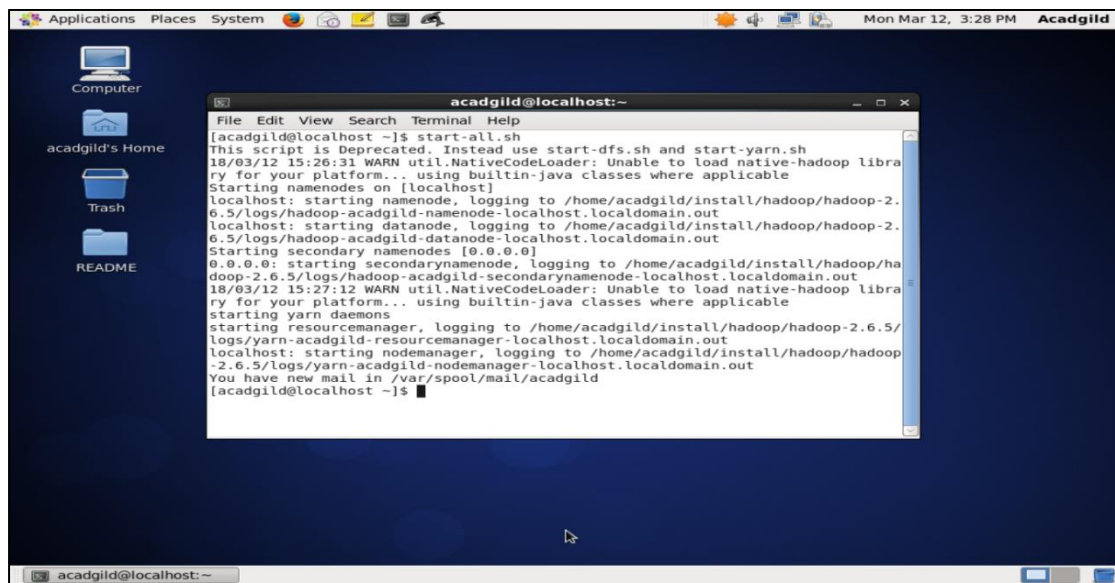


Session 1 - Bigdata and Hadoop - Introduction

Assignment 1

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

Start-all.sh command starts all Hadoop daemons, the namenode, datanodes, the jobtracker and tasktrackers.

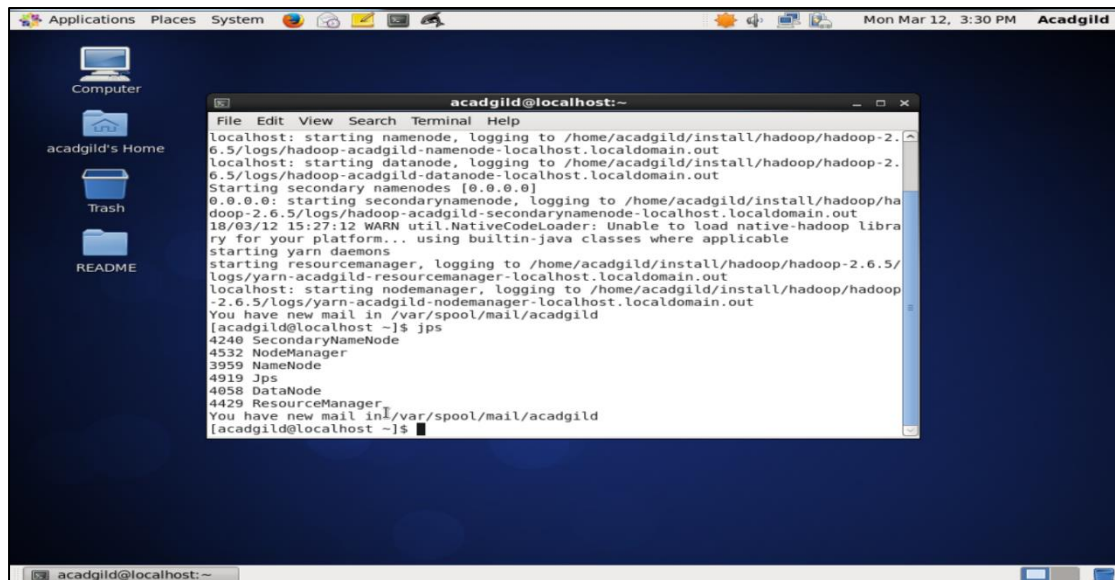


The screenshot shows a terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal output shows the execution of the 'start-all.sh' script. The script is deprecated and suggests using 'start-dfs.sh' and 'start-yarn.sh'. It logs the starting of the namenode, datanode, and secondary namenodes, along with the starting of yarn daemons (resourcemanager, nodemanager). The output also includes warnings about the NativeCodeLoader and a message about new mail in /var/spool/mail/acadgild.

```
acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/03/12 15:26:31 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
18/03/12 15:27:12 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/hadoop
-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
You have new mail in /var/spool/mail/acadgild
acadgild@localhost ~]$
```

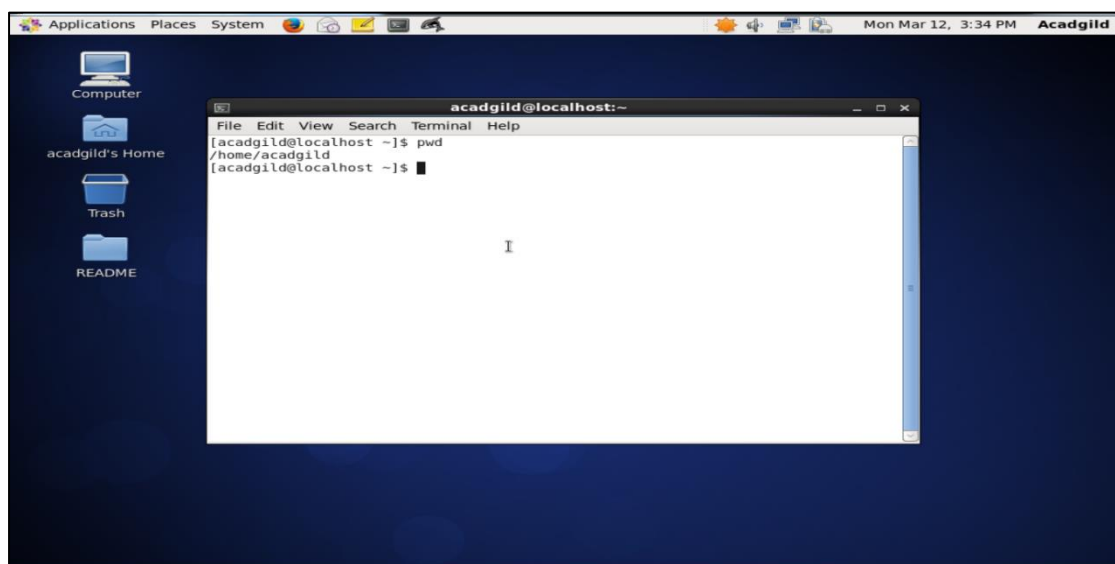
Task 2 : Run a JPS command to see if all Hadoop daemons are running.

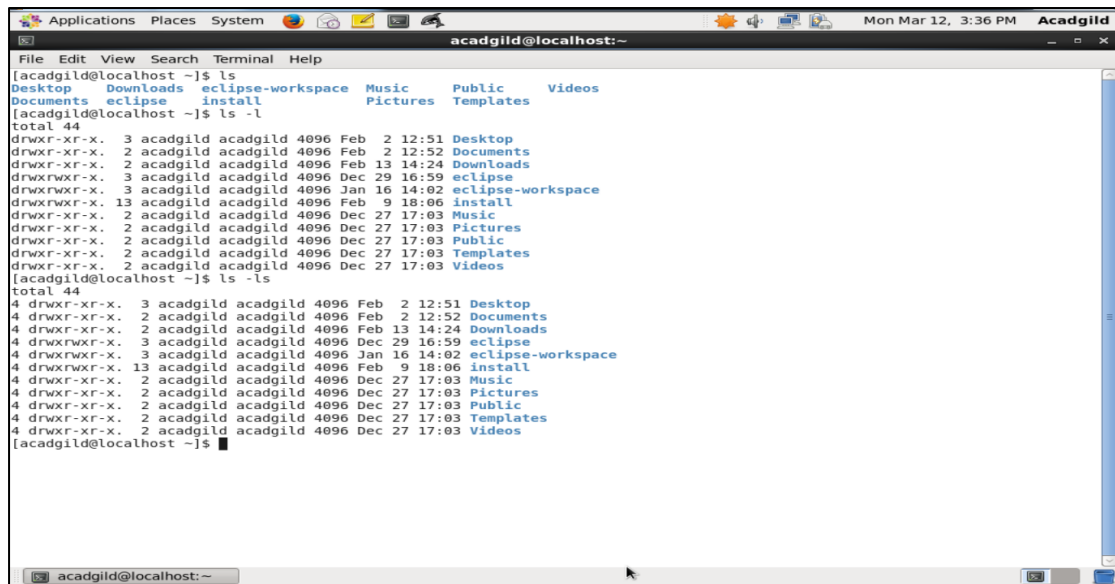
jps command is used to check all the Hadoop daemons like NameNode, DataNode, ResourceManager, NodeManager etc. which are running on the machine.



Task 3 : Run few Unix commands like `pwd`, `ls -ls` , etc.

- `pwd` : shows the path to the present working directory.
- `ls` : lists all the files and folders in the present directory.
- `ls -l` : lists the files and folders in the present directory along with the details of permissions for each file/folder.
- `ls -ls` : lists the files and folders in the present directory along with the details of file size.



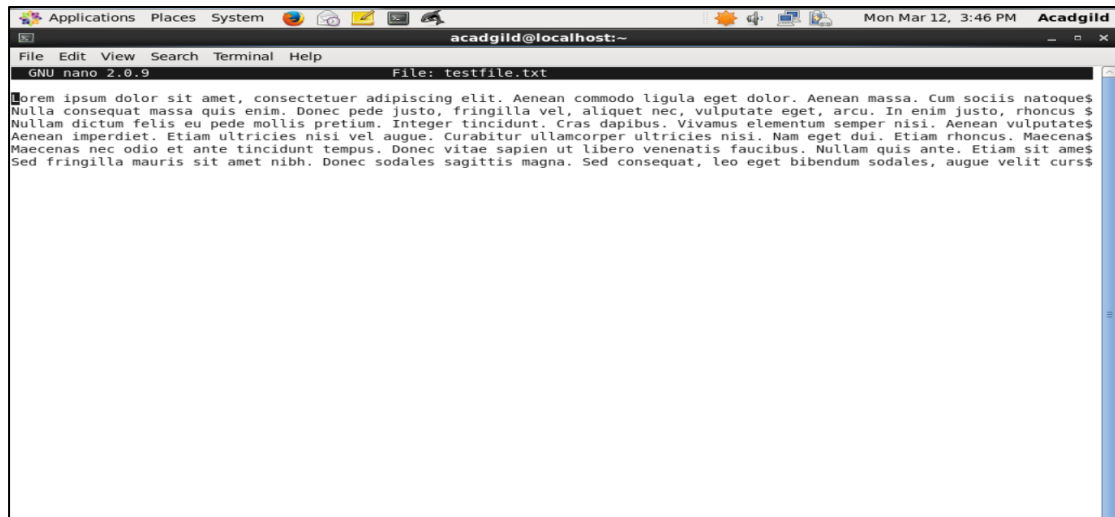
A terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The user has executed 'ls' and 'ls -l' commands. The output shows a list of files and directories with permissions, owner, group, size, date, and name. The files listed are Desktop, Documents, Downloads, eclipse, eclipse-workspace, install, Music, Pictures, Public, Templates, and Videos. The terminal shows the output of 'ls' and 'ls -l' commands.

```
[acadgild@localhost ~]$ ls
Desktop  Downloads  eclipse-workspace  Music      Public      Videos
Documents eclipse     install           Pictures   Templates
[acadgild@localhost ~]$ ls -l
total 44
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 -ls
total 44
4 drwxr-xr-x. 3 acadgild acadgild 4096 Feb  2 12:51 Desktop
4 drwxr-xr-x. 2 acadgild acadgild 4096 Feb  2 12:52 Documents
4 drwxr-xr-x. 2 acadgild acadgild 4096 Feb 13 14:24 Downloads
4 drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29 16:59 eclipse
4 drwxrwxr-x. 3 acadgild acadgild 4096 Jan 16 14:02 eclipse-workspace
4 drwxrwxr-x. 13 acadgild acadgild 4096 Feb  9 18:06 install
4 drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Music
4 drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Pictures
4 drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Public
4 drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Templates
4 drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Videos
[acadgild@localhost ~]$
```

Task 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.

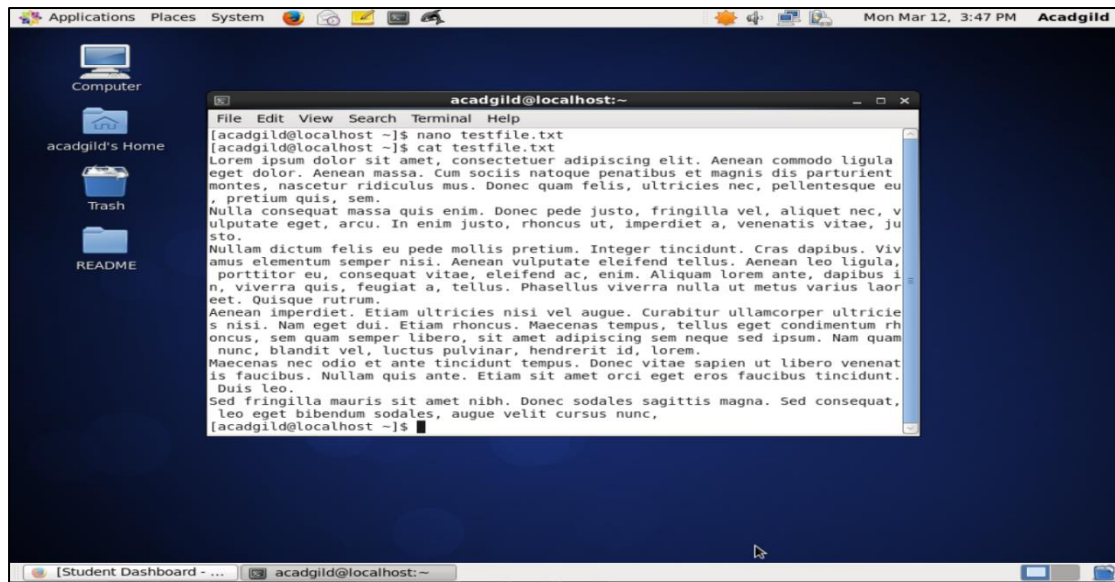
cat command to output the contents of file1 to the display.

Nano command :

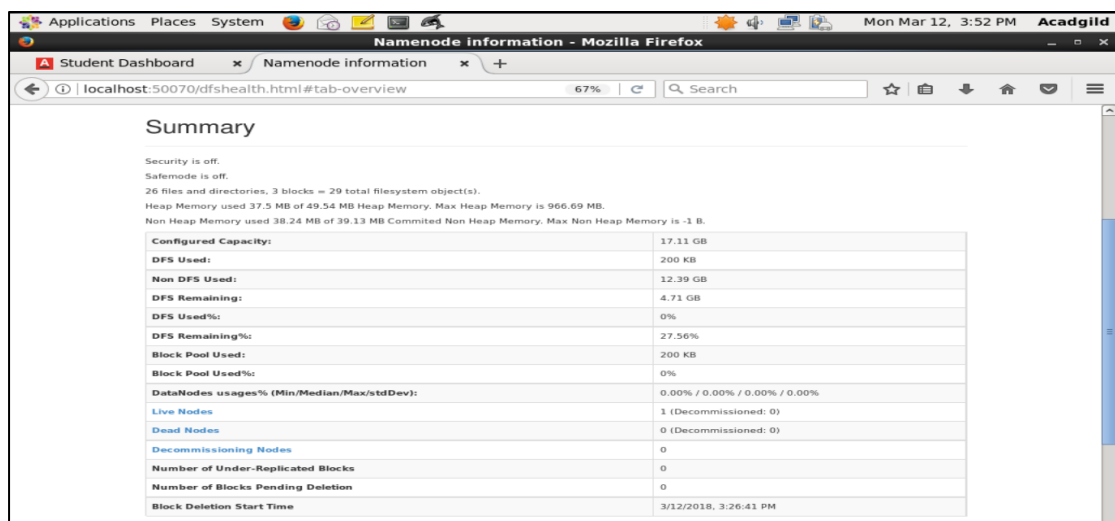
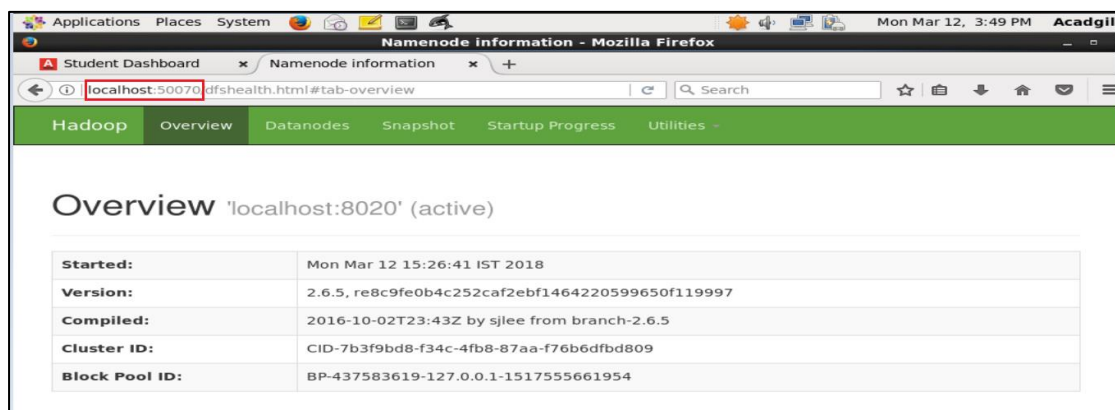
A terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The user has opened the nano editor with 'nano testfile.txt'. The editor shows a file named 'testfile.txt' with the following content: 'Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque...'. The nano editor interface includes a status bar at the bottom showing 'GNU nano 2.0.9' and 'File: testfile.txt'.

```
GNU nano 2.0.9 File: testfile.txt
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque$
Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus $
Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate$
Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas$
Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit ame$
Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit curs$
```

Cat command :



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



The web UI of Hadoop has the following information :

1. Summary of the HDFS file system
2. Name Node status and storage information
3. Data Nodes information
4. Snap shot summary

- We can browse file system from Utilities menu on screen.
- We can browse through each directory by clicking on each directory name.
- For each file, owner, group, size, replication factor, block size are listed in the table.
- We can have a look at the block information of each and download the files by clicking on each file.