Homework1

Name: Yang Cai
NUID:001632759

Weather Data Results:

1. For each version running without Fibonacci (17):
   Sequential: average = 2231, min = 2080, max = 2658
   No lock: average = 1307, min = 1233, max = 1381
   Coarse lock: average = 1299, min = 1185, max = 1696
   Fine lock: average = 1226, min = 1090, max = 1700
   No share: average = 1242, min = 1105, max = 1762

2. For each version running with Fibonacci (17) when modifying accumulation value:
   Sequential: average = 7895, min = 7492, max = 10220
   No lock: average = 3803, min = 3768, max = 3837
   Coarse lock: average = 6840, min = 6422, max = 9304
   Fine lock: average = 4201, min = 3749, max = 7026
   No share: average = 3876, min = 3772, max = 4432

3. 1). Using 3 threads, speedup of non- Fibonacci:
   No lock: 1.71; Coarse lock: 1.72; Fine lock: 1.82; No share: 1.80
   2). Using 3 threads, speedup of Fibonacci (17):
   No lock: 2.08; Coarse lock: 1.15; Fine lock: 1.88; No share: 2.04

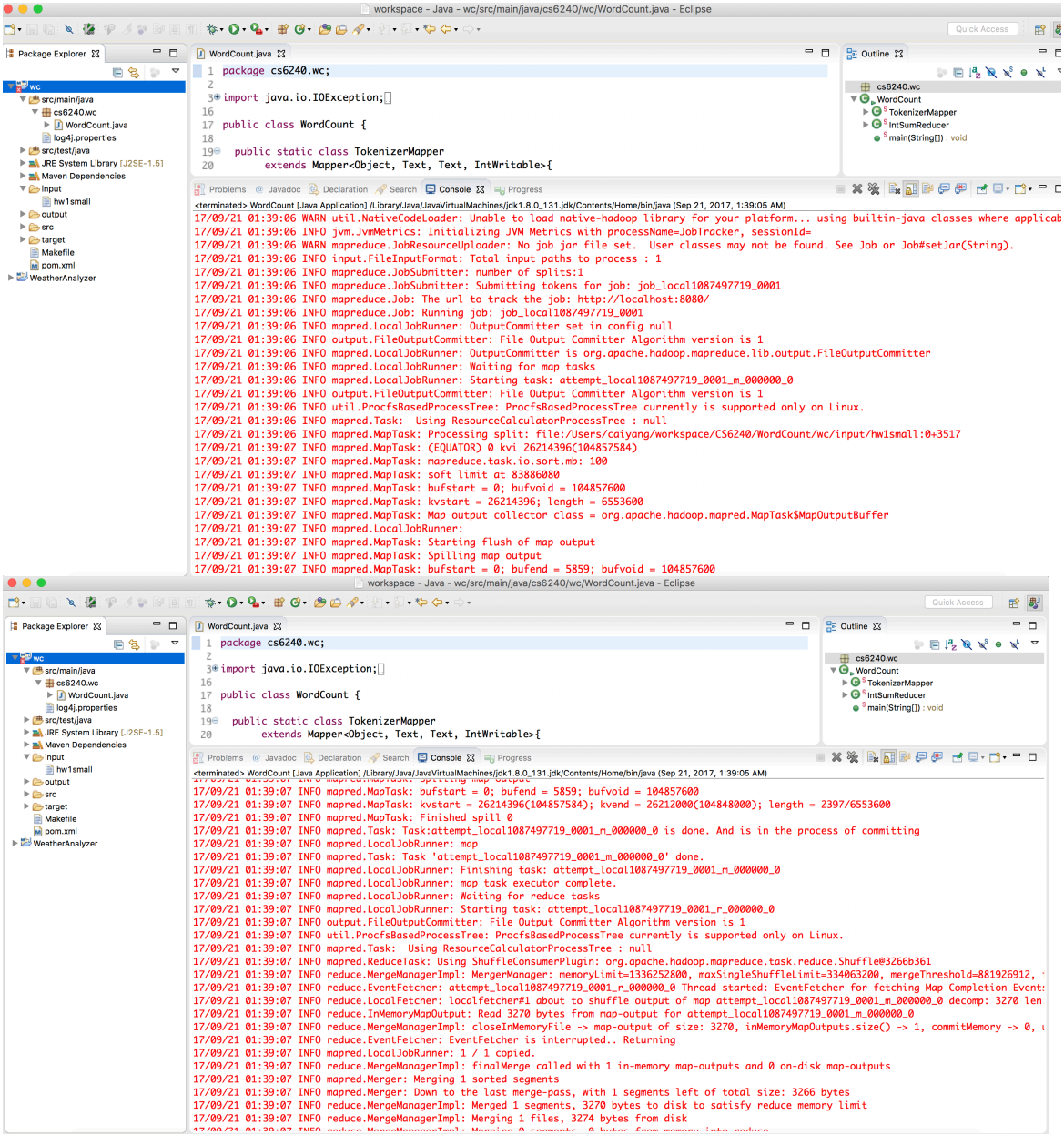
Weather Data Analysis:

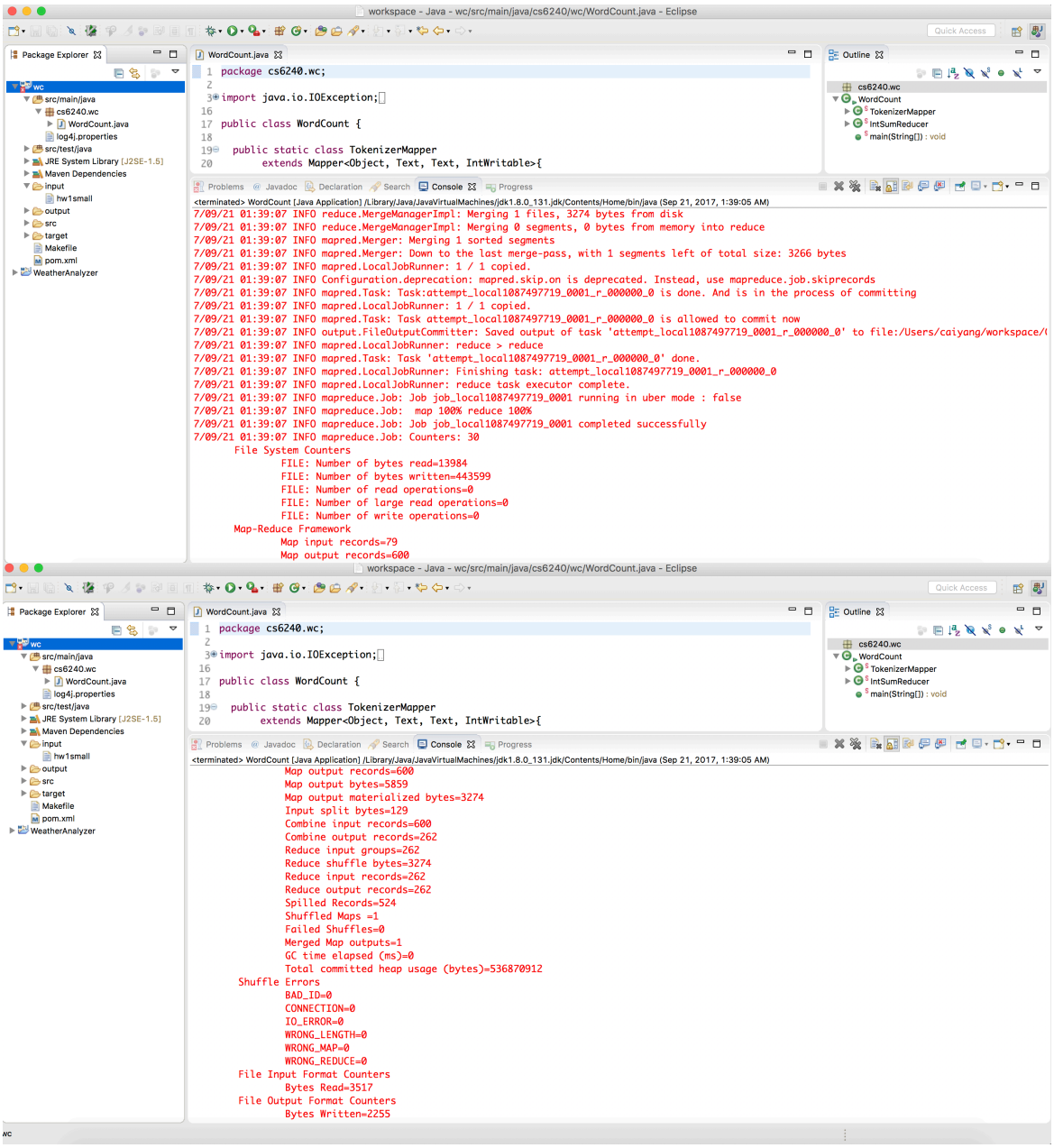
1. I would expect no lock to be the fastest, since there is no synchronization, and there are multiple threads working, also at the end of program it doesn't need aggregate all the result. The result of Fibonacci exactly confirms my expectation: the no lock runs fastest. However, in the result of non- Fibonacci, there isn't much difference between no lock and other parallel executions, it is because that the program runs really fast when there is no Fibonacci, so the lock contention doesn't have too much impact.
2. I would expect sequential to finish the slowest, since there are no parallel at all. The result with Fibonacci confirms this.
3. The no lock program has an incorrect result, it might crash sometimes. This is because there is no lock when executing concurrent program, and when race condition happens, there is an incorrect result. Also, when race condition happens, there might be NullPointerException, this might because when hashmap is resizing, another thread is trying to put/get, thus causing NullPointerException.
4. The sequential is slower than coarse in both B and C. Since there isn't any parallel in sequential, so it is much slower. As for the coarse lock, although its lock contention is

really severe, but at least, there is some parallel when all threads are parsing records and filtering records.

5. Fibonacci computation has more severe impact on COARSE LOCK than FINE LOCK. I think this is because COARSE LOCK will lock the whole table, and the Fibonacci computation has relative high cost, thus make COARSE LOCK performance like a single-thread process: all other threads are waiting for one thread to finish Fibonacci computation. However, in FINE LOCK, we only lock the entry, so all threads can still work as long as they operate on difference entries.

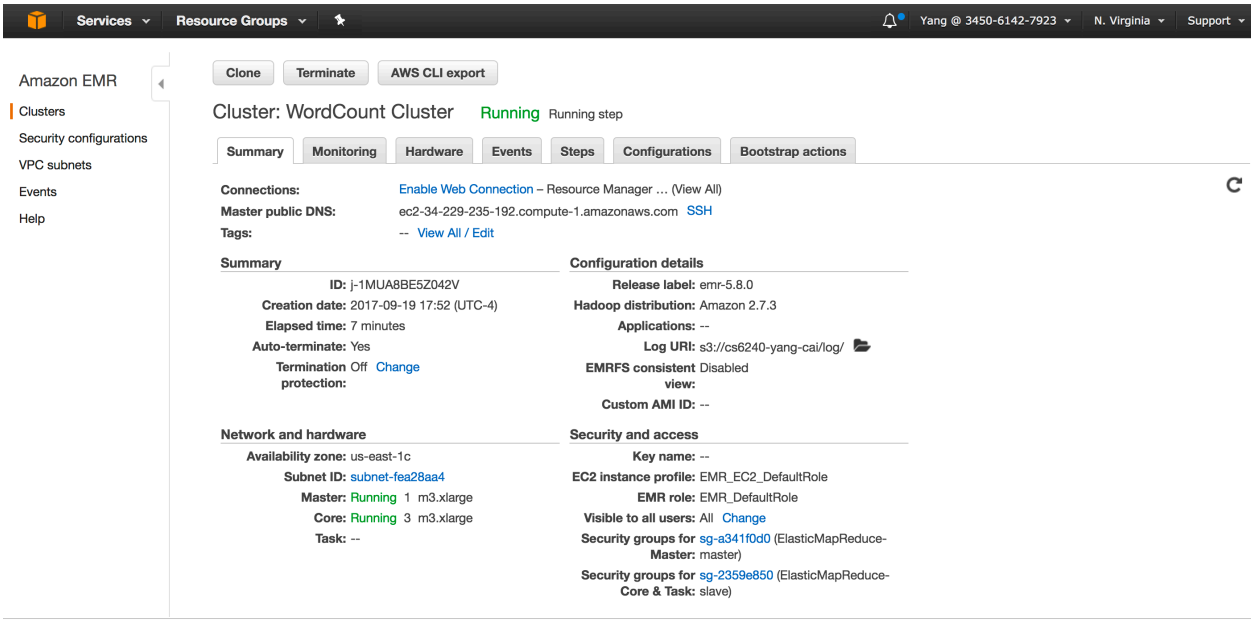
Word Count Local:




workspace - Java - wc/src/main/java/cs6240/wc/WordCount.java - Eclipse
Package Explorer
wc
src/main/java
cs6240.wc
WordCount.java
log4j.properties
src/test/java
JRE System Library [J2SE-1.5]
Maven Dependencies
input
hw1small
output
src
target
Makefile
pom.xml
WeatherAnalyzer
WordCount.java
1 package cs6240.wc;
2
3 import java.io.IOException;
16
17 public class WordCount {
18
19 public static class TokenizerMapper
20 extends Mapper<Object, Text, Text, IntWritable>{
Outline
cs6240.wc
WordCount
TokenizerMapper
IntSumReducer
main(String[]) : void
Problems Javadoc Declaration Search Console Progress
<terminated> WordCount [Java Application] /Library/Java/JavaVirtualMachines/jdk1.8.0_131.jdk/Contents/Home/bin/java (Sep 21, 2017, 1:39:05 AM)
7/09/21 01:39:07 INFO reduce.MergeManagerImpl: Merging 1 files, 3274 bytes from disk
7/09/21 01:39:07 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
7/09/21 01:39:07 INFO mapred.Merger: Merging 1 sorted segments
7/09/21 01:39:07 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 3266 bytes
7/09/21 01:39:07 INFO mapred.LocalJobRunner: 1 / 1 copied.
7/09/21 01:39:07 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
7/09/21 01:39:07 INFO mapred.Task: Task:attempt_local1087497719_0001_r_000000_0 is done. And is in the process of committing
7/09/21 01:39:07 INFO mapred.LocalJobRunner: 1 / 1 copied.
7/09/21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 is allowed to commit now
7/09/21 01:39:07 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1087497719_0001_r_000000_0' to file:/Users/caiyang/workspace/
7/09/21 01:39:07 INFO mapred.LocalJobRunner: reduce > reduce
7/09/21 01:39:07 INFO mapred.Task: Task 'attempt_local1087497719_0001_r_000000_0' done.
7/09/21 01:39:07 INFO mapred.LocalJobRunner: Finishing task: attempt_local1087497719_0001_r_000000_0
7/09/21 01:39:07 INFO mapred.LocalJobRunner: reduce task executor complete.
7/09/21 01:39:07 INFO mapreduce.Job: Job job_local1087497719_0001 running in uber mode : false
7/09/21 01:39:07 INFO mapreduce.Job: map 100% reduce 100%
7/09/21 01:39:07 INFO mapreduce.Job: Job job_local1087497719_0001 completed successfully
7/09/21 01:39:07 INFO mapreduce.Job: Counters: 30
File System Counters
FILE: Number of bytes read=13984
FILE: Number of bytes written=443599
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
Map-Reduce Framework
Map input records=79
Map output records=600
workspace - Java - wc/src/main/java/cs6240/wc/WordCount.java - Eclipse
Map output records=600
Map output bytes=5859
Map output materialized bytes=3274
Input split bytes=129
Combine input records=600
Combine output records=262
Reduce input groups=262
Reduce shuffle bytes=3274
Reduce input records=262
Reduce output records=262
Spilled Records=524
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=536870912
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=3517
File Output Format Counters
Bytes Written=2255

Word Count AWS

```
Yangs-MacBook-Pro:wc caiyang$ make cloud
mvn clean package
[INFO] Scanning for projects...
[INFO]
[INFO] ------------------------------------------------------------------------
[INFO] Building wc 1.0
[INFO] ------------------------------------------------------------------------
[INFO]
[INFO] --- maven-clean-plugin:2.5:clean (default-clean) @ wc ---
[INFO] Deleting /Users/caiyang/workspace/CS6240/WordCount/wc/target
[INFO]
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ wc ---
[WARNING] Using platform encoding (UTF-8 actually) to copy filtered resources, i.e. build is platform dependent!
[INFO] skip non existing resourceDirectory /Users/caiyang/workspace/CS6240/WordCount/wc/src/main/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.3:compile (default-compile) @ wc ---
[INFO] Changes detected - recompiling the module!
[WARNING] File encoding has not been set, using platform encoding UTF-8, i.e. build is platform dependent!
[INFO] Compiling 1 source file to /Users/caiyang/workspace/CS6240/WordCount/wc/target/classes
[INFO]
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ wc ---
[WARNING] Using platform encoding (UTF-8 actually) to copy filtered resources, i.e. build is platform dependent!
[INFO] skip non existing resourceDirectory /Users/caiyang/workspace/CS6240/WordCount/wc/src/test/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.3:testCompile (default-testCompile) @ wc ---
[INFO] Changes detected - recompiling the module!
[WARNING] File encoding has not been set, using platform encoding UTF-8, i.e. build is platform dependent!
[INFO] Compiling 1 source file to /Users/caiyang/workspace/CS6240/WordCount/wc/target/test-classes
[INFO]
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ wc ---
[INFO] Surefire report directory: /Users/caiyang/workspace/CS6240/WordCount/wc/target/surefire-reports

-------------------------------------------------------
 T E S T S
-------------------------------------------------------
Running cs6240.wc.AppTest
Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.036 sec

Results :

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0

[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ wc ---
[INFO] Building jar: /Users/caiyang/workspace/CS6240/WordCount/wc/target/wc-1.0.jar
[INFO]
[INFO] --- maven-shade-plugin:2.4.3:shade (default) @ wc ---
[INFO] Replacing original artifact with shaded artifact.
[INFO] Replacing /Users/caiyang/workspace/CS6240/WordCount/wc/target/wc-1.0.jar with /Users/caiyang/workspace/CS6240/WordCount/wc/target/wc-1.0-shaded.jar
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time: 3.322 s
[INFO] Finished at: 2017-09-19T17:52:47-04:00
[INFO] Final Memory: 28M/283M
[INFO] ------------------------------------------------------------------------

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0

[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ wc ---
[INFO] Building jar: /Users/caiyang/workspace/CS6240/WordCount/wc/target/wc-1.0.jar
[INFO]
[INFO] --- maven-shade-plugin:2.4.3:shade (default) @ wc ---
[INFO] Replacing original artifact with shaded artifact.
[INFO] Replacing /Users/caiyang/workspace/CS6240/WordCount/wc/target/wc-1.0.jar with /Users/caiyang/workspace/CS6240/WordCount/wc/target/wc-1.0-shaded.jar
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time: 3.322 s
[INFO] Finished at: 2017-09-19T17:52:47-04:00
[INFO] Final Memory: 28M/283M
[INFO] ------------------------------------------------------------------------
aws s3 cp target/wc-1.0.jar s3://cs6240-yang-cai
upload: target/wc-1.0.jar to s3://cs6240-yang-cai/wc-1.0.jar
aws s3 rm s3://cs6240-yang-cai/ --recursive --exclude "*" --include "output*"
aws emr create-cluster \
                --name "WordCount Cluster" \
                --release-label emr-5.8.0 \
                --instance-groups '[{"InstanceCount":3,"InstanceGroupType":"CORE","InstanceType":"m3.xlarge"},{"InstanceCount":1,"InstanceGroupType":"MASTER","InstanceType":"m3.xlarge"}]' \
            --applications Name=Hadoop \
            --steps '[{"Args":["cs6240.wc.WordCount","s3://cs6240-yang-cai/input","s3://cs6240-yang-cai/output"],"Type":"CUSTOM_JAR","Jar":"s3://cs6240-yang-cai/wc-1.0.jar","ActionOnFailure":"TERMINATE_CL
USTER","Name":"Custom JAR"}]' \
                --log-uri s3://cs6240-yang-cai/log \
                --service-role EMR_DefaultRole \
                --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-fea28aa4 \
                --region us-east-1 \
                --enable-debugging \
                --auto-terminate
j-1MUA8BESZ042V
Yangs-MacBook-Pro:wc caiyang$
```



Cluster: WordCount Cluster Terminated Steps completed

Summary | Monitoring | Hardware | Events | Steps | Configurations | Bootstrap actions

Time	Event description	Source ID	Source type	Event type	Severity	Full date & time
Sep 19 06:03 PM	Amazon EMR Cluster j-1MUA8BE5Z042V (WordCount Cluster) has terminated at 2017-09-19 22:02 UTC with a reason of ALL_STEPS_COMPLETED.	j-1MUA8BE5Z042V	Cluster	Cluster State Change	INFO	September 19, 2017 at 06:03:16 PM (UTC-4)
Sep 19 06:01 PM	Amazon EMR cluster j-1MUA8BE5Z042V (WordCount Cluster) finished running all pending steps at 2017-09-19 22:00 UTC.	j-1MUA8BE5Z042V	Cluster	Cluster State Change	INFO	September 19, 2017 at 06:01:21 PM (UTC-4)
Sep 19 06:01 PM	Step s-B9J92KNSEEDR (Custom JAR) in Amazon EMR cluster j-1MUA8BE5Z042V (WordCount Cluster) completed execution at 2017-09-19 22:00 UTC. The step started running at 2017-09-19 21:58 UTC and took 2 minutes to complete.	s-B9J92KNSEEDR	Step	Step State Change	INFO	September 19, 2017 at 06:01:20 PM (UTC-4)
Sep 19 05:58 PM	Step s-B9J92KNSEEDR (Custom JAR) in Amazon EMR cluster j-1MUA8BE5Z042V (WordCount Cluster) started running at 2017-09-19 21:58 UTC.	s-B9J92KNSEEDR	Step	Step State Change	INFO	September 19, 2017 at 05:58:46 PM (UTC-4)
Sep 19 05:58 PM	Step s-2O3YU93AJRQU2 (Setup Hadoop Debugging) in Amazon EMR cluster j-1MUA8BE5Z042V (WordCount Cluster) completed execution at 2017-09-19 21:58 UTC. The step started running at 2017-09-19 21:58 UTC and took 0 minutes to complete.	s-2O3YU93AJRQU2	Step	Step State Change	INFO	September 19, 2017 at 05:58:46 PM (UTC-4)
Sep 19 05:58 PM	Amazon EMR cluster j-1MUA8BE5Z042V (WordCount Cluster) began running steps at 2017-09-19 21:58 UTC.	j-1MUA8BE5Z042V	Cluster	Cluster State Change	INFO	September 19, 2017 at 05:58:29 PM (UTC-4)