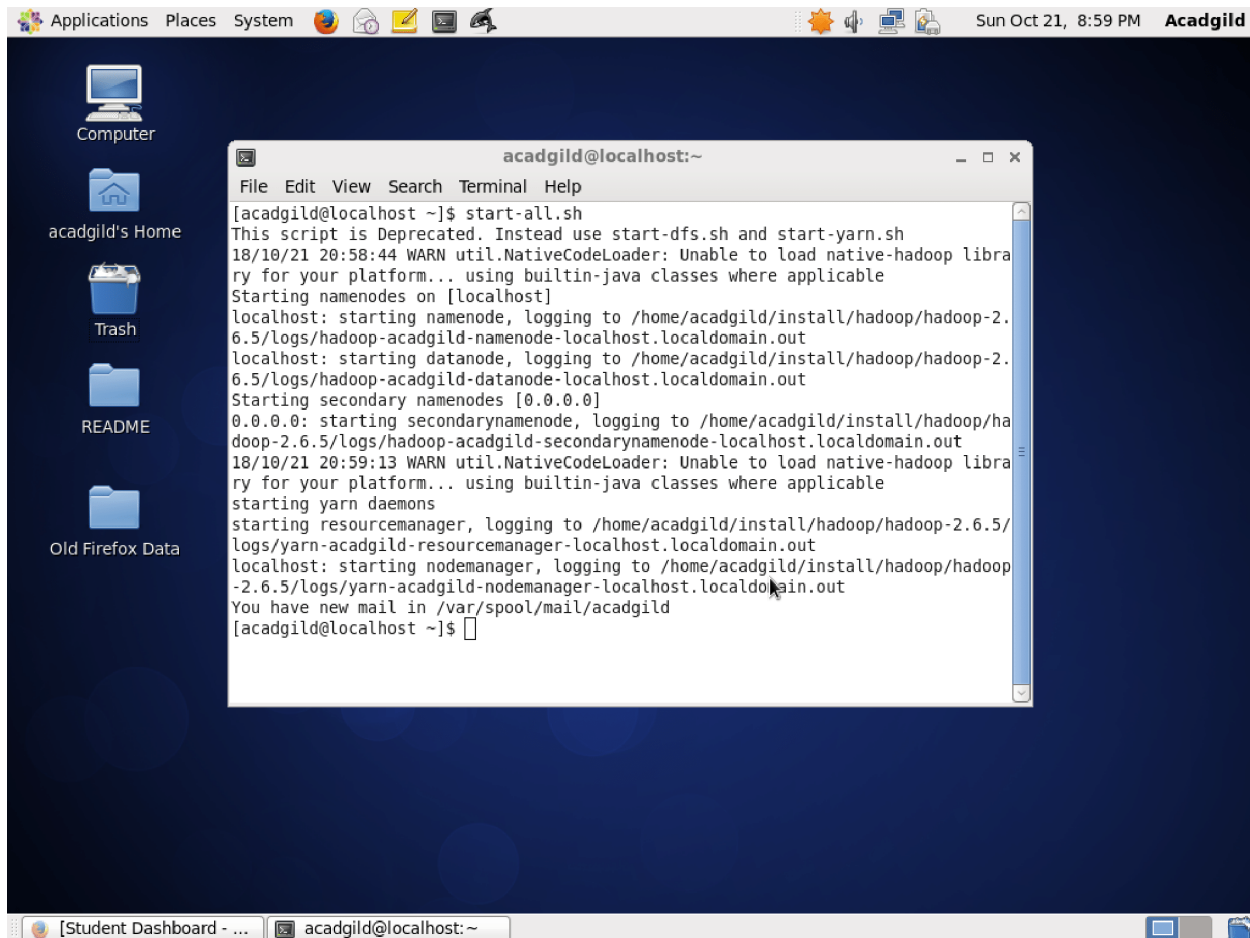


1) Start Hadoop single node on AcadGild VM. The command is start-all.sh.

Solution : when we run the command “start-all.sh” , this will startup a Namenode, Datanode, Jobtracker and a Tasktracker on your machine.

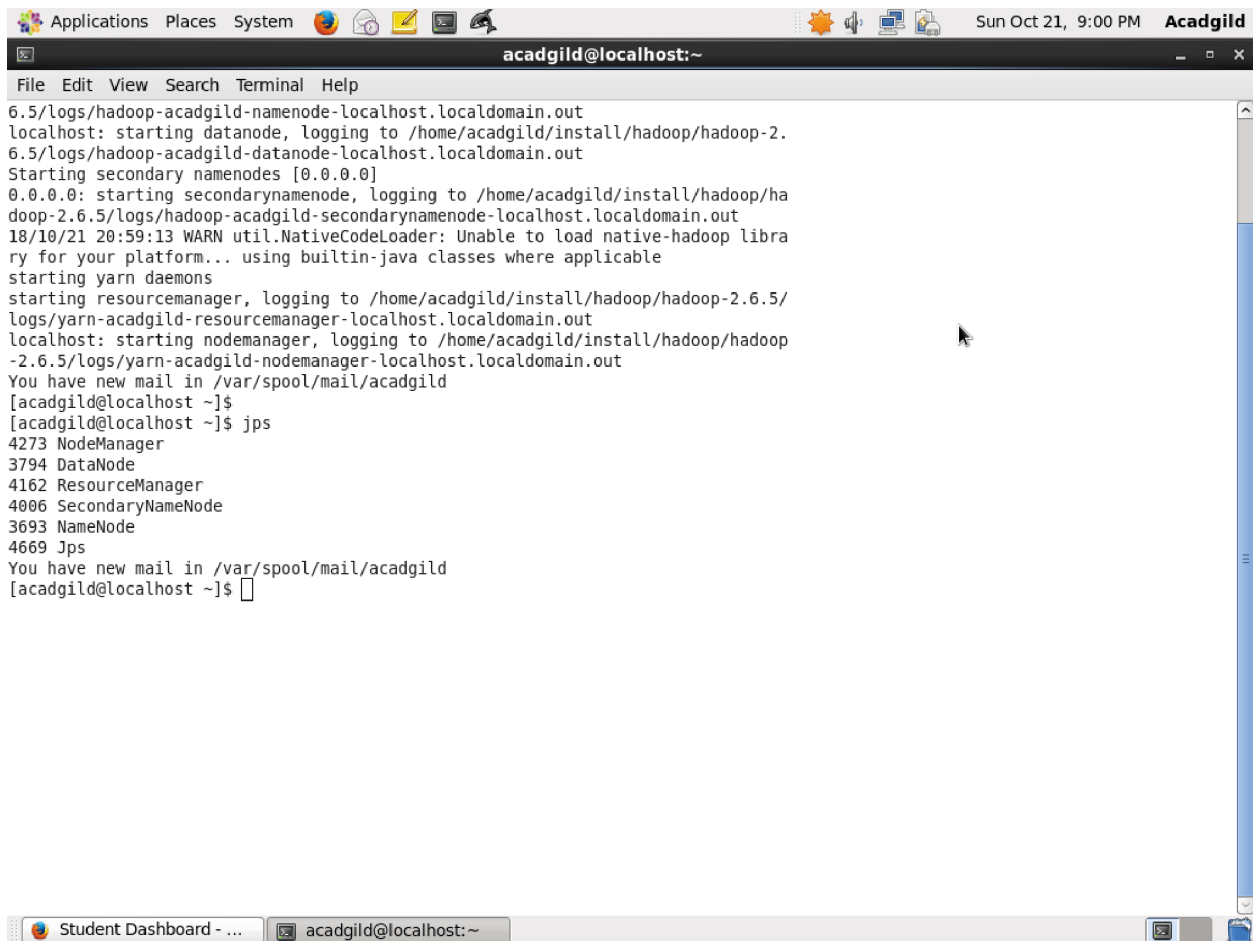


The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output is as follows:

```
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/10/21 20:58:44 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry 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/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.
6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/ha
doo-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/10/21 20:59:13 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry 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-acadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoo
p-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

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

Solution: JPS checks the java processes that are running. The **jps command** is found in the \$JAVA_HOME/bin directory.



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

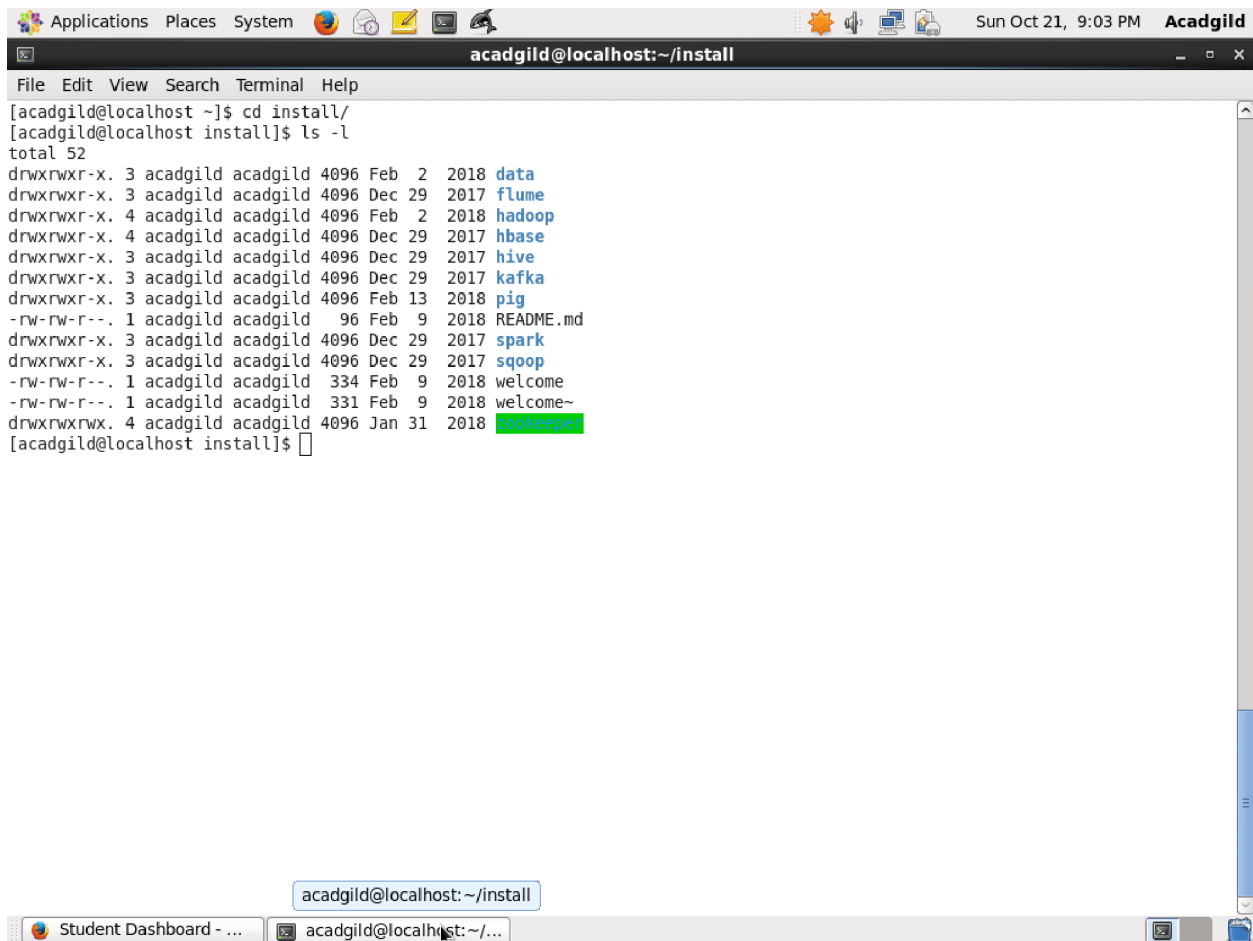
```
6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.
6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/10/21 20:59:13 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-acadgild-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 ~]$
[acadgild@localhost ~]$ jps
4273 NodeManager
3794 DataNode
4162 ResourceManager
4006 SecondaryNameNode
3693 NameNode
4669 Jps
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

The terminal window is part of a desktop environment with a taskbar at the top showing 'Applications', 'Places', 'System', and a clock for 'Sun Oct 21, 9:00 PM'. The bottom of the window shows a taskbar with 'Student Dashboard - ...' and 'acadgild@localhost:~'.

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

Solution : ls -l Lists the files and folders in the current directory.

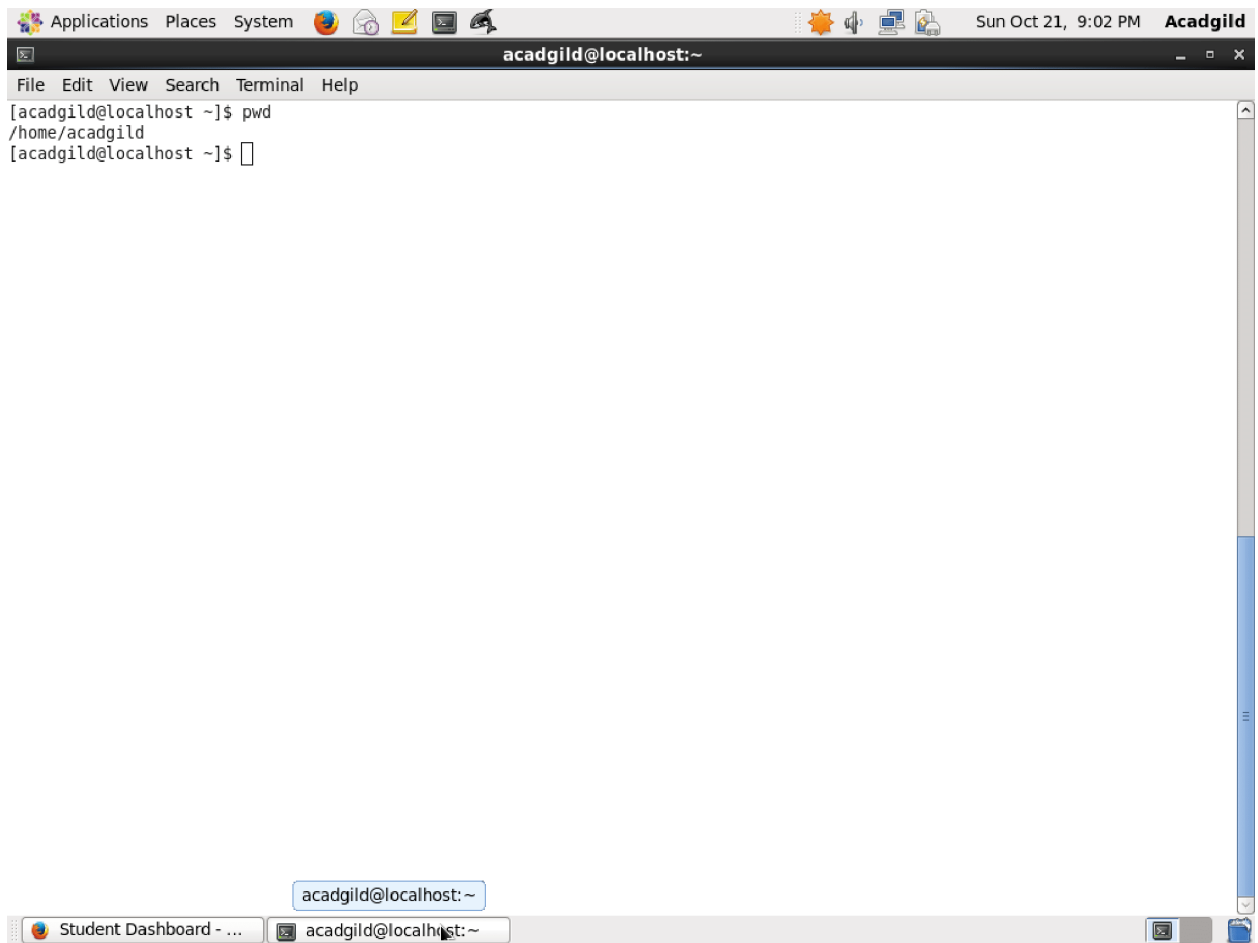
cd is used to change the directory.



The screenshot shows a Linux desktop environment with a terminal window open. The terminal title bar reads "acadgild@localhost: ~/install". The terminal content shows the user navigating to the "install" directory and listing its contents with the command "ls -l". The output lists various files and directories with their permissions, owner, group, size, date, and name. The files listed are: data, flume, hadoop, hbase, hive, kafka, pig, README.md, spark, sqoop, welcome, welcome~, and zookeeper. The terminal window is part of a taskbar at the bottom that includes a "Student Dashboard" icon and a "acadgild@localhost: ~/..." icon.

```
Applications Places System Sun Oct 21, 9:03 PM Acadgild
acadgild@localhost: ~/install
File Edit View Search Terminal Help
[acadgild@localhost ~]$ cd install/
[acadgild@localhost install]$ ls -l
total 52
drwxrwxr-x. 3 acadgild acadgild 4096 Feb  2  2018 data
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29  2017 flume
drwxrwxr-x. 4 acadgild acadgild 4096 Feb  2  2018 hadoop
drwxrwxr-x. 4 acadgild acadgild 4096 Dec 29  2017 hbase
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29  2017 hive
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29  2017 kafka
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29  2017 pig
-rw-rw-r--. 1 acadgild acadgild  96 Feb  9  2018 README.md
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29  2017 spark
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29  2017 sqoop
-rw-rw-r--. 1 acadgild acadgild 334 Feb  9  2018 welcome
-rw-rw-r--. 1 acadgild acadgild 331 Feb  9  2018 welcome~
drwxrwxrwx. 4 acadgild acadgild 4096 Jan 31  2018 zookeeper
[acadgild@localhost install]$
```

Next we type `pwd` to find out where you are within your computer. This command means “print working directory” and will tell you the exact working directory you are currently in.



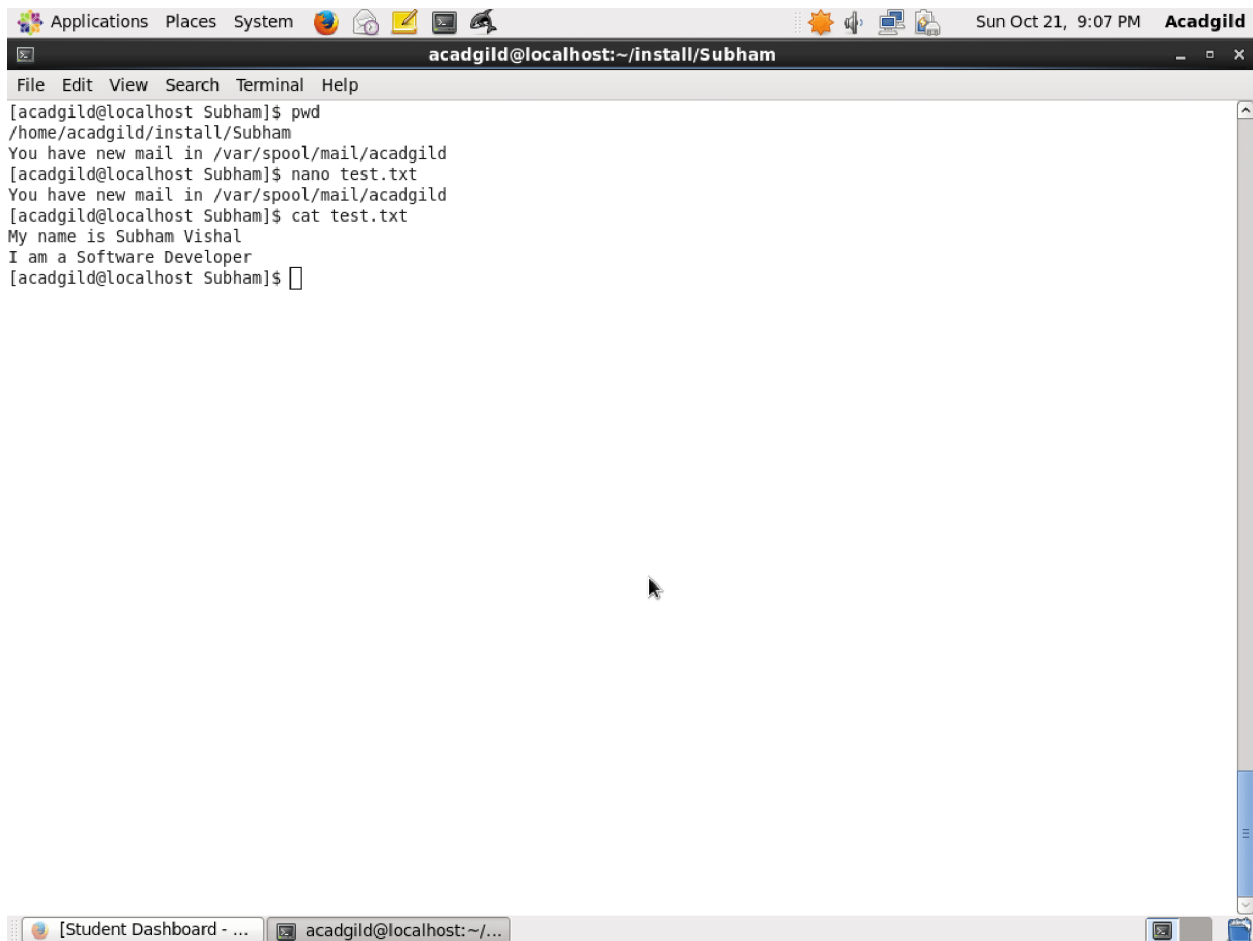
pwdx command shows current working directory of a process.

The screenshot shows a terminal window titled 'acadgild@localhost:~'. The terminal output displays the logs for starting Hadoop services. It includes messages for starting the datanode, namenodes, secondary namenode, resource manager, and nodemanager. A warning message from 'util.NativeCodeLoader' is also visible. The user then runs 'jps' to list the running processes, which shows 'NodeManager', 'DataNode', 'ResourceManager', 'SecondaryNameNode', 'NameNode', and 'Jps'. The user also runs 'pwdx' for each process ID to show their current working directories. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The system tray at the bottom shows 'Student Dashboard' and the terminal window itself.

```
6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.
6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/ha
dooop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/10/21 20:59:13 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resource manager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/
logs/yarn-acadgild-resource manager-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 ~]$
[acadgild@localhost ~]$ jps
4273 NodeManager
3794 DataNode
4162 ResourceManager
4006 SecondaryNameNode
3693 NameNode
4669 Jps
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ pwdx 4273
4273: /home/acadgild/install/hadoop/hadoop-2.6.5
[acadgild@localhost ~]$ pwdx 3794
3794: /home/acadgild/install/hadoop/hadoop-2.6.5
[acadgild@localhost ~]$ pwdx 4162
4162: /home/acadgild/install/hadoop/hadoop-2.6.5
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ pwdx 4006
4006: /home/acadgild/install/hadoop/hadoop-2.6.5
[acadgild@localhost ~]$ pwdx 3693
3693: /home/acadgild/install/hadoop/hadoop-2.6.5
[acadgild@localhost ~]$
```

4) 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.

Solution :



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

Solution : If all our hadoop services are up and running, then if we hit localhost:50070 in our web browser, we will see the page below. By default the UI port of HDFS cluster is 50070.

ApplicationsPlacesSystem

Namenode information - Mozilla Firefox

Student Dashboard x Namenode information x +

localhost:50070/dfshealth.html#tab-overview

Search

☆📁⬇️🏠🔍☰

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Overview 'localhost:8020' (active)

Started:	Sun Oct 21 20:58:52 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.

160 files and directories, 70 blocks = 230 total filesystem object(s).

Heap Memory used 31.78 MB of 49.18 MB Heap Memory. Max Heap Memory is 966.69 MB.

Here we can see the different attributes of the hdfs cluster, memory used, number of live nodes, dead nodes etc.

Applications Places System Sun Oct 21, 9:09 PM Acadgild

Namenode information - Mozilla Firefox

Student Dashboard x Namenode information x +

localhost:50070/dfshealth.html#tab-overview Search

non heap memory used 37.99 MB of 58.94 MB Committed non heap memory: max non heap memory is 1 b.

Configured Capacity:	17.11 GB
DFS Used:	22.81 MB
Non DFS Used:	11.42 GB
DFS Remaining:	5.67 GB
DFS Used%:	0.13%
DFS Remaining%:	33.13%
Block Pool Used:	22.81 MB
Block Pool Used%:	0.13%
DataNodes usages% (Min/Median/Max/stdDev):	0.13% / 0.13% / 0.13% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	16
Number of Blocks Pending Deletion	0
Block Deletion Start Time	10/21/2018, 8:58:52 PM

Namenode informatio... acadgild@localhost: ~/...

If we click on Datanodes in the header, we can see the information about the data nodes.

Namenode information - Mozilla Firefox

Student Dashboard x Namenode information x +

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)	0	In Service	17.11 GB	22.82 MB	11.42 GB	5.67 GB	68	22.82 MB (0.13%)	0	2.6.5

Decomissioning

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

Namenode informatio... acadgild@localhost: ~/...

We can browse through the directories of hdfs also.

Path does not exist on HDFS or WebHDFS is disabled. Please check your path or enable WebHDFS

/ Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	hbase
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	home
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	sqoopout111
drwxrwx---	acadgild	supergroup	0 B	0	0 B	tmp
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	user
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	usr