

# CASE STUDY



## 83% Faster Pipeline Execution with 100 ETL Pipelines Migrated to Databricks Unity Catalog

### BUSINESS CHALLENGES

Our client, a leading e-commerce platform, relied on a legacy data pipeline architecture that extracted data from Amazon Redshift, processed it, and wrote the results back into Redshift. This setup was becoming increasingly inefficient and costly. The client sought to modernize their data infrastructure by migrating approximately **100 ETL jobs** to a more scalable and cost-effective solution.

### OUR SOLUTION

We transitioned the client's data pipelines to Databricks, leveraging the Unity Catalog for unified data governance.

The migration involved:

- **Data Ingestion:** Rebuilding ETL workflows to read from Redshift and other sources and write to Unity Catalog-managed Delta tables
- **Data Transformation:** Refactoring existing transformations to utilize Databricks' optimized Spark engine
- **Data Governance:** Implementing fine-grained access controls and data lineage using Unity Catalog
- **Automation:** Utilizing Databricks' job scheduling and monitoring features to automate and oversee pipeline execution

### SERVICES PROVIDED

- Data Engineering
- ETL Pipeline Migration
- Databricks Implementation
- Unity Catalog Integration
- Data Governance and Security

### VALUE DELIVERED

- Migrated **100+ ETL jobs** with zero disruption
- **30% cut** in data processing costs led to higher efficiency and stronger business outcomes.
- Enhanced data governance and compliance through Unity Catalog
- Pipeline execution time reduced by **83%**
- Improved pipeline reliability and **performance by 3x**

