

Mastering Data Governance & Security In Modern Cloud Data Platforms

using Databricks And Snowflake



The Cambridge Analytica Scandal



Cambridge Analytica and Facebook: The Scandal and the Fallout So Far

Marriott Data Breach



Marriott's Data Breach One of the Biggest in History

Equifax Data Breach

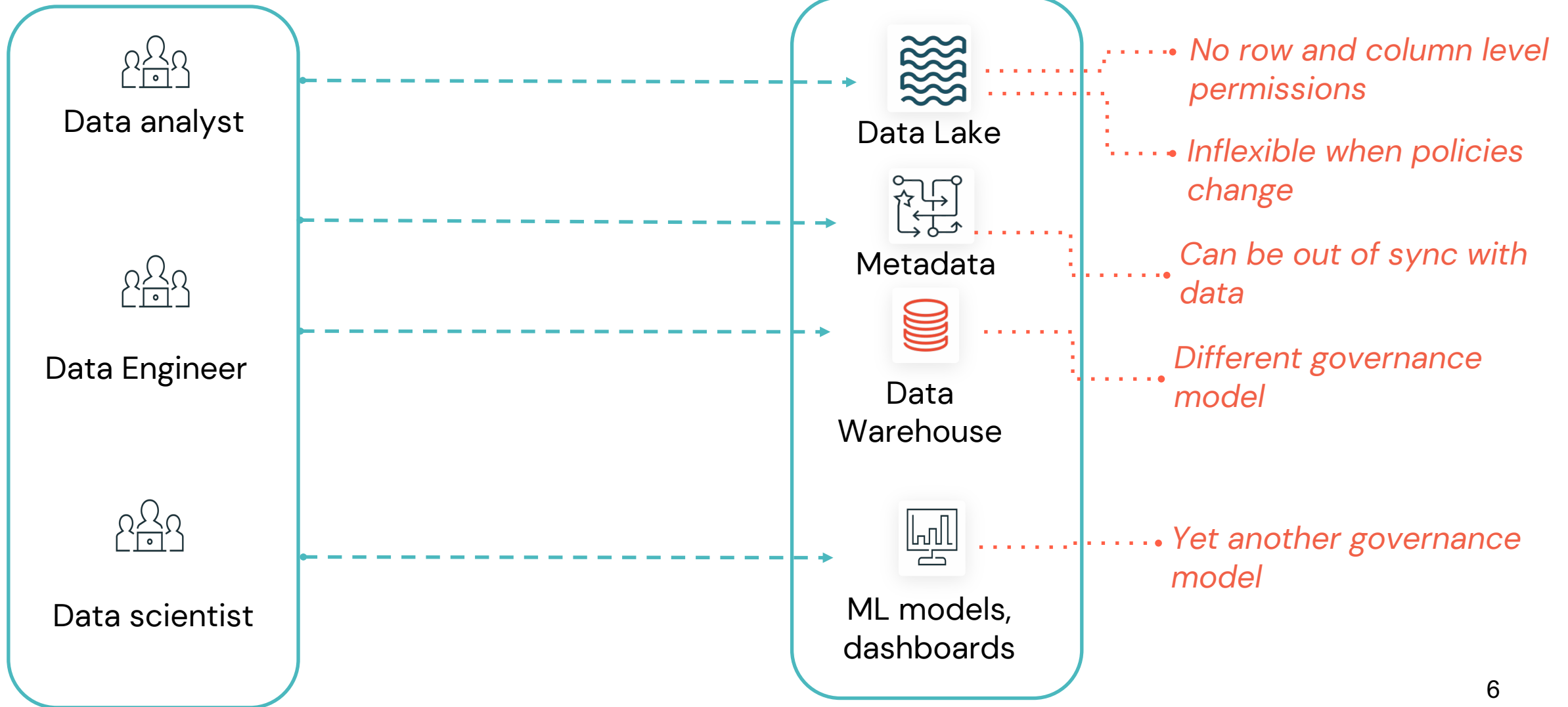


Facebook Data Leak



[Facebook Data Breach: How to Tell If Your Account Was Exposed | Fortune](#)

Why Data Governance & Data Security Matter





Philip Goldman

Sr. Data Solution Architect
EPAM

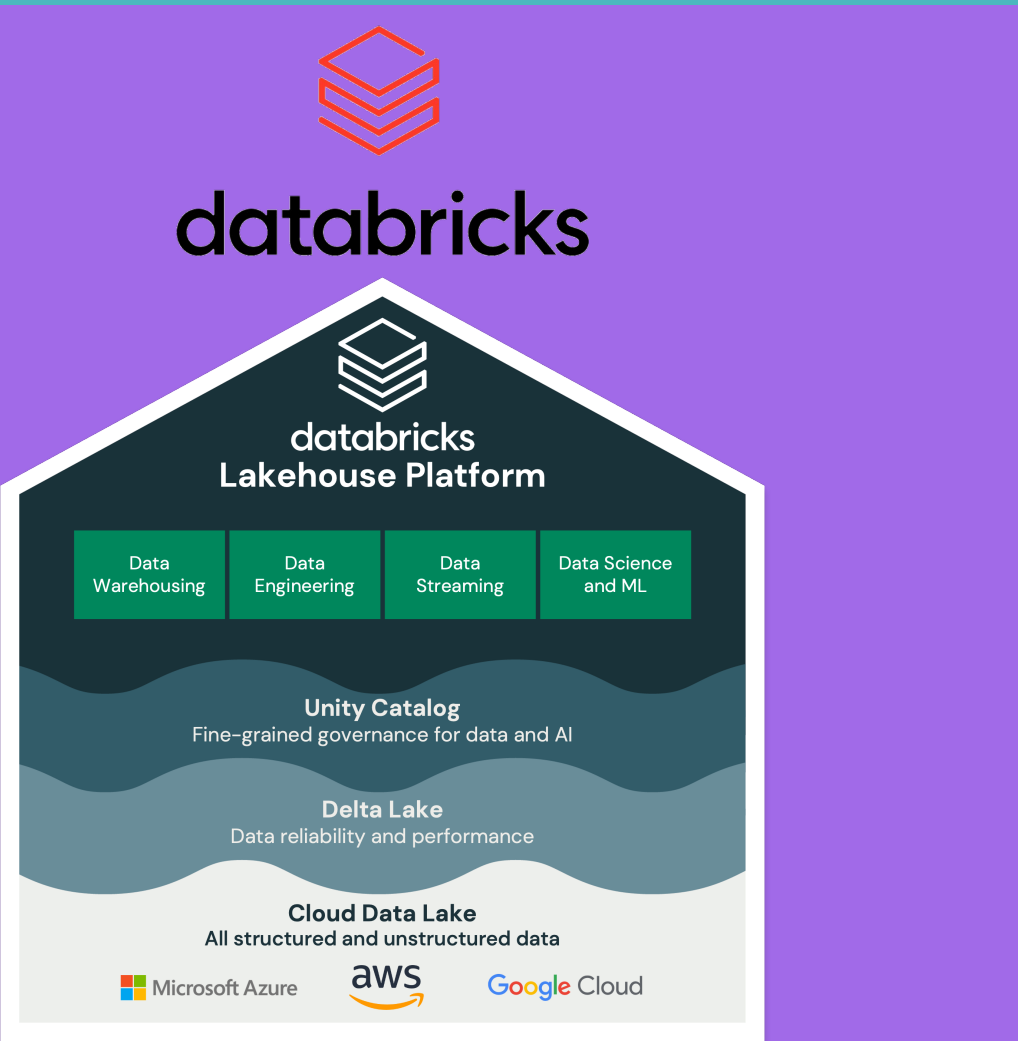
Navigating Today's Discussion

Elaborate on the data you want to discuss.



Choosing Databricks & Snowflake

Some of the benefits and reasons businesses choose Databricks or Snowflake



Magic Quadrant for Cloud Database Management Systems



Principles of Security and Data Governance



Principles of data governance

Data Governance

Unify data
management

Principles of data governance

Data Governance

Unify data
management

Unify Data
Security

Principles of data governance

Data Governance

Unify data
management

Unify Data
Security

Manage Data Quality

Principles of data governance

Data Governance

Unify data
management

Unify Data
Security

Manage Data Quality

Data Sharing

Principles of security, compliance, and privacy

Data Governance

Unify data
management

Unify Data
Security

Manage Data Quality

Data Sharing

Security, compliance, and privacy

Identity &
Privileges

Principles of security, compliance, and privacy

Data Governance

Unify data
management

Unify Data
Security

Manage Data Quality

Data Sharing

Security, compliance, and privacy

Identity &
Privileges

Data
security

Principles of security, compliance, and privacy

Data Governance

Unify data management

Unify Data Security

Manage Data Quality

Data Sharing

Security, compliance, and privacy

Identity & Privileges

Data security

Network security

Principles of security, compliance, and privacy

Data Governance

Unify data management

Unify Data Security

Manage Data Quality

Data Sharing

Security, compliance, and privacy

Identity & Privileges

Data security

Network security

Compliance & Privacy

Principles of security, compliance, and privacy

Data Governance

Unify data management

Unify Data Security

Manage Data Quality

Data Sharing

Security, compliance, and privacy

Identity & Privileges

Data security

Network security

Compliance & Privacy

Security Monitoring

How do we implement these principles in real-world cloud platforms like Databricks and Snowflake?



— Bring Principles with Practice

Security & Data Governance Best Practices



— In Snowflake



SNOWFLAKE SECURITY AT A GLANCE



Access

- All communication secured & encrypted
- TLS 1.2 encryption in both trusted and untrusted networks
- IP Whitelisting
- Private Link



Authentication

- Password Policy enforcement
- Multifactor Authentication
- SAML 2.0 support for Federated Authentication



Authorization

- Flexible user management
- Role-based access control for granular control
- RBAC for data and actions



Data

- Encrypted at rest
- Hierarchical key model rooted in Cloud HSM
- Automatic key rotation
- Time Travel 1-90 days
- Tri-Secret Secure
- Query statement encryption



Infrastructure

- AWS, Azure Physical Security
- AWS, Azure Redundancy
- Regional Data Centers
US
EU
AP

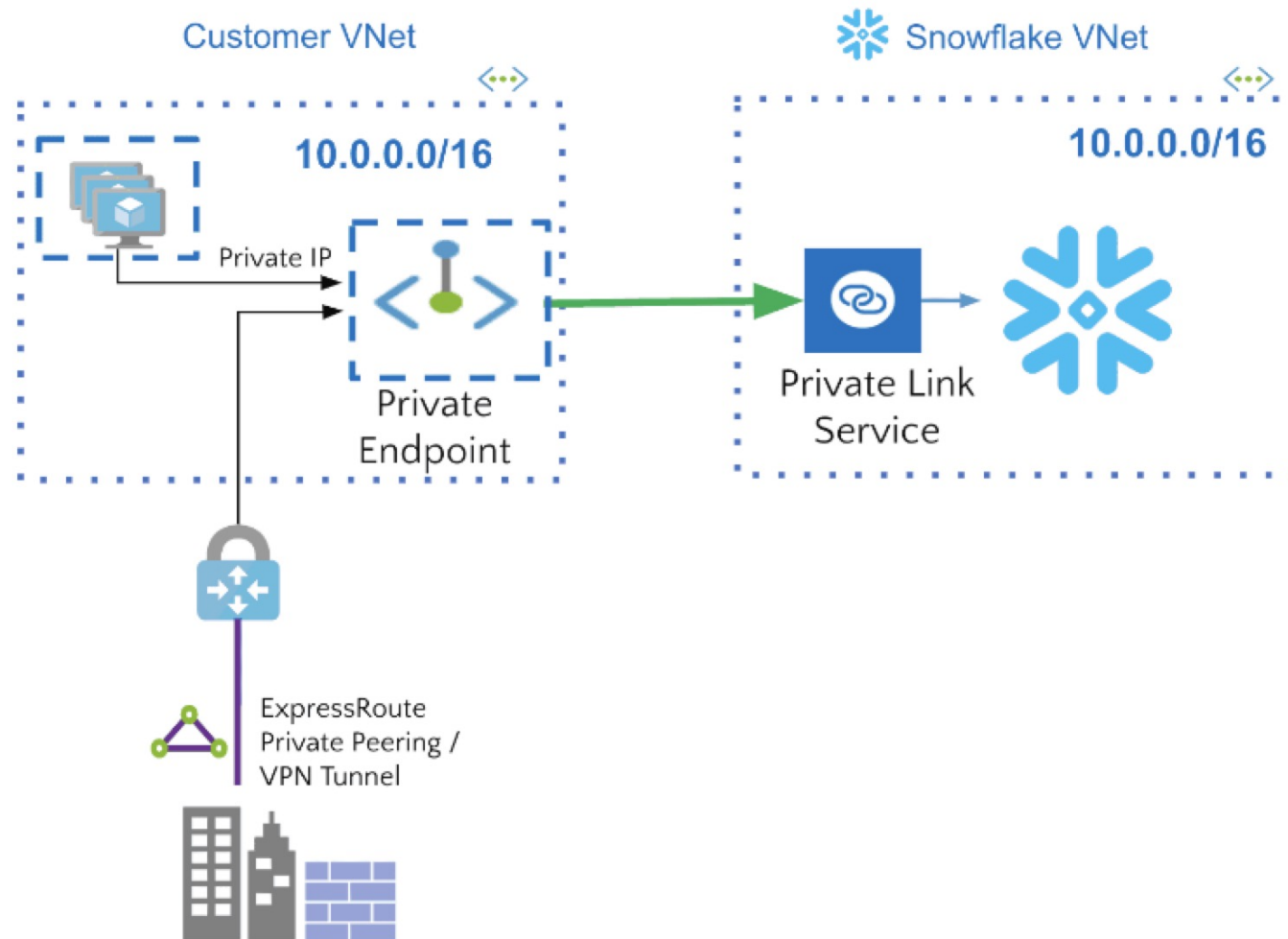


Snowflake Operational Controls

- NIST 800-53
- SOC2 Type 2
- HIPAA
- PCI
- FedRAMP



AWS/Azure/GCP Private Link



Snowflake Network Policies

Network policies can be applied at three levels



1. Snowflake Account



2. Outside Integration



3. User Specific

Edit network policy

MULTIVERSE as ACCOUNTADMIN

Enter a valid IPv4 address and optional CIDR or multiple addresses in a comma-separated list.

4 allowed IP addresses

3	. . . 2	×
7	. . . 26	×
1	. . . 17	×
1	. . . 31	×

0 blocked IP addresses

No blocked addresses

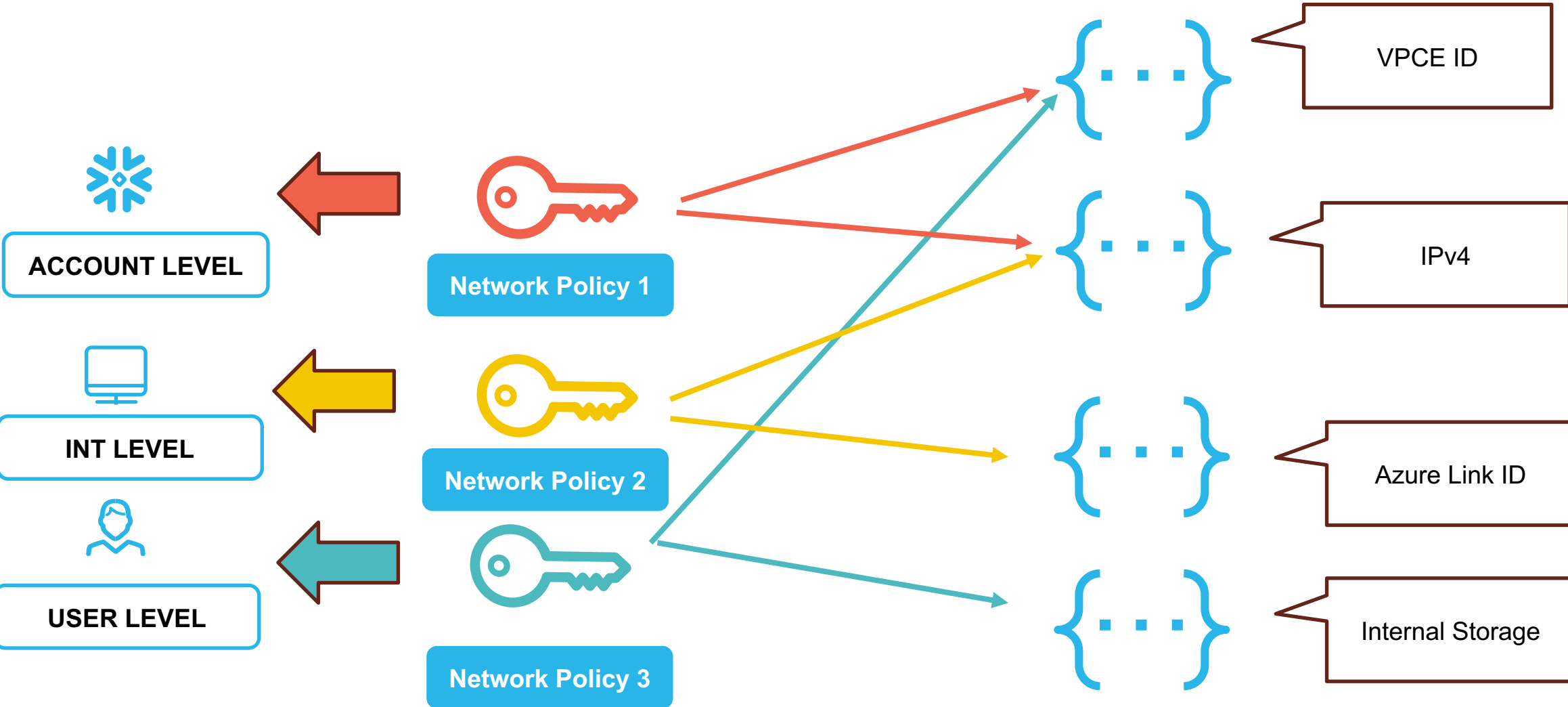
Add addresses that will be blocked from accessing your Snowflake account

Comment (optional)

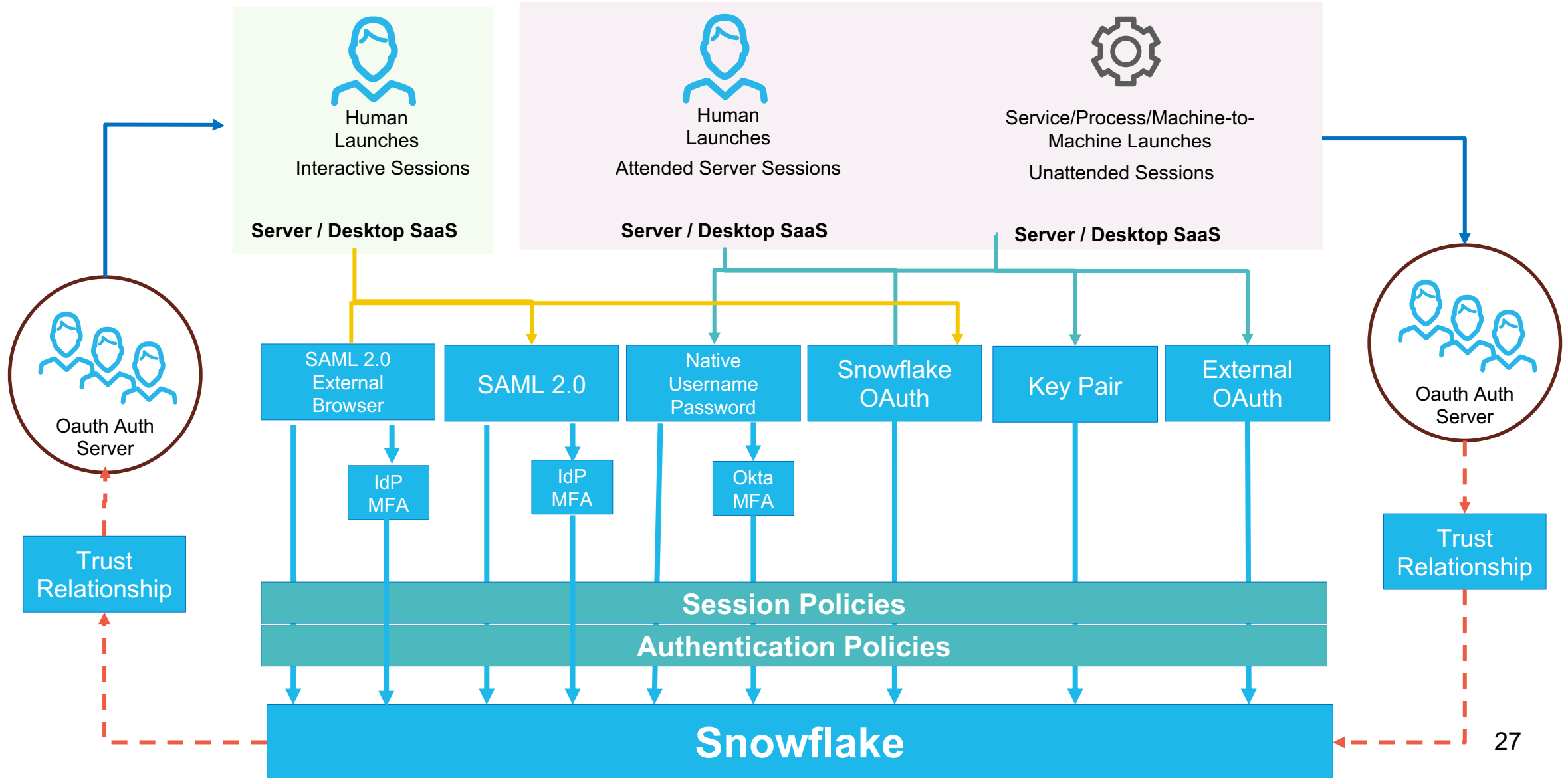
Cancel Save changes

The most specific policy always wins.

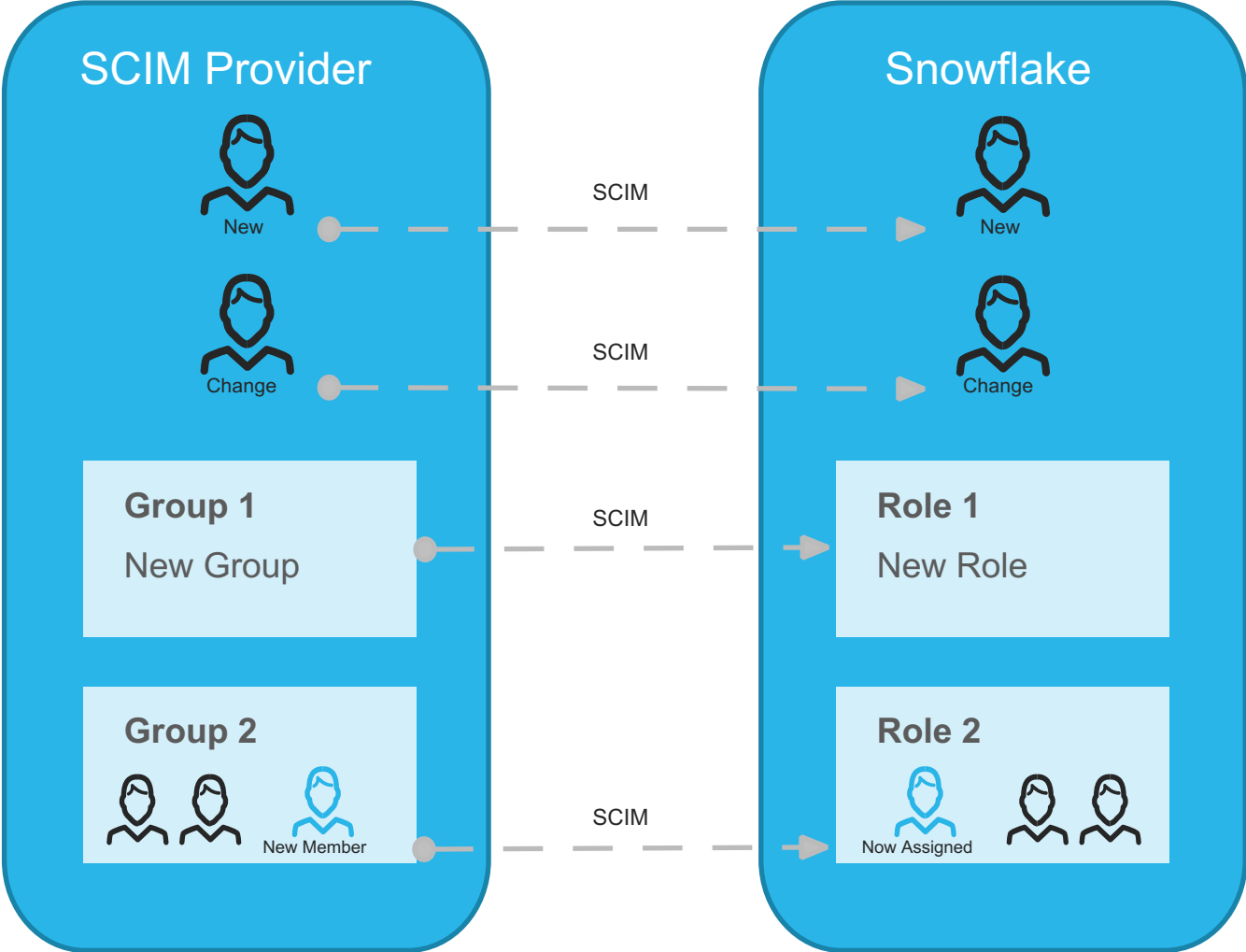
Snowflake Network rule



Snowflake Authentication

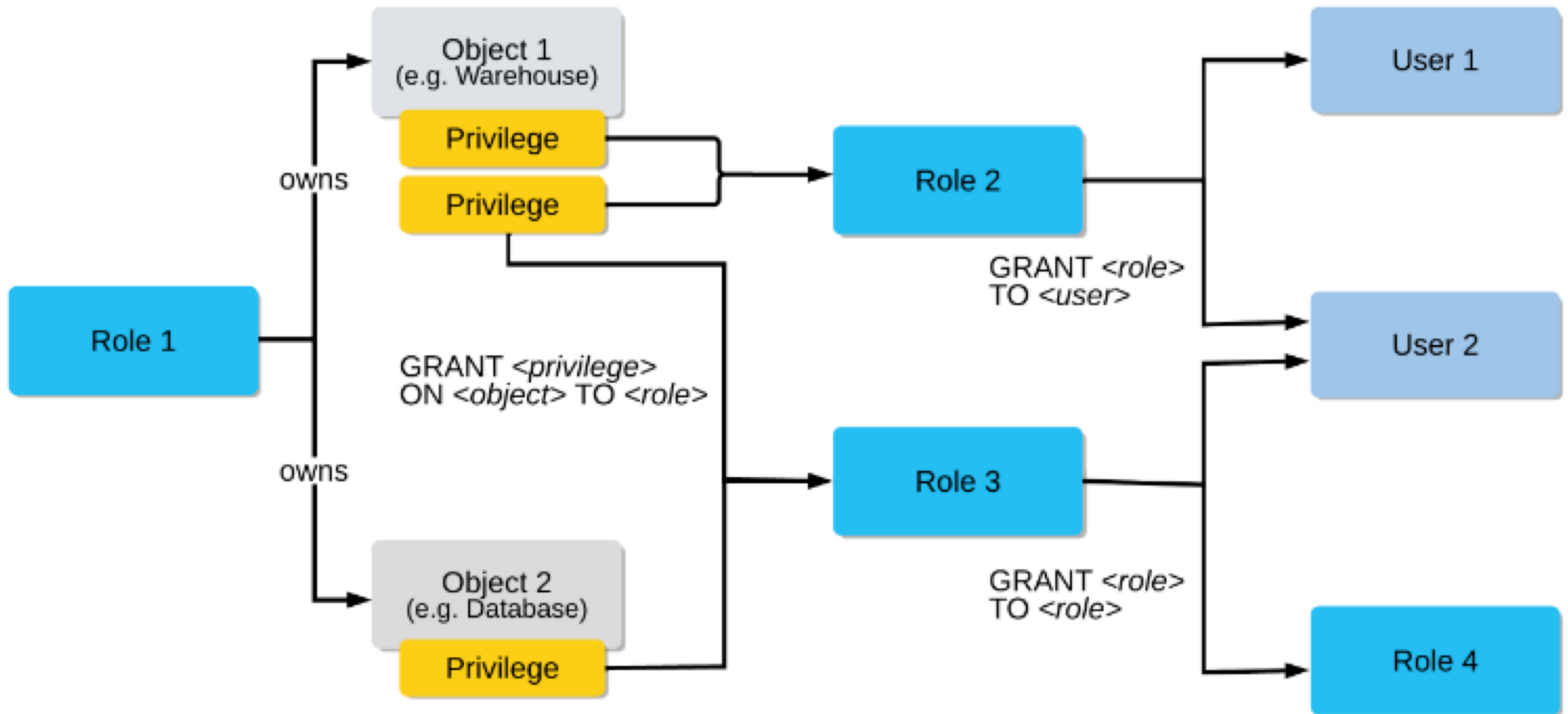


User Provisioning with SCIM



RBAC

Role-based access control



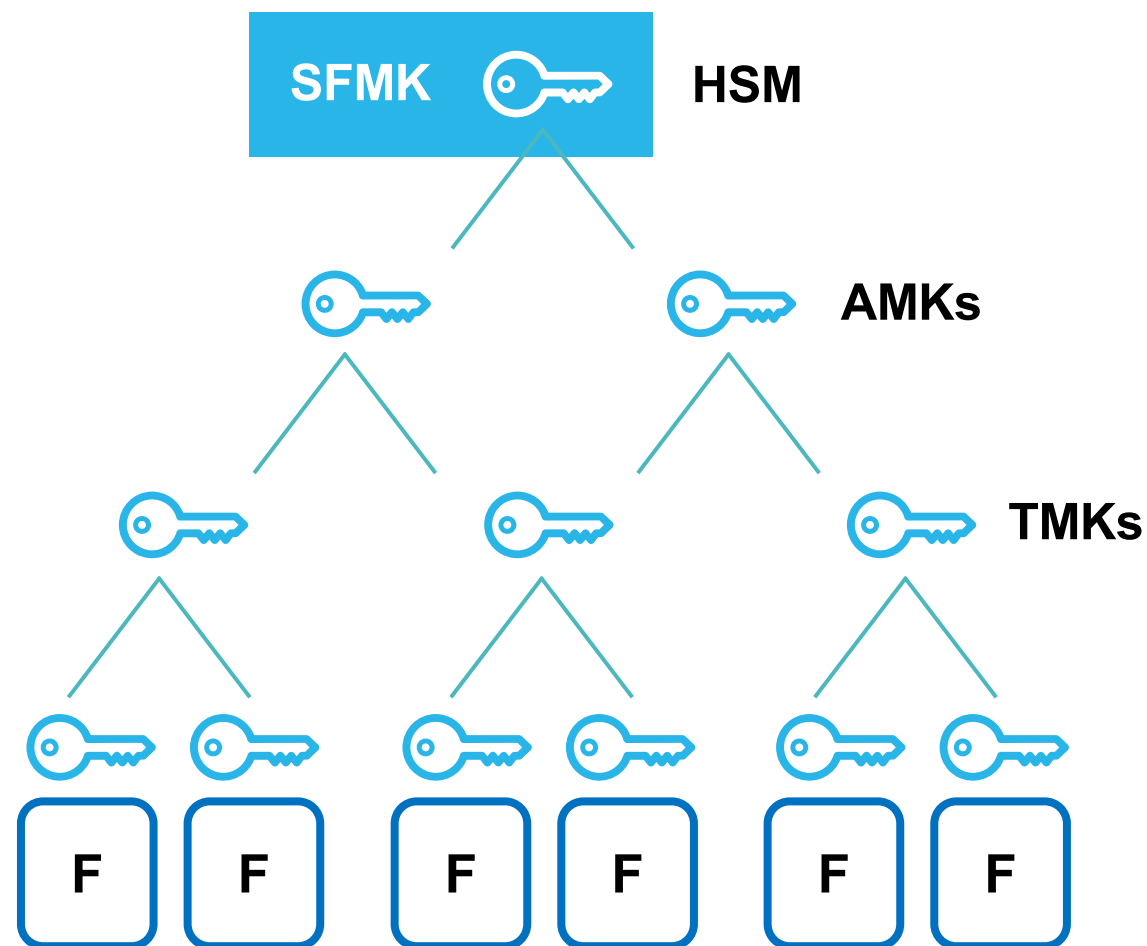
Tri-Secret Secure

Encrypt using customer managed key (CMK)



Customer Key

- Bring your own key (BYOK)
- Revoke at any time → no access to data
- Keys are rotated every 30 days
- Customers setup re-keying



Unified Governance



Know Your Data

What Where	Classification
	Object Tagging
	Object Dependencies
Who	Access History
	Account Usage





Protect Your Data

Row Access Policies	Dynamic Data Masking
External Tokenization	Tag-Based Policies
Conditional Masking	Anonymization (PrPr)



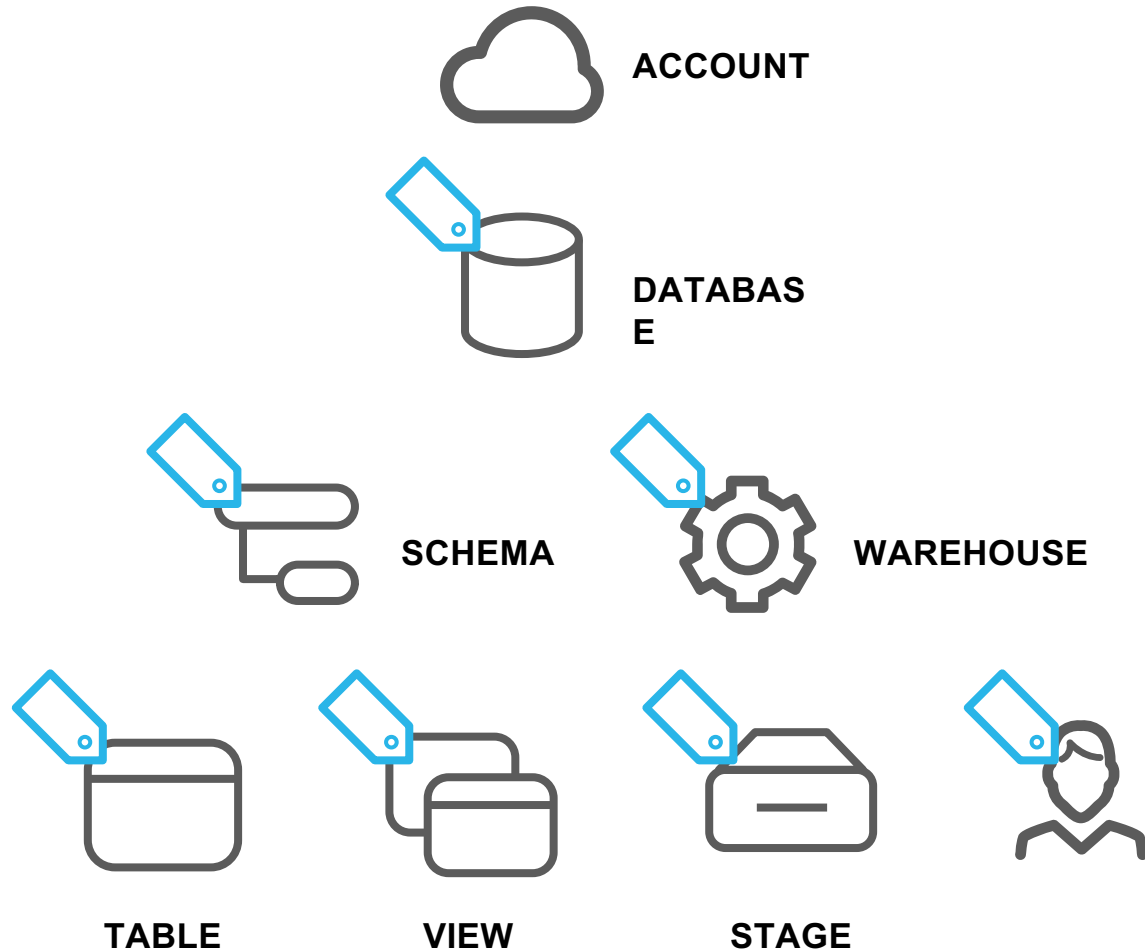
Connect Your Ecosystem

 COLLABORATION	 DATA GOVERNANCE ACCELERATED
Direct Secure Sharing	Pre-built Partner Integrations to Manage Entire Data Estate
Data Cleanrooms	
Data Marketplace	



Object Tagging

Track sensitive data and compute objects




Track sensitive and PII data

Track resource usage for cost visibility and attribution

Flexible privilege management options

Data Classification

Classify sensitive personal data



fname	gen	age	phone
Jane Doe	F	50	333-555-1236
Michael Gaines	M	75	666-666-1357
Ann Marshall	F	48	555-555-1234
John Smith	M	39	123-555-1234

Diagram illustrating data classification. A table with columns: fname, gen, age, phone. Rows: Jane Doe (F, 50, 333-555-1236), Michael Gaines (M, 75, 666-666-1357), Ann Marshall (F, 48, 555-555-1234), John Smith (M, 39, 123-555-1234). Tags are placed above the columns: NAME (above fname), GENDER (above gen), AGE (above age), PHONE_NUMBER (above phone). A magnifying glass icon is shown to the left of the table.

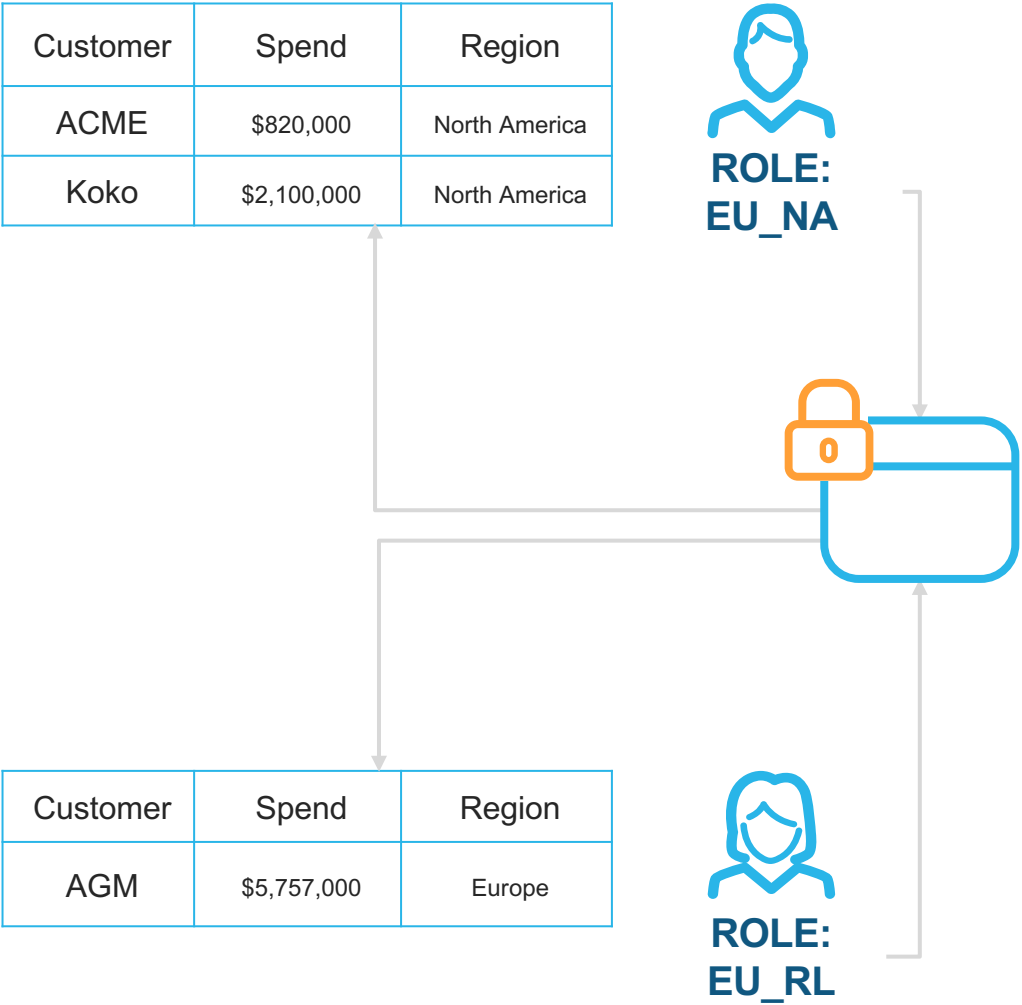
Analyze table columns for sensitive personal information

Get recommended tags using built-in machine learning

Apply tags to track and audit sensitive data

Row Access Policies

Row-level Security



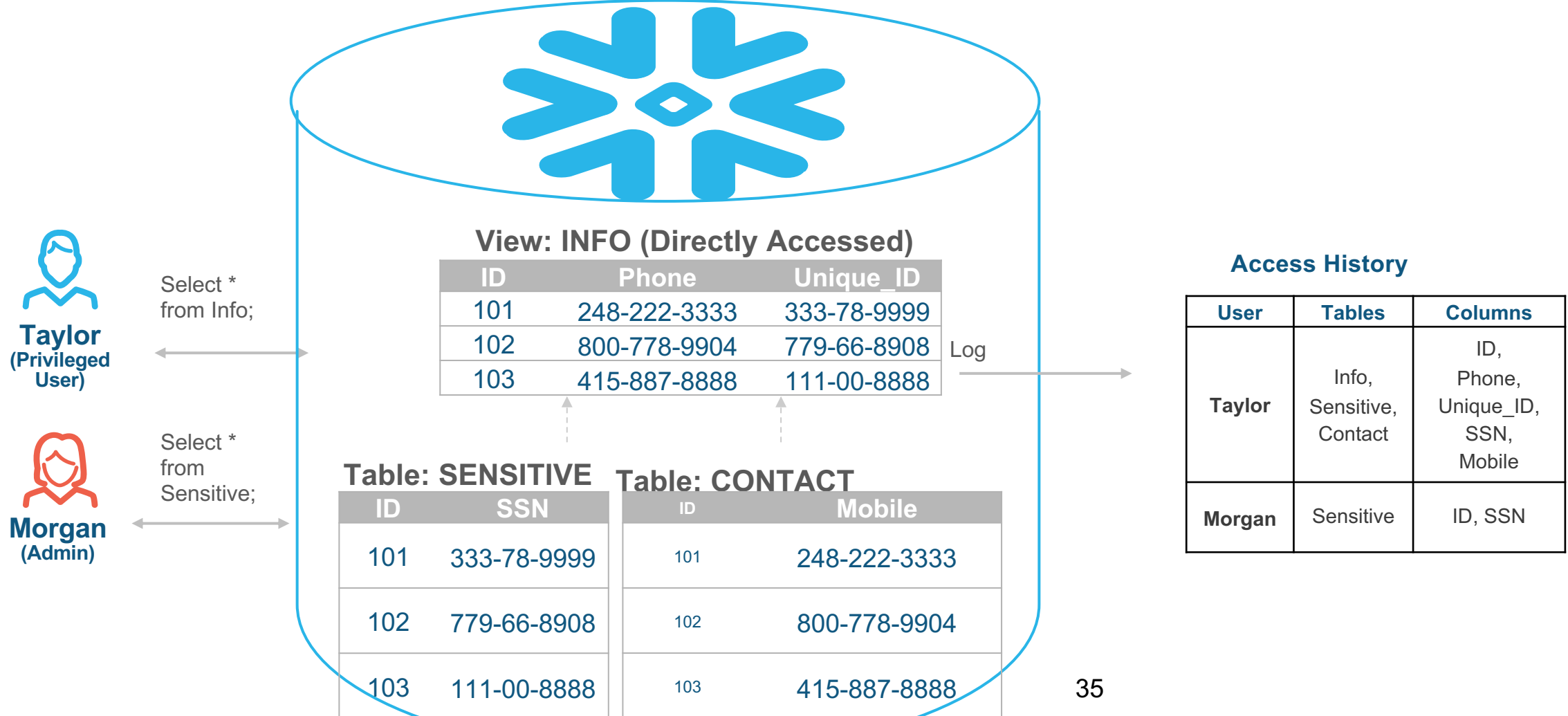
Filter unauthorized rows at query time

Use mapping tables for authorization

Apply one policy to many tables

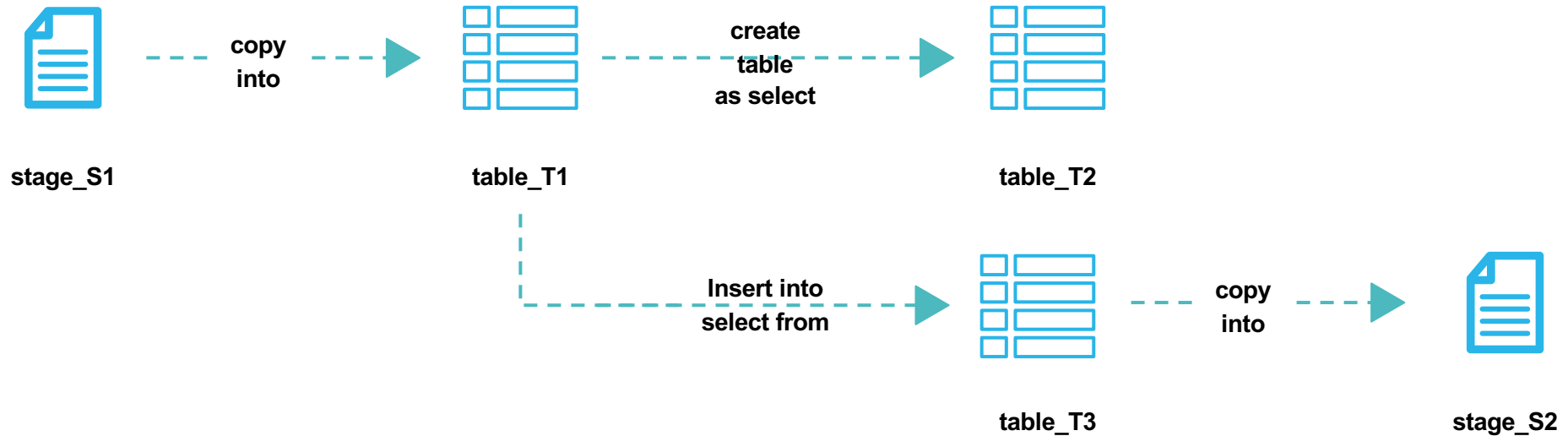
Access History (Reads)

Satisfy regulatory compliance, understand usage with column-level access visibility



Access History (Writes)

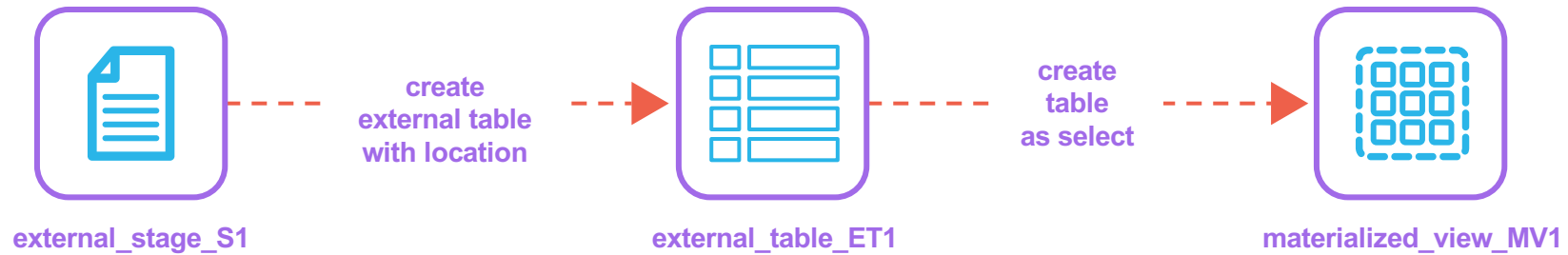
Know the lineage of data



PATH	TARGET_NAME	TARGET_DOMAIN	TARGET_COLUMNS
stage_S1->table_T1	table_T1	TABLE	["CONTENT"]
stage_S1->table_T1->table_T2	table_T2	TABLE	["ID","NAME"]
stage_S1->table_T1->table_T3	table_T3	TABLE	["NAME","ID"]
stage_S1->table_T1->table_T3->stage_S2	stage_S2	STAGE	[]

Object Dependencies

Identify dependencies and downstream impact



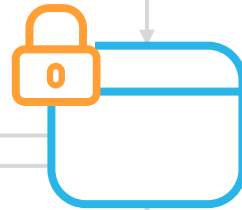
REFERENCED_OBJECT_NAME	REFERENCED_OBJECT_DOMAIN	REFERENCING_OBJECT_NAME	TREFERENCING_OBJECT_DOMAIN
external_stage_S1	STAGE	external_table_ET1	EXTERNAL TABLE
external_table_ET1	EXTERNAL_TABLE	materialized_view_MV1	MATERIALIZED_VIEW

Dynamic Data Masking

Column-level Security

ID	Phone	SSN
101	***-***-5534	*****
102	***-***-3564	*****
103	***-***-9787	*****


(Authorized
Access:
Restricted Data)



ID	Phone	SSN
101	408-123-5534	111-22-3333
102	510-335-3564	222-33-4444
103	214-553-9787	333-44-5555


(Authorized
Access:
Unrestricted Data)

Dynamically mask data at query time

Centralized policy management

Apply one policy to many columns

External Tokenization

Dynamically de-tokenize data for authorized users

Externally tokenize protected data at ingest

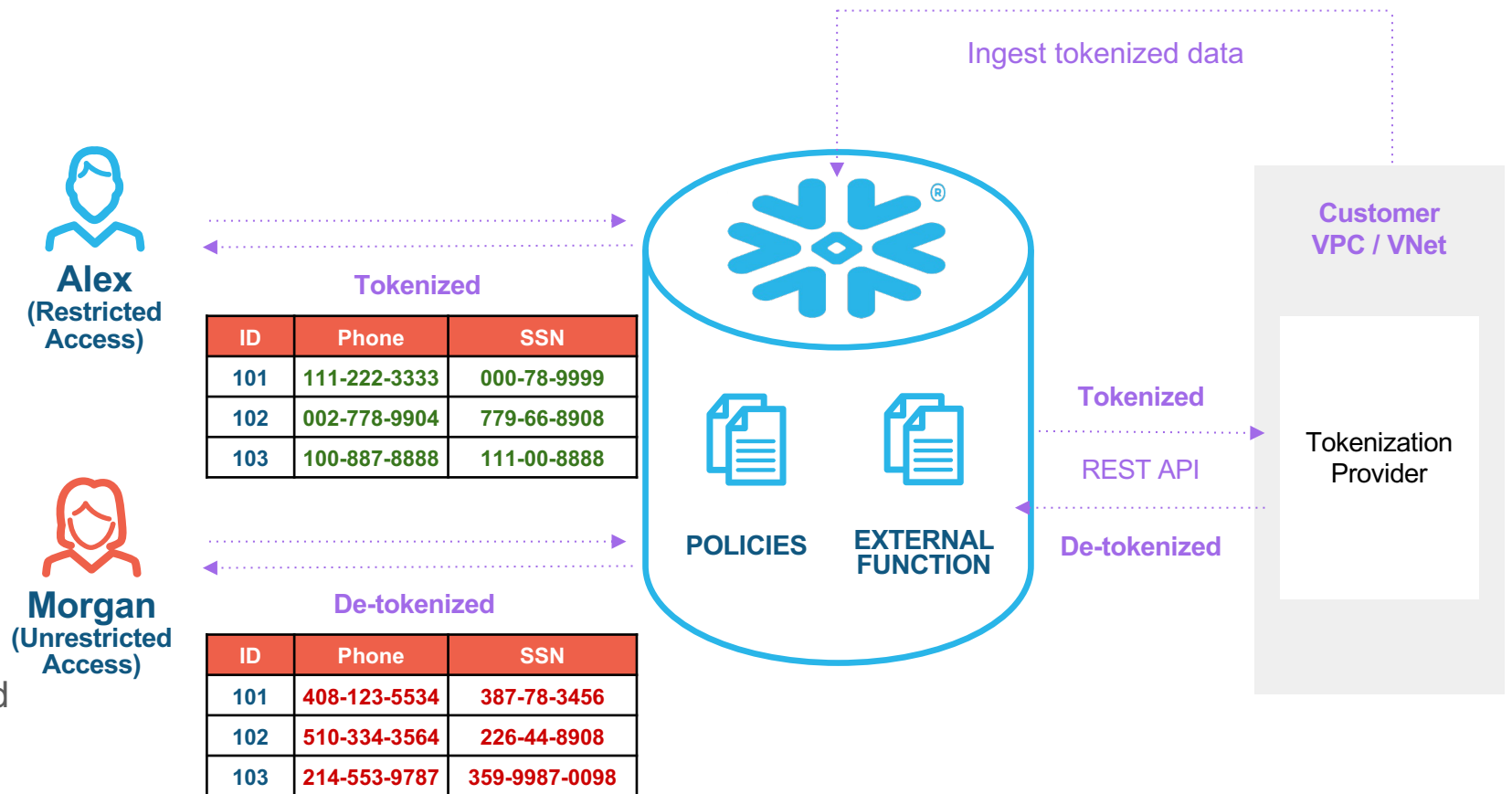
- Using tokenization provider agents on ETL tools

Dynamically de-tokenize at query time

- Call third-party service using external functions to de-tokenize data
- For unauthorized users, third-party service is not called

Policy Based Control

- Table/View owners and privileged users unauthorized by default
- Centralized policy management



Benefits of Snowflake Collaboration

Across Cloud & Region with Snowgrid

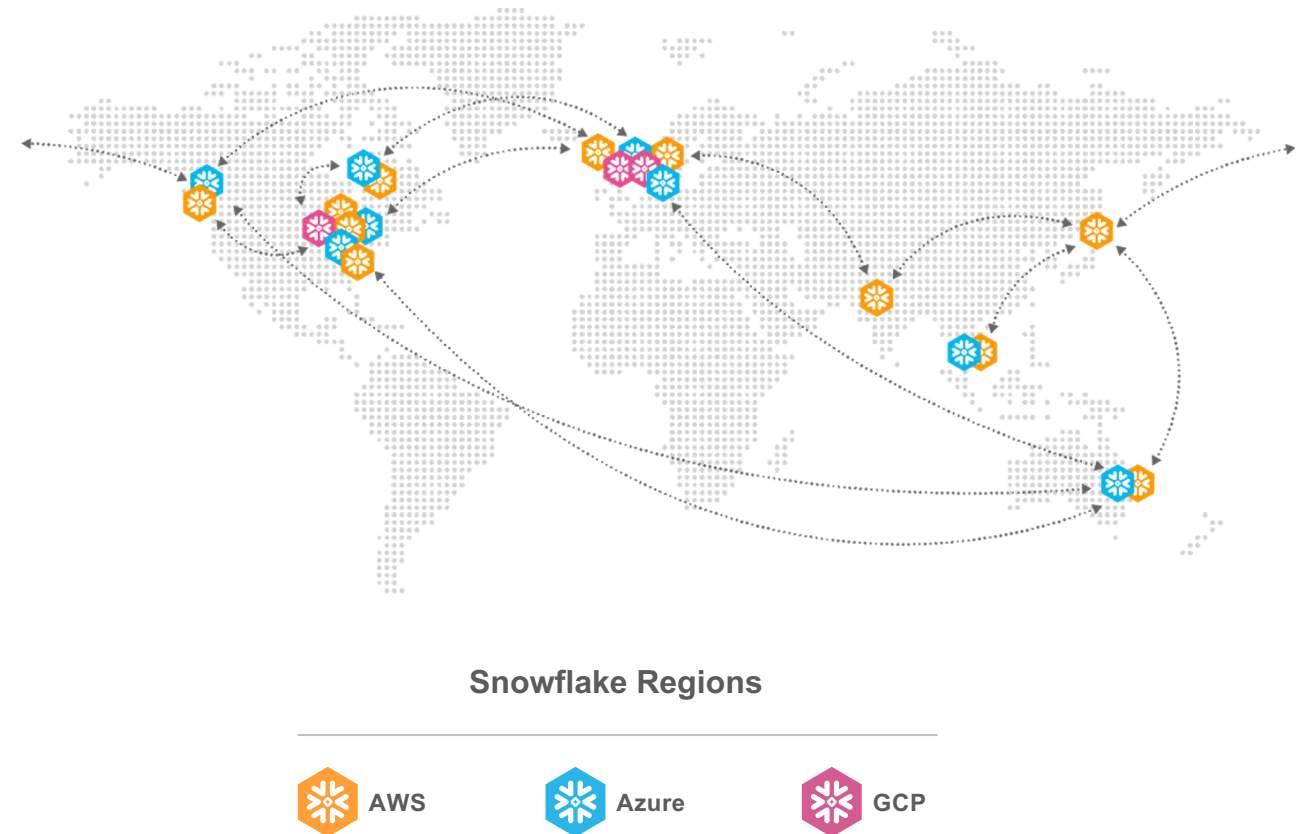
Delivers direct access to live, ready-to-query data across clouds and regions with auto-fulfillment and no ETL

More than Data

Snowflake enables customers to collaborate with data, data services and applications including built-in usage based monetization.

Robust Data Governance

Achieve privacy-preserving collaboration with targeted discovery, revocable access and custom event logging.



Benefits of Snowflake Collaboration

Across Cloud & Region with Snowgrid

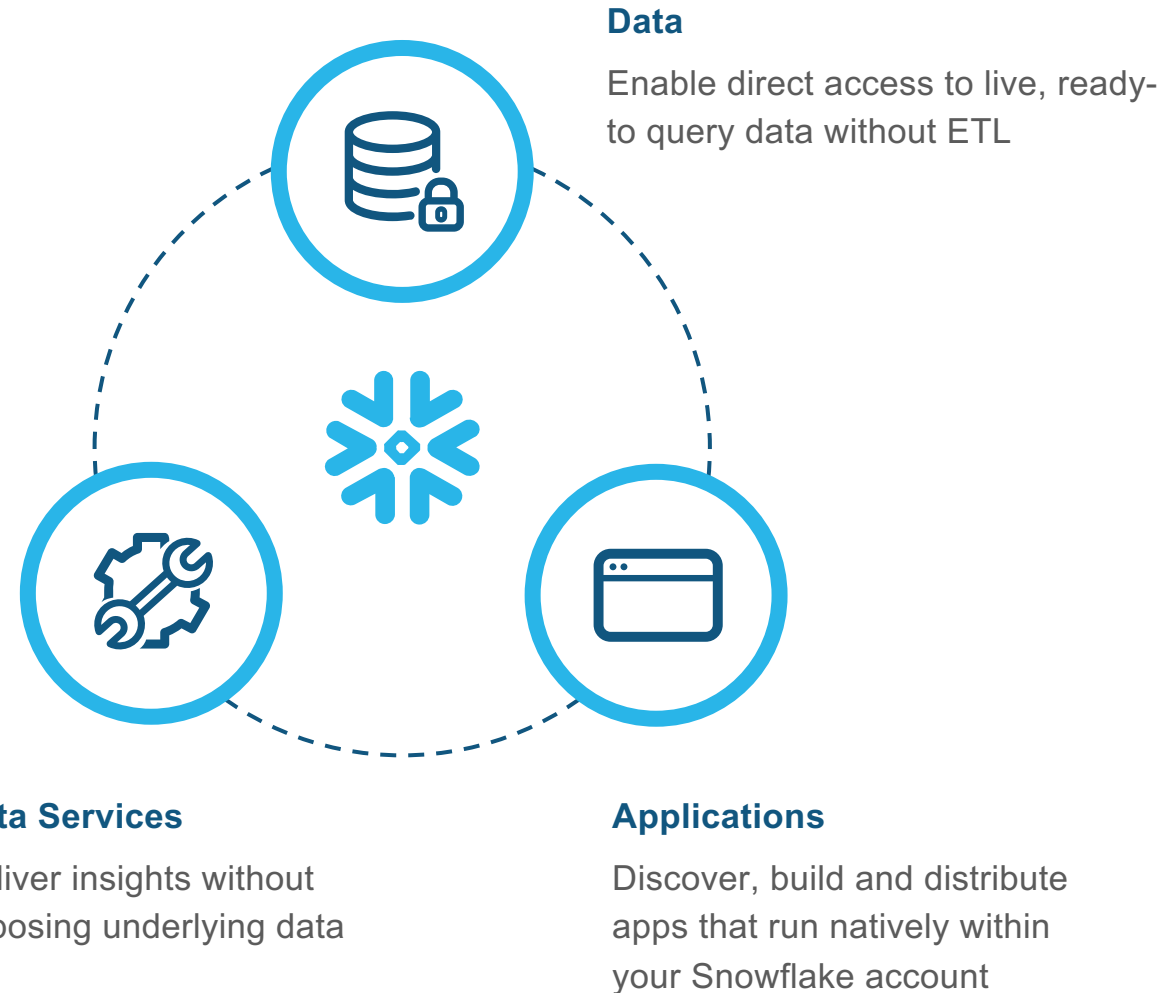
Delivers direct access to live, ready-to-query data across clouds and regions with auto-fulfillment and no ETL

More than Data

Snowflake enables customers to collaborate with data, data services and applications including built-in usage based monetization.

Robust Data Governance

Achieve privacy-preserving collaboration with targeted discovery, revocable access and custom event logging.



Benefits of Snowflake Collaboration

Across Cloud & Region with Snowgrid

Delivers direct access to live, ready-to-query data across clouds and regions with auto-fulfillment and no ETL

More than Data

Snowflake enables customers to collaborate with data, data services and applications including built-in usage based monetization.

Robust Data Governance

Achieve privacy-preserving collaboration with targeted discovery, revocable access and custom event logging.

Discovery

- Single Account
- Account Group
- Cloud Region(s)
- Public Marketplace

Audit

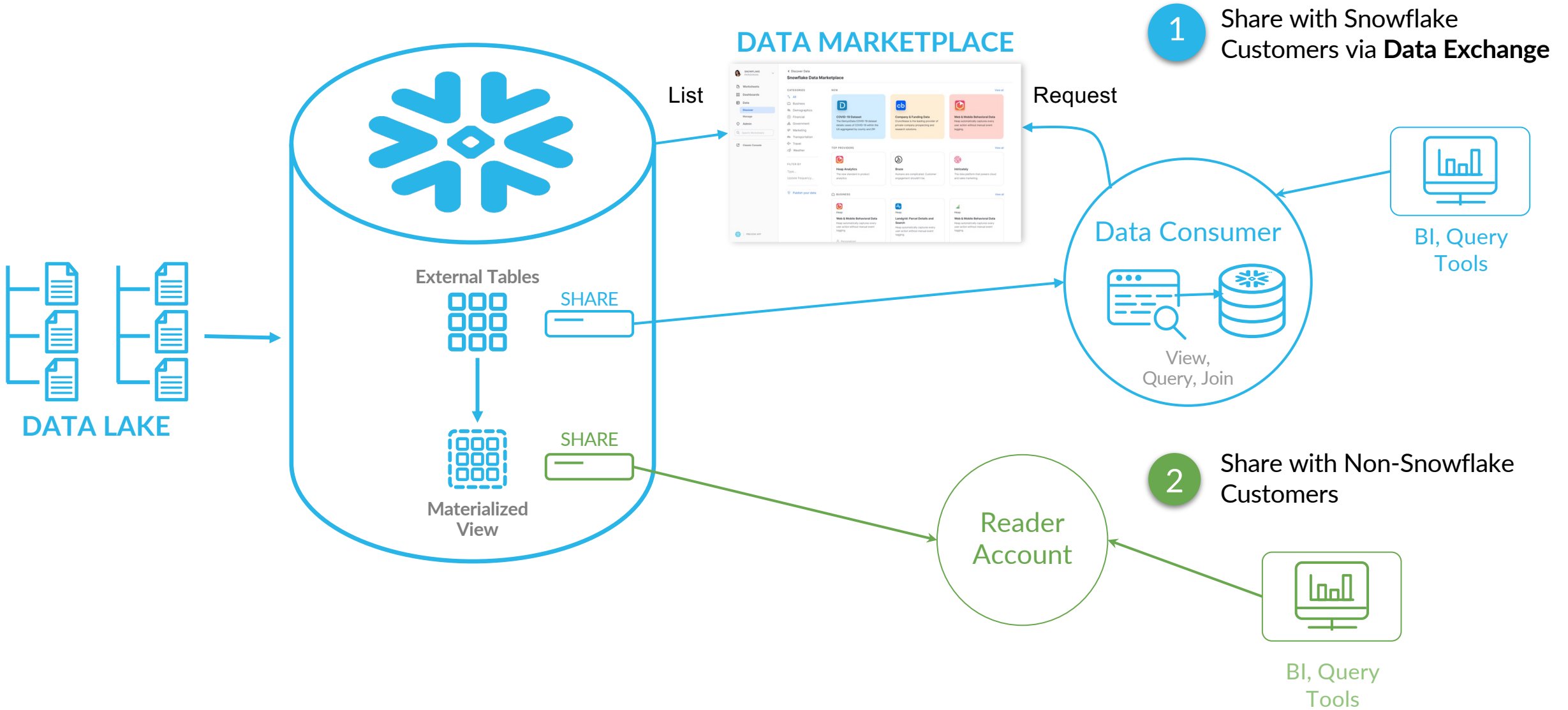
- Listing Views & Events
- Jobs Run by Consumer
- Object & Columns Accessed
- Custom Event Logging



Access

- Row Access Policies
- Dynamic Data Masking
- Conditional Masking
- Query Constraints*

SHARE YOUR LAKE!



Security & Data Governance Best Practices

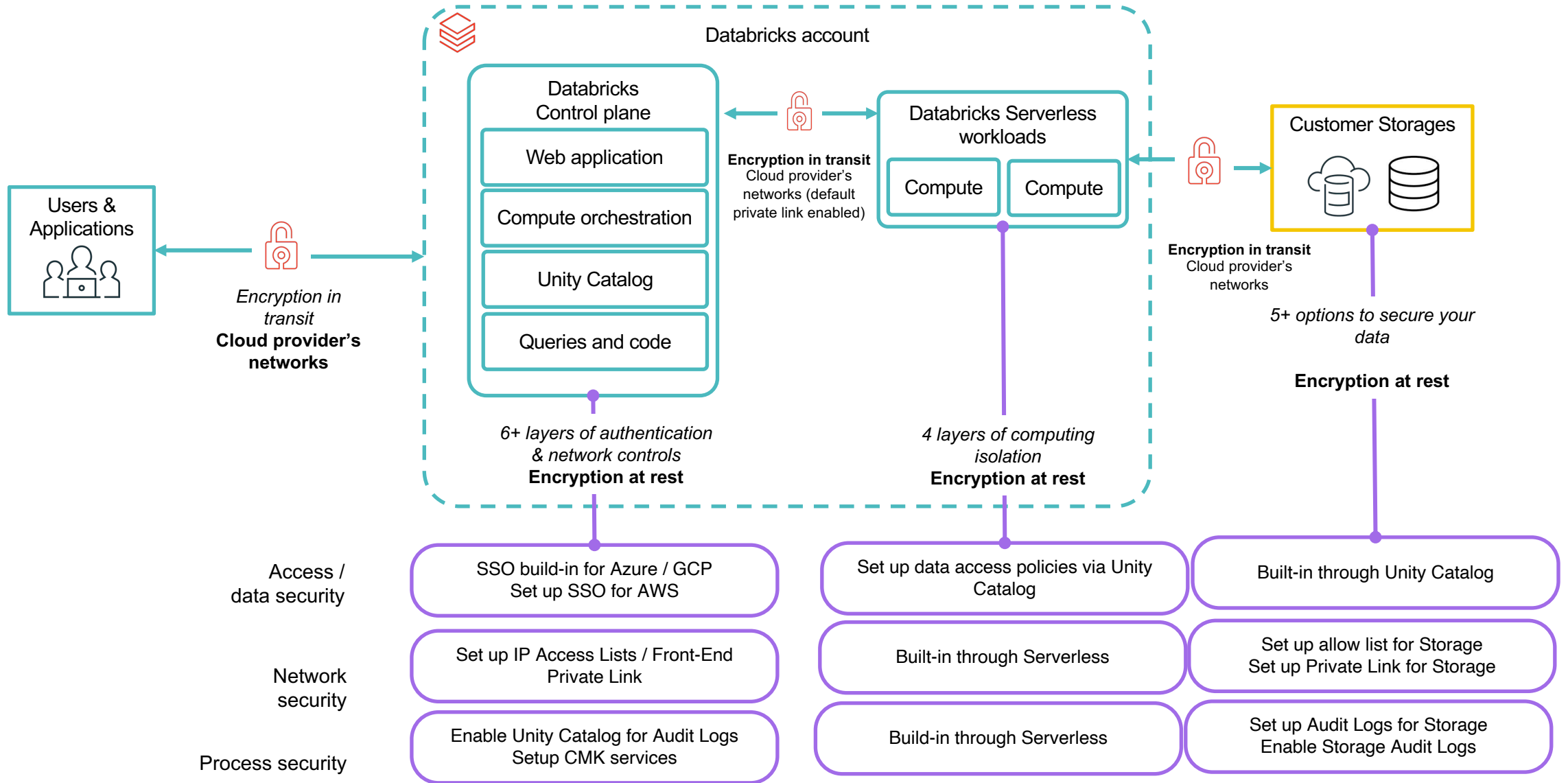


databricks



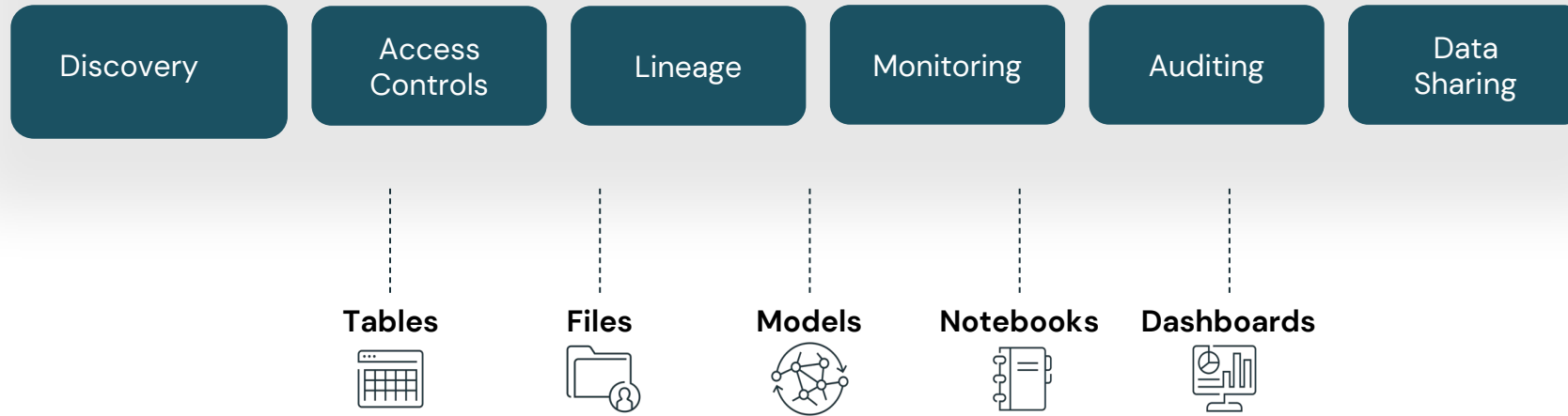
— In Databricks

Security Overview



What is Databricks Unity Catalog?

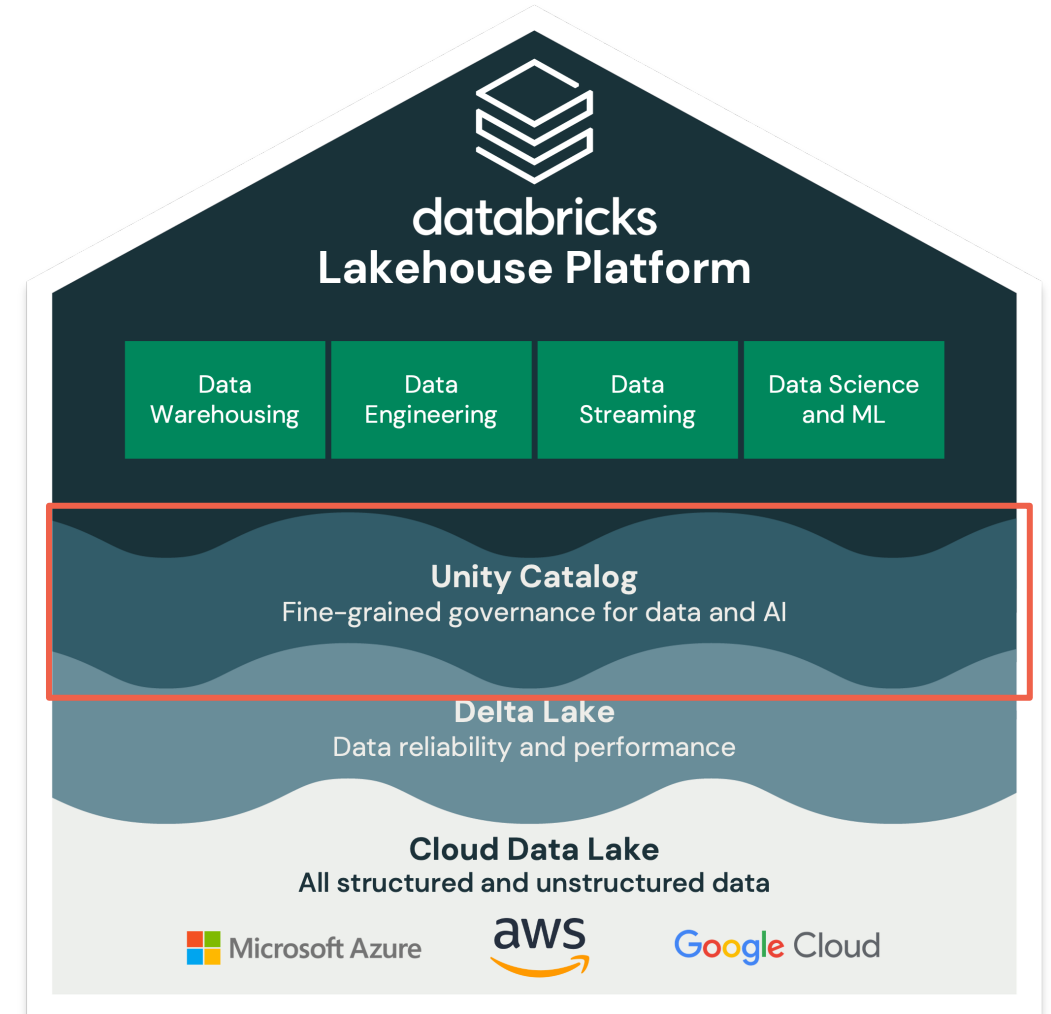
 Databricks Unity Catalog



Unity Catalog - Key Capabilities

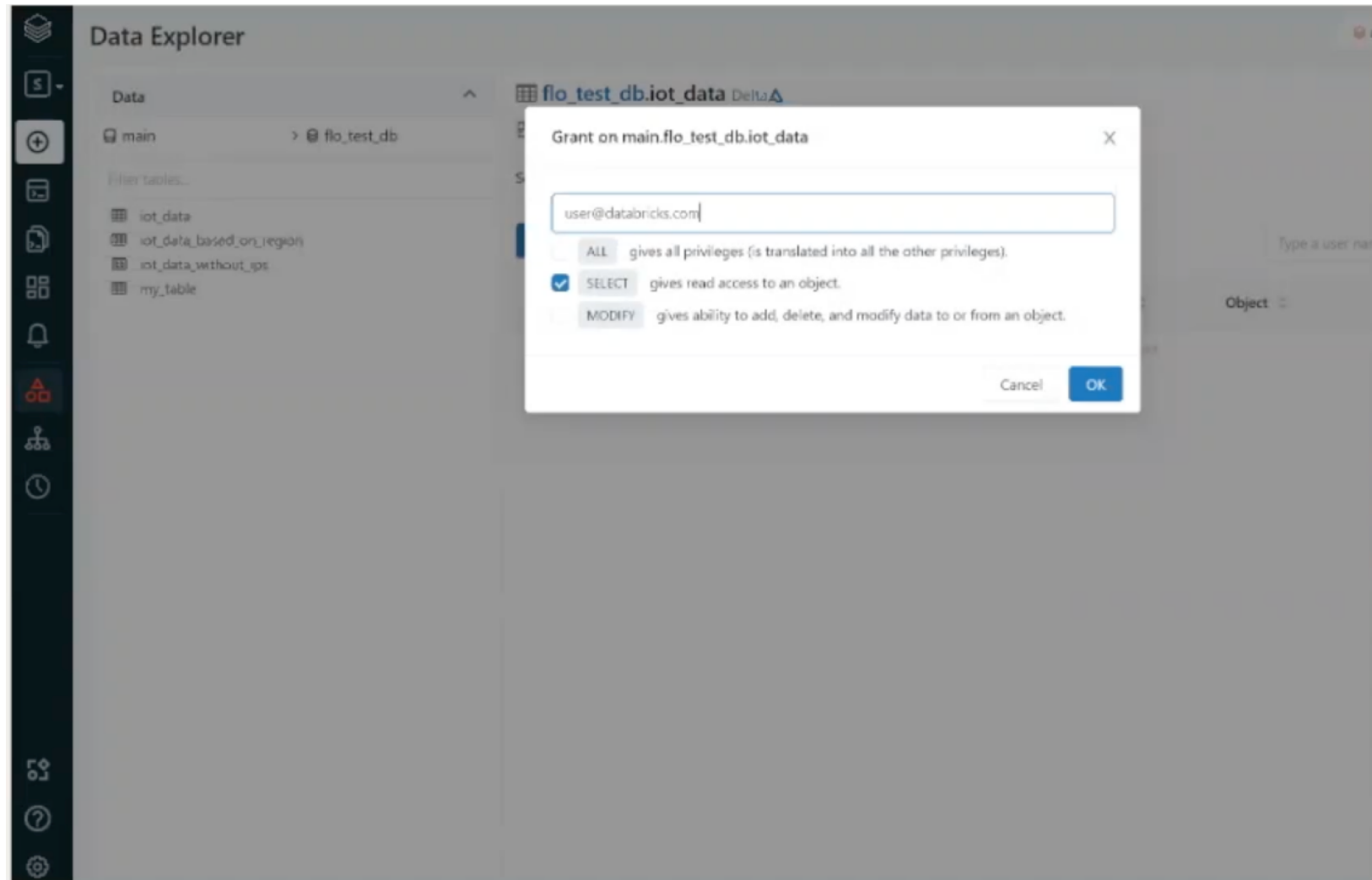
Unified governance for all data assets

- Centralized metadata and user management
- Centralized access controls
- Data lineage
- Data access auditing
- Data search and discovery
- Secure data sharing with Delta Sharing



Map, secure and audit data across clouds

- Catalog all your data, analytics and AI assets and create a unified view of you entire data estate
- Centrally manage access permissions and audit controls for files, tables across all workspaces and workloads using a familiar interface based on ANSI SQL
- Enable fine-grained access controls on rows, and columns



Open data sharing and collaboration



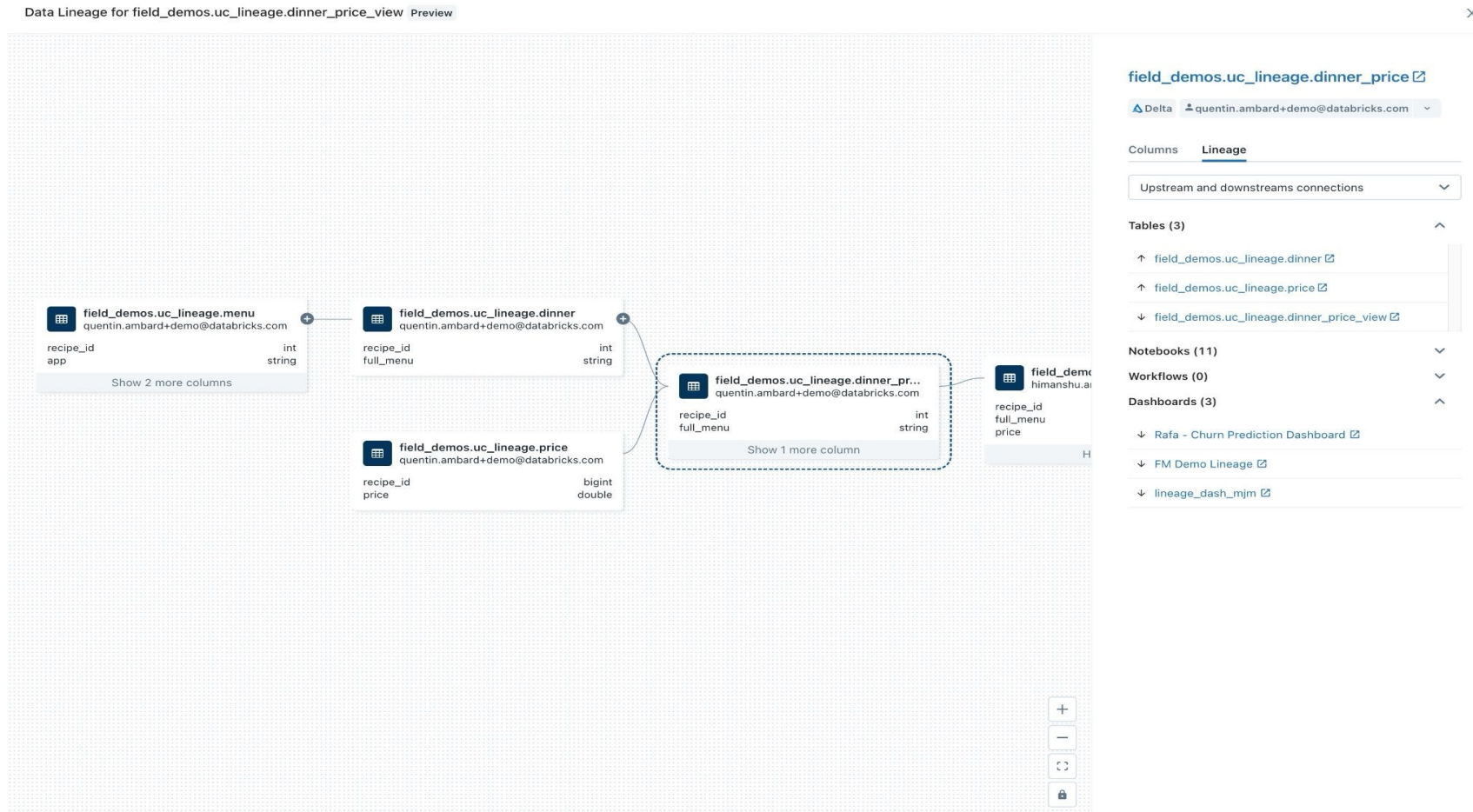
Data Discovery

The screenshot shows the Databricks Data Explorer interface. On the left, a sidebar lists data objects under 'Data Explorer unity-catalog-demo'. The main area is a search modal titled 'Search' with a 'Provide feedback' link and a close button. The search input contains 'customer' and a 'Search' button. Below the search bar are tabs for 'All', 'Tables', 'Queries', 'Dashboards', 'Alerts', and 'More'. Two dropdown menus are visible, labeled 'Catalog' and 'Schema'. The search results are listed under 'Tables' and include:

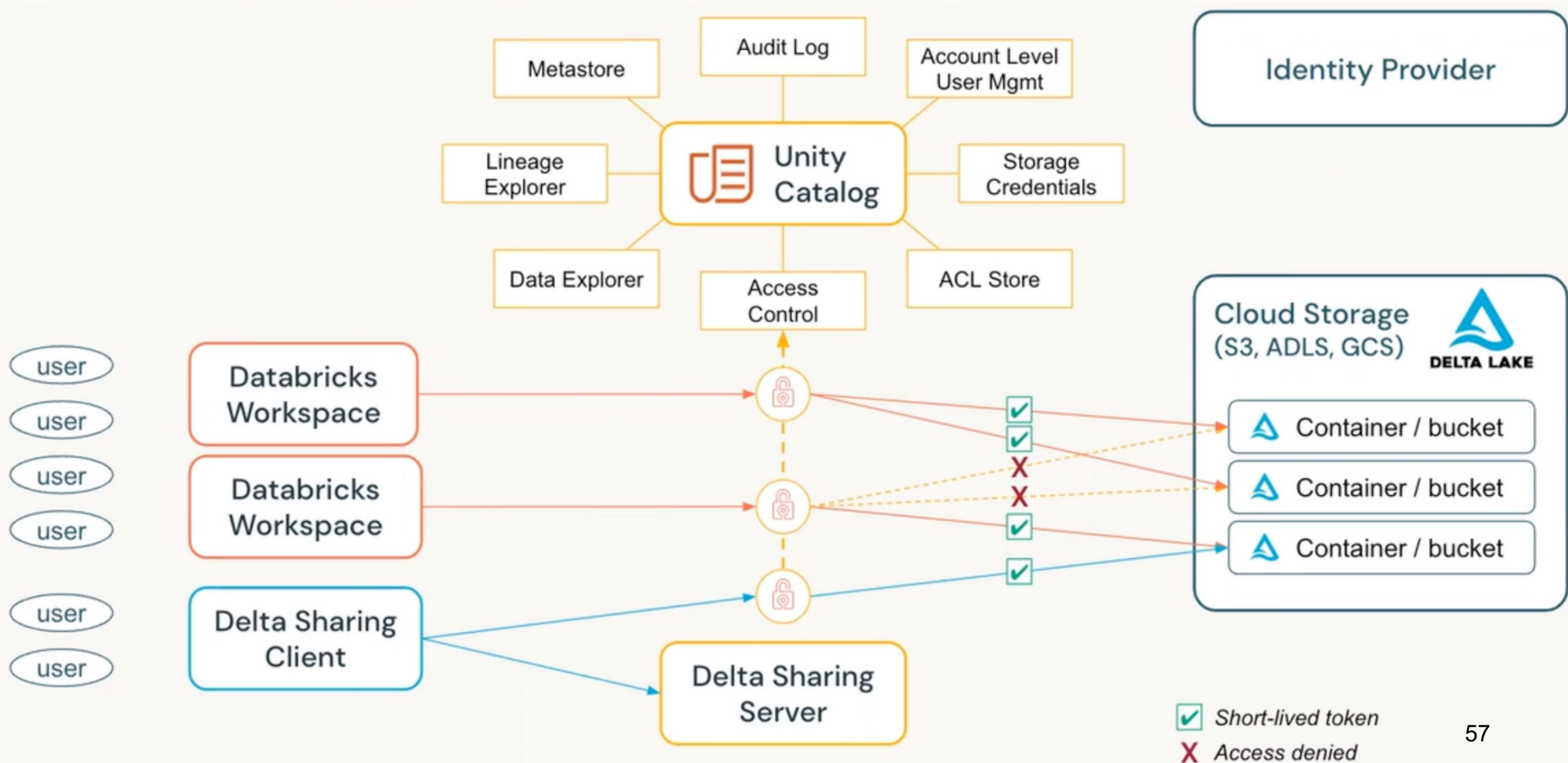
- customers2 (uc_demos_tim_stanton.uc_acl)
- customers3 (uc_demos_tim_stanton.uc_acl)
- customer (airties_bim1.tpch)
- dimcustomer (pritesh_dev._tpcdi_warehouse)

Automated lineage for all workloads

End-to-end visibility into how data flows and consumed in your organization

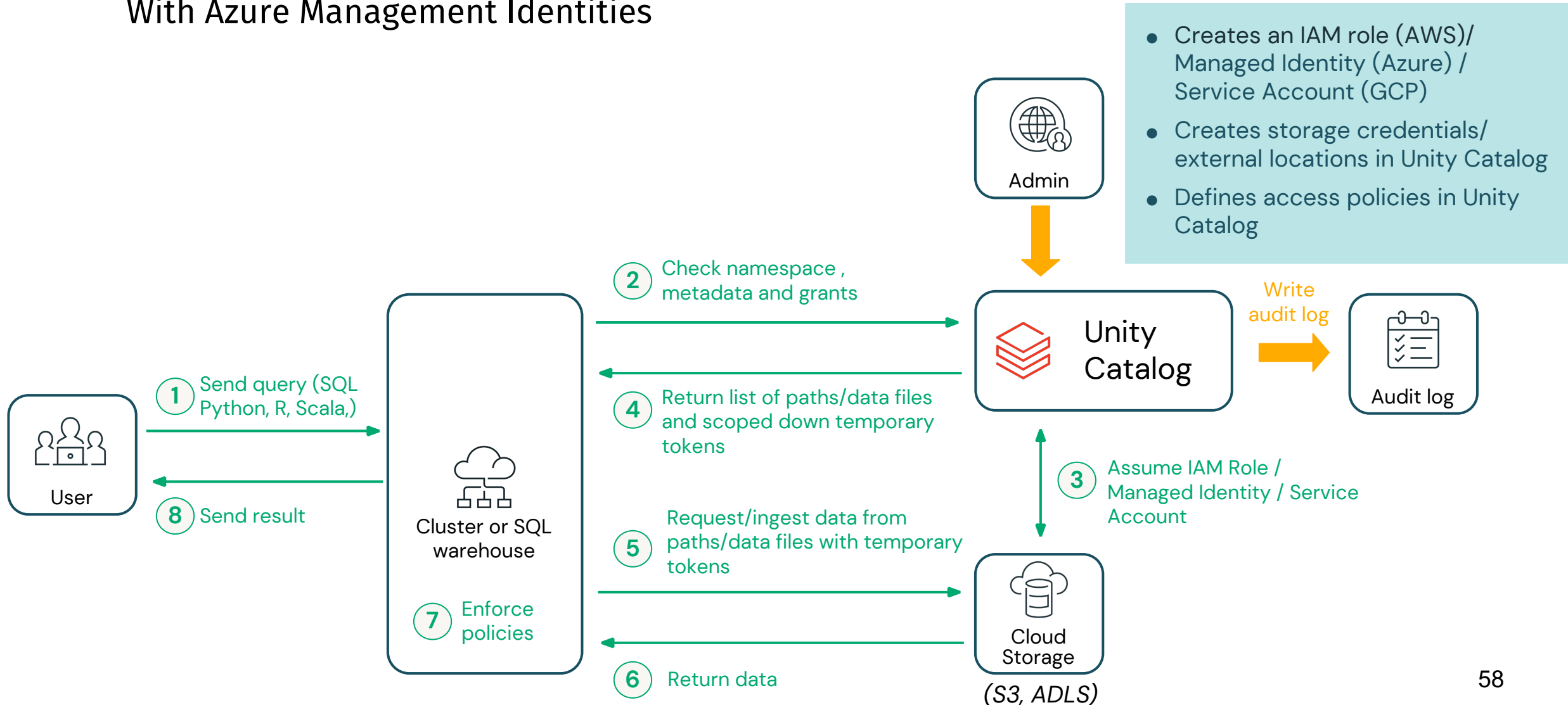


Unity Catalog - Architecture



Life of a query with Unity Catalog

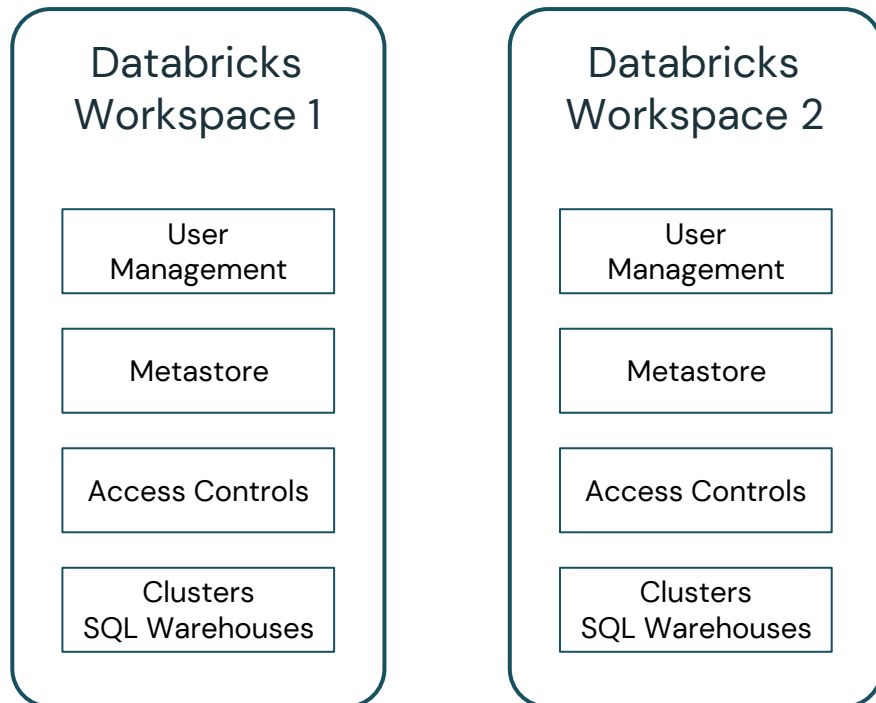
With Azure Management Identities



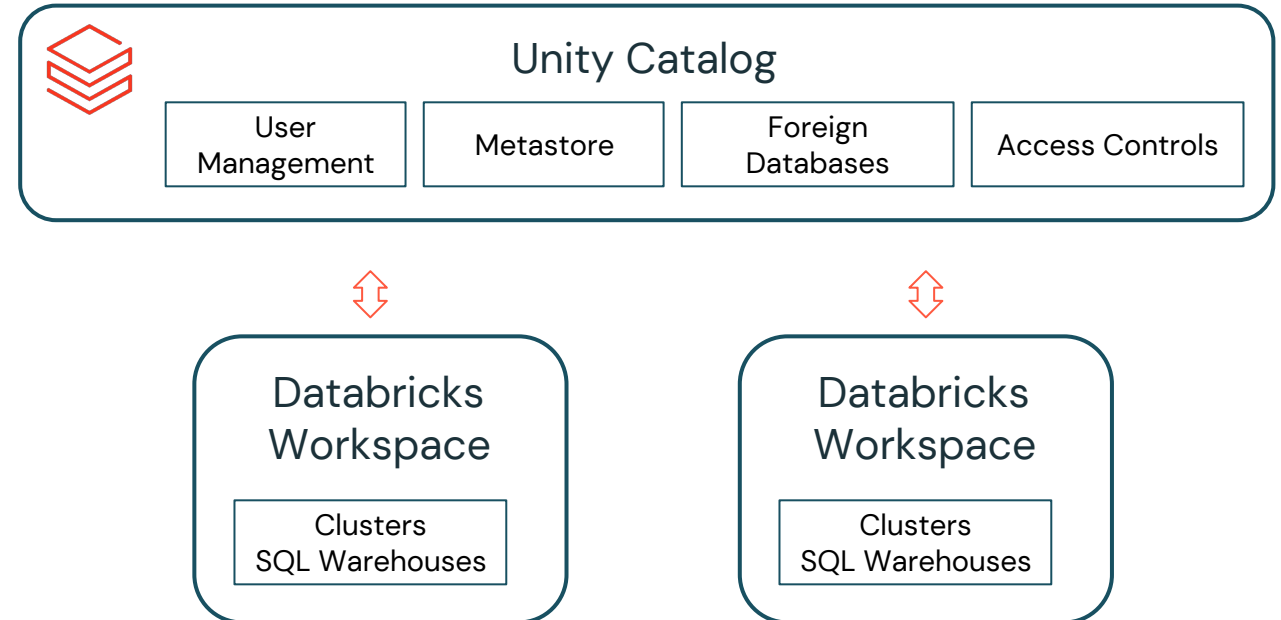
Centralized Metadata and User Management

Create a unified view of your data estate

Without Unity Catalog



With Unity Catalog

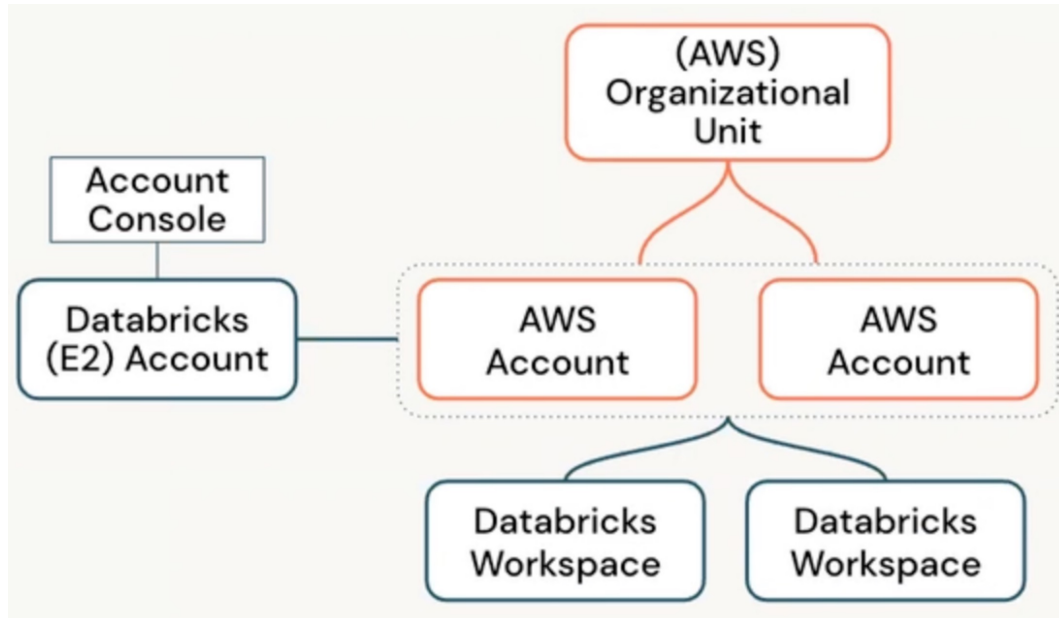


Databricks Account and the Cloud Provider Hierarchy

Create a unified view of your data estate

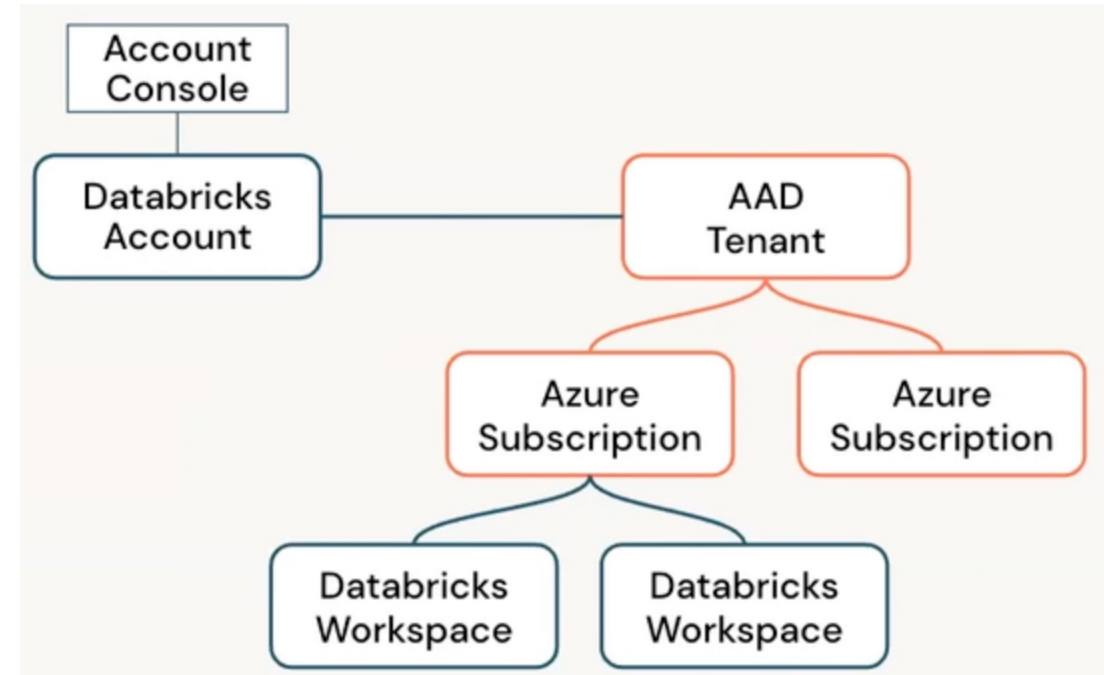
AWS

<https://accounts.cloud.databricks.com/>



Azure

<https://accounts.azure.databricks.net/>

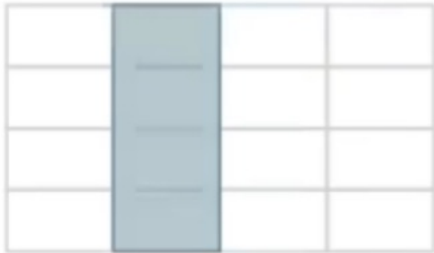


- Global Admin used initially for security

Dynamic View

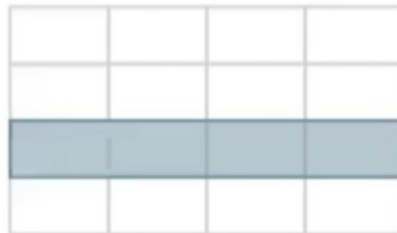
Limit access to columns

Omit column values from output



Limit access to rows

Omit rows from output



Data Masking

Obscure data

.....@databricks.com

Centralized Access Controls

Centrally grant and manage access permissions across workloads

Using ANSI SQL DCL

```
GRANT <privilege> ON <securable_type>  
<securable_name> TO `<principal>`
```

```
GRANT SELECT ON iot.events TO engineers
```

Choose
permission level

'Table'= collection of
files in S3/ADLS

Sync groups from
your identity
provider

Using UI

Data Explorer unity-catalog-demo

Catalogs > main > default >

main.default.department

Grant on main.default.department

ⓘ Users also require **USE CATALOG** and **USE SCHEMA** on the parent catalog and schema to perform actions in this table. [Learn more](#)

Users and groups

analysts x

Privileges

- SELECT** gives read access to an object
- MODIFY** gives ability to add, delete, and modify data to or from an object
- ALL PRIVILEGES** gives all privileges ⓘ

Cancel Grant

Row and Column Filtering

Fine-grained governance for the Lakehouse

Problem

Managing fine-grained **access controls on rows and columns** in tables is critical to ensure data security and meet compliance

Solution

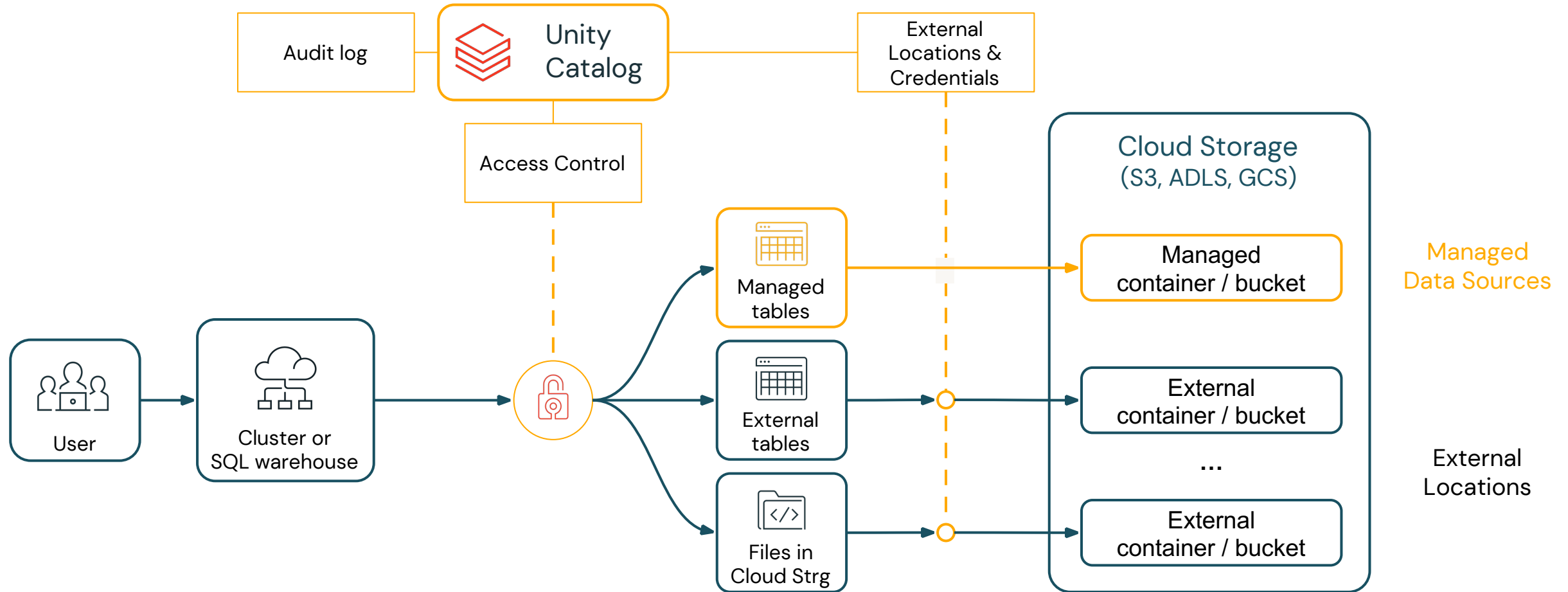
With Unity Catalog, you can use standard SQL functions to define **row filters** and **column masks**, allowing fine-grained access controls on rows and columns

```
// Row filtering  
  
CREATE FUNCTION us_filter(region STRING)  
RETURNS BOOLEAN  
  RETURN if(is_member('admin'), true, region='US')  
  
ALTER TABLE sales  
SET ROW FILTER us_filter ON (region)
```

```
// Column masking  
  
CREATE FUNCTION ssn_mask(ssn STRING)  
RETURNS STRING  
  RETURN if(is_member('admin'), ssn, '****')  
  
ALTER TABLE users  
ALTER COLUMN ssn SET MASK ssn_mask
```

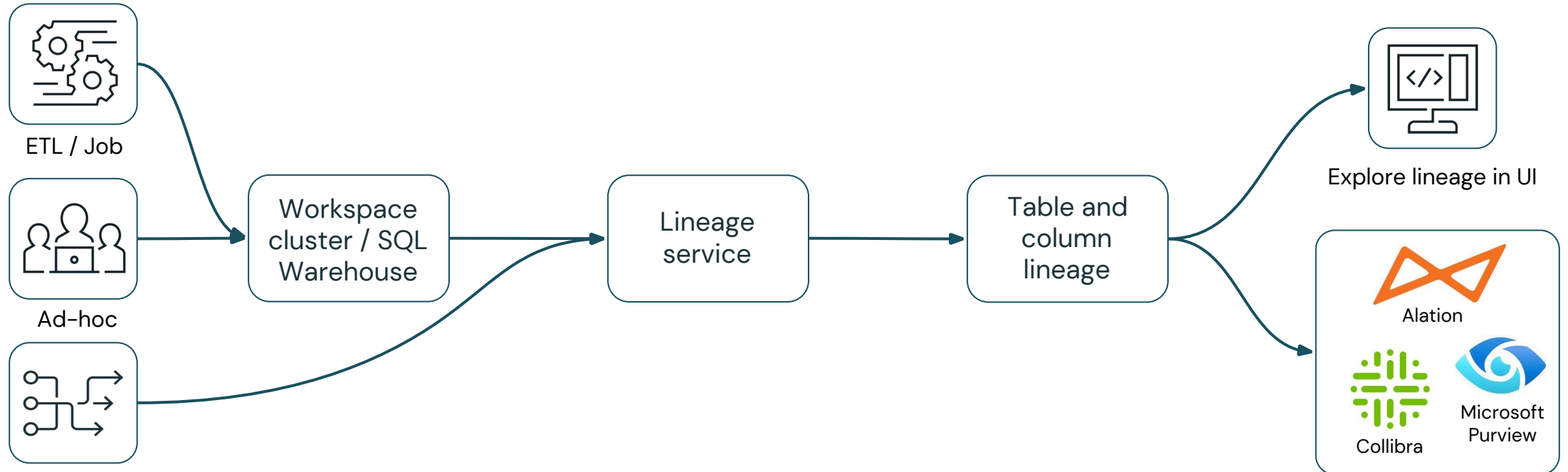
Manage Data Sources & External Locations

Simplify data access management across clouds



Data Lineage – How it works in Databricks

End-to-end visibility into how data flows and consumed in your organization



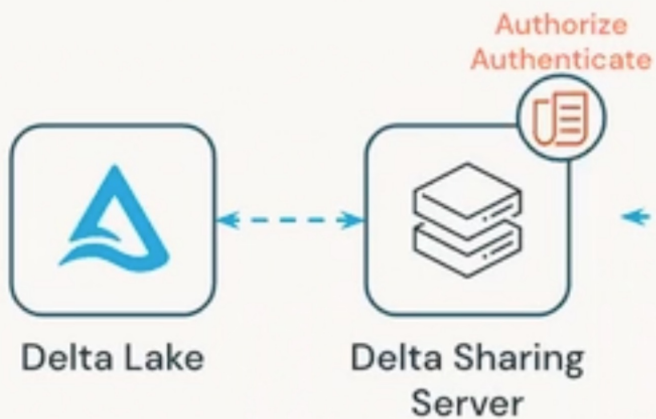
Code (any language) is submitted to a cluster or SQL warehouse or DLT* executes data flow


- Lineage service analyzes logs emitted from the cluster, and pulls metadata from DLT
- Assembles column and table level lineage

- Presented to the end user graphically in Databricks
- Lineage can be exported via API and imported into other 66 tool

Delta Sharing

Data Provider



 Delta Sharing protocol

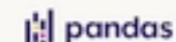
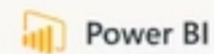
No replication
Easy to manage
Secure

Data Recipient

Any use case



Any platform



...

Any cloud

Google Cloud



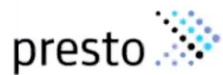
Microsoft Azure

On-Premise

Delta Sharing Ecosystem

Endorsed by many of Databricks partners with integration and connectors being developed

Open Source Clients



Commercial Clients

Business Intelligence



Analytics



Governance



Data Providers



Conclusion



Security, Compliance, and Privacy



Capability	Databricks	Snowflake
Automated Account Management	✓	✓
Role-Based Access Control (RBAC)	✓	✓
Multi-Factor Authentication (MFA)	✓	✓
Identity Federation	✓	✓
Data Classification and Encryption	✓	✓
Tokenization and Access Control for Sensitive Data <i>with partner integrations)</i>	✓	✓
Advanced Firewall and Intrusion Detection Systems (IDS)	✓	✓

Security, Compliance, and Privacy continues



Capability		
Endpoint Protection		
Clarity in Roles and Responsibilities		
Compliance Documentation and Assurance		
Automated Compliance Tools		
Data Localization and PII Management		
Automated Security Scanning		
Security Information and Event Management (SIEM)		

Data Governance



Capability		
Centralized Data Catalog		
Data Lifecycle Management		
Data Integration and Quality Tools		
Centralized Access Control		
Audit and Monitoring System		
Data Masking and Encryption		

Data Governance continues



Capability		
Data Quality Framework		
Automated Data Quality Tools		
Open and Secure Exchange Technology		
Real-time Data Sharing Platforms		
Data Usage Agreements and Compliance		

A teal shield-shaped graphic with a purple bookmark icon at the top right. The text "Thank you!" and "Q&A" is centered in white.

Thank you!
Q&A

Learn More



- [Security and Trust Center – Databricks](#)
- [Data governance guide | Databricks](#)
- [Security and compliance guide | Databricks](#)



- [Snowflake Security Overview and Best Practices](#)
- [Securing Snowflake | Snowflake Documentation](#)
- [Snowflake Security 101](#)