

Big Data Pipeline for Real-Time Stock Market Analysis

Binaya Sharma

bsharm6@lsu.edu

This project aims to create a big data pipeline for real-time stock market analysis using stock price data available from Yahoo Finance (<https://finance.yahoo.com>) and/or Alpha Vantage (<https://www.alphavantage.co/#about>). Utilizing Apache Kafka for data streaming, the application will feed the data into Apache Spark for processing and analysis. The processed and aggregated data will be stored in HBase or another database for later in-depth analysis. The goal is to provide a scalable pipeline for monitoring the stock market. Some of the parameters that will be calculated and used in analysis are opening price of the day, closing price of the day, highest price of the day, lowest price of the day, volume, etc. Real time indicators like Moving Average (MA) or Relative Strength Index (RSI) could also be calculated. Using the calculated parameters and indicators, graphical representations can be made and further analysis like buying/selling point can be suggested.