

Real-Time Batch Traceability: How Data Modernization Made a Bio-Pharma Leader Audit-Ready

INDUSTRY

Life Sciences

CAPABILITY

Data Modernization
& Analytics Platform

FUNCTION

Business
Operations

TECH STACK

Amazon Redshift,
Databricks
(PySpark, Spark
SQL), SAP HANA,
MES, LIMS,
QMS, Bitbucket

Client Overview

Our client is a global biopharmaceutical leader. Known for producing life-saving plasma-derived therapies, they specialize in Immunoglobulin (Ig) and Albumin product portfolios across global markets.

The Ask

The client planned to modernize its Data platform by transitioning from Amazon Redshift and legacy systems to Databricks cloud platform. The goal was to consolidate data silos, enable end-to-end batch traceability, automate data pipelines, and establish a governed, auditable single source of truth for manufacturing and quality analytics.

Challenges

Integrating siloed operational datasets from enterprise platforms, including SAP HANA, MES, LIMS, and QMS, into a unified data ecosystem.

Streamlining end-to-end batch genealogy and product mapping to transition from manual tracking to automated, high-visibility reporting.

Optimizing data ingestion and processing workflows to move from fragmented manual tasks to efficient, consistent, and automated pipelines.

Establishing comprehensive end-to-end data lineage and auditable pipelines to strengthen compliance frameworks and streamline regulatory processes.

Our Solution: Manufacturing & Quality Data Platform Modernization

Migrated Legacy Warehouse to Cloud Infrastructure:

Transitioned legacy SQL data warehouse components to Amazon Redshift to support high-performance SQL servicing and enterprise data queries.

Architected Medallion Lakehouse Structure:

Structured raw and processed data into Bronze (Raw), Silver (Standardized), and Gold (Business Model) layers using Azure Databricks to clean, harmonize, and enrich telemetry.

Automated Multi-Source Data Ingestion:

Implemented automated extraction pipelines using Apache NiFi, Fivetran, and Azure Databricks Spark to ingest data from SAP HANA, Salesforce Cloud, SharePoint Lists, Rapid Response, and Star LIMS.

Engineered Standardized Batch Data Products:

Developed PySpark and Spark SQL pipelines to create unified data entities, including Batch, Batch Genealogy, and Batch-to-Product models for cross-system harmonization.

Orchestrated Workflows with Automated DevOps:

Managed pipeline workflows with Databricks Native Scheduler and enforced continuous integration/deployment using Bitbucket.

Enabled Advanced BI & ML Sandbox:

Enabled self-service dashboards through Power BI Premium Datasets and established an MLflow sandbox environment to accelerate predictive analytics and machine learning use cases.

Impact Delivered

Provided real-time visibility into manufacturing, quality testing, and distribution processes across operating sites.

Drastically reduced reporting cycles by automating manual ingestion pipelines and standardizing enterprise data models.

Enabled continuous process validation, early detection of process deviations, and real-time quality trend monitoring.

Established fully auditable data pipelines, transparent data lineage, and governed access controls to satisfy biopharma compliance mandates.

Standardized data quality and validation checks and established trusted, reliable datasets for the Manufacturing, Quality, and Operations teams.

What Could Real-Time Batch Traceability Unlock for You?

[START A CONVERSATION](#)

About Us

Tiger Analytics is a global leader in AI and analytics, helping Fortune 1000 companies solve their toughest challenges. We offer full-stack AI and analytics services & solutions to help businesses achieve real outcomes and value at scale. We are on a mission to push the boundaries of what AI and analytics can do to help enterprises navigate uncertainty and move forward decisively. Our purpose is to provide certainty to shape a better tomorrow.

US | UK | Canada | Mexico | India | Spain | Singapore | Philippines | Malaysia | Australia

