Streaming Case Studies

ECommerce Case Study

Sheetz – Streaming Analytics

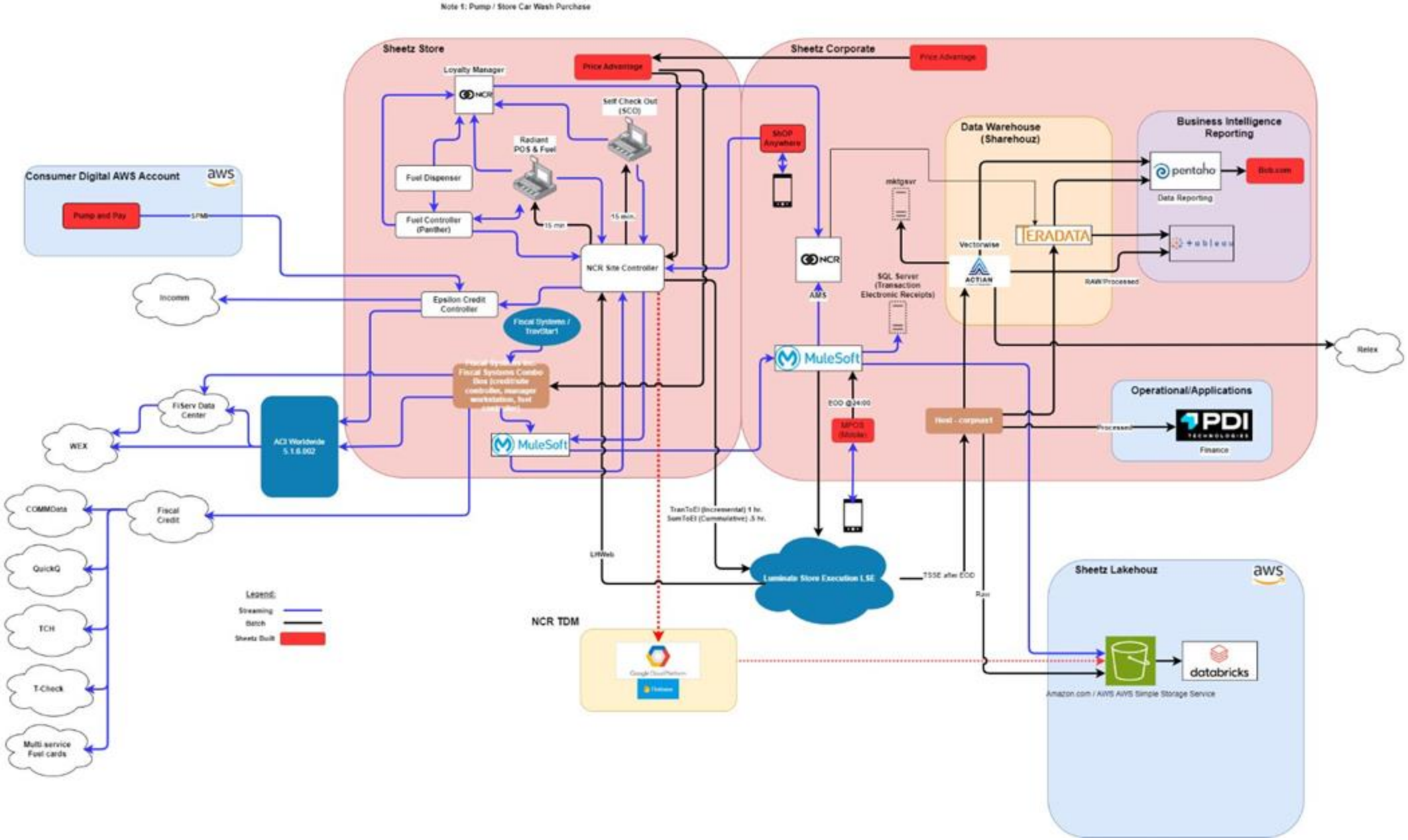
Problem Statement

1. **Fragmented Data Sources:** Sheetz's data systems were isolated, causing inefficiencies in integration and processing.
2. **Complex Data Flows:** The convoluted data flow among store systems and corporate applications led to delays and synchronization errors.
3. **Limited Scalability:** Legacy systems struggled to meet Sheetz's increasing data demands, particularly for real-time analytics.
4. **Inflexible Reporting Tools:** Outdated Business Intelligence tools hindered quick generation of actionable insights.
5. **High Maintenance Costs:** Managing separate systems and custom integrations was costly and resource-heavy.

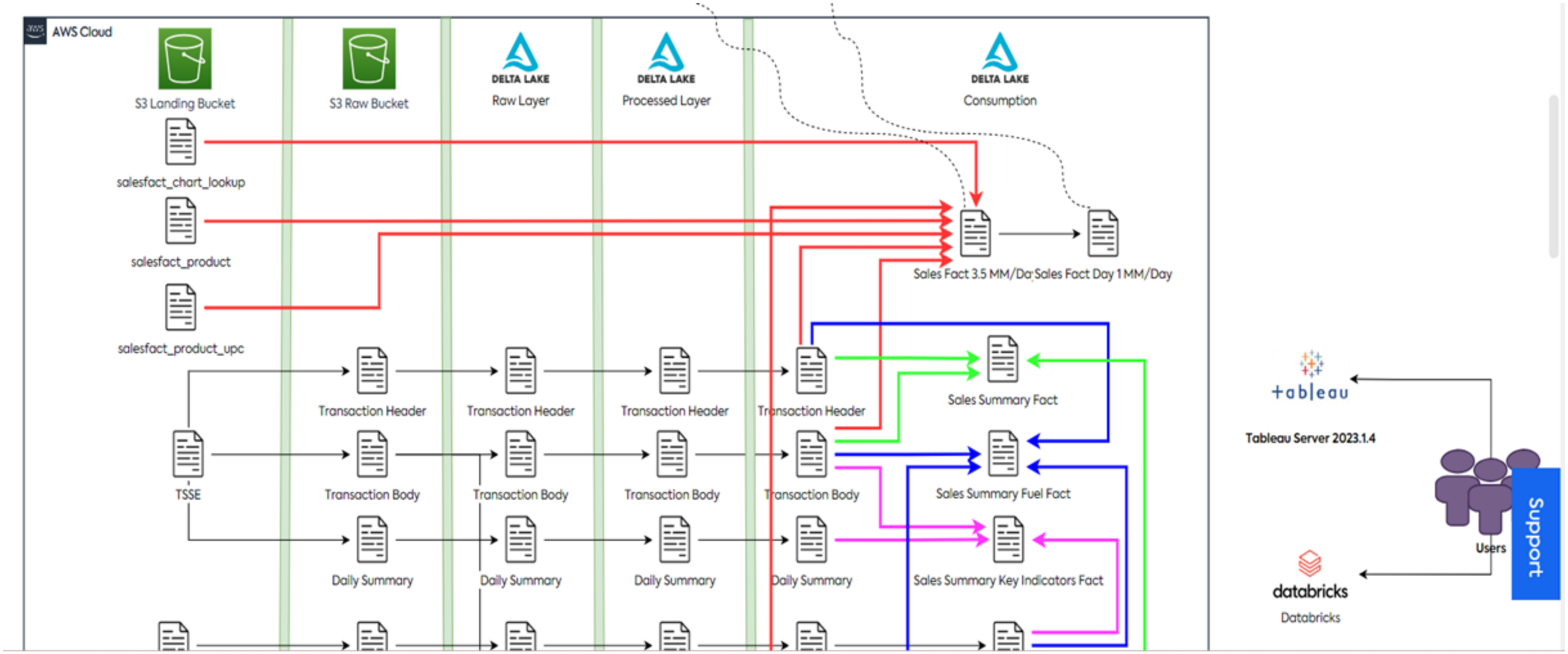
Solution Delivered

1. **Centralized Data Lakehouse**: Databricks consolidates structured and unstructured data, removing silos and integrating sources like Oracle, SQL Server, and streaming data.
2. **Real-Time Analytics**: It supports Change Data Capture (CDC) and batch processing for timely insights, with streaming capabilities for immediate updates.
3. **Enhanced Data Governance**: Databricks Unity Catalog provides detailed access control, improving compliance and enabling self-service analytics.
4. **Simplified Reporting Workflow**: Tableau Prep's integration with Databricks allows faster data preparation and analysis without legacy tools.
5. **Scalable Infrastructure**: Utilizing cloud-native technologies, Sheetz can effectively manage increasing data volumes.

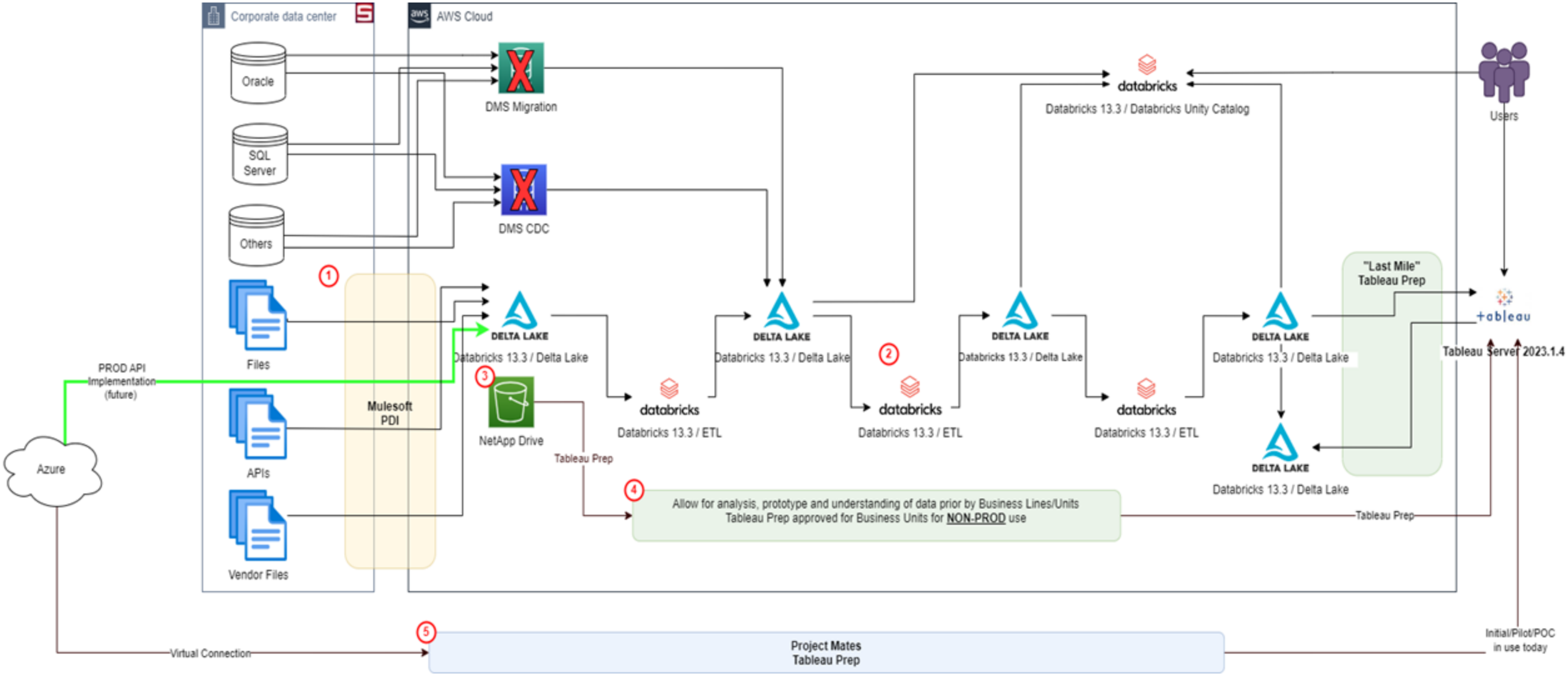
Sheetz – World before Databricks



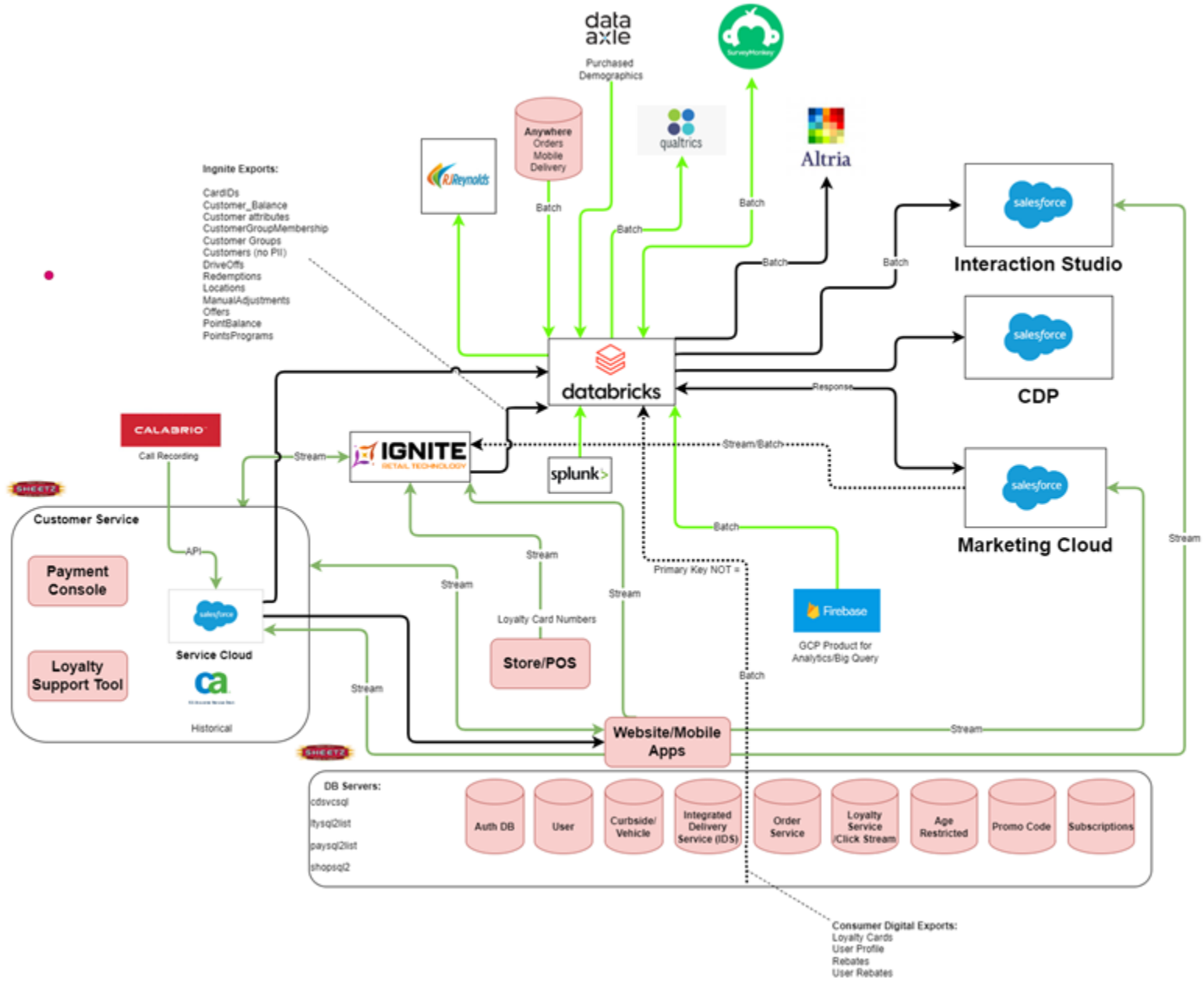
Sheetz – Adopting Databricks for POS Sales Data



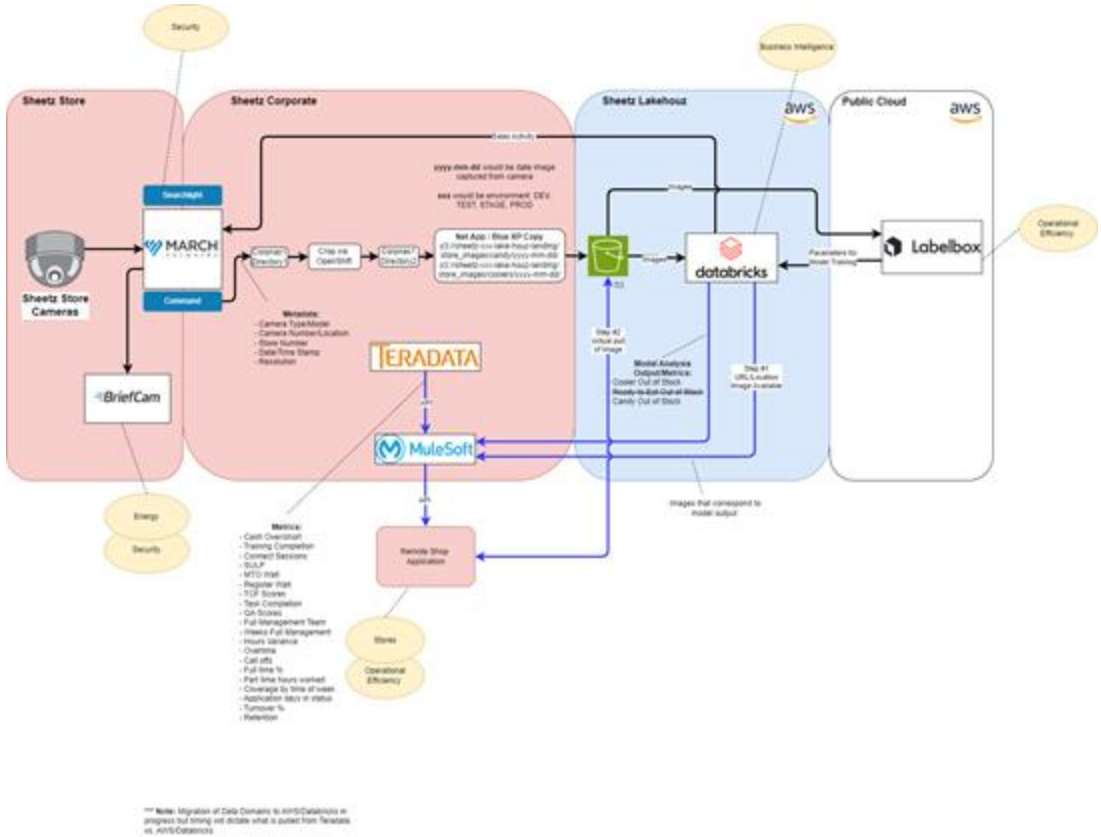
Sheetz – Embracing Databricks First



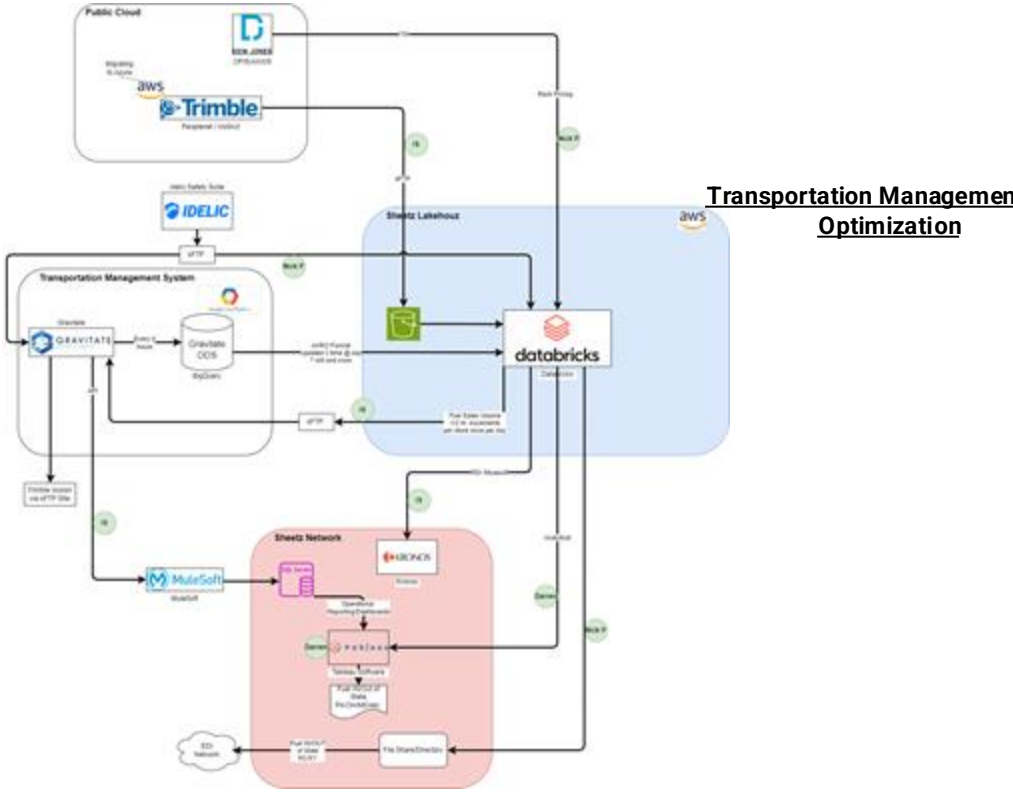
Sheetz – Modernized Enterprise View



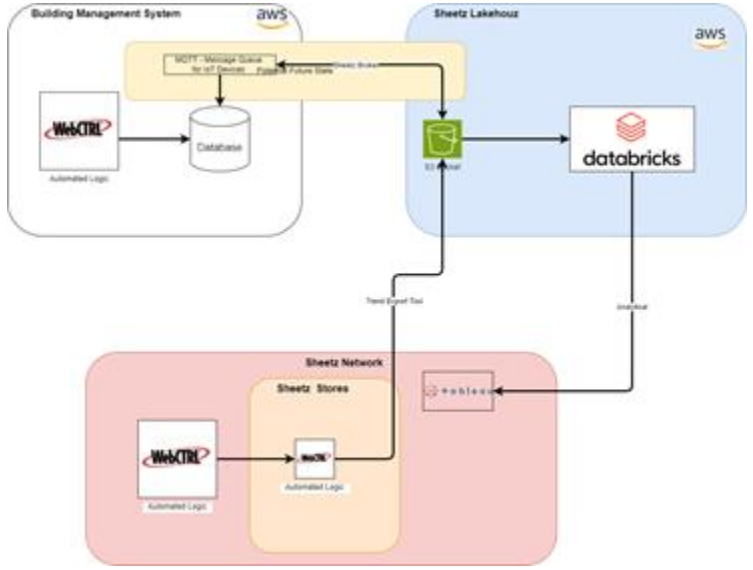
Sheetz Other Use Cases



Video Analytics



Transportation Management and Optimization



Building Energy Management

FUEL AUTOMATION™ STATION

motifworks | CASE STUDY

SOLUTION
IoT, Digital Platform

FOCUS AREAS
Process Automation, IT-OT, Operational Efficiency, Accounting

INDUSTRY
Oil & Gas, Logistics

Unique automated fuel delivery and management technology that uses advanced sensors to collect over 500,000 data points per day, providing real-time visibility of operations, guaranteed fuel supply and new levels of efficiency

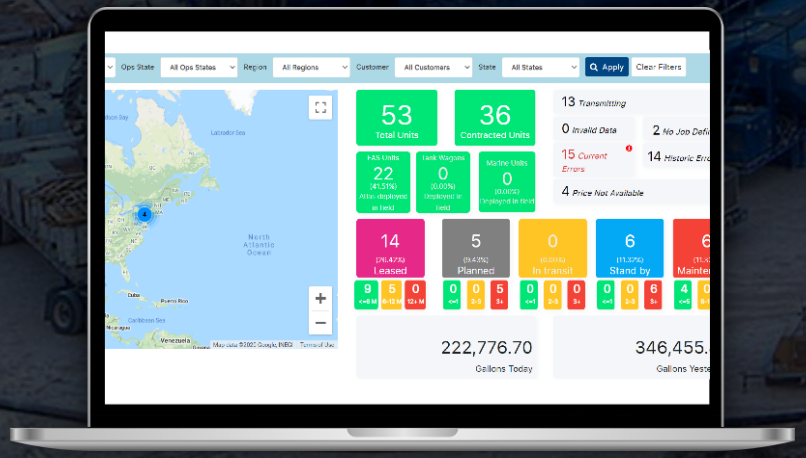
$1.4M
Net Operating Savings
in first 6 months

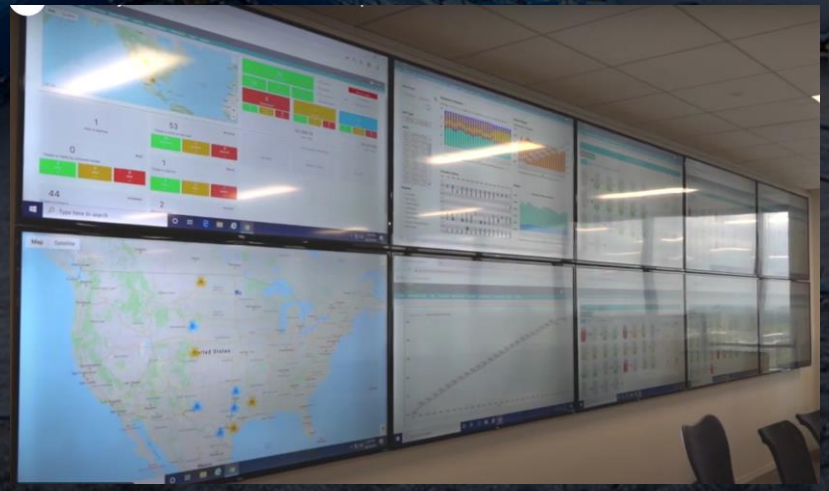
85%
Reduction in invoice times
Down to 2 days from 30 days

$7.5M
Working Capital Improvements
in first 6 months

- Digital Platform Development
- Product Development
- UI/UX Design
- IoT Platform
- Cloud-based SaaS Solution
- Data Analytics
- Machine Learning
- Power BI Development
- Azure Managed Services

1. FAS Units with IoT sensors
2. Automated fueling with 28 hoses
3. Remote Monitoring
4. Real-time tracking
5. Automated invoice generation
6. Electronic approval





Case Study: Dynamic Pricing Analysis

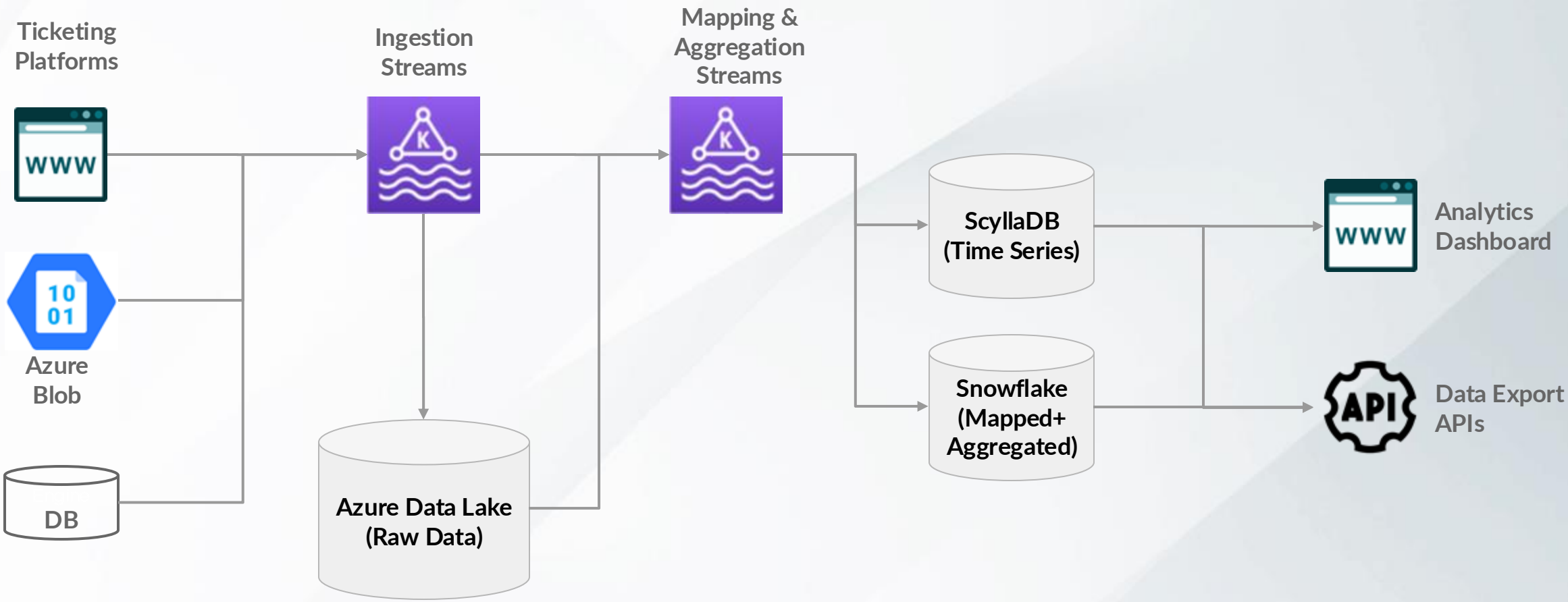
Requirements

- Platform ingests data from multiple ticketing engines to provide a dynamic pricing analysis to brokers
- Customers (brokers) require access to near real time streaming and batch analysis of pricing data
- Latency for pricing updates expected to be less than 1 minute
- Millions of data records per day to be ingested and processed, thousands of brokers accessing the platform using Analytics Dashboard as well as Export APIs

Solution

- Streaming ETL Engine using Spark on clusters with implicit data parallelism and fault tolerance
- Data Lake using Azure Data Lake (HDFS Parquet)
- Pricing Time Series Data in ScyllaDB and PostgreSQL containing mapped and aggregated data
- Orchestration using Airflow
- Customers consume near real-time data streams for various data sources per need
- “Pub/sub” model for data delivery, with available REST endpoint for data export

Data Flow



Data Sources → ETL & Data Lake → Analytical Data Marts → Real Time + Batch Data Export

Case Study: AI based Audience Projections

Background

Client would like to attain the capability to perform audience forecasting in a consistent and reliable manner for various demographics for six networks so that they can:

- Reduce the manual efforts of maintaining their projected estimates/revenue
- Price their inventory better
- Incorporate updated estimates into ongoing program scheduling decisions
- Track Performance on a daily basis - planned vs actual for Audience delivery to further enable current and projected audience delivery vs. revenue goals.

Solution

- To perform data ingestion and matching of the diverse data sources to populate data lake based structures designed to support the AI/ML models built in the cloud
- The predictions of the models provide a point and a range estimate for audience projections which is then exposed through API
- The APIs serve a frontend application with a projections workflow that allows ad-sales analysts to consume, edit and finalize projections for their viewership and associated ad-revenue.

Architecture Diagram