Computomic

Modernizing an Enterprise Datamart Platform with Databricks

Abstract

This case study describes how a global Professional Services Organization modernized a fragmented, on-premises datamart environment by migrating to Databricks. The engagement focused on consolidating siloed data, modernizing ETL pipelines, strengthening governance, and reducing operational complexity, while maintaining continuity for downstream analytics consumers. Through a phased migration approach and centralized data governance, the organization achieved improved performance, lower total cost of ownership, and a scalable foundation for future analytics and AI initiatives.

Overview

A global enterprise organization undertook a strategic initiative to modernize its legacy data platform by migrating multiple on-premises datamarts to a Databricks architecture.

The engagement focused on consolidating siloed data systems, improving performance and scalability, strengthening governance, and reducing operational complexity — while ensuring uninterrupted support for downstream analytics and reporting during the transition.

Business Challenges

The organization faced several challenges with its existing data platform:

- **Fragmented data landscape** with multiple independent datamarts
- **Legacy on-premises infrastructure** limiting scalability and agility
- **High operational cost and complexity** driven by multiple ETL tools and custom ingestion frameworks
- **Limited data governance and discovery**, making access control and compliance difficult
- **Downstream system constraints**, with some BI tools not yet compatible with modern SQL engines

Objectives

The primary goals of the initiative were to:

- Migrate legacy, on-premises datamarts to a cloud-native lakehouse
- Eliminate data silos by consolidating data into a single governed platform
- Improve performance, scalability, and operational efficiency
- Strengthen data governance, access control, and lineage
- Reduce total cost of ownership by simplifying the technology stack
- Enable a phased transition for downstream analytics consumers

Solution

The organization implemented a modern lakehouse architecture built on Databricks, supported by a phased migration and coexistence strategy.

Key elements of the solution included

- **Lakehouse Consolidation**
  All datamarts were migrated into a unified Delta Lake–based architecture.

- **ETL Modernization**
  Legacy ETL workflows were migrated to PySpark and SQL, replacing proprietary tooling and custom scripts.

- **Direct Source Ingestion**
  Data ingestion was simplified by directly connecting to source systems (files, databases, APIs, streaming platforms), eliminating intermediate processing layers.

- **Centralized Governance**
  Fine-grained access control, data discovery, lineage, and unified metadata management were implemented using Unity Catalog.

- **Dual-source Architecture (Temporary)**
  To support downstream systems not yet compatible with modern SQL engines, a controlled synchronization mechanism was introduced during the transition phase.

Implementation Highlights

- Migrated data from approximately 15 upstream source systems
- Centralized configuration management using a single, declarative configuration model
- Implemented scalable batch and streaming ingestion using Spark
- Migrated historical data using a staged, low-risk approach
- Introduced incremental data synchronization using change data capture
- Developed custom tooling to enforce data quality and integrity checks

Results & Impact

The engagement delivered measurable improvements across performance, cost, and operational efficiency:

- **Eliminated data silos** by consolidating all datamarts into a single lakehouse
- **Significantly improved performance,** reducing ETL runtimes and improving data availability
- **Simplified the technology stack,** decommissioning multiple legacy tools and scripts
- **Reduced operational overhead and TCO** through standardization
- **Improved governance and security** with centralized access control and lineage
- **Enabled future scalability,** positioning the platform for advanced analytics and AI use cases

Challenges Overcome

Automated Migration Quality
Initial automated conversions required refinement. Iterative tuning and validation ensured high-quality, production-ready pipelines.

Downstream Compatibility Constraints
A phased coexistence strategy ensured uninterrupted analytics access while enabling future modernization.

Platform Capability Gaps
Custom solutions were implemented to address temporary platform limitations around collations and constraint enforcement.

Why This Matters

This engagement demonstrates how enterprises can modernize complex, legacy datamart environments without disrupting downstream consumers — by combining a lakehouse architecture with a pragmatic, phased migration strategy.

The result is a scalable, governed, and cost-efficient data platform that accelerates analytics today while laying the foundation for future innovation.

Technologies Used

- Databricks Lakehouse Platform
- Delta Lake
- Unity Catalog
- Databricks SQL
- Spark (Batch & Streaming)
- JDBC / ODBC Connectivity

Key Takeaways

- A lakehouse architecture can successfully replace fragmented datamart ecosystems
- Phased migration strategies reduce risk while enabling modernization
- Centralized governance is critical for enterprise-scale data platforms
- Simplifying the technology stack delivers both cost and operational benefits

Computomic