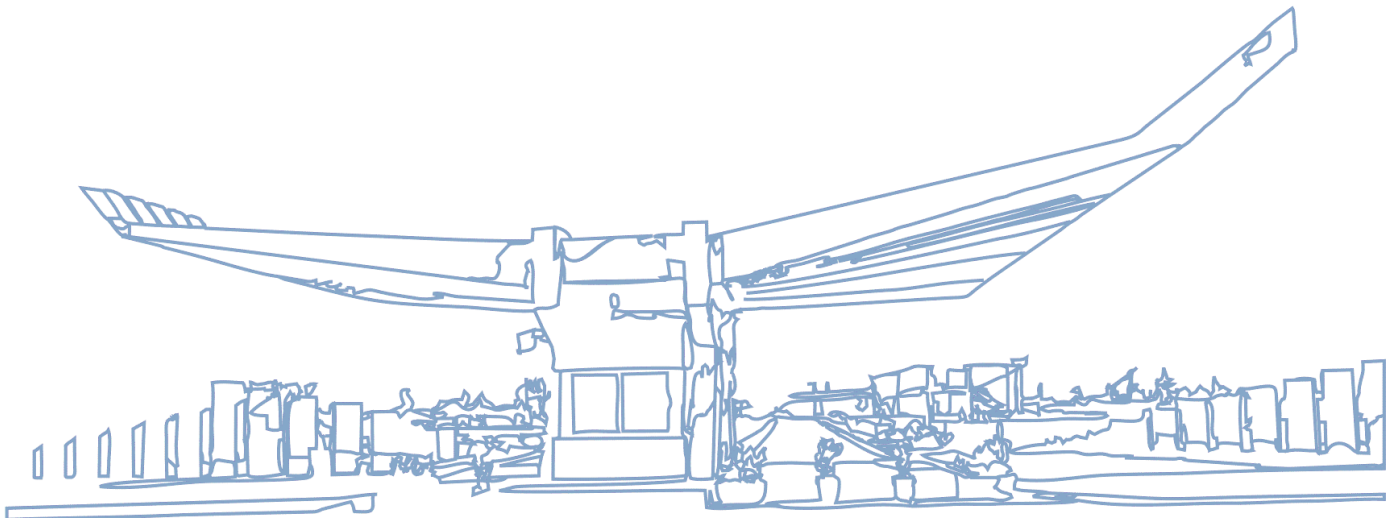


CEN 571 – Data Mining

Assignment 02 – Question 2



PREPARED:
Baftjar TABAKU

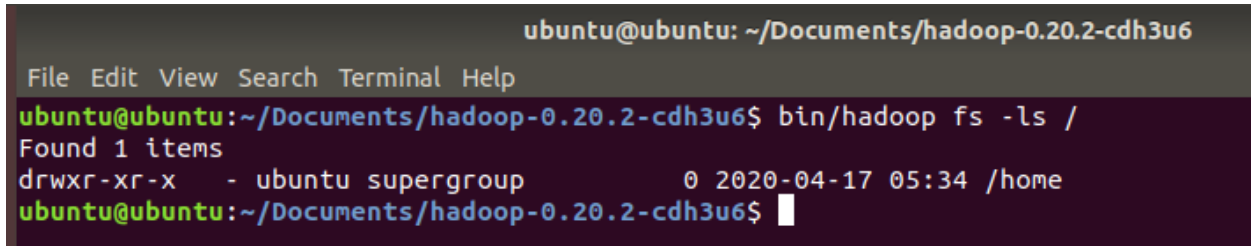
19.04.2020
Epoka University
Tirana, ALBANIA

ACCEPTED:
Prof.Dr. Arben Asllani

Question 2

1. Listing the directory contents.

By running the command 'Hadoop fs -ls' and all the files are displayed according to the respective directory as shown in the following screenshot.

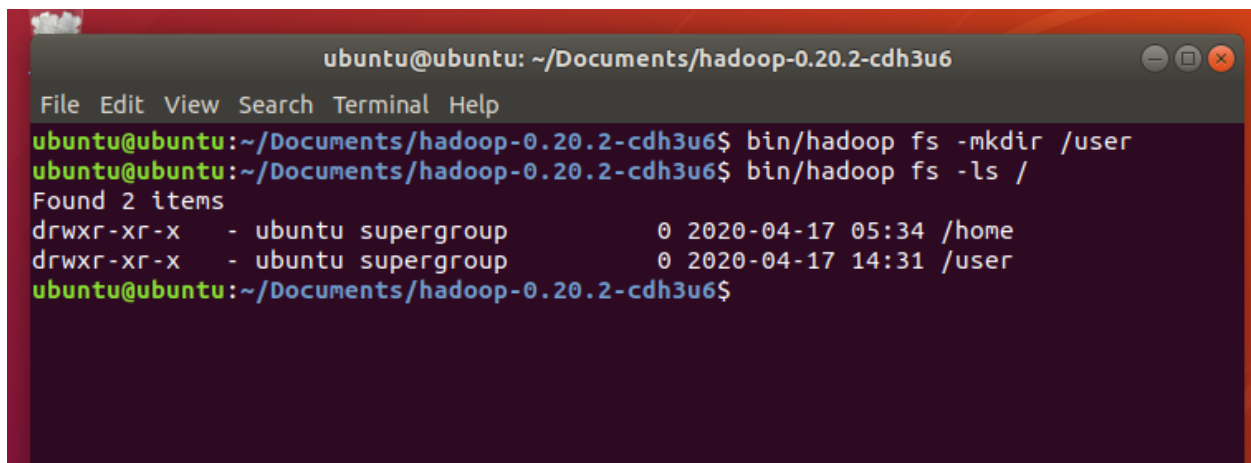


```

ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 1 items
drwxr-xr-x  - ubuntu supergroup          0 2020-04-17 05:34 /home
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

2. Creating a Directory, in Hadoop as the screenshot shows first by running the command to create a user directory named 'user', as it is shown listed above.

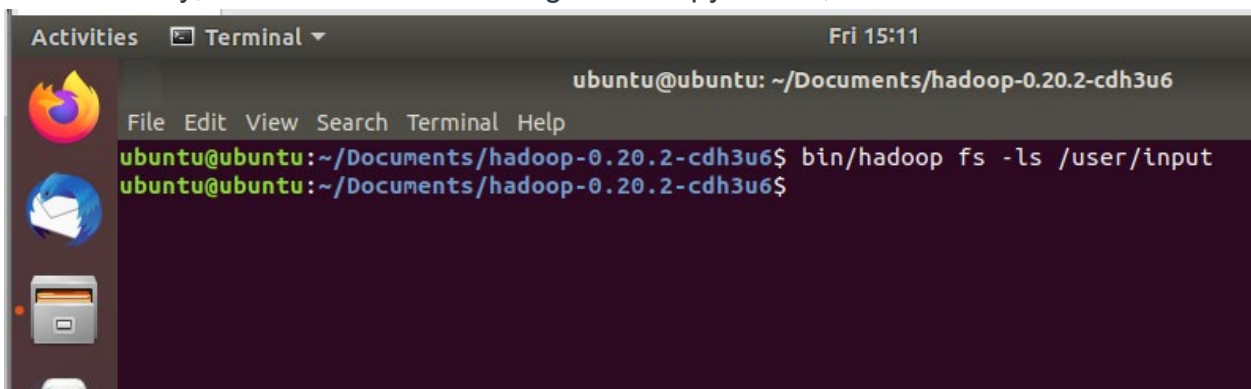


```

ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -mkdir /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 2 items
drwxr-xr-x  - ubuntu supergroup          0 2020-04-17 05:34 /home
drwxr-xr-x  - ubuntu supergroup          0 2020-04-17 14:31 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

3. Copying a file into the Hadoop directory it will happen only within the Hadoop directory, since there is no existing file to copy or edit, we create one.



```

Activities  Terminal
Fri 15:11
ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/input
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

Using the Hadoop commands using the '**touchz <path>**'.

For a file.txt created in the /user/input,

```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -touchz /user/input/file.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/input
Found 1 items
-rw-r--r-- 1 ubuntu supergroup 0 2020-04-17 15:16 /user/input/file.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

After creating the file.txt as shown in the below screenshot, now we copy it to the home directory (we could do it also in the /user directory but we want a different path).

```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -touchz /user/input/file.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/input
Found 1 items
-rw-r--r-- 1 ubuntu supergroup 0 2020-04-17 15:16 /user/input/file.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -cp /user/input/file.txt /home
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 0 2020-04-17 15:19 /home/file.txt
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

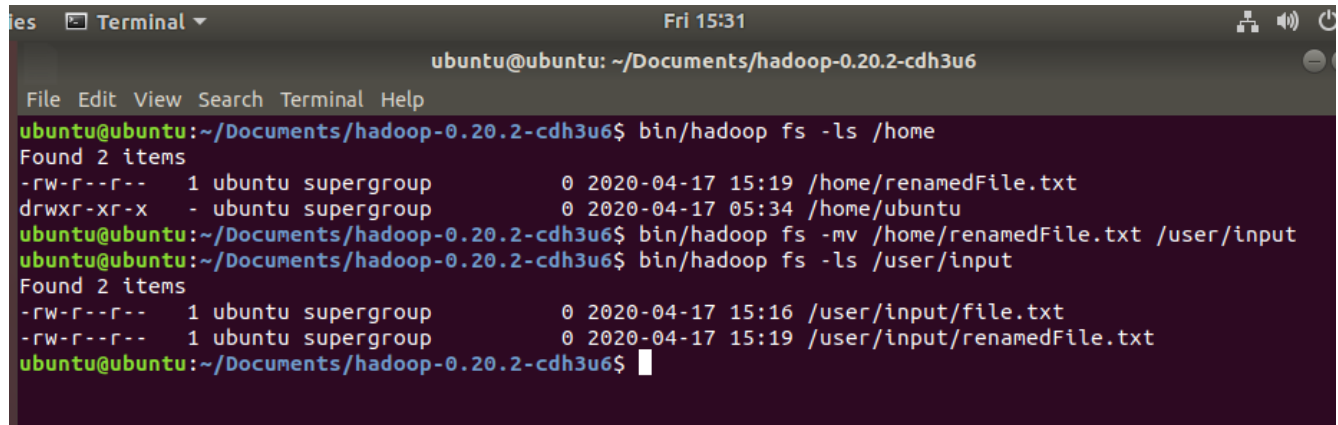
As it is shown with the respective date and all creation details after being copied from the /user/input to /home directory.

4. Renaming the file, according to the location using the Hadoop commands as the following screenshot shows.

```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -mv /home/file.txt /home/renamedFile.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 0 2020-04-17 15:19 /home/renamedFile.txt
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

According to the Hadoop, I can say that is very sensitive, everything must be clear, even if an operation will be done within the same directory, everything should be clear and the paths should be specified as the upper screenshot shows the file when they're renamed with the respective names.

- Moving the file from one directory to another, even if it is shown in the previous exercises, we will transfer the new renamed file of the exercise 4 from /home to /user/input directory as the following screenshot shows.



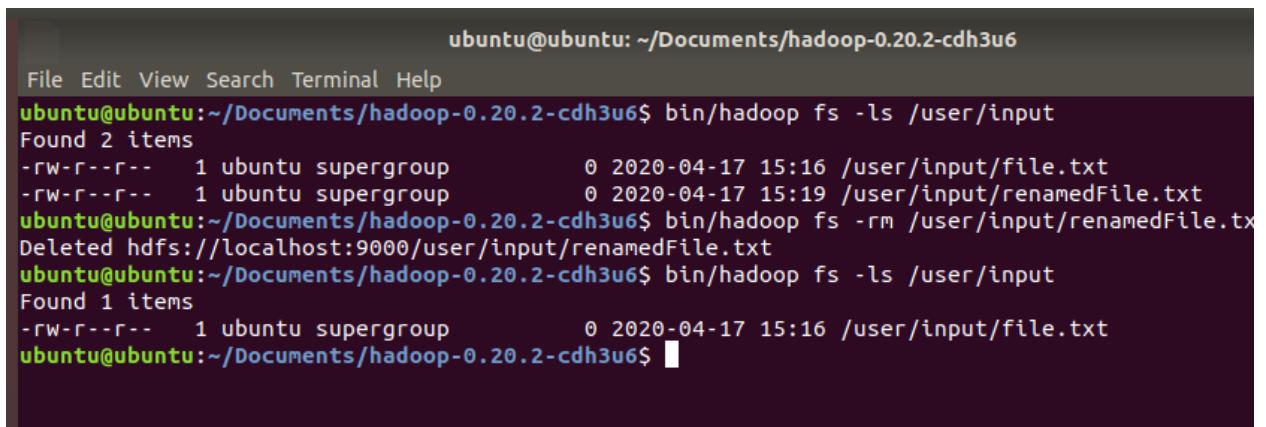
```

ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 2 items
-rw-r--r--  1 ubuntu supergroup          0 2020-04-17 15:19 /home/renamedFile.txt
drwxr-xr-x  - ubuntu supergroup          0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -mv /home/renamedFile.txt /user/input
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/input
Found 2 items
-rw-r--r--  1 ubuntu supergroup          0 2020-04-17 15:16 /user/input/file.txt
-rw-r--r--  1 ubuntu supergroup          0 2020-04-17 15:19 /user/input/renamedFile.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

And as it is shown, from /home everything is listed, then we moved the file 'renamedFile.txt' from home to the '/user/input/' and it was clean safe operation.

- About deleting a file or directory, in the Hadoop framework, as is shown in the following screenshot,



```

ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/input
Found 2 items
-rw-r--r--  1 ubuntu supergroup          0 2020-04-17 15:16 /user/input/file.txt
-rw-r--r--  1 ubuntu supergroup          0 2020-04-17 15:19 /user/input/renamedFile.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -rm /user/input/renamedFile.tx
Deleted hdfs://localhost:9000/user/input/renamedFile.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/input
Found 1 items
-rw-r--r--  1 ubuntu supergroup          0 2020-04-17 15:16 /user/input/file.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

The file successfully deleted, and impressing is that everything is done through the localhost, even if the files are within the same computer.

7. Viewing the files, in this case all the directory of Hadoop in general, as the following screenshot shows, before creating the file I set the content to a DataMining.txt file, writing some simple text 'Hello World'

```

ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 3 items
-rw-r--r--  1 ubuntu supergroup      0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x  - ubuntu supergroup      0 2020-04-17 15:31 /home
drwxr-xr-x  - ubuntu supergroup      0 2020-04-17 15:06 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ echo -e 'Hello\nWorld' | bin/hadoop fs -put - /DataMining.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup     12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup      0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x  - ubuntu supergroup      0 2020-04-17 15:31 /home
drwxr-xr-x  - ubuntu supergroup      0 2020-04-17 15:06 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -cat /DataMining.txt
Hello
World
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

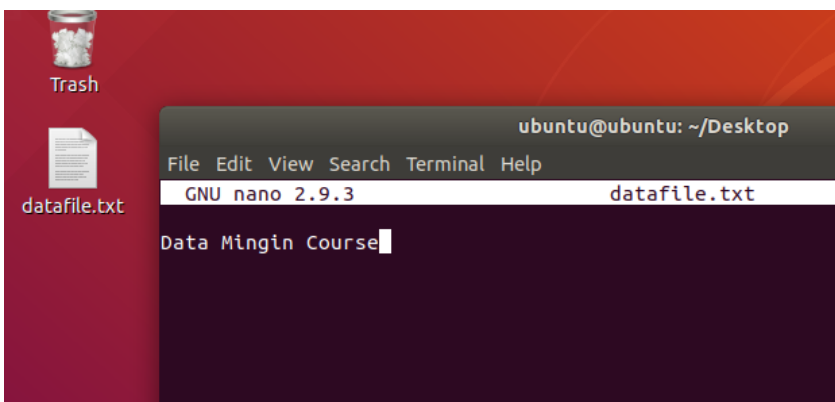
8. CopyFromLocal, the file is copied from the local file system into the specified HDFS file system, in our case the /user/input, as shown in the current screenshot above, but first we create a local path, according to Hadoop, then a simple file in the desktop, like data.txt, and proceeding as screenshots shows.

```

ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -mkdir /user/root/local
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user
Found 2 items
drwxr-xr-x  - ubuntu supergroup      0 2020-04-17 15:35 /user/input
drwxr-xr-x  - ubuntu supergroup      0 2020-04-17 16:19 /user/root
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$

```

After creating the necessary directories, we proceed copying as shown in the following



screenshot, making a simple 'datafile.txt' in the desktop.

Then we proceed copying from desktop to the Hadoop directory.

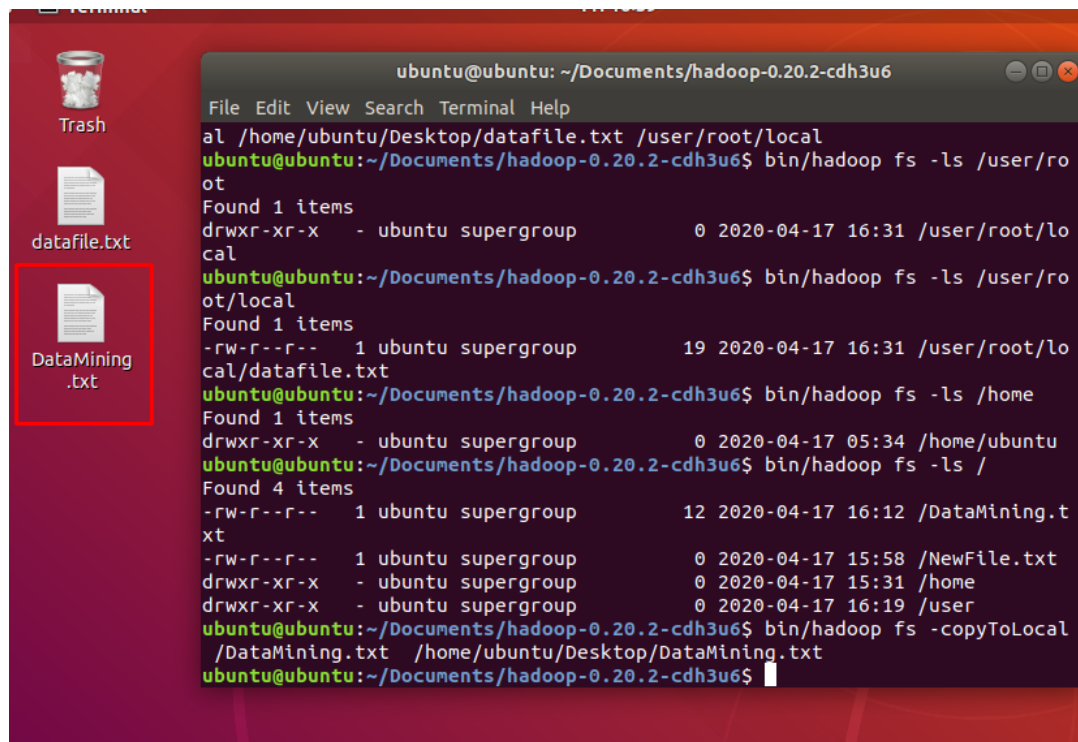
```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -copyFromLocal /home/ubuntu/Desktop/datafile.txt /user/root/local
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/root
Found 1 items
drwxr-xr-x - ubuntu supergroup          0 2020-04-17 16:31 /user/root/local
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/root/local
Found 1 items
-rw-r--r--  1 ubuntu supergroup          19 2020-04-17 16:31 /user/root/local/datafile.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

So, from Desktop to a specific Hadoop directory.

9. CopyToLocal in this case we will send some file.txt from Hadoop to the ubuntu /Desktop or '/home/ubuntu/Desktop/' directory, as shown even in the screenshot above,

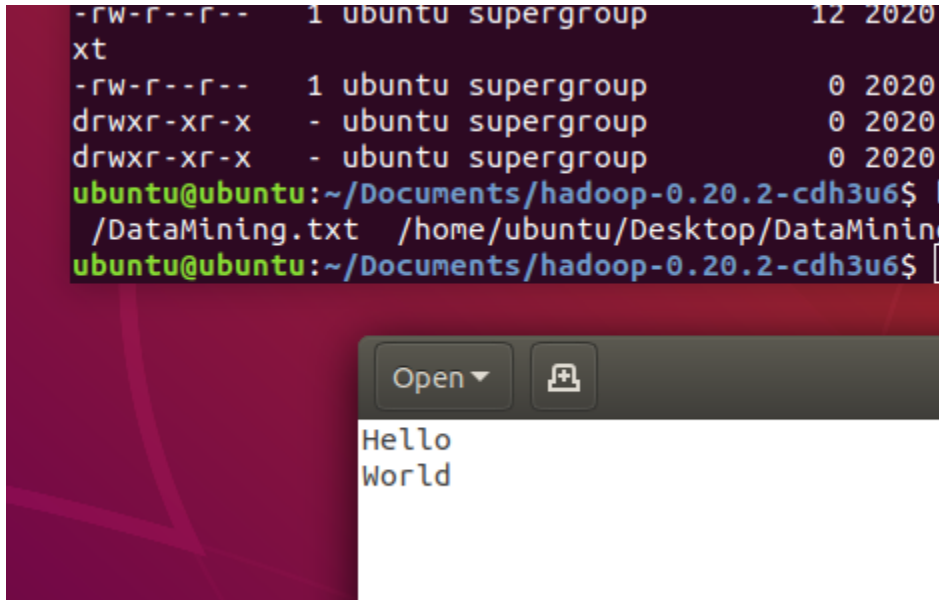
```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup          12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup           0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x - ubuntu supergroup           0 2020-04-17 15:31 /home
drwxr-xr-x - ubuntu supergroup           0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

So there is a simple file and we send it from Hadoop directory to the Desktop in the ubuntu,



```
ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
al /home/ubuntu/Desktop/datafile.txt /user/root/local
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/root
Found 1 items
drwxr-xr-x - ubuntu supergroup          0 2020-04-17 16:31 /user/root/local
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /user/root/local
Found 1 items
-rw-r--r--  1 ubuntu supergroup          19 2020-04-17 16:31 /user/root/local/datafile.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 1 items
drwxr-xr-x - ubuntu supergroup           0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup          12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup           0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x - ubuntu supergroup           0 2020-04-17 15:31 /home
drwxr-xr-x - ubuntu supergroup           0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -copyToLocal /DataMining.txt /home/ubuntu/Desktop/DataMining.txt
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```


And even the content is the same as the Hadoop framework had it in his directory,



```

-rw-r--r-- 1 ubuntu supergroup 12 2020
xt
-rw-r--r-- 1 ubuntu supergroup 0 2020
drwxr-xr-x - ubuntu supergroup 0 2020
drwxr-xr-x - ubuntu supergroup 0 2020
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ 
/DataMining.txt /home/ubuntu/Desktop/DataMining
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ 

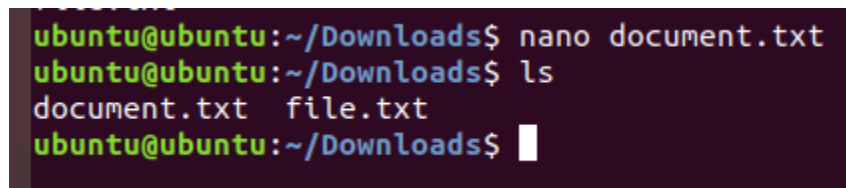
```

Open ▾

Hello
World

So, as it is shown the full text is shown, and the file is copied in the /Desktop.

10. Put, the put command is similar to *put* command, except that the source is **restricted to a local** file reference. This time a document.txt file will be created to the downloads in the ubuntu and it will be transferred using Hadoop to the Hadoop corresponding directories.

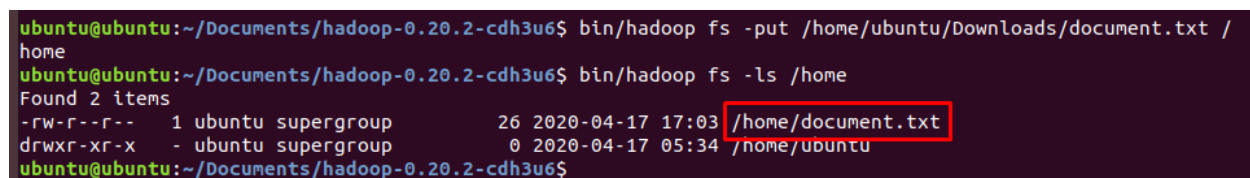


```

ubuntu@ubuntu:~/Downloads$ nano document.txt
ubuntu@ubuntu:~/Downloads$ ls
document.txt  file.txt
ubuntu@ubuntu:~/Downloads$ 

```

After creating the file where is written 'Data Mining is important'



```

ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -put /home/ubuntu/Downloads/document.txt /
home
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 26 2020-04-17 17:03 /home/document.txt
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ 

```

It is taken from '/home/ubuntu/Downloads/' to the '/home' directory inside the Hadoop directories.

11. Get, the Get command is similar to the copyToLocal, this time we will copy from Hadoop directory to the ubuntu /Desktop some file as shown in the following screenshot,

```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 26 2020-04-17 17:03 /home/document.txt
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

document.txt will be taken from Hadoop directory to the ubuntu /Desktop

```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /home
Found 2 items
-rw-r--r-- 1 ubuntu supergroup 26 2020-04-17 17:03 /home/document.txt
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 05:34 /home/ubuntu
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -get /home/document.txt /home/ubuntu/Desktop
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ cd
ubuntu@ubuntu:~$ cd Desktop
ubuntu@ubuntu:~/Desktop$ ls
datafile.txt DataMining.txt document.txt
ubuntu@ubuntu:~/Desktop$
```

And then its content

```
ubuntu@ubuntu:~/Desktop$ cat document.txt
Data Mining is important
ubuntu@ubuntu:~/Desktop$
```

12. Displaying the size of the files, as shown in the screenshot above, all the file root is displayed with the respective size in the left part, in each respective directory.

```
ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r-- 1 ubuntu supergroup 12 2020-04-17 16:12 /DataMining.txt
-rw-r--r-- 1 ubuntu supergroup 0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 17:03 /home
drwxr-xr-x - ubuntu supergroup 0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -du /
Found 4 items
12      hdfs://localhost:9000/DataMining.txt
0       hdfs://localhost:9000/NewFile.txt
30      hdfs://localhost:9000/home
19      hdfs://localhost:9000/user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```


13. Change group of file/directory, first of all we will list all the directories in the Hadoop framework, as shown in the following framework.

```
ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup      12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup       0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x  - ubuntu supergroup       0 2020-04-17 17:03 /home
drwxr-xr-x  - ubuntu supergroup       0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

As shown from the listing part they are all listed as part of the supergroup,

```
ubuntu@ubuntu: ~/Documents/hadoop-0.20.2-cdh3u6
File Edit View Search Terminal Help
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -chgrp -R cloudera
/user/
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls
ls: Cannot access .: No such file or directory.
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup      12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup       0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x  - ubuntu supergroup       0 2020-04-17 17:03 /home
drwxr-xr-x  - ubuntu cloudera        0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

Then the last directory the '/user' changed its group from superuser to '/user'.

14. Change permissions of file/directory, for the '/user' directory in this case, as shown in the following screenshot.

```
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup      12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup       0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x  - ubuntu supergroup       0 2020-04-17 17:03 /home
drwxr-xr-x  - ubuntu cloudera        0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -chmod -R 777 /user/
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$ bin/hadoop fs -ls /
Found 4 items
-rw-r--r--  1 ubuntu supergroup      12 2020-04-17 16:12 /DataMining.txt
-rw-r--r--  1 ubuntu supergroup       0 2020-04-17 15:58 /NewFile.txt
drwxr-xr-x  - ubuntu supergroup       0 2020-04-17 17:03 /home
drwxrwxrwx  - ubuntu cloudera        0 2020-04-17 16:19 /user
ubuntu@ubuntu:~/Documents/hadoop-0.20.2-cdh3u6$
```

All the permissions are set successfully on Hadoop.