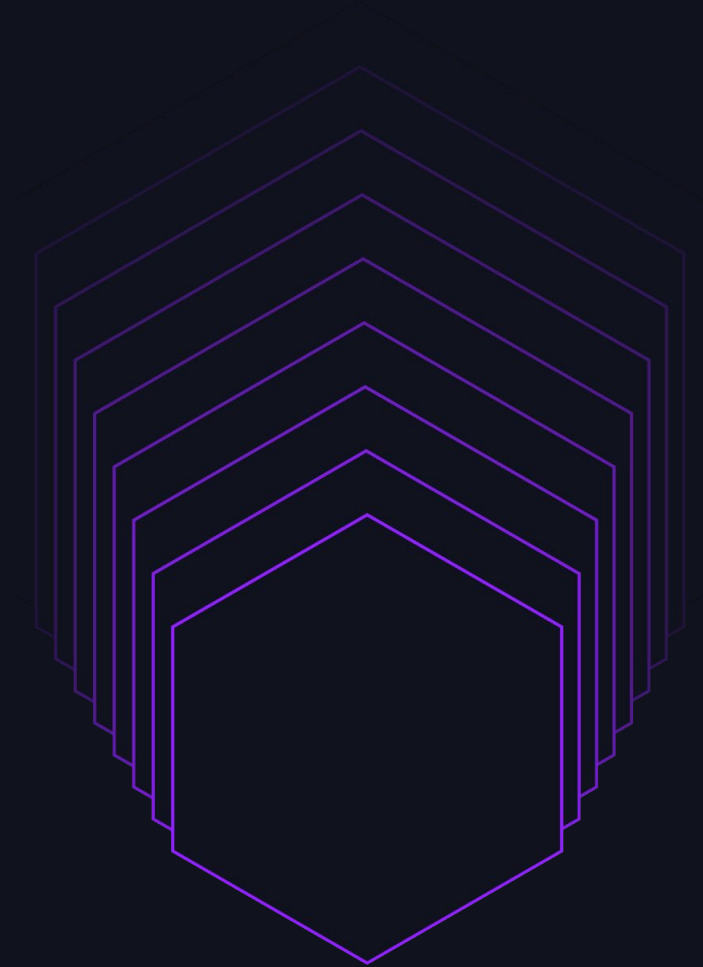


UNITY CATALOG MIGRATION IN EUROPE'S TOP SUPERMARKET CHAIN

Enrique Catalá | Karthik Subbarao
June 2024



Agenda

- Introducing Lidl loyalty program
- Motivation to migrate
- The journey towards Unity Catalog
- Outcome of the Migration
- UC Migration @ Enterprise Scale

Enrique Catalá Bañuls

Schwarz Global Services Barcelona

- ❑ Data Architect at SCRM Lidl International Hub
- ❑ 20+ years in the industry
- ❑ 10-year tenure as Microsoft Data Platform MVP
- ❑  enriquecatala



Karthikeya Sampa Subbarao

Databricks EMEA

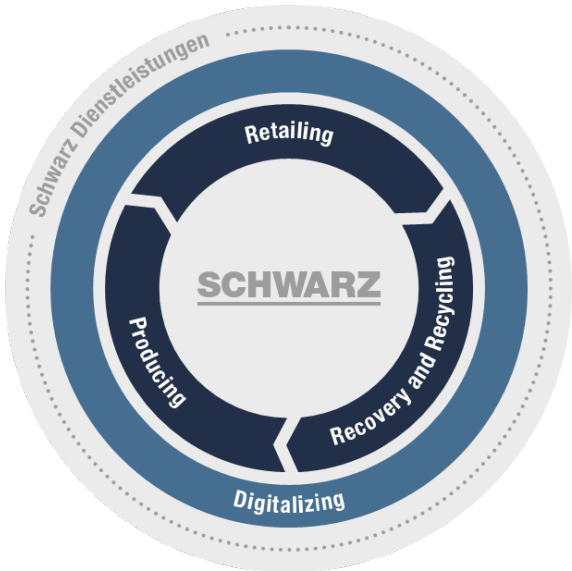
- ❑ Specialist Solutions Architect @ Databricks
- ❑ Experienced Software Engineer & Architect
- ❑ 10+ years in the Tech industry
- ❑ Data Architecture, Security & Governance SME



INTRODUCING LIDL LOYALTY PROGRAM

Schwarz Group

worldwide



Retail divisions comparison

Revenue rankings in fiscal year 2022

Top 5 worldwide

Rank	Company	Gross sales in billion EUR
1	Walmart	557.9
2	Amazon	240.5
3	Costco	221.7
4	Lidl and Kaufland	164.4
5	Kroger	141.7

Top 5 European-wide

Rank	Company	Gross sales in billion EUR
1	Lidl and Kaufland	161.2
2	Aldi	80.4
3	Amazon	66.2
4	Carrefour	60.1
5	Edeka	59.7

Source: Edge by Ascential; as of June 2023

Schwarz Group

Retail group as pioneer and pacemaker

Over

575,000

employees world-wide

of these, 204,000 are in Germany

Round

6.8 billion

store customer

at Lidl and Kaufland

Represented in 32 countries,
Lidl and Kaufland together have
round



13,700

stores

The companies of Schwarz Group achieved more than



154.1 billion euros

sales* in fiscal year 2022

**Net sales in fiscal year 2022 (includes all sales under commercial law incl. other revenues)*

Lidl



SCRM Lidl International Hub

Generate competitive advantages in a challenging market environment



Lidl Plus Today

12.000+

Stores



31

Countries



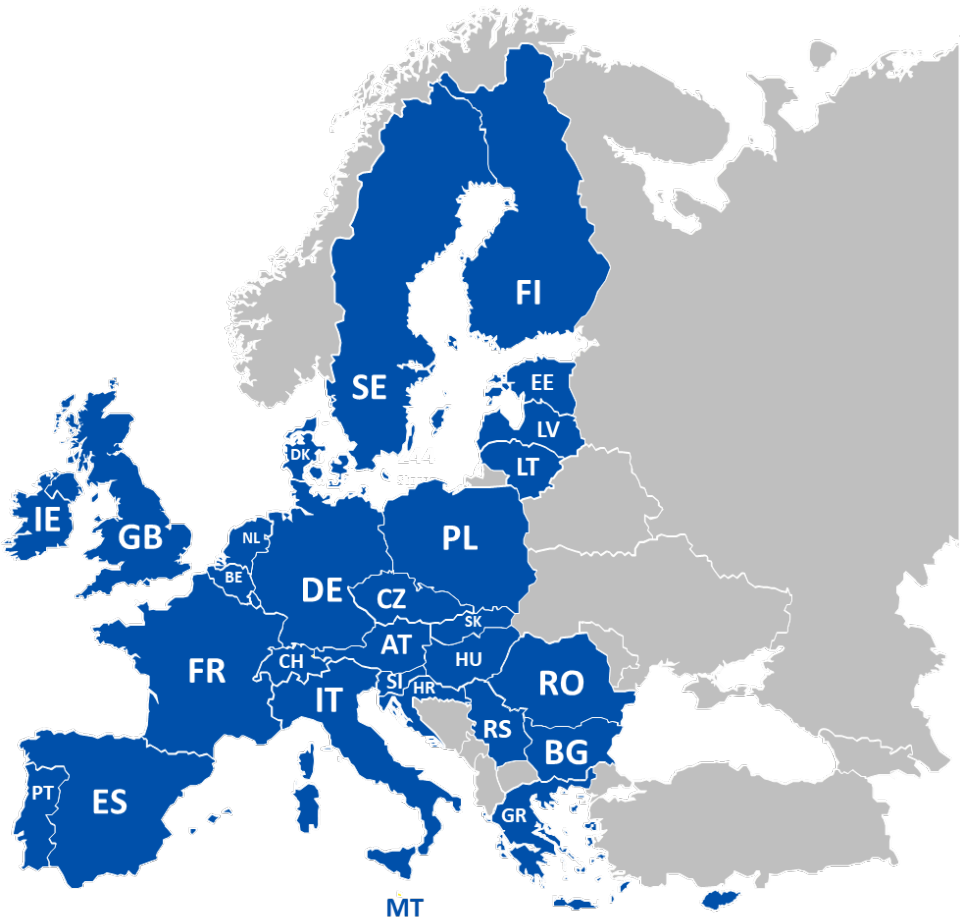
1T+

Daily Lidl Plus Events



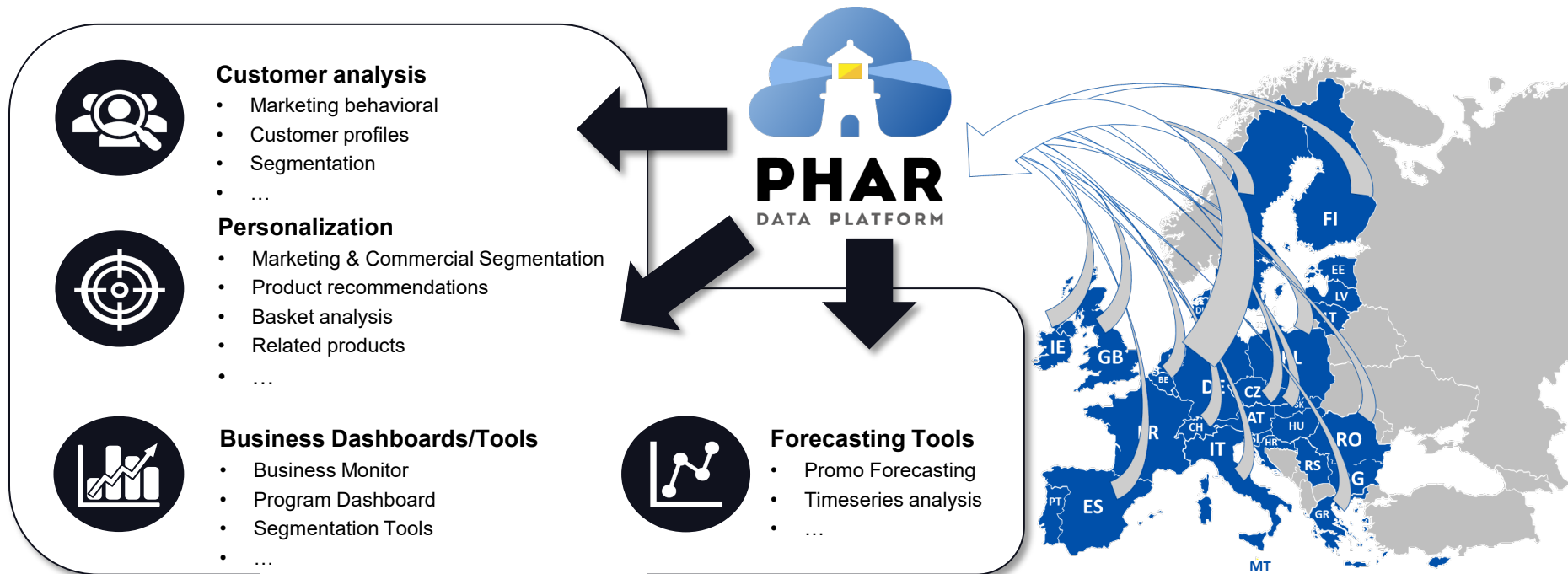
Pb

Petabyte scale



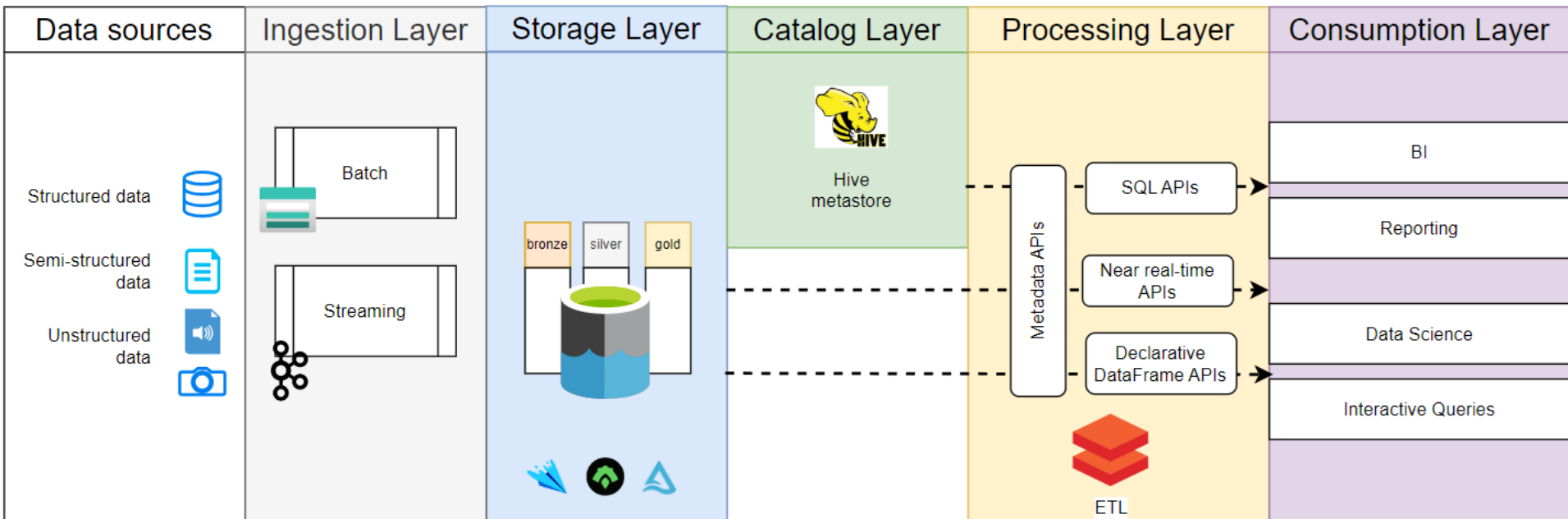
Lidl Plus Today

Phar data platform



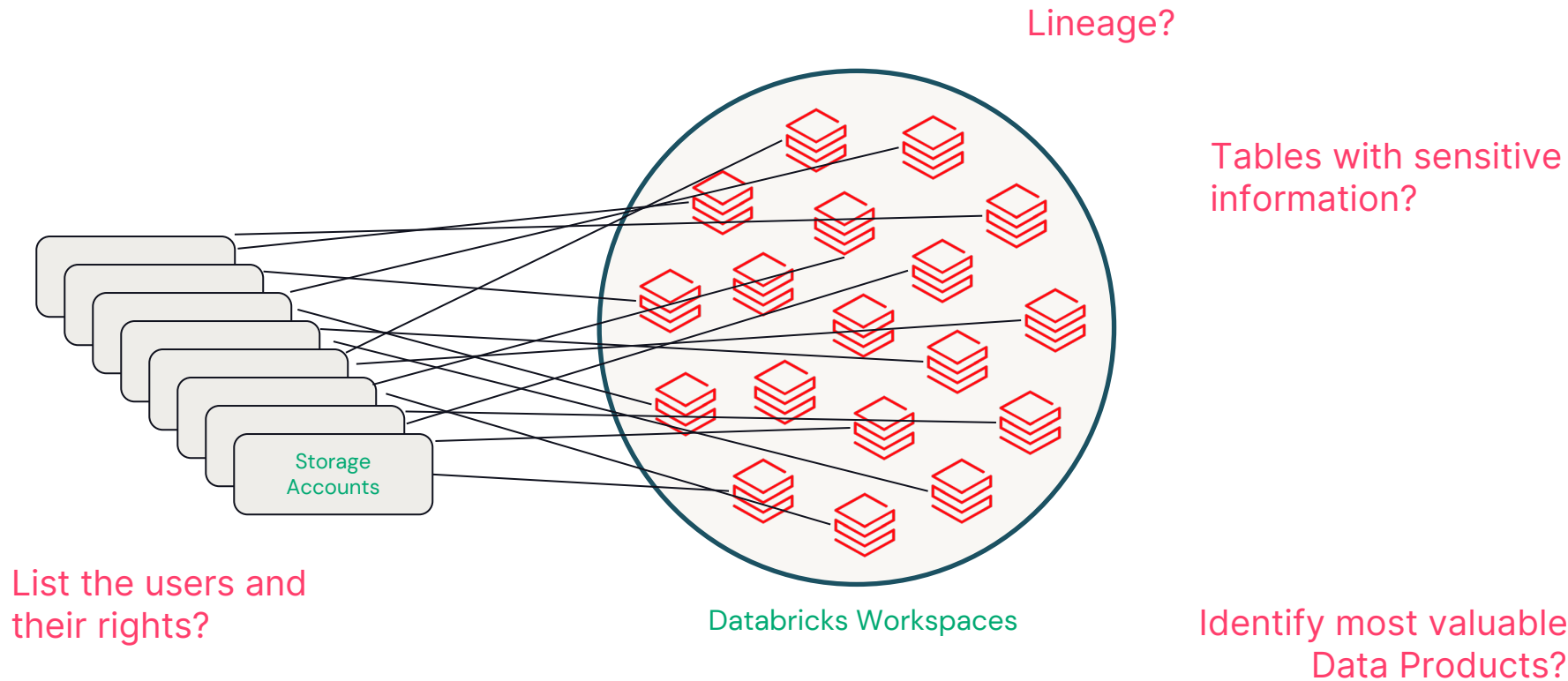
Phar architecture

How it started

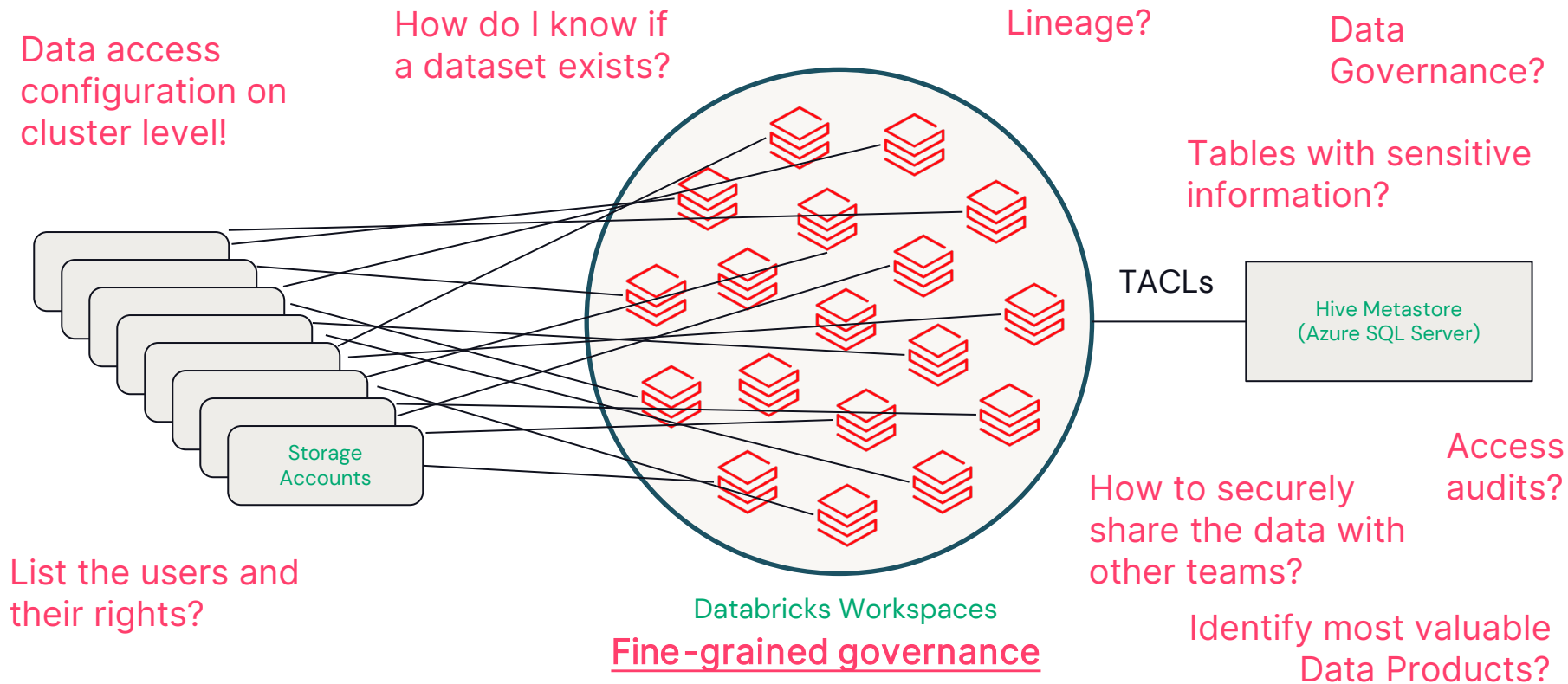


MOTIVATION TO MIGRATE

Governance Challenges Pre-UC

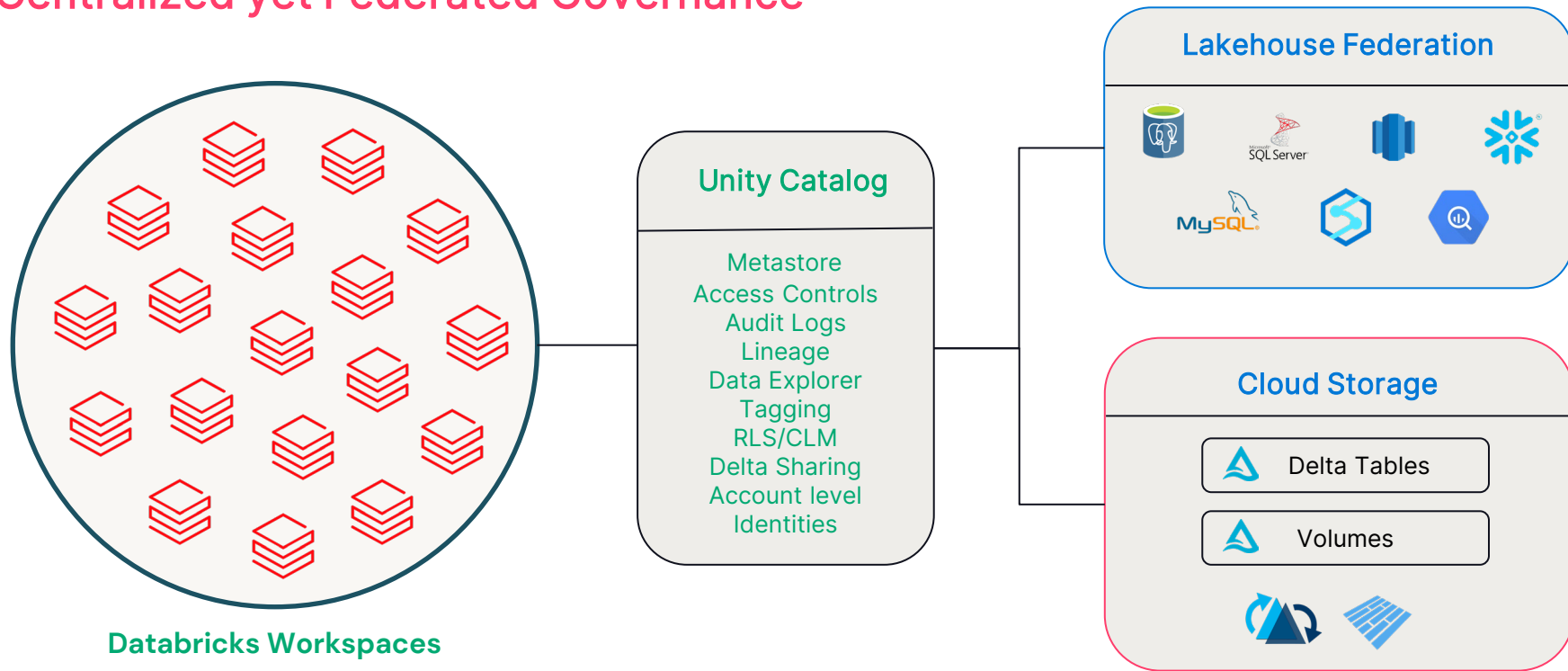


Governance Challenges Pre-UC



Simplified Governance with Unity Catalog

Centralized yet Federated Governance



UNITY CATALOG IS
NOT JUST A
CATALOG BUT IS
AN EVOLUTION OF
DATABRICKS

THE JOURNEY TOWARDS UNITY CATALOG

UC Upgrade Approach

Supported by Databricks



Assessment

- Analyze the existing Databricks setup and document
- A rough estimation of the upgrade effort



Feasibility Testing

- Identify potential issues and edge cases
- Find and document mitigation strategies



Designing and Scripting

- Design the target architecture and document it in detail
- Develop/Update scripts including automation scripts, ETL pipelines, etc.



Migrate Upstream Tables and Workloads

- Migrate Bronze and Silver tables to UC (keep them up to date) but retain the HMS tables until all the dependent downstream systems are upgraded to UC



Migrate Downstream Systems

- Gradual migration of gold tables, Dashboards, etc.



Decommission

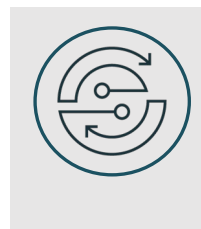
- Decommission HMS, non-UC workloads, clusters, etc.

UC Upgrade Process

Global setup (Org level)

For each team

For each workspace



Accounts Console setup

- Activate the Account Console
- Provision Identities at the Account Level

Architecture design

- Design the Catalog structure
- Design the Data Access Controls

Setup the Metastore

- Optionally setup the Metastore storage account and access connector
- Create the Metastore

Setup the Catalogs

- Create Storage Credentials and External Locations
- Create Catalogs and Schemas
- Setup the default Data Access Controls

Code changes

- Develop/Update orchestration scripts and libs
- Develop/Update infrastructure automation scripts

Enable the Workspace(s)

- Add the workspace to UC Metastore
- Assign users and groups to the workspace

Upgrade tables and workloads

- Upgrade tables to UC
- Upgrade clusters
- Upgrade workloads

Maintenance windows



UC Design Decisions

The decisions were focused on easier migration

One catalog per environment
i.e., phar_dev, phar_stg, & phar_prd

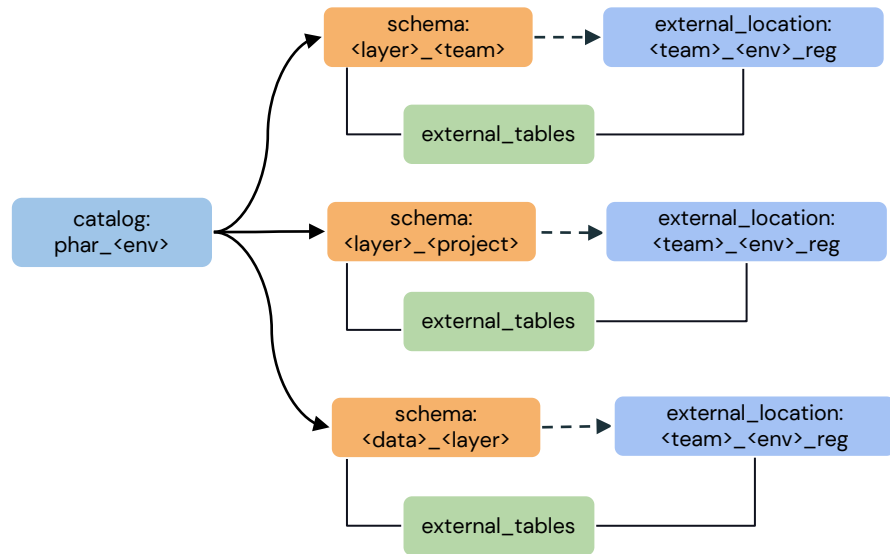
Use only External Tables

Managed location at the Schema level

One access connector per storage account

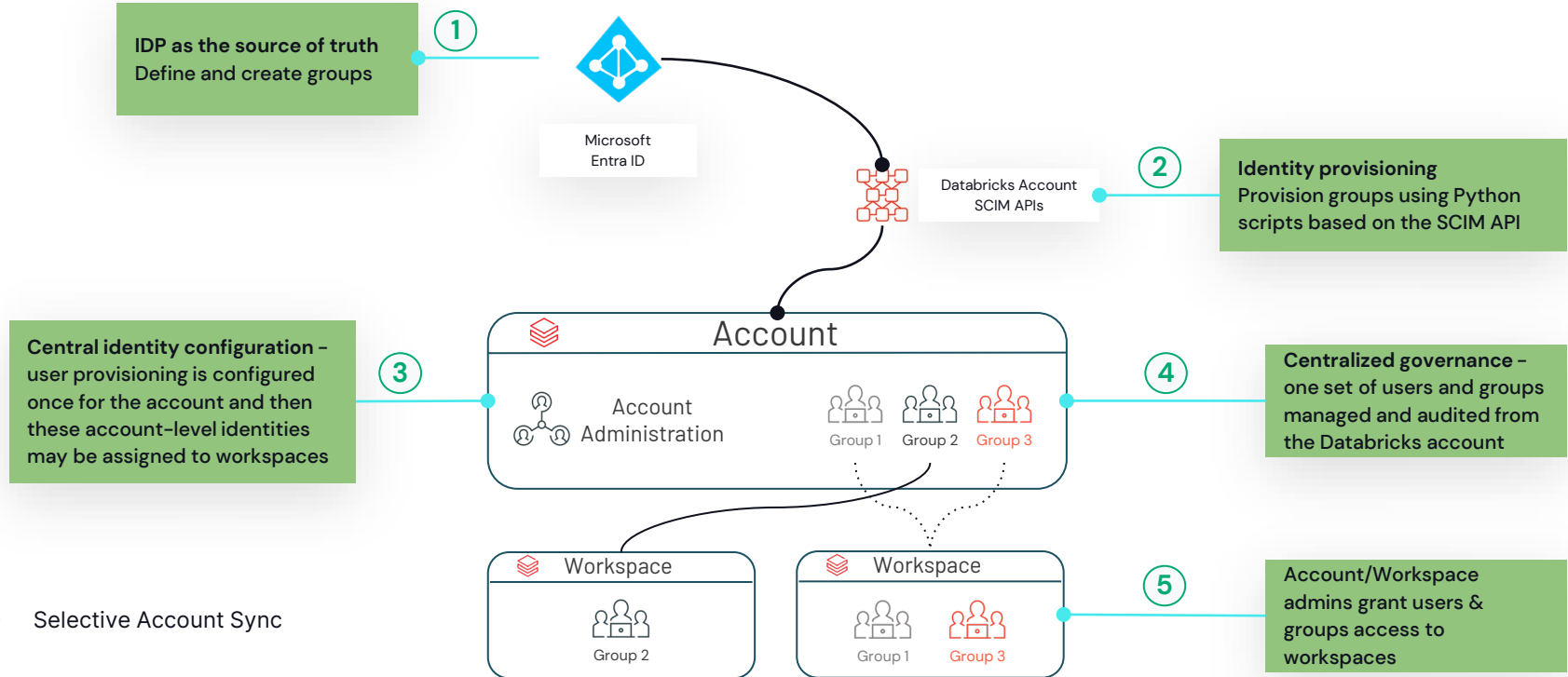
Data Platform team will own all the catalogs

The SP that creates the Catalog is set as its owner and the 'USE CATALOG' privilege is assigned to 'account users'



Identity Upgrade

Upgrading the identities to account level



Selective Account Sync

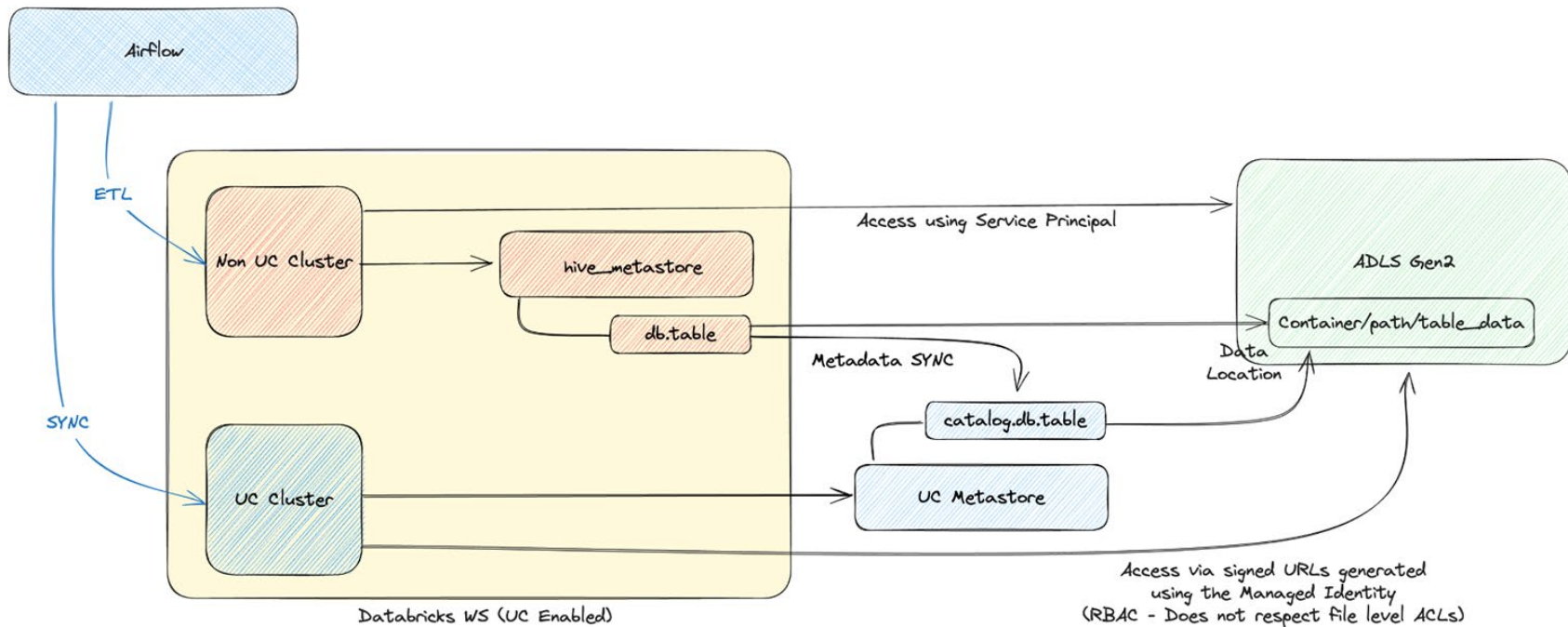


Test using Test Azure
Tenant / Databricks Account



Table & ETL Pipeline Upgrade

Data accessible via HMS as well as UC

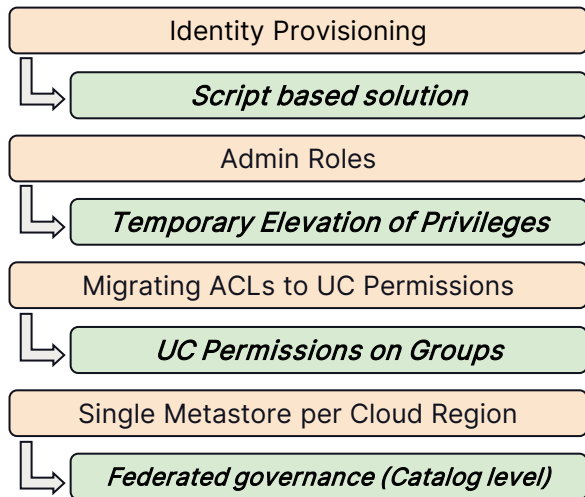


UC Migration Challenges

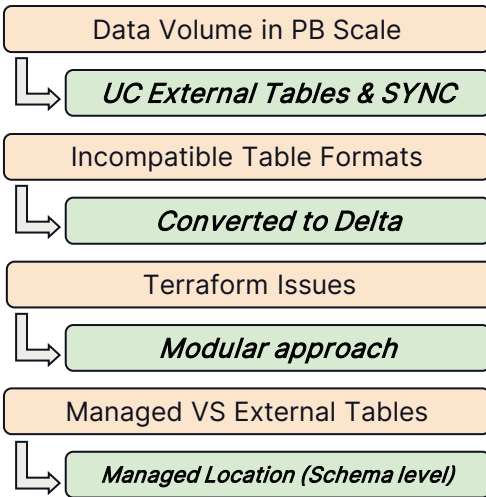
Migrating to UC is not without challenges



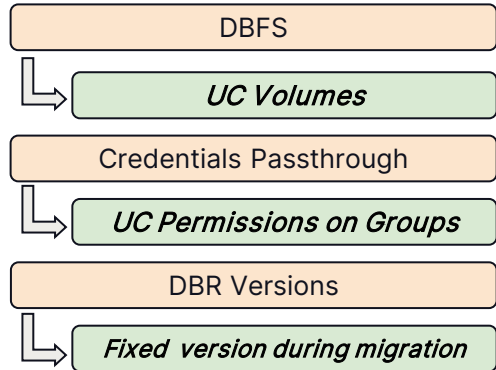
Identity and Access Management



Data and Transformation



Feature Deprecation

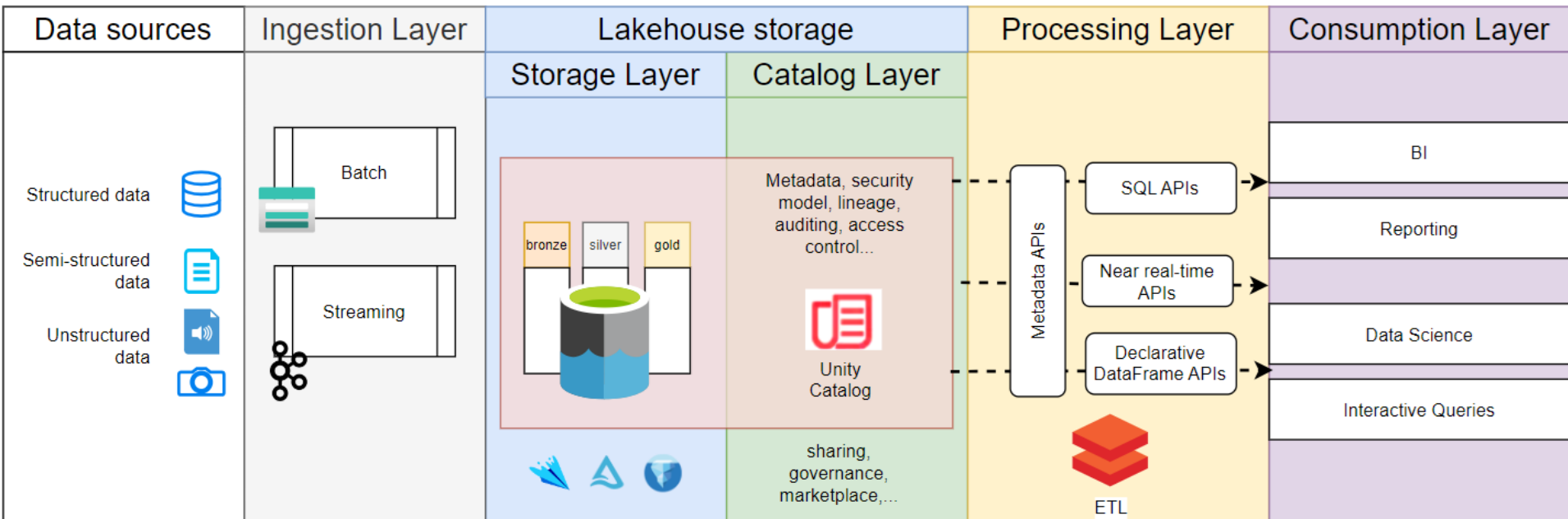


MOVING TO UNITY CATALOG CAN BE A GRADUAL PROCESS

THE OUTCOME OF UC MIGRATION

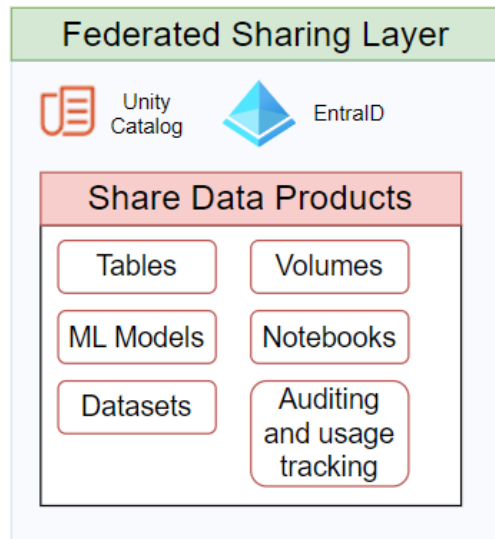
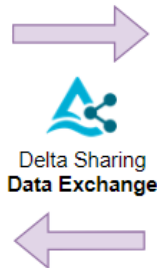
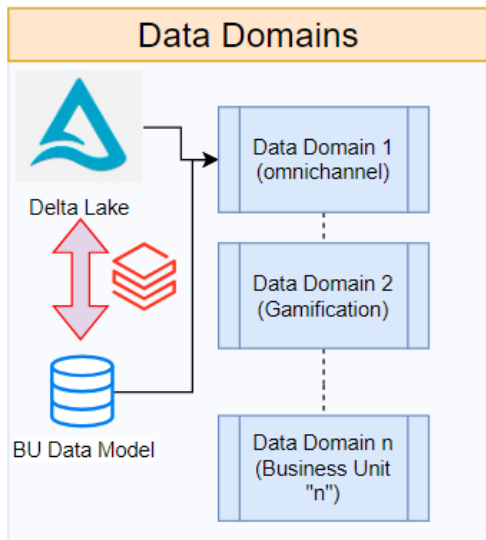
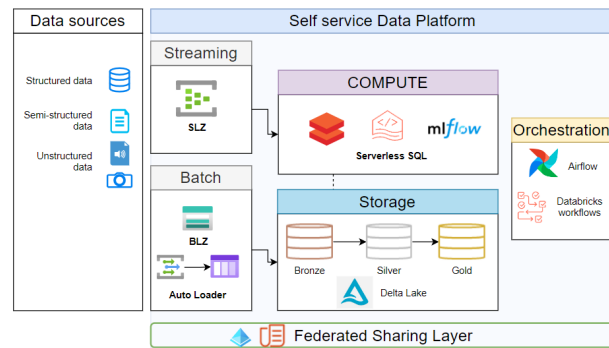
Phar architecture

How is it going



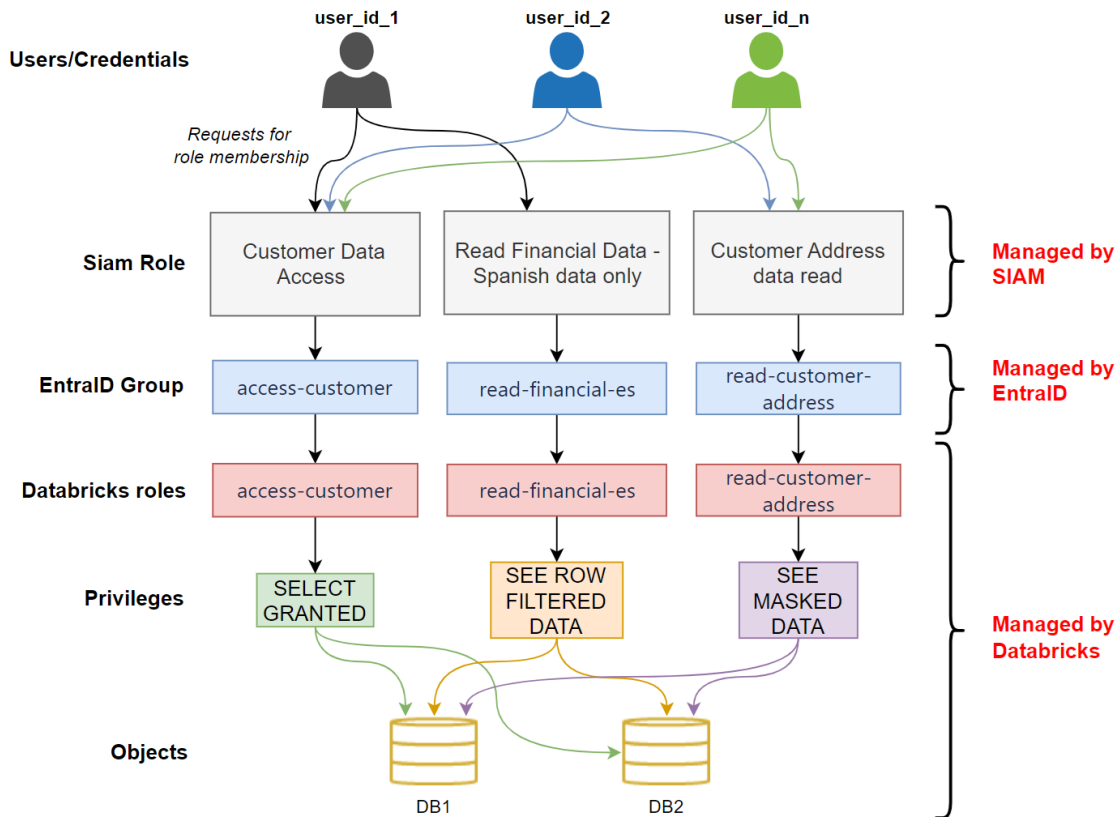
Use cases unlocked by UC

Federated sharing layer



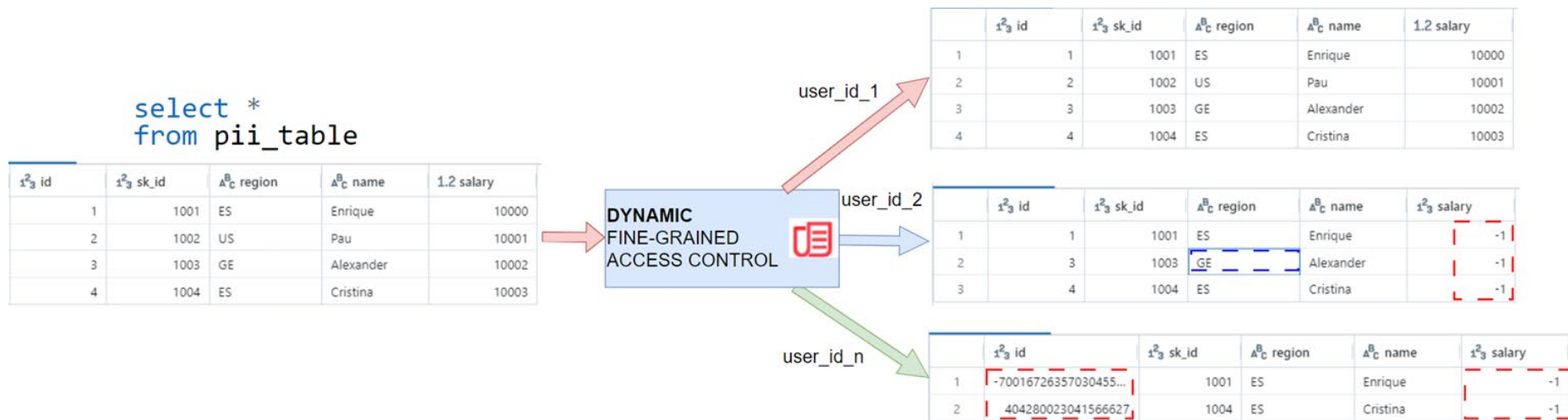
Use case: Sensitive Data management

Zero trust access management



Use case: Sensitive Data management

Fine-grained access control



Use case: Sensitive Data management

Federated governance model



Security & Access Control

- Fine-grained access control
- Zero trust model
- Data access democratization



Integration & Compliance

- Microsoft EntraID integration
- Multi-Cloud Management tool
- Security Protocols
 - Time-bound approvals
 - Auditing
 - ...

Additional benefits of moving to UC

Migrating to UC tends to reduce the technical debt



Reduction in Total Cost of Ownership (TCO)

- Reduction in operations overhead (~85%)
- Observability over the Data Product usage
- Ability to use Serverless further reduced the TCO
- Noticed up to 50% cost reduction in some edge scenarios



Improvements in Architecture

- Reduced number of workspaces
- Performance improvements w.r.t. Metastore access
- Improved dependence management using Volumes
- Reduction in number of zombie tables and files

UC MIGRATION @ ENTERPRISE SCALE

Unity Catalog Migration Readiness

Unity Catalog Migration requires change management

Assessment

- Do the assessment of the current Databricks setup (Now you can leverage [UCX](#))
- Document current setup (# of workspaces, tables etc)
- Identify and prioritize the use cases to migrate

Team Planning

- Dedicate a team for the UC Migration project
- Plan and coordinate with Databricks (Account Team, SSA, Professional Services)
- Work with partners to scale the efforts (if needed)

Internal Alignment

- Align with other business units and agree upon the account level configurations (Account Admin role, Metastore selection, SCIM Integration strategy)
- Agree upon the ownership of common assets
- Align with the infrastructure/cloud teams

Timelines

- Discuss and decide on the upgrade timelines
- Remember: **UC migration can be a gradual process!**
- Agree upon when and how to communicate the upgrade timelines with the respective stakeholders

THANK YOU

