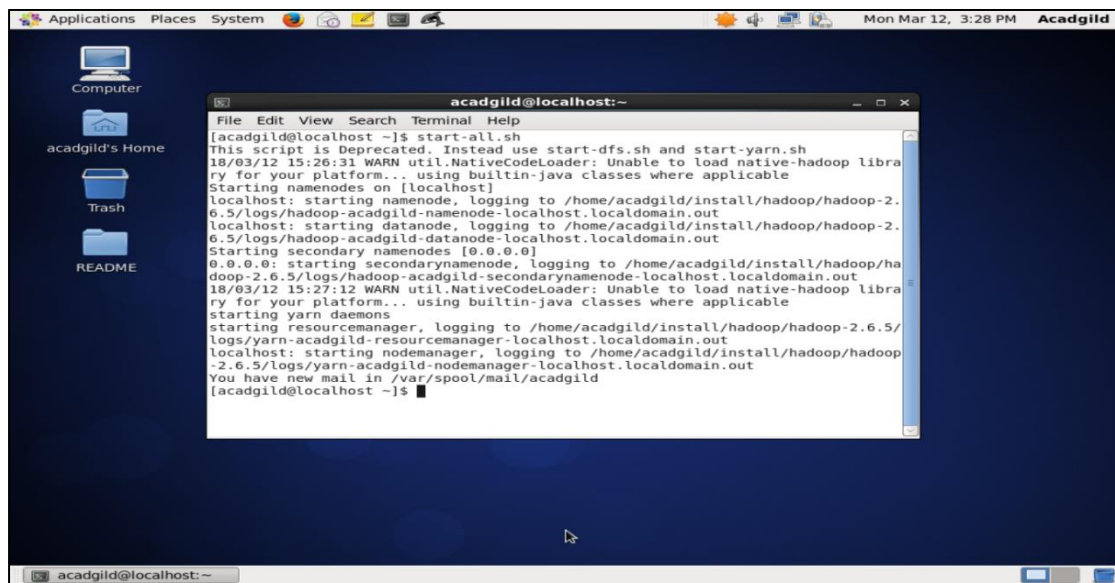# Session 1 - Bigdata and Hadoop - Introduction

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.
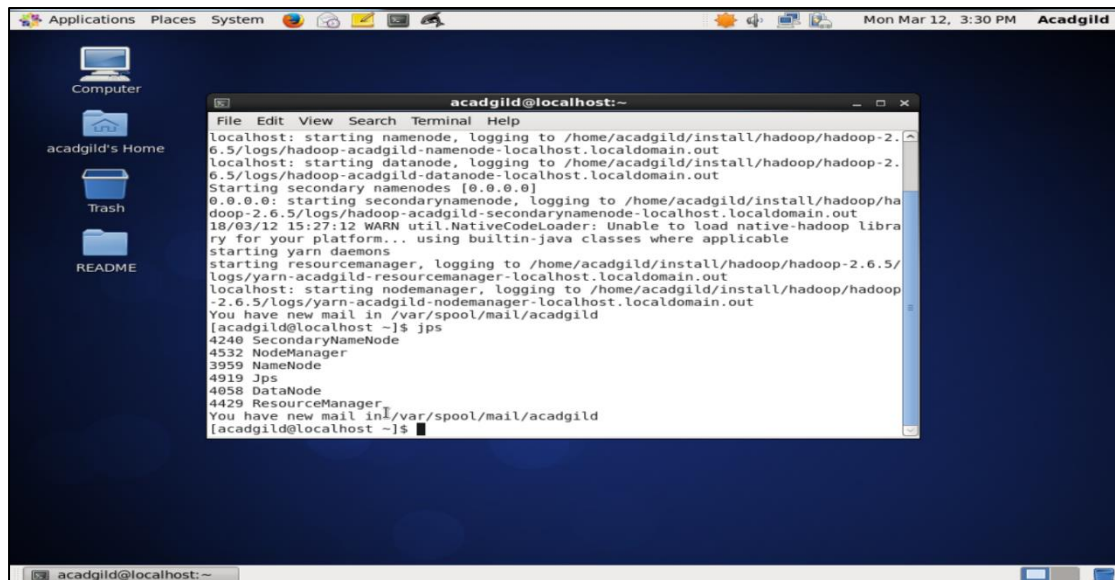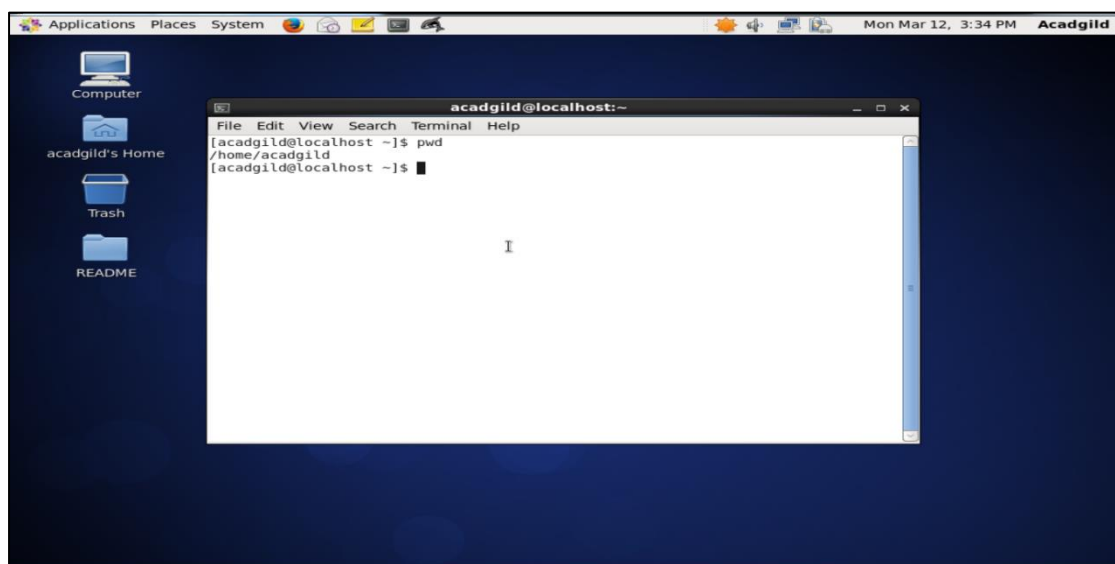

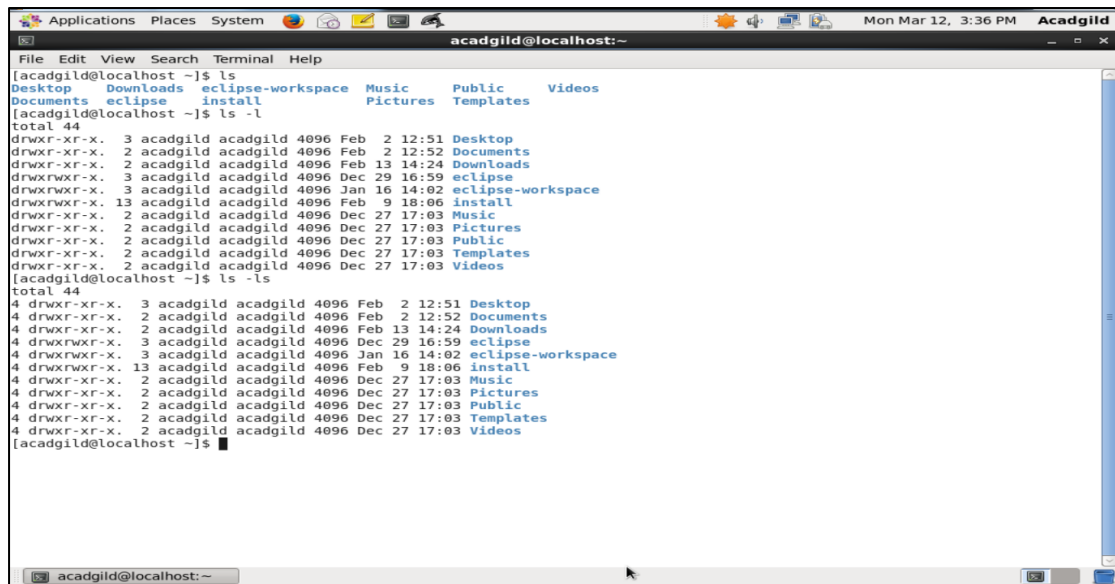**Task 2 :** Run a JPS command to see if all Hadoop daemons are running.

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.

- <u>pwd :</u> shows the path to the present working directory.

- <u>ls :</u>   lists all the files and folders in the present directory.

- <u>ls -l :</u> lists the files and folders in the present directory along with the details of permissions for each file/folder.

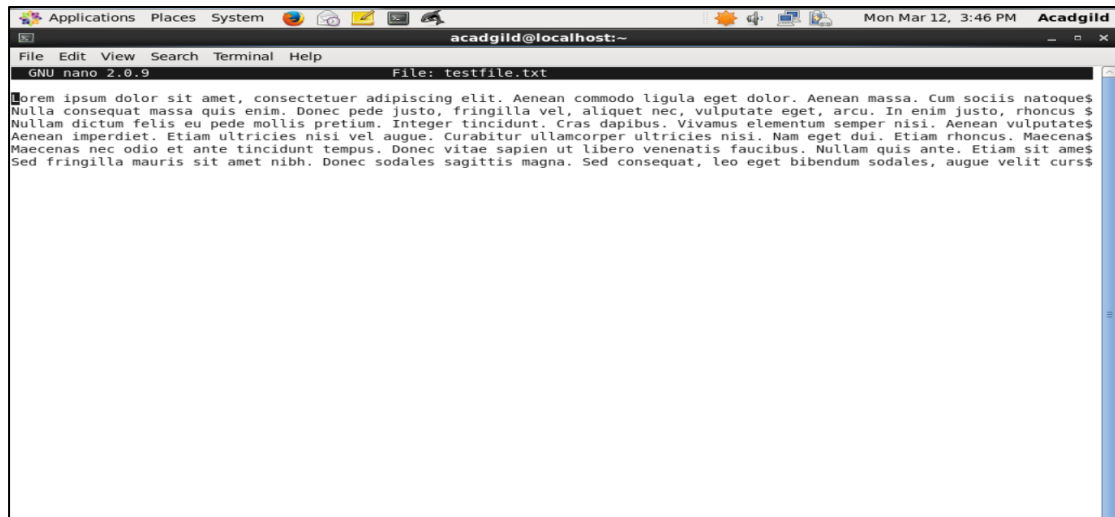- <u>ls -ls :</u> ists the files and folders in the present directory along with the details of file size.

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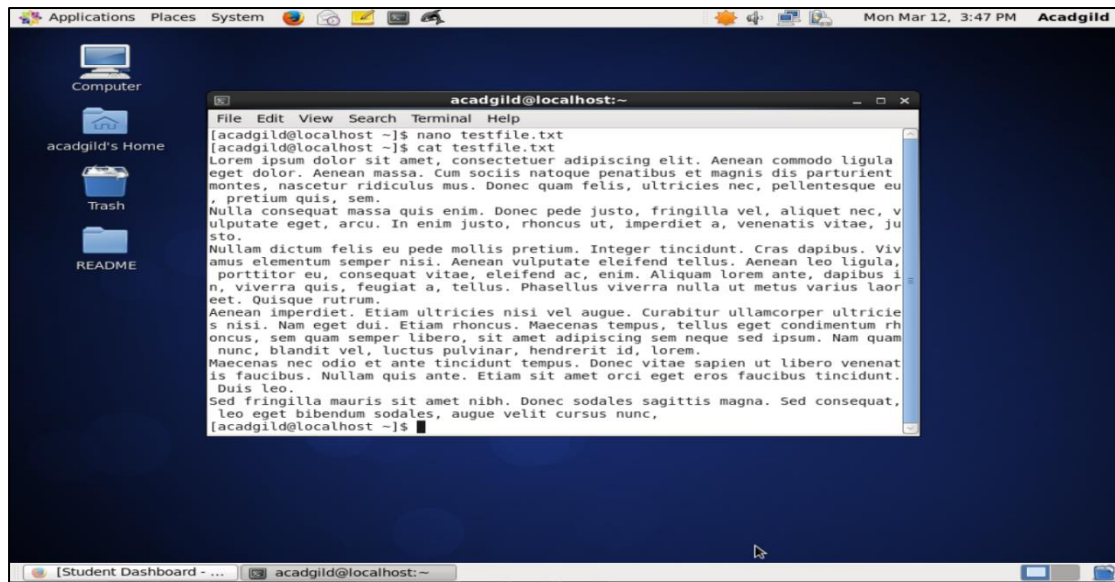
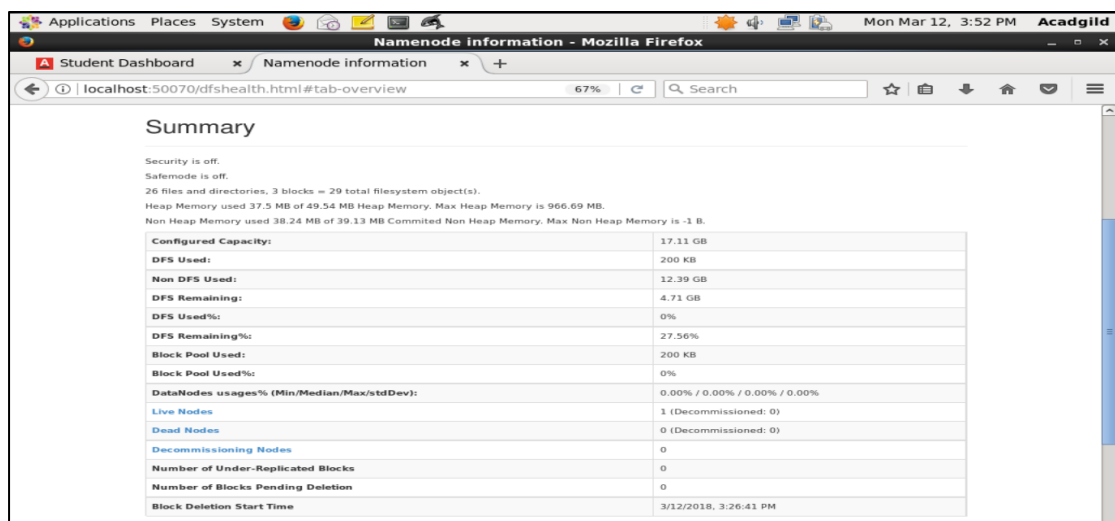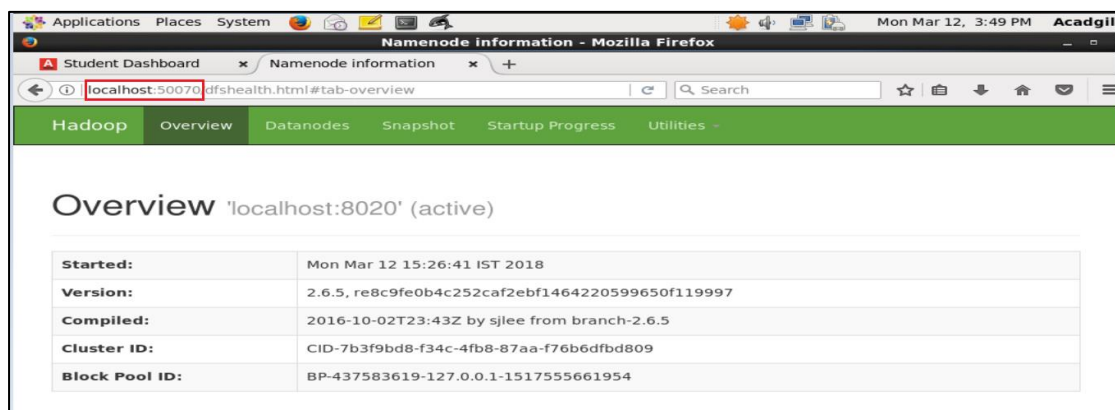cat command to output the contents of file1 to the display.

Nano command :

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.