In this exercise, we will learn about custom partiioners, MultipleOutputs and suppressing the result of reducer

Writing custom Partitioner

Custom partitioner should extend partitioner class and override getPartition()

```
import org.apache.hadoop.mapreduce.Partitioner;

public class customPartitioner extends Partitioner<patent,name> {
	@Override
	public int getPartition(patent p, name n, int numReducetasks) {
		// TODO Auto-generated method stub
		int res;
		res = Math.abs(p.country.hashCode())%numReducetasks;
		return res;
	}
}
```

Partitioner works on intermediate key – value pairs out from map phase and decide the number of reducers.

Above function generates hashcode on country field of output key(patent). Absolute value of this hashcode is taken and remainder after division with number of reducer tasks is returned. The returned value is between 0 and (number of Reduce tasks – 1).

This return value determines the reducer that the key-value pair will be sent to.

Reducer Class

```
public class runner {
	public static void main(String[] args) throws IOException, ClassNotFoundException,
InterruptedException {
		Configuration conf = new Configuration();
		conf.set("heading", "This involves custom writables and partitioners");

		Job job = new Job(conf);
		job.setJarByClass(runner.class);

		FileInputFormat.setInputPaths(job, args[0]);
		FileOutputFormat.setOutputPath(job, new Path(args[1]));

		job.setMapperClass(map_class.class);
		job.setReducerClass(reduce_class.class);

		job.setInputFormatClass(TextInputFormat.class);
		job.setOutputFormatClass(TextOutputFormat.class);

		job.setMapOutputKeyClass(patent.class);
		job.setMapOutputValueClass(name.class);

		job.setOutputKeyClass(Text.class);
		job.setOutputValueClass(Text.class);

		job.setPartitionerClass(customPartitioner.class);
		job.setNumReduceTasks(10);

		System.exit(job.waitForCompletion(true)?0:1);
	}}
```

Do set Partitioner class and number of reducers (default is 1).

Output

Contents of directory /user/training/MR/custom/out_partitioner

Goto : /user/training/MR/custom/ou go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
_SUCCESS	file	0 KB	1	64 MB	2014-03-08 07:38	rw-r--r--	training	supergroup
_logs	dir				2014-03-08 07:36	rwxr-xr-x	training	supergroup
part-r-00000	file	6.32 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00001	file	6.39 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00002	file	2.2 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00003	file	16.62 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00004	file	1.84 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00005	file	54.23 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00006	file	2.82 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00007	file	997.65 KB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00008	file	35.85 MB	1	64 MB	2014-03-08 07:37	rw-r--r--	training	supergroup
part-r-00009	file	142.08 MB	1	64 MB	2014-03-08 07:38	rw-r--r--	training	supergroup

```
----------------------------------------
(Bas-Rhin) 4151865
Alfred  Lehrer
----------------------------------------
----------------------------------------
00125 Roma 5181724
Vittorio  Spadoni
----------------------------------------
----------------------------------------
08006 Barcelona 5038536
Juan  De Mendoza Sans
----------------------------------------
----------------------------------------
103489 4320321
Vitaly V. Alexandrov
----------------------------------------
----------------------------------------
103498 4240086
Nina I. Komarovskikh
Vladimir I. Makhov
Petr E. Kandyba
----------------------------------------
----------------------------------------
103498 4272163
Alexandr N. Davydov
----------------------------------------
```

Above program does not solves the naming of output files. **"MultipleOutputs"** solves this problem

Reducer class for above program:

```
public class reduce_class extends Reducer<Text, name, NullWritable, Text> {
      public void reduce(Text key,Iterable<name> values,Context context) throws
IOException, InterruptedException{
            MultipleOutputs<NullWritable,Text> m = new
MultipleOutputs<NullWritable,Text>(context);
            long pat;
            String n;
            NullWritable out = NullWritable.get();
            TreeMap<Long,ArrayList<String>> map = new TreeMap<Long,ArrayList<String>>();
            for(name nn : values){
                  pat = nn.patent_No.get();
                  if(map.containsKey(pat))
                        map.get(pat).add(nn.getName().toString());
                  else{
                        map.put(pat,(new ArrayList<String>()));
                        map.get(pat).add(nn.getName().toString());}
            }
            for(Map.Entry entry : map.entrySet()){
                  n = entry.getKey().toString();
                  m.write(out, new Text("--------------------------"), key.toString());
                  m.write(out, new Text(n), key.toString());
                  ArrayList<String> names = (ArrayList)entry.getValue();
                  Iterator i = names.iterator();
                  while(i.hasNext()){
                        n = (String)i.next();
                        m.write(out, new Text(n), key.toString());
                  }
                  m.write(out, new Text("--------------------------"), key.toString());

            }
            m.close();
      }
}
```

MultipleOuts instance is created.

A tree map is invoked. For a particular patentno (key for tree map), all inventor names are its values.

Make sure to close the instance of MultipleOutputs:

Output:

```
--------------------------
3858241
Philip E. Durand
Lonnie H. Norris
--------------------------
--------------------------
3858242
Elwyn R. Gooding
--------------------------
--------------------------
3858244
Richard L. Mann
--------------------------
--------------------------
3858245
Michael A. Nate
Maurice A. Mann
--------------------------
--------------------------
3858247
Jack  Bauman
--------------------------
--------------------------
3858248
Sylvester L. Crowe
```

40019-r-00000	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
401 Tung Dist-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
40125 Bologna-r-00004	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
4019-r-00002	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
4074-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
41-r-00002	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
41002-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
4105-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
4113-r-00002	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
4128-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
417-r-00002	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
42-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
42000 Saint Entenne-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
42700 Firminy-r-00003	file	0.08 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
430-r-00002	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
430033-r-00004	file	0.07 KB	1	64 MB	2014-03-08 08:07	rw-r--r--	training	supergroup
4350 Quee-r-00002	file	0.07 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup
454106 Chelyabinsk-r-00002	file	0.08 KB	1	64 MB	2014-03-08 08:06	rw-r--r--	training	supergroup

Precaution:

Above method stores large amount of data in tree map during run time. Insufficient memory may cause java heap error. Increase memory allocated to JVM for hadoop by following command.

<name>mapred.child.java.opts</name>
  <value>-Xmx2048m</value>

Add above line to *mapred-site.xml* .

Lets see what happens if we set number of reducers to zero.

```
Job.setBumReducetasks(0);
```

This command sets number of reducers to zero.

Output:

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
_SUCCESS	file	0 KB	1	64 MB	2014-03-08 08:27	rw-r--r--	training	supergroup
_logs	dir				2014-03-08 08:26	rwxr-xr-x	training	supergroup
part-m-00000	file	14.96 MB	1	64 MB	2014-03-08 08:26	rw-r--r--	training	supergroup
part-m-00001	file	15.29 MB	1	64 MB	2014-03-08 08:26	rw-r--r--	training	supergroup
part-m-00002	file	15.39 MB	1	64 MB	2014-03-08 08:26	rw-r--r--	training	supergroup
part-m-00003	file	15.37 MB	1	64 MB	2014-03-08 08:26	rw-r--r--	training	supergroup
part-m-00004	file	2.59 MB	1	64 MB	2014-03-08 08:26	rw-r--r--	training	supergroup

5 outputs from mappers

```
US      name@16877f8
US      name@134eb84
FR      name@1522de2
FR      name@425743
KR      name@16ef705
US      name@1b7c76
US      name@883357
US      name@1e2afb2
US      name@17cfd38
US      name@d480ea
US      name@15b8520
JP      name@18105e8
JP      name@1aacd5f
US      name@913dc1
US      name@16d8196
US      name@56b93a
US      name@19abd2b
US      name@1f98d58
US      name@aea710
US      name@5a2eaa
US      name@adae91
US      name@9fe84e
US      name@4c71d2
US      name@c4bc34
US      name@2006a0
```

Notice for a particular map output, records are not sorted.