

author: id: collibra summary: This is a sample Snowflake Guide to show how a user can work with Collibra in integration with Snowflake. categories: Getting-Started environments: web status: Published feedback link: <https://github.com/Snowflake-Labs/sfguides/issues> tags: Getting Started, Data Science, Data Engineering, Twitter

Collibra Data Governance with Snowflake

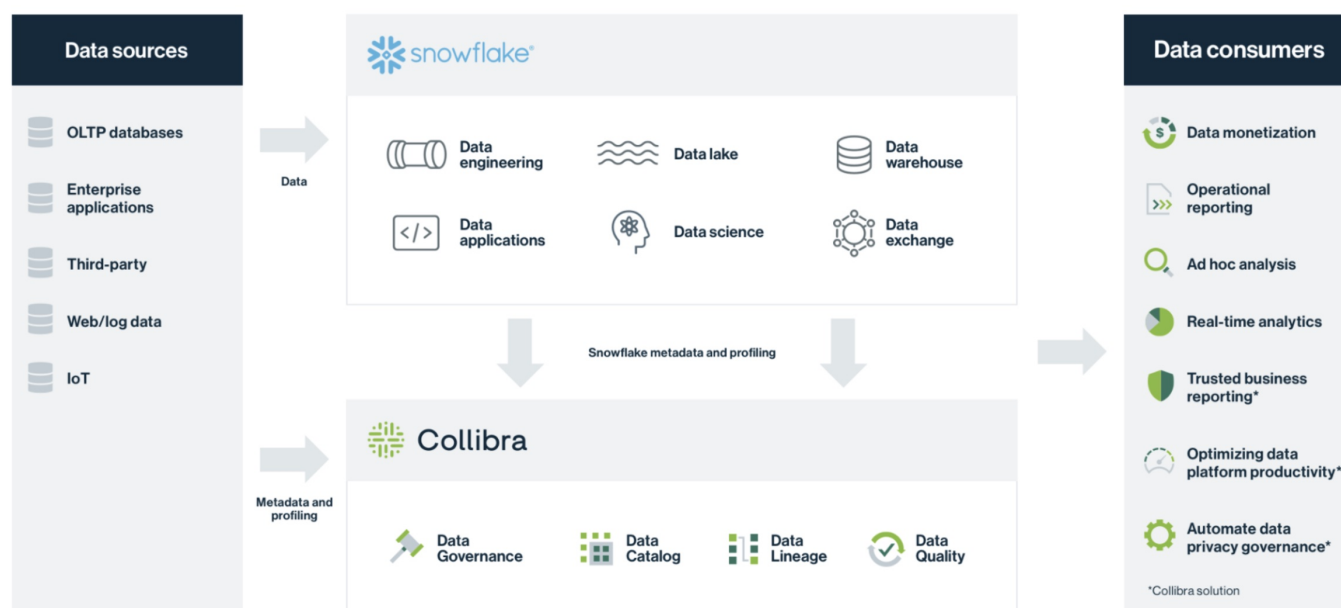
Overview

Duration: 1

This is a guide on how to catalog Snowflake data into Collibra, link the data to business and logical context, create and enforce policies. Also we will show how a user can search and find data in Collibra, request access and go directly to the data in Snowflake with access policies automatically enforced.

Collibra can enable controlled and secure data access across Snowflake databases from your analytic, machine learning and ETL services. Collibra and Snowflake enforces data access policies at run-time, so each user will only see the data they are authorized to view. Collibra integrates to Snowflake using a JDBC driver or API, providing the ability to automatically ingest technical metadata and lineage, as well as perform profiling and classification.

Collibra overall functionality in integration with Snowflake can be summarized with the following diagram:



Prerequisites

- Access to Snowflake database
- Access to a Collibra instance. Please reach out to:

What You'll Learn

- how to catalog Snowflake data in Collibra
- how to connect the metadata into a data domain and business and logical data attributes

- how to enforce data privacy policies on the Snowflake data
- how to search for data in Collibra
- how to request access of Snowflake data in Collibra and view data with policies enforced

What You'll Need

- A Snowflake instance
- A Collibra instance

What You'll Build

- Metadata of Snowflake, A Policy for this data, diagram that connects the data to its business context.

Metadata Configuration

Duration: 2

It is important to set the correct metadata for your Snowflake Guide. The metadata contains all the information required for listing and publishing your guide and includes the following:

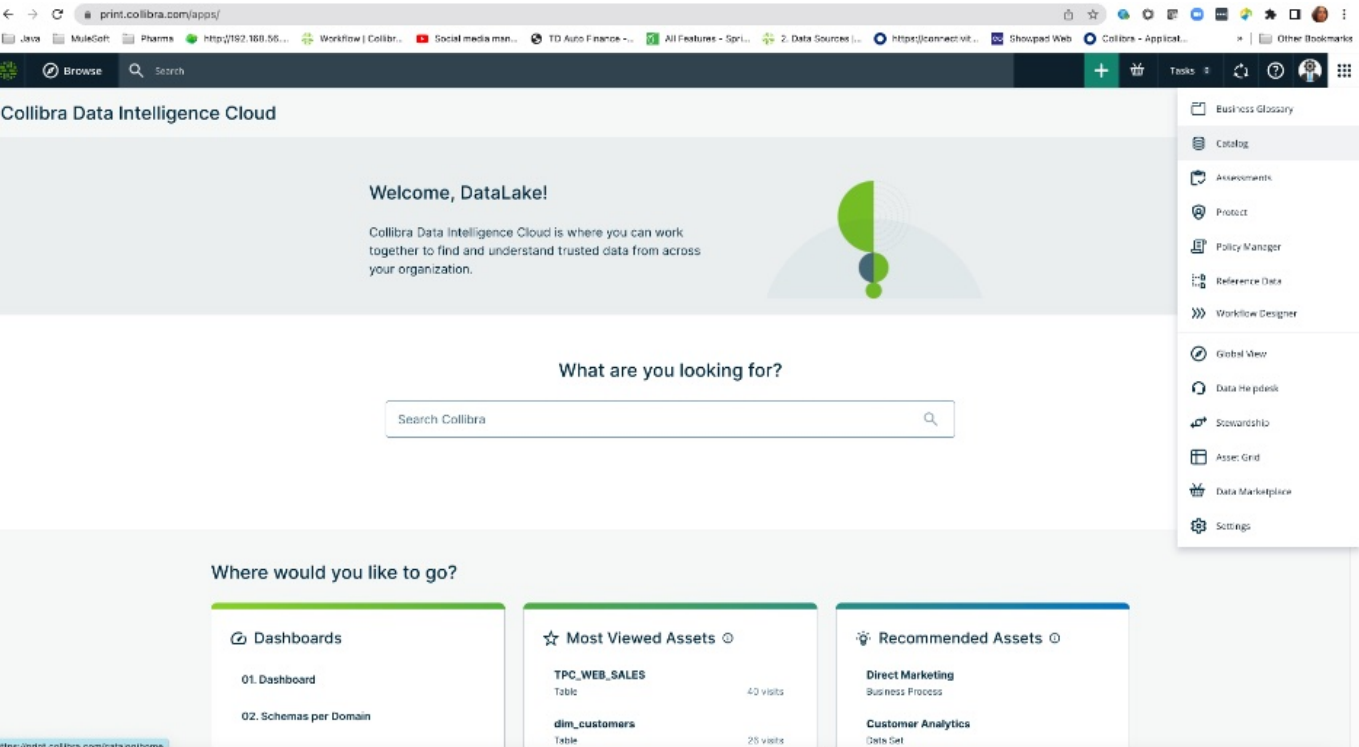
- **summary:**
 - This is a guide on how to Catalog and use Snowflake data using Collibra.
 - **categories:**
 - data governance
 - data intelligence
 - data mesh
 - data privacy
 - data science
 - **environments:**
 - **web** . All work is done on the web, and there is no need to install anything locally on your computer.
 - **status:** Draft
 - **feedback link:** <https://github.com/Snowflake-Labs/sfguides/issues>
 - **tags:** Getting Started, Data Science, Twitter, Data Governance, Data Intelligence, Data Privacy, Data Mesh
 - **authors:**
 - Vasiliki Nikolopoulou, Collibra
 - Tomer Steinberg, Snowflake
-

STEP 1: Catalog Snowflake into Collibra

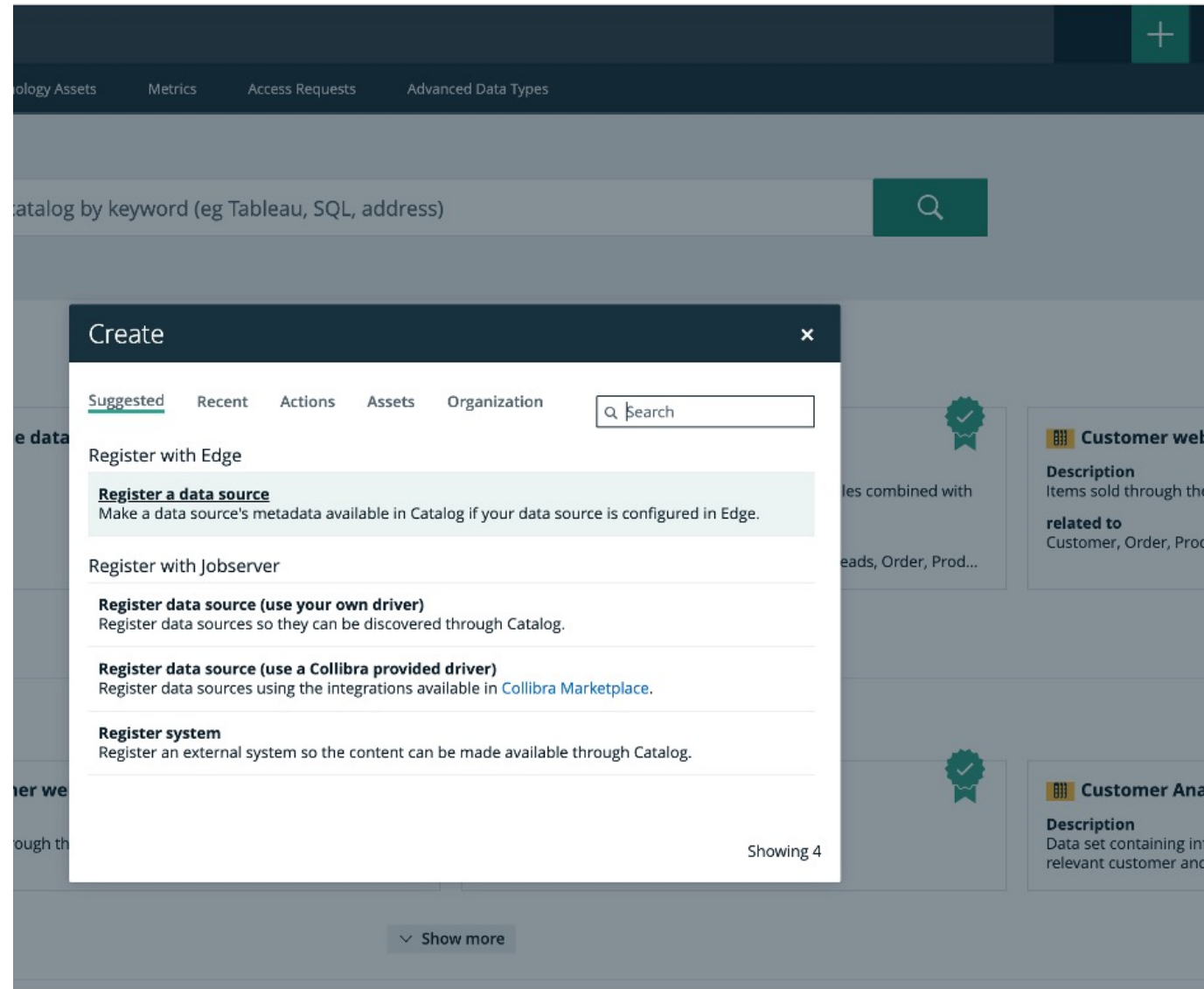
Duration: 2

In this step we will import into Collibra the technical metadata from a Snowflake database. We will use Collibra Edge, which is an Integration Hub residing within the customer firewall for safety. This integration hub retrieves the Snowflake metadata using native JDBC drivers.

Go to Collibra and click on Catalog to start the ingestion:



Click on the plus sign to choose the database:



Choose register through Edge and choose the Snowflake option (first option)

Add Database

Edge data source

D_CUSTOMER_snowflake

Community *

Snowflake

System *

Snowflake Software

Database name *

D_CUSTOMER

Description

Owner *

DataLake Admin

Cancel

Register

Go to the Snowflake Community and look for the D_CUSTOMER database:

Browse

Search

Business Analysts community

Snowflake

Type: Community

Overview

Organization

Comments

Assets

Responsibilities

History

Files

Description

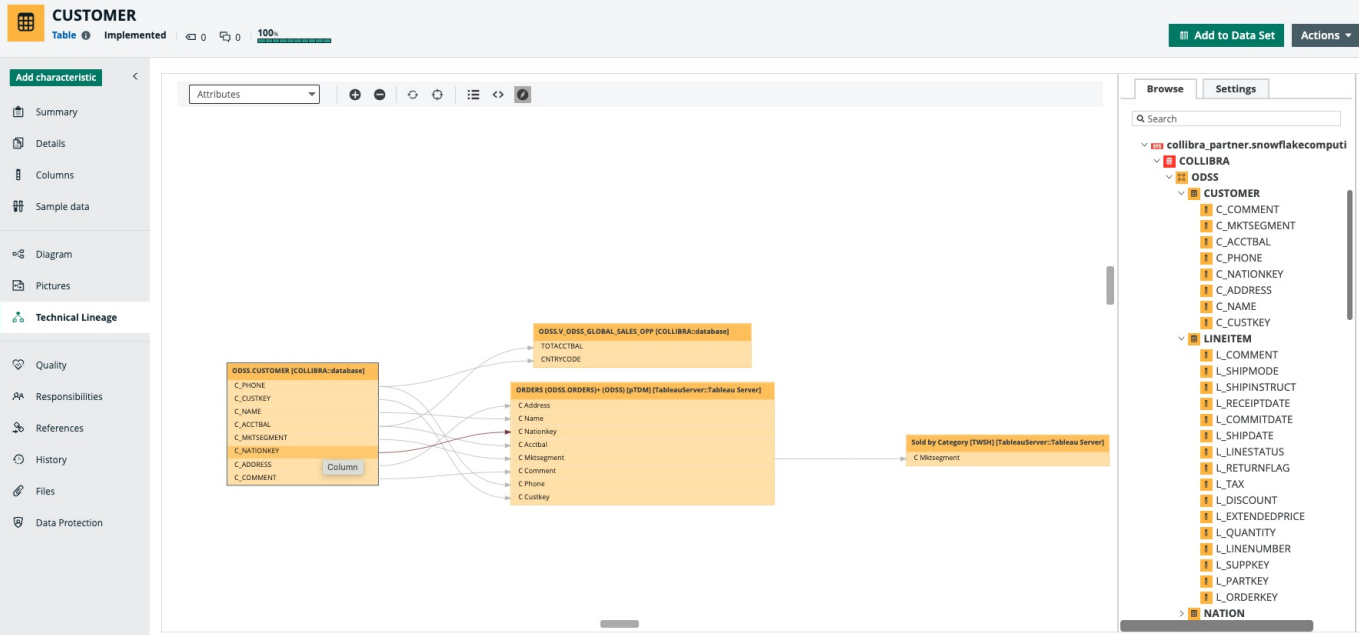
No value has been given yet. Double click or use the edit button.

Organization

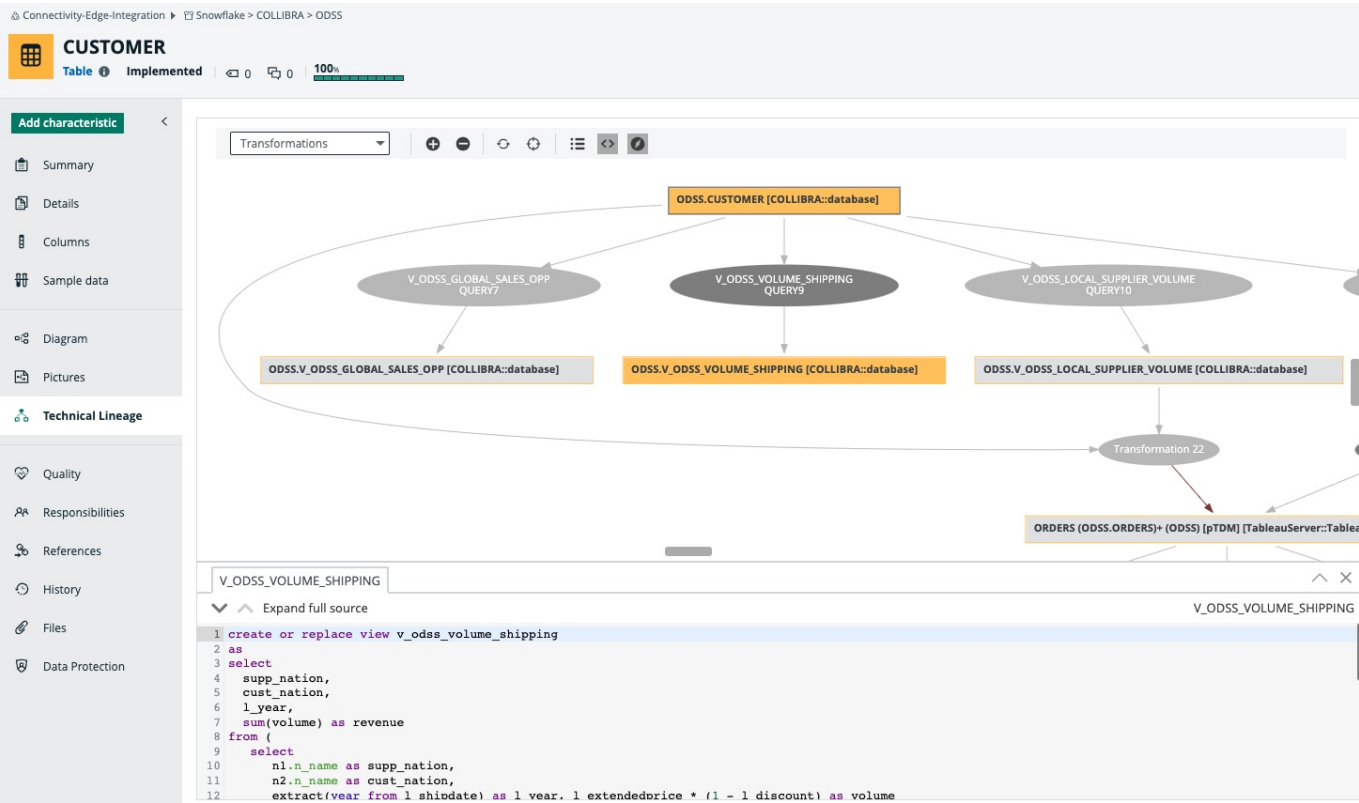
Name	Description	Domain Type	Owner	Stakeholder
Snowflake				
---- cltibike		Technology Asset Domain		vasiliki nikolopoulou
---- collibra		Physical Data Dictionary	vasiliki nikolopoulou	...
---- collibra > COLLIBRA > TPC_DS		Physical Data Dictionary		
---- collibra > COLLIBRA > TPCH_SF001		Physical Data Dictionary		
---- collibra > D_CUSTOMER > PUBLIC		Physical Data Dictionary		
---- collibra > WEATHER > PUBLIC		Physical Data Dictionary		
---- D_CUSTOMER_snowflake		Physical Data Dictionary	DataLake Admin	
---- D_CUSTOMER_snowflake > WEATHER > P...		Physical Data Dictionary		

Collibra, in addition to technical metadata it can also import technical lineage directly from Snowflake into Collibra using either SQL or API calls.

Technical lineage in Collibra looks for Snowflake data includes the SQL transformations. Here is an example of technical lineage of Snowflake data in Collibra:



Users can see the actual SQL transformations if needed:



Collibra also allows users to see profiling information and sampling (anonymized) if needed. Here is an example of profiling information:

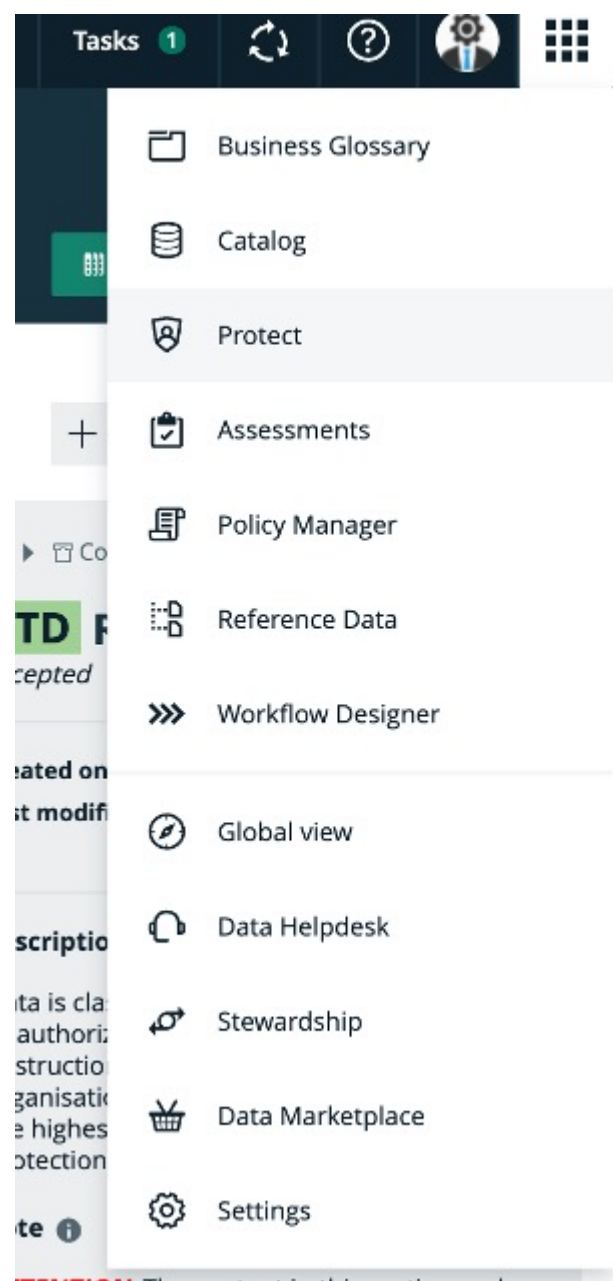


STEP 2: Data Privacy and Policy Enforcement

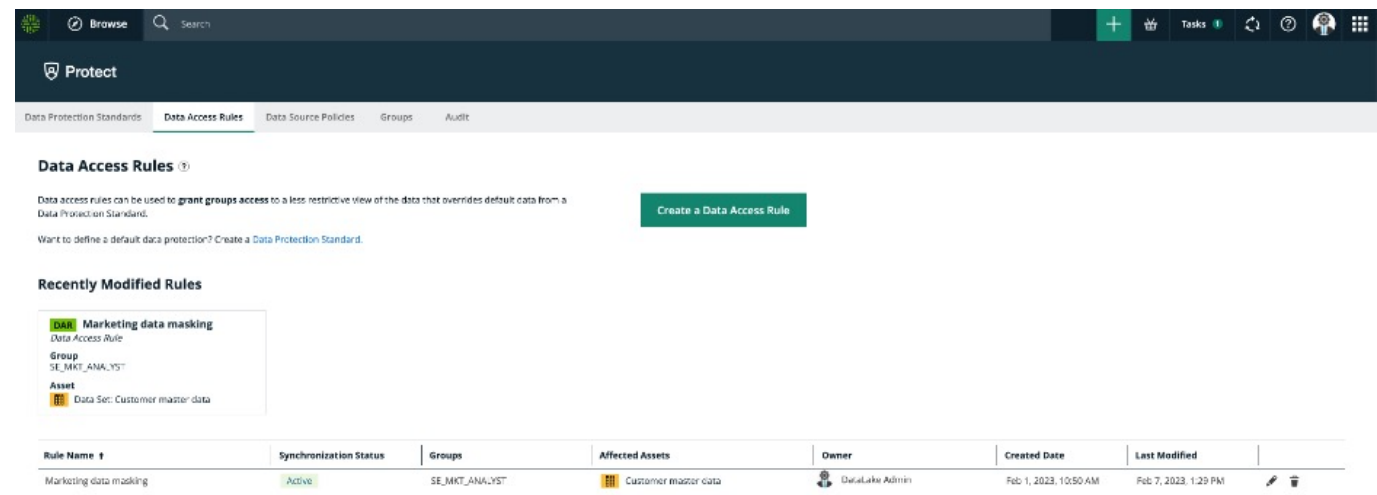
Duration: 2

In this stage, we will specifically define the Data Privacy policies and rules that will be automatically enforced in Snowflake directly from within Collibra.

Please click on the Collibra Protect option after clicking the nine dots all the way to the top right of the Collibra User Interface:



This will take you to the Collibra Protect environment, then click on the Data Access Rules. And then click on Create a Data Access Rule button:



A form opens where you can create the new rule.

Type the name of the rule, its description. This is a rule about Marketing people. All people within Marketing will have access to this customer dataset, however we will enforce the hiding or masking of the Last Name column and we will only allow users to see customers from the country Belgium. We will choose the user group BE_MKT_ANALYST, in which will belong all marketing analysts in Belgium. You can fill in as follows:

Create a Data Access Rule

Use data access rules to grant groups different access to the same data in data sets, business processes, or identified by data categories. You can mask or hide columns by their data category and you can also conditionally filter rows based on code set values.

Rule Name *

Belgian Marketing Analysts can see Belgian customers with Last Name hidden

Description

Country filtering for Belgium, hidden Last name.

Set rule for

group *

Belgian Marketing Analyst

+ -

asset *

Customer master data

+ -

☒ Grant access to the data linked to these assets.

By checking this box, additional access is given to the data tables or columns linked with the selected assets. If this box is unchecked, no access is given to the selected assets, but they can still be protected. **Note: once the rule granting access is saved and synchronized, access to these assets cannot be revoked through Collibra Protect. It can only be revoked by direct action on the data source.**

with ⓘ

Default masking

for

Data Category

Data Classification

Last name

+ -

and

Show

rows where

Country

has

Country Codes

BE

+ -

Summary

Grant access to Belgian Marketing Analyst

for Customer master data

with Default masking for Last name

and Show rows where Country has Country Codes: BE

Generate Preview

Cancel

Save Rule

At this point you can click on "Generate Preview" to see which columns will be viewed and which will be filtered and masked:

Create a Data Access Rule

Use data access rules to grant groups different access to the same data in data sets, business processes, or identified by data categories. You can mask or hide columns by their data category and you can also conditionally filter rows based on code set values.

Rule Name *

Belgian Marketing Analysts can see Belgian customers with Last Name hidden

Description

Country filtering for Belgium, hidden Last name.

Set rule for

group *

Belgian Marketing Analyst

+

-

asset *

Customer master data

+

-

☒ Grant access to the data linked to these assets.

By checking this box, additional access is given to the data tables or columns linked with the selected assets. If this box is unchecked, no access is given to the selected assets, but they can still be protected. **Note: once the rule granting access is saved and synchronized, access to these assets cannot be revoked through Collibra Protect. It can only be revoked by direct action on the data source.**

with ⓘ

Default masking

for

Data Category

Data Classification

Last name

+

-

and

Show

rows where

Country

has

Country Codes

BE

+

-

Summary

Grant access to Belgian Marketing Analyst
for [Customer master data](#)
with Default masking for [Last name](#)
and Show rows where [Country](#) has [Country Codes: BE](#)

Generate Preview

Cancel

Save Rule

Summary
Grant access to Belgian Marketing Analyst
for [Customer master data](#)
with Default masking for [Last name](#)
and Show rows where [Country](#) has [Country Codes: BE](#)

Generate Preview

Customer master data

Column ↑	Access	Masking Agent	Masking	Code Value
C_CUSTOMER_ADDRESS	✓ Full View			
C_FIRST_NAME	✓ Full View			
C_FIRST_SALES_DATE_SK	✓ Full View			
C_FIRST_SHIPTO_DATE_SK	✓ Full View			
C_LAST_NAME	⚠ Masked	Last name	0	
C_LAST_REVIEW_DATE	✓ Full View			
C_LOGIN	✓ Full View			
C_PREFERRED_CUST_FLAG	✓ Full View			
C_SALUTATION	✓ Full View			

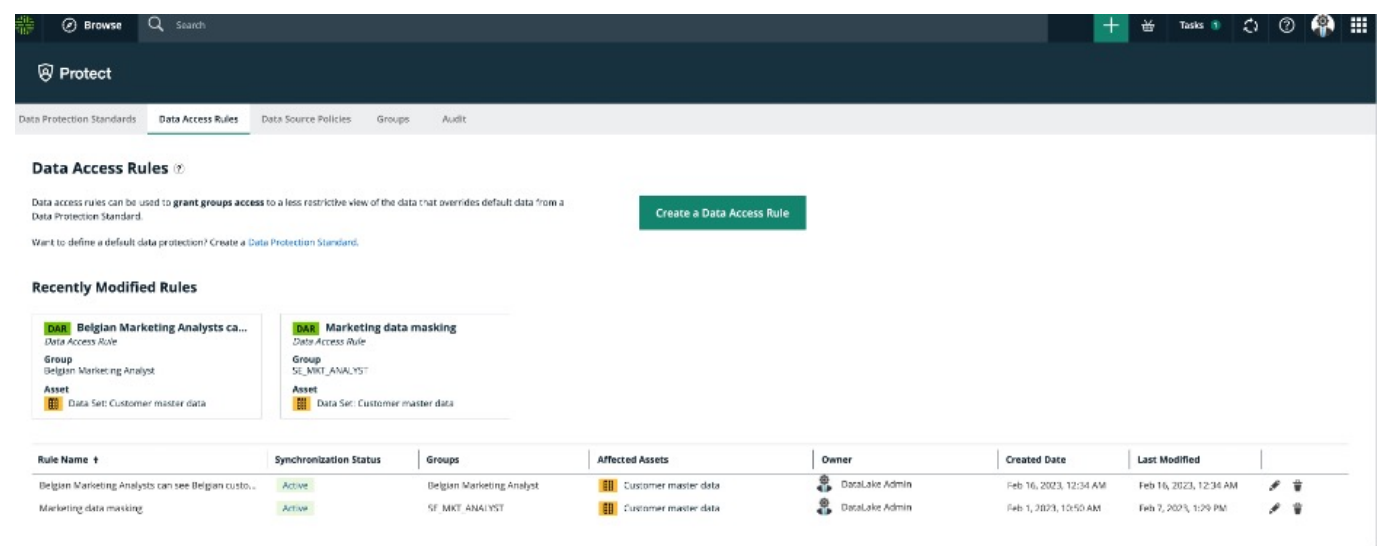
Cancel

Save Rule

Click on Save Rule for the rule to be save. At this point the creation of this rule is sent to Collibra and the synchronization via API to the Snowflake environment is in Pending state. The synchronization happens automatically every 60 minutes.

If you need to expedite the synchronization of the new rules, you can trigger them via a POST command as follows from an API posting software:

This synchronization is triggered and within less than a minute we see that the rule is no longer in pending state but it is in effect:



We can simply check this if we login to Snowflake and go to D_CUSTOMER database, choose the PUBLIC schema and open the CUSTOMER table. Go to Preview and choose the BE_MKT_WH and also choose the role BE_MKT_ANALYST. Then you should be able to see the columns except for Last Name which is masked while all the rows are filtered for country Belgium.

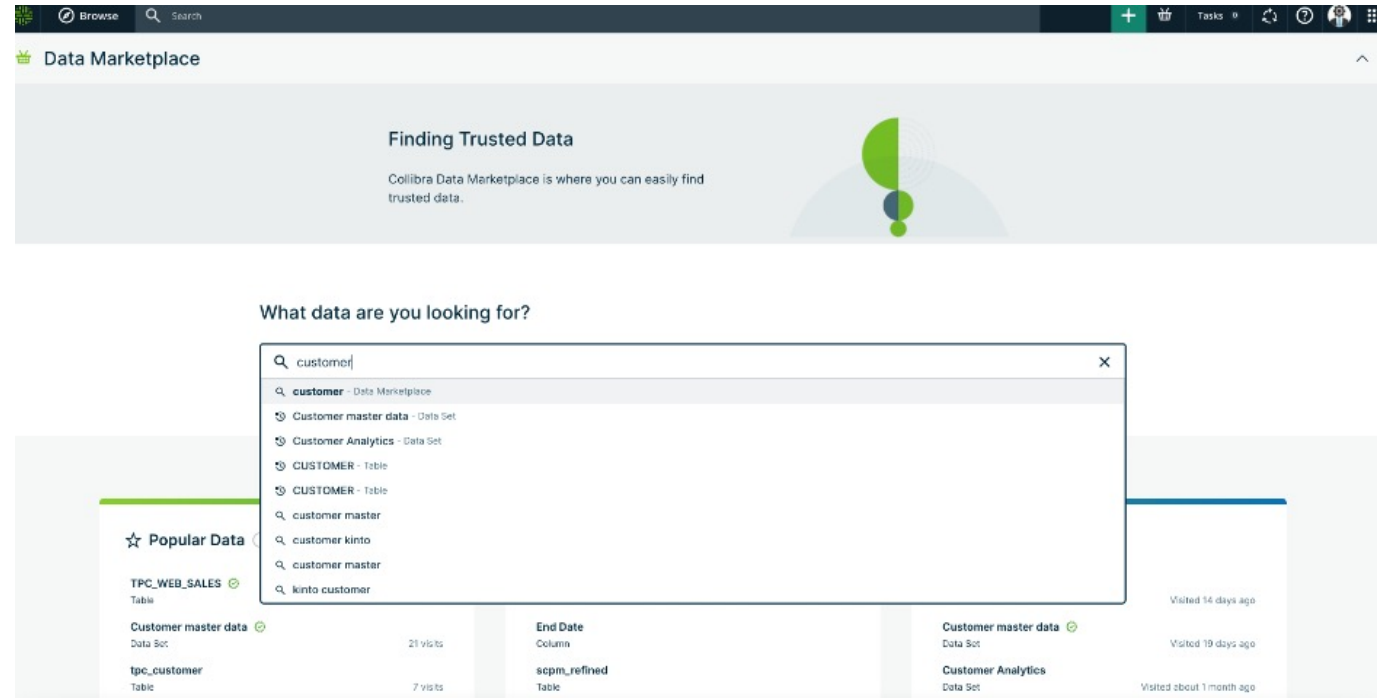
STEP 3: Data Consumption and Access

Duration: 2

In this chapter we will look into how a user can search and find trusted data to use and request access to the data while at the same time policies will be automatically enforced.

David, the marketing analyst from Sweden is looking for customer data that is trusted and contain email and first name and birthdate at the minimum.

In Colibra Data Marketplace the user is searching for customer data:



Filters the search by pressing the Data Sets tab and the Certified option in the filters to the left.

Q customer

X

Filters

AllData SetsReportsTablesMore

4 resultsRelevance

SELECTED (3)Clear All

Certified XData Set XBI Data Set X

DATA DOMAIN

☐ Customer (3)☐ Order (1)☐ Product (1)

DATA CONCEPT

Find Data Concept

☐ Country (3)☐ Customer Identifier (2)☐ Email Address (2)☐ First Name (2)☐ Last Name (2)☐ Address (1)

Show more

CERTIFIED

☒ Certified (4)

BUSINESS ASSET

Customer Analytics Anonymized

Data Set

Data set containing information on our internet sales combined with relevant **customer** and product data.

Status: AcceptedCertified

One or more search terms appear in Properties, Data Concept and more.

Customer master data

Data Set

Customer master data. It contains names, (PII) emails and other contact details, date of birth. Contains data from the **Customer** Store.

Status: AcceptedCertified

One or more search terms appear in Properties, Data Concept and Data Domain.

CUSTOMER WEB SALES

Data Set

Items sold through the web sales channel

Status: AcceptedCertified

One or more search terms appear in Data Domain.

The second dataset is the one that seems proper since in the description it states that it contains names, birthdates and emails.

customer

Customer master data

Status: Accepted

Certified

Share

Overview

Columns (18)

Details

Responsibilities

Actions

Description

Customer master data. It contains names, (PII) emails and other contact details, date of birth. Contains data from the Customer Store.

Owner

vasiliki nikolopoulou

See all (5)

Business Context

Customer

Source Lineage

Database: D_CUSTOMER / Schema: PUBLIC / Table: CUSTOMER

Go to Asset Details

Add to Basket

Status: Accepted

Certified

Clicking on the columns we confirm this is the right set:

customer

Customer master data

Data Set

Status: Accepted

Certified

Share

Overview

Columns (18)

Details

Responsibilities

Actions

Search

Data type

Classification

Name	Data type	Distinct values	Empty values	Classification	Table	DB
C_FIRST_NAME	VARCHAR	7887750	4%	First name	CUSTOMER	D_C
C_BIRTH_MONTH	NUMBER	12	4%		CUSTOMER	D_C
C_CUSTOMER_SK	NUMBER	64357810	0%		CUSTOMER	D_C
C_BIRTH_COUNTRY	VARCHAR	211	3%	Country	CUSTOMER	D_C
C_FIRST_SALES_DATE_SK	NUMBER	11817000	4%		CUSTOMER	D_C
C_LOGIN	VARCHAR	0	100%	user id	CUSTOMER	D_C
C_BIRTH_DAY	NUMBER	31	3%		CUSTOMER	D_C
C_EMAIL_ADDRESS	VARCHAR	62287379	4%	Email	CUSTOMER	D_C
C_LAST_NAME	VARCHAR	12701000	4%	Last name	CUSTOMER	D_C
C_SALUTATION	VARCHAR	6	4%	Title	CUSTOMER	D_C

Go to Asset Details

Add to Basket

Status: Accepted

Certified

Clicking on the Go to Asset Details we can discover more about this dataset. It belongs to the Customer Domain, we see the name of the Data Owner, the rating:

Customer master data

Data Set Accepted

5 stars (1)

Add to Data Basket

Actions

Add characteristic

Summary

Details

Data Elements

Sample data

Diagram

Pictures

Quality

Similar Data Sets

Responsibilities

References

History

Files

Description

Customer master data. It contains names, (PII) emails and other contact details, date of birth. Contains data from the Customer Store.

Business Context

Business Data Model's

DDOM Customer

Data Elements

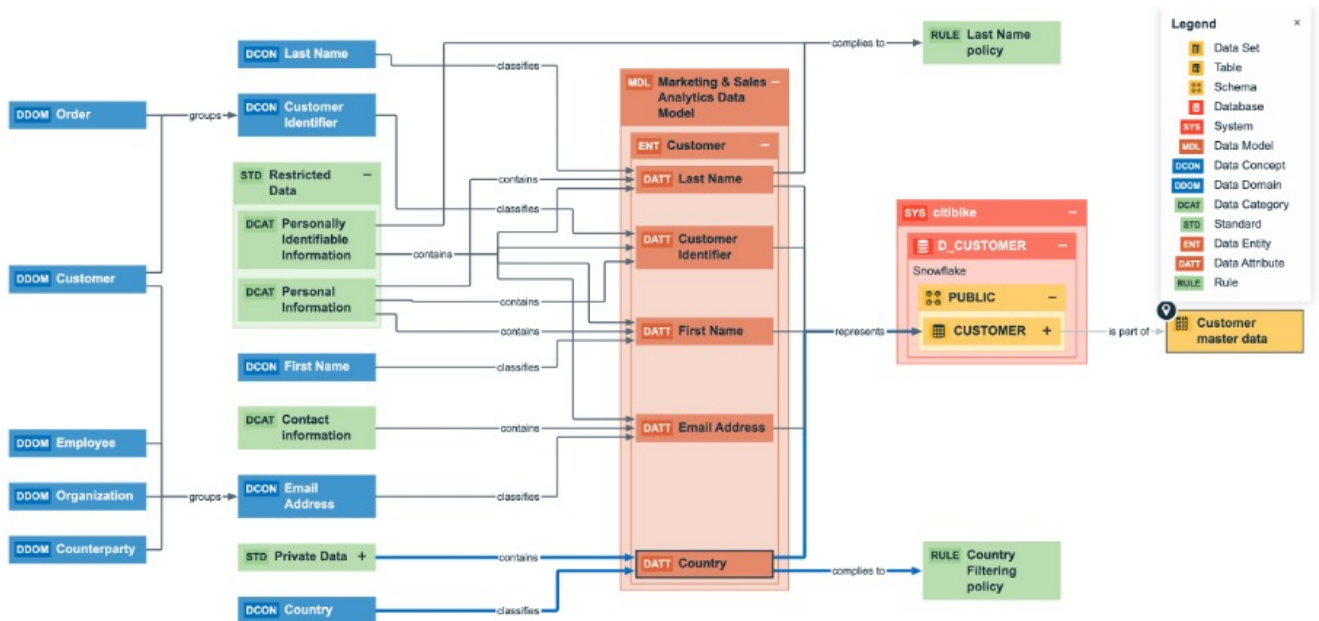
Name	Data Type	represented by
C_CUSTOMER_SK	Whole Number	
C_CUSTOMER_ID	Text	
C_CURRENT_CDEMO_SK	Whole Number	
C_CURRENT_HDEMO_SK	Whole Number	
C_CURRENT_ADDR_SK	Whole Number	

Owner

vasiliki nikropouleu

If the user clicks on the "Diagram" to the left it is possible to see the full lineage of the dataset, that includes the logical, the business, the data quality and data privacy context.

What we can see in this diagram, is the connection of the dataset to the columns and tables and systems it draws its data from. Then the connection of these technical metadata to the logical definitions and consequently to the business context. Also linked to the logical definitions we can see the data privacy rules.



Also you can click at the Data Protection tab all the way to the bottom left, and this will show the policies and restrictions associated with this dataset.

Browse

Search

Marketing community > Automotive Data Products > Data Products

Customer master data

Data Set

Accepted

(1)

0

1

Add characteristic

<

Summary

Details

Data Elements

Sample data

Diagram

Pictures

Quality

Similar Data Sets

Responsibilities

References

History

Files

Data Protection

Data Protection

Data Protection policies provide protection to sensitive data by hiding or masking the data for a specified group of people.
[Go to Collibra Protect to create and manage data protection policies.](#)

Q Search a policy by name

Filter by data protection group

Data Access Rules

Data access rules assign protection to columns. They grant groups access to a less restrictive view of data.

Belgian Marketing Analysts can see Belgian customers with Last Name hidden

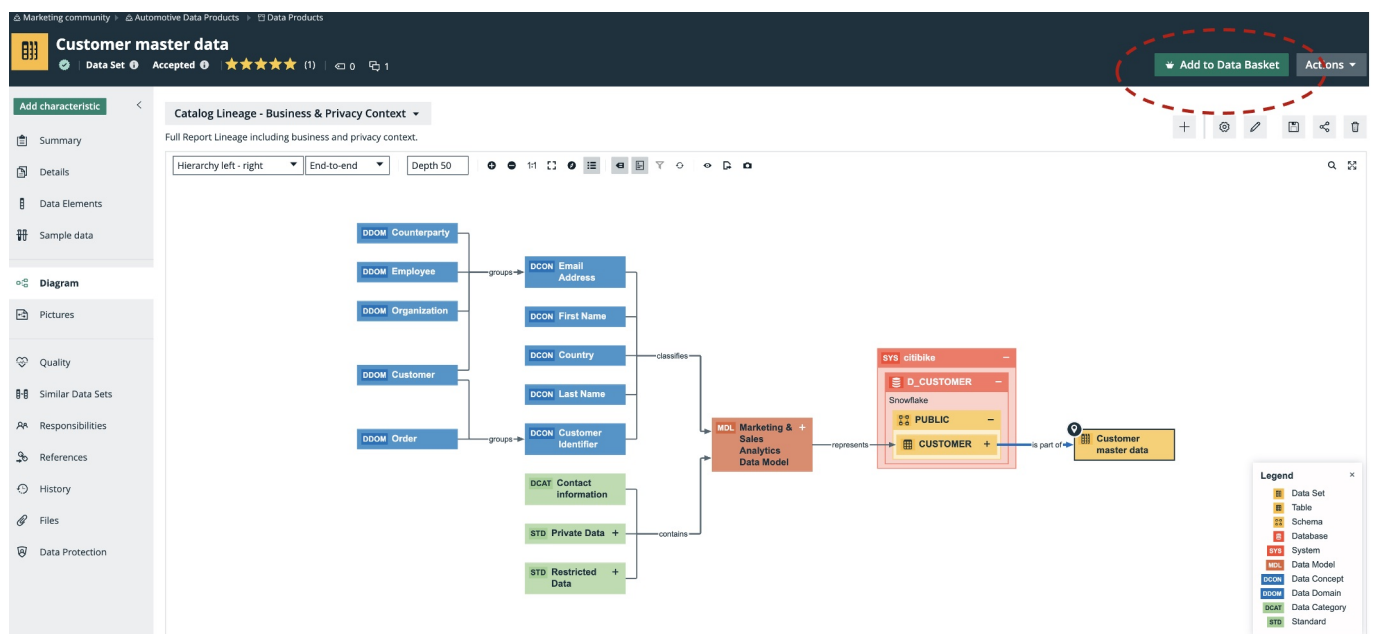
Grant access to Belgian Marketing Analyst
for [Customer master data](#)
with Default masking for [Last name](#)
and Show rows where [Country](#) has [Country Codes: BE](#)

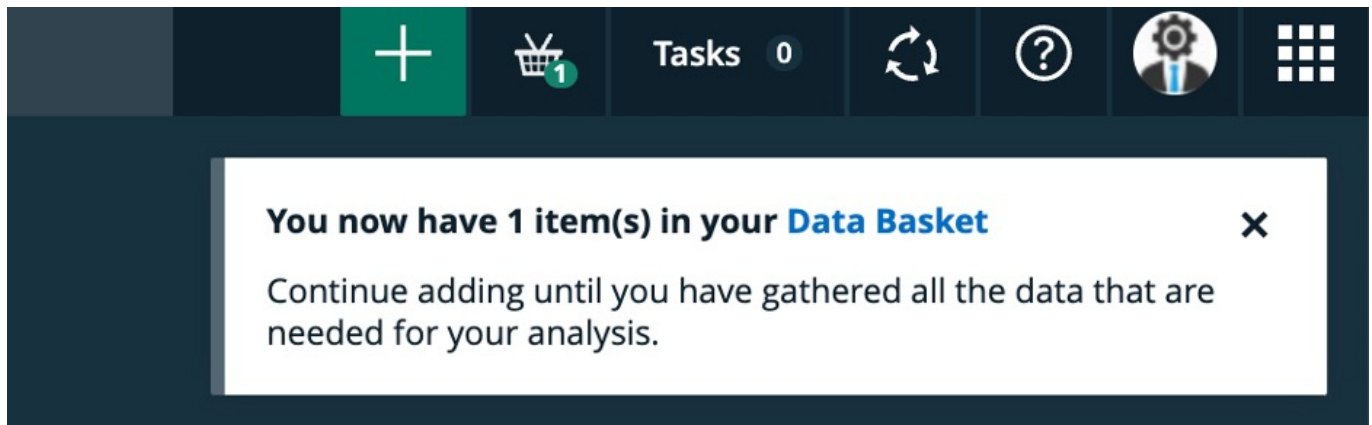
Marketing data masking

Grant access to SE_MKT_ANALYST
for [Customer master data](#)
with Default masking for [Last name](#)
and Show rows where [Country](#) has [Country Name: SWEDEN](#)

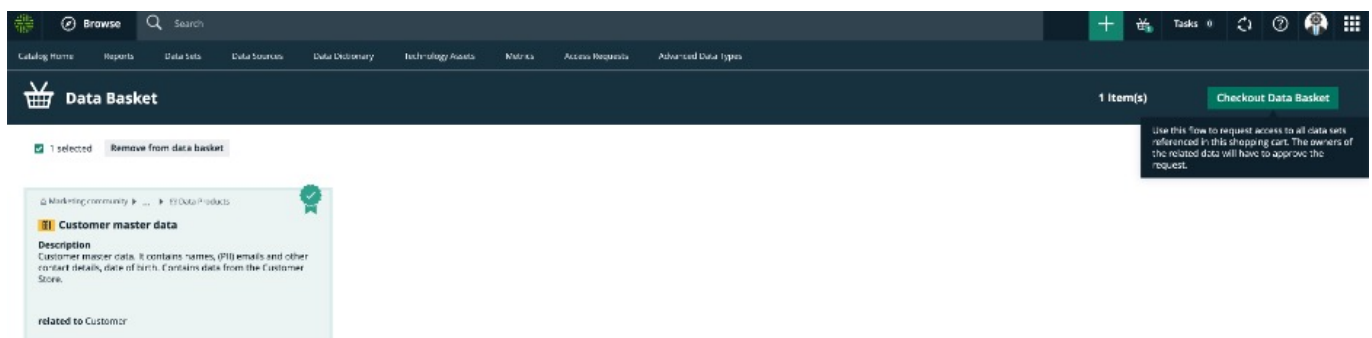
As you can see Collibra allows you to see the lineage of a dataset in relation to its technical contents and in extension to its business and data privacy context. For example we can see that the columns that represent Last Name and Country each has a policy.

The user can add the dataset into the Data Basket by pressing the "Add to Data Basket" button:





This brings the dataset to the user's personal Data Basket:



By clicking at the Data Basket icon we go to the area where we can start requesting the dataset(s).

The user can then choose the dataset and click on the "Checkout Data Basket" in order to start requesting access to the dataset.

Requests Data Sets Access

Paragraph

Campaign for marketing emails on customer birthdays.

I need this data by... *

2/16/2023

and will use it until... *

3/12/2023

Purpose *

Activities related to customer management

By clicking Submit at the end of this form, a workflow starts which sends a notification to the dataset Data Owner, requesting access to the dataset:

Browse

Search

All Tasks

My Tasks

My Overdue Tasks

Select

Name	Description	Related Resource
Approve	Please accept or reject this data usage request and add reasoning why you did so.	2023-02-16 #1

Due: 3/15/2023

Approve

Please accept or reject this data usage request and add reasoning why you did so.

Related to

DU 2023-02-16 #1

Reasoning

Paragraph

Please provide some more info about why you approved/rejected this data access request.

Reject

Approve

The Data Owner will see the request and will evaluate if this user can get access to the dataset.

The Data Owner will then go to the dataset asset to check about policies and groups associated with this dataset. This will take us to the Collibra Protect rule that the Data Owner can check in detail:

Data Access Rules ?

Data access rules can be used to **grant groups access** to a less restrictive view of the data that overrides default data from a Data Protection Standard.

Want to define a default data protection? Create a [Data Protection Standard](#).

Create a Data Access Rule

Recently Modified Rules

DAR

Marketing data masking

Data Access Rule

Group

Rule information


Marketing data masking

Summary

Grant access to SE_MKT_ANALYST for [Customer master data](#) with Default masking for Last name and Show rows where [Country](#) has [Country Name: SWEDEN](#)

Copy to Clipboard

Edit

Groups	Affected Assets
SE_MKT_ANALYST	 Customer master data

In this case we see that this rule is about granting access to the dataset Customer master data to all people who work in the Marketing department, while there is default masking policy for the Last Name and the only customer rows shown are filtered by country, in this case Sweden.

Therefore the Data Owner understands that the user should be granted the SE_MKT_ANALYST role in Snowflake in order to get access to this dataset.

The Data Owner sends an approval notice to the requester, asking the requester to first accept the terms and conditions of accessing this dataset:

20 / 23

Due: 3/15/2023

Approve

Please accept or reject this data usage request and add reasoning why you did so.

Related to

DU

2023-02-16 #1

Reasoning

Paragraph

B

I

U

A

...

Please accept terms and conditions and your access will then be granted.

Please provide some more info about why you approve/reject this data access request.

Reject

Approve

The requester will then receive the task to accept the terms and conditions in order for access to be granted. The user checks the "All of the above" radio button and clicks on the Accept Notice button to proceed.

At this point the Data Owner receives notification that the requester accepted the conditions and goes to Snowflake to grant to the requester the role "SE_MKT_ANALYST".

Go to Snowflake in the area Users and Roles and Search for the user to which we are going to grant access. Then click on the three dots to the right of the user and choose the Grant Role option in order to grant the Role to the User.

VN

Vasiliki Nikolop...

ACCOUNTADMIN

Users

Roles

+ User

1 of 77 Users

NAME ↑	DISPLAY NAME	STATUS	LAST LOGIN	MFA	OWNER
DAVID_SE	DAVID_SE	Enabled	2 months ago	No	ACCOUNTADMIN

Edit

Reset Password

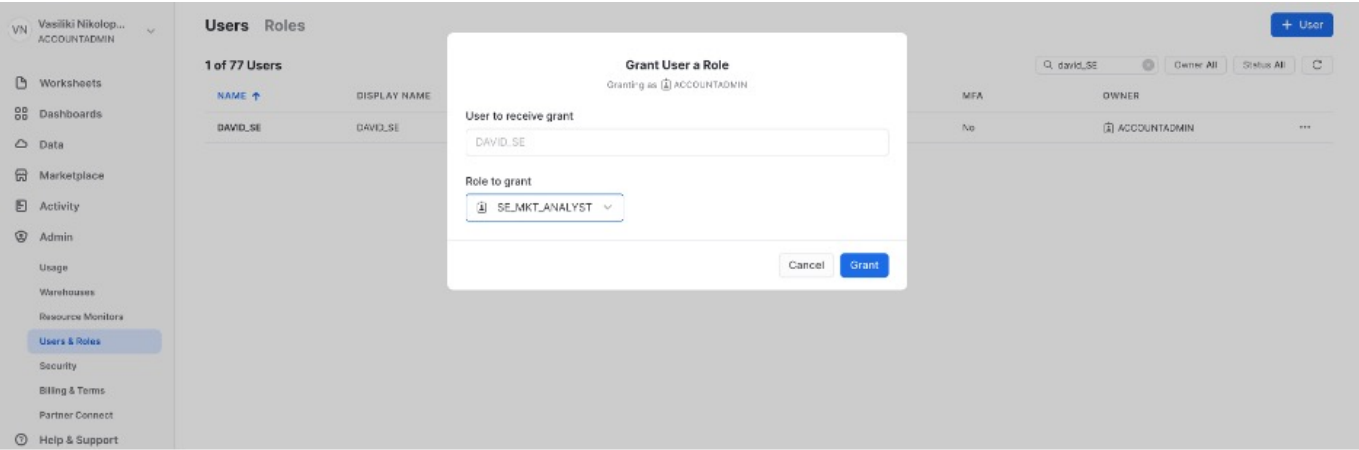
Disable User

Drop

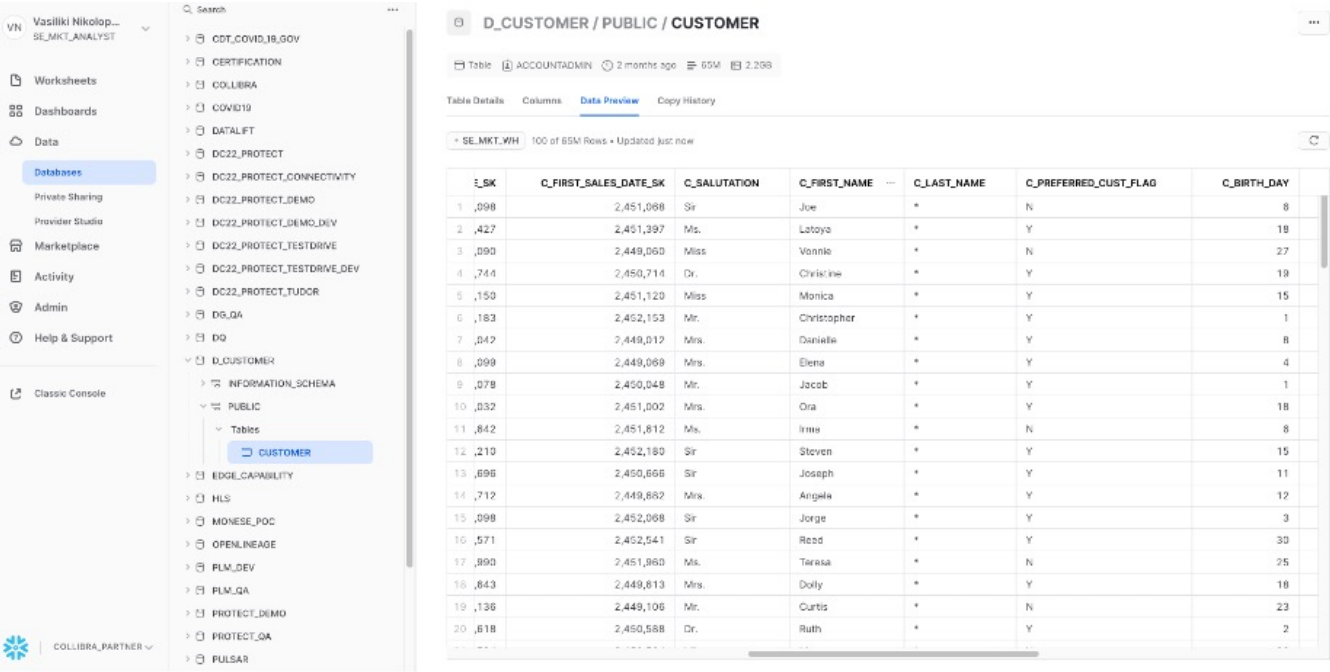
Grant a Role

Revoke a Role

At this point look for the Role in Collibra which gives access with enforced policies to a user, in this case the SE_MKT_ANALYST.



Now the user David will be able to login to Snowflake and access the table D_CUSTOMER with his role SE_MKT_ANALYST, and see the data with the enforced policies. First look for "Data", then choose D_CUSTOMER, then choose the schema "Public" then choose the warehouse SE_MKT_WH, then Preview Data. As you can see the Last name is hidden:



And also the country is filtered to include customers from Sweden only:

D_CUSTOMER / PUBLIC / CUSTOMER

Table
 ACCOUNTADMIN
 2 months ago
 65M
 2.2GB

Table Details
 Columns
 Data Preview
 Copy History

SE_MKT_WH
 100 of 65M Rows • Updated 3 minutes ago

	C_BIRTH_MONTH	C_BIRTH_YEAR	C_BIRTH_COUNTRY	C_LOGIN	C_EMAIL_ADDRESS	...	C_LAST_REVIEW_DATE
1	3	1,956	SWEDEN	null	Joe.Freeman@YOk.edu		2452363
2	8	1,988	SWEDEN	null	Latoya.Martinez@7QlZKeHdpqF.com		2452503
3	6	1,930	SWEDEN	null	Vonnie.Herrin@U3TP35I6zQ.edu		2452630
4	3	1,981	SWEDEN	null	Christine.Garcia@oq81HZHeYS.com		2452527
5	7	1,928	SWEDEN	null	Monica.Lane@leSQDfAZ2gz0b3MKtm.com		2452366
6	10	1,962	SWEDEN	null	Christopher.Hubbard@g6anlxh.edu		2452392
7	10	1,952	SWEDEN	null	Danielle.Martinez@DEvKkOOajBYV.org		2452301
8	7	1,951	SWEDEN	null	Elena.Rivas@i.edu		2452635
9	5	1,933	SWEDEN	null	Jacob.Bethel@Ad2F1Q3Na1LL2.org		2452644
10	7	1,938	SWEDEN	null	Ora.Walls@IsEDP125avl.com		2452432
11	11	1,988	SWEDEN	null	Irma.Nutter@YdkmpKITJP.com		2452586
12	11	1,984	SWEDEN	null	Steven.Brown@ymETJu9Pfztc.edu		2452505
13	5	1,974	SWEDEN	null	Joseph.Morin@OI7EXLJi1Y.com		2452519
14	2	1,936	SWEDEN	null	Angela.Wood@Bln3xmmp39jkPUO.edu		2452392
15	4	1,927	SWEDEN	null	Jorge.Weeks@2vMxyFBcollFH.edu		2452522
16	6	1,954	SWEDEN	null	Reed.Maloney@y1T.edu		2452560
17	12	1,963	SWEDEN	null	Teresa.Hill@4BZkk.edu		2452410
18	12	1,976	SWEDEN	null	Dolly.Shaw@LNNJBDypDm0Hqlav.org		2452604
19	3	1,969	SWEDEN	null	Curtis.Burkhart@rf4mZIC9EHIYI7n.org		2452365
20	7	1,965	SWEDEN	null	Ruth.Swann@aMRejVm.edu		2452524

Conclusion

Duration: 2

In this lab we learned how to catalog Snowflake data, how to link this data to its business context, how to enforce privacy policies on it and how to search and access it.

Collibra provides value for Snowflake users because it provides a unified environment to:

- see all technical metadata with its lineage
- with a full business context
- in a friendly user interface with:
 - powerful search, with filters
 - collaboration features (ratings, comments, messages),
 - automation (workflows, tasks, issue management)
- automatic policy creation and enforcement
- data quality assessment and metrics

Collibra is an ideal environment for all Snowflake users, business or technical, to view, search, understand, trust and collaborate with data.