

Department of Artificial Intelligence and Data Science

Retail & E-Commerce — Inventory Stockout Prediction and Auto-Replenishment

Dr.Suresh Kumar.S

Professor,

Artificial intelligence and Data science

**GAURAV RAMASUBRAMANIAM-
231801038**

HARISH TUTU YT-231801050

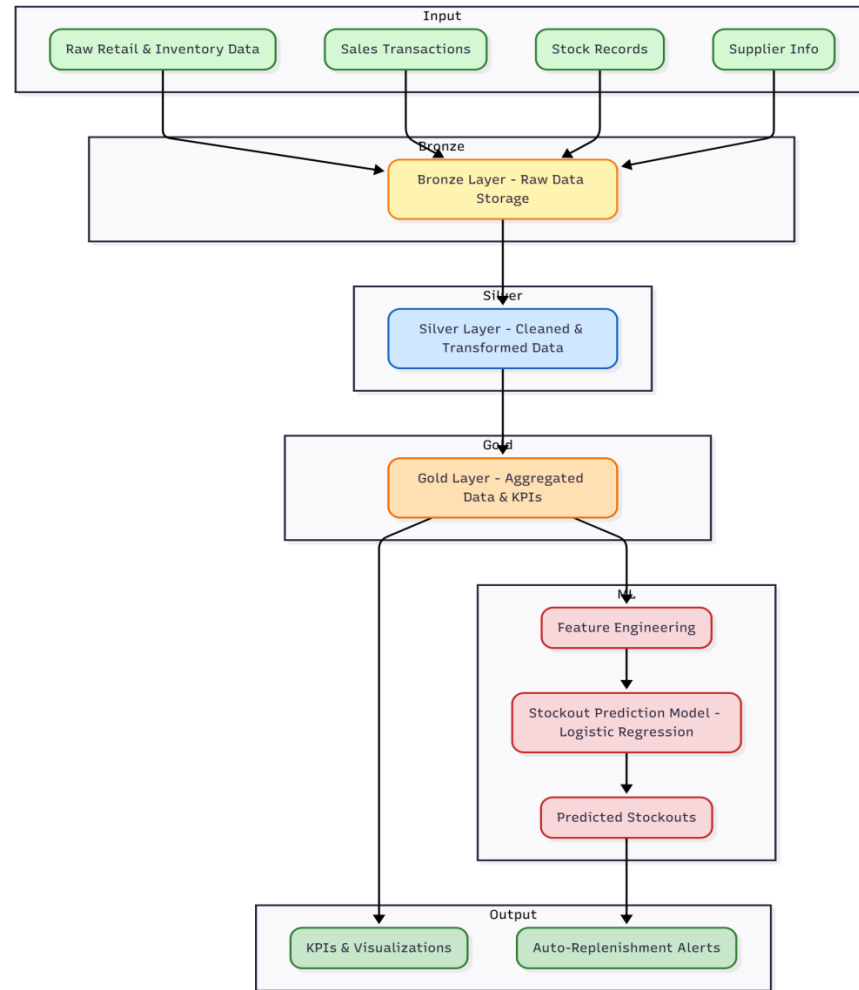
Abstract

□ This project focuses on building an **Inventory Analytics system** that transforms raw transactional data into actionable insights through a **multi-layered data pipeline**. Using PySpark, the system processes data into bronze, silver, and gold layers, enabling KPIs such as average daily demand, demand-to-stock ratio, and lead time analytics. The platform supports dashboards for inventory optimization, helping warehouses improve stock management and replenishment efficiency.

Literature Survey

- ❑ Traditional inventory systems rely on static reporting and often fail to predict stockouts.
- ❑ Modern analytics leverage **Big Data frameworks** (like Apache Spark) for **real-time processing** and **predictive insights**.
- ❑ KPI-driven dashboards help decision-makers visualize trends and optimize inventory replenishment.
- ❑ Our project builds upon layered ETL concepts (Bronze → Silver → Gold) to ensure **data quality and usability**.

Architecture Diagram



Modules

- ❑ **Data Ingestion Module:** Reads transactional data from CSV/Delta.
- ❑ **Data Cleaning & Transformation Module:** Handles nulls, type conversions, and aggregates features.
- ❑ **KPI Calculation Module:** Computes metrics like avg_demand_to_stock_ratio, avg_lead_time_days, and transaction counts.
- ❑ **Dashboard Module:** Generates visualizations for warehouse-level insights.
- ❑ **Access & Security Module:** Implements table permissions and ensures data sharing.

Implementation

- **Framework:** Apache Spark with PySpark APIs.
 - **Layers:**
 - . Bronze: Raw ingestion.
 - . Silver: Cleansing and type conversion.
 - . Gold: Aggregation and KPI computation.
 - **KPI Computation Examples:**
 - . Avg daily demand per warehouse
 - . Avg on-hand quantity per SKU
 - . Avg demand-to-stock ratio
 - **Dashboards:** Configured in Databricks for heatmaps, area charts, and KPI visualizations.
-

Results

- KPIs generated provide clear warehouse and SKU-level insights.
- **Visualizations include:**
 - Heatmaps for demand-to-stock ratio.
 - Area charts for daily demand vs. stock.
 - Bar charts for warehouse lead time comparisons.
- Gold layer ensures all metrics are reliable and ready for business decision-making.
- Performance metrics show the pipeline handles large datasets efficiently (~500,000+ rows).

Conclusion

- ❑ The project successfully implements a multi-layered inventory analytics pipeline.
- ❑ Clean, aggregated data in the gold layer enables accurate **KPIs** and effective visualization.
- ❑ **Dashboards** allow actionable decision-making for warehouse stock optimization.
- ❑ **Future work:** Incorporating predictive analytics, automatic reorder suggestions, and supplier performance scoring.

References

- Databricks, *Delta Lake Guide*, <https://docs.databricks.com/delta/index.html>
- Apache Spark Documentation, <https://spark.apache.org/docs/latest/>
- “Big Data Analytics for Inventory Management,” International Journal of Advanced Research in Computer Science, 2022
- Kimball, R., *The Data Warehouse Toolkit*, Wiley, 2013
- Tableau/Power BI Dashboard Best Practices, 2023

Thank You
