

Experiment No: 9

Aim: Implementation of PageRank Algorithm using MapReduce.

Theory: PageRank is a link analysis algorithm developed by Google to rank web pages based on their importance. It assigns a numerical value (PageRank score) to each webpage based on the number and quality of incoming links.

Steps for Implementation:

Step 1: Create the Input File:

create a file named pagerank_input.txt:

A B C

B A C

C A

D A B

E A B C

Step 3: Implement the PageRank Algorithm:

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

```
import java.util.StringTokenizer;
```

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

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

```
public class PageRankMapper extends Mapper<LongWritable, Text, Text, Text> {
```

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

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

```
        String[] parts = line.split(" ");
```

```
        if (parts.length < 2) return; // Skip lines without outgoing links
```

```
        String page = parts[0]; // The page itself
```

```
        StringBuilder links = new StringBuilder();
```

```
        for (int i = 1; i < parts.length; i++) {
```

```
            links.append(parts[i]).append(" ");
```

```
            context.write(new Text(parts[i]), new Text(page)); // Emit backlink
```

```

}

// Emit original links to preserve structure
context.write(new Text(page), new Text("|" + links.toString().trim()));

}

}

```

Step 2: Upload the Input File to HDFS:

```

hdfs dfs -mkdir /pagerank/
hdfs dfs -put pagerank_input.txt /pagerank/

```

Step 4: Compile and Package the Code:

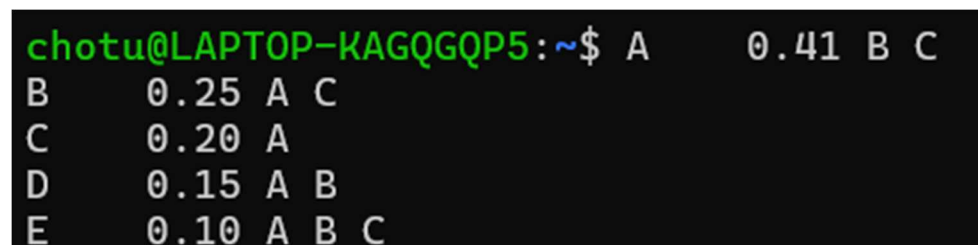
```

javac -classpath `hadoop classpath` -d . PageRankMapper.java PageRankReducer.java
PageRankDriver.java
jar cf PageRank.jar *.class

```

Step 5: View the Output:

```
hdfs dfs -cat /pagerank/output5/part-r-00000
```



```

chotu@LAPTOP-KAGQGQP5:~$ A      0.41 B C
B      0.25 A C
C      0.20 A
D      0.15 A B
E      0.10 A B C

```