

ASSIGNMENT 1

BIG DATA & HADOOP – INTRODUCTION

1. Start Hadoop single node on AcadGild VM

Command: start-all.sh

Starts the Hadoop DFS daemons, NameNode, DataNode, ResourceManager, NodeManager.
The below screenshot shows the daemons starting.

```
[acadgild@localhost ~]$ start-all.sh
This script is deprecated. Instead use start-dfs.sh and start-yarn.sh
18/07/29 11:42:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.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-secondarynamenode-localhost.localdomain.out
18/07/29 11:43:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
```

2. Run a JPS command

Command: jps

JPS (Java Virtual Machine Process Status Tool) is a command used to check all the Hadoop daemons like NameNode, DataNode, ResourceManager, NodeManager etc. which are running on the machine.

The below screenshot shows the output of **jps** command.

```
[acadgild@localhost ~]$ jps
6112 ResourceManager
5650 NameNode
5747 DataNode
5943 SecondaryNameNode
6216 NodeManager
6568 Jps
```

3. Unix commands

Command: pwd

The **pwd** command displays the full pathname of the current directory.

```
[acadgild@localhost ~]$ pwd
/home/acadgild
```

Command: ls

The ls command is used to list all the files in the current directory.

The “-l” option is used for long listing.

```
[acadgild@localhost ~]$ ls -l
total 44
drwxr-xr-x.  3 acadgild acadgild 4096 Feb  2 12:51 Desktop
drwxr-xr-x.  2 acadgild acadgild 4096 Feb  2 12:52 Documents
drwxr-xr-x.  2 acadgild acadgild 4096 Feb 13 14:24 Downloads
drwxrwxr-x.  3 acadgild acadgild 4096 Dec 29 2017 eclipse
drwxrwxr-x.  3 acadgild acadgild 4096 Jan 16 2018 eclipse-workspace
drwxrwxr-x. 13 acadgild acadgild 4096 Feb  9 18:06 install
drwxr-xr-x.  2 acadgild acadgild 4096 Dec 27 2017 Music
drwxr-xr-x.  2 acadgild acadgild 4096 Dec 27 2017 Pictures
drwxr-xr-x.  2 acadgild acadgild 4096 Dec 27 2017 Public
drwxr-xr-x.  2 acadgild acadgild 4096 Dec 27 2017 Templates
drwxr-xr-x.  2 acadgild acadgild 4096 Dec 27 2017 Videos
```

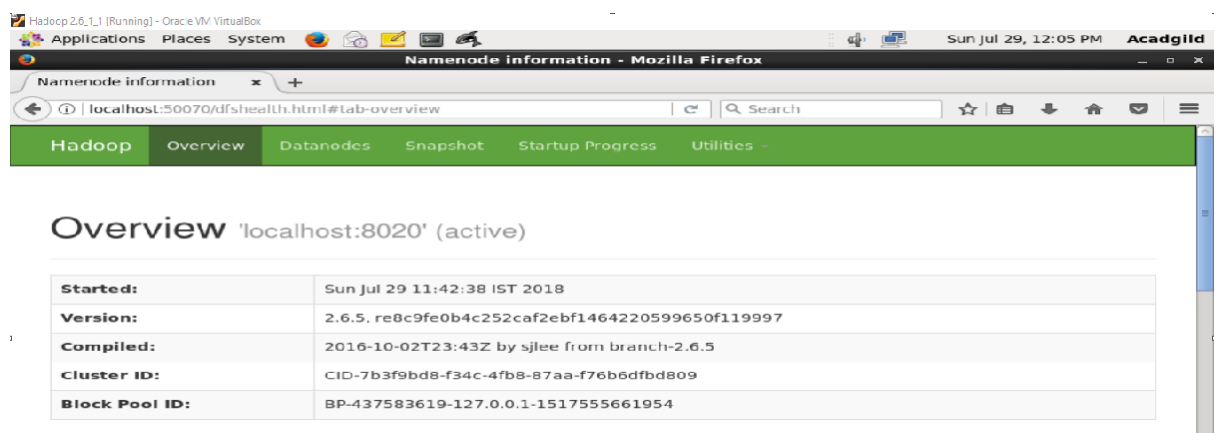
4. Create a file using Nano editor and view the contents of file

Nano is a text editor for Unix-like systems or any other environments using a command line interface. After creating a file, to view the contents of the file use **cat** command.

```
[acadgild@localhost ~]$ nano
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ cat test.txt
Welcome to Acadgild!
[acadgild@localhost ~]$
```

5. Open HDFS webpage

Enter the URL as <http://localhost:50070/> in web browser to open the HDFS web UI.



The screenshot shows a web browser window with the title "Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox". The address bar shows the URL "localhost:50070/dfshealth.html#tab-overview". The page content includes a navigation bar with tabs: "Hadoop", "Overview", "Datanodes", "Snapshot", "Startup Progress", and "Utilities". The "Overview" tab is selected, showing the "Overview 'localhost:8020' (active)" section. Below this, there is a table with the following information:

Started:	Sun Jul 29 11:42:38 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