

COL 733 A4

Group No. 9

Anupam Khandelwal(2013CS10212)

Prakhar Gupta(2013CS10245)

Sahil Jindal(2013CS10253)

Shubham Rawat(2013CS10258)

September 2016

1 WordCount program

1.1 Changes in configuration

- Change in mapred for setting port and ip for job tracker.
- Change in mapred-site.xml for setting mapreduce local directory
- Change in no of map tasks.
- Change in no of reduce tasks.

1.2 Writing MapReduce Code

- Wrote an extension to Mapper class specifying key and value for the map function. For map , input is line of words. The keys generated from this are words and value for each key is 1.
- Wrote an extension to Reduce class which takes the key and list of values as input, and maps that key to sum of the values.

1.3 Running MapReduce task

- Compiled Java code as a jar file and ran it on hadoop.
- Specified no of map tasks as 20 and no of reduce tasks as 4 in the config file.
- `jar cf <out_file> <filename>.class`
- `hadoop jar <out_file> <input_dir> <output_dir>`

1.4 Shutting down VMs

- While the job was running, we shut down one VM and observed that the job still completed successfully.
- In the background, the resource manager tried to connect to the shut down VM and realised that it was not reachable and rescheduled the tasks.
- The resource manager logs showed that it tried to reconnect to shutdown VM.

2 Program for Average Grade

- Wrote an extension to Mapper class specifying key and value for the map function. For map , input is a line containing one score. The keys generated from this are course codes and value for each key is the score in that course.
- Wrote an extension to Reduce class which takes the key and list of values (scores) as input, and maps that key to sum of the values divided by length of list of values (average).