ANALYSIS OF STOCK MARKET DATA

Idea

The idea is to stream real-time NYSE stocks data and perform stock analysis on it.

Approach

The Alpha Vantage API provides stock market data from various countries around the world. It covers a wide range of stock exchanges, including those in the United States, Europe, Asia, and other regions. Some of the major stock exchanges covered by the Alpha Vantage API include: New York Stock Exchange (NYSE) NASDAQ.

To access the Alpha Vantage API, you need to sign up on their website and obtain an API key. The API key is required to authenticate your requests and access the data.

Kafka producer uses the API key to access the stock data and sends it to the kafka topic. Pyspark streams the topic in integration with kafka and sends it to the. The file is retrieved from hdfs and various graphs and queries are implemented on the data.

Benefits:
The user can produce candlestick graphs and histograms to see the stock trends. The user can also get the historical data of the stocks and even query all time high and low for a particular stock of the day.

Implementation

The following steps are involved in implementing this approach:

1. Create a Kafka topic.

2. Set up the ALPHA VANTAGE API and the Kafka producer.

3. Set up the PySpark streaming engine.

4. Send the data from API to the topic using producer and stream it on pyspark engine.

5. Store the data to Hadoop HDFS.

6. Visualize the stock data by fetching the files from the HDFS and producing candlestick charts.

PLAN:

WEDNESDAY: CREATION OF KAFKA TOPIC AND KAFKA PRODUCER FILE.

THURSDAY: TEST STREAMING OF DATA USING KAFKA CONSUMER.

FRIDAY: EXTRACTING KAFKA TOPIC DATA USING PYSPARK AND STORING IT IN HDFS

SATURDAY: PERFORMING GRAPH ANALYSIS ON THE DATA