

<p style="text-align: center;"><u>Team 14</u></p> <p><u>Title:</u> Evaluation of framework performing Sentimental Analysis on Twitter Stream Data</p> <p><u>Description:</u> 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.</p> <p><u>Team Members:</u> Li Shi – lshi7 Navjot Singh – nsingh9 Shriyansh Yadav – scyadav</p>	<p style="text-align: center;"><u>Deliverables</u></p> <ol style="list-style-type: none"> 1. A web application hosted on AWS that will show the analysis in the form of charts/graphs <ol style="list-style-type: none"> a. Dynamic to choose the kind of chart/graph based on user b. Ability to show the analysis based on a particular sentiment 2. An architectural framework based on Hadoop Distributed File System to handle large stream data and client requests from the web application. 3. Implementation of statistical Natural Language Processing Toolkit for the categorization of sentiments in Tweets.
<p style="text-align: center;"><u>Dependencies</u></p> <p>Architecture will be supported by AWS</p> <ul style="list-style-type: none"> • EC2 instances running on AWS • Hadoop Distributed File System – underlying fault-tolerant, reliable file distribution system • YARN – A resource manager • Apache Spark • Apache Kafka • Twitter Stream API • REST API • NLTK/Wit.ai – A natural language processing toolkit to extract out the sentiment/intent of a text 	<p style="text-align: center;"><u>Issues</u></p> <ul style="list-style-type: none"> • Dealing with slow streaming of Twitter Data: <ul style="list-style-type: none"> ○ 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 <ul style="list-style-type: none"> ○ 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 <ul style="list-style-type: none"> ○ Planning to substitute VCLs with longer reservation period.