

PRATEEK ARYAN (1BM17CS065) - Average Hadoop Program

AverageDriver.java

```
package Average;
import org.apache.hadoop.io.*;
import org.apache.hadoop.fs.*;
import org.apache.hadoop.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class AverageDriver
{
    public static void main (String[] args) throws Exception
    {
        if (args.length != 2)
        {
            System.err.println("Please Enter the input and output parameters");
            System.exit(-1);
        }

        Job job = new Job();
        job.setJarByClass(AverageDriver.class);
        job.setJobName("Max temperature");

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

        job.setMapperClass(AverageMapper.class);
        job.setReducerClass(AverageReducer.class);

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

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

AverageMapper.java

```
package Average;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.*;
import java.io.IOException;

public class AverageMapper extends Mapper <LongWritable, Text, Text, IntWritable>
{
    public 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 temperature;
        if (line.charAt(87)=='+')
            temperature = Integer.parseInt(line.substring(88, 92));
        else
            temperature = Integer.parseInt(line.substring(87, 92));

        String quality = line.substring(92, 93);
        if(temperature != MISSING && quality.matches("[01459]"))
            context.write(new Text(year),new IntWritable(temperature));
    }
}
```

```
}  
}
```

AverageReducer.java

```
package Average;  
import org.apache.hadoop.io.IntWritable;  
import org.apache.hadoop.io.Text;  
import org.apache.hadoop.mapreduce.*;  
import java.io.IOException;  
  
public class AverageReducer extends Reducer <Text, IntWritable,Text, IntWritable >  
{  
    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,  
        InterruptedException  
    {  
        int max_temp = 0;  
        int count = 0;  
        for (IntWritable value : values)  
        {  
            max_temp += value.get();  
            count+=1;  
        }  
        context.write(key, new IntWritable(max_temp/count));  
    }  
}
```

Manifest.txt

```
Main-Class: Average.AverageDriver
```

Notice we have added `package Average;` to all the java files.

```
cd  
  
mkdir lab8  
  
cd lab8  
  
touch 1901  
  
touch AverageDriver.java  
  
touch AverageMapper.java  
  
touch AverageReducer.java  
  
touch Manifest.txt  
  
export CLASSPATH="/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.6.0.jar:/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.6.0.jar:/usr/local/hadoop/share/hadoop/common/hadoop-common-2.6.0.jar:/home/hduser/lab8/*:/usr/local/hadoop/lib/*"  
  
javac -d . AverageMapper.java AverageReducer.java AverageDriver.java  
  
jar cfm avg.jar Manifest.txt Average/*.class
```

```
cd /usr/local/hadoop/sbin/

./stop-all.sh

./start-dfs.sh

./start-yarn.sh
```

```
cd

cd lab8

hadoop fs -mkdir /lab8

hadoop fs -copyFromLocal /home/hduser/lab8/1901 /lab8/1901

hadoop fs -ls /lab8

hadoop jar /home/hduser/lab8/avg.jar /lab8/1901 /lab8/output

hadoop fs -ls /lab8/output

hadoop fs -cat /lab8/output/part-r-00000
```

```
hduser@ip-172-31-90-179:~/lab8$ cd /usr/local/hadoop/sbin/
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ ./stop-all.sh
This script is Deprecated. Instead use stop-dfs.sh and stop-yarn.sh
20/12/10 12:31:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Stopping namenodes on [localhost]
localhost: stopping namenode
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
0.0.0.0: stopping secondarynamenode
20/12/10 12:32:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
stopping yarn daemons
stopping resourcemanager
localhost: stopping nodemanager
localhost: nodemanager did not stop gracefully after 5 seconds: killing with kill -9
no proxyserver to stop
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ ./start-dfs.sh
20/12/10 12:32:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-ip-17
2-31-90-179.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-ip-17
2-31-90-179.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-seconda
rynamenode-ip-172-31-90-179.out
20/12/10 12:32:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-ip-1
72-31-90-179.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-i
p-172-31-90-179.out
```

```
hduser@ip-172-31-90-179:/usr/local/hadoop/sbin$ cd
hduser@ip-172-31-90-179:~$ cd lab8
```

```
hduser@ip-172-31-90-179:~/lab8$ ls -l
total 884
-rw-rw-r-- 1 ubuntu ubuntu 888190 Dec 10 12:20 1901
-rwxr-xr-x 1 hduser hadoop 1429 Dec 10 11:58 AverageDriver.java
-rwxr-xr-x 1 hduser hadoop 1199 Dec 10 11:58 AverageMapper.java
-rwxr-xr-x 1 hduser hadoop 835 Dec 10 11:58 AverageReducer.java
-rw-r--r-- 1 hduser hadoop 34 Dec 10 12:00 Manifest.txt
```

```
hduser@ip-172-31-90-179:~/lab8$ export CLASSPATH="/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.6.0.jar:/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.6.0.jar:/usr/local/hadoop/share/hadoop/common/hadoop-common-2.6.0.jar:/home/hduser/lab8/*:/usr/local/hadoop/lib/*"
```

```
hduser@ip-172-31-90-179:~/lab8$ javac -d . AverageMapper.java AverageReducer.java AverageDriver.java
Note: AverageDriver.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
hduser@ip-172-31-90-179:~/lab8$ ls
1901 AverageDriver.java AverageReducer.java
Average AverageMapper.java Manifest.txt
```

```
hduser@ip-172-31-90-179:~/lab8$ ls -l Average
total 12
-rw-r--r-- 1 hduser hadoop 1417 Dec 10 12:23 AverageDriver.class
-rw-r--r-- 1 hduser hadoop 1870 Dec 10 12:23 AverageMapper.class
-rw-r--r-- 1 hduser hadoop 1624 Dec 10 12:23 AverageReducer.class
```

```
hduser@ip-172-31-90-179:~/lab8$ jar cfm avg.jar Manifest.txt Average/*.class
hduser@ip-172-31-90-179:~/lab8$ ls -l
total 892
-rw-rw-r-- 1 ubuntu ubuntu 888190 Dec 10 12:20 1901
drwxr-xr-x 2 hduser hadoop 4096 Dec 10 12:23 Average
-rwxr-xr-x 1 hduser hadoop 1429 Dec 10 11:58 AverageDriver.java
-rwxr-xr-x 1 hduser hadoop 1199 Dec 10 11:58 AverageMapper.java
-rwxr-xr-x 1 hduser hadoop 835 Dec 10 11:58 AverageReducer.java
-rw-r--r-- 1 hduser hadoop 34 Dec 10 12:00 Manifest.txt
-rw-r--r-- 1 hduser hadoop 3110 Dec 10 12:24 avg.jar
```

```
hduser@ip-172-31-90-179:~/lab8$ hadoop fs -mkdir /lab8
20/12/10 12:26:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@ip-172-31-90-179:~/lab8$ hadoop fs -copyFromLocal /home/hduser/lab8/1901 /lab8/1901
20/12/10 12:27:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@ip-172-31-90-179:~/lab8$
hduser@ip-172-31-90-179:~/lab8$ hadoop fs -ls /lab8
20/12/10 12:27:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 hduser supergroup 888190 2020-12-10 12:27 /lab8/1901
```

```

hduser@ip-172-31-90-179:~/lab8$ hadoop jar /home/hduser/lab8/avg.jar /lab8/1901 /lab8/output
20/12/10 12:28:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
20/12/10 12:28:21 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs
.metrics.session-id
20/12/10 12:28:21 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker,
sessionId=
20/12/10 12:28:21 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not perfor
med. Implement the Tool interface and execute your application with ToolRunner to remedy thi
s.
20/12/10 12:28:21 INFO input.FileInputFormat: Total input paths to process : 1
20/12/10 12:28:21 INFO mapreduce.JobSubmitter: number of splits:1
20/12/10 12:28:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local205642446
8_0001
20/12/10 12:28:22 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
20/12/10 12:28:22 INFO mapreduce.Job: Running job: job_local2056424468_0001
20/12/10 12:28:22 INFO mapred.LocalJobRunner: OutputCommitter set in config null
20/12/10 12:28:22 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce
.lib.output.FileOutputCommitter
20/12/10 12:28:22 INFO mapred.LocalJobRunner: Waiting for map tasks
20/12/10 12:28:22 INFO mapred.LocalJobRunner: Starting task: attempt_local2056424468_0001_m_
0000000_0
20/12/10 12:28:22 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
20/12/10 12:28:22 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/lab8/1901:0+
888190
20/12/10 12:28:22 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
20/12/10 12:28:22 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
20/12/10 12:28:22 INFO mapred.MapTask: soft limit at 83886080
20/12/10 12:28:22 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
20/12/10 12:28:22 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
20/12/10 12:28:22 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred
.MapTask$MapOutputBuffer

```

```

HDFS: Number of bytes written=8
HDFS: Number of read operations=13
HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Map-Reduce Framework
  Map input records=6565
  Map output records=6564
  Map output bytes=59076
  Map output materialized bytes=72210
  Input split bytes=97
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=72210
  Reduce input records=6564
  Reduce output records=1
  Spilled Records=13128
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=53
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
  Total committed heap usage (bytes)=242360320
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=888190
File Output Format Counters
  Bytes Written=8
hduser@ip-172-31-90-179:~/lab8$

```

```
hduser@ip-172-31-90-179:~/lab8$ hadoop fs -ls /lab8/output
20/12/10 12:29:29 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 hduser supergroup          0 2020-12-10 12:28 /lab8/output/_SUCCESS
-rw-r--r--    1 hduser supergroup          8 2020-12-10 12:28 /lab8/output/part-r-00000
hduser@ip-172-31-90-179:~/lab8$ hadoop fs -cat /lab8/output/part-r-00000
20/12/10 12:29:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
1901    46
```