BDA Lab

Assignment-3

Q Write a map reduce program to count the frequency of given word in a given file.

Step 1 Start Hadoop Services

```
vboxuser@Ubuntu-hadoop:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Ubuntu-hadoop]
vboxuser@Ubuntu-hadoop:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
vboxuser@Ubuntu-hadoop:~$
```

Step 2 create file

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab$ mkdir MaxTemp
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab$ cd MaxTemp/
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ nano MaxTemperature.java
```

Step 3 Write word counting program

```
import java.io.IOException;

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Job;

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

```
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class MaxTemperature {

   // Mapper Class
   public static class TempMapper
         extends Mapper<Object, Text, Text, IntWritable> {

      private Text year = new Text();
      private IntWritable temperature = new IntWritable();

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

         String line = value.toString();
         String[] parts = line.split(",");

         year.set(parts[0]);
         temperature.set(Integer.parseInt(parts[1]));

         context.write(year, temperature);
      }
   }
   // Reducer Class
   public static class TempReducer
         extends Reducer<Text, IntWritable, Text, IntWritable> {
```

```
    public void reduce(Text key, Iterable<IntWritable> values,
                Context context)
        throws IOException, InterruptedException {

      int maxTemp = Integer.MIN_VALUE;

      for (IntWritable val : values) {
        if (val.get() > maxTemp) {
          maxTemp = val.get();
        }
      }
      context.write(key, new IntWritable(maxTemp));
    }
}
// Driver Code
public static void main(String[] args) throws Exception {

    Configuration conf = new Configuration();
    Job job = Job.getInstance(conf, "Max Temperature");

    job.setJarByClass(MaxTemperature.class);
    job.setMapperClass(TempMapper.class);
    job.setReducerClass(TempReducer.class);

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

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

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

Step 4 Compile Program and create Jar

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ export HADOOP_CLASSPATH=$(hado
op classpath)
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ javac -classpath ${HADOOP_CLAS
SPATH} -d MaxTemperature MaxTemperature.java
```

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ jar -cvf maxtemp.jar -C MaxTem
perature/ .
added manifest
adding: MaxTemperature$TempMapper.class(in = 1764) (out= 726)(deflated 58%)
adding: MaxTemperature$TempReducer.class(in = 1709) (out= 723)(deflated 57%)
adding: MaxTemperature.class(in = 1483) (out= 801)(deflated 45%)
```

Step 5 Create Input file and upload to hdfs

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ nano input.txt
```

1990,23

1990,45

1991,34

1991,67

1990,12

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ hdfs dfs -mkdir /input
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ hdfs dfs -put input.txt /input
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ hdfs dfs -cat /input/input.txt
1990,23
1990,45
1991,34
1991,67
1990,12
```

Step 6 Run MapReduce Job

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ hadoop jar maxtemp.jar MaxTemp
erature /input /output
```

```
2026-02-15 19:24:55,868 INFO mapreduce.Job:  map 0% reduce 0%
2026-02-15 19:25:07,691 INFO mapreduce.Job:  map 100% reduce 0%
2026-02-15 19:25:20,313 INFO mapreduce.Job:  map 100% reduce 100%
2026-02-15 19:25:21,416 INFO mapreduce.Job: Job job_1771180586441_0005 completed
 successfully
```

Step 7 View Output

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/MaxTemp$ hdfs dfs -cat /output/part-r-0
0000
1990    45
1991    67
```