Exp No: 11 **Date:**

HADOOP
IMPLEMENT THE MAX TEMPERATURE MAPREDUCE PROGRAM TO IDENTIFY THE YEAR WISE MAXIMUM TEMPERATURE FROM SENSOR DATA

AIM

To implement the Max temperature MapReduce program to identify the year-wise maximum temperature from the sensor data.

Description

Sensors senses weather data in big text format containing station ID, year, date, time, temperature, quality etc. from each sensor and store it in a single line. Suppose thousands of data sensors are there, then we have thousands of records with no particular order. We require only a year and maximum temperature of particular quality in that year.

For example:

Input string from sensor:
0029029070999991902010720004+64333+023450
FM-12+
000599999V0202501N027819999999N0000001N9-00331+
99999098351ADDGF102991999999999999999999
Here: 1902 is year
0033 is temperature
1 is measurement quality (Range between 0 or 1 or 4 or 5 or 9)

Here each mapper takes the input key as "byte offset of line" and value as "one weather sensor read i.e one line". and parse each line and produce an intermediate key "year" and intermediate value as "temperature of certain measurement qualities" for that year.

The combiner will form set values of temperature. Year and set of values of temperatures is given as input <key, value> to reducer and Reducer will produce year and maximum temperature for that year from the set of temperature values.

PROGRAM

*/

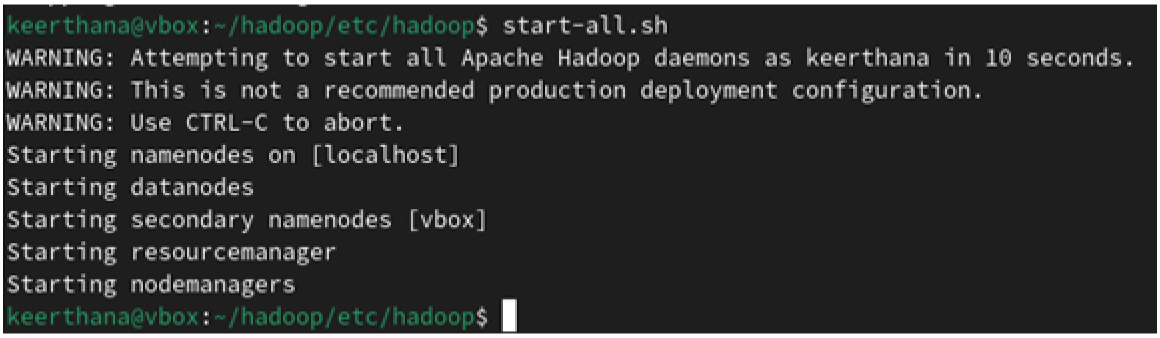
```
import  org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
importorg.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;
importorg.apache.hadoop.mapreduce.Reducer;

//Mapper class


class MaxTemperatureMapper
extends Mapper<LongWritable, Text, Text, IntWritable> { private static final int MISSING

= 9999;

@Override
public void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException {

String line = value.toString(); String year = line.substring(15, 19); int airTemperature;
if (line.charAt(87) == '+') { // parseInt doesn't like leading plus signs airTemperature =
Integer.parseInt(line.substring(88, 92));
} else {
airTemperature = Integer.parseInt(line.substring(87, 92));
}
String quality = line.substring(92, 93);
if (airTemperature != MISSING && quality.matches("[01459]")) { context.write(new
Text(year), new IntWritable(airTemperature));
}
}
}

//Reducer class
class MaxTemperatureReducer
extends Reducer<Text, IntWritable, Text, IntWritable> {

@Override
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));
}
}
//Driver Class

public class MaxTemperature {

public static void main(String[] args) throws Exception { if (args.length != 2) {
System.err.println("Usage: MaxTemperature <input path=""> <output path>"); System.exit(-
1);
}

Job job = Job.getInstance(new Configuration()); job.setJarByClass(MaxTemperature.class);
job.setJobName("Max temperature");

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

job.setMapperClass(MaxTemperatureMapper.class);
job.setReducerClass(MaxTemperatureReducer.class);

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

job.submit();
}
}
```

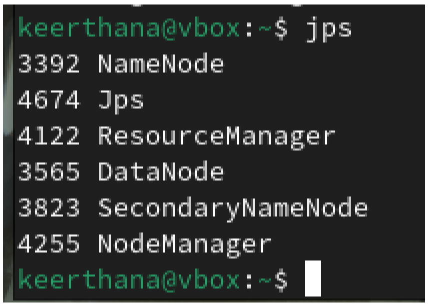
OUTPUT:

Input for String :

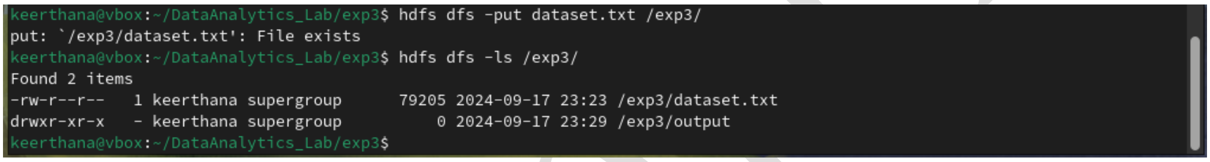
```
0029029070999991902010720004+64333+023450FM-12+
000599999V0202501N027819999999N0000001N9-00331+
99999098351ADDGF102991999999999999999’
```



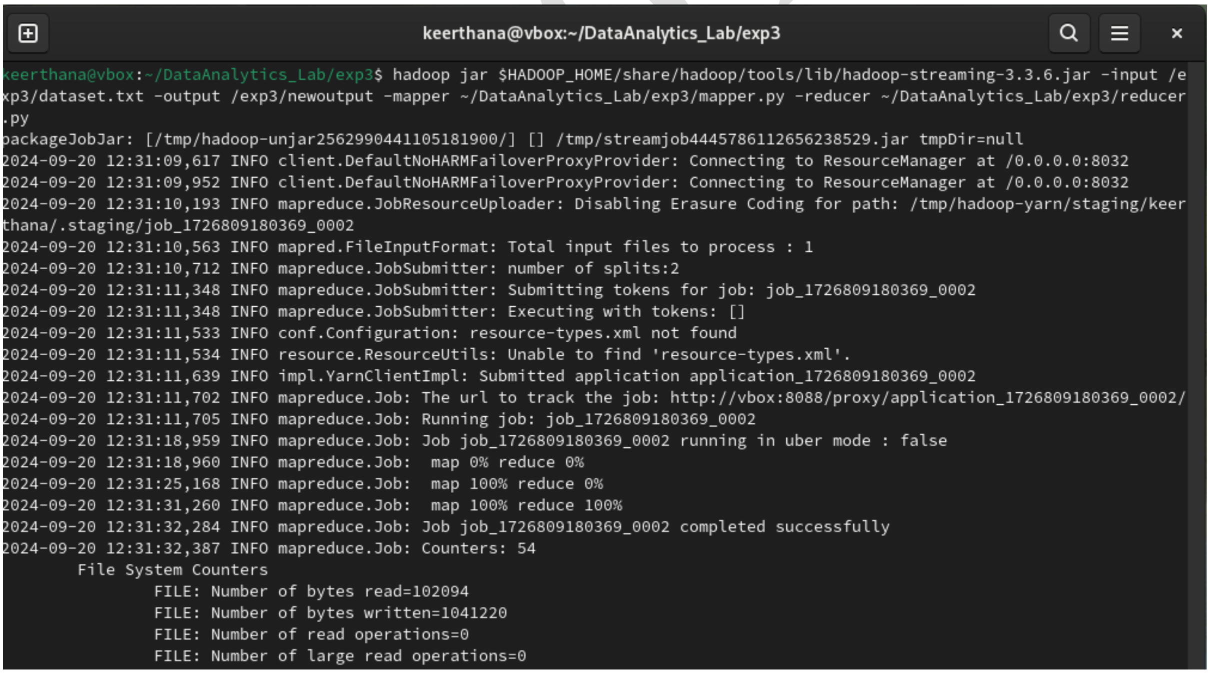
```
keerthana@vbox:~/hadoop/etc/hadoop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as keerthana in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [vbox]
Starting resourcemanager
Starting nodemanagers
keerthana@vbox:~/hadoop/etc/hadoop$
```



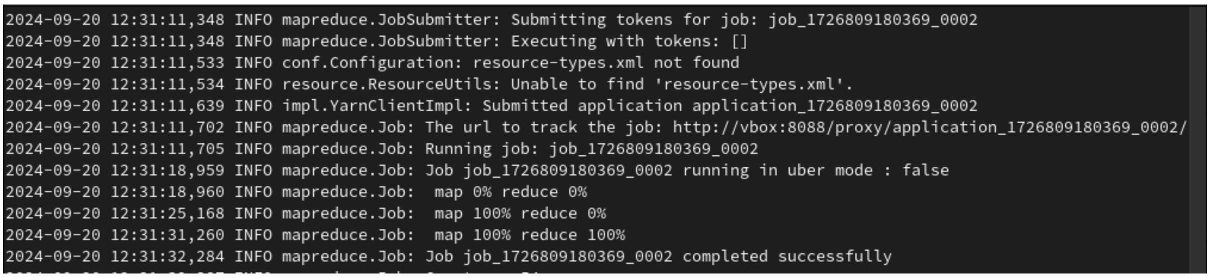
```
keerthana@vbox:~$ jps
3392 NameNode
4674 Jps
4122 ResourceManager
3565 DataNode
3823 SecondaryNameNode
4255 NodeManager
keerthana@vbox:~$
```



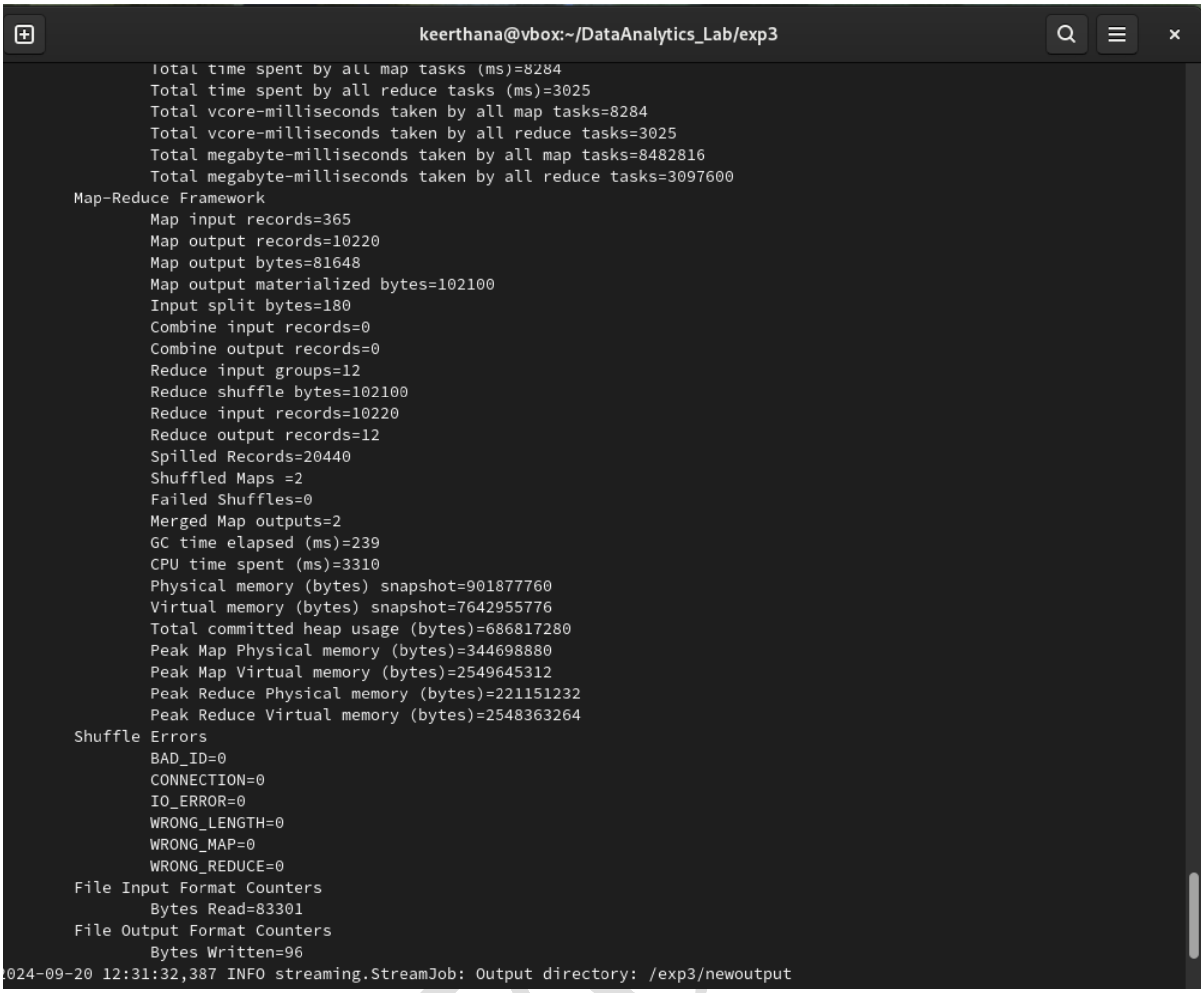
```
keerthana@vbox:~/DataAnalytics_Lab/exp3$ hdfs dfs -put dataset.txt /exp3/
put: `/exp3/dataset.txt': File exists
keerthana@vbox:~/DataAnalytics_Lab/exp3$ hdfs dfs -ls /exp3/
Found 2 items
-rw-r--r--   1 keerthana supergroup      79205 2024-09-17 23:23 /exp3/dataset.txt
drwxr-xr-x   - keerthana supergroup          0 2024-09-17 23:29 /exp3/output
keerthana@vbox:~/DataAnalytics_Lab/exp3$
```



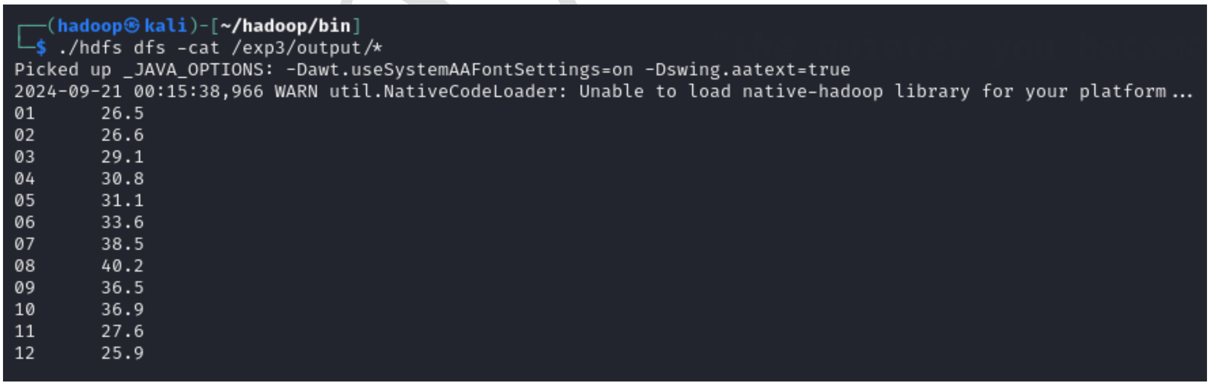
```
keerthana@vbox:~/DataAnalytics_Lab/exp3$ hadoop jar $HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar -input /e
xp3/dataset.txt -output /exp3/newoutput -mapper ~/DataAnalytics_Lab/exp3/mapper.py -reducer ~/DataAnalytics_Lab/exp3/reducer
.py
packageJobJar: [/tmp/hadoop-unjar2562990441105181900/] [] /tmp/streamjob4445786112656238529.jar tmpDir=null
2024-09-20 12:31:09,617 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-20 12:31:09,952 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-20 12:31:10,193 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/keer
thana/.staging/job_1726809180369_0002
2024-09-20 12:31:10,563 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-20 12:31:10,712 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-20 12:31:11,348 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1726809180369_0002
2024-09-20 12:31:11,348 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-20 12:31:11,533 INFO conf.Configuration: resource-types.xml not found
2024-09-20 12:31:11,534 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-20 12:31:11,639 INFO impl.YarnClientImpl: Submitted application application_1726809180369_0002
2024-09-20 12:31:11,702 INFO mapreduce.Job: The url to track the job: http://vbox:8088/proxy/application_1726809180369_0002/
2024-09-20 12:31:11,705 INFO mapreduce.Job: Running job: job_1726809180369_0002
2024-09-20 12:31:18,959 INFO mapreduce.Job: Job job_1726809180369_0002 running in uber mode : false
2024-09-20 12:31:18,960 INFO mapreduce.Job:  map 0% reduce 0%
2024-09-20 12:31:25,168 INFO mapreduce.Job:  map 100% reduce 0%
2024-09-20 12:31:31,260 INFO mapreduce.Job:  map 100% reduce 100%
2024-09-20 12:31:32,284 INFO mapreduce.Job: Job job_1726809180369_0002 completed successfully
2024-09-20 12:31:32,387 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=102094
                FILE: Number of bytes written=1041220
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
```



```
2024-09-20 12:31:11,348 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1726809180369_0002
2024-09-20 12:31:11,348 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-20 12:31:11,533 INFO conf.Configuration: resource-types.xml not found
2024-09-20 12:31:11,534 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-20 12:31:11,639 INFO impl.YarnClientImpl: Submitted application application_1726809180369_0002
2024-09-20 12:31:11,702 INFO mapreduce.Job: The url to track the job: http://vbox:8088/proxy/application_1726809180369_0002/
2024-09-20 12:31:11,705 INFO mapreduce.Job: Running job: job_1726809180369_0002
2024-09-20 12:31:18,959 INFO mapreduce.Job: Job job_1726809180369_0002 running in uber mode : false
2024-09-20 12:31:18,960 INFO mapreduce.Job:  map 0% reduce 0%
2024-09-20 12:31:25,168 INFO mapreduce.Job:  map 100% reduce 0%
2024-09-20 12:31:31,260 INFO mapreduce.Job:  map 100% reduce 100%
2024-09-20 12:31:32,284 INFO mapreduce.Job: Job job_1726809180369_0002 completed successfully
```



```
                Total time spent by all map tasks (ms)=8284
                Total time spent by all reduce tasks (ms)=3025
                Total vcore-milliseconds taken by all map tasks=8284
                Total vcore-milliseconds taken by all reduce tasks=3025
                Total megabyte-milliseconds taken by all map tasks=8482816
                Total megabyte-milliseconds taken by all reduce tasks=3097600
        Map-Reduce Framework
                Map input records=365
                Map output records=10220
                Map output bytes=81648
                Map output materialized bytes=102100
                Input split bytes=180
                Combine input records=0
                Combine output records=0
                Reduce input groups=12
                Reduce shuffle bytes=102100
                Reduce input records=10220
                Reduce output records=12
                Spilled Records=20440
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=239
                CPU time spent (ms)=3310
                Physical memory (bytes) snapshot=901877760
                Virtual memory (bytes) snapshot=7642955776
                Total committed heap usage (bytes)=686817280
                Peak Map Physical memory (bytes)=344698880
                Peak Map Virtual memory (bytes)=2549645312
                Peak Reduce Physical memory (bytes)=221151232
                Peak Reduce Virtual memory (bytes)=2548363264
        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=83301
        File Output Format Counters
                Bytes Written=96
024-09-20 12:31:32,387 INFO streaming.StreamJob: Output directory: /exp3/newoutput
```



```
┌──(hadoop㉿kali)-[~/hadoop/bin]
└─$ ./hdfs dfs -cat /exp3/output/*
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:15:38,966 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
01      26.5
02      26.6
03      29.1
04      30.8
05      31.1
06      33.6
07      38.5
08      40.2
09      36.5
10      36.9
11      27.6
12      25.9
```

RESULT

Thus a java program has been implemented to identify the year-wise maximum temperature from the sensor data.