

Exercise-3: Map Reduce (Programs)

Use the Hadoop framework to write a custom MapReduce program to perform word count operation on a custom data set .

JAVA CODE:

```
package shwethahebbbar152;

import java.io.IOException;
import java.util.*;
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.mapred.FileInputFormat;
import org.apache.hadoop.mapred.FileOutputFormat;
import org.apache.hadoop.mapred.JobClient;
import org.apache.hadoop.mapred.JobConf;
import org.apache.hadoop.mapred.MapReduceBase;
import org.apache.hadoop.mapred.Mapper;
import org.apache.hadoop.mapred.OutputCollector;
import org.apache.hadoop.mapred.Reducer;
import org.apache.hadoop.mapred.Reporter;
import org.apache.hadoop.mapred.TextOutputFormat;
import org.apache.hadoop.mapred.TextInputFormat;

public class WordCount {
    public static class Map extends MapReduceBase implements
    Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable>
        output, Reporter reporter) throws IOException {
            String line = value.toString();
            StringTokenizer tokenizer = new StringTokenizer(line);
            while (tokenizer.hasMoreTokens()) {
                word.set(tokenizer.nextToken());
                output.collect(word, one);
            }
        }
    }
}
```

```

public static class Reduce extends MapReduceBase implements
Reducer<Text, IntWritable, Text, IntWritable> {
public void reduce(Text key, Iterator<IntWritable> values,
OutputCollector<Text, IntWritable> output, Reporter reporter)
throws IOException {
int sum = 0;
while (values.hasNext()) {
sum += values.next().get();
}
output.collect(key, new IntWritable(sum));
}
}

public static void main(String[] args) throws Exception {
JobConf conf = new JobConf(WordCount.class);
conf.setJobName("wordcount");
conf.setOutputKeyClass(Text.class);
conf.setOutputValueClass(IntWritable.class);
conf.setMapperClass(Map.class);
conf.setCombinerClass(Reduce.class);
conf.setReducerClass(Reduce.class);
conf.setInputFormat(TextInputFormat.class);
conf.setOutputFormat(TextOutputFormat.class);
FileInputFormat.setInputPaths(conf, new Path(args[0]));
FileOutputFormat.setOutputPath(conf, new Path(args[1]));
JobClient.runJob(conf);
}

```

}

WordCount.java

```
1 package shwethahebbbar152;
2
3 import java.io.IOException;
4 import java.util.*;
5 import org.apache.hadoop.fs.Path;
6 import org.apache.hadoop.io.IntWritable;
7 import org.apache.hadoop.io.LongWritable;
8 import org.apache.hadoop.io.Text;
9 import org.apache.hadoop.mapred.FileInputFormat;
10 import org.apache.hadoop.mapred.FileOutputFormat;
11 import org.apache.hadoop.mapred.JobClient;
12 import org.apache.hadoop.mapred.JobConf;
13 import org.apache.hadoop.mapred.MapReduceBase;
14 import org.apache.hadoop.mapred.Mapper;
15 import org.apache.hadoop.mapred.OutputCollector;
16 import org.apache.hadoop.mapred.Reducer;
17 import org.apache.hadoop.mapred.Reporter;
18 import org.apache.hadoop.mapred.TextOutputFormat;
19 import org.apache.hadoop.mapred.TextInputFormat;
20
21 public class WordCount {
22     public static class Map extends MapReduceBase implements
23     Mapper<LongWritable, Text, Text, IntWritable> {
24         private final static IntWritable one = new IntWritable(1);
25         private Text word = new Text();
26     public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable>
27     output, Reporter reporter) throws IOException {
28         String line = value.toString();
29         StringTokenizer tokenizer = new StringTokenizer(line);
30         while (tokenizer.hasMoreTokens()) {
31             word.set(tokenizer.nextToken());
32             output.collect(word, one);
33         }
34     }
35 }
```

```

WordCount.java
29 StringTokenizer tokenizer = new StringTokenizer(line);
30 while (tokenizer.hasMoreTokens()) {
31     word.set(tokenizer.nextToken());
32     output.collect(word, one);
33 }
34 }
35 }
36 public static class Reduce extends MapReduceBase implements
37     Reducer<Text, IntWritable, Text, IntWritable> {
38 public void reduce(Text key, Iterator<IntWritable> values,
39     OutputCollector<Text, IntWritable> output, Reporter reporter)
40     throws IOException {
41     int sum = 0;
42     while (values.hasNext()) {
43         sum += values.next().get();
44     }
45     output.collect(key, new IntWritable(sum));
46 }
47 }
48 public static void main(String[] args) throws Exception {
49     JobConf conf = new JobConf(WordCount.class);
50     conf.setJobName("wordcount");
51     conf.setOutputKeyClass(Text.class);
52     conf.setOutputValueClass(IntWritable.class);
53     conf.setMapperClass(Map.class);
54     conf.setCombinerClass(Reduce.class);
55     conf.setReducerClass(Reduce.class);
56     conf.setInputFormat(TextInputFormat.class);
57     conf.setOutputFormat(TextOutputFormat.class);
58     FileInputFormat.setInputPaths(conf, new Path(args[0]));
59     FileOutputFormat.setOutputPath(conf, new Path(args[1]));
60     JobClient.runJob(conf);
61 }
62 }

```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME/sbin
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop 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 [admin1-HP-280-G4-MT-Business-PC]
Starting resourcemanager
Starting nodemanagers
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
16690 DataNode
17122 ResourceManager
16532 NameNode
16933 SecondaryNameNode
15545 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
17628 Jps
17278 NodeManager
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs fsck /
Connecting to namenode via http://localhost:9870/fsck?ugi=hadoop&path=%2F
FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Fri Jun 17 12:38:39 IST 2022

Status: HEALTHY
Number of data-nodes: 1
Number of racks: 1
Total dirs: 20
Total symlinks: 0

Replicated Blocks:
Total size: 100446138 B
Total files: 45
Total blocks (validated): 42 (avg. block size 2391574 B)
Minimally replicated blocks: 42 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0

```

```
Minimally replicated blocks: 42 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Missing blocks: 0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
```

Erasure Coded Block Groups:

```
Total size: 0 B
Total files: 0
Total block groups (validated): 0
Minimally erasure-coded block groups: 0
Over-erasure-coded block groups: 0
Under-erasure-coded block groups: 0
Unsatisfactory placement block groups: 0
Average block group size: 0.0
Missing block groups: 0
Corrupt block groups: 0
Missing internal blocks: 0
```

FSCK ended at Fri Jun 17 12:38:39 IST 2022 in 14 milliseconds

The filesystem under path '/' is HEALTHY

```
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -ls /
```

Found 6 items

```
drwxr-xr-x - hadoop supergroup 0 2022-05-24 10:18 /int19is080
drwxr-xr-x - hadoop supergroup 0 2022-06-13 15:07 /int19is119
drwxr-xr-x - hadoop supergroup 0 2022-05-24 09:30 /manvesh
drwxr-xr-x - hadoop supergroup 0 2022-06-13 14:27 /test
drwxrwxr-x - hadoop supergroup 0 2021-07-28 12:58 /tmp
drwxr-xr-x - hadoop supergroup 0 2021-07-28 12:56 /user
```

```
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /shwetha
```

```
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -touchz /shwetha/first.txt
```

```
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - /shwetha/text.txt
```

the girl in the town is the only girl in the town

```
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - /shwetha/text.txt
```

the girl in the town is the only girl in the town

```
2022-06-17 12:50:44,010 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/shwetha.jar /shwetha /out
2022-06-17 12:52:02,102 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 12:52:02,239 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-17 12:52:02,384 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
2022-06-17 12:52:02,401 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_165
5449652406_0001
2022-06-17 12:52:02,490 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 12:52:02,589 INFO mapred.FileInputFormat: Total input files to process : 2
2022-06-17 12:52:02,625 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 12:52:02,667 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 12:52:03,092 INFO mapreduce.JobSubmitter: number of splits:3
2022-06-17 12:52:03,192 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-06-17 12:52:03,208 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655449652406_0001
2022-06-17 12:52:03,208 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-17 12:52:03,332 INFO conf.Configuration: resource-types.xml not found
2022-06-17 12:52:03,332 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-17 12:52:03,678 INFO impl.YarnClientImpl: Submitted application application_1655449652406_0001
2022-06-17 12:52:03,736 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655449652
406_0001/
2022-06-17 12:52:03,737 INFO mapreduce.Job: Running job: job_1655449652406_0001
2022-06-17 12:52:08,805 INFO mapreduce.Job: Job job_1655449652406_0001 running in uber mode : false
2022-06-17 12:52:08,807 INFO mapreduce.Job: map 0% reduce 0%
2022-06-17 12:52:13,876 INFO mapreduce.Job: map 100% reduce 0%
2022-06-17 12:52:17,898 INFO mapreduce.Job: map 100% reduce 100%
2022-06-17 12:52:17,919 INFO mapreduce.Job: Job job_1655449652406_0001 completed successfully
2022-06-17 12:52:17,995 INFO mapreduce.Job: Counters: 55
File System Counters
FILE: Number of bytes read=67
FILE: Number of bytes written=903415
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=346
HDFS: Number of bytes written=37
```

```

2022-06-17 12:52:17,919 INFO mapreduce.Job: Job job_1655449652406_0001 completed successfully
2022-06-17 12:52:17,995 INFO mapreduce.Job: Counters: 55
  File System Counters
    FILE: Number of bytes read=67
    FILE: Number of bytes written=903415
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=346
    HDFS: Number of bytes written=37
    HDFS: Number of read operations=14
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
    HDFS: Number of bytes read erasure-coded=0
  Job Counters
    Launched map tasks=3
    Launched reduce tasks=1
    Other local map tasks=1
    Data-local map tasks=2
    Total time spent by all maps in occupied slots (ms)=7150
    Total time spent by all reduces in occupied slots (ms)=1512
    Total time spent by all map tasks (ms)=7150
    Total time spent by all reduce tasks (ms)=1512
    Total vcore-milliseconds taken by all map tasks=7150
    Total vcore-milliseconds taken by all reduce tasks=1512
    Total megabyte-milliseconds taken by all map tasks=7321600
    Total megabyte-milliseconds taken by all reduce tasks=1548288
  Map-Reduce Framework
    Map input records=1
    Map output records=12
    Map output bytes=98
    Map output materialized bytes=79
    Input split bytes=271
    Combine input records=12
    Combine output records=6
    Reduce input groups=6
    Reduce shuffle bytes=79
    Reduce input records=6
    Reduce output records=6
    Combine output records=6
    Reduce input groups=6
    Reduce shuffle bytes=79
    Reduce input records=6
    Reduce output records=6
    Spilled Records=12
    Shuffled Maps =3
    Failed Shuffles=0
    Merged Map outputs=3
    GC time elapsed (ms)=354
    CPU time spent (ms)=1480
    Physical memory (bytes) snapshot=1119911936
    Virtual memory (bytes) snapshot=10147336192
    Total committed heap usage (bytes)=1164443648
    Peak Map Physical memory (bytes)=336830464
    Peak Map Virtual memory (bytes)=2534244352
    Peak Reduce Physical memory (bytes)=190541824
    Peak Reduce Virtual memory (bytes)=2544775168
  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=75
  File Output Format Counters
    Bytes Written=37
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /out/part*
2022-06-17 12:52:45,602 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
girl      2
in        2
is        1
only      1
the       4
town      2
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$

```

```

hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME/sbin
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop 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 [admin1-HP-280-G4-MT-Business-PC]

Starting resourcemanager

Starting nodemanagers

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ jps

16690 DataNode

17122 ResourceManager

16532 NameNode

16933 SecondaryNameNode

15545 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar

17628 Jps

17278 NodeManager

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hdfs fsck /

Connecting to namenode via http://localhost:9870/fsck?ugi=hadoop&path=%2F

FSCK started by hadoop (auth:SIMPLE) from /127.0.0.1 for path / at Fri Jun 17 12:38:39 IST 2022

Status: HEALTHY

Number of data-nodes: 1

Number of racks: 1

Total dirs: 20

Total symlinks: 0

Replicated Blocks:

Total size: 100446138 B

Total files: 45

Total blocks (validated): 42 (avg. block size 2391574 B)

Minimally replicated blocks: 42 (100.0 %)

Over-replicated blocks: 0 (0.0 %)

Under-replicated blocks: 0 (0.0 %)

Mis-replicated blocks: 0 (0.0 %)

Default replication factor: 1

Average block replication: 1.0

Missing blocks: 0

Corrupt blocks: 0

Missing replicas: 0 (0.0 %)

Erasure Coded Block Groups:

Total size: 0 B

Total files: 0

Total block groups (validated): 0

Minimally erasure-coded block groups: 0

Over-erasure-coded block groups: 0

Under-erasure-coded block groups: 0

Unsatisfactory placement block groups: 0

Average block group size: 0.0

Missing block groups: 0

Corrupt block groups: 0

Missing internal blocks: 0

FSCK ended at Fri Jun 17 12:38:39 IST 2022 in 14 milliseconds

The filesystem under path '/' is HEALTHY

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hdfs dfs -ls /

Found 6 items

drwxr-xr-x - hadoop supergroup 0 2022-05-24 10:18 /Int19is080

drwxr-xr-x - hadoop supergroup 0 2022-06-13 15:07 /Int19is119

drwxr-xr-x - hadoop supergroup 0 2022-05-24 09:30 /manvesh

drwxr-xr-x - hadoop supergroup 0 2022-06-13 14:27 /test

drwxrwxr-x - hadoop supergroup 0 2021-07-28 12:58 /tmp

drwxr-xr-x - hadoop supergroup 0 2021-07-28 12:56 /user

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hdfs dfs -mkdir -p /shwetha

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hdfs dfs -touchz

/shwetha/first.txt

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hdfs dfs -appendToFile -

/shwetha/text.txt

the girl in the town is the only girl in the town

2022-06-17 12:50:44,010 INFO sasl.SaslDataTransferClient: SASL encryption trust check:

localhostTrusted = false, remoteHostTrusted = false

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hadoop jar

/home/hadoop/Desktop/shwetha.jar /shwetha /out

2022-06-17 12:52:02,102 INFO client.RMProxy: Connecting to ResourceManager at

/127.0.0.1:8032

2022-06-17 12:52:02,239 INFO client.RMProxy: Connecting to ResourceManager at

/127.0.0.1:8032

2022-06-17 12:52:02,384 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

2022-06-17 12:52:02,401 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1655449652406_0001

2022-06-17 12:52:02,490 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-06-17 12:52:02,589 INFO mapred.FileInputFormat: Total input files to process : 2

2022-06-17 12:52:02,625 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-06-17 12:52:02,667 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-06-17 12:52:03,092 INFO mapreduce.JobSubmitter: number of splits:3

2022-06-17 12:52:03,192 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false

2022-06-17 12:52:03,208 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655449652406_0001

2022-06-17 12:52:03,208 INFO mapreduce.JobSubmitter: Executing with tokens: []

2022-06-17 12:52:03,332 INFO conf.Configuration: resource-types.xml not found

2022-06-17 12:52:03,332 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2022-06-17 12:52:03,678 INFO impl.YarnClientImpl: Submitted application application_1655449652406_0001

2022-06-17 12:52:03,736 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655449652406_0001/

2022-06-17 12:52:03,737 INFO mapreduce.Job: Running job: job_1655449652406_0001

2022-06-17 12:52:08,805 INFO mapreduce.Job: Job job_1655449652406_0001 running in uber mode : false

2022-06-17 12:52:08,807 INFO mapreduce.Job: map 0% reduce 0%

2022-06-17 12:52:13,876 INFO mapreduce.Job: map 100% reduce 0%

2022-06-17 12:52:17,898 INFO mapreduce.Job: map 100% reduce 100%

2022-06-17 12:52:17,919 INFO mapreduce.Job: Job job_1655449652406_0001 completed successfully

2022-06-17 12:52:17,995 INFO mapreduce.Job: Counters: 55

File System Counters

- FILE: Number of bytes read=67
- FILE: Number of bytes written=903415
- FILE: Number of read operations=0
- FILE: Number of large read operations=0
- FILE: Number of write operations=0
- HDFS: Number of bytes read=346

HDFS: Number of bytes written=37
HDFS: Number of read operations=14
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0

Job Counters

Launched map tasks=3
Launched reduce tasks=1
Other local map tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=7150
Total time spent by all reduces in occupied slots (ms)=1512
Total time spent by all map tasks (ms)=7150
Total time spent by all reduce tasks (ms)=1512
Total vcore-milliseconds taken by all map tasks=7150
Total vcore-milliseconds taken by all reduce tasks=1512
Total megabyte-milliseconds taken by all map tasks=7321600
Total megabyte-milliseconds taken by all reduce tasks=1548288

Map-Reduce Framework

Map input records=1
Map output records=12
Map output bytes=98
Map output materialized bytes=79
Input split bytes=271
Combine input records=12
Combine output records=6
Reduce input groups=6
Reduce shuffle bytes=79
Reduce input records=6
Reduce output records=6
Spilled Records=12
Shuffled Maps =3
Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=354
CPU time spent (ms)=1480
Physical memory (bytes) snapshot=1119911936
Virtual memory (bytes) snapshot=10147336192
Total committed heap usage (bytes)=1164443648
Peak Map Physical memory (bytes)=336830464

Peak Map Virtual memory (bytes)=2534244352
Peak Reduce Physical memory (bytes)=190541824
Peak Reduce Virtual memory (bytes)=2544775168

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=75

File Output Format Counters

Bytes Written=37

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$ hdfs dfs -cat /out/part*

2022-06-17 12:52:45,602 INFO sasl.SaslDataTransferClient: SASL encryption trust check:

localhostTrusted = false, remoteHostTrusted = false

girl 2

in 2

is 1

only 1

the 4

town 2

hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin\$