

Session 2 - HDFS - Introduction

Assignment 1

Task 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.

```
acadmild@localhost:~$ hadoop fs -ls /user
1
18/03/13 22:29:59 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user/hive
2
[acadmild@localhost ~]$ hadoop fs -mkdir /user/acadmild
3
18/03/13 22:30:18 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[acadmild@localhost ~]$ hadoop fs -ls /user
4
18/03/13 22:30:30 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - acadgild supergroup 0 2018-03-13 22:30 /user/acadmild
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user/hive
5
[acadmild@localhost ~]$ hadoop fs -mkdir /user/acadmild/hadoop
6
18/03/13 22:30:40 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[acadmild@localhost ~]$ hadoop fs -ls /user/acadmild
7
18/03/13 22:30:52 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x - acadgild supergroup 0 2018-03-13 22:30 /user/acadmild/ha
doo
8
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$
```

With reference to the screenshot above,

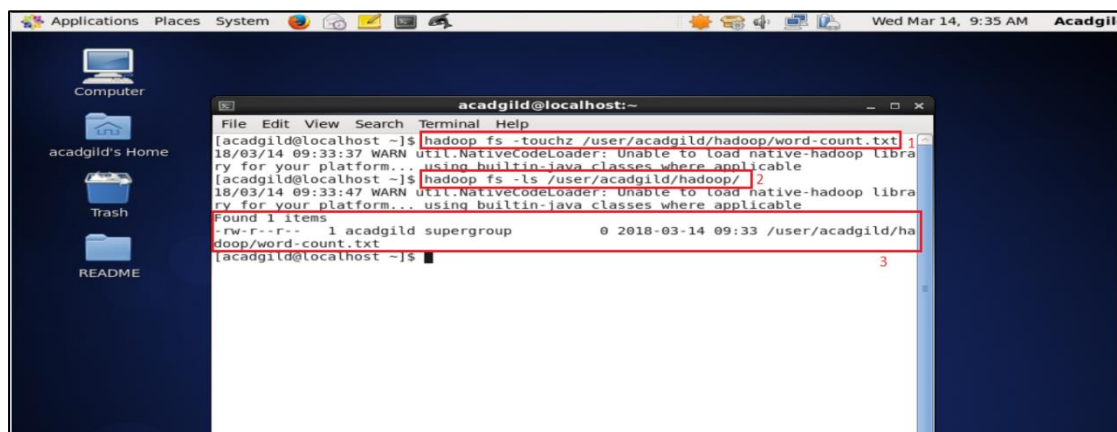
- 1 : Listed the contents of the '/user' directory.
- 2 : '/user' contained only one folder and did not have 'acadgild' directory.
- 3 : created directory 'acadgild' using following command :
hadoop fs -mkdir /user/acadmild
- 4 : Listed the contents of '/user' folder again
- 5 : '/user/acadmild' directory is created.
- 6 : created directory 'hadoop' using following command :
hadoop fs -mkdir /user/acadmild/hadoop
- 7 : listed the contents of the '/user/acadmild' directory
- 8 : '/user/acadmild/hadoop' directory is created.

Task 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.

Creation of the file :



```
acadgild@localhost:~$ hadoop fs -touchz /user/acadgild/hadoop/word-count.txt
18/03/14 09:33:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
acadgild@localhost:~$ hadoop fs -ls /user/acadgild/hadoop/
18/03/14 09:33:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadgild supergroup          0 2018-03-14 09:33 /user/acadgild/hadoop/word-count.txt
acadgild@localhost:~$
```

With reference to the screenshot above,

1 : Created a file in hdfs using

`hadoop fs -touchz /user/acadgild/hadoop/word-count.txt`

2 : Listing the contents of '/user/acadgild/hadoop' directory

3 : The file id created in the directory.

Appending the entered text to the file :

```
acadmild@localhost:~$ echo "Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs." | hadoop fs -appendToFile - /user/acadmild/hadoop/word-count.txt
18/03/14 10:39:29 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/acadmild
acadmild@localhost:~$ echo "Big Data and Hadoop" | hadoop fs -appendToFile - /user/acadmild/hadoop/word-count.txt
18/03/14 10:39:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
acadmild@localhost:~$ hadoop fs -ls /user/acadmild/hadoop/
18/03/14 10:39:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 acadmild supergroup 289 2018-03-14 10:39 /user/acadmild/hadoop/word-count.txt
acadmild@localhost:~$ (hadoop fs -cat /user/acadmild/hadoop/word-count.txt)
18/03/14 10:40:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Hadoop is an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs.
Big Data and Hadoop
acadmild@localhost:~$
```

With reference to the screenshot above,

1 : appending text to the file using the following command.

```
echo "---any text---" | hadoop fs -appendToFile - /user/acadmild/hadoop/word-count.txt
```

2 : Adding one more line to the file in the same manner.

3 : When the directory contents are listed, the length of the file changes from 0(refer previous screenshot) to 289 indicating the text is appended to the file.

4 : using cat command to view the contents of the file.

5 : The text passed above is displayed. The text added by adding 2nd time is appended after the text initially added.

Task 3 : Create a file max-temp.txt in local FS.

Put some 10-15 records of date and temperature example:

dd-mm-yyyy,temperature

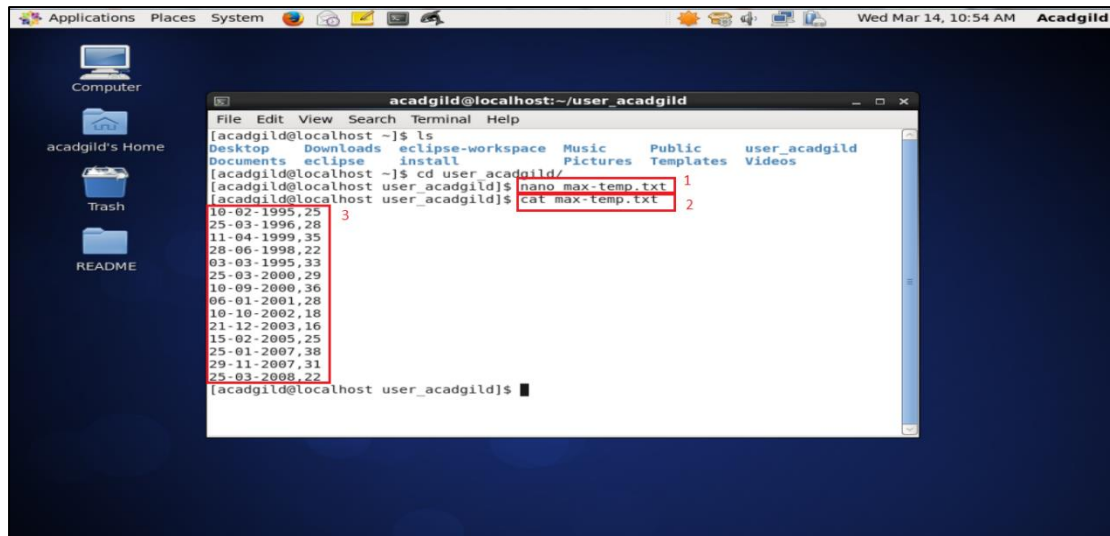
Example:

10-01-1990,10

10-02-1991,20

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

Creation and population of file in local FS :



```
acadgild@localhost:~/user_acadgild
File Edit View Search Terminal Help
[acadgild@localhost ~]$ ls
Desktop  Downloads  eclipse-workspace  Music  Public  user_acadgild
Documents  eclipse  install  Pictures  Templates  Videos
[acadgild@localhost ~]$ cd user acadgild/
[acadgild@localhost user_acadgild]$ nano max-temp.txt
[acadgild@localhost user_acadgild]$ cat max-temp.txt
10-02-1995,25
25-03-1996,28
11-04-1999,35
28-06-1998,22
03-03-1995,33
25-03-2000,29
10-09-2000,36
06-01-2001,28
10-10-2002,18
21-12-2003,16
15-02-2005,25
25-01-2007,38
29-11-2007,31
25-03-2008,22
[acadgild@localhost user_acadgild]$
```

With reference to the screenshot above,

- 1 : Creating the file locally (without using hadoop fs commands) using nano. This opens up the nano editor, I added few entries to the file.
- 2 : After saving the file, ran a cat command to view its contents.
- 3 : The rows added are shown.

Moving the file to HDFS :

```
acadmild@localhost:~/user_acadmild
File Edit View Search Terminal Help
[acadmild@localhost user_acadmild]$ hadoop fs -put max-temp.txt /user/acadmild/hadoop
18/03/14 10:56:31 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/acadmild
[acadmild@localhost user_acadmild]$ hadoop fs -ls /user/acadmild/hadoop/
18/03/14 10:56:58 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 196 2018-03-14 10:56 /user/acadmild/hadoop/max-temp.txt
-rw-r--r-- 1 acadgild supergroup 289 2018-03-14 10:39 /user/acadmild/hadoop/word-count.txt
[acadmild@localhost user_acadmild]$ hadoop fs -cat /user/acadmild/hadoop/max-temp.txt
18/03/14 10:57:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
10-02-1995,25
25-03-1996,28
11-04-1999,35
28-06-1998,22
03-03-1995,33
25-03-2000,29
10-09-2000,36
06-01-2001,28
10-10-2002,18
21-12-2003,16
15-02-2005,25
25-01-2007,38
29-11-2007,31
25-03-2008,22
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost user_acadmild]$
```

With reference to the screenshot above,

1 : Hadoop put command is used to move files from local file system to hdfs.

`hadoop fs -put <location in local> <location on hdfs>`

`Hadoop fs -put max-temp.txt /user/acadmild/hadoop`

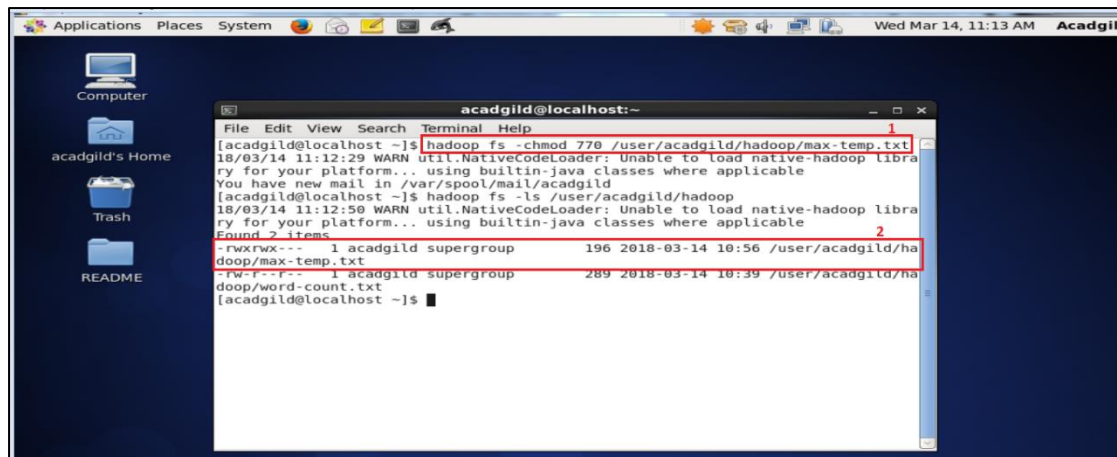
2 : When listed the contents of the specific folder in hdfs, the max-temp.txt file is shown.

3 : On performing a cat on the file on hdfs, the contents are displayed.

Task 4 : Change the permission of the file /user/acadmild/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.



With reference to the screenshot above,

1 : For changing the permissions on a file, chmod command is used.

chmod is paired with a 3 digit numeric notation.

1st digit for the owner of the file, 2nd for the group and 3rd for everyone else.

Each digit is calculated as follows.

Each of the 3 entities : owner , group and others have read(r), write(w) and execute(x) permissions on the file.

These permissions have a particular number assigned to them.

r = 4

w = 2

x = 1

So give full permissions (r+w+x) to an entity, we must assign (4+2+1 = 7) number to that particular entity.

So to give full permissions to the owner and the group, and to restrict any access by others, we assign :

$(4+2+1)(4+2+1)(0+0+0) = 770$

So the command sums up to be :

`hadoop fs - chmod 770 /user/acadgild/hadoop/max-temp.txt`

2 : When listed, the permissions of this file show the difference : 'rwxrwx---'