

Aayush Shah

19BCE245

7 October 2022

Big Data Analytics

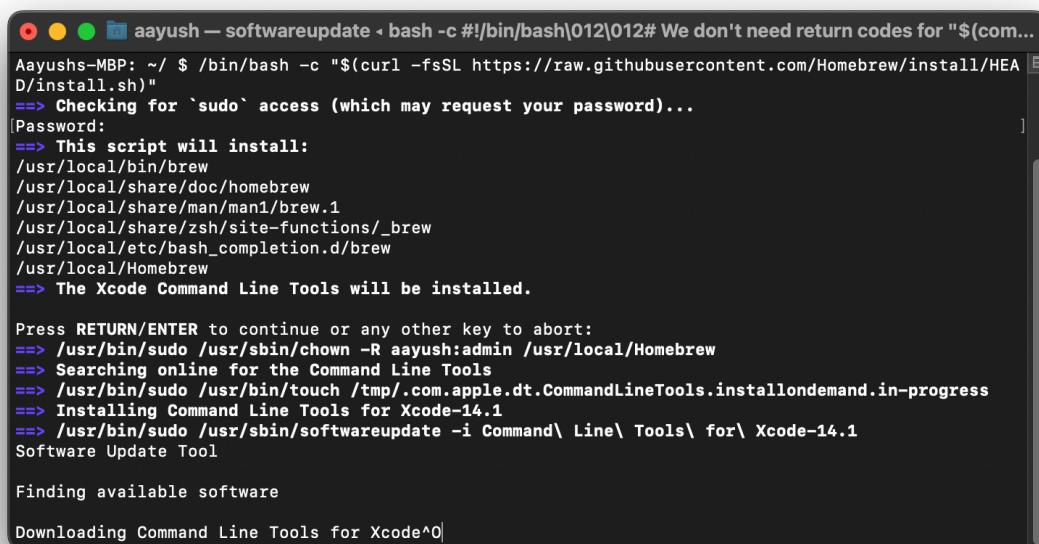
Practical 3

Aim

Setup single node Hadoop cluster and apply HDFS commands on single node Hadoop Cluster.

Hadoop installation on Mac

1. Installed brew via `/bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"`



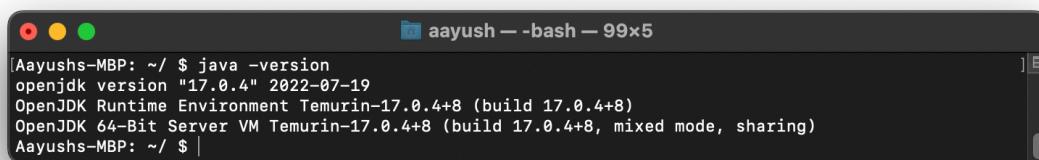
```
aayush — softwareupdate ✘ bash -c #!/bin/bash|012|012# We don't need return codes for "$(com...
Aayushs-MBP: ~ $ /bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEA D/install.sh)"
==> Checking for `sudo` access (which may request your password)...
[Password:
==> This script will install:
/usr/local/bin/brew
/usr/local/share/doc/homebrew
/usr/local/share/man/man1/brew.1
/usr/local/share/zsh/site-functions/_brew
/usr/local/etc/bash_completion.d/brew
/usr/local/Homebrew
==> The Xcode Command Line Tools will be installed.

Press RETURN/ENTER to continue or any other key to abort:
==> /usr/bin/sudo /usr/sbin/chown -R aayush:admin /usr/local/Homebrew
==> Searching online for the Command Line Tools
==> /usr/bin/sudo /usr/bin/touch /tmp/.com.apple.dt.CommandLineTools.installondemand.in-progress
==> Installing Command Line Tools for Xcode-14.1
==> /usr/bin/sudo /usr/sbin/softwareupdate -i Command\ Line\ Tools\ for\ Xcode-14.1
Software Update Tool

Finding available software

Downloading Command Line Tools for Xcode^0|
```

2. Check java version with `java -version`



```
aayush — -bash — 99x5
Aayushs-MBP: ~ $ java -version
openjdk version "17.0.4" 2022-07-19
OpenJDK Runtime Environment Temurin-17.0.4+8 (build 17.0.4+8)
OpenJDK 64-Bit Server VM Temurin-17.0.4+8 (build 17.0.4+8, mixed mode, sharing)
Aayushs-MBP: ~/ $ |
```

3. Installed older java version with *brew cask install java8*

4. Installed Hadoop with *Brew install hadoop*

```
aayush — git-remote-https ✘ brew.sh install hadoop — 99x5
OpenJDK Runtime Environment Temurin-17.0.4+8 (build 17.0.4+8)
OpenJDK 64-Bit Server VM Temurin-17.0.4+8 (build 17.0.4+8, mixed mode, sharing)
[Aayushs-MBP: ~] $ brew install hadoop
Running `brew update --auto-update`...
```

5. Configuration Changes in various files :

```
Shell Script
Language Run Stop Run Settings... Back/Forward View
practical3_own.py hadoop-installatio...ninal-snapshot.txt hadoop-env.sh +
```

```
51
52 # The java implementation to use. By default, this environment
53 # variable is REQUIRED on ALL platforms except OS X!
54 # export JAVA_HOME=
55
56 export JAVA_HOME="/Library/Java/JavaVirtualMachines/
  adoptopenjdk-8.jdk/Contents/Home"
57
58
59 # Location of Hadoop. By default, Hadoop will attempt to determine
60 # this location based upon its execution path.
61 # export HADOOP_HOME=
```

```
Syntax: XML
Language Run Stop Run Settings... Back/Forward View
practical3_own.py hadoop-inst...-snapshot.txt hadoop-env.sh core-site.xml +
```

```
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. -->

<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

```
Syntax: XML
Language Run Stop Run Settings... Back/Forward View
practical3_own.py hadoop-i...pshot.txt hadoop-env.sh core-site.xml mapred-site.xml +
```

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.application.classpath</name>
    <value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:
      $HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
  </property>
</configuration>
```

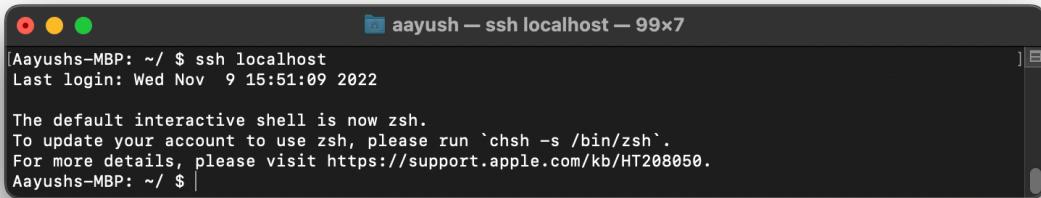
```
Syntax: XML
Language Run Stop Run Settings... Back/Forward View
practical3_own.py hadoop-i...pshot.txt hadoop-env.sh core-site.xml mapred-site.xml hdfs-site.xml +
```

```
limitations under the License. See accompanying LICENSE file.

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

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

6. Configure SSH with *ssh localhost*



```
[Aayushs-MBP: ~/ $ ssh localhost
Last login: Wed Nov  9 15:51:09 2022

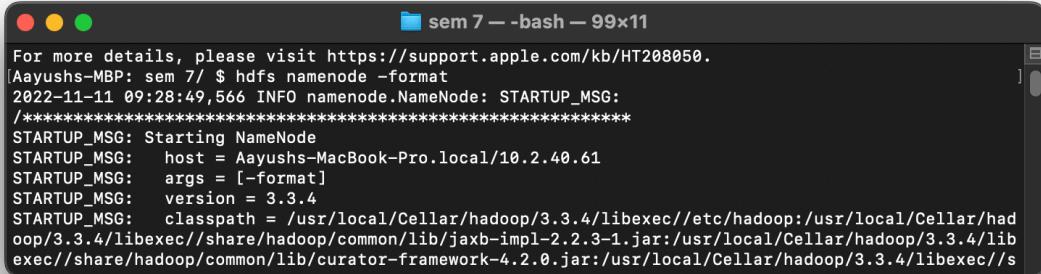
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
Aayushs-MBP: ~/ $ |
```

7. Configure SSH

```
ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 0600 ~/.ssh/authorized_keys
```

8. Format Namenode

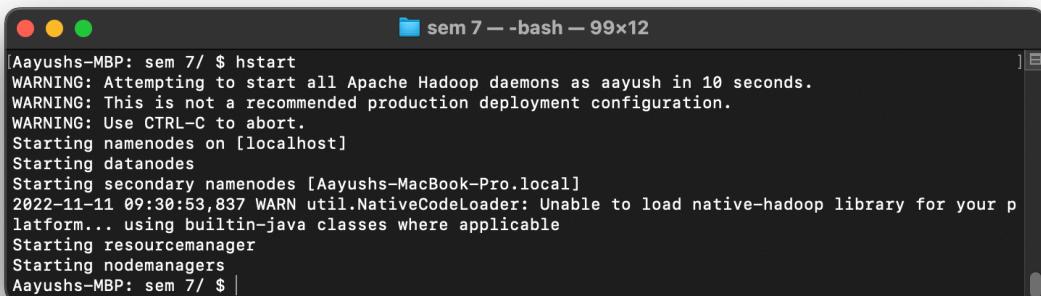
```
cd /usr/local/cellar/hadoop/3.2.1/libexec/bin
hdfs namenode -format
```



```
For more details, please visit https://support.apple.com/kb/HT208050.
[Aayushs-MBP: sem 7/ $ hdfs namenode -format
2022-11-11 09:28:49,566 INFO namenode.NameNode: STARTUP_MSG:
*****STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = Aayushs-MacBook-Pro.local/10.2.40.61
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.3.4
STARTUP_MSG: classpath = /usr/local/Cellar/hadoop/3.3.4/libexec//etc/hadoop:/usr/local/Cellar/hadoop/3.3.4/libexec//share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/usr/local/Cellar/hadoop/3.3.4/libexec//share/hadoop/common/lib/curator-framework-4.2.0.jar:/usr/local/Cellar/hadoop/3.3.4/libexec//s
```

9. Run Hadoop

```
cd /usr/local/cellar/hadoop/3.2.1/libexec/sbin
./start-all.sh
jps
```



```
[Aayushs-MBP: sem 7/ $ hstart
WARNING: Attempting to start all Apache Hadoop daemons as aayush 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 [Aayushs-MacBook-Pro.local]
2022-11-11 09:30:53,837 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers
Aayushs-MBP: sem 7/ $ |
```

10. WordCount executing

```

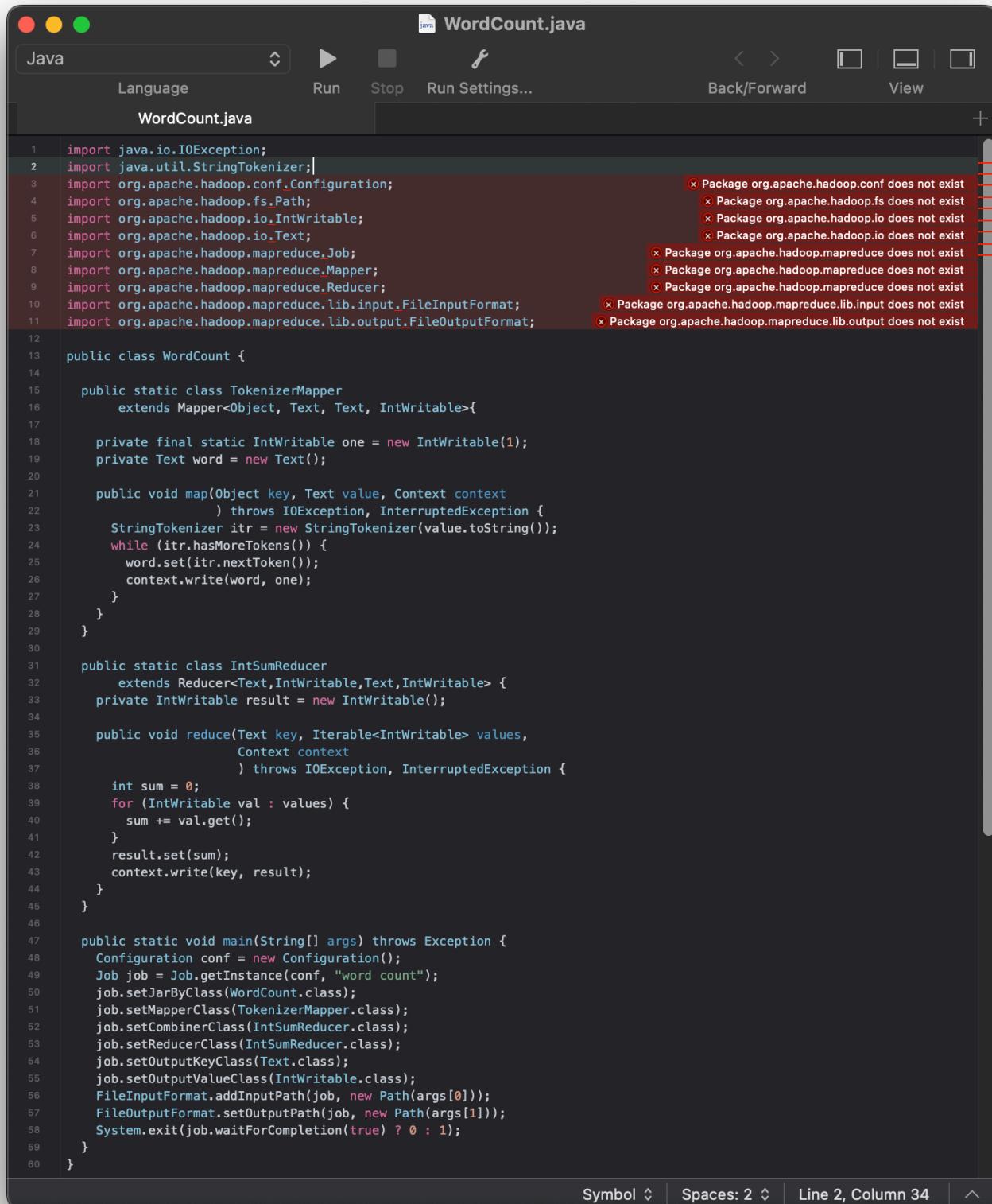
Practical 3 -- bash -- 99x7
[Aayushs-MBP: Practical 3/ $ hadoop com.sun.tools.javac.Main WordCount.java
[Aayushs-MBP: Practical 3/ $ jar cf wc.jar WordCount*.class
[Aayushs-MBP: Practical 3/ $ ls
README.md           WordCount.class          sample1.txt
README.md.tmp.html WordCount.java          snapshot.md
WordCount$IntSumReducer.class first          wc.jar
WordCount$TokenizerMapper.class input

Practical 3 -- bash -- 99x60
[Aayushs-MBP: Practical 3/ $ hadoop fs -put input /prac3
2022-11-11 09:53:56,663 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
[Aayushs-MBP: Practical 3/ $ hadoop fs -ls /prac3
2022-11-11 09:54:00,899 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
Found 4 items
-rw-r--r--  1 aayush supergroup      6148 2022-11-11 09:53 /prac3/.DS_Store
-rw-r--r--  1 aayush supergroup       22 2022-11-11 09:53 /prac3/file01.txt
-rw-r--r--  1 aayush supergroup       40 2022-11-11 09:53 /prac3/file02.txt
drwxr-xr-x - aayush supergroup        0 2022-11-11 09:53 /prac3/input
Aayushs-MBP: Practical 3/ $ hadoop fs -rm /prac3/input/.DS_Store

2022-11-11 09:54:15,281 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
Deleted /prac3/input/.DS_Store
Aayushs-MBP: Practical 3/ $
[Aayushs-MBP: Practical 3/ $ hadoop fs -ls /prac3/input
2022-11-11 09:54:24,108 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 aayush supergroup      22 2022-11-11 09:53 /prac3/input/file01.txt
-rw-r--r--  1 aayush supergroup      40 2022-11-11 09:53 /prac3/input/file02.txt
[Aayushs-MBP: Practical 3/ $ hadoop fs -put file01.txt /prac3/input
2022-11-11 09:54:45,591 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
put: `file01.txt': No such file or directory
[Aayushs-MBP: Practical 3/ $ cd input
[Aayushs-MBP: input/ $ hadoop fs -put file02.txt /prac3/input
2022-11-11 09:54:59,715 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
put: `/prac3/input/file02.txt': File exists
[Aayushs-MBP: input/ $ hadoop fs -ls /prac3/input
2022-11-11 09:55:08,292 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 aayush supergroup      22 2022-11-11 09:53 /prac3/input/file01.txt
-rw-r--r--  1 aayush supergroup      40 2022-11-11 09:53 /prac3/input/file02.txt
[Aayushs-MBP: input/ $ cd ..
[Aayushs-MBP: Practical 3/ $ hadoop jar wc.jar WordCount prac3/input prac3/output
2022-11-11 09:55:30,121 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
2022-11-11 09:55:30,676 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-11 09:55:31,281 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not
performed. Implement the Tool interface and execute your application with ToolRunner to remedy this
.
2022-11-11 09:55:31,293 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp
/hadoop-yarn/staging/aayush/.staging/job_1668139259002_0001
2022-11-11 09:55:31,490 INFO mapreduce.JobSubmitter: Cleaning up the staging area /tmp/hadoop-yarn/
staging/aayush/.staging/job_1668139259002_0001
Exception in thread "main" org.apache.hadoop.mapreduce.lib.input.InvalidInputException: Input path
does not exist: hdfs://localhost:9000/user/aayush/prac3/input
        at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.singleThreadedListStatus(FileInput
Format.java:340)
        at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.listStatus(FileInputFormat.java:27
9)
        at org.apache.hadoop.mapreduce.lib.input.FileInputFormat.getSplits(FileInputFormat.java:404
)
        at org.apache.hadoop.mapreduce.JobSubmitter.writeNewSplits(JobSubmitter.java:310)

```

11. WordCount.java file :



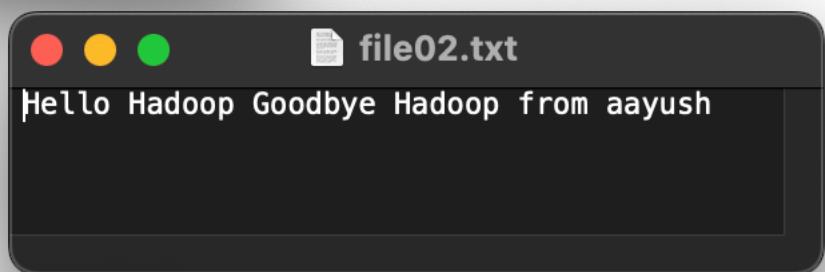
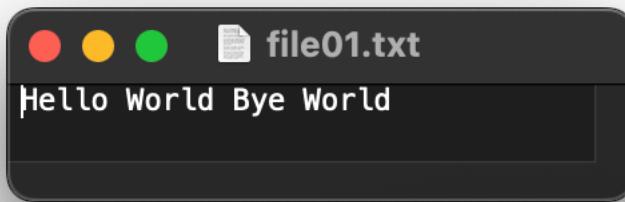
```

1 import java.io.IOException;
2 import java.util.StringTokenizer;
3 import org.apache.hadoop.conf.Configuration;
4 import org.apache.hadoop.fs.Path;
5 import org.apache.hadoop.io.IntWritable;
6 import org.apache.hadoop.io.Text;
7 import org.apache.hadoop.mapreduce.Job;
8 import org.apache.hadoop.mapreduce.Mapper;
9 import org.apache.hadoop.mapreduce.Reducer;
10 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
11 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
12
13 public class WordCount {
14
15     public static class TokenizerMapper
16         extends Mapper<Object, Text, Text, IntWritable>{
17
18         private final static IntWritable one = new IntWritable(1);
19         private Text word = new Text();
20
21         public void map(Object key, Text value, Context context
22                         ) throws IOException, InterruptedException {
23             StringTokenizer itr = new StringTokenizer(value.toString());
24             while (itr.hasMoreTokens()) {
25                 word.set(itr.nextToken());
26                 context.write(word, one);
27             }
28         }
29     }
30
31     public static class IntSumReducer
32         extends Reducer<Text,IntWritable,Text,IntWritable> {
33         private IntWritable result = new IntWritable();
34
35         public void reduce(Text key, Iterable<IntWritable> values,
36                           Context context
37                           ) throws IOException, InterruptedException {
38             int sum = 0;
39             for (IntWritable val : values) {
40                 sum += val.get();
41             }
42             result.set(sum);
43             context.write(key, result);
44         }
45     }
46
47     public static void main(String[] args) throws Exception {
48         Configuration conf = new Configuration();
49         Job job = Job.getInstance(conf, "word count");
50         job.setJarByClass(WordCount.class);
51         job.setMapperClass(TokenizerMapper.class);
52         job.setCombinerClass(IntSumReducer.class);
53         job.setReducerClass(IntSumReducer.class);
54         job.setOutputKeyClass(Text.class);
55         job.setOutputValueClass(IntWritable.class);
56         FileInputFormat.addInputPath(job, new Path(args[0]));
57         FileOutputFormat.setOutputPath(job, new Path(args[1]));
58         System.exit(job.waitForCompletion(true) ? 0 : 1);
59     }
60 }

```

The code is annotated with numerous red error markers pointing to package names that do not exist: org.apache.hadoop.conf, org.apache.hadoop.fs, org.apache.hadoop.io, org.apache.hadoop.mapreduce, org.apache.hadoop.mapreduce.lib.input, and org.apache.hadoop.mapreduce.lib.output.

12. Given input files and generated output file



part-r-00000	
Bye	1
Goodbye	1
Hadoop	2
Hello	2
World	2
aayush	1
from	1

Conclusion

From this practical, we learned the installation of Hadoop and performed basic WordCount program through which we learned the simple Hadoop commands like ls, putting file inside Hadoop file system, executing java programs.