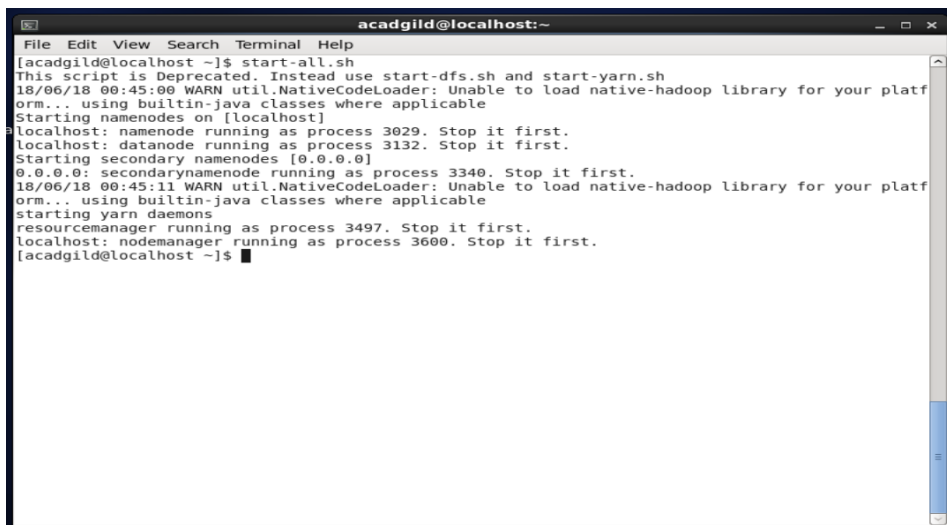


# Big Data Assignment – Session 1

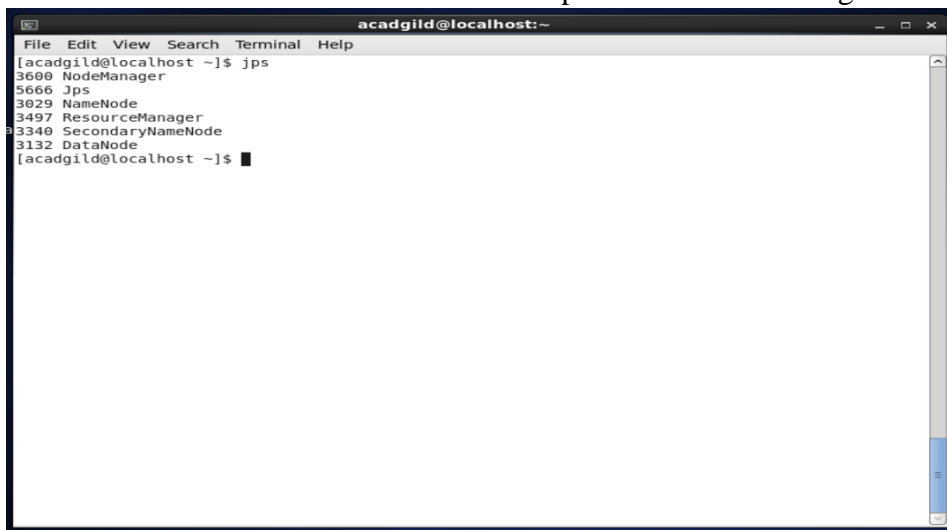
## Task 1:

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



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ start-all.sh  
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh  
18/06/18 00:45:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf  
orm... using builtin-java classes where applicable  
Starting namenodes on [localhost]  
localhost: namenode running as process 3029. Stop it first.  
localhost: datanode running as process 3132. Stop it first.  
Starting secondary namenodes [0.0.0.0]  
0.0.0.0: secondarynamenode running as process 3340. Stop it first.  
18/06/18 00:45:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platf  
orm... using builtin-java classes where applicable  
Starting yarn daemons  
resourceemanager running as process 3497. Stop it first.  
localhost: nodemanager running as process 3600. Stop it first.  
[acadgild@localhost ~]$
```

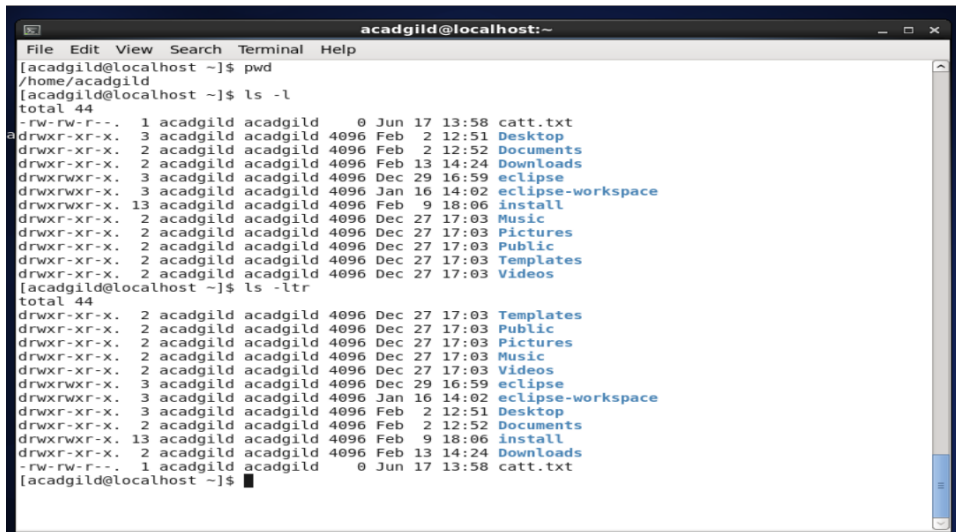
2. Run a JPS command to see if all the Hadoop daemons are running.



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ jps  
3600 NodeManager  
5666 Jps  
3029 NameNode  
3497 ResourceManager  
3340 SecondaryNameNode  
3132 DataNode  
[acadgild@localhost ~]$
```

# Big Data Assignment – Session 1

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



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ pwd  
/home/acadgild  
[acadgild@localhost ~]$ ls -l  
total 44  
-rw-rw-r--. 1 acadgild acadgild  0 Jun 17 13:58 catt.txt  
drwxr-xr-x. 3 acadgild acadgild 4096 Feb  2 12:51 Desktop  
drwxr-xr-x. 2 acadgild acadgild 4096 Feb  2 12:52 Documents  
drwxr-xr-x. 2 acadgild acadgild 4096 Feb 13 14:24 Downloads  
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29 16:59 eclipse  
drwxrwxr-x. 3 acadgild acadgild 4096 Jan 16 14:02 eclipse-workspace  
drwxrwxr-x. 13 acadgild acadgild 4096 Feb  9 18:06 install  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Music  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Pictures  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Public  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Templates  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Videos  
[acadgild@localhost ~]$ ls -ltr  
total 44  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Templates  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Public  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Pictures  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Music  
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Videos  
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29 16:59 eclipse  
drwxrwxr-x. 3 acadgild acadgild 4096 Jan 16 14:02 eclipse-workspace  
drwxr-xr-x. 3 acadgild acadgild 4096 Feb  2 12:51 Desktop  
drwxr-xr-x. 2 acadgild acadgild 4096 Feb  2 12:52 Documents  
drwxrwxr-x. 13 acadgild acadgild 4096 Feb  9 18:06 install  
drwxr-xr-x. 2 acadgild acadgild 4096 Feb 13 14:24 Downloads  
-rw-rw-r--. 1 acadgild acadgild  0 Jun 17 13:58 catt.txt  
[acadgild@localhost ~]$
```

# Big Data Assignment – Session 1

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.



The image consists of three vertically stacked screenshots of a terminal window. The terminal title bar shows 'acadgild@localhost:~'. The first screenshot shows the nano editor in its initial state with an empty file named 'text.txt'. The second screenshot shows the same editor with a C program saved in 'text.txt'. The third screenshot shows the terminal after running 'nano test.txt' and 'cat test.txt', displaying the saved program content.

```
acadgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadgild@localhost:~  
GNU nano 2.0.9 File: text.txt  
  
[ New File ]  
^G Get Help ^O WriteOut ^R Read File ^Y Prev Page ^K Cut Text ^C Cur Pos  
^X Exit ^J Justify ^W Where Is ^V Next Page ^U UnCut Text ^T To Spell
```

```
acadgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadgild@localhost:~  
GNU nano 2.0.9 File: text.txt Modified  
  
#include<stdio.h>  
void main()  
{  
    int a = 10;  
    int b = 20;  
    printf("When you add 10 with 20, the sum will be %d.", a+b);  
}
```

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ nano test.txt  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ cat test.txt  
#include<stdio.h>  
void main()  
{  
    int a = 10;  
    int b = 20;  
    printf("When you add 10 with 20, the sum will be %d.", a+b);  
}  
[acadgild@localhost ~]$
```

# Big Data Assignment – Session 1

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

The screenshot displays the Hadoop NameNode web interface in a Mozilla Firefox browser. The address bar shows 'localhost:50070/dfshealth.html#tab-overview'. The page title is 'Namenode information - Mozilla Firefox'. The main content area is titled 'Overview 'localhost:8020' (active)'. Below this, there is a table with details about the NameNode's state and configuration. The 'Summary' section provides a high-level overview of the HDFS state, including security status, file and directory counts, and memory usage. The 'Configured Capacity' section shows the total capacity and usage of the NameNode, including DFS used, non-DFS used, and block pool usage. The 'Live Nodes' section shows the number of live nodes and decommissioning nodes.

Started:	Sun Jul 01 20:19:56 IST 2018
Version:	2.6.5, re8c9fe0b4c252caf2ebf1464220599650f119997
Compiled:	2016-10-02T23:43Z by sjlee from branch-2.6.5
Cluster ID:	CID-7b3f9bd8-f34c-4fb8-87aa-f76b6dfbd809
Block Pool ID:	BP-437583619-127.0.0.1-1517555661954

**Summary**

Security is off.  
Safemode is off.  
177 files and directories, 95 blocks = 272 total filesystem object(s).  
Heap Memory used 28.3 MB of 49.43 MB Heap Memory. Max Heap Memory is 966.69 MB.  
Non Heap Memory used 36.91 MB of 38 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	17.11 GB
DFS Used:	6.8 MB
Non DFS Used:	11.72 GB
DFS Remaining:	5.38 GB
DFS Used%:	0.04%
DFS Remaining%:	31.46%
Block Pool Used:	6.8 MB
Block Pool Used%:	0.04%
DataNodes usages% (Min/Median/Max/stdDev):	0.04% / 0.04% / 0.04% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	7/1/2018, 8:19:56 PM

# Big Data Assignment – Session 1

Hadoop 2.6.1\_1 [Running] - Oracle VM VirtualBox

Applications Places System

Sun Jul 1, 8:24 PM Acadgild

## Namenode information - Mozilla Firefox

Namenode information x Problem loading page x +

localhost:50070/dfshealth.html#tab-overview

Number of Blocks Pending Deletion	0
Block Deletion Start Time	7/1/2018, 8:19:56 PM

## NameNode Journal Status

Current transaction ID: 3191

Journal Manager	State
FileJournalManager(root=/home/acadgild/install/data/dfs/name)	EditLogFileOutputStream(/home/acadgild/install/data/dfs/name/current/edits_inprogress_00000000000000003191)

## NameNode Storage

Storage Directory	Type	State
/home/acadgild/install/data/dfs/name	IMAGE_AND_EDITS	Active

Hadoop, 2016. Legacy UI

Namenode informatio... acadgild@localhost: ~

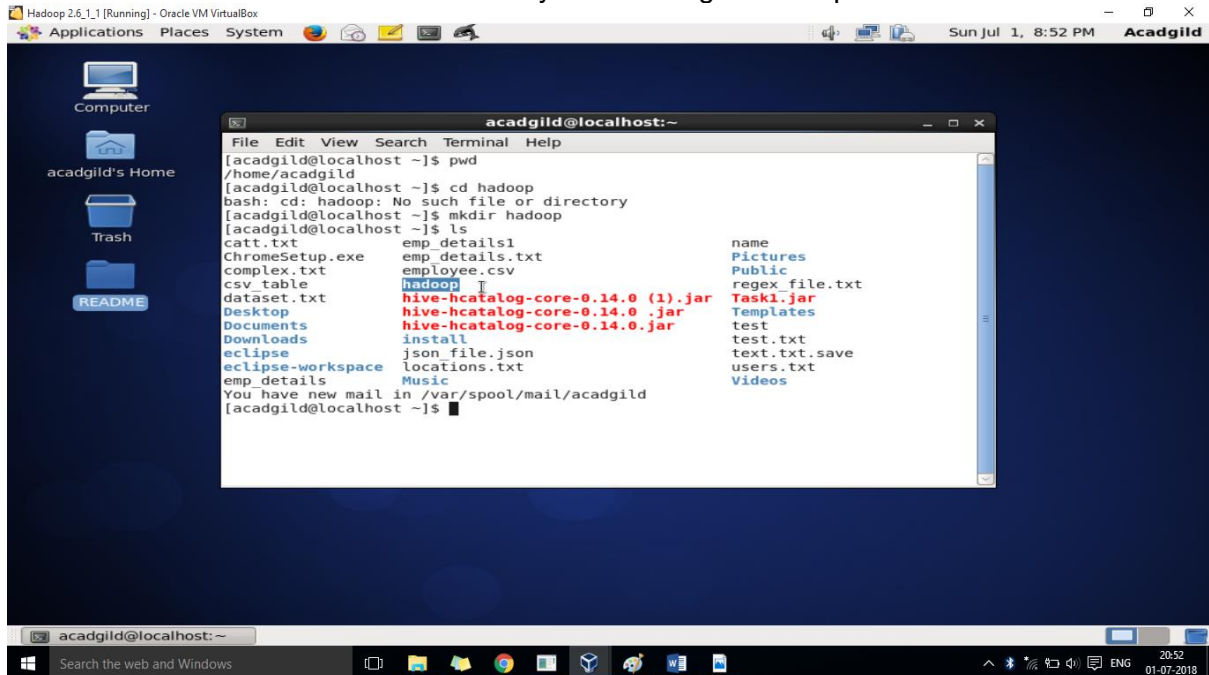
Search the web and Windows

20:24 01-07-2018

# Big Data Assignment – Session 1

## Task 2:

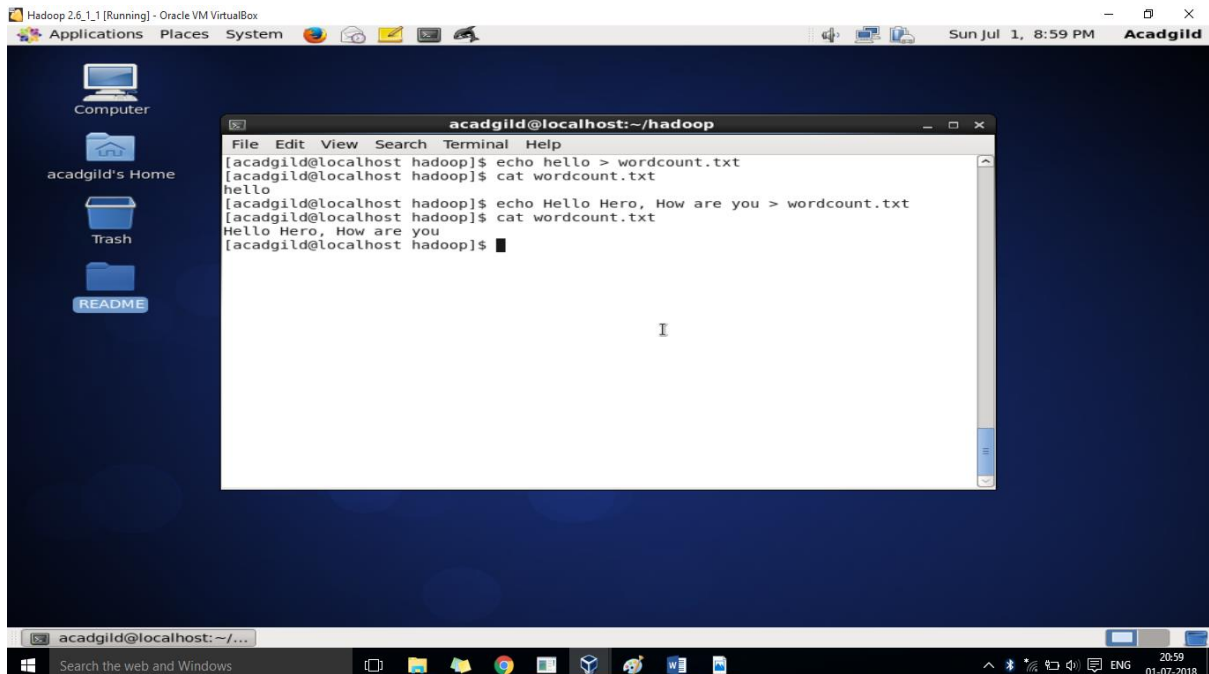
1. Check whether /user/acadgild directory exists or not in the HDFS. If it doesn't exist, then create this. Create a directory /user/acadgild/hadoop.



The screenshot shows a terminal window titled 'acadgild@localhost:~' with the following commands and output:

```
[acadgild@localhost ~]$ pwd
/home/acadgild
[acadgild@localhost ~]$ cd hadoop
bash: cd: hadoop: No such file or directory
[acadgild@localhost ~]$ mkdir hadoop
[acadgild@localhost ~]$ ls
catt.txt      emp_details1      name
ChromeSetup.exe  emp_details.txt  Pictures
complex.txt     employee.csv      Public
csv_table       hadoop            regex file.txt
dataset.txt     hive-hcatalog-core-0.14.0 (1).jar Task1.jar
Desktop        hive-hcatalog-core-0.14.0 .jar Templates
Documents      hive-hcatalog-core-0.14.0 .jar test
Downloads      install           test.txt
eclipse        json_file.json   text.txt.save
eclipse-workspace locations.txt      users.txt
emp_details    Music             Videos
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

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.



The screenshot shows a terminal window titled 'acadgild@localhost:~/hadoop' with the following commands and output:

```
[acadgild@localhost hadoop]$ echo hello > wordcount.txt
[acadgild@localhost hadoop]$ cat wordcount.txt
hello
[acadgild@localhost hadoop]$ echo Hello Hero, How are you > wordcount.txt
[acadgild@localhost hadoop]$ cat wordcount.txt
Hello Hero, How are you
[acadgild@localhost hadoop]$
```

# Big Data Assignment – Session 1

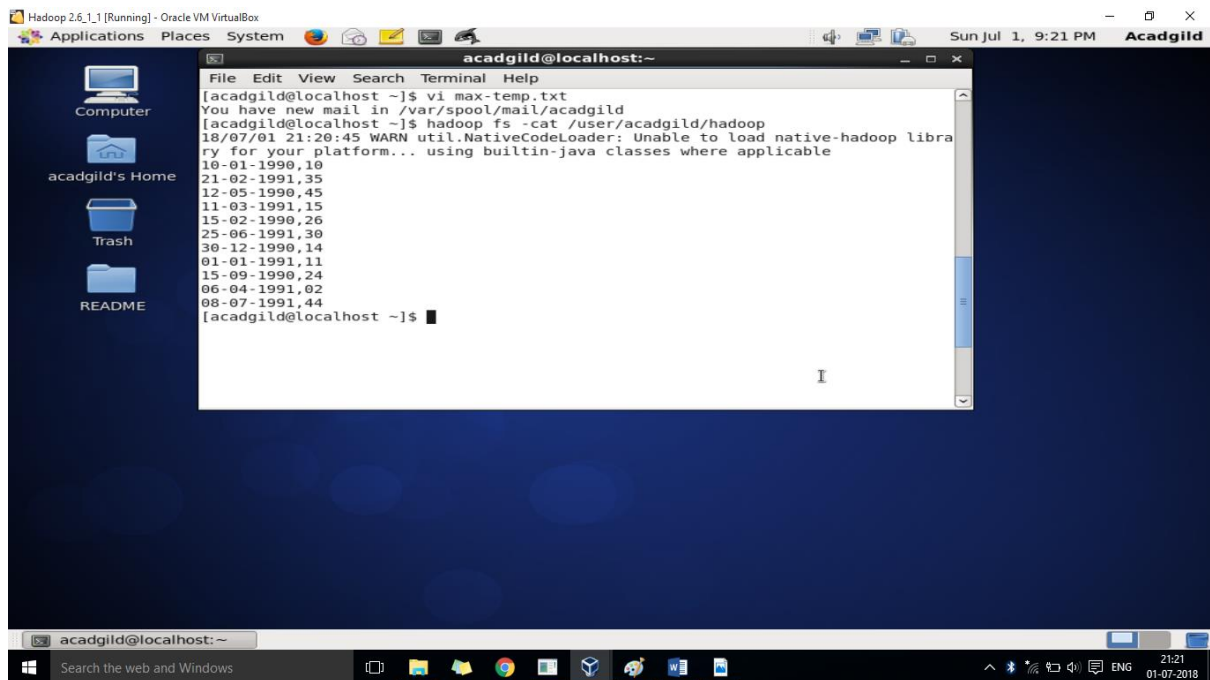
3. Create a file max-temp.txt in local FS. Put some 10-15 records of date and temperature example: dd-mm-yyyy,temperature

Example:

10-01-1990,10

10-02-1991,20

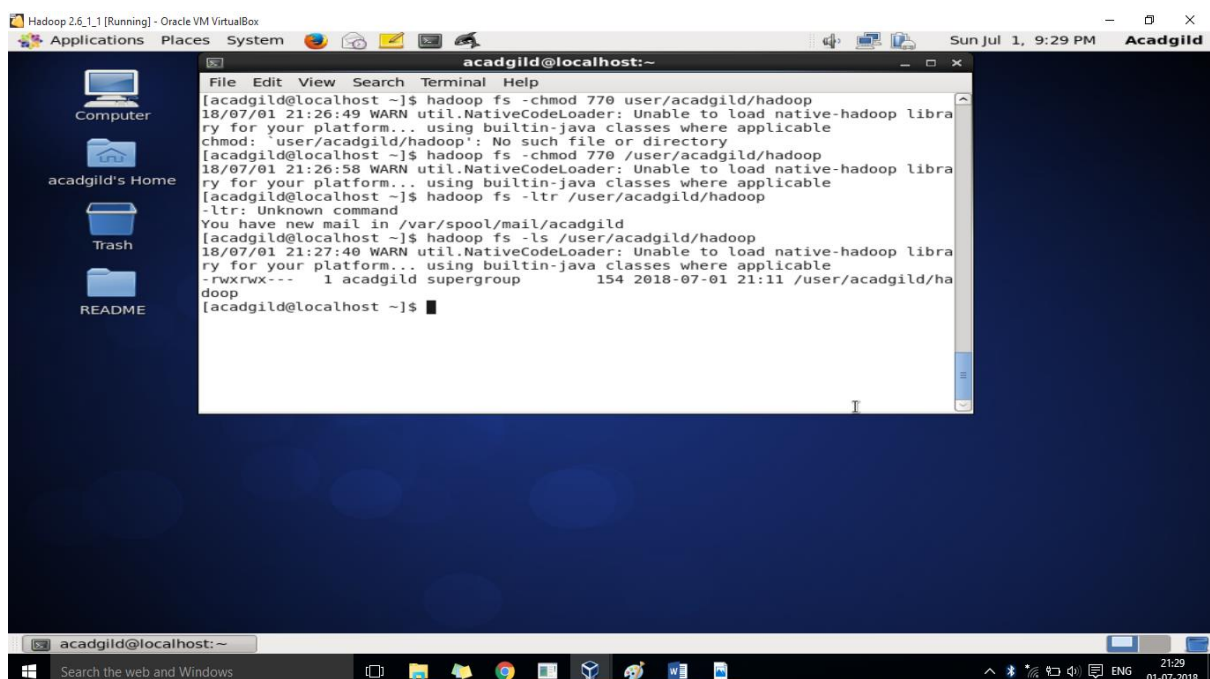
Move this file to HDFS at /user/acadgild/hadoop.



The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output is as follows:

```
[acadgild@localhost ~]$ vi max-temp.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop
18/07/01 21:20:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
10-01-1990,10
21-02-1991,35
12-05-1990,45
11-03-1991,15
15-02-1990,26
25-06-1991,30
30-12-1990,14
01-01-1991,11
15-09-1990,24
06-04-1991,02
08-07-1991,44
[acadgild@localhost ~]$
```

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.



The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output is as follows:

```
[acadgild@localhost ~]$ hadoop fs -chmod 770 user/acadgild/hadoop
18/07/01 21:26:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hadoop fs -chmod 770 /user/acadgild/hadoop
18/07/01 21:26:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hadoop fs -ltr /user/acadgild/hadoop
-ltr: Unknown command
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop
18/07/01 21:27:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
-rwxrwx--- 1 acadgild supergroup 154 2018-07-01 21:11 /user/acadgild/hadoop
[acadgild@localhost ~]$
```

# 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.

```
localhost: Starting ResourceManager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-ResourceManager-localhost.localdomain.out
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /wordcount.txt /output_file1
18/07/08 20:29:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/08 20:29:43 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/08 20:29:45 INFO input.FileInputFormat: Total input paths to process : 1
18/07/08 20:29:46 INFO mapreduce.JobSubmitter: number of splits:1
18/07/08 20:29:46 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531061966747_0001
18/07/08 20:29:47 INFO impl.YarnClientImpl: Submitted application application_1531061966747_0001
18/07/08 20:29:47 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531061966747_0001/
18/07/08 20:29:47 INFO mapreduce.Job: Running job: job_1531061966747_0001
18/07/08 20:30:01 INFO mapreduce.Job: Job job_1531061966747_0001 running in uber mode : false
18/07/08 20:30:01 INFO mapreduce.Job: map 0% reduce 0%
18/07/08 20:30:10 INFO mapreduce.Job: map 100% reduce 0%
18/07/08 20:30:19 INFO mapreduce.Job: map 100% reduce 100%
18/07/08 20:30:20 INFO mapreduce.Job: Job job_1531061966747_0001 completed successfully
18/07/08 20:30:20 INFO mapreduce.Job: Counters: 49

File System Counters
  FILE: Number of bytes read=39
  FILE: Number of bytes written=215401
  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=18
  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)=5780
  Total time spent by all reduces in occupied slots (ms)=6530
  Total time spent by all map tasks (ms)=5780
  Total time spent by all reduce tasks (ms)=6530
  Total vcore-milliseconds taken by all map tasks=5780
  Total vcore-milliseconds taken by all reduce tasks=6530
  Total megabyte-milliseconds taken by all map tasks=5918720
  Total megabyte-milliseconds taken by all reduce tasks=6686720

Map-Reduce Framework
  Map input records=3
  Map output records=12
  Map output bytes=174
  Map output materialized bytes=39
  Input split bytes=100
  Combine input records=12
  Combine output records=2
  Reduce input groups=2
  Reduce shuffle bytes=39
  Reduce input records=2
```

```

  Total vcore-milliseconds taken by all map tasks=5780
  Total vcore-milliseconds taken by all reduce tasks=6530
  Total megabyte-milliseconds taken by all map tasks=5918720
  Total megabyte-milliseconds taken by all reduce tasks=6686720

Map-Reduce Framework
  Map input records=3
  Map output records=12
  Map output bytes=174
  Map output materialized bytes=39
  Input split bytes=100
  Combine input records=12
  Combine output records=2
  Reduce input groups=2
  Reduce shuffle bytes=39
  Reduce input records=2
  Reduce output records=2
  Spilled Records=4
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=176
  CPU time spent (ms)=1680
  Physical memory (bytes) snapshot=298831872
  Virtual memory (bytes) snapshot=4118224896
  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=18

The mean is: 7.0
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /output_file1
18/07/08 20:32:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-07-08 20:30 /output_file1/ SUCCESS
-rw-r--r-- 1 acadgild supergroup 18 2018-07-08 20:30 /output_file1/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat ^C
[acadgild@localhost ~]$ hadoop fs -cat /output_file1/part-r-00000
18/07/08 20:33:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count 6
length 42
[acadgild@localhost ~]$
```



# 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 this.
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 20:37:55 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-milliseonds taken by all map tasks=5133
  Total vcore-milliseonds taken by all reduce tasks=5438
  Total megabyte-milliseonds taken by all map tasks=5256192
  Total megabyte-milliseonds 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=418220800
  Total committed heap usage (bytes)=170094480
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 ~)$
```

```
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-milliseonds taken by all map tasks=5133
Total vcore-milliseonds taken by all reduce tasks=5438
Total megabyte-milliseonds taken by all map tasks=5256192
Total megabyte-milliseonds 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=418220800
  Total committed heap usage (bytes)=170094480
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-milliseonds taken by all map tasks=4837
  Total vcore-milliseonds taken by all reduce tasks=5450
  Total megabyte-milliseonds taken by all map tasks=4953088
  Total megabyte-milliseonds 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=19
  Combine output records=3
  Reduce input groups=3
```

# Big Data Assignment - Session 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 reduce 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=4933088
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
  Reduce shuffle bytes=56
  Reduce input records=3
  Reduce output records=3
  Spilled Records=6
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=459
  CPU time spent (ms)=1510
  Physical memory (bytes) snapshot=302469120
  Virtual memory (bytes) snapshot=418228992
  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=29

The standard deviation is: 1.6329931619554514
[acadgild@localhost ~]$ hadoop fs -cat /output_file3/part-r-000000
18/07/08 20:43:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count      6
length    42
square    310
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```