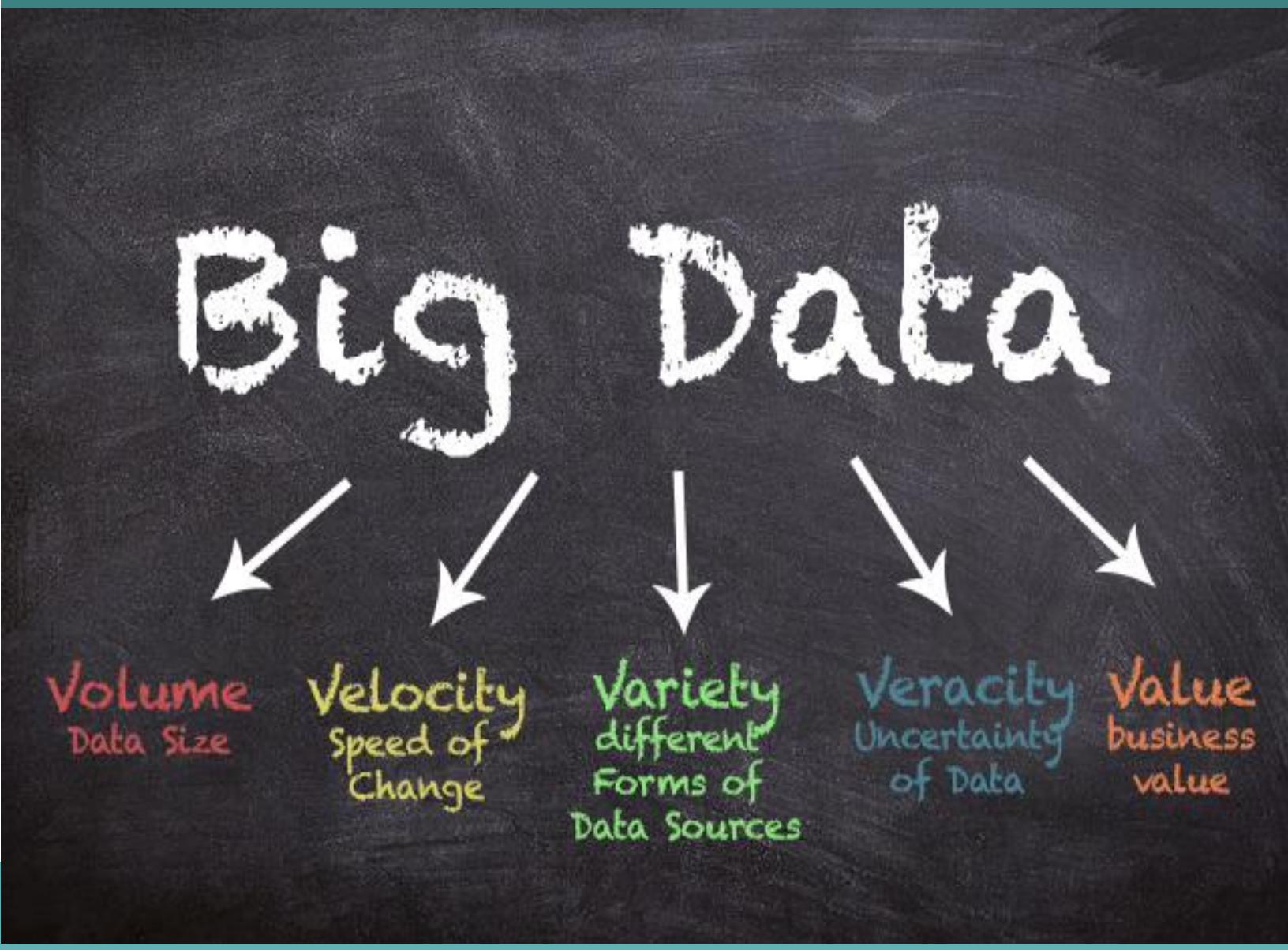
06/13/2018 | By: Ajit K Prasad




# Big Data Engineering with Hadoop & Spark

Assignment on Introduction to Big Data

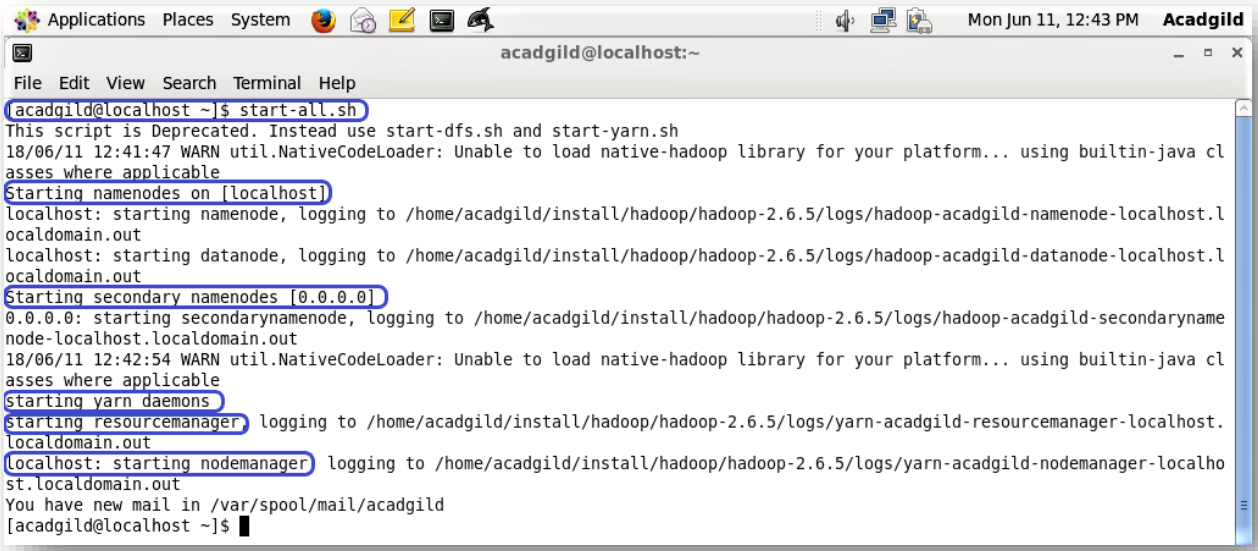
Acadgild

# Session 1: Assignment 1.1

This assignment is aimed at consolidating the concepts that was learned during the opening session of the course.

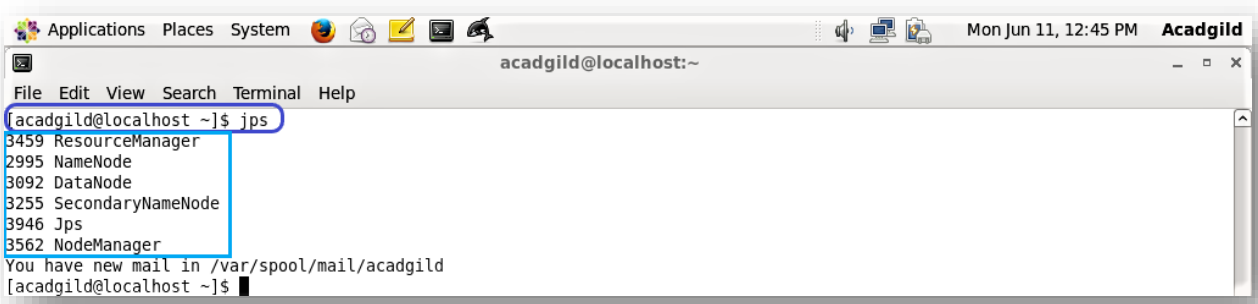
# Task 1:

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




# Task 2:

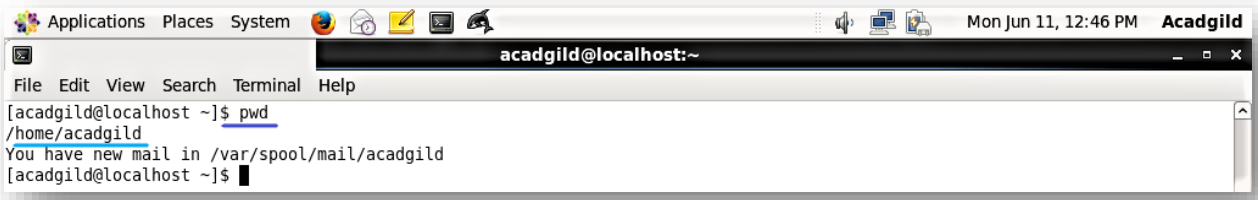
Run a ***JPS*** command to see if all Hadoop daemons are running.



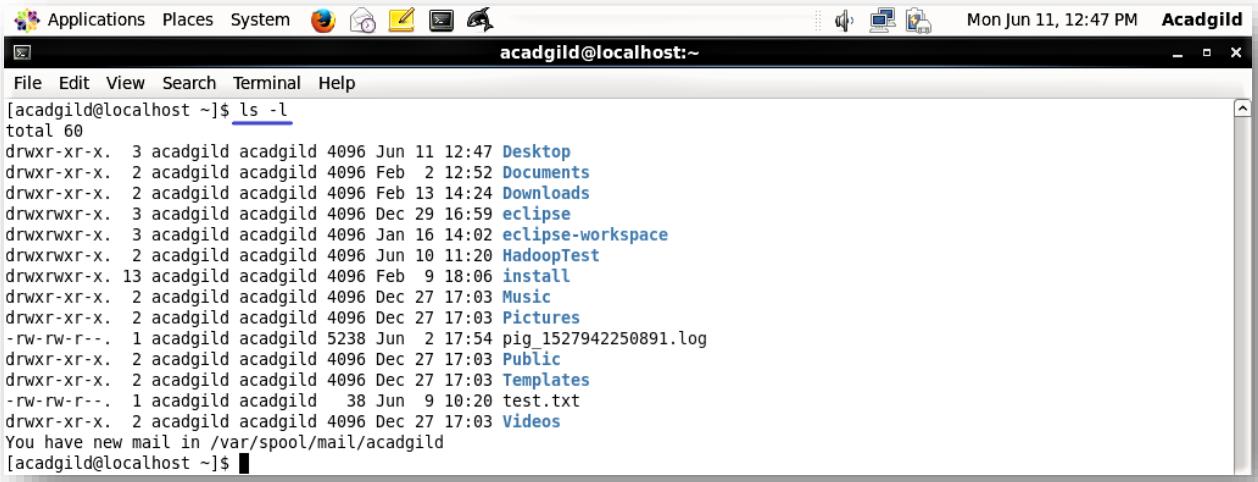
# Task 3:

Run few Unix commands.

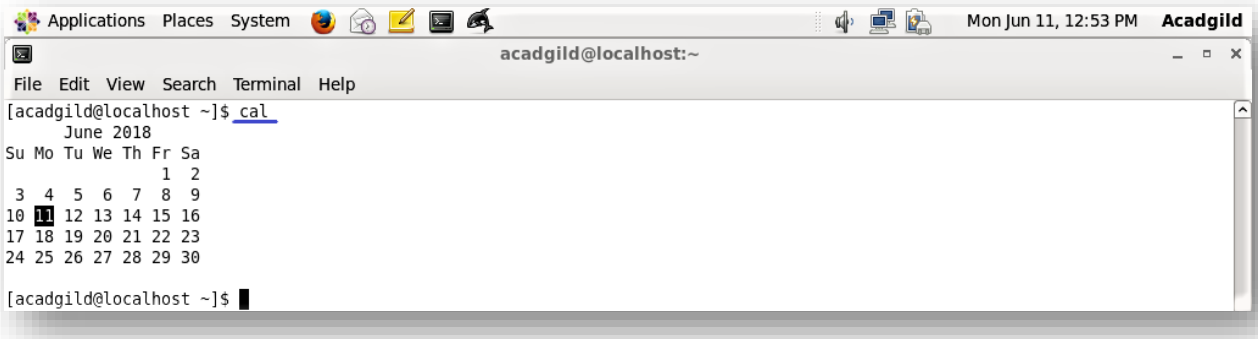
1. pwd: When you first open the terminal, you are in the home directory of your user. To know which directory you are in, you can use the *"pwd"* command. It gives us the absolute path, which means the path that starts from the root. The root is the base of the Linux file system. It is denoted by a forward slash( / ). The user directory is usually something like "/home/username". Here the username is "acadgild".



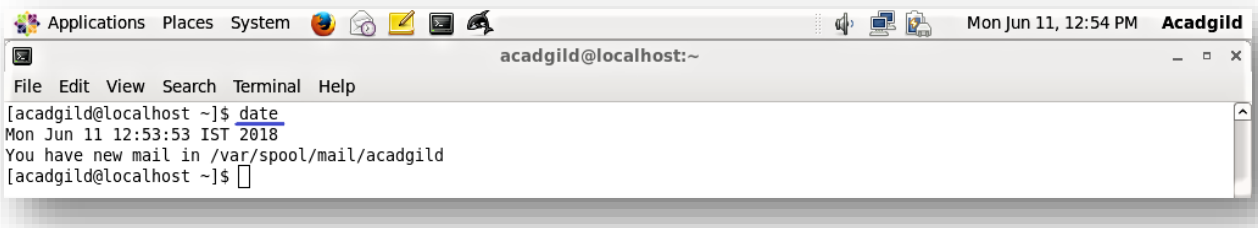

2. ls: Use the *"ls"* command to know what files are in the directory you are in. *"-l"* is a quick and easy way to list a file's permissions are with the long listing option of the ls command.



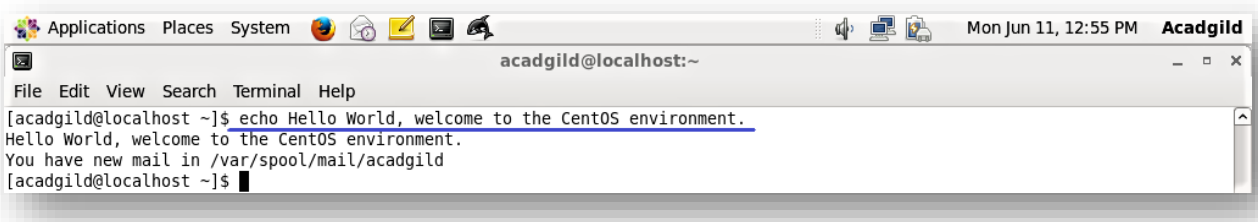
**3.** cal: The *"cal"* command displays a simple, formatted calendar in your terminal.



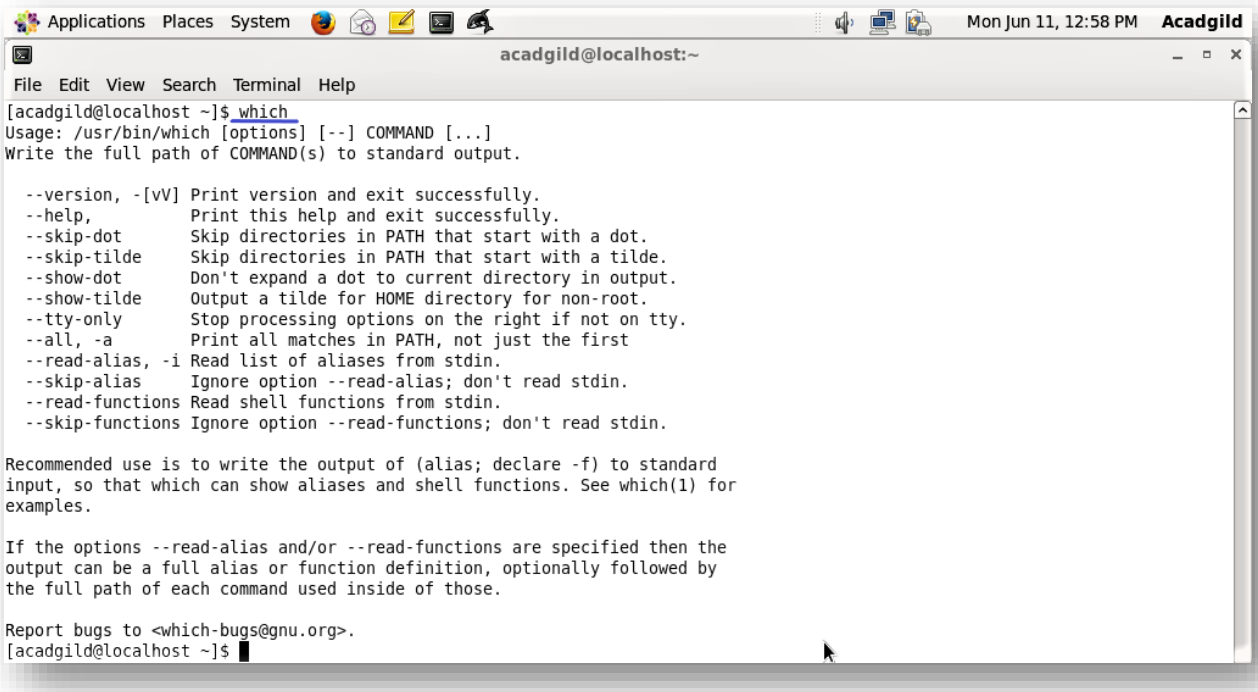

**4.** date: The *"date"* command is used to display the system date and time.



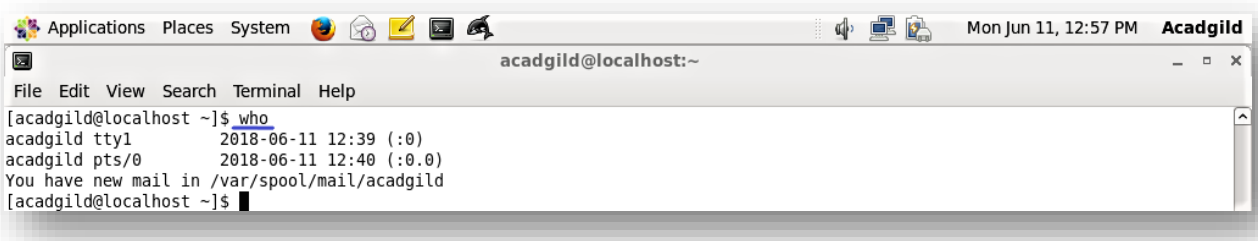

**5.** echo: The *"echo"* command is used to input a line of text and display on standard output.



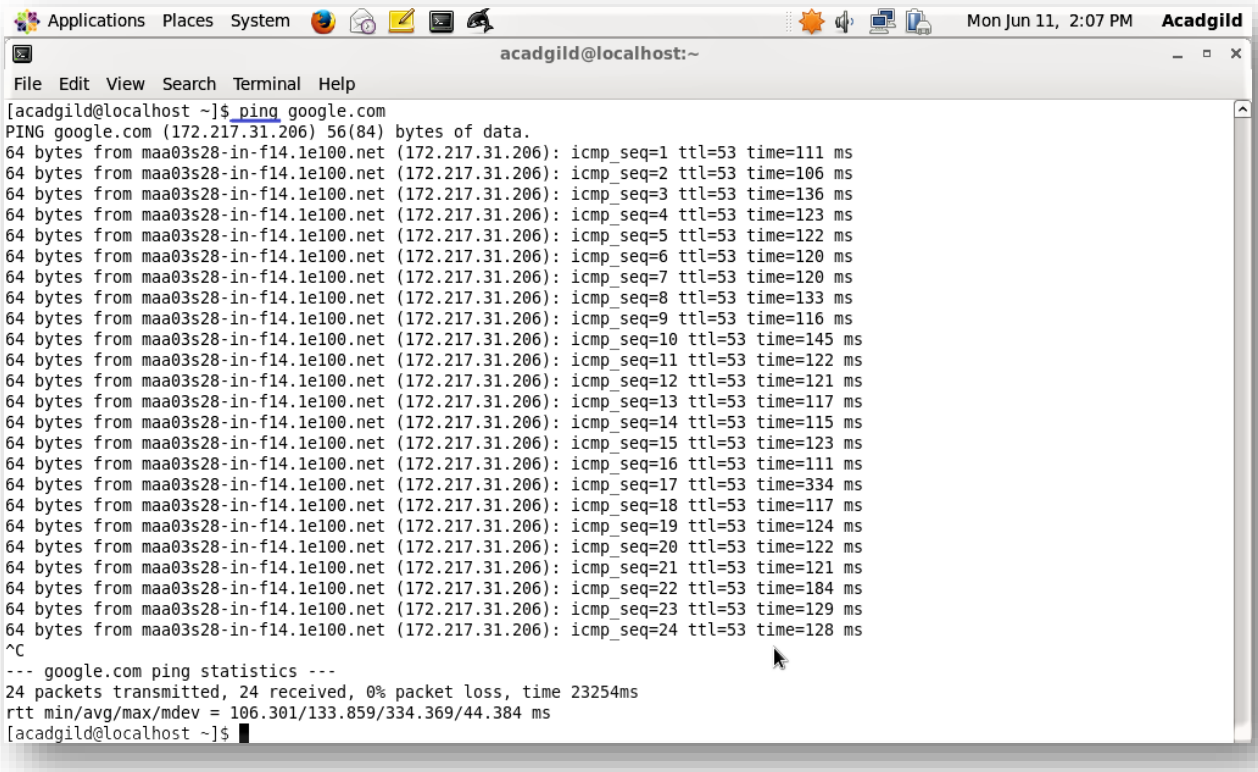
6. which: The *"which"* command is used to identify the location of executables. The command takes one or more arguments; for each of these arguments, it prints the full path of the executable to standard output that would have been executed if this argument had been entered into the shell.



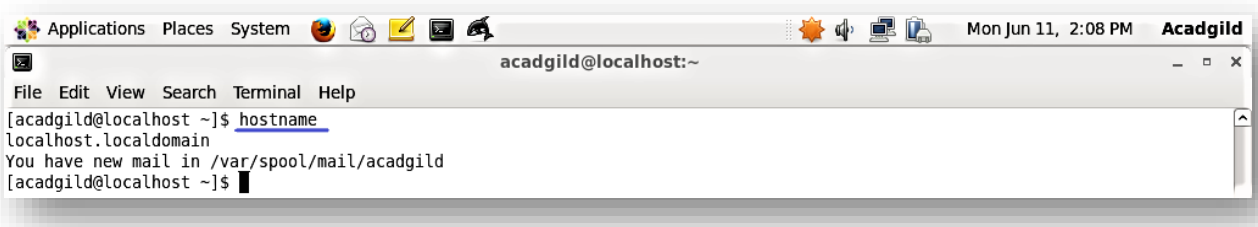

7. who: The *"who"* command is used to get information about currently logged in user on to system.



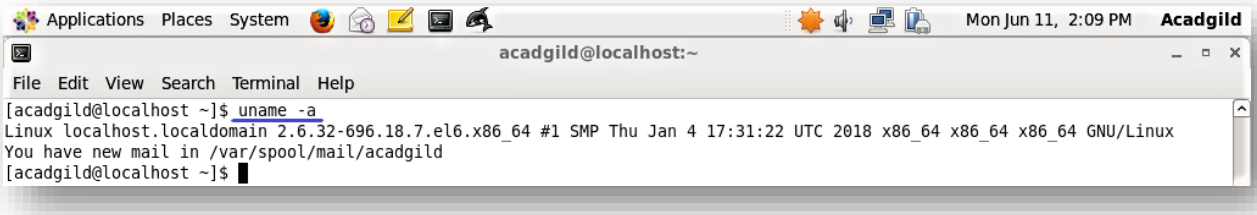
**8.** ping: The *"ping"* command is used to check your connection to a server. When you type in, for example, "**ping google.com**", it checks if it can connect to the server and come back. It measures this round-trip time and gives you the details about it. The use of this command for simple users like us is to check your internet connection. If it pings the Google server (in this case), you can confirm that your internet connection is active!



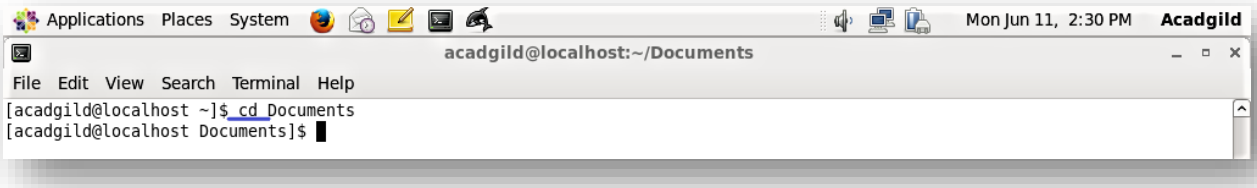

**9.** hostname: The *"hostname"* command is used to know your name in your host or network.



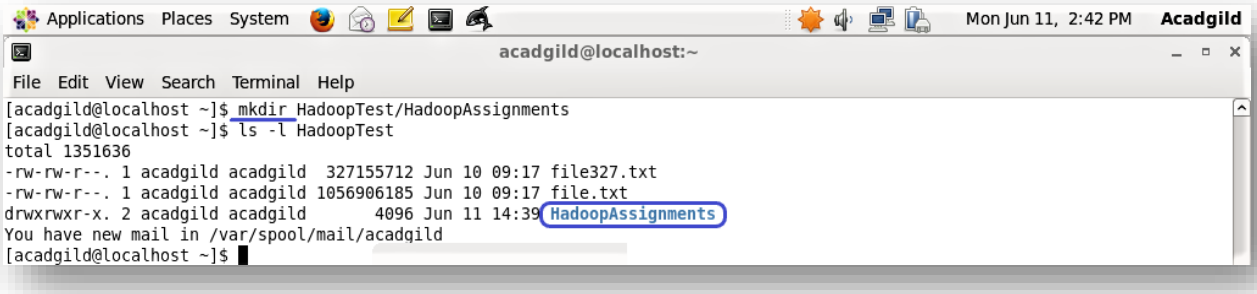
**10.** uname: The *"uname"* command shows the information about the system your Linux distro is running. Using the command "**uname -a**" prints most of the information about the system. This prints the kernel release date, version, processor type, etc.



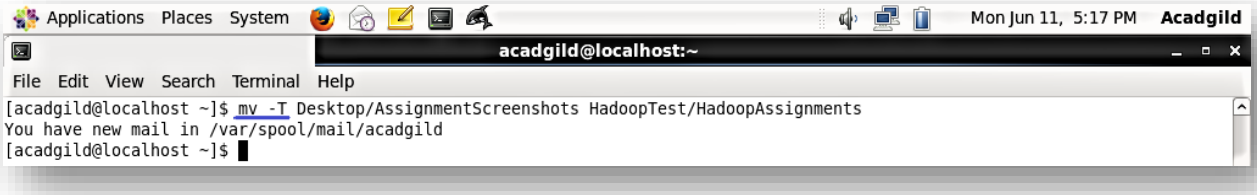

**11.** cd: The *"cd"* command is used to go to a directory.




**12.** mkdir: The *"mkdir"* command is used to create a folder or a directory.




**13.** mv: The *"mv"* command is used to move files through the command line.

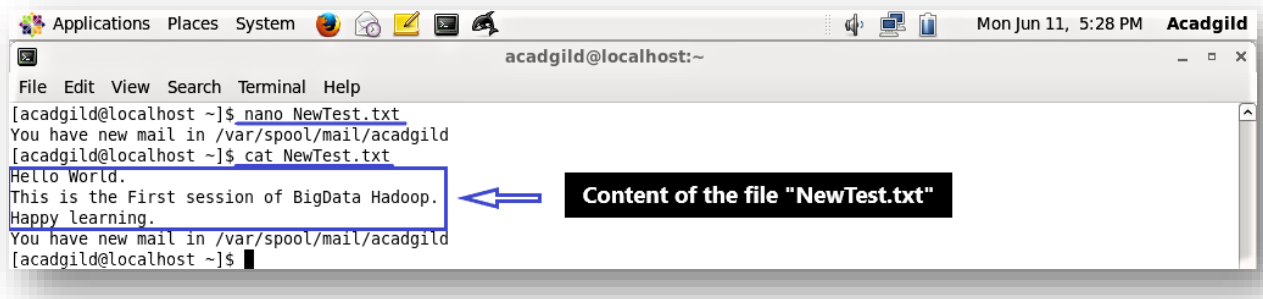


# Task 4:

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.

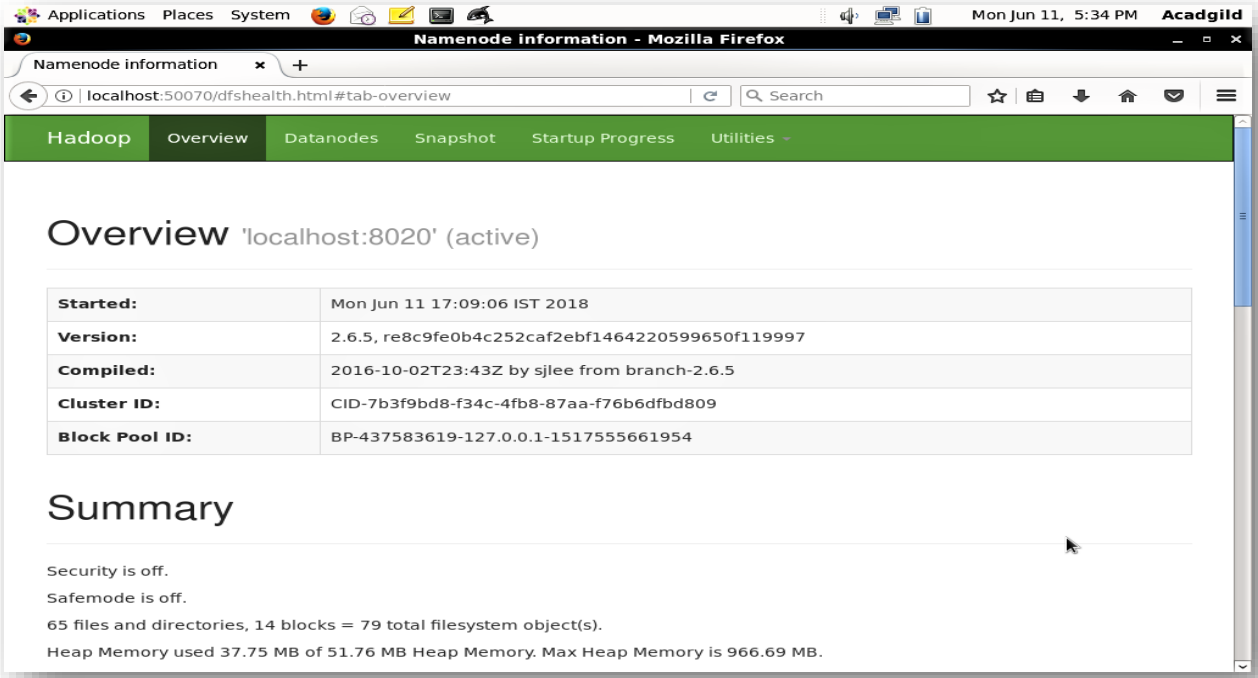
<u>Solution:</u>

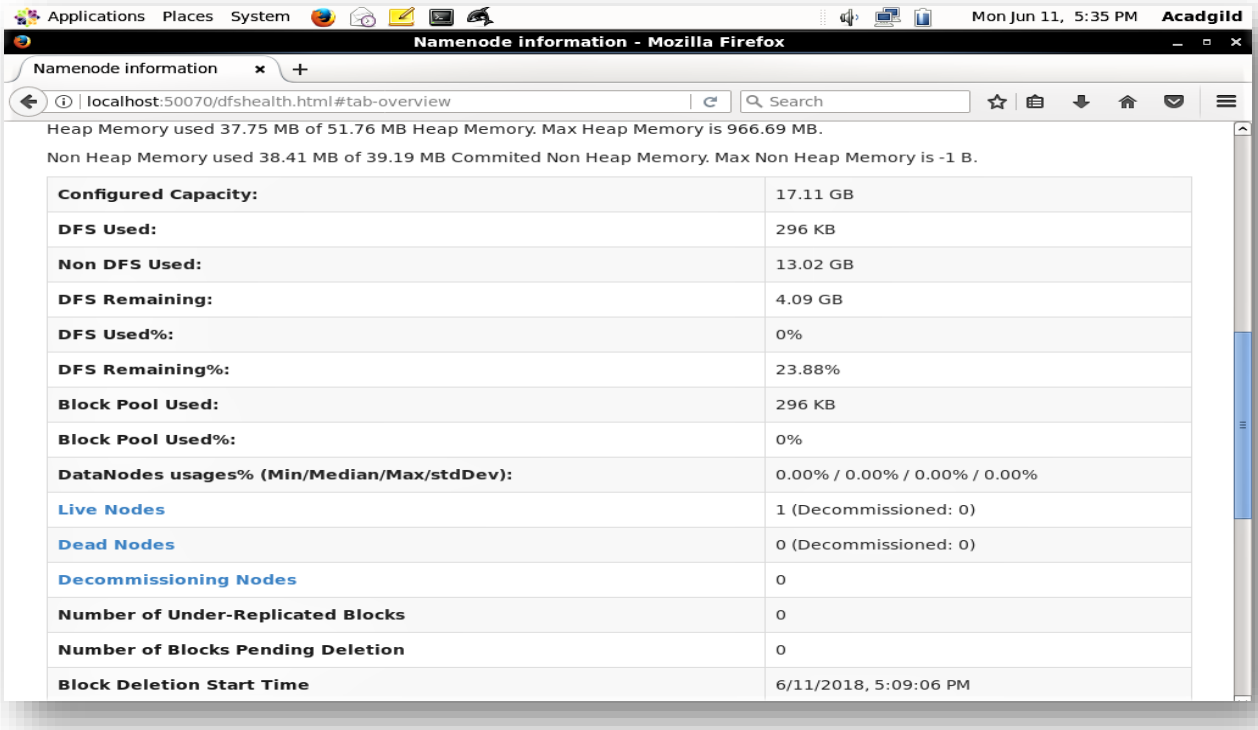
- A file named *“NewTest.txt”* was created using *“nano editor”* with the help of *“nano”* command on command line.
- Text content was added on to the editor and was saved.
- From command line *“cat”* command was used to check in the text content was saved on the file or not.

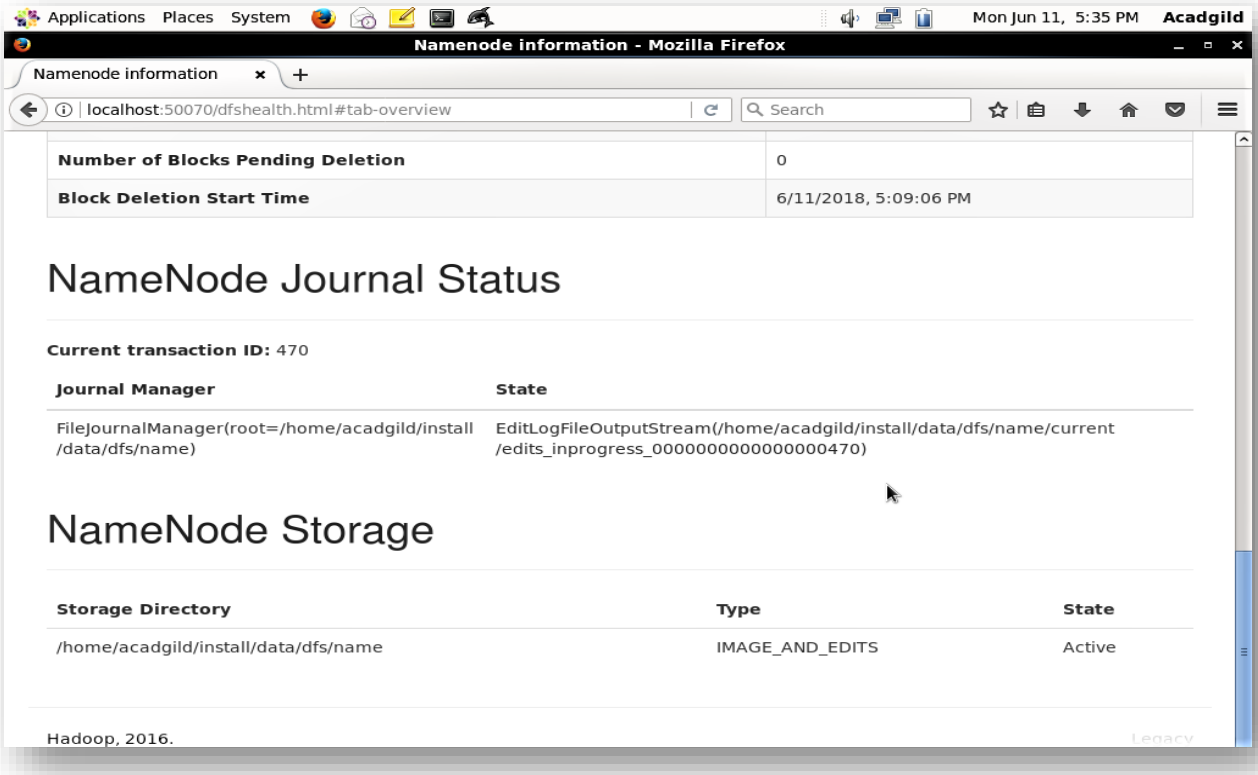


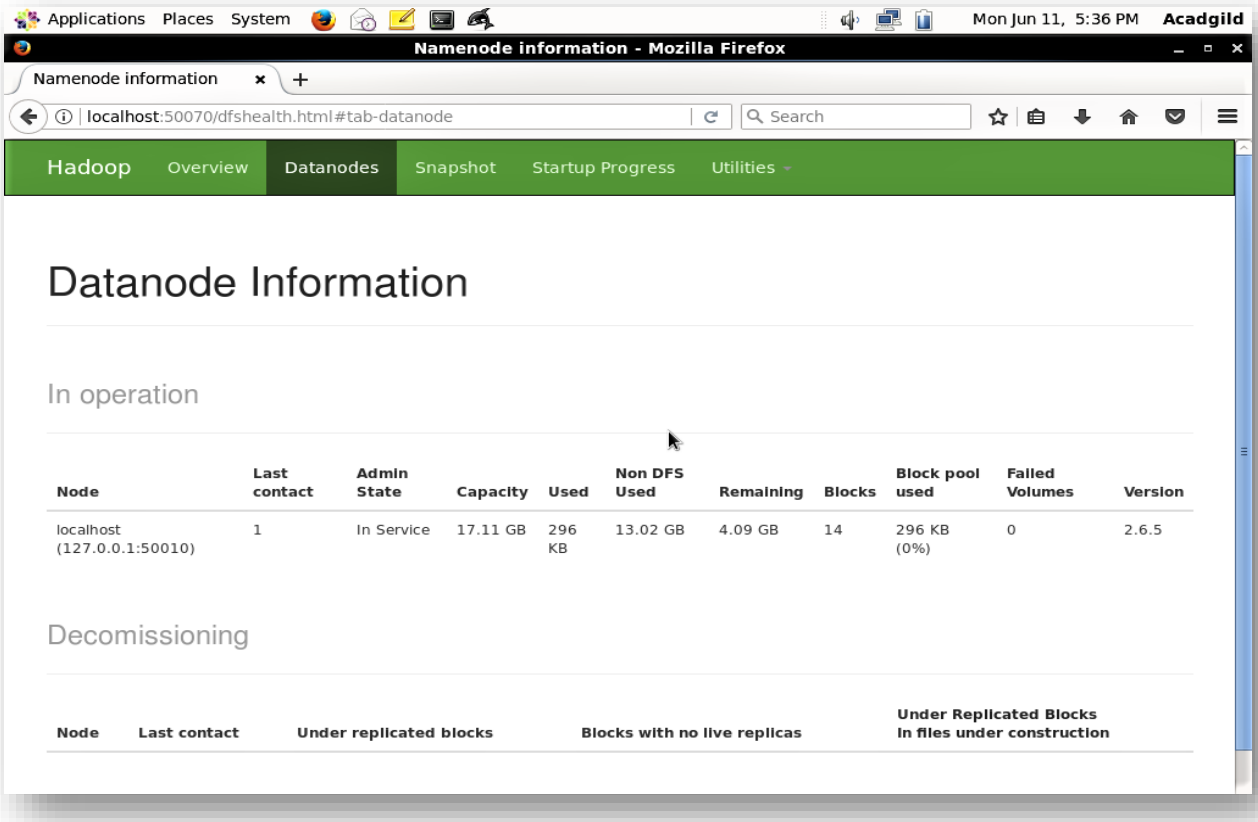
# Task 5:

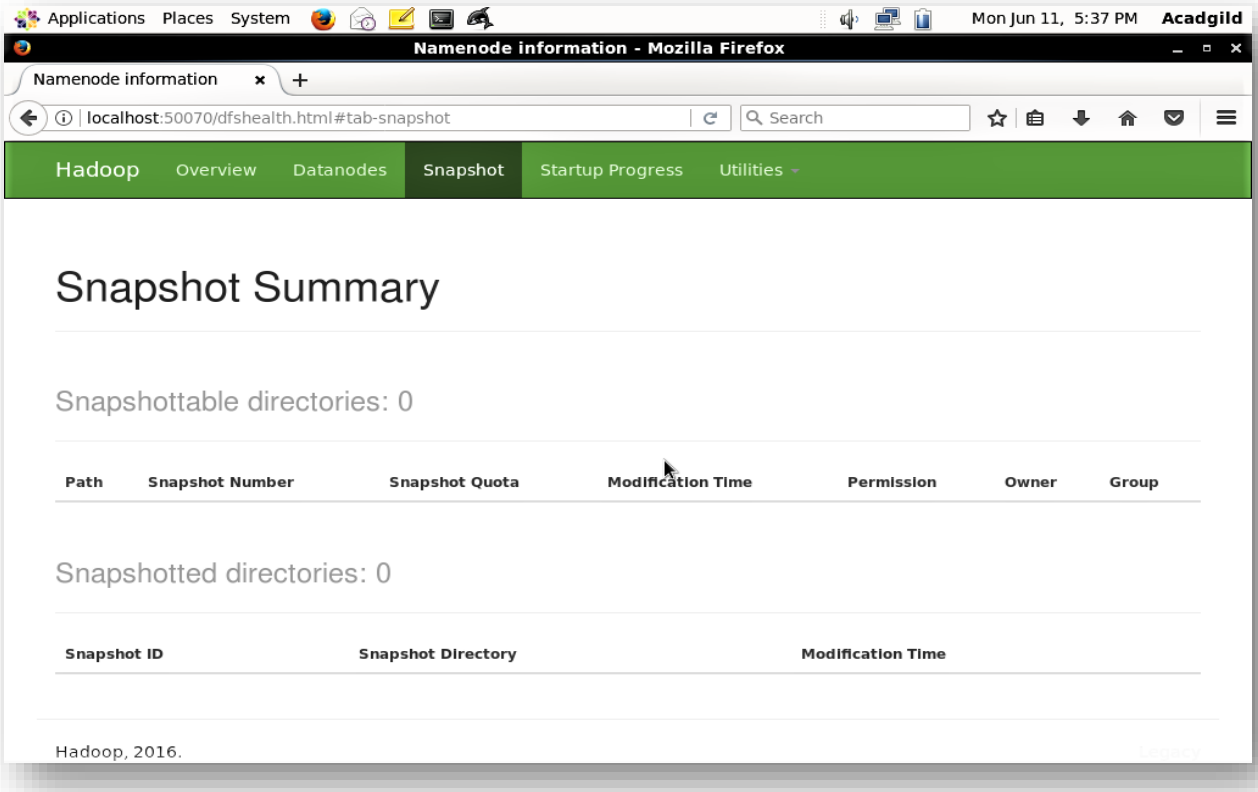
Open the hdfs web page by typing *localhost:50070* in the browser. Check all the details of the HDFS.

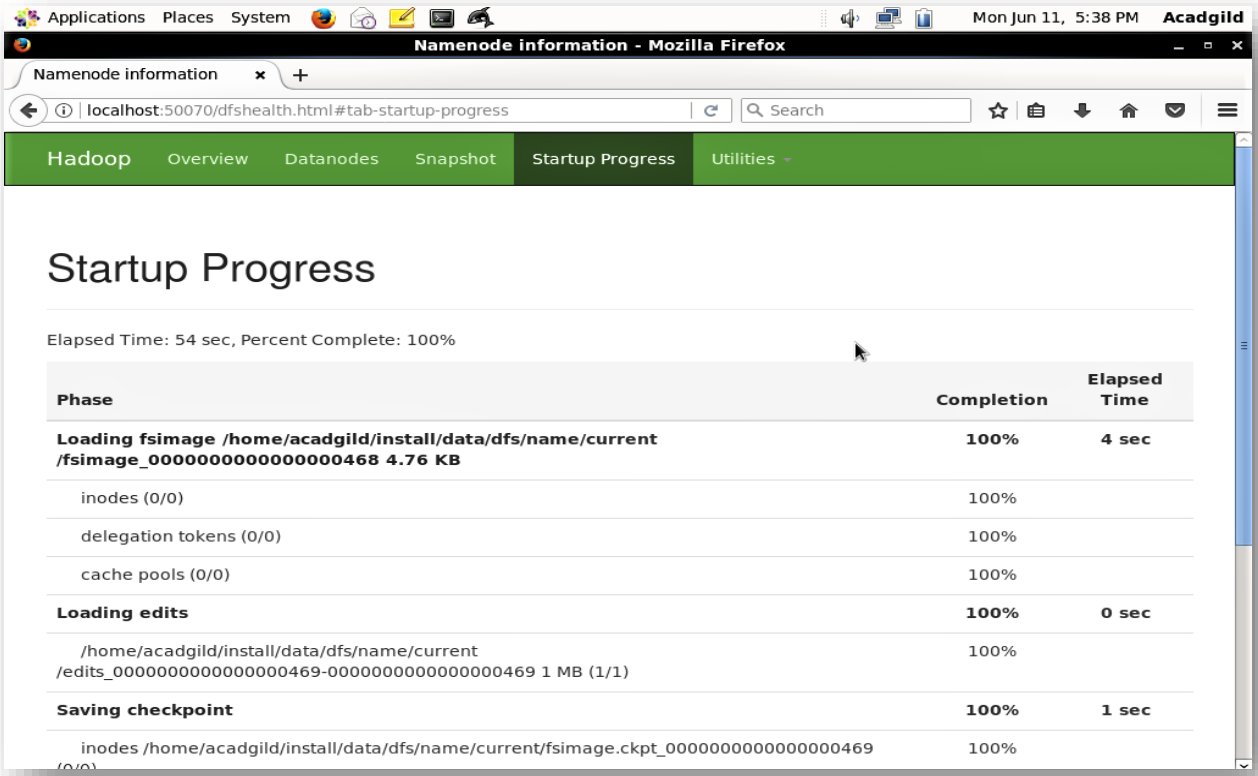


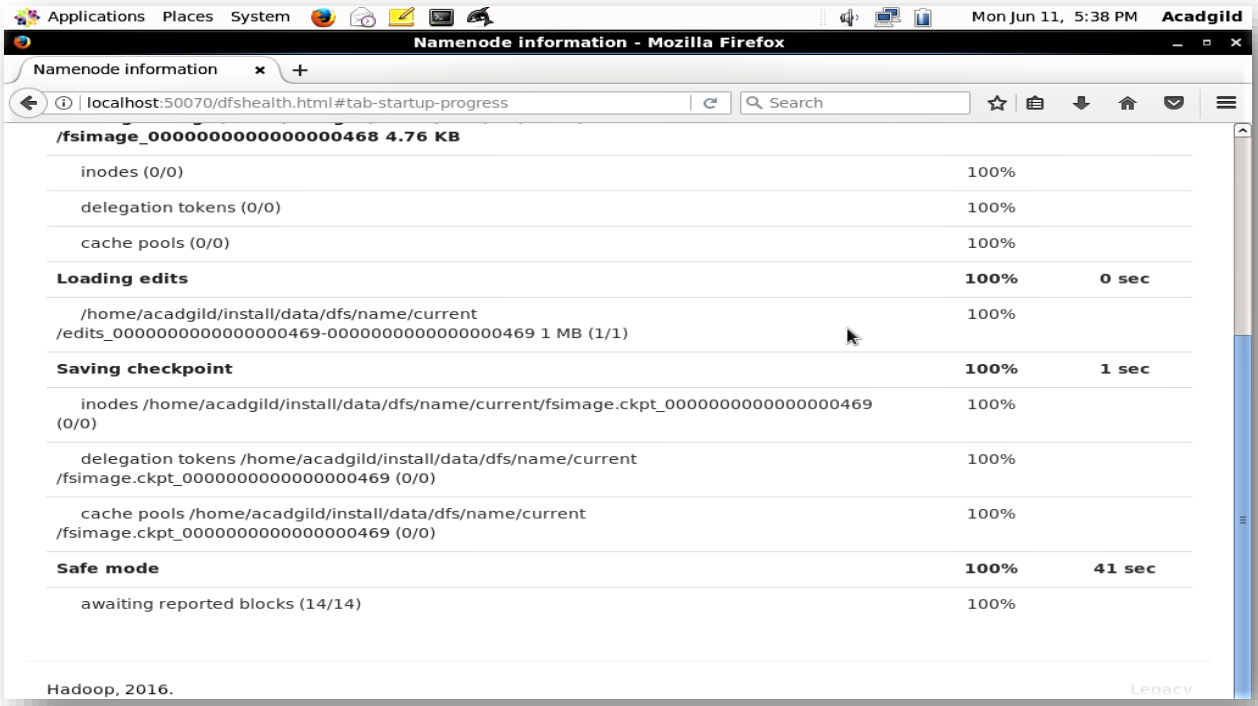



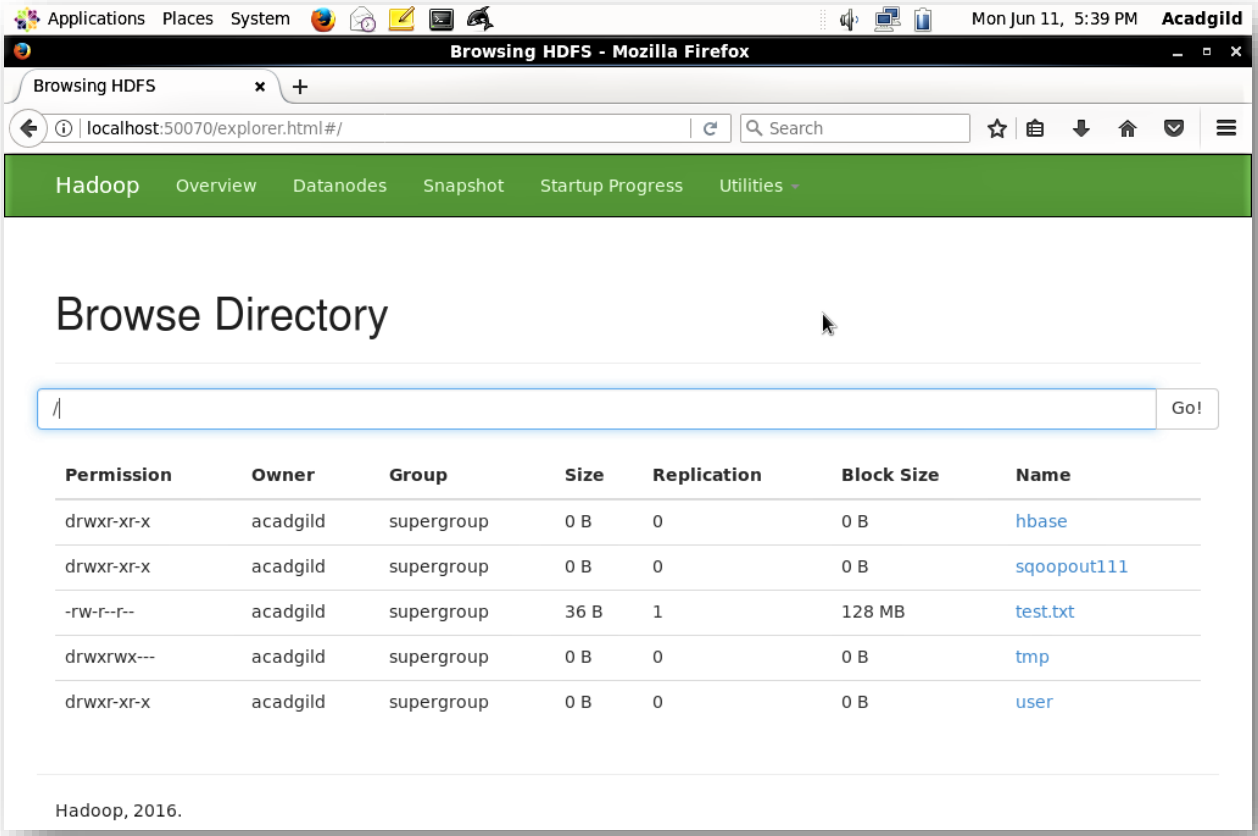

Applications Places System
Mon Jun 11, 5:35 PM Acadgild
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-overview
Search
Number of Blocks Pending Deletion 0
Block Deletion Start Time 6/11/2018, 5:09:06 PM
NameNode Journal Status
Current transaction ID: 470
Journal Manager
State
FileJournalManager(root=/home/acadgild/install/data/dfs/name)
EditLogFileOutputStream(/home/acadgild/install/data/dfs/name/current/edits_inprogress_0000000000000000470)
NameNode Storage
Storage Directory
Type
State
/home/acadgild/install/data/dfs/name
IMAGE_AND_EDITS
Active
Hadoop, 2016.


Applications Places System
Mon Jun 11, 5:36 PM Acadgild
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-datanode
Search
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
296 KB
13.02 GB
4.09 GB
14
296 KB (0%)
0
2.6.5
Decomissioning
Node
Last contact
Under replicated blocks
Blocks with no live replicas
Under Replicated Blocks In files under construction


Applications Places System
Mon Jun 11, 5:37 PM Acadgild
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-snapshot
Search
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Snapshot Summary
Snapshottable directories: 0
Path Snapshot Number Snapshot Quota Modification Time Permission Owner Group
Snapshotted directories: 0
Snapshot ID Snapshot Directory Modification Time
Hadoop, 2016.


Applications Places System
Mon Jun 11, 5:38 PM Acadgild
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-startup-progress
Search
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Startup Progress
Elapsed Time: 54 sec, Percent Complete: 100%
Phase Completion Elapsed Time
Loading fsimage /home/acadgild/install/data/dfs/name/current /fsimage_0000000000000000468 4.76 KB 100% 4 sec
inodes (0/0) 100%
delegation tokens (0/0) 100%
cache pools (0/0) 100%
Loading edits 100% 0 sec
/home/acadgild/install/data/dfs/name/current /edits_0000000000000000469-0000000000000000469 1 MB (1/1) 100%
Saving checkpoint 100% 1 sec
inodes /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_0000000000000000469 100%


Applications Places System
Mon Jun 11, 5:38 PM Acadgild
Namenode information - Mozilla Firefox
Namenode information
localhost:50070/dfshealth.html#tab-startup-progress
Search
/fsimage_0000000000000000468 4.76 KB
inodes (0/0) 100%
delegation tokens (0/0) 100%
cache pools (0/0) 100%
Loading edits 100% 0 sec
/home/acadgild/install/data/dfs/name/current /edits_0000000000000000469-0000000000000000469 1 MB (1/1) 100%
Saving checkpoint 100% 1 sec
inodes /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_0000000000000000469 (0/0) 100%
delegation tokens /home/acadgild/install/data/dfs/name/current /fsimage.ckpt_0000000000000000469 (0/0) 100%
cache pools /home/acadgild/install/data/dfs/name/current /fsimage.ckpt_0000000000000000469 (0/0) 100%
Safe mode 100% 41 sec
awaiting reported blocks (14/14) 100%
Hadoop, 2016.


Applications Places System
Mon Jun 11, 5:39 PM Acadgild
Browsing HDFS - Mozilla Firefox
Browsing HDFS
localhost:50070/explorer.html#/
Search
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Browse Directory
/
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 sqoopout111
-rw-r--r-- acadgild supergroup 36 B 1 128 MB test.txt
drwxrwx--- acadgild supergroup 0 B 0 0 B tmp
drwxr-xr-x acadgild supergroup 0 B 0 0 B user
Hadoop, 2016.