




Modernizing SAP Data for Enterprise-Scale Analytics on Databricks

Building a governed, real-time data foundation for healthcare insights





As a leading U.S.-based healthcare organization expanded its distribution of critical-care biopharmaceuticals, vaccines, and plasma products, the scale of its operations began to outpace the capabilities of its existing data environment. Data lived across SAP and other enterprise systems, and while these platforms reliably supported day-to-day transactions, they were increasingly strained when it came to analytics and reporting. Decision-makers needed faster, more reliable access to insights, but the underlying data architecture was no longer built for that demand.

Recognizing that incremental fixes would not be enough, the organization set out to modernize its data landscape and establish a scalable foundation that could support both operational reporting and longer-term strategic analytics.

Moving Beyond Legacy Reporting Architectures

The modernization effort centered on migrating data from SAP and other enterprise sources into Databricks, replacing an existing environment running on GCP. This was not approached as a simple platform swap. The objective was to create a cohesive, modern data platform, one that could handle ingestion, transformation, modeling, and reporting within a single, governed architecture. To support consistent and automated data movement, Fivetran was selected as the primary ingestion mechanism, ensuring reliable data replication while preserving accuracy across systems.

This shift laid the groundwork for a platform that could evolve alongside the business, rather than constrain it.

Tackling SAP Data Complexity

One of the most significant hurdles emerged from SAP's highly standardized and normalized data models. While well suited for transactional integrity, these structures were not designed for analytics consumption. Transforming SAP data into meaningful, business-ready views required extensive de-normalization, careful table joins, and performance-aware transformations. At the same time, the platform needed to support

near-real-time data synchronization from source systems and real-time processing within Databricks, raising the stakes for stability and consistency across downstream workloads.

Solving this challenge required both architectural discipline and a deep understanding of how SAP data behaves at scale.

Structuring Data for Analytics at Scale

Brillio designed and implemented a Databricks-based modern data platform using a medallion architecture - with bronze, silver, and gold layers, to systematically manage data refinement. Raw data from SAP, Salesforce, Azure SQL, and other on-premise and SaaS sources was ingested into the Bronze layer using Fivetran as the primary ingestion mechanism, along with Databricks-native connectors.

From there, data was cleansed, enriched, and de-normalized into the Silver layer, where models

were tailored specifically to the client's operational workflows and reporting requirements. This structure ensured that data quality improved at each stage while remaining traceable back to its source.

The Gold layer was then built to support analytics and reporting use cases, providing curated, business-ready datasets that functioned as trusted data marts across the organization.

Embedding Governance and Enabling Self-Service Analytics

From the outset, governance was treated as a core architectural requirement rather than an afterthought. Brillio implemented Unity Catalog to provide centralized access control, data lineage, and metadata management across the platform. This ensured that as data volumes grew and more users accessed the platform, security and traceability remained intact.

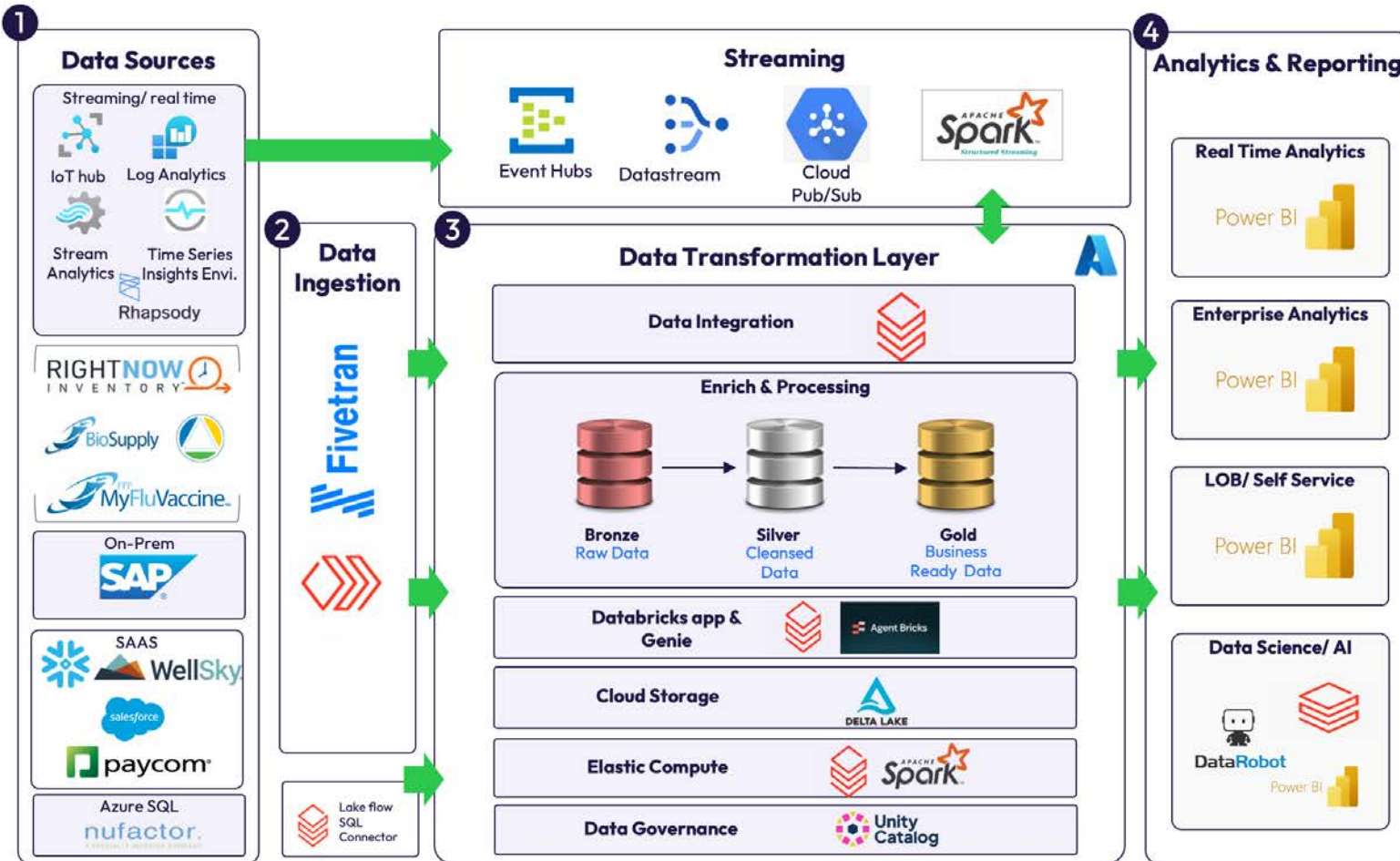
Power BI was integrated as the primary analytics and visualization layer, supporting real-time analytics, enterprise reporting, and self-service exploration for business users. The platform was

also designed to support advanced analytics and data science use cases through native Databricks capabilities, creating a foundation for future innovation.

Further extending the platform's capabilities, Brillio enabled Agentic AI to support reasoning and parsing across enterprise data artifacts. Leveraging LLMs with Retrieval-Augmented Generation (RAG) and Databricks Genie, the solution supports the extraction and summarization of business KPIs related to shipments and warehouse operations.

Solution Architecture of Modern Data Platform

Onsite + Offshore Resources Mix (3 months and ongoing)



Key Highlights

- 1 **Data Sources:**
 - Comprehensive intake of diverse datatypes coming from various origins of data, including real-time streaming sources like IoT Hub and Log Analytics, alongside traditional on-premise and SaaS applications.
- 2 **Data Ingestion:**
 - Fivetran and other integration tools are used to efficiently move data for further processing.
- 3 **Data Transformation Layer**
 - Bronze layer itself acts as a Landing layer.
 - Transformation Layer leverages Spark and Delta Lake for scalable data processing, storage.
 - Real-time data processing in Databricks using streaming delta tables
- 4 **Consumption:**
 - Power BI is mainly used for visualization and analytics and reporting on the data while DataRobot enables advanced data science.
- 5 **Agentic AI**
 - Enabled Agentic AI for reasoning, parsing various artifacts using LLMs with RAG, memory, and execution to automate business use cases end-to-end.
 - Multi-agent solution using Genie to extract and summarize business KPIs related to shipments and warehouse operations.
- 6 **Security & Governance**
 - Robust data governance is ensured via Unity Catalog.

Driving Measurable Improvements in Data Reliability and Insight

With the new platform in place, integrated with Power BI for analytics and visualization, and DataRobot supporting advanced data science use cases, the organization has significantly improved the accuracy, availability, and timeliness of its reporting and analytics.

Automated ingestion and streamlined transformation pipelines reduced manual effort and improved consistency across datasets, enabling faster access to reliable insights. Strong governance controls enhanced data security and transparency, while the structured migration approach accelerated the move away from legacy reporting systems.

Together, these improvements delivered a scalable, future-ready data foundation, one that supports better decision-making today while positioning the organization to grow with confidence.

- Modernized the enterprise data platform by migrating from a legacy GCP environment to Databricks
- Improved reporting accuracy, availability, and timeliness through automated ingestion and structured data modeling
- Strengthened data governance, security, and lineage with Unity Catalog
- Enabled self-service analytics and advanced data science using Power BI and Databricks



ABOUT BRILLIO

Brillio is a digital technology services company that drives AI-first engineering and design-led experiences for global enterprises. Born digital in 2014, its consulting-led services span Customer Experience, Data & AI, Product Engineering, and Digital Infrastructure. With an industry-leading NPS of 71, Brillio accelerates time to market through its proprietary BrillioOne.ai platform, powered by AI-ready talent with deep domain expertise.

Brillio is the official Digital Transformation Partner and the official Data and AI Services Provider of Atlassian Williams Racing. Brillio partners with leading technology providers including Microsoft, AWS, Google Cloud, Salesforce, Adobe, Databricks, and Snowflake and operates with 6,000+ “Brillians” across 15 global delivery centers. Consistently recognized as a Great Place to Work® since 2021, Brillio blends innovation, talent, and purpose to deliver measurable outcomes for clients and fulfilling careers for employees.



<https://www.brillio.com/>

Contact Us: info@brillio.com