

## Getting in

```
1
ssh -i ~/Downloads/s4005768-cosc2637.pem ec2-user@s4005768.jump.cosc2637.route53.aws.rmit.edu.au
2
./create_cluster.sh
3
ssh hadoop@s4005768.emr.cosc2637.route53.aws.rmit.edu.au -i s4005768-cosc2637.pem
```

## Getting off

```
1
exit
2
./terminate_cluster.sh
```

## Files

### Example copying file from local machine to jumphost

```
- Copying "folder" week3 to jumphost, the folder is in downloads
scp -r -i ~/Downloads/s4005768-cosc2637.pem ~/Downloads/week3*
ec2-user@s4005768.jump.cosc2637.route53.aws.rmit.edu.au:/home/ec2-user/

Last login: Mon Aug  4 11:15:29 on ttys001
(base) aidenaiden@Aidens-MacBook-Air ~ % scp -r -i ~/Downloads/s4005768-cosc2637
.pem ~/Downloads/week3 ec2-user@s4005768.jump.cosc2637.route53.aws.rmit.edu.au:/
home/ec2-user/

mapper.py                                100% 470      1.5KB/s   00:00
reducer.py                              100% 640      2.0KB/s   00:00
run-3.sh                                100% 222      0.9KB/s   00:00
(base) aidenaiden@Aidens-MacBook-Air ~ % scp -r -i ~/Downloads/s4005768-cosc2637
.pem ~/Downloads/week3 ec2-user@s4005768.jump.cosc2637.route53.aws.rmit.edu.au:/
home/ec2-user/
```

### Example copying folder from jumphost to masternode

```
- Below code copies all files in the "week3 folder" into masternode
scp -i ~/s4005768-cosc2637.pem -r ~/week3*
hadoop@s4005768.emr.cosc2637.route53.aws.rmit.edu.au:/home/hadoop/

[ec2-user@ip-192-168-6-143 ~]$ ls
boost_cluster.sh  instructions  student  week3
create_cluster.sh s4005768-cosc2637.pem terminate_cluster.sh
[ec2-user@ip-192-168-6-143 ~]$ scp -i ~/s4005768-cosc2637.pem -r ~/week3 hadoop@
s4005768.emr.cosc2637.route53.aws.rmit.edu.au:/home/hadoop/
mapper.py                                100% 470      88.8KB/s   00:00
reducer.py                              100% 640      359.2KB/s   00:00
run-3.sh                                100% 222      100.0KB/s   00:00
```

### Moving all the files from folder into home directory

```
- Moved week3 files out of folder into home directory
mv ~/week3/* ~/
```

### Copying from master node to HDFS

```
- Moved the 3littlepigs.txt to HDFS (data files)
hadoop fs -copyFromLocal 3littlepigs.txt /

[hadoop@ip-192-168-26-107 ~]$ hadoop fs -ls /
Found 5 items
-rw-r--r-- 1 hadoop hdfsadmin group 5485 2025-08-04 01:39 /3littlepigs.tx
t
drwxr-xr-x - hdfs hdfsadmin group 0 2025-08-04 00:37 /apps
drwxrwxrwt - hdfs hdfsadmin group 0 2025-08-04 00:40 /tmp
drwxr-xr-x - hdfs hdfsadmin group 0 2025-08-04 00:37 /user
drwxr-xr-x - hdfs hdfsadmin group 0 2025-08-04 00:37 /var
[hadoop@ip-192-168-26-107 ~]$
```

### Removing folders in jumphost

```
- rm -r folder_name
```

### Running code

Chmod +x [run.sh](#)

[./run.sh](#)

### Checking output

hadoop fs -cat /output/part-00000

hadoop fs -cat /output/part-00001

Run.sh file

```
#!/bin/bash
```

```
arg=4
```

define an argument

```
hadoop fs -rm -f -r /input  
hadoop fs -rm -f -r /output
```

Delete folders on HDFS

```
hadoop fs -mkdir /input
```

```
hadoop fs -copyFromLocal ./3littlepigs.txt /input/3littlepigs.txt
```

Create a new folder on HDFS and copy a file to the folder

```
hadoop jar ./hadoop-streaming-3.1.4.jar \  
-D mapred.reduce.tasks=3 \  
-file ./mapperArgument.py \  
-mapper "./mapperArgument.py $arg" \  
-file ./reducer.py \  
-reducer ./reducer.py \  
-input /input/3littlepigs.txt \  
-output /output
```

define an argument

Pig

- Enter pig

pig -x mapreduce

- Exit pig

grunt>quit

Example running sample.pig

- pig -x mapreduce **sample.pig**