

# **Maharishi University of Management**

**CS522 – Big Data**

**Prof. Premchand Nair**

**Set Up a Single Node Cluster**

**January 30<sup>th</sup>, 2017**

Author	ID	Date
Sovichea Cheth	985421	01/30/2017

In this manual, we're going to see how to set up a single node cluster and optionally an eclipse development environment to create and test your programs on Microsoft Windows.

## I. Install Oracle Virtual Box

First, go to <https://www.virtualbox.org/wiki/Downloads>

Under VirtualBox 5.1.14 platform packages, click on Windows hosts to download setup file for Windows.

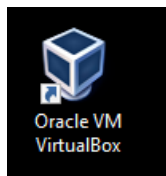


The screenshot shows the VirtualBox website's download page. On the left is a sidebar with links: About, Screenshots, Downloads, Documentation (with sub-links for End-user docs and Technical docs), Contribute, and Community. The main content area has the VirtualBox logo and the heading "Download VirtualBox". Below this, it says "Here, you will find links to VirtualBox binaries and its source code." and "VirtualBox binaries". A note states: "By downloading, you agree to the terms and conditions of the respective license." There are three main bullet points: 1. "VirtualBox 5.1.14 platform packages. The binaries are released under the terms of the GPL version 2." with sub-links for Windows hosts, OS X hosts, Linux distributions, and Solaris hosts. A red arrow points to "Windows hosts". 2. "VirtualBox 5.1.14 Oracle VM VirtualBox Extension Pack" with a link to "All supported platforms". It describes support for USB 2.0 and 3.0, RDP, disk encryption, NVMe, and PXE boot. It mentions the VirtualBox Personal Use and Evaluation License (PUEL) and provides instructions for installation. 3. "VirtualBox 5.1.14 Software Developer Kit (SDK)" with a link to "All platforms".

Then after download, double click on setup file to perform the VirtualBox installation

Name	Date modified	Type	Size
SetUp_Documentation	1/30/2017 10:18 AM	Microsoft Word D...	987 KB
VirtualBox-5.1.14-112924-Win	1/30/2017 10:38 AM	Application	120,433 KB

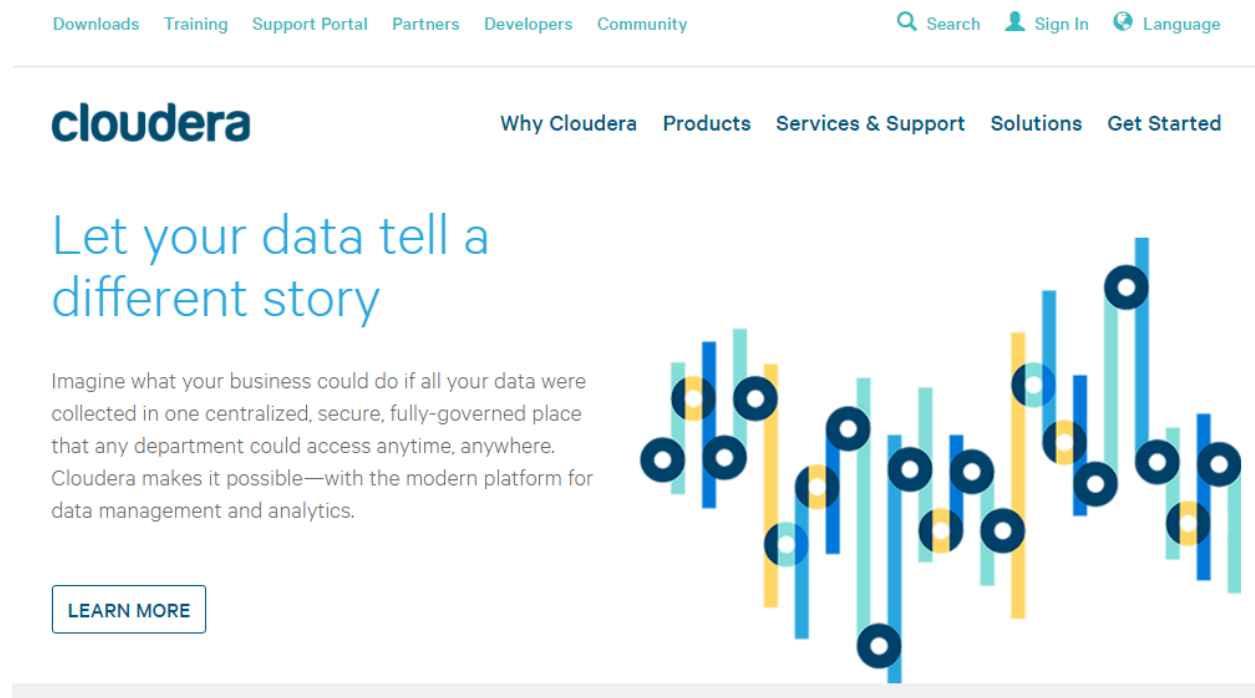
After finished installation, you will able to run the program.



## II. Download & Install Cloudera

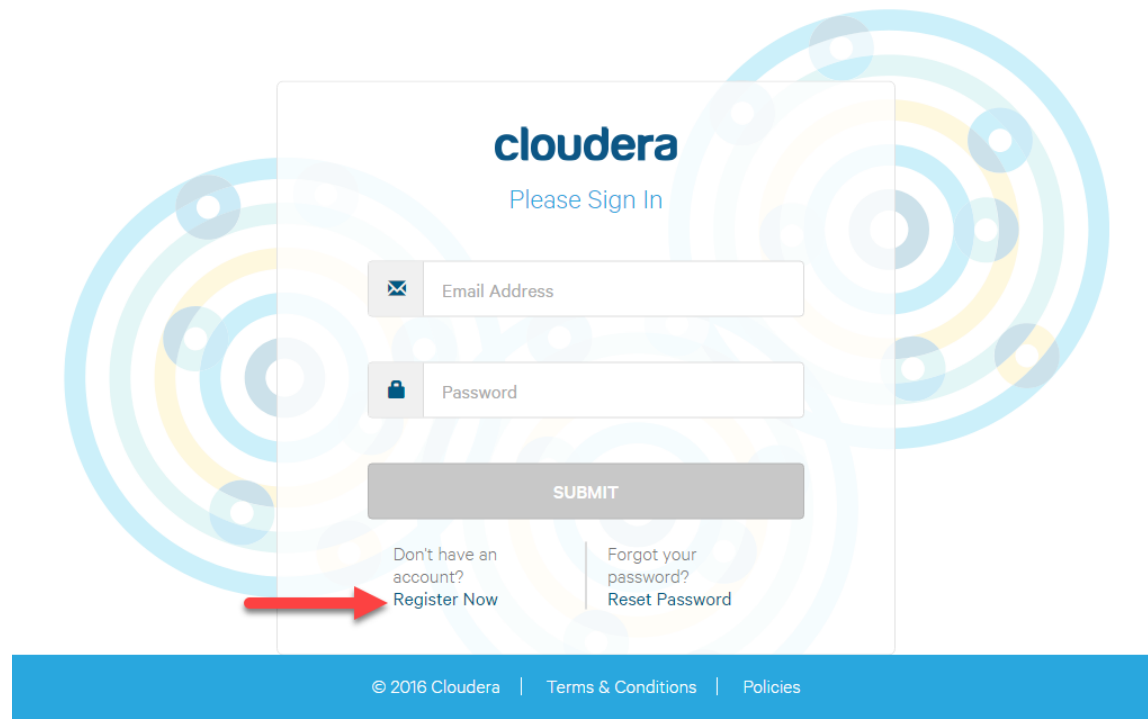
### a. Get Cloudera

First, go to <http://www.cloudera.com/>



The screenshot shows the Cloudera website homepage. At the top, there is a navigation bar with links: Downloads, Training, Support Portal, Partners, Developers, Community, Search, Sign In, and Language. Below this is the Cloudera logo and a secondary navigation bar with links: Why Cloudera, Products, Services & Support, Solutions, and Get Started. The main content area features the headline "Let your data tell a different story" and a paragraph: "Imagine what your business could do if all your data were collected in one centralized, secure, fully-governed place that any department could access anytime, anywhere. Cloudera makes it possible—with the modern platform for data management and analytics." To the right of the text is a decorative graphic of vertical bars and circles in blue, yellow, and teal. Below the text is a "LEARN MORE" button.

Click on **Sign In**, you will need an account to download Cloudera, if you don't have an account yet, click on **Register Now**.



The screenshot shows the Cloudera Sign In form. The form is titled "cloudera Please Sign In". It contains two input fields: "Email Address" and "Password". Below these fields is a "SUBMIT" button. At the bottom of the form, there are two links: "Don't have an account? Register Now" and "Forgot your password? Reset Password". A red arrow points to the "Register Now" link. The background of the form is a light blue and yellow circular pattern.

After finished Register or Sign In, go to <http://www.cloudera.com/downloads.html> to download Cloudera. Under Quick Start, click on **Download Now**

Downloads Training Support Portal Partners Developers Community

Search Sovichea Language

# cloudera

Why Cloudera Products Services & Support Solutions Get Started

## Download Cloudera Enterprise

Local, On Premise, or Cloud-based Apache Hadoop Management

### QuickStarts

Get Started on your local machine using a QuickStart VM or Docker Image.

**DOWNLOAD NOW**

[Learn More](#)

### Cloudera Manager

A unified interface to manage your enterprise data hub. Express and Enterprise editions available.

**DOWNLOAD NOW**

[Learn More](#)

### Cloudera Director

Self-service, reliable experience for CDH and Cloudera Enterprise in the cloud

**DOWNLOAD NOW**

[Learn More](#)

Select Version and Platform: Virtual Box, then click **Get It Now**

Downloads Training Support Portal Partners Developers Community

Search Sovichea Language

# cloudera

Why Cloudera Products Services & Support Solutions Get Started

## QuickStarts for CDH 5.8

Virtualized clusters for easy installation on your desktop!

Cloudera QuickStart for Docker (multi-node cluster) and Cloudera QuickStart VM (single-node cluster) make it easy to quickly get hands-on with CDH for testing, demo, and self-learning purposes, and include Cloudera Manager for managing your cluster. Cloudera QuickStart VM also includes a tutorial, sample data, and scripts for getting started.

Cloudera QuickStarts, deployed via Docker containers or VMs, are not intended or supported for use in production.

### Get Started Now

Version


QuickStarts for CDH 5.8

Platform



Virtual Box

**GET IT NOW →**

Download will start, you will get the zip file after finished

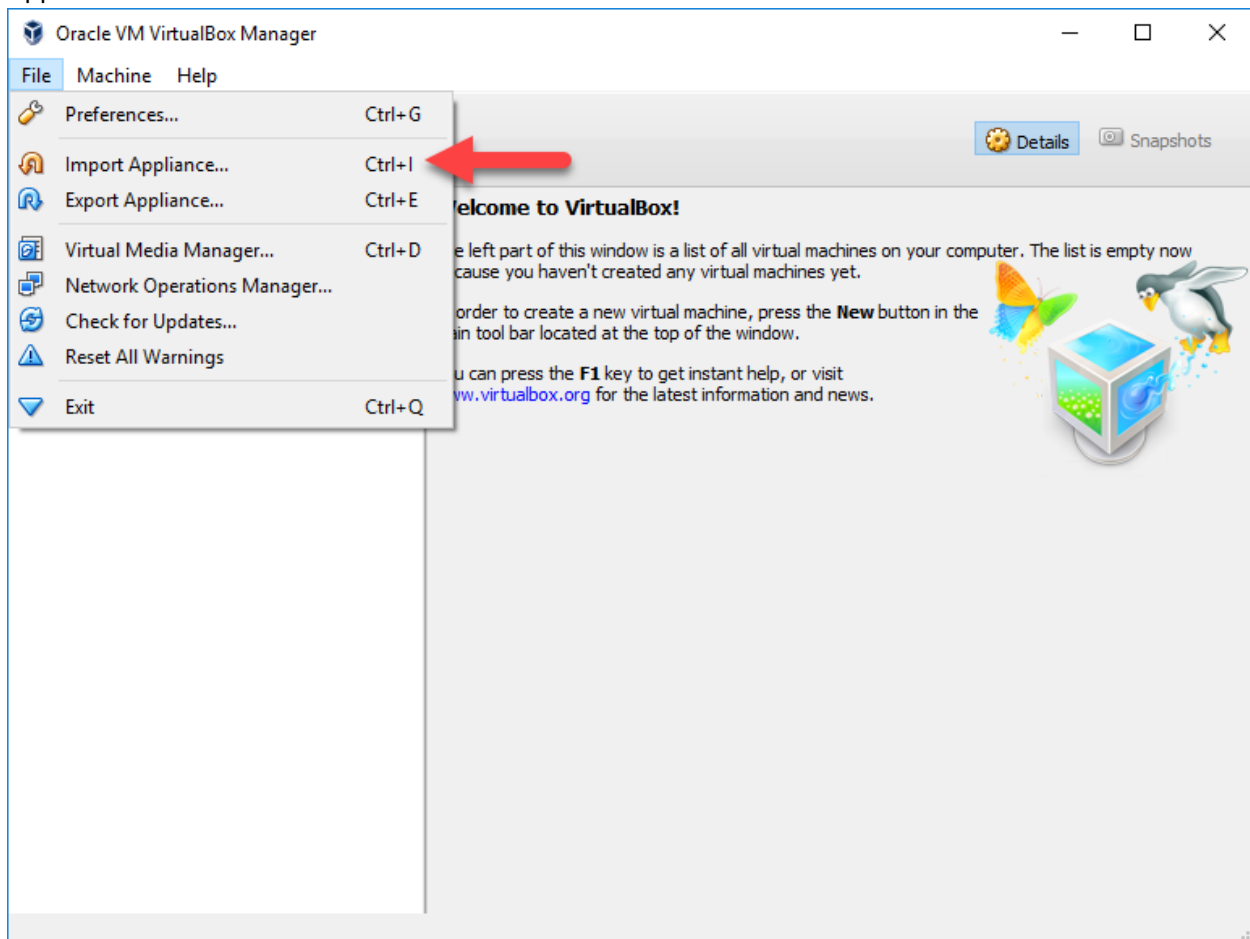
Name	Date modified	Type	Size
 cloudera-quickstart-vm-5.8.0-0-virtualbox	1/30/2017 1:33 AM	WinZip File	5,148,003 KB

Extract the zip file, you will get VirtualBox image file

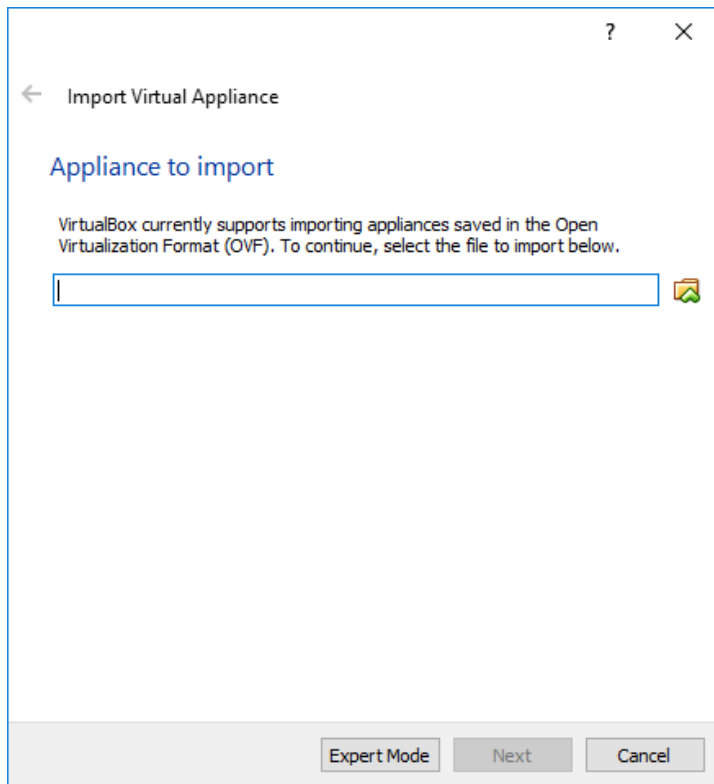
Name	Date modified	Type	Size
 cloudera-quickstart-vm-5.8.0-0-virtualbox	8/10/2016 9:20 PM	Open Virtualization...	15 KB
 cloudera-quickstart-vm-5.8.0-0-virtualbo...	8/10/2016 9:25 PM	Virtual Machine Di...	5,199,931 KB

## b. Install Cloudera

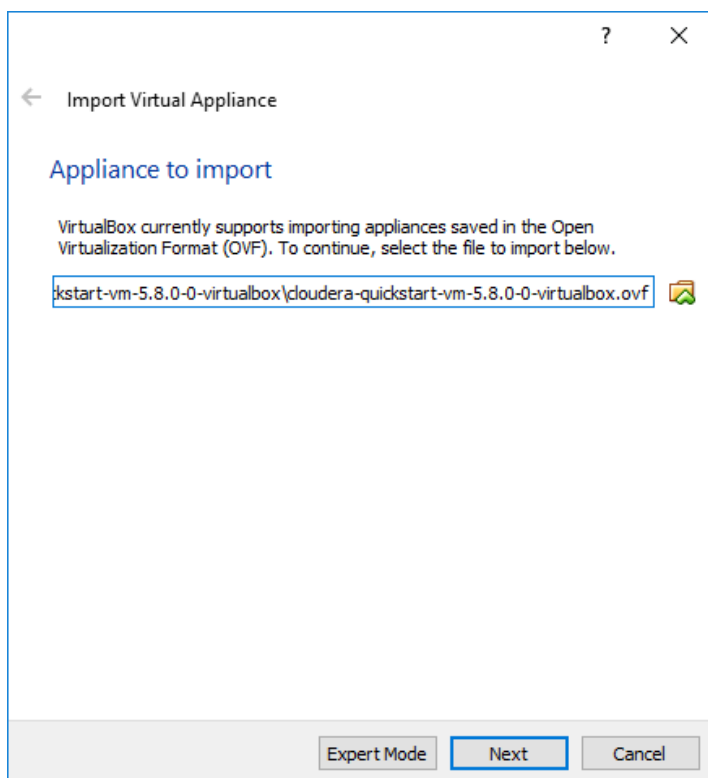
After you VirtualBox image file, you can import it to VirtualBox. Run VirtualBox, then click on Import Appliance



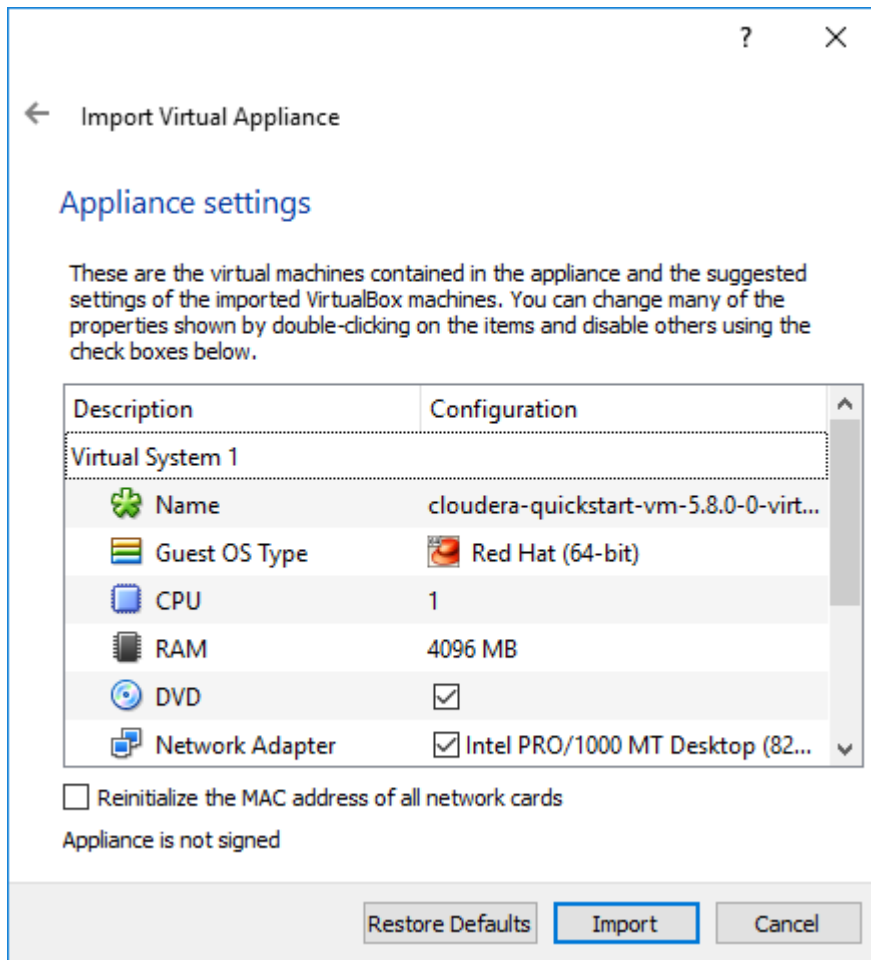
Select your VirtualBox image file location



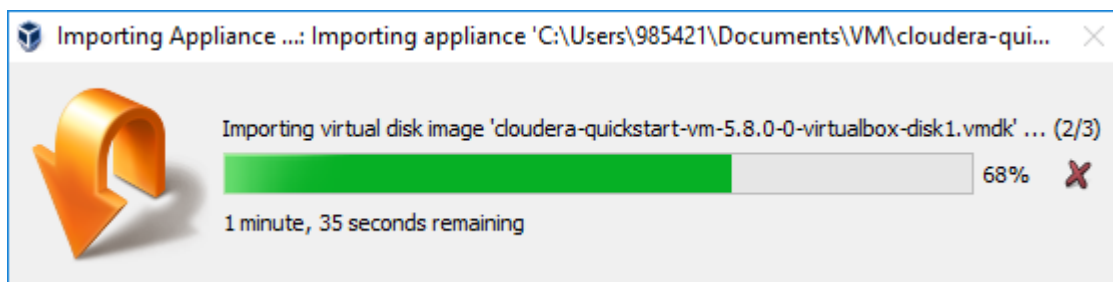
After that click Next



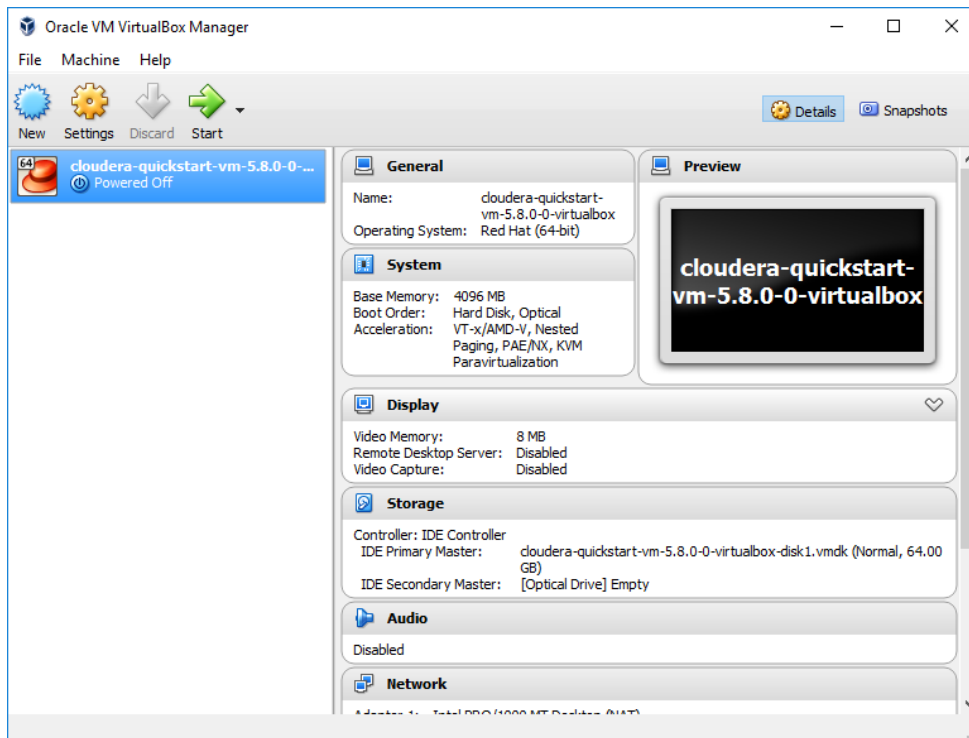
Configure your setting. Minimum RAM requirement for this VM is 4GB. Then click on Import.



VirtualBox will start importing the appliance.



After finished import, you will able to see your Virtual Machine. Click Start to start the VM.

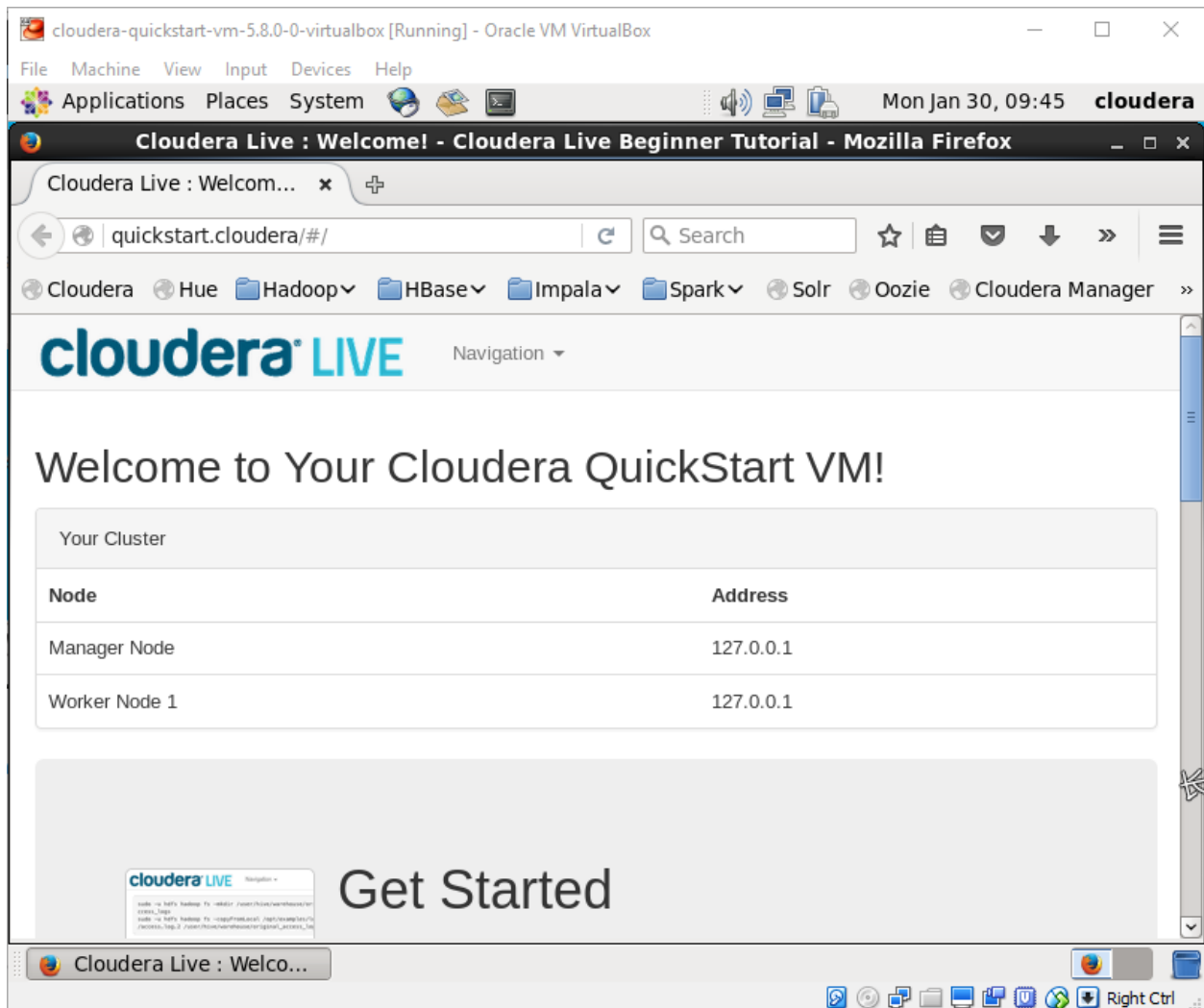


You will able to get to Red Hat CentOS after finish running

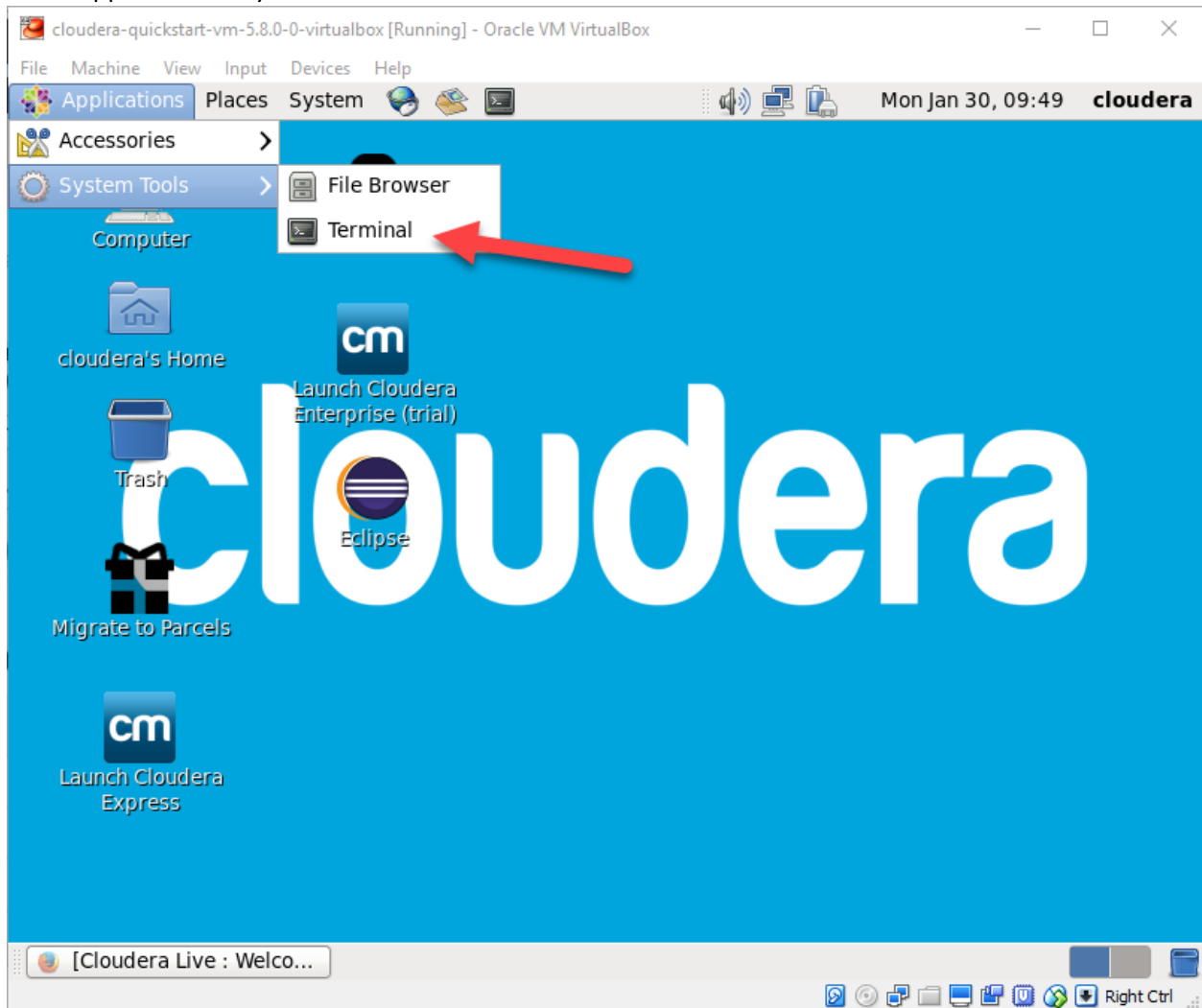




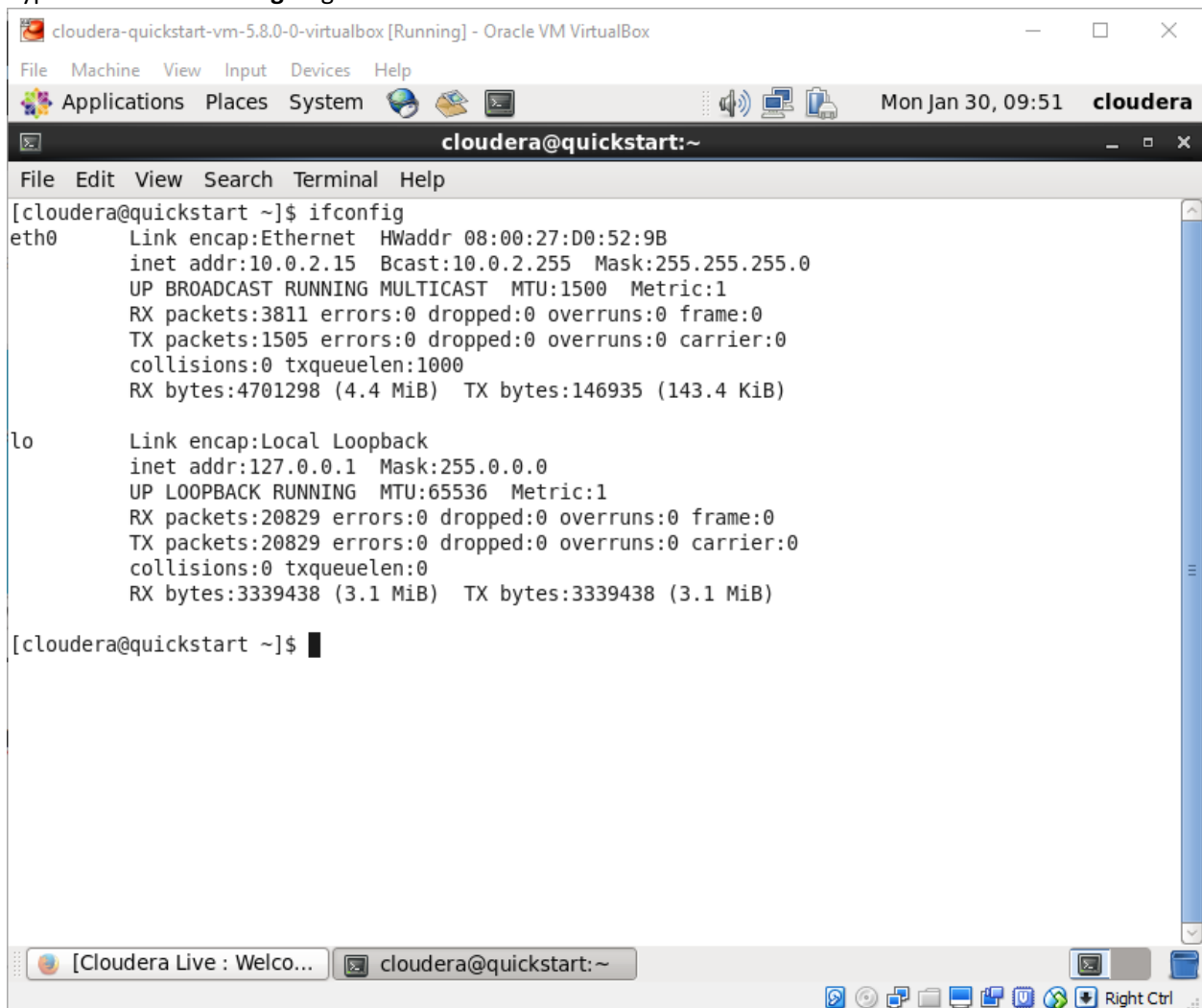
Cloudera Welcome page will pop up automatically.



Go to Application -> System Tools -> Terminal



Type command: **ifconfig** to get the IP and network information



The screenshot shows a terminal window titled "cloudera-quickstart-vm-5.8.0-0-virtualbox [Running] - Oracle VM VirtualBox". The terminal displays the output of the `ifconfig` command, showing details for the `eth0` and `lo` interfaces. The `eth0` interface has an IP address of `10.0.2.15` and a netmask of `255.255.255.0`. The `lo` interface has an IP address of `127.0.0.1` and a netmask of `255.0.0.0`. The terminal window also shows a taskbar at the bottom with a "Cloudera Live" window and a "cloudera@quickstart:~" terminal window.

```
[cloudera@quickstart ~]$ ifconfig
eth0      Link encap:Ethernet  HWaddr 08:00:27:D0:52:9B
          inet addr:10.0.2.15  Bcast:10.0.2.255  Mask:255.255.255.0
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:3811 errors:0 dropped:0 overruns:0 frame:0
          TX packets:1505 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:4701298 (4.4 MiB)  TX bytes:146935 (143.4 KiB)

lo        Link encap:Local Loopback
          inet addr:127.0.0.1  Mask:255.0.0.0
          UP LOOPBACK RUNNING  MTU:65536  Metric:1
          RX packets:20829 errors:0 dropped:0 overruns:0 frame:0
          TX packets:20829 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:3339438 (3.1 MiB)  TX bytes:3339438 (3.1 MiB)

[cloudera@quickstart ~]$
```

### III. Test Run WordCount

#### a. Download Hadoop Libraries

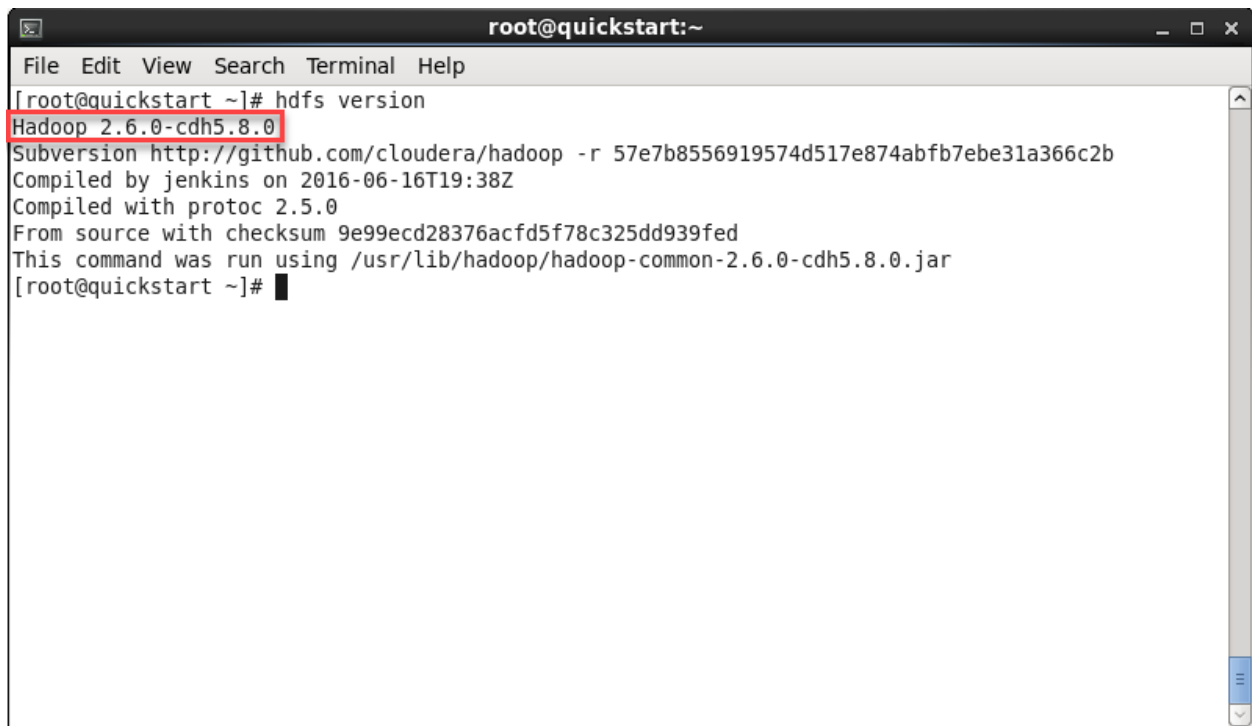
You will require hadoop libraries to run WordCount.

First, check your Hadoop version. Go to Terminal -> run command

hdfs version

Or

hadoop version

A terminal window titled 'root@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '[root@quickstart ~]# hdfs version' is entered. The output is: 'Hadoop 2.6.0-cdh5.8.0' (highlighted with a red box), 'Subversion http://github.com/cloudera/hadoop -r 57e7b8556919574d517e874abfb7ebe31a366c2b', 'Compiled by jenkins on 2016-06-16T19:38Z', 'Compiled with protoc 2.5.0', 'From source with checksum 9e99ecd28376acfd5f78c325dd939fed', and 'This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.8.0.jar'. The prompt '[root@quickstart ~]#' is followed by a cursor.

```
root@quickstart:~  
File Edit View Search Terminal Help  
[root@quickstart ~]# hdfs version  
Hadoop 2.6.0-cdh5.8.0  
Subversion http://github.com/cloudera/hadoop -r 57e7b8556919574d517e874abfb7ebe31a366c2b  
Compiled by jenkins on 2016-06-16T19:38Z  
Compiled with protoc 2.5.0  
From source with checksum 9e99ecd28376acfd5f78c325dd939fed  
This command was run using /usr/lib/hadoop/hadoop-common-2.6.0-cdh5.8.0.jar  
[root@quickstart ~]#
```

In this case, Hadoop version is 2.6.0, so we will need to download libraries exactly as this version.

You can download Hadoop library from this website

<http://mirrors.ibiblio.org/apache/hadoop/common/>

Index of /apache/hadoop/con

Cloudera Li... WordCount ... Hadoop ... Y hadoop ... rewrite -... Running... Y

mirrors.ibiblio.org/apache/hadoop/common/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloude

## Index of /apache/hadoop/common

Name	Last modified	Size	Description
<a href="#">Parent Directory</a>		-	
<a href="#">current/</a>	25-Jan-2017 16:32	-	
<a href="#">hadoop-1.2.1/</a>	17-Feb-2015 18:46	-	
<a href="#">hadoop-2.5.2/</a>	17-Feb-2015 18:46	-	
<a href="#">hadoop-2.6.0/</a>	17-Feb-2015 18:46	-	
<a href="#">hadoop-2.6.1/</a>	23-Sep-2015 18:34	-	
<a href="#">hadoop-2.6.2/</a>	30-Oct-2015 17:05	-	
<a href="#">hadoop-2.6.3/</a>	17-Dec-2015 21:49	-	
<a href="#">hadoop-2.6.4/</a>	12-Feb-2016 07:37	-	
<a href="#">hadoop-2.6.5/</a>	10-Oct-2016 23:24	-	
<a href="#">hadoop-2.7.0/</a>	21-Apr-2015 12:47	-	
<a href="#">hadoop-2.7.1/</a>	06-Jul-2015 20:33	-	
<a href="#">hadoop-2.7.2/</a>	25-Jan-2016 20:29	-	
<a href="#">hadoop-2.7.3/</a>	25-Aug-2016 15:25	-	

Choose the correct version and file

Index of /apache/hadoop/common/hadoop-2.6.0 - Mozilla Fire

Cloudera Li... WordCount ... Hadoop ... Y hadoop ... rewrite -... Running... Y tar -xvzf... Index of ... x

mirrors.ibiblio.org/apache/hadoop/common/hadoop-2.6.0/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Starte

## Index of /apache/hadoop/common/hadoop-2.6.0

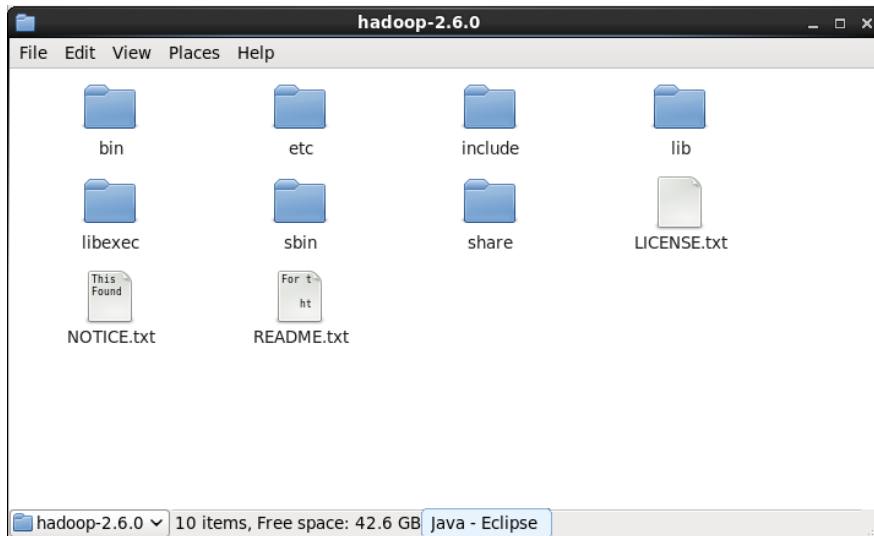
Name	Last modified	Size	Description
<a href="#">Parent Directory</a>		-	
<a href="#">hadoop-2.6.0-src.tar.gz</a>	30-Nov-2014 18:52	17M	
<a href="#">hadoop-2.6.0-src.tar.gz.mds</a>	30-Nov-2014 18:52	1.1K	
<a href="#">hadoop-2.6.0.tar.gz</a>	30-Nov-2014 18:52	186M	
<a href="#">hadoop-2.6.0.tar.gz.mds</a>	30-Nov-2014 18:52	958	

Apache Server at mirrors.ibiblio.org Port 80

After that, extract the file with this command

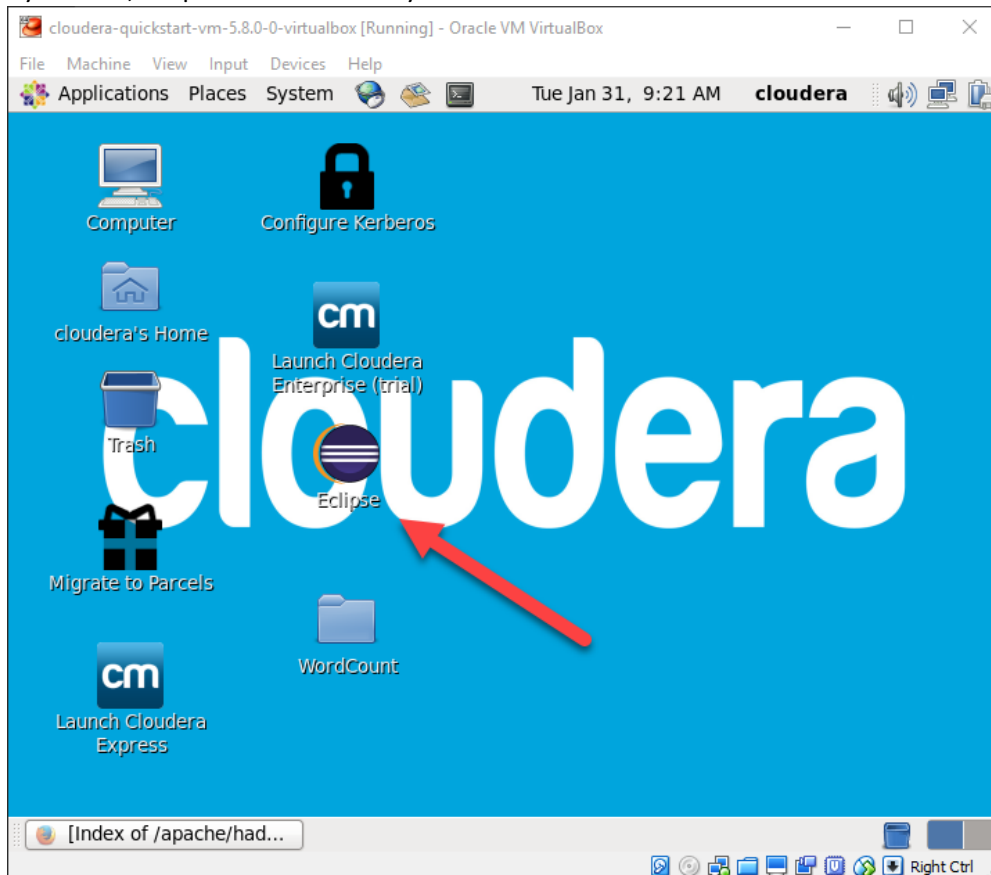
```
tar -xvzf filename.tar.gz
```

You will get the file as below, and you will need it to import to Eclipse later

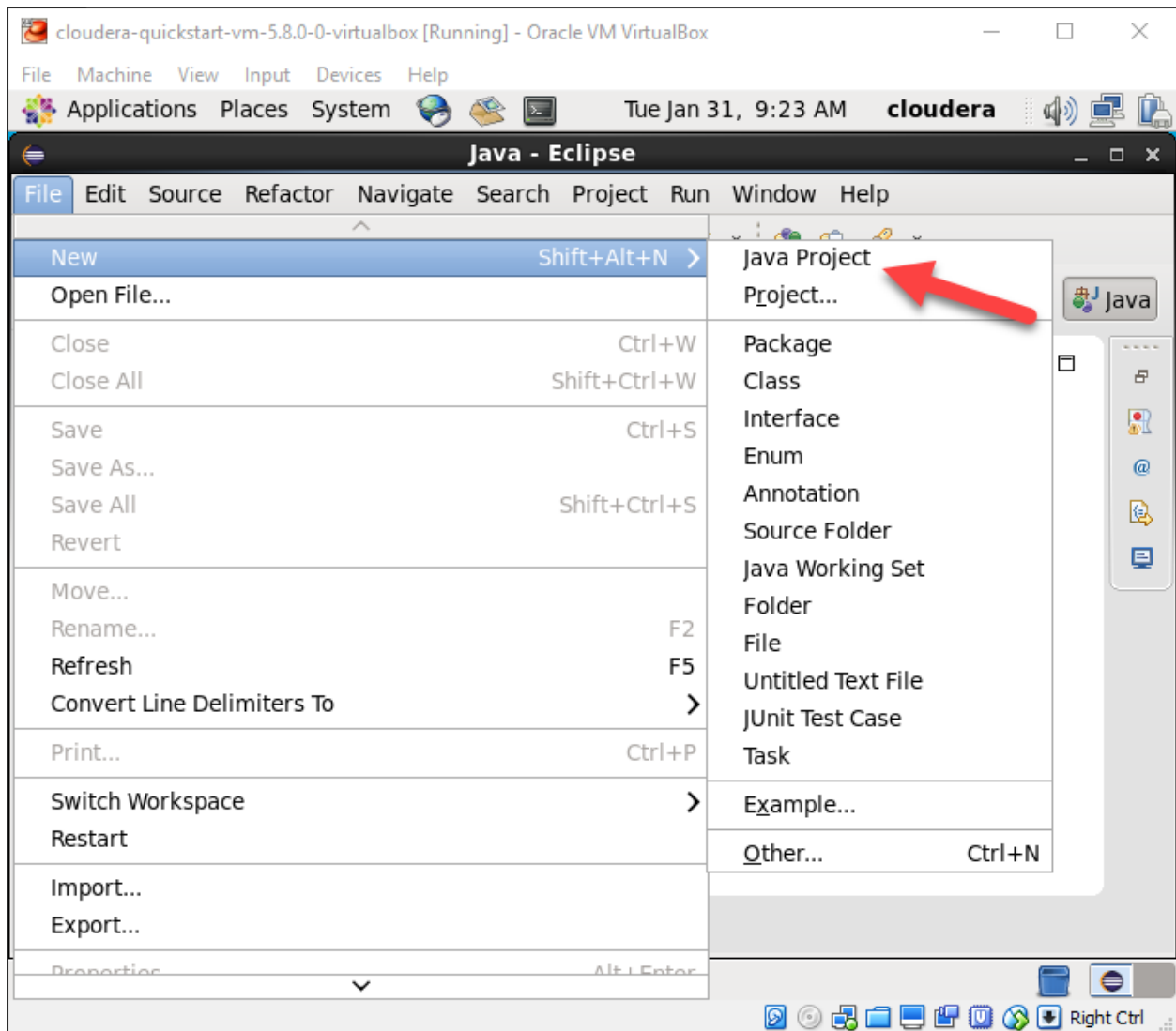


## b. Eclipse Project Setup

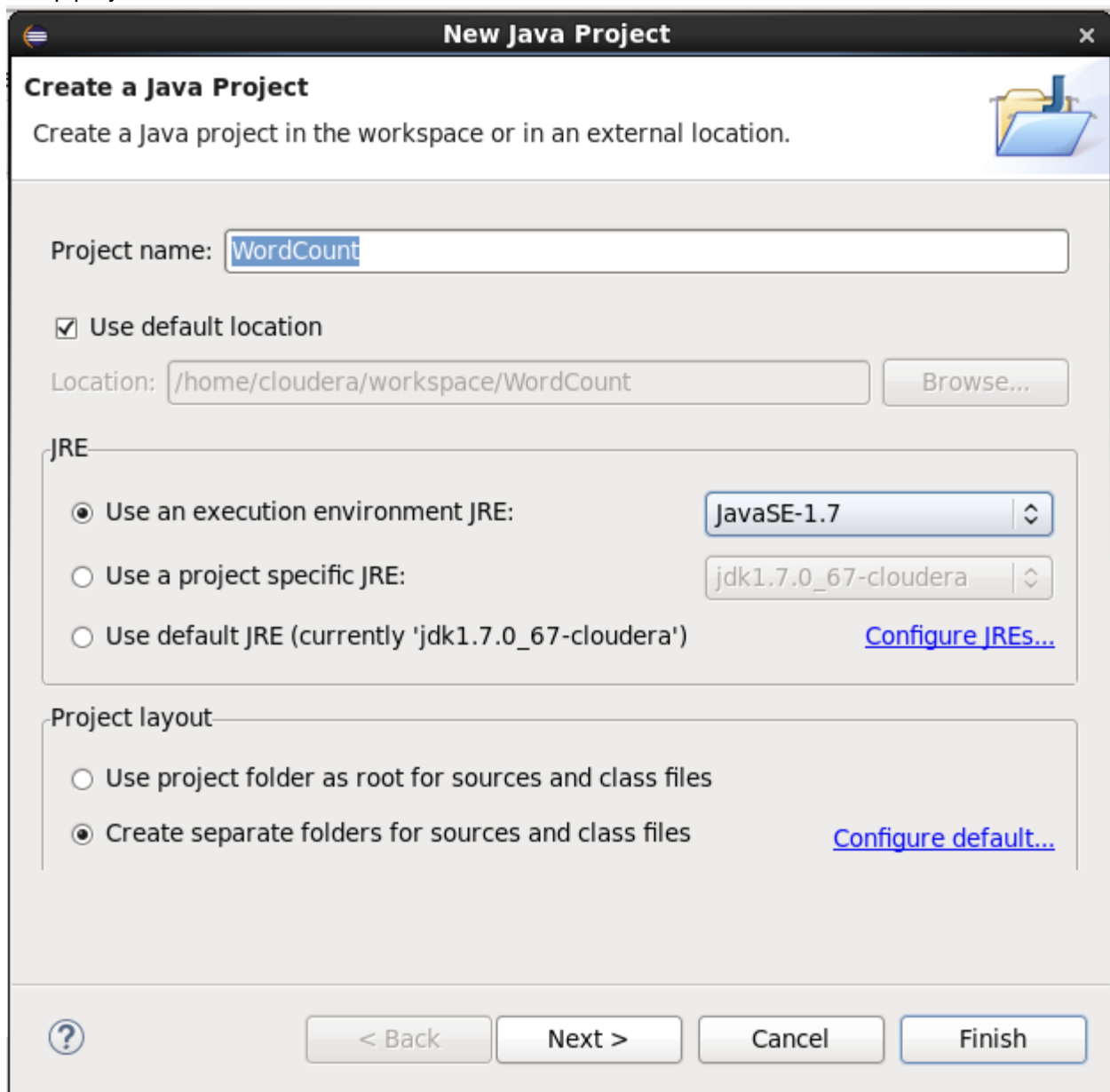
By default, Eclipse is available on your VM machine



Run the Eclipse IDE and then create Java Project



Setup project name and JRE



The image shows a 'New Java Project' dialog box with a title bar containing a menu icon, the text 'New Java Project', and a close button. The main area is titled 'Create a Java Project' and includes a folder icon. Below the title is the instruction 'Create a Java project in the workspace or in an external location.' The 'Project name' field contains 'WordCount'. The 'Use default location' checkbox is checked, and the 'Location' field shows '/home/cloudera/workspace/WordCount' with a 'Browse...' button. The 'JRE' section has three radio buttons: 'Use an execution environment JRE:' (selected), 'Use a project specific JRE:', and 'Use default JRE (currently 'jdk1.7.0\_67-cloudera')'. The first option has a dropdown menu showing 'JavaSE-1.7'. The second option has a dropdown menu showing 'jdk1.7.0\_67-cloudera'. A 'Configure JREs...' link is next to the third option. The 'Project layout' section has two radio buttons: 'Use project folder as root for sources and class files' and 'Create separate folders for sources and class files' (selected). A 'Configure default...' link is next to the second option. At the bottom are a help icon, '< Back' button, 'Next >' button, 'Cancel' button, and 'Finish' button.

**New Java Project**

**Create a Java Project**

Create a Java project in the workspace or in an external location.

Project name:

☒ Use default location

Location:  [Browse...](#)

**JRE**

☒ Use an execution environment JRE:

☐ Use a project specific JRE:

☐ Use default JRE (currently 'jdk1.7.0\_67-cloudera') [Configure JREs...](#)

**Project layout**

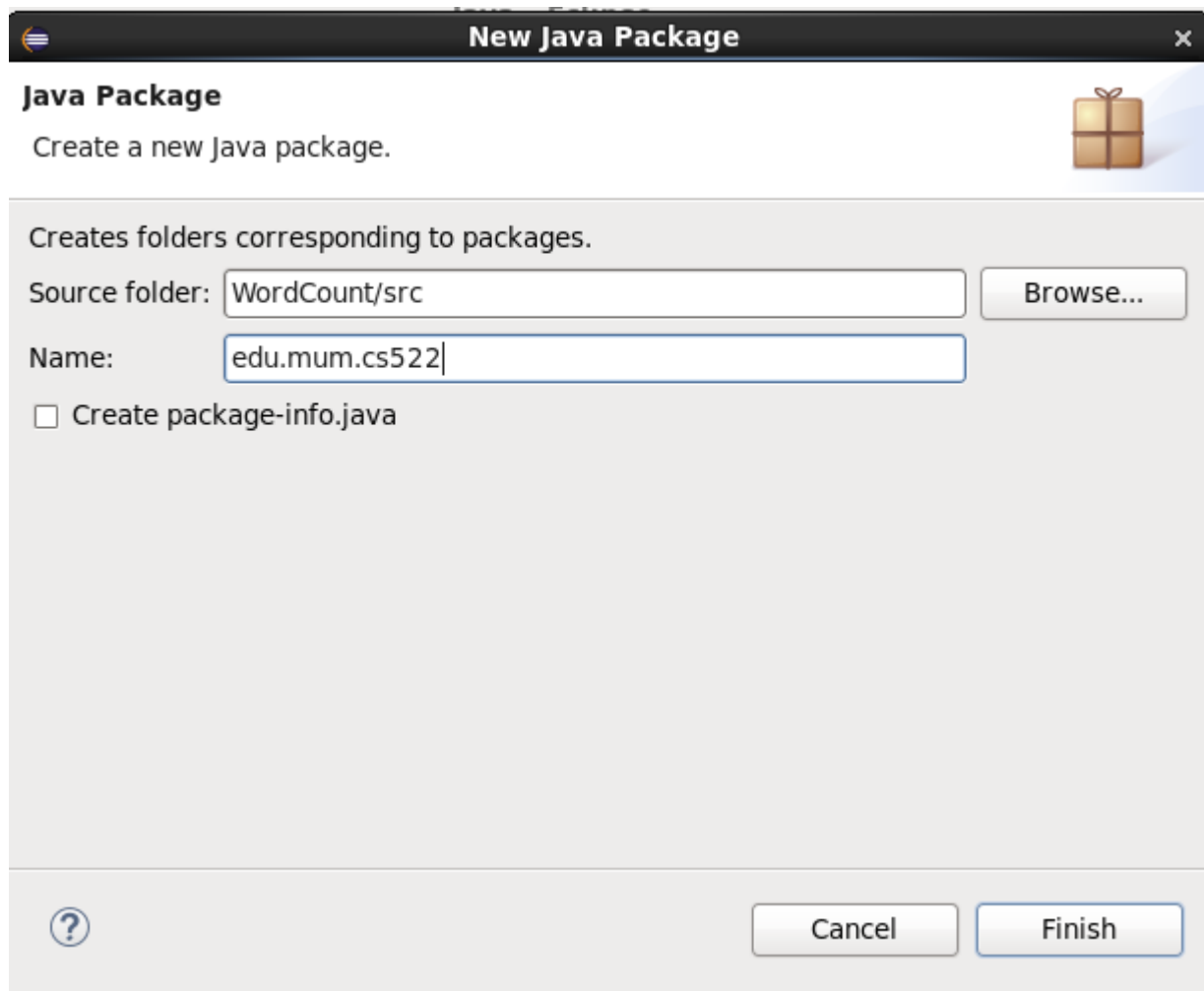
☐ Use project folder as root for sources and class files

☒ Create separate folders for sources and class files [Configure default...](#)

[?](#) [< Back](#) [Next >](#) [Cancel](#) [Finish](#)



Create Package name



The image shows a 'New Java Package' dialog box from an IDE. The title bar is dark with a blue icon on the left and a close button on the right. The main area has a light gray background. At the top, it says 'Java Package' in bold, followed by 'Create a new Java package.' and a small gift icon. Below this, it says 'Creates folders corresponding to packages.' There are two text input fields: 'Source folder:' with the value 'WordCount/src' and a 'Browse...' button to its right; and 'Name:' with the value 'edu.mum.cs522'. Below the 'Name:' field is a checkbox labeled 'Create package-info.java'. At the bottom, there is a help icon (a question mark in a circle) on the left, and 'Cancel' and 'Finish' buttons on the right.

**New Java Package**

**Java Package**  
Create a new Java package.

Creates folders corresponding to packages.

Source folder: WordCount/src Browse...

Name: edu.mum.cs522

☐ Create package-info.java

? Cancel Finish

Create WordCount class

**New Java Class**

Create a new Java class.

Source folder: WordCount/src Browse...

Package: edu.mum.cs522 Browse...

☐ Enclosing type: Browse...

Name: WordCount

Modifiers: ☒ public ☐ package ☐ private ☐ protected  
☐ abstract ☐ final ☐ static

Superclass: java.lang.Object Browse...

Interfaces: Add...

Which method stubs would you like to create?

☐ public static void main(String[] args)

☐ Constructors from superclass

? Cancel Finish

### c. WordCount Source Code and Library Setup

WordCount maps (extract) words from an input source and reduces the result, return a count of each word. You can find source code below or on the internet

```

package edu.mum.cs522;

import java.io.IOException;
import java.util.regex.Pattern;
import org.apache.hadoop.conf.Configured;
import org.apache.hadoop.util.Tool;
import org.apache.hadoop.util.ToolRunner;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;

import org.apache.log4j.Logger;

public class WordCount extends Configured implements Tool {

    private static final Logger LOG = Logger.getLogger(WordCount.class);

    public static void main(String[] args) throws Exception {
        int res = ToolRunner.run(new WordCount(), args);
        System.exit(res);
    }

    public int run(String[] args) throws Exception {
        Job job = Job.getInstance(getConf(), "wordcount");
        job.setJarByClass(this.getClass());
        // Use TextInputFormat, the default unless job.setInputFormatClass is used
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.setMapperClass(Map.class);
        job.setReducerClass(Reduce.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        return job.waitForCompletion(true) ? 0 : 1;
    }

    public static class Map extends Mapper<LongWritable, Text, Text, IntWritable> {
        private final static IntWritable one = new IntWritable(1);
        private Text word = new Text();
        private long numRecords = 0;
        private static final Pattern WORD_BOUNDARY = Pattern.compile("\\s*\\b\\s*");

        public void map(LongWritable offset, Text lineText, Context context)
            throws IOException, InterruptedException {
            String line = lineText.toString();
            Text currentWord = new Text();
            for (String word : WORD_BOUNDARY.split(line)) {
                if (word.isEmpty()) {
                    continue;
                }
                currentWord = new Text(word);
            }
        }
    }

```

```

        context.write(currentWord, one);
    }
}

public static class Reduce extends Reducer<Text, IntWritable, Text, IntWritable> {
    @Override
    public void reduce(Text word, Iterable<IntWritable> counts, Context context)
        throws IOException, InterruptedException {
        int sum = 0;
        for (IntWritable count : counts) {
            sum += count.get();
        }
        context.write(word, new IntWritable(sum));
    }
}
}

```

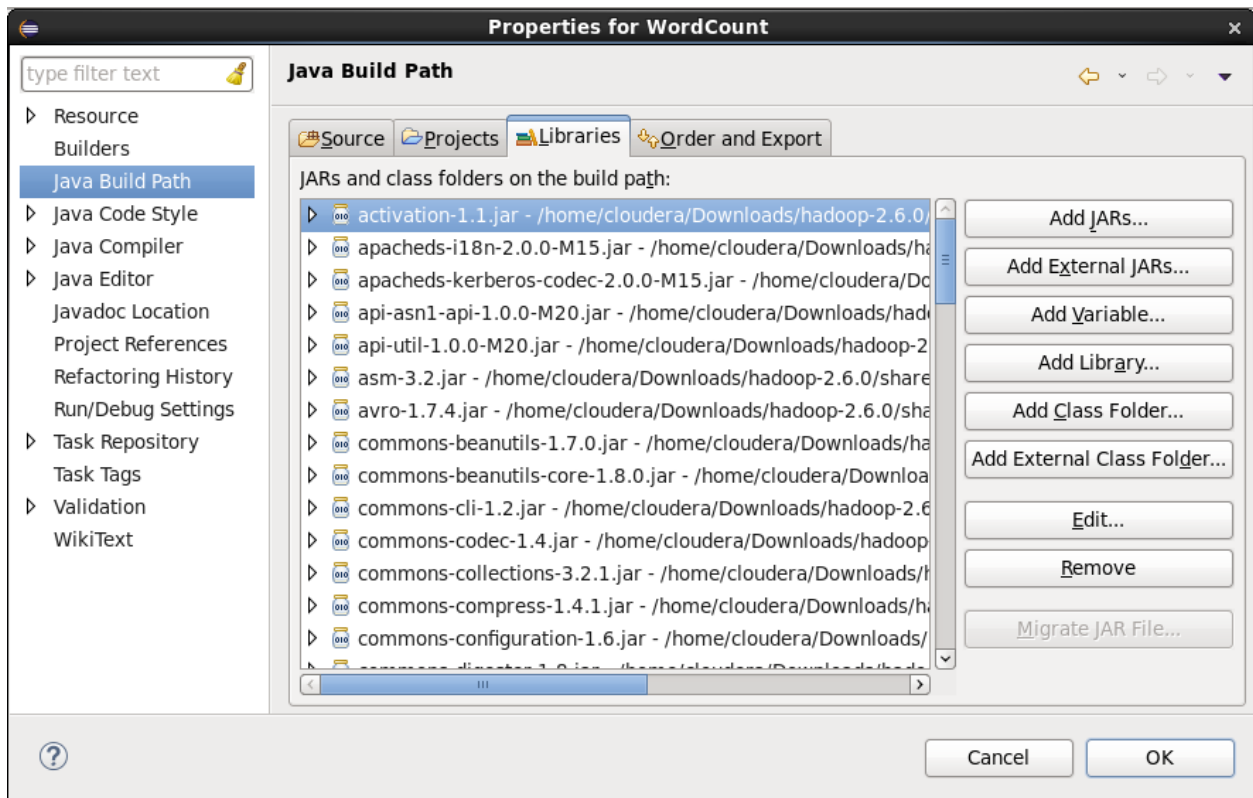
Add all required libraries as below

share/hadoop/common/hadoop-common-\*.jar

share/hadoop/mapreduce/hadoop-mapreduce-client-core-\*.jar

share/hadoop/mapreduce/hadoop-mapreduce-client-jobclient-\*.jar

share/hadoop/common/hadoop-common-lib-\*.jar (all files)



#### d. Running WordCount

Before you run, you must create input and output locations in HDFS. Use the following commands to create input directory /user/cloudera/wordcount/input in HDFS:

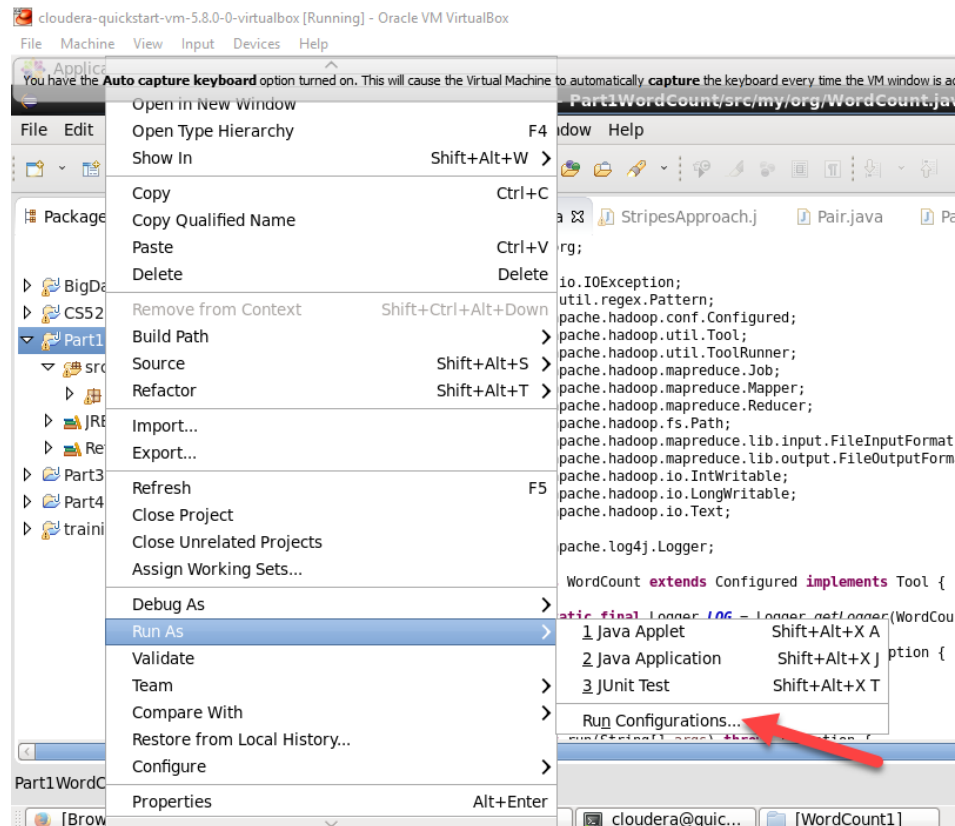
```
$ sudo su hdfs
$ hadoop fs -mkdir /user/cloudera
$ hadoop fs -chown cloudera /user/cloudera
$ exit$ sudo su cloudera
$ hadoop fs -mkdir /user/cloudera/wordcount /user/cloudera/wordcount/input
```

Create sample text files to use as input, and move them to the /user/cloudera/wordcount/input directory in HDFS. You can use any files you choose; for convenience, the following shell commands create a few small input files for illustrative purposes.

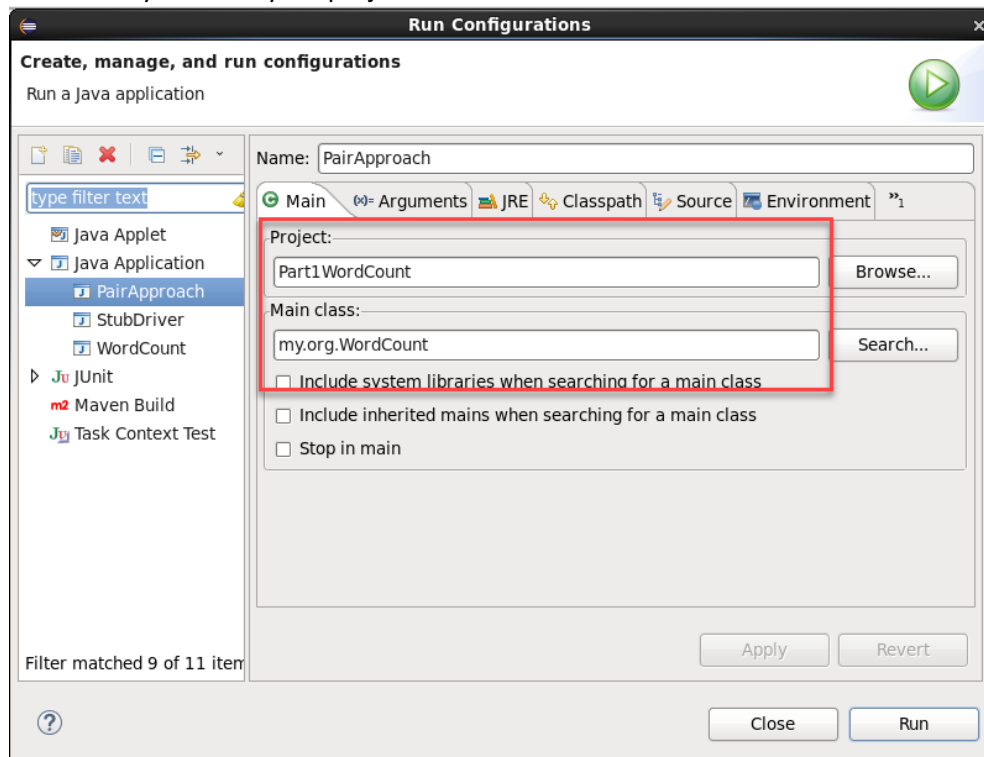
```
$ echo "Hadoop is an elephant" > file0
$ echo "Hadoop is as yellow as can be" > file1
$ echo "Oh what a yellow fellow is Hadoop" > file2
$ hadoop fs -put file* /user/cloudera/wordcount/input
```

Compile WordCount class.

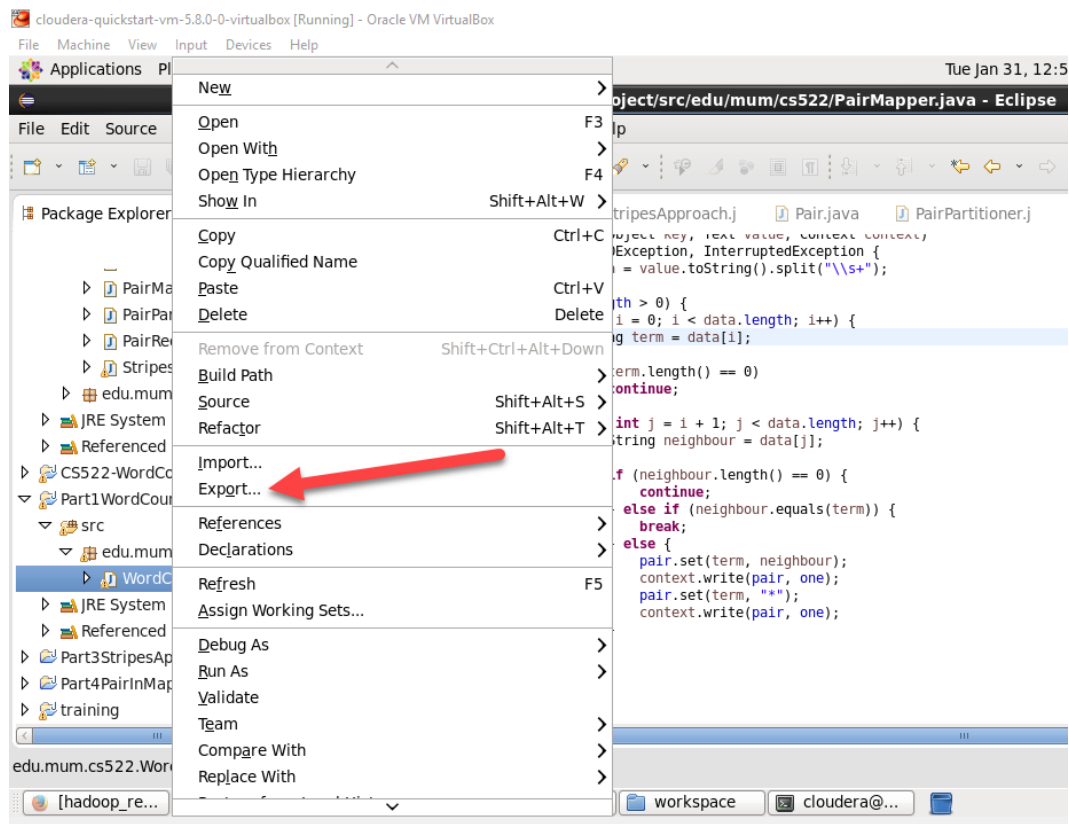
Check your project Run Configuration



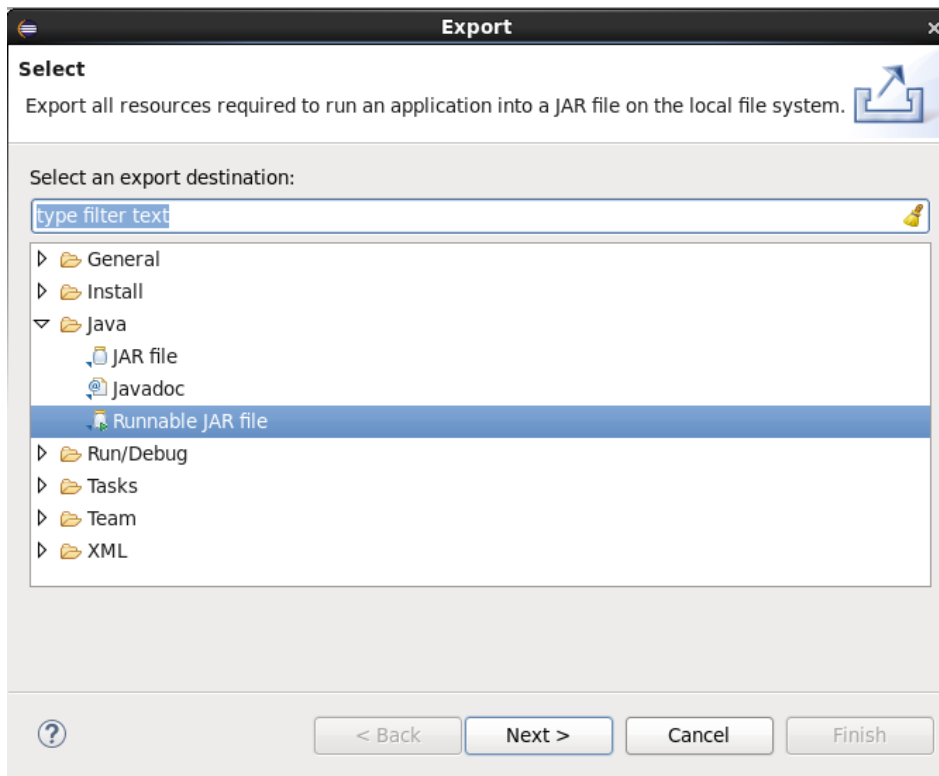
Make sure you check your project name and main class



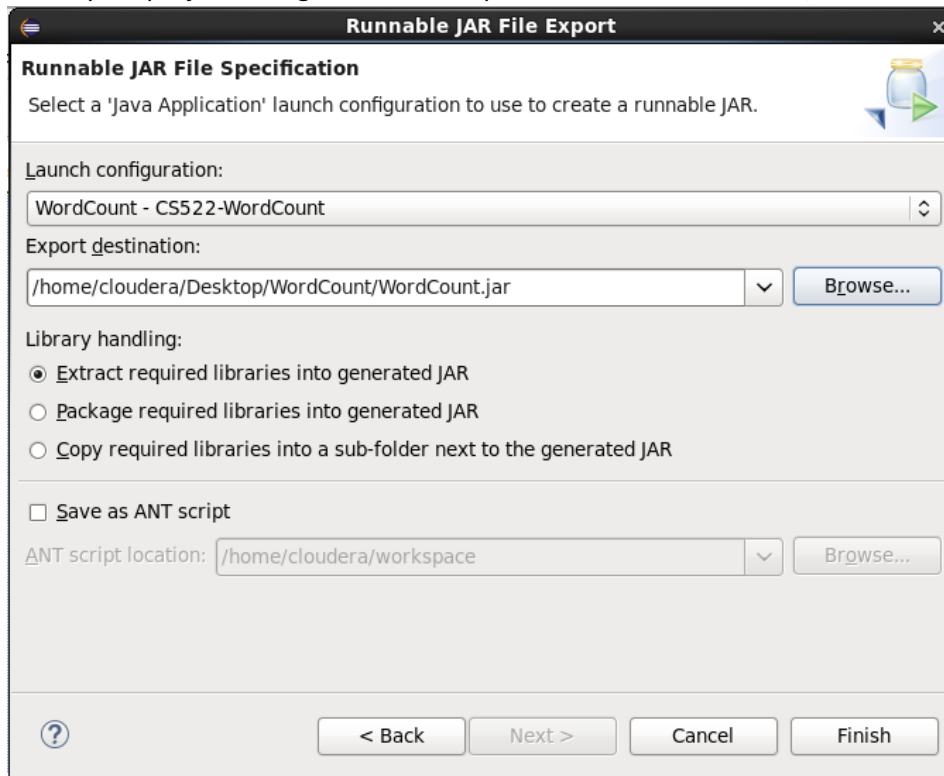
In Eclipse, right click on your class file -> Choose Export



Select Java -> Runnable JAR file



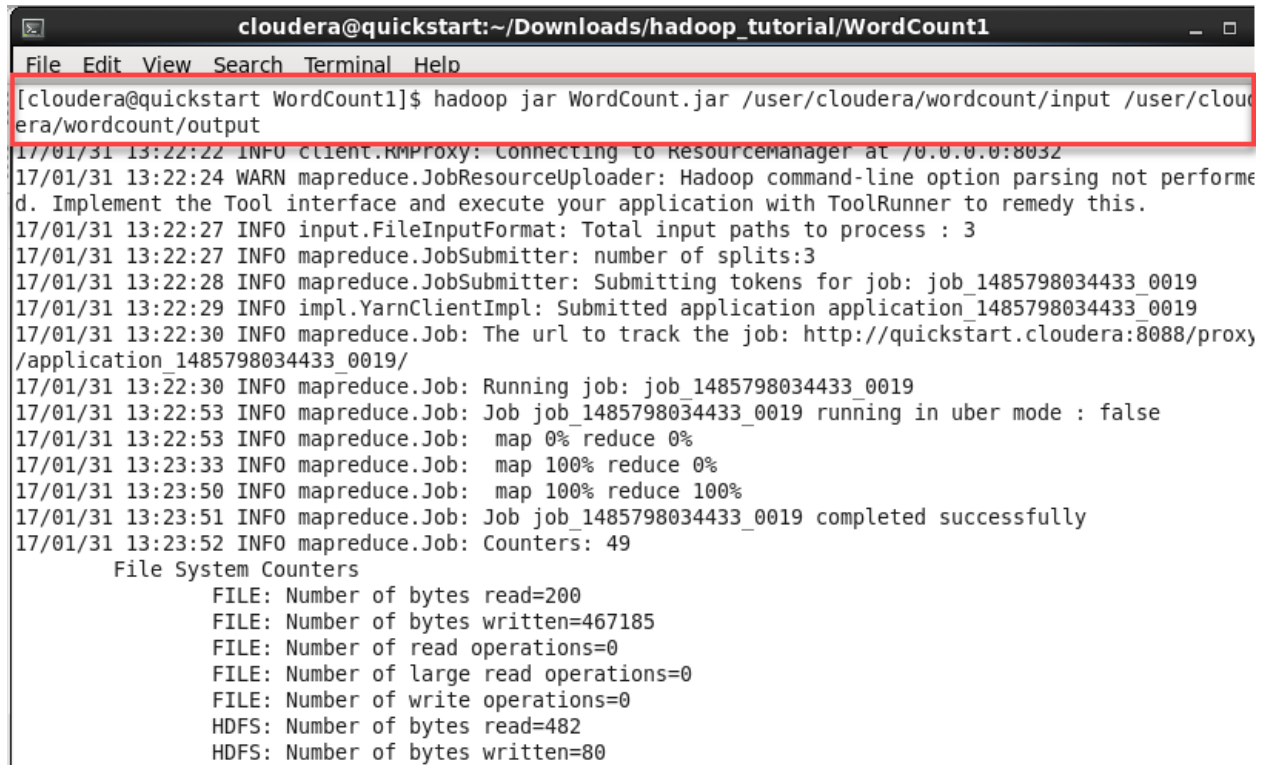
Select your project configuration and export destination. After that, click Finish.



To run the WordCount application from JAR file, passing the paths to the input and output directories in HDFS.

Run command

```
$ hadoop jar WordCount.jar /user/cloudera/wordcount/input  
/user/cloudera/wordcount/output
```



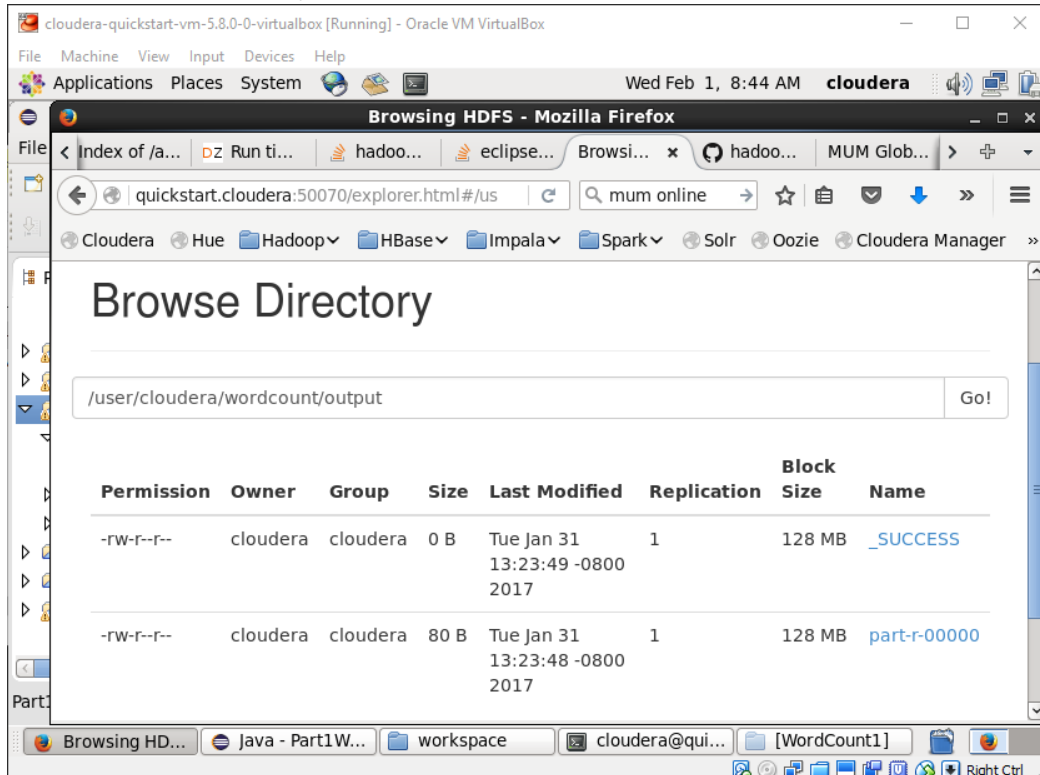
```
cloudera@quickstart:~/Downloads/hadoop_tutorial/WordCount1
File Edit View Search Terminal Help
[cloudera@quickstart WordCount1]$ hadoop jar WordCount.jar /user/cloudera/wordcount/input /user/cloudera/wordcount/output
17/01/31 13:22:22 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
17/01/31 13:22:24 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
17/01/31 13:22:27 INFO input.FileInputFormat: Total input paths to process : 3
17/01/31 13:22:27 INFO mapreduce.JobSubmitter: number of splits:3
17/01/31 13:22:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1485798034433_0019
17/01/31 13:22:29 INFO impl.YarnClientImpl: Submitted application application_1485798034433_0019
17/01/31 13:22:30 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1485798034433_0019/
17/01/31 13:22:30 INFO mapreduce.Job: Running job: job_1485798034433_0019
17/01/31 13:22:53 INFO mapreduce.Job: Job job_1485798034433_0019 running in uber mode : false
17/01/31 13:22:53 INFO mapreduce.Job:  map 0% reduce 0%
17/01/31 13:23:33 INFO mapreduce.Job:  map 100% reduce 0%
17/01/31 13:23:50 INFO mapreduce.Job:  map 100% reduce 100%
17/01/31 13:23:51 INFO mapreduce.Job: Job job_1485798034433_0019 completed successfully
17/01/31 13:23:52 INFO mapreduce.Job: Counters: 49
    File System Counters
      FILE: Number of bytes read=200
      FILE: Number of bytes written=467185
      FILE: Number of read operations=0
      FILE: Number of large read operations=0
      FILE: Number of write operations=0
      HDFS: Number of bytes read=482
      HDFS: Number of bytes written=80
```



You can see some output result as below

```
cloudera@quickstart: ~/Downloads/hadoop_tutorial/WordCount1
File Edit View Search Terminal Help
17/01/31 13:23:51 INFO mapreduce.Job: Job job_1485798034433_0019 completed successfully
17/01/31 13:23:52 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=200
    FILE: Number of bytes written=467185
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=482
    HDFS: Number of bytes written=80
    HDFS: Number of read operations=12
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2
  Job Counters
    Launched map tasks=3
    Launched reduce tasks=1
    Data-local map tasks=3
    Total time spent by all maps in occupied slots (ms)=110973
    Total time spent by all reduces in occupied slots (ms)=14377
    Total time spent by all map tasks (ms)=110973
    Total time spent by all reduce tasks (ms)=14377
    Total vcore-seconds taken by all map tasks=110973
    Total vcore-seconds taken by all reduce tasks=14377
    Total megabyte-seconds taken by all map tasks=113636352
    Total megabyte-seconds taken by all reduce tasks=14722048
Map-Reduce Framework
```

You also can browse output file on the web browser



To download the output file to text file run below command

```
cloudera@quickstart:~/Desktop/Part1-WordCount
File Edit View Search Terminal Help
Reduce output records=12
Spilled Records=36
Shuffled Maps =3
Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=4689
CPU time spent (ms)=6070
Physical memory (bytes) snapshot=730394624
Virtual memory (bytes) snapshot=6007341056
Total committed heap usage (bytes)=557592576
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=86
File Output Format Counters
Bytes Written=80
[cloudera@quickstart Part1-WordCount]$ hadoop fs -get /user/cloudera/wordcount/output/part-r-00000 output.txt
[cloudera@quickstart Part1-WordCount]$
```

Output.txt

```
cloudera@quickstart:~/Desktop/Part1-WordCount
File Edit View Search Terminal Help
Reduce output records=12
Spilled Records=36
Shuffled Maps =3
Failed Shuffles=0
Merged Map outputs=3
GC time elapsed (ms)=4689
CPU time spent (ms)=6070
Physical memory (bytes) snapshot=730394624
Virtual memory (bytes) snapshot=6007341056
Total committed heap usage (bytes)=557592576
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=86
File Output Format Counters
Bytes Written=80
[cloudera@quickstart Part1-WordCount]$ hadoop fs -get /user/cloudera/wordcount/output/part-r-00000 output.txt
[cloudera@quickstart Part1-WordCount]$
```

To clear the output directory run below command

```
[cloudera@quickstart Part1-WordCount]$ hadoop fs -rm -r /user/cloudera/wordcount/output  
Deleted /user/cloudera/wordcount/output
```