

Task1:

```

Total committed heap usage (bytes)=361660416
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=115
File Output Format Counters
Bytes Written=95
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output$ ls -ltr
total 48
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_0
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_1
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_2
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_3
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_4
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_5
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_6
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_7
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_8
drwxr-xr-x 2 ping ping 4096 Jun 23 23:24 iteration_9
drwxr-xr-x 2 ping ping 4096 Jun 23 23:25 iteration_10
drwxr-xr-x 2 ping ping 4096 Jun 23 23:25 result
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output$ cd result/
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/result$ ls -ltr
total 4
-rw-r--r-- 1 ping ping 83 Jun 23 23:25 part-r-00000
-rw-r--r-- 1 ping ping 0 Jun 23 23:25 _SUCCESS
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/result$ more part-r-00000
A      0.333251953125
D      0.22224934895833331
C      0.22224934895833331
B      0.22224934895833331
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/result$

```

Code changes:

*****HadoopPageRank class*****

*****place holder part*****

```

resultJob.setMapperClass(HadoopPageRankResultMapper.class);
resultJob.setReducerClass(HadoopPageRankResultReducer.class);
resultJob.setOutputKeyClass(DoubleWritable.class);
resultJob.setOutputValueClass(Text.class);
resultJob.setInputFormatClass(TextInputFormat.class);

```

*****HadoopPageRankResultMapper class*****

```

import java.io.IOException;
import org.apache.hadoop.io.DoubleWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Mapper.Context;

```

```

public class HadoopPageRankResultMapper extends Mapper<LongWritable, Text, DoubleWritable,
Text>{
    public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException {
        if ( value == null || value.getLength() == 0 ) {
            return;
        }
    }
}

```

```

        int tabIdx1 = value.find("\t");
        int tabIdx2 = value.find("\t", tabIdx1 + 1);
        String link = Text.decode(value.getBytes(), 0, tabIdx1);
        String pageRank = Text.decode(value.getBytes(), tabIdx1 + 1, tabIdx2 - (tabIdx1 + 1));
        double PR = -1*Double.parseDouble(pageRank.toString());
        context.write(new DoubleWritable(PR), new Text(link));
    }
}

```

*****HadoopPageRankResultReducer class*****

```

import java.io.IOException;
import org.apache.hadoop.io.DoubleWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class HadoopPageRankResultReducer extends Reducer<DoubleWritable, Text, Text, DoubleWritable> {
    public void reduce(DoubleWritable key, Iterable<Text> values, Context context) throws
IOException, InterruptedException {
        double temp = 0.0;
        temp = key.get() * -1;
        for (Text value : values) {
            context.write(value, new DoubleWritable(temp));
        }
    }
}

```

Task2:

```
Activities Terminal Jun 24 00:03 ping@ping-VirtualBox: ~/Desktop/Big Data/HW3/pageRank_output/result

Total committed heap usage (bytes)=361512960
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=119
File Output Format Counters
Bytes Written=99
ping@ping-VirtualBox:~/Desktop/Big Data/HW3$ cd pageRank_output/
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output$ ls -ltr
total 48
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_0
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_1
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_2
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_3
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_4
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_5
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_6
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_7
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_8
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_9
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 iteration_10
drwxr-xr-x 2 ping ping 4096 Jun 24 00:02 result
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output$ cd result/
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/result$ ls -ltr
total 4
-rw-r--r-- 1 ping ping 87 Jun 24 00:02 part-r-00000
-rw-r--r-- 1 ping ping 0 Jun 24 00:02 _SUCCESS
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/result$ more part-r-00000
C 0.6401342646255148
D 0.12903270916213994
B 0.12903270916213994
A 0.10180031705020576
ping@ping-VirtualBox:~/Desktop/Big Data/HW3/pageRank_output/result$
```

Code changes:

*****HadoopPageRankMainJobReducer class*****

```
// I changed the calculation function of receivedContribution
```

```
// with the updated flow model calculation
```

```
import java.io.IOException;
```

```
import org.apache.hadoop.io.Text;
```

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

```
public class HadoopPageRankMainJobReducer extends Reducer<Text, Text, Text, Text> {
```

```
    @Override
```

```
    public void reduce(Text key, Iterable<Text> values, Context context)
```

```
        throws IOException, InterruptedException {
```

```
        if ( values == null ) {
```

```
            return;
```

```
        }
```

```
        String links = "";
```

```
        double receivedContribution = 0.0;
```

```
        for (Text value : values) {
```

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

```
            if (content.startsWith("|")) {
```

```
                links += content.substring("|".length());
```

```
            } else {
```

```
        receivedContribution += Double.parseDouble(content);
    }

}

//the Google solution for spider traps:
receivedContribution = 0.80*receivedContribution+0.20*0.25;//here is the change
double newPageRank = receivedContribution;
context.write(key, new Text(newPageRank + "\t" + links));

}

}
```

Task3:

3 After iteration 4, $b = 1/2$