS Lake Formation database. The **Data Source Type** attribute containing the value **Amazon Athena** is added to the database asset only after the Catalog JDBC ingestion process is complete.



BigQuery

1. Download the JDBC driver for Google BigQuery.
2. Create a JDBC connection from your Edge site to Google BigQuery.

   Tip When creating the connection, in the **Connection provider** field, select **Generic JDBC connection**. In the **Connection properties** section, set the value of the **Other** connection property to **SupportNativeDataType=True**.

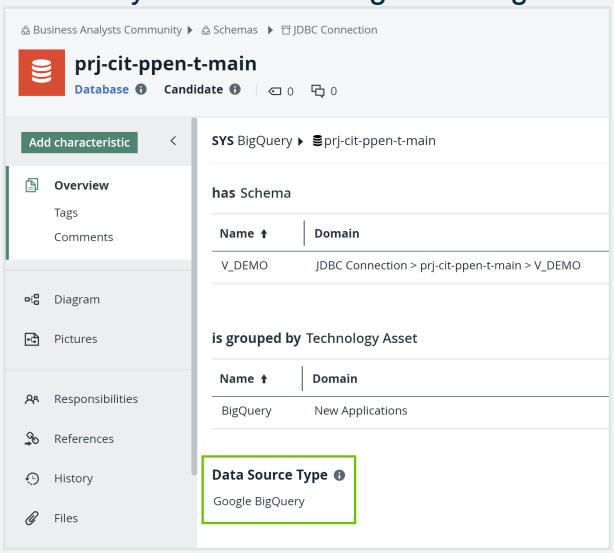
3. Add the Catalog JDBC ingestion capability to the Edge site.

   Tip
   When adding the capability:

   - In the **Capability template** field, select **Catalog JDBC Ingestion**.
   - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

Tip The following image shows an ingested BigQuery database. The **Data Source Type** attribute containing the value **Google BigQuery** is added to the database asset only after the Catalog JDBC ingestion process is complete.



Watch a video

Databricks

1. Download the JDBC driver for Databricks.
2. Create a JDBC connection from your Edge site to Databricks.

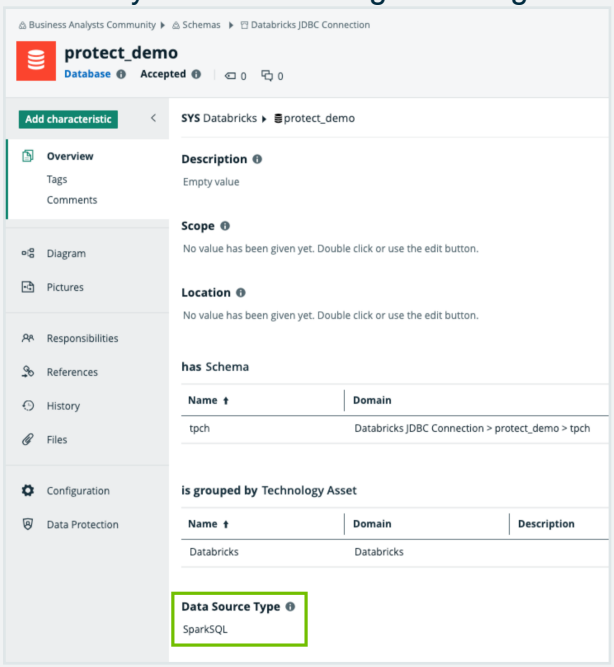
> Tip When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.

3. Add the Catalog JDBC ingestion capability to the Edge site.

   > Tip
   > When adding the capability:
   >
   > - In the **Capability template** field, select **Catalog JDBC Ingestion**.
   > - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

> Tip The following image shows an ingested Databricks database. The **Data Source Type** attribute containing the value **SparkSQL** is added to the database asset only after the Catalog JDBC ingestion process is complete.



Snowflake

1. Download the JDBC driver for Snowflake.
2. Create a JDBC connection from your Edge site to Snowflake.

   > Tip When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.

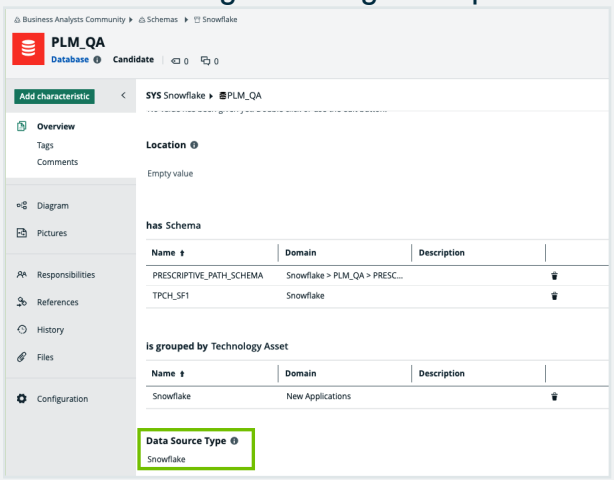
3. Add the Catalog JDBC ingestion capability to the Edge site.

   Tip
   When adding the capability:

   - In the **Capability template** field, select **Catalog JDBC Ingestion**.
   - In the **JDBC Connection** field, select the connection that you created in Step 2.

4. Register and synchronize the data source.

Tip The following image shows an ingested Snowflake database. The **Data Source Type** attribute containing the value **Snowflake** is added to the database asset only after the Catalog JDBC ingestion process is complete.

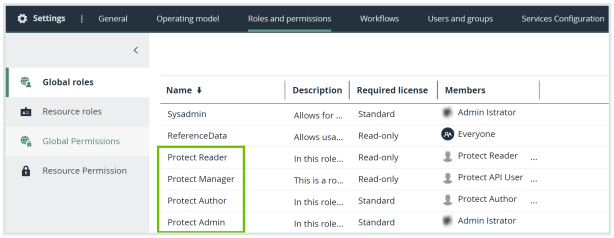


Steps

AWS Lake Formation

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.

2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create an AWS connection from the Edge site to Amazon Athena.

> Tip
> - When creating the connection, in the **Connection provider** field, select **AWS connection**.
> - Ensure that the user associated with the Access Key ID used in the connection has the required permissions.

4. Add the Protect for AWS Lake Formation capability to the Edge site.

> Tip
> - When adding the capability:
>   - In the **Capability template** field, select **Collibra Protect for AWS Lake Formation**.
>   - In the **Connection** field, select the connection that you created in Step 3.
> - Do not add more than one Protect for AWS Lake Formation capability to the Edge site.

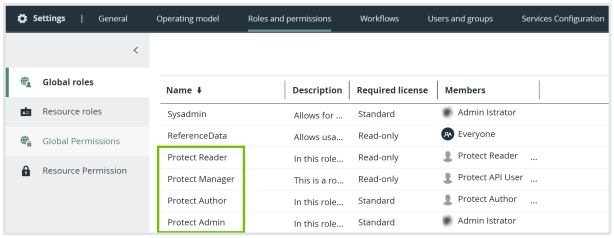
» Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

BigQuery

> Note Apart from the JDBC connection created for the Catalog ingestion, Protect for BigQuery requires an extra connection, which is the GCP connection. The GCP connection is necessary because Protect requires access to certain GCP APIs that cannot be reached through the JDBC connection alone. The GCP connection ensures that data protection is enforced.

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.

2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create a GCP connection from the Edge site to Google BigQuery.

> Tip
> - When creating the connection, in the **Connection provider** field, select **GCP connection**.
> - Ensure that the user associated with the GCP Service Account used in the connection has the required permissions.

4. Add the Protect for BigQuery capability to the Edge site.

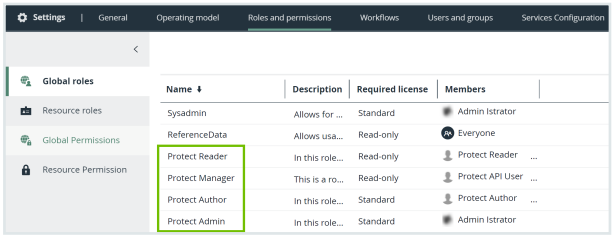
> Tip
> - When adding the capability:
>   - In the **Capability template** field, select **Collibra Protect for Google BigQuery**.
>   - In the **Connection** field, select the connection that you created in Step 3.
> - Do not add more than one Protect for BigQuery capability to the Edge site.
> - If the version of the capability is **1.97.1**, then ensure that the JSON content in the **GCP Service Account** field in the GCP connection you created is Base64 encoded. You can find the version of the capability in the **Version** column on the **Capabilities** tab.

» Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

Watch a video

Databricks

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.
2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create a Username/Password JDBC connection from the Edge site to Databricks.

   Tip
   - When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.
   - Ensure that the user associated with the Databricks role used in the connection has the required privileges.

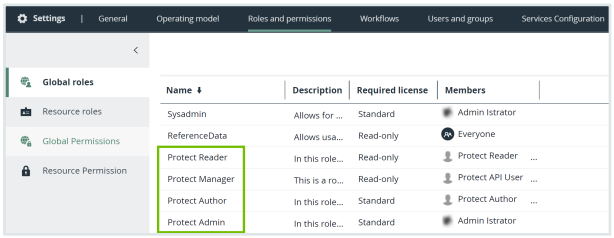
4. Add the Protect for Databricks capability to the Edge site.

   Tip
   - When adding the capability:
     - In the **Capability template** field, select **Collibra Protect for Databricks**.
     - In the **Connection** field, select the connection that you created in Step 3.
   - Do not add more than one Protect for Databricks capability to the Edge site.

   » Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

Snowflake

1. Contact Collibra Support or your representative to enable Protect on your Collibra environment.
2. Ensure that the Protect global roles and global permissions are correctly set.



3. Create a Username/Password JDBC connection from the Edge site to Snowflake.

   Tip
   - When creating the connection, in the **Connection provider** field, select **Username/Password JDBC connection**.
   - Ensure that the user associated with the Snowflake role used in the connection has the required privileges.

4. Add the Protect for Snowflake capability to the Edge site.

   Tip
   - When adding the capability:
     - In the **Capability template** field, select **Collibra Protect for Snowflake**.
     - In the **Connection** field, select the connection that you created in Step 3.
   - Do not add more than one Protect for Snowflake capability to the Edge site.

   » Protect is set up. On the main menu, if you click ⠿, **Protect** is shown.

Protect global roles and permissions

The following tables describe the global roles and global permissions that are specific to Collibra Protect.

Global role	Description
Protect Reader	A user who can view Protect with read-only access to data protection standards and data access rules.
Protect Author	A user who can: • Create standards and rules. • Modify or delete only the standards and rules that they created. • View imported policies. • View groups. • Generate audit logs as an individual contributor.
Protect Admin	A user who has the same permissions as a Protect Author. In addition, this user can modify or delete the standards and rules created by others, and access additional APIs.

> Note The **Protect Manager** global role is intended only for the Protect system user.

Global permission	Description
Product Rights > Protect	The **Read-only** license is required for this permission. The permission allows a user to access Collibra Protect. All Protect global roles and the **Edge site** global role have this permission.

Global permission	Description
Protect > Edit	The **Standard** license is required for this permission. The permission allows a user to: • Create standards and rules. • Modify only the standards and rules that they created. • Delete only the standards and rules that they created.
Protect > Administration	The **Standard** license is required for this permission. The permission allows a user to: • Create standards and rules. • Modify all standards and rules. • Delete all standards and rules.

Open Protect

This topic describes how to open Collibra Protect and what is shown on the **Protect** landing page.

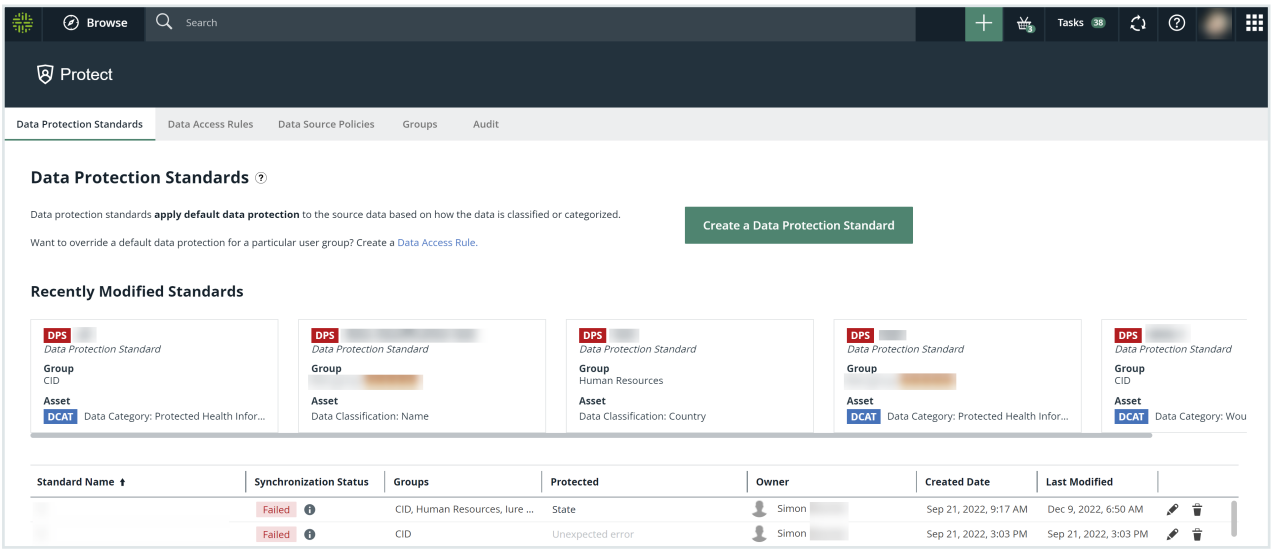
Requirements and permissions

You have a global role that has the **Protect** global permission.

Steps

On the main menu, click , and then click **Protect**.

» The **Protect** landing page opens.



Protect landing page

The following table describes the tabs that are shown on the **Protect** landing page depending on your role.

Tab	Description
Data Protection Standards	Data protection standards to define data source access to data types based on data categories, data attributes, or data classifications.
Data Access Rules	Data access rules to grant specific groups different accesses to the same data in business processes, data categories, or data sets. Note Data access rules take priority over data protection standards.
Data Source Policies	Policies that are active in the data source tables.
Groups	Groups that are mapped to the roles in data sources for use in data protection standards and data access rules.
Audit	Option to generate an audit log of the ingested data from the data sources.

Protect groups

You must create at least one Protect group before creating a data protection standard or a data access rule. Each Protect group is associated with a role in the data source provider.

> Note In BigQuery, *roles* are referred to as *principals*.

The **Groups** tab in Protect contains an overview of the Protect groups that are created for standards and rules. The table on the **Groups** tab contains the Protect groups that are active in the data source.

This topic describes how to create a Protect group and what is shown on the **Groups** tab in Protect.

Create a Protect group

Requirements and permissions

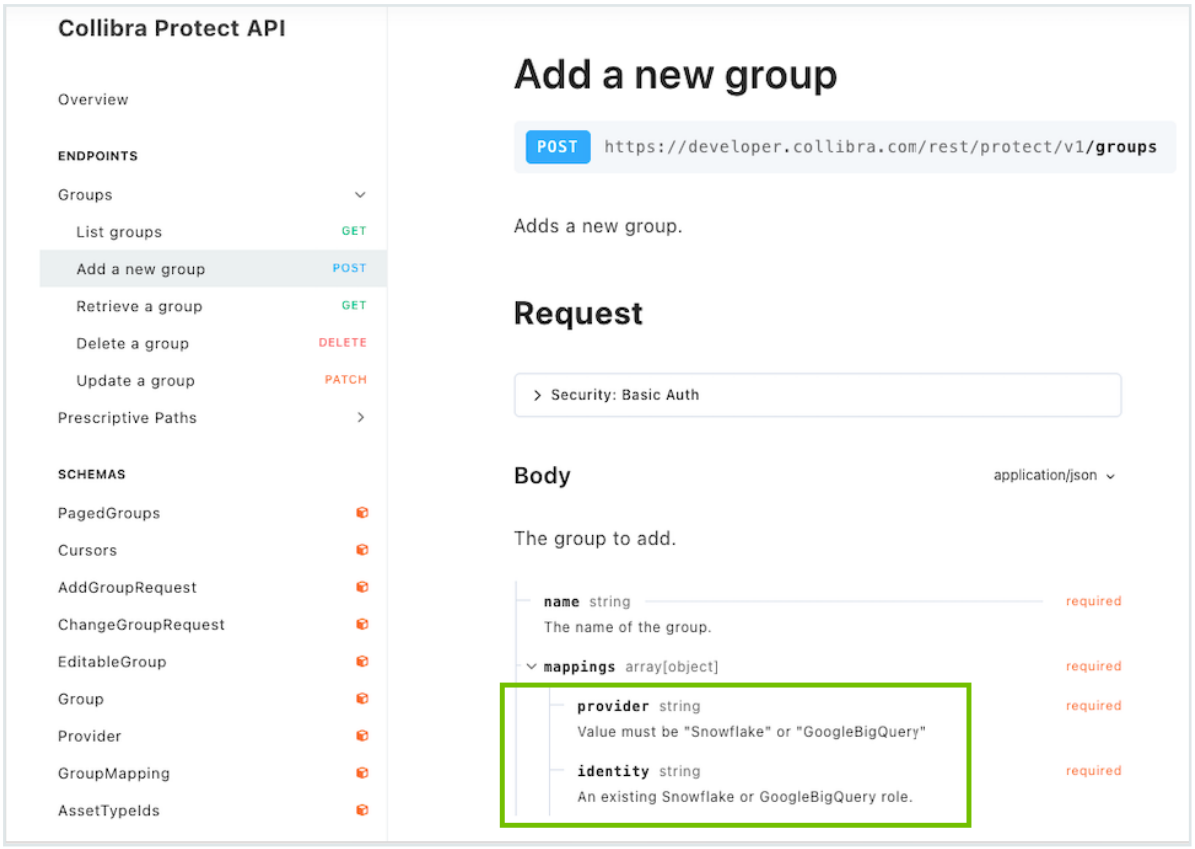
You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

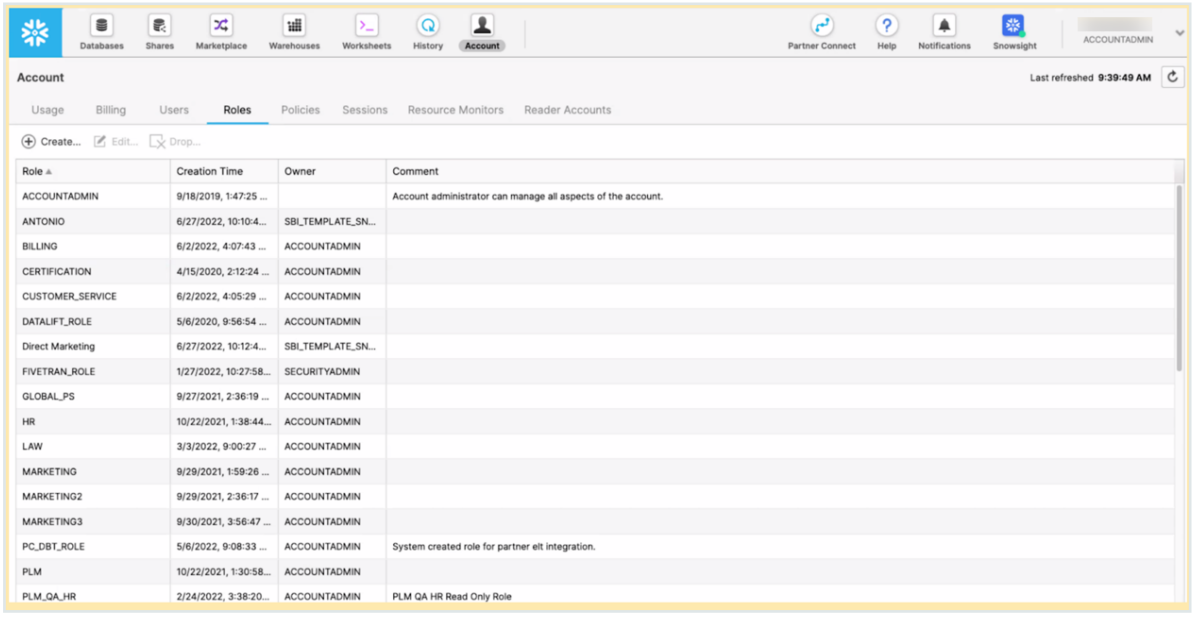
1. Open Protect.
2. On the **Groups** tab, click **Collibra Protect Group API**.
3. For the next steps, go to Add a new group.

Details

- When creating a Protect group, you are prompted to specify the data source provider (**AWSLakeFormation**, **Databricks**, **GoogleBigQuery**, or **Snowflake**) and the existing role from the provider to map the role to the group.

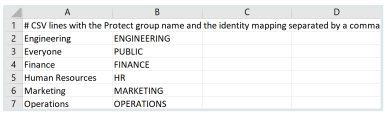


- The following image shows the roles in Snowflake.

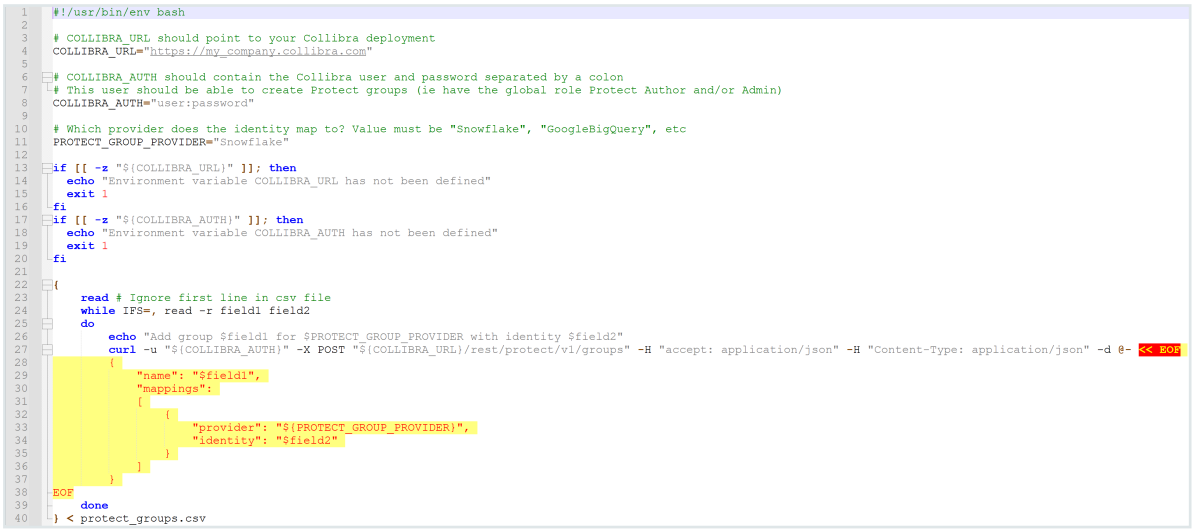


- The following images show a CSV file (named **protect_groups.csv**) that contains Protect groups to be added to Collibra, and a bash script that adds those Protect

groups to Collibra for Snowflake.



	A	B	C	D
1	# CSV lines with the Protect group name and the identity mapping separated by a comma			
2	Engineering	ENGINEERING		
3	Everyone	PUBLIC		
4	Finance	FINANCE		
5	Human Resources	HR		
6	Marketing	MARKETING		
7	Operations	OPERATIONS		



```
#!/usr/bin/env bash

# COLLIBRA_URL should point to your Collibra deployment
COLLIBRA_URL="https://my_company.collibra.com"

# COLLIBRA_AUTH should contain the Collibra user and password separated by a colon
# This user should be able to create Protect groups (ie have the global role Protect Author and/or Admin)
COLLIBRA_AUTH="user:password"

# Which provider does the identity map to? Value must be "Snowflake", "GoogleBigQuery", etc
PROTECT_GROUP_PROVIDER="Snowflake"

if [[ -z "${COLLIBRA_URL}" ]]; then
    echo "Environment variable COLLIBRA_URL has not been defined"
    exit 1
fi
if [[ -z "${COLLIBRA_AUTH}" ]]; then
    echo "Environment variable COLLIBRA_AUTH has not been defined"
    exit 1
fi

{
    read # Ignore first line in csv file
    while IFS=, read -r field1 field2
    do
        echo "Add group $field1 for $PROTECT_GROUP_PROVIDER with identity $field2"
        curl -u "${COLLIBRA_AUTH}" -X POST "${COLLIBRA_URL}/rest/protect/v1/groups" -H "accept: application/json" -H "Content-Type: application/json" -d @- << EOF
        {
            "name": "$field1",
            "mappings":
            [
                {
                    "provider": "${PROTECT_GROUP_PROVIDER}",
                    "identity": "$field2"
                }
            ]
        }
EOF
    done
} < protect_groups.csv
```

Groups tab

The following table describes the columns that are shown in the table on the **Groups** tab.

Column	Description
Group Name	The name of the group.
System Reference	References to identify the data source provider and the native identifier associated with the group.
Created By	The name of the user who created the group.
Created Date	The date when the group was created.

> Note
> - Multiple Protect groups can be mapped to the same data source identity.
> - Within a single Protect group, only one mapping per data source is supported.

Data protection standards

Data protection standards in Collibra Protect protect your data by masking similar types of data wherever it is stored, through column-based protection.

Create a data protection standard

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.
- You have the **Catalog** global role. This role is required to view data classifications for selection in a data protection standard.

Before you begin

Ensure that Protect groups have been created.

Steps

1. Open Protect.
2. Click the **Data Protection Standards** tab.
3. Click **Create a Data Protection Standard**.
   » The **Create a Data Protection Standard** dialog box appears.
4. Enter the required information.

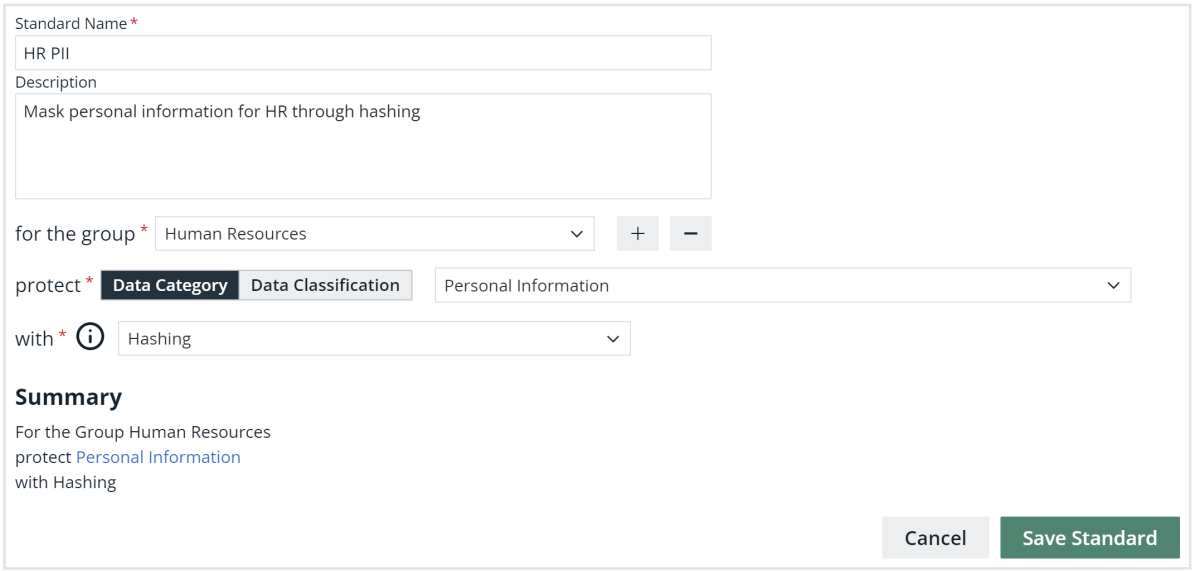
   Details

Field	Description
Standard Name	Enter a name for the data protection standard.
Optional: Description	Enter a description for the data protection standard.

Field	Description
Group	Select the group for the data protection standard. Tip You can add more groups by using the plus icon.
Protect (Data Category/Data Classification)	Click **Data Category** or **Data Classification**, and then select the data category or data classification that you want to protect. Note If the association between the data classification and a column is not yet accepted yet, the standard ignores the column.

Field	Description
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. Tip This field contains the following options: ◦ **Default masking** ◦ **Hashing** ◦ **Show last**

Tip
- The **Summary** section shows a summary of the standard.



5. Click **Save Standard**.
   » A message appears stating that the standard is sent to source, and the standard is shown in the table on the **Data Protection Standards** tab.

Modify a data protection standard

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

  > Note If you have the the **Protect** > **Edit** global permission, you can modify only the data protection standard that you created. If you have the **Protect** > **Administration** global permission, you can modify any data protection standard.

- You have the **Catalog** global role. This role is required to view data classifications for selection in a data protection standard.
- You have the permissions to view the assets that are associated with the data protection standard. Otherwise, the **Unauthorized Asset** value is shown to you when you modify the standard.

Steps

1. Open Protect.
2. In the table, in the row containing the standard that you want to modify, click ✎.
   » The **Edit a Data Protection Standard** dialog box appears.
3. Modify the required information.

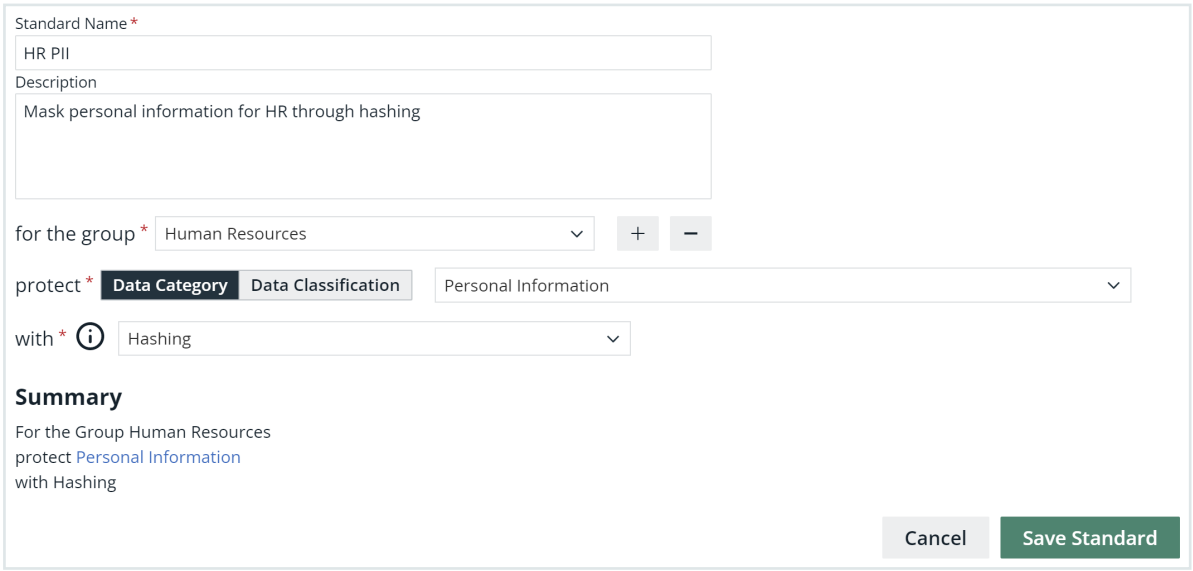
Details

Field	Description
Standard Name	Enter a name for the data protection standard.

Field	Description
Optional: Description	Enter a description for the data protection standard.
Group	Select the group for the data protection standard. Tip You can add more groups by using the plus icon.
Protect (Data Category/Data Classification)	Click **Data Category** or **Data Classification**, and then select the data category or data classification that you want to protect. Note If the association between the data classification and a column is not yet accepted yet, the standard ignores the column.

Field	Description
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. Tip This field contains the following options: ◦ **Default masking** ◦ **Hashing** ◦ **Show last**

Tip
- The **Summary** section shows a summary of the standard.



4. Click **Save Standard**.
   » A message appears stating that the standard is sent to source, and the standard is shown in the table on the **Data Protection Standards** tab.

Delete a data protection standard

Requirements and permissions

You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

1. Open Protect.
2. Click the **Data Protection Standards** tab.
3. In the table, in the row containing the standard that you want to delete, click 🗑.
   » The **Delete Data Protection Standard** dialog box appears.
4. Click **Delete**.
   » A message appears stating that the request to delete the standard is received.

> Tip You can check the status of the standard in the **Synchronization Status** column in the table on the **Data Protection Standards** tab.

Data Protection Standards tab

The **Data Protection Standards** tab in Protect contains an overview of data protection standards. The **Recently Modified Standards** section on the tab shows the 5 last modified data protection standards.

The following table describes the columns that are shown in the table on the **Data Protection Standards** tab.

Column	Description
Standard Name	The name of the standard.
Synchronization Status	The status of synchronization between the standard in Protect and that in the data source.
Groups	The groups for which the standard is created.
Protected	The assets that the standard protects. Tip Depending on your role, you can view the details of an asset by clicking the asset link in this column.
Owner	The name of the user who created the standard.
Created Date	The date and time when the standard was created.
Last Modified	The date and time when the standard was last modified.

Synchronization status

The following table describes the statuses that may be shown in the **Synchronization status** column on the **Data Protection Standard** tab.

> Tip To view the status of the data protection standard in the data source, go to the database of the data source provider.

Synchronization Status	Description
Active	The standard is enforced in the data source.
Pending	The standard is created or modified and is pending synchronization.
Failed	The synchronization of the standard has failed. Tip For more information about the error, click next to the status.
Delete Pending	The standard will be deleted during the next synchronization.
Not Deleted	The standard could not be deleted. Tip For more information about the error, click next to the status.

> Note Protect periodically synchronizes with your data source providers to update the status of the data protection standards in Collibra, except if the status is **Failed**. For more information, go to Synchronization.

Data access rules

Data access rules in Collibra Protect protect your data by managing access and enhancing protection for specific usages. They protect your data by:

- Managing access to the data (access-based protection)
- Masking the data (column-based protection)
- Filtering the data (row-based protection)

Create a data access rule

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.
- You have the **Catalog** global role. This role is required to view data classifications for selection in a data access rule.

Before you begin

Ensure that Protect groups have been created.

Steps

1. Open Protect.
2. Click the **Data Access Rules** tab.
3. Click **Create a Data Access Rule**.
   » The **Create a Data Access Rule** dialog box appears.
4. Enter the required information.

   Details

Field	Description
Rule Name	Enter a name for the data access rule.
Optional: Description	Enter a description for the data access rule.

Field	Description
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.
Asset	Select the data asset that the rule is protecting. Tip ◦ This field contains Business Process, Data Category, and Data Set assets, in addition to assets of custom asset types. ◦ For more information, go to Technical background and Prescriptive paths. ◦ You can add more groups by using the plus icon.

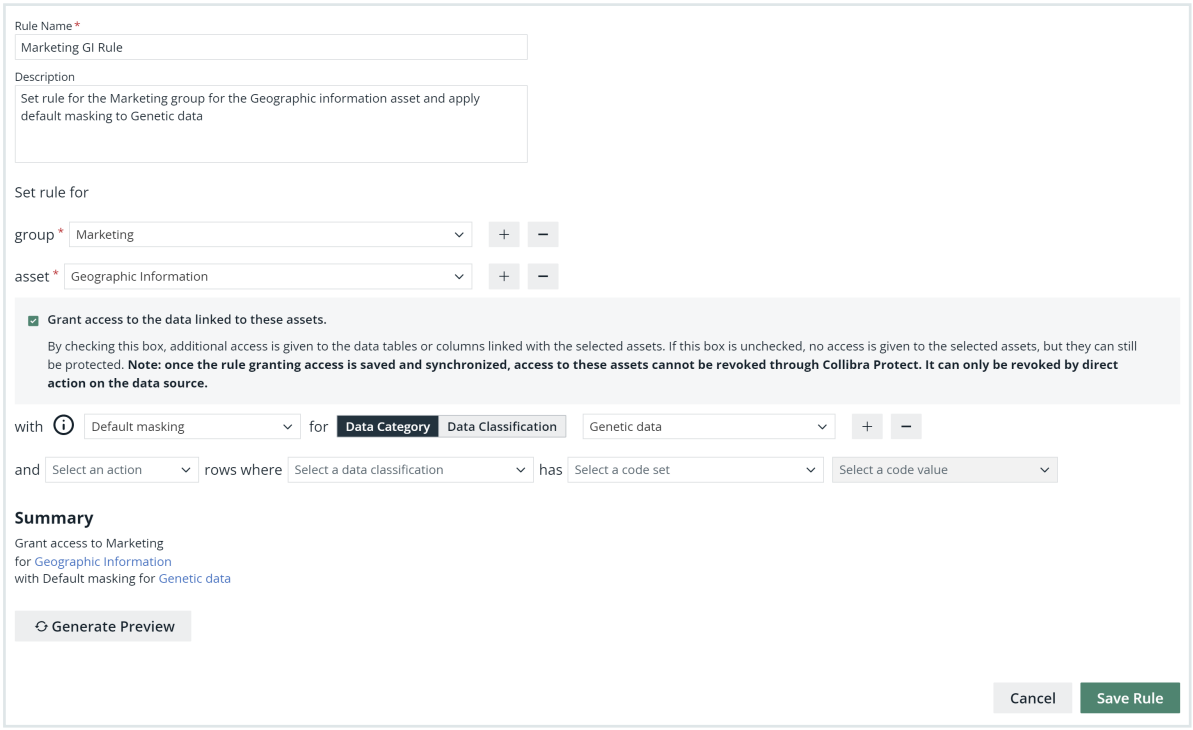
Field	Description
Optional: With (masking option)	◦ Select the type of masking that you want to apply to a data category or data classification. Tip This field contains the following options: ■ **Default masking** ■ **Hashing** ■ **Show last** ■ **No masking** ◦ Click **Data Category** or **Data Classification**, and then select the data category or data classification for the selected masking option. Note If the association between the data classification and a column is not accepted yet, the rule ignores the column.

Field	Description
	Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	a. Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. Tip This field contains the following options: ◦ **Show** ◦ **Hide** b. In the **rows where** field, select the data classification that you want to show or hide. c. In the **has** field, select the code set for the selected data classification. d. In the next field, select the code value for the selected code set. Tip You can add more data classifications for row-filtering by using the plus icon.

Tip

- The **Grant access to the data linked to these assets** checkbox, which is selected by default, is applicable to only Databricks and Snowflake. A selected checkbox indicates that you are allowing the selected groups to access those tables and columns in the database that are linked to the selected assets. If you do not want the selected groups to have this level of access, clear the checkbox.
- The **Summary** section shows a summary of the rule.



5. To preview the rule, click **Generate Preview**.

   Tip The preview shows only the first 1,000 affected columns. The drop-down list box below the **Generate Preview** button is used to switch between the assets that you selected in the rule. Each asset has its own preview table.

6. Click **Save Rule**.

   » A message appears stating that the rule is sent to source, and the rule is shown in the table on the **Data Access Rules** tab.

Modify a data access rule

Requirements and permissions

- You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

  > Note If you have the the **Protect** > **Edit** global permission, you can modify only the data access rule that you created. If you have the **Protect** > **Administration** global permission, you can modify any data access rule.

- You have the **Catalog** global role. This role is required to view data classifications for selection in a data access rule.
- You have the permissions to view the assets that are associated with the data access rule. Otherwise, the **Unauthorized Asset** value is shown to you when you modify the rule.

Steps

1. Open Protect.
2. In the table, in the row containing the rule that you want to modify, click ✎.
   » The **Edit a Data Access Rule** dialog box appears.
3. Modify the required information.

   Details

Field	Description
Rule Name	Enter a name for the data access rule.
Optional: Description	Enter a description for the data access rule.

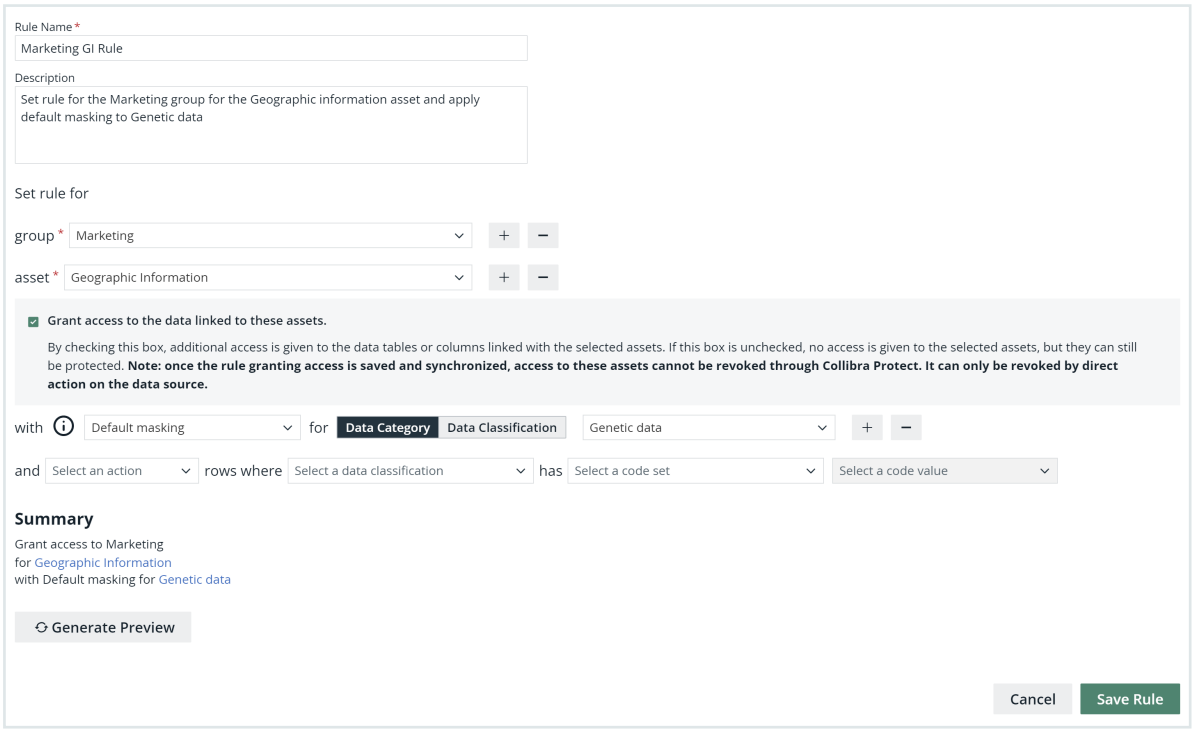
Field	Description
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.
Asset	Select the data asset that the rule is protecting. Tip ◦ This field contains Business Process, Data Category, and Data Set assets, in addition to assets of custom asset types. ◦ For more information, go to Technical background and Prescriptive paths. ◦ You can add more groups by using the plus icon.

Field	Description
Optional: With (masking option)	◦ Select the type of masking that you want to apply to a data category or data classification. Tip This field contains the following options: ■ **Default masking** ■ **Hashing** ■ **Show last** ■ **No masking** ◦ Click **Data Category** or **Data Classification**, and then select the data category or data classification for the selected masking option. Note If the association between the data classification and a column is not accepted yet, the rule ignores the column.

Field	Description
	Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	a. Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. Tip This field contains the following options: ◦ **Show** ◦ **Hide** b. In the **rows where** field, select the data classification that you want to show or hide. c. In the **has** field, select the code set for the selected data classification. d. In the next field, select the code value for the selected code set. Tip You can add more data classifications for row-filtering by using the plus icon.

> Tip
> - A selected checkbox indicates that you are allowing the selected groups to access those tables and columns in the database that are linked to the selected assets. If you do not want the selected groups to have this level of access, clear the checkbox.
> - The **Summary** section shows a summary of the rule.



4. To preview the rule, click **Generate Preview**.

> Tip The preview shows only the first 1,000 affected columns. The drop-down list box below the **Generate Preview** button is used to switch between the assets that you selected in the rule. Each asset has its own preview table.

5. Click **Save Rule**.
   » A message appears stating that the rule is sent to source, and the rule is shown in the table on the **Data Access Rules** tab.

Delete a data access rule

Requirements and permissions

You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

1. Open Protect.
2. Click the **Data Access Rules** tab.
3. In the table, in the row containing the rule that you want to delete, click 🗑.
   » The **Delete Data Access Rule** dialog box appears.
4. Click **Delete**.
   » A message appears stating that the request to delete the rule is received.

> Tip You can check the status of the rule in the **Synchronization Status** column in the table on the **Data Access Rules** tab.

Data Access Rules tab

The **Data Access Rules** tab in Protect contains an overview of data access rules. The **Recently Modified Rules** section on the tab shows the 5 last modified data access rules.

The following table describes the columns that are shown in the table on the **Data Access Rules** tab.

Column	Description
Rule Name	The name of the rule.
Synchronization Status	The status of synchronization between the rule in Protect and that in the data source.
Groups	The groups for which the rule is created.
Affected Assets	The assets that the rule protects. Tip Depending on your role, you can view the details of an asset by clicking the asset link in this column.
Owner	The name of the user who created the rule.
Created Date	The date and time when the rule was created.
Last Modified	The date and time when the rule was last modified.

Synchronization status

The following table describes the statuses that may be shown in the **Synchronization status** column on the **Data Access Rule** tab.

> Tip To view the status of the data access rule in the data source, go to the database of the data source provider.

Synchronization Status	Description
Active	The rule is enforced in the data source.
Pending	The rule is created or modified and is pending synchronization.
Failed	The synchronization of the rule has failed. Tip For more information about the error, click next to the status.
Delete Pending	The rule will be deleted during the next synchronization.
Not Deleted	The rule could not be deleted. Tip For more information about the error, click next to the status.

> Note Protect periodically synchronizes with your data source providers to update the status of the data access rules in Collibra, except if the status is **Failed**. For more information, go to Synchronization.

Data source policies (beta)

Data source policies are the policies that are native to a data source, for example, AWS Lake Formation data filters, BigQuery policy tags, and Snowflake masking policies. Data protection standards and data access rules created in Protect result in policies in the data sources. Protect enforces its standards and rules by creating and applying the data source policies on the physical data layer (tables and columns).

Import data source policies

Requirements and permissions

- You have the **Protect Author** or **Protect Admin** global role.
- The **Manage all resources** global permission is assigned to the **Edge site** global role.

Steps

You can import policies from your data source to Protect by using the Collibra Protect Data Source Policies API. The following is a template of a cURL command that you can use.

```
curl --location --request POST 'https://<collibra-environment-
url>/rest/protect/v1/policies/import' --header 'Authorization:
Basic <user:password encoded in base64>' --header 'Content-
Type: application/json' -d '{"databaseId": "<database-asset-
ID>"}' -v
```

> Note
> In the template:
>
> - Replace the placeholders indicated by "<>" with the actual values for your Collibra environment.
> - *database-asset-ID* refers to the ID of the database asset in Collibra that maps to the database in your data source.

Data Source Policies tab

The **Data Source Policies** tab contains an overview of the native data source policies. The table on the tab contains the policies that are active in the data source. These include both the policies that already exist in your data source and the policies that are automatically created by Protect in your data source.

The following table describes the columns that are shown in the table on the **Data Source Policies** tab.

Column	Description
Policy Name	The name of the policy in the data source.
Policy Logic	The logic that the data source uses to enforce the policy. For example, Snowflake runs an SQL script when you try to access protected data.
Tags	The names of the tags associated with the policy.
Data Source	The data source provider.

Data source providers

Collibra Protect periodically synchronizes with an aggregation of all data protection standards and data access rules. These standards and rules form a data source-agnostic representation containing all databases, schemas, tables, and columns, as well as their protections and accesses. The synchronization process then triggers the Edge capabilities, such as **Collibra Protect for Snowflake**. These Edge capabilities are responsible for translating the representation to actions toward the data source provider using their technology. This process might involve JDBC and REST calls to perform low-level operations to guarantee that the protections and accesses are applied.

Protect for AWS Lake Formation

To protect your AWS Lake Formation data, Protect uses AWS Lake Formation's permissions and data filters. The name of the data category or data classification selected in a data protection standard becomes an AWS Lake Formation tag (LF-tag) with the same name. The tag is then applied to all affected columns.

AWS Lake Formation policies

AWS Lake Formation protects your data by either granting access to or revoking access from one or more columns via permissions and data filters.

> Note AWS Lake Formation does not support data masking.

When you create a data protection standard or data access rule, one or more permissions and data filters are created in AWS Lake Formation. Each permission includes a data filter

to control access to data. Additionally, for a data protection standard, AWS Lake Formation tags (LF-tags) are created and assigned to columns.

> Note In this topic, the term *policies* refers to AWS Lake Formation permissions and data filters.

Data filters

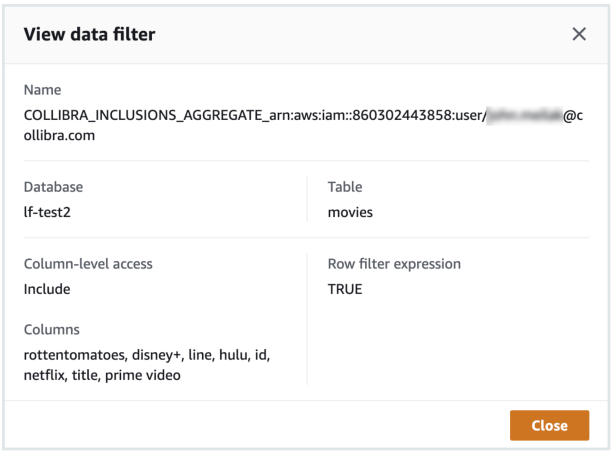
The following table contains the equivalent AWS Lake Formation data filter for a given Protect masking type.

Protect masking type	Equivalent AWS Lake Formation data filter
Default masking	Exclude
Hashing	Exclude
Show last	Exclude
No masking	Include

Each data filter belongs to a specific table in your AWS Data Catalog.

A data filter includes the following information:

- **Name:** The name of the data filter.
- **Table:** The name of the table whose columns are included or excluded.
- **Database:** The name of the database that contains the table.
- **Columns:** A list of columns to include or exclude in query results.
- **Column-level access:** The type of access—either include or exclude—for the columns.
- **Row filter expression:** An expression that specifies the rows to include in query results. The value **TRUE** indicates that all the rows in the table are shown.



> Note Protect safeguards your data in AWS Lake Formation by aggregating all the data protection standards and rules so that a single data filter is created in AWS Lake Formation per table per group. If multiple standards or rules exist for excluding columns, a single data filter with all the columns excluded is created. If a rule is then created for including columns, a data filter with all the columns included is created and the previously excluded columns are no longer considered.

Revoking existing policies for an effective data protection

To effectively protect your AWS Lake Formation data using Protect, you must first revoke any existing AWS Lake Formation policies. Data protection standards and access rules control access to tables and columns for IAM users by creating policies in AWS Lake Formation. To ensure that these policies work as intended, any previous policies granted to those users must be revoked.

> Example Suppose that Joe has full access to the **customers** table. If a data protection standard that hides PII is created and synchronized with AWS Lake Formation, policies are created for Joe. Those policies allow Joe only limited access to the **customers** table by excluding the PII columns. However, the policies will not work if Joe's existing full access to the **customers** table is not first revoked.

AWS Lake Formation group mapping

The Protect group mapping for AWS Lake Formation must follow the syntax for IAM identifiers.

Suppose that you want to create a Protect group named **Sales** that maps to the AWS IAM user **arn:aws:iam::000000000000:user/sales@example.com**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "AWSLakeFormation",
     "identity":
"arn:aws:iam::000000000000:user/sales@example.com"
    }
  ]
}
```

AWS Lake Formation permissions

To perform actions in AWS Lake Formation, Protect uses an AWS connection. This AWS connection must be configured with an AWS IAM user that has the following permissions on all the specified services.

```
{
    "Version": "2012-10-17",
    "Statement":
    [
        {
            "Effect": "Allow",
            "Action":
            [
                "athena:ListDataCatalogs",
                "athena:GetQueryExecution",
                "athena:StartQueryExecution",
                "cloudtrail:LookupEvents",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:GetTableVersions",
                "glue:GetTables",
                "lakeformation:AddLFTagsToResource",
                "lakeformation:CreateDataCellsFilter",
                "lakeformation:CreateLFTag",
                "lakeformation:DeleteDataCellsFilter",
                "lakeformation:DeleteLFTag",
                "lakeformation:GetLFTag",
                "lakeformation:GetResourceLFTags",
                "lakeformation:GrantPermissions",
                "lakeformation:ListDataCellsFilter",
                "lakeformation:ListLFTags",
                "lakeformation:ListPermissions",
                "lakeformation:RemoveLFTagsFromResource",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:GetBucketAcl"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Deny",
            "Action":
            [
                "lakeformation:PutDataLakeSettings"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS APIs

The following table explains the functions of the AWS APIs that are used by Protect for AWS Lake Formation.

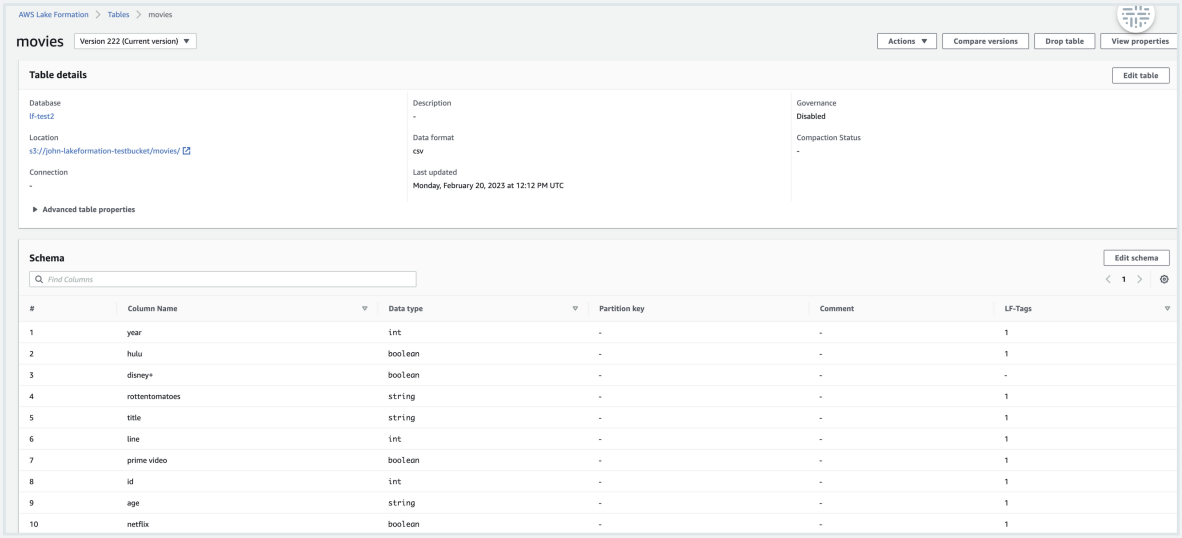
AWS API	Function
athena	Gets information from the AWS Glue Data Catalog. Note Catalog ingestion for AWS databases is performed by using the Amazon Athena service. However, not all the databases ingested from Athena are AWS Lake Formation databases. Hence, Protect needs to identify if a database ingested from Athena is also recognized by AWS Lake Formation. This can be achieved by making an API call to Athena's ListDataCatalogs.
cloudtrail	Shows the audit log in Protect.
glue	Gets a list of tables for a database.
lakeformation	• Creates, deletes, and lists an AWS Lake Formation tag (LF-tag). • Adds and removes an LF-Tag from a resource (column). • Creates, deletes, and lists data filters. • Adds and removes permissions from a resource (table).

AWS Lake Formation examples

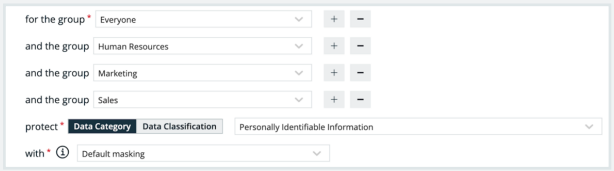
This topic contains examples of how AWS Lake Formation behaves with respect to certain data protection standards and data access rules.

Example

Suppose that a table named **movies** exists in AWS Lake Formation. This table contains Personally Identifiable Information (PII). The PII data category contains all the columns from **movies**, except for **age** and **year**.

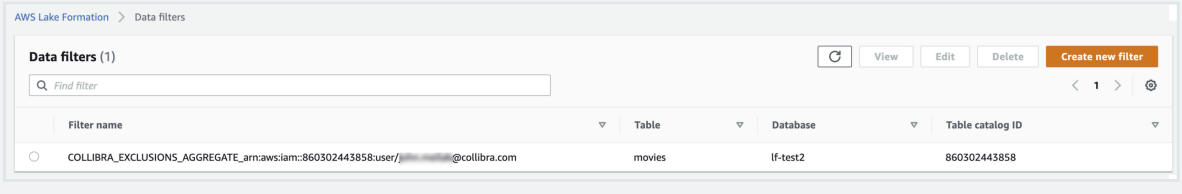


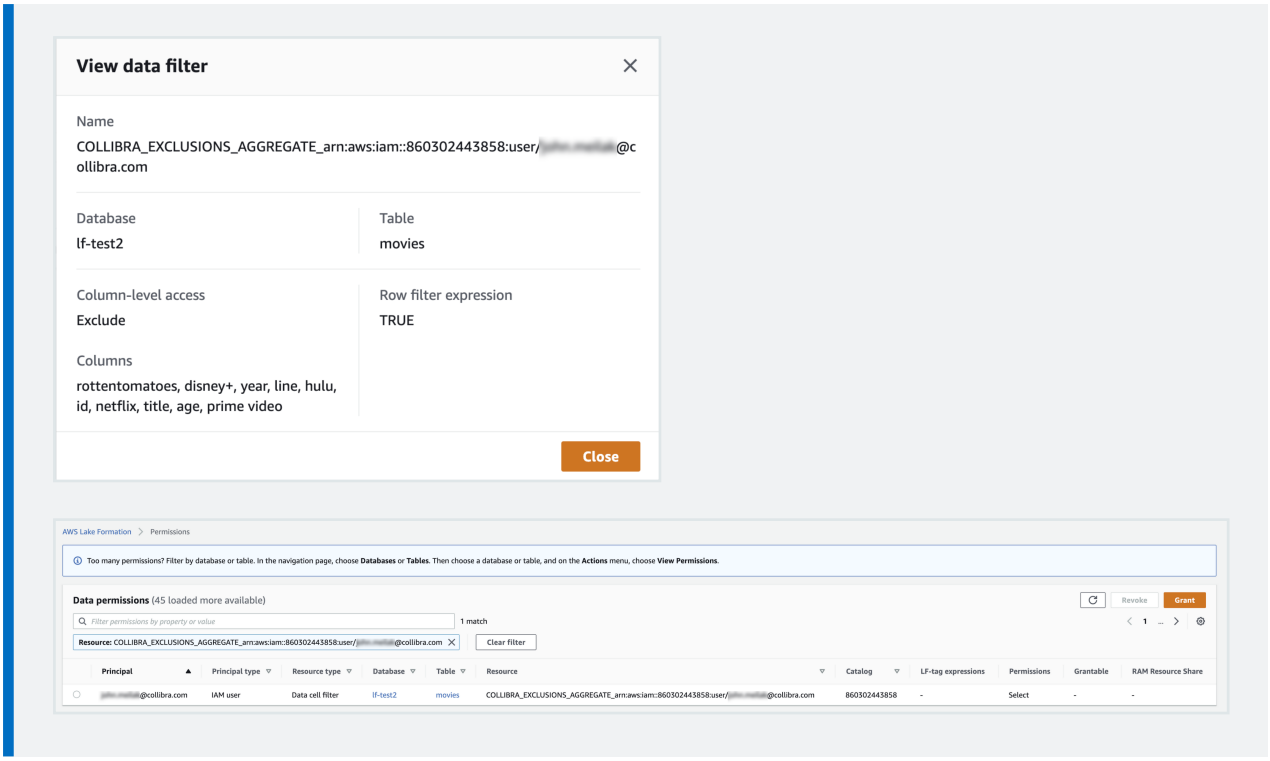
A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



Behavior

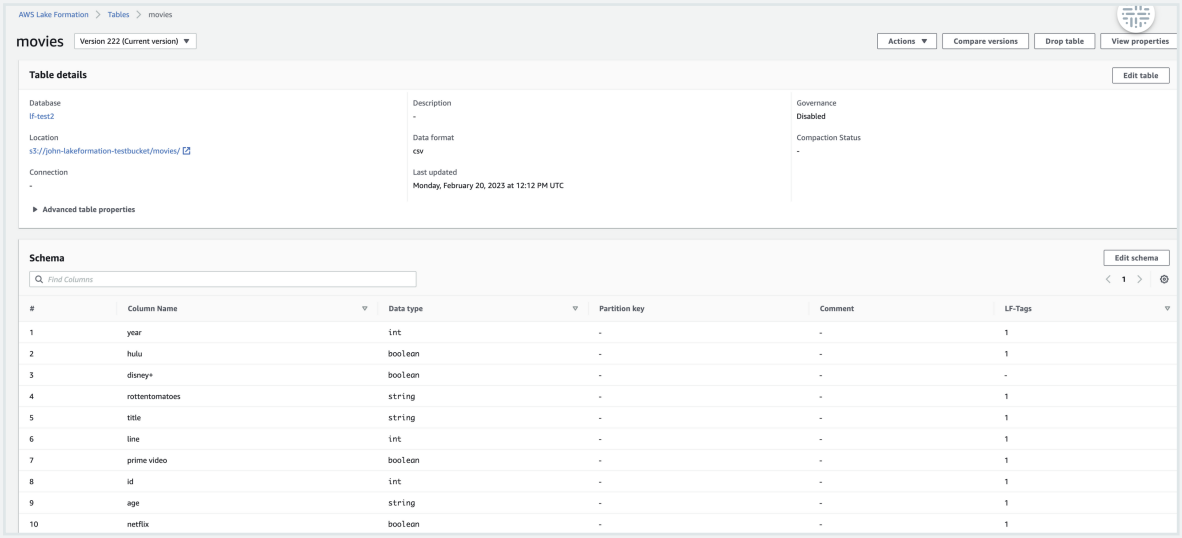
When the standard is synchronized and active, an exclusion data filter is created in AWS Lake Formation. This exclusion data filter hides all the PII columns from the specified groups. The exclusion data filter is named COLLIBRA_EXCLUSIONS_AGGREGATE_/<arn>.



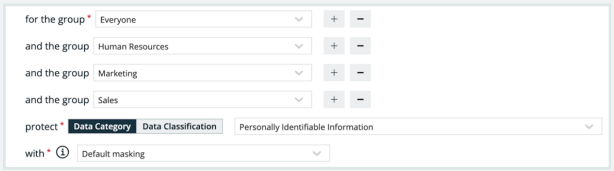

View data filter
Name
COLLIBRA_EXCLUSIONS_AGGREGATE_arn:aws:iam::860302443858:user/[redacted]@collibra.com
Database
lf-test2
Table
movies
Column-level access
Exclude
Row filter expression
TRUE
Columns
rottentomatoes, disney+, year, line, hulu, id, netflix, title, age, prime video
Close
AWS Lake Formation
Permissions
Too many permissions? Filter by database or table. In the navigation page, choose Databases or Tables. Then choose a database or table, and on the Actions menu, choose View Permissions.
Data permissions (45 loaded more available)
Revoke
Grant
Filter permissions by property or value
1 match
Resource: COLLIBRA_EXCLUSIONS_AGGREGATE_arn:aws:iam::860302443858:user/[redacted]@collibra.com
Clear filter
Principal
Principal type
Resource type
Database
Table
Resource
Catalog
LF-tag expressions
Permissions
Grantable
RAM Resource Share
[redacted]@collibra.com
IAM user
Data cell filter
lf-test2
movies
COLLIBRA_EXCLUSIONS_AGGREGATE_arn:aws:iam::860302443858:user/[redacted]@collibra.com
860302443858
-
Select
-
-

Example

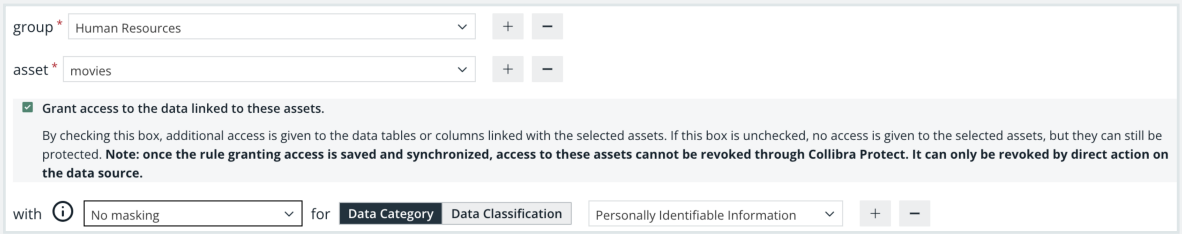
Suppose that a table named **movies** exists in AWS Lake Formation. This table contains Personally Identifiable Information (PII). The PII data category contains all the columns from **movies**, except for **age** and **year**.



A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.

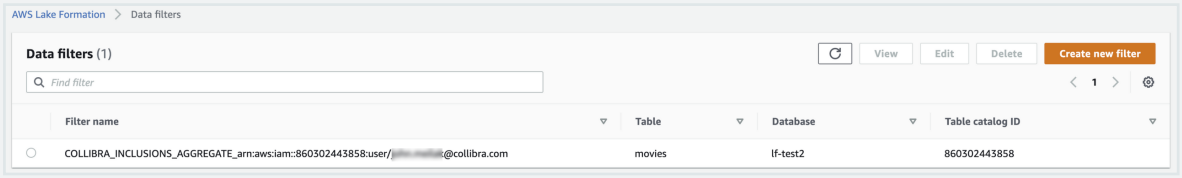


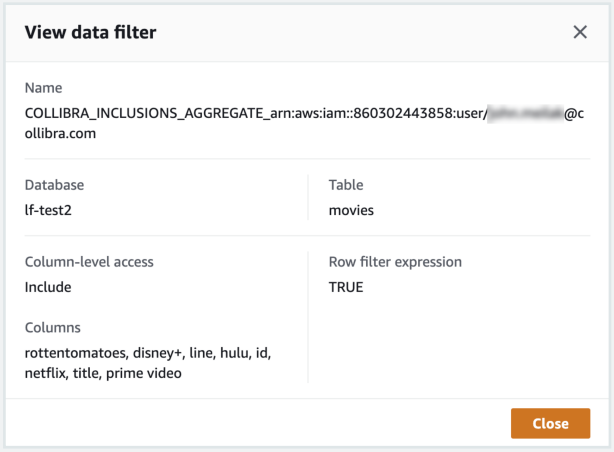
However, a rule that applies to the **Human Resources** group has been created. This rule does not require any masking for the PII columns in **movies**.



Behavior

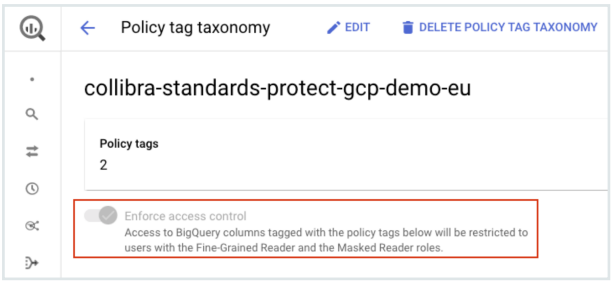
Because the rule takes priority over the standard, when the standard and the rule are synchronized and active, an inclusion data filter resulting from the rule is created in AWS Lake Formation, instead of an exclusion data filter resulting from the standard. This inclusion data filter shows all the PII columns in the **movies** table to the **Human Resources** group. The inclusion data filter is named COLLIBRA_INCLUSIONS_AGGREGATE_/<arn>.





Protect for BigQuery

To protect your BigQuery data, Protect uses Google's policy tags to create tags and assign the tags to the BigQuery columns. These tags control who can access the tagged data. Only the Protect groups specified in your data protection standards and data access rules can access the tagged BigQuery columns.



BigQuery masking rules

Each Protect masking type has an equivalent counterpart in BigQuery called a masking rule. As such, masking rules in BigQuery correspond to masking types in Protect.

> Note The BigQuery masking rules are not the same as the Protect data access rules.

The following table contains the equivalent BigQuery masking rule for a given Protect masking type.

Protect masking type	Equivalent BigQuery masking rule
Default masking	Default masking value
Hashing	Hash (SHA256) > Note BigQuery supports the **Hash (SHA256)** masking rule only for certain columns depending on their data types. If **Hash (SHA256)** cannot be applied to a certain column due to the data type of the column, the following masking rule is applied instead: **Default masking value**.
Show last	Default masking value > Note BigQuery does not support the **Show last** masking type. The **Show last** masking type is supported only on Snowflake.

Protect masking type	Equivalent BigQuery masking rule
No masking	Fine-Grained Reader Note Each Protect group to which you assign standards has an equivalent counterpart in BigQuery called a GCP principal. BigQuery grants the **Fine-Grained Reader** role to the assigned GCP principal to allow the GCP principal to view the data to which no masking is applied in Protect.

BigQuery data types

The following table contains the BigQuery masking rule that Protect supports for a given BigQuery data type.

Summary

- Protect supports the BigQuery **Default masking value** rule for all types of columns.
- Protect does not support the BigQuery **Nullify** rule for any type of column.
- Protect supports the BigQuery **Hash (SHA256)** rule only for the following types of columns: BYTES, STRING.

BigQuery data type	BigQuery masking rule supported by Protect
ARRAY	Default masking value
BIGNUMERIC	Default masking value
BOOL	Default masking value
BYTES	• Default masking value • Hash (SHA256)
DATE	Default masking value
DATETIME	Default masking value
FLOAT64	Default masking value

BigQuery data type	BigQuery masking rule supported by Protect
GEOGRAPHY	Default masking value
INT64	Default masking value
INTERVAL	Default masking value
JSON	Default masking value
NUMERIC	Default masking value
STRING	• Default masking value • Hash (SHA256)
STRUCT	Default masking value
TIME	Default masking value
TIMESTAMP	Default masking value

BigQuery group mapping

The Protect group mapping for BigQuery must follow the syntax for principal identifiers.

Suppose that you want to create a Protect group named **Sales** that maps to the BigQuery group email address **sales@example.com**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "GoogleBigQuery",
     "identity": "group:sales@example.com"
    }
  ]
}
```

BigQuery permissions

To perform actions in BigQuery, Protect uses a GCP connection. This GCP connection must be configured with a service account that has the following permissions.

- `bigquery.dataPolicies.create`
- `bigquery.dataPolicies.delete`
- `bigquery.dataPolicies.get`
- `bigquery.dataPolicies.getIamPolicy`
- `bigquery.dataPolicies.list`
- `bigquery.dataPolicies.setIamPolicy`
- `bigquery.dataPolicies.update`
- `bigquery.datasets.get`
- `bigquery.datasets.getIamPolicy`
- `bigquery.jobs.create`
- `bigquery.rowAccessPolicies.create`
- `bigquery.rowAccessPolicies.delete`
- `bigquery.rowAccessPolicies.list`
- `bigquery.rowAccessPolicies.setIamPolicy`
- `bigquery.rowAccessPolicies.update`
- `bigquery.tables.get`
- `bigquery.tables.getData`
- `bigquery.tables.list`
- `bigquery.tables.setCategory`
- `bigquery.tables.update`
- `datacatalog.categories.getIamPolicy`
- `datacatalog.categories.setIamPolicy`
- `datacatalog.taxonomies.create`
- `datacatalog.taxonomies.get`
- `datacatalog.taxonomies.list`
- `datacatalog.taxonomies.update`
- `logging.logEntries.list`
- `resourcemanager.projects.get`

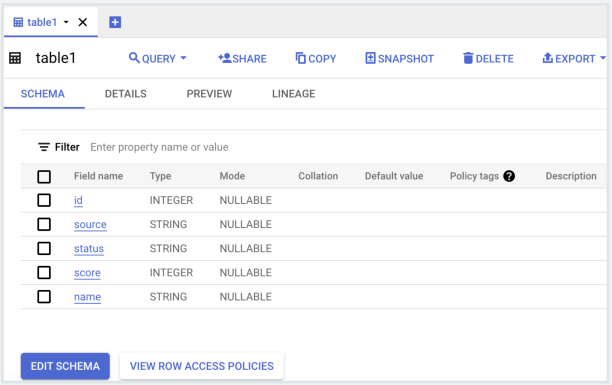
In addition, ensure that the following APIs are enabled for the GCP projects used by Protect:

- BigQuery API
- BigQuery Data Policy API
- Google Cloud Data Catalog API
- Cloud Logging API

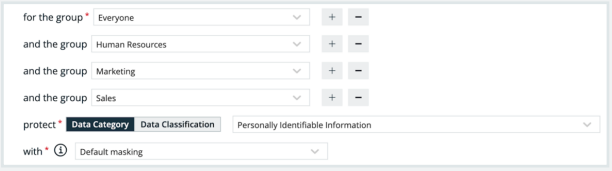
BigQuery examples

This topic contains examples of how BigQuery behaves with respect to certain data protection standards and data access rules.

Example
Suppose that a table named **table1** exists in BigQuery. This table contains Personally Identifiable Information (PII). The PII data category contains all the columns from **table1**.

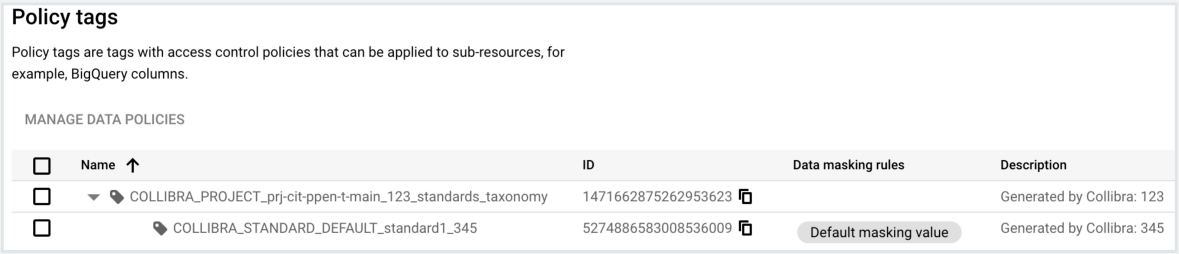


A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.

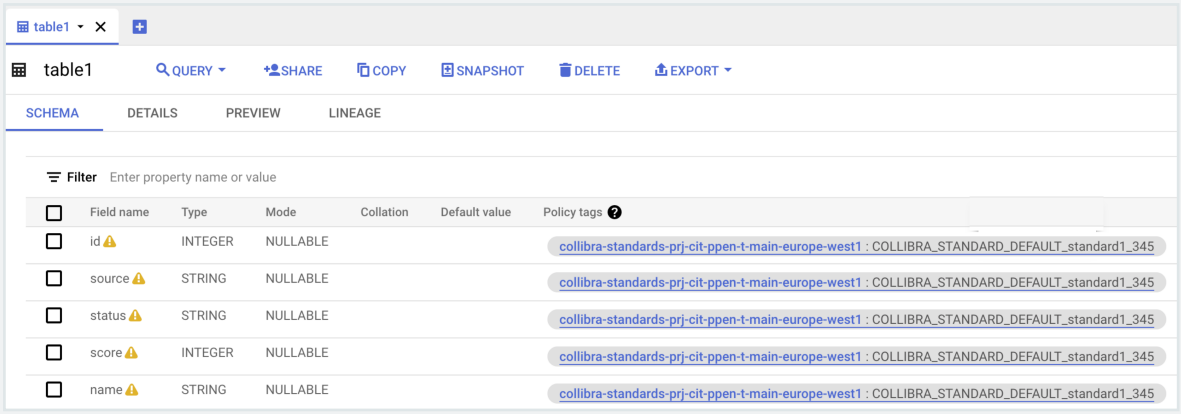


Behavior

When the standard is synchronized and active, a standard policy tag is created in BigQuery's taxonomy. The standard policy tag is named COLLIBRA_STANDARD_DEFAULT_<data protection standard name><data protection standard ID>.

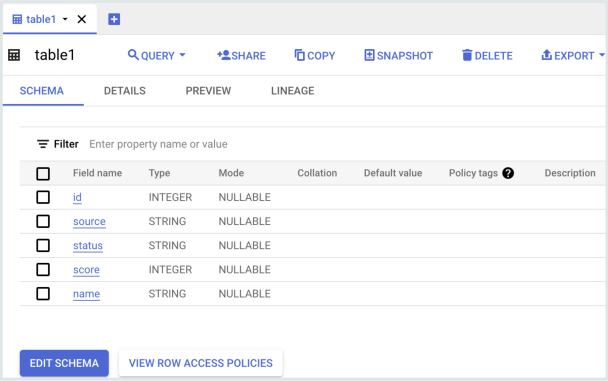


The following image shows how the policy tags are applied to the columns in **table1**.

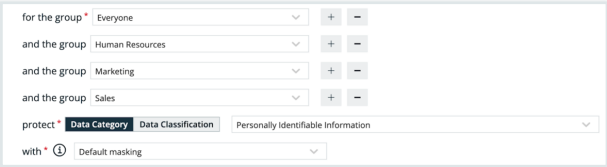


All the columns are assigned the same standard policy tag and are protected by default masking because they belong to the PII data category (selected in the standard).

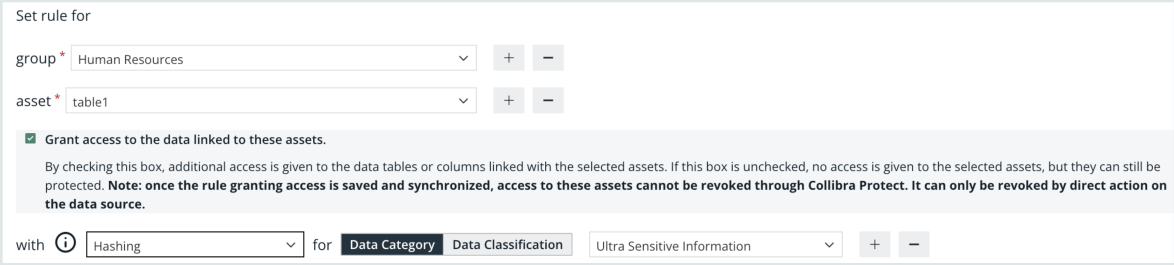
Example
Suppose that a table named **table1** exists in BigQuery. This table contains Personally Identifiable Information (PII) and Ultra Sensitive Information (USI). The PII data category contains all the columns from **table1**, except for **id** and **source**. The USI data category contains only the **status** column.



A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



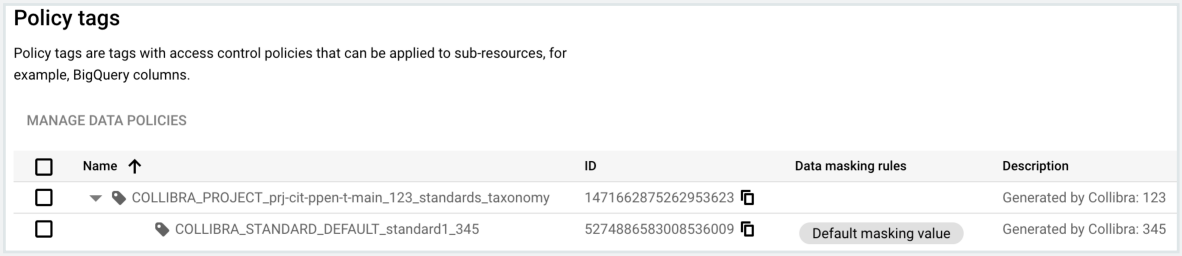
However, a rule that applies to the **Human Resources** group has been created. This rule requires hashing for the USI columns in **table1**.

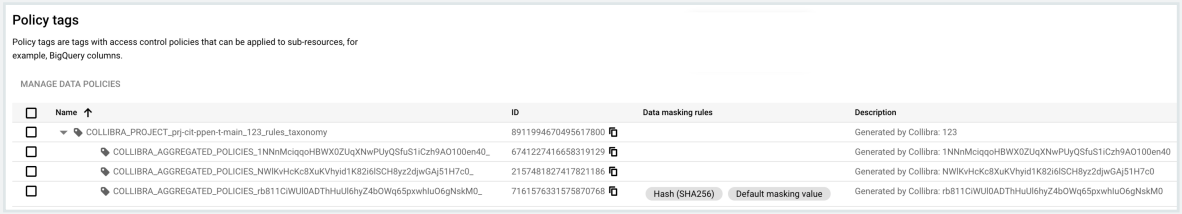


Behavior

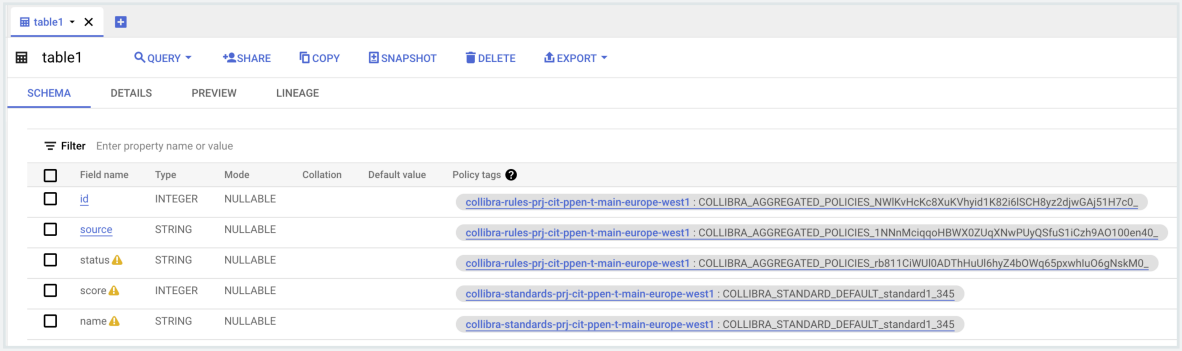
When the standard and rule are synchronized and active, policy tags are created in BigQuery's taxonomy. The standard policy tag is named COLLIBRA_STANDARD_DEFAULT_<data protection standard name><data protection standard ID>. The

rule policy tag is named COLLIBRA_AGGREGATED_POLICIES_ <rulesaccesshash>.





The following image shows how the policy tags are applied to the columns in **table1**.



- The **id** and **source** columns do not belong to the PII data category (selected in the standard) or the USI data category (selected in the rule). Therefore, they are not protected by either the standard or the rule. However, they are still assigned a rule policy tag with the Fine-Grained Reader access to allow users to view the original data.
- The **name** and **score** columns belong to the PII data category (selected in the standard). They are assigned the same standard policy tag and are protected by default masking.
- The **status** column belongs to both the PII data category (selected in the standard) and the USI data category (selected in the rule). Because the rule takes priority over the standard, the **status** column is assigned only the rule policy tag and is protected by hashing.

Protect for Databricks

To protect your Databricks data, Protect uses Databricks's column-based masking functions. These masking functions are applied to columns to enforce data protection.

> Note The Databricks functions are in the public preview stage. Protect for Databricks will be generally available (GA) whenever the Databricks functions reach the GA stage.

Databricks policies

Databricks has the following types of policies:

- Column-based
- Row-based

Each of these policy types can be created either in Protect or on Databricks.

Data access standards created in Protect result in column-based policies on Databricks. Column-based policies are applied directly to the columns on Databricks.

Row filters in data access rules result in row-based policies on Databricks. Row-based policies are applied to the tables on Databricks.

Databricks data types

Databricks provides several functions to transform the data. This topic describes how Databricks transforms the data for a given Protect masking type.

- **Default masking:** Databricks does not support this masking type. Protect, however, uses the default masking type to apply protection to a wide range of data types. A default masking value is applied to each column according to the data type of the column.
  Default masking values for data types

Column data type	Databricks data type	Default masking value
NUMERIC	BIGINT	bigint('0')
BIGNUMERIC	BIGINT	bigint('0')
BYTEINT	BIGINT	bigint('0')
BIGINT	BIGINT	bigint('0')
BINARY	BINARY	binary('00')
VARBINARY	BINARY	binary('00')
BYTES	BINARY	binary('00')
BOOLEAN	BOOLEAN	false
DATE	DATE	1970-01-01
DATETIME	DATE	1970-01-01
DECIMAL	DECIMAL(p,s)	decimal('0.0')
DOUBLE	DOUBLE	double('0.0')
DOUBLE PRECISION	DOUBLE	double('0.0')
REAL	DOUBLE	double('0.0')
FLOAT	FLOAT	float('0.0')
FLOAT4	FLOAT	float('0.0')
FLOAT8	FLOAT	float('0.0')
INT	INT	int('0')
NUMBER	NUMBER	int('0')

Column data type	Databricks data type	Default masking value
BIT	INT	int('0')
INTEGER	INT	int('0')
SMALLINT	SMALLINT	smallint('0')
STRING	STRING	mask('S','*')
CHAR	STRING	mask('S','*')
CHARACTER	STRING	mask('S','*')
VARCHAR	VARCHAR	mask('S','*')
TEXT	STRING	mask('S','*')
TIMESTAMP	TIMESTAMP	1970-01-01 00:00:00.000
TIME	TIMESTAMP	1970-01-01 00:00:00.000
TIMESTAMP_NTZ	TIMESTAMP	1970-01-01 00:00:00.000
TIMESTAMP_LTZ	TIMESTAMP	1970-01-01 00:00:00.000
TIMESTAMP_TZ	TIMESTAMP	1970-01-01 00:00:00.000
TINYINT	TINYINT	tinyint('0')
ARRAY	ARRAY <elementType >	array()
MAP	MAP < keyType,valueType >	map()

Column data type	Databricks data type	Default masking value
STRUCT	STRUCT < [fieldName : fieldType [NOT NULL] [COMMENT str][, ...]] >	struct(0) or struct(0,0) Tip The dynamic value depends on how many fields are defined for the STRUCT datatype.

- **Hashing:** Uses the following Databricks functions:
  - `SHA2` (for strings)
  - `HASH` (for numbers)
  - `right(hash(value), (precision - scale))` (for decimals)
- **Show last:** Uses the following expressions:
  - `right(value,n)` (for strings)
  - `mod(value, cast(power(10,n) AS INT))` (for integers)
  - `regexp_replace(substr(string(value), length(value) - (n-1), n), '^$', '0')` (for floating-point numbers and decimals)

    Tip In the expressions, *value* indicates the content and *n* indicates the number of characters to be shown.
- **No masking:** Returns the raw content.

Note
- You can apply the **Hashing** and **Show last** masking types to only the following Databricks data types: BIGINT, DECIMAL, DOUBLE, FLOAT, INT, SMALLINT, STRING, and TINYINT.
- If a selected masking type cannot be applied to a certain data type–for example, when you attempt to apply the **Hashing** masking type to the DATE data type–the **Default masking** type is applied to the data type to guarantee protection.

Databricks group mapping

The Protect group mapping for Databricks must follow the syntax for principals.

Suppose that you want to create a Protect group named **Sales** that maps to the Databricks group **SALES**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "Databricks",
     "identity": "SALES"
    }
  ]
}
```

Databricks privileges

To perform actions in Databricks, Protect uses an Edge connection. This Edge connection must be configured with a role that is the owner of the catalog or schema in Databricks.

Databricks examples

This topic contains examples of how Databricks behaves with respect to certain data protection standards and data access rules.

Example
Suppose that:

- The **Personally Identifiable Information** (PII) and **Personal Information** (PI) data categories exist in Databricks. These two data categories contain a column named **DOB**.
- A standard that applies to the **HR** group has been created. This standard requires hashing for the PII data category.
- A standard that applies to the **Marketing** group has been created. This standard requires default masking for the PI data category.

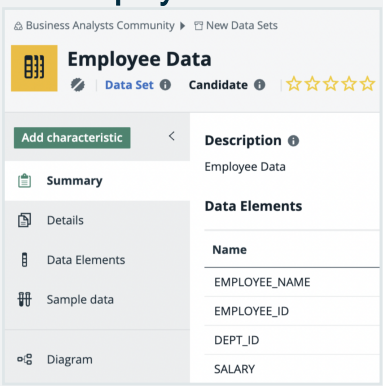
Behavior

When the standards are synchronized and active, a function is created in Databricks for each standard and linked to the **DOB** column. A single column masking policy that combines the two policies is then created and applied to the **DOB** column. This column masking policy includes the protection defined in each standard.

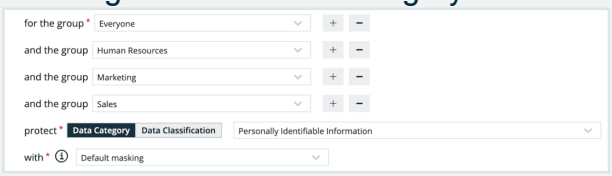
```
CASE
  WHEN (
    current_user() == 'HR'
    or is_account_group_member('HR')
  ) THEN hash(val)

WHEN (
    current_user() == 'Marketing'
    or is_account_group_member('Marketing')
  ) THEN 0
  ELSE val
END
```

Example
Suppose that:

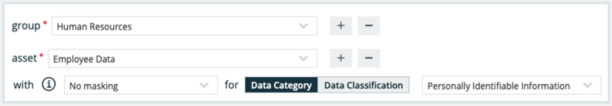
- The **Personally Identifiable Information** (PII) data category exists in Databricks.
- The **Employee Data** data set exists in Databricks. This data set contains PII.



- A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



- A rule that applies to the **Human Resources** group has been created. This rule does not require any masking for the PII columns in the **Employee Data** asset.



Behavior

When the standard is synchronized and active, masking policies are created in Databricks—one policy for each column. The masking functions are named collibra_masking_policy_<asset ID>.

Column	Type	Comment	Tags	Mask
EMPLOYEE_NAME	string			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_9d293821_f1fa_4564_bc20_6fb33331256c
EMPLOYEE_ID	int			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_f0cc1791_5238_4404_9314_ab13f226c605
DEPT_ID	int			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_14cbeca4_0c58_42a2_944b_88838469d140
SALARY	decimal(10,0)			Function name: protect_dev_catalog.tpch_dev.collibra_masking_policy_cfd4ff71_6940_4736_b2d4_8cbc50e51a5b

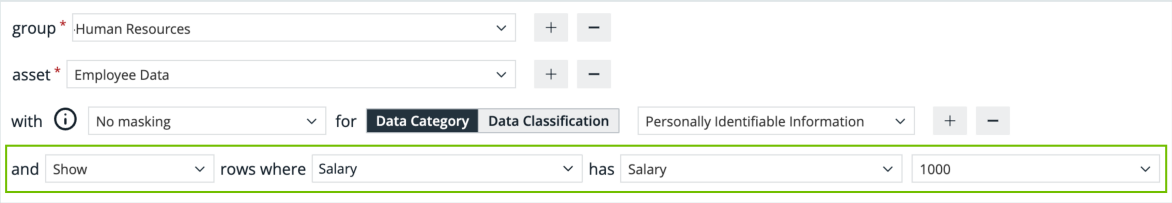
The following image shows a masking policy for the STRING data type. The data that is shown in the policy depends on the masking type selected in the standard and rule. In the policy, `val` indicates the value as it is stored in the table.

```
  WHEN (
    current_user() == 'HR'
    or is_account_group_member('HR')
  ) THEN val
  WHEN (
    current_user() == 'Everyone'
    or is_account_group_member('Everyone')
  ) THEN mask('S', '*')
  WHEN (
    current_user() == 'Marketing'
    or is_account_group_member('Marketing')
  ) THEN mask('S', '*')
  WHEN (
    current_user() == 'Sales'
    or is_account_group_member('Sales')
  ) THEN mask('S', '*')
  ELSE val
```

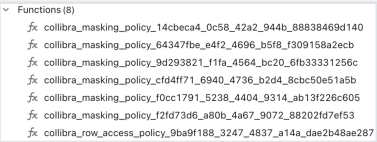
According to the standard, the **Everyone**, **Human Resources**, **Marketing**, and **Sales** groups have masked access to the data. However, according to the rule, the **Human Resources** group has unmasked access to the data. As a result:

- The column is not masked for the **Human Resources** group.
- The column is masked for the **Everyone**, **Marketing**, and **Sales** groups.

Example
Consider the above example with the row filter added, as shown in the following image.



Behavior




```
CREATE
OR REPLACE FUNCTION protect_dev_catalog.tpch_dev.COLLIBRA_
ROW_ACCESS_POLICY_9ba9f188_3247_4837_a14a_dae2b48ae287
(SALARY decimal(10, 0)) RETURN IF(
  (
    (
      current_user() == 'HR'
      or is_account_group_member('HR')
    )
    and SALARY IN (1000)
  ),
  true,
  false
)
```

The row access functions are named collibra_row_access_policy_<asset ID>. The masking and row access policy functions are created at the schema level in Databricks.

Note Protect for Databricks supports Databricks external tables.

Protect for Snowflake

To protect your Snowflake data, Protect uses Snowflake's tag-based masking policies. The name of the data category or data classification selected in a data protection standard becomes a tag with the same name. The tag is then applied to all affected columns to enforce data protection.

Snowflake policies

Snowflake has the following types of policies:

- Column-based
- Row-based
- Tag-based

Each of these policy types can be created either in Protect or on Snowflake.

Data access rules created in Protect result in column-based policies on Snowflake. Column-based policies are applied directly to the columns on Snowflake.

Row filters in data access rules result in row-based policies on Snowflake. Row-based policies are applied to the tables on Snowflake.

Data protection standards created in Protect result in tag-based policies on Snowflake. The tags are subsequently applied to the columns on Snowflake.

Snowflake data types

Snowflake provides several functions to transform the data. This topic describes how Snowflake transforms the data for a given Protect masking type.

- **Default masking:** Snowflake does not support this masking type. Protect, however, uses the default masking type to apply protection to a wide range of data types. A default masking value is applied to each column according to the data type of the column.

Default masking values for data types

Column data type	Snowflake data type	Default masking value
NUMBER	NUMBER	0
DECIMAL	NUMBER	0
NUMERIC	NUMBER	0
INT	NUMBER	0
INTEGER	NUMBER	0
BIGINT	NUMBER	0
SMALLINT	NUMBER	0
TINYINT	NUMBER	0
BYTEINT	FLOAT	0
FLOAT	FLOAT	0
FLOAT4	FLOAT	0
FLOAT8	FLOAT	0
DOUBLE	FLOAT	0

Column data type	Snowflake data type	Default masking value
DOUBLE PRECISION	FLOAT	0
REAL	FLOAT	0
VARCHAR	VARCHAR	*
CHAR	VARCHAR	*
CHARACTER	VARCHAR	*
STRING	VARCHAR	*
TEXT	VARCHAR	*
BINARY	BINARY	00
VARBINARY	BINARY	00
BOOLEAN	BOOLEAN	false
DATE	DATE	1970-01-01
DATETIME	TIMESTAMP_NTZ	1970-01-01 00:00:00.000
TIME	TIME	00:00:00
TIMESTAMP	TIMESTAMP_NTZ	1970-01-01 00:00:00.000
TIMESTAMP_LTZ	TIMESTAMP_LTZ	1969-12-31 16:00:00.000-0800 Note This may change depending on the time zone.

Column data type	Snowflake data type	Default masking value
TIMESTAMP_NTZ	TIMESTAMP_NTZ	1970-01-01 00:00:00.000
TIMESTAMP_TZ	TIMESTAMP_TZ	1969-12-31 16:00:00.000-0800 Note This may change depending on the time zone.
VARIANT	VARIANT	0
OBJECT	OBJECT	{}
ARRAY	ARRAY	[]
GEOGRAPHY	GEOGRAPHY	{"coordinates": [0,0],"type": "Point"} (aka point(0, 0) and visualization can change based on user preferences)

- **Hashing:** Uses the following Snowflake functions:
  - `SHA2` (for strings)
  - `HASH` (for numbers)
- **Show last:** Uses the following expressions:
  - `substr(to_varchar(value), length(value) - n, n)` (for strings)
  - `mod(value, power(10,n))` (for numbers)
  - Tip In the expressions, *value* indicates the content and *n* indicates the number of characters to be shown.
- **No masking:** Returns the raw content.

> Note
> - You can apply the **Hashing** and **Show last** masking types to only the following Snowflake data types: FLOAT, NUMBER, and STRING.
> - If a selected masking type cannot be applied to a certain data type–for example, when you attempt to apply the **Hashing** masking type to the DATE data type–the **Default masking** type is applied to the data type to guarantee protection.

Snowflake group mapping

The Protect group mapping for Snowflake must follow the syntax for identifiers.

Suppose that you want to create a Protect group named **Sales** that maps to the Snowflake role **SALES**. Then, the Protect API to add a new group should have the following syntax.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "Snowflake",
     "identity": "SALES"
    }
  ]
}
```

Snowflake privileges

To perform actions in Snowflake, Protect uses an Edge connection. This Edge connection must be configured with a role that has the following privileges in Snowflake.

Snowflake privilege	Description
`[APPLY MASKING POLICY]`	To apply masking policies. Required for the role performing the actions.

Snowflake privilege	Description
`[APPLY ROW ACCESS POLICY]`	To apply row access policies. Required for the role performing the actions.
`[APPLY TAG]`	To apply tags. Required for the role performing the actions.
`[IMPORTED PRIVILEGES]`	To import privileges. Required for the role performing the actions.
`[MANAGE GRANTS]`	To manage access privileges. Required for the role performing the actions.
`[USAGE]`	To manage usage access on databases and schemas involved in the protection. Required on each database and schema where policies are applied to the role performing the actions.
`[CREATE MASKING POLICY]`	To create masking policies. Required on each schema where policies are applied to the role performing the actions.
`[CREATE ROW ACCESS POLICY]`	To create row access policies. Required on each schema where policies are applied to the role performing the actions.
`[CREATE TAG]`	To create tags. Required on each schema where policies are applied to the role performing the actions.

Example Suppose that a role named **PROTECT** exists in Snowflake and this role is responsible for managing access privileges on all schemas within a database named **DEMO**. To enable the Snowflake **PROTECT** role to perform an action in Snowflake, the following statements can be used.

```
GRANT APPLY MASKING POLICY ON ACCOUNT TO ROLE PROTECT;
GRANT APPLY ROW ACCESS POLICY ON ACCOUNT TO ROLE PROTECT;
GRANT APPLY TAG ON ACCOUNT TO ROLE PROTECT;
GRANT MANAGE GRANTS ON ACCOUNT TO ROLE PROTECT;
GRANT IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE TO ROLE
PROTECT;
GRANT USAGE ON DATABASE DEMO TO ROLE PROTECT;
GRANT USAGE ON ALL SCHEMAS IN DATABASE DEMO TO ROLE
PROTECT;
GRANT CREATE MASKING POLICY ON ALL SCHEMAS IN DATABASE DEMO
TO ROLE PROTECT;
GRANT CREATE ROW ACCESS POLICY ON ALL SCHEMAS IN DATABASE
DEMO TO ROLE PROTECT;
GRANT CREATE TAG ON ALL SCHEMAS IN DATABASE DEMO TO ROLE
PROTECT
```

Snowflake examples

This topic contains examples of how Snowflake behaves with respect to certain data protection standards and data access rules.

Example
Suppose that:

- The **Personally Identifiable Information** (PII) and **Personal Information** (PI) data categories exist in Snowflake. These two data categories contain a column named **DOB**.
- A standard that applies to the **HR** group has been created. This standard requires hashing for the PII data category.
- A standard that applies to the **Marketing** group has been created. This standard requires default masking for the PI data category.

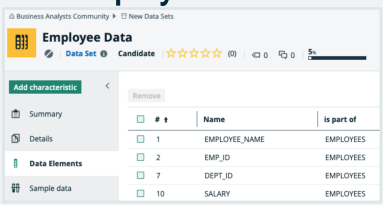
Behavior

When the standards are synchronized and active, a tag policy is created in Snowflake for each standard and linked to the **DOB** column. A single column masking policy that combines the two tag policies is then created and applied to the **DOB** column. This column masking policy includes the protection defined in each standard.

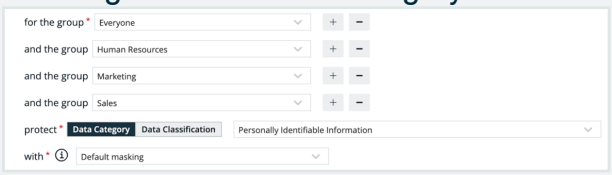
```
CASE
        WHEN CURRENT_ROLE() = 'HR' THEN hash(val)::NUMBER
        WHEN CURRENT_ROLE() = 'MARKETING' THEN 0
        ELSE val
    END
```

Example
Suppose that:

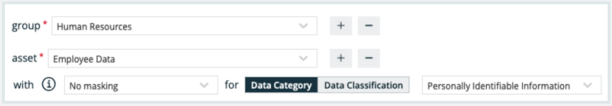
- The **Personally Identifiable Information** (PII) data category exists in Snowflake.
- The **Employee Data** data set exists in Snowflake. This data set contains PII.



- A standard that applies to the following groups has been created: **Everyone**, **Human Resources**, **Marketing**, and **Sales**. This standard requires default masking for the PII data category.



- A rule that applies to the **Human Resources** group has been created. This rule does not require any masking for the PII columns in the **Employee Data** asset.



Behavior

Standard

When the standard is synchronized and active, 14 masking policies are created in Snowflake–one policy for each Snowflake data type. These masking policies are associated with the **Personally Identifiable Information** tag and are created at the schema level. The tag is assigned to those columns that need to be protected. The masking policies are named COLLIBRA/MASKING_POLICY/<asset ID>/<Snowflake type>.

Results Data Preview

Query ID SQL 84ms 18 rows

Filter result... Copy

Row	created_on	name	database_name	schema_name	kind	owner
1	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/ARRAY	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
2	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/BINARY	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
3	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/BOOLEAN	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
4	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/DATE	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
5	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/FLOAT	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
6	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/GEOGRAPHY	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
7	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/NUMBER	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
8	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/OBJECT	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
9	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/STRING	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
10	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIME	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
11	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIMESTAMP	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
12	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIMESTAMP_LTZ	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
13	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/TIMESTAMP_TZ	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN
14	2022-09-06 03:41:13....	COLLIBRA/MASKING_POLICY/28d226cc-0ab0-4d23-b912-985312fb36b1/VARIANT	PROTECT_QA	DEMO	MASKING_POLICY	ACCOUNTADMIN

At runtime, Snowflake fetches the right masking policy based on the column data type.

```
SHOW TAGS;
```

Results Data Preview

Query ID SQL 48ms 2 rows

Filter result... Copy

Row	created_on	name	database_name	schema_name	owner	comment
1	2022-09-06 03:46:10.054...	Personally Identifiable Information	PROTECT_QA	DEMO	ACCOUNTADMIN	Generated by Collibra: 28d226cc-0ab0-4d23-b912-985312fb36b1

The following image shows a masking policy for the STRING data type. The data that is shown in the policy depends on the masking type selected in the standard. In the policy, `val` indicates the value as it is stored in the table.

Details

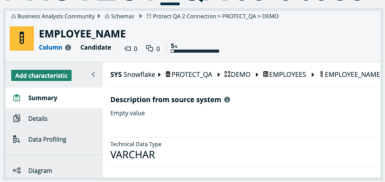
```
CASE
        WHEN CURRENT_ROLE() = 'PUBLIC' THEN '*'
        WHEN CURRENT_ROLE() = 'HR' THEN '*'
        WHEN CURRENT_ROLE() = 'MARKETING' THEN '*'
        WHEN CURRENT_ROLE() = 'SALES' THEN '*'
        ELSE val
    END
```

Rule

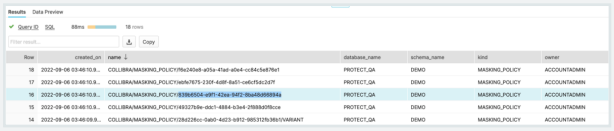
A rule results in a combination of grant instructions, dynamic masking, and row access policies.

The rule grants access of the **Employee Data** data set to the **Human Resources** group, as indicated by the selected **Grant access...** checkbox in the rule. Then, the corresponding Snowflake role for the group can access each database, schema, and table in the data set. In addition, the column masking policy is applied to those columns that need to be protected.

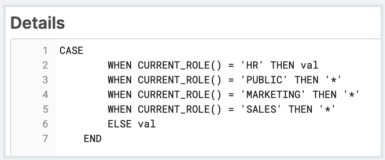
Consider the **EMPLOYEE_NAME** column in the **Employee Data** data set. This column belongs to the **EMPLOYEES** table within the **DEMO** schema in the **PROTECT_QA** database.



In Snowflake, each column that is categorized as PII within the **Employee Data** dataset inherits the masking policy that is applied to the column in Protect. The masking policies created at the schema level are named COLLIBRA/MASKING_POLICY/<asset ID>.



The following image shows the masking policy created for the **EMPLOYEE_NAME** column.



Summary

According to the standard, the **Everyone**, **Human Resources**, **Marketing**, and **Sales** groups have masked access to the data. However, according to the rule, the **Human Resources** group has unmasked access to the data. As a result, the **EMPLOYEE_NAME** column has both a policy tag and a column masking policy applied to it via the standard and the rule, respectively.

In Snowflake, if both a policy tag and a column masking policy exist for the same column, the column masking policy takes priority and the policy tag is not assigned to the column. To ensure that the protection defined in the standard is not ignored, the column masking policy also considers the conditions defined in the standard (policy tag).

Thus, when a standard is created for the **Human Resources**, **Marketing**, and **Sales** groups to mask a column, and when a rule is created for the **Human Resources** group to not mask the same column, the result is as follows:

- The column is not masked for the **Human Resources** group.
- The column is masked for the **Marketing** and **Sales** groups.

Protect audit (beta)

An audit log in Protect contains information about the queries that were run to access the data and the data that was accessed.

This topic describes how to generate an audit log for Protect and what is shown in an audit log.

> Tip
> The information in this topic varies depending on the data source that you select.
>
> Data source
>
> 

Generate an audit log

You can generate an audit log of access records from the data source on the **Audit** page.

> Note The time that it takes for the actions performed in a data source to appear in an audit log in Protect varies from several minutes to hours, depending on the data source.

Requirements and permissions

You have a global role that has the **Protect** > **Edit** or **Protect** > **Administration** global permission.

Steps

1. Open Protect.
2. Click the **Audit** tab.

3. Click **BigQueryDatabricksLake FormationSnowflake**.
4. In the **AWS Region** field, select the hosting region for your Amazon Web Services.
5. Click one of the following buttons: **Today**, **Yesterday**, **A week ago**, **30 days ago**.

   > Tip The start date corresponding to the button that you clicked is shown in the **Start Date** field. Alternatively, you can enter or select a date in the **Start Date** field.

6. Click **Generate Log**.
   » The audit log is generated.

> Important
> - The generation of an audit log may take up to a minute. After clicking **Generate Log**, do not navigate away from the **Audit** page because doing so cancels the audit log generation.
> - The audit log contains the first 1,000 records from the selected start date. If you want to view the remaining records, contact your data source administrator.

Audit log data

The following table describes the columns that are shown in an audit log.

AWS Lake FormationDatabricksBigQuerySnowflake

Column	Description
Query ID	The ID of the query in Snowflake.
Query Start Time	The date and time of the query in Snowflake.
Source User Name	The name of the user in Snowflake who ran the query to access the data.
Direct Objects Accessed	The database object (a table or a view) that was used to access the data.
Base Objects Accessed	The database object that was accessed.

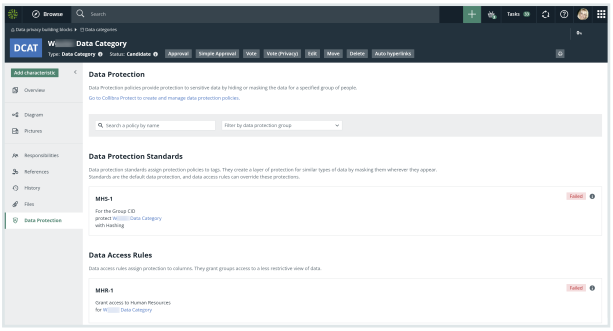
Column	Description
Event Name	The name of the event in AWS Lake Formation.
Date	The date and time of the event in AWS Lake Formation.
Source User Name	The name of the user in AWS Lake Formation who ran the event to access the data.
Event Source	The source of the event, for example, AWS Athena.
Resources	The resources that were accessed.
Method Name	The name of the method in BigQuery.
Date	The date and time of the method in BigQuery.
Principal	The name of the user in BigQuery who ran the method to access the data.
Resource Name	The resource that was accessed.
Action Name	The name of the action in Databricks.
Objects Accessed	The objects that were used to access the data.
Email	The email address of the user in Databricks who ran the action to access the data.
Query Start Time	The date and time of the action in Databricks.

Asset data protection

The asset pages for the following asset types contain the **Data Protection** tab to allow you to view, filter, create, and manage data protection standards and data access rules:

- Business Process
- Data Category
- Data Set
- Custom asset types such as Column, Database, Schema, and Table, derived from the aforementioned asset types via prescriptive paths

> Note Data protection standards support only Data Category assets and data classifications.



View or filter standards and rules

Requirements and permissions

You have the **Protect Reader** global role.

Steps

1. On the asset page (for the one of the aforementioned asset types), click the **Data Protection** tab.
   » Data protection standards and data access rules that are linked to the asset are shown.

Tip
- To filter the standards and rules by name, in the **Search a policy by name** field, enter the name of the standard or rule that you want to view.
- To filter the standards and rules by group, in the **Filter by data protection group** field, select the group for which you want to view the standard or rule.

Create or manage standards and rules

Requirements and permissions

You have the **Protect Author** and **Protect Admin** global roles.

Steps

1. On the asset page (for the one of the aforementioned asset types), click the **Data Protection** tab.
2. Click the following link: **Go to Collibra Protect to create and manage data protection policies.**

Tip For information about how to create and manage data protection standards and data access rules, go to Data Protection Standards tab and Data Access Rules tab.

Why certain standards and rules fail

Certain data protection standards or data access rules may fail due to logical errors. This section describes some of the common scenarios that cause them to fail.

Different types of masking affecting the same column

This topic contains examples to describe how data protection standards and data access rules behave when different types of masking affect the same column.

> Note In the topic, the term *agent* refers to a data category or a data classification.

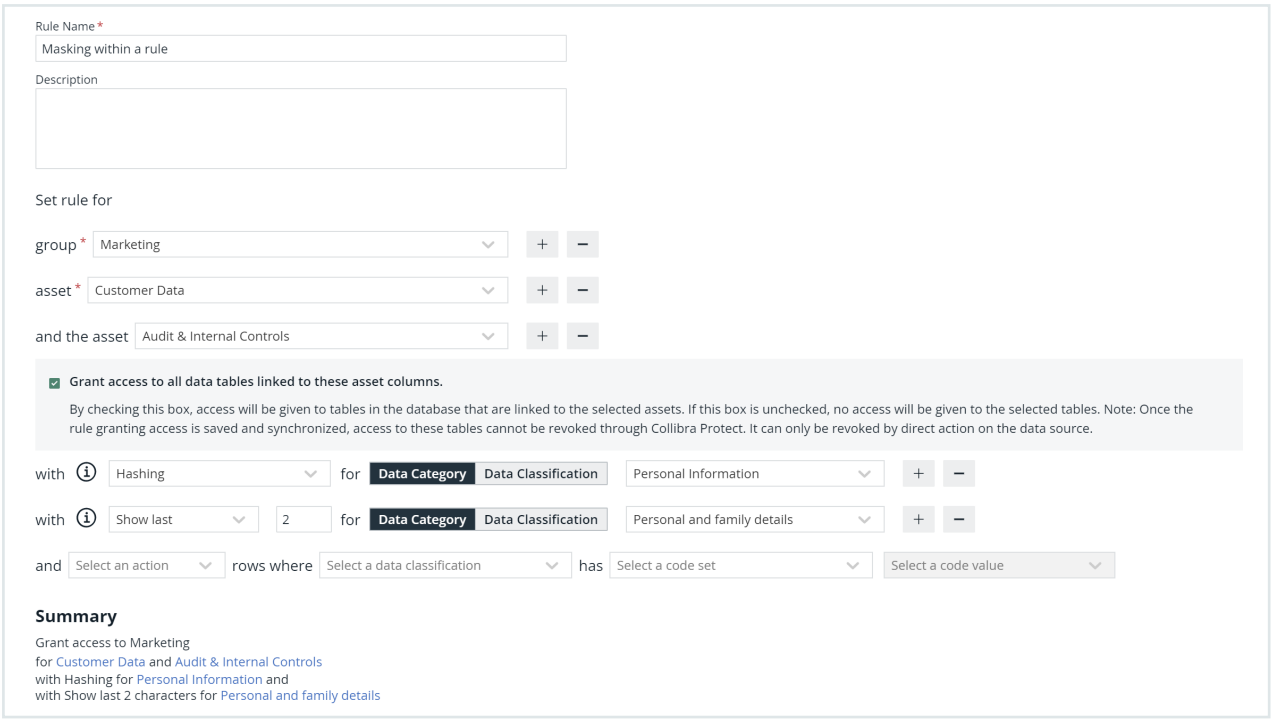
Masking within a rule

Scenario

A rule that is set for a group masks multiple agents using different types of masking, and the agents share the same column. This scenario is applicable regardless of whether the rule applies to a single asset or multiple assets.

Example

Consider a rule that is set for the **Marketing** group. The rule masks the **Personal Information** data category by hashing and masks the **Personal and family details** data category by showing only the last two digits. Suppose that both these data categories share the same column. Then, the rule will fail because the same column cannot be masked using two different masking types for a given group.

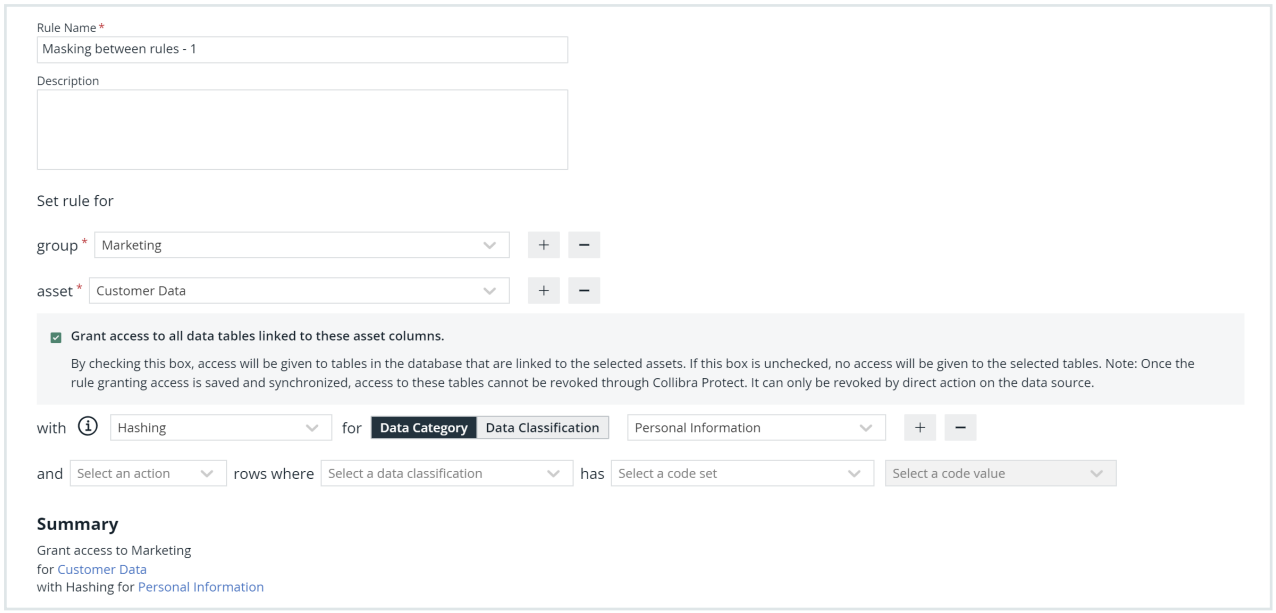


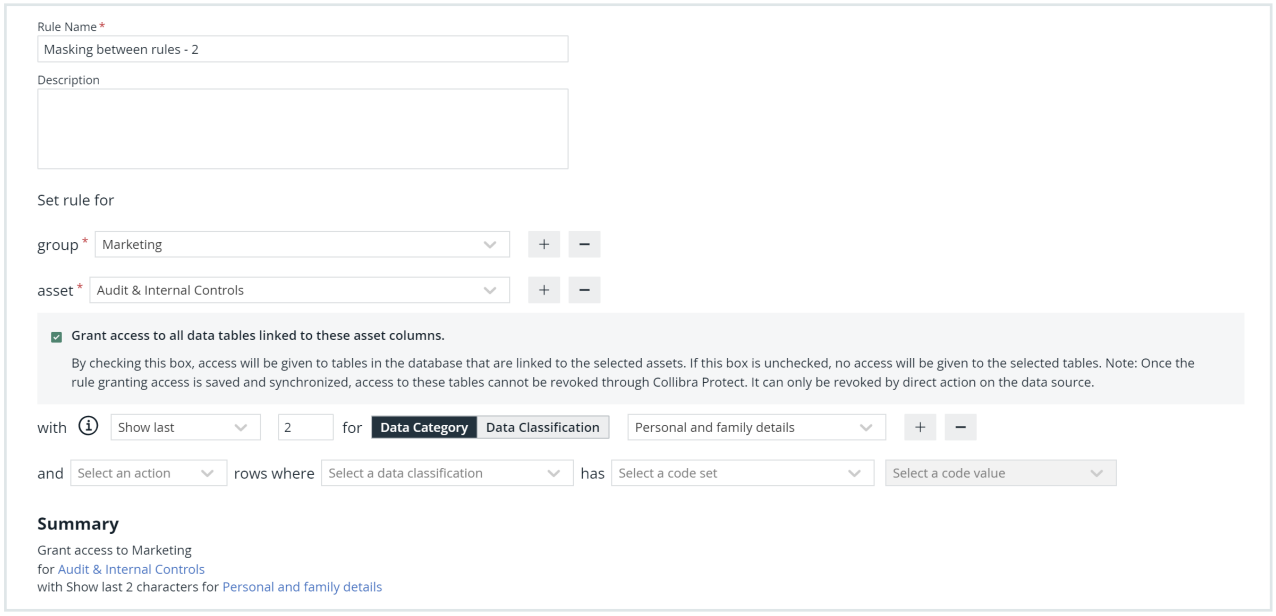
Masking between rules

This scenario is similar to the previous scenario except that this scenario considers two rules, instead of one, that are set for the same group. The masking types for the agents in the two rules are different, and both the agents share the same column. Then, a conflict occurs because the same column cannot be masked using two different masking types for a given group.

When two rules conflict with each other, if the synchronization status of only one of them is **Active**, then the other rule fails. If, however, the synchronization status of both the rules is **Active** or **Pending**, then both of them fail.

This scenario is applicable regardless of whether the agents are the same or different, and regardless of whether the rule applies to a single asset or multiple assets.





Conflicting filters affecting the same column

This topic contains examples to describe how data protection standards and data access rules behave when conflicting filters affect the same column.

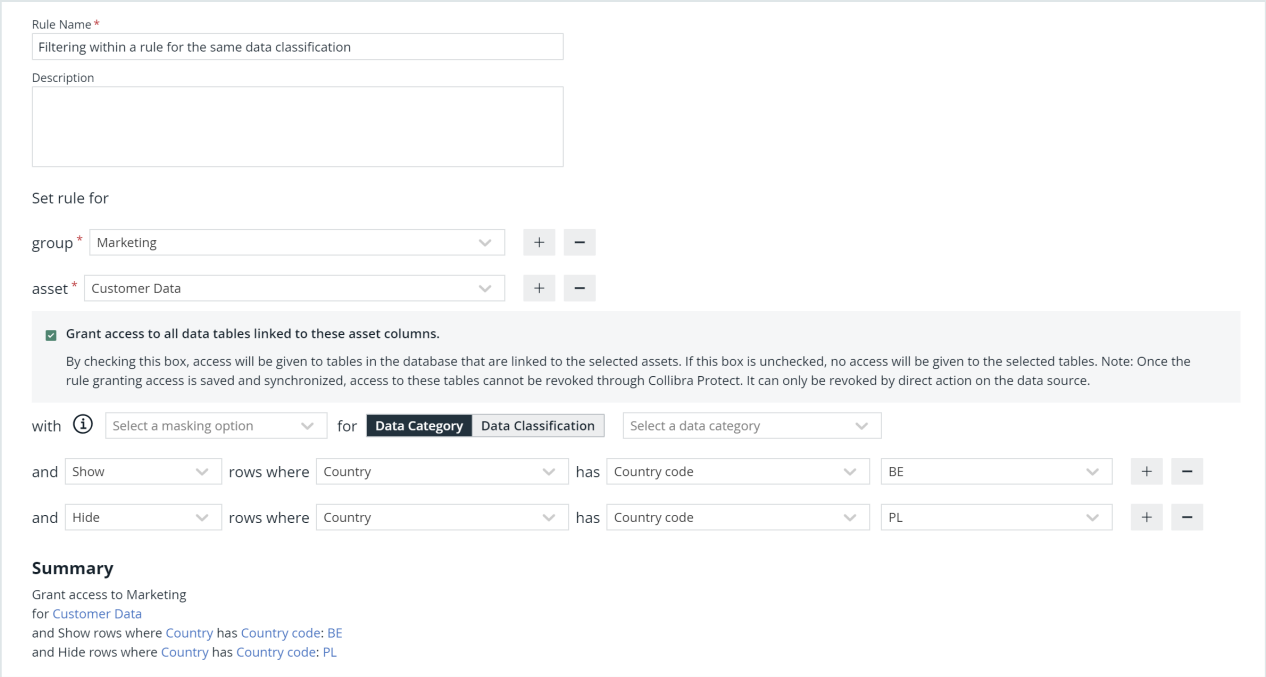
Filtering within a rule for the same data classification

Scenario

A rule that is set for a group contains conflicting filters for the same data classification.This scenario is applicable regardless of whether the rule applies to a single asset or multiple assets.

Example

Consider a rule that is set for the **Marketing** group and the **Customer Data** asset. The rule contains two filters for the **Country** data classification.



If any of the tables in the asset contain a column that is classified as **Country**:

- The first filter shows the rows that contain **BE** in that column.
- The second filter hides the rows that contain **PL** in that column.

Then, this rule will fail because two conflicting filters affect the same column.
When applying a filter for a specific data classification, you must select only one type of action. That is, you can choose to either show rows based on one or more values or hide rows based on one or more values. You must not use the show and hide filter actions together for the same data classification.

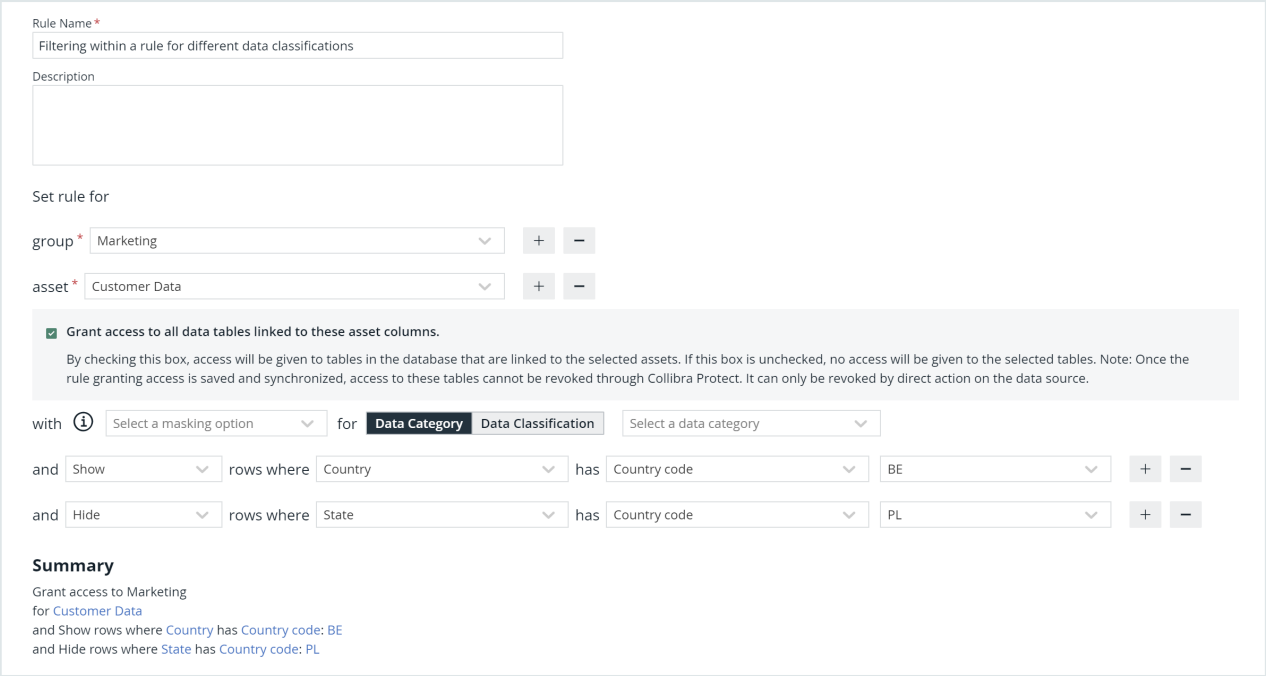
Filtering within a rule for different data classifications

Scenario

A rule that is set for a group contains conflicting filters for different data classifications that share the same column. This scenario is applicable regardless of whether the rule applies to a single asset or multiple assets.

Example

Consider a rule that is set for the **Marketing** group and the **Customer Data** asset. The rule contains two filters: one for the **Country** data classification, and another for the **State** data classification.



If any of the tables in the asset contain columns that are classified as **Country**, the first filter shows only the rows that contain **BE** in those columns.

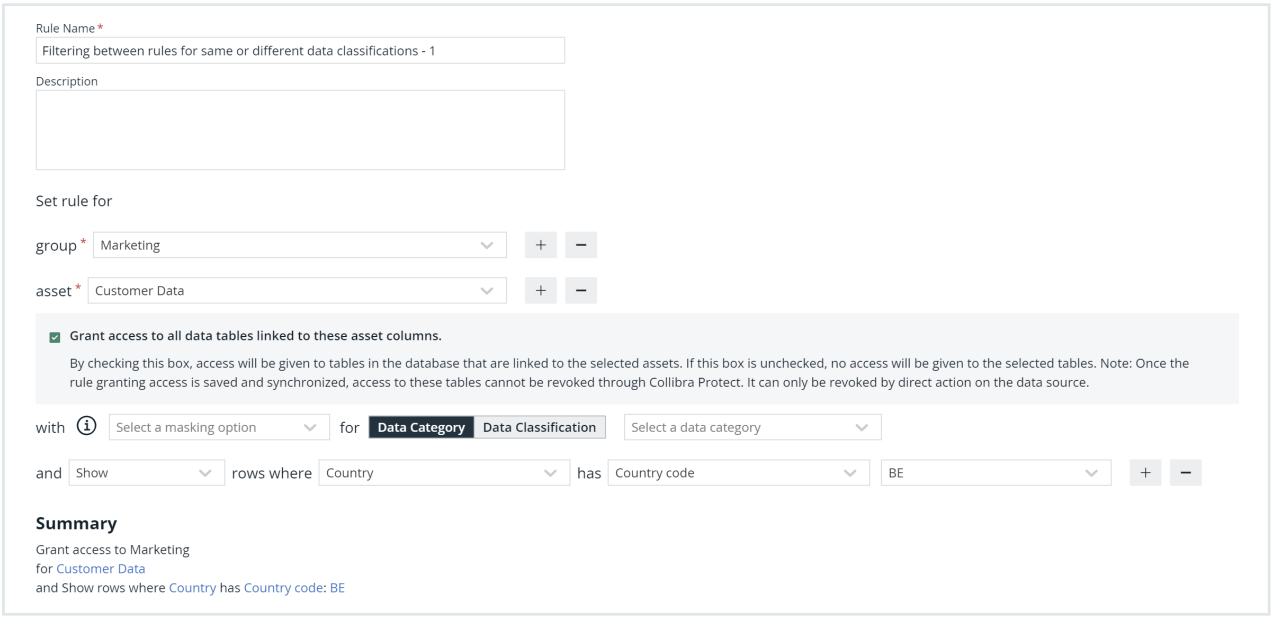
If any of the tables in the asset contain columns that are classified as **State**, the second filter hides only the rows that contain **PL** in those columns.

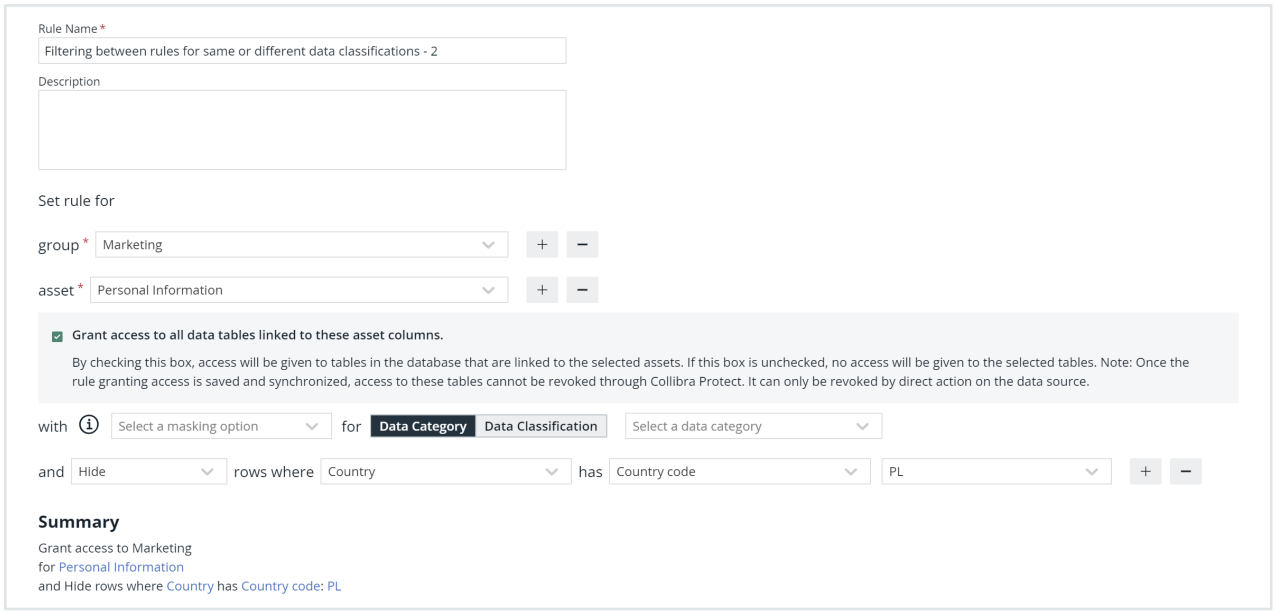
Suppose that a column is classified as both **Country** and **State**. That is, data classifications **Country** and **State** share the same column. Then, this rule will fail because two conflicting filters affect the same column.

Filtering between rules for same or different data classifications

This scenario is similar to the previous scenarios except that this scenario considers two rules, instead of one, that are set for the same group. The filter in one rule is different from the filter in the other rule, and both the filters affect the same column. Then, a conflict occurs because two conflicting filters affect the same column.

When two rules conflict with each other, if the synchronization status of only one of them is **Active**, then the other rule fails. If, however, the synchronization status of both the rules is **Active** or **Pending**, then both of them fail.




Rule Name *
Filtering between rules for same or different data classifications - 2
Description
Set rule for
group * Marketing
asset * Personal Information
Grant access to all data tables linked to these asset columns.
By checking this box, access will be given to tables in the database that are linked to the selected assets. If this box is unchecked, no access will be given to the selected tables. Note: Once the rule granting access is saved and synchronized, access to these tables cannot be revoked through Collibra Protect. It can only be revoked by direct action on the data source.
with
Select a masking option
for
Data Category
Data Classification
Select a data category
and
Hide
rows where
Country
has
Country code
PL
Summary
Grant access to Marketing
for Personal Information
and Hide rows where Country has Country code: PL