Bansilal Ramnath Agarwal Charitable Trust's

Vishwakarma Institute of Information Technology

Department of

Artificial Intelligence and Data Science

Name: Siddhesh Dilip Khairnar		
Class: **TY**	Division: **B**	Roll No: **372028**
Semester: **6th**	Academic Year: **2023-24**	
Subject Name & Code: **Data Science & ADUA32202**		
Title of Assignment: Hadoop with MapReduce Examples.		

ASSIGNMENT NO: - 6

Input

```
Student Name, Year of Admission, Mode of Admission, City
John, 2021, Entrance, New York
Alice, 2020, Direct, Boston
Bob, 2021, Entrance, Boston
Sam, 2020, Direct, New York
Lucy, 2021, Entrance, Boston
David, 2020, Direct, New York
```

Steps:

1. Create a new Java project in Eclipse and add the Hadoop libraries to the project.

2. Create a new Java class named "CityMapper" and implement the Mapper interface to read the input data and emit city name as key and value as 1.

```
public class CityMapper extends Mapper<LongWritable, Text, Text, Int
    private final static IntWritable one = new IntWritable(1);
    private Text city = new Text();

    public void map(LongWritable key, Text value, Context context)
            throws IOException, InterruptedException {
        String[] line = value.toString().split(",");
        city.set(line[3].trim());
        context.write(city, one);
    }
}
```

3. Create a new Java class named "CityReducer" and implement the Reducer interface to receive city names and count the number of students for each city.

```
public class CityReducer extends Reducer<Text, IntWritable, Text, IntWrita
    private IntWritable count = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values, Context con
            throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get();
        }
        count.set(sum);
        context.write(key, count);
    }
}
```

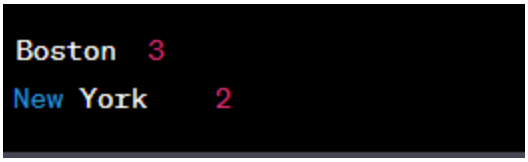
4. Create a new Java class named "CityCount" and configure the MapReduce job by setting the input and output paths, mapper and reducer classes, and the output key-value classes.

```
public class CityCount {
    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        Job job = Job.getInstance(conf, "City Count");
        job.setJarByClass(CityCount.class);
        job.setMapperClass(CityMapper.class);
        job.setReducerClass(CityReducer.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);
    }
}
```

5. Run the MapReduce job using the following command:

   Hadoop jar <jar_file_path> CityCount <input_path> <output_path>

6. Check the output file generated in the output path, which will contain the number of students admitted from different cities.



Conclusion: In this assignment, we used Hadoop MapReduce to find out the number of students admitted from different cities based on the given input data containing student information.