**DistinctWord Program:**

**MapperCode:**

import java.io.IOException;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Mapper;

public class DistinctRecordMapper extends Mapper<Object, Text, Text, Text> {

private Text keyValue = new Text();

private StringBuilder keyFormat=new StringBuilder();

@Override

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

keyFormat.setLength(0);

String data = value.toString();

String[] field = data.split(",", -1);

if (null != field && field.length > 10) {

keyFormat.append(field[0]+field[1]+field[3]);

keyValue.set(keyFormat.toString());

context.write(keyValue, value);

}

}

}

**ReducerCode:**

import java.io.IOException;

import org.apache.hadoop.io.NullWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Reducer;

public class DistinctRecordReducer extends Reducer<Text, Text, NullWritable, Text> {

@Override

public void reduce(Text text, Iterable<Text> values, Context context) throws IOException,

InterruptedException {

for (Text value : values) {

context.write(NullWritable.get(),value);

break;

}

}

}

**DriverCode:**

import java.io.File;

import org.apache.commons.io.FileUtils;

import org.apache.hadoop.conf.Configuration;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Job;

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

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

import com.hadoop.design.summarization.blog.ConfigurationFactory;

public class DriverDistinctRecord {

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

/\*

\* I have used my local path in windows change the path as per your

\* local machine

\*/

args = new String[] { "Replace this string with Input Path location",

"Replace this string with output Path location" };

/\* delete the output directory before running the job \*/

FileUtils.deleteDirectory(new File(args[1]));

/\* set the hadoop system parameter \*/

System.setProperty("hadoop.home.dir", "Replace this string with hadoop home directory location");

if (args.length != 2) {

System.err.println("Please specify the input and output path");

System.exit(-1);

}

Configuration conf = ConfigurationFactory.getInstance();

Job job = Job.getInstance(conf);

job.setJarByClass(DriverDistinctRecord.class);

job.setJobName("Distinct\_Record");

FileInputFormat.addInputPath(job, new Path(args[0]));

FileOutputFormat.setOutputPath(job, new Path(args[1]));

job.setMapperClass(DistinctRecordMapper.class);

job.setReducerClass(DistinctRecordReducer.class);

job.setOutputKeyClass(Text.class);

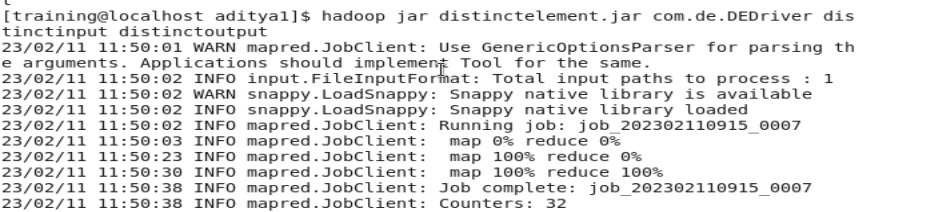
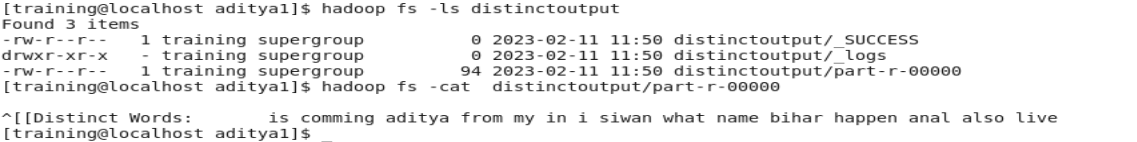
job.setOutputValueClass(Text.class);

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

}

}

**Output:**

****