**Assignment-2: Frequent Itemset Mining Program**

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

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 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 input key as Text: trans\_i and the input value is the rest content in each line. The output key is Text: the singleton or doubleton or tripleton itemsets. The output value is IntWritable(1). Here is the example output of Mapper-1:

1 2 5 ⇒ (1,1), (2,1), (5,1), ((1,2),1), ((1,5),1), ((2,5),1), ((1,2,5),1)

2 3 ⇒ (2,1), (3,1), ((2,3),1)

2 4 5 ⇒ (2,1), (4,1), (5,1), ((2,4),1), ((2,5),1), ((4,5),1), ((2,4,5),1)

1 2 ⇒ (1,1), (2,1), ((1,2),1)

1 5 ⇒ (1,1), (5,1), ((1,5),1)

This output is given to Reducer-1, which counts the sum for each unique pair.

The output of the reducer-1 is key-item sets and value-sum of each itemset.

In the example, case, it will generate the following outputs:

(1,3),(2,4),(3,1),(4,1),(5,3),((1,2),2),((1,5),2),((2,3),1),((2,4),1),((2,5),2),((4,5),1) ((1,2,5),1), ((2,4,5),1)

The output of the Reducer-1 is then stored in a local directory folder using command line.

1. **MapResults Class**

This class reads all the values from the text files from the intermediate output and calculates the confidence for all singletons, doubletons and tripletons. This is the final output of the program.

The final output looks like this:

999,8816-->6656 1.0

6656,8816-->999 1.0

(999)-->(6858) 0.2

(6858)-->(999) 0.1

999,6858-->7895 1.0

999,7895-->6858 1.0

6858,7895-->999 1.0

(999)-->(7153) 0.2

(7153)-->(999) 0.2

999,7153-->8030 1.0

999,8030-->7153 1.0

7153,8030-->999 1.0

999,7153-->8816 1.0

999,8816-->7153 1.0

7153,8816-->999 1.0

(999)-->(7652) 0.2

(7652)-->(999) 0.25

(999)-->(7895) 0.2

(7895)-->(999) 0.0625

(999)-->(8030) 0.2

(8030)-->(999) 0.16666666666666666

999,8030-->8816 1.0

999,8816-->8030 1.0

8030,8816-->999 1.0

(999)-->(8347) 0.2

(8347)-->(999) 1.0

1. **Partitioner Class**

The partitioner class creates 3 partitions based on the itemsets whether it is singleton, doubleton and tripleton. This partitioner is used for the first job, which means it acts between Mapper-1 and Reducer-1.

**Hadoop Commands to run the Frequent Itemset Mining Program**

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

**Step1:** **go to Hadoop directory**

cd Documents/Softwares/hadoop-1.0.3/

**Step2: start hadoop**

./bin/start-all.sh

**Step3:** **copy the input file to hdfs**

./bin/hadoop fs -copyFromLocal /Users/himansubadhai/Documents/input/sample/sample.txt /user/himansubadhai/input

**Step4:** **check if the input file is imported properly**

./bin/hadoop fs -ls input

Desired Result:

Found 2 items

-rw-r--r-- 3 himansubadhai supergroup 1600 2014-09-16 20:15 /user/himansubadhai/input/inputMapReduce.txt

-rw-r--r-- 3 himansubadhai supergroup 29 2014-10-02 00:00 /user/himansubadhai/input/sample.txt

**Step5:** **run the Driver class using the jar file**

./bin/hadoop jar /Users/himansubadhai/Documents/input/sample/FrequentItemsetMining.jar edu.uic.ids594.assignment2.DriverClass

**Step6: check if the intermediate output file is generated**

./bin/hadoop fs –ls

Desired Output:

Found 3 items

drwxr-xr-x - himansubadhai supergroup 0 2014-09-24 18:27 /user/himansubadhai/hadoop-mahout

drwxr-xr-x - himansubadhai supergroup 0 2014-10-02 00:00 /user/himansubadhai/input

drwxr-xr-x - himansubadhai supergroup 0 2014-10-02 00:10 /user/himansubadhai/intermediate\_output

**Step7: copy the generated intermediate output to local**

./bin/hadoop fs -copyToLocal /user/himansubadhai/intermediate\_output /Users/himansubadhai/Documents/input/sample/intermediate\_output

**Step8:** **stop hadoop**

Stop Hadoop- ./bin/stop-all.sh