RAJALAKSHMI ENGINEERING COLLEGE

Department of Artificial Intelligence and Data Science

FINANCE & BANKING - ATM CASH DEMAND FORECASTING

Dr. Suresh Kumar S

RAMESH M (2116231801134)

SARATH KUMAR P (2116231801156)

SRICHARAN N (2116231801170)

Problem Statement and Motivation

- Banks face challenges in accurately forecasting daily ATM cash demand across multiple locations.Existing manual or statistical systems fail to account for complex seasonal patterns, holiday effects, and regional variations.

- Inefficient forecasting results in cash shortages, overstocking, and increased operational costs.

Motivation

- Growing transaction volumes and digital banking require data-driven decisions for efficient cash logistics.
- The need to move from intuition-based forecasting to an automated, scalable, and intelligent system.

Existing System

Manual Forecasting:

- Based on experience and simple spreadsheets.
- Inconsistent, error-prone, and not scalable for large ATM networks.

Simple Statistical Models:

- Techniques like moving averages or exponential smoothing.
- Cannot capture multiple seasonalities or holiday impacts.

Legacy IT Systems:

- Outdated architecture with limited data integration and analytic capability.
- Lack support for multi-source datasets and real-time analytics.

Objectives

Project Objective:

To design and deploy a robust, end-to-end big data pipeline that accurately forecasts ATM cash demand and supports data-driven decision-making in financial institutions.

Specific Objectives:

- **Build a Scalable Architecture:** Implement a distributed pipeline using HDFS (Databricks Volumes), Apache Spark, and Apache Hive for large-scale data processing and storage.

- **Integrate Multi-Source Data:** Ingest and unify 65 raw datasets into a single, high-quality time-series dataset ready for analysis.

- **Develop Intelligent Forecasting:** Train a Prophet-based model to forecast daily cash demand for the next 90 days, capturing trends, seasonality, and holiday effects.

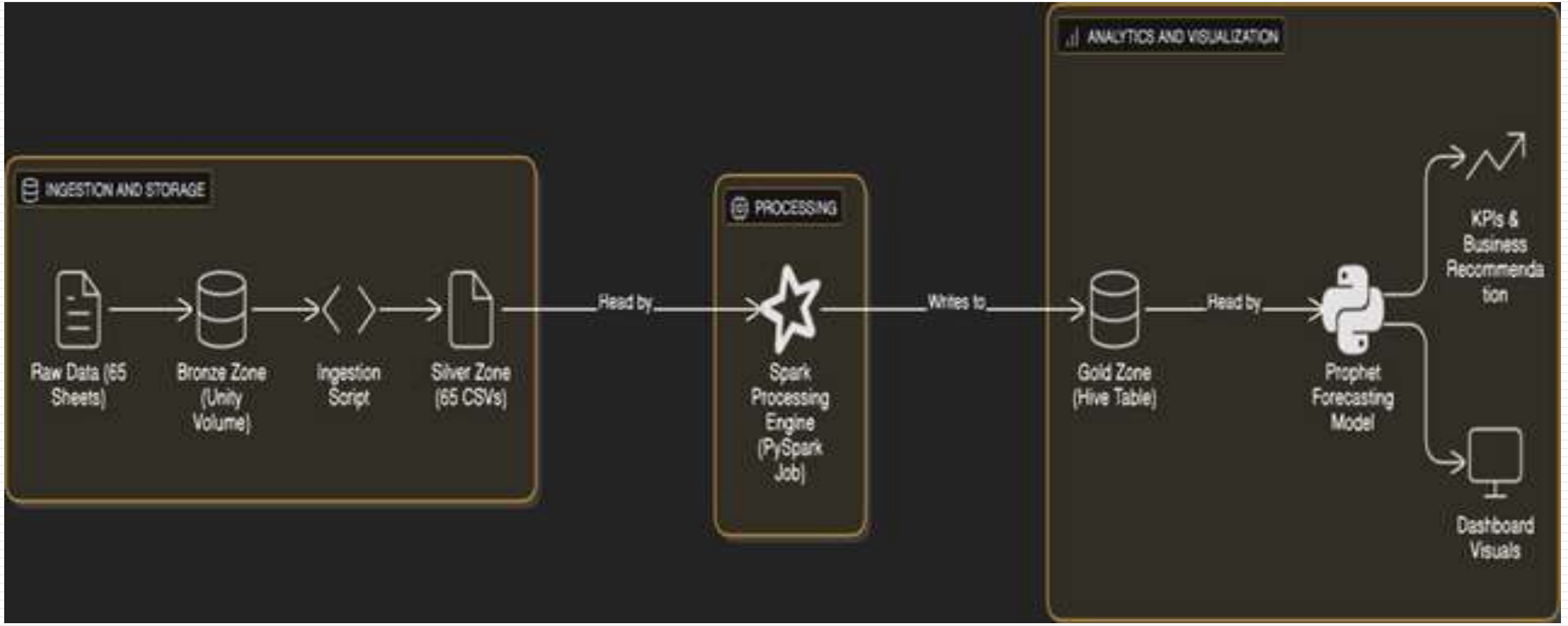
Tools Used

Apache Spark – Core distributed processing engine for large-scale data transformation and aggregation.
Apache Hive – Acts as a data warehouse layer to store and query structured data in the Gold zone using HiveQL.
Databricks Lakehouse Platform – Cloud-based unified environment for data engineering, analytics, and machine learning.
Prophet (Meta AI) – For accurate time-series forecasting with trend, seasonality, and holiday effects.
Pandas handled data ingestion and cleaning.
PySpark managed distributed processing.
Matplotlib was used to visualize forecasts and insights clearly.

Abstract

This project focuses on developing a scalable and intelligent solution for forecasting daily ATM cash demand using advanced data engineering and predictive analytics. The system is built on a two-stage big data architecture implemented within the Databricks Lakehouse Platform, leveraging Apache Spark, Hive, and the Prophet forecasting model. In the first stage, 65 heterogeneous datasets are ingested, cleaned, and structured through a Bronze–Silver–Gold data pipeline, ensuring high data quality and consistency. In the second stage, the refined data is used to train the Prophet model, which accurately captures long-term trends, weekly and yearly seasonal patterns, and the impact of regional holidays. The outcome is a 90-day cash withdrawal forecast, visualized through dashboards and key performance indicators (KPIs) that empower financial institutions to make data-driven cash management decisions. This system transforms traditional manual forecasting into an automated, predictive, and cost-efficient process, significantly improving operational planning and service reliability across ATM networks.

ARCHITECTURE DIAGRAM



Implementation

Setup Environment – Configured Databricks workspace, created Spark cluster, and installed required Python libraries.
Data Ingestion – Imported 65 Excel sheets (ATM data) from Bronze zone and converted them into clean CSV files in Silver zone.
Data Processing – Used PySpark to clean, transform, and merge all datasets into a unified time-series DataFrame.
Data Warehousing – Stored processed Gold-layer data in a managed Hive table for querying and analysis.
Forecasting – Trained Prophet model on Gold data to predict 90-day ATM cash demand with seasonal and holiday effects.
Visualization & KPIs – Created forecast plots, heatmaps, and extracted key metrics for business insights.

Conclusion & Future Scope

Developed a complete end-to-end big data pipeline for forecasting ATM cash demand.Implemented a scalable architecture using Databricks, Apache Spark, and Hive for large-scale data handling.Consolidated 65 multi-source datasets into a clean, structured time-series dataset.Applied the Prophet model to generate accurate 90-day forecasts with trend, seasonality, and holiday analysis.Enabled data-driven decision-making for cash logistics, reducing operational costs and improving ATM uptime.Demonstrated the power of modern data engineering and AI forecasting in financial applications.

Future Scope

- Add External Data Sources
- Real-Time Forecasting
- Interactive Dashboards

References

1.Taylor S.J., Letham B. (2018). Prophet: Forecasting at Scale. PeerJ Preprints.

2.Zaharia M. et al. (2016). Apache Spark: A Unified Engine for Big Data Processing. ommunications of the ACM.

3.Shvachko K. et al. (2010). The Hadoop Distributed File System. IEEE Symposium on Mass Storage Systems and Technologies.

4.Thusoo A. et al. (2009). Hive: A Warehousing Solution Over a Map-Reduce Framework. Proceedings of VLDB Endowment.

5.Databricks Documentation (2025). Unity Catalog Volumes. Retrieved from docs.databricks.com

.

Thank You