end-to-end-data-flow

TALABAT

Comprehensive Analysis & Pipeline Design

Prepared by:
Amatalrahman Sayed

Data Engineering Use Case – Talabat

Comprehensive Analysis & Pipeline Design

1. Introduction

Talabat is one of the leading food and grocery delivery platforms in the MENA region. Its operations rely on a robust data engineering ecosystem that ensures seamless ordering, fast delivery, and personalized user experiences. This document provides a structured breakdown of Talabat's potential data architecture based on known industry practices and publicly available information.

2. Data Sources

Talabat collects data from multiple layers of its ecosystem, including:

• Application events: user profiles, interactions, searches, app clicks, and session logs.

• Order lifecycle data: timestamps, items, delivery address, pricing, promotions.

• Delivery fleet data: real-time rider GPS, route progress, delivery status events.

• Vendor & restaurant systems: menus, prices, discounts, preparation time, availability.

• Payment gateways: wallet activity, credit card transactions, refunds, failed payments.

• External services: traffic APIs, weather feeds, and location services.

3. Storage Layer

Talabat utilizes multiple specialized storage systems for performance, scalability, and flexibility:

Data Lake: Stores raw structured and unstructured data, including logs, clickstreams, and rider telemetry. Ideal for large-scale storage and future-use analytics.

Data Warehouse: Holds structured and optimized analytical datasets used for dashboards, KPIs, financial reports, and operational analysis.

Operational Databases: Support real-time transactional workflows like order creation, tracking, and user authentication. These often include SQL stores, NoSQL systems, and caching layers.

4. Processing Layer

Talabat's processing workflows include:

Real-Time Stream Processing: Used for live rider tracking, ETA updates, fraud detection, and service availability monitoring.

Batch Processing: Enables daily analytics, machine-learning training, demand forecasting, and customer segmentation.

Orchestration: Tools like Airflow schedule, monitor, and manage data pipelines efficiently.

5. Serving Layer

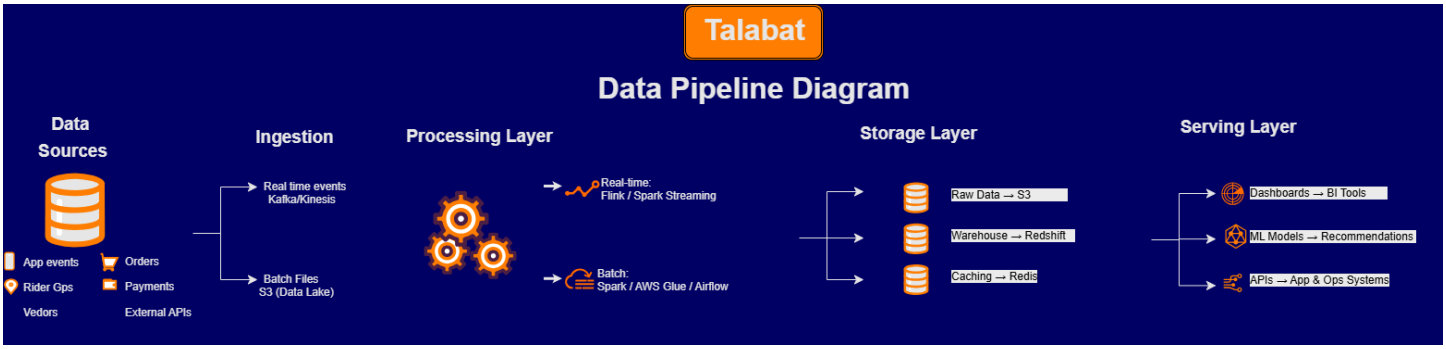
• Dashboards for operations, finance, supply chain, and marketing teams.

• Machine learning models powering recommendations, ETA predictions, and fraud detection.

• APIs supporting real-time order monitoring, vendor insights, and customer notifications.

6. High-Level Data Pipeline Diagram

The following diagram illustrates the main layers of Talabat's data flow:



7. Use Cases

• Accurate ETA prediction using historical data and real-time GPS telemetry.

• Personalized product recommendations based on behavioral and contextual data.

• Automated fraud detection integrating transaction patterns and user behavior.

• Operational optimization such as demand forecasting and delivery fleet balancing.

References

- Delivery Hero Data Engineering Articles

- AWS Case Study for Talabat

- Miracuves – Talabat App Architecture Summary

- Actowiz – Talabat Data Sources Analysis