

Accelerating
**Unity Catalog
migration with
UCX & UCMX**



UCX & UC-MX – Accelerate Unity Catalog Migration

What

Purpose

What is Unity Catalog (UC)?

UC is a unified governance solution for data & AI governance on Databricks platform

What is UC Migration ?

Migration of Databricks objects from Hive Metastore to Unity Catalog

How

Process

How is Unity Catalog Migration executed ?

UC migration is a four-step process: Plan, Prepare, Execute, and Validate

It uses UCX (Databricks Accelerator) and UC-MX (Capgemini Accelerator)

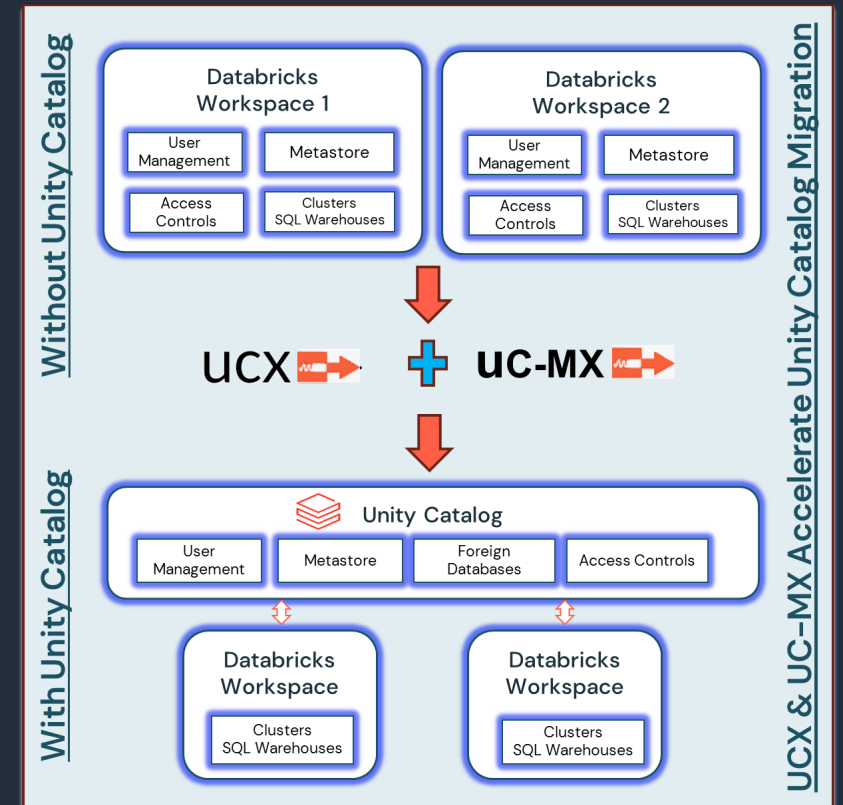
Why

Result

Why use UCX and UC-MX ?

It enables faster go-to-market and metadata-driven scalable solution

UCX and UC-MX accelerate UC migration and reduces migration efforts by ~30%





UCX & UC-MX – Value Add and Architecture

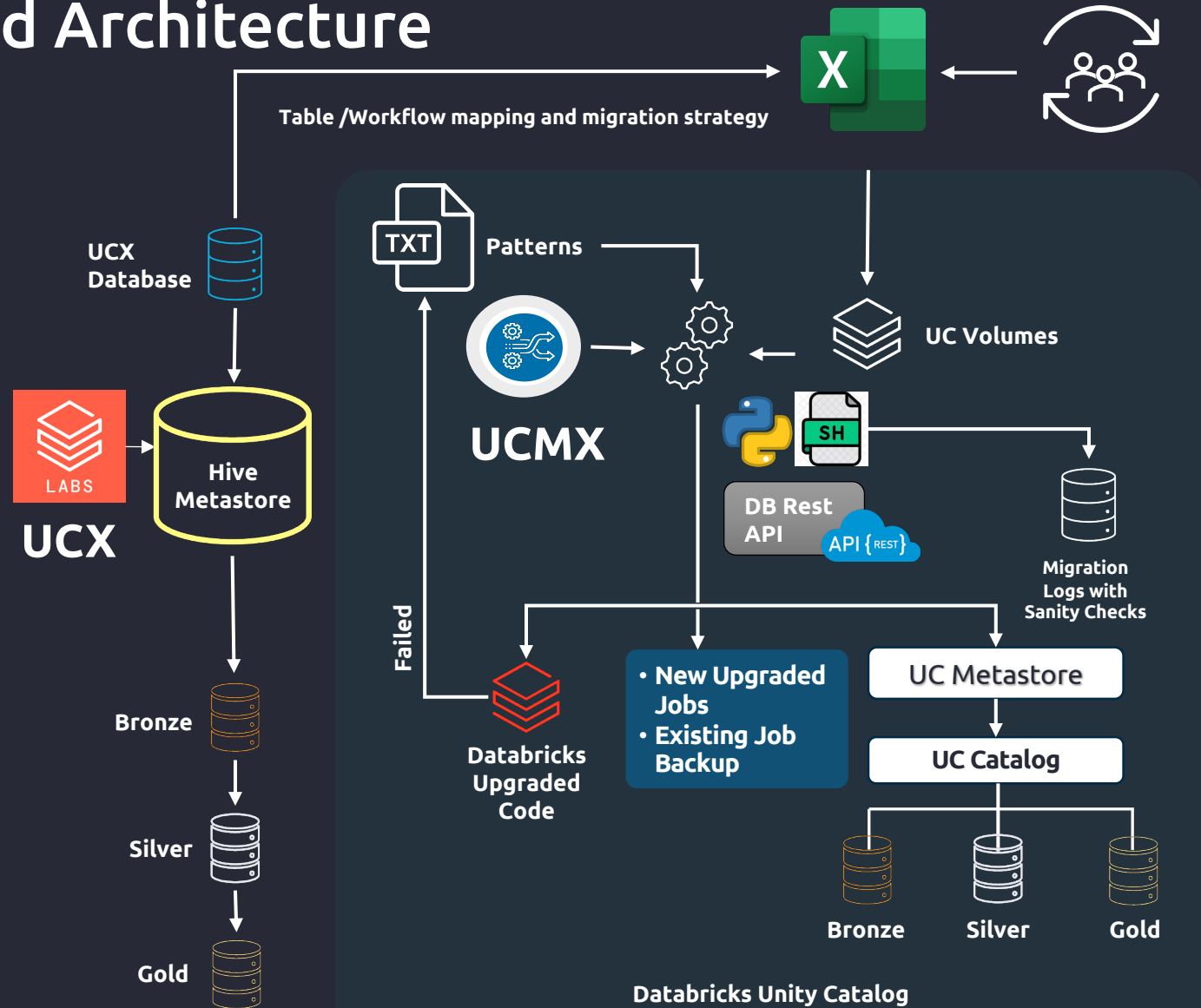
1 Easy Metadata driven Migration
Its a metadata-driven solution that accelerates upgrades

2 More than 25% Effort reduction
facilitating the parallel migration of all workflows without manual interventions.

3 Migration Status and easy Rollbacks
Logs comments, migration status, and back ups the existing jobs for easy rollbacks

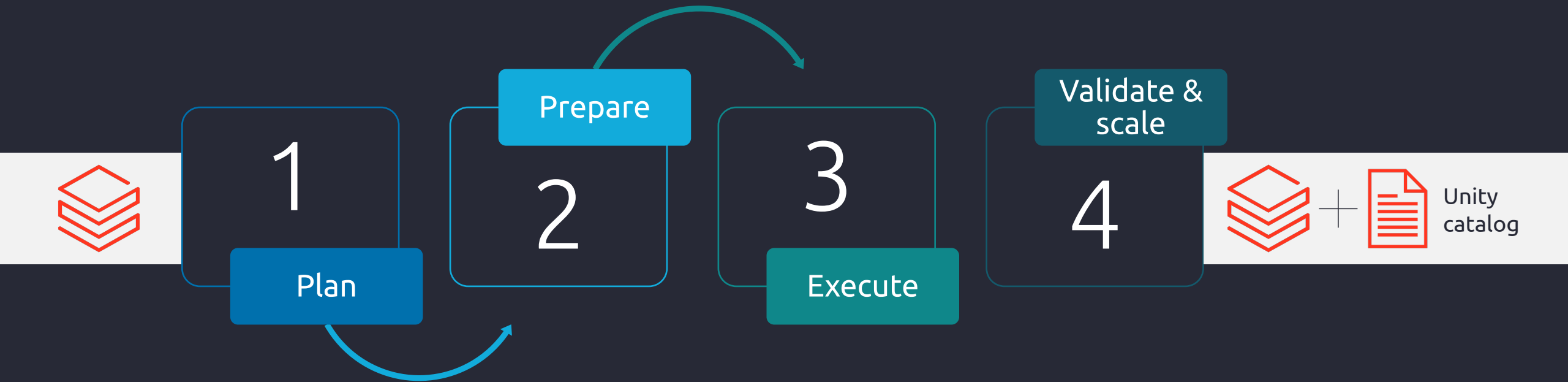
4 Single Solution for Pyspark and SQL
seamlessly works with both Python and SQL-based notebooks, utilizing Pyspark and Spark SQL.

5 Can be easily enhanced for more configs
Offers unparalleled flexibility by allowing to add new configurations easily and quickly.



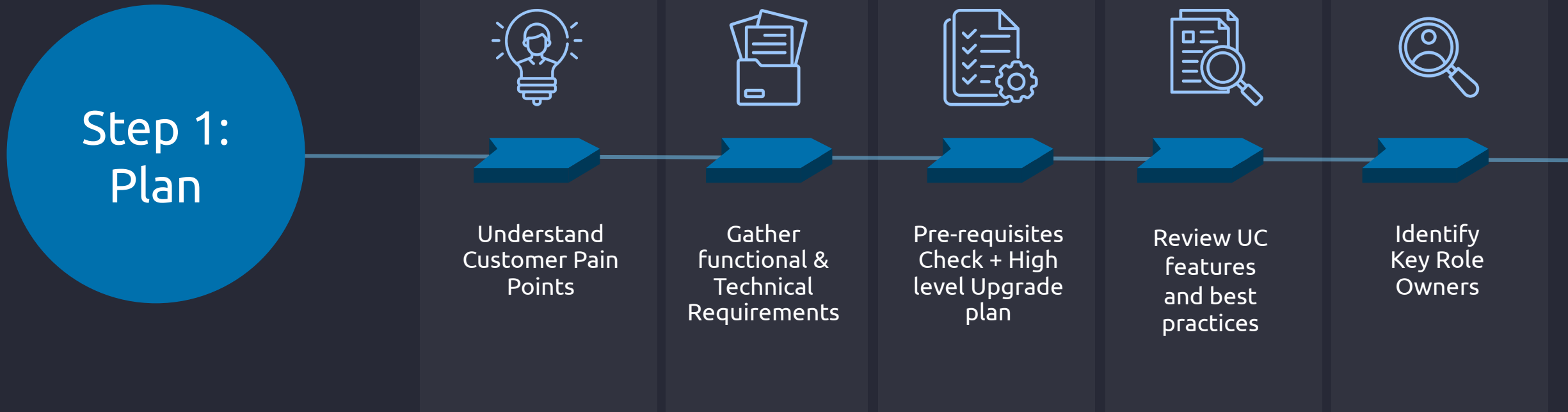


Our Migration Approach



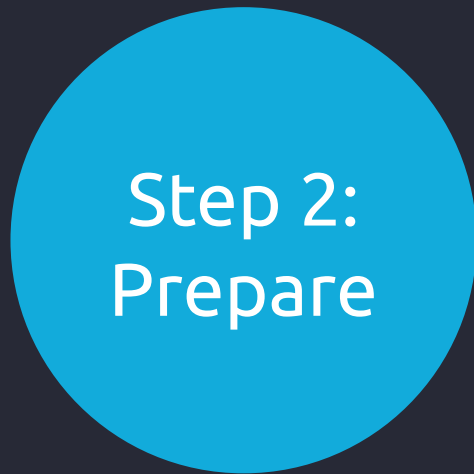


Migration deep dive: Step 1





Migration deep dive: Step 2



Run
Assessment

1. Validate migration level (low, medium, high)
2. Present assessment results
3. Choose migration strategy
4. Prioritize the application based on Customer need
5. Review Application end to end (UAT) testing strategy
6. Review Roll back plan
7. Design Catalog Structures Impact analysis on downstream



Migration deep dive: Step 3

Step 3:
Execute

Capgemini
Databricks
Automation

Activities	Efforts	Capgemini Databricks Accelerators
Set up SCIM and Elevate groups at account level	L	Capgemini accelerator to migrate tables
Set Up Metastores and Create Catalog	L	
Create account level groups & run [UCX] migrate-groups workflow	M	Databricks accelerator to migrate groups
Create required storage credential for SPN	L	
Create external locations for Mounts	L	Capgemini accelerator to migrate tables
Create source and target tables mapping	H	
Upgrade External & Managed tables	H	Capgemini accelerator to migrate tables
Setup access controls for Data and Catalog	M	Databricks accelerator to migrate groups
Upgrade Views	H	Capgemini accelerator to migrate views
Upgrade Interactive Clusters	L	
Upgrade Notebooks	H	Capgemini accelerator to migrate Notebooks
Upgrade Workflows/Job Clusters	M	Capgemini accelerator to migrate workflows & Job Clusters
Migrate DLT Pipelines	H	
Upgrade ML workloads and DBSQL workloads	M	Capgemini accelerator to migrate ML Notebooks
Set up Delta sharing (Optional)	L	



Migration deep dive: Step 4

Step 4:
Validate
& scale



Validation

- ✓ Source and Target table Match
- ✓ Successful Execution of Jobs
- ✓ Successful Execution of Workflow







Scale Accelerate &
Deploy Approach on
Broader Scope

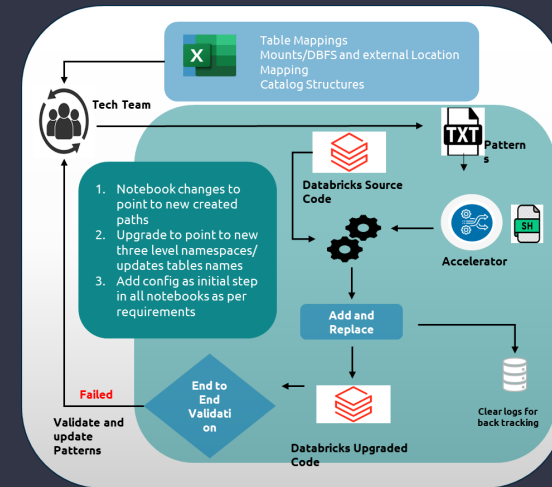


Client context

A leading energy infrastructure company based in Calgary, Alberta. The company connects millions of people to energy across North America through its extensive network of natural gas, oil, and renewable power assets.

 Business objective	 Capgemini approach & solution	 Tools & technology tenets	 Benefits/value delivered
<p>Challenge</p> <ul style="list-style-type: none"> Centralized Data Management: Storing Land Ownership data for Alberta and British Columbia in a centralized storage. Compliance: Meet compliance issues for LOM data. Data Lineage Tracking: Tracking Data Lineage in a user-friendly way. <p>Goal</p> <ul style="list-style-type: none"> Customer's primary goal is to safely and efficiently connect millions of people to the energy they rely on daily while maintaining a strong commitment to sustainability and environmental responsibility. 	<ul style="list-style-type: none"> Migrated existing notebooks from existing DataBricks Non UC workspace to New DataBricks UC workspace using Import/Export functionality. Modified the notebooks to have mount points replaced with External locations by manually replacing them in the notebooks. Created the delta tables in UC by pointing to the external locations. Create Cluster with UC access by selecting Access Mode as Shared and Installed required libraries like-access-parser, com.creatytics:spark-excel_2.12:0.13.7, fsspec, geopandas, openpyxl, pandas-access, pyodbc 	<ul style="list-style-type: none"> Databricks Unity Catalog Azure Data Factory 	<ul style="list-style-type: none"> Centralized Metadata Management Fine-Grained Access Control (UC supports role-based access control (RBAC) and attribute-based access control (ABAC), allowing organizations to enforce detailed security policies. This ensures that only authorized users can access specific data sets) Data Lineage Data Quality and Compliance

Notebook Migration



Using Capgemini Accelerators

About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, generative AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2024 global revenues of €22.1 billion.

Get the future you want | www.capgemini.com



This presentation contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright © 2025 Capgemini. All rights reserved.