

**GROUP-9**  
**Assignment 3**  
**Mohd Imran, Sagar Patidar, Rhythm Gupta, Rahul Patidar**

1. **File Size:** 4.1 MB (Attached)
2. **Steps to configure Mapreduce:** In file etc/hadoop/mapred-site.xml

parameter name	value
mapreduce.framework.name	yarn

3. Word Count Program (**PFA - WordCount.java**): Counts the frequency of words.
4. Output on shutting down one node(Complete Output - **failure output.txt**): We shut down this slave - 172.16.5.44 during the execution of program.

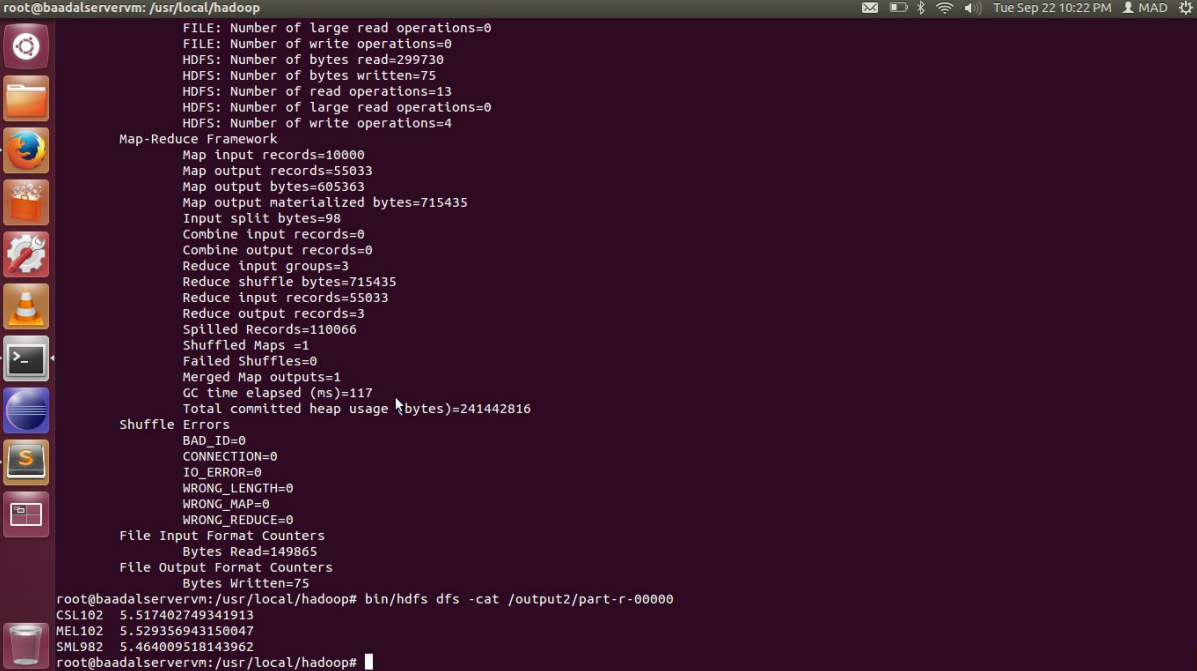
```
java.io.IOException: Got error, status message , ack with firstBadLink as 172.16.5.44:50010 at
org.apache.hadoop.hdfs.protocol.datatransfer.DataTransferProtoUtil.checkBlockOpStatus(DataT
ransferProtoUtil.java:140) at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.createBlockOutputStream(DFSOutputStream
.java:1334) at
org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.nextBlockOutputStream(DFSOutputStream.j
ava:1237
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:449)
15/09/21 23:29:52 INFO hdfs.DFSClient: Abandoning
BP-772401265-127.0.1.1-1442850580167:blk_1073741839_1015
15/09/21 23:29:52 INFO hdfs.DFSClient: Excluding datanode
DatanodeInfoWithStorage[172.16.5.44:50010,DS-f55f915a-3e52-4f9f-8c67-30e8b5d50f1f,DISK]
```

#### **Observation:**

NameNode periodically receives a Heartbeat and a Blockreport from each of the DataNodes in the cluster. Receipt of a Heartbeat implies that the DataNode is functioning properly. A Blockreport contains a list of all blocks on a DataNode. When NameNode notices that it has not recieved a hearbeat message from a data node after a certain amount of time, the data node is marked as dead. Since blocks will be under replicated the system begins replicating the blocks that were stored on the dead datanode. The NameNode Orchestrates the replication of data blocks from one datanode to another. The replication data transfer happens directly between data nodes and the data never passes through the namenode.

**GROUP-9**  
**Assignment 3**  
**Mohd Imran, Sagar Patidar, Rhythm Gupta, Rahul Patidar**

5. Average Count Program (**PFA - WordCount2.java**): Counts the average of grades. Program file along with the input file attached. The following screenshot the average of grades.



```
root@baadalservervm: /usr/local/hadoop
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=299730
HDFS: Number of bytes written=75
HDFS: Number of read operations=13
HDFS: Number of large read operations=0
HDFS: Number of write operations=4

Map-Reduce Framework
  Map input records=10000
  Map output records=55033
  Map output bytes=605363
  Map output materialized bytes=715435
  Input split bytes=98
  Combine input records=0
  Combine output records=0
  Reduce input groups=3
  Reduce shuffle bytes=715435
  Reduce input records=55033
  Reduce output records=3
  Spilled Records=110066
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=117
  Total committed heap usage (bytes)=241442816

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=149865
File Output Format Counters
  Bytes Written=75

root@baadalservervm: /usr/local/hadoop# bin/hdfs dfs -cat /output2/part-r-00000
CSL102  5.517402749341913
MEL102  5.529356943150047
SML982  5.464009518143962
root@baadalservervm: /usr/local/hadoop#
```