



Courses ▾

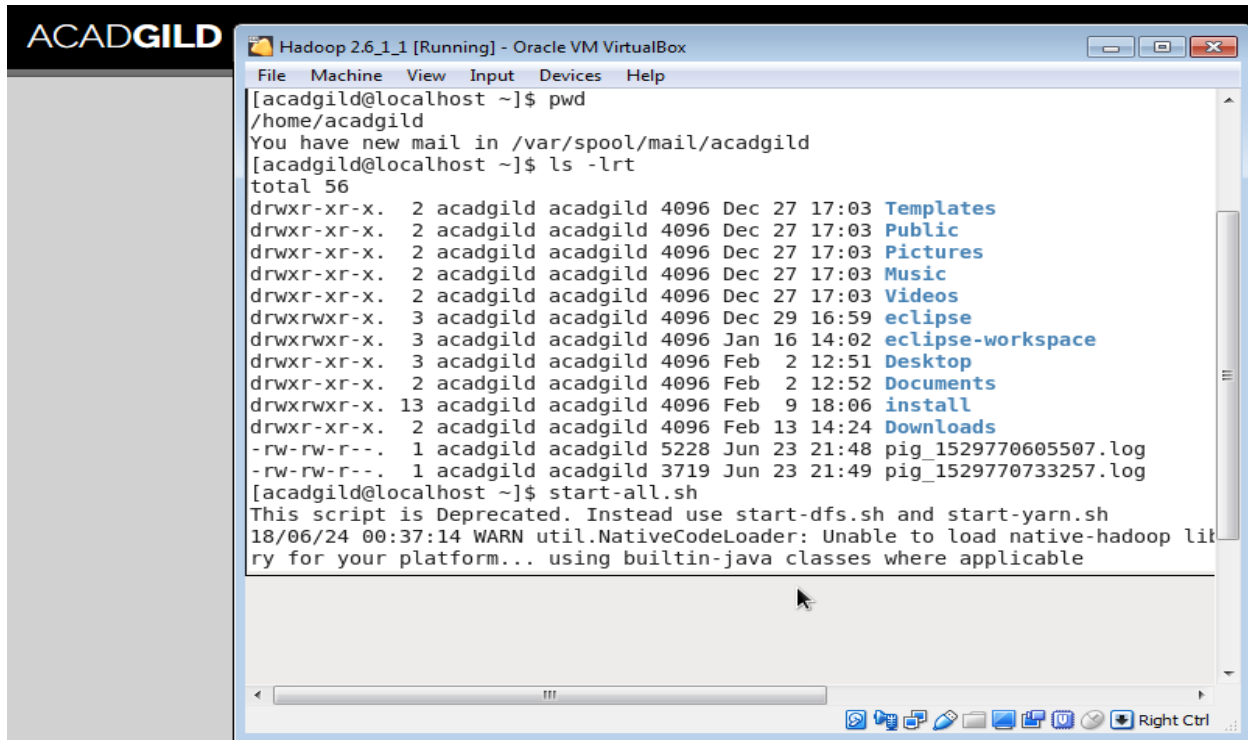


BIG DATA
DEVELOPMENT

ACADGILD

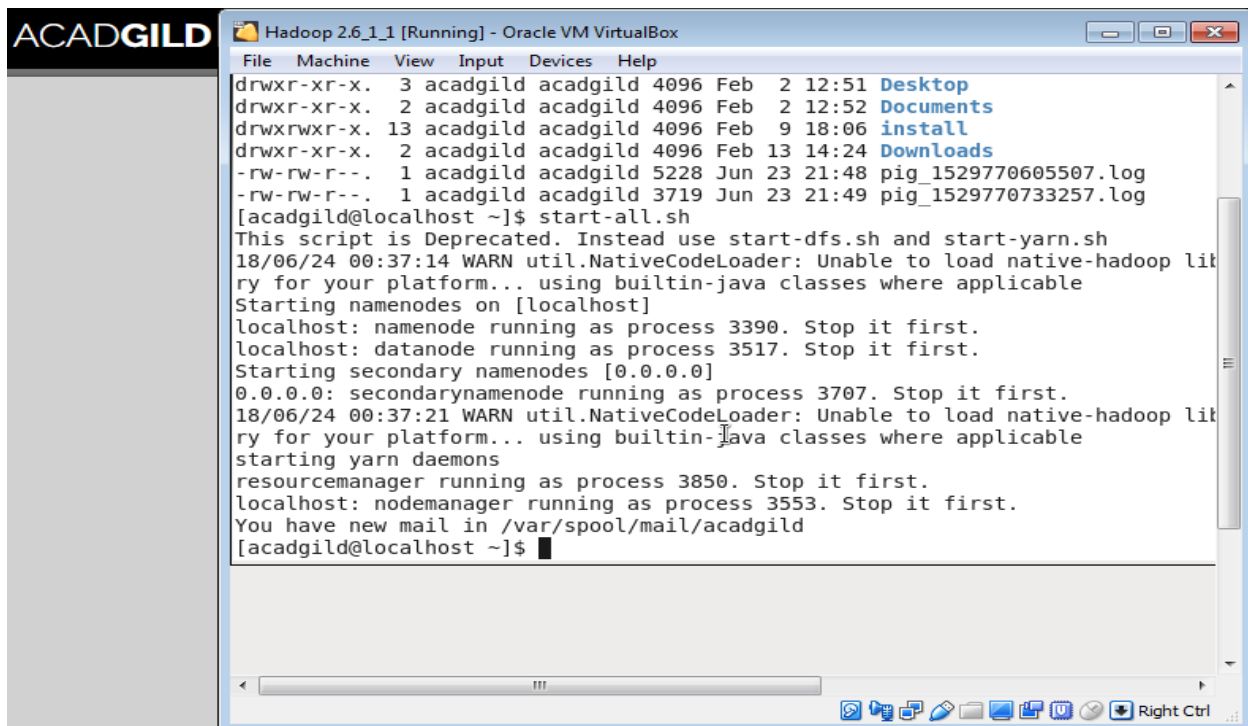
Session 1: BIG DATA & HADOOP – INTRODUCTION & CASE STUDIES

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



The screenshot shows a terminal window with the title "Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox". The terminal output is as follows:

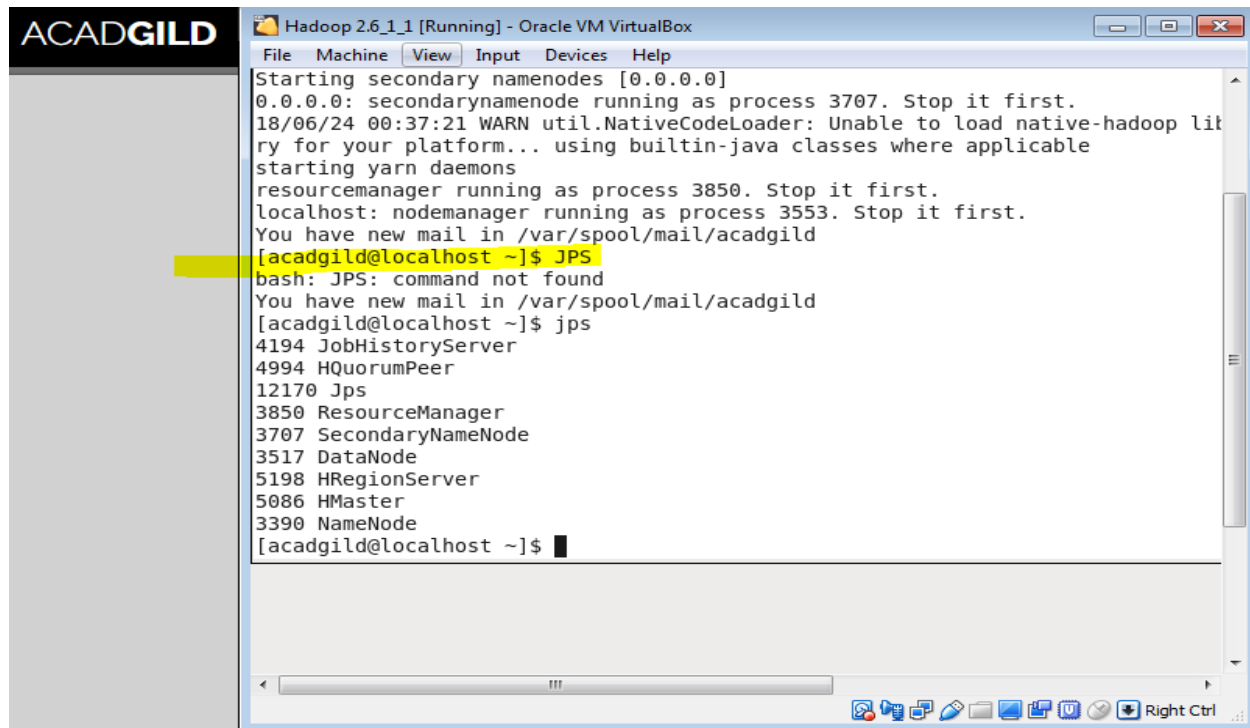
```
[acadgild@localhost ~]$ pwd
/home/acadgild
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ ls -lrt
total 56
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Templates
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Public
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Pictures
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Music
drwxr-xr-x. 2 acadgild acadgild 4096 Dec 27 17:03 Videos
drwxrwxr-x. 3 acadgild acadgild 4096 Dec 29 16:59 eclipse
drwxrwxr-x. 3 acadgild acadgild 4096 Jan 16 14:02 eclipse-workspace
drwxr-xr-x. 3 acadgild acadgild 4096 Feb 2 12:51 Desktop
drwxr-xr-x. 2 acadgild acadgild 4096 Feb 2 12:52 Documents
drwxrwxr-x. 13 acadgild acadgild 4096 Feb 9 18:06 install
drwxr-xr-x. 2 acadgild acadgild 4096 Feb 13 14:24 Downloads
-rw-rw-r--. 1 acadgild acadgild 5228 Jun 23 21:48 pig_1529770605507.log
-rw-rw-r--. 1 acadgild acadgild 3719 Jun 23 21:49 pig_1529770733257.log
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/06/24 00:37:14 WARN util.NativeCodeLoader: Unable to load native-hadoop lib
ry for your platform... using builtin-java classes where applicable
```



The screenshot shows the continuation of the terminal window. The output continues from the previous state:

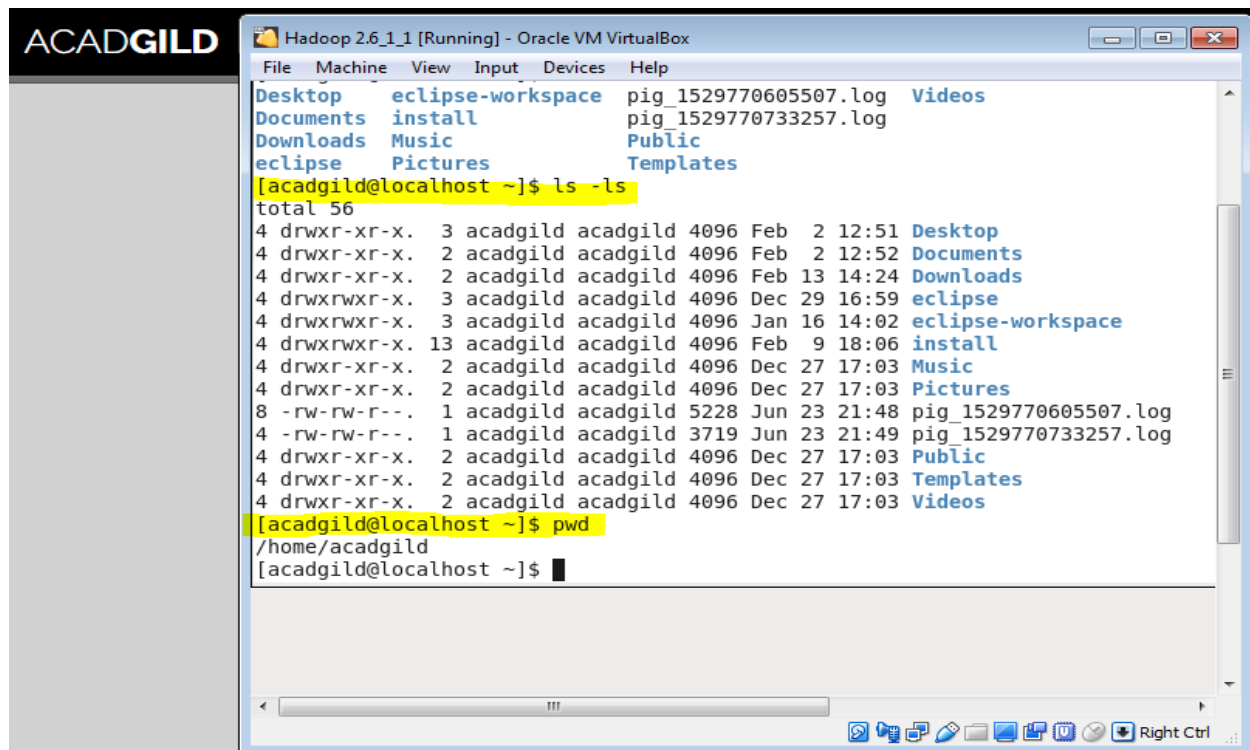
```
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/06/24 00:37:14 WARN util.NativeCodeLoader: Unable to load native-hadoop lib
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 3390. Stop it first.
localhost: datanode running as process 3517. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 3707. Stop it first.
18/06/24 00:37:21 WARN util.NativeCodeLoader: Unable to load native-hadoop lib
ry for your platform... using builtin-java classes where applicable
starting yarn daemons
resourcemanager running as process 3850. Stop it first.
localhost: nodemanager running as process 3553. Stop it first.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

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



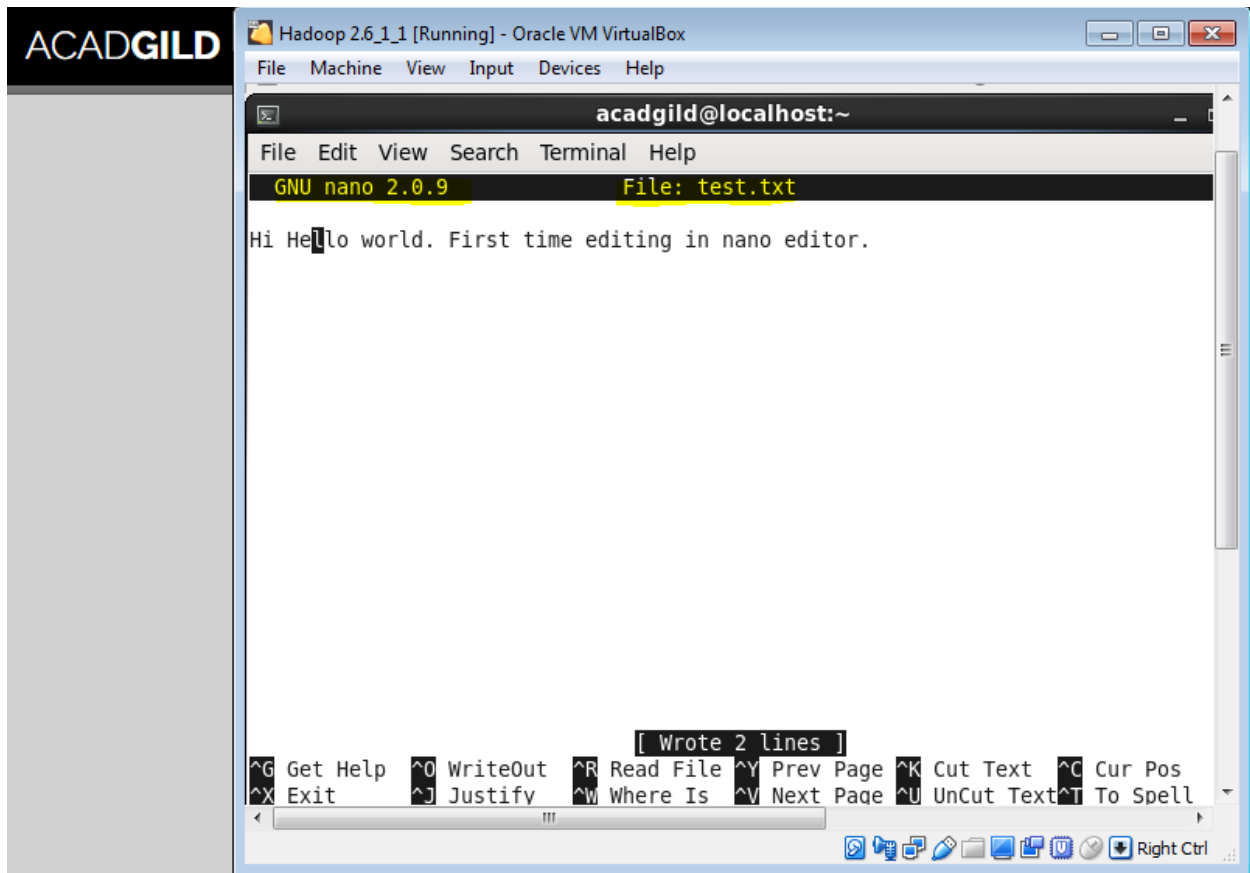
```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 3707. Stop it first.
18/06/24 00:37:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
resourcemanager running as process 3850. Stop it first.
localhost: nodemanager running as process 3553. Stop it first.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ JPS
bash: JPS: command not found
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ jps
4194 JobHistoryServer
4994 HQuorumPeer
12170 Jps
3850 ResourceManager
3707 SecondaryNameNode
3517 DataNode
5198 HRegionServer
5086 HMaster
3390 NameNode
[acadgild@localhost ~]$
```

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



```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Desktop eclipse-workspace pig_1529770605507.log Videos
Documents install pig_1529770733257.log
Downloads Music Public
eclipse Pictures Templates
[acadgild@localhost ~]$ ls -ls
total 56
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
8 -rw-rw-r--. 1 acadgild acadgild 5228 Jun 23 21:48 pig_1529770605507.log
4 -rw-rw-r--. 1 acadgild acadgild 3719 Jun 23 21:49 pig_1529770733257.log
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 ~]$ pwd
/home/acadgild
[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.



The screenshot shows a terminal window titled "Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox". The terminal prompt is "acadgild@localhost:~". The nano editor is open, editing "File: test.txt". The text "Hi Hello world. First time editing in nano editor." is visible. The nano editor's status bar at the bottom indicates "[Wrote 2 lines]". The nano editor's help text is visible at the bottom of the terminal window.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
GNU nano 2.0.9 File: test.txt  
Hi Hello world. First time editing in nano editor.  
[ Wrote 2 lines ]  
^G Get Help ^O WriteOut ^R Read File ^Y Prev Page ^K Cut Text ^C Cur Pos  
^X Exit ^J Justify ^W Where Is ^V Next Page ^U UnCut Text ^T To Spell
```

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

The screenshot shows the 'Overview' tab of the Hadoop NameNode web interface. The browser address bar shows 'localhost:50070/dfshealth.html#tab-overview'. The page has a green header with navigation links: Hadoop, Overview, Datanodes, Snapshot, Startup Progress, and Utilities. The main content area is titled 'Overview 'localhost:8020' (active)'. Below the title is a table with the following information:

Started:	Sat Jun 23 21:43:57 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

Below the table is a section titled 'Summary' with the following text:

Security is off.
Safemode is off.

The screenshot shows the 'Summary' tab of the Hadoop NameNode web interface. The browser address bar shows 'localhost:50070/dfshealth.html#tab-overview'. The page displays the following summary information:

Security is off.
Safemode is off.
63 files and directories, 13 blocks = 76 total filesystem object(s).
Heap Memory used 33.08 MB of 49.21 MB Heap Memory. Max Heap Memory is 966.69 MB.
Non Heap Memory used 46.75 MB of 47.69 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	17.11 GB
DFS Used:	152 KB
Non DFS Used:	11.78 GB
DFS Remaining:	5.33 GB
DFS Used%:	0%
DFS Remaining%:	31.15%
Block Pool Used:	152 KB
Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0)

Applications Places System Sun Jun 24, 1:01 AM Acadgild

Namenode information - Mozilla Firefox

Namenode information x +

localhost:50070/dfshealth.html#tab-overview

Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	6/23/2018, 9:43:57 PM

NameNode Journal Status

Current transaction ID: 519

Journal Manager	State
FileJournalManager(root=/home/acadgild/install/data/dfs/name)	EditLogOutputStream(/home/acadgild/install/data/dfs/name/current/edits_inprogress_0000000000000000497)

NameNode Storage

Applications Places System Sun Jun 24, 1:01 AM Acadgild

Namenode information - Mozilla Firefox

Namenode information x +

localhost:50070/dfshealth.html#tab-overview

Number of Blocks Pending Deletion	0
Block Deletion Start Time	6/23/2018, 9:43:57 PM

NameNode Journal Status

Current transaction ID: 519

Journal Manager	State
FileJournalManager(root=/home/acadgild/install/data/dfs/name)	EditLogOutputStream(/home/acadgild/install/data/dfs/name/current/edits_inprogress_0000000000000000497)

NameNode Storage

Storage Directory	Type	State
/home/acadgild/install/data/dfs/name	IMAGE_AND_EDITS	Active