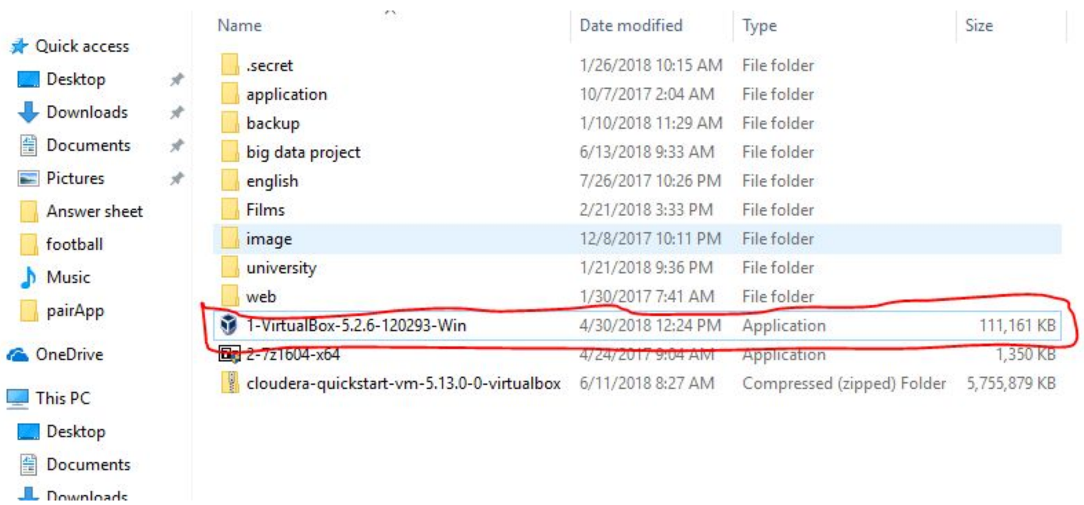
Documentation

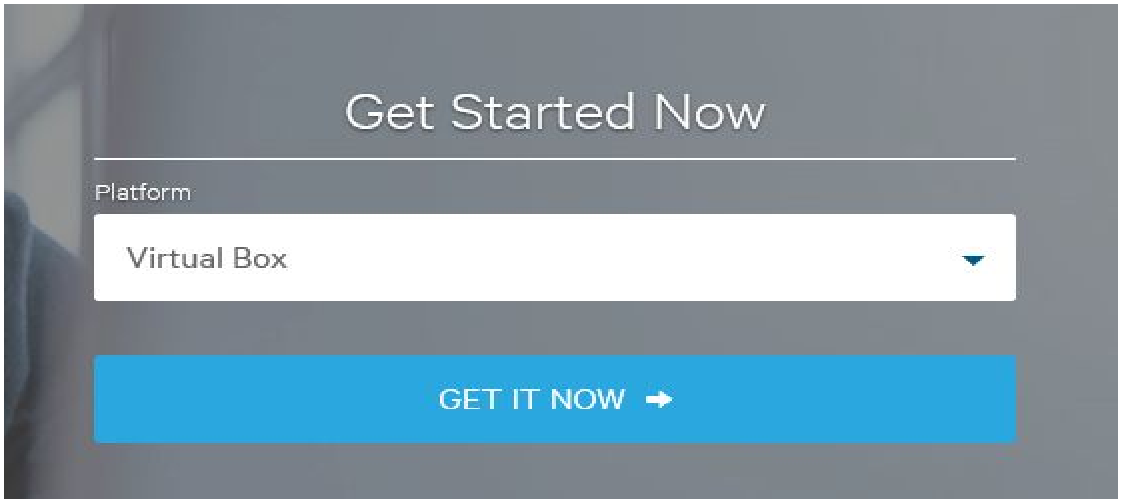
First thing to do is to download VirtualBox.
From the link provided below we can download the file.
http://www.oracle.com/technetwork/server-storage/virtualbox/downloads/index.html?ssSourceSiteId=otnjp



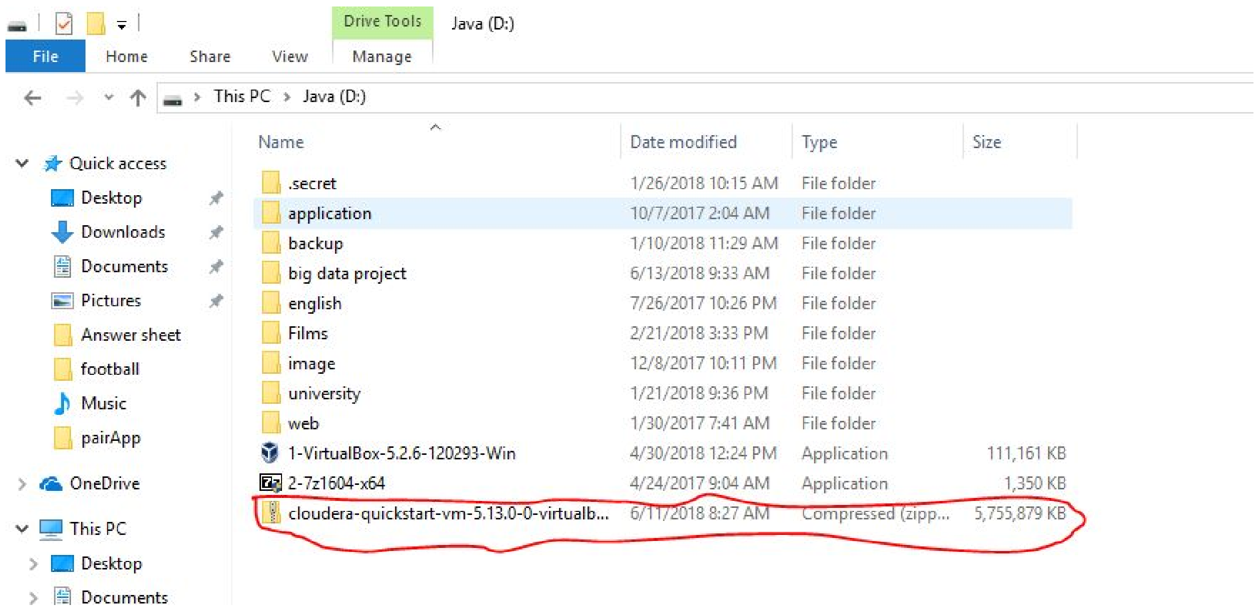
Quick access
Desktop
Downloads
Documents
Pictures
Answer sheet
football
Music
pairApp
OneDrive
This PC
Desktop
Documents
Downloads

Name	Date modified	Type	Size
.secret	1/26/2018 10:15 AM	File folder	
application	10/7/2017 2:04 AM	File folder	
backup	1/10/2018 11:29 AM	File folder	
big data project	6/13/2018 9:33 AM	File folder	
english	7/26/2017 10:26 PM	File folder	
Films	2/21/2018 3:33 PM	File folder	
image	12/8/2017 10:11 PM	File folder	
university	1/21/2018 9:36 PM	File folder	
web	1/30/2017 7:41 AM	File folder	
1-VirtualBox-5.2.6-120293-Win	4/30/2018 12:24 PM	Application	111,161 KB
2-7z1604-x64	4/24/2017 9:04 AM	Application	1,350 KB
cloudera-quickstart-vm-5.13.0-0-virtualbox	6/11/2018 8:27 AM	Compressed (zipped) Folder	5,755,879 KB

After download VirtualBox file we also need to download cloudera-quickstart-vm-5.13.0-0-virtualbox file. Here is a link provided to get it.
https://www.cloudera.com/downloads/quickstart_vms/5-13.html

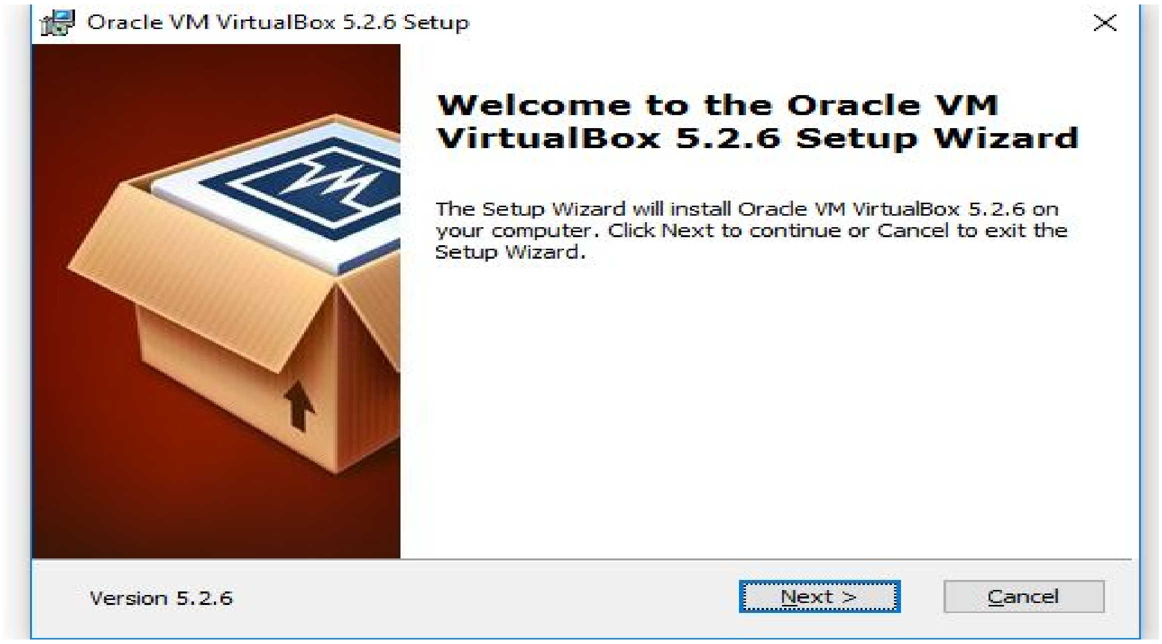


Platform
Virtual Box



Name	Date modified	Type	Size
.secret	1/26/2018 10:15 AM	File folder	
application	10/7/2017 2:04 AM	File folder	
backup	1/10/2018 11:29 AM	File folder	
big data project	6/13/2018 9:33 AM	File folder	
english	7/26/2017 10:26 PM	File folder	
Films	2/21/2018 3:33 PM	File folder	
image	12/8/2017 10:11 PM	File folder	
university	1/21/2018 9:36 PM	File folder	
web	1/30/2017 7:41 AM	File folder	
1-VirtualBox-5.2.6-120293-Win	4/30/2018 12:24 PM	Application	111,161 KB
2-7z1604-x64	4/24/2017 9:04 AM	Application	1,350 KB
cloudera-quickstart-vm-5.13.0-0-virtualb...	6/11/2018 8:27 AM	Compressed (zipp...	5,755,879 KB

After having these two files we can start installing VirutalBox by showing a path of Cloudera file.

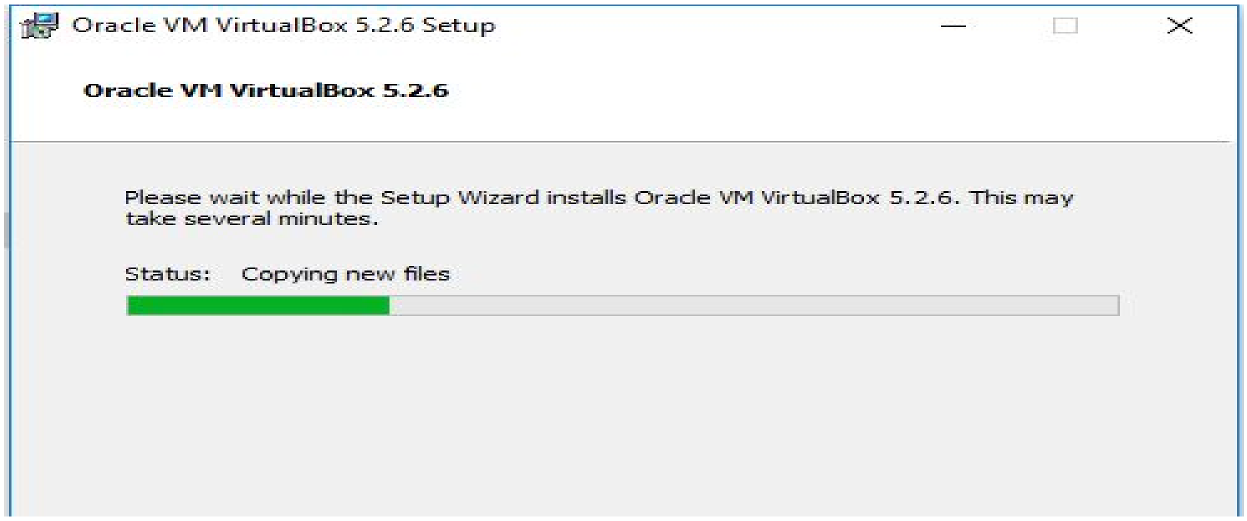


Oracle VM VirtualBox 5.2.6 Setup

Welcome to the Oracle VM VirtualBox 5.2.6 Setup Wizard

The Setup Wizard will install Oracle VM VirtualBox 5.2.6 on your computer. Click Next to continue or Cancel to exit the Setup Wizard.

Version 5.2.6

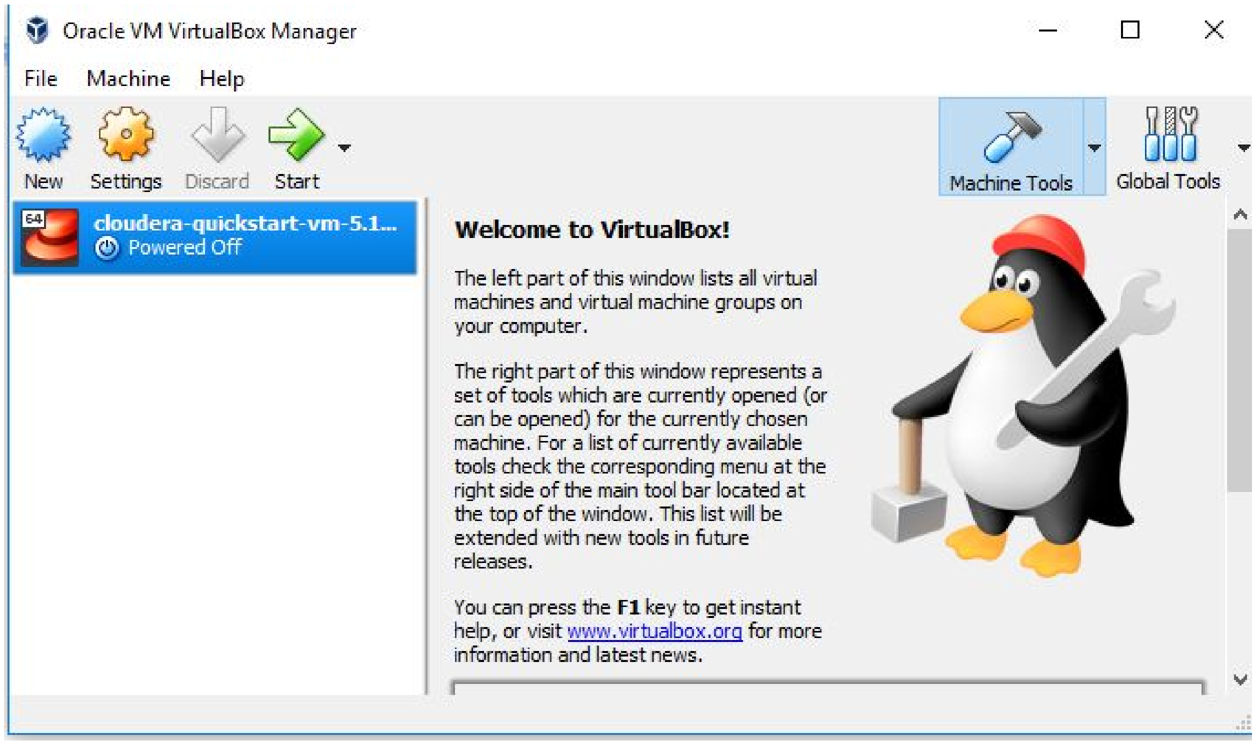

Oracle VM VirtualBox 5.2.6 Setup

Oracle VM VirtualBox 5.2.6

Please wait while the Setup Wizard installs Oracle VM VirtualBox 5.2.6. This may take several minutes.

Status: Copying new files

When we get installed, this window comes up and it means we have successfully installed VirtualBox as well as Cloudera. So let’s run it.


Oracle VM VirtualBox Manager

Welcome to VirtualBox!

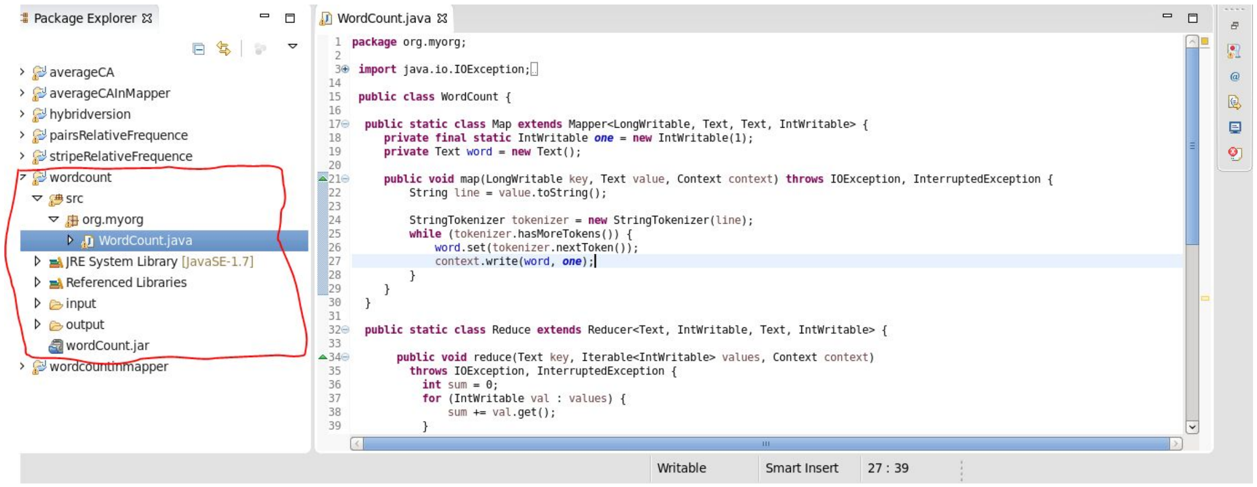
The left part of this window lists all virtual machines and virtual machine groups on your computer.

The right part of this window represents a set of tools which are currently opened (or can be opened) for the currently chosen machine. For a list of currently available tools check the corresponding menu at the right side of the main tool bar located at the top of the window. This list will be extended with new tools in future releases.

You can press the **F1** key to get instant help, or visit www.virtualbox.org for more information and latest news.

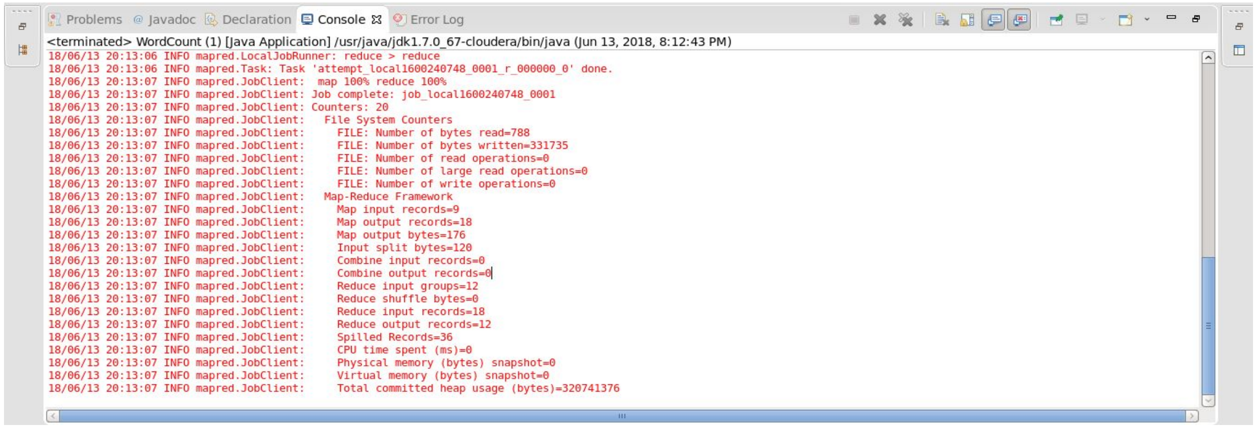
And Main view

Next thing we need to do is create the simple WordCount example and run it on the hadoop. For that we need eclipse to write code and make jar file.



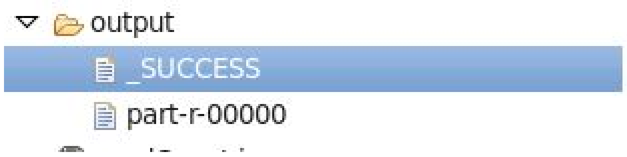
```java
package org.myorg;

import java.io.IOException;

public class WordCount {

 public static class Map extends 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, Context context) throws IOException, InterruptedException {
        String line = value.toString();

        StringTokenizer tokenizer = new StringTokenizer(line);
        while (tokenizer.hasMoreTokens()) {
            word.set(tokenizer.nextToken());
            context.write(word, one);
        }
    }
 }

 public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {

      public void reduce(Text key, Iterable<IntWritable> values, Context context)
        throws IOException, InterruptedException {
          int sum = 0;
          for (IntWritable val : values) {
              sum += val.get();
          }
```

The last step is to run the project to check if it is working.

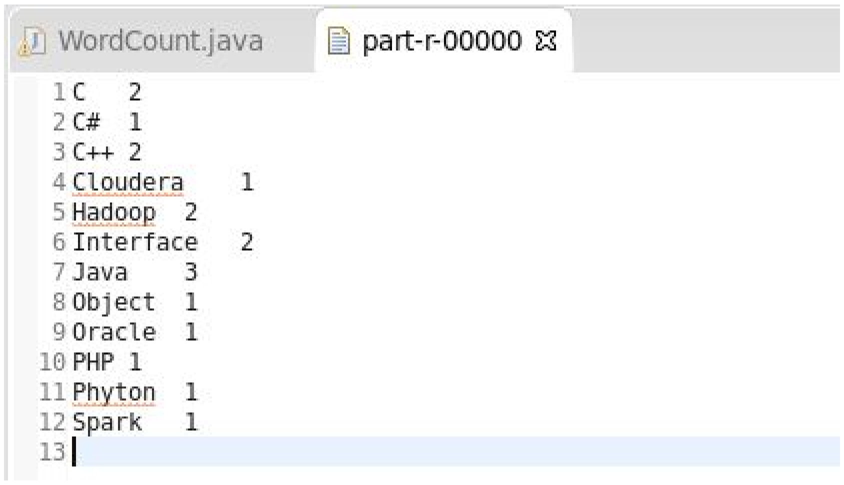


```
<terminated> WordCount (1) [Java Application] /usr/java/jdk1.7.0_67-cloudera/bin/java (Jun 13, 2018, 8:12:43 PM)
18/06/13 20:13:06 INFO mapred.LocalJobRunner: reduce > reduce
18/06/13 20:13:06 INFO mapred.Task: Task 'attempt_local1600240748_0001_r_000000_0' done.
18/06/13 20:13:07 INFO mapred.JobClient:  map 100% reduce 100%
18/06/13 20:13:07 INFO mapred.JobClient: Job complete: job_local1600240748_0001
18/06/13 20:13:07 INFO mapred.JobClient: Counters: 20
18/06/13 20:13:07 INFO mapred.JobClient:   File System Counters
18/06/13 20:13:07 INFO mapred.JobClient:     FILE: Number of bytes read=788
18/06/13 20:13:07 INFO mapred.JobClient:     FILE: Number of bytes written=331735
18/06/13 20:13:07 INFO mapred.JobClient:     FILE: Number of read operations=0
18/06/13 20:13:07 INFO mapred.JobClient:     FILE: Number of large read operations=0
18/06/13 20:13:07 INFO mapred.JobClient:     FILE: Number of write operations=0
18/06/13 20:13:07 INFO mapred.JobClient:   Map-Reduce Framework
18/06/13 20:13:07 INFO mapred.JobClient:     Map input records=9
18/06/13 20:13:07 INFO mapred.JobClient:     Map output records=18
18/06/13 20:13:07 INFO mapred.JobClient:     Map output bytes=176
18/06/13 20:13:07 INFO mapred.JobClient:     Input split bytes=120
18/06/13 20:13:07 INFO mapred.JobClient:     Combine input records=0
18/06/13 20:13:07 INFO mapred.JobClient:     Combine output records=0
18/06/13 20:13:07 INFO mapred.JobClient:     Reduce input groups=12
18/06/13 20:13:07 INFO mapred.JobClient:     Reduce shuffle bytes=0
18/06/13 20:13:07 INFO mapred.JobClient:     Reduce input records=18
18/06/13 20:13:07 INFO mapred.JobClient:     Reduce output records=12
18/06/13 20:13:07 INFO mapred.JobClient:     Spilled Records=36
18/06/13 20:13:07 INFO mapred.JobClient:     CPU time spent (ms)=0
18/06/13 20:13:07 INFO mapred.JobClient:     Physical memory (bytes) snapshot=0
18/06/13 20:13:07 INFO mapred.JobClient:     Virtual memory (bytes) snapshot=0
18/06/13 20:13:07 INFO mapred.JobClient:     Total committed heap usage (bytes)=320741376
```

It has successfully worked and created output file.



To make sure if it has produced the correct result open the file and see the result.



```
C	2
C#	1
C++	2
Cloudera	1
Hadoop	2
Interface	2
Java	3
Object	1
Oracle	1
PHP	1
Phyton	1
Spark	1
```

Code of WordCount.java

```
package org.myorg;

 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.mapreduce.*;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;

 public class WordCount {

  public static class Map extends 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, Context context) throws
IOException, InterruptedException {
        String line = value.toString();

        StringTokenizer tokenizer = new StringTokenizer(line);
        while (tokenizer.hasMoreTokens()) {
        word.set(tokenizer.nextToken());
        context.write(word, one);
        }
        }
  }

  public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {

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

        public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();

        @SuppressWarnings("deprecation")
Job job = new Job(conf, "wordcount");
        job.setJarByClass(WordCount.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);

        job.setMapperClass(Map.class);
        job.setReducerClass(Reduce.class);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));

        job.waitForCompletion(true);
        }
 }
```