BDA Lab

Assignment-4

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 StudentAvg1
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab$ cd StudentAvg1
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ nano StudentAverage.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.*;

import org.apache.hadoop.mapreduce.*;

import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;

import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

```java
public class StudentAverage {

  // Mapper Class
  public static class AvgMapper
    extends Mapper<LongWritable, Text, Text, FloatWritable> {

    private Text studentId = new Text();
    private FloatWritable grade = new FloatWritable();
    public void map(LongWritable key, Text value, Context context)
        throws IOException, InterruptedException {
      String line = value.toString();
      String[] parts = line.split(",");
      if (parts.length == 3) {
        studentId.set(parts[0]);
        grade.set(Float.parseFloat(parts[2]));
        context.write(studentId, grade);
      }
    }
  }
  // Reducer Class
  public static class AvgReducer
    extends Reducer<Text, FloatWritable, Text, FloatWritable> {
    private FloatWritable result = new FloatWritable();
    public void reduce(Text key, Iterable<FloatWritable> values,
            Context context)
        throws IOException, InterruptedException {

      float sum = 0;
```

```java
      int count = 0;

      for (FloatWritable val : values) {

        sum += val.get();

        count++;

      }

      float average = sum / count;

      result.set(average);

      context.write(key, result);

    }

  }

  // Driver Class

  public static void main(String[] args) throws Exception {

    Configuration conf = new Configuration();

    Job job = Job.getInstance(conf, "Average Grade");

    job.setJarByClass(StudentAverage.class);

    job.setMapperClass(AvgMapper.class);

    job.setReducerClass(AvgReducer.class);


    job.setOutputKeyClass(Text.class);

    job.setOutputValueClass(FloatWritable.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/StudentAvg$ javac -classpath `hadoop cl
asspath` -d StudentAvg_classes StudentAverage.java
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ jar -cvf studentavg.jar -C
StudentAvg_classes/ .
added manifest
adding: StudentAverage.class(in = 1479) (out= 806)(deflated 45%)
adding: StudentAverage$AvgReducer.class(in = 1805) (out= 776)(deflated 57%)
adding: StudentAverage$AvgMapper.class(in = 1948) (out= 798)(deflated 59%)
```

Step 5 Create Input file and upload to hdfs

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ hdfs dfs -mkdir /studentinp
ut
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ hdfs dfs -put input.txt /st
udentinput
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ hdfs dfs -ls /studentinput
Found 1 items
-rw-r--r--   3 vboxuser supergroup         66 2026-02-16 10:36 /studentinput/inp
ut.txt
```

101,Math,85

101,Science,90

102,Math,78

102,Science,82

103,Math,88

Step 6 Run MapReduce Job

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ hadoop jar studentavg.jar S
tudentAverage /studentinput /studentoutput
```

```
2026-02-16 10:39:06,173 INFO mapreduce.Job:  map 0% reduce 0%
2026-02-16 10:40:02,561 INFO mapreduce.Job:  map 100% reduce 0%
2026-02-16 10:40:39,011 INFO mapreduce.Job:  map 100% reduce 100%
2026-02-16 10:41:09,548 INFO mapreduce.Job: Job job_1771237549532_0001 completed
 successfully
```

Step 7 View Output

```
vboxuser@Ubuntu-hadoop:~/Desktop/BDA Lab/StudentAvg$ hdfs dfs -cat /studentoutpu
t/part-r-00000
101    87.5
102    80.0
103    88.0
```