**Assignment 3- Report**

**Task 1**

Q1: Can you figure out what are the main steps do we need to run a hadoop mapreduce task (i.e., wordcount here)?

Ans: To run a Hadoop MapReduce task:

First, we create the input files, and put them in the HDFS input directory.

Second, we execute the wordcount java program taking input and output as arguments.

Third, we print the input files used for reference.

Then, the output is displayed from the HDFS output directory.

Q2: What does this command mean — “hdfs dfs -put ./input/\* input”?

Ans: This command copies the content of the local directory named “input” in HDFS directory “input”.

Q3:How many mappers and reducers are launched for executing the above wordcount program?

Ans: 2 mappers and 1 Reducers

Q4: How much time do mappers and reducers spend for the above tasks, separately?

Ans: Mapper: 17318 ms | Reducers: 4391 ms

Q5: After execution, what are the files in the output folder in HDFS, and what content do they contain?

Ans: Contents: part-r-00000 and \_SUCCESS

part-r-00000: This file is the output file created after the execution. Here,

r represents “reducer” operation is performed and 00000 represents the number of reducers which is 1 (count starts from 00000)

\_SUCCESS- File that indicates that the execution is successful.

**Task II**

Q6: How many master and slave containers do you launch separately this time?

Ans: 1 master and 4 slaves.

Q7: Please figure out what a master container/node and a slave container/node are used for.

Ans: Master node manages operations like opening, closing, and renaming files and directories and determines the mapping of blocks to DataNodes. It also manages access to files by clients.

Slaves nodes perform read and write requests from the file system and perform block creation, deletion, and replication upon instruction from the master.

Q8: How many mappers and reducers are launched for executing the above wordcount program?

Ans: 3 Mappers 1 Reducer

Q9: How much time do mappers and reducers spend for the above tasks, separately?

Ans: Mapper: 33578 ms | Reducer: 4724 ms

Q10: What are the two most frequently occurring words, and how many times do they occur?

Ans: the - 42 times | of- 27 times

**Task III**

Q11: Please describe the basic steps in the map function of WordCount.java

Ans: The map function takes one line at a time. It then splits each line into tokens. Then it gives a word and its number of occurrence.

Q12: Please describe the basic steps in the reduce function of WordCount.java.

Ans: The reduce function sums up the occurrences of each each word from map.

Q13: How many mappers and reducers are launched for executing the above wordcount program?

Ans: Mappers 2 | Reducers 1

Q14: How much time do mappers and reducers spend for the above tasks, separately?

Ans: Mappers: 14752 ms | Reducers: 4192 ms