databricks01.1 - Unity Catalog

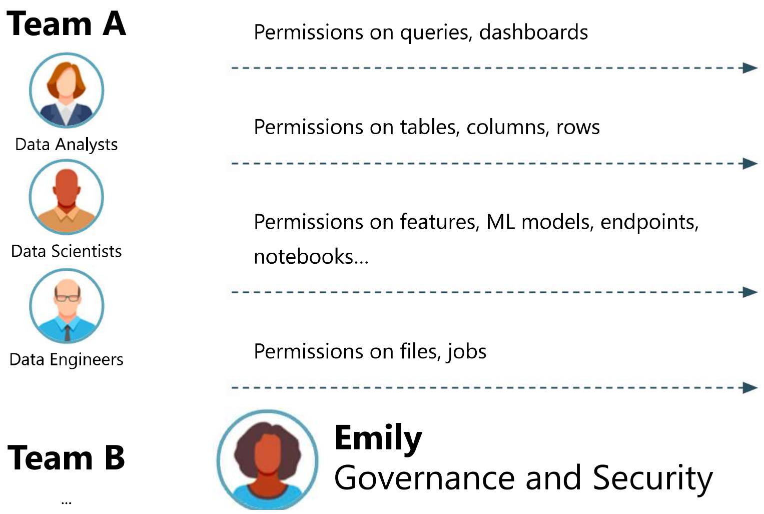
(https://databricks.com)

Ensuring Governance and security for our C360 lakehouse

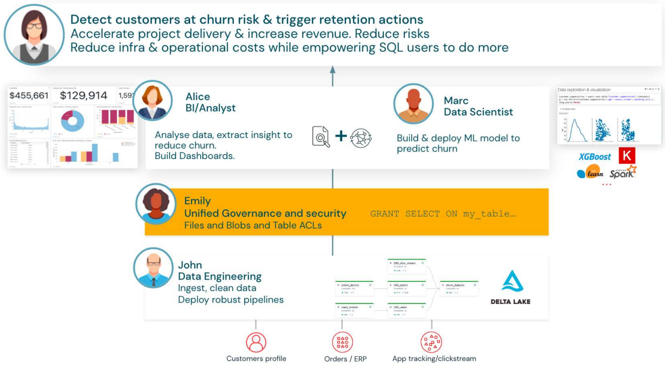
Data governance and security is hard when it comes to a complete Data Platform. SQL GRANT on tables isn't enough and security must be enforced for multiple data assets (dashboards, Models, files etc).

To reduce risks and driving innovation, Emily's team needs to:

- Unify all data assets (Tables, Files, ML models, Features, Dashboards, Queries)
- Onboard data with multiple teams
- Share & monetize assets with external Organizations


Team A
Data Analysts
Data Scientists
Data Engineers
Team B
...
Permissions on queries, dashboards
Permissions on tables, columns, rows
Permissions on features, ML models, endpoints, notebooks...
Permissions on files, jobs
Emily
Governance and Security

Implementing a global data governance and security with Unity Catalog


Detect customers at churn risk & trigger retention actions
Accelerate project delivery & increase revenue. Reduce risks
Reduce infra & operational costs while empowering SQL users to do more
$455,661
$129,914
1,59
Alice
BI/Analyst
Analyse data, extract insight to reduce churn.
Build Dashboards.
Marc
Data Scientist
Build & deploy ML model to predict churn
XGBoost
K
learn
Spark
...
Emily
Unified Governance and security
Files and Blobs and Table ACLs
GRANT SELECT ON my_table…
John
Data Engineering
Ingest, clean data
Deploy robust pipelines
DELTA LAKE
Customers profile
Orders / ERP
App tracking/clickstream

Let's see how the Lakehouse can solve this challenge leveraging Unity Catalog.

Our Data has been saved as Delta Table by our Data Engineering team. The next step is to secure this data while allowing cross team to access it.
A typical setup would be the following:

- Data Engineers / Jobs can read and update the main data/schemas (ETL part)
- Data Scientists can read the final tables and update their features tables
- Data Analyst have READ access to the Data Engineering and Feature Tables and can ingest/transform additional data in a separate schema.
- Data is masked/anonymized dynamically based on each user access level

This is made possible by Unity Catalog. When tables are saved in the Unity Catalog, they can be made accessible to the entire

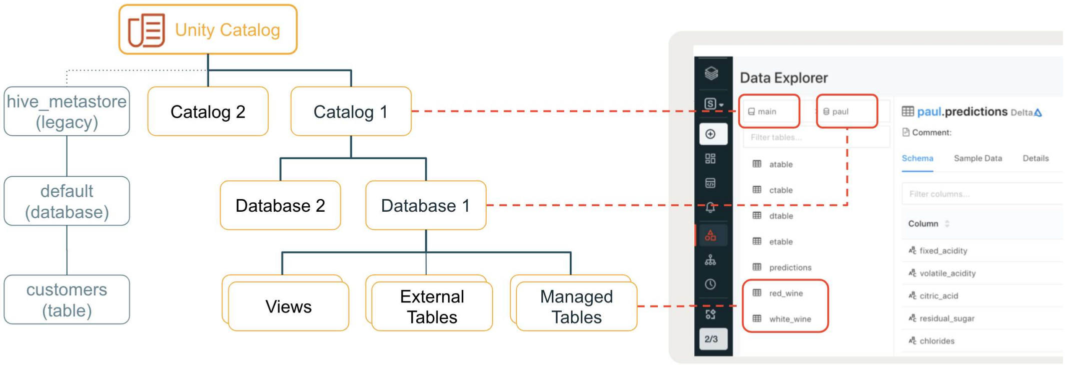
Exploring our Customer360 database

Let's review the data created.

Unity Catalog works with 3 layers:

- CATALOG
- SCHEMA (or DATABASE)
- TABLE

All unity catalog is available with SQL

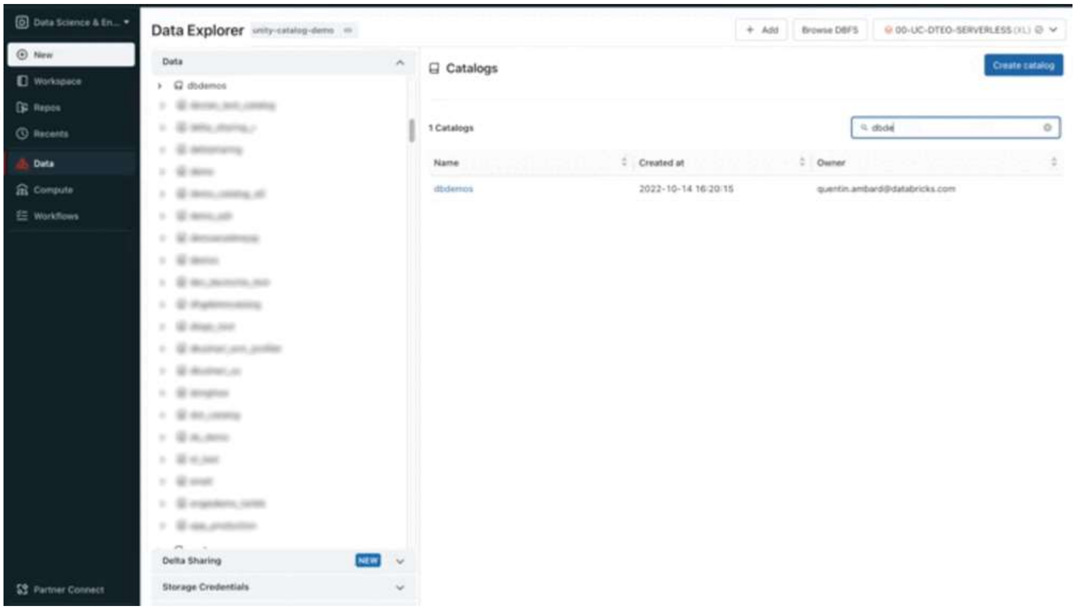


Let's review the tables we created under our schema

Unity Catalog provides a comprehensive Data Explorer that you can access on the left menu.

You'll find all your tables, and can use it to access and administrate your tables.

They'll be able to create extra table into this schema.



```
%run ./includes/SetupLab
```

The above script sets up the default catalog and schema

```
print("Working catalog and schema: " + labContext.catalogAndSchema())
```

```
Working catalog and schema: cloud_lakehouse_labs.odl_user_1237583_databrickslabs_com_retail
```

```
%sql
-- This lab has been set up to use a specific catalog, if it has been set up, or "main".
SELECT CURRENT_CATALOG();
```

Table

	current_catalog()
1	cloud_lakehouse_labs

1 row

```
%sql
-- Shows the schemas (databases) in the the current catalog
SHOW DATABASES;
```

Table

	databaseName
1	information_schema
2	odl_instructor_1036177_databrickslabs_com_retail
3	odl_user_1225167_databrickslabs_com_retail
4	odl_user_1237583_databrickslabs_com_retail
5	odl_user_1238721_databrickslabs_com_retail
6	odl_user_1238938_databrickslabs_com_retail

6 rows

```
%sql
-- Shows the tables in the current database
SHOW TABLES;
```

Table

	database	tableName	isTemporary
1	odl_user_1237583_databrickslabs_com_retail	churn_app_events	false
2	odl_user_1237583_databrickslabs_com_retail	churn_features	false
3	odl_user_1237583_databrickslabs_com_retail	churn_orders	false
4	odl_user_1237583_databrickslabs_com_retail	churn_orders_bronze	false
5	odl_user_1237583_databrickslabs_com_retail	churn_users	false
6	odl_user_1237583_databrickslabs_com_retail	churn_users_bronze	false

6 rows

Give access to your schema to other users / or groups

```
%sql
-- FILL IN <SCHEMA> and <TABLE>
GRANT USE SCHEMA ON SCHEMA <SCHEMA> TO `account users`;
GRANT SELECT ON TABLE <SCHEMA>.<TABLE> TO `account users`;
```

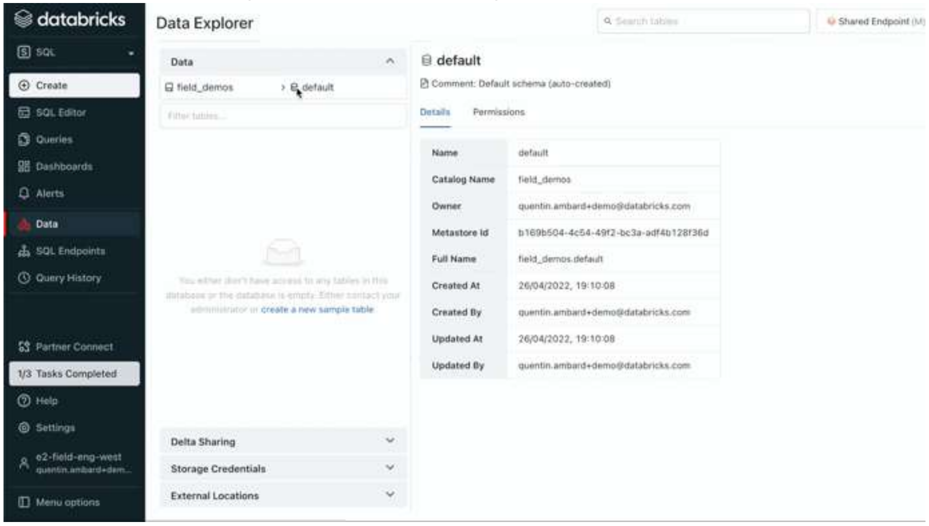
```
ParseException:
Operation not allowed: GRANT.(line 2, pos 0)

== SQL ==
-- FILL IN <SCHEMA> and <TABLE>
GRANT USE SCHEMA ON SCHEMA <SCHEMA> TO `account users`
^^^
```

Going further with Data governance & security

By bringing all your data assets together, Unity Catalog let you build a complete and simple governance to help you scale your teams.

Unity Catalog can be leveraged from simple GRANT to building a complete datamesh organization.



Comment: Default schema (auto-created)

You either don't have access to any tables in this database or the database is empty. Either contact your administrator or create a new sample table

Name	default
Catalog Name	field_demos
Owner	quentin.ambard+demo@databricks.com
Metastore Id	b169b504-4c54-49f2-bc3a-adf4b128f36d
Full Name	field_demos.default
Created At	26/04/2022, 19:10:08
Created By	quentin.ambard+demo@databricks.com
Updated At	26/04/2022, 19:10:08
Updated By	quentin.ambard+demo@databricks.com

Fine-grained ACL

Need more advanced control? You can chose to dynamically change your table output based on the user permissions.

Secure external location (S3/ADLS/GCS)

Unity Catatalog let you secure your managed table but also your external locations.

Lineage

UC automatically captures table dependencies and let you track how your data is used, including at a row level.

This leat you analyze downstream impact, or monitor sensitive information across the entire organization (GDPR).

Audit log

UC captures all events. Need to know who is accessing which data? Query your audit log.

This leat you analyze downstream impact, or monitor sensitive information across the entire organization (GDPR)