myDriver.java

**import java.io.IOException;**

**import java.util.StringTokenizer;**

**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 myDriver {**

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

**Configuration conf = new Configuration();**

**Job job = Job.getInstance(conf, "word count");**

**job.setJarByClass(myDriver.class);**

**job.setMapperClass(myMapper.class);**

**job.setCombinerClass(myReducer.class);**

**job.setReducerClass(myReducer.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);**

**}**

**}**

---------------------------------------------------------

myMapper.java

**import java.io.IOException;**

**import java.util.StringTokenizer;**

**import org.apache.hadoop.conf.Configuration;**

**import org.apache.hadoop.fs.Path;**

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

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

**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 myMapper extends Mapper<LongWritable, Text, Text, IntWritable>{**

**private final static IntWritable one = new IntWritable(1);**

**private Text word = new Text();**

**private static final int MISSING = 9999;**

**public void map(LongWritable key, Text value, Context context**

**) throws IOException, InterruptedException {**

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

**String year = line.substring(15,19);**

**int temp = Integer.parseInt(line.substring(87,92));**

**if (line.charAt(87)=='+')**

**{**

**temp=Integer.parseInt(line.substring(88,92));**

**}**

**if (line.charAt(87)=='-')**

**{**

**temp=Integer.parseInt(line.substring(87,92));**

**}**

**String quality = line.substring(92, 93);**

**if (temp!= MISSING&& quality.matches("[01459]"))**

**{**

**context.write(new Text(year), new IntWritable(temp));**

**}**

**}**

**}**

---------------------------------------------------------

myReducer.java

import java.io.IOException;

import java.util.StringTokenizer;

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 myReducer extends Reducer<Text,IntWritable,Text,IntWritable> {

private IntWritable result = new IntWritable();

public void reduce(Text key, Iterable<IntWritable> values,

Context context

) throws IOException, InterruptedException {

int maxValue = Integer.MIN\_VALUE;

for (IntWritable value : values)

{

maxValue = Math.max(maxValue, value.get());

}

context.write(key, new IntWritable(maxValue));

}

}

OUTPUT

hadoop\_varad@c05l0516:~$ hadoop jar /home/hadoop\_varad/Desktop/wea.jar /varad\_weather/weather/ /varad\_weather/output

19/02/07 15:11:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

19/02/07 15:11:30 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

19/02/07 15:11:30 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

19/02/07 15:11:31 INFO input.FileInputFormat: Total input files to process : 5

19/02/07 15:11:31 INFO mapreduce.JobSubmitter: number of splits:5

19/02/07 15:11:31 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled

19/02/07 15:11:31 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1549527274684\_0008

19/02/07 15:11:32 INFO impl.YarnClientImpl: Submitted application application\_1549527274684\_0008

19/02/07 15:11:32 INFO mapreduce.Job: The url to track the job: http://c05l0516:8088/proxy/application\_1549527274684\_0008/

19/02/07 15:11:32 INFO mapreduce.Job: Running job: job\_1549527274684\_0008

19/02/07 15:11:38 INFO mapreduce.Job: Job job\_1549527274684\_0008 running in uber mode : false

19/02/07 15:11:38 INFO mapreduce.Job: map 0% reduce 0%

19/02/07 15:12:19 INFO mapreduce.Job: map 100% reduce 0%

19/02/07 15:12:41 INFO mapreduce.Job: map 100% reduce 100%

19/02/07 15:12:42 INFO mapreduce.Job: Job job\_1549527274684\_0008 completed successfully

19/02/07 15:12:42 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=61

FILE: Number of bytes written=1191763

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=4439563

HDFS: Number of bytes written=45

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Launched map tasks=5

Launched reduce tasks=1

Data-local map tasks=5

Total time spent by all maps in occupied slots (ms)=190290

Total time spent by all reduces in occupied slots (ms)=18675

Total time spent by all map tasks (ms)=190290

Total time spent by all reduce tasks (ms)=18675

Total vcore-milliseconds taken by all map tasks=190290

Total vcore-milliseconds taken by all reduce tasks=18675

Total megabyte-milliseconds taken by all map tasks=194856960

Total megabyte-milliseconds taken by all reduce tasks=19123200

Map-Reduce Framework

Map input records=32827

Map output records=32783

Map output bytes=295047

Map output materialized bytes=85

Input split bytes=565

Combine input records=32783

Combine output records=5

Reduce input groups=5

Reduce shuffle bytes=85

Reduce input records=5

Reduce output records=5

Spilled Records=10

Shuffled Maps =5

Failed Shuffles=0

Merged Map outputs=5

GC time elapsed (ms)=649

CPU time spent (ms)=6080

Physical memory (bytes) snapshot=1481351168

Virtual memory (bytes) snapshot=5064704000

Total committed heap usage (bytes)=1065353216

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=4438998

File Output Format Counters

Bytes Written=45

hadoop\_varad@c05l0516:~$ hadoop fs -ls /varad\_weather/output

19/02/07 15:13:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 2 items

-rw-r--r-- 1 hadoop\_varad supergroup 0 2019-02-07 15:12 /varad\_weather/output/\_SUCCESS

-rw-r--r-- 1 hadoop\_varad supergroup 45 2019-02-07 15:12 /varad\_weather/output/part-r-00000

hadoop\_varad@c05l0516:~$ hadoop fs -cat /varad\_weather/output/part-r-00000

19/02/07 15:14:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

1901 317

1902 244

1903 289

1904 256

1905 283