Amitkumar Sandesara
ass503
N11389150

Assignment No. 1

➢ Mapper

```
import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

public 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

```
import java.io.IOException;

import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public 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));
 }
```

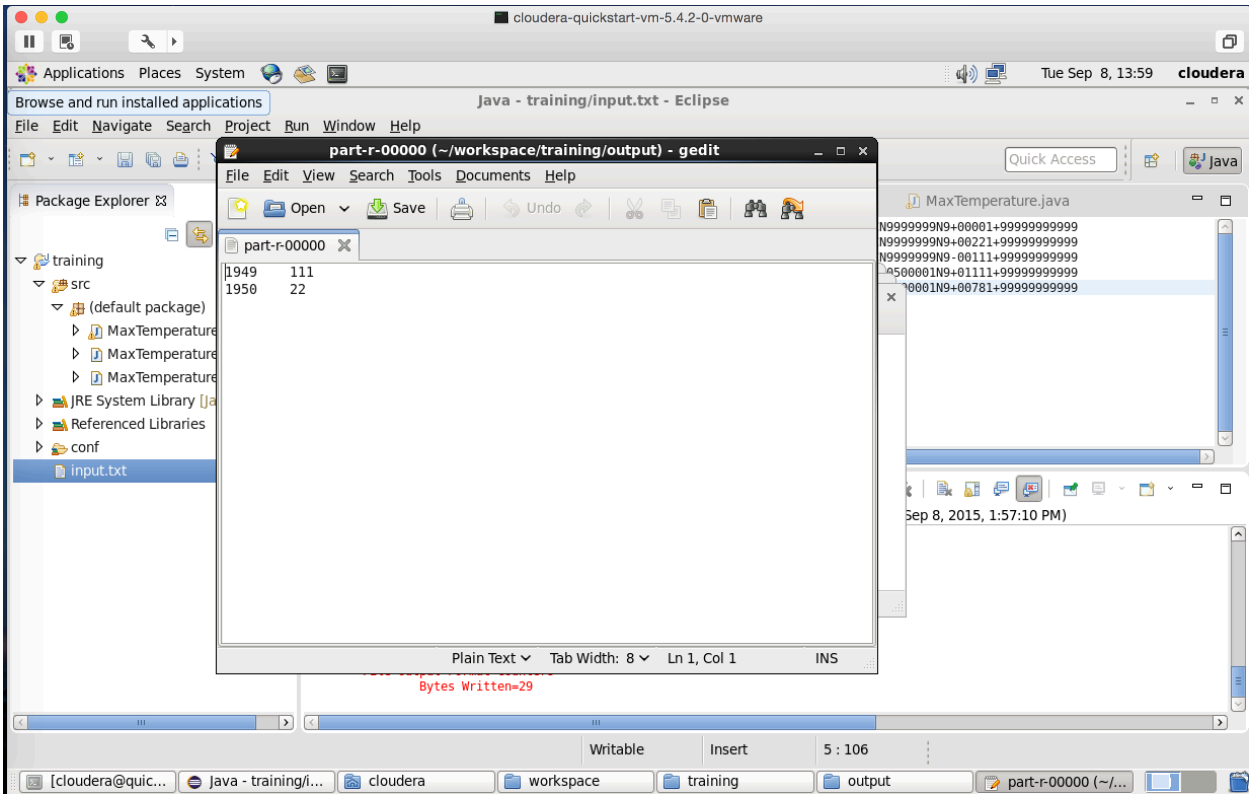
```
}
```

- **MaxTemperature.java**

```
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;

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 = new Job();
  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);

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

Amitkumar Sandesara
ass503
N11389150

Output File:



Console Output:

SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
15/09/08 14:00:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
15/09/08 14:00:45 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
15/09/08 14:00:45 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
15/09/08 14:00:45 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
15/09/08 14:00:46 WARN mapreduce.JobSubmitter: No job jar file set. User classes may not be found. See Job or Job#setJar(String).
15/09/08 14:00:46 INFO input.FileInputFormat: Total input paths to process : 1
15/09/08 14:00:46 INFO mapreduce.JobSubmitter: number of splits:1
15/09/08 14:00:46 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local603973379_0001
15/09/08 14:00:47 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
15/09/08 14:00:47 INFO mapreduce.Job: Running job: job_local603973379_0001
15/09/08 14:00:47 INFO mapred.LocalJobRunner: OutputCommitter set in config null
15/09/08 14:00:47 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
15/09/08 14:00:47 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
15/09/08 14:00:47 INFO mapred.LocalJobRunner: Waiting for map tasks
15/09/08 14:00:47 INFO mapred.LocalJobRunner: Starting task: attempt_local603973379_0001_m_000000_0
15/09/08 14:00:47 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
15/09/08 14:00:47 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]

Amitkumar Sandesara
ass503
N11389150

```
15/09/08 14:00:47 INFO mapred.MapTask: Processing split: file:/home/cloudera/workspace/training/input.txt:0+529
15/09/08 14:00:48 INFO mapreduce.Job: Job job_local603973379_0001 running in uber mode : false
15/09/08 14:00:48 INFO mapreduce.Job:  map 0% reduce 0%
15/09/08 14:00:48 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
15/09/08 14:00:48 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
15/09/08 14:00:48 INFO mapred.MapTask: soft limit at 83886080
15/09/08 14:00:48 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
15/09/08 14:00:48 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
15/09/08 14:00:48 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
15/09/08 14:00:48 INFO mapred.LocalJobRunner:
15/09/08 14:00:48 INFO mapred.MapTask: Starting flush of map output
15/09/08 14:00:48 INFO mapred.MapTask: Spilling map output
15/09/08 14:00:48 INFO mapred.MapTask: bufstart = 0; bufend = 45; bufvoid = 104857600
15/09/08 14:00:48 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214380(104857520); length =
17/6553600
15/09/08 14:00:48 INFO mapred.MapTask: Finished spill 0
15/09/08 14:00:48 INFO mapred.Task: Task:attempt_local603973379_0001_m_000000_0 is done. And is in the process of
committing
15/09/08 14:00:48 INFO mapred.LocalJobRunner: map
15/09/08 14:00:48 INFO mapred.Task: Task 'attempt_local603973379_0001_m_000000_0' done.
15/09/08 14:00:48 INFO mapred.LocalJobRunner: Finishing task: attempt_local603973379_0001_m_000000_0
15/09/08 14:00:48 INFO mapred.LocalJobRunner: map task executor complete.
15/09/08 14:00:48 INFO mapred.LocalJobRunner: Waiting for reduce tasks
15/09/08 14:00:48 INFO mapred.LocalJobRunner: Starting task: attempt_local603973379_0001_r_000000_0
15/09/08 14:00:48 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
15/09/08 14:00:48 INFO mapred.Task:  Using ResourceCalculatorProcessTree : [ ]
15/09/08 14:00:48 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@552d7308
15/09/08 14:00:48 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=681154944,
maxSingleShuffleLimit=170288736, mergeThreshold=449562272, ioSortFactor=10, memToMemMergeOutputsThreshold=10
15/09/08 14:00:48 INFO reduce.EventFetcher: attempt_local603973379_0001_r_000000_0 Thread started: EventFetcher for
fetching Map Completion Events
15/09/08 14:00:48 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
attempt_local603973379_0001_m_000000_0 decomp: 57 len: 61 to MEMORY
15/09/08 14:00:48 INFO reduce.InMemoryMapOutput: Read 57 bytes from map-output for
attempt_local603973379_0001_m_000000_0
15/09/08 14:00:49 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 57,
inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->57
15/09/08 14:00:49 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
15/09/08 14:00:49 INFO mapred.LocalJobRunner: 1 / 1 copied.
15/09/08 14:00:49 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-
outputs
15/09/08 14:00:49 INFO mapred.Merger: Merging 1 sorted segments
15/09/08 14:00:49 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50 bytes
15/09/08 14:00:49 INFO reduce.MergeManagerImpl: Merged 1 segments, 57 bytes to disk to satisfy reduce memory limit
15/09/08 14:00:49 INFO reduce.MergeManagerImpl: Merging 1 files, 61 bytes from disk
15/09/08 14:00:49 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
15/09/08 14:00:49 INFO mapred.Merger: Merging 1 sorted segments
15/09/08 14:00:49 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 50 bytes
15/09/08 14:00:49 INFO mapred.LocalJobRunner: 1 / 1 copied.
15/09/08 14:00:49 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
15/09/08 14:00:49 INFO mapred.Task: Task:attempt_local603973379_0001_r_000000_0 is done. And is in the process of
committing
15/09/08 14:00:49 INFO mapred.LocalJobRunner: 1 / 1 copied.
15/09/08 14:00:49 INFO mapred.Task: Task attempt_local603973379_0001_r_000000_0 is allowed to commit now
15/09/08 14:00:49 INFO output.FileOutputCommitter: Saved output of task 'attempt_local603973379_0001_r_000000_0' to
file:/home/cloudera/workspace/training/output/_temporary/0/task_local603973379_0001_r_000000
```

Amitkumar Sandesara
ass503
N11389150

```
15/09/08 14:00:49 INFO mapred.LocalJobRunner: reduce > reduce
15/09/08 14:00:49 INFO mapred.Task: Task 'attempt_local603973379_0001_r_000000_0' done.
15/09/08 14:00:49 INFO mapred.LocalJobRunner: Finishing task: attempt_local603973379_0001_r_000000_0
15/09/08 14:00:49 INFO mapred.LocalJobRunner: reduce task executor complete.
15/09/08 14:00:49 INFO mapreduce.Job:  map 100% reduce 100%
15/09/08 14:00:49 INFO mapreduce.Job: Job job_local603973379_0001 completed successfully
15/09/08 14:00:49 INFO mapreduce.Job: Counters: 33
        File System Counters
                FILE: Number of bytes read=1550
                FILE: Number of bytes written=518542
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
        Map-Reduce Framework
                Map input records=5
                Map output records=5
                Map output bytes=45
                Map output materialized bytes=61
                Input split bytes=113
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=61
                Reduce input records=5
                Reduce output records=2
                Spilled Records=10
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=73
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=331489280
        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=529
        File Output Format Counters
                Bytes Written=29
```

Hadoop.log file is also attached in this folder, named as hadoop.log.
Evidence is attached inside the file named part-r-00000 in this folder.