A report describing the design and implementation of your tool/software along with details on testing

**Design:**

This entire project can be divided into three modules – Data collection, data processing and Data Visualization. The incremental steps of this project are presented in the following image.



* **Data collection stage:** Inthis stage, data is collected using theTwitter API. To access the twitter api from the java code we need the access tokens. The access tokens are generated for the every application by the twitter. Using these tokens and twitter API is called using the java code. We can collect the tweets based on various selection criteria. We used twitter hashtags to collect the data. We collected the tweets regarding the Nepal earth quake. The hash tags we used are - #nepal, #nepalnarthquake, #prayfornepal, #earthquake, #nepalinneed, #terremotonepal, #earthquakevictims, #helpnepal, #pray\_for\_nepal, #kathmandu, #katmandu, #prayersfornepal, #nepalvictims, #nepalearthquake2015 and #nepalquickrelief .We collected 0.2 million tweets, which accounted for 1GB of data. Loading and analyzing this size amount of data is not feasible using RDBMS is difficult. So, we used hadoop and hive to load and analyze this data.
* **Data Processing:** The tweets collected in the data collection stage are in JSON format. Loading and processing the data in JSON format is difficult. So, this JSON format tweets are converted to CSV format using java. This CSV file is loaded into hive using the hive queries mentioned in the queries section.
* **Data Analysis and Visualization:** The entire data loaded into hive is extracted based on the visualization criteria. The extracted queries and the visualization criteria are provided in the queries and output sections. The extracted and analyzed queries are displayed in the graphical formats using the IBMBigSheets.

**Implementation of tool:**

* Eclipse is used to collect the tweets and convert the JSON file to the CSV file
* IBMBigInsights snapshot is downloaded and is opened in VMWare
* IBMBigInsights comes with preinstalled Hadoop ecosystem.
* Hive in IBMBigInSights is used for data storage, processing and extraction
* IBMBigSheets are used for data visualization

**Queries:**

****

**Testing Details:**

****

**Outputs Screens:**

****

**Team Members:**

Keerthi Yanda

Sandeep Ballu

Rakesh Vistarakula