Advanced Database Management

Topic

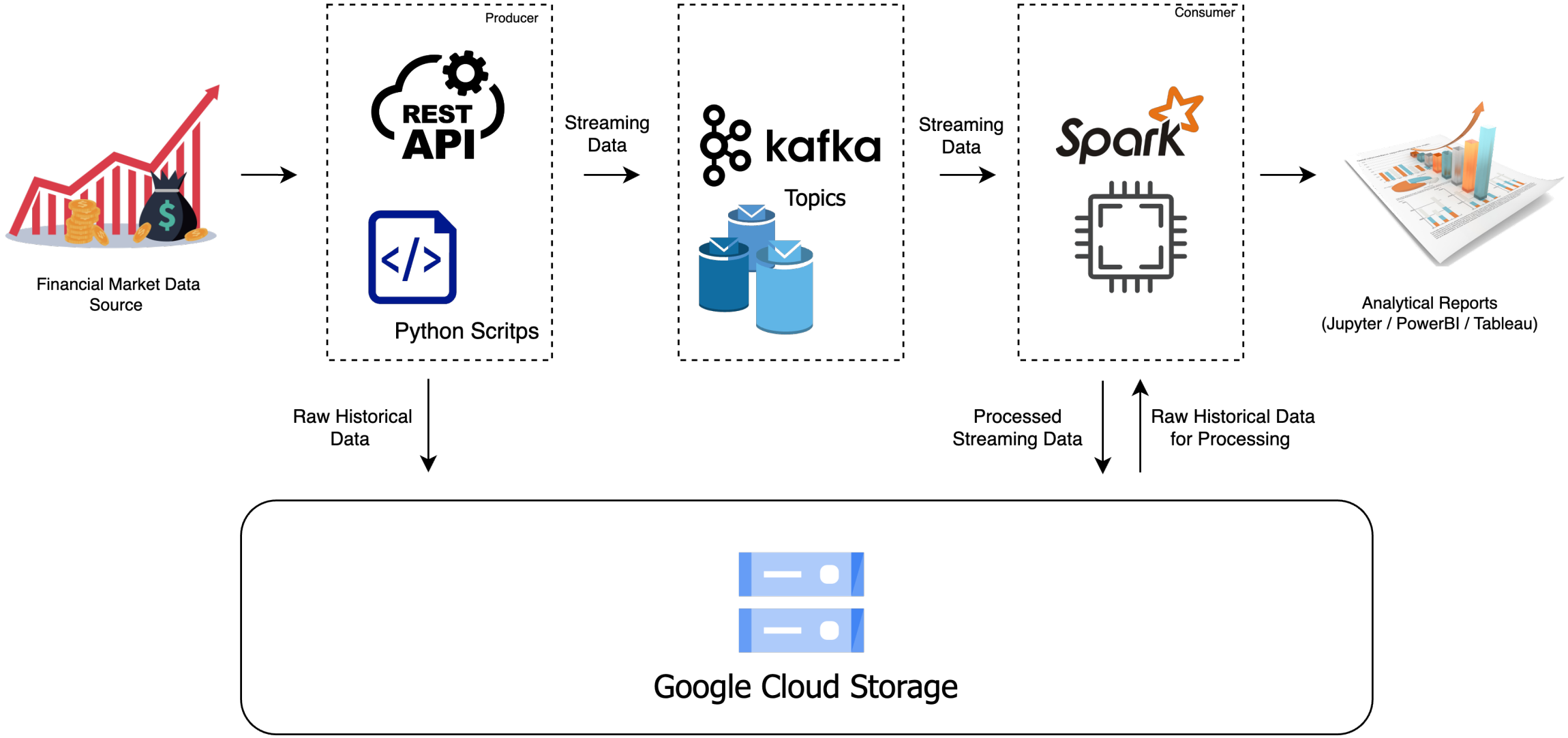
Financial Market Data Analysis

Team Members

Tushar Yadav
Dhruti Dilipbhai Patel

Problem Statement

Architecture Design



Architecture Design

1. **Financial Market Data API**:
   This component represents the external API you're using to fetch financial market data. For this we are using Alpha Vantage.

2. **Python Script**:
   A Python script or application responsible for making API requests, retrieving streaming data, and potentially transforming it. Instead of saving the data locally, it processes and sends the data to Apache Kafka topics.

3. **Apache Kafka**:
   Kafka receives the streaming data from the Python Script and acts as a message broker. It stores the data temporarily and allows multiple consumers (e.g., Spark, other applications, etc.) to subscribe to and process the data.

4. **Apache Spark** :
   Apache Spark will be used for more complex data processing, analysis, and transformation. It can read data from Kafka and perform real-time or batch processing as needed and also grab historical data from data storage and process data.

Architecture Design

5. Google Cloud Storage:

A Google Cloud Storage bucket where the processed financial market data and unprocessed historical data is stored. This serves as a durable and scalable big data storage solution.

6. Data Processing and Analysis :

This represents the component or tools (e.g., Python, Spark, etc.) used for analyzing and visualizing the stored data in AWS S3. It could involve generating reports, charts, and insights from the data.

7. Analytics and Representation :

This layer is the dashboard or presentation layer where we try to understand and analyze the final processed data using graphs and charts.

Streaming Data Architecture Design

DATASOURCE

SAMPLE DATA

REFERENCES