

Big Data Hadoop Training- Assignment 1

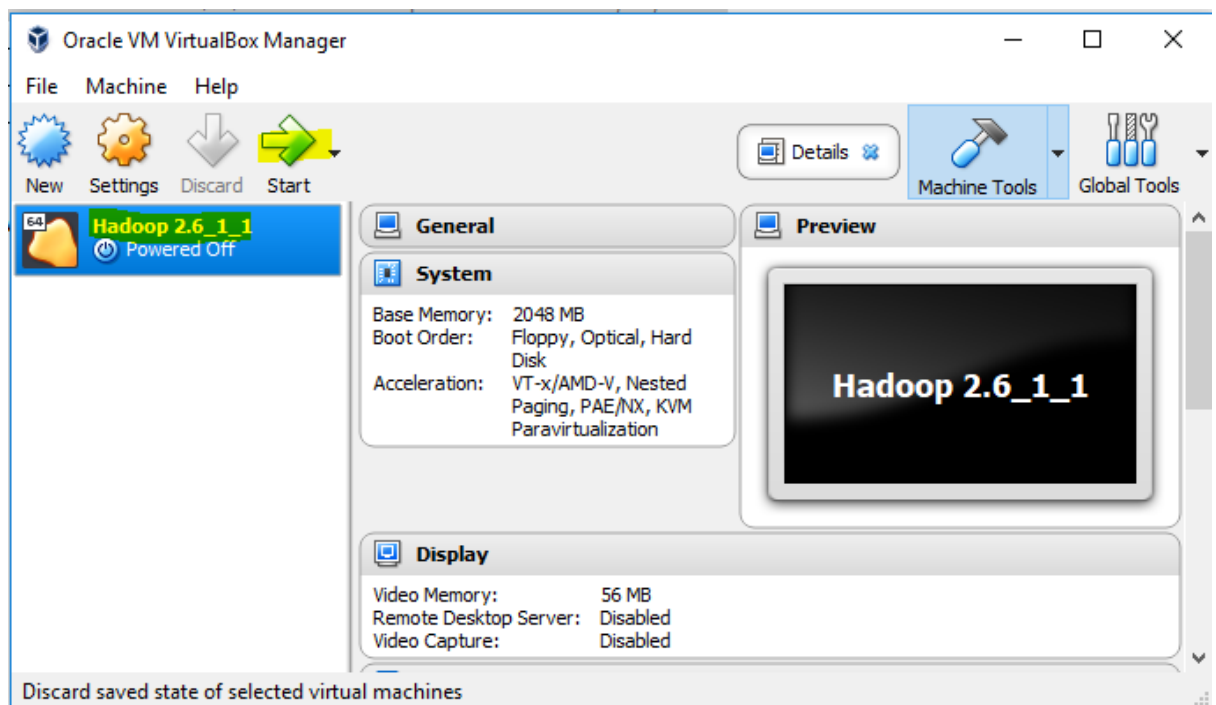
Task 1:

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

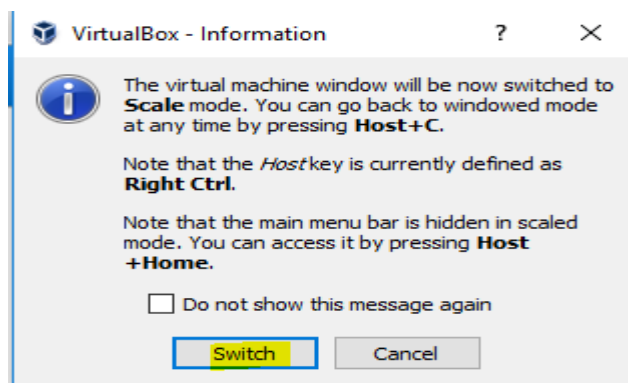
Step 1: Select Acadgild_64bit_2.6 from the folder where it has been placed. Priorly VirtualBox of Oracle has to be installed and the AcadgildVm should be opened.

VirtualBox-5.2.12-122591-Win	16/06/2018 13:09	Application	111,106 KB
Acadgild_64bit_2.6	16/06/2018 11:26	Open Virtualizatio...	9,567,606 KB

Step 2: Open Acadgild_64bit where can find Hadoop 2.6_1_1. Select it and click on -> Start button.

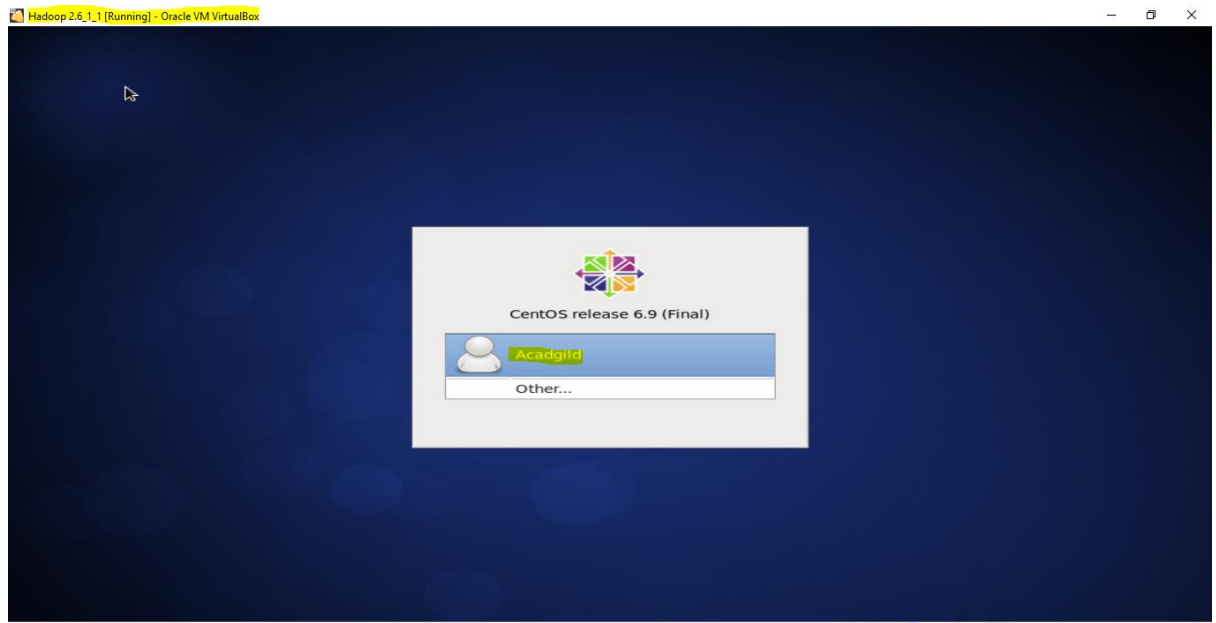


Step 3: Click on “Switch” icon in the pop-up

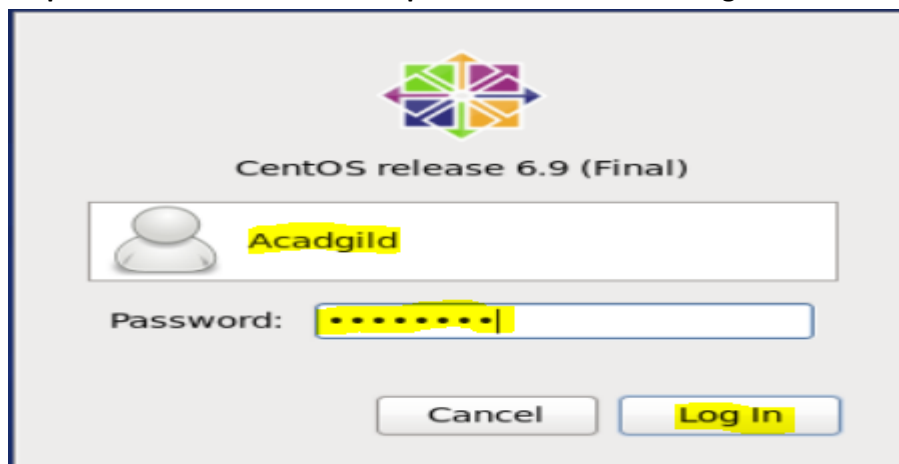


Big Data Hadoop Training- Assignment 1

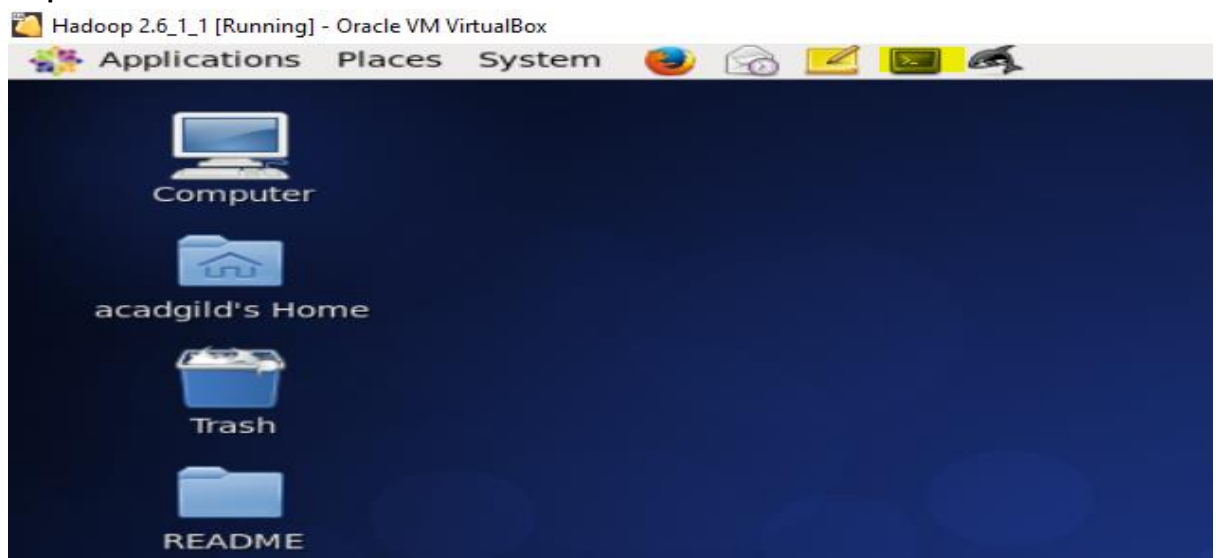
Step 4: Hadoop Oracle Virtual VM box will open.



Step 5: Asks for Password. Enter password and click on "Login"

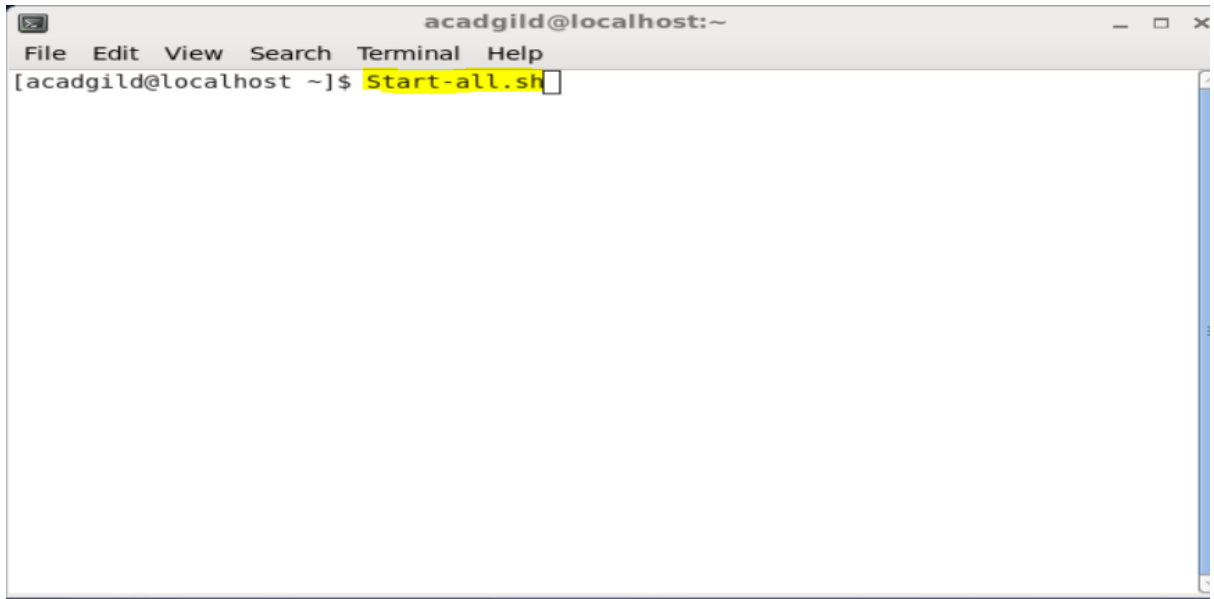


Step 6: Click on "Terminal" icon.

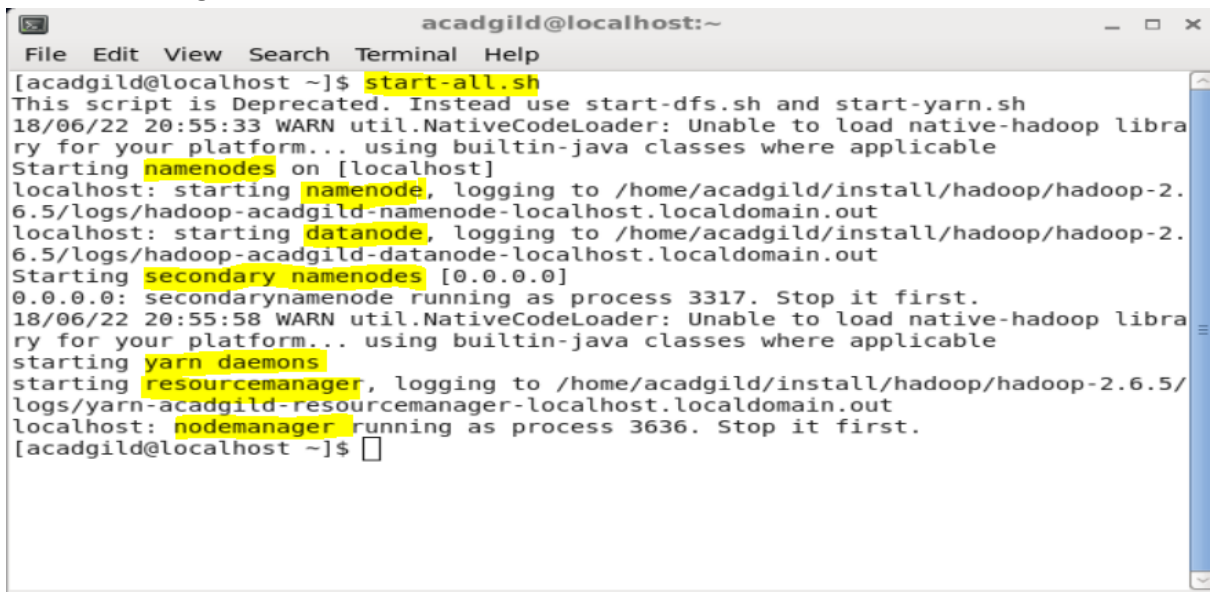


Big Data Hadoop Training- Assignment 1

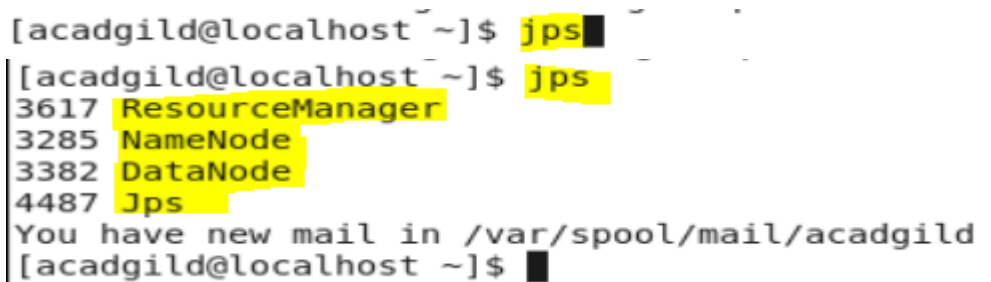
Step 7: Enter "Start-all.sh" and click "Enter"

A terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command prompt shows '[acadgild@localhost ~]\$ Start-all.sh' with the command highlighted in yellow.

Step 8: Which Starts Namenode, Datanode, Secondary name nodes, Yarn daemons and Resource Manager

A terminal window titled 'acadgild@localhost:~' showing the output of the 'start-all.sh' command. The output includes warnings about deprecated scripts and native-hadoop library loading, followed by the starting of namenode, datanode, secondary namenodes, yarn daemons, and the resourcemanager. The command prompt returns to '[acadgild@localhost ~]\$'.

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

A terminal window showing the output of the 'jps' command. The output lists four processes: 3617 ResourceManager, 3285 NameNode, 3382 DataNode, and 4487 Jps. Below this, a message states 'You have new mail in /var/spool/mail/acadgild'. The command prompt returns to '[acadgild@localhost ~]\$'.

Big Data Hadoop Training- Assignment 1

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

I. Ls -lt command – displays list of file along with date and time.

```
[acadgild@localhost ~]$ ls -lt
total 68
-rw-rw-r--. 1 acadgild acadgild  55 Jun 24 15:00 locations.txt
-rw-rw-r--. 1 acadgild acadgild 115 Jun 24 14:59 users.txt
-rw-rw-r--. 1 acadgild acadgild 108 Jun 24 14:47 employee.csv
-rw-rw-r--. 1 acadgild acadgild 170 Jun 24 14:31 complex_data.txt
-rw-rw-r--. 1 acadgild acadgild  62 Jun 24 12:06 emp_data1.txt
-rw-rw-r--. 1 acadgild acadgild  97 Jun 24 12:00 emp_data.txt
drwxr-xr-x. 2 acadgild acadgild 4096 Feb 13 14:24 Downloads
drwxrwxr-x. 13 acadgild acadgild 4096 Feb  9 18:06 install
drwxr-xr-x. 2 acadgild acadgild 4096 Feb  2 12:52 Documents
drwxr-xr-x. 2 acadgild acadgild 4096 Feb  2 12:51 Desktop
drwxrwxr-x. 3 acadgild acadgild 4096 Jan 16 14:02 eclipse-workspace
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29 16:59 eclipse
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Videos
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
```

II. Creating a file and inserting data into it and checking data present in the file using “cat” command

acadgild@localhost:~

```
[acadgild@localhost ~]$ vi data.txt
```

acadgild@localhost:~

```
[acadgild@localhost ~]$ vi data.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ cat data.txt
Happy world
Have a nice day
Be smiling.
[acadgild@localhost ~]$
```

Big Data Hadoop Training- Assignment 1

III. Creating a directory and listing the subdirectories

```
acadgild@localhost:~  
[acadgild@localhost ~]$ mkdir -p tech/net/quest  
[acadgild@localhost ~]$ ls tech  
net  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ ls tech/net  
quest  
[acadgild@localhost ~]$
```

```
acadgild@localhost:~/tech/net/quest  
[acadgild@localhost ~]$ cd tech/net/quest  
[acadgild@localhost quest]$
```

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

```
[acadgild@localhost ~]$ nano data.txt  
You have new mail in /var/spool/mail/acadgild
```

```
GNU nano 2.0.9 File: data.txt Modified  
  
Happy world  
Have a nice day  
Be smiling.  
  
[ Read 4 lines ]  
^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 ~]$ cat data.txt  
  
Happy world  
Have a nice day  
Be smiling.  
  
You have new mail in /var/spool/mail/acadgild
```

Big Data Hadoop Training- Assignment 1

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

Step1 : Main page which displays overview

The screenshot shows the Hadoop NameNode Overview page. The browser address bar displays 'localhost:50070/dfshealth.html#tab-overview'. The page has a green navigation bar with tabs: Hadoop, Overview (selected), Datanodes, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Overview 'localhost:8020' (active)'. Below the title is a table with the following details:

Started:	Thu Jun 28 23:17:58 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

Below the table is a 'Summary' section with the following text:

Security is off.
Safemode is off.
85 files and directories, 48 blocks = 133 total filesystem object(s).
Heap Memory used 30.89 MB of 62.06 MB Heap Memory. Max Heap Memory is 966.69 MB.

Step2: Datanode Information

The screenshot shows the Hadoop NameNode Datanode Information page. The browser address bar displays 'localhost:50070/dfshealth.html#tab-datanode'. The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes (selected), Snapshot, Startup Progress, and Utilities. The main content area is titled 'Datanode Information'. Below the title is a section 'In operation' with a table showing the status of the datanode:

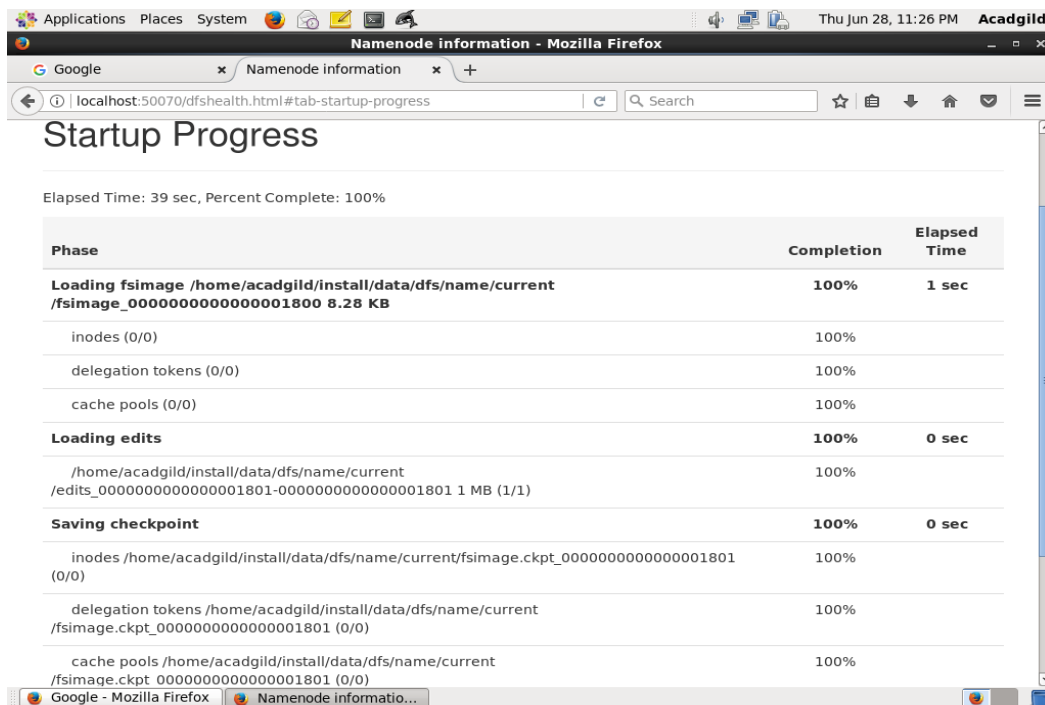
Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
localhost (127.0.0.1:50010)	1	In Service	17.11 GB	4.17 MB	11.74 GB	5.36 GB	48	4.17 MB (0.02%)	0	2.6.5

Below the 'In operation' section is a 'Decommissioning' section with a table showing the status of decommissioning:

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks in files under construction
------	--------------	-------------------------	------------------------------	---

Big Data Hadoop Training- Assignment 1

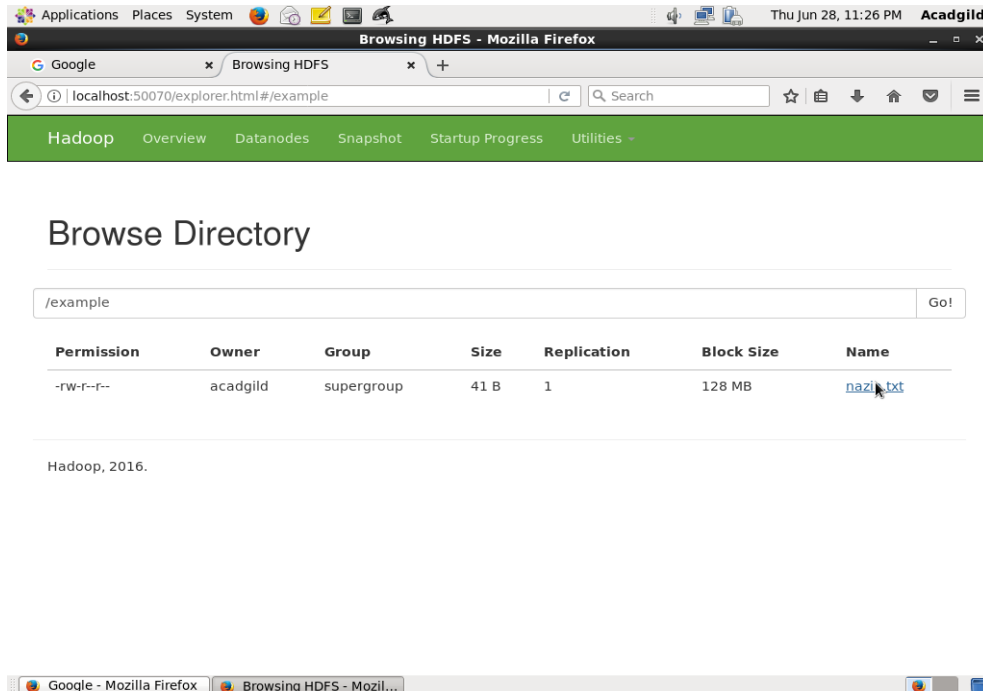
Step 3: Edit log details



The screenshot shows the 'Startup Progress' page of the Hadoop NameNode. The browser is Mozilla Firefox, and the URL is localhost:50070/dfshealth.html#tab-startup-progress. The page title is 'Startup Progress'. Below the title, it says 'Elapsed Time: 39 sec, Percent Complete: 100%'. A table lists the phases of the startup process, their completion status, and elapsed time.

Phase	Completion	Elapsed Time
Loading fsimage /home/acadgild/install/data/dfs/name/current/fsimage_00000000000000001800 8.28 KB	100%	1 sec
inodes (0/0)	100%	
delegation tokens (0/0)	100%	
cache pools (0/0)	100%	
Loading edits	100%	0 sec
/home/acadgild/install/data/dfs/name/current/edits_00000000000000001801-00000000000000001801 1 MB (1/1)	100%	
Saving checkpoint	100%	0 sec
inodes /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_00000000000000001801 (0/0)	100%	
delegation tokens /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_00000000000000001801 (0/0)	100%	
cache pools /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_00000000000000001801 (0/0)	100%	

Step 4: Directory Details



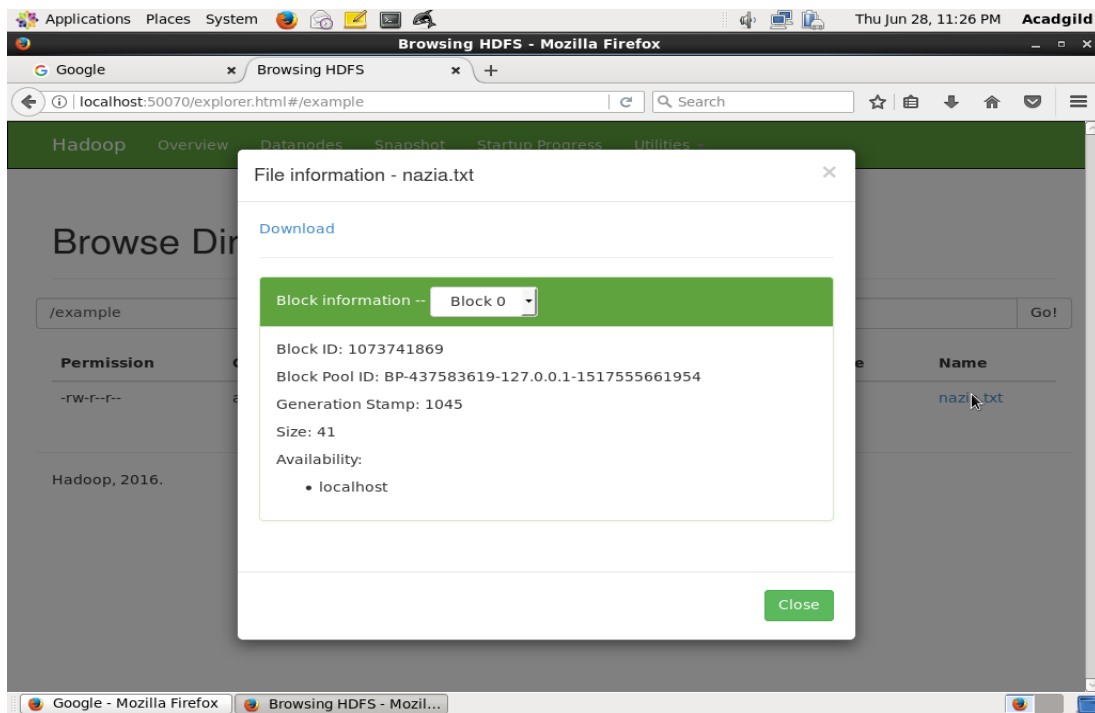
The screenshot shows the 'Browse HDFS' page of the Hadoop NameNode. The browser is Mozilla Firefox, and the URL is localhost:50070/explorer.html#/example. The page title is 'Browse HDFS'. Below the title, there is a navigation bar with links: Hadoop, Overview, Datanodes, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Browse Directory'. It shows a search bar with the text '/example' and a 'Go!' button. Below the search bar, there is a table listing the details of the directory contents.

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	acadgild	supergroup	41 B	1	128 MB	nazi.txt

Hadoop, 2016.

Big Data Hadoop Training- Assignment 1

Step 5: File details with number of blocks created, block id , size and so on.



Big Data Hadoop Training- Assignment 1

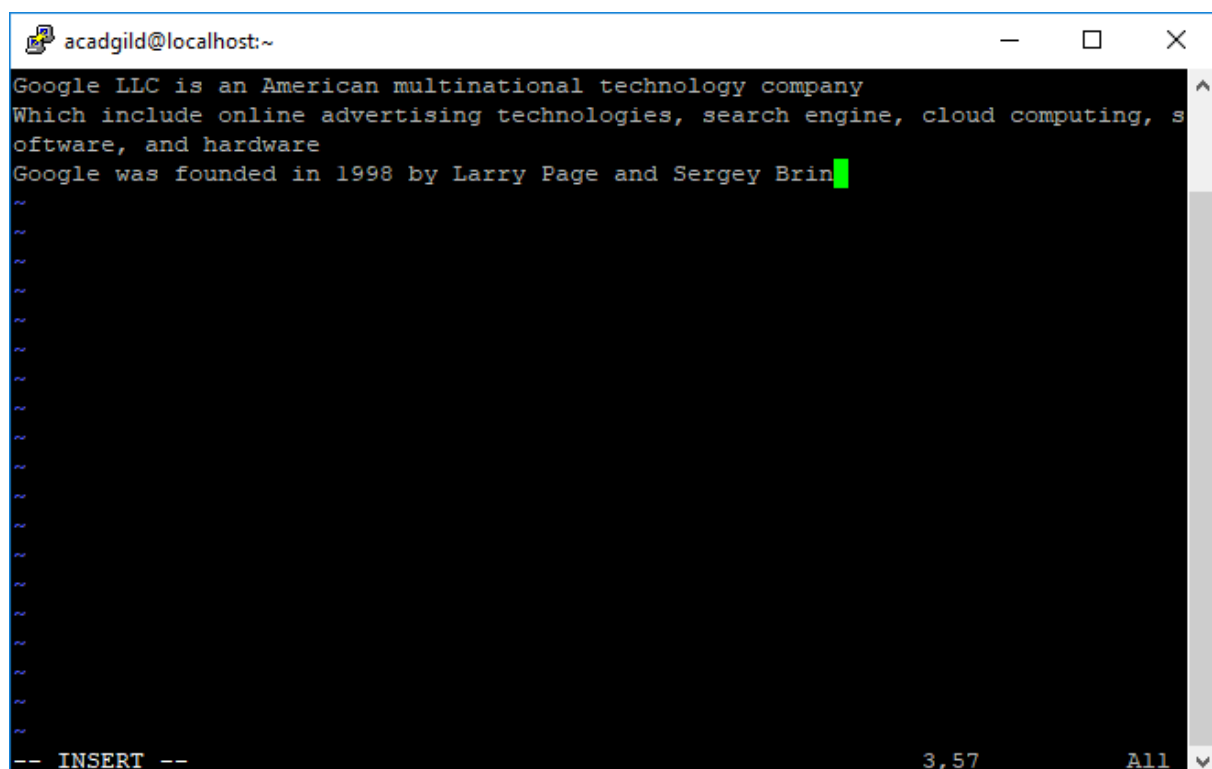
Task 2:

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[This directory already exists in the HDFS]

```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild
18/07/07 18:11:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - acadgild supergroup          0 2018-07-07 18:10 /user/acadgild/hadoop
drwxr-xr-x - acadgild supergroup          0 2018-06-24 13:03 /user/acadgild/hive
```

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.

A few lines are written on the screen and saved in a file. This file content is appended into a new file name 'word-count.txt' in the directory /user/acadgild/Hadoop using "-appendToFile" command.



The screenshot shows a terminal window titled 'acadgild@localhost:~'. The terminal displays the following text:

```
Google LLC is an American multinational technology company
Which include online advertising technologies, search engine, cloud computing, s
oftware, and hardware
Google was founded in 1998 by Larry Page and Sergey Brin
```

The text is being typed line by line, and a green cursor is visible at the end of the last line. The terminal window has standard window controls (minimize, maximize, close) in the top right corner. At the bottom of the terminal, there is a status bar that reads '-- INSERT --' on the left, '3,57' in the center, and 'All' on the right.

Big Data Hadoop Training- Assignment 1

```
acadgild@localhost:~  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ vi wordcount.txt  
[acadgild@localhost ~]$ hadoop fs -appendToFile wordcount.txt /user/acadgild/hadoop/wordcount.txt  
18/07/14 22:43:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop  
18/07/14 22:44:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Found 1 items  
-rw-r--r--  1 acadgild supergroup          218 2018-07-14 22:43 /user/acadgild/hadoop/wordcount.txt  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/wordcount.txt  
18/07/14 22:44:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Google LLC is an American multinational technology company  
Which include online advertising technologies, search engine, cloud computing, software, and hardware  
Google was founded in 1998 by Larry Page and Sergey Brin  
[acadgild@localhost ~]$
```

3. Create a file max-temp.txt in local FS. Put some 10-15 records of date and temperature.

Example: dd-mm-yyyy, temperature

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

maxtemp.txt file is created in local FS and then the file is moved to HDFS using -put command. The -put command creates a new file with the file name max-temp.txt itself in path /user/acadgild/Hadoop.

```
[acadgild@localhost ~]$ vi maxtemp.txt  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ ls  
complex_data.txt      hadoop-mapreduce-examples-2.6.5.jar  regex1.txt  
csvdata.txt           hive-hcatalog-core-0.14.0.jar         regex.txt  
datal.txt             install                               sample_input.txt  
datainput.txt         locations.txt                          tech  
data.txt              maxtemp.txt                           Templates  
Desktop               Music                                 testdata.txt  
Documents             myjar.jar                             users.txt  
Downloads             parking_violation_filter.csv          Videos  
eclipse               Pictures                              wordcount.txt  
eclipse-workspace     pig                                  wordmean.txt  
emp_datal.txt         pig_1530346642754.log                wordstd.txt  
emp_data.txt          pis-useHCatalog  
employee.csv          Public
```

Big Data Hadoop Training- Assignment 1

```
acadgild@localhost:~  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop fs -put maxtemp.txt /user/acadgild/hadoop  
18/07/14 22:53:48 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop  
18/07/14 22:54:28 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
Found 2 items  
-rw-r--r-- 1 acadgild supergroup 141 2018-07-14 22:53 /user/acadgild/ha  
doop/maxtemp.txt  
-rw-r--r-- 1 acadgild supergroup 218 2018-07-14 22:43 /user/acadgild/ha  
doop/wordcount.txt  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/hadoop/maxtemp.txt  
18/07/14 22:55:06 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
10-02-1990,23  
23-03-1991,28  
24-05-1990,34  
25-07-1992,19  
25-07-2000,20  
20-10-1994,30  
20-09-2000,30  
26-01-1990,40  
23-08-1995,60  
18-08-2008,80
```

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 permission access is changed by using command “-chmod”.

```
[acadgild@localhost ~]$ hadoop fs -chmod 777 /user/acadgild/hadoop/maxtemp.txt  
18/07/14 22:58:04 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop  
18/07/14 22:58:27 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
Found 2 items  
-rwxrwxrwx 1 acadgild supergroup 141 2018-07-14 22:53 /user/acadgild/ha  
doop/maxtemp.txt  
-rw-r--r-- 1 acadgild supergroup 218 2018-07-14 22:43 /user/acadgild/ha  
doop/wordcount.txt  
[acadgild@localhost ~]$ hadoop fs -chmod 770 /user/acadgild/hadoop/maxtemp.txt  
18/07/14 22:59:03 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop  
18/07/14 22:59:24 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
Found 2 items  
-rwxrwx--- 1 acadgild supergroup 141 2018-07-14 22:53 /user/acadgild/ha  
doop/maxtemp.txt  
-rw-r--r-- 1 acadgild supergroup 218 2018-07-14 22:43 /user/acadgild/ha  
doop/wordcount.txt
```

Big Data Hadoop Training- Assignment 1

Task 3:

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

a) Wordmedian

```
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /data1.txt /dataout.txt
18/07/08 17:23:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/08 17:23:40 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/08 17:23:41 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 17:23:42 INFO input.FileInputFormat: Total input paths to process : 1
18/07/08 17:23:42 INFO mapreduce.JobSubmitter: number of splits:1
18/07/08 17:23:44 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531033752353_0002
18/07/08 17:23:46 INFO impl.YarnClientImpl: Submitted application application_1531033752353_0002
18/07/08 17:23:46 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531033752353_0002/
18/07/08 17:23:46 INFO mapreduce.Job: Running job: job_1531033752353_0002
18/07/08 17:23:59 INFO mapreduce.Job: Job job_1531033752353_0002 running in uber mode : false
18/07/08 17:23:59 INFO mapreduce.Job: map 0% reduce 0%
18/07/08 17:24:07 INFO mapreduce.Job: map 100% reduce 0%
18/07/08 17:24:15 INFO mapreduce.Job: map 100% reduce 100%
18/07/08 17:24:16 INFO mapreduce.Job: Job job_1531033752353_0002 completed successfully
18/07/08 17:24:17 INFO mapreduce.Job: Counters: 49
```

```
The median is: 5
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /
```

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/07/08 17:24:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 11 items
-rw-r--r-- 1 acadgild supergroup 73 2018-07-08 17:23 /data1.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-08 17:24 /dataout.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-01 11:21 /hbase
-rw-r--r-- 1 acadgild supergroup 125 2018-07-08 12:41 /out_file
-rw-r--r-- 1 acadgild supergroup 105 2018-07-05 05:08 /output_file.txt
-rw-r--r-- 1 acadgild supergroup 125 2018-07-05 05:03 /sample_input
-rw-r--r-- 1 acadgild supergroup 105 2018-07-05 05:07 /sample_input.txt
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sgoopout111
-rw-r--r-- 1 acadgild supergroup 44 2018-07-08 13:41 /testout.txt
drwxrwx--- - acadgild supergroup 0 2018-06-24 12:34 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-06-24 13:01 /user
```

```
[acadgild@localhost ~]$ hadoop fs -ls /dataout.txt
18/07/08 17:24:40 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 17:24 /dataout.txt/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 16 2018-07-08 17:24 /dataout.txt/part-r-000000
[acadgild@localhost ~]$ hadoop fs -cat /dataout.txt/part-r-000000
18/07/08 17:25:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
4      2
5      6
6      1
9      2
```

Big Data Hadoop Training- Assignment 1

2. WordMean

```
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /wordmean1.txt /wordmeanout.txt
18/07/14 18:22:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/14 18:22:10 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/14 18:22:11 INFO input.FileInputFormat: Total input paths to process : 1
18/07/14 18:22:11 INFO mapreduce.JobSubmitter: number of splits:1
18/07/14 18:22:12 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531571980861_0002
18/07/14 18:22:12 INFO impl.YarnClientImpl: Submitted application application_1531571980861_0002
18/07/14 18:22:12 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531571980861_0002/
18/07/14 18:22:12 INFO mapreduce.Job: Running job: job_1531571980861_0002
18/07/14 18:22:23 INFO mapreduce.Job: Job job_1531571980861_0002 running in uber mode : false
18/07/14 18:22:23 INFO mapreduce.Job: map 0% reduce 0%
18/07/14 18:22:31 INFO mapreduce.Job: map 100% reduce 0%
18/07/14 18:22:39 INFO mapreduce.Job: map 100% reduce 100%
```

```

The mean is: 6.4
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/07/14 18:25:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 14 items
-rw-r--r--  1 acadgild supergroup      73 2018-07-08 17:23 /data1.txt
drwxr-xr-x  - acadgild supergroup      0 2018-07-08 17:24 /dataout.txt
drwxr-xr-x  - acadgild supergroup      0 2018-07-01 11:21 /hbase
-rw-r--r--  1 acadgild supergroup    125 2018-07-08 12:41 /out_file
-rw-r--r--  1 acadgild supergroup    105 2018-07-05 05:08 /output_file.txt
-rw-r--r--  1 acadgild supergroup    125 2018-07-05 05:03 /sample_input
-rw-r--r--  1 acadgild supergroup    105 2018-07-05 05:07 /sample_input.txt
drwxr-xr-x  - acadgild supergroup      0 2018-02-02 12:49 /sqoopout111
-rw-r--r--  1 acadgild supergroup     44 2018-07-08 13:41 /testout.txt
drwxrwx---  - acadgild supergroup      0 2018-06-24 12:34 /tmp
drwxr-xr-x  - acadgild supergroup      0 2018-06-24 13:01 /user
-rw-r--r--  1 acadgild supergroup     37 2018-07-14 18:20 /wordmean1.txt
drwxr-xr-x  - acadgild supergroup      0 2018-07-14 18:22 /wordmeanout.txt
drwxr-xr-x  - acadgild supergroup      0 2018-07-14 18:10 /wordout.txt
```

```
[acadgild@localhost ~]$ hadoop fs -ls /wordmeanout.txt
18/07/14 18:27:38 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 2018-07-14 18:22 /wordmeanout.txt/
_SUCCESS
-rw-r--r--  1 acadgild supergroup    18 2018-07-14 18:22 /wordmeanout.txt/
part-r-00000
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -cat /wordmeanout.txt/part-r-00000
18/07/14 18:30:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
count    5
length  32
```


Big Data Hadoop Training- Assignment 1

3. WordStandard Deviation

```
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstand
arddeviation /wordstdl.txt /wordstdout.txt
18/07/14 18:34:10 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
18/07/14 18:34:11 INFO client.RMPProxy: Connecting to ResourceManager at localhos
t/127.0.0.1:8032
18/07/14 18:34:13 INFO input.FileInputFormat: Total input paths to process : 1
18/07/14 18:34:13 INFO mapreduce.JobSubmitter: number of splits:1
18/07/14 18:34:13 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_15
31571980861_0003
18/07/14 18:34:13 INFO impl.YarnClientImpl: Submitted application application_15
31571980861_0003
18/07/14 18:34:13 INFO mapreduce.Job: The url to track the job: http://localhost
:8088/proxy/application_1531571980861_0003/
18/07/14 18:34:13 INFO mapreduce.Job: Running job: job_1531571980861_0003
18/07/14 18:34:25 INFO mapreduce.Job: Job job_1531571980861_0003 running in uber
mode : false
18/07/14 18:34:25 INFO mapreduce.Job: map 0% reduce 0%
18/07/14 18:34:32 INFO mapreduce.Job: map 100% reduce 0%
18/07/14 18:34:40 INFO mapreduce.Job: map 100% reduce 100%
```

```
The standard deviation is: 1.2583057392117925
You have new mail in /var/spool/mail/acadgild
```

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/07/14 18:36:37 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 16 items
-rw-r--r-- 1 acadgild supergroup 73 2018-07-08 17:23 /datal.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-08 17:24 /dataout.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-01 11:21 /hbase
-rw-r--r-- 1 acadgild supergroup 125 2018-07-08 12:41 /out_file
-rw-r--r-- 1 acadgild supergroup 105 2018-07-05 05:08 /output_file.txt
-rw-r--r-- 1 acadgild supergroup 125 2018-07-05 05:03 /sample_input
-rw-r--r-- 1 acadgild supergroup 105 2018-07-05 05:07 /sample_input.txt
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopoutlll
-rw-r--r-- 1 acadgild supergroup 44 2018-07-08 13:41 /testout.txt
drwxrwx--- - acadgild supergroup 0 2018-06-24 12:34 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-06-24 13:01 /user
-rw-r--r-- 1 acadgild supergroup 37 2018-07-14 18:20 /wordmeanl.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-14 18:22 /wordmeanout.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-14 18:10 /wordout.txt
-rw-r--r-- 1 acadgild supergroup 46 2018-07-14 18:33 /wordstdl.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-14 18:34 /wordstdout.txt
```

```
[acadgild@localhost ~]$ hadoop fs -ls /wordstdout.txt
18/07/14 18:37:53 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-07-14 18:34 /wordstdout.txt/_
SUCCESS
-rw-r--r-- 1 acadgild supergroup 29 2018-07-14 18:34 /wordstdout.txt/p
art-r-00000
You have new mail in /var/spool/mail/acadgild
```

```
[acadgild@localhost ~]$ hadoop fs -cat /wordstdout.txt/part-r-00000
18/07/14 18:39:31 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
count 6
length 39
square 263
```