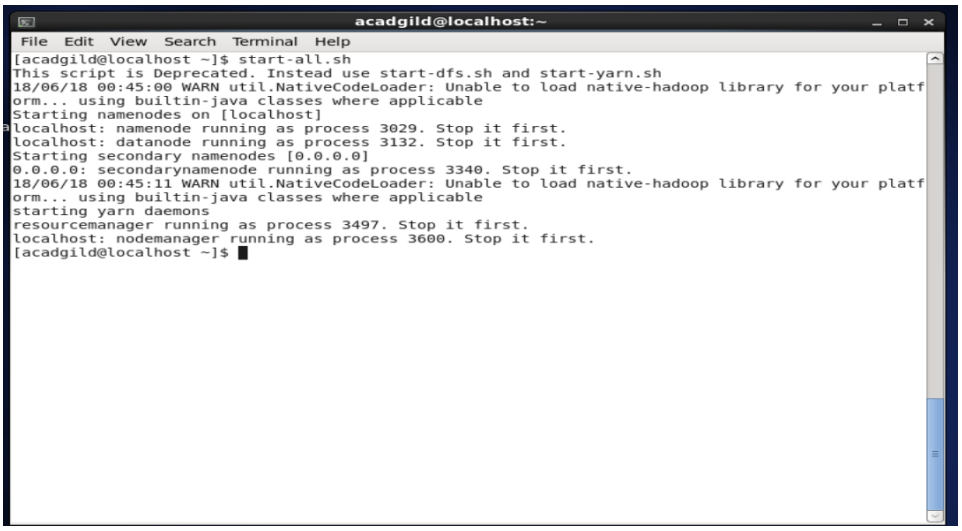
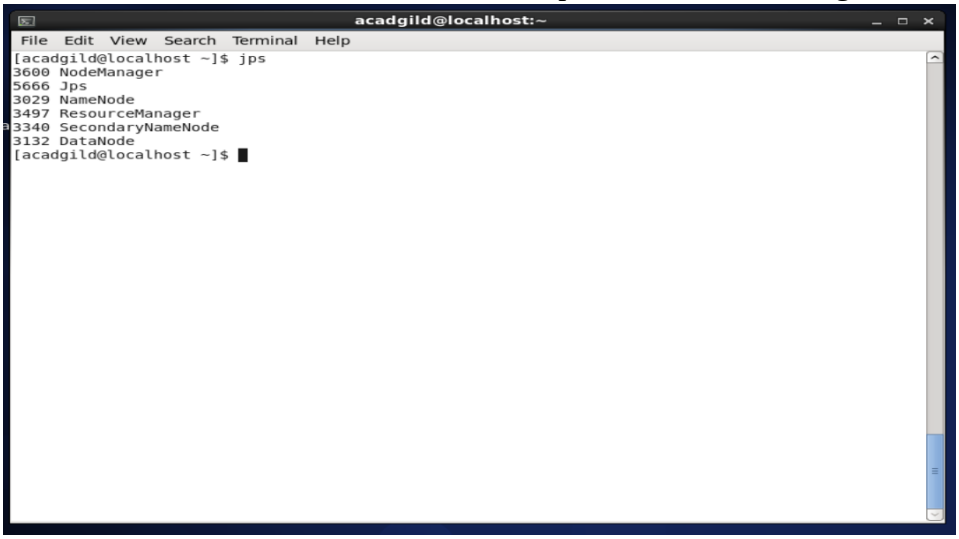
Big Data Assignment – Session 1

Task 1:

1. Start Hadoop single node on AcadGild VM (The command is start-all.sh)

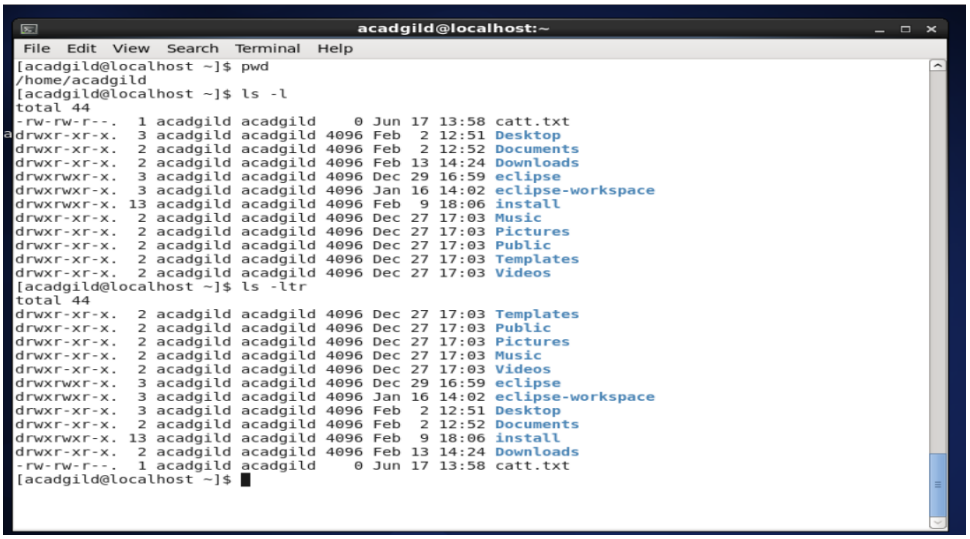


2. Run a JPS command to see if all the Hadoop daemons are running.

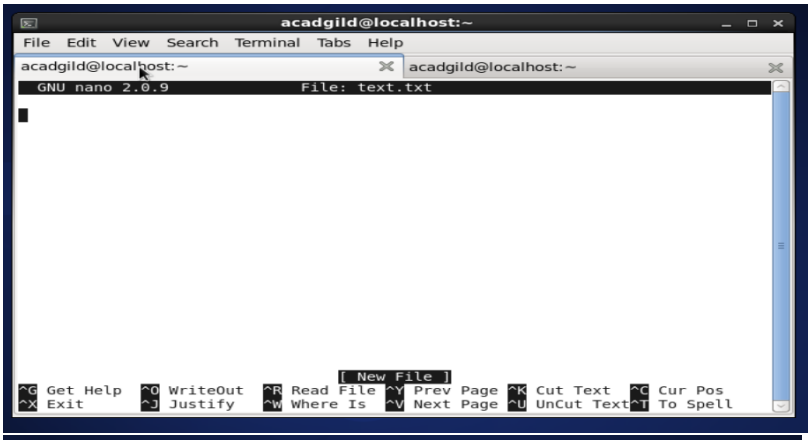


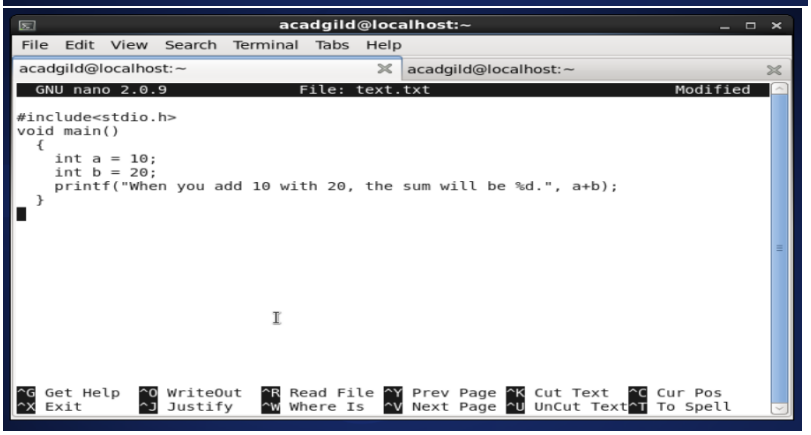
Big Data Assignment – Session 1

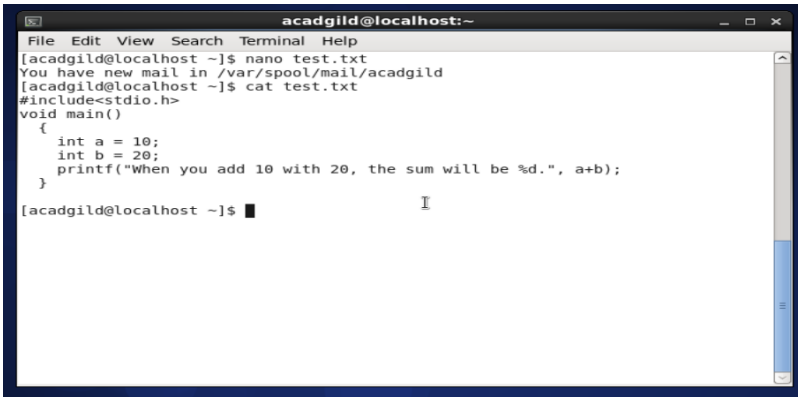
3. Run a couple of Unix commands like pwd, ls -ls etc.



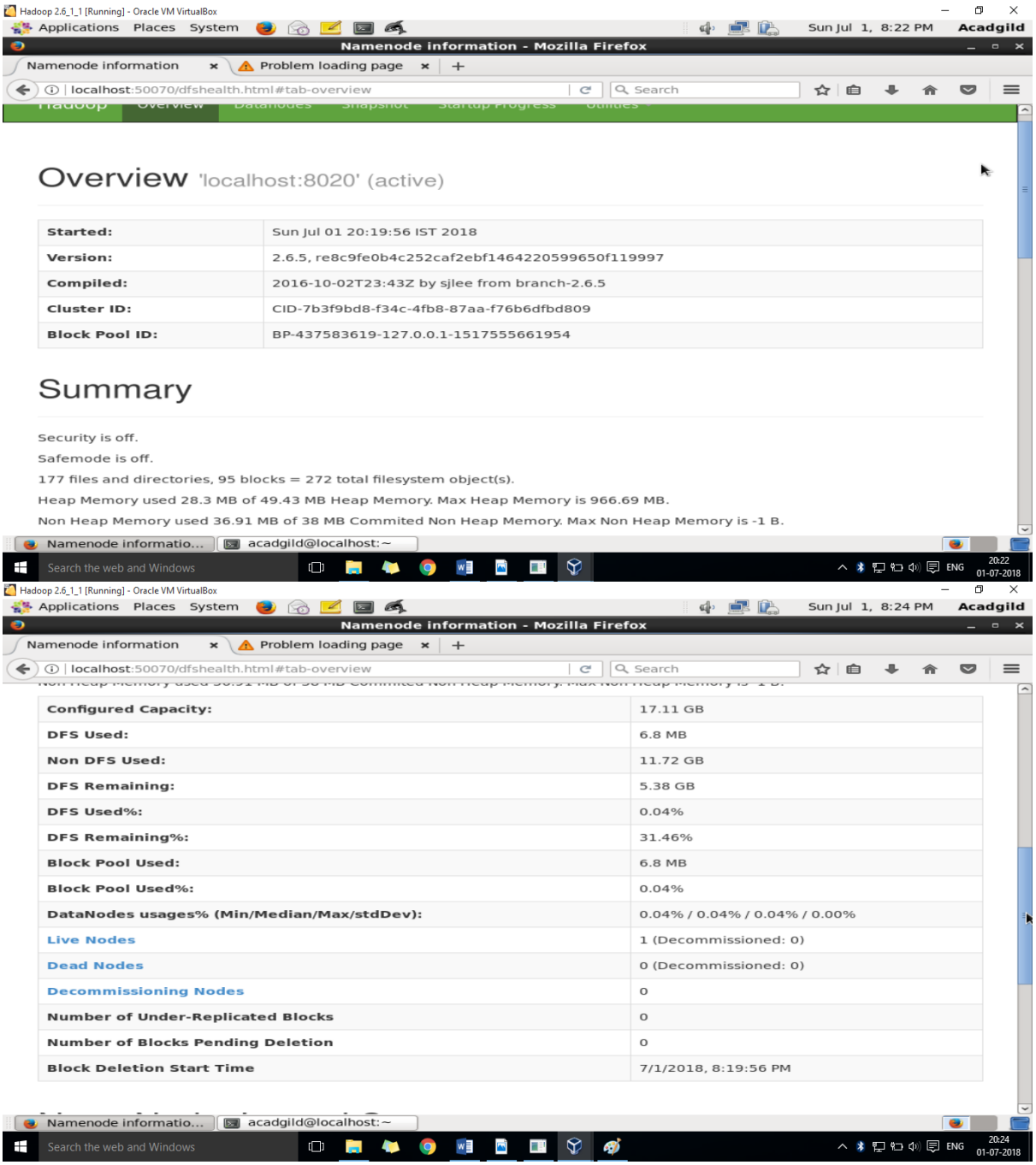
4. Create a file from the terminal using nano editor (for example nano test.txt) and put some content in it. Cat it to see if the content is saved.



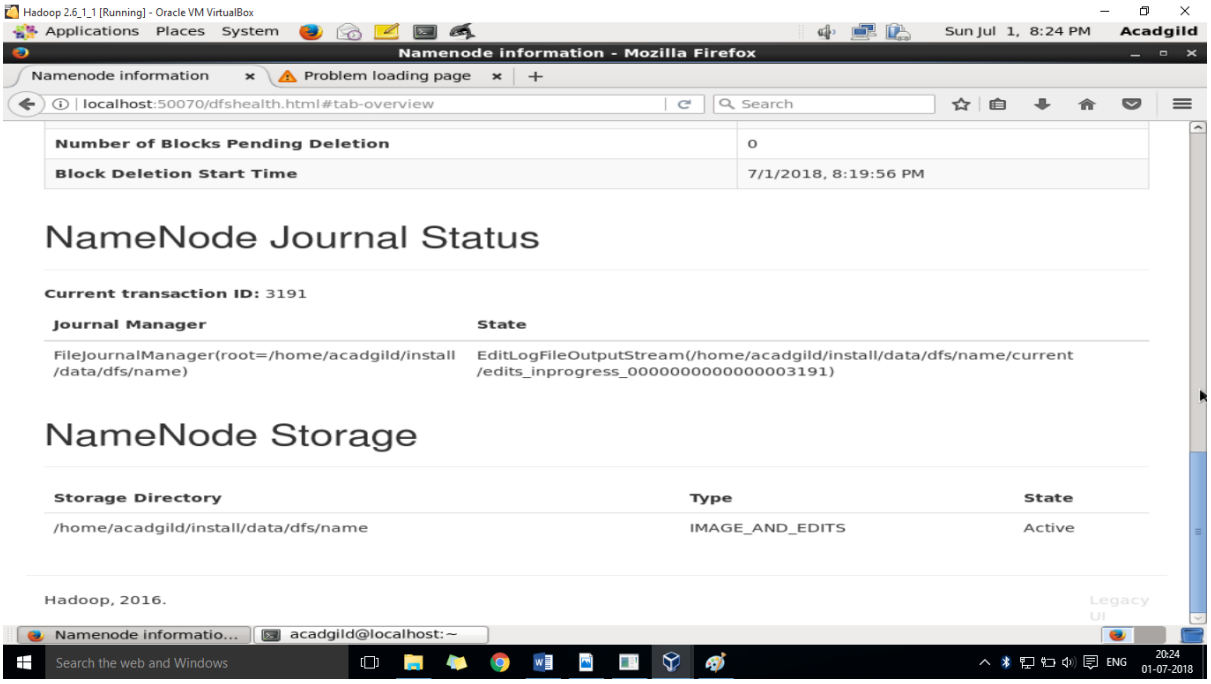




5. Open the hdfs web page from the browser by typing localhost:50070 in the browser. Check all the details of the HDFS.



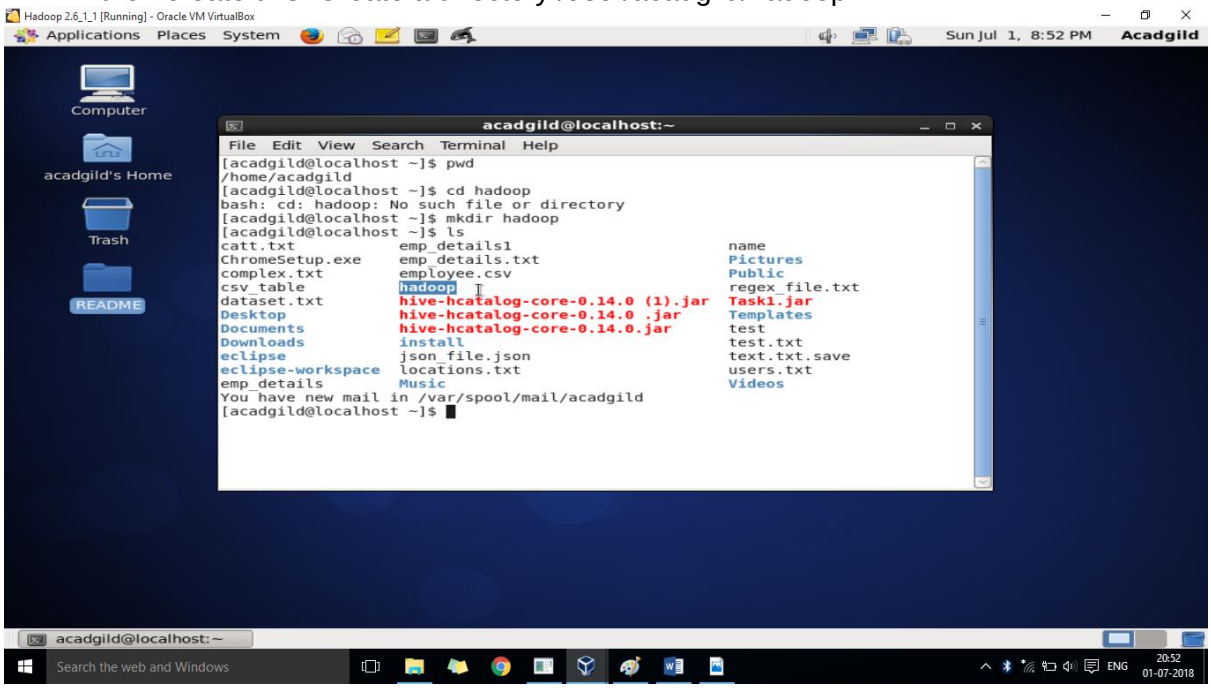
Big Data Assignment – Session 1



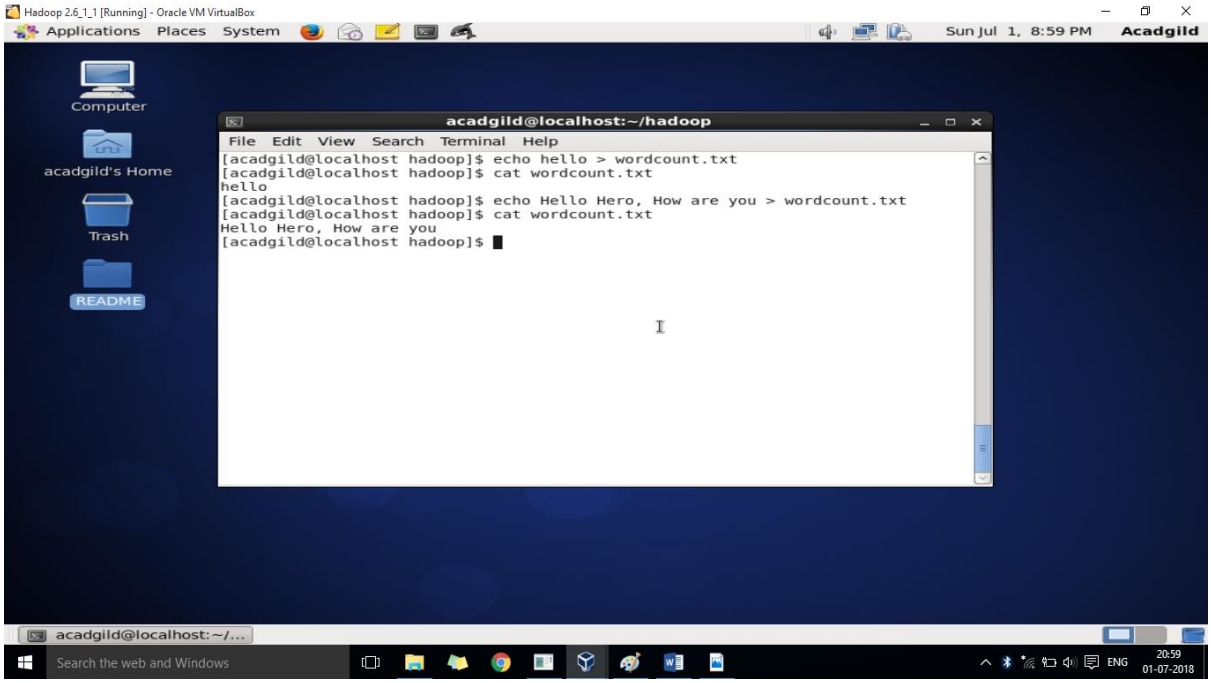
Big Data Assignment – Session 1

Task 2:

1. Check whether /user/acadgild directory exists or not in the HDFS. If it doesn't exist, then create this. Create a directory /user/acadgild/hadoop.



2. Create a file in HDFS under directory /user/acadgild/hadoop, with name word-count.txt. Whatever we type on screen should get appended to the file. Try to type (on screen) few lines from any online article or textbook.



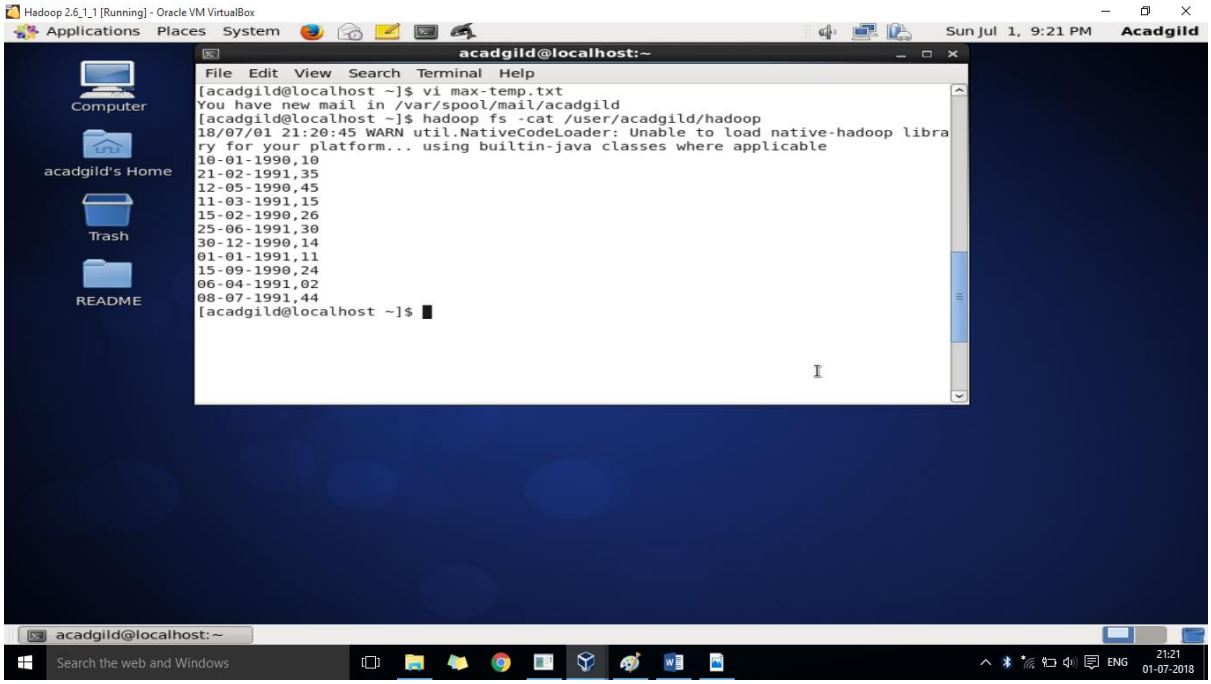
Big Data Assignment – Session 1

3. Create a file max-temp.txt in local FS. Put some 10-15 records of date and temperature example: dd-mm-yyyy,temperature

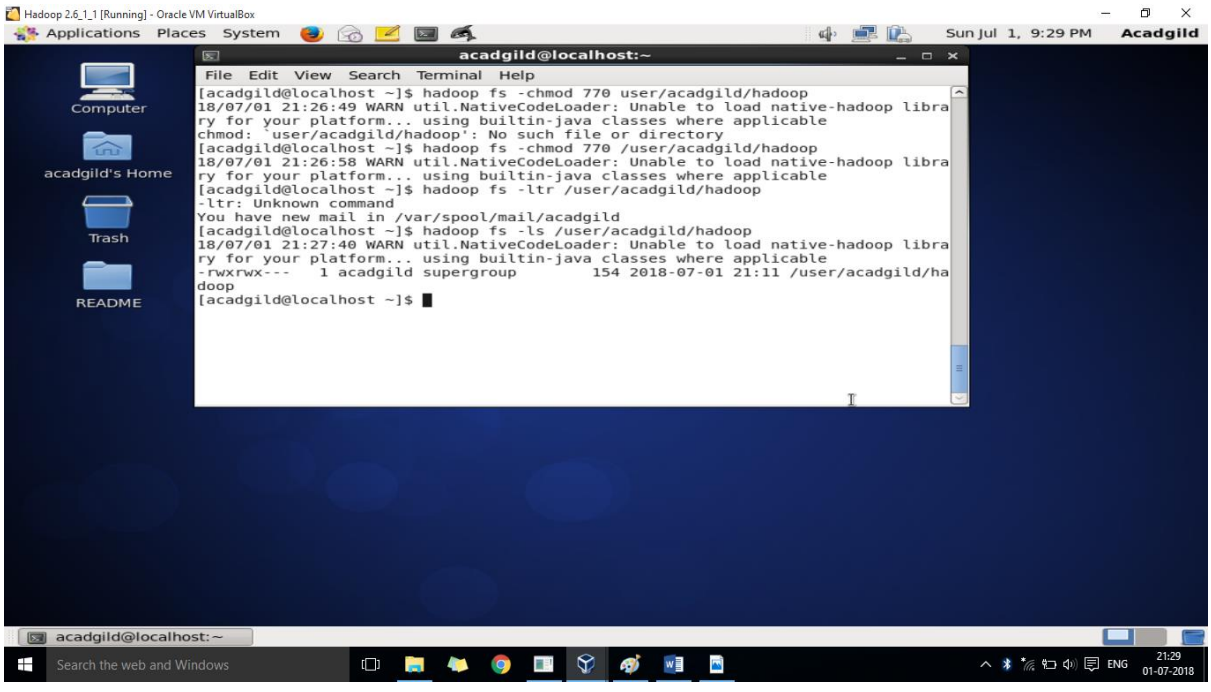
Example:

10-01-1990,10
10-02-1991,20

Move this file to HDFS at /user/acadgild/hadoop.



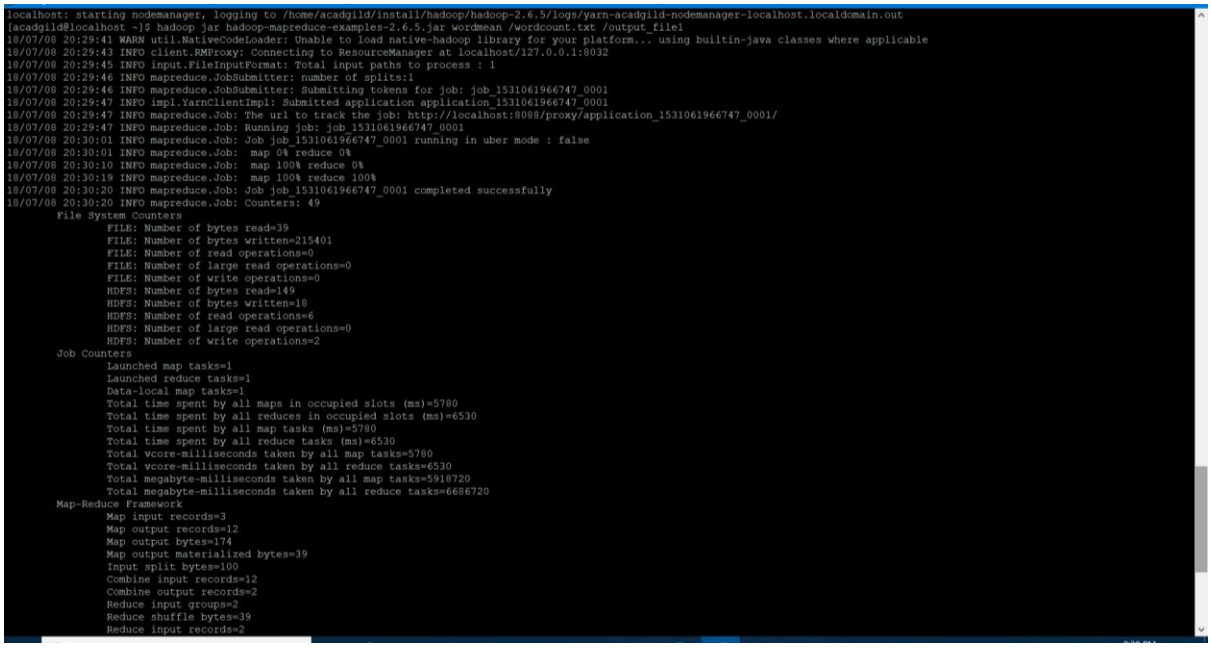
4. Change the permission of the file /user/acadgild/hadoop/max-temp.txt, such that only the owner and the group members have full control over the file. Others do not have any control over it.

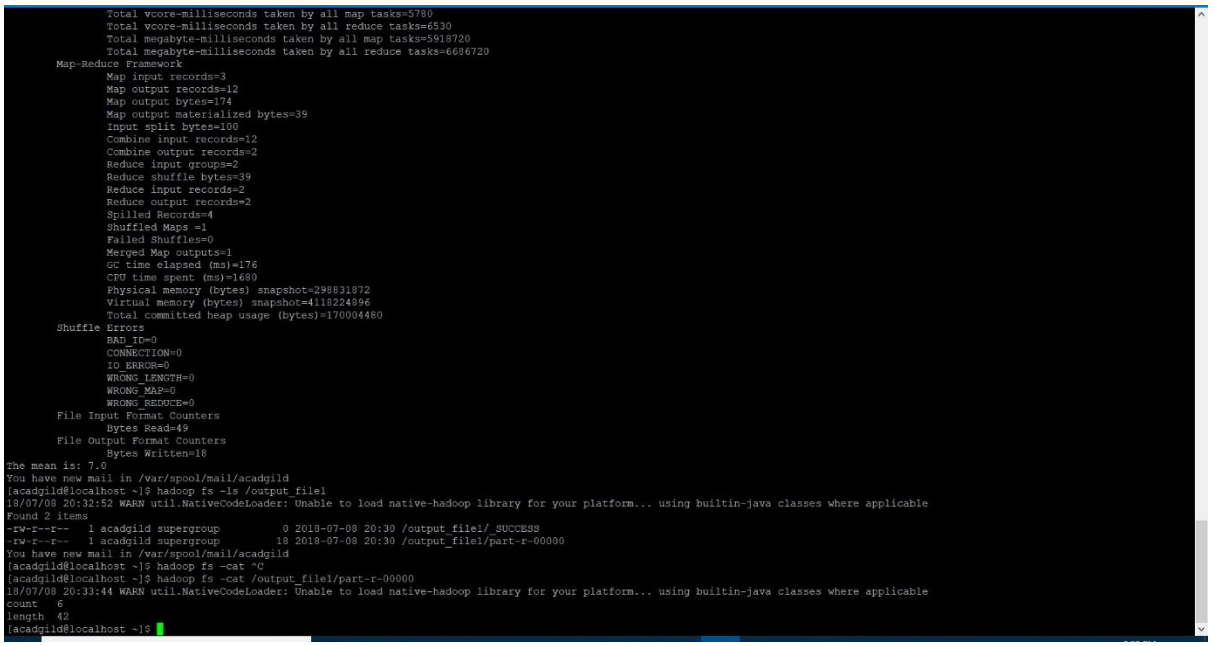


Big Data Assignment – Session 1

Task 3:

1. Execute **WordMedian** , **WordMean** , **WordStandardDeviation** programs using hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM.

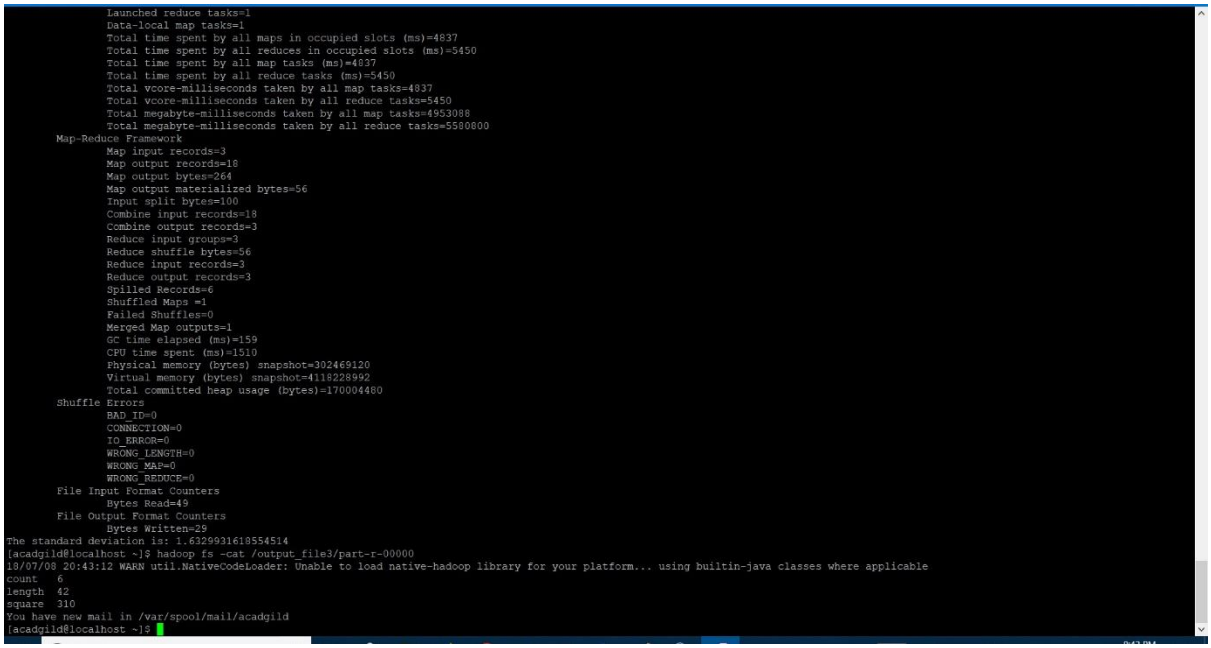




Big Data Assignment – Session 1

```
length  42
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /wordcount.txt /output_file2
18/07/08 20:37:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/08 20:37:31 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/08 20:37:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy t
his.
18/07/08 20:37:32 INFO input.FileInputFormat: Total input paths to process : 1
18/07/08 20:37:33 INFO mapreduce.JobSubmitter: number of splits:1
18/07/08 20:37:33 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531061966747_0002
18/07/08 20:37:33 INFO impl.YarnClientImpl: Submitted application application_1531061966747_0002
18/07/08 20:37:33 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531061966747_0002/
18/07/08 20:37:33 INFO mapreduce.Job: Running job: job_1531061966747_0002
18/07/08 20:37:44 INFO mapreduce.Job: Job job_1531061966747_0002 running in uber mode : false
18/07/08 20:37:44 INFO mapreduce.Job:  map 0% reduce 0%
18/07/08 20:37:51 INFO mapreduce.Job:  map 100% reduce 0%
18/07/08 20:37:59 INFO mapreduce.Job:  map 100% reduce 100%
18/07/08 20:38:00 INFO mapreduce.Job: Job job_1531061966747_0002 completed successfully
18/07/08 20:38:00 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=215171
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=149
                HDFS: Number of bytes written=20
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=5133
                Total time spent by all reduces in occupied slots (ms)=5438
                Total time spent by all map tasks (ms)=5133
                Total time spent by all reduce tasks (ms)=5438
                Total vcore-milliseconds taken by all map tasks=5133
                Total vcore-milliseconds taken by all reduce tasks=5438
                Total megabyte-milliseconds taken by all map tasks=5256192
                Total megabyte-milliseconds taken by all reduce tasks=5568512
        Map-Reduce Framework
                Map input records=3
                Map output records=6
                Map output bytes=48
                Map output materialized bytes=56
                Input split bytes=100
                Combine input records=6
                Combine output records=5
                Reduce input groups=5
```

```
                Total time spent by all maps in occupied slots (ms)=5133
                Total time spent by all reduces in occupied slots (ms)=5438
                Total time spent by all map tasks (ms)=5133
                Total time spent by all reduce tasks (ms)=5438
                Total vcore-milliseconds taken by all map tasks=5133
                Total vcore-milliseconds taken by all reduce tasks=5438
                Total megabyte-milliseconds taken by all map tasks=5256192
                Total megabyte-milliseconds taken by all reduce tasks=5568512
        Map-Reduce Framework
                Map input records=3
                Map output records=6
                Map output bytes=48
                Map output materialized bytes=56
                Input split bytes=100
                Combine input records=6
                Combine output records=5
                Reduce input groups=5
                Reduce shuffle bytes=56
                Reduce input records=5
                Reduce output records=5
                Spilled Records=10
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=148
                CPU time spent (ms)=1590
                Physical memory (bytes) snapshot=301129728
                Virtual memory (bytes) snapshot=4118220800
                Total committed heap usage (bytes)=170004480
        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=49
        File Output Format Counters
                Bytes Written=20
The median is: 7
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /output_file2/part-r-00000
18/07/08 20:38:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
4       1
6       1
7       1
8       2
9       1
[acadgild@localhost ~]$
```

```
        at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /wordcount.txt /output_file3
18/07/08 20:41:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/08 20:42:00 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/08 20:42:02 INFO input.FileInputFormat: Total input paths to process : 1
18/07/08 20:42:02 INFO mapreduce.JobSubmitter: number of splits:1
18/07/08 20:42:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531061966747_0003
18/07/08 20:42:03 INFO impl.YarnClientImpl: Submitted application application_1531061966747_0003
18/07/08 20:42:03 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531061966747_0003/
18/07/08 20:42:03 INFO mapreduce.Job: Running job: job_1531061966747_0003
18/07/08 20:42:13 INFO mapreduce.Job: Job job_1531061966747_0003 running in uber mode : false
18/07/08 20:42:13 INFO mapreduce.Job:  map 0% reduce 0%
18/07/08 20:42:20 INFO mapreduce.Job:  map 100% reduce 0%
18/07/08 20:42:28 INFO mapreduce.Job:  map 100% reduce 100%
18/07/08 20:42:29 INFO mapreduce.Job: Job job_1531061966747_0003 completed successfully
18/07/08 20:42:30 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=56
                FILE: Number of bytes written=215595
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=149
                HDFS: Number of bytes written=29
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=4837
                Total time spent by all reduces in occupied slots (ms)=5450
                Total time spent by all map tasks (ms)=4837
                Total time spent by all reduce tasks (ms)=5450
                Total vcore-milliseconds taken by all map tasks=4837
                Total vcore-milliseconds taken by all reduce tasks=5450
                Total megabyte-milliseconds taken by all map tasks=4953088
                Total megabyte-milliseconds taken by all reduce tasks=5580800
        Map-Reduce Framework
                Map input records=3
                Map output records=18
                Map output bytes=264
                Map output materialized bytes=56
                Input split bytes=100
                Combine input records=18
                Combine output records=3
                Reduce input groups=3
```

Big Data Assignment – Session 1