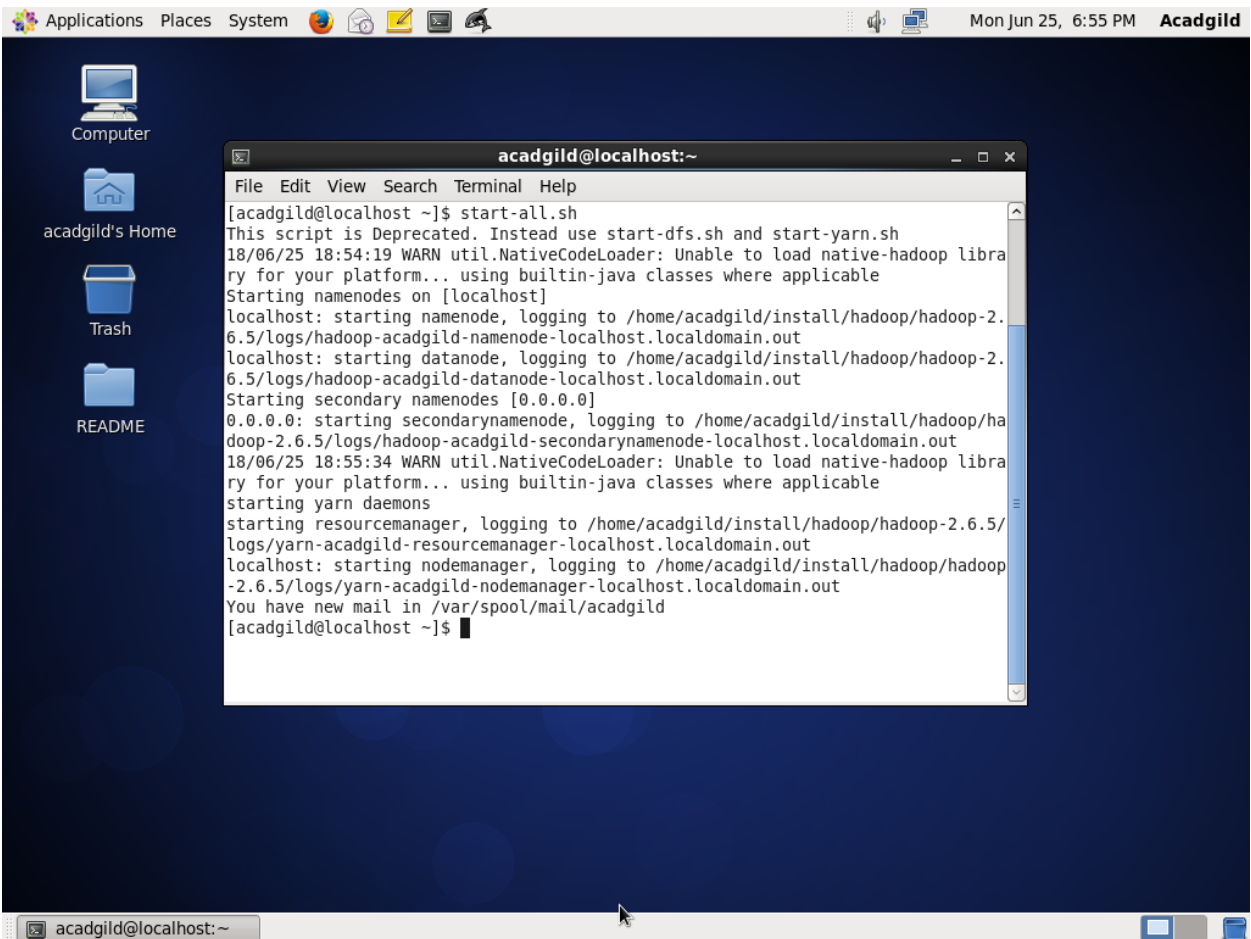


Session 1

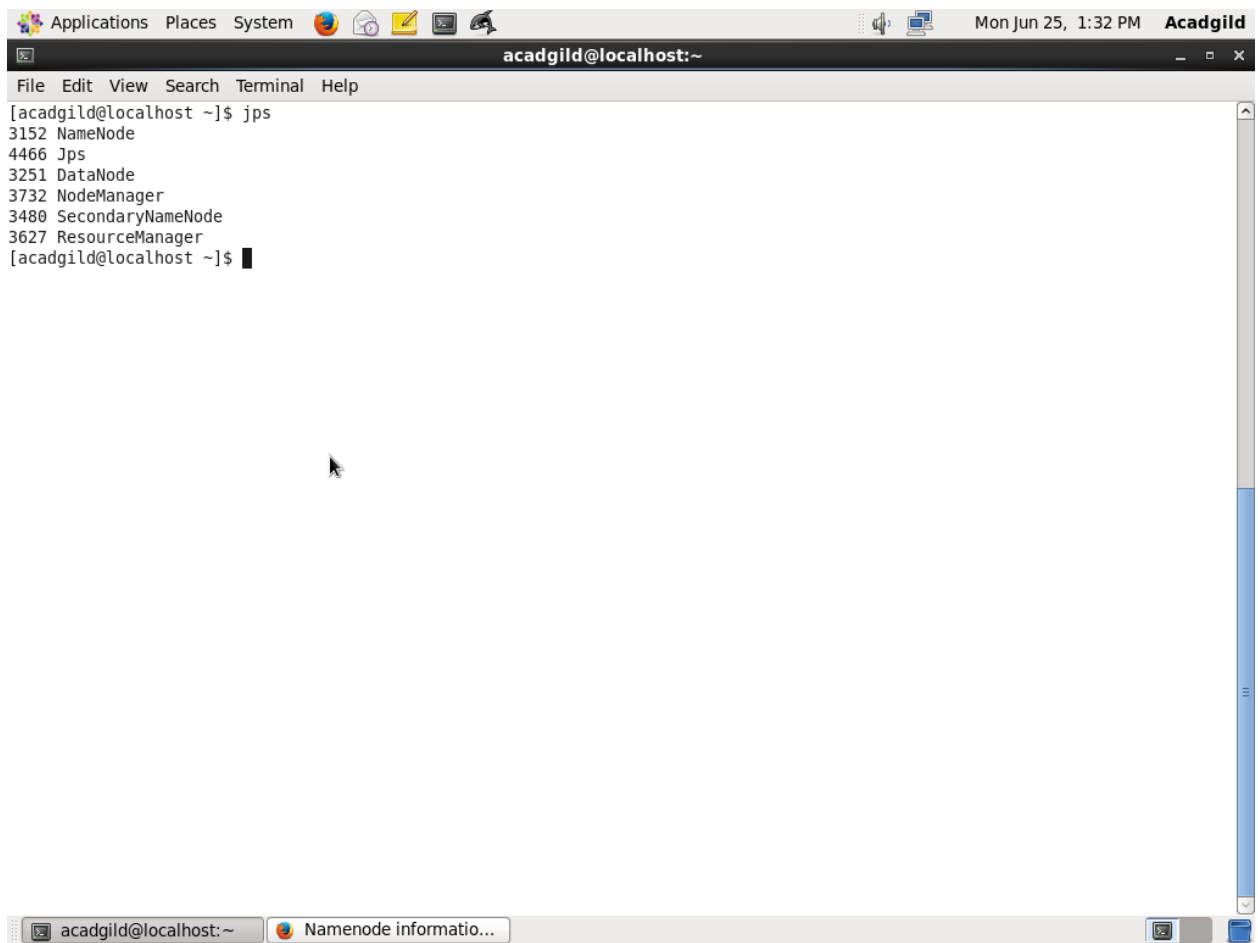
BIG DATA & HADOOP

INTRODUCTION & CASE STUDIES

Assignment 1



Starting Hadoop Cluster

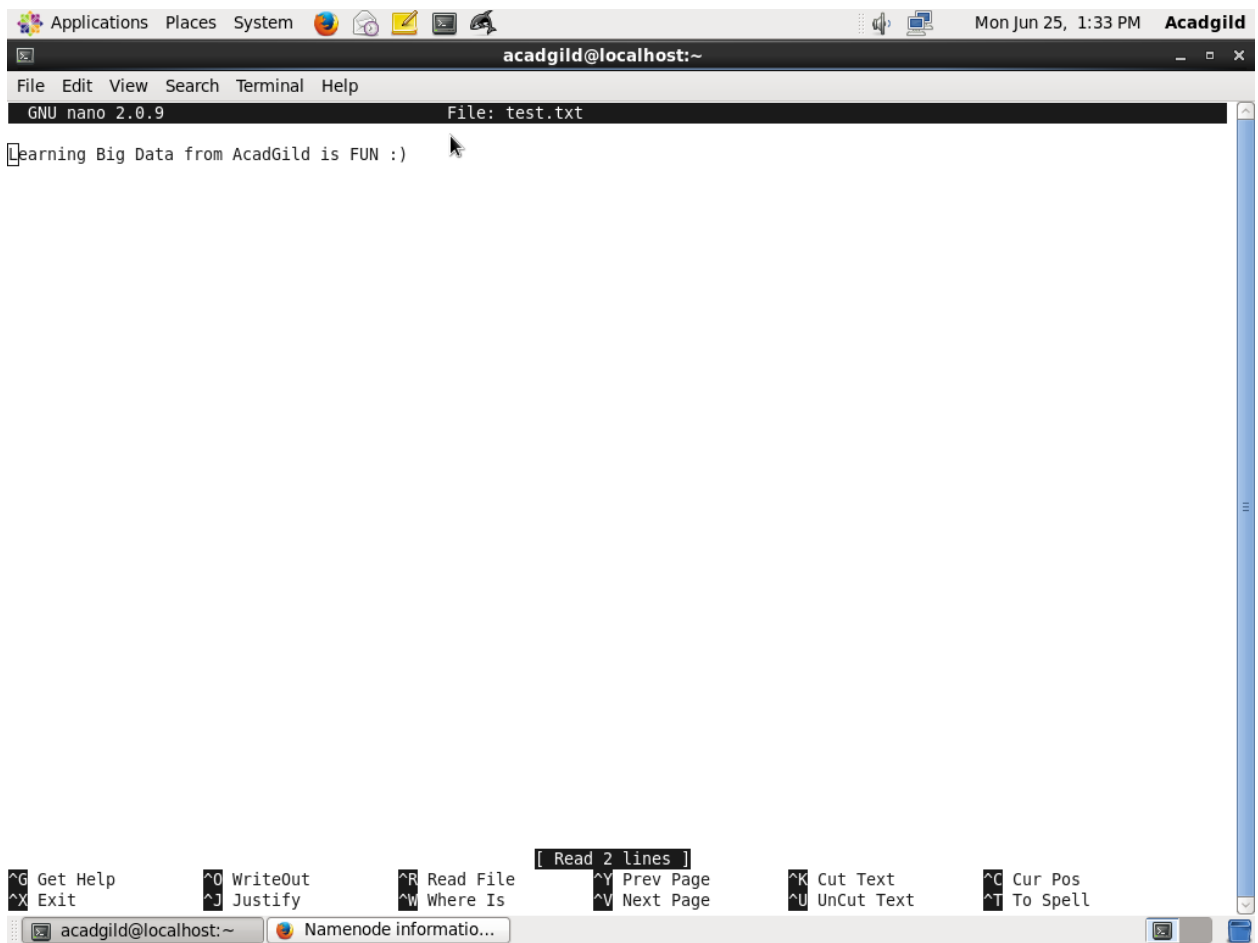


Checking if all processes are running

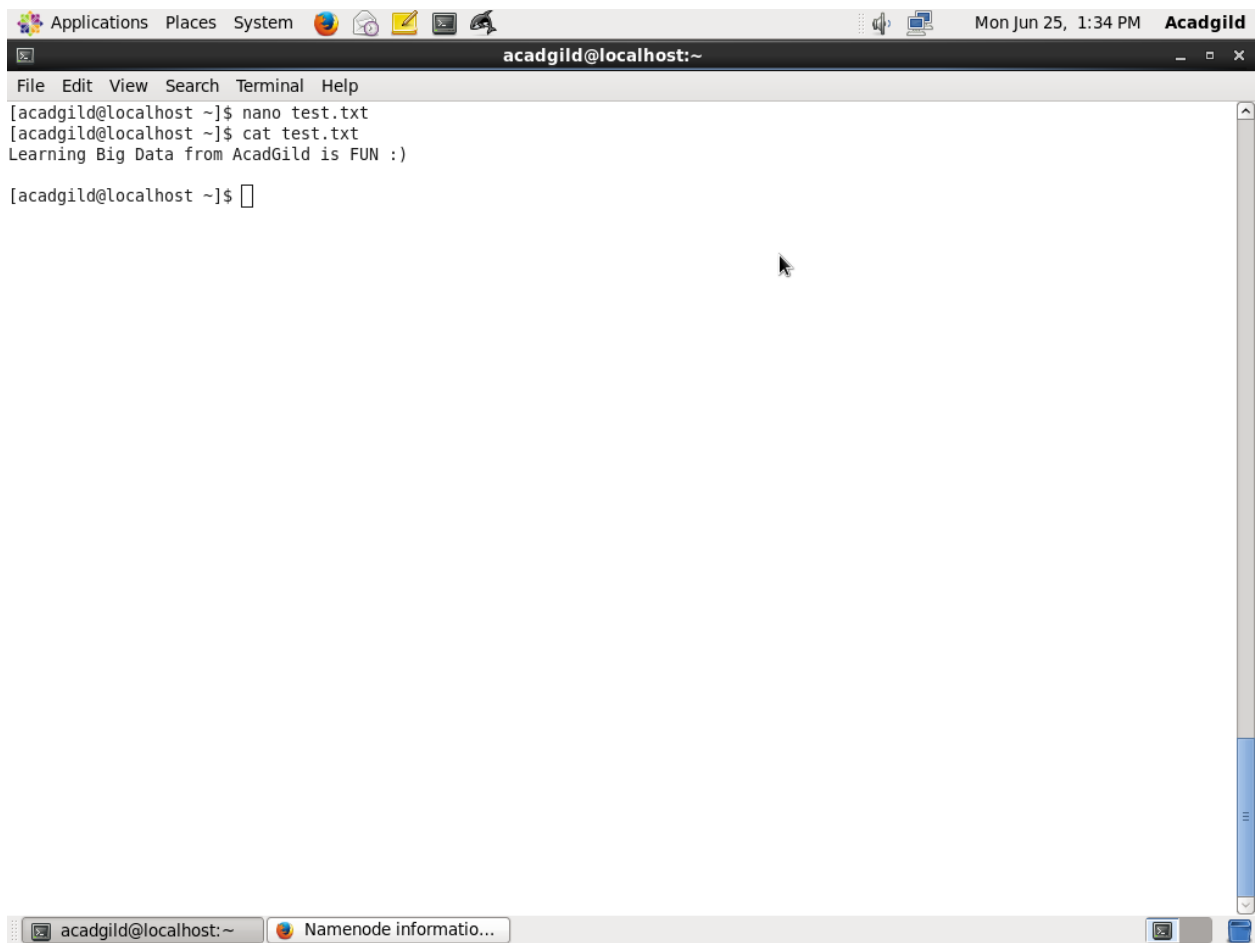
The screenshot shows a Linux terminal window titled "acadgild@localhost:~". The window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The terminal output shows the user running the "pwd" command, which returns "/home/acadgild". Then, the user runs "ls -ls", which lists the contents of the home directory with permissions, owner, group, size, date, and filename. The files listed are Desktop, Documents, Downloads, eclipse, eclipse-workspace, install, Music, Pictures, Public, Templates, and Videos. The window also shows a taskbar at the bottom with icons for the terminal and a file manager window titled "Namenode informatio...".

```
acadmild@localhost:~  
File Edit View Search Terminal Help  
[acadmild@localhost ~]$ pwd  
/home/acadmild  
[acadmild@localhost ~]$ ls -ls  
total 44  
4 drwxr-xr-x. 3 acadmild acadmild 4096 Jun 25 13:32 Desktop  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Feb 2 12:52 Documents  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Feb 13 14:24 Downloads  
4 drwxrwxr-x. 3 acadmild acadmild 4096 Dec 29 16:59 eclipse  
4 drwxrwxr-x. 3 acadmild acadmild 4096 Jan 16 14:02 eclipse-workspace  
4 drwxrwxr-x. 13 acadmild acadmild 4096 Feb 9 18:06 install  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 17:03 Music  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 17:03 Pictures  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 17:03 Public  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 17:03 Templates  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 17:03 Videos  
[acadmild@localhost ~]$
```

Sample Linux Commands



Creating a file test.txt using nano



Displaying the contents of test.txt

Applications Places System Mon Jun 25, 1:35 PM Acadgild

Namenode information - Mozilla Firefox

Hotspot > status x Namenode information x +

localhost:50070/dfshealth.html#tab-overview Search ☆ ⌵ ⌶ ⌵ ⌵ ⌵ ⌵

Hadoop Overview Datanodes Snapshot Startup Progress Utilities ▾

Overview 'localhost:8020' (active)

Started:	Mon Jun 25 18:55:02 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.

26 files and directories, 3 blocks = 29 total filesystem object(s).

Heap Memory used 30.34 MB of 96 MB Heap Memory. Max Heap Memory is 889 MB.

[acadgild@localhost: ~] Namenode informatio...

Accessing Hadoop over localhost

ApplicationsPlacesSystem

Mon Jun 25, 1:35 PMAcadgild

Namenode information - Mozilla Firefox

Hotspot > statusNamenode information

localhost:50070/dfshealth.html#tab-datanode

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Datanode Information

In operation

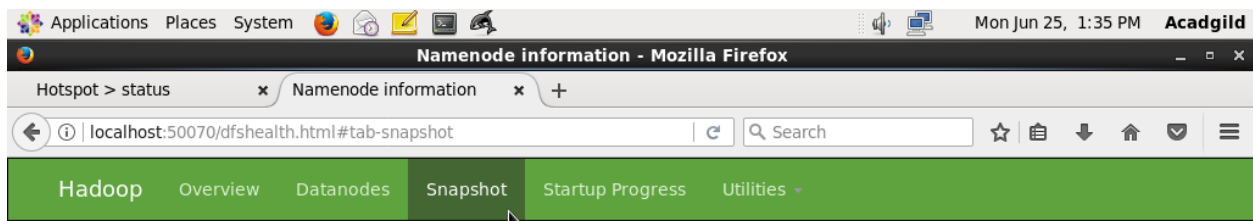
Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Volumes	Version
localhost (127.0.0.1:50010)	-19553	In Service	17.11 GB	200 KB	11.61 GB	5.5 GB	3	200 KB (0%)	0	2.6.5

Decomissioning

Node	Last contact	Under replicated blocks	Blocks with no live replicas	Under Replicated Blocks In files under construction
------	--------------	-------------------------	------------------------------	---

[acadgild@localhost: ~]Namenode informatio...

Datanode Information



Snapshot Summary

Snapshottable directories: 0

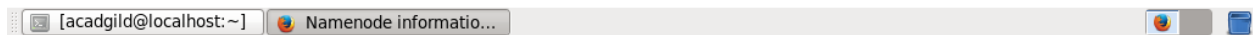
Path	Snapshot Number	Snapshot Quota	Modification Time	Permission	Owner	Group
------	-----------------	----------------	-------------------	------------	-------	-------

Snapshotted directories: 0

Snapshot ID	Snapshot Directory	Modification Time
-------------	--------------------	-------------------

Hadoop, 2016.

Legacy
UI



Snapshot Summary

ApplicationsPlacesSystem

Mon Jun 25, 1:35 PMAcadgild

Namenode information - Mozilla Firefox

Hotspot > statusNamenode information

localhost:50070/dfshealth.html#tab-startup-progress70%Search

HadoopOverviewDatanodesSnapshotStartup ProgressUtilities

Startup Progress

Elapsed Time: 32 sec, Percent Complete: 100%






Phase	Completion	Elapsed Time
Loading fsimage /home/acadgild/install/data/dfs/name/current/fsimage_0000000000000000332 2.07 KB	100%	0 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_0000000000000000333-0000000000000000333 1 MB (1/1)	100%	
Saving checkpoint	100%	0 sec
inodes /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_0000000000000000333 (0/0)	100%	
delegation tokens /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_0000000000000000333 (0/0)	100%	
cache pools /home/acadgild/install/data/dfs/name/current/fsimage.ckpt_0000000000000000333 (0/0)	100%	
Safe mode	100%	31 sec
awaiting reported blocks (3/3)	100%	

Hadoop, 2016.

Legacy UI

[acadgild@localhost: ~]Namenode informatio...

Startup Progress

Applications Places System      Mon Jun 25, 1:35 PM Acadgild

Directory: /logs/ - Mozilla Firefox

Hotspot > status x Directory: /logs/ x +

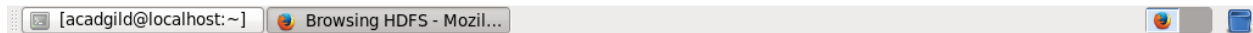
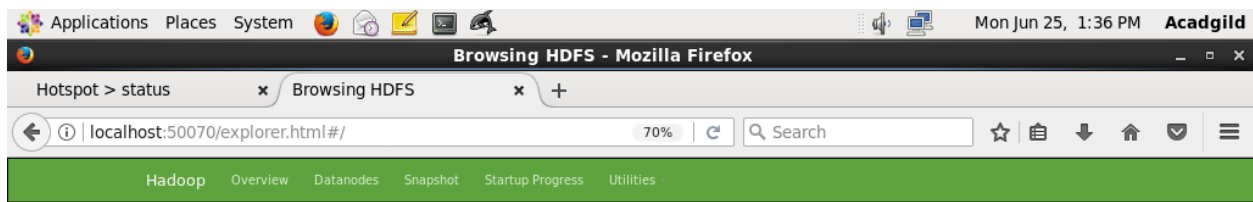
localhost:50070/logs/ 70% Search ☆ ⌵ ⌶ ⌵ ⌵ ⌵ ⌵

Directory: /logs/

SecurityAuth-acadgild.audit	0 bytes	Feb 2, 2018 12:44:52 PM
hadoop-acadgild-datanode-localhost.localdomain.log	199109 bytes	Jun 25, 2018 6:55:06 PM
hadoop-acadgild-datanode-localhost.localdomain.out	719 bytes	Jun 25, 2018 6:54:26 PM
hadoop-acadgild-datanode-localhost.localdomain.out.1	719 bytes	Feb 12, 2018 3:07:07 PM
hadoop-acadgild-datanode-localhost.localdomain.out.2	719 bytes	Feb 9, 2018 2:45:15 PM
hadoop-acadgild-datanode-localhost.localdomain.out.3	719 bytes	Feb 9, 2018 10:52:39 AM
hadoop-acadgild-datanode-localhost.localdomain.out.4	720 bytes	Feb 2, 2018 12:44:57 PM
hadoop-acadgild-namenode-localhost.localdomain.log	296276 bytes	Jun 25, 2018 1:35:48 PM
hadoop-acadgild-namenode-localhost.localdomain.out	719 bytes	Jun 25, 2018 6:54:21 PM
hadoop-acadgild-namenode-localhost.localdomain.out.1	719 bytes	Feb 12, 2018 3:07:02 PM
hadoop-acadgild-namenode-localhost.localdomain.out.2	719 bytes	Feb 9, 2018 2:45:09 PM
hadoop-acadgild-namenode-localhost.localdomain.out.3	719 bytes	Feb 9, 2018 10:52:33 AM
hadoop-acadgild-namenode-localhost.localdomain.out.4	720 bytes	Feb 2, 2018 12:44:52 PM
hadoop-acadgild-secondarynamenode-localhost.localdomain.log	136795 bytes	Jun 25, 2018 6:56:12 PM
hadoop-acadgild-secondarynamenode-localhost.localdomain.out	719 bytes	Jun 25, 2018 6:55:29 PM
hadoop-acadgild-secondarynamenode-localhost.localdomain.out.1	719 bytes	Feb 12, 2018 3:07:16 PM
hadoop-acadgild-secondarynamenode-localhost.localdomain.out.2	719 bytes	Feb 9, 2018 2:45:23 PM
hadoop-acadgild-secondarynamenode-localhost.localdomain.out.3	719 bytes	Feb 9, 2018 10:52:49 AM
hadoop-acadgild-secondarynamenode-localhost.localdomain.out.4	720 bytes	Feb 2, 2018 12:45:05 PM
mapred-acadgild-historyserver-localhost.localdomain.log	61440 bytes	Feb 13, 2018 2:30:51 PM
mapred-acadgild-historyserver-localhost.localdomain.out	0 bytes	Feb 13, 2018 2:30:17 PM
mapred-acadgild-historyserver-localhost.localdomain.out.1	0 bytes	Feb 9, 2018 10:54:14 AM
mapred-acadgild-historyserver-localhost.localdomain.out.2	0 bytes	Feb 2, 2018 12:48:35 PM
userlogs/	4096 bytes	Jun 25, 2018 6:59:42 PM
yarn-acadgild-nodemanager-localhost.localdomain.log	214028 bytes	Jun 25, 2018 6:56:26 PM
yarn-acadgild-nodemanager-localhost.localdomain.out	701 bytes	Jun 25, 2018 6:55:37 PM
yarn-acadgild-nodemanager-localhost.localdomain.out.1	701 bytes	Feb 12, 2018 3:07:26 PM
yarn-acadgild-nodemanager-localhost.localdomain.out.2	701 bytes	Feb 9, 2018 2:45:34 PM
yarn-acadgild-nodemanager-localhost.localdomain.out.3	701 bytes	Feb 9, 2018 10:52:59 AM
yarn-acadgild-nodemanager-localhost.localdomain.out.4	702 bytes	Feb 2, 2018 12:45:15 PM
yarn-acadgild-resourcemanager-localhost.localdomain.log	227422 bytes	Jun 25, 2018 6:56:26 PM
yarn-acadgild-resourcemanager-localhost.localdomain.out	701 bytes	Jun 25, 2018 6:55:35 PM
yarn-acadgild-resourcemanager-localhost.localdomain.out.1	701 bytes	Feb 12, 2018 3:07:24 PM
yarn-acadgild-resourcemanager-localhost.localdomain.out.2	701 bytes	Feb 9, 2018 2:45:31 PM

[acadgild@localhost:~] Directory: /logs/ - Mozi...

Accessing Logs



Browsing HDFS from a browser

ApplicationsPlacesSystem

Mon Jun 25, 1:36 PMAcadgild

Namenode information - Mozilla Firefox

Hotspot > statusNamenode information

localhost:50070/dfshealth.html#tab-overview50%Search

Summary

Security is off.
SafeMode is off.
26 files and directories, 3 blocks = 29 total filesystem object(s).
Heap Memory used 33.92 MB of 90.5 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 45.08 MB of 45.88 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

Configured Capacity:	17.11 GB
DFS Used:	200 KB
Non DFS Used:	11.61 GB
DFS Remaining:	5.5 GB
DFS Used%:	0%
DFS Remaining%:	32.14%
Block Pool Used:	200 KB
Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
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	6/25/2018, 6:55:02 PM

NameNode Journal Status

Current transaction ID: 334

Journal Manager	State
FileJournalManager{root=/home/acadgild/install/data/dfs/name}	EditLogFileOutputStream{home/acadgild/install/data/dfs/name/current/edits_inprogress_0000000000000000334}

NameNode Storage

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

[acadgild@localhost: ~]Namenode informatio...

Namenode information