POLESTAR ANALYTICS

# **Unlocking ERP Data at Scale for a Leading Global Footwear & Accessories Retailer:** An M3-to-Databricks Migration Story

---

How a metadata-driven Medallion framework moved 143 Infor M3 ERP tables into a governed Azure Databricks Lakehouse, replacing ad-hoc extraction with automated daily pipelines cutting refresh time by 50%.



# 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 **M3 ERP** on Azure Cloud SQL Server serving as the system of record for sales, inventory, order processing, purchasing, customer master, and distribution.

In a trend-driven category, inventory and order decisions can't wait for data. Yet all 143 M3 tables still relied on direct queries and ad-hoc extracts, with no automation, no history, and no governance. Every new question meant another manual pull.

To modernize its ERP data landscape, the client partnered with Polestar Analytics to migrate its M3 data into a governed Azure Databricks Lakehouse, anchored in a **Medallion (Bronze, Silver) architecture**. The objective was to replace manual extraction with an **automated, governed, and auditable foundation for enterprise ERP data.**



POLESTAR ANALYTICS

# Key Challenges

Four structural gaps were limiting how much value the business could draw from its ERP data. Each had to be closed before M3 data could serve as a trusted foundation for analytics.

## ERP data locked in manual, ad-hoc extraction

- Teams relied on direct queries and one-off extracts to move M3 data into downstream analytics.
- Every new data request meant repeat effort, and results varied depending on who pulled them.
- Direct querying placed avoidable load on the production ERP database.

## A large, complex table estate with no standard processing

- 143 tables across sales, inventory, orders, purchasing, customer, and distribution, each with its own structure.
- No common cleansing, data type standards, or deduplication rules were applied across tables.
- Building and maintaining pipelines table by table would not scale with the size of the estate.

## No incremental loading or historical tracking

- Without change detection, every refresh meant reprocessing full tables.
- Record updates overwrote prior values, so there was no view of how data looked at an earlier point in time.
- Late or failed extracts went unnoticed until the business flagged them.

## No governance, lineage, or audit trail

- No central catalog showed what data existed, who owned it, or who could access it.
- Lineage from the M3 source to downstream consumption could not be traced.
- Credentials and access were managed outside a controlled framework.

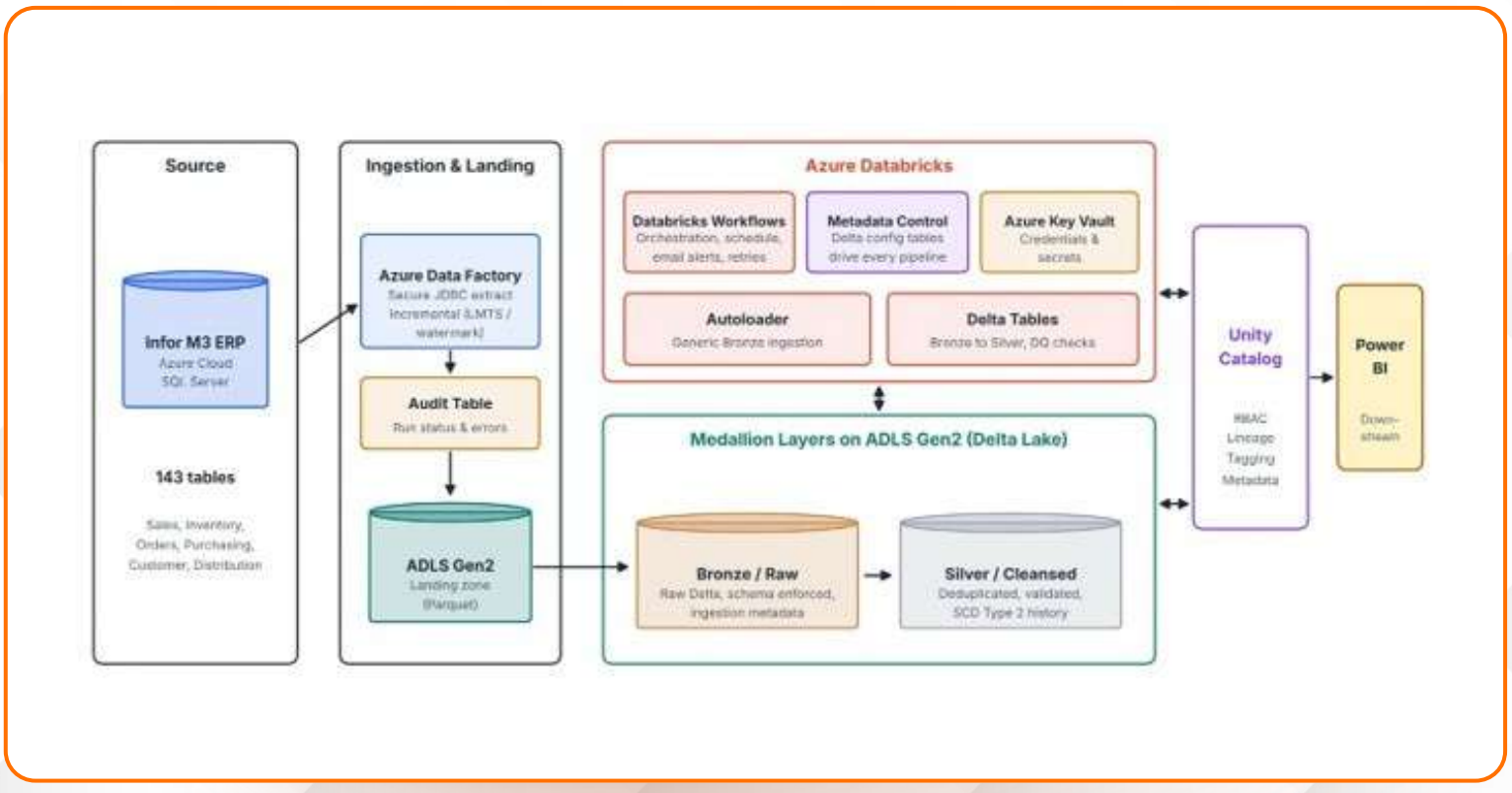
# Our Approach

## A Metadata-Driven Medallion Foundation

Framework design came before pipeline build. Polestar Analytics mapped every M3 table field by field, then built one metadata-driven Medallion framework on Azure Databricks to load all 143 tables from a single configuration.

## The approach rested on four design choices that kept the migration fast, consistent, and low-risk:

- **Field-level STM** for every table, agreed with the business before build.
- **Metadata-driven, generic pipelines** in place of table-by-table code.
- **Incremental extraction** using last-modified timestamps to protect the source system.
- **Phased go-live** in three batches, with build and testing running in parallel.



# The Solution

With the framework defined, attention turned to the platform build. **Polestar Analytics** delivered the M3 Cloud Data Migration on a focused **Azure Databricks Lakehouse** stack, engineered around **automation, data quality, governance, and operational resilience**.

## Secure, Incremental Ingestion

- M3 data extracted from Azure Cloud SQL Server through **Azure Data Factory** over secure JDBC connectivity.
- **Incremental loads using last-modified timestamp (LMTS) and watermark columns**, so only changed records move on each run.
- Data landed in **ADLS Gen2 as Parquet** and ingested into Bronze Delta tables by a generic **Autoloader** framework, with an audit table tracking every run.

## Cleansed, Historized Silver Layer

- **Delta Tables (DT)** pipelines turn each Bronze table into a standardized Silver table, applying the agreed STM rules.
- Deduplication, null handling, and **data quality checks with threshold-based error flagging** at the Silver boundary.
- **SCD Type 2** history preserves every change to a record, with timestamp-level traceability.

## Governance and Security by Design

- **Unity Catalog** provides one governed catalog with metadata management, table tagging, and **role-based access control**.
- End-to-end **lineage** from source through Bronze and Silver is tracked natively.
- Credentials and secrets are managed in **Azure Key Vault**, never in code.

## Automated, Production-Grade Operations

- **Databricks Workflows** orchestrate daily runs with dependency management, retry logic, and **email-based failure alerts.**
- **YAML-based CI/CD** promotes code consistently from Dev to Production.
- Go-live completed in three phases, followed by **48-hour hypercare** and a full operations runbook handover.

# Business Impact

## 143 tables live in production

Every M3 table now flows in automatically, no manual extracts.

## Refresh time cut by 40–50%

Incremental loads now finish in ~50 minutes, down from 90.

## Live in 10 weeks

One reusable framework, three phased go-lives, no table-by-table rebuilds.

## Full history, fully governed

Change history, lineage, and access control across every table.

POLESTAR ANALYTICS

# 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.

**Reach out to us today!**

