

Driving Near Real-Time POS Analytics for a Leading Global Footwear & Accessories Retailer: A Databricks Medallion Architecture Story

How a modern Lakehouse foundation cut dashboard latency from 30–120 minutes to under 10 min, and unlocked near real-time retail decision-making.



Client Overview

Our client is a leading **global fashion footwear and accessories retailer**, operating across a wide brick-and-mortar footprint and a fast-growing e-commerce channel, with high-volume omnichannel POS transactions flowing through PredictSpring mPOS.

Operating in a trend-driven, promotion-intensive retail category, the organization manages compressed decision cycles across pricing, promotions, inventory, and store operations. As transaction volumes climbed and reporting cadences fell behind, the business needed a POS analytics foundation that could match the pace of the retail floor.

To modernize its POS analytics landscape, the client partnered with Polestar Analytics to build a Near Real-Time POS Analytics Solution on Azure Databricks, anchored in Medallion (Bronze, Silver, Gold) architecture. The objective was to replace batch-based reporting with a **governed, near real-time single source of truth for retail performance**.



Key Challenges

Four structural challenges were constraining the retail operating rhythm, not just the ability to report on it. Each had to be addressed before near real-time decisioning could become the norm.

Four Structural Challenges Limiting Real-Time Retail Decisioning

Slow, batch-based reporting

- ✓ Merchandising and store operations were reacting to insights that had already aged out of relevance.
- ✓ Peak-hour traffic patterns, the exact moments the business needed to see, were the ones most delayed by batch cycles.
- ✓ Dashboards were being consumed, but rarely acted on within the same trading window.

High-volume, complex POS data

- ✓ JSON payloads from mPOS carried nested structures that stressed downstream parsing.
- ✓ Scale was compounding: as store traffic grew, the fixed processing window was becoming a hard constraint.
- ✓ No governed landing zone existed to standardize the shape of incoming data before it hit reporting.

Heavy reliance on legacy ERP

- ✓ The ERP could not stream, so any request for near real-time visibility routed back into batch.
- ✓ Business teams lacked a modern, queryable single source of truth for POS performance.
- ✓ Even accurate reporting from the ERP arrived too late to change the day.

Limited pipeline stability and cost efficiency

- ✓ DBU consumption and cluster sizing were not tuned to the shape of the workload.
- ✓ JSON parsing relied on repeated explodes and cross joins, taxing both cost and runtime.
- ✓ Failure recovery and monitoring were manual, eroding confidence in freshness.

Our Approach

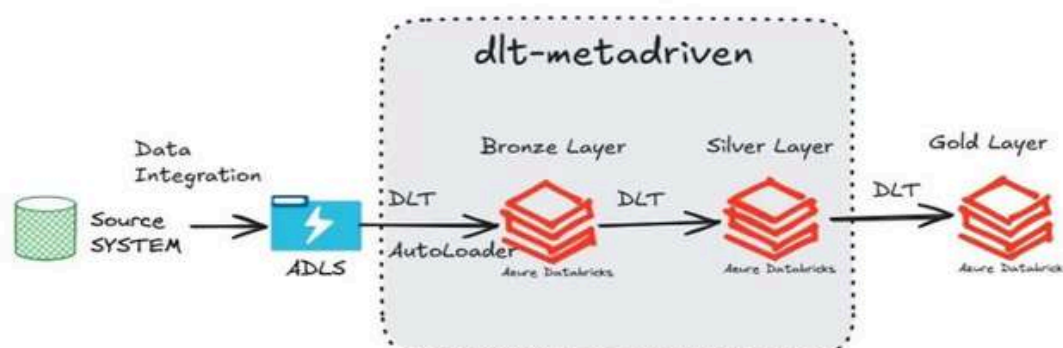
A Medallion-Led Lakehouse Foundation

Architecture design came before pipeline build. Polestar Analytics mapped the POS data journey end to end, from mPOS event to business decision, and anchored the engagement in a **Medallion (Bronze, Silver, Gold) architecture on Azure Databricks**. The result is a governed, layered foundation that gives the organization one agreed reference point for how raw POS data becomes decision-ready insight.

The framework concentrated on four data domains that together describe the health of the retail floor:

- ✓ Cash TLog (transaction-level sales)
- ✓ Product, Traffic, and Store performance
- ✓ Inventory and Employee data
- ✓ Cross-domain KPIs feeding Power BI

Batch/Streaming Data Pipelines using Databricks DLT



The Solution

With the architectural framework defined, attention turned to the platform build. **Polestar Analytics** delivered the Near Real-Time POS Analytics Solution on a focused **Azure Databricks Lakehouse** stack, engineered around four anchoring principles: **near real-time freshness, pipeline stability, cost efficiency, and governed extensibility.**

Medallion Architecture on Databricks

- ✓ Bronze, Silver, and Gold pipelines implemented on Azure Databricks using Autoloader and Delta Live Tables (DLT).
- ✓ POS data (Cash TLog, Product, Traffic, Store, Inventory, Employee) ingested from Blob storage into ADLS Gen2 as a single governed landing.
- ✓ Orchestration and scheduling managed through Databricks Workflows, giving the pipeline a repeatable, observable operating rhythm.

Refined and Modeled POS Data

- ✓ Nested JSON payloads standardized into cleaned, structured Silver tables ready for business consumption.
- ✓ A domain-aligned Gold layer built so merchandising, operations, and finance query the same, trusted numbers.
- ✓ Business functions now analyze POS performance without first having to reconcile it.

Performance and Cost Optimization

- ✓ JSON parsing logic re-engineered to reduce explodes and eliminate cross joins, lowering compute cost and improving runtime accuracy.
- ✓ DBU usage stabilized to 16 DBU/hour through pipeline tuning and cluster right-sizing.
- ✓ Ingestion pipeline re-engineered to handle materially larger file volumes within the same processing window.

Power BI Integration for Decision-Ready Dashboards

- ✓ Power BI connected directly to the Databricks Lakehouse, removing intermediate handoffs.
- ✓ Dashboards designed for merchandising, store operations, and leadership rather than analysts.
- ✓ Frontline teams move from POS data to a pricing, promotion, or inventory decision with materially less friction.

Business Impact



From hours to minutes.

POS insight refreshes in 6 to 10 minutes, not 30 to 120, while the trading window is still open.



70 to 80% faster processing

Bigger volumes, same window. Scale becomes a tailwind, not a bottleneck.



4,500 files in 15 minutes.

Throughput that scales with store growth instead of buckling under it.



Predictable cost at scale.

DBU usage held steady at ~16 DBU/hour through disciplined parsing and right-sized clusters.

About Polestar Analytics

Polestar Analytics is a data, analytics, AI, and planning partner helping enterprises move from fragmented data estates to production-grade data, AI, and agents on the Databricks Data Intelligence Platform. With 10+ years of delivery experience and a large pool of Databricks-certified engineers, we modernize data platforms, optimize ETL and Lakehouse architectures, and deliver industry-specific analytics and agentic AI solutions that drive measurable impact in CPG and retail, life sciences, manufacturing, and TMT/FinOps.

Reach out to us today!