**README**

Deliverables

1. Source code for No Combiner, Combiner, In Mapper Combiner and Secondary Sort as a zipped file named CS6240\_Neha\_Pradhan\_2
2. Report
3. The log files (syslog) for the above programs on AWS
4. Output file(s) of the above programs run on AWS.

Running the source code:

1. Extract the src.zip file from the Source Code folder.
2. Create a new Maven project in the IDE and copy the hw2 folder from the extracted folder into the src folder of the newly created Maven project.
3. Copy the pom.xml and MakeFile to the newly created Maven Project.
4. Create an input folder in the project for the input files.
5. Please type the below commands to run the programs after installation of AWS CLI:

|  |  |  |
| --- | --- | --- |
| Program | Local Run | AWS Run |
| No Combiner | make alone-no-combiner | make cloud-no-combiner |
| Combiner | make alone-combiner | make cloud-combiner |
| In Mapper Combiner | make alone-inmapper-combiner | make cloud-inmapper-combiner |
| Secondary Sort | make alone-secondary-sort | make cloud-secondary-sort |

1. The output will be created in folders named according to the algorithms implemented

Report:

Report contains the pseudo code for all the Map Reduce algorithms implemented as part of this assignment along with the performance comparison of the run-time of the different algorithms.

Log and Output Files:

The logs and output files are present in the EMR\_Results folder. The output part files and logs are found in each of the folders named after algorithms implemented. They are present in the following format:

* Folder for Output Files: <Algorithm Name>\_output
* Logs\_1 contains the syslog for the first run.
* Logs\_2 contains the syslog for the second run as well as the output part files for the second run.