**README**

Deliverables

1. Source code for PageRank in Spark as a zipped file
2. Report
3. The log files (stderr) 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 add the extracted folder into the newly created Maven project and add the hierarchy to the Java build path. Ex: The project should have the following folder structure src/main/scala and src/main/java configured in the Java build path. Place the hw4 folder provided in the same hierarchical folder structure in the zipped file.
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. For local execution, please uncomment the configuration line in the Driver.scala program where the setMaster has argument local[\*] and comment out the one with argument yarn.
6. Please type the below commands to run the programs after installation of AWS CLI:

make cloud – for run on 6 machines on AWS

make cloud11 – for run on 11 machines on AWS

make alone – for local execution

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

output\_spark – for run on 6 machines on AWS

output\_spark\_11machines – for run on 11 machines in AWS

output\_spark – for local execution

Report:

Report contains the program discussion along with the performance comparison of the run-time of PageRank evaluation on different machines.

Log and Output Files:

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