

# Assignment-1

MTCS-202(P) - Hadoop Assignment

Aryan Sai Arvapelly, Regd. No - 23352, I MTech CS

## Question - 1

- Create two directories 'dir1' and 'dir2' using a single hdfs command inside home directory of hdfs. The dir2 should be subdirectory of dir1.

```
hdfs dfs -mkdir -p /dir1/dir2
```

### Browse Directory

Show  entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
No data available in table								
Showing 0 to 0 of 0 entries								
								<div>PreviousNext</div>

Hadoop, 2023.

- Verify that the two folders have been created in the above path Inside dir 2

```
hdfs dfs -ls /
```

```

huser@aryan-OptiPlex-9020:~$ hdfs dfs -ls /
Found 9 items
drwxr-xr-x - huser supergroup 0 2024-01-16 21:24 /average
drwxr-xr-x - huser supergroup 0 2024-02-26 10:36 /chain
drwxr-xr-x - huser supergroup 0 2024-02-26 10:15 /charcount
drwxr-xr-x - huser supergroup 0 2024-02-26 10:47 /dir1
drwxr-xr-x - huser supergroup 0 2024-01-17 10:45 /matrixaddition
drwxr-xr-x - huser supergroup 0 2024-01-29 21:03 /select
drwxr-xr-x - huser supergroup 0 2024-01-13 13:46 /tmp
drwxr-xr-x - huser supergroup 0 2024-01-17 11:51 /transpose
drwxr-xr-x - huser supergroup 0 2024-01-13 13:46 /wordcount

```

- Create an empty file, file1.txt

```
hdfs dfs -touch /dir1/dir2/file1.txt
```

## Browse Directory

Show  entries
 Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	<a href="#">huser</a>	<a href="#">supergroup</a>	0 B	Feb 26 11:27	<a href="#">1</a>	128 MB	<a href="#">file1.txt</a>	

Showing 1 to 1 of 1 entries

Hadoop, 2023.

- Create a file file2.txt in local filesystem with some text inside it

```
touch Desktop/file2.txt
```

```

huser@aryan-OptiPlex-9020:~$ touch Desktop/file2.txt
huser@aryan-OptiPlex-9020:~$ ls
Desktop  Downloads  file2  hadoop  myfile.txt  Public  Templates  Videos
Documents  file1      file3  Music  Pictures    snap    test

```

- Copy file2.txt from local to hdfs inside dir2.

```
hdfs dfs -put /home/huser/Desktop/file2.txt /dir1/dir2/
```

## Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	huser	supergroup	0 B	Feb 26 11:27	1	128 MB	<a href="#">file1.txt</a>	
<input type="checkbox"/>	-rw-r--r--	huser	supergroup	13 B	Feb 26 11:33	1	128 MB	<a href="#">file2.txt</a>	

Showing 1 to 2 of 2 entries

Hadoop, 2023.

- List the subdirectories and files inside dir1 recursively

```
hdfs dfs -ls -R /dir1
```

```
huser@aryan-OptiPlex-9020:~$ hdfs dfs -ls -R /dir1
drwxr-xr-x  - huser supergroup      0 2024-02-26 11:33 /dir1/dir2
-rw-r--r--  1 huser supergroup      0 2024-02-26 11:27 /dir1/dir2/file1.txt
-rw-r--r--  1 huser supergroup    13 2024-02-26 11:33 /dir1/dir2/file2.txt
huser@aryan-OptiPlex-9020:~$
```

- List the files inside dir2 ,sorted by size but size should be displayed in KBs/MBs and not bytes

```
hdfs dfs -ls -S -h /dir1/dir2
```

```
huser@aryan-OptiPlex-9020:~$ hdfs dfs -ls -S -h /dir1/dir2
Found 2 items
-rw-r--r--  1 huser supergroup    13 2024-02-26 11:33 /dir1/dir2/file2.txt
-rw-r--r--  1 huser supergroup      0 2024-02-26 11:27 /dir1/dir2/file1.txt
```

- Rename the file, file2.txt to file3.txt

```
hdfs dfs -mv /dir1/dir2/file2.txt /dir1/dir2/file3.txt
```

## Browse Directory

Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	huser	supergroup	0 B	Feb 26 11:27	1	128 MB	<a href="#">file1.txt</a>	
<input type="checkbox"/>	-rw-r--r--	huser	supergroup	13 B	Feb 26 11:33	1	128 MB	<a href="#">file3.txt</a>	

Showing 1 to 2 of 2 entries

Hadoop, 2023.

- Remove the directory dir1 using a single command.

```
hadoop fs -rm -r /dir1
```

```
huser@aryan-OptiPlex-9020:~$ hadoop fs -rm -r /dir1
Deleted /dir1
huser@aryan-OptiPlex-9020:~$
```

## Question 2:

- Suppose there is file of size 514 MB stored in HDFS (Hadoop 2.x) using default block size configuration and default replication factor:
- How many blocks will be created in total?
- What will be the size of each block?

### Answer:

In Hadoop 2.x, the default block size is 128 MB, and the default replication factor is 3.

To calculate the number of blocks created and the size of each block:

1. Number of blocks = Total file size / Block size
2. Size of each block = Total file size / Number of blocks

Given that the file size is 514 MB, the block size is 128 MB, and the replication factor is 3:

1. Number of blocks = 514 MB / 128 MB = 4.015625 blocks (approx)

Since the number of blocks must be a whole number, Hadoop rounds up to the nearest whole block. Therefore, the number of blocks created will become 5.

2. Size of each block = 514 MB / 5 blocks  $\approx$  128 MB

### Question 3:

- Create a directory inside home directory of local filesystem named 'test'
- Create few empty files inside the test directory namely a.pdf, b.html, c.xml
- List the files in reverse alphabetical order of file name
- Display only the file which ends with .html extension

```
huser@aryan-OptiPlex-9020:~$ mkdir test
huser@aryan-OptiPlex-9020:~$ cd test
huser@aryan-OptiPlex-9020:~/test$ touch a.pdf
huser@aryan-OptiPlex-9020:~/test$ touch b.html
huser@aryan-OptiPlex-9020:~/test$ touch c.xml
huser@aryan-OptiPlex-9020:~/test$ ls -r
c.xml b.html a.pdf
huser@aryan-OptiPlex-9020:~/test$ ls *.html
b.html
huser@aryan-OptiPlex-9020:~/test$
```

### Question 4:

- Create two new text files, file1 and file2 , with following content using cat command in your linux home directory:
  - file1: This is from file1
  - file2: This is from file2
- Display the contents of the file 1 and file2 using cat command
- Concatenate the contents of the two files and put them into a new file file3 and
- display the results.
- Count the number of lines and number of words in the file3.

```

huser@aryan-OptiPlex-9020:~/test$ cat > ~/file1 <<EOF
> This is from file1
> EOF
huser@aryan-OptiPlex-9020:~/test$ cat > ~/file2 <<EOF
> This is from file2
> EOF
huser@aryan-OptiPlex-9020:~/test$ cat file1
cat: file1: No such file or directory
huser@aryan-OptiPlex-9020:~/test$ cd ..
huser@aryan-OptiPlex-9020:~$ cat > ~/file1 <<EOF
This is from file1
EOF
huser@aryan-OptiPlex-9020:~$ cat > ~/file2 <<EOF
This is from file2
EOF
huser@aryan-OptiPlex-9020:~$ cat file1
This is from file1
huser@aryan-OptiPlex-9020:~$ cat file2
This is from file2
huser@aryan-OptiPlex-9020:~$ cat ~/file1 ~/file2 > ~/file3
huser@aryan-OptiPlex-9020:~$ wc ~/file3
  2  8 38 /home/huser/file3
huser@aryan-OptiPlex-9020:~$

```

## Question 5:

- Create a text file myfile.txt with 5 lines in home directory of local filesystem:
- Display last 3 lines of that file.
- Display all lines of that text file except first line.

```

huser@aryan-OptiPlex-9020:~$ echo "Line 1" > ~/myfile.txt
huser@aryan-OptiPlex-9020:~$ echo "Line 2" >> ~/myfile.txt
huser@aryan-OptiPlex-9020:~$ echo "Line 3" >> ~/myfile.txt
huser@aryan-OptiPlex-9020:~$ echo "Line 4" >> ~/myfile.txt
huser@aryan-OptiPlex-9020:~$ echo "Line 5" >> ~/myfile.txt
huser@aryan-OptiPlex-9020:~$ tail -n 3 ~/myfile.txt
Line 3
Line 4
Line 5
huser@aryan-OptiPlex-9020:~$ sed -n '2,$p' ~/myfile.txt
Line 2
Line 3
Line 4
Line 5
huser@aryan-OptiPlex-9020:~$

```