Abstract

Twitter, one of the largest and famous social media site receives millions of tweets every day on variety of important topic. This large amount of raw data can be used for industrial, Social, Economic, Government policies or business purpose by organizing according to our need and processing. Hadoop is one of the best tool options for twitter data analysis and hadoop works for distributed Big data, Streaming data, Time Stamped data, text data etc. This paper discuss how to use FLUME for extracting twitter data and store it into HDFS for analysis, and after that we are use apache hive for analysing these data. We perform analysis on twitter data to find the number of tweets are posted location wise and also finds the keywords on which maximum and minimum tweets are posted.

References


Index Terms

Computer Science  Distributed Systems

Keywords
Hadoop, twitter, Flume, social analysis, hive, HWI, JSON