

Task 01

```
root@07c2f64b580a:/# cd $HADOOP_HOME/etc/hadoop
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# nano core-site.xml
bash: nano: command not found
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# cat core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "<property>
>   <name>fs.defaultFS</name>
>   <value>hdfs://localhost:9000</value>
> </property>" >> core-site.xml
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# cat hdfs-site.xml  
<?xml version="1.0" encoding="UTF-8"?>  
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
<!--  
    Licensed under the Apache License, Version 2.0 (the "License");  
    you may not use this file except in compliance with the License.  
    You may obtain a copy of the License at  
  
        http://www.apache.org/licenses/LICENSE-2.0  
  
    Unless required by applicable law or agreed to in writing, software  
    distributed under the License is distributed on an "AS IS" BASIS,  
    WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
    See the License for the specific language governing permissions and  
    limitations under the License. See accompanying LICENSE file.  
-->  
    <!-- Put site-specific property overrides in this file -->
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "<property>  
>   <name>dfs.replication</name>  
>   <value>2</value>  
> </property>  
> <property>  
>   <name>dfs.blocksize</name>  
>   <value>134217728</value> <!-- 128 MB -->  
bash: !-: event not found  
> </property>" >> hdfs-site.xml  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# jps  
707 ResourceManager  
533 NodeManager  
1910 Jps  
393 DataNode  
239 NameNode  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "export HADOOP_HOME=/opt/hadoop  
-3.2.1" >> ~/.bashrc  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "export PATH=\$PATH:\$HADOOP_HO  
ME/bin:\$HADOOP_HOME/sbin" >> ~/.bashrc  
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```

root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "export HADOOP_HOME=/opt/hadoop-3.2.1" >> ~/.bashrc
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "export PATH=\$PATH:\$HADOOP_HOME/bin:\$HADOOP_HOME/sbin" >> ~/.bashrc
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# source ~/.bashrc
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo $PATH
/opt/hadoop-3.2.1/bin:/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/opt/hadoop-3.2.1/bin:/opt/hadoop-3.2.1/sbin
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# hadoop-daemon.sh stop namenode
WARNING: Use of this script to stop HDFS daemons is deprecated.
WARNING: Attempting to execute replacement "hdfs --daemon stop" instead.
2024-11-18 09:07:35,787 ERROR namenode.NameNode: RECEIVED SIGNAL 15: SIGTERM
2024-11-18 09:07:36,046 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****

```

```

root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# hadoop-daemon.sh stop datanode
WARNING: Use of this script to stop HDFS daemons is deprecated.
WARNING: Attempting to execute replacement "hdfs --daemon stop" instead.
2024-11-18 09:07:37,051 ERROR datanode.DataNode: RECEIVED SIGNAL 15: SIGTERM
2024-11-18 09:07:37,060 INFO datanode.DataNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down DataNode at 07c2f64b580a/172.17.0.2
*****/

```

```

y state
2024-11-18 09:07:38,553 INFO resourcemanager.ResourceManager: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down ResourceManager at 07c2f64b580a/172.17.0.2
*****/
2024-11-18 09:07:38,575 ERROR conf.Configuration: error parsing conf core-site.xml
com.ctc.wstx.exc.WstxParsingException: Illegal to have multiple roots (start tag in
epilog?).
  at [row,col,system-id]: [22,2,"file:/opt/hadoop-3.2.1/etc/hadoop/core-site.xml"]
  at com.ctc.wstx.sr.StreamScanner.constructWfcException(StreamScanner.java:62

```

```

root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# hadoop-daemon.sh start namenode
WARNING: Use of this script to start HDFS daemons is deprecated.
WARNING: Attempting to execute replacement "hdfs --daemon start" instead.
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# hadoop-daemon.sh start datanode
WARNING: Use of this script to start HDFS daemons is deprecated.
WARNING: Attempting to execute replacement "hdfs --daemon start" instead.
ERROR: Cannot set priority of datanode process 2520
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# yarn-daemon.sh start resourcemanager
WARNING: Use of this script to start YARN daemons is deprecated.
WARNING: Attempting to execute replacement "yarn --daemon start" instead.
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# yarn-daemon.sh start nodemanager
WARNING: Use of this script to start YARN daemons is deprecated.
WARNING: Attempting to execute replacement "yarn --daemon start" instead.
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#

```

```
2024-11-18 10:56:27,379 INFO mapred.Task: Final Counters for attempt_local1430197592_0001_m_000000_0: Counters: 24
```

File System Counters

```
FILE: Number of bytes read=317864
FILE: Number of bytes written=848358
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=11392
HDFS: Number of bytes written=0
HDFS: Number of read operations=5
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
```

Map-Reduce Framework

```
Map input records=265
Map output records=1310
Map output bytes=15783
Map output materialized bytes=5170
Input split bytes=125
Combine input records=1310
Combine output records=256
Spilled Records=256
```

```
HDFS: Number of bytes written=0
HDFS: Number of read operations=7
HDFS: Number of large read operations=0
HDFS: Number of write operations=1
HDFS: Number of bytes read erasure-coded=0
```

Map-Reduce Framework

```
Map input records=220
Map output records=832
Map output bytes=10690
Map output materialized bytes=6333
Input split bytes=130
Combine input records=832
Combine output records=356
Spilled Records=356
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=0
Total committed heap usage (bytes)=391118848
```

File Input Format Counters

```
Bytes Read=8260
```

```
2024-11-18 10:56:27,468 INFO mapred.LocalJobRunner: Finishing task: attempt_local1430197592_0001_m_000001_0
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# echo "<property>
```

```
> <name>fs.defaultFS</name>
```

```
> <value>hdfs://localhost:9000</value>
```

```
> </property>" >> core-site.xml
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop#
```

```
File System Counters
  FILE: Number of bytes read=3264130
  FILE: Number of bytes written=8618331
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=243235
  HDFS: Number of bytes written=10327
  HDFS: Number of read operations=143
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=12
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=751
  Map output records=3063
  Map output bytes=38519
  Map output materialized bytes=21327
  Input split bytes=1104
  Combine input records=3063
  Combine output records=1223
  Reduce input groups=588
  Reduce shuffle bytes=21327
```

```
  Reduce input groups=588
  Reduce shuffle bytes=21327
  Reduce input records=1223
  Reduce output records=588
  Spilled Records=2446
  Shuffled Maps =9
  Failed Shuffles=0
  Merged Map outputs=9
  GC time elapsed (ms)=9
  Total committed heap usage (bytes)=7580155904
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=28589
File Output Format Counters
  Bytes Written=10327
@07c2f64b580a:/#
```

```
root@07c2f64b580a:/opt/hadoop-3.2.1/etc/hadoop# stop-all.sh
bash: stop-all.sh: command not found
```

Task 02

```
root@07c2f64b580a:/# rm CustomWordCount.java
root@07c2f64b580a:/# cat > CustomWordCount.java <<EOF
> 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;
>
> import java.io.IOException;
>
> public class CustomWordCount {
>     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 {
>             String[] tokens = value.toString().split("\\s+");
>             for (String token : tokens) {
```

```
>                 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(CustomWordCount.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);
>         }
>     }
> }
```

```
root@07c2f64b580a:/# ls -l CustomWordCount.java
-rw-r--r-- 1 root root 2123 Nov 18 11:18 CustomWordCount.java
root@07c2f64b580a:/# javac -cp $(hadoop classpath) -d . CustomWordCount.java
root@07c2f64b580a:/# jar cf CustomWordCount.jar CustomWordCount/*.class
CustomWordCount/*.class : no such file or directory
root@07c2f64b580a:/# hadoop jar CustomWordCount.jar CustomWordCount <input_path> <output_path>
bash: syntax error near unexpected token `<'
root@07c2f64b580a:/# hdfs dfs -mkdir /user/hadoop/custom_input
mkdir: '/user/hadoop/custom_input': File exists
root@07c2f64b580a:/# hdfs dfs -put /path/to/input_data.txt /user/hadoop/custom_input
put: '/path/to/input_data.txt': No such file or directory
root@07c2f64b580a:/# hadoop jar CustomWordCount.jar CustomWordCount /user/hadoop/custom_input /user/hadoop/custom_output
2024-11-18 11:20:53,219 INFO impl.MetricsConfig: Loaded properties from hadoop-metrics2.properties
2024-11-18 11:20:53,332 INFO impl.MetricsSystemImpl: Scheduled Metric snapshot period at 10 second(s).
2024-11-18 11:20:53,333 INFO impl.MetricsSystemImpl: JobTracker metrics system started
2024-11-18 11:20:53,538 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2024-11-18 11:20:53,544 WARN mapreduce.JobResourceUploader: No job jar file set. User classes may not be found. See Job or Job#setJar(String)
```

```
File System Counters
  FILE: Number of bytes read=186
  FILE: Number of bytes written=522088
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=40
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=5
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=1
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=1
  Map output records=6
  Map output bytes=64
  Map output materialized bytes=82
  Input split bytes=129
  Combine input records=6
  Combine output records=6
  Spilled Records=6
  Failed Shuffles=0
```

```
Reduce input groups=6
Reduce shuffle bytes=82
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)=21
Total committed heap usage (bytes)=657981440
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=40
File Output Format Counters
  Bytes Written=52
```