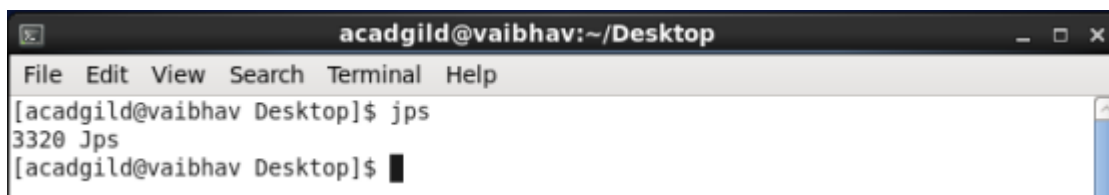


Start Hadoop single node on AcadGild VM. The command is *“start-all.sh.”*

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



```
acadgild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadgild@vaibhav Desktop]$ jps
3320 Jps
[acadgild@vaibhav Desktop]$
```

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.



```
acadgild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadgild@vaibhav Desktop]$ jps
3320 Jps
[acadgild@vaibhav Desktop]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/05/05 18:18:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-vaibhav.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-vaibhav.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondaryname
node-vaibhav.out
18/05/05 18:19:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-vaibhav.ou
t
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-vaibhav
.out
```

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.

```

Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadmild@vaibhav Desktop]$ jps
3320 Jps
[acadmild@vaibhav Desktop]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/05/05 18:18:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-namenode-vaibhav.out
localhost: starting datanode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-datanode-vaibhav.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadmild-secondaryname
node-vaibhav.out
18/05/05 18:19:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/yarn-acadmild-resourcemanager-vaibhav.ou
t
localhost: starting nodemanager, logging to /home/acadmild/install/hadoop/hadoop-2.6.5/logs/yarn-acadmild-nodemanager-vaibhav
.out
You have new mail in /var/spool/mail/acadmild
[acadmild@vaibhav Desktop]$ jps
3780 DataNode
4533 Jps
3686 NameNode
4139 ResourceManager
3996 SecondaryNameNode
4239 NodeManager
[acadmild@vaibhav Desktop]$

```

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

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

```

Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadmild@vaibhav Desktop]$ pwd
/home/acadmild/Desktop
[acadmild@vaibhav Desktop]$

```

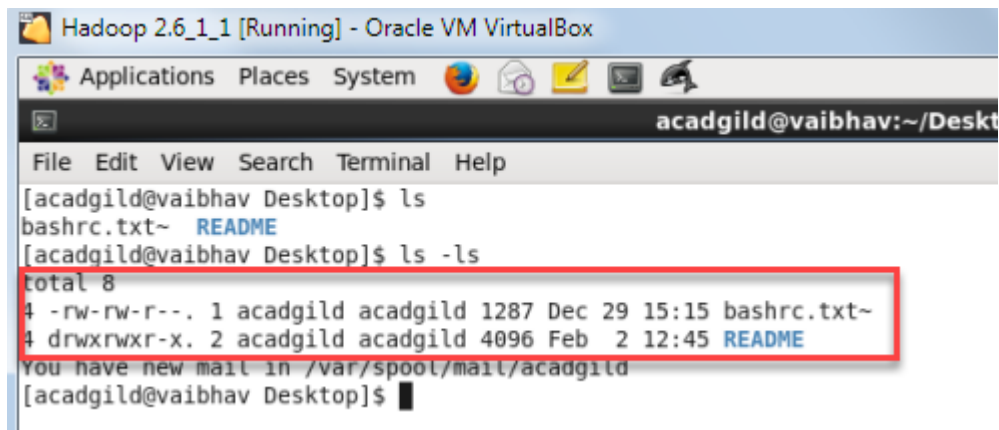
- ii) **ls**: Lists out all directories present in the current location

```

Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadmild@vaibhav Desktop]$ ls
bashrc.txt~ README
[acadmild@vaibhav Desktop]$

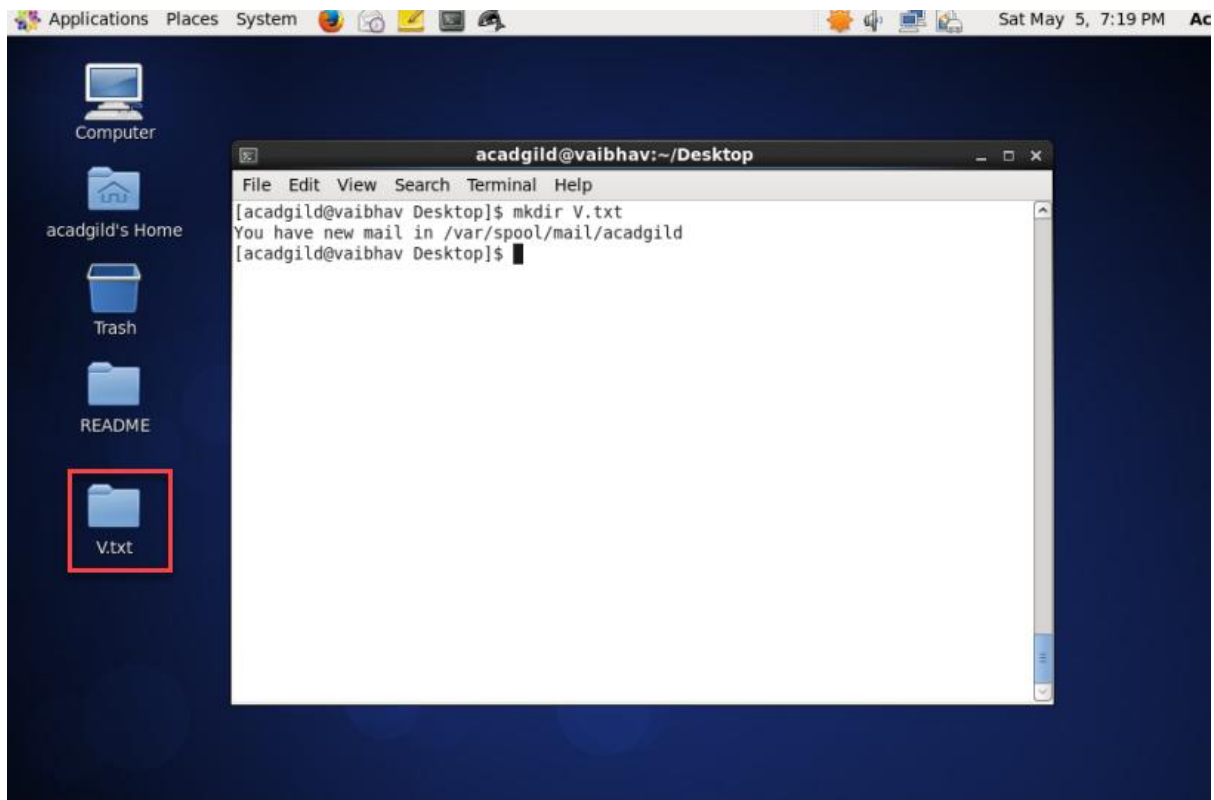
```

- 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.

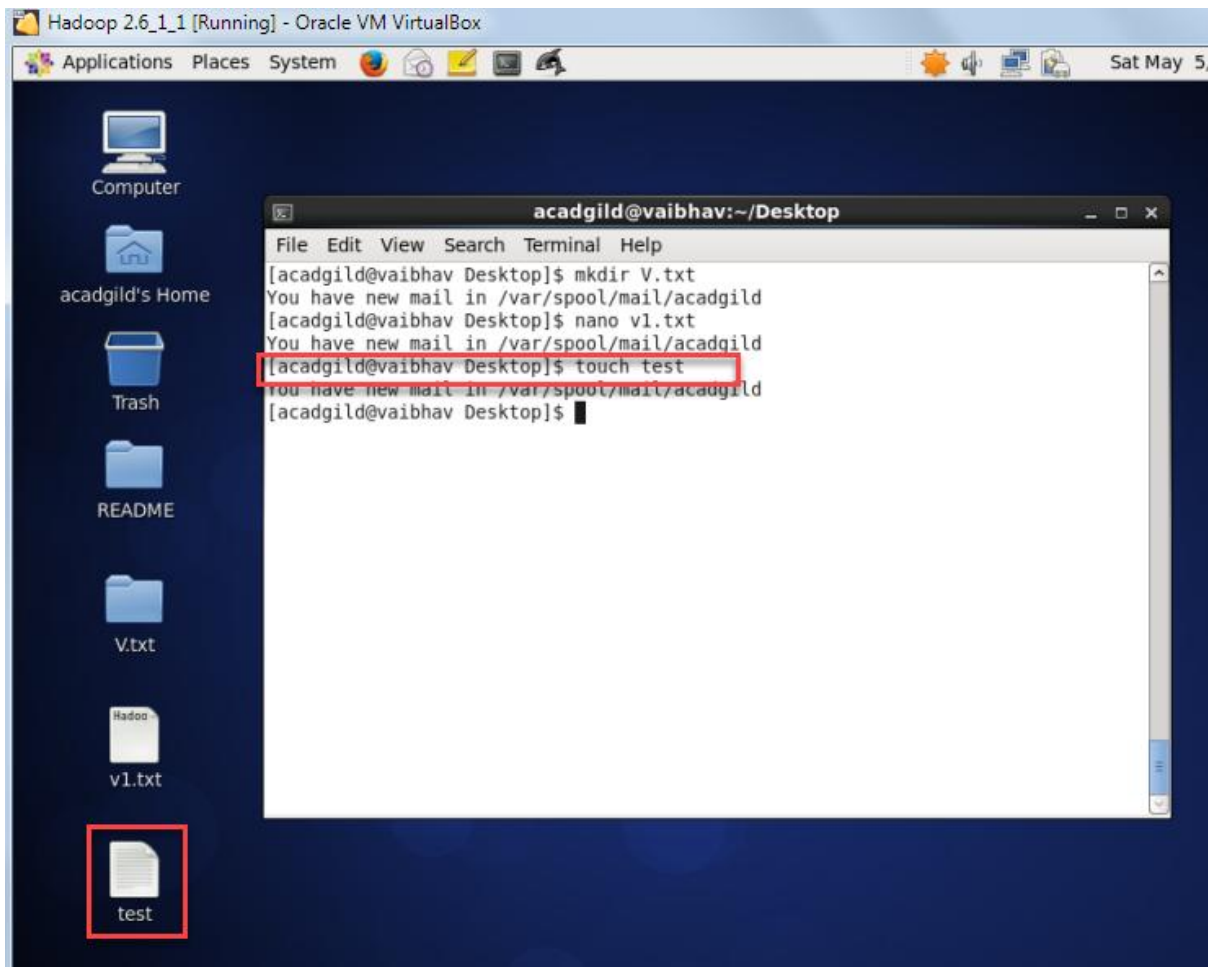


```
Hadoop 2.6_1_1 [Running] - Oracle VM VirtualBox
acadgild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadgild@vaibhav Desktop]$ ls
bashrc.txt~  README
[acadgild@vaibhav Desktop]$ ls -ls
total 8
4 -rw-rw-r--. 1 acadgild acadgild 1287 Dec 29 15:15 bashrc.txt~
4 drwxrwxr-x. 2 acadgild acadgild 4096 Feb  2 12:45  README
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav Desktop]$
```

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



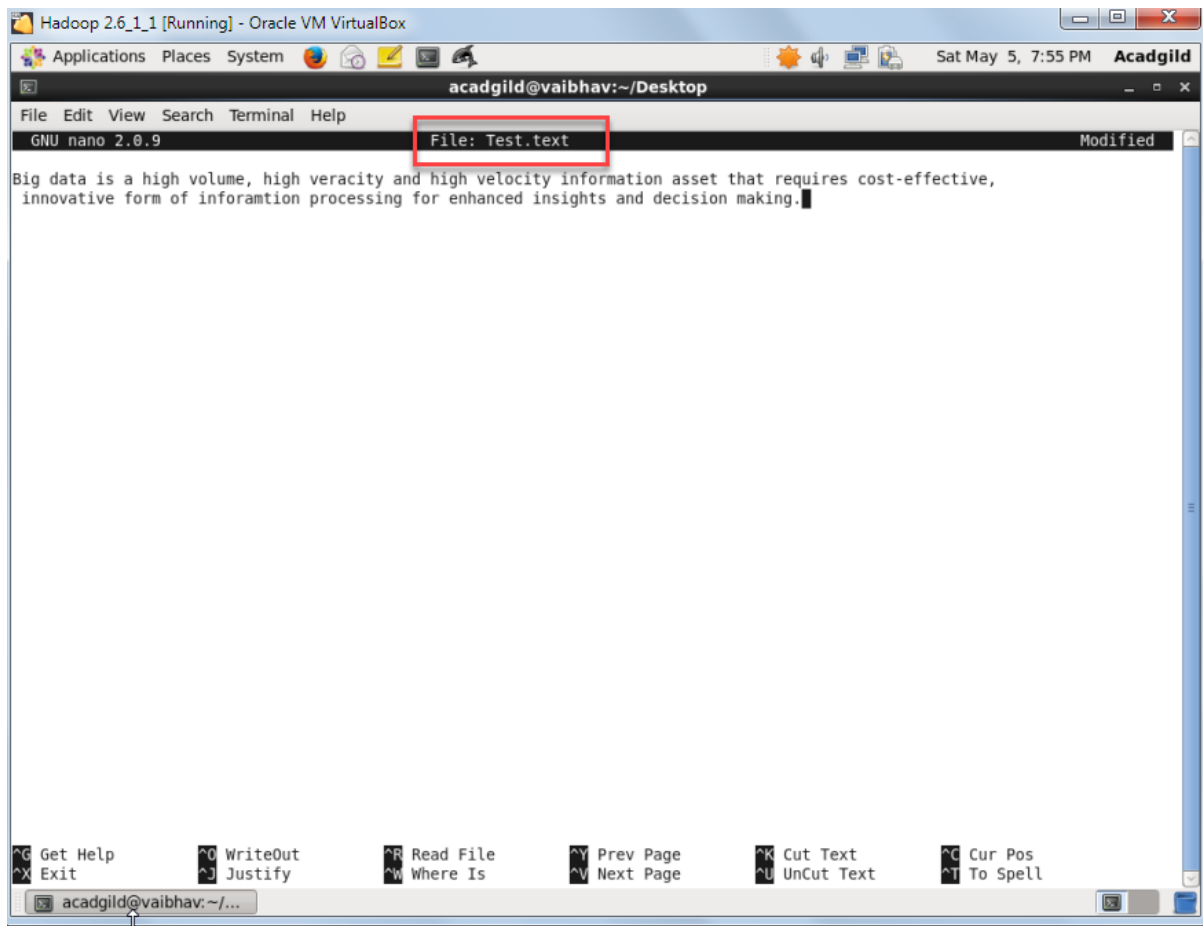
- v) **touch:** The command is used to create an empty file in the current location. Eg. Here the “test” named file is created in the current location.



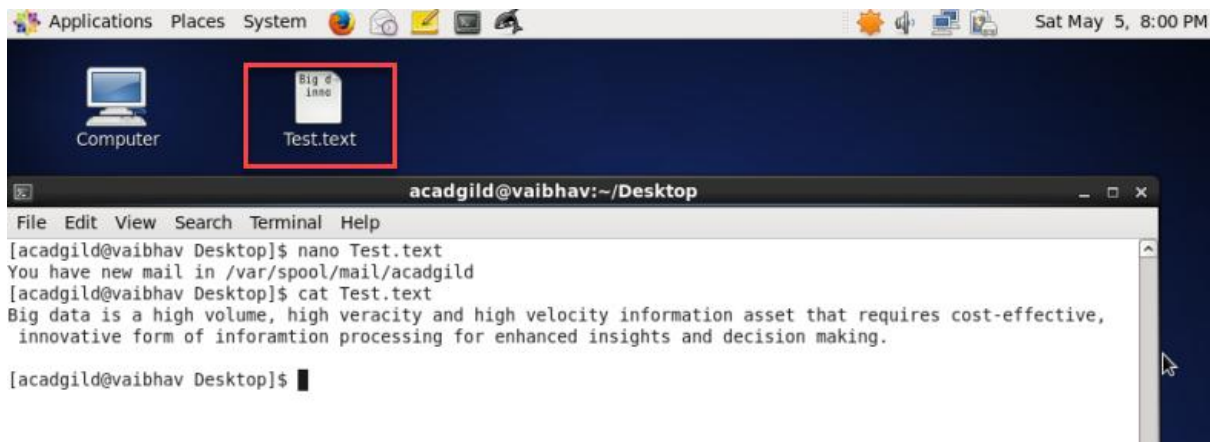
Create a file from the terminal using nano editor (example: nano test.txt), and add some content in it. Cat it to see if the content is saved.

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.

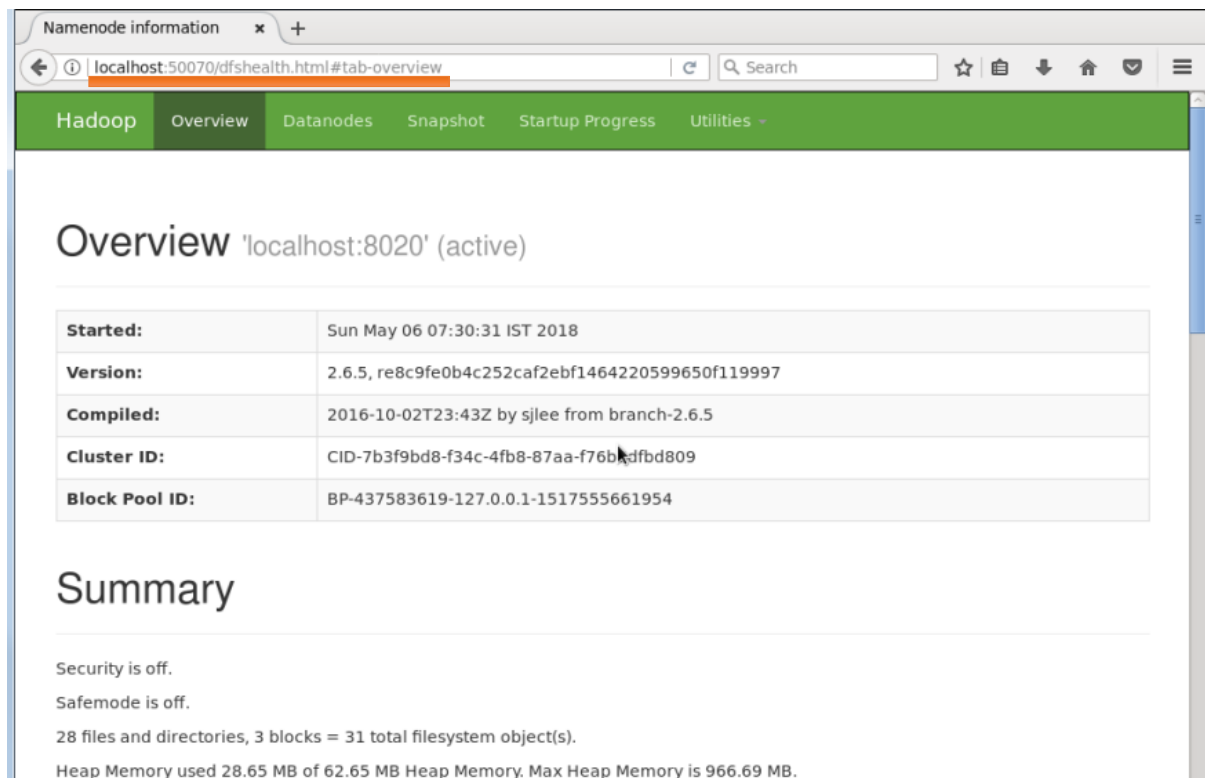


Open the hdfs web page by typing *localhost: 50070* in the browser. Check all the details of the HDFS.

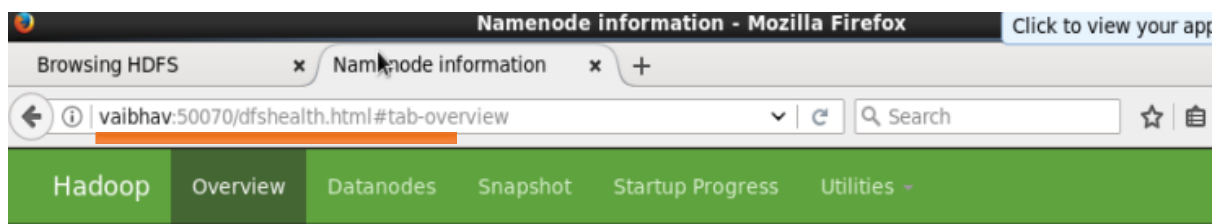
Apart from the command line interface. Hadoop provides Web UI interface to HDFS resource manager.

- 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

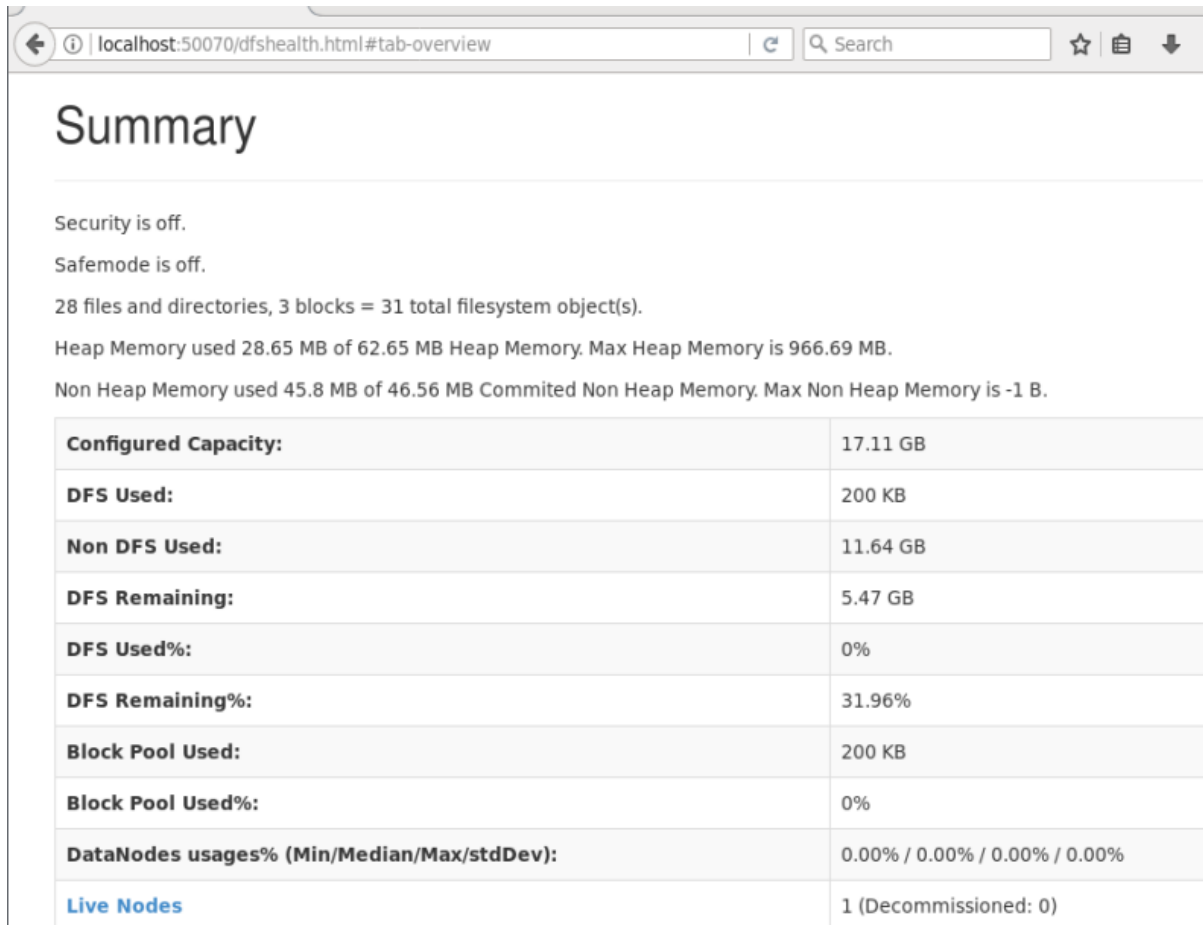
Security is off.

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:



The screenshot shows the Hadoop NameNode Overview tab in a web browser. The browser address bar shows the URL: localhost:50070/dfshealth.html#tab-overview. The page title is "Summary". Below the title, there is a section with status information: "Security is off.", "Safemode is off.", "28 files and directories, 3 blocks = 31 total filesystem object(s).", "Heap Memory used 28.65 MB of 62.65 MB Heap Memory. Max Heap Memory is 966.69 MB.", and "Non Heap Memory used 45.8 MB of 46.56 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.". Below this section is a table with HDFS metrics.

| Metric | Value |
|--------------------------------------------|-------------------------------|
| Configured Capacity: | 17.11 GB |
| DFS Used: | 200 KB |
| Non DFS Used: | 11.64 GB |
| DFS Remaining: | 5.47 GB |
| DFS Used%: | 0% |
| DFS Remaining%: | 31.96% |
| 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) |

- Namenode information and storage status is also listed.

NameNode Journal Status

Current transaction ID: 349

| Journal Manager | State |
|---------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| FileJournalManager(root=/home/acadgild/install/data/dfs/name) | EditLogFileOutputStream(/home/acadgild/install/data/dfs/name/current/edits_inprogress_00000000000000000349) |

NameNode Storage

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

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

The screenshot shows the Hadoop NameNode web interface with the 'Datanodes' tab selected. The page title is 'Datanode Information'. Below the title, there is a section for 'In operation' which contains a table of datanode details. The table has columns for Node, Last contact, Admin State, Capacity, Used, Non DFS Used, Remaining, Blocks, Block pool used, Failed Volumes, and Version. One datanode is listed: vaibhav (127.0.0.1:50010) with a capacity of 17.11 GB, 200 KB used, 11.64 GB non-DFS used, 5.47 GB remaining, 3 blocks, 200 KB block pool used (0%), 0 failed volumes, and version 2.6.5. Below this is a section for 'Decommissioning' which is currently empty.

| Node | Last contact | Admin State | Capacity | Used | Non DFS Used | Remaining | Blocks | Block pool used | Failed Volumes | Version |
|---------------------------|--------------|-------------|----------|--------|--------------|-----------|--------|-----------------|----------------|---------|
| vaibhav (127.0.0.1:50010) | 0 | In Service | 17.11 GB | 200 KB | 11.64 GB | 5.47 GB | 3 | 200 KB (0%) | 0 | 2.6.5 |

- 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.

The screenshot shows the Hadoop NameNode web interface with the 'Browse Directory' tab selected. The page title is 'Browse Directory'. Below the title, there is a search bar with a 'Go!' button. Below the search bar, there is a table of directory details. The table has columns for Permission, Owner, Group, Size, Replication, Block Size, and Name. Three directories are listed: sqoopout111, tmp, and user, all owned by acadgild and supergroup, with a size of 0 B and 0 replication.

| Permission | Owner | Group | Size | Replication | Block Size | Name |
|------------|----------|------------|------|-------------|------------|-------------|
| drwxr-xr-x | acadgild | supergroup | 0 B | 0 | 0 B | sqoopout111 |
| drwxrwx--- | acadgild | supergroup | 0 B | 0 | 0 B | tmp |
| drwxr-xr-x | acadgild | supergroup | 0 B | 0 | 0 B | user |

Hadoop, 2016.