Assignment-1

2023-11-08

Architecture of Big Data System

Assignment 1

HDFS Basic Commands

Command 1 - `ls`

`ls` : This command is used to list all the files. Use lsr for recursive approach. It is useful when we want a hierarchy of a folder.

List Root Directory

```
#List Root Directory
hdfs dfs -ls /
```

```
Found 8 items
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:00 /Orders
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:32 /Orders_new
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:53 /TestOrdersid
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:25 /Test_DB
drwxr-xr-x   - sois supergroup          0 2023-10-18 11:37 /mydir
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:10 /order_folder
drwxrwxr-x   - sois supergroup          0 2023-01-15 16:51 /tmp
drwxr-xr-x   - sois supergroup          0 2023-01-15 16:26 /user
```

List all files under 'Orders'

```
#List all files under 'Orders'
hdfs dfs -ls /Orders
```

```
Found 5 items
-rw-r--r--   1 sois supergroup          0 2023-10-25 11:00 /Orders/_SUCCESS
-rw-r--r--   1 sois supergroup     741614 2023-10-25 11:00 /Orders/part-m-00000
-rw-r--r--   1 sois supergroup     753022 2023-10-25 11:00 /Orders/part-m-00001
-rw-r--r--   1 sois supergroup     752368 2023-10-25 11:00 /Orders/part-m-00002
-rw-r--r--   1 sois supergroup     752940 2023-10-25 11:00 /Orders/part-m-00003
```

Command 2 - `mkdir`

`mkdir` : To create a directory. In Hadoop dfs there is no home directory by default.

Make a new folder in root called 'Assignment_1'

```
#Create New Directory
hdfs dfs -mkdir /Assignment_1

#Check if new directory is created
hdfs dfs -ls /
```

```
Found 9 items
drwxr-xr-x   - sois supergroup          0 2023-11-04 13:57 /Assignment_1
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:00 /Orders
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:32 /Orders_new
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:53 /TestOrdersid
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:25 /Test_DB
drwxr-xr-x   - sois supergroup          0 2023-10-18 11:37 /mydir
drwxr-xr-x   - sois supergroup          0 2023-10-25 11:10 /order_folder
drwxrwxr-x   - sois supergroup          0 2023-01-15 16:51 /tmp
drwxr-xr-x   - sois supergroup          0 2023-01-15 16:26 /user
```

Command 3 - `touchz`

`touchz` : It creates an empty file.

```
#Create new empty file
hdfs dfs -touchz /Assignment_1/sample.txt

#Check if new file is created
hdfs dfs -ls /Assignment_1
```

```
Found 1 items
-rw-r--r--   1 sois supergroup          0 2023-11-04 14:04 /Assignment_1/sample.txt
```

Command 4 - `copyFromLocal` or `put`

`copyFromLocal` (or) `put` : To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.

```

#Use put to copy test1.py from local machine to hdfs
hdfs dfs -put /home/sois/test1.py /Assignment_1
#Check if file is copied to hdfs
hdfs dfs -ls /Assignment_1
```

```
Found 2 items
-rw-r--r--   1 sois supergroup          0 2023-11-04 14:04 /Assignment_1/sample.txt
-rw-r--r--   1 sois supergroup        257 2023-11-04 14:14 /Assignment_1/test1.py
```

Command 5 - `cat`

`cat` : To print file contents

```
#print contents of test1.py
hdfs dfs -cat /Assignment_1/test1.py
```

```
2021-11-04 14:19:41,060 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false,
remoteHostTrusted = false
#Arithmatic Functions

def add(a,b):
    return a+b;
def mul(a,b):
    return a*b;
def diff(a,b):
    return a-b;
def div(a,b):
    return a/b;

#Arithmatic Operations

a = 10;
b = 20;
print(add(a,b));
print(mul(a,b));
print(diff(a,b));
print(div(a,b));
```

Command 6 - `copyToLocal` or `get`

`copyToLocal` (or) `get` : To copy files/folders from hdfs store to local file system

```
#Copy sample.txt to local machine
hdfs dfs -get /Assignment_1/sample.txt /home/sois/copy_sample.txt
#Check if file is copied
ls /home/sois
```

```
copy_sample.txt  Documents          examples.desktop     input         orders.java  snap           test1.py
derby.log        Downloads          get-pip.py           metastore_db  Pictures     Students.java  Videos
Desktop          eclipse-workspace  hadoop-3.2.1.tar.gz  Music         Public       Templates
```

Command 7 - `moveFromLocal`

`moveFromLocal` : This command will move file from local to hdfs

```

#Move file from local machine to hdfs
hdfs dfs -moveFromLocal /home/sois/copy_sample.txt /Assignment_1
#Check if file has been moved
hdfs dfs -ls /Assignment_1
```

```
Found 3 items
-rw-r--r--   1 sois supergroup          0 2023-11-04 14:28 /Assignment_1/copy_sample.txt
-rw-r--r--   1 sois supergroup          0 2023-11-04 14:04 /Assignment_1/sample.txt
-rw-r--r--   1 sois supergroup        257 2023-11-04 14:14 /Assignment_1/test1.py
```

Command 8 - `cp`

`cp` : This command is used to copy files within hdfs

```
#Make a directory to store copied file
hdfs dfs -mkdir /Assignment_1/Backup

#Copy file within hdfs
hdfs dfs -cp /Assignment_1/test1.py /Assignment_1/Backup

#check if file has been copied
hdfs dfs -ls /Assignment_1/Backup
```

```
Found 1 items
-rw-r--r--   1 sois supergroup        257 2023-11-04 14:38 /Assignment_1/Backup/test1.py
```

Command 9 - `mv`

`mv` : This command is used to move files within hdfs.

```
#Move Files within hdfs
hdfs dfs -mv /Assignment_1/sample.txt /Assignment_1/Backup
#Check if file has been moved
hdfs dfs -ls /Assignment_1/Backup
```

```
Found 2 items
-rw-r--r--   1 sois supergroup          0 2023-11-04 14:04 /Assignment_1/Backup/sample.txt
-rw-r--r--   1 sois supergroup        257 2023-11-04 14:38 /Assignment_1/Backup/test1.py
```

Command 10 - `rm r`

`rm r` : This command deletes a file from HDFS recursively. It is very useful command when you want to delete a non-empty directory.

```
#Remove files recursively
hdfs dfs -rm -r /Assignment_1/Backup
```

```
Deleted /Assignment_1/Backup
```

Command 11 - `du`

`du` : It will give the size of each file in directory

```
#Size of each file in 'Assignment_1'
hdfs dfs -du /Assignment_1
```

```
0    0    /Assignment_1/copy_sample.txt
257  257  /Assignment_1/test1.py
```

Command 12 - `du s`

`du s` : This command will give the total size of directory/file.

```
#Size of  'Assignment_1'
hdfs dfs -du -s /Assignment_1
```

```
257  257  /Assignment_1
```

Command 13 - `stat`

`stat` : It will give the last modified time of directory or path. In short it will give stats of the directory or file.

```
#last modified time
hdfs dfs -stat /Assignment_1
```

```
2023-11-04 09:19:42
```