Real Time E-Commerce Analytics

Members:

1. Sai Sri Harsha Kanamanapalli
2. Srikara Krishna Kanduri
3. Srinivas Parasa

Idea

The idea of this project is to handle the high-volume real-time ecommerce streaming data and perform analytics on the orders made from e-commerce websites.

Working

1. Synthesize dummy data equivalent to real time e-commerce data and expose a streaming API.
2. Fetch the data from streaming API.
3. Preprocess the data and convert it into required format.
4. Perform aggregations and analytics to identify the trends in purchases.

In this project, we try to perform aggregations on ratings of a particular product and also run various aggregations on same products from different companies. This way, we would like to display the trends in purchasing and the quality of same products from different companies.

Here, we try to run the aggregations on millions of orders at the same time and the results are displayed on a consolidated dashboard. The backend part of the streaming is processed from kafka engine and the intermediate files are stored in a spark cluster.

Hardware: System capable of running software like Kafka, Apache Spark and Mongo DB and should be able to process large number of requests without having a significant impact on the performance.

Software: Kafka, Apache spark and MongoDB

Dataset: E-commerce streaming APIs containing consumer purchase data. Since, most of the e-commerce firms do not provide the consumer purchase data to external developers, we choose to synthesize our data with different parameters.

What is success for this project?

If the system is able to handle large amounts of data without showing a significant downgrade in the performance and is able to identify the trends in purchase, then