

Findings:

1. The source code is taken from hadoop-1.2.1 “src/examples” folder.
2. AWS EC2 works well with java 1.7. I initially ran the code with java 1.8 and it failed then I google the error and it says that you need to have 1.6 or 1.7 to ran the hadoop code.

Local Execution

The screenshot shows an IDE with the following components:

- Package Explorer:** Shows the project structure with 'WordCount' and 'src' folders. The 'src' folder contains 'org.apache.hadoop.examples' and 'WordCount.java'.
- WordCount.java:** The source code is displayed, showing a map-reduce job. The code includes a `main` method that takes command-line arguments and a `mapreduce` method that implements the map-reduce logic.
- Console:** Displays the output of the program. The output shows the progress of the map-reduce job, including the number of map and reduce tasks, the amount of data processed, and the final result.

```

15/01/21 23:21:33 INFO mapred.LocalJobRunner:
15/01/21 23:21:33 INFO mapred.Task: Task:attempt_local1634717174_0001_r_000000_0 is done. And is in the process of committing
15/01/21 23:21:33 INFO mapred.LocalJobRunner:
15/01/21 23:21:33 INFO mapred.Task: Task:attempt_local1634717174_0001_r_000000_0 is allowed to commit now
15/01/21 23:21:33 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1634717174_0001_r_000000_0' to out
15/01/21 23:21:33 INFO mapred.LocalJobRunner: reduce > reduce
15/01/21 23:21:33 INFO mapred.Task: Task:attempt_local1634717174_0001_r_000000_0 done.
15/01/21 23:21:33 INFO mapred.JobClient: map 100% reduce 100%
15/01/21 23:21:33 INFO mapred.JobClient: Job complete: job_local1634717174_0001
15/01/21 23:21:33 INFO mapred.JobClient: Counters: 17
15/01/21 23:21:33 INFO mapred.JobClient:   Map-Reduce Framework
15/01/21 23:21:33 INFO mapred.JobClient:     Spilled Records=366
15/01/21 23:21:33 INFO mapred.JobClient:     Map output materialized bytes=2454
15/01/21 23:21:33 INFO mapred.JobClient:     Reduce input records=183
15/01/21 23:21:33 INFO mapred.JobClient:     Map input records=7
15/01/21 23:21:33 INFO mapred.JobClient:     SPLIT_RAW_BYTES=135
15/01/21 23:21:33 INFO mapred.JobClient:     Map output bytes=3018
15/01/21 23:21:33 INFO mapred.JobClient:     Reduce shuffle bytes=0
15/01/21 23:21:33 INFO mapred.JobClient:     Reduce input groups=183
15/01/21 23:21:33 INFO mapred.JobClient:     Combine output records=183
15/01/21 23:21:33 INFO mapred.JobClient:     Reduce output records=183
15/01/21 23:21:33 INFO mapred.JobClient:     Map output records=294
15/01/21 23:21:33 INFO mapred.JobClient:     Combine input records=294
15/01/21 23:21:33 INFO mapred.JobClient:     Total committed heap usage (bytes)=458227712
15/01/21 23:21:33 INFO mapred.JobClient: File Input Format Counters
15/01/21 23:21:33 INFO mapred.JobClient:   Bytes Read=1845
15/01/21 23:21:33 INFO mapred.JobClient: File System Counters
15/01/21 23:21:33 INFO mapred.JobClient:   FILE_BYTES_WRITTEN=109384
15/01/21 23:21:33 INFO mapred.JobClient:   FILE_BYTES_READ=6526
15/01/21 23:21:33 INFO mapred.JobClient: File Output Format Counters
15/01/21 23:21:33 INFO mapred.JobClient:   Bytes Written=1744

```

AWS Execution

The screenshot shows the AWS Elastic MapReduce console for a cluster named 'wordcount'. The cluster is in a 'Waiting' state, waiting for the step to complete. The console displays various details about the cluster, including connections, master public DNS, tags, summary, configuration details, network and hardware, and security and access.

Cluster: wordcount **Waiting** Waiting after step completed

Connections: [Enable Web Connection](#) – Job Tracker ... (View All)

Master public DNS: ec2-54-173-209-103.compute-1.amazonaws.com [SSH](#)

Tags: -- [View All / Edit](#)

Summary

- ID: j-3NDV83P9J0ZYM
- Creation date: 2015-01-22 02:05 (UTC-5)
- Elapsed time: 24 minutes
- Auto-terminate: No
- Termination protection: [Change](#)

Configuration Details

- AMI version: 2.4.8
- Hadoop Amazon 1.0.3
- distribution: --
- Applications: --
- Log URI: s3://cs6240jalpan/output/hw1/log/
- EMRFS Disabled
- consistent view:

Network and Hardware [Resize](#)

- Availability zone: us-east-1d
- Subnet ID: subnet-f1851086
- Master: **Running** 1 m1.small
- Core: **Running** 2 m1.small
- Task: --

Security and Access

- Key name: jalpanccs
- EC2 instance profile: EMR_EC2_DefaultRole
- EMR role: EMR_DefaultRole
- Visible to all: All [Change](#)
- users:
- Security groups: sg-026ebc66 for Master: (ElasticMapReduce-master)
- Security groups: sg-036ebc67 for Core & Task: (ElasticMapReduce-slave)

Monitoring

Hardware

Steps

[Add step](#) [Clone step](#)

Steps [View all interactive jobs](#) | [View all jobs](#)

Filter: All steps [Filter steps ...](#) 2 steps (all loaded)

| ID | Name | Status | Start time (UTC-5) | Elapsed time | Log files | Actions |
|---------------|-----------|---------|--------------------------|--------------|---|---------------------------|
| s-H00IWBHODDL | wordcount | Running | 2015-01-22 02:12 (UTC-5) | 8 minutes | controller syslog stderr stdout | View jobs |

JAR location: s3://cs6240jalpan/input/wordcount.jar

Main class: None

Arguments: s3://cs6240jalpan/input/hw1.txt s3://cs6240jalpan/output/hw1out

Action on failure: Terminate cluster

syslog.txt controller.txt jalpanccs.pem [Show All](#)