

Homework Assignment 1

Name: Quach, Hoa

NetID: gt4789

1. Work *Example 3-3* on Amazon Elastic MapReduce. Type out all the commands in each step of the process and print out a screenshot of the final results in Amazon EMR.

List of Commands

```
javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-client-core-2.6.1.jar:commons-cli-2.0.jar -d . FileSystemDoubleCat.java
```

```
jar -cvf FileSystemDoubleCat.jar FileSystemDoubleCat.class
```

```
hdfs dfs -mkdir /user/hadoop/input
```

```
hdfs dfs -ls /user/hadoop
```

```
hdfs dfs -copyFromLocal quangle.txt .
```

```
hadoop jar FileSystemDoubleCat.jar FileSystemDoubleCat /user/hadoop/quangle.txt /user/hadoop/output
```

OR

```
hadoop FileSystemDoubleCat s3://gt4789pro1/input/quangle.txt
```

```
hadoop@ip-172-31-1-33:~  
But his face you could not see,  
On account of his Beaver Hat.  
On the top of the Crumpetty Tree  
The Quangle Wangle sat,  
But his face you could not see,  
On account of his Beaver Hat.  
16/04/25 02:06:18 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: true maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1461547721572  
16/04/25 02:06:18 INFO metrics.MetricsSaver: Created MetricsSaver j-1C2R1IX5RRK6M:i-024fdecf296d549a8:RunJar:24660 period:60 /mnt/var/em/raw/i-024fdecf296d549a8_20160425_RunJar_24660_raw.bin  
[hadoop@ip-172-31-1-33 ~]$ hdfs dfs -ls /user/hadoop/output  
ls: '/user/hadoop/output': No such file or directory  
[hadoop@ip-172-31-1-33 ~]$ hadoop jar FileSystemDoubleCat.jar FileSystemDoubleCat /user/hadoop/quangle.txt /user/hadoop/output  
On the top of the Crumpetty Tree  
The Quangle Wangle sat,  
But his face you could not see,  
On account of his Beaver Hat.  
On the top of the Crumpetty Tree  
The Quangle Wangle sat,  
But his face you could not see,  
On account of his Beaver Hat.  
16/04/25 02:07:58 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: true maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1461547721572  
16/04/25 02:07:58 INFO metrics.MetricsSaver: Created MetricsSaver j-1C2R1IX5RRK6M:i-024fdecf296d549a8:RunJar:25183 period:60 /mnt/var/em/raw/i-024fdecf296d549a8_20160425_RunJar_25183_raw.bin  
[hadoop@ip-172-31-1-33 ~]$ hadoop FileSystemDoubleCat s3://gt4789ch3/input/quangle.txt  
16/04/25 02:10:14 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon.ws.emr.hadoop.fs.s3n.S3NativeFileSystem as filesystem implementation  
16/04/25 02:10:14 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instanceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: true maxMemoryMb: 3072 maxInstanceCount: 500 lastModified: 1461547721572  
16/04/25 02:10:14 INFO metrics.MetricsSaver: Created MetricsSaver j-1C2R1IX5RRK6M:i-024fdecf296d549a8:FileSystemDoubleCat:25766 period:60 /mnt/var/em/raw/i-024fdecf296d549a8_20160425_FileSystemDoubleCat_25766_raw.bin  
16/04/25 02:10:15 INFO s3n.S3NativeFileSystem: Opening 's3://gt4789ch3/input/quangle.txt' for reading  
16/04/25 02:10:16 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSaver 1  
On the top of the Crumpetty Tree  
The Quangle Wangle sat,  
But his face you could not see,  
On account of his Beaver Hat.  
16/04/25 02:10:16 INFO s3n.S3NativeFileSystem: Stream for key 'input/quangle.txt' seeking to position '0'  
On the top of the Crumpetty Tree  
The Quangle Wangle sat,  
But his face you could not see,  
On account of his Beaver Hat.  
[hadoop@ip-172-31-1-33 ~]$ █
```

2. Work *Example 3-4* on Amazon Elastic MapReduce. Type out all the commands in each step of the process and print out a screenshot of the final results (showing the file directory) in Amazon EMR.

List of Commands

```
javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-client-core-2.6.1.jar:commons-cli-2.0.jar -d . FileCopyWithProgress.java
```

```
jar -cvf FileCopyWithProgress.jar FileCopyWithProgress.class
```

```
hdfs dfs -mkdir /user/hadoop/input (folder already exists from Example 3-3)
```

```
hdfs dfs -copyFromLocal 1400-8.txt .
```

```
hadoop jar FileCopyWithProgress.jar FileCopyWithProgress /home/hadoop/1400-8.txt  
/home/hadoop/output4
```

```
hadoop FileCopyWithProgress 1400-8.txt s3://chapter3-3-4/output/1400-8.txt
```

```
hadoop@ip-172-31-1-33:~  
Exception in thread "main" java.lang.ArrayIndexOutOfBoundsException: 1  
    at FileCopyWithProgress.main(FileCopyWithProgress.java:18)  
[hadoop@ip-172-31-1-33 ~]$ hadoop FileCopyWithProgress input/1400-8.txt s3://chapter3-3-4/input/  
1400-8.txt  
Exception in thread "main" java.io.FileNotFoundException: input/1400-8.txt (No such file or dire  
ctory)  
    at java.io.FileInputStream.open(Native Method)  
    at java.io.FileInputStream.<init>(FileInputStream.java:146)  
    at java.io.FileInputStream.<init>(FileInputStream.java:101)  
    at FileCopyWithProgress.main(FileCopyWithProgress.java:20)  
[hadoop@ip-172-31-1-33 ~]$ hadoop FileCopyWithProgress input/1400-8.txt s3://chapter3-3-4/input/  
Exception in thread "main" java.io.FileNotFoundException: input/1400-8.txt (No such file or dire  
ctory)  
    at java.io.FileInputStream.open(Native Method)  
    at java.io.FileInputStream.<init>(FileInputStream.java:146)  
    at java.io.FileInputStream.<init>(FileInputStream.java:101)  
    at FileCopyWithProgress.main(FileCopyWithProgress.java:20)  
[hadoop@ip-172-31-1-33 ~]$ hadoop FileCopyWithProgress input/1400-8.txt s3://chapter3-3-4/output  
/1400-8.txt  
Exception in thread "main" java.io.FileNotFoundException: input/1400-8.txt (No such file or dire  
ctory)  
    at java.io.FileInputStream.open(Native Method)  
    at java.io.FileInputStream.<init>(FileInputStream.java:146)  
    at java.io.FileInputStream.<init>(FileInputStream.java:101)  
    at FileCopyWithProgress.main(FileCopyWithProgress.java:20)  
[hadoop@ip-172-31-1-33 ~]$ hdfs dfs -ls  
Found 4 items  
-rw-r--r-- 1 hadoop hadoop 1033751 2016-04-25 02:33 1400-8.txt  
drwxr-xr-x - hadoop hadoop 0 2016-04-25 01:54 input  
-rw-r--r-- 1 hadoop hadoop 119 2016-04-25 02:29 output  
-rw-r--r-- 1 hadoop hadoop 119 2016-04-25 02:00 quangle.txt  
[hadoop@ip-172-31-1-33 ~]$ pwd  
/home/hadoop  
[hadoop@ip-172-31-1-33 ~]$ hadoop FileCopyWithProgress 1400-8.txt s3://chapter3-3-4/output/1400-  
8.txt  
16/04/25 02:41:12 INFO fs.EmrFileSystem: Consistency disabled, using com.amazon.ws.emr.hadoop.fs  
.s3n.S3NativeFileSystem as filesystem implementation  
16/04/25 02:41:13 INFO metrics.MetricsSaver: MetricsConfigRecord disabledInCluster: false instan  
ceEngineCycleSec: 60 clusterEngineCycleSec: 60 disableClusterEngine: true maxMemoryMb: 3072 maxI  
nstanceCount: 500 lastModified: 1461547721572  
16/04/25 02:41:13 INFO metrics.MetricsSaver: Created MetricsSaver j-1C2R1IX5RRK6M:i-024fdecf296d  
549a8:FileCopyWithProgress:04263 period:60 /mnt/var/em/raw/i-024fdecf296d549a8_20160425_FileCopy  
WithProgress_04263_raw.bin  
16/04/25 02:41:13 INFO metrics.MetricsSaver: Thread 1 created MetricsLockFreeSaver 1  
16/04/25 02:41:13 INFO s3n.MultipartUploadOutputStream: close closed:false s3://chapter3-3-4/out  
put/1400-8.txt  
.....  
.....16/04/25 02:41:14 INFO s3n.Je  
ts3tNativeFileSystemStore: s3.putObject chapter3-3-4 output/1400-8.txt 1033751  
[hadoop@ip-172-31-1-33 ~]$
```

AWS

Services

Edit

Hoa Quach

Global

Support

Upload

Create Folder

Actions

Search by prefix

None

Properties

Transfers

All Buckets / chapter3-3-4 / output

Name	Storage Class	Size	Last Modified
1400-8.txt	Standard	1009.5 KB	Sun Apr 24 19:41:14 GMT-700 2016