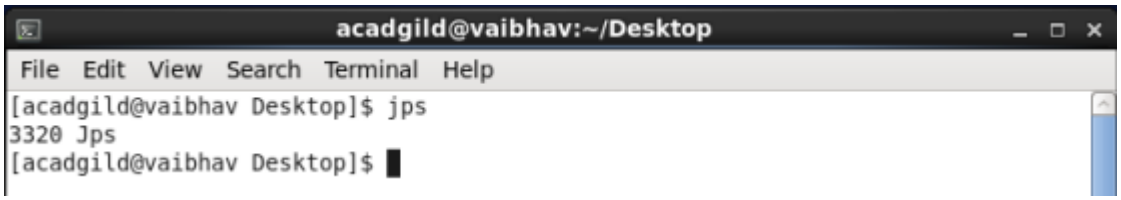
Start Hadoop single node on AcadGild VM. The command is *"start-all.sh."*

Command "start-all.sh":

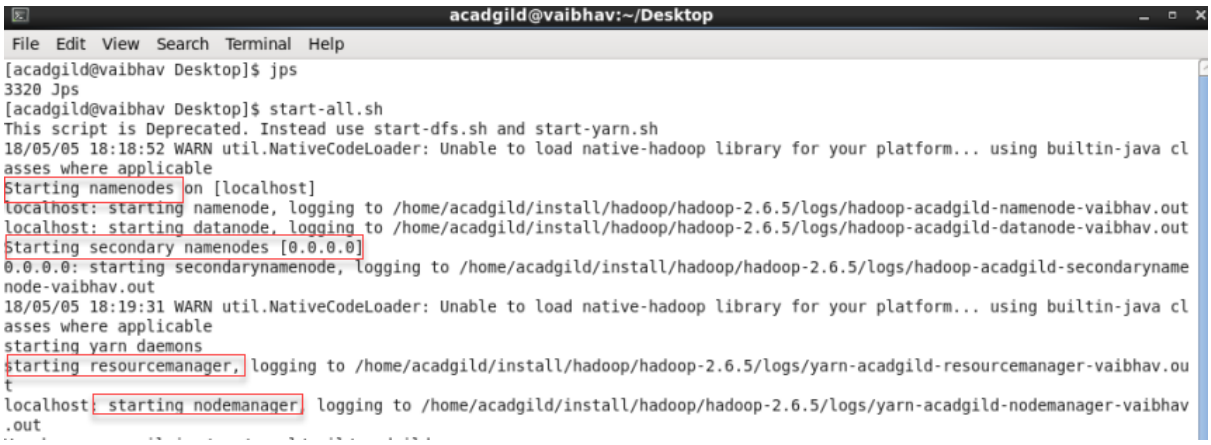
- The command start-all.sh will start all the hadoop daemons (back ground processes) which includes five daemon processes name nodes, Job tracker, task tracker and data nodes.
- "Jps" (Java Virtual Machine Process Status) command is used to list out the daemons running on the current machine.

Before executing start-all.sh



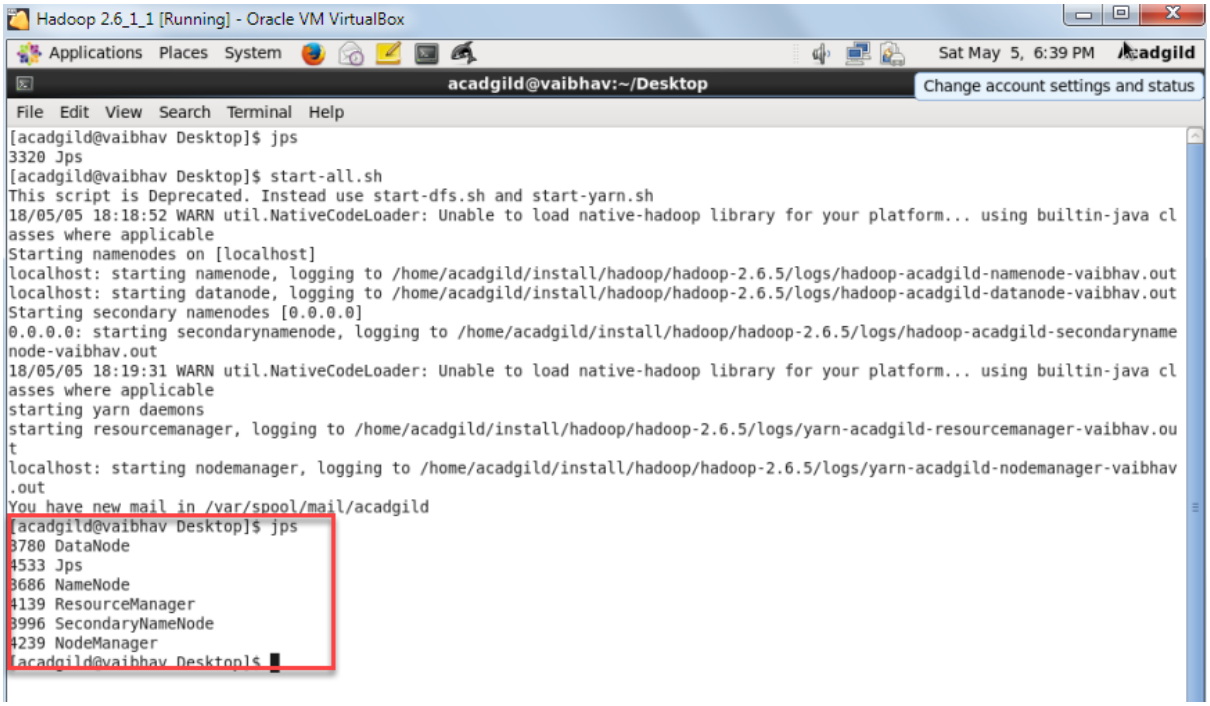
After executing start-all.sh

- This command starts total five daemons in the machine that is Name Node, Secondary name node, Node Manager, Source Manager, Data Node.



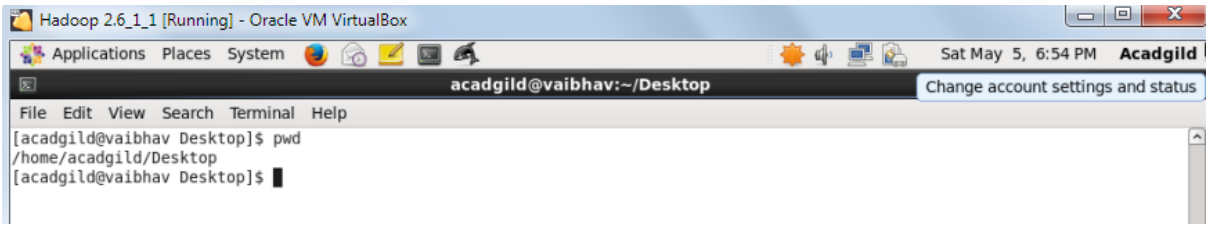
Run a JPS command to see if all Hadoop daemons are running.

- After executing the start-all.sh when JPS command was executed it shows following results:
- It is observed that 5 new daemons are running now in the background including JPS.
- JPS command provides the details of the block location and the name of the daemon running in the background Eg. 3780 Data node where 3780 is the block location post that is the name of the daemon running i.e. Data-node in this case.

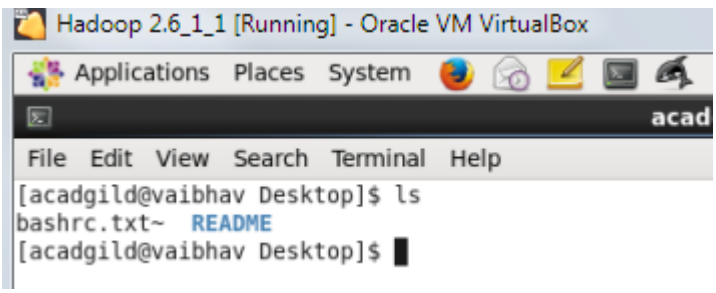


Run few UNIX commands like *pwd, ls -ls,* etc.

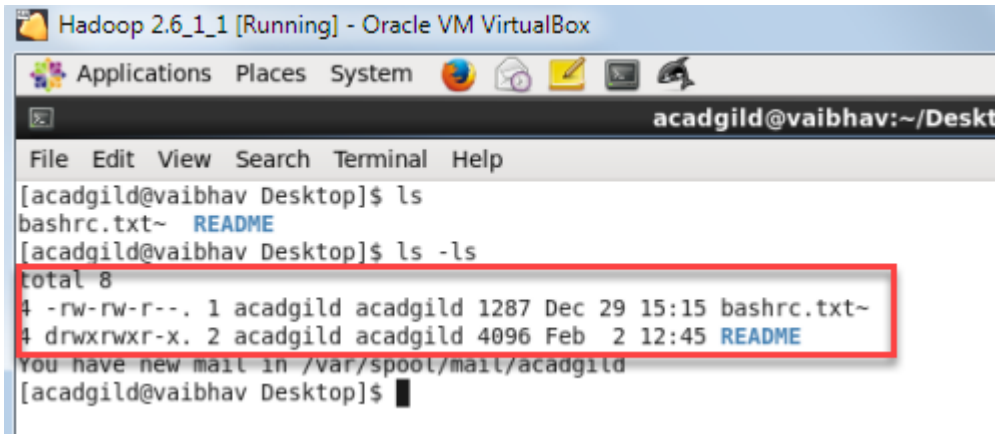
i) **pwd:** Provides the location path of the current directory.



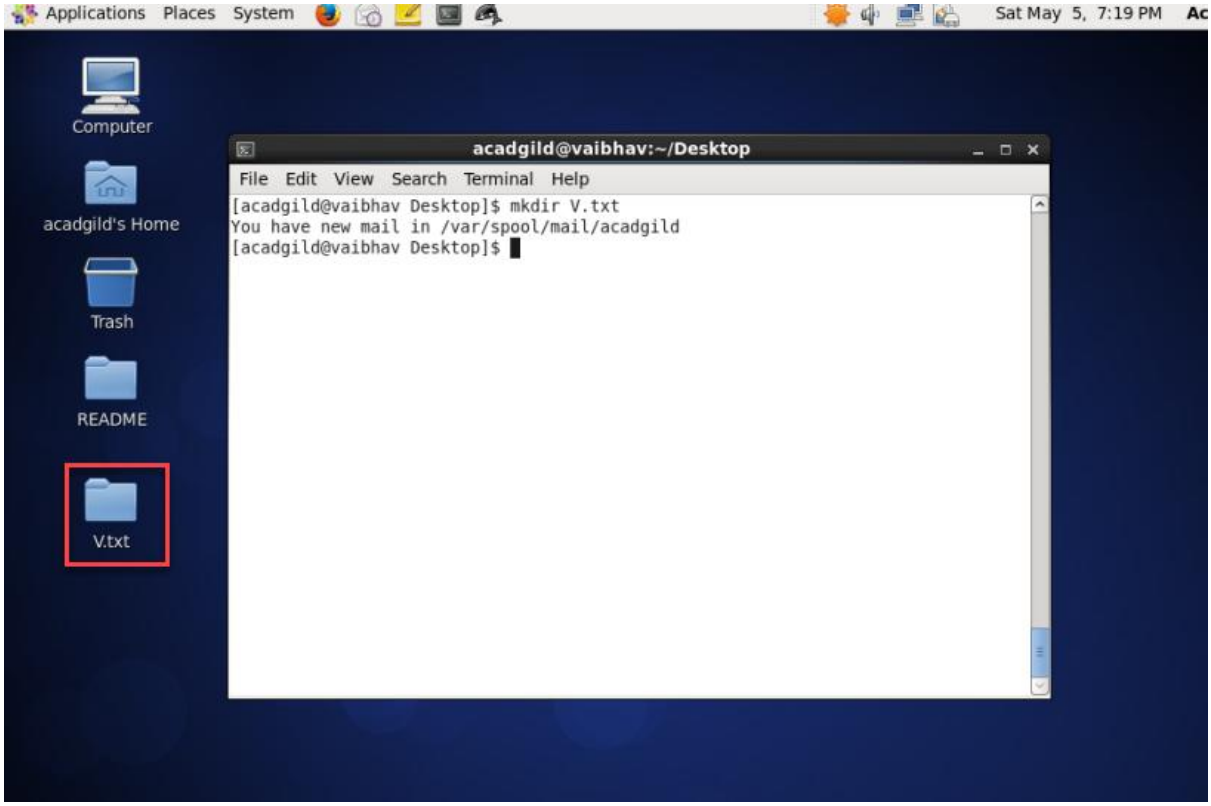
ii) **ls:** Lists out all directories present in the current location



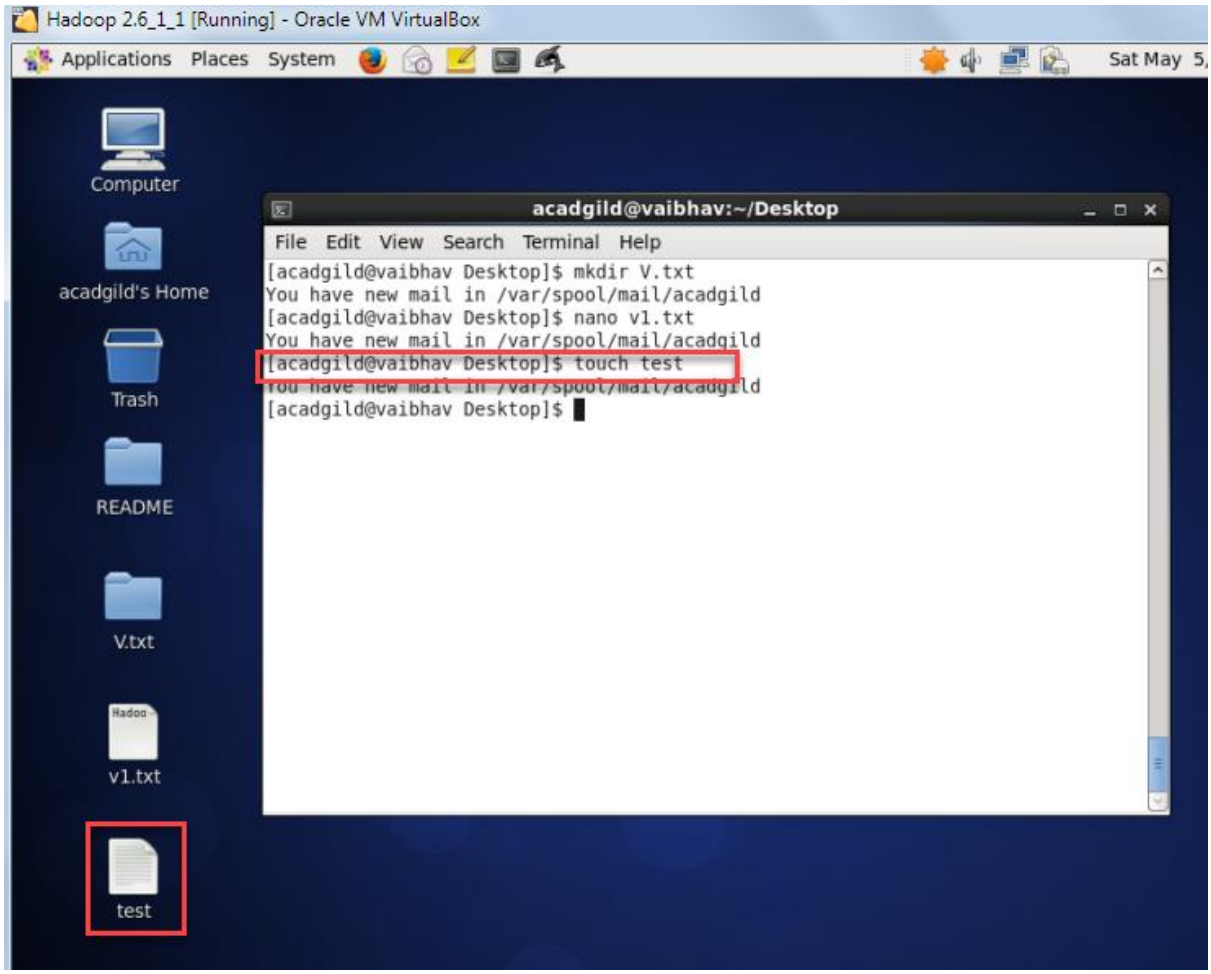
iii) **ls –ls :** used to list out directories with full details about file size , file permissions , block location, date created and time and the format of the file respectively.



iv) **mkdir:** creates a directory/Folder in current location.



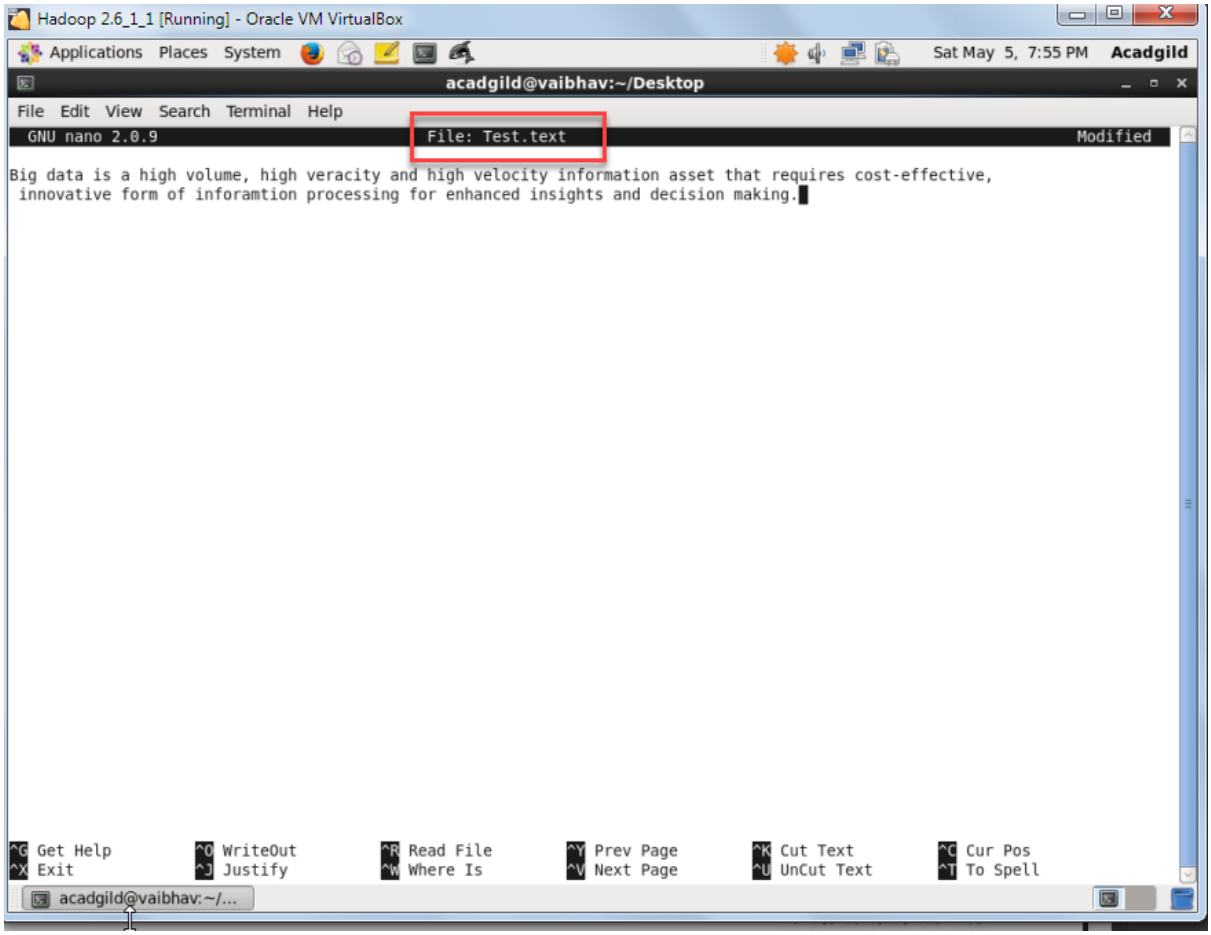
v) **touch:** The command is used create empty file in the current location. Eg. Here the “test” named file is created in the current location.



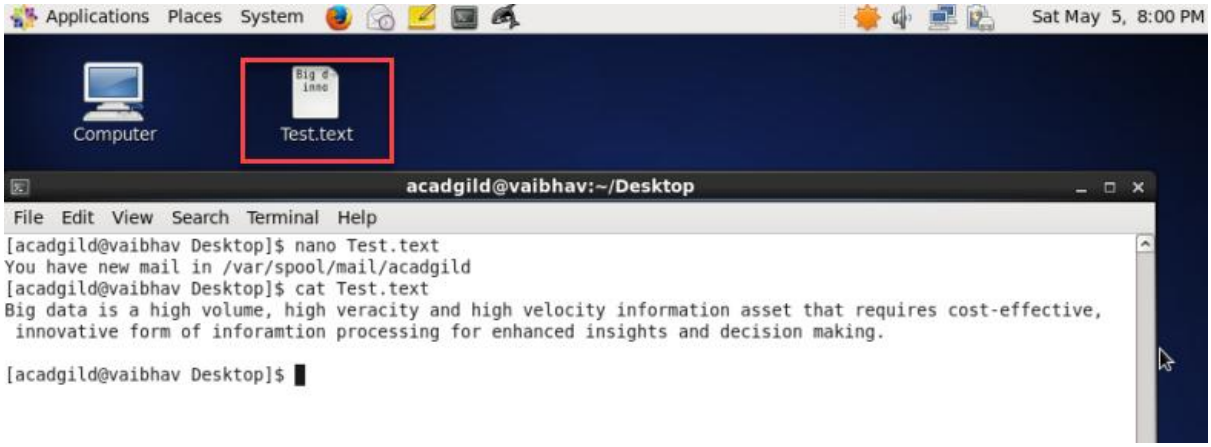
Create a file from the terminal using nano editor (example: nano test.txt), and add some content in it. Cat it to see if the content is saved.

Nano Editor: Nano editor is text editor used in linux operating systems through command line interface.

- Editing a file “Test.txt” through nano-editor adding the definition of big data in file.



- File saved in the current location. And cat command is used to read the content added to the file which is the same as added through the nano editor hence the file is saved in the current directory successfully.

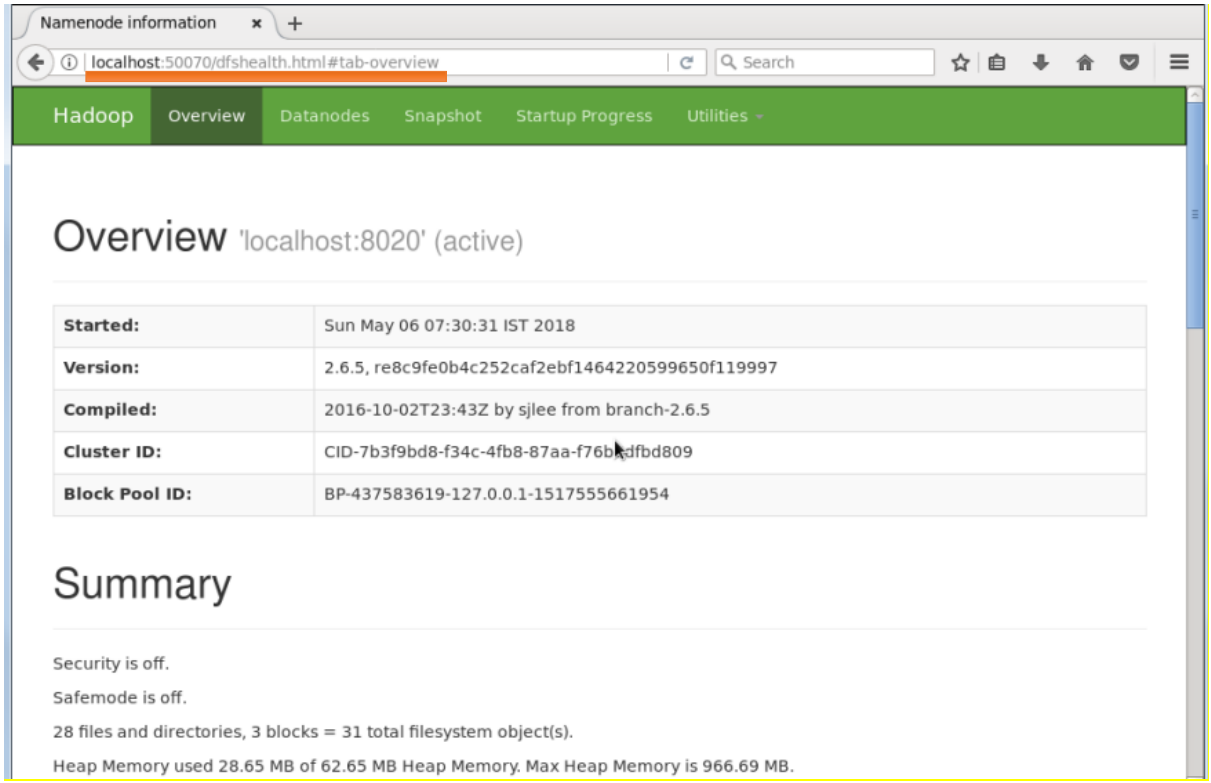


Open the hdfs web page by typing *localhost: 50070* in the browser. Check all the details of the HDFS.

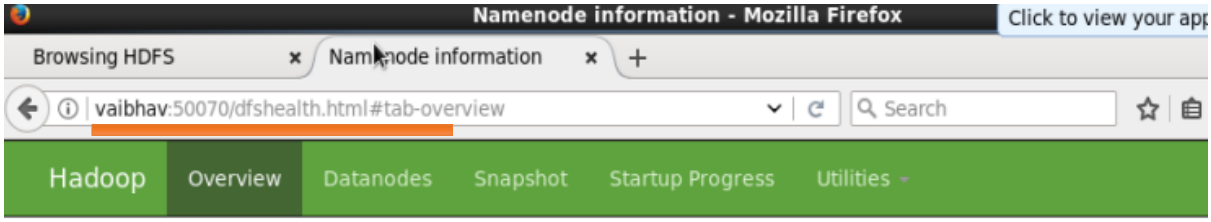
Apart from the command line interface. Hadoop provides Web UI interface to HDFS resource manager.

- Enter URL "Localhost: 50070" in web browser to point to the port 50070 in pseudo distributed mode. For fully distributed mode, "localhost" can be replaced by the actual host name of machine in cluster.

Pseudo distributed mode:



Fully Distributed Mode:



Overview 'localhost:8020' (active)

Started:	Sun May 06 07:30:31 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

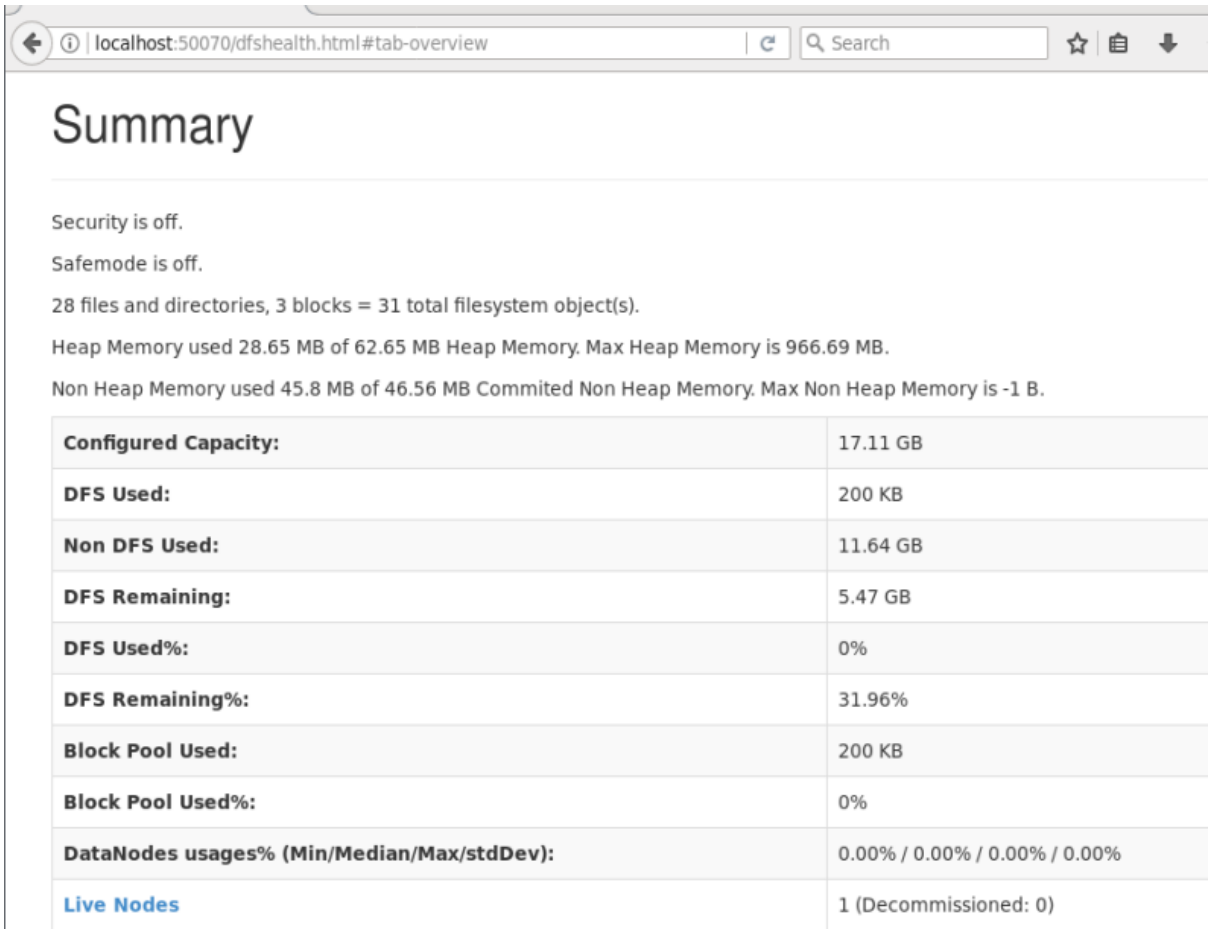
Security is off.

Safemode is off.

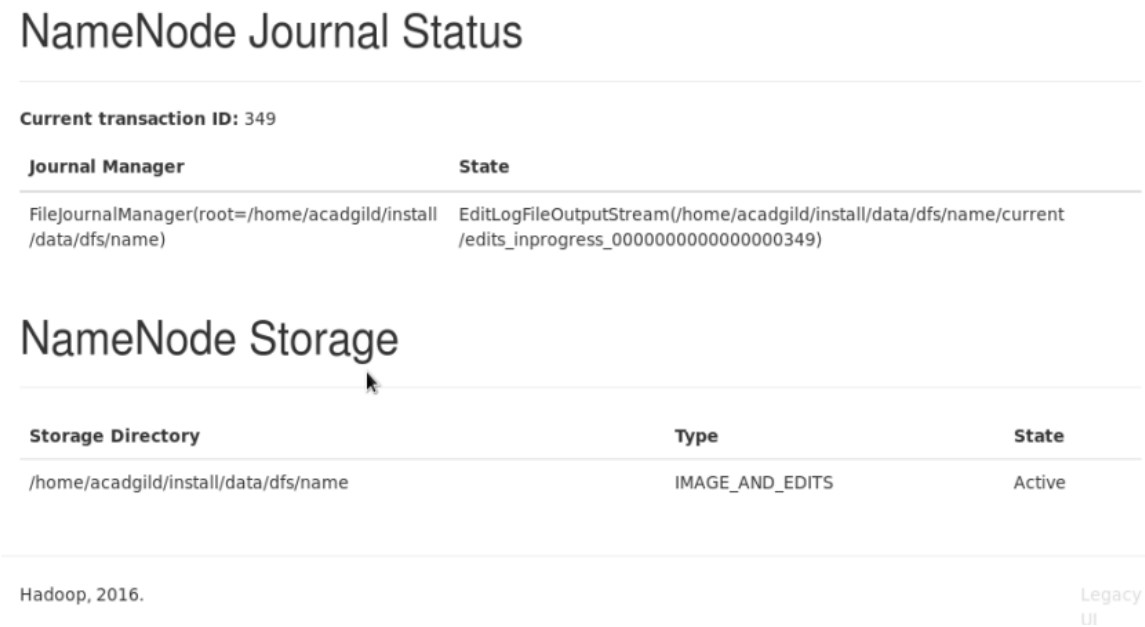
28 files and directories, 3 blocks = 31 total filesystem object(s).

Heap Memory used 28.35 MB of 62.65 MB Heap Memory. Max Heap Memory is 966.69 MB.

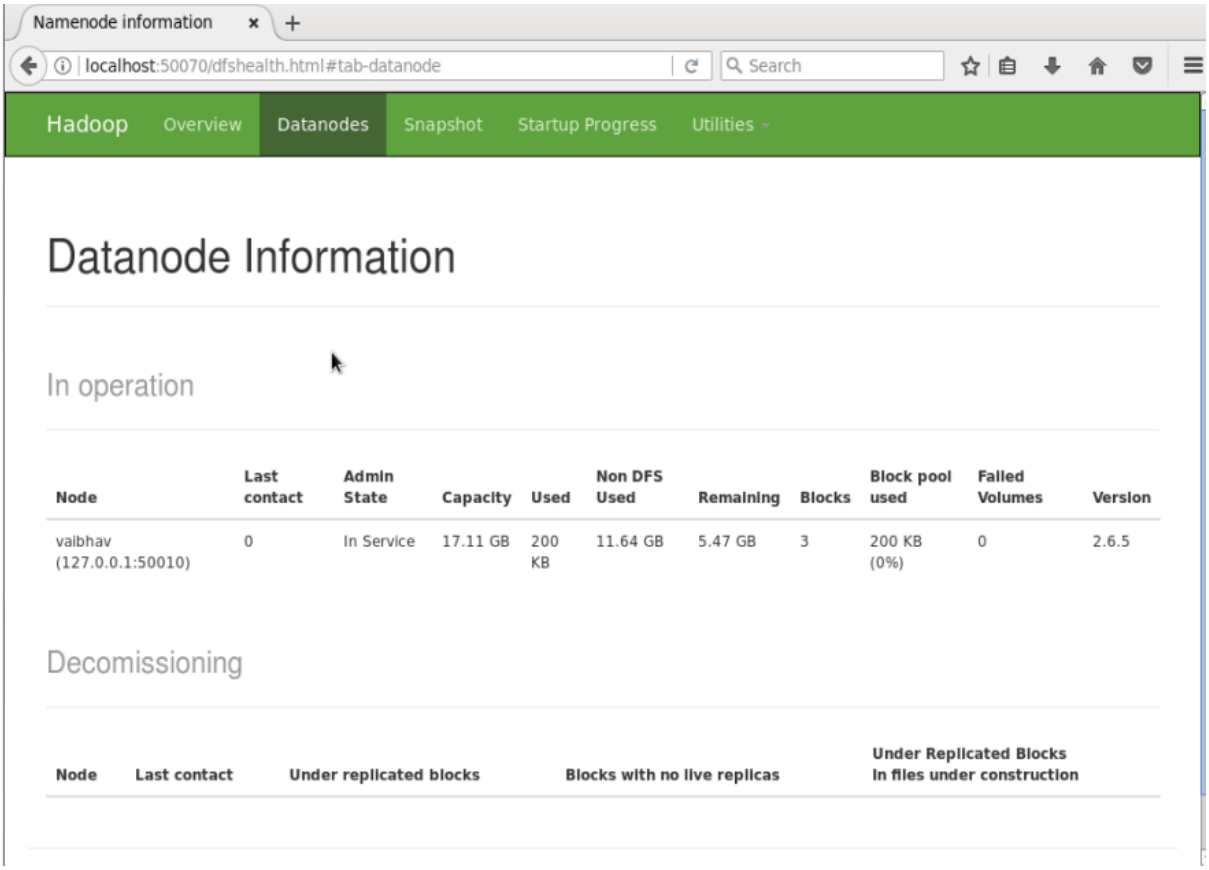
- Summary of the HDFS file system of the current machine is listed in overview tab as shown below:



- Namenode information and storage status is also listed.



- Datanode information is provided in the datanode tab which inlcudes the name of the node capacity and ,blocks details etc.



- We can also browse through the directories in the hadoop ecosystem through the browse directory tab. The directory names are listed in a tabular form which includes details of permission, owner.