

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.

- 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:

The image shows two side-by-side screenshots of the Eclipse IDE interface. Both screenshots display the same Java code for a Word Count application, which reads from a file named 'hw1small' and prints word counts to the console. The code uses a static class TokenizerMapper that extends Mapper<Object, Text, IntWritable>. The execution logs below the code show the output of the job for each Java version.

Java Version 1 (Screenshot Top):

```

1 package cs6240.wc;
2
3 import java.io.IOException;
4
5 public class WordCount {
6
7     public static class TokenizerMapper
8         extends Mapper<Object, Text, IntWritable>{
9
10        @Override
11        protected void map(Object key, Text value, Context context) throws IOException, InterruptedException {
12            StringTokenizer itr = new StringTokenizer(value.toString());
13            while (itr.hasMoreTokens()) {
14                context.write(itr.nextToken(), new IntWritable(1));
15            }
16        }
17
18        @Override
19        protected void reduce(Text key, IntWritable value, Context context) throws IOException, InterruptedException {
20            context.write(key, value);
21        }
22    }
23
24    public static void main(String[] args) throws Exception {
25        Configuration conf = new Configuration();
26        Job job = new Job(conf, "WordCount");
27        job.setMapperClass(TokenizerMapper.class);
28        job.setCombinerClass(IntSumReducer.class);
29        job.setReducerClass(IntSumReducer.class);
30        job.setOutputKeyClass(Text.class);
31        job.setOutputValueClass(IntWritable.class);
32        FileInputFormat.addInputPath(job, new Path(args[0]));
33        FileOutputFormat.setOutputPath(job, new Path(args[1]));
34        job.waitForCompletion(true);
35    }
36}

```

Java Version 2 (Screenshot Bottom):

```

1 package cs6240.wc;
2
3 import java.io.IOException;
4
5 public class WordCount {
6
7     public static class TokenizerMapper
8         extends Mapper<Object, Text, IntWritable>{
9
10        @Override
11        protected void map(Object key, Text value, Context context) throws IOException, InterruptedException {
12            StringTokenizer itr = new StringTokenizer(value.toString());
13            while (itr.hasMoreTokens()) {
14                context.write(itr.nextToken(), new IntWritable(1));
15            }
16        }
17
18        @Override
19        protected void reduce(Text key, IntWritable value, Context context) throws IOException, InterruptedException {
20            context.write(key, value);
21        }
22    }
23
24    public static void main(String[] args) throws Exception {
25        Configuration conf = new Configuration();
26        Job job = new Job(conf, "WordCount");
27        job.setMapperClass(TokenizerMapper.class);
28        job.setCombinerClass(IntSumReducer.class);
29        job.setReducerClass(IntSumReducer.class);
30        job.setOutputKeyClass(Text.class);
31        job.setOutputValueClass(IntWritable.class);
32        FileInputFormat.addInputPath(job, new Path(args[0]));
33        FileOutputFormat.setOutputPath(job, new Path(args[1]));
34        job.waitForCompletion(true);
35    }
36}

```

Execution Log for Java 8.0.131 (Top Screenshot):

```

17/09/21 01:39:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
17/09/21 01:39:06 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
17/09/21 01:39:06 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
17/09/21 01:39:06 INFO input.FileInputFormat: Total input paths to process : 1
17/09/21 01:39:06 INFO mapred.JobSubmitter: number of splits:1
17/09/21 01:39:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1087497719_0001
17/09/21 01:39:06 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
17/09/21 01:39:06 INFO mapreduce.Job: Running job: job_local1087497719_0001
17/09/21 01:39:06 INFO mapred.LocalJobRunner: OutputCommitter set in config null
17/09/21 01:39:06 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/21 01:39:06 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
17/09/21 01:39:06 INFO mapred.LocalJobRunner: Waiting for map tasks
17/09/21 01:39:06 INFO mapred.LocalJobRunner: Starting task: attempt_local1087497719_0001_m_000000_0
17/09/21 01:39:06 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/21 01:39:06 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
17/09/21 01:39:06 INFO mapred.Task: Using ResourceCalculatorProcessTree : null
17/09/21 01:39:06 INFO mapred.MapTask: Processing split: file:/Users/cayyang/workspace/CS6240/WordCount/wc/input/hw1small:0+3517
17/09/21 01:39:07 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(10485784)
17/09/21 01:39:07 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/09/21 01:39:07 INFO mapred.MapTask: soft limit at 83886080
17/09/21 01:39:07 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/09/21 01:39:07 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
17/09/21 01:39:07 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
17/09/21 01:39:07 INFO mapred.LocalJobRunner:
17/09/21 01:39:07 INFO mapred.MapTask: Starting Flush of map output
17/09/21 01:39:07 INFO mapred.MapTask: Spilling map output
17/09/21 01:39:07 INFO mapred.MapTask: bufstart = 0; bufend = 5859; bufvoid = 104857600

```

Execution Log for Java 1.8.0_131 (Bottom Screenshot):

```

17/09/21 01:39:07 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
17/09/21 01:39:07 INFO mapred.MapTask: soft limit at 83886080
17/09/21 01:39:07 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
17/09/21 01:39:07 INFO mapred.MapTask: kvstart = 26214396(10485784); kvend = 26212000(104848000); length = 2397/6553600
17/09/21 01:39:07 INFO mapred.MapTask: finished spill 0
17/09/21 01:39:07 INFO mapred.Task: Task:attempt_local1087497719_0001_m_000000_0 is done. And is in the process of committing
17/09/21 01:39:07 INFO mapred.LocalJobRunner: map
17/09/21 01:39:07 INFO mapred.Task: Task 'attempt_local1087497719_0001_m_000000_0' done.
17/09/21 01:39:07 INFO mapred.LocalJobRunner: Finishing task: attempt_local1087497719_0001_m_000000_0
17/09/21 01:39:07 INFO mapred.LocalJobRunner: map task executor complete.
17/09/21 01:39:07 INFO mapred.LocalJobRunner: Waiting for reduce tasks
17/09/21 01:39:07 INFO mapred.LocalJobRunner: Starting task: attempt_local1087497719_0001_r_000000_0
17/09/21 01:39:07 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
17/09/21 01:39:07 INFO util.ProcfsBasedProcessTree: ProcfsBasedProcessTree currently is supported only on Linux.
17/09/21 01:39:07 INFO mapred.Task: Using ResourceCalculatorProcessTree : null
17/09/21 01:39:07 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@3266b361
17/09/21 01:39:07 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=1336252800, maxSingleShuffleLimit=334063200, mergeThreshold=881926912, mergeThreshold=881926912
17/09/21 01:39:07 INFO reduce.EventFetcher: attempt_local1087497719_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Event:
17/09/21 01:39:07 INFO reduce.LocalFetcher: localFetcher#1 about to shuffle output of map attempt_local1087497719_0001_m_000000_0 decmp: 3270 len: 17/09/21 01:39:07 INFO reduce.InMemoryMapOutput: Read 3270 bytes from map-output for attempt_local1087497719_0001_m_000000_0
17/09/21 01:39:07 INFO reduce.InMemoryMapOutput: closeInMemoryFile -> map-output of size: 3270, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, inMemoryMapOutputs.size() -> 1
17/09/21 01:39:07 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
17/09/21 01:39:07 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
17/09/21 01:39:07 INFO mapred.Merger: Merging 1 sorted segments
17/09/21 01:39:07 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 3266 bytes
17/09/21 01:39:07 INFO reduce.MergeManagerImpl: Merged 1 segments, 3270 bytes to disk to satisfy reduce memory limit
17/09/21 01:39:07 INFO reduce.MergeManagerImpl: Merging 1 files, 3274 bytes from disk
17/09/21 01:39:07 INFO reduce.MergeManagerImpl: Merging 1 files, 3274 bytes from disk

```

The screenshot shows two instances of the Eclipse IDE running on a Mac OS X desktop. Both instances have the same workspace setup, featuring a 'WordCount' Java project with a 'src/main/java' directory containing a 'WordCount.java' file and a 'pom.xml' file. The 'pom.xml' file includes dependencies for 'JRE System Library [J2SE-1.5]' and 'Maven Dependencies'. The 'WordCount.java' file contains the code for a MapReduce job to count word occurrences.

Top Window (Left):

```
1 package cs6240.wc;
2
3 import java.io.IOException;
4
5 public class WordCount {
6
7     public static class TokenizerMapper
8         extends Mapper<Object, Text, Text, IntWritable>{
9
10        @Override
11        protected void map(Object key, Text value, Context context) throws IOException, InterruptedException {
12            StringTokenizer itr = new StringTokenizer(value.toString());
13            while (itr.hasMoreTokens()) {
14                context.write(itr.nextToken(), new IntWritable(1));
15            }
16        }
17
18        @Override
19        protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
20            int sum = 0;
21            for (IntWritable val : values) {
22                sum += val.get();
23            }
24            context.write(key, new IntWritable(sum));
25        }
26    }
27
28    public static class IntSumReducer
29        extends Reducer<Text, IntWritable, Text, IntWritable> {
30
31        @Override
32        protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
33            int sum = 0;
34            for (IntWritable val : values) {
35                sum += val.get();
36            }
37            context.write(key, new IntWritable(sum));
38        }
39    }
40
41    public static void main(String[] args) throws Exception {
42        Configuration config = new Configuration();
43        config.set("mapreduce.job.reduces", "1");
44        Job job = new Job(config, "WordCount");
45        job.setMapperClass(TokenizerMapper.class);
46        job.setCombinerClass(IntSumReducer.class);
47        job.setReducerClass(IntSumReducer.class);
48        job.setOutputKeyClass(Text.class);
49        job.setOutputValueClass(IntWritable.class);
50
51        FileInputFormat.addInputPath(new Path(args[0]), job);
52        FileOutputFormat.setOutputPath(new Path(args[1]), job);
53
54        job.waitForCompletion(true);
55    }
56}
```

Bottom Window (Right):

```
1 package cs6240.wc;
2
3 import java.io.IOException;
4
5 public class WordCount {
6
7     public static class TokenizerMapper
8         extends Mapper<Object, Text, Text, IntWritable>{
9
10        @Override
11        protected void map(Object key, Text value, Context context) throws IOException, InterruptedException {
12            StringTokenizer itr = new StringTokenizer(value.toString());
13            while (itr.hasMoreTokens()) {
14                context.write(itr.nextToken(), new IntWritable(1));
15            }
16        }
17
18        @Override
19        protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
20            int sum = 0;
21            for (IntWritable val : values) {
22                sum += val.get();
23            }
24            context.write(key, new IntWritable(sum));
25        }
26    }
27
28    public static class IntSumReducer
29        extends Reducer<Text, IntWritable, Text, IntWritable> {
30
31        @Override
32        protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {
33            int sum = 0;
34            for (IntWritable val : values) {
35                sum += val.get();
36            }
37            context.write(key, new IntWritable(sum));
38        }
39    }
40
41    public static void main(String[] args) throws Exception {
42        Configuration config = new Configuration();
43        config.set("mapreduce.job.reduces", "1");
44        Job job = new Job(config, "WordCount");
45        job.setMapperClass(TokenizerMapper.class);
46        job.setCombinerClass(IntSumReducer.class);
47        job.setReducerClass(IntSumReducer.class);
48        job.setOutputKeyClass(Text.class);
49        job.setOutputValueClass(IntWritable.class);
50
51        FileInputFormat.addInputPath(new Path(args[0]), job);
52        FileOutputFormat.setOutputPath(new Path(args[1]), job);
53
54        job.waitForCompletion(true);
55    }
56}
```

Logs:

Top Window Log:

```
2017-09-21 01:39:07 INFO reduce.MergeManagerImpl: Merging 1 files, 3274 bytes from disk
2017-09-21 01:39:07 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2017-09-21 01:39:07 INFO mapred.Merger: Merging 1 sorted segments
2017-09-21 01:39:07 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 3266 bytes
2017-09-21 01:39:07 INFO mapred.LocalJobRunner: 1 / 1 copied.
2017-09-21 01:39:07 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2017-09-21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 is done. And is in the process of committing
2017-09-21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 is allowed to commit now
2017-09-21 01:39:07 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1087497719_0001_r_000000_0' to file:/Users/caiyang/workspace/cs6240/wc/_temp/_local/_attempt_local1087497719_0001_r_000000_0
2017-09-21 01:39:07 INFO mapred.LocalJobRunner: reduce > reduce
2017-09-21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 done.
2017-09-21 01:39:07 INFO mapred.LocalJobRunner: Finishing task: attempt_local1087497719_0001_r_000000_0
2017-09-21 01:39:07 INFO mapreduce.Job: Job job_local1087497719_0001 running in uber mode : false
2017-09-21 01:39:07 INFO mapreduce.Job: map 100% reduce 100%
2017-09-21 01:39:07 INFO mapreduce.Job: Job job_local1087497719_0001 completed successfully
2017-09-21 01:39:07 INFO mapreduce.Job: Counters
  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
  Map Reduce Counters
    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
```

Bottom Window Log:

```
2017-09-21 01:39:07 INFO reduce.MergeManagerImpl: Merging 1 files, 3274 bytes from disk
2017-09-21 01:39:07 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
2017-09-21 01:39:07 INFO mapred.Merger: Merging 1 sorted segments
2017-09-21 01:39:07 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 3266 bytes
2017-09-21 01:39:07 INFO mapred.LocalJobRunner: 1 / 1 copied.
2017-09-21 01:39:07 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2017-09-21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 is done. And is in the process of committing
2017-09-21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 is allowed to commit now
2017-09-21 01:39:07 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1087497719_0001_r_000000_0' to file:/Users/caiyang/workspace/cs6240/wc/_temp/_local/_attempt_local1087497719_0001_r_000000_0
2017-09-21 01:39:07 INFO mapred.LocalJobRunner: reduce > reduce
2017-09-21 01:39:07 INFO mapred.Task: Task attempt_local1087497719_0001_r_000000_0 done.
2017-09-21 01:39:07 INFO mapred.LocalJobRunner: Finishing task: attempt_local1087497719_0001_r_000000_0
2017-09-21 01:39:07 INFO mapreduce.Job: Job job_local1087497719_0001 running in uber mode : false
2017-09-21 01:39:07 INFO mapreduce.Job: map 100% reduce 100%
2017-09-21 01:39:07 INFO mapreduce.Job: Job job_local1087497719_0001 completed successfully
2017-09-21 01:39:07 INFO mapreduce.Job: Counters
  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
  Map Reduce Counters
    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:3.0: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:3.0:testCompile (default-testCompile) @ wc ---
[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

-----
TESTS

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] --- 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] --- 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_CLUSTER"}]'
  --log-uri s3://cs6240-yang-cai/log \
  --ec2-attributes EmrDefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-fea28aa4 \
  --region us-east-1 \
  --enable-debugging \
  --auto-terminate
j-1MUABE5Z042V
Yangs-MacBook-Pro:wc caiyang$
```

Services ▾ Resource Groups ▾

Clone Terminate AWS CLI export

Amazon EMR

Clusters Security configurations VPC subnets Events Help

Cluster: WordCount Cluster Running Running step

Summary Monitoring Hardware Events Steps Configurations Bootstrap actions

Connections: Enable Web Connection – Resource Manager ... (View All)

Master public DNS: ec2-34-229-235-192.compute-1.amazonaws.com SSH

Tags: -- View All / Edit

Summary

ID: j-1MUA8BE5Z042V
Creation date: 2017-09-19 17:52 (UTC-4)
Elapsed time: 7 minutes
Auto-terminate: Yes
Termination Off Change protection:
Custom AMI ID: --

Configuration details

Release label: emr-5.8.0
Hadoop distribution: Amazon 2.7.3
Applications: --
Log URI: s3://cs6240-yang-cai/log/

EMRFS consistent view:
Disabled

Network and hardware

Availability zone: us-east-1c
Subnet ID: subnet-fea28aa4
Master: Running 1 m3.xlarge
Core: Running 3 m3.xlarge
Task: --

Security and access

Key name: --
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole
Visible to all users: All Change
Security groups for sg-a341fdd0 (ElasticMapReduce-Master: master)
Security groups for sg-2359e850 (ElasticMapReduce-Core & Task: slave)

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)