

```
kelly@yunyy98 [?] ~/Desktop/2019_Summer/Ve572/
hw/hw2 [?] ~/hadoop/bin/hadoop jar ./s3distcp-
TRUNK-SNAPSHOT.jar --src=hdfs://localhost:9000/
csv --dest=hdfs://localhost:9000/output/ --groupBy
'./gradebook\_(\..csv)'
2019-06-10 23:25:54,837 INFO s3distcp.S3DistCp:
Running with args: [Ljava.lang.String;@4ffe3d42
2019-06-10 23:25:54,981 WARN
util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes
where applicable
2019-06-10 23:25:55,433 INFO s3distcp.S3DistCp:
Using output path 'hdfs://tmp/2dc96ad2-1690-4c2f-
b80e-cf2e4433e794/output'
2019-06-10 23:26:00,703 INFO s3distcp.S3DistCp:
GET http://169.254.169.254/latest/meta-data/
placement/availability-zone exception
org.apache.commons.httpclient.ConnectTimeoutExcept
ion: The host did not accept the connection within
timeout of 5000 ms
    at
org.apache.commons.httpclient.protocol.ReflectionSoc
ketFactory.createSocket(ReflectionSocketFactory.jav
a:155)
    at
org.apache.commons.httpclient.protocol.DefaultProtoc
olSocketFactory.createSocket(DefaultProtocolSocket
Factory.java:125)
    at
org.apache.commons.httpclient.HttpConnection.open(H
ttpConnection.java:707)
    at
```

```
org.apache.commons.httpclient.HttpMethodDirector.executeWithRetry(HttpMethodDirector.java:387)
    at
org.apache.commons.httpclient.HttpMethodDirector.executeMethod(HttpMethodDirector.java:171)
    at
org.apache.commons.httpclient.HttpClient.executeMethod(HttpClient.java:397)
    at
org.apache.commons.httpclient.HttpClient.executeMethod(HttpClient.java:323)
    at
com.amazon.external.elasticmapreduce.s3distcp.S3DistCp.isGovCloud(S3DistCp.java:185)
    at
com.amazon.external.elasticmapreduce.s3distcp.S3DistCp.run(S3DistCp.java:296)
    at
com.amazon.external.elasticmapreduce.s3distcp.S3DistCp.run(S3DistCp.java:205)
    at
org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
    at
org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:90)
    at
com.amazon.external.elasticmapreduce.s3distcp.Main.main(Main.java:14)
    at java.base/
jdk.internal.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at java.base/
```

```
jdk.internal.reflect.NativeMethodAccessorImpl.invoke(
NativeMethodAccessorImpl.java:62)
    at java.base/
jdk.internal.reflect.DelegatingMethodAccessorImpl.inv
oke(DelegatingMethodAccessorImpl.java:43)
    at java.base/
java.lang.reflect.Method.invoke(Method.java:567)
    at
org.apache.hadoop.util.RunJar.run(RunJar.java:323)
    at
org.apache.hadoop.util.RunJar.main(RunJar.java:236)
Caused by: java.net.SocketTimeoutException: connect
timed out
    at java.base/
java.net.PlainSocketImpl.socketConnect(Native
Method)
    at java.base/
java.net.AbstractPlainSocketImpl.doConnect(Abstract
PlainSocketImpl.java:399)
    at java.base/
java.net.AbstractPlainSocketImpl.connectToAddress(A
bstractPlainSocketImpl.java:242)
    at java.base/
java.net.AbstractPlainSocketImpl.connect(AbstractPla
inSocketImpl.java:224)
    at java.base/
java.net.SocksSocketImpl.connect(SocksSocketImpl.jav
a:403)
    at java.base/
java.net.Socket.connect(Socket.java:591)
    at java.base/
jdk.internal.reflect.NativeMethodAccessorImpl.invoke0
(Native Method)
```

```
    at java.base/
jdk.internal.reflect.NativeMethodAccessorImpl.invoke(
NativeMethodAccessorImpl.java:62)
    at java.base/
jdk.internal.reflect.DelegatingMethodAccessorImpl.in
voke(DelegatingMethodAccessorImpl.java:43)
    at java.base/
java.lang.reflect.Method.invoke(Method.java:567)
    at
org.apache.commons.httpclient.protocol.ReflectionSoc
ketFactory.createSocket(ReflectionSocketFactory.jav
a:140)
```

... 18 more

2019-06-10 23:26:00,713 INFO

s3distcp.FileInfoListing: Opening new file: hdfs:/tmp/
2dc96ad2-1690-4c2f-b80e-cf2e4433e794/files/1

2019-06-10 23:26:00,807 INFO s3distcp.S3DistCp:
Created 1 files to copy 3 files

2019-06-10 23:26:00,925 INFO impl.MetricsConfig:
Loaded properties from hadoop-metrics2.properties

2019-06-10 23:26:01,002 INFO

impl.MetricsSystemImpl: Scheduled Metric snapshot
period at 10 second(s).

2019-06-10 23:26:01,002 INFO

impl.MetricsSystemImpl: JobTracker metrics system
started

2019-06-10 23:26:01,015 WARN

impl.MetricsSystemImpl: JobTracker metrics system
already initialized!

2019-06-10 23:26:01,182 INFO

mapred.FileInputFormat: Total input files to process :
1

2019-06-10 23:26:01,223 INFO

mapreduce.JobSubmitter: number of splits:1  
2019-06-10 23:26:01,346 INFO  
mapreduce.JobSubmitter: Submitting tokens for job:  
job\_local1022848299\_0001  
2019-06-10 23:26:01,346 INFO  
mapreduce.JobSubmitter: Executing with tokens: []  
2019-06-10 23:26:01,447 INFO mapreduce.Job: The  
url to track the job: http://localhost:8080/  
2019-06-10 23:26:01,449 INFO mapreduce.Job:  
Running job: job\_local1022848299\_0001  
2019-06-10 23:26:01,449 INFO  
mapred.LocalJobRunner: OutputCommitter set in  
config null  
2019-06-10 23:26:01,451 INFO  
mapred.LocalJobRunner: OutputCommitter is  
org.apache.hadoop.mapred.FileOutputCommitter  
2019-06-10 23:26:01,454 INFO  
output.FileOutputCommitter: File Output Committer  
Algorithm version is 2  
2019-06-10 23:26:01,455 INFO  
output.FileOutputCommitter: FileOutputCommitter skip  
cleanup \_temporary folders under output  
directory:false, ignore cleanup failures: false  
2019-06-10 23:26:01,494 INFO  
mapred.LocalJobRunner: Waiting for map tasks  
2019-06-10 23:26:01,496 INFO  
mapred.LocalJobRunner: Starting task:  
attempt\_local1022848299\_0001\_m\_000000\_0  
2019-06-10 23:26:01,513 INFO  
output.FileOutputCommitter: File Output Committer  
Algorithm version is 2  
2019-06-10 23:26:01,513 INFO  
output.FileOutputCommitter: FileOutputCommitter skip

cleanup \_temporary folders under output  
directory:false, ignore cleanup failures: false  
2019-06-10 23:26:01,519 INFO  
util.ProcfsBasedProcessTree:  
ProcfsBasedProcessTree currently is supported only  
on Linux.  
2019-06-10 23:26:01,520 INFO mapred.Task: Using  
ResourceCalculatorProcessTree : null  
2019-06-10 23:26:01,522 INFO mapred.MapTask:  
Processing split: hdfs://localhost:9000/tmp/  
2dc96ad2-1690-4c2f-b80e-cf2e4433e794/files/  
1:0+472  
2019-06-10 23:26:01,551 INFO mapred.MapTask:  
numReduceTasks: 1  
2019-06-10 23:26:01,599 INFO mapred.MapTask:  
(EQUATOR) 0 kvi 26214396(104857584)  
2019-06-10 23:26:01,599 INFO mapred.MapTask:  
mapreduce.task.io.sort.mb: 100  
2019-06-10 23:26:01,599 INFO mapred.MapTask:  
soft limit at 83886080  
2019-06-10 23:26:01,599 INFO mapred.MapTask:  
bufstart = 0; bufvoid = 104857600  
2019-06-10 23:26:01,599 INFO mapred.MapTask:  
kvstart = 26214396; length = 6553600  
2019-06-10 23:26:01,603 INFO mapred.MapTask:  
Map output collector class =  
org.apache.hadoop.mapred.MapTask\$MapOutputBuffer  
r  
2019-06-10 23:26:01,607 INFO  
mapred.LocalJobRunner:  
2019-06-10 23:26:01,607 INFO mapred.MapTask:  
Starting flush of map output  
2019-06-10 23:26:01,607 INFO mapred.MapTask:

Spilling map output

2019-06-10 23:26:01,607 INFO mapred.MapTask:

bufstart = 0; bufend = 324; bufvoid = 104857600

2019-06-10 23:26:01,607 INFO mapred.MapTask:

kvstart = 26214396(104857584); kvend =

26214388(104857552); length = 9/6553600

2019-06-10 23:26:01,630 INFO mapred.MapTask:

Finished spill 0

2019-06-10 23:26:01,642 INFO mapred.Task:

Task:attempt\_local1022848299\_0001\_m\_000000\_0

is done. And is in the process of committing

2019-06-10 23:26:01,644 INFO

mapred.LocalJobRunner: hdfs://localhost:9000/tmp/

2dc96ad2-1690-4c2f-b80e-cf2e4433e794/files/

1:0+472

2019-06-10 23:26:01,645 INFO mapred.Task: Task

'attempt\_local1022848299\_0001\_m\_000000\_0'

done.

2019-06-10 23:26:01,649 INFO mapred.Task: Final

Counters for

attempt\_local1022848299\_0001\_m\_000000\_0:

Counters: 23

File System Counters

FILE: Number of bytes read=13331847

FILE: Number of bytes written=13952295

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=472

HDFS: Number of bytes written=472

HDFS: Number of read operations=8

HDFS: Number of large read operations=0

HDFS: Number of write operations=4

HDFS: Number of bytes read erasure-coded=0

Map-Reduce Framework

Map input records=3

Map output records=3

Map output bytes=324

Map output materialized bytes=336

Input split bytes=122

Combine input records=0

Spilled Records=3

Failed Shuffles=0

Merged Map outputs=0

GC time elapsed (ms)=5

Total committed heap usage  
(bytes)=214958080

File Input Format Counters

Bytes Read=472

2019-06-10 23:26:01,649 INFO

mapred.LocalJobRunner: Finishing task:

attempt\_local1022848299\_0001\_m\_000000\_0

2019-06-10 23:26:01,650 INFO

mapred.LocalJobRunner: map task executor complete.

2019-06-10 23:26:01,651 INFO

mapred.LocalJobRunner: Waiting for reduce tasks

2019-06-10 23:26:01,652 INFO

mapred.LocalJobRunner: Starting task:

attempt\_local1022848299\_0001\_r\_000000\_0

2019-06-10 23:26:01,657 INFO

output.FileOutputCommitter: File Output Committer

Algorithm version is 2

2019-06-10 23:26:01,657 INFO

output.FileOutputCommitter: FileOutputCommitter skip  
cleanup \_temporary folders under output



directory:false, ignore cleanup failures: false  
2019-06-10 23:26:01,657 INFO  
util.ProcfsBasedProcessTree:  
ProcfsBasedProcessTree currently is supported only  
on Linux.  
2019-06-10 23:26:01,657 INFO mapred.Task: Using  
ResourceCalculatorProcessTree : null  
2019-06-10 23:26:01,660 INFO mapred.ReduceTask:  
Using ShuffleConsumerPlugin:  
org.apache.hadoop.mapreduce.task.reduce.Shuffle@4  
45e461d  
2019-06-10 23:26:01,661 WARN  
impl.MetricsSystemImpl: JobTracker metrics system  
already initialized!  
2019-06-10 23:26:01,671 INFO  
reduce.MergeManagerImpl: MergerManager:  
memoryLimit=3006477056,  
maxSingleShuffleLimit=751619264,  
mergeThreshold=1984274944, ioSortFactor=10,  
memToMemMergeOutputsThreshold=10  
2019-06-10 23:26:01,673 INFO  
reduce.EventFetcher:  
attempt\_local1022848299\_0001\_r\_000000\_0  
Thread started: EventFetcher for fetching Map  
Completion Events  
2019-06-10 23:26:01,691 INFO reduce.LocalFetcher:  
localfetcher#1 about to shuffle output of map  
attempt\_local1022848299\_0001\_m\_000000\_0  
decomp: 332 len: 336 to MEMORY  
2019-06-10 23:26:01,692 INFO  
reduce.InMemoryMapOutput: Read 332 bytes from  
map-output for  
attempt\_local1022848299\_0001\_m\_000000\_0

2019-06-10 23:26:01,692 INFO  
reduce.MergeManagerImpl: closeInMemoryFile -> map-  
output of size: 332, inMemoryMapOutputs.size() -> 1,  
commitMemory -> 0, usedMemory -> 332  
2019-06-10 23:26:01,693 INFO  
reduce.EventFetcher: EventFetcher is interrupted..  
Returning  
2019-06-10 23:26:01,694 INFO  
mapred.LocalJobRunner: 1 / 1 copied.  
2019-06-10 23:26:01,694 INFO  
reduce.MergeManagerImpl: finalMerge called with 1  
in-memory map-outputs and 0 on-disk map-outputs  
2019-06-10 23:26:01,702 INFO mapred.Merger:  
Merging 1 sorted segments  
2019-06-10 23:26:01,702 INFO mapred.Merger:  
Down to the last merge-pass, with 1 segments left of  
total size: 325 bytes  
2019-06-10 23:26:01,708 INFO  
reduce.MergeManagerImpl: Merged 1 segments, 332  
bytes to disk to satisfy reduce memory limit  
2019-06-10 23:26:01,708 INFO  
reduce.MergeManagerImpl: Merging 1 files, 336 bytes  
from disk  
2019-06-10 23:26:01,709 INFO  
reduce.MergeManagerImpl: Merging 0 segments, 0  
bytes from memory into reduce  
2019-06-10 23:26:01,709 INFO mapred.Merger:  
Merging 1 sorted segments  
2019-06-10 23:26:01,709 INFO mapred.Merger:  
Down to the last merge-pass, with 1 segments left of  
total size: 325 bytes  
2019-06-10 23:26:01,710 INFO  
mapred.LocalJobRunner: 1 / 1 copied.

2019-06-10 23:26:01,725 INFO  
s3distcp.CopyFilesReducer: tempPath:hdfs:/tmp/  
2dc96ad2-1690-4c2f-b80e-cf2e4433e794/  
tempSpace/  
068429b3-5393-49e1-87c5-6be44a56a9ce  
interPath:hdfs://localhost:9000/output  
2019-06-10 23:26:01,726 INFO  
s3distcp.CopyFilesReducer: Processing object:  
hdfs://localhost:9000/csv/gradebook\_3.csv  
2019-06-10 23:26:01,726 INFO  
s3distcp.CopyFilesReducer: Processing object:  
hdfs://localhost:9000/csv/gradebook\_2.csv  
2019-06-10 23:26:01,726 INFO  
s3distcp.CopyFilesReducer: Processing object:  
hdfs://localhost:9000/csv/gradebook\_1.csv  
2019-06-10 23:26:01,726 INFO  
s3distcp.CopyFilesReducer: Processing 3 files  
2019-06-10 23:26:01,728 INFO  
s3distcp.CopyFilesRunnable: Creating  
CopyFilesRunnable hdfs:/tmp/2dc96ad2-1690-4c2f-  
b80e-cf2e4433e794/tempSpace/  
068429b3-5393-49e1-87c5-6be44a56a9ce:hdfs://  
localhost:9000/output/0.csv  
2019-06-10 23:26:01,729 INFO  
s3distcp.CopyFilesRunnable: Opening temp file: hdfs:/  
tmp/2dc96ad2-1690-4c2f-b80e-cf2e4433e794/  
tempSpace/  
068429b3-5393-49e1-87c5-6be44a56a9ce1b7269  
03-03e0-401e-897d-129922adbb3b  
2019-06-10 23:26:01,737 INFO  
s3distcp.CopyFilesRunnable: Starting download of  
hdfs://localhost:9000/csv/gradebook\_3.csv to  
hdfs:/tmp/2dc96ad2-1690-4c2f-b80e-

cf2e4433e794/tempspace/  
068429b3-5393-49e1-87c5-6be44a56a9ce1b7269  
03-03e0-401e-897d-129922addb3b  
2019-06-10 23:26:01,743 INFO  
s3distcp.CopyFilesRunnable: Copied 1720 bytes  
2019-06-10 23:26:01,743 INFO  
s3distcp.CopyFilesRunnable: Finished downloading  
hdfs://localhost:9000/csv/gradebook\_3.csv  
2019-06-10 23:26:01,743 INFO  
s3distcp.CopyFilesRunnable: Starting download of  
hdfs://localhost:9000/csv/gradebook\_2.csv to  
hdfs:/tmp/2dc96ad2-1690-4c2f-b80e-  
cf2e4433e794/tempspace/  
068429b3-5393-49e1-87c5-6be44a56a9ce1b7269  
03-03e0-401e-897d-129922addb3b  
2019-06-10 23:26:01,746 INFO  
s3distcp.CopyFilesRunnable: Copied 212 bytes  
2019-06-10 23:26:01,746 INFO  
s3distcp.CopyFilesRunnable: Finished downloading  
hdfs://localhost:9000/csv/gradebook\_2.csv  
2019-06-10 23:26:01,746 INFO  
s3distcp.CopyFilesRunnable: Starting download of  
hdfs://localhost:9000/csv/gradebook\_1.csv to  
hdfs:/tmp/2dc96ad2-1690-4c2f-b80e-  
cf2e4433e794/tempspace/  
068429b3-5393-49e1-87c5-6be44a56a9ce1b7269  
03-03e0-401e-897d-129922addb3b  
2019-06-10 23:26:01,750 INFO  
s3distcp.CopyFilesRunnable: Copied 344 bytes  
2019-06-10 23:26:01,750 INFO  
s3distcp.CopyFilesRunnable: Finished downloading  
hdfs://localhost:9000/csv/gradebook\_1.csv  
2019-06-10 23:26:01,779 INFO

s3distcp.CopyFilesRunnable: Renaming hdfs:/tmp/  
2dc96ad2-1690-4c2f-b80e-cf2e4433e794/  
tempespace/  
068429b3-5393-49e1-87c5-6be44a56a9ce1b7269  
03-03e0-401e-897d-129922addb3b to hdfs://  
localhost:9000/output/0.csv  
2019-06-10 23:26:01,788 INFO  
s3distcp.CopyFilesReducer: commit hdfs://localhost:  
9000/csv/gradebook\_3.csv  
2019-06-10 23:26:01,789 INFO  
s3distcp.CopyFilesReducer: commit hdfs://localhost:  
9000/csv/gradebook\_2.csv  
2019-06-10 23:26:01,789 INFO  
s3distcp.CopyFilesReducer: commit hdfs://localhost:  
9000/csv/gradebook\_1.csv  
2019-06-10 23:26:01,790 INFO  
s3distcp.CopyFilesReducer: CopyFilesReducer  
uncommitted file 0  
2019-06-10 23:26:01,797 INFO mapred.Task:  
Task:attempt\_local1022848299\_0001\_r\_000000\_0  
is done. And is in the process of committing  
2019-06-10 23:26:01,798 INFO  
mapred.LocalJobRunner: 1 / 1 copied.  
2019-06-10 23:26:01,799 INFO mapred.Task: Task  
attempt\_local1022848299\_0001\_r\_000000\_0 is  
allowed to commit now  
2019-06-10 23:26:01,809 INFO  
output.FileOutputCommitter: Saved output of task  
'attempt\_local1022848299\_0001\_r\_000000\_0' to  
hdfs://localhost:9000/tmp/2dc96ad2-1690-4c2f-  
b80e-cf2e4433e794/output  
2019-06-10 23:26:01,809 INFO  
mapred.LocalJobRunner: reduce > reduce

2019-06-10 23:26:01,810 INFO mapred.Task: Task  
'attempt\_local1022848299\_0001\_r\_000000\_0'  
done.

2019-06-10 23:26:01,810 INFO mapred.Task: Final  
Counters for  
attempt\_local1022848299\_0001\_r\_000000\_0:  
Counters: 30

#### File System Counters

FILE: Number of bytes read=13332551  
FILE: Number of bytes written=13952631  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=2748  
HDFS: Number of bytes written=2748  
HDFS: Number of read operations=16  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=10  
HDFS: Number of bytes read erasure-  
coded=0

#### Map-Reduce Framework

Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=336  
Reduce input records=3  
Reduce output records=0  
Spilled Records=3  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=1  
Total committed heap usage

(bytes)=101711872

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Output Format Counters

Bytes Written=0

2019-06-10 23:26:01,810 INFO

mapred.LocalJobRunner: Finishing task:

attempt\_local1022848299\_0001\_r\_000000\_0

2019-06-10 23:26:01,811 INFO

mapred.LocalJobRunner: reduce task executor  
complete.

2019-06-10 23:26:02,456 INFO mapreduce.Job: Job  
job\_local1022848299\_0001 running in uber mode :  
false

2019-06-10 23:26:02,458 INFO mapreduce.Job:  
map 100% reduce 100%

2019-06-10 23:26:02,460 INFO mapreduce.Job: Job  
job\_local1022848299\_0001 completed successfully

2019-06-10 23:26:02,468 INFO mapreduce.Job:

Counters: 36

File System Counters

FILE: Number of bytes read=26664398

FILE: Number of bytes written=27904926

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=3220

HDFS: Number of bytes written=3220

HDFS: Number of read operations=24  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=14  
HDFS: Number of bytes read erasure-coded=0

#### Map-Reduce Framework

Map input records=3  
Map output records=3  
Map output bytes=324  
Map output materialized bytes=336  
Input split bytes=122  
Combine input records=0  
Combine output records=0  
Reduce input groups=1  
Reduce shuffle bytes=336  
Reduce input records=3  
Reduce output records=0  
Spilled Records=6  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=6  
Total committed heap usage  
(bytes)=316669952

#### Shuffle Errors

BAD\_ID=0  
CONNECTION=0  
IO\_ERROR=0  
WRONG\_LENGTH=0  
WRONG\_MAP=0  
WRONG\_REDUCE=0

#### File Input Format Counters

Bytes Read=472



## File Output Format Counters

Bytes Written=0

2019-06-10 23:26:02,468 INFO s3distcp.S3DistCp:

Try to recursively delete hdfs:/tmp/

2dc96ad2-1690-4c2f-b80e-cf2e4433e794/

tempespace

kelly@yunyy98 [?] ~/Desktop/2019\_Summer/

Ve572/hw/hw2 [?] ~/hadoop/bin/hadoop fs -ls

hdfs://localhost:9000/output/

2019-06-10 23:26:08,656 WARN

util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes  
where applicable

Found 1 items

-rw-r--r-- 1 kelly supergroup 2276

2019-06-10 23:26 hdfs://localhost:9000/output/

0.csv

kelly@yunyy98 [?] ~/Desktop/2019\_Summer/

Ve572/hw/hw2 [?] ~/hadoop/bin/hadoop fs -cat

hdfs://localhost:9000/output/0.csv > read

2019-06-10 23:26:37,371 WARN

util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes  
where applicable