

Big Data Program - 04

Name : vishal kumar H K

Usn : 1NT19IS189

```
TransactionCount.java
1 package prog4;
2 import java.io.IOException;
3
4
5 public class TransactionCount {
6 //MAPPER CODE
7
8 public static class Map extends MapReduceBase implements
9 Mapper<LongWritable, Text, Text, IntWritable> {
10 private final static IntWritable one = new IntWritable(1);
11 //private Text word = new Text();
12
13 public void map(LongWritable key, Text value, OutputCollector<Text,
14 IntWritable> output, Reporter reporter) throws IOException {
15 String myString = value.toString();
16 String[] userCount = myString.split(",");
17 output.collect(new Text(userCount[3]), one);
18 }
19 }
20
21 //REDUCER CODE
22 public static class Reduce extends MapReduceBase implements
23 Reducer<Text, IntWritable, Text, IntWritable> {
24
25 public void reduce(Text key, Iterator<IntWritable> values,
26 OutputCollector<Text, IntWritable> output, Reporter reporter) throws
27 IOException {
28 int finaluserCount = 0 ;
29 Text mykey = key ;
30 while(values.hasNext()) {
31 IntWritable value = values.next();
32 finaluserCount += value.get();
33 }
34 output.collect(mykey, new IntWritable(finaluserCount));
35 }
36 }
37
38 //DRIVER CODE
39 public static void main(String[] args) throws Exception {
40 JobConf conf = new JobConf(TransactionCount.class);
41 conf.setJobName("wordcount");
42 conf.setOutputKeyClass(Text.class);
43 conf.setOutputValueClass(IntWritable.class);
44 conf.setMapperClass(Map.class);
45 conf.setCombinerClass(Reduce.class);
46 conf.setReducerClass(Reduce.class);
47 conf.setInputFormat(TextInputFormat.class);
48 conf.setOutputFormat(TextOutputFormat.class);
49 FileInputFormat.setInputPaths(conf, new Path(args[0]));
50 }
```

Command :

Cd \$HADOOP_HOME

Cd sbin

Start-all.sh

Hdfs dfs -mkdir -p/input123

Hdfs dfs

-copyFromLocal/home/hdoop/myfiles11/bank_data.csv/input101

Hadoop jar /home/hdoop/desktop/vishal.jar/input101/output1011

Hdfs dfs -cat /output1011/part*

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin
hdoop@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
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
5585 NodeManager
4277 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
6023 Jps
5002 DataNode
5434 ResourceManager
4861 NameNode
5246 SecondaryNameNode
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p /input155
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -copyFromLocal /home/hdoop/myfiles11/bank_data.csv /input101
copyFromLocal: '/input101': File exists
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -copyFromLocal /home/hdoop/myfiles11/bank_data.csv /input101
copyFromLocal: '/input101': File exists
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hdoop/Desktop/vishal.jar /input101 /output101
2022-06-28 14:19:38,286 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-28 14:19:38,419 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
Exception in thread "main" org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory hdfs://127.0.0.1:9000/output101 already exist
s
    at org.apache.hadoop.mapred.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:131)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:279)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:143)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1570)
    at org.apache.hadoop.mapreduce.Job$11.run(Job.java:1567)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1730)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:576)
```

```

    at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hadoop jar /home/hadoop/Desktop/vishal.jar /input101 /output101
2022-06-28 14:22:31,227 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-28 14:22:31,346 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-28 14:22:31,478 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and
execute your application with ToolRunner to remedy this.
2022-06-28 14:22:31,518 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_165
6405214921_0001
2022-06-28 14:22:31,665 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-28 14:22:32,218 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-28 14:22:32,294 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-28 14:22:32,751 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-28 14:22:33,168 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-28 14:22:33,483 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
2022-06-28 14:22:33,901 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1656405214921_0001
2022-06-28 14:22:33,901 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-28 14:22:34,014 INFO conf.Configuration: resource-types.xml not found
2022-06-28 14:22:34,015 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-28 14:22:34,199 INFO impl.YarnClientImpl: Submitted application application_1656405214921_0001
2022-06-28 14:22:34,227 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1656405214
921_0001/
2022-06-28 14:22:34,228 INFO mapreduce.Job: Running job: job_1656405214921_0001
2022-06-28 14:22:40,323 INFO mapreduce.Job: Job job_1656405214921_0001 running in uber mode : false
2022-06-28 14:22:40,325 INFO mapreduce.Job: map 0% reduce 0%
2022-06-28 14:22:43,388 INFO mapreduce.Job: map 50% reduce 0%
2022-06-28 14:22:44,399 INFO mapreduce.Job: map 100% reduce 0%
2022-06-28 14:22:48,440 INFO mapreduce.Job: map 100% reduce 100%
2022-06-28 14:22:49,467 INFO mapreduce.Job: Job job_1656405214921_0001 completed successfully
2022-06-28 14:22:49,529 INFO mapreduce.Job: Counters: 54
  File System Counters
    FILE: Number of bytes read=63
    FILE: Number of bytes written=677558
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=334
    HDFS: Number of bytes written=30
    HDFS: Number of read operations=11
    HDFS: Number of write operations=0

```

```

    HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=2
  Launched reduce tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=3546
  Total time spent by all reduces in occupied slots (ms)=2013
  Total time spent by all map tasks (ms)=3546
  Total time spent by all reduce tasks (ms)=2013
  Total vcore-milliseconds taken by all map tasks=3546
  Total vcore-milliseconds taken by all reduce tasks=2013
  Total megabyte-milliseconds taken by all map tasks=3631104
  Total megabyte-milliseconds taken by all reduce tasks=2061312
Map-Reduce Framework
  Map input records=5
  Map output records=5
  Map output bytes=47
  Map output materialized bytes=69
  Input split bytes=164
  Combine input records=5
  Combine output records=5
  Reduce input groups=4
  Reduce shuffle bytes=69
  Reduce input records=5
  Reduce output records=4
  Spilled Records=10
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=141
  CPU time spent (ms)=1200
  Physical memory (bytes) snapshot=878354432
  Virtual memory (bytes) snapshot=7615954944
  Total committed heap usage (bytes)=882900992
  Peak Map Physical memory (bytes)=341823488
  Peak Map Virtual memory (bytes)=2537697280
  Peak Reduce Physical memory (bytes)=239497216
  Peak Reduce Virtual memory (bytes)=2543996928

```

```

    Input split bytes=164
    Combine input records=5
    Combine output records=5
    Reduce input groups=4
    Reduce shuffle bytes=69
    Reduce input records=5
    Reduce output records=4
    Spilled Records=10
    Shuffled Maps =2
    Failed Shuffles=0
    Merged Map outputs=2
    GC time elapsed (ms)=141
    CPU time spent (ms)=1200
    Physical memory (bytes) snapshot=878354432
    Virtual memory (bytes) snapshot=7615954944
    Total committed heap usage (bytes)=882900992
    Peak Map Physical memory (bytes)=341823488
    Peak Map Virtual memory (bytes)=2537697280
    Peak Reduce Physical memory (bytes)=239497216
    Peak Reduce Virtual memory (bytes)=2543996928
  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=170
  File Output Format Counters
    Bytes Written=30
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat /output1011/part*
2022-06-28 14:23:39,707 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localhostTrusted = false, remoteHostTrusted = false
1200 2
1500 1
1700 1
Amount 1
hadoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$

```

```

hive> select * from bank1;
OK
NULL      Card name      NaMe      NULL
1         AXIS      Rajesh    1200
2         CANARA    Pratheek      1500
3         Syndicate      Praneeth      1200
4         SBI      Sohan      1700
Time taken: 0.101 seconds, Fetched: 5 row(s)
hive>

```