

Vansh Sharma
1NT19IS181
C2 BATCH

Date: 09/06/2022

EXPERIMENT 3

-> BUILD PATH :-

/home/hadoop/hadoop-3.2.1/share/hadoop/common/hadoop-common-3.2.1.jar
/home/hadoop/hadoop-3.2.1/share/hadoop/mapreduce/hadoop-mapreduce-client-core-3.2.1.jar

WordCount.java File

```
package vansh;

import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
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.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class WordCount {
    public static class TokenizerMapper
    extends Mapper<Object, Text, Text, IntWritable>
    {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        public void map(Object key, Text value, Context context
        ) throws IOException, InterruptedException {
            StringTokenizer itr = new StringTokenizer(value.toString());
```

```

while (itr.hasMoreTokens()) {
word.set(itr.nextToken());
context.write(word, one);
}
}
}

public static class IntSumReducer
extends Reducer<Text,IntWritable,Text,IntWritable> {
private IntWritable result = new IntWritable();
public void reduce(Text key, Iterable<IntWritable> values,
Context context
) throws IOException, InterruptedException {
int sum = 0;
for (IntWritable val : values) {
sum += val.get();
}
result.set(sum);
context.write(key, result);
}
}

public static void main(String[] args) throws Exception {
Configuration conf = new Configuration();
Job job = Job.getInstance(conf, "word count");

job.setJarByClass(WordCount.class);
job.setMapperClass(TokenizerMapper.class);
job.setCombinerClass(IntSumReducer.class);
job.setReducerClass(IntSumReducer.class);
job.setOutputKeyClass(Text.class);
job.setOutputValueClass(IntWritable.class);
FileInputFormat.addInputPath(job, new Path(args[0]));
FileOutputFormat.setOutputPath(job, new Path(args[1]));
System.exit(job.waitForCompletion(true) ? 0 : 1);
}
}

```

Commands:-

doop@admin1-HP-280-G4-MT-Business-PC:~\$ **cd \$HDOOP_HOME**

```
hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd sbin
bash: cd: sbin: No such file or directory
hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd /hadoop/sbin
bash: cd: /hadoop/sbin: No such file or directory
hadoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME/sbin
hadoop@admin1-HP-280-G4-MT-Business-PC:/sbin$ start -all.sh
```

Command 'start' not found, did you mean:

- command 'stars' from snap stars (2.7jrc3)
- command 'rstart' from deb x11-session-utils (7.7+4)
- command 'tart' from deb tart (3.10-1build1)
- command 'stat' from deb coreutils (8.30-3ubuntu2)
- command 'startx' from deb xinit (1.4.1-0ubuntu2)

See 'snap info <snapname>' for additional versions.

```
hadoop@admin1-HP-280-G4-MT-Business-PC:/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]
localhost: namenode is running as process 29392. Stop it first.
Starting datanodes
localhost: datanode is running as process 29540. Stop it first.
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
admin1-HP-280-G4-MT-Business-PC: secondarynamenode is running as process
29778. Stop it first.
Starting resourcemanager
resourcemanager is running as process 29965. Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 30115. Stop it first.
hadoop@admin1-HP-280-G4-MT-Business-PC:/sbin$ hdfs dfs -mkdir -p ~/input
hadoop@admin1-HP-280-G4-MT-Business-PC:/sbin$ hdfs dfs -appendToFile -
~/input/test.txt
hello world
world
shrey hello
2022-06-17 12:34:10,938 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localhostTrusted = false, remoteHostTrusted = false
```

```
hadoop@admin1-HP-280-G4-MT-Business-PC:/sbin$ hadoop jar /home/hadoop/shrey.jar  
~/input ~/out
```

```
Exception in thread "main" java.lang.ClassNotFoundException: /home/hadoop/input  
  at java.lang.Class.forName0(Native Method)  
  at java.lang.Class.forName(Class.java:348)  
  at org.apache.hadoop.util.RunJar.run(RunJar.java:316)  
  at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
```

```
hadoop@admin1-HP-280-G4-MT-Business-PC:/sbin$ hadoop jar  
/home/hadoop/shrey.jar ~/input ~/out
```

```
2022-06-17 12:37:34,125 INFO client.RMProxy: Connecting to ResourceManager at  
/127.0.0.1:8032
```

```
2022-06-17 12:37:34,419 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 12:37:34,435 INFO mapreduce.JobResourceUploader: Disabling Erasure  
Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1655449081065_0001
```

```
2022-06-17 12:37:34,524 INFO sasl.SaslDataTransferClient: SASL encryption trust  
check: localhostTrusted = false, remoteHostTrusted = false
```

```
2022-06-17 12:37:34,622 INFO input.FileInputFormat: Total input files to process : 1
```

```
2022-06-17 12:37:34,695 INFO sasl.SaslDataTransferClient: SASL encryption trust  
check: localhostTrusted = false, remoteHostTrusted = false
```

```
2022-06-17 12:37:34,761 INFO sasl.SaslDataTransferClient: SASL encryption trust  
check: localhostTrusted = false, remoteHostTrusted = false
```

```
2022-06-17 12:37:34,785 INFO mapreduce.JobSubmitter: number of splits:1
```

```
2022-06-17 12:37:34,901 INFO sasl.SaslDataTransferClient: SASL encryption trust  
check: localhostTrusted = false, remoteHostTrusted = false
```

```
2022-06-17 12:37:35,319 INFO mapreduce.JobSubmitter: Submitting tokens for job:  
job_1655449081065_0001
```

```
2022-06-17 12:37:35,319 INFO mapreduce.JobSubmitter: Executing with tokens: []
```

```
2022-06-17 12:37:35,464 INFO conf.Configuration: resource-types.xml not found
```

```
2022-06-17 12:37:35,464 INFO resource.ResourceUtils: Unable to find  
'resource-types.xml'.
```

```
2022-06-17 12:37:35,635 INFO impl.YarnClientImpl: Submitted application  
application_1655449081065_0001
```

```
2022-06-17 12:37:35,663 INFO mapreduce.Job: The url to track the job:
```

```
http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1655449081065_00  
01/
```

```
2022-06-17 12:37:35,664 INFO mapreduce.Job: Running job:  
job_1655449081065_0001
```

2022-06-17 12:37:40,732 INFO mapreduce.Job: Job job_1655449081065_0001
running in uber mode : false

2022-06-17 12:37:40,734 INFO mapreduce.Job: map 0% reduce 0%

2022-06-17 12:37:44,806 INFO mapreduce.Job: map 100% reduce 0%

2022-06-17 12:37:48,835 INFO mapreduce.Job: map 100% reduce 100%

2022-06-17 12:37:49,854 INFO mapreduce.Job: Job job_1655449081065_0001
completed successfully

2022-06-17 12:37:49,911 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=69

FILE: Number of bytes written=451021

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=175

HDFS: Number of bytes written=39

HDFS: Number of read operations=8

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=1

Launched reduce tasks=1

Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=1404

Total time spent by all reduces in occupied slots (ms)=1515

Total time spent by all map tasks (ms)=1404

Total time spent by all reduce tasks (ms)=1515

Total vcore-milliseconds taken by all map tasks=1404

Total vcore-milliseconds taken by all reduce tasks=1515

Total megabyte-milliseconds taken by all map tasks=1437696

Total megabyte-milliseconds taken by all reduce tasks=1551360

Map-Reduce Framework

Map input records=5

Map output records=12

Map output bytes=111

Map output materialized bytes=69

Input split bytes=112

Combine input records=12

Combine output records=6

Reduce input groups=6
Reduce shuffle bytes=69
Reduce input records=6
Reduce output records=6
Spilled Records=12
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=66
CPU time spent (ms)=690
Physical memory (bytes) snapshot=490770432
Virtual memory (bytes) snapshot=5081649152
Total committed heap usage (bytes)=517472256
Peak Map Physical memory (bytes)=298610688
Peak Map Virtual memory (bytes)=2537594880
Peak Reduce Physical memory (bytes)=192159744
Peak Reduce Virtual memory (bytes)=2544054272

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=63

File Output Format Counters

Bytes Written=39

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

2022-06-17 12:38:29,847 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localhostTrusted = false, remoteHostTrusted = false

hello 4

how 1

r 1

shrey 2

u? 1

world 3