Amazon Binning of Ratings

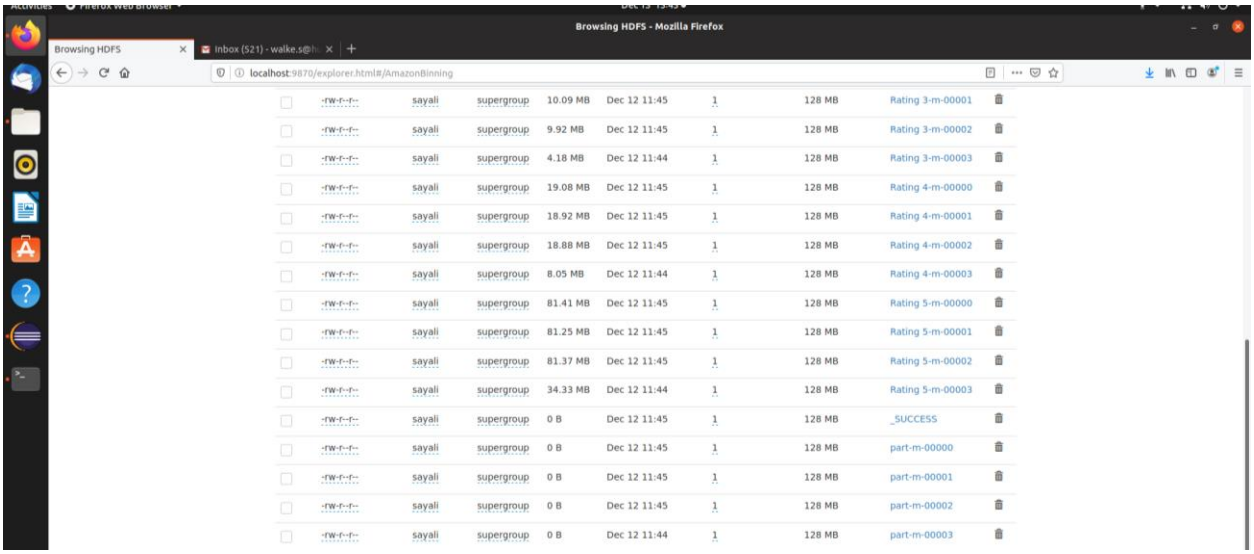
(Organization Pattern)

This is similar to the partitioning pattern described above but the only difference is that this is a mapper-only job and the bins have to be specified in the mapper setup method.
Here I am separating all the products into 5 different bins based on rating(1,2,3,4,5).

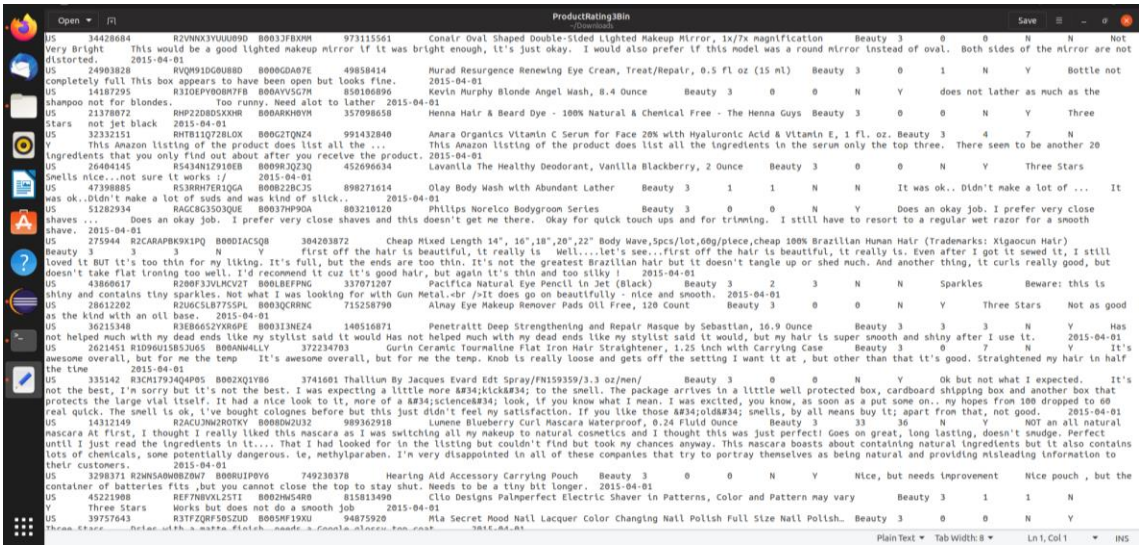
OUTPUT:

./hadoop jar /home/sayali/Desktop/AmazonBinning.jar sayali.AmazonReviews7.AmazonBinningOfRating /AmazonReviews/AmazonReviews.tsv /AmazonBinning

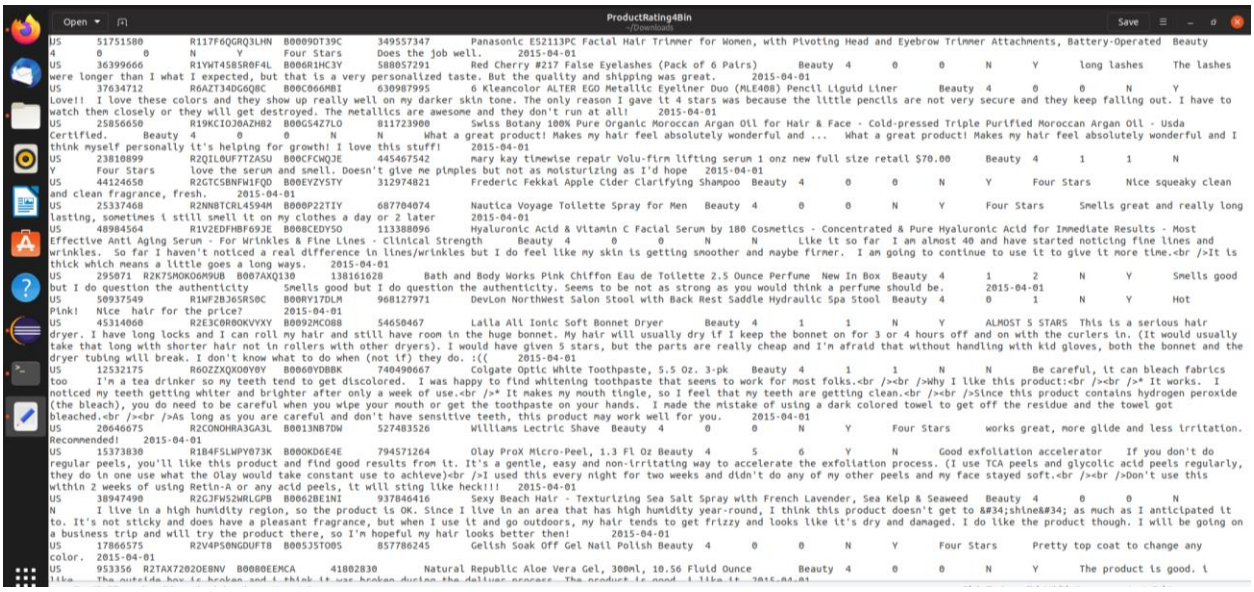
There are 5 Bins for each rating:



Product Rating 3 Bin:



Product Rating 4 Bin:



Map Reduce Code

```
package sayali.AmazonReviews7;
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.NullWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.MultipleOutputs;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
```

Mapper:

```
public class AmazonBinningMapper extends Mapper<LongWritable, Text, Text, NullWritable> {
private MultipleOutputs<Text, NullWritable> mos = null;
		protected void setup(Context context) {
		mos = new MultipleOutputs(context); }
protected void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException {
		if (key.get() == 0) {
			return;
		}
		String[] token = value.toString().split("\\t");
		// for(String rate: token)
		{
```

```
                String rating = token[7].trim();
                if (rating.equals("1")) {
                        mos.write("bins", value, NullWritable.get(), "Rating 1");
                }
                if (rating.equals("2")) {
                        mos.write("bins", value, NullWritable.get(), "Rating 2");
                }
                if (rating.equals("3")) {
                        mos.write("bins", value, NullWritable.get(), "Rating 3");
                }
                if (rating.equals("4")) {
                        mos.write("bins", value, NullWritable.get(), "Rating 4");
                }
                if (rating.equals("5")) {
                        mos.write("bins", value, NullWritable.get(), "Rating 5");
                }
        }
        protected void cleanup(Context context) throws IOException, InterruptedException {
                mos.close();
        }
}
```

Driver:

```
public class AmazonBinningOfRating {
 public static void main(String[] args)throws IOException, InterruptedException,
ClassNotFoundException {
      Configuration conf = new Configuration();
      Job job = Job.getInstance(conf, "Binning Hour");
      job.setJarByClass(AmazonBinningOfRating.class);
      //Setting Mapper Class and the output key and value
      job.setMapperClass(AmazonBinningMapper.class);
      job.setMapOutputKeyClass(Text.class);
      job.setMapOutputValueClass(NullWritable.class);
      //No combiner, partitioner or reducer is used in this pattern!
      job.setNumReduceTasks(0);
      TextInputFormat.addInputPath(job, new Path(args[0]));
      FileOutputFormat.setOutputPath(job, new Path(args[1]));
      MultipleOutputs.addNamedOutput(job, "bins", TextOutputFormat.class, Text.class,
NullWritable.class);
      MultipleOutputs.setCountersEnabled(job, true);
     System.exit(job.waitForCompletion(true) ? 0 : 1);
   }}
```