**Part 2: Real-Time E-commerce Data Processing and Analysis**

**Objective**

Develop a system to process and analyze e-commerce data in real-time, focusing on customer behavior and transaction analysis.

Tools and Technologies

* Stream Processing: Apache Kafka and (Apache Flink or Apache Spark Streaming)
* Database: Any suitable database for storing processed data (e.g., PostgreSQL, MongoDB)

**Dataset**

Real-time e-commerce transaction data (simulate using a script or use a real-time data generator)

**Tasks**

1. Data Ingestion and Streaming: Set up Kafka to ingest real-time e-commerce transaction data.
2. Stream Processing: Use (Flink or Spark Streaming) for real-time data processing – filtering, aggregating, and enriching the data.
3. Data Storage: Store the processed data (summaries and insights) in a **HADOOP HDFS** for further analysis and historical record-analysis.
4. Generate visual reports with data summaries and insights.

**Implementation**

* Implement a Kafka producer to simulate real-time e-commerce transactions.
* Set up Kafka topics to categorize different types of transaction data.
* Use Apache Flink to process the streams – calculate metrics, identify patterns, and detect anomalies in real-time.
* Store the processed data in a database.