Talabat Data Engineering Pipeline Analysis

1. Data Sources

Main sources of data in Talabat's ecosystem:

1.1 Mobile apps & website

order events, cart actions, search, location data, talabat Pro subscriptions

1.2 Restaurants & vendors

menu updates, pricing, inventory status, prep time

1.3 Delivery riders:

real-time GPS tracking, delivery status changes, ETA updates

1.4 Payment gateways

transaction confirmations, refunds, disputes, Talabat Pay events

1.5 Grocery & pharmacy partners

fulfillment confirmations, inventory sync

2. Data Ingestion Layer

2.1 Real-time ingestion

- Apache Kafka (or Amazon MSK) as the central event streaming platform for all live events (new order, status change, rider location)
- API Gateway + webhooks from external partners (vendors, payment providers)

2.2 Batch ingestion

- Daily CSV/JSON files from legacy vendors uploaded via SFTP → S3
- Scheduled pulls orchestrated by Apache Airflow

3. Storage Layer

3.1 Data Lake (Raw Zone)

- Amazon S3 – raw event data stored in Parquet + JSON Lines

3.2 Data Warehouse (Processed Zone)

- Google BigQuery (primary analytical warehouse) or Snowflake – cleaned and modeled data

3.3 Operational Databases

- Amazon Aurora PostgreSQL + DynamoDB – low-latency reads for order status, rider location, inventory

3.4 Cold Archive

- S3 Glacier Deep Archive – compliance and audit logs older than 2–3 years

4. Processing Layer

4.1 Pipeline type

- Primarily ELT: raw data lands in S3/BigQuery → transformations via dbt

4.2 Batch processing

- Apache Spark on EMR/Dataproc for heavy aggregations, daily settlements, revenue reporting
- Orchestration with Apache Airflow

4.3 Real-time processing

- Kafka Streams / ksqlDB or Apache Flink for live ETA calculation, fraud signals, leaderboard updates.

4.4 Cleaning & transformation

- Deduplication, timestamp standardization (UTC), geolocation enrichment, ID mapping (user-rider-vendor).

5. Serving Layer

5.1 Dashboards & BI

- Looker (main BI tool) connected to BigQuery – order volume, delivery performance, vendor KPIs, cancellation rates, talabat Pro metrics.

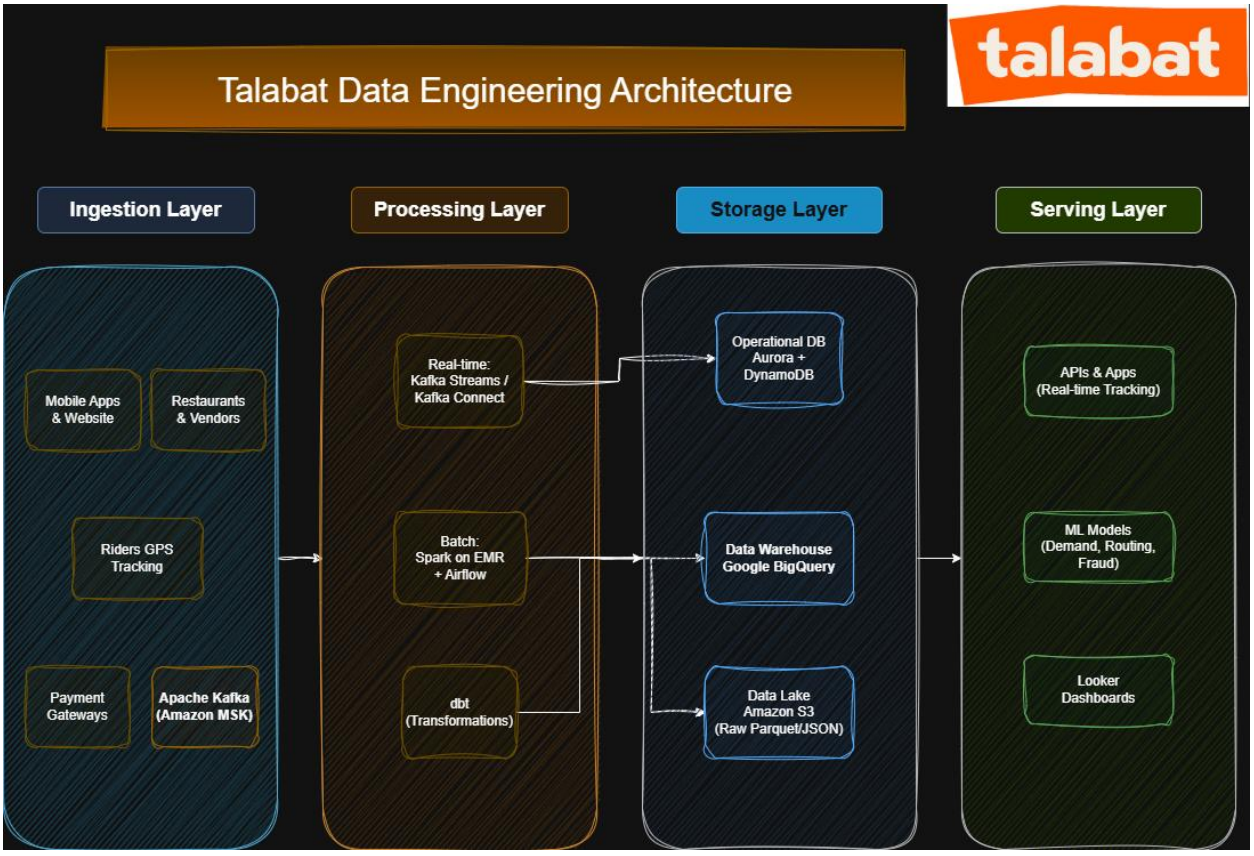
5.2 Machine Learning models

- Demand forecasting, route optimization, recommendation engine, fraud detection, churn prediction (deployed on Vertex AI or SageMaker) .

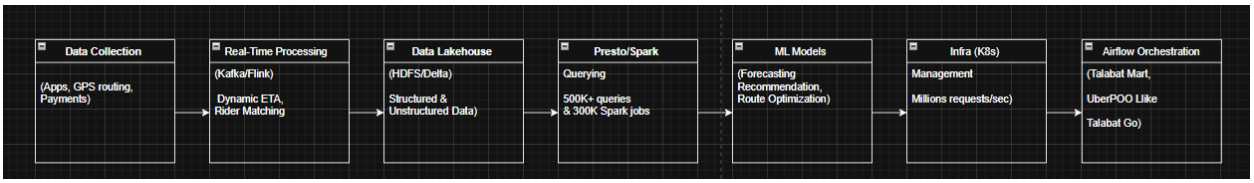
5.3 Application services & APIs

- REST/GraphQL APIs (API Gateway + Lambda) powering real-time order tracking, order history, vendor and rider dashboards.

6. Pipeline



talabat abstracted data pipeline flow



Use Case

Example: Placing a Food Order on Talabat

1. The user browses the app/website, adds items to the cart, and places an order (food, groceries, or pharmacy).

2. The order event is instantly captured and pushed to Apache Kafka as a real-time stream.

3. Real-time processors (Kafka Streams / Flink) validate the order, check inventory with the restaurant, assign the nearest available rider, calculate dynamic ETA, and process payment via the gateway or Talabat Pay.

4. The raw order event (including GPS, items, prices, timestamps) is immediately landed in the Data Lake (Amazon S3 – Parquet/JSON).

5. Batch jobs (Spark on EMR + Airflow) run every few hours/daily: clean the data, enrich with geolocation and vendor categories, deduplicate, and load the transformed data into Google BigQuery (Data Warehouse).

6. Looker dashboards automatically refresh showing updated metrics: order volume, average delivery time, cancellation rate, revenue per city, restaurant performance, and rider efficiency.

7. The restaurant receives instant preparation notification, the rider gets the pickup assignment with optimized route, and the user sees live tracking; settlement files are generated and sent to restaurants and payment partners at the end of the day.

References

- **Talabat Official Website – Company Overview** https://corporate.talabat.com/
- **Talabat Tech Blog (Medium) – Engineering and Data Insights** https://medium.com/talabat-tech
- **Delivery Hero Tech Blog – Talabat Technology and Innovation** https://tech.deliveryhero.com **(articles on Talabat's tech practices and events)**
- **Talabat Integrated Annual Report 2024 – Company Performance and Market Overview** https://ir.talabat.com/wp-content/uploads/2025/04/20250404_Talabat-Annual-Report-2024-En.pdf
- **Khaleej Times – Talabat Fintech and Innovation Profile** https://www.khaleejtimes.com/business-technology-review/fintech-innovation-remains-a-major-focus-area-for-talabat