Submission is a single ZIP file named CS6240\_SHABBIR\_SAIFEE\_3.ZIP

Folders and Files:

1. src folder which contains source code,pom.xml and makefile
2. Run-1 and Run-2 : part1-00000 which aws output file,syslog and output file(part-r-00000)

-----------------------------------------------------------------------------------------------------------------

**How to executeStand alone application:**

­Open Makefile and change all the variables before “# Pseudo-Cluster Execution” namely hadoop root directory, job.name(specified inside driver class of java file), and jar name specified in pom.xml file where artifactId and groupId form the jar name.

Open terminal.

Then we navigate to the src directory where the java programs and the makefile is present

Type: make alone

**To run programs on AWS**

Open MakeFile and under “# AWS EMR Execution”:

aws.emr.release=emr-5.2.1(version of emr)

aws.region=us-east-2(specify the region under AWS)

aws.bucket.name=shabbirbucket(specify the bucket name in S3 directory)

aws.subnet.id=subnet-6356553a

aws.input=s3://shabbirbucket/hw2/IO/input/1991.csv(specify the input file from S3)

aws.output=s3://shabbirbucket/hw2/IO/output1(specify the output file from S3)

aws.log.dir=s3:// shabbirbucket /hw2/og1((specify the log file from S3))

aws.num.nodes=6(specify the number of nodes to run the program on)

aws.instance.type=m4.large(specify the type of machine to run it on)

Open terminal.

Type: make cloud

and that will create a cluster and give the output in S3. So we navigate to s3 and see the output file after successful termination of the program.