Please follow the below steps to build and execute both the standalone and AWS Elastic MapReduce (EMR) versions of my program:

Step 1 - On unzipping my solution you will get following deliverables in one folder:

1_Report - pdf report as required

2_SourceCode – source code (Java Maven Project) & Makefile (see step 2 for running the code)

3_AWSSyslogFiles - plain text syslog files

- Wikipedia-full-html-syslog-6-machines (1 Master & 5 Workers)
- Wikipedia-full-html-syslog-11-machines (1 Master & 10 Workers)
- Wikipedia-simple-html-syslog-6-machines (1 Master & 5 Workers)

4_AWSOutputFiles - output files (Top-100 pages with pagerank)

- Wikipedia-full-html-output-6-machines (1 Master & 5 Workers)
- Wikipedia-full-html-output-11-machines (1 Master & 10 Workers)
- Wikipedia-simple-html-output-6-machines (1 Master & 5 Workers)

Step 2: Steps to build and execute the program on local and AWS (location: 2_SourceCode/PageRank)

2.1 Copy the input file(s) into input folder (PageRank/input)

2.2 Open terminal and cd to this PageRank directory

2.3 You can configure the value of alpha, noOfIteration and kForTopK in Makefile:
- args.alpha : alph to be used in PageRank calculation
- args.noOfIteration : no of PageRank iterations to run
- args.kForTopK : K for getting Top-K records

2.4 Execute in Local
- Run the below command to execute the program in local
  "make alone"
- Output files will be generated in the same directory.

2.5 Execute on AWS
- Run the below commands to execute the program on AWS
  "make upload-input-aws" : command to upload the input to aws (it will copy the input files into dspatel28 bucket)
  "make cloud" : to execute the program/job on AWS

- After execution, you can find the output results and logs in dspatel28 bucket
- At last, you can delete all the data from dspatel28 bucket by running below command:
  "make delete-s3-aws"