Many people visit online websites to buy clothes. Every website generates log data which stores the number of clicks made by a particular user at every instance of time. Companies use this log data to understand and keep track of customer behaviour.

As the number of users of a website is large so the log data generated will also be very large. Companies find it very difficult to store and process such a large amount of data which is generated in GBs per minute. If not managed properly the data may get lost or such high volume of data can crash the system or computer being used. To overcome these challenges big data technologies like Apache Spark and Apache Kafka are used to extract, process and load data to its destination in real time and also with high fault tolerance.

This application will use Kafka and Spark to pull, process and push the data in real time with minimum time delay and it enables flexibility that anyone with proper and correct credentials can access the final processed data.

**UDHAV MOHATA**

**1SI16TE045**