

Big Data Technologies (CSP554) Assignment - 2

```
$ ssh -i keypair.pem hadoop@ec2-3-145-106-212.us-east-2.compute.amazonaws.com
The authenticity of host 'ec2-3-145-106-212.us-east-2.compute.amazonaws.com (3.1
45.106.212)' can't be established.
ED25519 key fingerprint is SHA256:zdDwbGTHsCDDfJXIAFnHG9HIkG70G0qNiNPfKgsP4+E.
This key is not known by any other names.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-3-145-106-212.us-east-2.compute.amazonaws.com' (ED25519) to the list of known hosts.

 _|_(_|-_) / Amazon Linux 2 AMI
 __|_\_|_|

https://aws.amazon.com/amazon-linux-2/
29 package(s) needed for security, out of 67 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE MMMMMMM MBBBBBBBBB RRRRRRRRRRRRRR
E:::::::::::E M:::::M M:::::M R:::::R
EE:::::EEEEEEE:::E M:::::M M:::::M R:::::RRRRR:::::R
 E::::E     EEEEE M:::::M M:::::M RR:::::R     R:::::R
 E::::E     M:::::M:::::M M:::M:::::M R:::::R     R:::::R
 E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::::RRRRR:::::R
 E:::::::::::E M:::::M M:::::M M:::::M R:::::R:::::RR
 E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::::RRRRR:::::R
 E::::E     M:::::M M:::::M M:::::M R:::::R     R:::::R
 E:::::E     EEEEE M:::::M M:::::M M:::::M R:::::R
EE:::::EEEEEEE:::E M:::::M M:::::M R:::::R     R:::::R
E:::::::::::E M:::::M M:::::M RR:::::R     R:::::R
EEEEEEEEEEEEEEEEEE MMMMMMM MBBBBBBBBB RRRRRRRR


```

9. hadoop fs -ls /

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /
ls: Call From ip-172-31-8-215.us-east-2.compute.internal/172.31.8.215 to ip-172-31-8-215.us-east-2.compute.internal:8020 failed on connection exception: java.net.ConnectException: Connection refused; For more details see: http://wiki.apache.org/hadoop/ConnectionRefused
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /
Found 4 items
drwxr-xr-x  - hdfs hdfsadmingroup      0 2023-09-14 03:54 /apps
drwxrwxrwt  - hdfs hdfsadmingroup      0 2023-09-14 03:54 /tmp
drwxr-xr-x  - hdfs hdfsadmingroup      0 2023-09-14 03:54 /user
drwxr-xr-x  - hdfs hdfsadmingroup      0 2023-09-14 03:54 /var
```

10. hadoop fs -ls /user

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user
Found 6 items
drwxrwxrwx  - hadoop hdfsadmingroup      0 2023-09-14 03:54 /user/hadoop
drwxr-xr-x  - mapred mapred              0 2023-09-14 03:54 /user/history
drwxrwxrwx  - hdfs hdfsadmingroup      0 2023-09-14 03:54 /user/hive
drwxrwxrwx  - hue  hue                  0 2023-09-14 03:54 /user/hue
drwxrwxrwx  - oozie oozie              0 2023-09-14 03:54 /user/oozie
drwxrwxrwx  - root  hdfsadmingroup      0 2023-09-14 03:54 /user/root
```

11. hadoop fs -mkdir /user/csp554

```
hadoop fs -ls /user
```

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -mkdir /user/csp554
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user
Found 7 items
drwxr-xr-x  - hadoop hdfsadmingroup          0 2023-09-14 03:56 /user/csp554
drwxrwxrwx  - hadoop hdfsadmingroup          0 2023-09-14 03:54 /user/hadoop
drwxr-xr-x  - mapred mapred                 0 2023-09-14 03:54 /user/history
drwxrwxrwx  - hdfs  hdfsadmingroup          0 2023-09-14 03:54 /user/hive
drwxrwxrwx  - hue   hue                      0 2023-09-14 03:54 /user/hue
drwxrwxrwx  - oozie  oozie                  0 2023-09-14 03:54 /user/oozie
drwxrwxrwx  - root   hdfsadmingroup          0 2023-09-14 03:54 /user/root
```

12. hadoop fs -mkdir /user/csp554-5

```
hadoop fs -ls /user
```

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -mkdir /user/csp554-5
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user
Found 8 items
drwxr-xr-x  - hadoop hdfsadmingroup          0 2023-09-14 03:56 /user/csp554
drwxr-xr-x  - hadoop hdfsadmingroup          0 2023-09-14 03:58 /user/csp554-5
drwxrwxrwx  - hadoop hdfsadmingroup          0 2023-09-14 03:54 /user/hadoop
drwxr-xr-x  - mapred mapred                 0 2023-09-14 03:54 /user/history
drwxrwxrwx  - hdfs  hdfsadmingroup          0 2023-09-14 03:54 /user/hive
drwxrwxrwx  - hue   hue                      0 2023-09-14 03:54 /user/hue
drwxrwxrwx  - oozie  oozie                  0 2023-09-14 03:57 /user/oozie
drwxrwxrwx  - root   hdfsadmingroup          0 2023-09-14 03:54 /user/root
```

13. hadoop fs -put ~/lavanyanarthu.txt /user/csp554

```
hadoop fs -ls /user/csp554
```

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -put ~/lavanyanarthu.txt /user/csp554
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user/csp554
Found 1 items
-rw-r--r--  1 hadoop hdfsadmingroup      21 2023-09-14 03:59 /user/csp554/lavany
anarthu.txt
```

14. hadoop distcp /user/csp554/lavanyanarthu.txt /user/csp554-5

```
[hadoop@ip-172-31-8-215 ~]$ hadoop distcp /user/csp554/lavanyanarthu.txt /user/csp55
4-5
2023-09-14 04:00:42,696 INFO tools.DistCp: Input Options: DistCpOptions{atomicCommit
=false, syncFolder=false, deleteMissing=false, ignoreFailures=false, overwrite=false
, append=false, useDiff=false, useRdiff=false, fromSnapshot=null, toSnapshot=null, s
kipCRC=false, blocking=true, numListStatusThreads=0, maxMaps=20, mapBandwidth=0.0, c
opyStrategy='uniformsize', preserveStatus=[], atomicWorkPath=null, logPath=null, sou
rceFileListing=null, sourcePaths=[/user/csp554/lavanyanarthu.txt], targetPath=/user/
csp554-5, filtersFile='null', blocksPerChunk=0, copyBufferSize=8192, verboseLog=false
, directWrite=false, useIterator=false}, sourcePaths=[/user/csp554/lavanyanarthu.tx
t], targetPathExists=true, preserveRawXattrs=false
2023-09-14 04:00:43,010 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting t
o ResourceManager at ip-172-31-8-215.us-east-2.compute.internal/172.31.8.215:8032
2023-09-14 04:00:43,186 INFO client.AHSProxy: Connecting to Application History serv
er at ip-172-31-8-215.us-east-2.compute.internal/172.31.8.215:10200
```

```
Peak Map Virtual memory (bytes)=3074936832
File Input Format Counters
    Bytes Read=310
File Output Format Counters
    Bytes Written=0
DistCp Counters
    Bandwidth in Bytes=21
    Bytes Copied=21
    Bytes Expected=21
    Files Copied=1
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user/csp554-5
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmingroup          21 2023-09-14 04:00 /user/csp554-5/lava
yanarthu.txt
```

15. aws s3 cp s3://csp554/a20516764.txt /home/hadoop/a20516764.txt

ls /home/hadoop

```
[hadoop@ip-172-31-8-215 ~]$ aws s3 cp s3://a20516764-bucket/a20516764.txt /home/hadoop
fatal error: An error occurred (404) when calling the HeadObject operation: Key "a20516764.txt" does not exist
[hadoop@ip-172-31-8-215 ~]$ aws s3 cp s3://csp554/a20516764.txt /home/hadoop/a20516764.txt
download: s3://csp554/a20516764.txt to ./a20516764.txt
[hadoop@ip-172-31-8-215 ~]$ ls /home/hadoop
a20516764.txt  lavanyanarthu.txt
```

16. hadoop fs -cp s3://csp554/a20516764.txt hdfs:///user/csp554-5

hadoop fs -ls /user/csp554-5

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -cp s3://csp554/a20516764.txt hdfs:///user/csp554-5
2023-09-14 04:20:49,214 INFO s3n.S3NativeFileSystem: Opening 's3://csp554/a20516764.txt' for reading
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user/csp554-5
Found 2 items
-rw-r--r-- 1 hadoop hdfsadmingroup          19 2023-09-14 04:20 /user/csp554-5/a20516764.txt
-rw-r--r-- 1 hadoop hdfsadmingroup          21 2023-09-14 04:00 /user/csp554-5/lavanyanarthu.txt
```

17. hadoop fs -cat /user/csp554-5/a20516764.txt

```
[hadoop@ip-172-31-8-215 ~]$
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -cat /user/csp554-5/a20516764.txt
this is the id file[hadoop@ip-172-31-8-215 ~]$
[hadoop@ip-172-31-8-215 ~]$
```

18. hadoop fs -rm /user/csp554-5//a20516764.txt

```
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -rm /user/csp554-5//a20516764.txt
Deleted /user/csp554-5/a20516764.txt
[hadoop@ip-172-31-8-215 ~]$ hadoop fs -ls /user/csp554-5
Found 1 items
-rw-r--r-- 1 hadoop hdfsadmin group 21 2023-09-14 04:00 /user/csp554-5/lava
```

**Submitted by:
Sailavanya Narthu
A20516764
snarthu@hawk.iit.edu**