|  |  |
| --- | --- |
| **Project**  **Team 14**  Title: Evaluation of framework performing Sentiment Analysis on Twitter Stream Data  Description: Our project aims to perform sentiment analysis depending on the chosen topic (which will be an input to the system) and find out the major talked about issues in relation to that. We will use Twitter’s real-time streaming data and perform the visualization on the results.  Team Members:  Li Shi – lshi7  Navjot Singh – nsingh9  Shriyansh Yadav – scyadav | **Deliverables**   1. Apache Kafka - to buffer the incoming Twitter Stream data 2. An architectural framework Hadoop ecosystem 3. Components of the ecosystem consisting of HDFS, YARN, and Apache Spark on Amazon’s EC2 instances 4. An application hosted on AWS that will show the analysis in the form of charts/graphs    1. Dynamic to choose the kind of chart/graph based on user    2. Ability to show the analysis based on a particular sentiment |
| **Status**  Architecture will be supported by AWS   * EC2 instances running on AWS * Completed installing a master node with 2 data nodes * with the setup of HDFS architecture on | **Issues**   * Dealing with slow streaming of Twitter Data:   + Since, there is a limit on the number of requests that can be made to the Twitter Stream API we plan to use downloaded datasets * Finding proper techniques to overcome shortcomings in analyzing a Tweet   + Complete expression of thought is limited because of limit on the length of message   + Presence of grammatical/spelling errors, use of emoticons, and colloquialism in the Tweets * Dealing with limit on the number of Free Tier EC2 instances   + Planning to substitute VCLs with longer reservation period. |