**README**

Deliverables

1. Source code for Preprocessor, PageRank and TopK evaluation as a zipped file
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 on AWS:

1. Extract the src.zip file from the Source Code folder.
2. Create a new Maven project in the IDE and copy the hw3 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:

make cloud – for run on 6 machines

make cloud11 – for run on 11 machines

1. The output will be created in folders named as follows:

Output folder name of run on 6 machines – output\_6machines

Output folder name of run on 11 machines – output\_11machines

There are sub folders created during intermediate phases between jobs. Output of the preprocessing stage can be found in the folder output\_0. Output of 10 iterations of page rank evaluation is found in folders named output\_<i> where i = 1, 2, …, 10. Output of the TopK evaluation is found in folder output\_TopKResults

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 processes.

Log and Output Files:

The logs and top-k output files are present in the Outputs folder.