Fawry Data Engineering Pipeline Analysis

1. Data Sources

Fawry operates one of the largest digital payment networks in Egypt, which means it collects data from several major sources:

- **POS Terminals and Agent Network:** Payment requests, transaction logs, timestamps, service types, and merchant IDs.
- **Mobile App & Website (myFawry):** User account data, wallet top-ups, bills, e-commerce payments, and behavioral events.
- **Merchant & Corporate Integrations:** Invoices, settlements, reconciliation files, and B2B payment data.
- **Banking & Financial Partners:** Transaction confirmations, settlement reports, and compliance-related data.
- **Service Providers (e.g., utilities, telecom):** Bill status updates, payment confirmations, and service charges.

2. Storage Layer

Fawry would typically rely on a multi-layer storage architecture to handle its massive transaction volume:

- **Data Lake (Raw Storage):** Stores unprocessed data from POS, app logs, API events, and external integrations in formats like JSON, CSV, or Parquet.
- **Data Warehouse:** Cleaned and structured data used for analytics, financial reports, dashboards, and regulatory compliance.
- **Operational Databases / NoSQL Stores:** Used for real-time operations such as wallet balance checks, instant confirmations, and fraud flagging.
- **Cold Storage / Archive:** Long-term storage for older transaction logs, audit trails, and compliance data.

3. Processing Layer

Fawry requires both batch and real-time processing capabilities:

- **Real-Time Stream Processing:** Handles instant validation of payments, fraud detection, balance updates, and event-driven alerts.
- **Batch Processing:** Daily or hourly ETL/ELT jobs to transform raw data into the Data Warehouse and generate settlement reports.
- **Data Cleaning and Enrichment:** Removing duplicates, fixing inconsistencies, and enriching transactions with merchant details, geolocation, or service category.

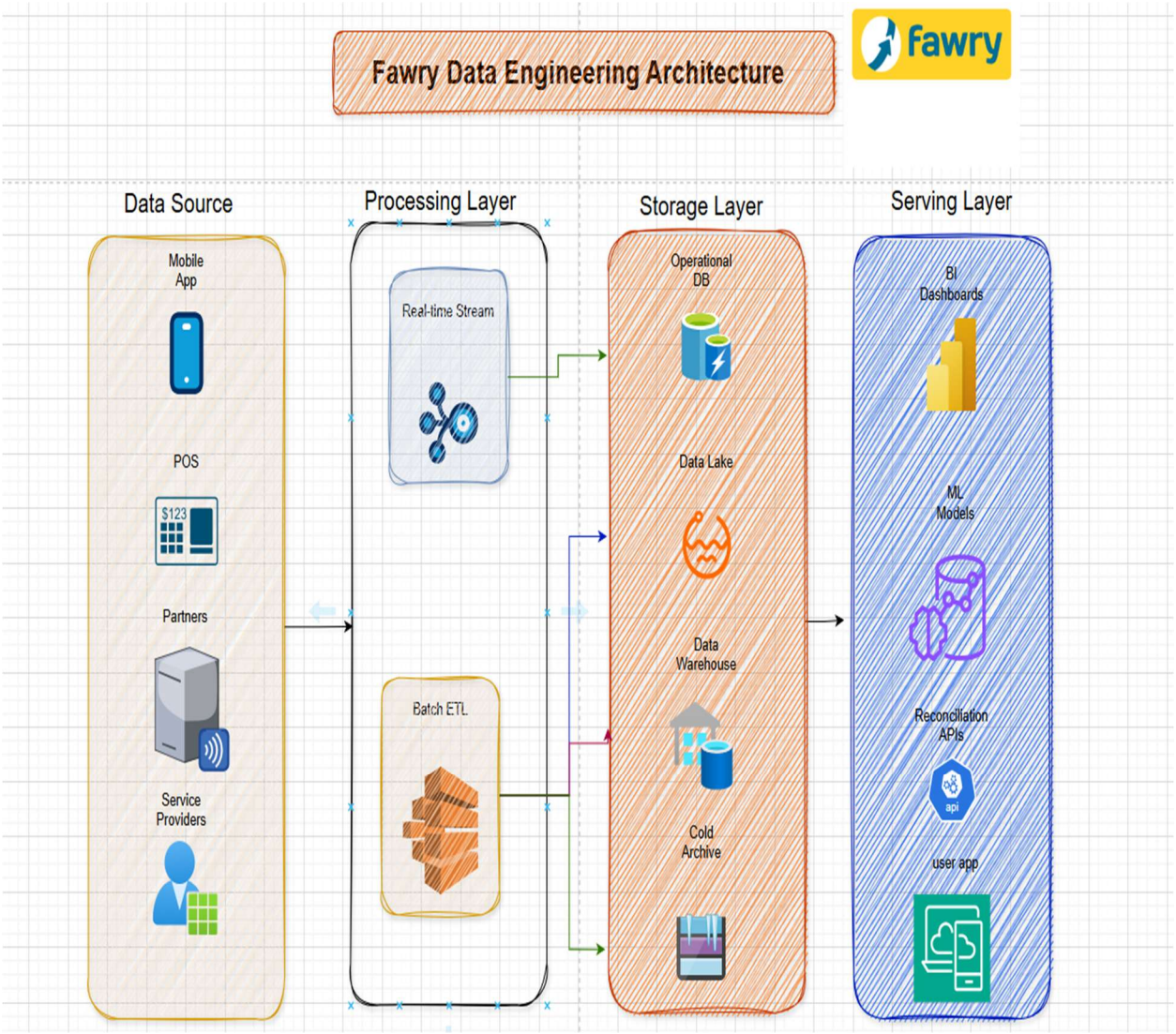
- **Aggregation:** Creating summary tables for daily transactions, merchant performance, and service usage trends.

4. Serving Layer

Once processed, the data supports several business functions:

- **Dashboards & BI Reports:** Used by management, finance, and operations to track revenue, transaction volumes, agent performance, and service usage.
- **Real-Time Application Services:** Wallet balance display, transaction history, instant confirmations, and agent portal updates.
- **ML-Driven Features:** Fraud alerts, personalized recommendations, transaction risk analysis, and predictive analytics.
- **User App / Website:** Financial reconciliation, settlement processing, auditing, and compliance reporting.

5. Pipeline



Use Case

Example: Paying an electricity bill through Fawry

1. The user pays via agent or the mobile app.
2. The transaction is captured instantly and sent to the system.
3. Real-time processors validate the payment and update balances.
4. The raw transaction is stored in the Data Lake.
5. Batch jobs later move the cleaned version into the Data Warehouse.
6. Dashboards update with daily transaction volumes and settlement info.
7. Utility providers receive confirmation and settlement files.

References

- Fawry Official Website – Company Overview: https://www.fawry.com/about/who-we-are/
- Fawry Partnership Announcements and Transaction Insights: https://www.fawry.com/2025/03/19/fawry-and-contact-join-forces-to-transform-egypts-e-payment-system/
- Fawry – Egypt's leading digital payments platform
- Fawry for Banking Technology and Electronic Payments - The Middle East's Fintech 50 - Forbes Lists