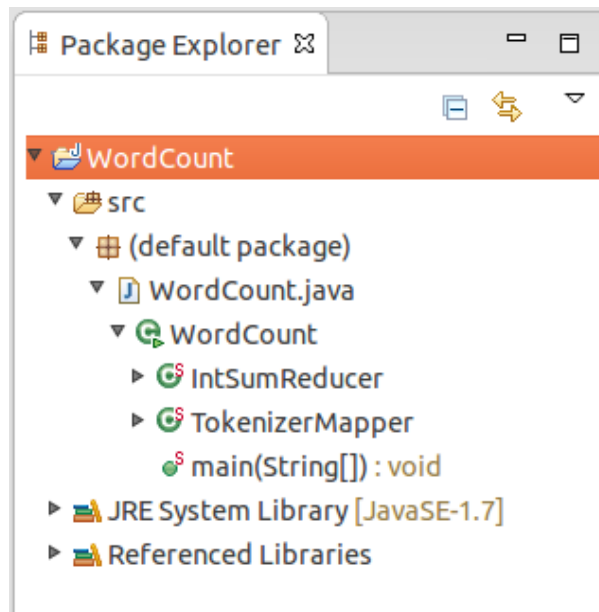


Report

Homework 1
Mukul Sharma
Class : CS 6240 - 02

Directory Structure :



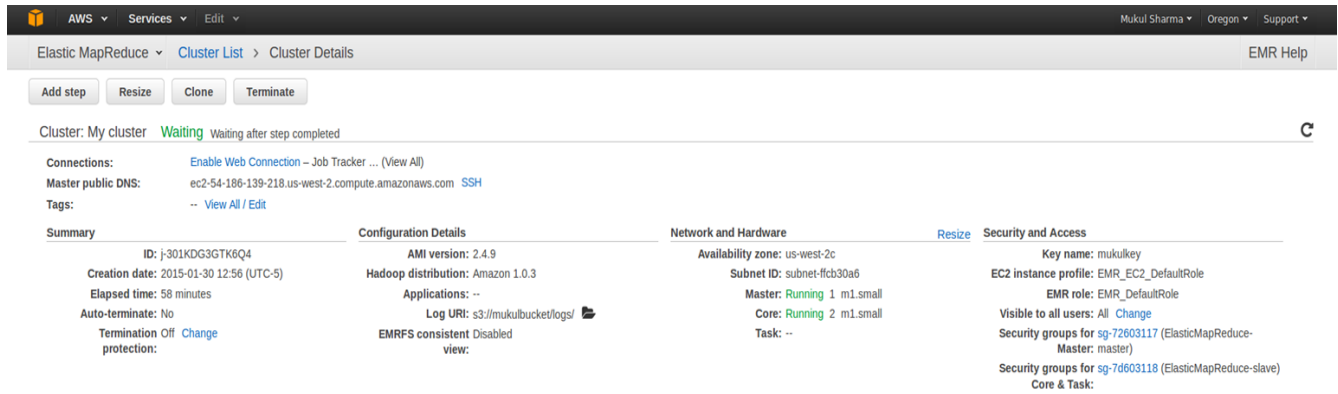
Console output- eclipse :

```
<terminated> WordCount [Java Application] /usr/lib/jvm/java-7-openjdk-amd64/bin/java (Jan 29, 2015, 7:44:06 PM)
15/01/29 19:48:55 INFO mapred.JobClient: Job complete: job_local1899387034_0001
15/01/29 19:48:55 INFO mapred.JobClient: Counters: 20
15/01/29 19:48:55 INFO mapred.JobClient:   File Output Format Counters
15/01/29 19:48:55 INFO mapred.JobClient:     Bytes Written=73395
15/01/29 19:48:55 INFO mapred.JobClient:   FileSystemCounters
15/01/29 19:48:55 INFO mapred.JobClient:     FILE_BYTES_READ=37420290510
15/01/29 19:48:55 INFO mapred.JobClient:     FILE_BYTES_WRITTEN=2838076571
15/01/29 19:48:55 INFO mapred.JobClient:   File Input Format Counters
15/01/29 19:48:55 INFO mapred.JobClient:     Bytes Read=1454183628
15/01/29 19:48:55 INFO mapred.JobClient: Map-Reduce Framework
15/01/29 19:48:55 INFO mapred.JobClient:   Reduce input groups=5273
15/01/29 19:48:55 INFO mapred.JobClient:   Map output materialized bytes=3265636
15/01/29 19:48:55 INFO mapred.JobClient:   Combine output records=5258005
15/01/29 19:48:55 INFO mapred.JobClient:   Map input records=21907700
15/01/29 19:48:55 INFO mapred.JobClient:   Reduce shuffle bytes=0
15/01/29 19:48:55 INFO mapred.JobClient:   Physical memory (bytes) snapshot=0
15/01/29 19:48:55 INFO mapred.JobClient:   Reduce output records=5273
15/01/29 19:48:55 INFO mapred.JobClient:   Spilled Records=8861318
15/01/29 19:48:55 INFO mapred.JobClient:   Map output bytes=2418234700
15/01/29 19:48:55 INFO mapred.JobClient:   CPU time spent (ms)=0
15/01/29 19:48:55 INFO mapred.JobClient:   Total committed heap usage (bytes)=36595826688
15/01/29 19:48:55 INFO mapred.JobClient:   Virtual memory (bytes) snapshot=0
15/01/29 19:48:55 INFO mapred.JobClient:   Combine input records=253969493
15/01/29 19:48:55 INFO mapred.JobClient:   Map output records=248943500
15/01/29 19:48:55 INFO mapred.JobClient:   SPLIT_RAW_BYTES=5412
15/01/29 19:48:55 INFO mapred.JobClient:   Reduce input records=232012
```

Report

Homework 1
Mukul Sharma
Class : CS 6240 - 02

Successful run on AWS:




The screenshot shows the AWS Elastic MapReduce console for a cluster named 'My cluster'. The cluster is in a 'Waiting' state. The page is divided into several sections: Summary, Configuration Details, Network and Hardware, and Security and Access. The Summary section shows the cluster ID, creation date, and elapsed time. The Configuration Details section shows the AMI version, Hadoop distribution, and applications. The Network and Hardware section shows the availability zone, subnet, and instance types. The Security and Access section shows the key name, EC2 instance profile, and security groups.

Cluster: My cluster **Waiting** Waiting after step completed

Connections: [Enable Web Connection](#) - Job Tracker ... (View All)
Master public DNS: ec2-54-186-139-218.us-west-2.compute.amazonaws.com SSH
Tags: -- [View All / Edit](#)

Summary ID: j-301KDG3GTK6Q4
Creation date: 2015-01-30 12:56 (UTC-5)
Elapsed time: 58 minutes
Auto-terminate: No
Termination protection: Off [Change](#)

Configuration Details AMI version: 2.4.9
Hadoop distribution: Amazon 1.0.3
Applications: --
Log URI: s3://mukulbucket/logs/ 
EMRFS consistent view: Disabled

Network and Hardware Availability zone: us-west-2c
Subnet ID: subnet-fcb30a6
Master: **Running** 1 m1.small
Core: **Running** 2 m1.small
Task: --

Security and Access Key name: mukulkey
EC2 instance profile: EMR_EC2_DefaultRole
EMR role: EMR_DefaultRole
Visible to all users: All [Change](#)
Security groups for sg-72603117 (ElasticMapReduce-Master: master)
Security groups for sg-7d603118 (ElasticMapReduce-slave)
Core & Task:

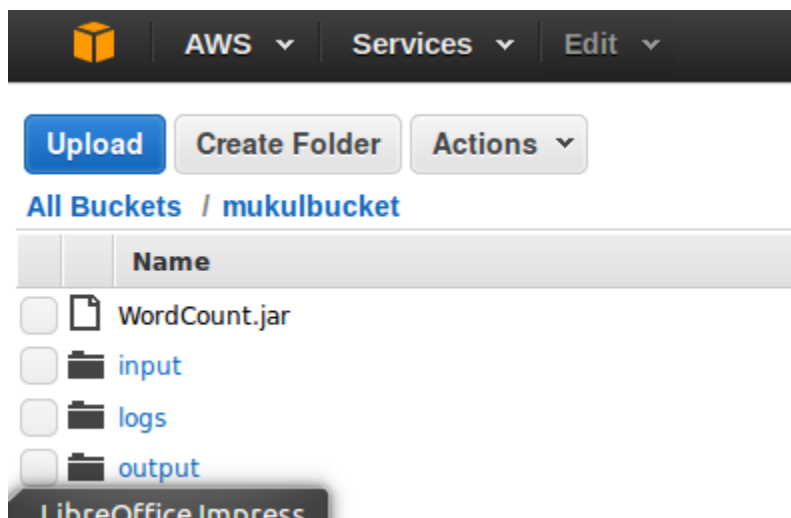
Job Completion Aws:

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

Filter: All steps 2 steps (all loaded)

ID	Name	Status	Start time (UTC-5)	Elapsed time	Log files	Actions
s-2VGAJZNOM22OK	Custom JAR	Completed	2015-01-30 13:02 (UTC-5)	13 minutes	controller syslog stderr stdout	View jobs
s-49GFSQWFXY24	Setup hadoop debugging	Completed	2015-01-30 13:01 (UTC-5)	26 seconds	controller syslog stderr stdout	View jobs

S3 Directory:



The screenshot shows the AWS S3 console interface. At the top, there are buttons for 'Upload', 'Create Folder', and 'Actions'. Below these, the path 'All Buckets / mukulbucket' is displayed. The main area shows a list of objects in the bucket. The first object is 'WordCount.jar', followed by folders named 'input', 'logs', and 'output'. The 'LibreOffice Impress' folder is partially visible at the bottom.

Upload **Create Folder** **Actions**

All Buckets / mukulbucket

	Name
<input type="checkbox"/>	WordCount.jar
<input type="checkbox"/>	input
<input type="checkbox"/>	logs
<input type="checkbox"/>	output
<input type="checkbox"/>	LibreOffice Impress

Report

Homework 1
Mukul Sharma
Class : CS 6240 - 02

Output:

