Lab Assignment No. 3
(Grp - B)

Sinhgad Institutes

Date

= Title:-

Locate dataset for working on weather data which reads the text input files & finds average for temperature, dew point & wind speed.

= Objectives:-

Students should be able to perform the MapReduce operation in Hadoop.

= Prerequisites:-

1. Hadoop installatn
2. Basics of Hadoop : HDFS, map reduce.
3. Java programming.

= Theory:-

MapReduce :-

MapReduce is a processing technique & a program model for distribute computing based on java.

= Algorithm:-

- Generally, MapReduce paradigm is based on sending the computer to where the data ~~model~~ resides.
- MapReduce program executes in three stages namely map stage, shuffle stage.
- Map-stage:- mapper's job is to process the input data.

- Input file is passed to mapper function. line by line.
- Reduce stage:- This stage of combination of shuffle stage & reduce stage. The reducer's job is to process data that comes from mapper.
- Hadoop sends map & reduce tasks to the appropriate servers in cluster.
- After completion of the given tasks, the cluster collects & reduces data to form an appropriate results.

= Inputs & outputs:-

- MapReduce framework operates on <key, pair> pair., the framework views the input to the job as a set of <key, value> pairs.
- Key & value classes should be in serialized manner by framework & hence, need to implement the writable interface.

=conclusion:-

In this manner we have successfully located dataset & worked on weather dataset which reads text input files & finds average for temperature, dew point & wind speed.