**Assignment-1: MapReduce Program**

Following is a brief description of what each class does in the MapReduce Program.

1. **File Generator Class**

The file generator class uses BufferedWriter and FileWriter class to generate file with random data. The data consists of userid, brandid, timestamp each separated by tab.

1. **Driver Class**

The driver class sets the configuration for all the Mappers, Reducers and Partitioner through the Job configuration object. Hadoop uses those configurations to run the map-reduce programs. All file input and output paths are also defined in Driver class. There can be as many mappers and reducers, but for this assignment I have used 2 mappers, 2 reducers and 1 partitioner class.

1. **Mappers and Reducers Classes**

The first mapper takes the random data as input file and processes it to prepare a key-value pair. Usually the key is the offset of the line and the data becomes the value. This can be customized based on requirement. Here we take the userid as the key and the combination of timestamp and brandid as the value, which is being collected by the collector and passed to the first reducer.

The reducer-1 then sorts the values based on increasing order of timestamp for each userid. The output of the reducer-1 is key-userid and value-sorted values of brand ids. This output is given as an input to the next mapper.

Mapper-2 creates pairs of brands for and within the same userid. The output becomes key- brand pairs, value- 1. This value has to be LongWritable.

Reducer-2 sums up all the values for each of the pairs. The output of Reducer-2 is the final output.

1. **Partitioner Class**

The partitioner class creates 3 partitions based on the values of userids. This partitioner is used for the first job, which means it acts between Mapper-1 and Reducer-1.

**Hadoop Commands to run the MapReduce Program**

Go to the installed Hadoop 1.0.3 directory and run the following commands in sequence.

1. Start Hadoop- ./bin/start-all.sh
2. Import input file to Hadoop directory- ./bin/Hadoop fs –copyFromLocal /Users/himansubadhai/Documents/input/ inputMapReduce.txt /Users/himansubadhai/input
3. Export the jar file and run it- ./bin/hadoop jar Users/himansubadhai/Documents/eclipse/assignment1.jar com.mapper.DriverClass edu.uic.ids594.DriverClass
4. Get the Output and store in local- ./bin/hadoop fs -copyToLocal /Users/himansubadhai/output.txt /Users/himansubadhai/Documents /Output
5. Stop Hadoop- ./bin/stop-all.sh