TASK 3: Architecture Design Challenge

1 .Traditional Data Warehouse:

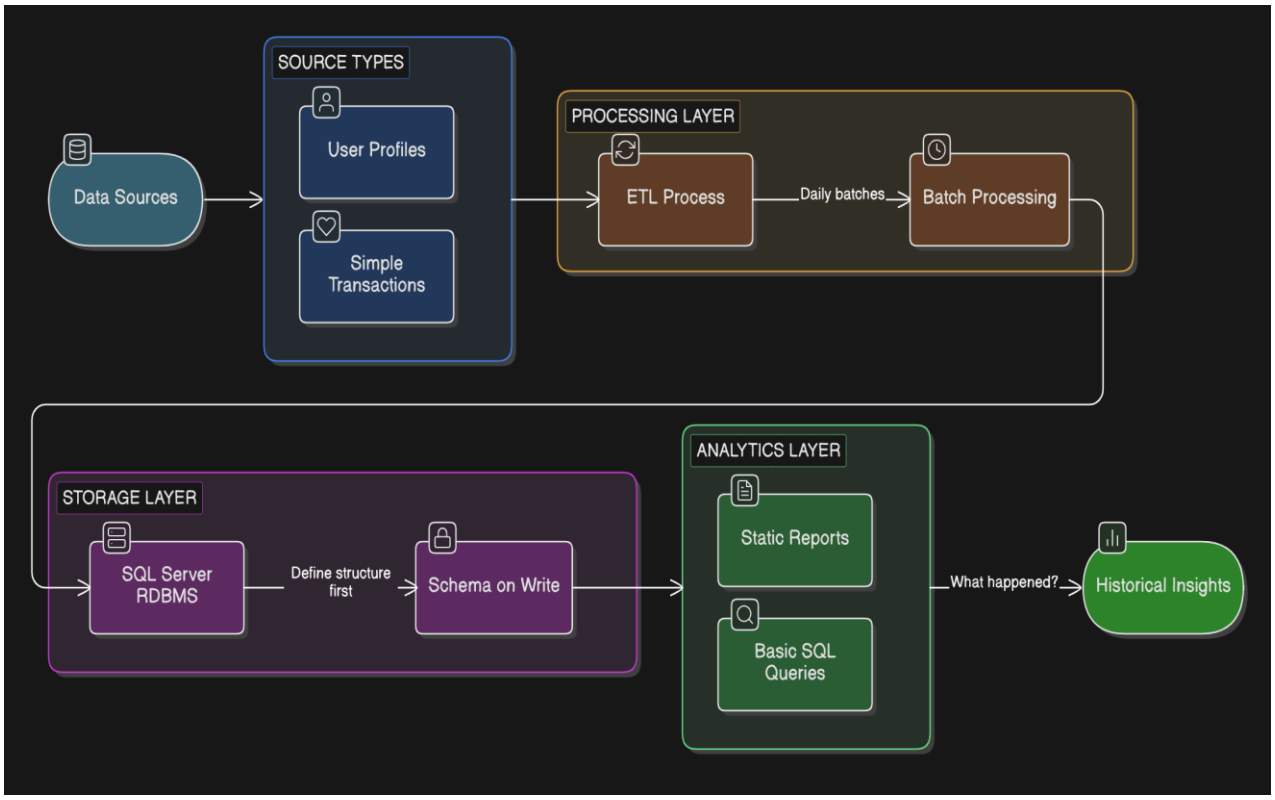

Fig 1.1

- **Data Sources:** Transactional databases (User login info, profile settings) and manual log entries.
- **Processing Layer: ETL (Extract, Transform, Load)**. Data is cleaned and converted into rows and columns in slow, nightly batches.

- **Storage: RDBMS** (Relational Database Management System) like Oracle or SQL Server. It uses **Schema-on-Write**, meaning the "shape" of the data is fixed.
- **Analytics Layer:** Basic **SQL Queries** and static BI dashboards

2. Hadoop based Big data Architecture

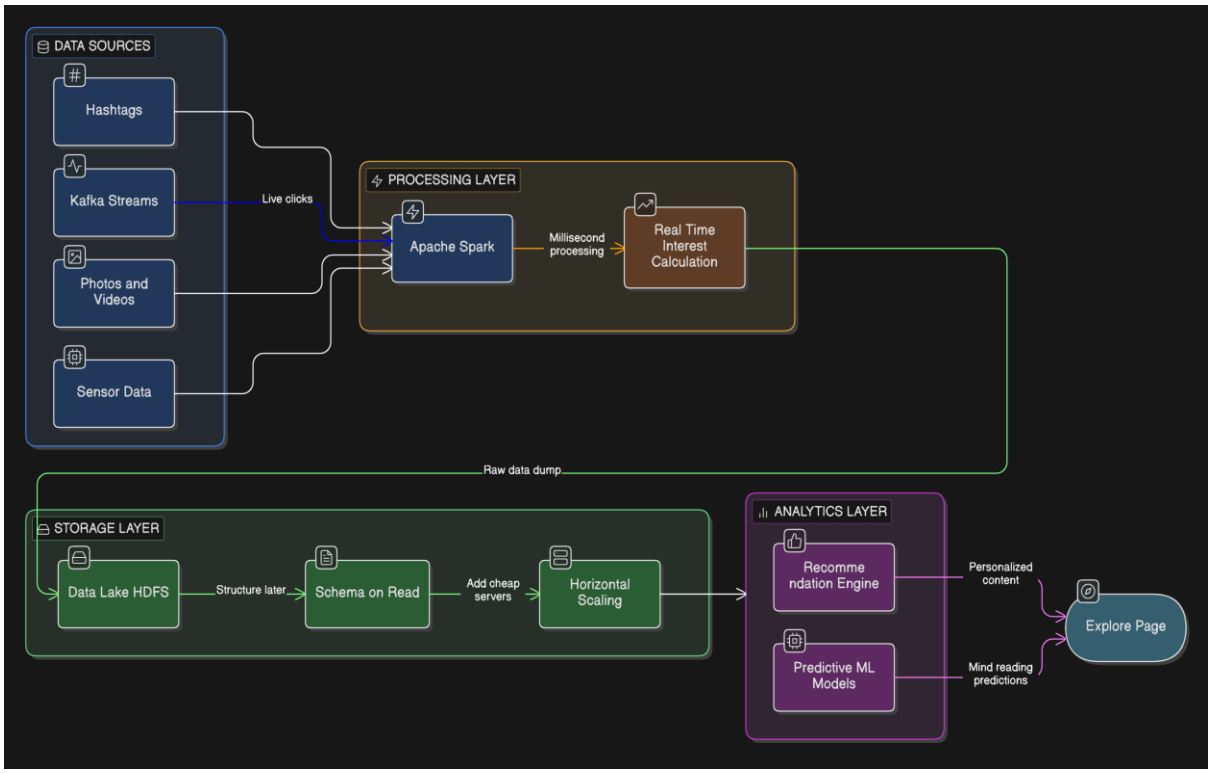

Fig 1.2

- **Data Sources:**

  Structured: Profile info, billing.

  Unstructured/Semi-Structured: High-res photos, Reels (Video), JSON activity logs (what you clicked), and hashtags.

- **Processing Layer:** Apache Spark: For real-time processing (updating your feed instantly).

  MapReduce: For massive overnight calculations across thousands of servers.

- **Storage**: HDFS (Hadoop Distributed File System) and NoSQL (like Cassandra). This is a Data Lake using Schema-on-Read—you store the raw data first and ask questions later.
- **Analytics Layer:** Machine Learning Models (Predicting which ads you'll click) and Recommendation Engines (The "Explore" page)