BIG DATA ENGINEERING WITH HADOOP & SPARK

Assignment 1

Task 1:

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

```
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/06/21 13:25:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 3111. Stop it first.
localhost: datanode running as process 3211. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 3434. Stop it first.
18/06/21 13:25:37 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 3584. Stop it first.
localhost: nodemanager running as process 3688. Stop it first.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
[acadgild@localhost ~]$
```

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

```
[acadgild@localhost ~]$
[acadgild@localhost ~]$
[acadgild@localhost ~]$ jps
3584 ResourceManager
9618 Jps
3111 NameNode
3688 NodeManager
3434 SecondaryNameNode
3211 DataNode
[acadgild@localhost ~]$
```

3. Run a couple of Unix commands like pwd, ls -ls etc.

```
[acadgild@localhost ~]$
[acadgild@localhost ~]$ pwd
/home/acadgild
[acadgild@localhost ~]$ ls -ltr
total 48
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   68 Jun 21 10:58 test.txt
[acadgild@localhost ~]$ ls -a
.               .cache        eclipse-workspace  .hivehistory      .nautilus         .scala_history  .viminfo
..              .config       .esd_auth          .ICEauthority     .oracle_jre_usage  .scalaide       .webclipse
.abrt           .dbus         .fontconfig        .imsettings.log   .p2               .ssh            .webclipse.properties
.bash_history   Desktop       .gconf             install           Pictures          .swt            .Xauthority
.bash_logout    .dmrc         .gnome2            .local            .pig_history      Templates       .xsession-errors
.bash_profile   Documents     .gnote             .m2               .pki              test.txt        .xsession-errors.old
.bashrc         Downloads     .gnupg             .mozilla          Public            .thumbnails
.bashrc~        eclipse       .gtk-bookmarks     Music             .pulse            .tooling
.beeline        .eclipse      .gvfs              .mysql_history    .pulse-cookie     Videos
[acadgild@localhost ~]$
```
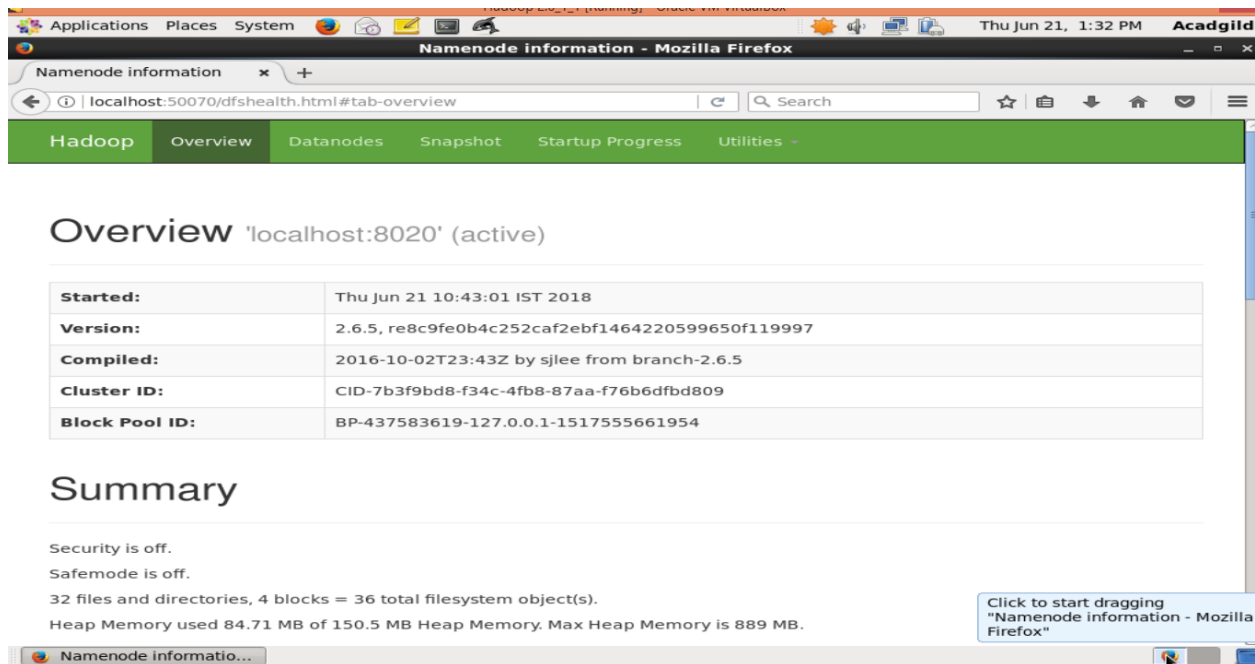
4. Create a file from the terminal using nano editor (for example nano test.txt) and put some content in it. Cat it to see if the content is saved.

```
[acadgild@localhost ~]$ ls
Desktop  Documents  Downloads  eclipse  eclipse-workspace  install  Music  Pictures  Public  Templates  test.txt  Videos
[acadgild@localhost ~]$ cat test.txt
This is a test file created for the exercise.
To check the output of the file, need to do a cat on the file.

[acadgild@localhost ~]$
```

Task 2:

1. Check whether /user/acadgild directory exists or not in the HDFS.

If it doesn't exist, then create this.

Create a directory /user/acadgild/hadoop.

2. Create a file in HDFS under directory /user/acadgild/hadoop, with name word-count.txt.

Whatever we type on screen should get appended to the file. Try to type (on screen) few lines

from any online article or textbook.

```
[acadgild@localhost ~]$ echo "Hello World" | hadoop fs -appendToFile - /user/acadgild/hadoop/word-count.txt
18/06/21 13:53:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
[acadgild@localhost ~]$ echo "Hello World" | hadoop fs -appendToFile^[[2~^C /user/acadgild/hadoop/word-count.txt
[acadgild@localhost ~]$
[acadgild@localhost ~]$
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoop/word-count.txt
18/06/21 13:53:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable

Hello World
[acadgild@localhost ~]$
```

3. Create a file max-temp.txt in local FS. Put some 10-15 records of date and temperature

example: dd-mm-yyyy,temperature

Example:

10-01-1990,10

10-02-1991,20

Move this file to HDFS at /user/acadgild/hadoop.

```
[acadgild@localhost ~]$
[acadgild@localhost ~]$
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ pwd
/home/acadgild
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/max-temp.txt /user/acadgild/hadoop
18/06/21 13:57:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
[acadgild@localhost ~]$
[acadgild@localhost ~]$ hdfs -dfs -ls /user/acadgild/hadoop
Unrecognized option: -dfs
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop
18/06/21 13:57:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup         28 2018-06-21 13:57 /user/acadgild/hadoop/max-temp.txt
-rw-r--r--   1 acadgild supergroup         13 2018-06-21 13:53 /user/acadgild/hadoop/word-count.txt
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/max-temp.txt
18/06/21 13:58:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
cat: `/user/acadgild/max-temp.txt': No such file or directory
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoop/max-temp.txt
18/06/21 13:58:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
10-01-1990,10
10-02-1991,20
[acadgild@localhost ~]$
```

4. Change the permission of the file /user/acadgild/hadoop/max-temp.txt, such that only the

owner and the group members have full control over the file.

Others do not have any control over it.

```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop/max-temp.txt
18/06/21 14:00:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
-rw-r--r--   1 acadgild supergroup         28 2018-06-21 13:57 /user/acadgild/hadoop/max-temp.txt
[acadgild@localhost ~]$
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -chmod 770 /user/acadgild/hadoop/max-temp.txt
18/06/21 14:06:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop/max-temp.txt
18/06/21 14:07:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
 classes where applicable
-rwxrwx---   1 acadgild supergroup         28 2018-06-21 13:57 /user/acadgild/hadoop/max-temp.txt
[acadgild@localhost ~]$
```