

## Assignment # 1.1 – INTRODUCTION TO BIG DATA

Date: 22-Oct-2018

---

### TASK 1:

#### Task:

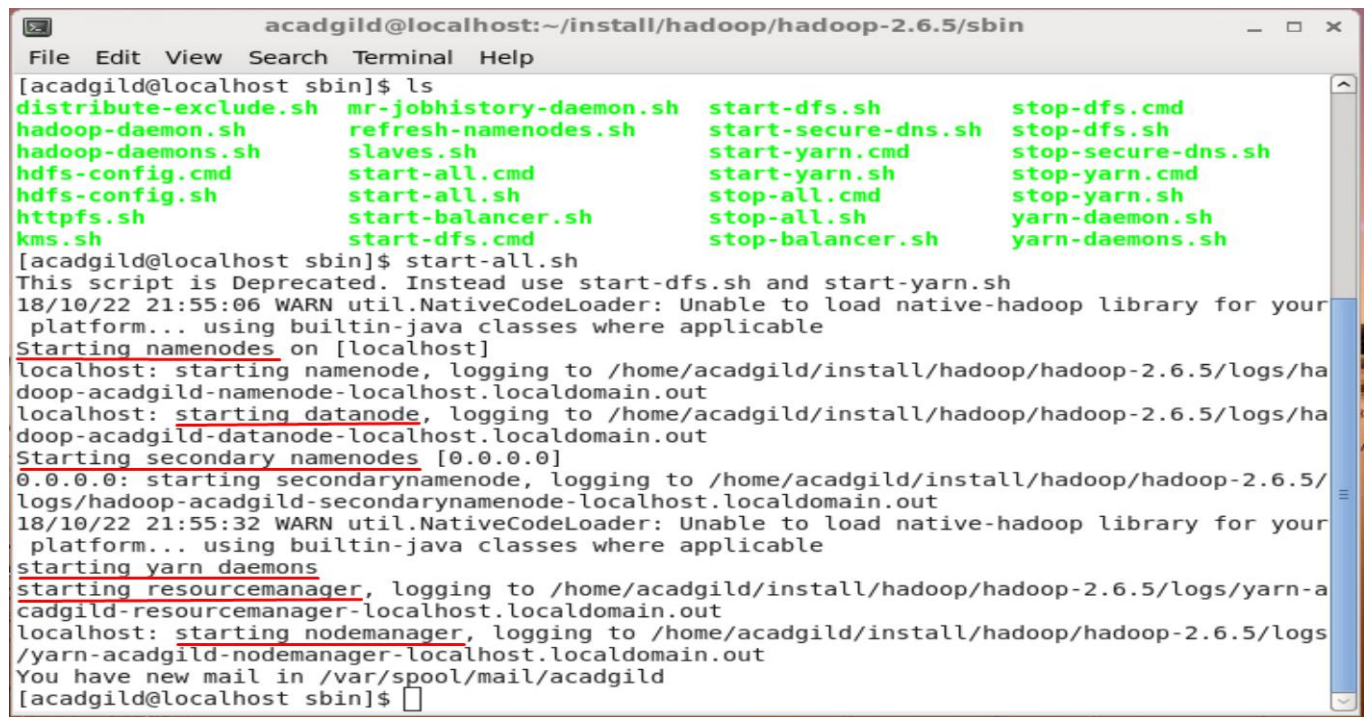
Start Hadoop single node on AcadGild VM. The command is *start-all.sh*.

#### Explanation:

Start-all.sh is a shell script that starts HDFS and Yarn framework. As part of HDFS, Namenode, Secondary Namenode and Datanode are started. And, as part of Yarn framework, Resource manager, Node manager are also started.

#### Output:

Below is the screenshot of start-all.sh command executed in terminal.



```
acadgild@localhost: ~/install/hadoop/hadoop-2.6.5/sbin
File Edit View Search Terminal Help
[acadgild@localhost sbin]$ ls
distribute-exclude.sh  mr-jobhistory-daemon.sh  start-dfs.sh  stop-dfs.cmd
hadoop-daemon.sh      refresh-namenodes.sh     start-secure-dns.sh  stop-dfs.sh
hadoop-daemons.sh    slaves.sh                 start-yarn.cmd  stop-secure-dns.sh
hdfs-config.cmd       start-all.cmd            start-yarn.sh    stop-yarn.cmd
hdfs-config.sh        start-all.sh             stop-all.cmd    stop-yarn.sh
httpfs.sh              start-balancer.sh         stop-all.sh     yarn-daemon.sh
kms.sh                 start-dfs.cmd             stop-balancer.sh  yarn-daemons.sh
[acadgild@localhost sbin]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/10/22 21:55:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/ha
doop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/ha
doop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/
logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/10/22 21:55:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-a
cadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs
/yarn-acadgild-nodemanager-localhost.localdomain.out
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$
```

## TASK 2:

### Task:

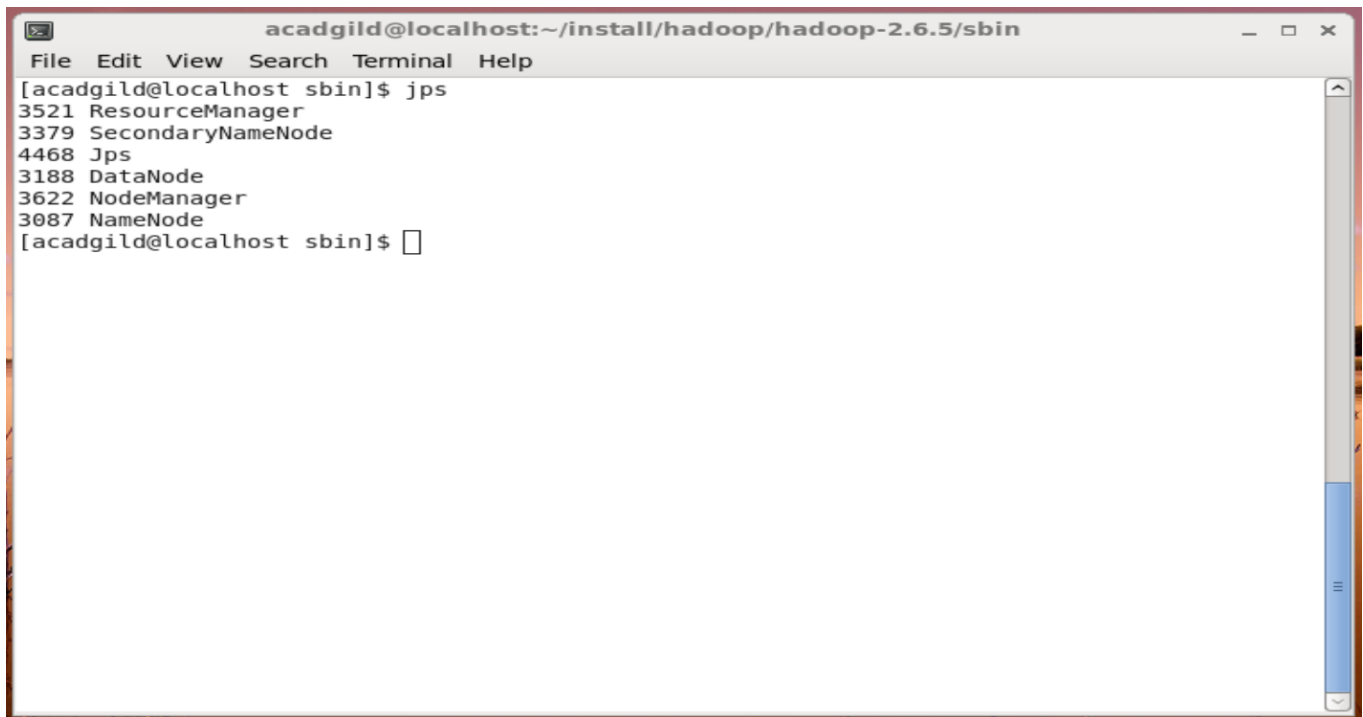
Run a JPS command to see if all Hadoop daemons are running.

### Explanation:

JPS command will show all the active java processes currently running in the virtual machine.

### Output:

Below is the screenshot of the command executed in terminal.

A screenshot of a terminal window titled 'acadgild@localhost:~/install/hadoop/hadoop-2.6.5/sbin'. The terminal shows the command '[acadgild@localhost sbin]\$ jps' and its output: '3521 ResourceManager', '3379 SecondaryNameNode', '4468 Jps', '3188 DataNode', '3622 NodeManager', and '3087 NameNode'. The prompt '[acadgild@localhost sbin]\$' is followed by a cursor. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'.

```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/sbin
File Edit View Search Terminal Help
[acadgild@localhost sbin]$ jps
3521 ResourceManager
3379 SecondaryNameNode
4468 Jps
3188 DataNode
3622 NodeManager
3087 NameNode
[acadgild@localhost sbin]$
```

### TASK 3:

#### Task:

Run few Unix commands like pwd, ls -ls , etc.

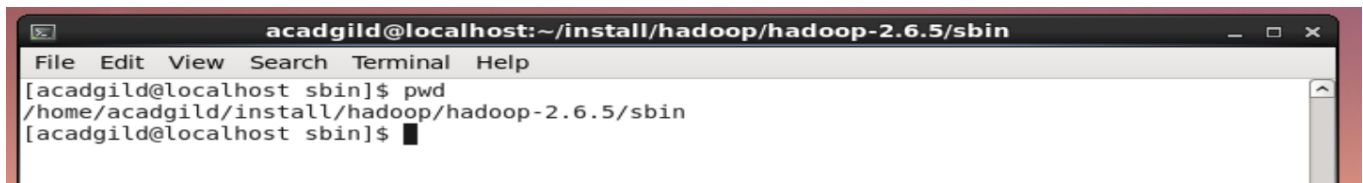
#### Explanation:

- pwd – This command will show the current working directory in the terminal
- ls – This command will list all the files and folders in the current working directory of the terminal.
- ls -ls – This command will provide the detailed list of all files and folders in the current working directory of the terminal
- mkdir <directory\_name> - This command will create a new directory in the current working directory.
- rmdir <directory\_name> - This command will delete the specified directory from the current working directory.

#### Output:

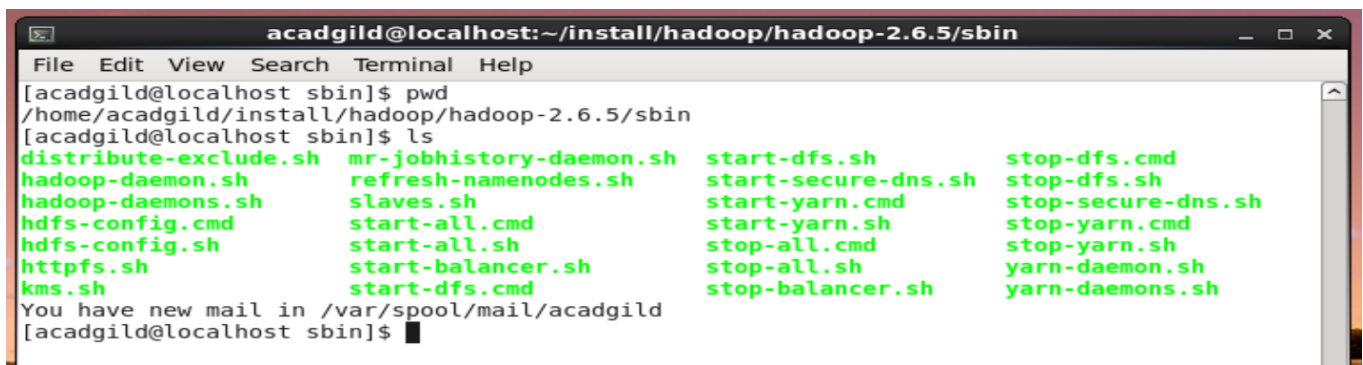
Below is the screenshot of the command executed in the terminal.

Command: pwd



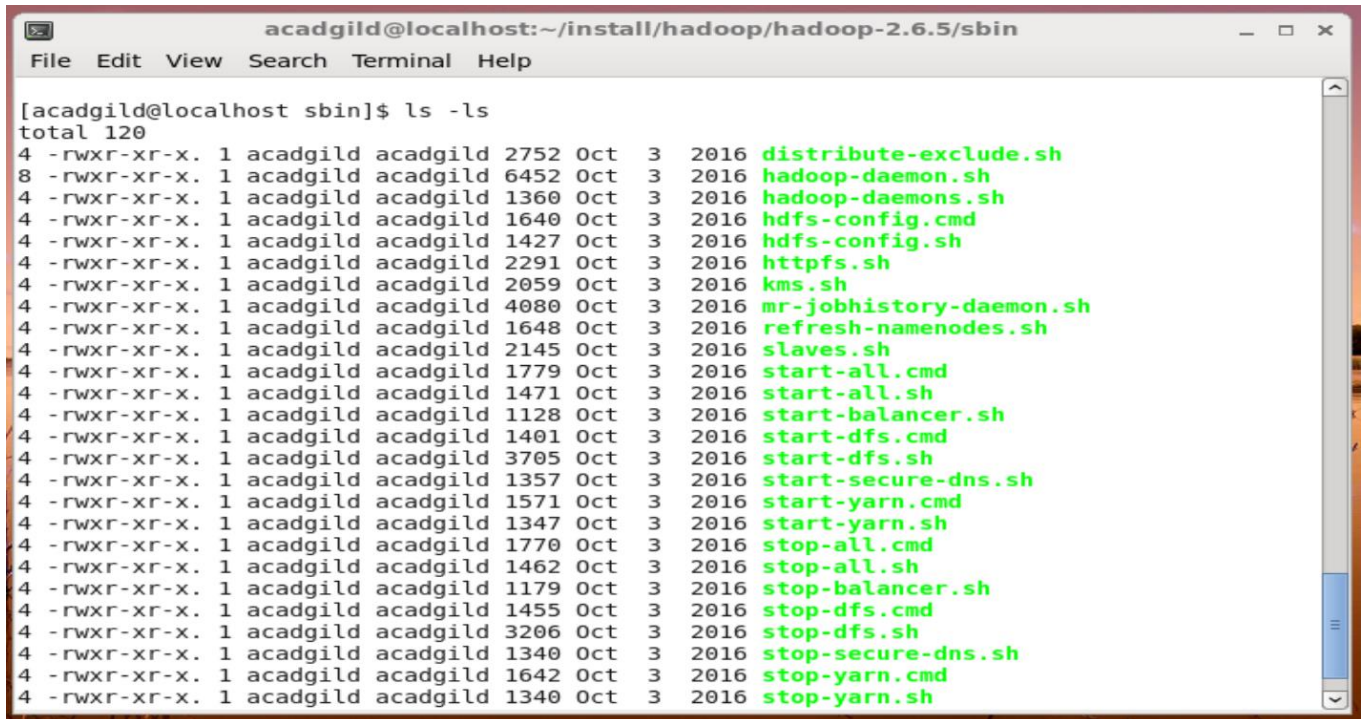
```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/sbin
File Edit View Search Terminal Help
[acadgild@localhost sbin]$ pwd
/home/acadgild/install/hadoop/hadoop-2.6.5/sbin
[acadgild@localhost sbin]$
```

Command: ls



```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/sbin
File Edit View Search Terminal Help
[acadgild@localhost sbin]$ pwd
/home/acadgild/install/hadoop/hadoop-2.6.5/sbin
[acadgild@localhost sbin]$ ls
distributed-exclude.sh  mr-jobhistory-daemon.sh  start-dfs.sh  stop-dfs.cmd
hadoop-daemon.sh       refresh-namenodes.sh     start-secure-dns.sh  stop-secure-dns.sh
hadoop-daemons.sh     slaves.sh                 start-yarn.cmd       stop-yarn.cmd
hdfs-config.cmd        start-all.cmd            start-yarn.sh        stop-yarn.sh
hdfs-config.sh         start-all.sh            stop-all.cmd        stop-yarn.sh
httpfs.sh              start-balancer.sh        stop-all.sh         yarn-daemon.sh
kms.sh                 start-dfs.cmd            stop-balancer.sh     yarn-daemons.sh
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$
```

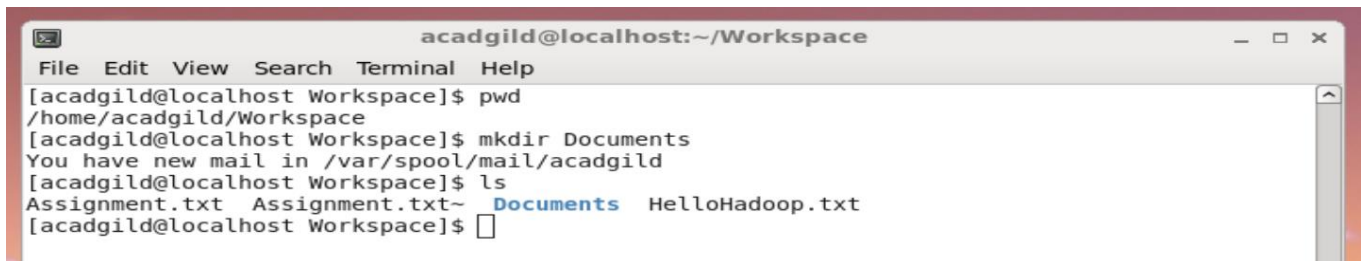
Command: `ls -ls`



```
acadgild@localhost:~/install/hadoop/hadoop-2.6.5/sbin
File Edit View Search Terminal Help

[acadgild@localhost sbin]$ ls -ls
total 120
4 -rwxr-xr-x. 1 acadgild acadgild 2752 Oct 3 2016 distribute-exclude.sh
8 -rwxr-xr-x. 1 acadgild acadgild 6452 Oct 3 2016 hadoop-daemon.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1360 Oct 3 2016 hadoop-daemons.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1640 Oct 3 2016 hdfs-config.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 1427 Oct 3 2016 hdfs-config.sh
4 -rwxr-xr-x. 1 acadgild acadgild 2291 Oct 3 2016 httpfs.sh
4 -rwxr-xr-x. 1 acadgild acadgild 2059 Oct 3 2016 kms.sh
4 -rwxr-xr-x. 1 acadgild acadgild 4080 Oct 3 2016 mr-jobhistory-daemon.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1648 Oct 3 2016 refresh-namenodes.sh
4 -rwxr-xr-x. 1 acadgild acadgild 2145 Oct 3 2016 slaves.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1779 Oct 3 2016 start-all.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 1471 Oct 3 2016 start-all.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1128 Oct 3 2016 start-balancer.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1401 Oct 3 2016 start-dfs.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 3705 Oct 3 2016 start-dfs.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1357 Oct 3 2016 start-secure-dns.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1571 Oct 3 2016 start-yarn.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 1347 Oct 3 2016 start-yarn.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1770 Oct 3 2016 stop-all.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 1462 Oct 3 2016 stop-all.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1179 Oct 3 2016 stop-balancer.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1455 Oct 3 2016 stop-dfs.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 3206 Oct 3 2016 stop-dfs.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1340 Oct 3 2016 stop-secure-dns.sh
4 -rwxr-xr-x. 1 acadgild acadgild 1642 Oct 3 2016 stop-yarn.cmd
4 -rwxr-xr-x. 1 acadgild acadgild 1340 Oct 3 2016 stop-yarn.sh
```

Command: `mkdir`



```
acadgild@localhost:~/Workspace
File Edit View Search Terminal Help

[acadgild@localhost Workspace]$ pwd
/home/acadgild/Workspace
[acadgild@localhost Workspace]$ mkdir Documents
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Workspace]$ ls
Assignment.txt Assignment.txt~ Documents HelloHadoop.txt
[acadgild@localhost Workspace]$
```

Command: `rmdir`



```
acadgild@localhost:~/Workspace
File Edit View Search Terminal Help

[acadgild@localhost Workspace]$ ls
Assignment.txt Assignment.txt~ Documents HelloHadoop.txt TestDocuments
[acadgild@localhost Workspace]$ rmdir TestDocuments
[acadgild@localhost Workspace]$ ls
Assignment.txt Assignment.txt~ Documents HelloHadoop.txt
[acadgild@localhost Workspace]$
```

## TASK 4:

### Task:

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

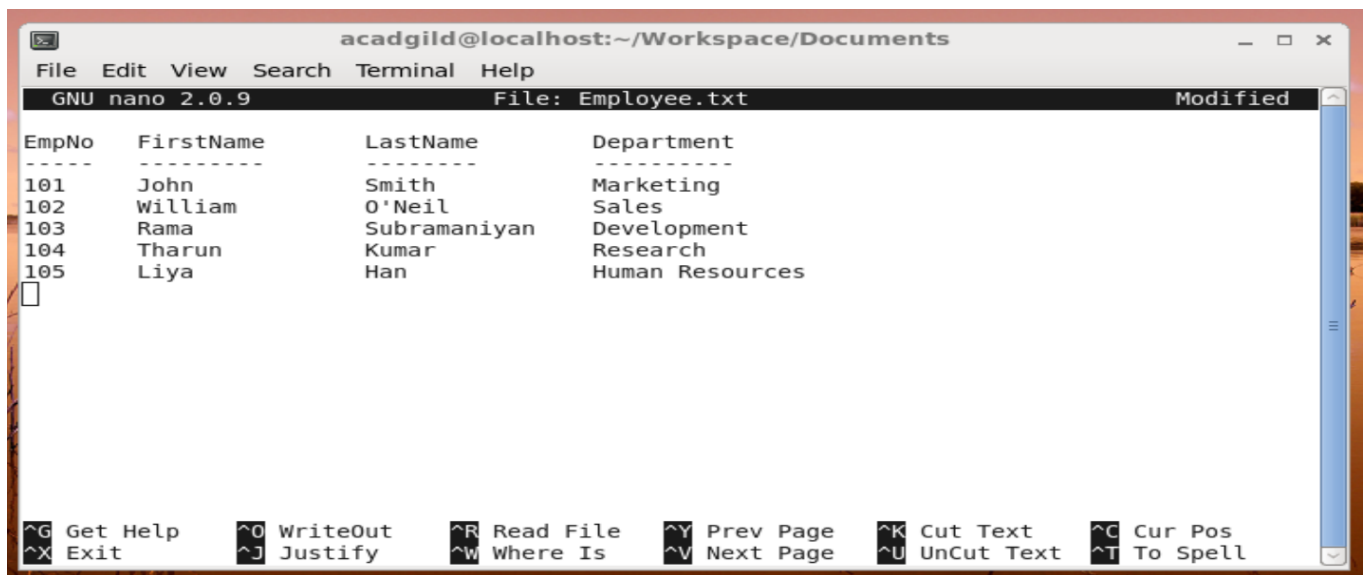
### Explanation:

- Command nano will be used to create a new file in nano editor
- Input the contents of the file and save it by ^O that will save the file. Press ^X will exit the editor window and return to command prompt.
- To retrieve the contents of the file, command cat can be used.

### Output:

Below are the screenshots of commands executed in terminal

i. Creating a new file called Employee.txt with nano command

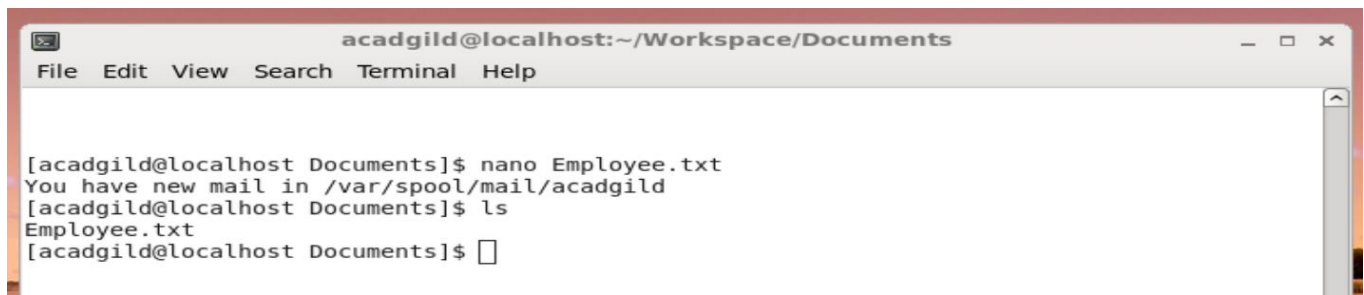


The screenshot shows the nano editor interface. The title bar indicates the user is 'acadgild@localhost' in the directory '~/Workspace/Documents'. The menu bar includes File, Edit, View, Search, Terminal, and Help. The status bar shows 'GNU nano 2.0.9', 'File: Employee.txt', and 'Modified'. The main text area contains a table with employee data:

EmpNo	FirstName	LastName	Department
101	John	Smith	Marketing
102	William	O'Neil	Sales
103	Rama	Subramaniyan	Development
104	Tharun	Kumar	Research
105	Liya	Han	Human Resources

The bottom status bar displays various keyboard shortcuts: ^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, and ^T To Spell.

ii. Saved the file and list the file created under the folder Documents



The screenshot shows a terminal window with the following commands and output:

```
[acadgild@localhost Documents]$ nano Employee.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Documents]$ ls
Employee.txt
[acadgild@localhost Documents]$
```

iii. Retrieve the contents of text file Employee.txt with cat command

```
acadgild@localhost:~/Workspace/Documents
File Edit View Search Terminal Help

[acadgild@localhost Documents]$ nano Employee.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Documents]$ ls
Employee.txt
[acadgild@localhost Documents]$ cat Employee.txt
EmpNo      FirstName      LastName      Department
-----
101        John          Smith         Marketing
102        William       O'Neil        Sales
103        Rama          Subramaniyan  Development
104        Tharun        Kumar         Research
105        Liya          Han           Human Resources

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost Documents]$
```

## TASK 5:

### Task:

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

### Explanation:

The URL "localhost:50070" provides the health status of namenode and datanode running in the cluster.

### Output:

The screenshot shows the 'Namenode information - Mozilla Firefox' window. The address bar displays 'localhost:50070/dfshealth.html#tab-overview'. The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes, Snapshot, Startup Progress, and Utilities. The 'Overview' tab is selected, showing the title 'Overview 'localhost:8020' (active)'. Below the title is a table with the following data:

Started:	Mon Oct 22 21:55:10 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.  
29 files and directories, 4 blocks = 33 total filesystem object(s).  
Heap Memory used 36.72 MB of 50.01 MB Heap Memory. Max Heap Memory is 966.69 MB.

In the bottom right corner, there is a small 'Activate Windows' watermark.



Heap Memory used 36.72 MB of 50.01 MB Heap Memory. Max Heap Memory is 966.69 MB.	
Non Heap Memory used 39.73 MB of 40.5 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.	
<b>Configured Capacity:</b>	17.11 GB
<b>DFS Used:</b>	208 KB
<b>Non DFS Used:</b>	11.13 GB
<b>DFS Remaining:</b>	5.97 GB
<b>DFS Used%:</b>	0%
<b>DFS Remaining%:</b>	34.92%
<b>Block Pool Used:</b>	208 KB
<b>Block Pool Used%:</b>	0%
<b>DataNodes usages% (Min/Median/Max/stdDev):</b>	0.00% / 0.00% / 0.00% / 0.00%
<b>Live Nodes</b>	1 (Decommissioned: 0)
<b>Dead Nodes</b>	0 (Decommissioned: 0)
<b>Decommissioning Nodes</b>	0
<b>Number of Under-Replicated Blocks</b>	0
<b>Number of Blocks Pending Deletion</b>	0
<b>Block Deletion Start Time</b>	10/22/2018, 9:55:10 PM

Namenode information	
localhost:50070/dfshealth.html#tab-datanode	
Hadoop	Overview
Datanodes	Snapshot
Startup Progress	Utilities

## Datanode Information

### In operation

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	208 KB	11.13 GB	5.97 GB	4	208 KB (0%)	0	2.6.5

### Decommissioning

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