Collibra

Collibra Data Intelligence Cloud

Collibra Protect

Version 2023.03

Collibra

Collibra Data Intelligence CloudCollibra Data Governance Center - Collibra Protect

Release date: Thursday, March 2, 2023

Revision date: March 02, 2023

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

https://productresources.collibra.com/docs/collibra/latest/Content/to_collibra-protect.htm

Contents

Contents ii

Scenarios iii

Essentials xxiii

Install Protect xliii

Protect global roles and permissions xlv

Open Protect xlvi

Protect groups xlviii

Data protection standards li

Data access rules lix

Data source policies lxxiv

Data source providers lxxvi

Audit xci

Asset data protection xciii

Why rules or standards may fail xcv

Collibra Protect ciii

Scenarios ciii

Essentials cxxiii

Install Protect cxliii

Protect global roles and permissions cxlv

Open Protect cxlvi

Protect groups cxlviii

Data protection standards cli

Data access rules clix

Data source policies clxxiv
Data source providers clxxvi
Audit cxci
Asset data protection cxciii
Why rules or standards may fail cxcv

Scenarios

This topic describes how Collibra Protect helps you to:

- Use the metamodel graph to build and enforce protection policies on Business Processes, Data Categories, and Data Sets.
- Use classifications to apply a broad coverage of protection mechanisms at the data source.
- Support privacy preferences such as consent management, data subject requests such as access requests, and the right to be forgotten through row-filtering mechanisms.
- Perform an audit of applicable protection at the data source and use reporting to demonstrate compliance where data is stored and consumed.

Discover and classify personal information

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

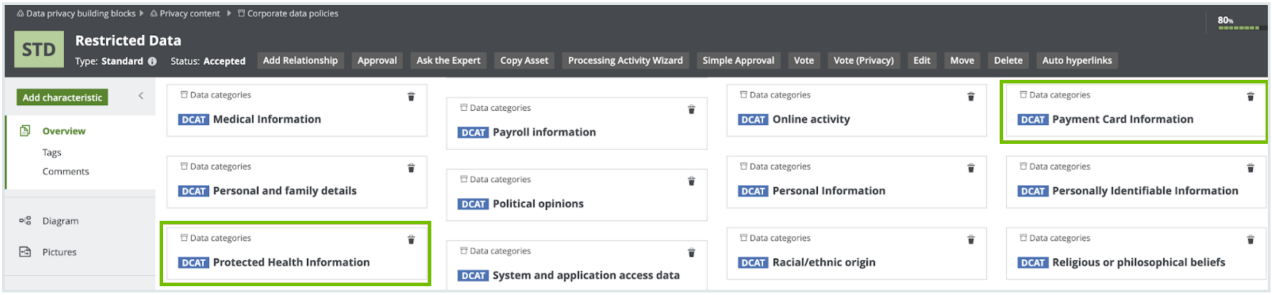
To achieve this, typically, your Privacy team sets up the Data Classification Policy, where they classify the data used in the organization based on the sensitivity or the business criticality of the data. This determines the required levels of security for the applications that store that data or the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

Consider the following three classifications for sensitivity:

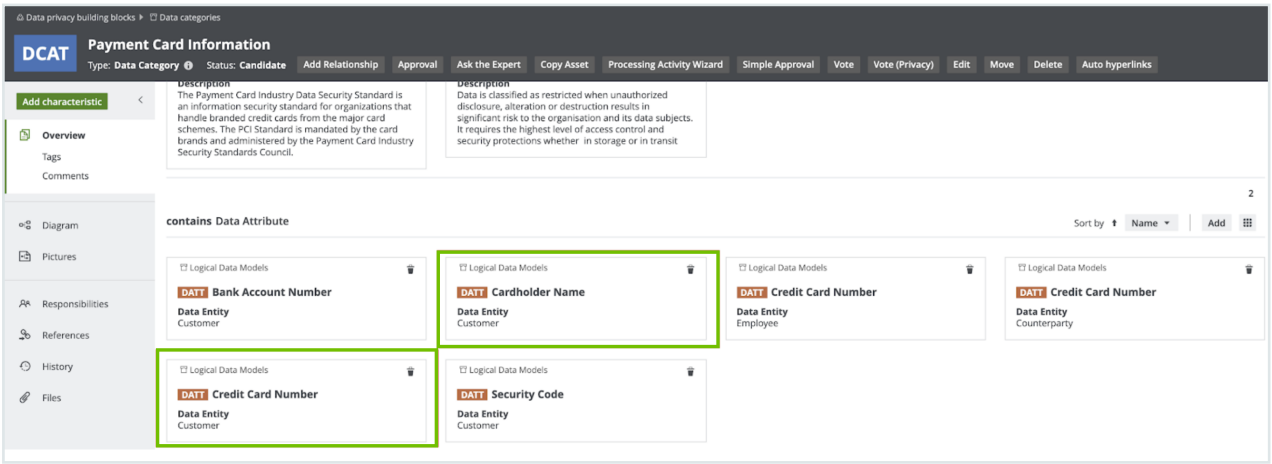
- Public data. which is least sensitive.
- Private data, which is slightly more sensitive than the public data.
- Restricted data, which is the most sensitive data and therefore requires the highest level of access controls and security protection.

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.

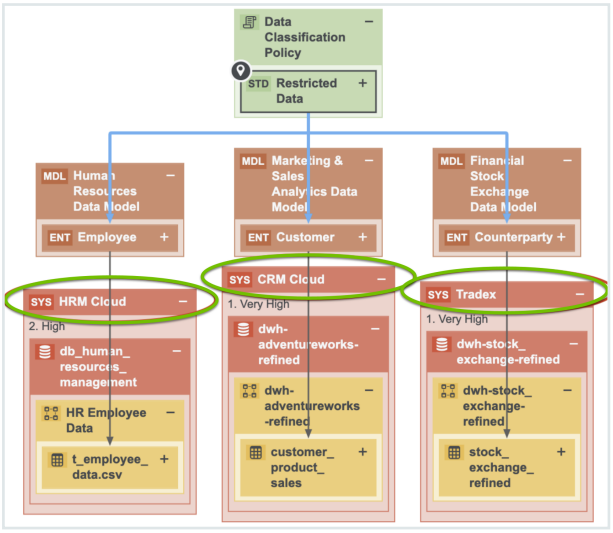


The Privacy team determines the sensitivity and the required security at the data category level as opposed to the column level. At the data category level, the Privacy team then determines what data elements belong to the identified data categories. For example, the Payment Card Information data category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

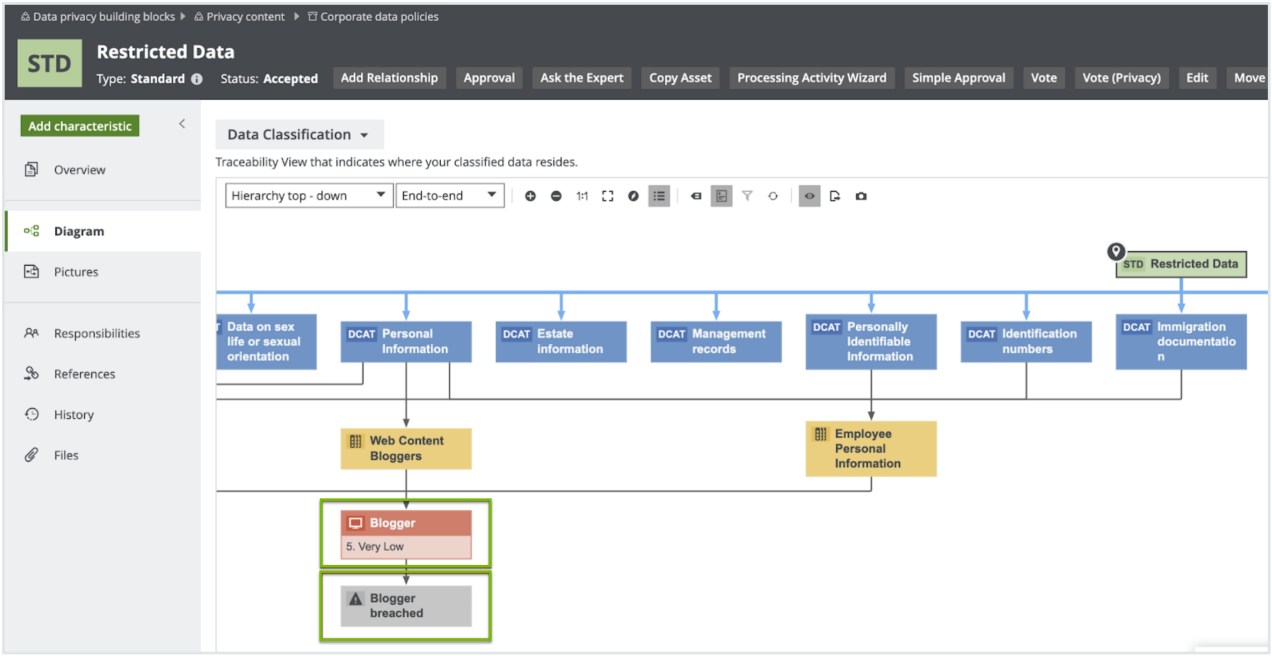
Privacy team and the Governance team. In addition, this allows the automated data classification of the organisation's personal information, which makes views such as the Restricted Data Overview diagram available at the most sensitive data category, Standard Restricted Data.



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 image, an investigation is already ongoing on the potential data breach on Blogger.

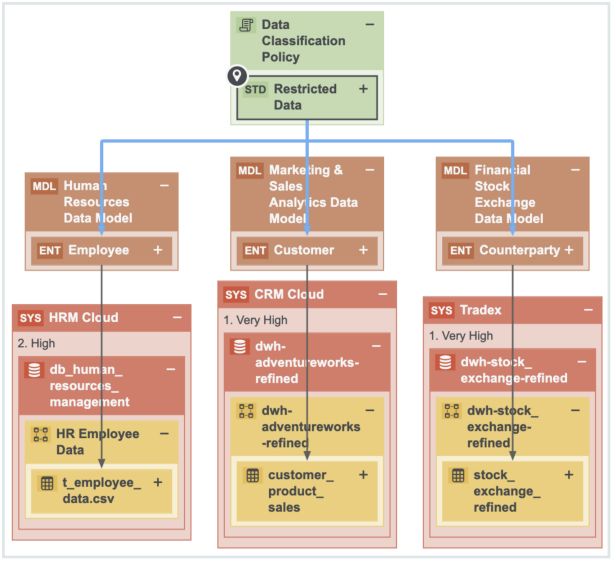


Data classification capabilities and guided stewardship

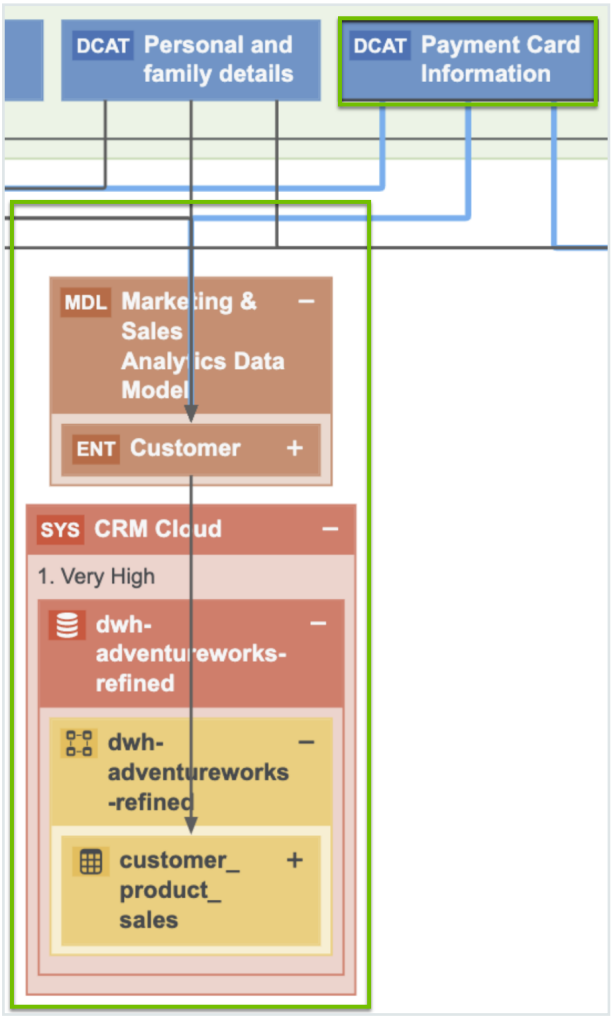
This section describes how Collibra Privacy and Risk leverages the data classification capabilities in Catalog. Thus far, we learned that the Restricted Data standard groups Data Categories, which group Data Attributes. In the example, the Payment Card Information data category contains the Credit Card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

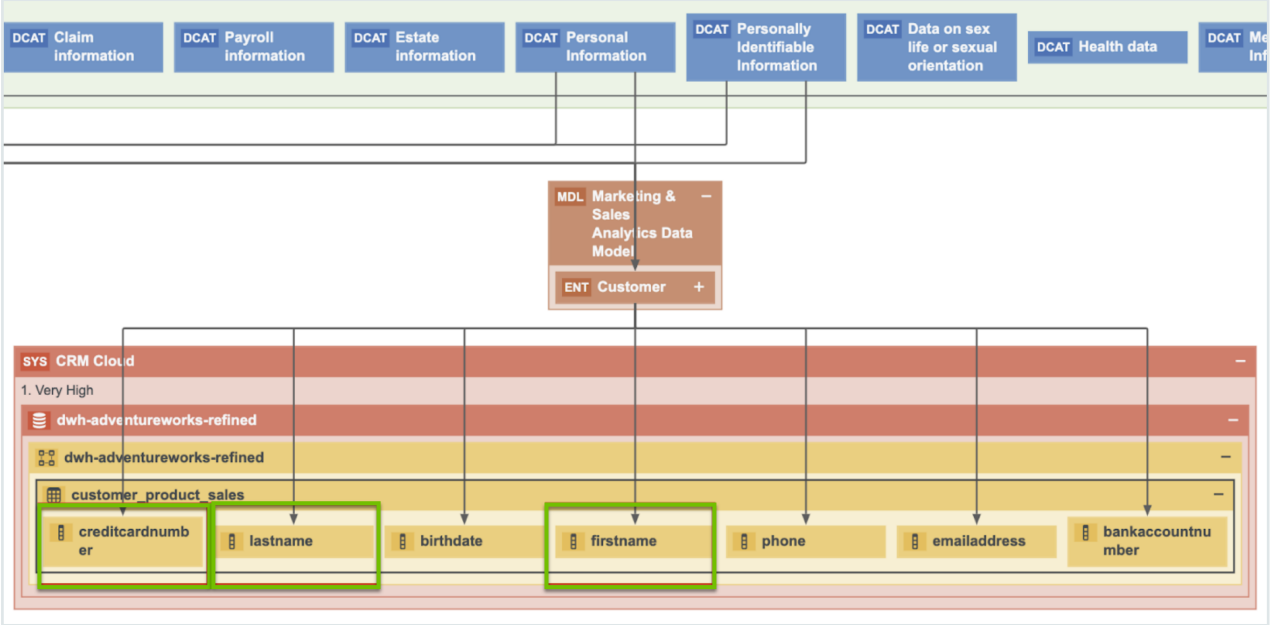


Restricted Data groups multiple data categories. The following image shows the data attributes that the Payment Card Information data category groups.



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.



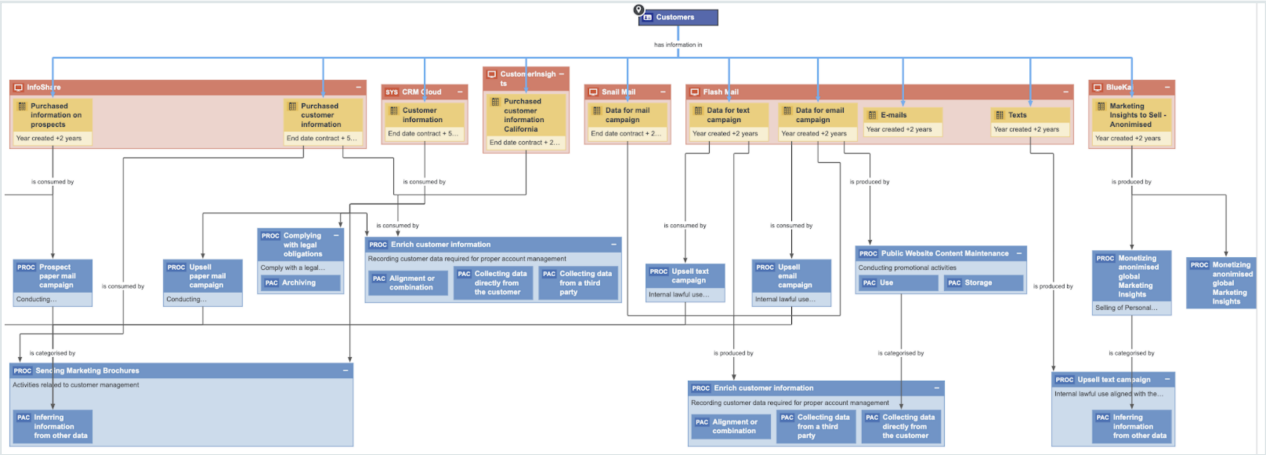
This is an example of how guided stewardship, Catalog's data classification combined with guided stewardship and CollibraPrivacy and Risk, gives you a vertical view on where Personal Information resides.

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.

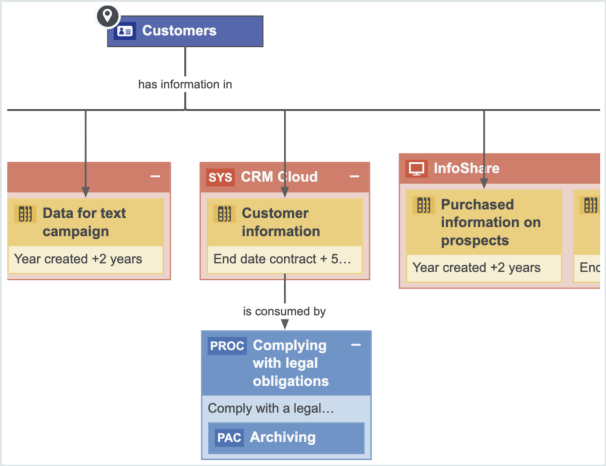
Consider the customer data. Collibra knows where this data resides and how it is being used. This is an outcome of obtaining input from the business users during the onboard of the Business Processes where the users are asked what data they use, which applications they use, for what purpose the data is used. 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.



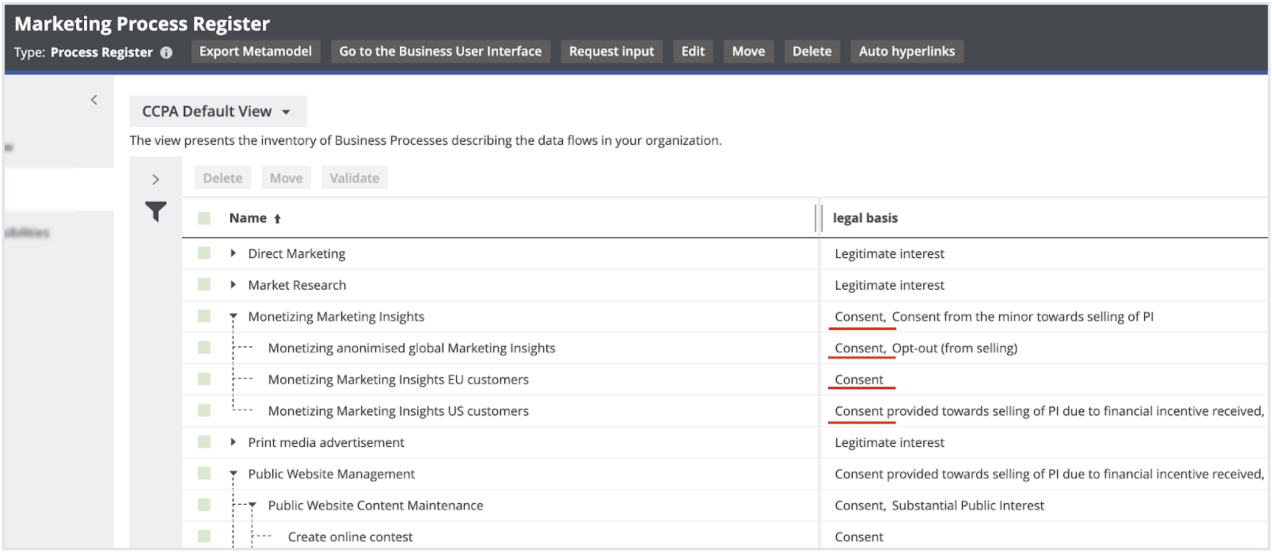
The above image shows where data resides and why it is used. It shows all the applications that contain customer data, and also the related retention periods, which can

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 the data subject requests. 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 process register, so that you can see all the processes that rely on the consent and the data categories for which you need consent.

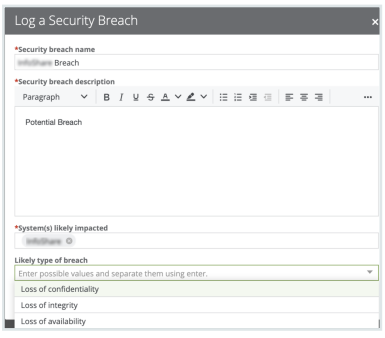


These are stored in the data sets that can also contain granular information, such as the individual data elements for which you want to obtain consent—this combines the information about which business processes require consent and the data categories for which you need consent to process all information 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

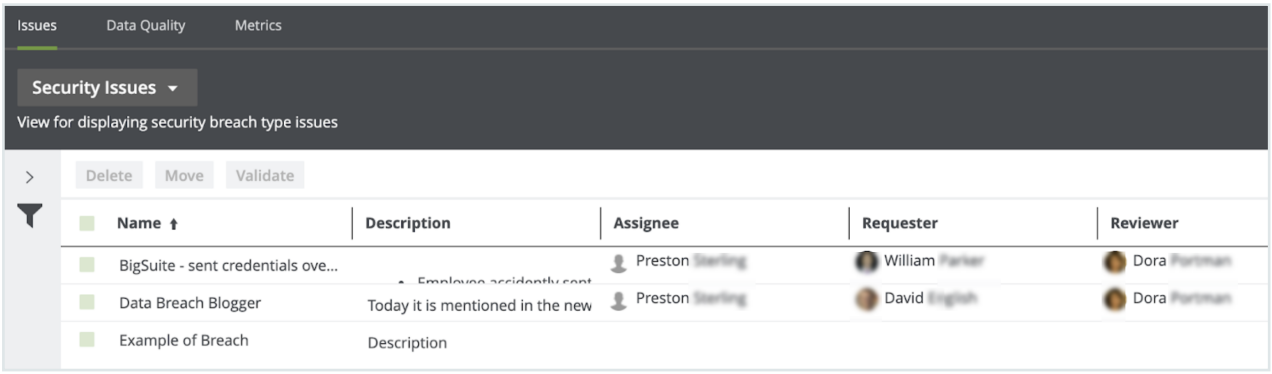
This section describes how Collibra helps when a data breach occurs.

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 the 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 the 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 Process register and Business Process discovery capabilities, data categorization and classification, and different prescriptive paths for reaching out 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

Process Register is an essential part of privacy compliance, foreseen directly by GDPR article 30 as a Record of Processing Activities 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

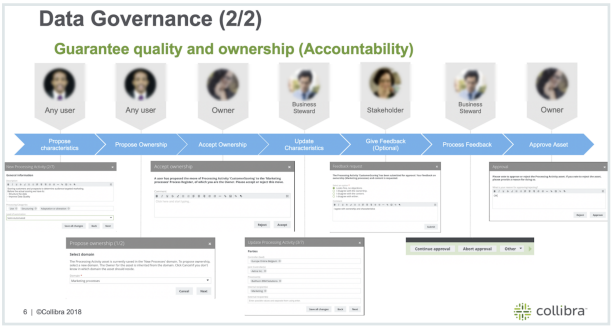
Business Processes may be onboarded by business users as well as privacy stewards through dedicated workflow implementing guided stewardship principle in Collibra Data Privacy. During onboarding, multiple roles collaborate in providing content to the onboarded Business Process. Because of the dedicated 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.

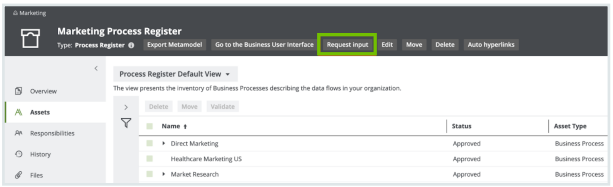
Consider a high-level overview of Collibra Privacy and Risk Business Process discovery capabilities. It commences with the Business Users describing the Business Processes in their terms. They will describe the data being used, applications being used, and any third parties with which they share information. 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

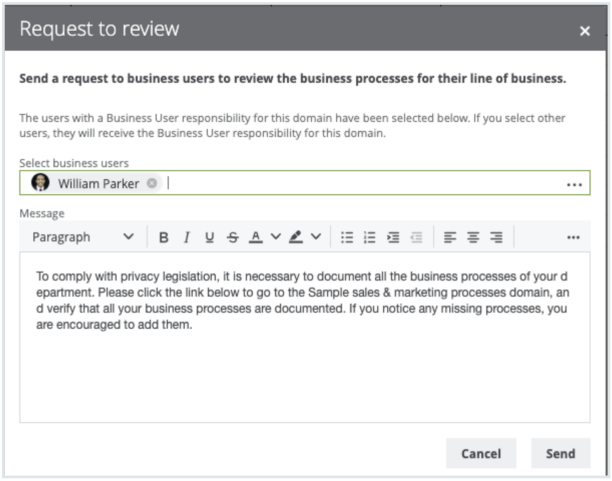
Process is accurate and actionable because that Business Process provides business context on how we use personal information and we must ensure that the description is accurate. Therefore, in principle, you will have the Business Steward, Privacy Steward, and Data Steward, each adding business metadata, 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 with a dedicated interface

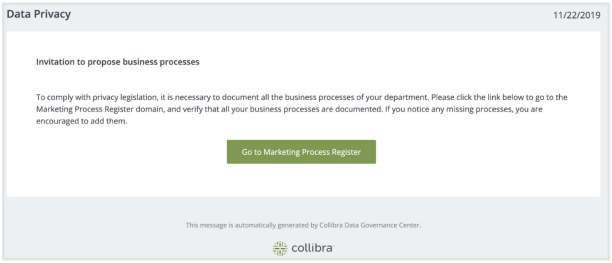
The information related to Business Processes may be requested from the Business User directly from Collibra Privacy and Risk Process Register. Typically, this will be done by those who work on the Privacy program. With the Request input button, email will be generated for the selected business users, which can provide relevant information on the business side of the process through a dedicated interface. 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.



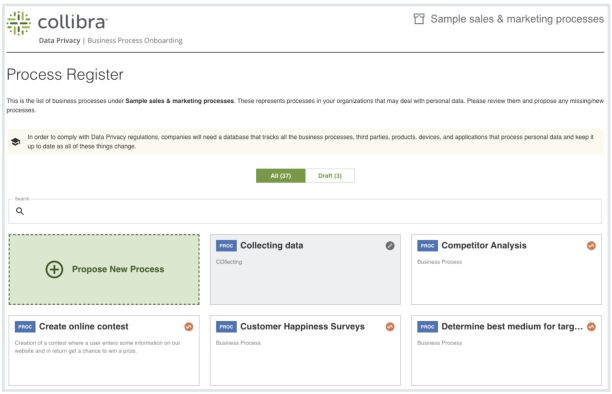


Onboarding Business Process with a business user interface

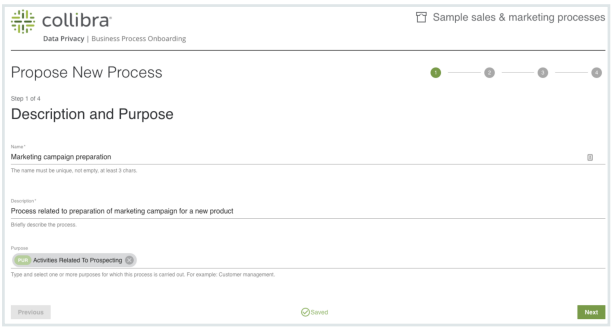
The Business User receives an email message asking them to verify that all the processes are in their domain.



When the Business User clicks Go to Marketing Process Register in the email, a page showing all the Business Processes for their department appears to allow the Business User to contribute to the Process Register.



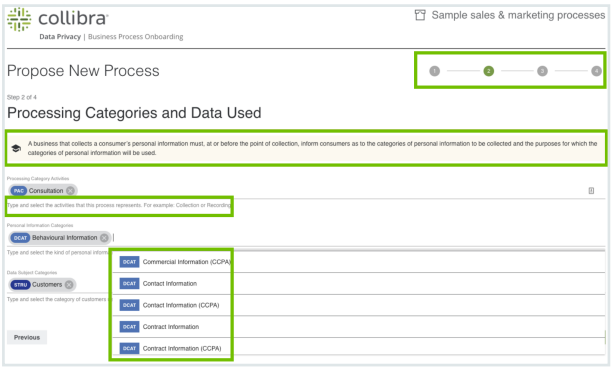
The link provided in the email message directs the User to a survey where they can describe the business processes that they perform on a daily basis. If the Business User cannot find the Business Process that was onboarded was in the process of being onboarded, they can propose a new Business Process using the Propose Business Process button. When proposing a Business Process, they can describe the Business Process, provide a unique name, description, and purpose.



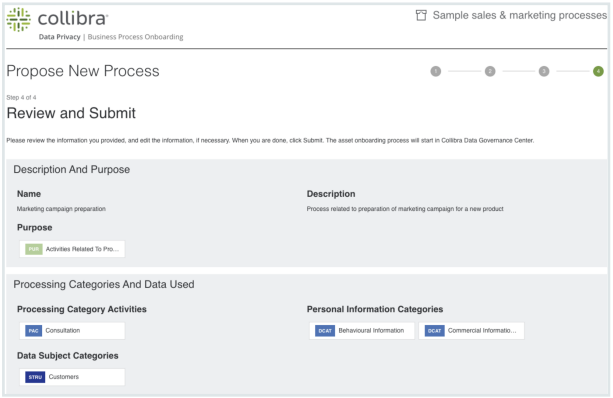
The next step involves covering Process Categories such as use, collection, adaptation, and alteration. The Business User defines the types of data that they are using, for example, behavioral information, contact information, or contract information. Finally, they determine what type of customer's data they are using, such as the customers covered by CCPA or GDPR. There can be also an indication on other types of data subjects, such as, employees and candidates. The Business User can select values only from predefined lists—this reduces the scope of errors as there is no ambiguity around the values that the Business User is able to provide. These values have been predefined by the Privacy team and have legal implications. They show how the organization complies with the privacy regulations. Because, when you collect data directly from customers or from a third parties—by using sensitive information, public information, or customer or employee information—the distinctions will have an impact on how you comply with the regulations. For example, the employee information is temporarily exempted from the CCPA. Therefore, it is considered better for the Business User to select from the drop-down list, as opposed to providing free text. This also prevents common issues such as spelling errors. In addition, if there is any uncertainty about the meaning of these values, the Business User can look up the definitions of these values in Collibra. In the next steps, the lines of business and third parties involved can be described, applications used can be indicated, and the level of automation in the Process can be determined.

The wizard is prescriptive:

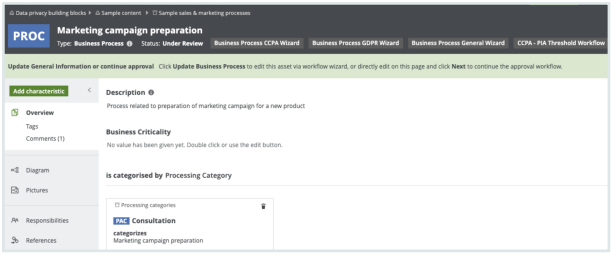
- It shows the user the steps that they have completed and how many steps are remaining, by visually indicating the progress.
- The help text below the question describes what is required from a particular question.
- The ability to open a side panel that provides additional educational information such as the wordings from the law or video content from the Collibra university.
- Smart suggestion based on what the user has already filled and the domain to which they belong.



After entering the information, the Business User can review it before submitting it.

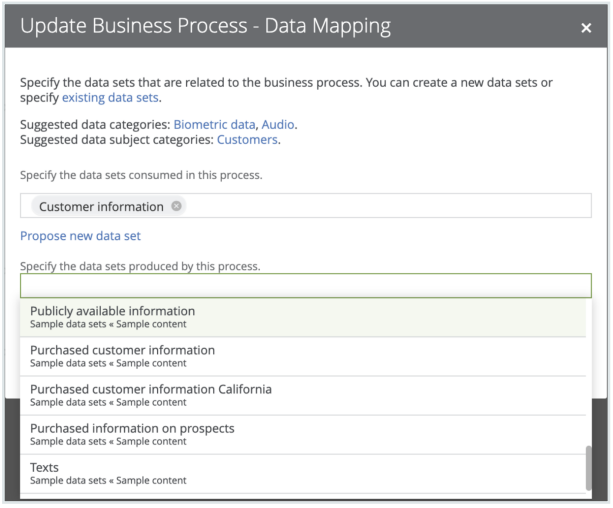


After the Business User provides their Business Process, they can submit the process for further onboarding. The next step is for the owner to accept the ownership of that Business Process. A new task is generated for the owner after they accept the ownership of the Business Process in the Process Register. Based on the metadata, the owner can determine that the Business Process belongs to their Process Register. The ownership can be accepted or rejected. As a result, the status of the asset is changed and the justification is added in the Comments section of the Business Process.

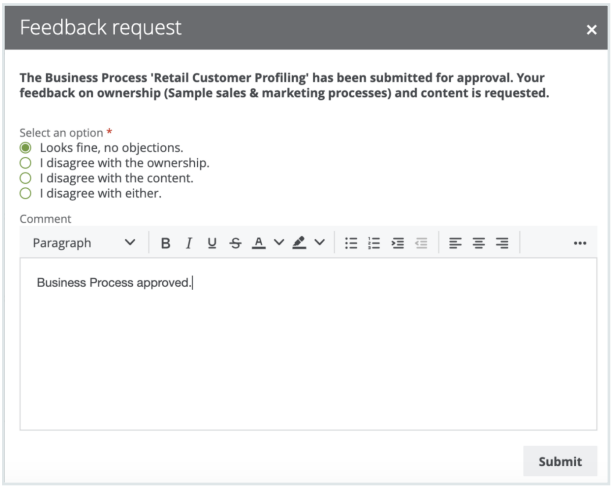


As a next step, the experts or the stewards will be consulted to ensure that the metadata is accurate and complete and the Business Process is going to be mapped to the data. In the following steps, relevant tasks will be created with the request to review and update attributes when necessary. Among others, the Data Mapping task is performed by Data Stewards. The form contains contextual help with suggestions on the relevant Data Sets consumed and produced by the Business Process. Because one of the data categories used in the process is behavioral information, you can click and review it and view the related data sets in the categories. Based on this, Data Stewards can ensure that Data Mapping has been correctly performed.

The last expert who needs to contribute to the Business Process is the Privacy Steward. After opening the task, the first step is to define the regulation that applies, be it GDPR, CCPA, or others. In addition, a purpose needs to be validated, and legal bases, controllers and processors need to be defined. Very specific information on regulation shall be specified, for example, on GDPR, we define cross-border transfers, safeguards, consent collection method, and automated decision-making confirmation. On the CCPA side, we are asked about the collection directly from customers or third parties and whether the data is being sold to third parties.



After the Stewards finish updating the Business Process, we ask the Stakeholders for final feedback. If the feedback is positive, we send the task to the Owner for the final approval.

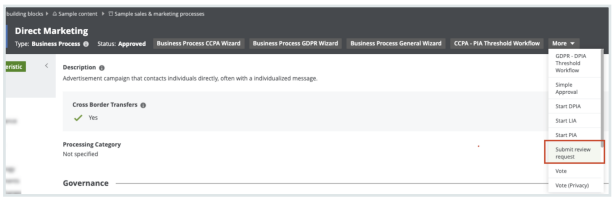


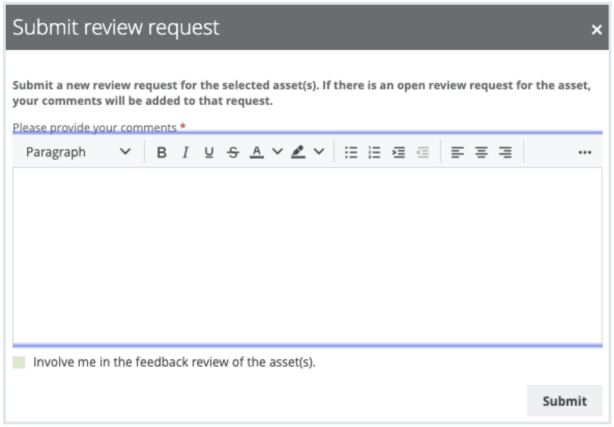
Maintain Process Register over time with review requests

Whereas 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.

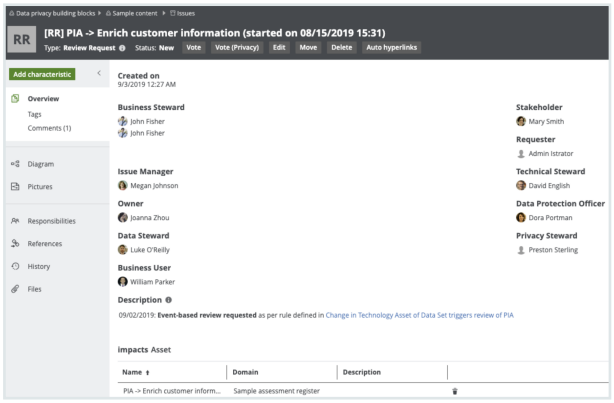
You may have many reasons to review an approved asset. Collibra Data Privacy groups such reasons into three categories and offers three corresponding means to trigger a review request:

- 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.



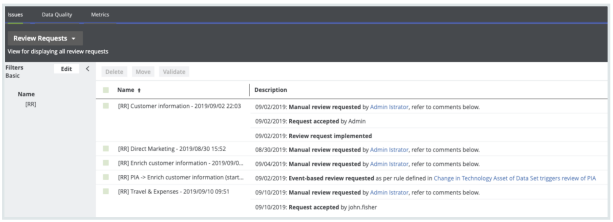


- Time-based: A trigger that is automatically actioned at a specified frequency. This is useful for assessment assets for which you might be required to review periodically to comply with a regulation.



- 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 the Data Helpdesk.



Perform Assessments

Conduct PIA and DPIA

If a business process is likely to introduce a level of risk to the rights and freedoms 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 freedoms of natural persons to risk is significant. Privacy Impact Assessments (PIA) and Data Privacy Impact Assessments (DPIA) assess the risks to the rights and freedoms 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:

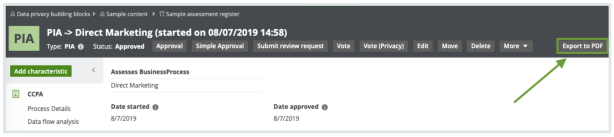
- PIA, if complying with CCPA
- DPIA, if complying with GDPR

Print assessment results

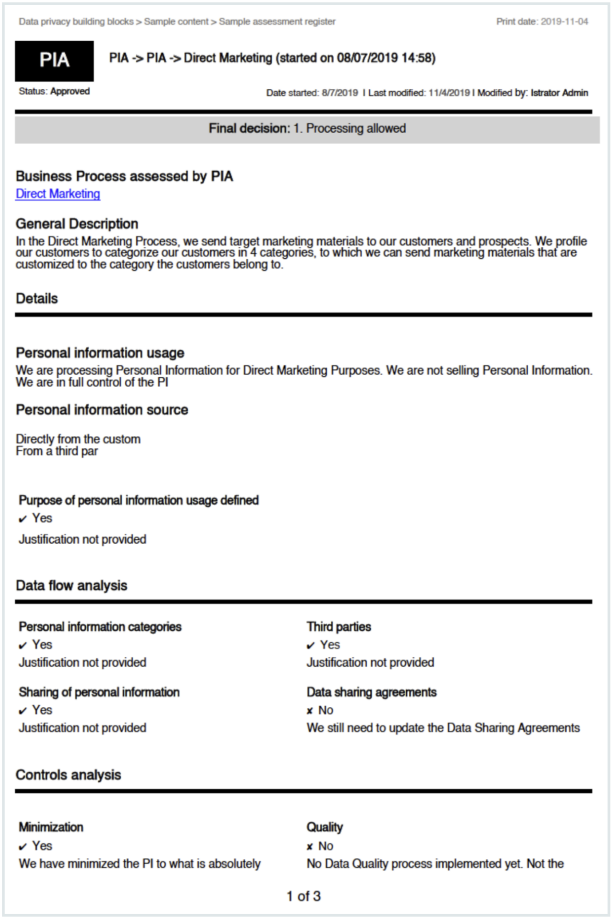
Assessments are a way for an organization to demonstrate compliance. You can export and print the PIA results in a unified way. You can also download a PIA asset page as a printable PDF, regardless of the status of the PIA asset.

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 Protect to the best of its ability.

Types of data protection

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.

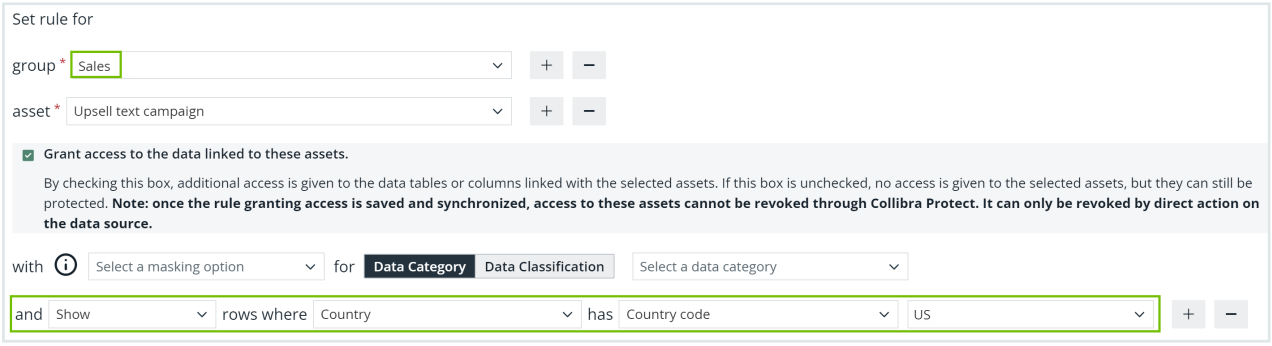
Suppose that you want the Human Resources (HR) group to be able to access a data set of US-based customers. Suppose that certain parts of the data set need to be hidden from the HR group because they contain restricted data, such as personally identifiable information (PII). Then, you can further protect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

Row-based protection

Row-based protection allows 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.

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.

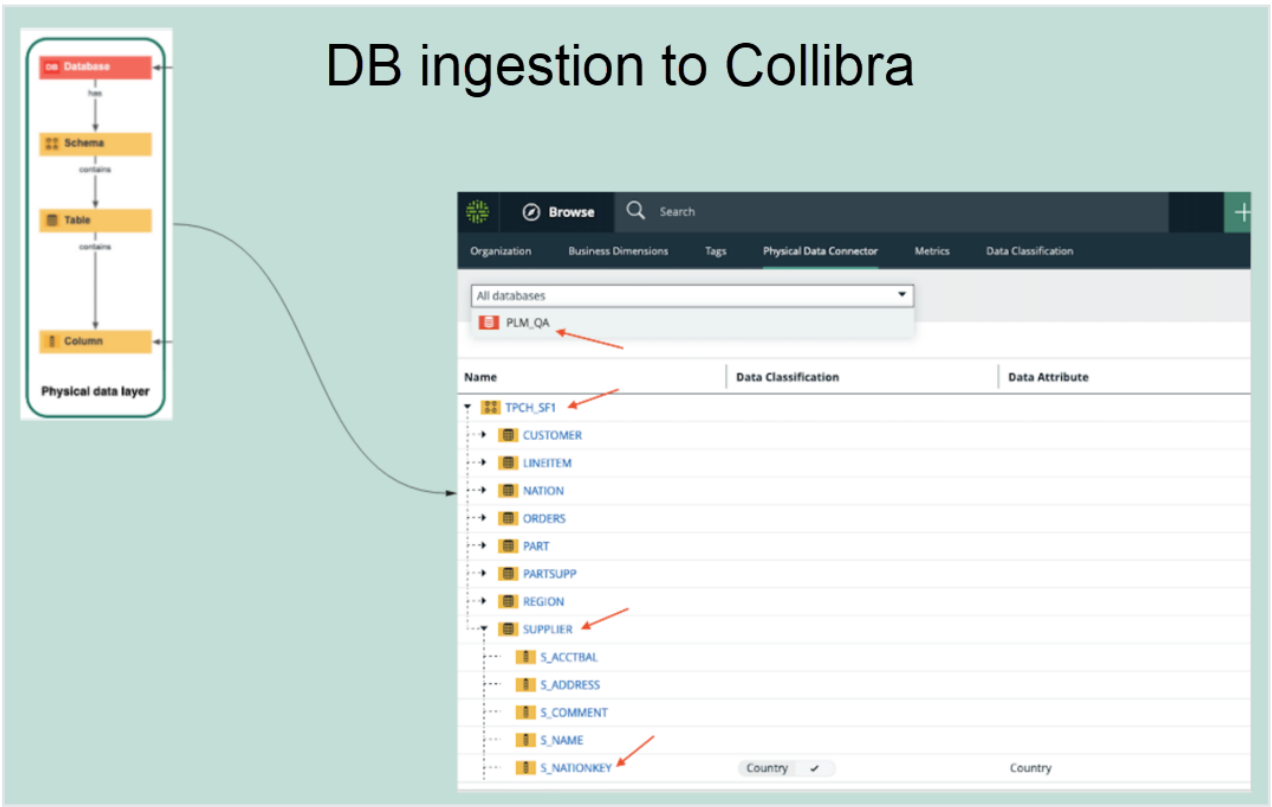


Technical background

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.

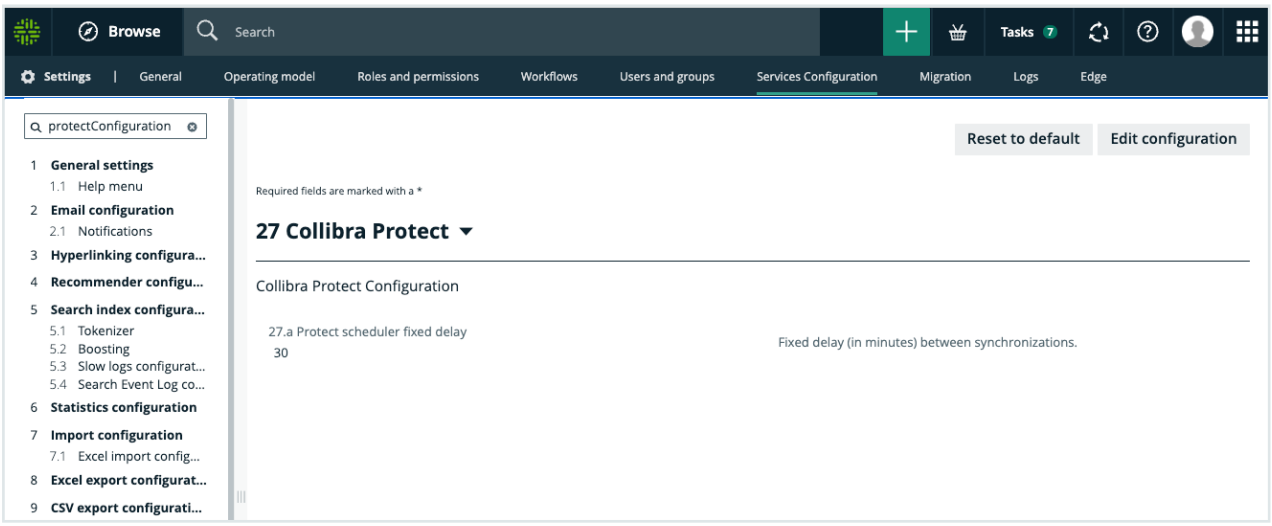


After the physical layer is created in Collibra, the logical layer can be created on top of the physical layer, as follows:

- 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.

Synchronization

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

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 they reside, 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.

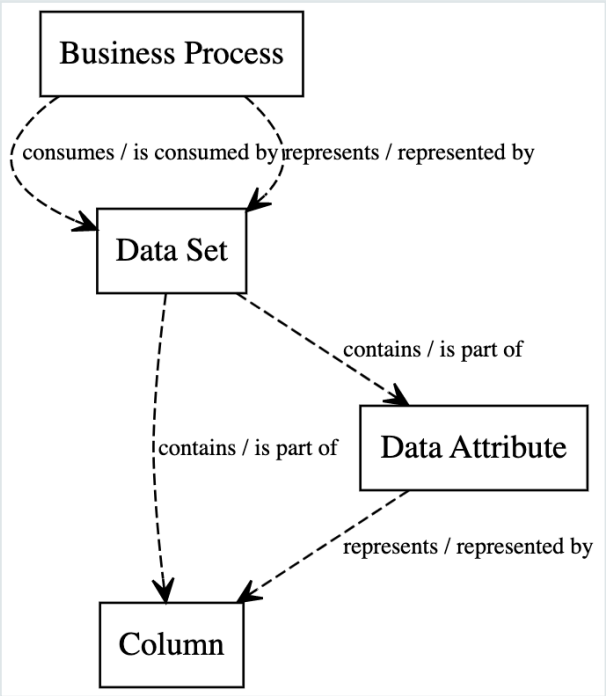
Prescriptive paths

You can use 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.

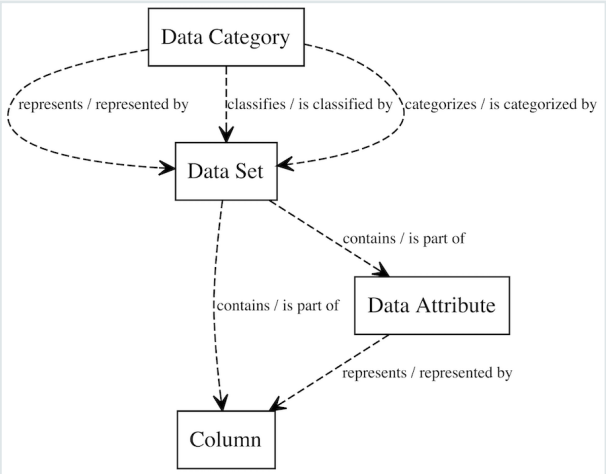
The assets that you use to create data protection standards and data access rules are related to the physical data layer, such as tables and columns, through a set of relations and intermediate assets. Protect uses these relationships and intermediate assets to search the knowledge graph to find the physical data layer assets that it needs to protect. The traversal of the knowledge graph 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.

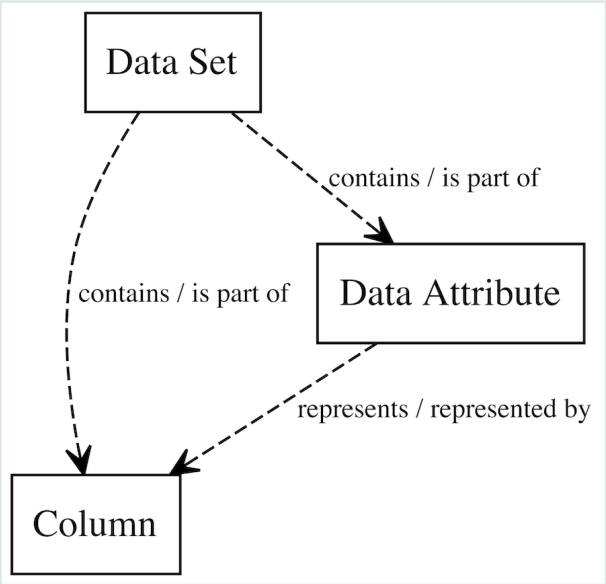
Business Process



Data Category



Data Set



Customization of prescriptive paths

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.

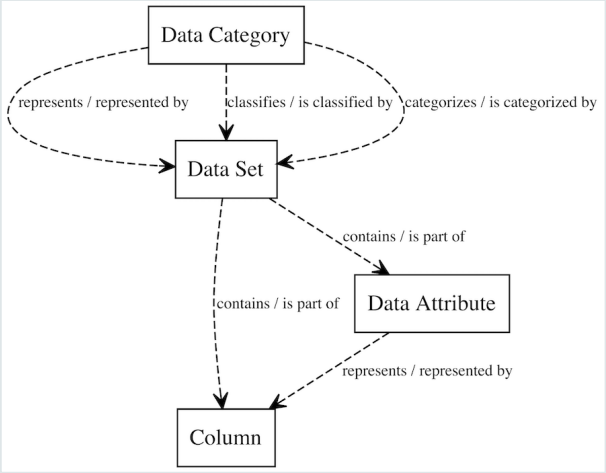
If you have the **Protect** > **Administration** global permission, you can customize the prescriptive paths for the asset types through APIs. The customization may include creating, modifying, or deleting the prescriptive paths: for example, adding or modifying the prescriptive paths for packaged and custom asset types, defining how the asset types relate to columns, 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 data protection standard or a data access 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"
      }
```

Install Protect

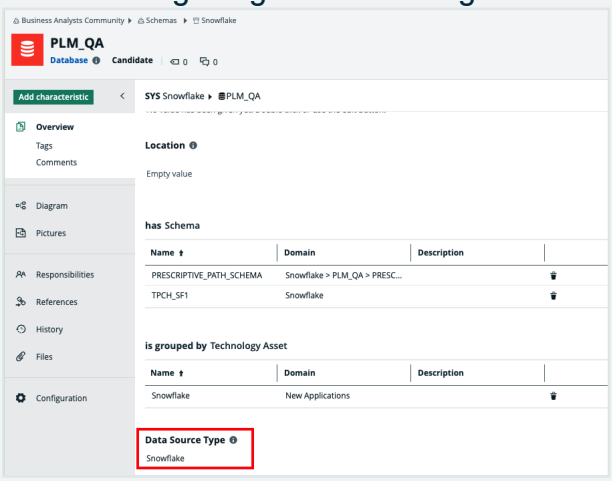
This topic describes how to install Protect.

Before you begin

1. Add the Catalog ingestion capability to the Edge site and register the required data source.
2. Add the required Protect capability to the Edge site. When adding the capability, in the **Capability template** field, select one of the following values, as applicable: **Collibra Protect for Google BigQuery** or **Collibra Protect for Snowflake**

Tip

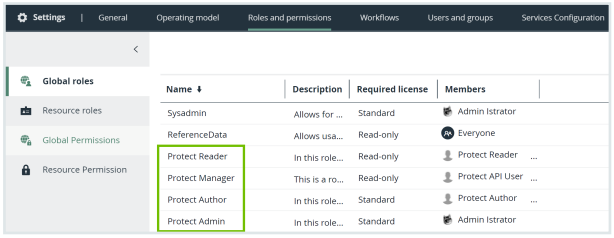
- If you selected **Collibra Protect for Snowflake**, in the **Snowflake Connection** field, you can select the Snowflake connection that was used for performing catalog ingestion, if the connection allows Protect to write to your database.
- The following image shows an ingested Snowflake database.



The **Data Source Type** attribute is added to the database asset only after the catalog ingestion process is complete.

Steps

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



» Protect is installed. 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 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 data protection standards and data access rules, modify only the standards and rules that they created, view imported policies, view groups, and 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 all data access rules and data protection standards and also access additional APIs.
Protect Manager	A user who manages background processes and configures Protect.This role is intended only for the Protect system user.

Global permission	Description
Protect > Edit	Allows a user to create data protection standards and data access rules and modify only the standards and rules that they created.
Protect > Administration	Allows a user to create data protection standards and data access rules and modify all standards and rules.

Open Protect

This topic describes how to open Protect, including how you can use the tabs on the Protect landing page.

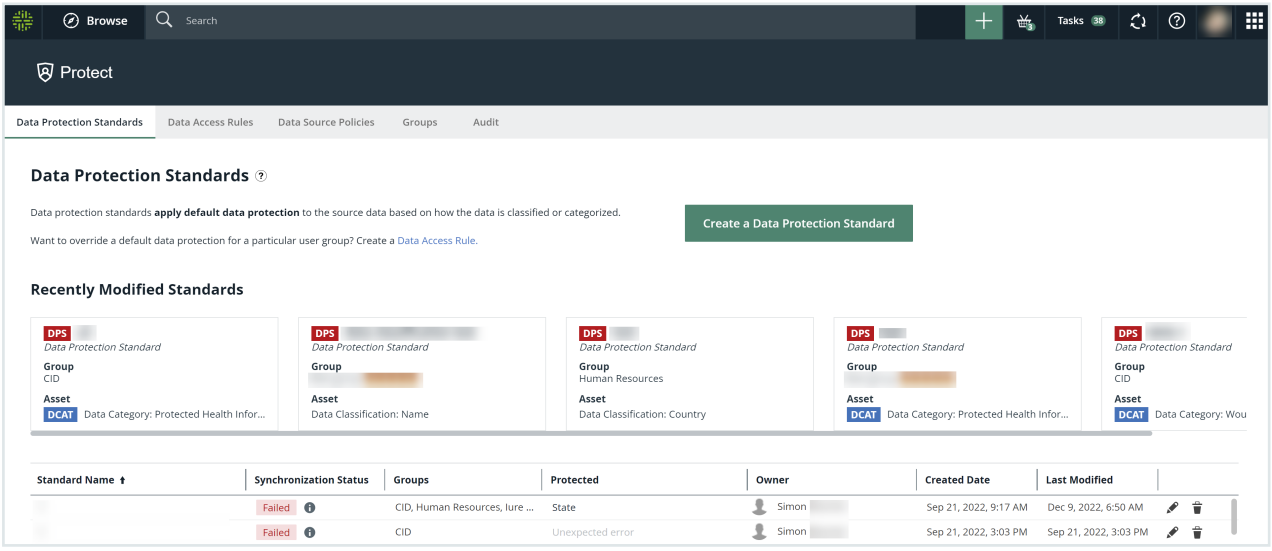
Requirements and permissions

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 is shown.



Protect landing page

On the Protect landing page, depending on your role, the following tabs are shown.

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 data protection standards and data access 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

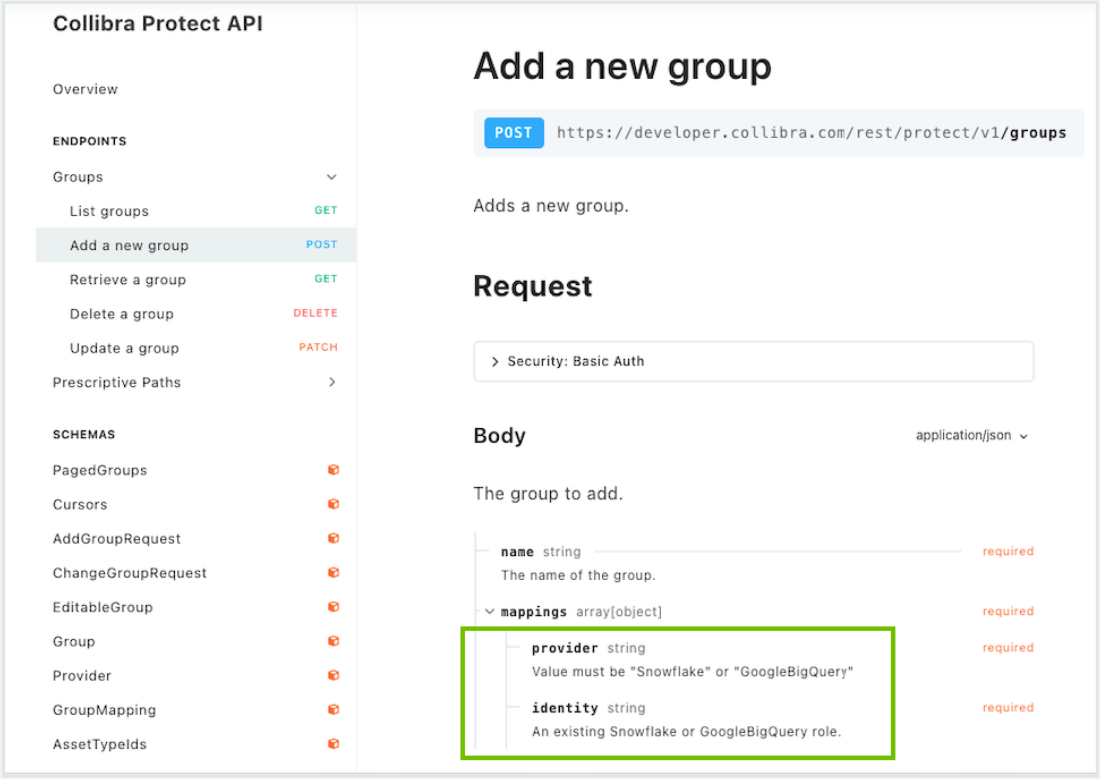
You have the **Protect Author** or **Protect Admin** global role.

Steps

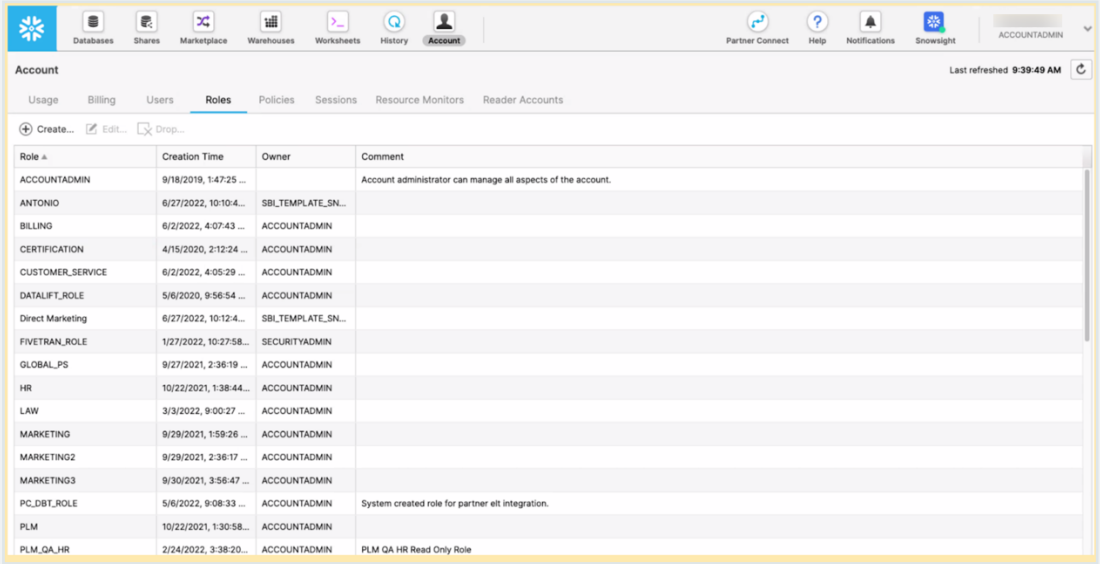
1. Open Protect.
2. On the **Groups** tab, click **Collibra Protect Group API**.
3. For the next steps, go to Add a new group.

Tip

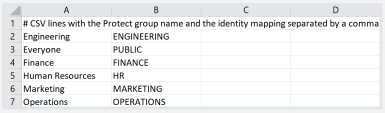
- When creating a Protect group, you are prompted to specify the data source provider (for example, **Snowflake**) and the existing role from the data source provider to map the role to the group.



- The following image shows the roles in the Snowflake data source provider.



- 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 (that is, they should have the global role Protect Author and/or Protect Admin)
COLLIBRA_AUTH="user:password"

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 the first line in the CSV file
    while IFS=, read -r field1 field2 field3
    do
        echo "Add group $field1 with Snowflake role $field2 and GCP $field3"
        curl -u "${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": "Snowflake",
                    "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 role associated with the group.
Created By	The name of the user who created the group.
Created Date	The date when the group was created.

Data protection standards

Data protection standards create a primary layer of protection for similar types of data by masking the data wherever they reside. They protect your data by masking the data (column-based protection).

Create a data protection standard

Requirements and permissions

- You have the **Protect Author** or **Protect Admin** global role.
- 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

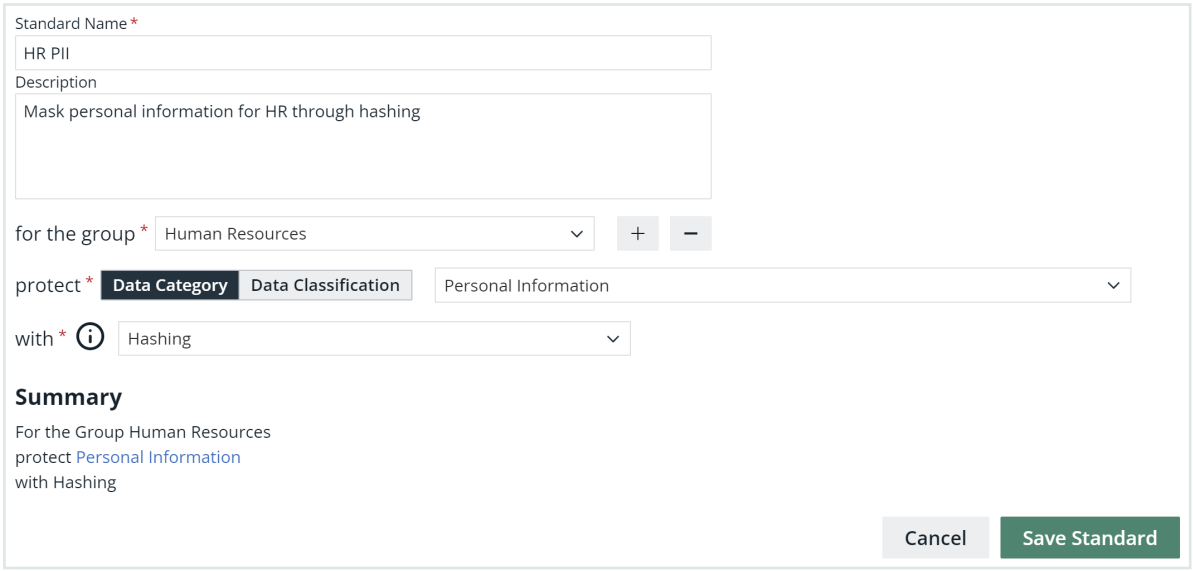
1. Open Protect.
2. Click the **Data Protection Standards** tab.
3. Click **Create a Data Protection Standard**.
   » The **Create a Data Protection Standard** dialog box appears.
4. Enter the required information.

Details

Field	Description
Standard Name	Enter a name to identify 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.
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. The following options are available: ◦ **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 the **Protect Author** or **Protect Admin** global role.

  > Note If you have the **Protect Author** global role, you can modify only the data protection standard that you created. If you have the **Protect Admin** global role, 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

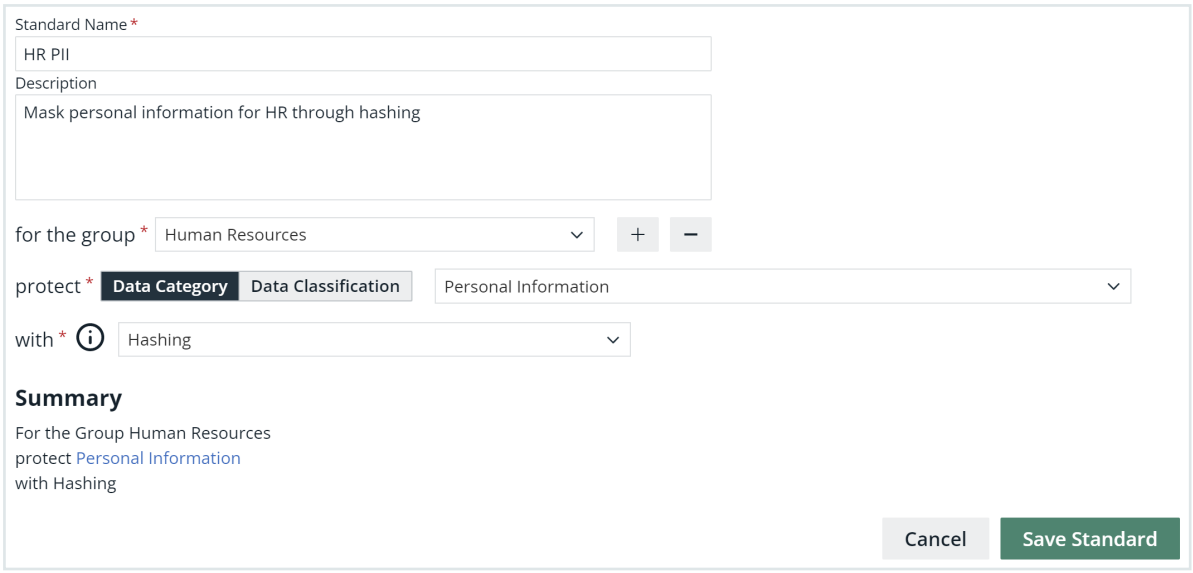
1. Open Protect.
2. In the table, in the row containing the standard that you want to modify, click ✎.
   » The **Edit a Data Protection Standard** dialog box appears.
3. Modify the required information.

Details

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

Field	Description
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.
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. The following options are available: ◦ **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 the **Protect Author** or **Protect Admin** global role.

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 If you have the **DataSteward** global 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 create an additional layer of protection 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 the **Protect Author** or **Protect Admin** global role.
- 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 to identify the data access rule.
Optional: Description	Enter a description for the data access rule.
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.

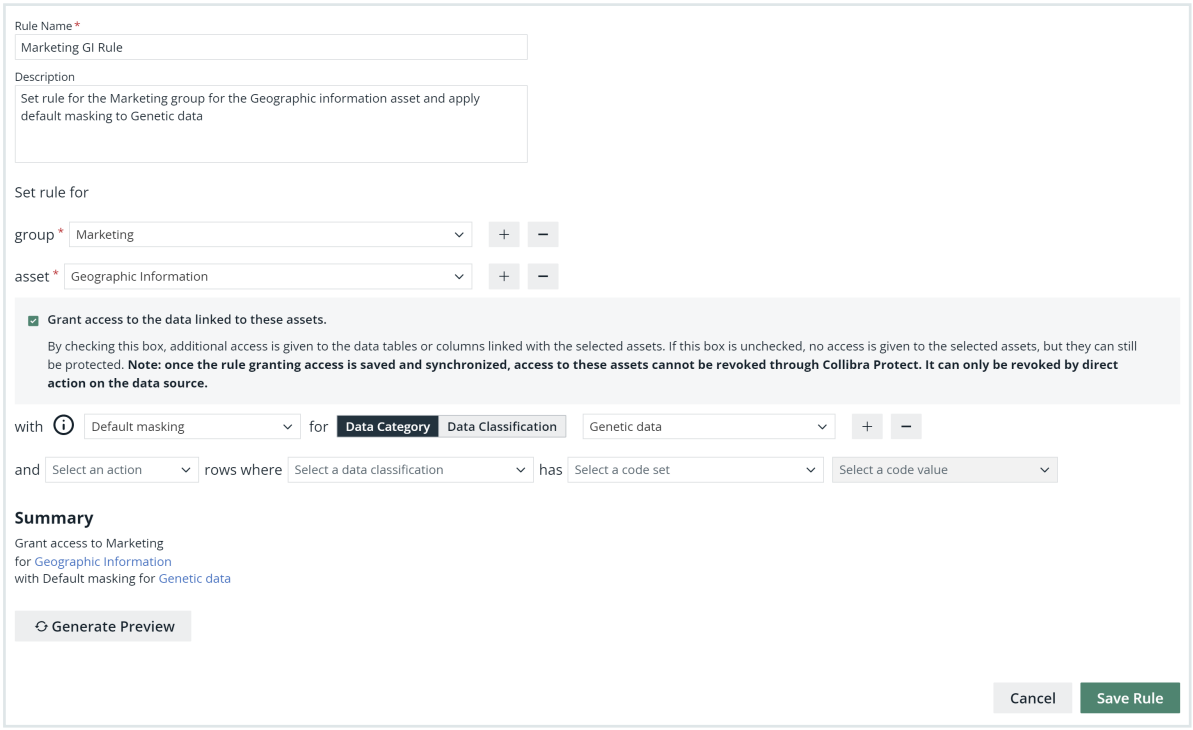
Field	Description
Asset	Select the data asset that the rule is protecting. This field contains Business Process, Data Category, and Data Set assets, as well as assets of custom asset types. Tip ◦ For more information, go to Technical background and Prescriptive paths. ◦ You can add more groups by using 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. The following options are available: ▪ **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. Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. The following actions are available: ◦ **Show** ◦ **Hide** a. In the **rows where** field, select the data classification that you want to show or hide. b. In the **has** field, select the code set for the selected data classification. c. 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 is selected by default. A selected **Grant access to the data linked to these assets** checkbox indicates that you are granting access to the tables and the columns in the database that are linked to the selected assets to the groups that you selected in the rule. If you do not want to grant this level of access to the selected groups, 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 the **Protect Author** or **Protect Admin** global role.

  > Note If you have the **Protect Author** global role, you can modify only the data access rule that you created. If you have the **Protect Admin** global role, 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 to identify the data access rule.
Optional: Description	Enter a description for the data access rule.

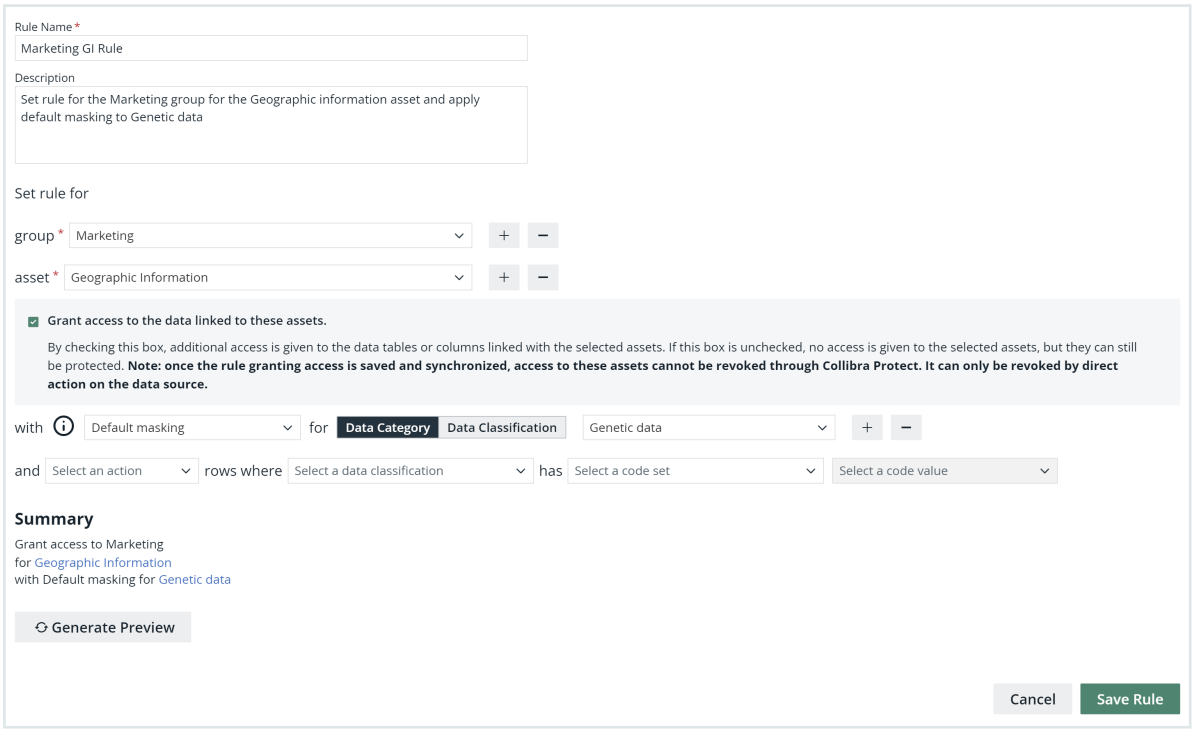
Field	Description
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.
Asset	Select the data asset that the rule is protecting. This field contains Business Process, Data Category, and Data Set assets, as well as assets of custom asset types. Tip ◦ For 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. The following options are available: ▪ **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. Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. The following actions are available: ◦ **Show** ◦ **Hide** a. In the **rows where** field, select the data classification that you want to show or hide. b. In the **has** field, select the code set for the selected data classification. c. 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 **Grant access to the data linked to these assets** checkbox indicates that you are granting access to the tables and the columns in the database that are linked to the selected assets to the groups that you selected in the rule. If you do not want to grant this level of access to the selected groups, 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 the **Protect Author** or **Protect Admin** global role.

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 If you have the **DataSteward** global 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

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

The **Data Source Policies** tab in Protect contains an overview of the native data source policies. To view the policies, you need the **Protect Author** or **Protect Admin** global role.

> Note Contact Collibra support to import policies from the data source using the Collibra Protect Data Source Policies API.

The table on the **Data Source Policies** tab contains the policies that are active in the data source. These include both the policies that are manually created in the data source and the policies that are generated in the data source as a result of the data protection standards and data access rules in Protect.

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.

Column	Description
Data Source	The data source provider.

Data source providers

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 Google BigQuery** and **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 Snowflake

Data protection standards in Collibra Protect rely on the tag-based masking policies of Snowflake. The name of the data category or data classification selected in a standard becomes a tag with the same name. The tag is applied to all the affected columns to enforce data protection. For more information, go to Snowflake examples.

Types of policies on Snowflake

There are three types of policies on Snowflake: Column-based policies, row access policies, and tag-based policies. Each type can be created in Collibra Protect or on Snowflake.

For rules, policies are created directly on the column level. Row access policies are created when row filters are specified. For standards, the policy is created, attached to a Snowflake tag, and attached to the tab on any affected column.

Snowflake examples

This topic contains examples to describe how Snowflake behaves in relation to certain data protection standards and data access rules.

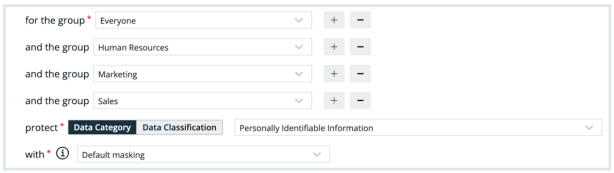
Example 1

Introduction

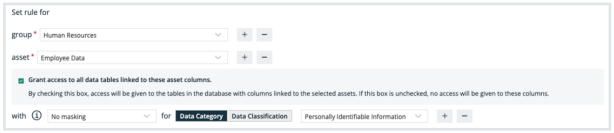
This example describes the behavior in Snowflake when a standard is applied to a data category and a rule is applied to a data set with categorized columns in Protect.

The example considers the following:

- A standard created for the **Everyone**, **Human Resources**, **Marketing**, and **Sales** groups, to protect the columns in the **Personally Identifiable Information** data category by default masking.

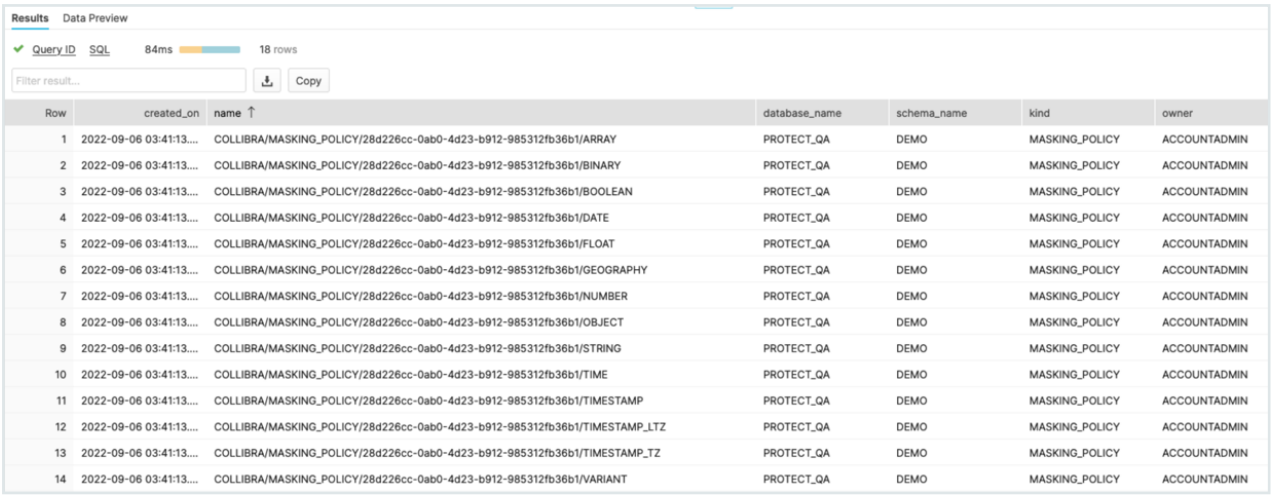


- A rule created for the **Human Resources** group and the **Employee Data** asset, without any protection applied to the columns in the **Personally Identifiable Information** data category.



Standard

When the standard is synchronized and active, the standard results in 14 masking policies—one policy for each Snowflake data type. The masking policies are created at the schema level with the following naming convention: 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

All the masking policies are then associated with the **Personally Identifiable Information** tag, which is created at the schema level and assigned to those columns that need to be protected. At runtime, Snowflake fetches the right masking policy based on the column data type.

```
35 SHOW TAGS;
36
```

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.

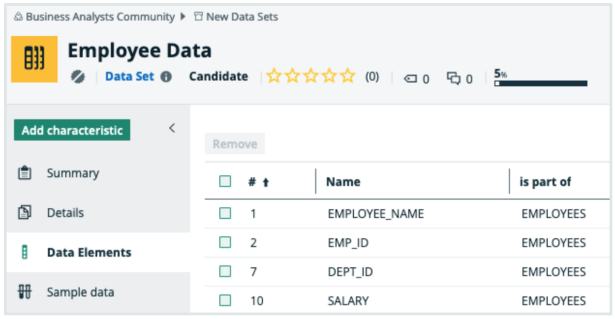
Details

```
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

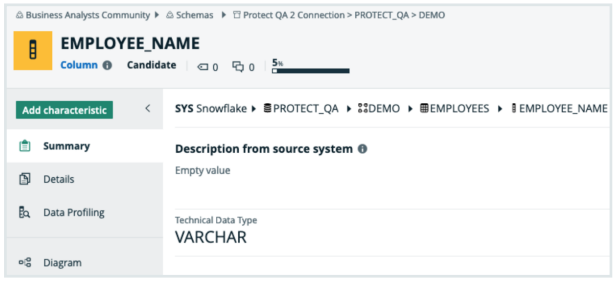
A rule results in a combination of grant instructions, dynamic masking, and row access policies.

Suppose that the **Employee Data** data set selected in the rule contains sensitive columns categorized as **Personally Identifiable Information**.

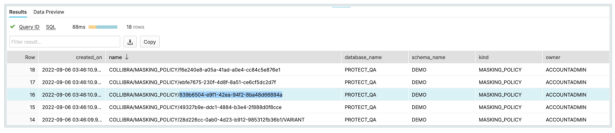


The rule grants access of the **Employee Data** data set to the **Human Resources** group, as indicated by the selected **Grant access...** checkbox in the rule. Then, the corresponding Snowflake role for the group can access each database, schema, and table in the data set. In addition, the column masking 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 **Personally Identifiable Information** 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 use the following naming convention: COLLIBRA/MASKING_POLICY/<asset ID>.



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

Details

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

Behavior

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 a column, the column masking policy takes precedence and the policy tag is not assigned to the column. To mitigate this behavior and 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 the **Personally Identifiable Information** column by default masking, 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 via default masking.

Example 2

Introduction

This example describes the behavior in Snowflake when multiple standards affect the same column without conflict.

The example considers the following:

- A standard created for the **HR** group to protect the columns in the **Personally Identifiable Information** data category by hashing.
- A standard created for the **Marketing** group to protect the columns in the **Personal Information** data category by default masking.
- The **Personally Identifiable Information** and **Personal Information** data categories share the same column named **DOB**.

Behavior

Protect creates a tag for each standard and adds a policy to each tag. The two tags are then linked to the **DOB** column. In addition, Protect creates a masking policy that is an aggregation of the policies from the two tags. This aggregated masking policy, which is then applied to the **DOB** column, thus contains the content of both the tag policies.

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

When a policy exists for the **DOB** column, Snowflake considers only the column masking policy, ignoring all the tag policies associated with the column. Because the column masking policy is an aggregation of all the tag policies, the protection that is defined in the two standards is not ignored.

Thus, Protect handles multiple standards with tag policies for Snowflake by creating a column masking policy, which considers the protection defined in the standards.

Snowflake masking and 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 privileges

To perform actions in Snowflake, Collibra Protect uses an Edge connection that must be configured with a user and a role that can manage grants; create and assign masking policies, row access policies, and tags; and manage usage access on databases and schemas involved in the protection. This enforcement role requires the following Snowflake privileges.

Snowflake privilege	Description
`[APPLY MASKING], [APPLY ROW ACCESS], [APPLY TAG], [MANAGE GRANTS], [IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE]`	Required for the role performing the actions.
`[USAGE]`	Required on each database and schemas where policies are applied to the role performing the actions.

Snowflake privilege	Description
`[CREATE MASKING POLICY], [CREATE ROW ACCESS POLICY], [CREATE TAG]`	Required on each schema where policies are applied to the role performing the actions.

Example

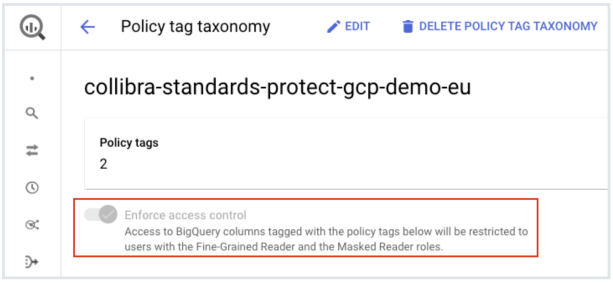
Suppose that a role named PROTECT exists in Snowflake and is responsible for managing access on all schemas within a database named DEMO. Then, the following statements can be used to enable the Snowflake PROTECT role to perform the enforcement.

```
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;
```

Protect for BigQuery

Collibra Protect uses Google's Policy tag taxonomies to create tags and assign the tags to your BigQuery columns. Policy tag taxonomies inherently apply access control. This

means that the tags applied to your BigQuery columns will be accessible only by the Protect groups configured in your data protection standards and data access rules.



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**.

Protect masking type	Equivalent BigQuery masking rule
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.
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)

BigQuery data type	BigQuery masking rule supported by Protect
DATE	Default masking value
DATETIME	Default masking value
FLOAT64	Default masking value
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 Collibra Protect group mapping for BigQuery must follow the syntax for principal identifiers. For example, the Protect group, **Sales**, maps to the BigQuery group email address, **sales@example.com**.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "GoogleBigQuery",
```

```
      "identity": "group:sales@example.com"
    }
  ]
}
```

BigQuery permissions

To perform actions in BigQuery, Collibra Protect uses a GCP connection that must be configured with a service account having 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.jobs.create
- bigquery.rowAccessPolicies.create
- bigquery.tables.get
- 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

Audit

An audit log 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 and what is shown in an audit log.

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 Collibra Protect varies from several minutes to hours, depending on the data source.

Requirements and permissions

You have the **Protect Author** or **Protect Admin** global role.

Steps

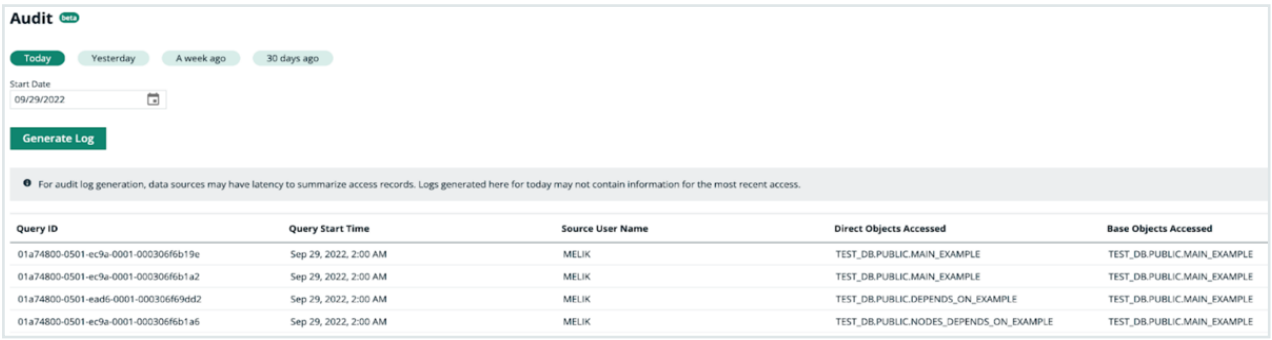
1. Open Protect.
2. Click the **Audit** tab.
3. Depending on your data source, click **BigQuery** or **Snowflake**.
4. Click one of the following buttons: **Today**, **Yesterday**, **A week ago**, **30 days ago**.

   > 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, and then 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.

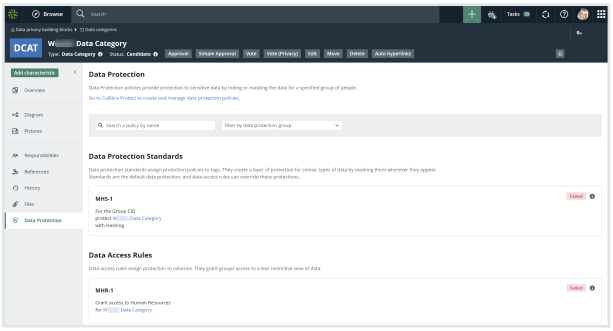
Column	Description
Query ID	The ID of the query in the source database.
Query Start Time	The date and time of the query in the source database.
Source User Name	The name of the user in the source database who ran the query to access the data.
Direct Object Accessed	The database object (a table or a view)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

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.

Chapter 13

Why rules or standards may 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.

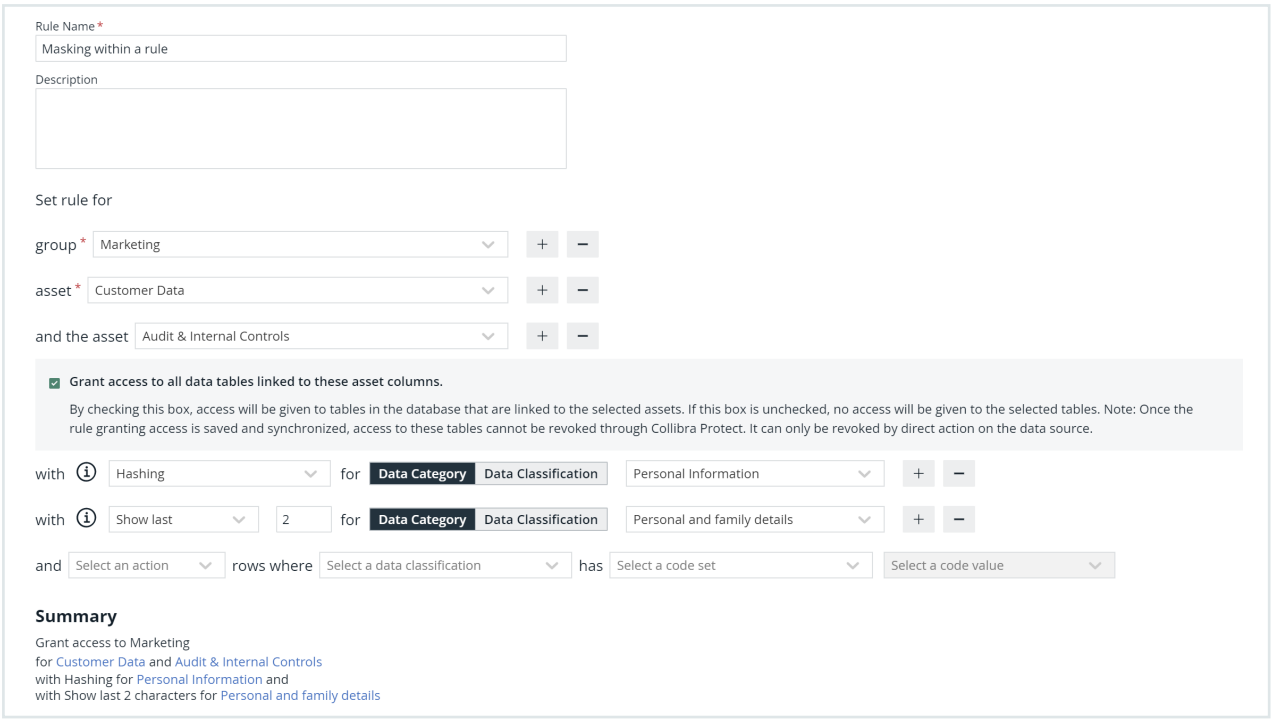
Masking within a rule

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.

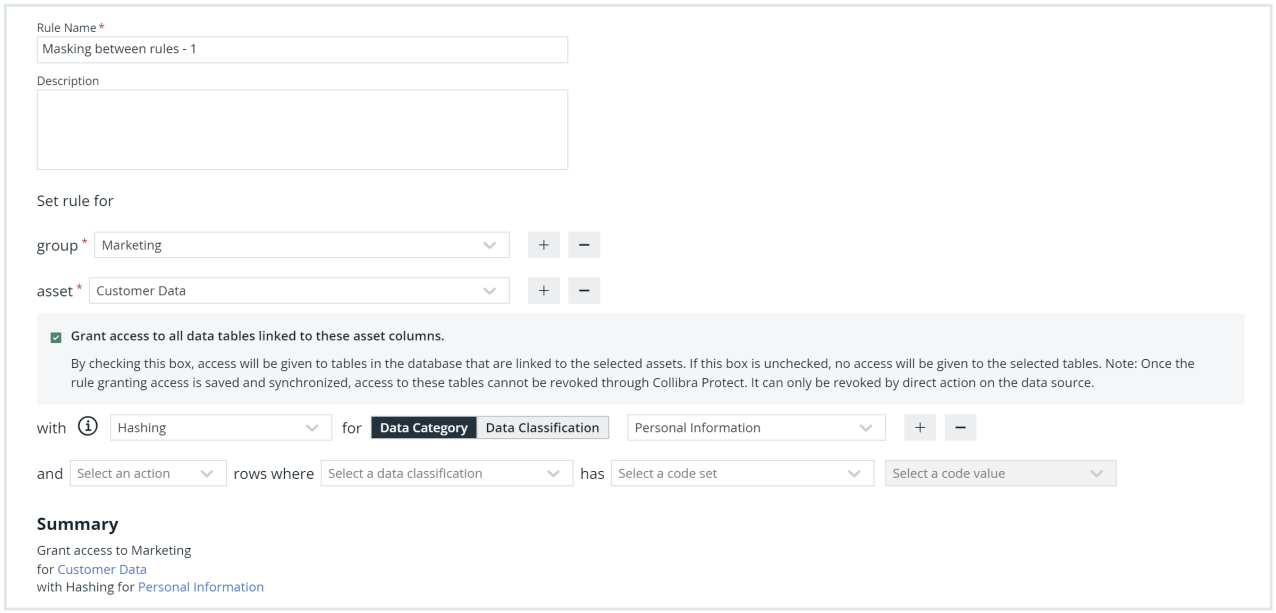


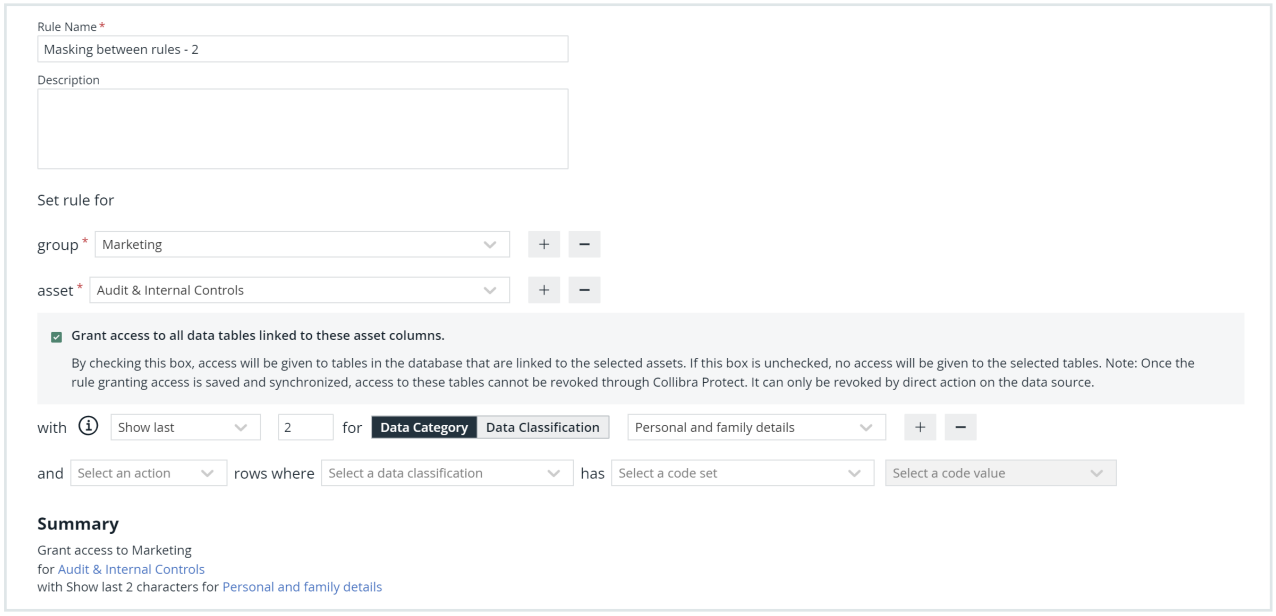
Masking between rules

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.

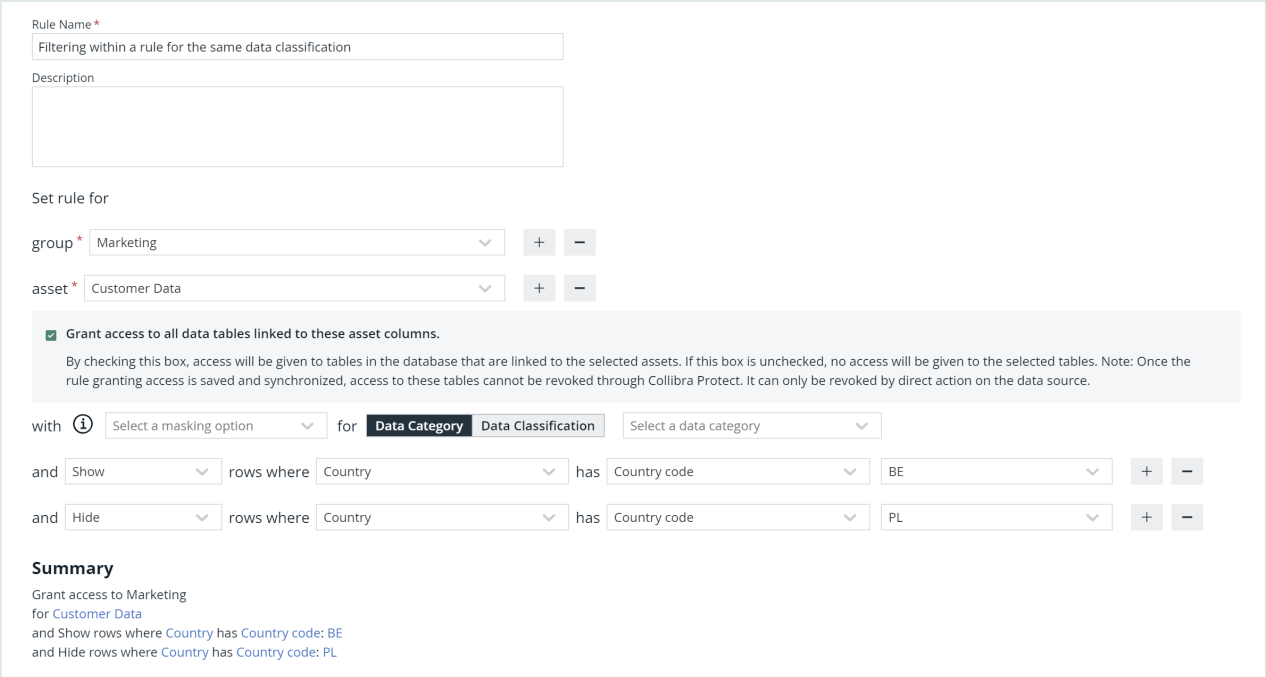
Filtering within a rule for the same data classification

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.

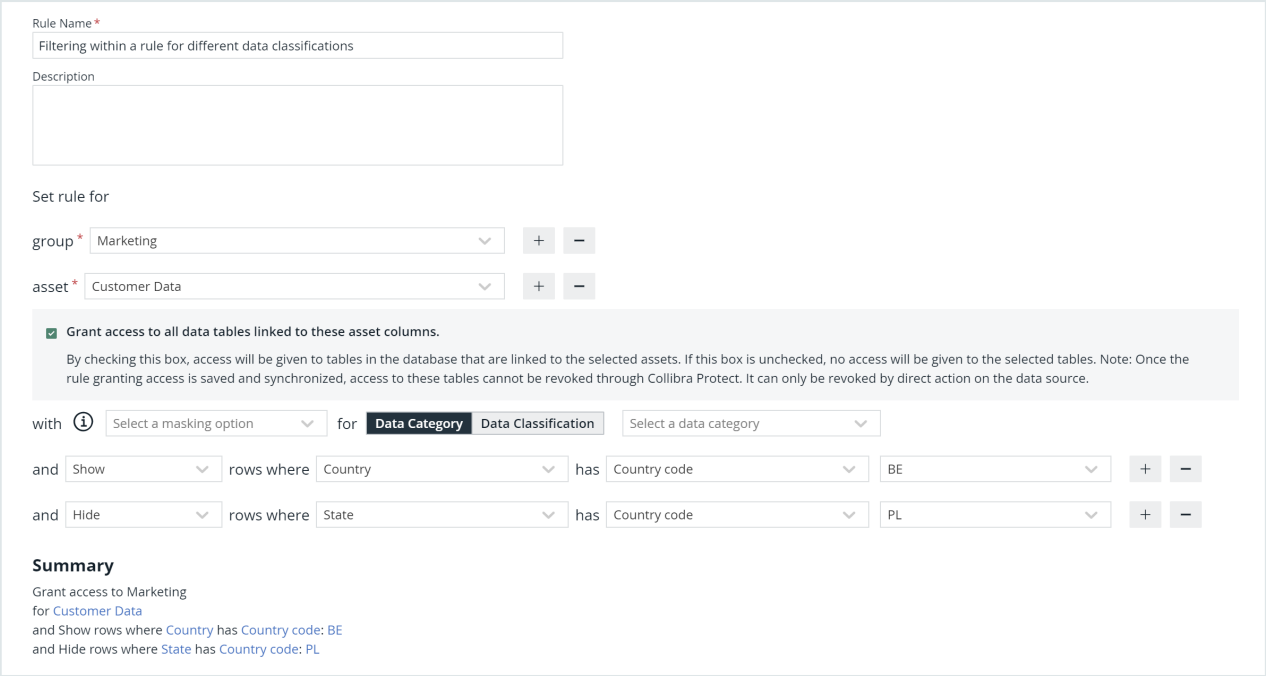
Filtering within a rule for different data classifications

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.

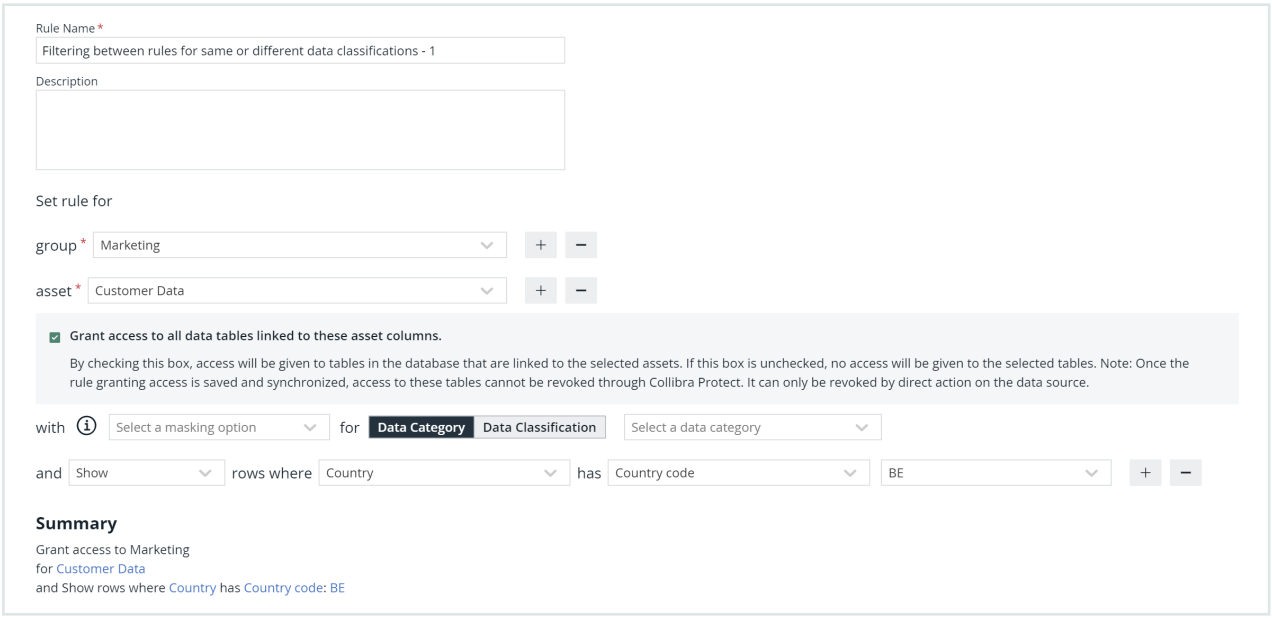
If any of the tables in the asset contain columns that are classified as **State**, the second filter hides only the rows that contain **PL** in those columns.

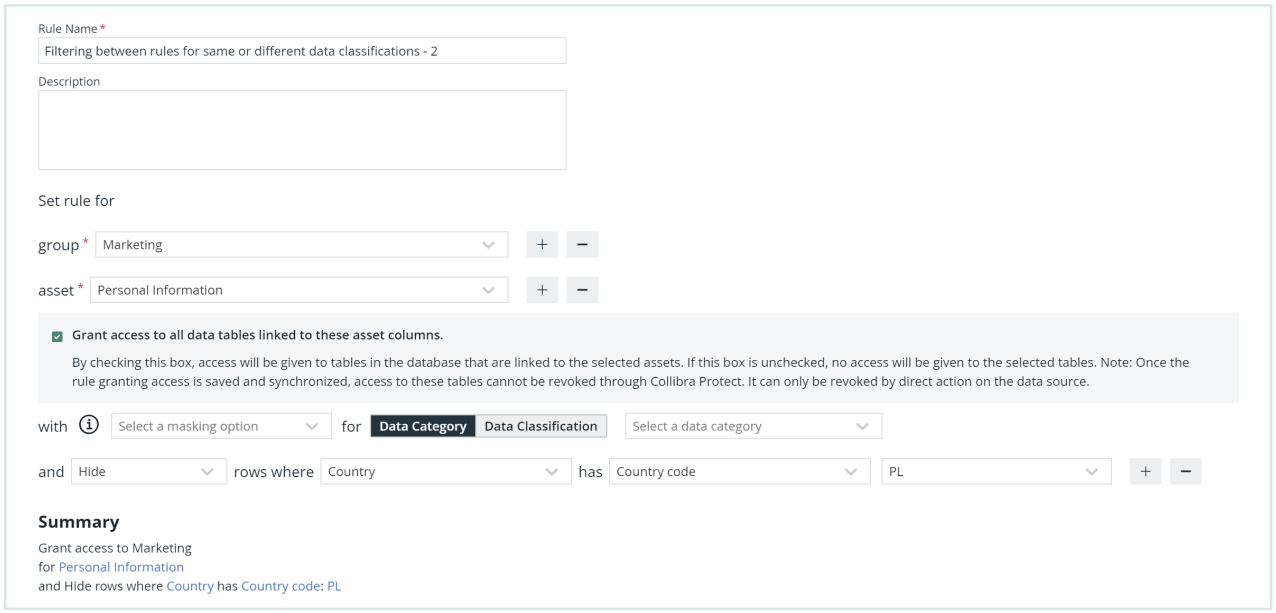
Suppose that a column is classified as both **Country** and **State**. That is, data classifications **Country** and **State** share the same column. Then, this rule will fail because two conflicting filters affect the same column.

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

Collibra Protect

Collibra Protect is a capability of the Collibra Data Intelligence Cloud 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. You can enforce data protection at the source database level directly from the Collibra 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:
>
> - BigQuery
> - Snowflake

Scenarios

This topic describes how Collibra Protect helps you to:

- Use the metamodel graph to build and enforce protection policies on Business Processes, Data Categories, and Data Sets.
- Use classifications to apply a broad coverage of protection mechanisms at the data source.
- Support privacy preferences such as consent management, data subject requests such as access requests, and the right to be forgotten through row-filtering mechanisms.
- Perform an audit of applicable protection at the data source and use reporting to demonstrate compliance where data is stored and consumed.

Discover and classify personal information

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

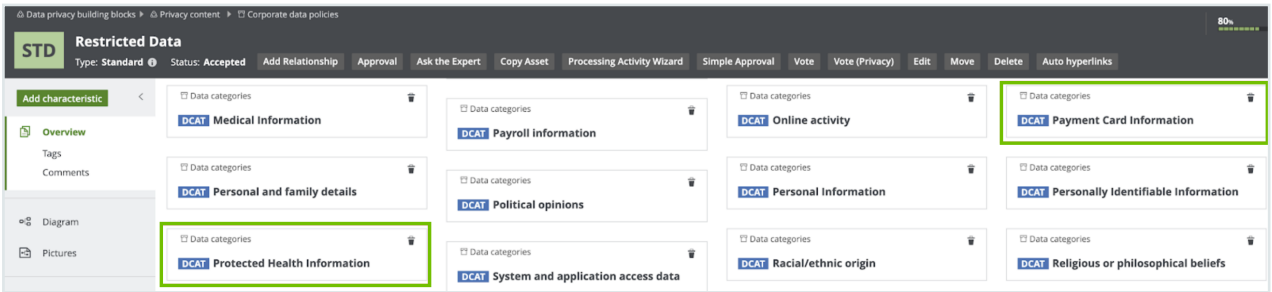
To achieve this, typically, your Privacy team sets up the Data Classification Policy, where they classify the data used in the organization based on the sensitivity or the business criticality of the data. This determines the required levels of security for the applications that store that data or the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

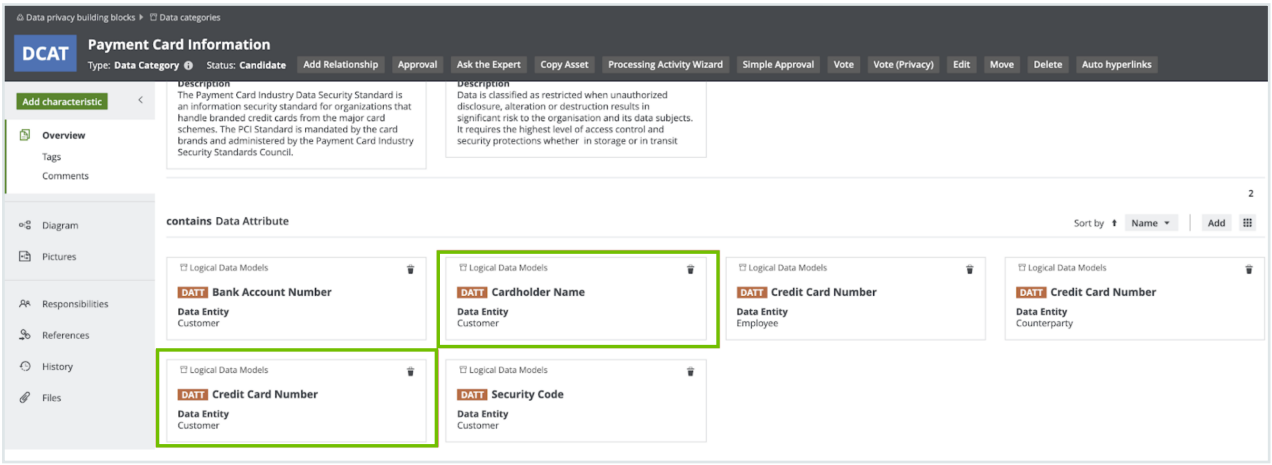
The following image shows the standard subassets of the Data Classification policy.

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.

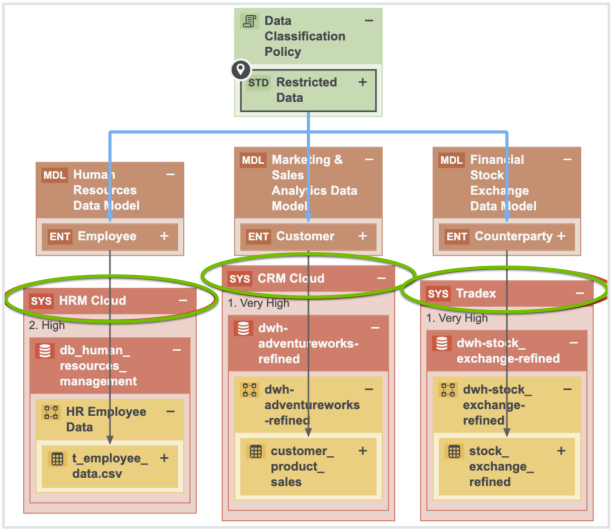


The Privacy team determines the sensitivity and the required security at the data category level as opposed to the column level. At the data category level, the Privacy team then determines what data elements belong to the identified data categories. For example, the

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 organisation's personal information, which makes views such as the Restricted Data Overview diagram available at the most sensitive data category, Standard Restricted Data.

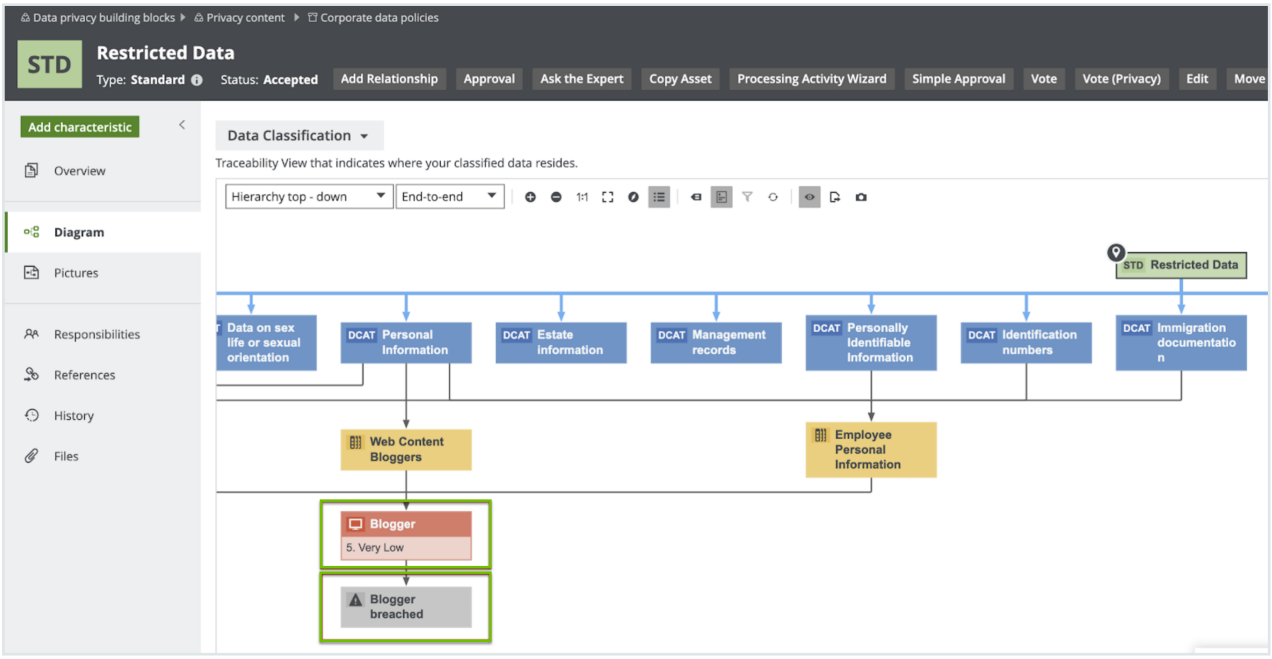


In the above image, the applications in which the restricted data resides are highlighted.

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 image, an investigation is already ongoing on the potential data breach on Blogger.



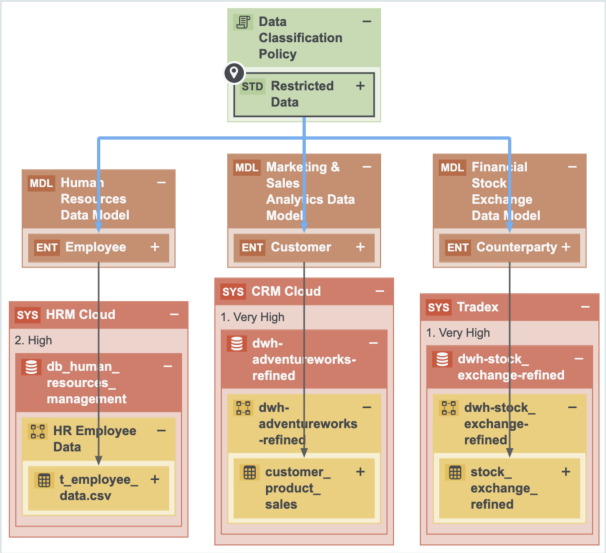
Data classification capabilities and guided stewardship

This section describes how Collibra Privacy and Risk 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.

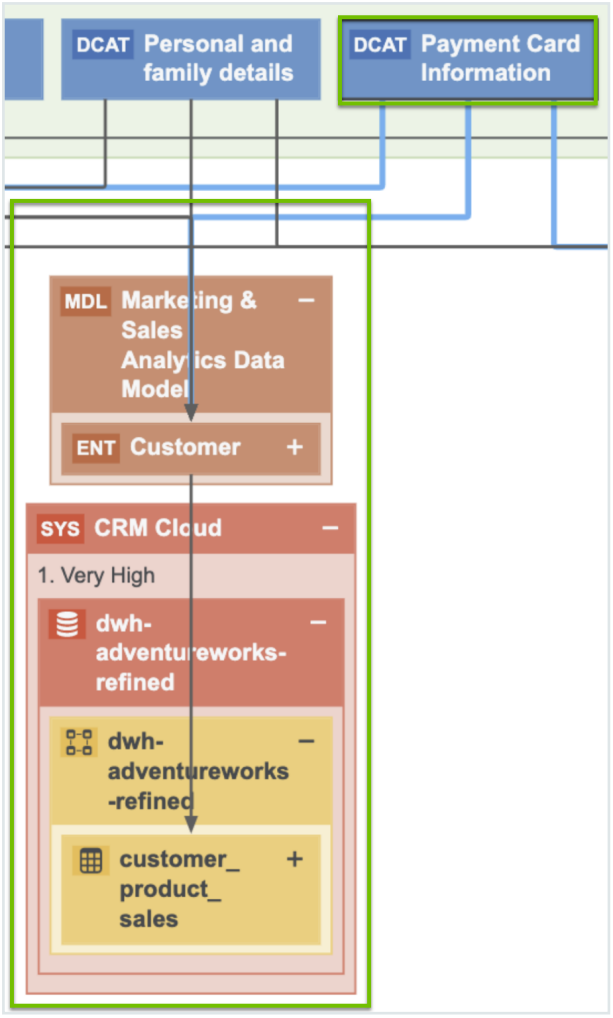
Guided stewardship is a semi-automated process of mapping columns and tables to logical data attributes. It enables content tables to be mapped to data attributes. After scanning a table and then applying guided stewardship in which the steward selects

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.

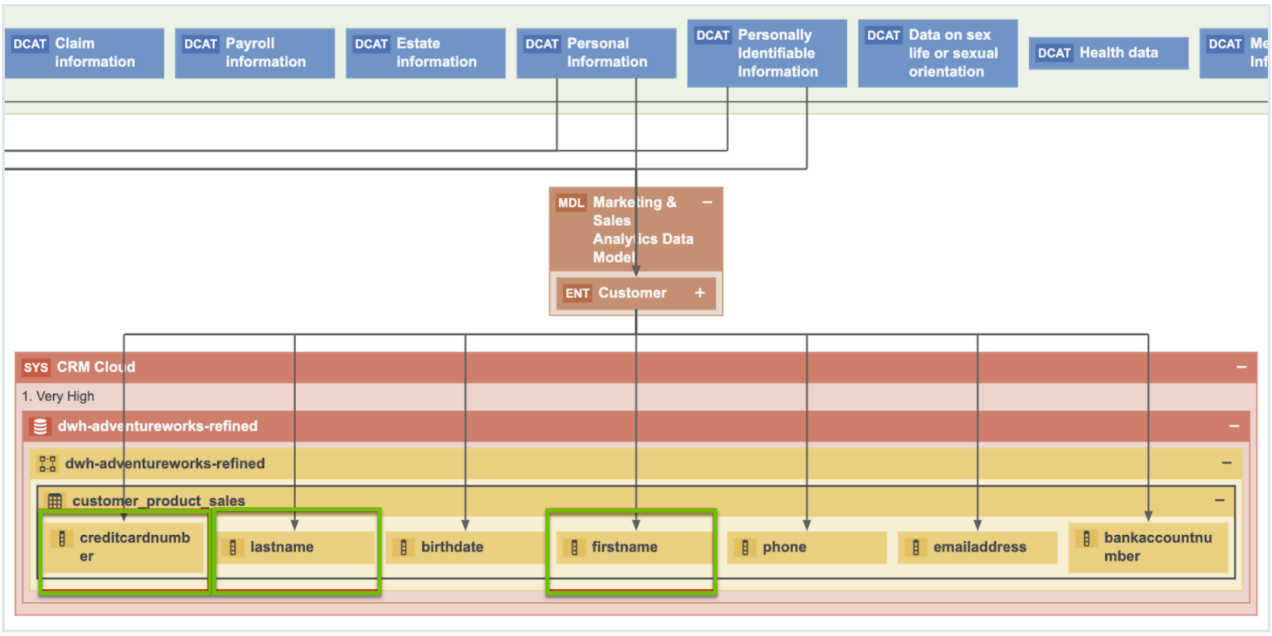


Restricted Data groups multiple data categories. The following image shows the data attributes that the Payment Card Information data category groups.



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 CollibraPrivacy and Risk, gives you a vertical view on where Personal Information resides.

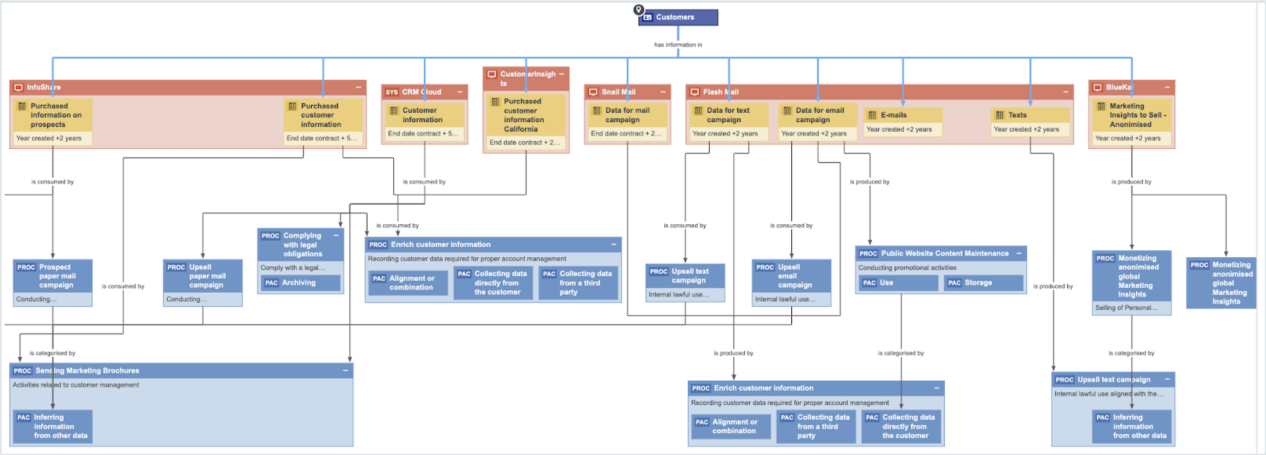
Customer requests and consent management

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.

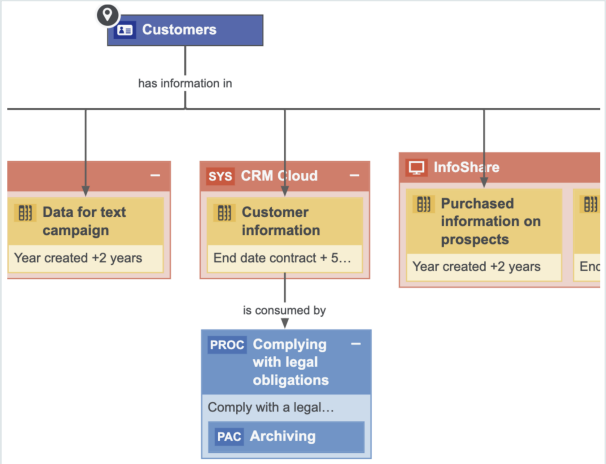
Consider the customer data. Collibra knows where this data resides and how it is being used. This is an outcome of obtaining input from the business users during the onboard of the Business Processes where the users are asked what data they use, which applications they use, for what purpose the data is used. 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.



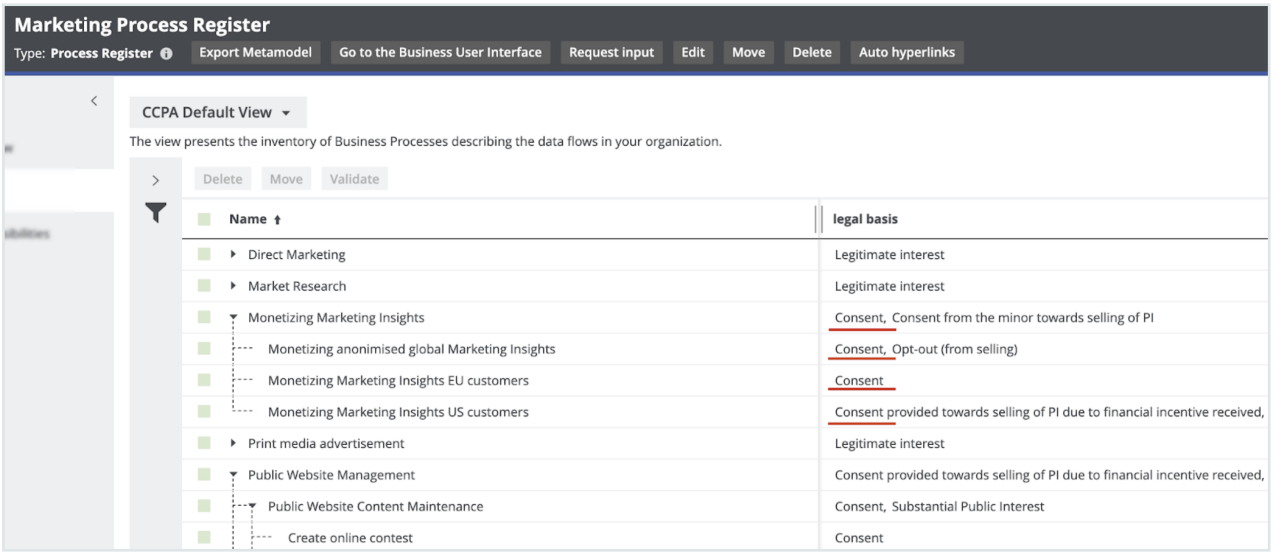
The above image shows where data resides and why it is used. It shows all the applications that contain customer data, and also the related retention periods, which can be imported when a customer wants to exercise their right to be forgotten. Collibra knows in which applications the data resides and 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 the data subject requests. 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 process register, so that you can see all the processes that rely on the consent and the data categories for which you need consent.



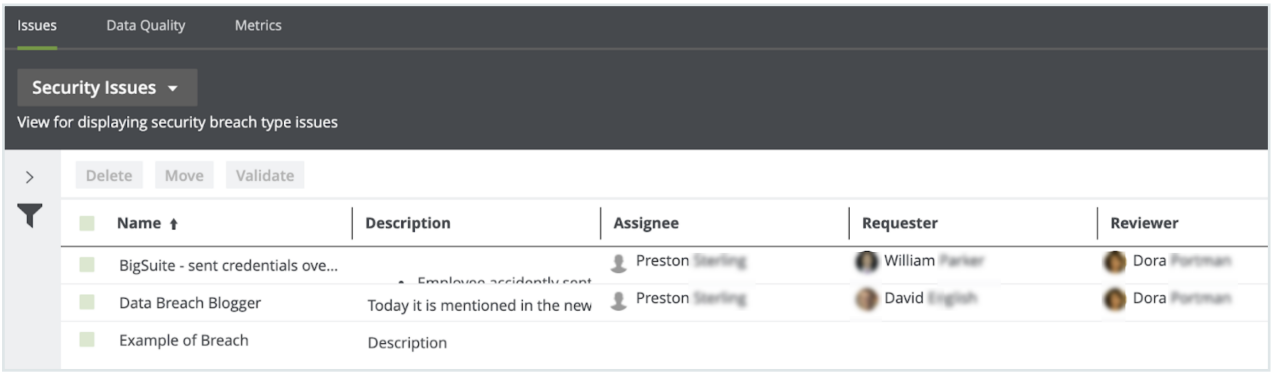
These are stored in the data sets that can also contain granular information, such as the individual data elements for which you want to obtain consent—this combines the information about which business processes require consent and the data categories for which you need consent to process all information 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 the 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 the 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 Process register and Business Process discovery capabilities, data categorization and classification, and different prescriptive paths for reaching out 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

Process Register is an essential part of privacy compliance, foreseen directly by GDPR article 30 as a Record of Processing Activities 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

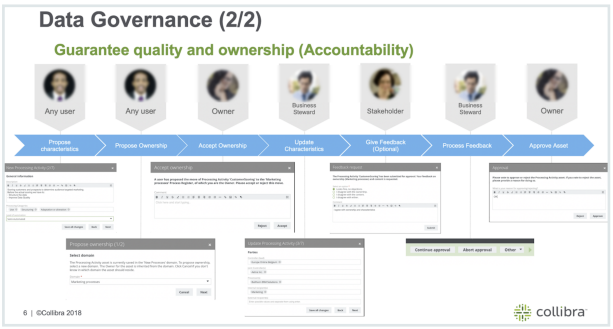
Business Processes may be onboarded by business users as well as privacy stewards through dedicated workflow implementing guided stewardship principle in Collibra Data Privacy. During onboarding, multiple roles collaborate in providing content to the onboarded Business Process. Because of the dedicated 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

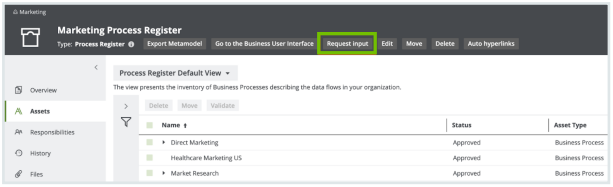
and it may not be easy to centralize this information and ensure that the information is up to date. To help you with this task, CollibraData Privacy comes with the Business Process discovery capabilities.

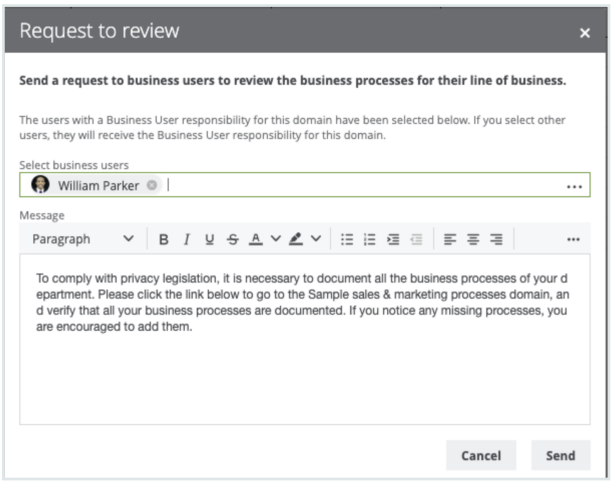
Consider a high-level overview of Collibra Privacy and Risk Business Process discovery capabilities. It commences with the Business Users describing the Business Processes in their terms. They will describe the data being used, applications being used, and any third parties with which they share information. 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 with a dedicated interface

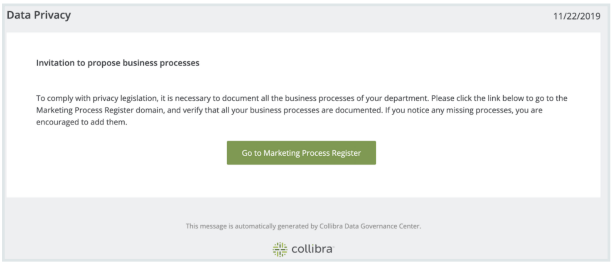
The information related to Business Processes may be requested from the Business User directly from Collibra Privacy and Risk Process Register. Typically, this will be done by those who work on the Privacy program. With the Request input button, email will be generated for the selected business users, which can provide relevant information on the business side of the process through a dedicated interface. 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.



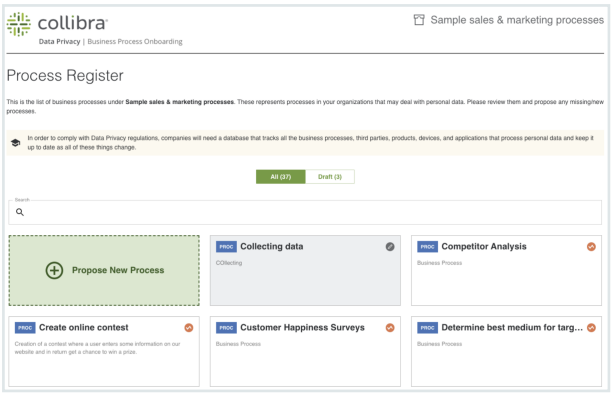


Onboarding Business Process with a business user interface

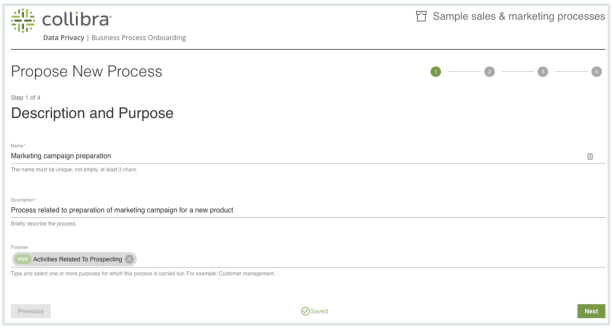
The Business User receives an email message asking them to verify that all the processes are in their domain.



When the Business User clicks Go to Marketing Process Register in the email, a page showing all the Business Processes for their department appears to allow the Business User to contribute to the Process Register.



The link provided in the email message directs the User to a survey where they can describe the business processes that they perform on a daily basis. If the Business User cannot find the Business Process that was onboarded was in the process of being onboarded, they can propose a new Business Process using the Propose Business Process button. When proposing a Business Process, they can describe the Business Process, provide a unique name, description, and purpose.

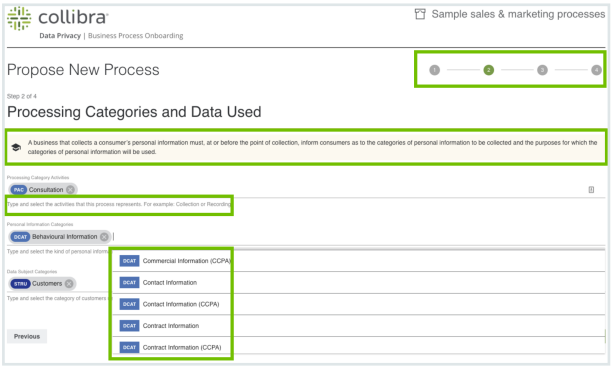


The next step involves covering Process Categories such as use, collection, adaptation, and alteration. The Business User defines the types of data that they are using, for example, behavioral information, contact information, or contract information. Finally, they determine what type of customer's data they are using, such as the customers covered by CCPA or GDPR. There can be also an indication on other types of data subjects, such as, employees and candidates. The Business User can select values only from predefined lists—this reduces the scope of errors as there is no ambiguity around the values that the Business User is able to provide. These values have been predefined by the Privacy team and have legal implications. They show how the organization complies with the privacy regulations. Because, when you collect data directly from customers or from a third parties—by using sensitive information, public information, or customer or employee information—the distinctions will have an impact on how you comply with the regulations.

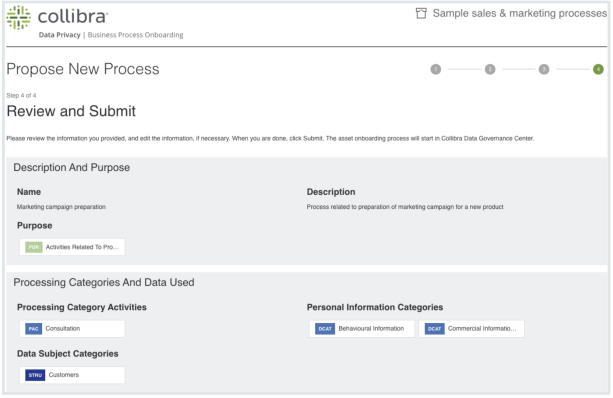
For example, the employee information is temporarily exempted from the CCPA. Therefore, it is considered better for the Business User to select from the drop-down list, as opposed to providing free text. This also prevents common issues such as spelling errors. In addition, if there is any uncertainty about the meaning of these values, the Business User can look up the definitions of these values in Collibra. In the next steps, the lines of business and third parties involved can be described, applications used can be indicated, and the level of automation in the Process can be determined.

The wizard is prescriptive:

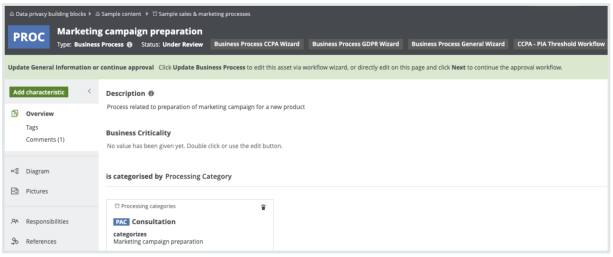
- It shows the user the steps that they have completed and how many steps are remaining, by visually indicating the progress.
- The help text below the question describes what is required from a particular question.
- The ability to open a side panel that provides additional educational information such as the wordings from the law or video content from the Collibra university.
- Smart suggestion based on what the user has already filled and the domain to which they belong.



After entering the information, the Business User can review it before submitting it.

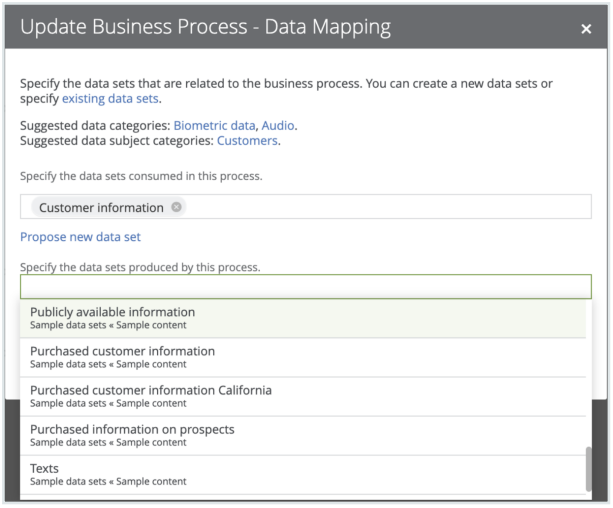


After the Business User provides their Business Process, they can submit the process for further onboarding. The next step is for the owner to accept the ownership of that Business Process. A new task is generated for the owner after they accept the ownership of the Business Process in the Process Register. Based on the metadata, the owner can determine that the Business Process belongs to their Process Register. The ownership can be accepted or rejected. As a result, the status of the asset is changed and the justification is added in the Comments section of the Business Process.

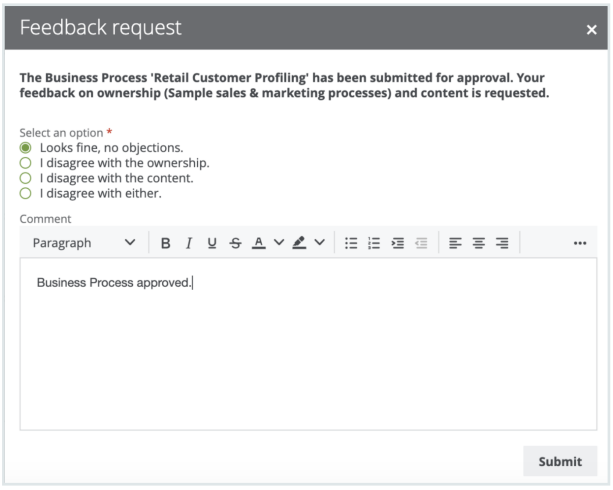


As a next step, the experts or the stewards will be consulted to ensure that the metadata is accurate and complete and the Business Process is going to be mapped to the data. In the following steps, relevant tasks will be created with the request to review and update attributes when necessary. Among others, the Data Mapping task is performed by Data Stewards. The form contains contextual help with suggestions on the relevant Data Sets consumed and produced by the Business Process. Because one of the data categories used in the process is behavioral information, you can click and review it and view the related data sets in the categories. Based on this, Data Stewards can ensure that Data Mapping has been correctly performed.

The last expert who needs to contribute to the Business Process is the Privacy Steward. After opening the task, the first step is to define the regulation that applies, be it GDPR, CCPA, or others. In addition, a purpose needs to be validated, and legal bases, controllers and processors need to be defined. Very specific information on regulation shall be specified, for example, on GDPR, we define cross-border transfers, safeguards, consent collection method, and automated decision-making confirmation. On the CCPA side, we are asked about the collection directly from customers or third parties and whether the data is being sold to third parties.



After the Stewards finish updating the Business Process, we ask the Stakeholders for final feedback. If the feedback is positive, we send the task to the Owner for the final approval.

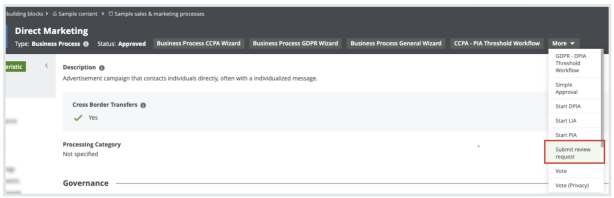


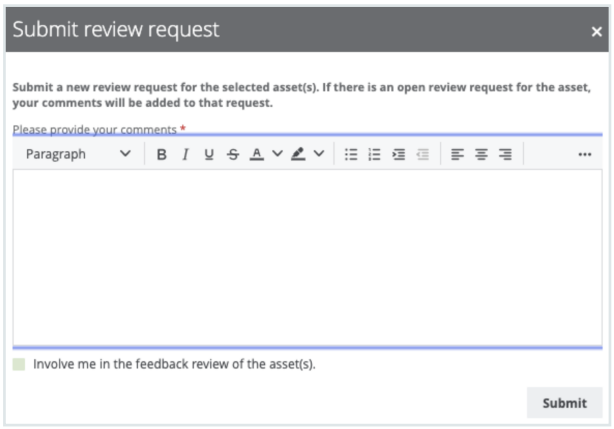
Maintain Process Register over time with review requests

Whereas 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.

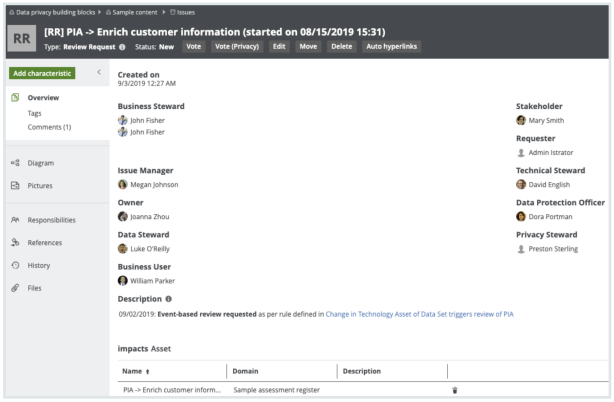
You may have many reasons to review an approved asset. Collibra Data Privacy groups such reasons into three categories and offers three corresponding means to trigger a review request:

- 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.



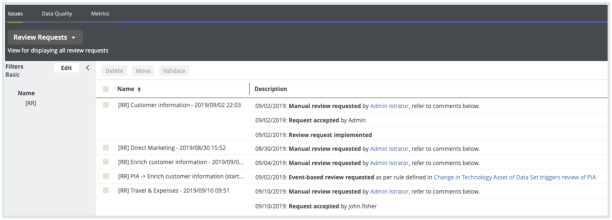


- Time-based: A trigger that is automatically actioned at a specified frequency. This is useful for assessment assets for which you might be required to review periodically to comply with a regulation.



- 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 the Data Helpdesk.



Perform Assessments

Conduct PIA and DPIA

If a business process is likely to introduce a level of risk to the rights and freedoms 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 freedoms of natural persons to risk is significant. Privacy Impact Assessments (PIA) and Data Privacy Impact Assessments (DPIA) assess the risks to the rights and freedoms 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:

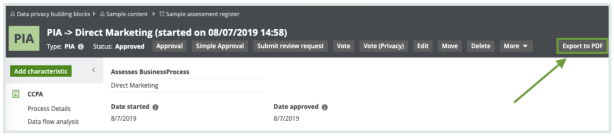
- PIA, if complying with CCPA
- DPIA, if complying with GDPR

Print assessment results

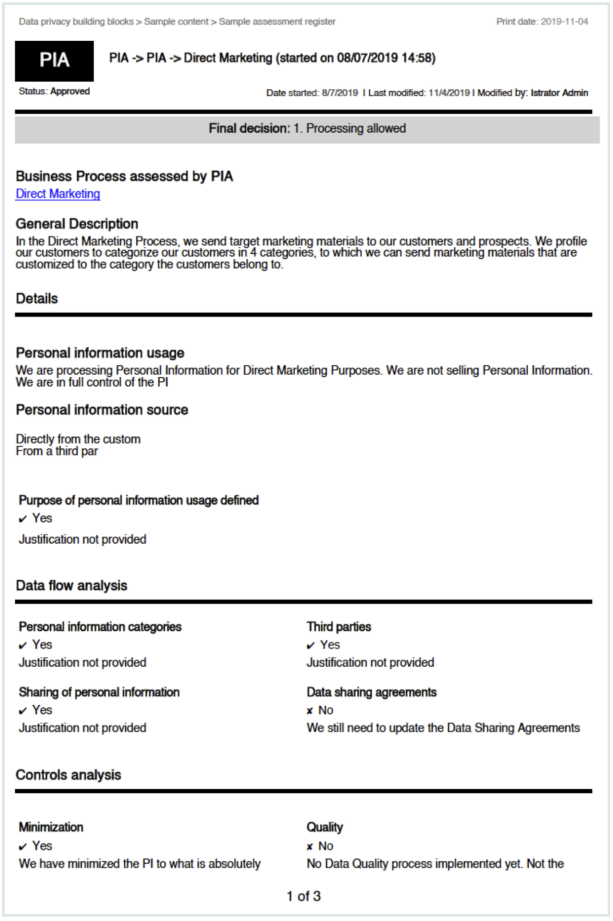
Assessments are a way for an organization to demonstrate compliance. You can export and print the PIA results in a unified way. You can also download a PIA asset page as a printable PDF, regardless of the status of the PIA asset.

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 Protect to the best of its ability.

Types of data protection

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.

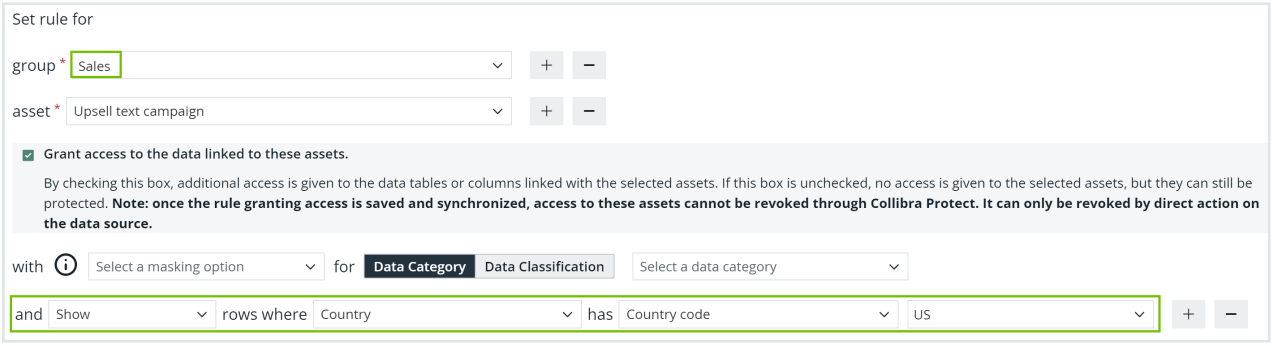
Suppose that you want the Human Resources (HR) group to be able to access a data set of US-based customers. Suppose that certain parts of the data set need to be hidden from the HR group because they contain restricted data, such as personally identifiable information (PII). Then, you can further protect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

Row-based protection

Row-based protection allows 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.

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.

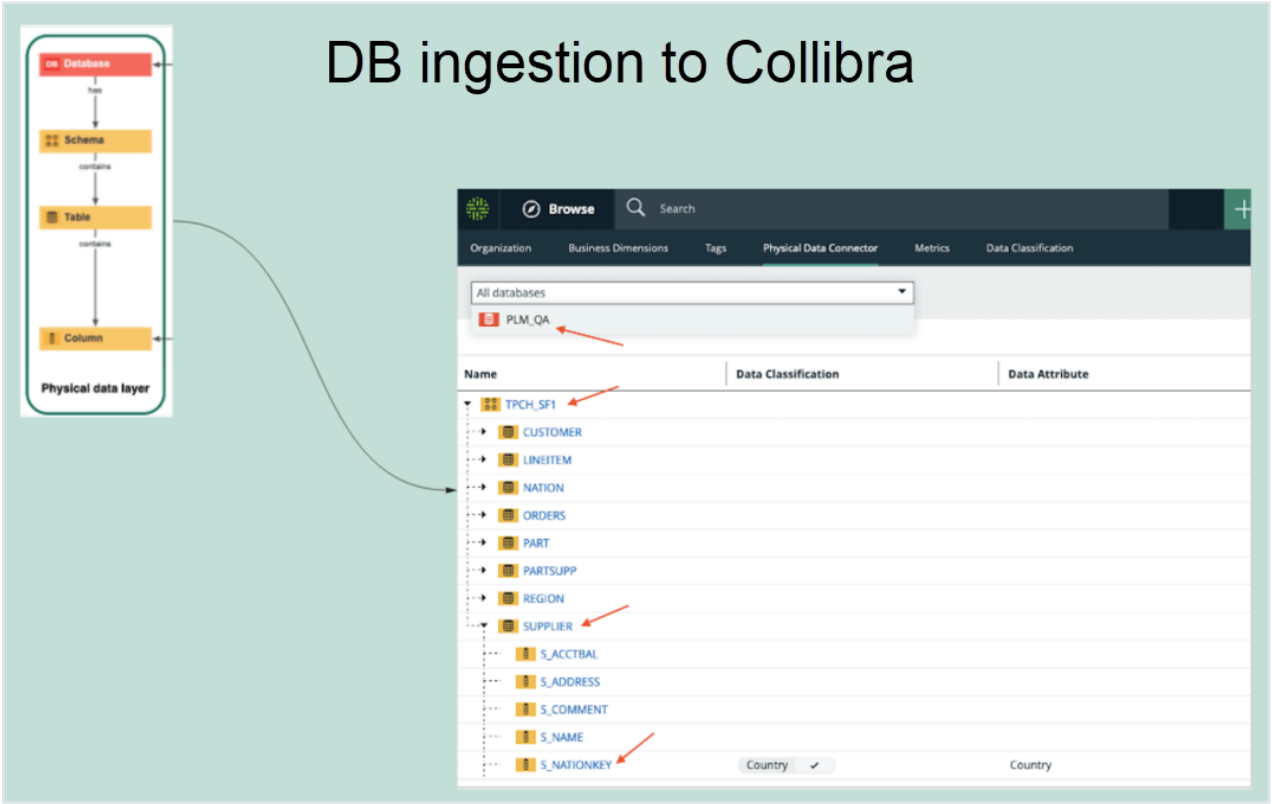


Technical background

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.

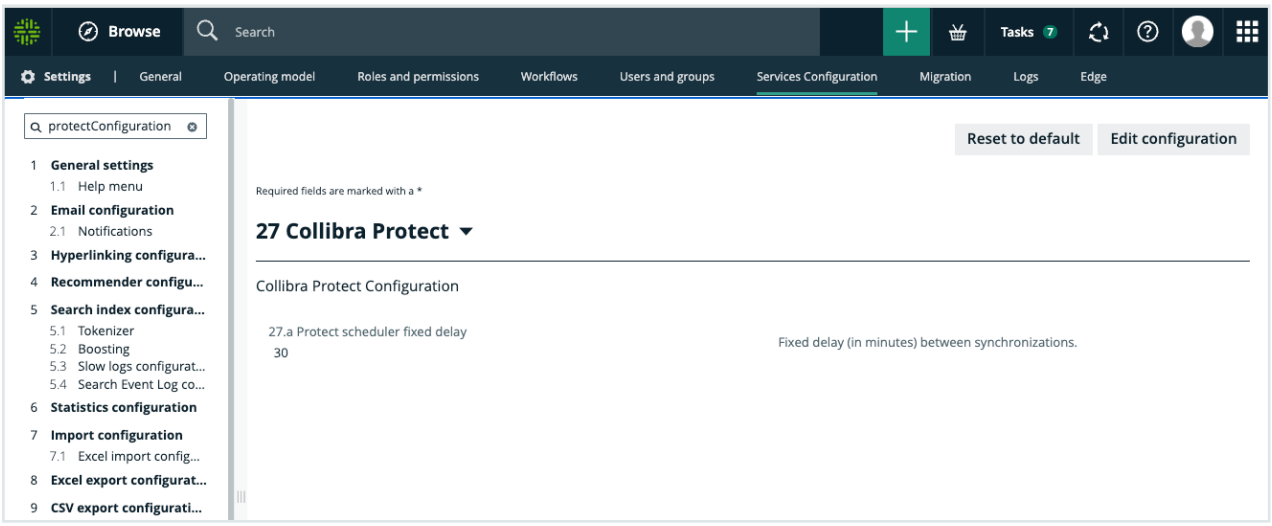


After the physical layer is created in Collibra, the logical layer can be created on top of the physical layer, as follows:

- 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.

Synchronization

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

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 they reside, 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.

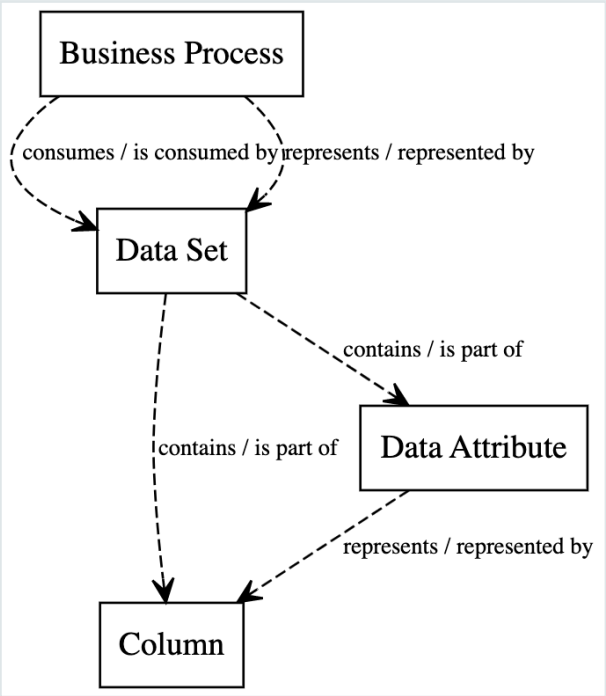
Prescriptive paths

You can use 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.

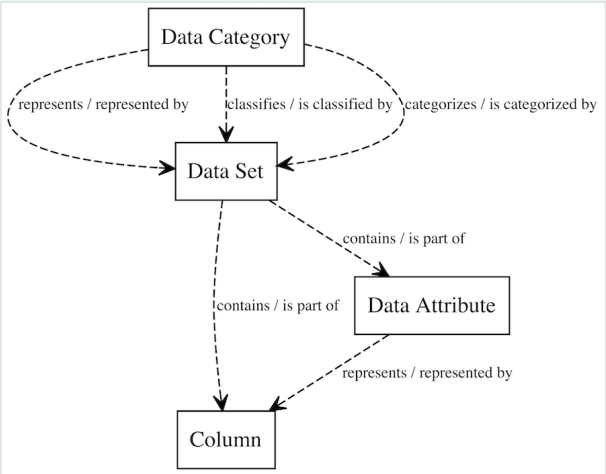
The assets that you use to create data protection standards and data access rules are related to the physical data layer, such as tables and columns, through a set of relations and intermediate assets. Protect uses these relationships and intermediate assets to search the knowledge graph to find the physical data layer assets that it needs to protect. The traversal of the knowledge graph 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.

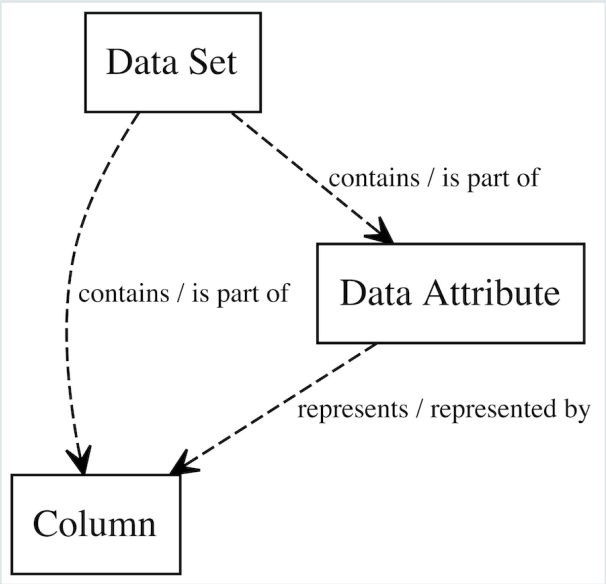
Business Process



Data Category



Data Set



Customization of prescriptive paths

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.

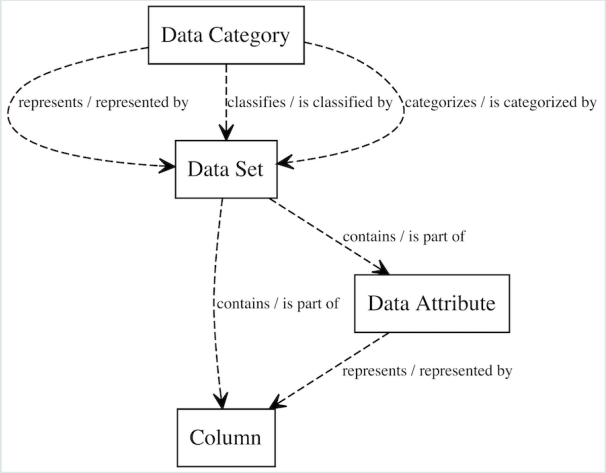
If you have the **Protect** > **Administration** global permission, you can customize the prescriptive paths for the asset types through APIs. The customization may include creating, modifying, or deleting the prescriptive paths: for example, adding or modifying the prescriptive paths for packaged and custom asset types, defining how the asset types relate to columns, 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 data protection standard or a data access 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"
      }
```

Install Protect

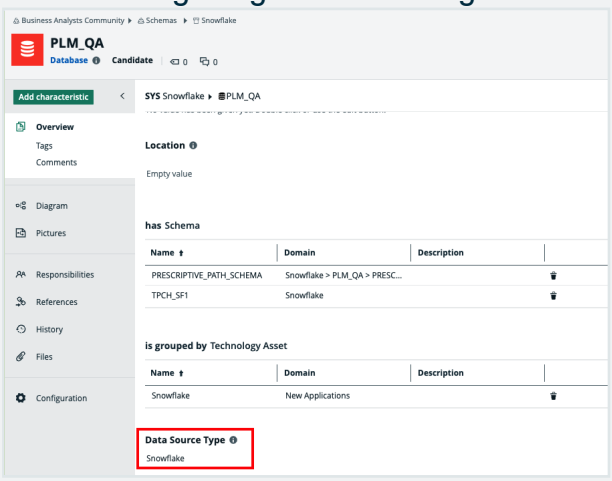
This topic describes how to install Protect.

Before you begin

1. Add the Catalog ingestion capability to the Edge site and register the required data source.
2. Add the required Protect capability to the Edge site. When adding the capability, in the **Capability template** field, select one of the following values, as applicable: **Collibra Protect for Google BigQuery** or **Collibra Protect for Snowflake**

Tip

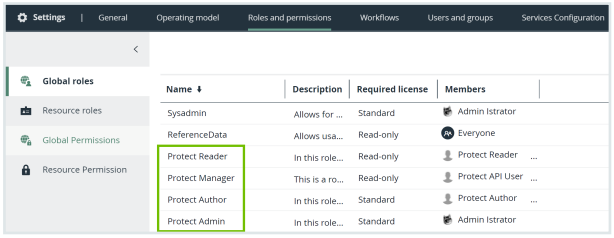
- If you selected **Collibra Protect for Snowflake**, in the **Snowflake Connection** field, you can select the Snowflake connection that was used for performing catalog ingestion, if the connection allows Protect to write to your database.
- The following image shows an ingested Snowflake database.



The **Data Source Type** attribute is added to the database asset only after the catalog ingestion process is complete.

Steps

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



» Protect is installed. 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 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 data protection standards and data access rules, modify only the standards and rules that they created, view imported policies, view groups, and 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 all data access rules and data protection standards and also access additional APIs.
Protect Manager	A user who manages background processes and configures Protect.This role is intended only for the Protect system user.

Global permission	Description
Protect > Edit	Allows a user to create data protection standards and data access rules and modify only the standards and rules that they created.
Protect > Administration	Allows a user to create data protection standards and data access rules and modify all standards and rules.

Open Protect

This topic describes how to open Protect, including how you can use the tabs on the Protect landing page.

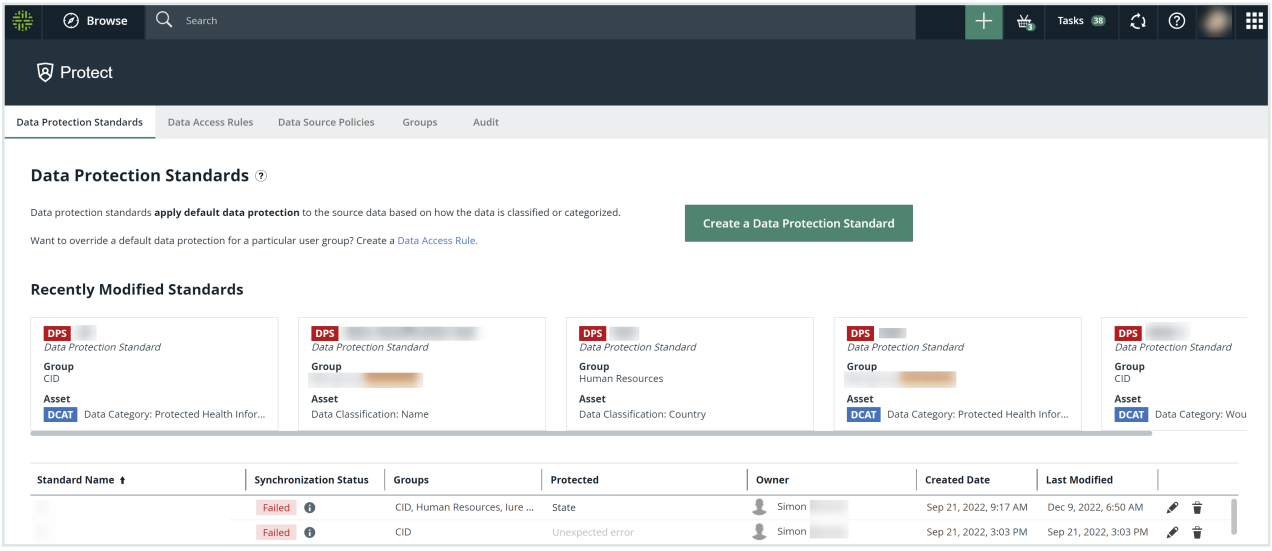
Requirements and permissions

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 is shown.



Protect landing page

On the Protect landing page, depending on your role, the following tabs are shown.

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 data protection standards and data access 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

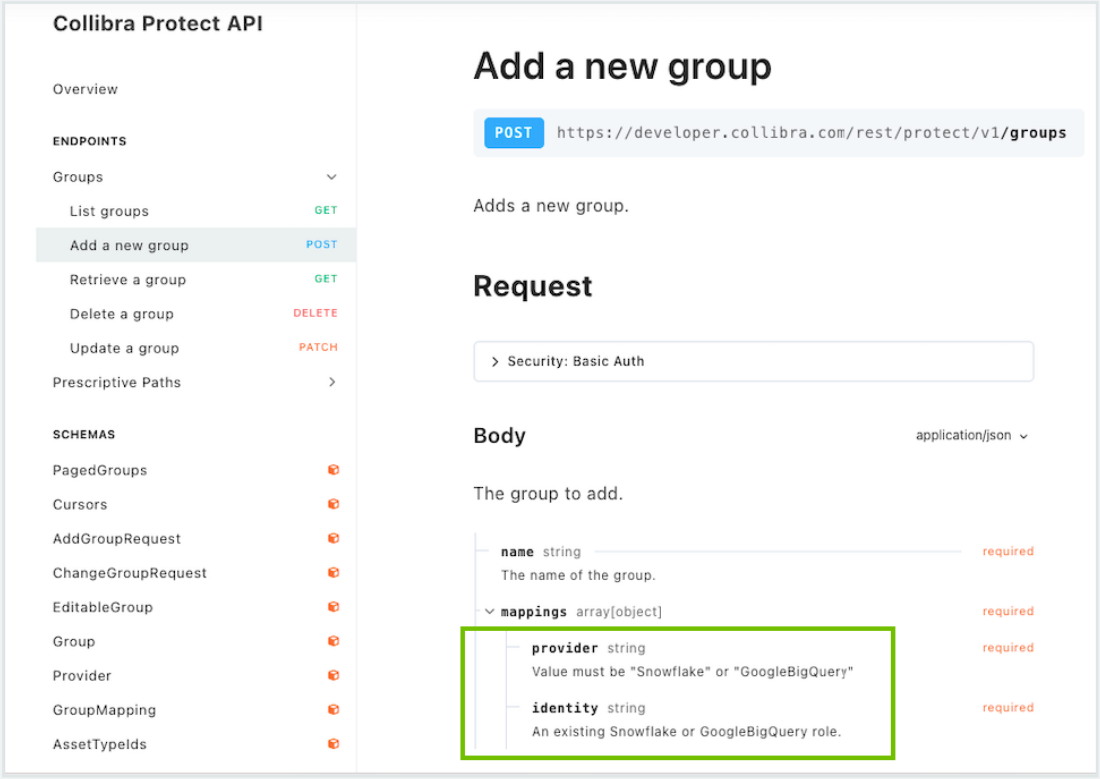
You have the **Protect Author** or **Protect Admin** global role.

Steps

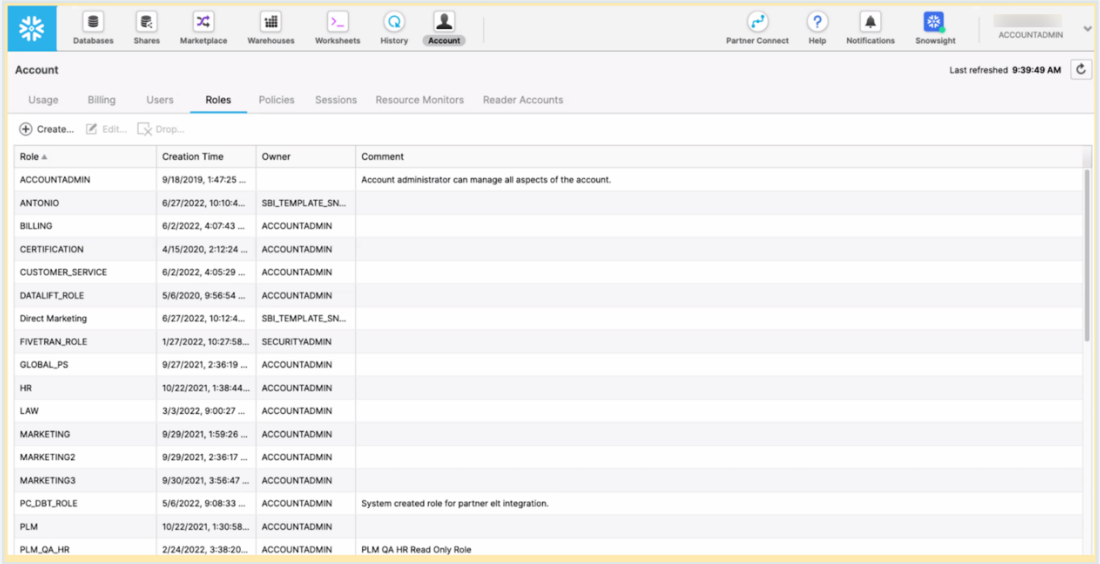
1. Open Protect.
2. On the **Groups** tab, click **Collibra Protect Group API**.
3. For the next steps, go to Add a new group.

Tip

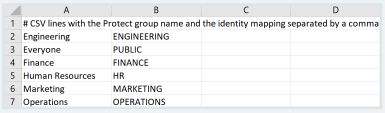
- When creating a Protect group, you are prompted to specify the data source provider (for example, **Snowflake**) and the existing role from the data source provider to map the role to the group.



- The following image shows the roles in the Snowflake data source provider.



- 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 (that is, they should have the global role Protect Author and/or Protect Admin)
COLLIBRA_AUTH="user:password"

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 the first line in the CSV file
    while IFS=, read -r field1 field2 field3
    do
        echo "Add group $field1 with Snowflake role $field2 and GCP $field3"
        curl -u "${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": "Snowflake",
                    "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 role associated with the group.
Created By	The name of the user who created the group.
Created Date	The date when the group was created.

Data protection standards

Data protection standards create a primary layer of protection for similar types of data by masking the data wherever they reside. They protect your data by masking the data (column-based protection).

Create a data protection standard

Requirements and permissions

- You have the **Protect Author** or **Protect Admin** global role.
- 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

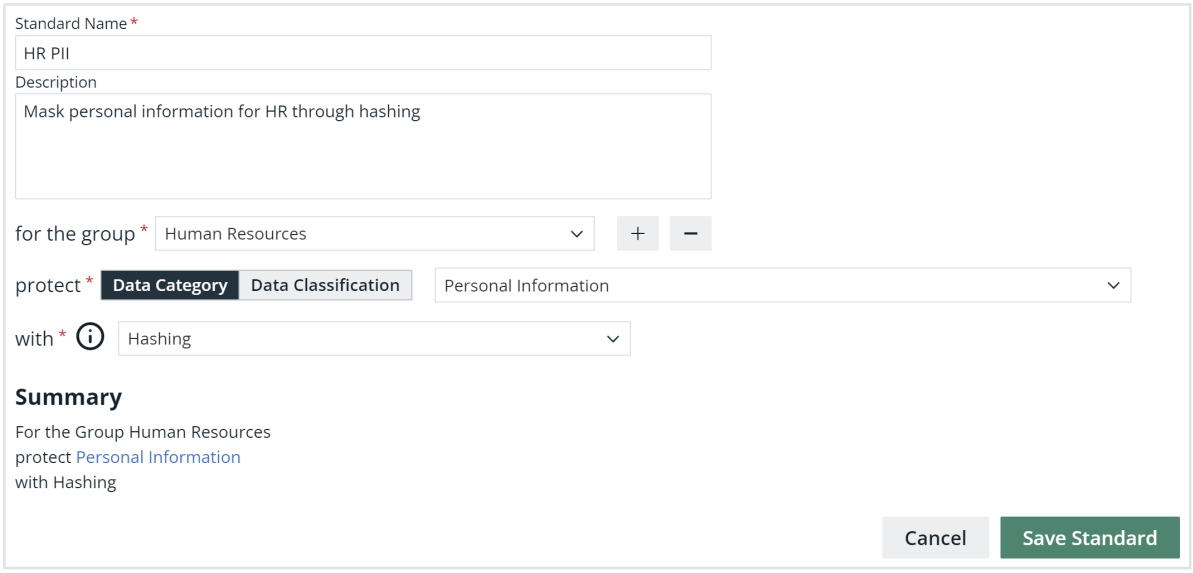
1. Open Protect.
2. Click the **Data Protection Standards** tab.
3. Click **Create a Data Protection Standard**.
   » The **Create a Data Protection Standard** dialog box appears.
4. Enter the required information.

Details

Field	Description
Standard Name	Enter a name to identify 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.
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. The following options are available: ◦ **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 the **Protect Author** or **Protect Admin** global role.

  > Note If you have the **Protect Author** global role, you can modify only the data protection standard that you created. If you have the **Protect Admin** global role, 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

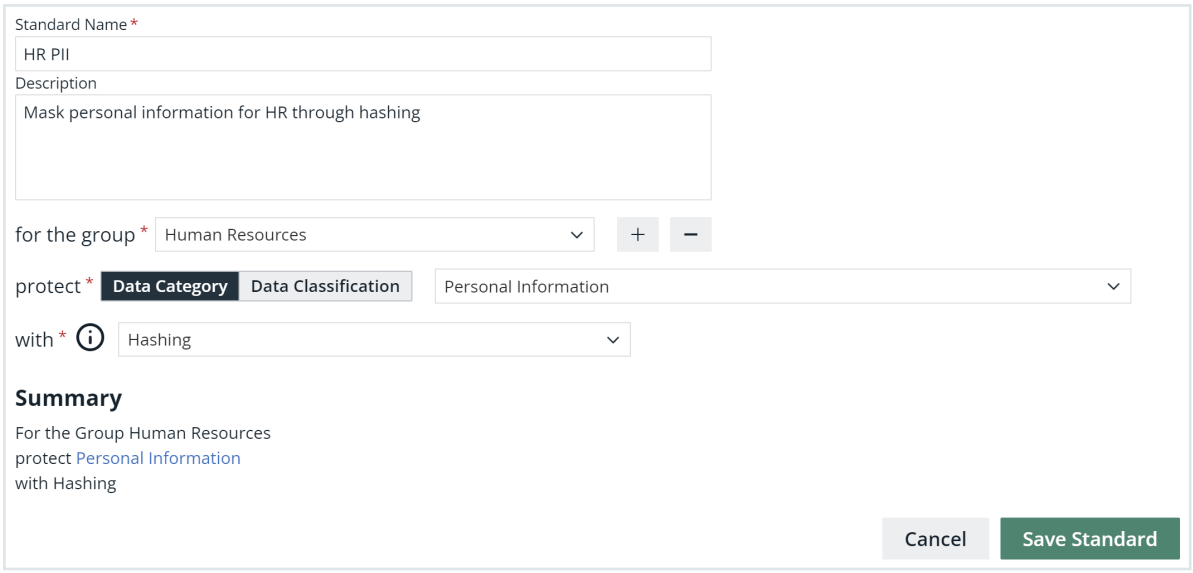
1. Open Protect.
2. In the table, in the row containing the standard that you want to modify, click ✎.
   » The **Edit a Data Protection Standard** dialog box appears.
3. Modify the required information.

Details

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

Field	Description
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.
With (masking option)	Select the type of masking that you want to apply to the selected data category or data classification for protection. The following options are available: ◦ **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 the **Protect Author** or **Protect Admin** global role.

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 If you have the **DataSteward** global 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 create an additional layer of protection 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 the **Protect Author** or **Protect Admin** global role.
- 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 to identify the data access rule.
Optional: Description	Enter a description for the data access rule.
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.

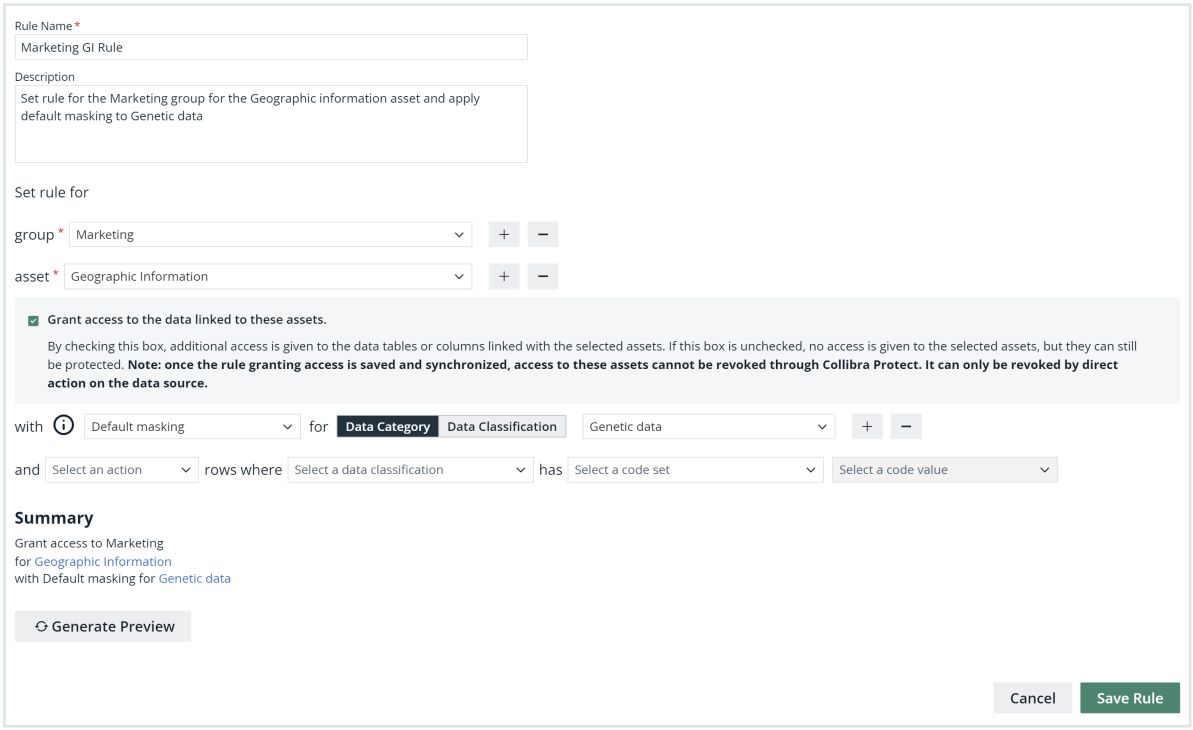
Field	Description
Asset	Select the data asset that the rule is protecting. This field contains Business Process, Data Category, and Data Set assets, as well as assets of custom asset types. Tip ◦ For more information, go to Technical background and Prescriptive paths. ◦ You can add more groups by using 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. The following options are available: ▪ **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. Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. The following actions are available: ◦ **Show** ◦ **Hide** a. In the **rows where** field, select the data classification that you want to show or hide. b. In the **has** field, select the code set for the selected data classification. c. 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 is selected by default. A selected **Grant access to the data linked to these assets** checkbox indicates that you are granting access to the tables and the columns in the database that are linked to the selected assets to the groups that you selected in the rule. If you do not want to grant this level of access to the selected groups, 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 the **Protect Author** or **Protect Admin** global role.

  > Note If you have the **Protect Author** global role, you can modify only the data access rule that you created. If you have the **Protect Admin** global role, 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 to identify the data access rule.
Optional: Description	Enter a description for the data access rule.

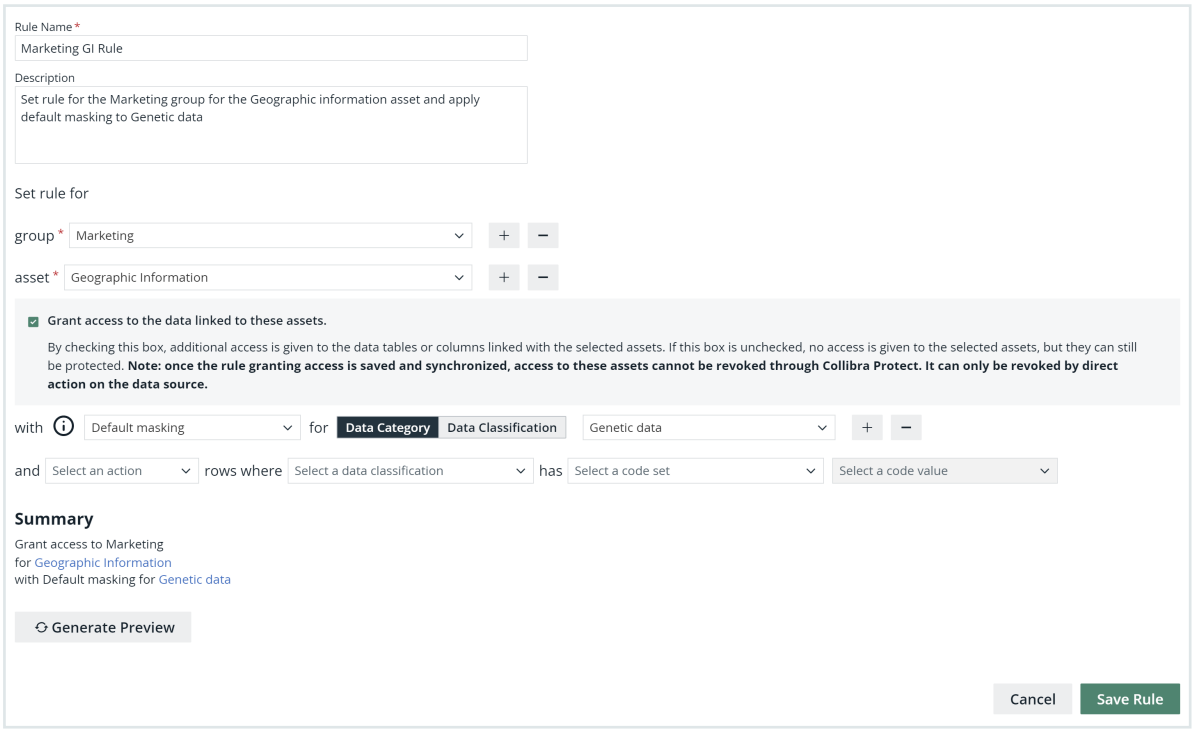
Field	Description
Group	Select the group for the data access rule. Tip You can add more groups by using the plus icon.
Asset	Select the data asset that the rule is protecting. This field contains Business Process, Data Category, and Data Set assets, as well as assets of custom asset types. Tip ◦ For 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. The following options are available: ▪ **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. Tip You can add more data categories and data classifications for masking by using the plus icon.

Field	Description
Optional: And (action)	Select the type of row-filtering action that you want to apply to a data classification with a specific code set and code value. The following actions are available: ◦ **Show** ◦ **Hide** a. In the **rows where** field, select the data classification that you want to show or hide. b. In the **has** field, select the code set for the selected data classification. c. 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 **Grant access to the data linked to these assets** checkbox indicates that you are granting access to the tables and the columns in the database that are linked to the selected assets to the groups that you selected in the rule. If you do not want to grant this level of access to the selected groups, 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 the **Protect Author** or **Protect Admin** global role.

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 If you have the **DataSteward** global 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

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

The **Data Source Policies** tab in Protect contains an overview of the native data source policies. To view the policies, you need the **Protect Author** or **Protect Admin** global role.

> Note Contact Collibra support to import policies from the data source using the Collibra Protect Data Source Policies API.

The table on the **Data Source Policies** tab contains the policies that are active in the data source. These include both the policies that are manually created in the data source and the policies that are generated in the data source as a result of the data protection standards and data access rules in Protect.

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.

Column	Description
Data Source	The data source provider.

Data source providers

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 Google BigQuery** and **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 Snowflake

Data protection standards in Collibra Protect rely on the tag-based masking policies of Snowflake. The name of the data category or data classification selected in a standard becomes a tag with the same name. The tag is applied to all the affected columns to enforce data protection. For more information, go to Snowflake examples.

Types of policies on Snowflake

There are three types of policies on Snowflake: Column-based policies, row access policies, and tag-based policies. Each type can be created in Collibra Protect or on Snowflake.

For rules, policies are created directly on the column level. Row access policies are created when row filters are specified. For standards, the policy is created, attached to a Snowflake tag, and attached to the tab on any affected column.

Snowflake examples

This topic contains examples to describe how Snowflake behaves in relation to certain data protection standards and data access rules.

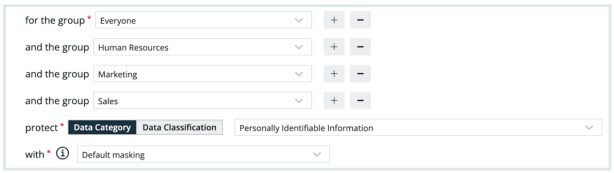
Example 1

Introduction

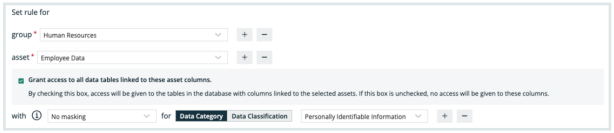
This example describes the behavior in Snowflake when a standard is applied to a data category and a rule is applied to a data set with categorized columns in Protect.

The example considers the following:

- A standard created for the **Everyone**, **Human Resources**, **Marketing**, and **Sales** groups, to protect the columns in the **Personally Identifiable Information** data category by default masking.

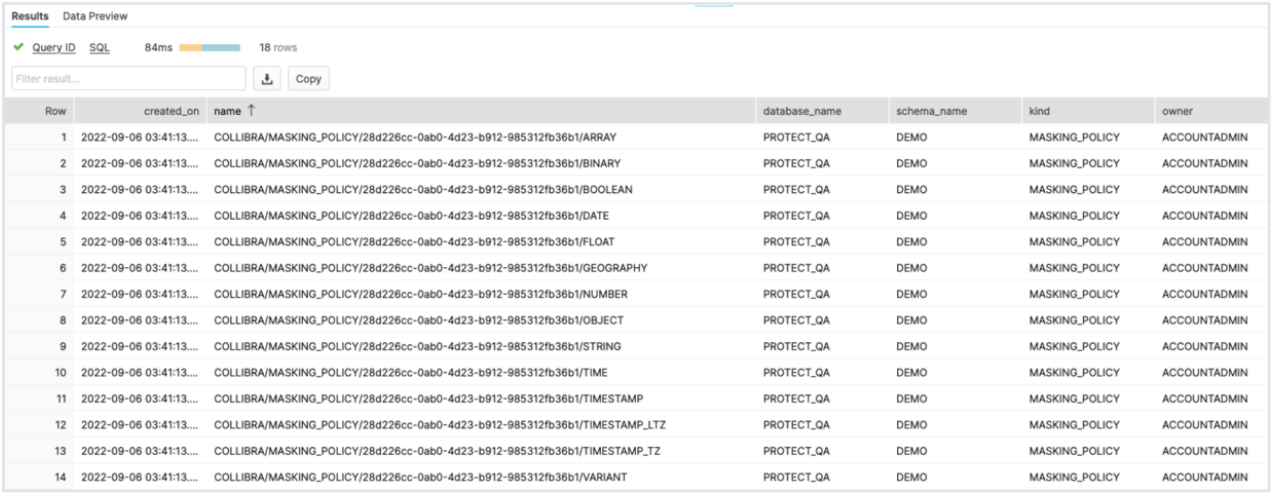


- A rule created for the **Human Resources** group and the **Employee Data** asset, without any protection applied to the columns in the **Personally Identifiable Information** data category.



Standard

When the standard is synchronized and active, the standard results in 14 masking policies—one policy for each Snowflake data type. The masking policies are created at the schema level with the following naming convention: 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

All the masking policies are then associated with the **Personally Identifiable Information** tag, which is created at the schema level and assigned to those columns that need to be protected. At runtime, Snowflake fetches the right masking policy based on the column data type.

```
35 SHOW TAGS;
36
```

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.

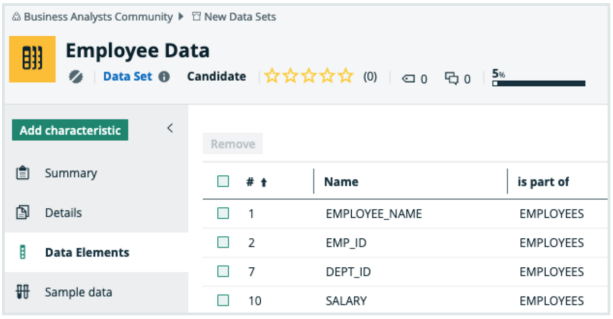
Details

```
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

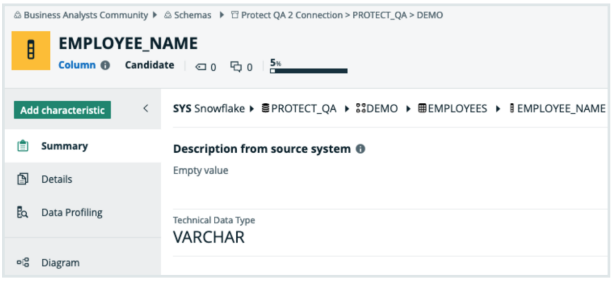
A rule results in a combination of grant instructions, dynamic masking, and row access policies.

Suppose that the **Employee Data** data set selected in the rule contains sensitive columns categorized as **Personally Identifiable Information**.

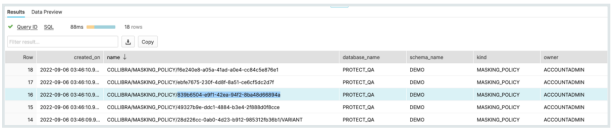


The rule grants access of the **Employee Data** data set to the **Human Resources** group, as indicated by the selected **Grant access...** checkbox in the rule. Then, the corresponding Snowflake role for the group can access each database, schema, and table in the data set. In addition, the column masking 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 **Personally Identifiable Information** 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 use the following naming convention: COLLIBRA/MASKING_POLICY/<asset ID>.



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

Details

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

Behavior

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 a column, the column masking policy takes precedence and the policy tag is not assigned to the column. To mitigate this behavior and 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 the **Personally Identifiable Information** column by default masking, 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 via default masking.

Example 2

Introduction

This example describes the behavior in Snowflake when multiple standards affect the same column without conflict.

The example considers the following:

- A standard created for the **HR** group to protect the columns in the **Personally Identifiable Information** data category by hashing.
- A standard created for the **Marketing** group to protect the columns in the **Personal Information** data category by default masking.
- The **Personally Identifiable Information** and **Personal Information** data categories share the same column named **DOB**.

Behavior

Protect creates a tag for each standard and adds a policy to each tag. The two tags are then linked to the **DOB** column. In addition, Protect creates a masking policy that is an aggregation of the policies from the two tags. This aggregated masking policy, which is then applied to the **DOB** column, thus contains the content of both the tag policies.

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

When a policy exists for the **DOB** column, Snowflake considers only the column masking policy, ignoring all the tag policies associated with the column. Because the column masking policy is an aggregation of all the tag policies, the protection that is defined in the two standards is not ignored.

Thus, Protect handles multiple standards with tag policies for Snowflake by creating a column masking policy, which considers the protection defined in the standards.

Snowflake masking and 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 privileges

To perform actions in Snowflake, Collibra Protect uses an Edge connection that must be configured with a user and a role that can manage grants; create and assign masking policies, row access policies, and tags; and manage usage access on databases and schemas involved in the protection. This enforcement role requires the following Snowflake privileges.

Snowflake privilege	Description
`[APPLY MASKING], [APPLY ROW ACCESS], [APPLY TAG], [MANAGE GRANTS], [IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE]`	Required for the role performing the actions.
`[USAGE]`	Required on each database and schemas where policies are applied to the role performing the actions.

Snowflake privilege	Description
`[CREATE MASKING POLICY], [CREATE ROW ACCESS POLICY], [CREATE TAG]`	Required on each schema where policies are applied to the role performing the actions.

Example

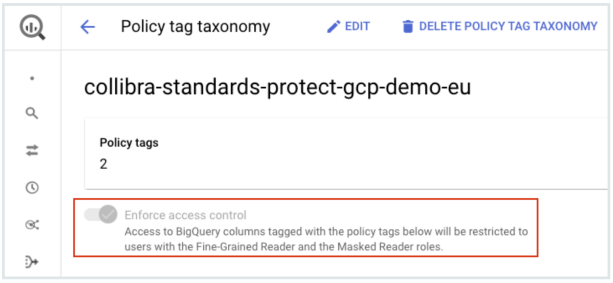
Suppose that a role named PROTECT exists in Snowflake and is responsible for managing access on all schemas within a database named DEMO. Then, the following statements can be used to enable the Snowflake PROTECT role to perform the enforcement.

```
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;
```

Protect for BigQuery

Collibra Protect uses Google's Policy tag taxonomies to create tags and assign the tags to your BigQuery columns. Policy tag taxonomies inherently apply access control. This

means that the tags applied to your BigQuery columns will be accessible only by the Protect groups configured in your data protection standards and data access rules.



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**.

Protect masking type	Equivalent BigQuery masking rule
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.
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)

BigQuery data type	BigQuery masking rule supported by Protect
DATE	Default masking value
DATETIME	Default masking value
FLOAT64	Default masking value
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 Collibra Protect group mapping for BigQuery must follow the syntax for principal identifiers. For example, the Protect group, **Sales**, maps to the BigQuery group email address, **sales@example.com**.

```
{
  "name": "Sales",
  "mappings":
  [
    {
     "provider": "GoogleBigQuery",
```

```
      "identity": "group:sales@example.com"
    }
  ]
}
```

BigQuery permissions

To perform actions in BigQuery, Collibra Protect uses a GCP connection that must be configured with a service account having 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.jobs.create
- bigquery.rowAccessPolicies.create
- bigquery.tables.get
- 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

Audit

An audit log 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 and what is shown in an audit log.

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 Collibra Protect varies from several minutes to hours, depending on the data source.

Requirements and permissions

You have the **Protect Author** or **Protect Admin** global role.

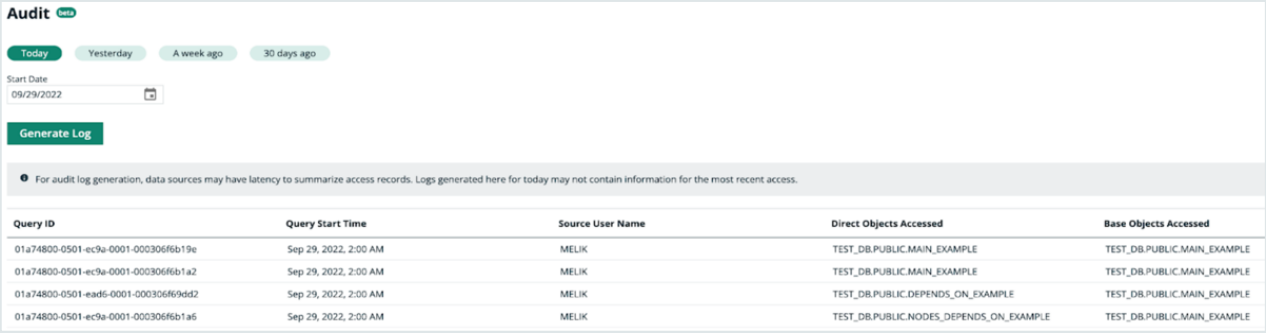
Steps

1. Open Protect.
2. Click the **Audit** tab.
3. Depending on your data source, click **BigQuery** or **Snowflake**.
4. 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, and then 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.

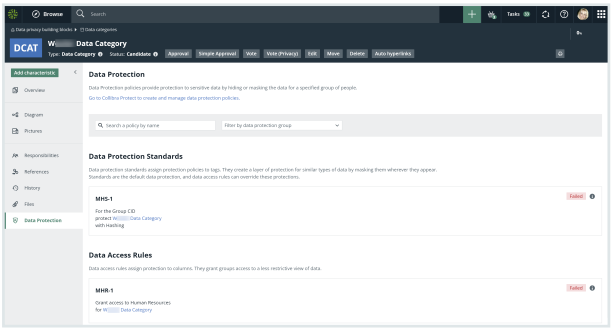
Column	Description
Query ID	The ID of the query in the source database.
Query Start Time	The date and time of the query in the source database.
Source User Name	The name of the user in the source database who ran the query to access the data.
Direct Object Accessed	The database object (a table or a view)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

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.

Why rules or standards may 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.

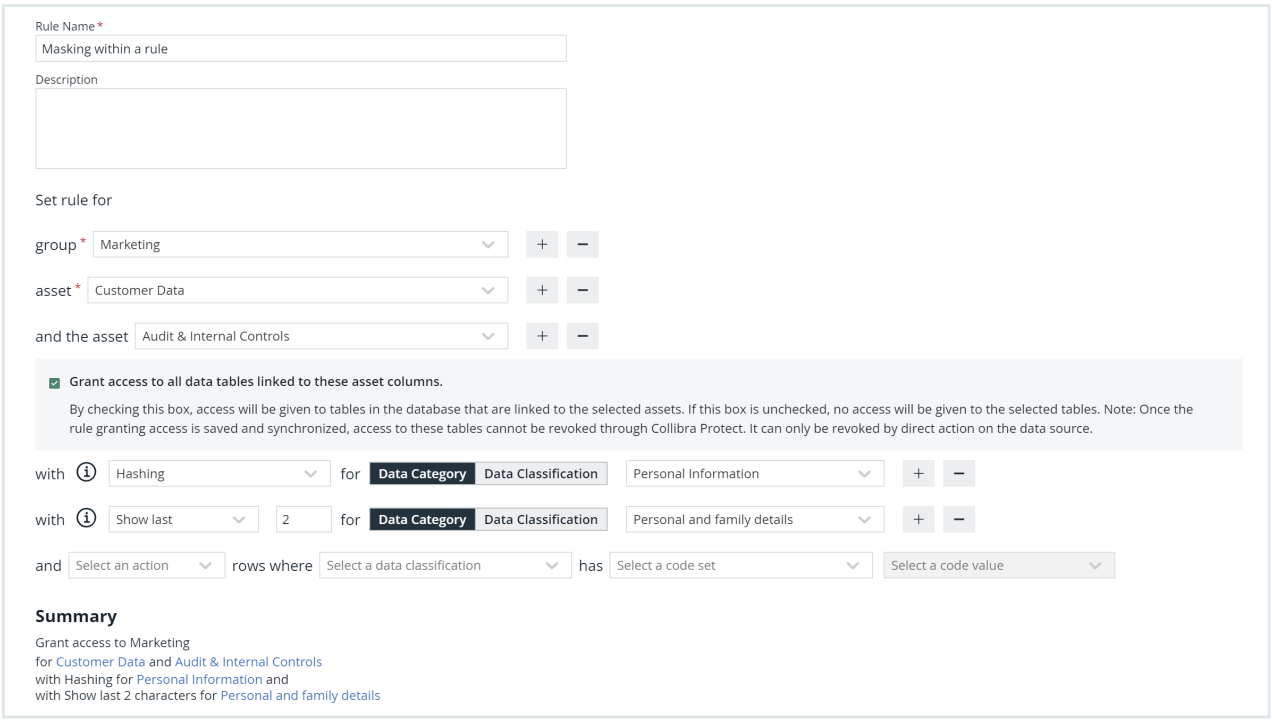
Masking within a rule

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.

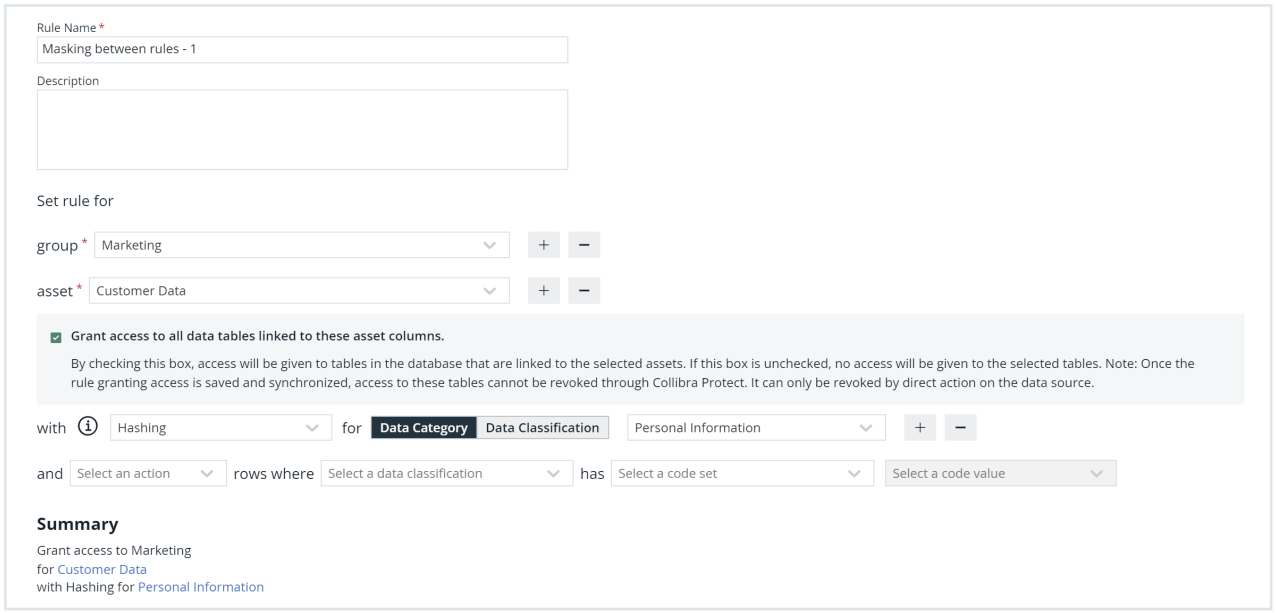


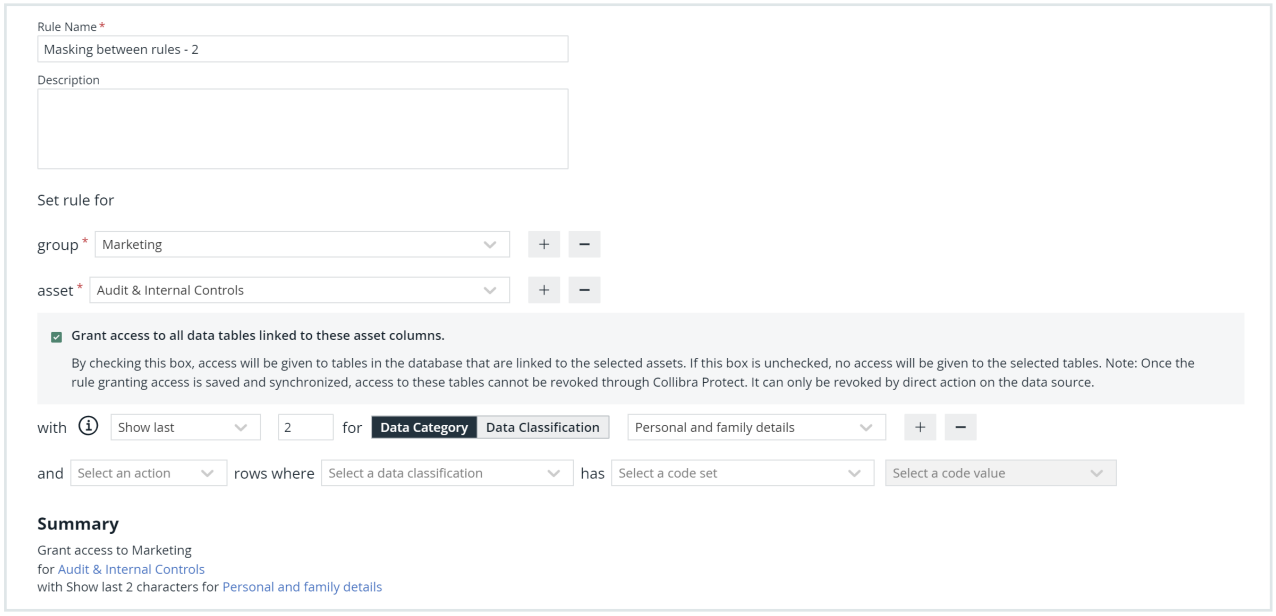
Masking between rules

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.

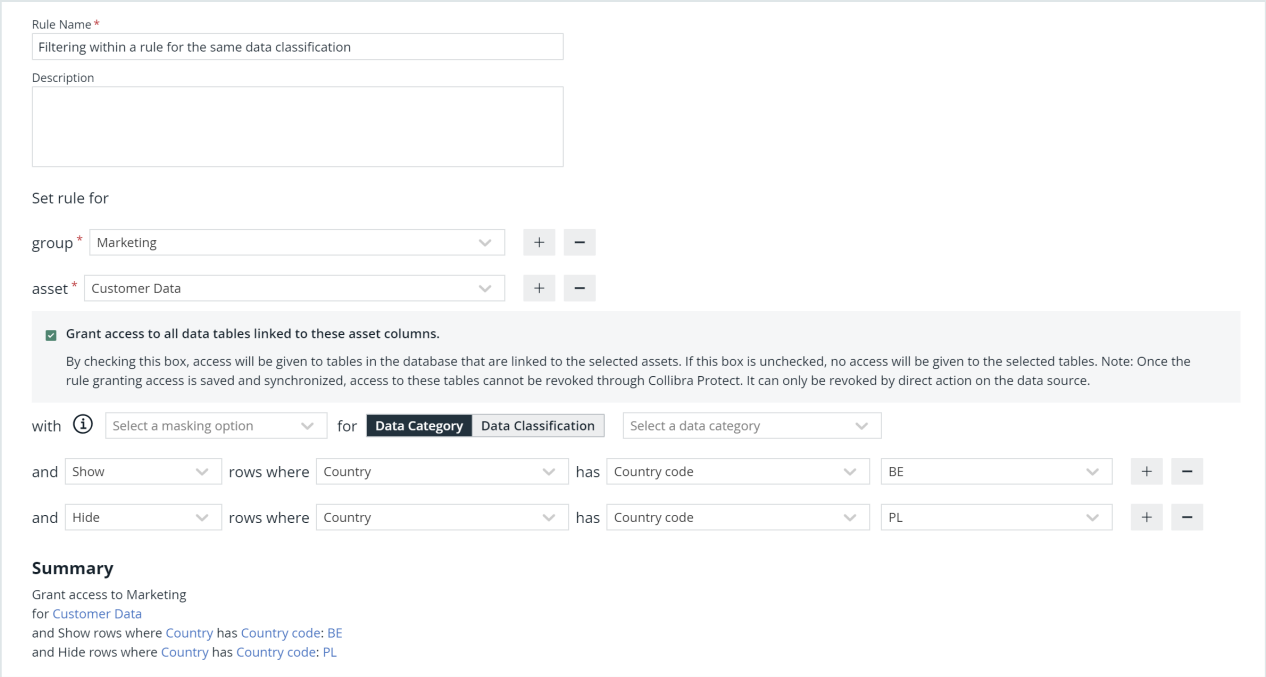
Filtering within a rule for the same data classification

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.

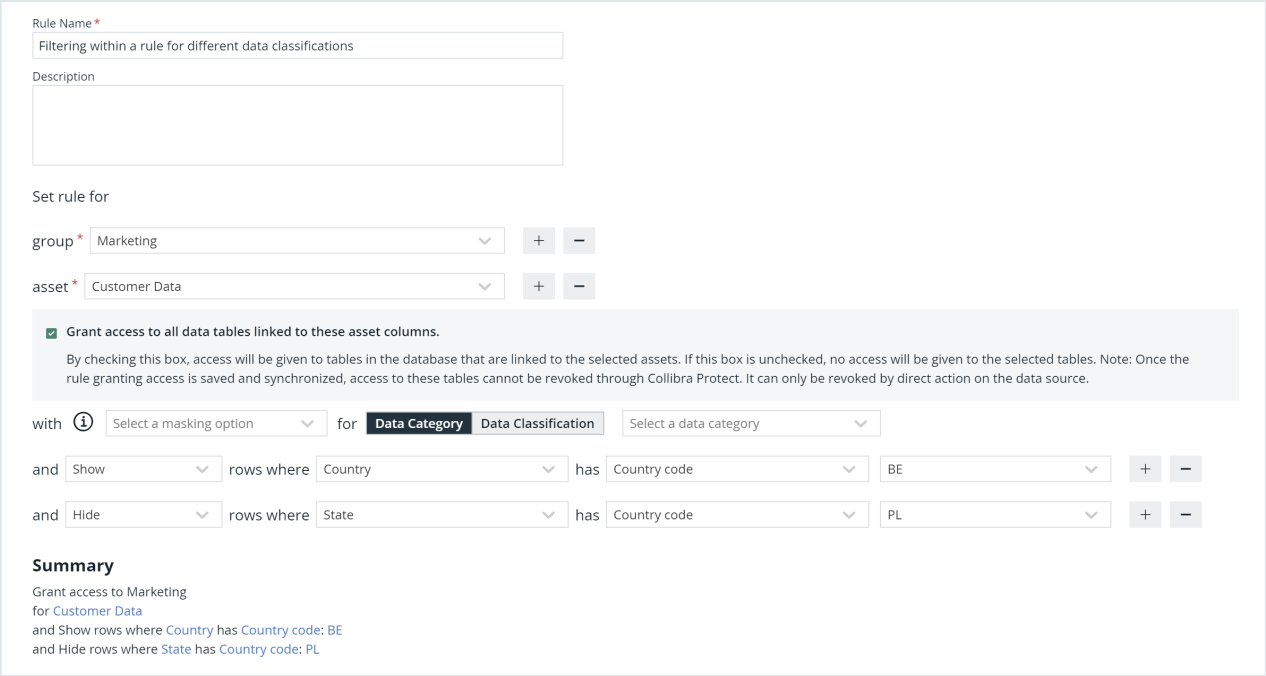
Filtering within a rule for different data classifications

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.

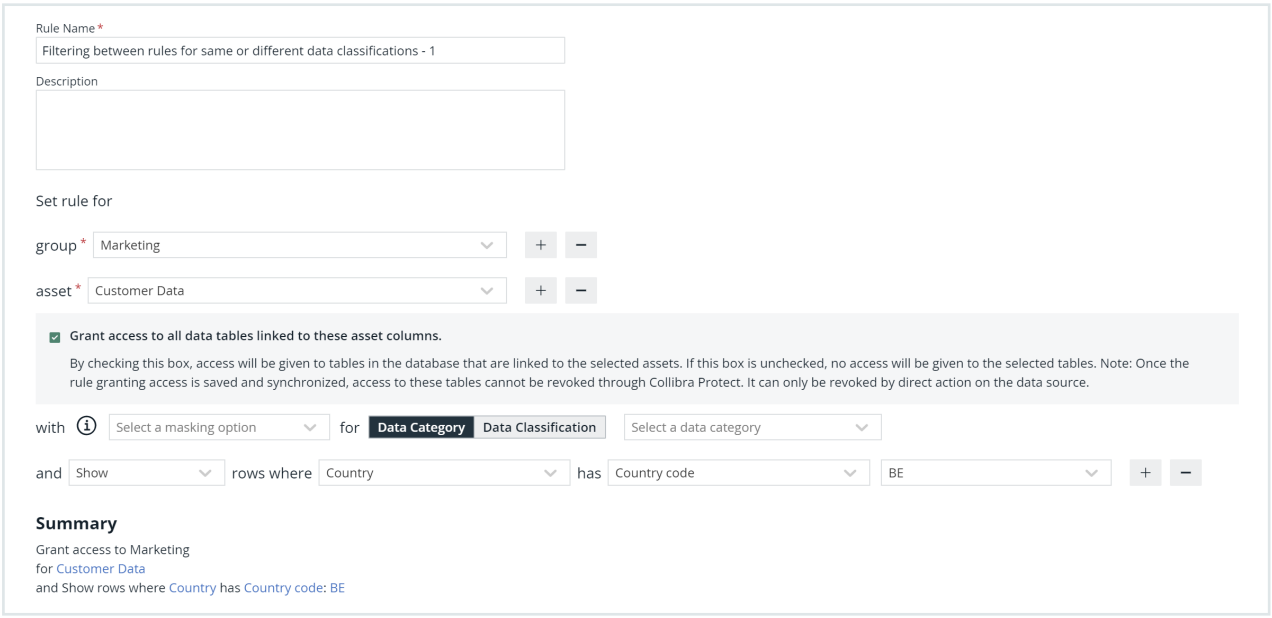
If any of the tables in the asset contain columns that are classified as **State**, the second filter hides only the rows that contain **PL** in those columns.

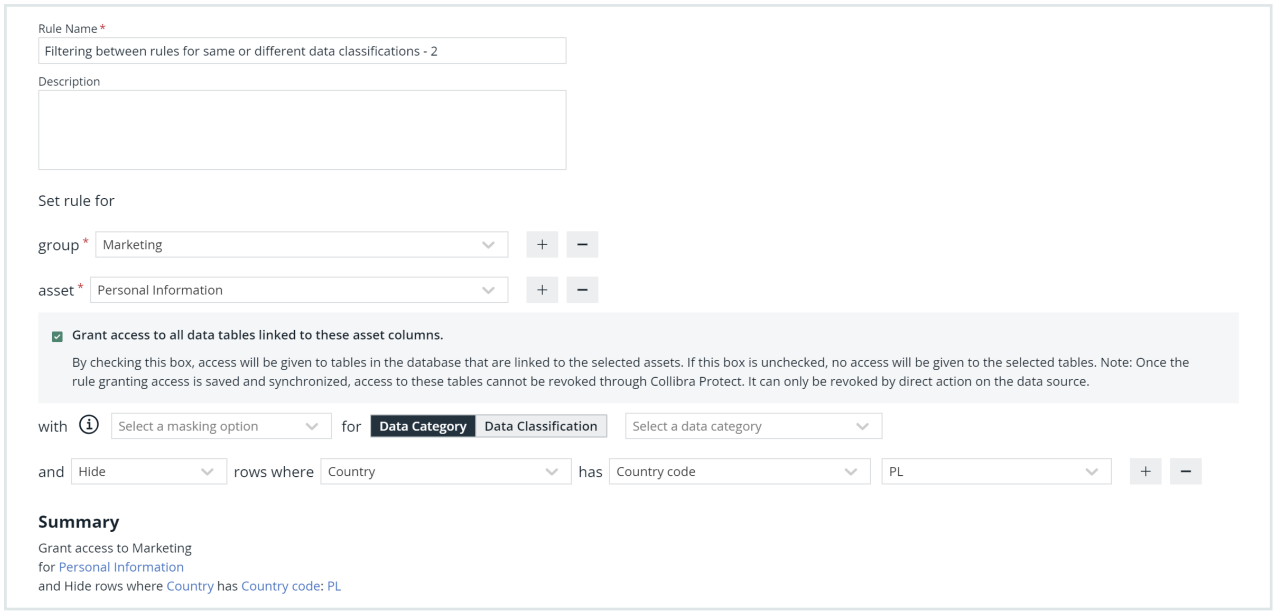
Suppose that a column is classified as both **Country** and **State**. That is, data classifications **Country** and **State** share the same column. Then, this rule will fail because two conflicting filters affect the same column.

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