PRINCIPLES OF BIG DATA

Phase 1

WORD COUNT PROGRAM FOR TWITTER TWEETS

By Team Members:

1) **Harshini Medikonda**
2) **Sri Naga Sarvani Jakkula**
3) **Sadhana Edukulla**

TASKS ACHIEVED:

1. Collected tweets on the topic 'elections' running the code in python shell.
2. Installed Spark on Ubuntu platform.
3. Installed Scala and SBT required for the spark successful installation.
4. Gave the twitter tweets file as input to the scala word count program.
5. Wrote the word count logic in Scala using map and reduce keywords.
6. Got the output files and success files.
7. At the same time, merged the output into single output file instead of multiple using coalesce function.

TECHNOLOGIES USED:

Framework – Apache Spark

Programming – Scala, Python

Other Tools: SBT

NEW THINGS LEARNT:

1) Apache Spark – Scala
2) Collection of tweets using Streaming API
3) Resolving possible occurrences of errors.

BIBILIOGRAPHY:

http://blog.prabeeshk.com/blog/2014/10/31/install-apache-spark-on-ubuntu-14-dot-04/

http://www-03.ibm.com/software/products/en/ibm-biginsights-for-apache-hadoop

http://www.tutorialspoint.com/scala/