

# LARGE SCALE DATA PROCESSING (CSE 3025)

## WIN SEMESTER 2017-18

NAME: B.SHUBANKAR

REGNO: 15BCE1123

PROFESSOR: DR. Maheswari N

---

### LAB-9

1. Create the employee details(eid,ename,designation) file and the department details(eid,department) file in HDFS. Let the department details can be stored using distributed cache. Write the employee id, name, designation and department in the HDFS. Use map side join.

#### PROGRAM:

```
import org.apache.hadoop.filecache.DistributedCache;
import org.apache.hadoop.fs.Path ;
import org.apache.hadoop.conf.* ;
import org.apache.hadoop.io.* ;
import org.apache.hadoop.mapreduce.* ;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.lib.output.*;
import java.io.BufferedReader;
import java.io.FileReader;
```

---

---

```
import java.io.IOException;
import java.util.ArrayList;

    public class distcache {

        public static class Map extends
Mapper<LongWritable,Text,Text,Text> {

    Path[] cfile=new Path[0];
    ArrayList<Text> empl=new ArrayList<Text>();
        private BufferedReader reader;
public void setup(Context context)
{
    Configuration conf=context.getConfiguration();
    try
    {
        cfile = DistributedCache.getLocalCacheFiles(conf);
        reader = new BufferedReader(new FileReader(cfile[0].toString()));
        String line;
        while ((line=reader.readLine())!=null)
        {
            Text tt=new Text(line);
            empl.add(tt);
        }
    }
}
```

---

---

```
}
```

```
catch(IOException e)
```

```
{
```

```
    e.printStackTrace();
```

```
    }
```

```
}
```

```
public void map(LongWritable key, Text value, Context context) throws  
IOException, InterruptedException {
```

```
    String line2 = value.toString();
```

```
    String[] elements=line2.split(",");
```

```
for(Text e:empl)
```

```
{
```

```
    String[] line1 = e.toString().split(",");
```

```
    if(elements[0].equals(line1[0]))
```

```
{
```

```
        context.write(new Text(elements[0]),new  
Text(elements[1]+","+elements[2]+","+line1[1]));
```

```
    }
```

```
}
```

---

```
}  
  
}  
  
public static void main(String[] args) throws Exception {  
    Configuration conf = new Configuration();  
    Job job = new Job(conf, "distcache");  
    job.setJarByClass(distcache.class);  
    job.setOutputKeyClass(Text.class);  
    job.setOutputValueClass(Text.class);  
    job.setMapperClass(Map.class);  
    job.setInputFormatClass(TextInputFormat.class);  
    job.setOutputFormatClass(TextOutputFormat.class);  
    DistributedCache.addCacheFile(new  
Path(args[0]).toUri(),job.getConfiguration());  
    FileInputFormat.addInputPath(job, new Path(args[1]));  
    FileOutputFormat.setOutputPath(job, new Path(args[2]));  
    job.waitForCompletion(true);  
}  
}
```

**2.**Create the customer details(cid,name) and the transaction details(cid, transaction amount) file in hdfs. Write the cname, number of transactions and transaction amount in hdfs. Use reduce side join.

### **PROGRAM:**

---

```
import java.io.IOException;
//import java.util.*;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.lib.input.MultipleInputs;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.*;
import org.apache.hadoop.mapreduce.lib.output.*;
public class reduce_side_join {
    public static class empmapper extends
Mapper<LongWritable,Text,Text,Text>{
        public void map(LongWritable key, Text value, Context context) throws
IOException, InterruptedException{
            String[] line = value.toString().split(",");
            context.write(new Text(line[0]), new Text("cust"+" "+line[1]));
        }
    }
    public static class depmapper extends Mapper<LongWritable,Text,Text,Text>
    {
        public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException{
            String[] line = value.toString().split(",");
            context.write(new Text(line[0]), new Text("trans"+" "+line[1]));
        }
    }
}
```

---

```
    }

    public static class jreducer extends Reducer<Text,Text,Text,Text>
    {
        String st1;

        public void reduce(Text key, Iterable<Text> values, Context context )
        throws IOException, InterruptedException
        {
            int c=0,amt=0;
            for(Text val:values)
            {
                String[] line = val.toString().split(",");
                if (line[0].equals("trans"))
                {
                    c=c+1;
                    amt+= Integer.parseInt(line[1]);
                }
                else if (line[0].equals("cust"))
                {
                    st1 = line[1];
                }
            }

            context.write(new Text(st1), new Text(c+","+amt));
        }
    }
}
```

---

---

```
    }  
  
    public static void main(String[] args) throws Exception { //Driver  
        Configuration conf = new Configuration();  
        Job job = new Job(conf, "Reduce-side join");  
        job.setJarByClass(reduce_side_join.class);  
        job.setOutputKeyClass(Text.class);  
        job.setOutputValueClass(Text.class);  
        job.setReducerClass(jreducer.class);  
        job.setInputFormatClass(TextInputFormat.class);  
        job.setOutputFormatClass(TextOutputFormat.class);  
        MultipleInputs.addInputPath(job, new Path(args[0]), TextInputFormat.class,  
            empmapper.class);  
        MultipleInputs.addInputPath(job, new Path(args[1]), TextInputFormat.class,  
            depmapper.class);  
        FileOutputFormat.setOutputPath(job, new Path(args[2]));  
        job.waitForCompletion(true);  
    }  
}
```

## **OUTPUT:**

---

**File: [/lab9/customer\\_details](#)**

---

Goto :

---

[Go back to dir listing](#)  
[Advanced view/download options](#)

---

```
a,snehaa  
b,bernice  
c,sharlin  
d,vandhana  
e,padmaja  
f,keerthana  
g,sona  
h,varsha
```

---

**File: [/lab9/transaction\\_details](#)**

---

Goto :

---

[Go back to dir listing](#)  
[Advanced view/download options](#)

---

```
a,900  
b,800  
c,1200  
d,1500  
e,50  
f,300  
g,15000  
h,350  
e,240  
f,380  
c,245  
d,120  
c,345
```



---

**File: [/lab9/reduce\\_side\\_join/part-r-00000](#)**

---

Goto :

---

[Go back to dir listing](#)

[Advanced view/download options](#)

---

```
snehaa 1,900
bernice 1,800
sharlin 3,1790
vandhana 2,1620
padmaja 2,290
keerthana 2,680
sona 1,15000
varsha 1,350
```