Distributed Data Analytics Platform for a Leading MedTech Company

Executive Summary

A leading MedTech company aimed to build a democratized, unified data consortium for clinical researchers by integrating data from EMR systems, mobile health devices, scanners, anesthesia machines, ventilators, and ECG devices within an ICU setting. The objective was to create high-fidelity data streams from on-premise systems to the cloud, handling various data formats like HL7, FHIR, waveforms, CSV, and proprietary formats. Leveraging a cloud-based Health Lakehouse with standardized data modeling following OHDSI/OMOP principles, the platform empowers clinical research, enhances patient care continuity, and streamlines population health initiatives.

Challenges

- Fragmented data sources across multiple medical devices and departments
- Lack of standardized data models hindering interoperability
- Difficulty in managing real-time data from diverse sources
- Data silos limiting insights for clinical researchers
- Complexities in ensuring data quality and consistency

Objectives

- Create a unified data platform for diverse healthcare data sources
- Enable real-time, high-fidelity data ingestion and processing
- Build a Health Lakehouse following OHDSI/OMOP principles
- Enhance data accessibility for clinical research and patient care
- Ensure data quality, security, and scalability

Solution Approach

- Unified Data Ecosystem: Shifted from a device-centric to a data-centric model, treating data as an asset.

- Health Lakehouse on Cloud: Built using Databricks to replace traditional data lakes and warehouses, streamlining data management.

- Real-Time Data Streaming: Employed Kafka for real-time data ingestion and processing.

- Scalable Analytics Framework: Leveraged Apache Spark and Delta Lake for speed, scalability, and reliability.

- Advanced Analytics Integration: Applied NLP and other techniques to standardize and enrich diverse data formats.

- Medallion Architecture: Implemented a multi-layered approach to ensure Atomicity, Consistency, Isolation, and Durability (ACID) properties.

- Data Quality Dashboard: Developed to maintain high-quality data federation and monitor data integrity.

- Compliance & Security: Established unified controls for data access and auditing.

Results & Impact

- Enhanced Interoperability: Eliminated departmental and device silos, fostering data collaboration.

- Data Liberation: Empowered providers and researchers with comprehensive, accessible data

- .Improved Patient Care: Delivered real-time patient and device data, enabling better clinical decisions.

- Future-Proof Scalability: Multi-tenant architecture supports easy integration of new applications.

- Faster Insights: Standardized data models accelerated app onboarding and data analysis.

- Reliable Data Streams: Low-latency, high-quality data streams optimized for research and patient care.