

Readme

The submission contains the following files:

1. **HW2:** Contains the source code of all programs.
2. **Output:** Contains the output from a successful run for each of the programs. The output folders inside are named after the programs e.g. output folder for Combiner is named Combiner.
3. **Syslogs:** Contains the syslogs from one successful run for each of the programs. The syslogs are named after the programs e.g. syslog for Combiner is named Combiner.
4. **Report:** The pdf containing the report and findings.
5. **Readme:** This file

How to Compile

Kindly go into the HW2 folder and then make the project. This will also create a copy of the jar named "hw2.jar" and place it in this folder.

The MakeFile contains all the builds necessary for local and EMR run for all the programs. The MakeFile has been commented to help with the same. Kindly refer to it.

Some useful commands:

make: builds the jar

make local-noCombiner: Runs NoCombiner on local machine

make local-Combiner: Runs Combiner on local machine

make local-inMapCombiner: Runs In Map Combiner on local machine

make local-secSort: Runs Secondary sort on local machine

make s3-all: creates s3 bucket, uploads the jar and 1991.csv file to s3

make s3-all-secSort: creates s3 bucket, uploads the jar and the entire directory containing the files 1880.csv to 1889.csv

make s3-remove: deletes the bucket

make emr-noCombiner: creates and runs EMR instance for No Combiner

make emr-combiner: creates and runs EMR instance for Combiner

make emr-inMapCombiner: creates and runs EMR instance for in Map Combiner

make emr-secSort: creates and runs EMR instance for Secondary sort