

Vinayak P Bhupal

INT19IS187

BIG DATA EXP

```
import java.io.IOException;
import java.util.*;

import org.apache.hadoop.fs.Path;
import org.apache.hadoop.conf.*;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapred.*;
import org.apache.hadoop.util.*;

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);
    }
}
```

cd \$HADOOP_HOME

cd sbin

jps

start-all.sh

```
hadoop admin1-HP-280-G4-MT-Business-PC ~ cd $HADOOP_HOME
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1 cd sbin
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin jps
4144 NameNode
4721 ResourceManager
12932 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
4533 SecondaryNameNode
13334 Jps
4877 NodeManager
4287 DataNode
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]
localhost: namenode is running as process 4144. Stop it first.
Starting datanodes
localhost: datanode is running as process 4287. 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 4533. Stop it first.
Starting resourcemanager
resourcemanager is running as process 4721. Stop it first.
```

hdfs dfs- mkdir -p ~ /input

hdfs dfs -appendToFile - ~/input/txt.txt

This is me this is you this is all we need

```
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]
localhost: namenode is running as process 4144. Stop it first.
Starting datanodes
localhost: datanode is running as process 4287. 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 4533. Stop it first.
Starting resourcemanager
resourcemanager is running as process 4721. Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 4877. Stop it first.
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -mkdir -p ~/input
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -appendToFile - ~/input/test.txt
this is you this me this all we need 2022-06-18 12:33:39,665 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localH
ostTrusted = false, remoteHostTrusted = false
```

```

2022-06-18 12:35:45,942 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-18 12:35:46,074 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-18 12:35:46,223 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool i
interface and execute your application with ToolRunner to remedy this.
2022-06-18 12:35:46,239 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.sta
ging/job_1655528231040_0001
2022-06-18 12:35:46,328 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted
= false
2022-06-18 12:35:46,427 INFO mapred.FileInputFormat: Total input files to process : 1
2022-06-18 12:35:46,473 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted
= false
2022-06-18 12:35:46,531 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted
= false
2022-06-18 12:35:46,555 INFO mapreduce.JobSubmitter: number of splits:2
2022-06-18 12:35:46,671 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted
= false
2022-06-18 12:35:46,680 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1655528231040_0001
2022-06-18 12:35:46,680 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-06-18 12:35:46,822 INFO conf.Configuration: resource-types.xml not found
2022-06-18 12:35:46,822 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-06-18 12:35:46,974 INFO impl.YarnClientImpl: Submitted application application_1655528231040_0001
2022-06-18 12:35:47,013 INFO mapreduce.Job: The url to track the job: http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/applicatio
n_1655528231040_0001/
2022-06-18 12:35:47,014 INFO mapreduce.Job: Running job: job_1655528231040_0001
2022-06-18 12:35:53,164 INFO mapreduce.Job: Job job_1655528231040_0001 running in uber mode : false
2022-06-18 12:35:53,166 INFO mapreduce.Job: map 0% reduce 0%
2022-06-18 12:35:57,220 INFO mapreduce.Job: map 100% reduce 0%
2022-06-18 12:36:00,251 INFO mapreduce.Job: map 100% reduce 100%
2022-06-18 12:36:01,280 INFO mapreduce.Job: Job job_1655528231040_0001 completed successfully
2022-06-18 12:36:01,271 INFO mapreduce.Job: Counters: 54

```

```

hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -mkdir -p ~/input123/testing.txt
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -mkdir -p ~/vinayak
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -appendToFile ~/vinayak/file1.txt
blah blah blah okaay okay 2022-06-18 13:17:40,387 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/vinayak.jar ~/vinayak ~/out
2022-06-18 13:18:25,321 INFO client.RMPProxy: Connecting to ResourceManager at /127.0.0.1:8032
2022-06-18 13:18:25,446 INFO client.RMPProxy: 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/home/hadoop/
out already exists
    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.mapreduce.security.UserGroupInformation.doAs(UserGroupInformation.java:1730)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1567)
    at org.apache.hadoop.mapreduce.JobClient$1.run(JobClient.java:576)
    at org.apache.hadoop.mapreduce.JobClient$1.run(JobClient.java:571)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:422)
    at org.apache.hadoop.mapreduce.security.UserGroupInformation.doAs(UserGroupInformation.java:1730)
    at org.apache.hadoop.mapreduce.JobClient.submitJobInternal(JobClient.java:571)
    at org.apache.hadoop.mapreduce.JobClient.submitJob(JobClient.java:562)
    at org.apache.hadoop.mapreduce.JobClient.runJob(JobClient.java:873)
    at Sanjeev.WordCount.main(WordCount.java:55)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:323)

```

```

Bytes Written=41
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -cat ~/sanjeev/part*
cat: '/home/hadoop/sanjeev/part*': No such file or directory
hadoop admin1-HP-280-G4-MT-Business-PC ~/hadoop-3.2.1/sbin hdfs dfs -cat ~/out/part*
2022-06-18 12:37:23,789 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted
= false
all 1
is 1
ne 1
need 1
this 3
ve 1
you 1

```