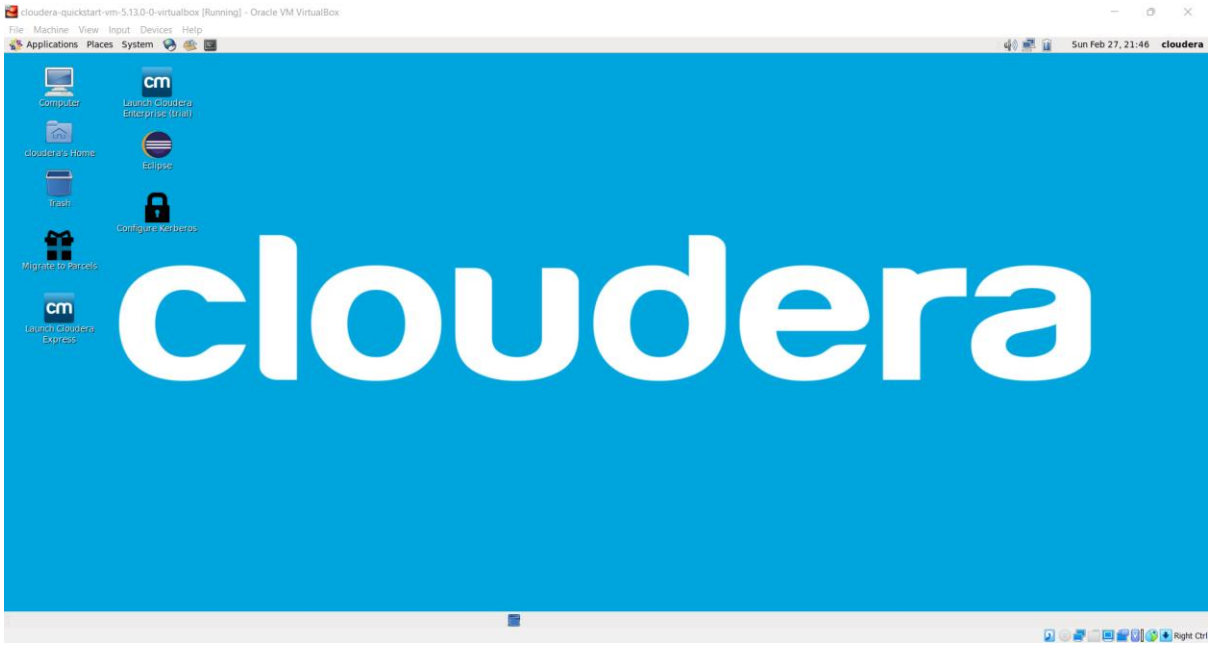
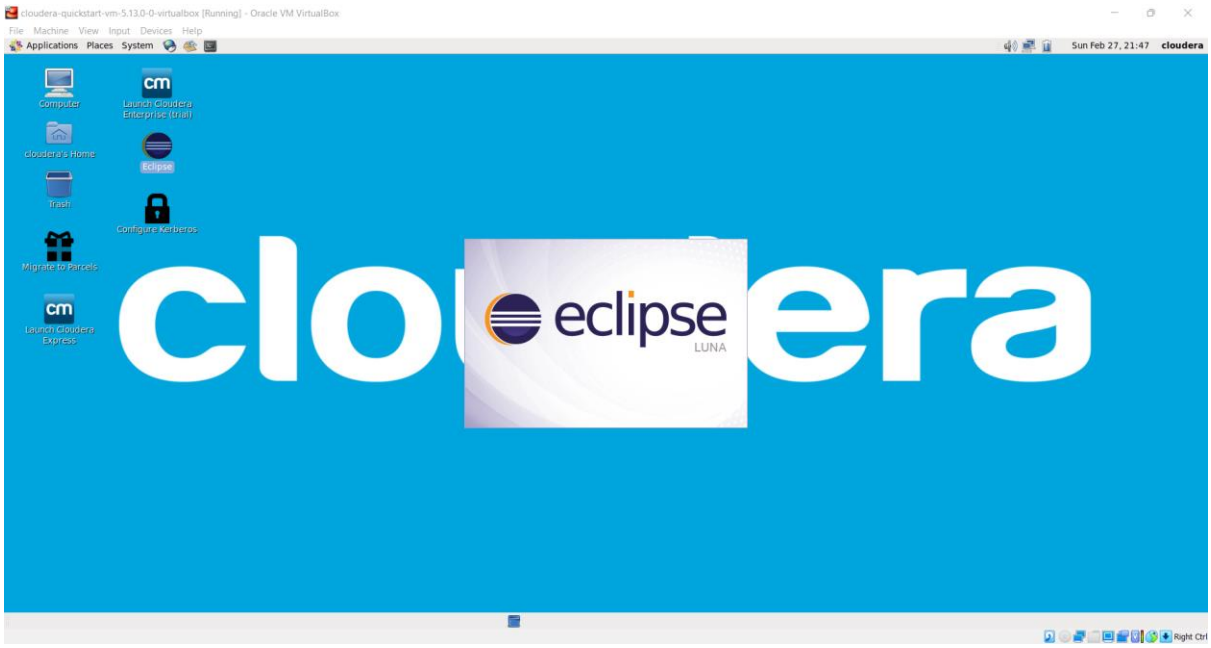
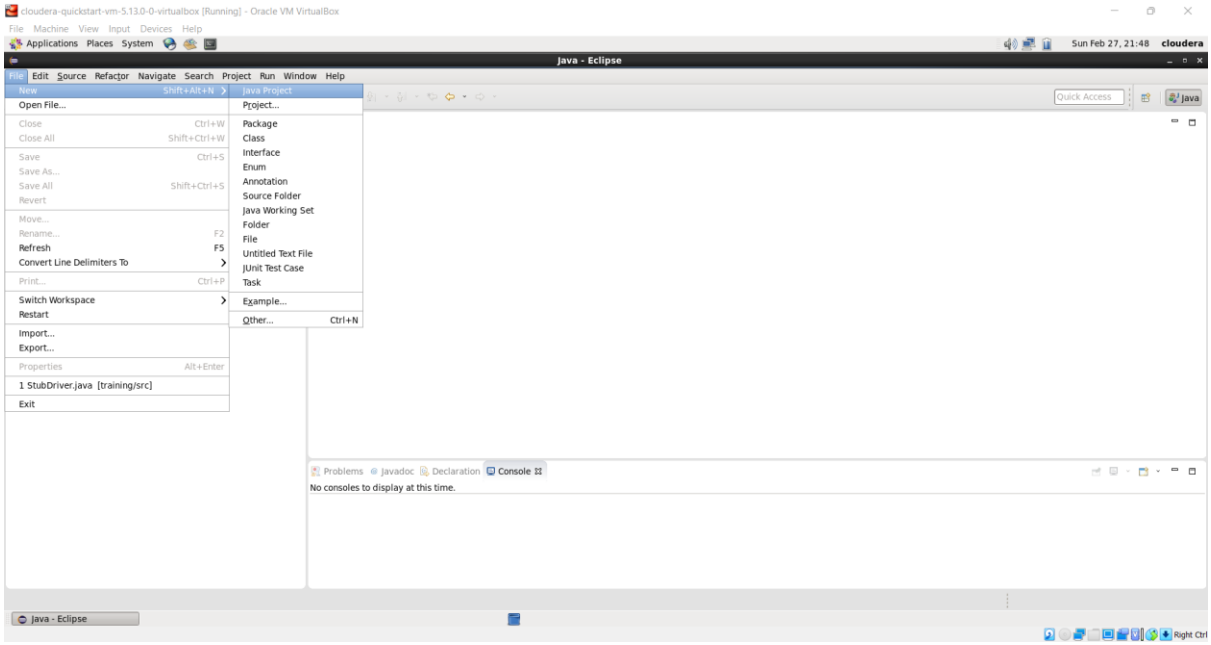
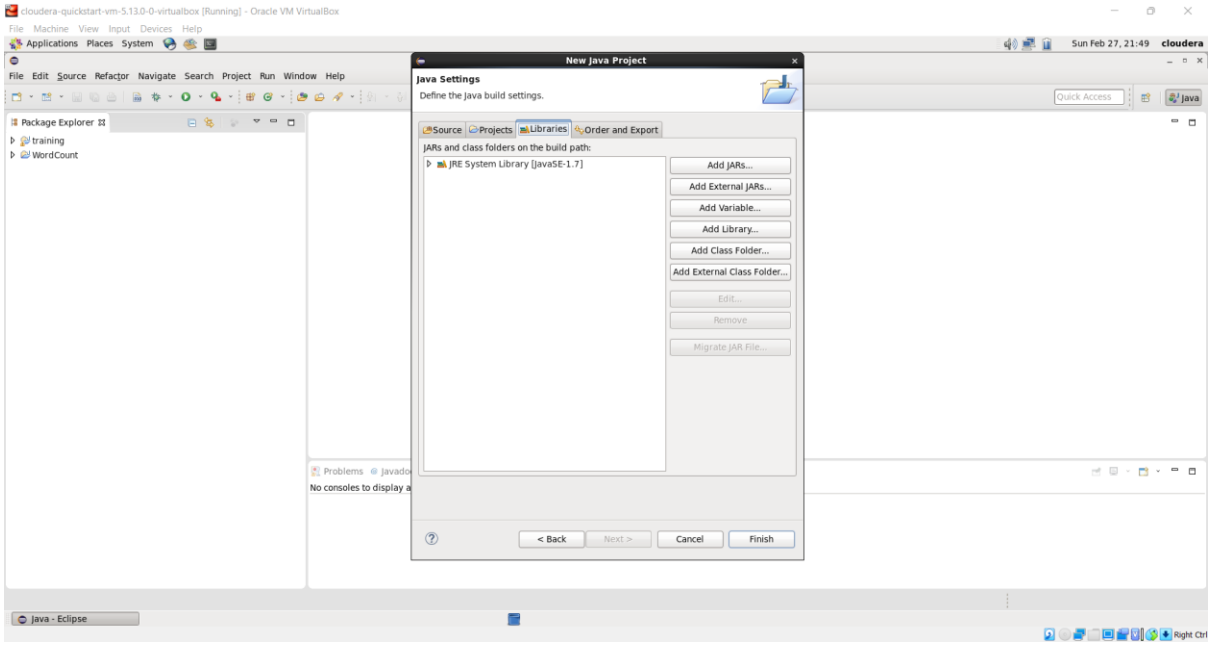
Practical 3
To Implement WordCount problem using Hadoop MapReduce in Eclipse: (Without Combiner & With Combiner)

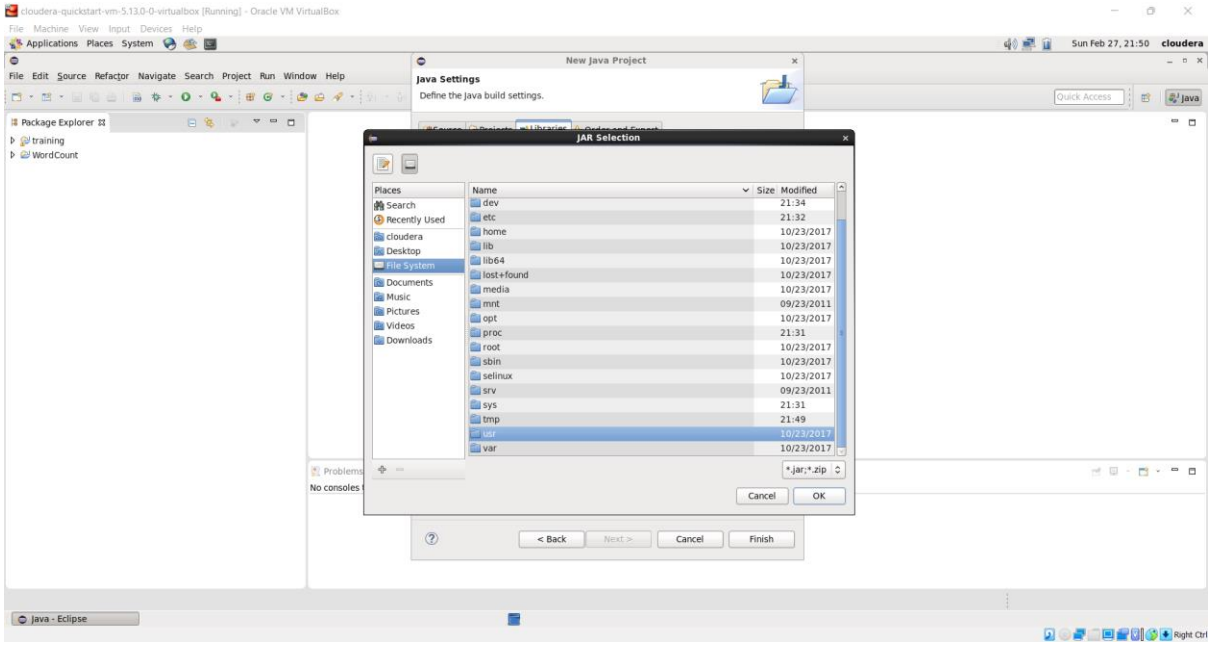
Word Count in Cloudera: (Without Combiner):

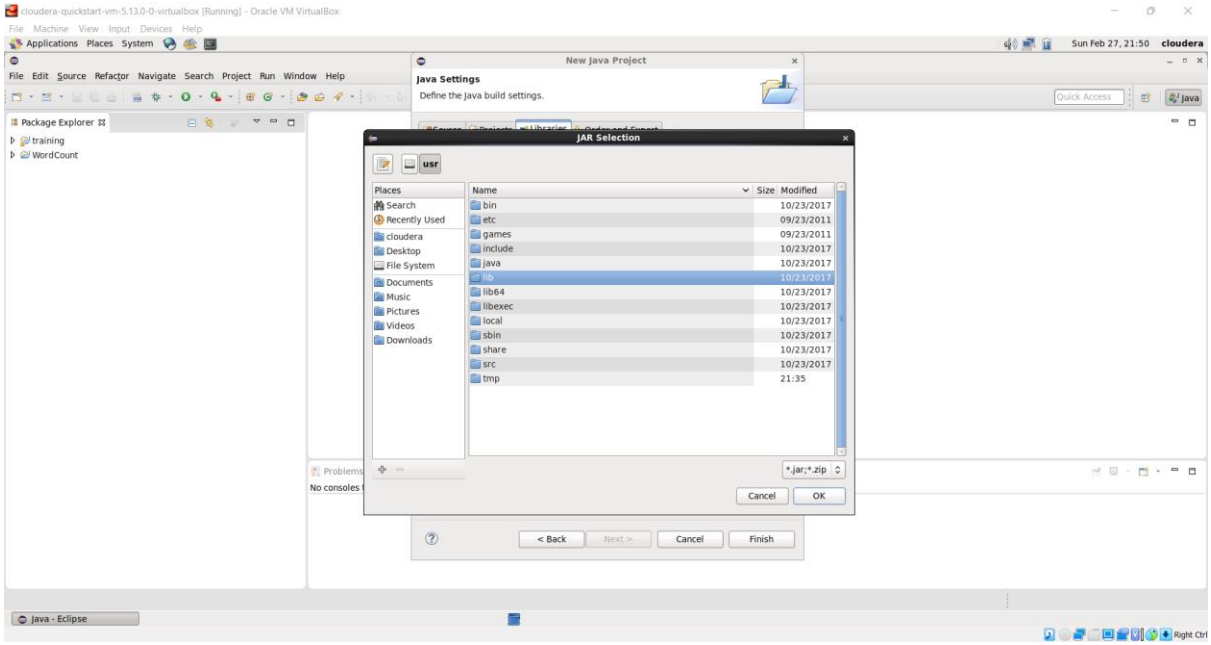


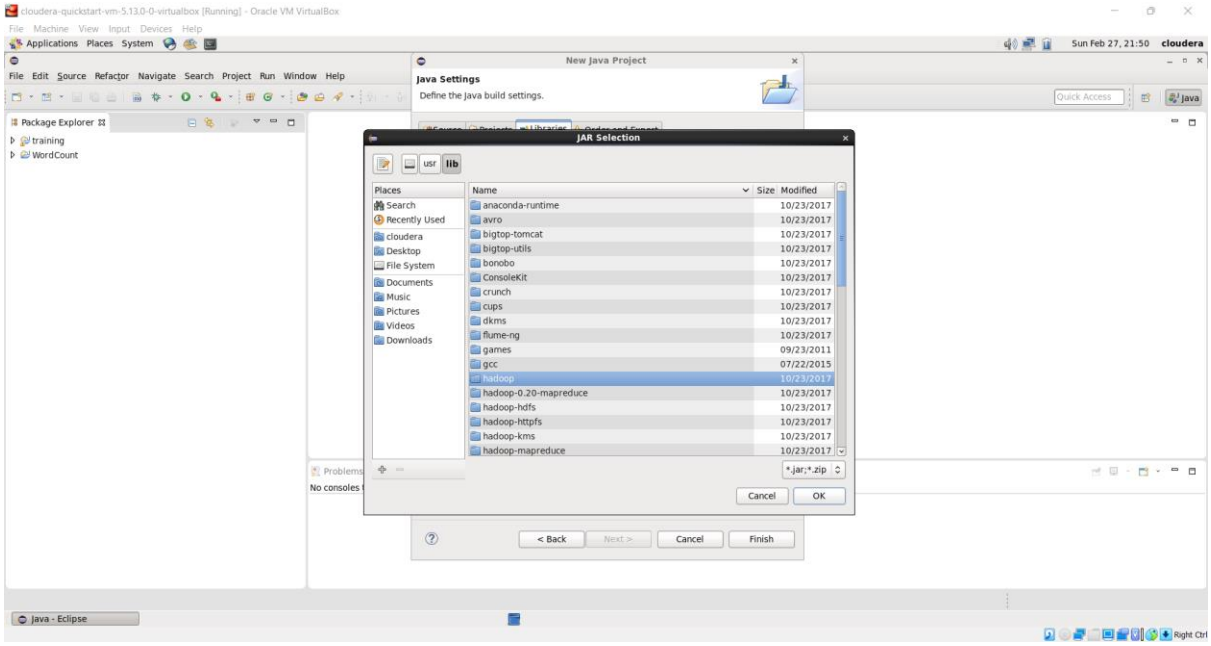


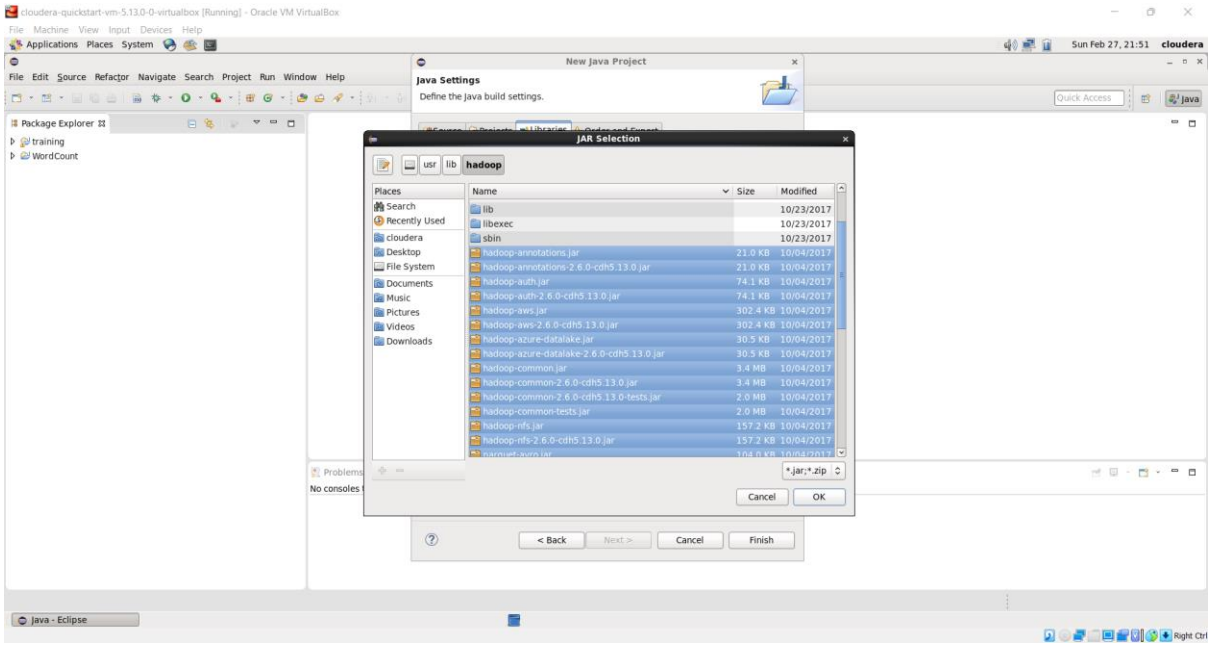

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:48 cloudera
Java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
New Shift+Alt+N
Open File...
Close Ctrl+W
Close All Shift+Ctrl+W
Save Ctrl+S
Save As...
Save All Shift+Ctrl+S
Revert
Move...
Rename... F2
Refresh F5
Convert Line Delimiters To
Print... Ctrl+P
Switch Workspace
Restart
Import...
Export...
Properties Alt+Enter
1 StubDriver.java [training/src]
Exit
Java Project
Project...
Package
Class
Interface
Enum
Annotation
Source Folder
Java Working Set
Folder
File
Untitled Text File
JUnit Test Case
Task
Example...
Other... Ctrl+N
Quick Access
Java
Problems Javadoc Declaration Console
No consoles to display at this time.
Java - Eclipse
Right Ctrl

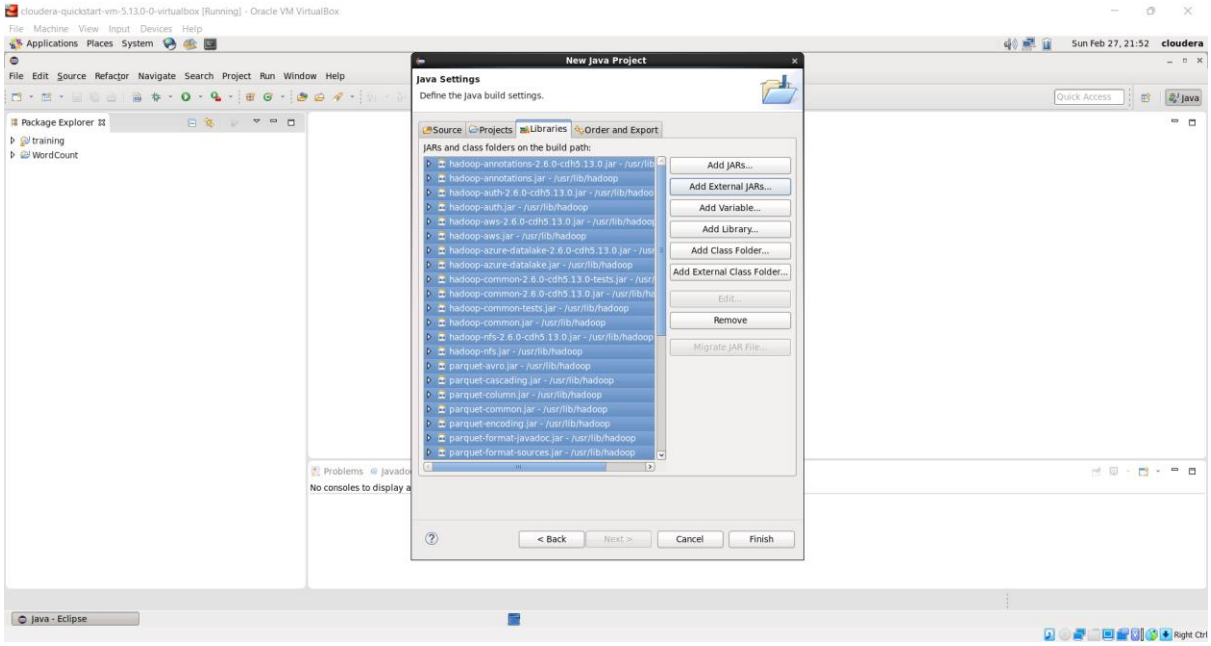

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:49 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
New Java Project
Java Settings
Define the Java build settings.
Source Projects Libraries Order and Export
JARs and class folders on the build path:
JRE System Library [JavaSE-1.7]
Add JARs...
Add External JARs...
Add Variable...
Add Library...
Add Class Folder...
Add External Class Folder...
Edit...
Remove
Migrate JAR File...
< Back
Next >
Cancel
Finish
Quick Access
Java
Problems
No consoles to display a
Java - Eclipse
Right Ctrl

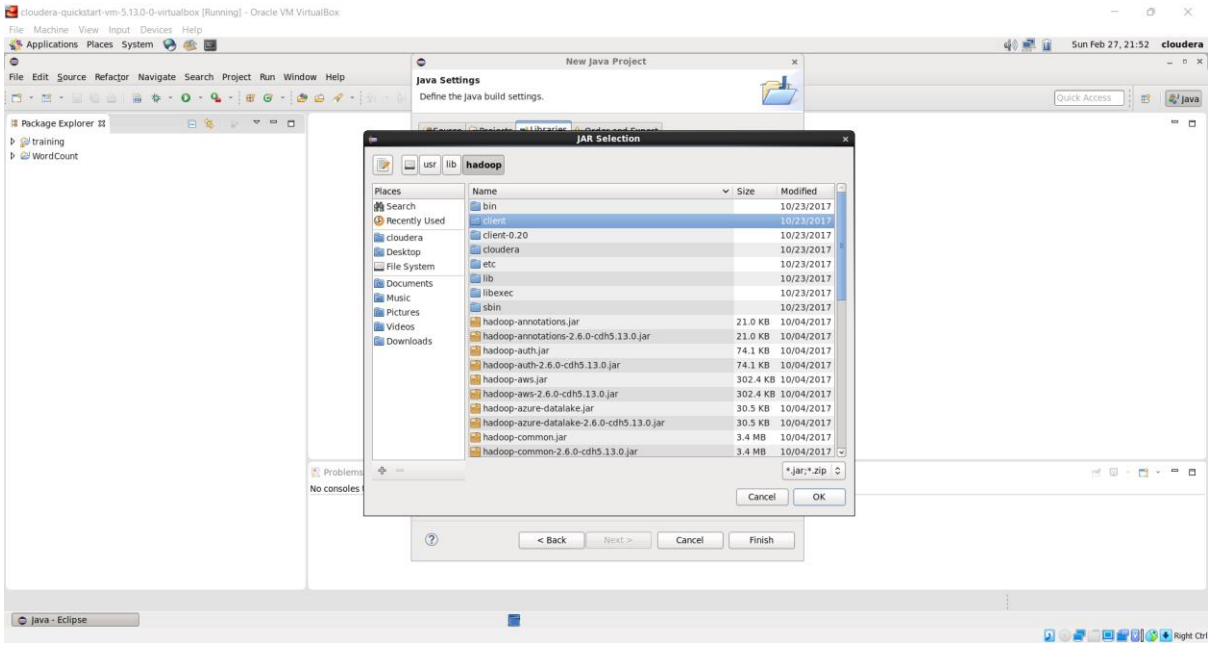

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:50 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
New Java Project
Java Settings
Define the Java build settings.
Quick Access
Java
Package Explorer
training
WordCount
JAR Selection
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name
Size
Modified
dev 21:34
etc 21:32
home 10/23/2017
lib 10/23/2017
lib64 10/23/2017
lost+found 10/23/2017
media 10/23/2017
mnt 09/23/2011
opt 10/23/2017
proc 21:31
root 10/23/2017
sbin 10/23/2017
selinux 10/23/2017
srv 09/23/2011
sys 21:31
tmp 21:49
usr 10/23/2017
var 10/23/2017
.jar;.zip
Cancel
OK
< Back
Next >
Cancel
Finish
Problems
No consoles
Java - Eclipse
Right Ctrl

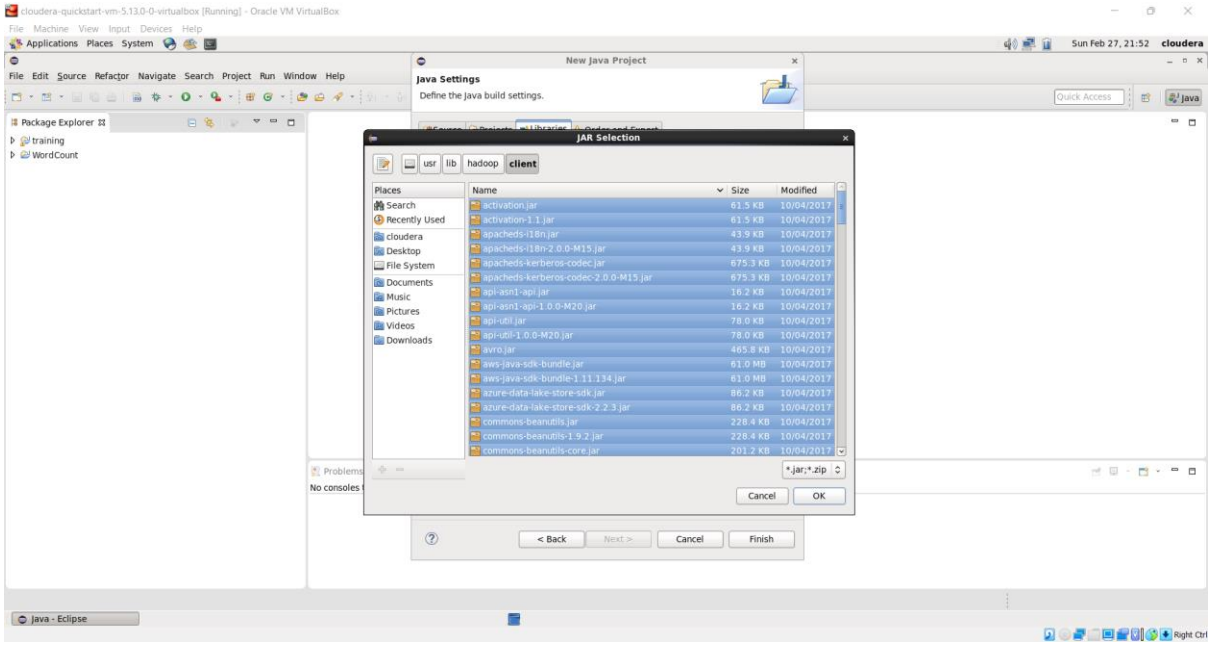

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:50 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
New Java Project
Java Settings
Define the Java build settings.
Quick Access
Java
Package Explorer
training
WordCount
JAR Selection
usr
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name
Size
Modified
bin 10/23/2017
etc 09/23/2011
games 09/23/2011
include 10/23/2017
java 10/23/2017
lib 10/23/2017
lib64 10/23/2017
libexec 10/23/2017
local 10/23/2017
sbin 10/23/2017
share 10/23/2017
src 10/23/2017
tmp 21:35
.jar;.zip
Cancel
OK
< Back
Next >
Cancel
Finish
Problems
No consoles
Java - Eclipse
Right Ctrl

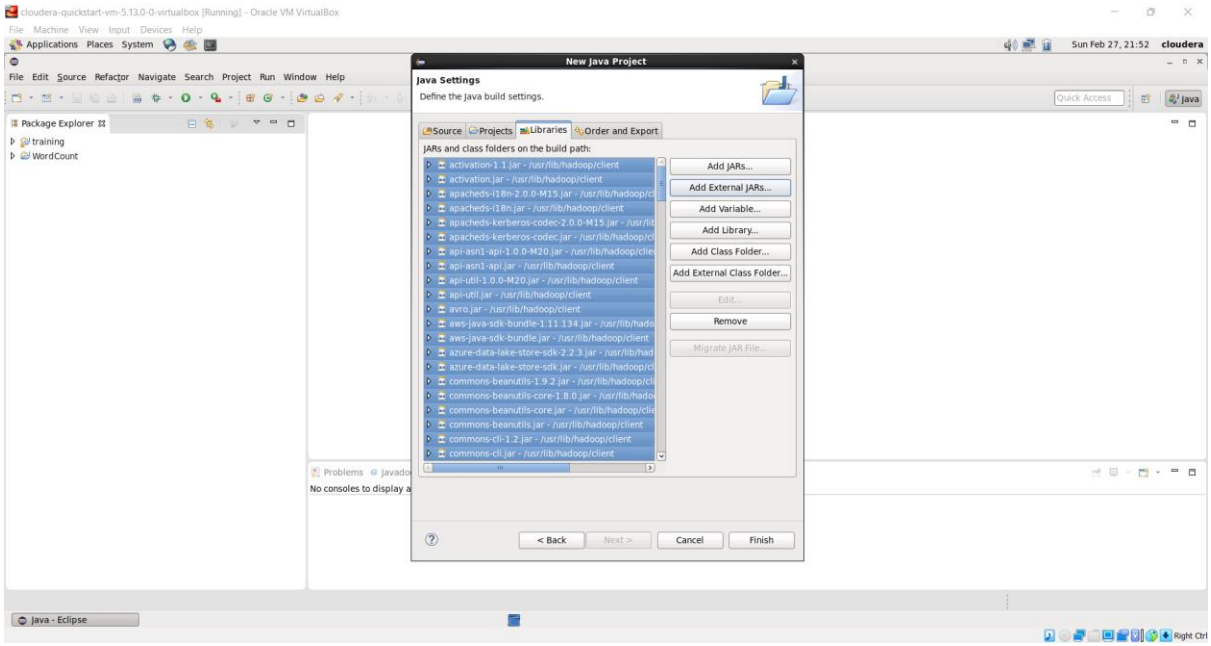

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 21:50 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
New Java Project
Java Settings
Define the Java build settings.
Package Explorer
training
WordCount
JAR Selection
usr lib
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name Size Modified
anaconda-runtime 10/23/2017
avro 10/23/2017
bigtop-tomcat 10/23/2017
bigtop-utils 10/23/2017
bonobo 10/23/2017
ConsoleKit 10/23/2017
crunch 10/23/2017
cups 10/23/2017
dkms 10/23/2017
flume-ng 10/23/2017
games 09/23/2011
gcc 07/22/2015
hadoop 10/23/2017
hadoop-0.20-mapreduce 10/23/2017
hadoop-hdfs 10/23/2017
hadoop-httpfs 10/23/2017
hadoop-kms 10/23/2017
hadoop-mapreduce 10/23/2017
.jar;.zip
Cancel OK
< Back Next > Cancel Finish
Java - Eclipse

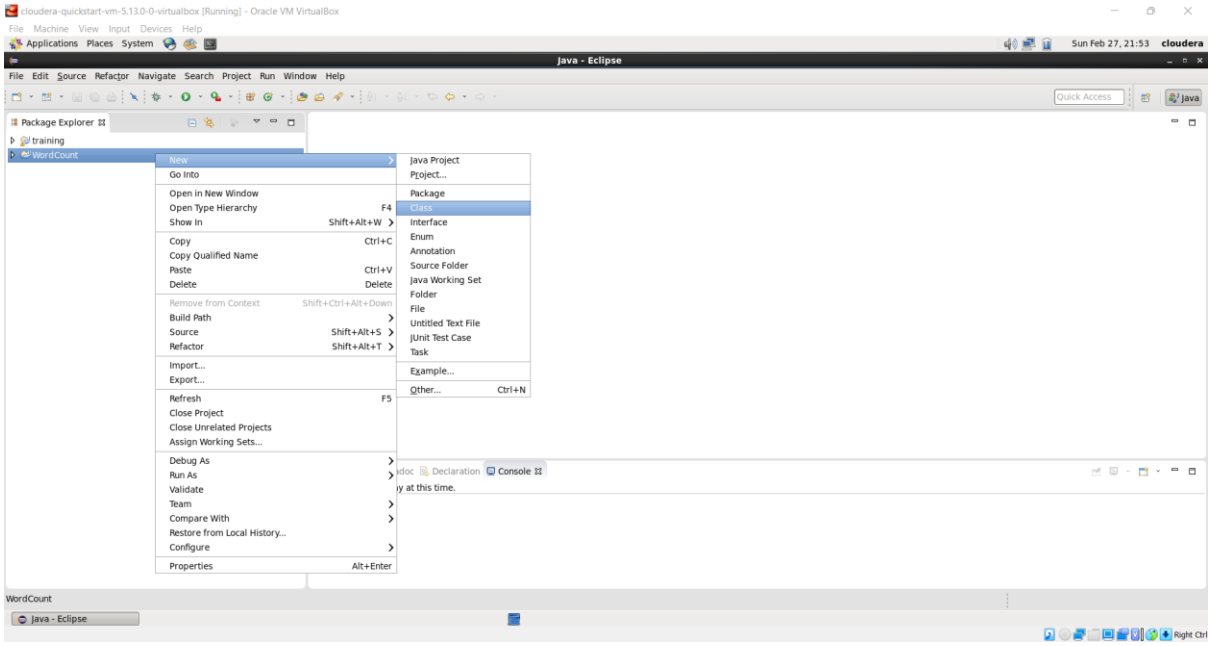

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 21:51 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
New Java Project
Java Settings
Define the Java build settings.
Package Explorer
training
WordCount
JAR Selection
usr lib hadoop
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name Size Modified
lib 10/23/2017
libexec 10/23/2017
sbin 10/23/2017
hadoop-annotations.jar 21.0 KB 10/04/2017
hadoop-annotations-2.6.0-cdh5.13.0.jar 21.0 KB 10/04/2017
hadoop-auth.jar 74.1 KB 10/04/2017
hadoop-auth-2.6.0-cdh5.13.0.jar 74.1 KB 10/04/2017
hadoop-aws.jar 302.4 KB 10/04/2017
hadoop-aws-2.6.0-cdh5.13.0.jar 302.4 KB 10/04/2017
hadoop-azure-datalake.jar 30.5 KB 10/04/2017
hadoop-azure-datalake-2.6.0-cdh5.13.0.jar 30.5 KB 10/04/2017
hadoop-common.jar 3.4 MB 10/04/2017
hadoop-common-2.6.0-cdh5.13.0.jar 3.4 MB 10/04/2017
hadoop-common-2.6.0-cdh5.13.0-tests.jar 2.0 MB 10/04/2017
hadoop-common-tests.jar 2.0 MB 10/04/2017
hadoop-nfs.jar 157.2 KB 10/04/2017
hadoop-nfs-2.6.0-cdh5.13.0.jar 157.2 KB 10/04/2017
.jar;.zip
Cancel OK
< Back Next > Cancel Finish
Java - Eclipse

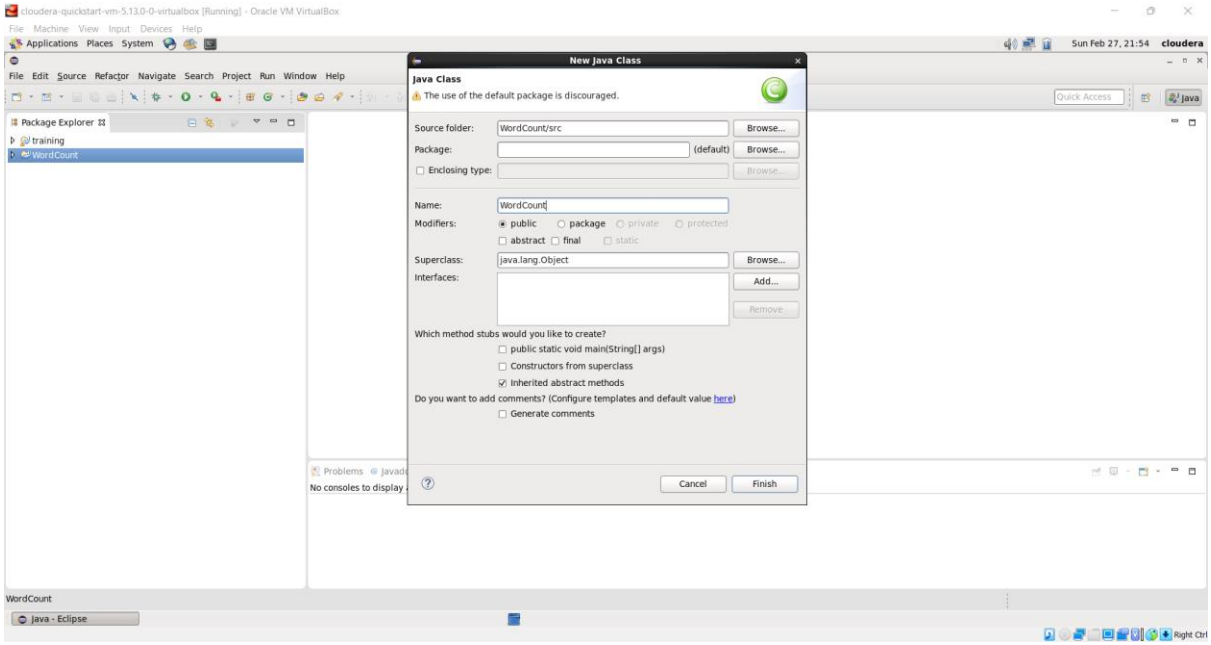

New Java Project
Java Settings
Define the Java build settings.
Source
Projects
Libraries
Order and Export
JARs and class folders on the build path:
Add JARs...
Add External JARs...
Add Variable...
Add Library...
Add Class Folder...
Add External Class Folder...
Edit...
Remove
Migrate JAR File...
< Back
Next >
Cancel
Finish

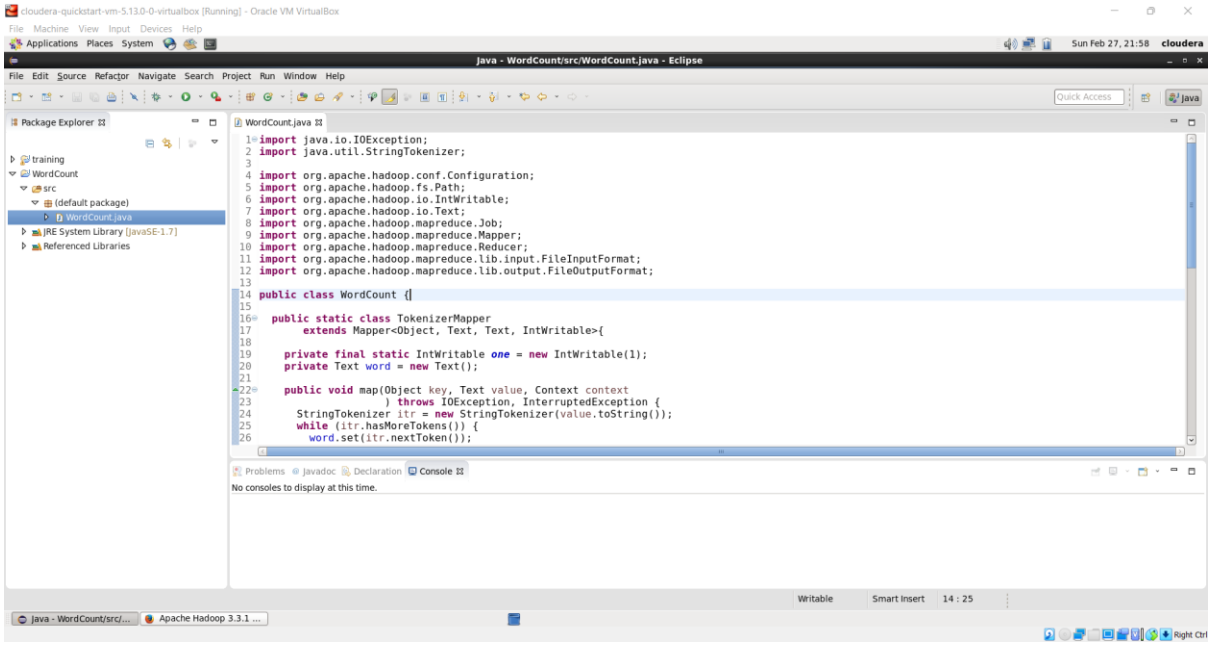

New Java Project
Java Settings
Define the Java build settings.
JAR Selection
usr
lib
hadoop
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name
Size
Modified
bin
client
client-0.20
cloudera
etc
lib
libexec
sbin
hadoop-annotations.jar
hadoop-annotations-2.6.0-cdh5.13.0.jar
hadoop-auth.jar
hadoop-auth-2.6.0-cdh5.13.0.jar
hadoop-aws.jar
hadoop-aws-2.6.0-cdh5.13.0.jar
hadoop-azure-datalake.jar
hadoop-azure-datalake-2.6.0-cdh5.13.0.jar
hadoop-common.jar
hadoop-common-2.6.0-cdh5.13.0.jar
.jar;.zip
Cancel
OK
< Back
Next >
Cancel
Finish

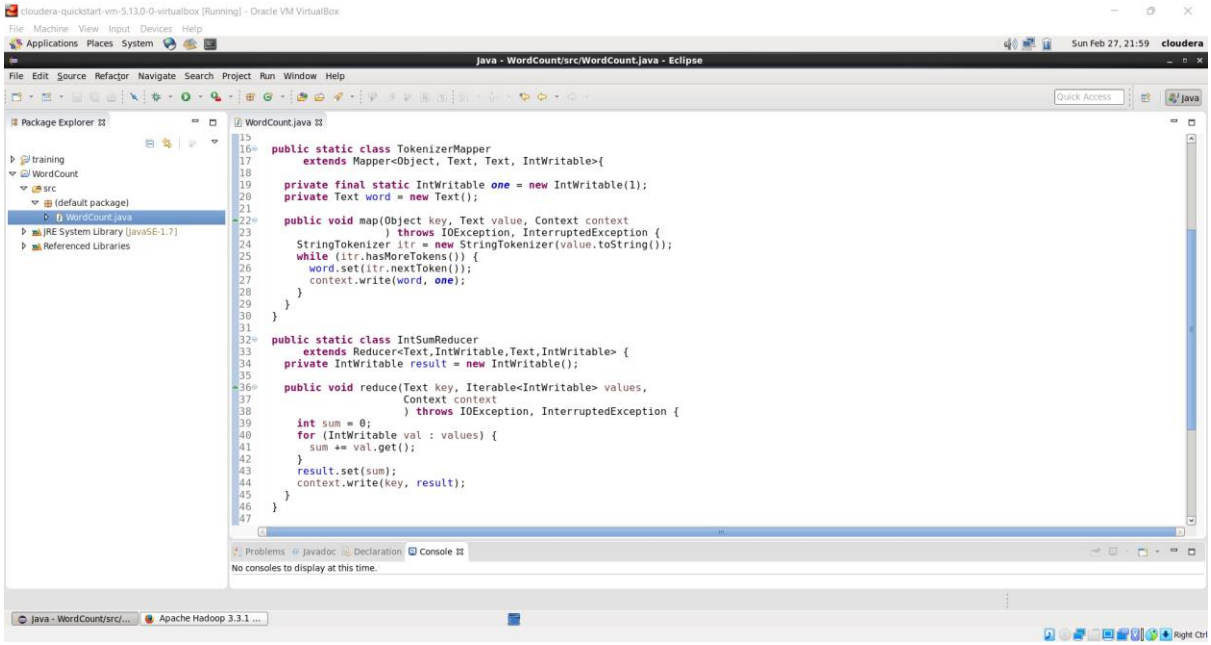

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:52 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
Java
Package Explorer
training
WordCount
New Java Project
Java Settings
Define the Java build settings.
JAR Selection
usr lib hadoop client
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name Size Modified
activation.jar 61.5 KB 10/04/2017
activation-1.1.jar 61.5 KB 10/04/2017
apacheds-i18n.jar 43.9 KB 10/04/2017
apacheds-i18n-2.0.0-M15.jar 43.9 KB 10/04/2017
apacheds-kerberos-codec.jar 675.3 KB 10/04/2017
apacheds-kerberos-codec-2.0.0-M15.jar 675.3 KB 10/04/2017
api-asn1-api.jar 16.2 KB 10/04/2017
api-asn1-api-1.0.0-M20.jar 16.2 KB 10/04/2017
api-util.jar 78.0 KB 10/04/2017
api-util-1.0.0-M20.jar 78.0 KB 10/04/2017
avro.jar 465.8 KB 10/04/2017
aws-java-sdk-bundle.jar 61.0 MB 10/04/2017
aws-java-sdk-bundle-1.11.134.jar 61.0 MB 10/04/2017
azure-data-lake-store-sdk.jar 86.2 KB 10/04/2017
azure-data-lake-store-sdk-2.2.3.jar 86.2 KB 10/04/2017
commons-beanutils.jar 228.4 KB 10/04/2017
commons-beanutils-1.9.2.jar 228.4 KB 10/04/2017
commons-beanutils-core.jar 201.2 KB 10/04/2017
.jar;.zip
Cancel
OK
Problems
No consoles
< Back
Next >
Cancel
Finish
Java - Eclipse
Right Ctrl

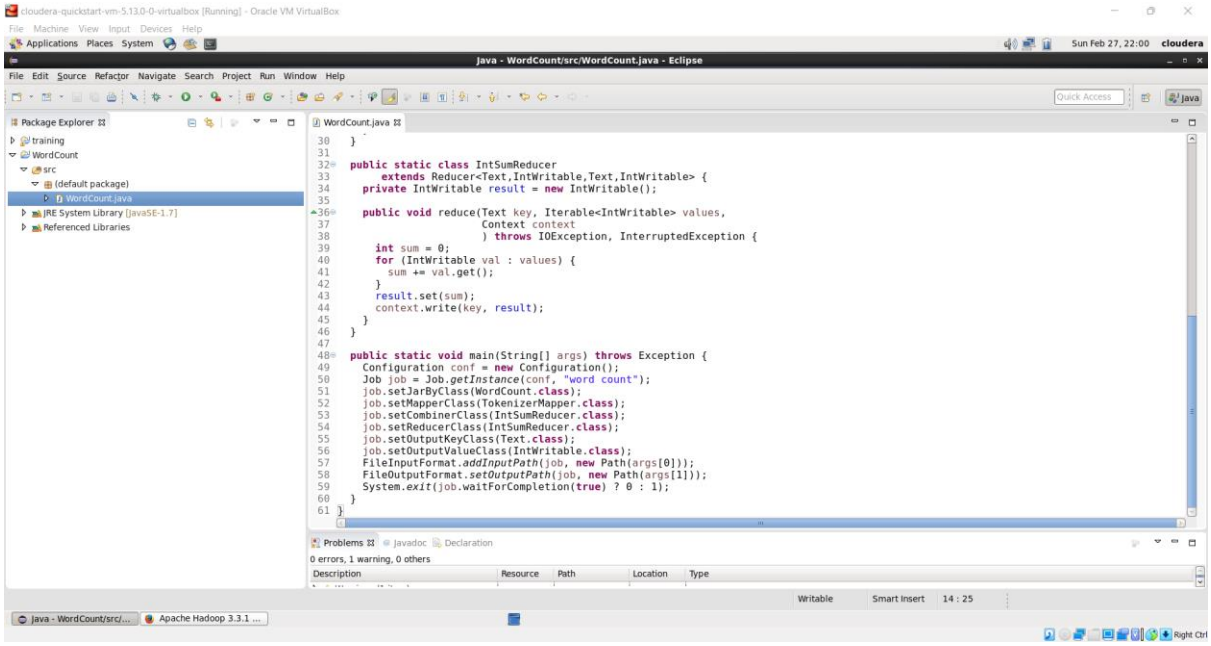

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:52 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
Java
Package Explorer
training
WordCount
New Java Project
Java Settings
Define the Java build settings.
Source
Projects
Libraries
Order and Export
JARs and class folders on the build path:
activation-1.1.jar - /usr/lib/hadoop/client
activation.jar - /usr/lib/hadoop/client
apacheds-i18n-2.0.0-M15.jar - /usr/lib/hadoop/cl
apacheds-i18n.jar - /usr/lib/hadoop/client
apacheds-kerberos-codec-2.0.0-M15.jar - /usr/lib
apacheds-kerberos-codec.jar - /usr/lib/hadoop/cl
api-asn1-api-1.0.0-M20.jar - /usr/lib/hadoop/clie
api-asn1-api.jar - /usr/lib/hadoop/client
api-util-1.0.0-M20.jar - /usr/lib/hadoop/client
api-util.jar - /usr/lib/hadoop/client
avro.jar - /usr/lib/hadoop/client
aws-java-sdk-bundle-1.11.134.jar - /usr/lib/hado
aws-java-sdk-bundle.jar - /usr/lib/hadoop/client
azure-data-lake-store-sdk-2.2.3.jar - /usr/lib/had
azure-data-lake-store-sdk.jar - /usr/lib/hadoop/cl
commons-beanutils-1.9.2.jar - /usr/lib/hadoop/cl
commons-beanutils-core-1.8.0.jar - /usr/lib/hado
commons-beanutils-core.jar - /usr/lib/hadoop/clie
commons-beanutils.jar - /usr/lib/hadoop/client
commons-cli-1.2.jar - /usr/lib/hadoop/client
commons-cli.jar - /usr/lib/hadoop/client
Add JARs...
Add External JARs...
Add Variable...
Add Library...
Add Class Folder...
Add External Class Folder...
Edit...
Remove
Migrate JAR File...
Problems
No consoles to display a
< Back
Next >
Cancel
Finish
Java - Eclipse
Right Ctrl

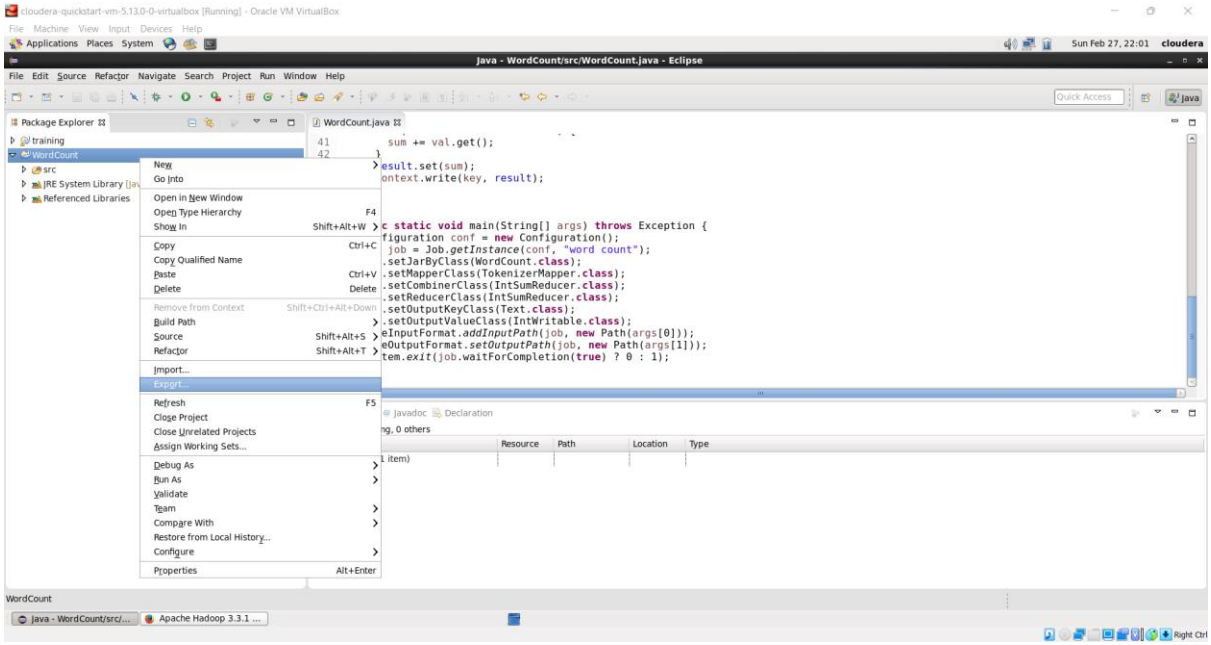

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:53 cloudera
Java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
Java
Package Explorer
training
WordCount
New
Go Into
Open in New Window
Open Type Hierarchy F4
Show In Shift+Alt+W
Copy Ctrl+C
Copy Qualified Name
Paste Ctrl+V
Delete Delete
Remove from Context Shift+Ctrl+Alt+Down
Build Path
Source Shift+Alt+S
Refactor Shift+Alt+T
Import...
Export...
Refresh F5
Close Project
Close Unrelated Projects
Assign Working Sets...
Debug As
Run As
Validate
Team
Compare With
Restore from Local History...
Configure
Properties Alt+Enter
Java Project
Project...
Package
Class
Interface
Enum
Annotation
Source Folder
Java Working Set
Folder
File
Untitled Text File
JUnit Test Case
Task
Example...
Other... Ctrl+N
Declaration
Console
WordCount
Java - Eclipse

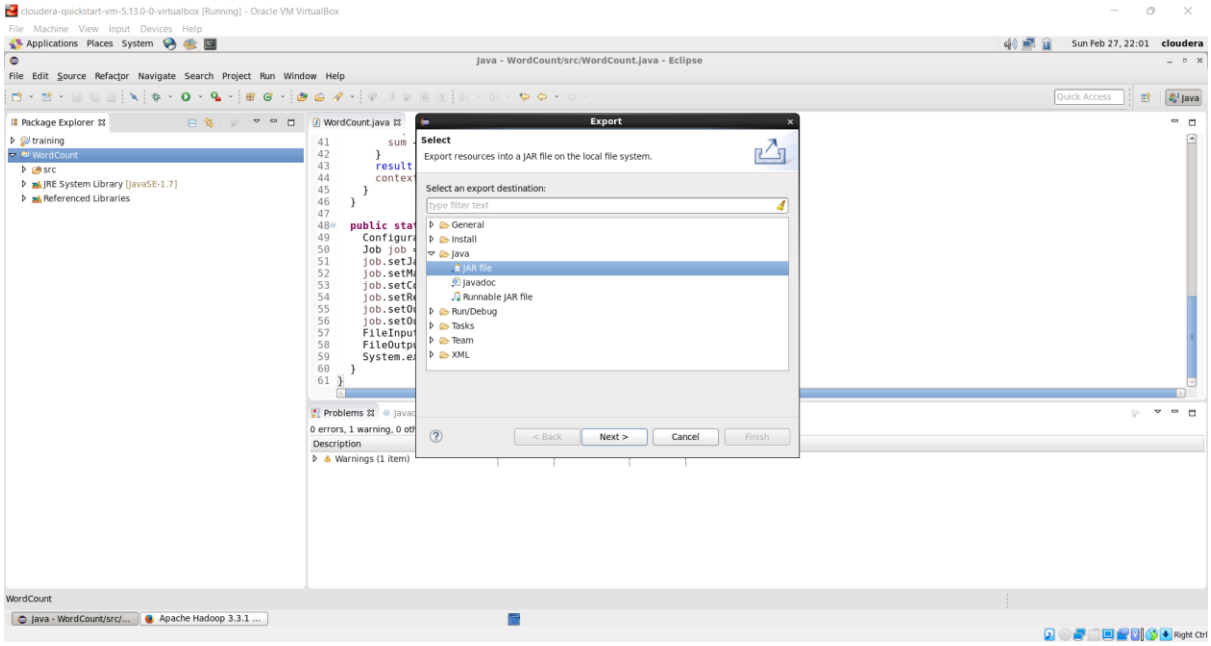

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Applications Places System
Sun Feb 27, 21:54 cloudera
File Edit Source Refactor Navigate Search Project Run Window Help
Quick Access
Java
Package Explorer
training
WordCount
New Java Class
Java Class
The use of the default package is discouraged.
Source folder: WordCount/src
Browse...
Package:
(default)
Enclosing type:
Name: WordCount
Modifiers: public package private protected
abstract final static
Superclass: java.lang.Object
Interfaces:
Add...
Remove
Which method stubs would you like to create?
public static void main(String[] args)
Constructors from superclass
Inherited abstract methods
Do you want to add comments? (Configure templates and default value here)
Generate comments
Cancel
Finish
Problems
No consoles to display
WordCount
Java - Eclipse

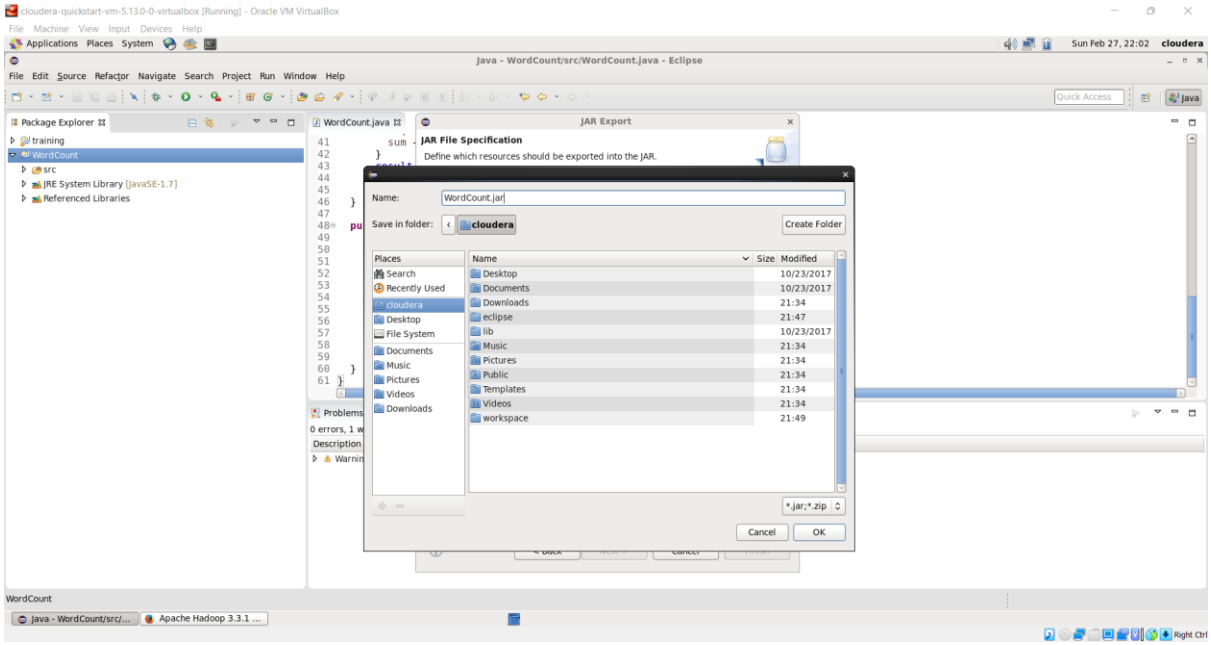

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 21:58 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
(default package)
WordCount.java
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
import java.io.IOException;
import java.util.StringTokenizer;
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;
public class WordCount {
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 {
StringTokenizer itr = new StringTokenizer(value.toString());
while (itr.hasMoreTokens()) {
word.set(itr.nextToken());
Problems Javadoc Declaration Console
No consoles to display at this time.
Writable Smart Insert 14 : 25
Java - WordCount/src/... Apache Hadoop 3.3.1 ...
Right Ctrl

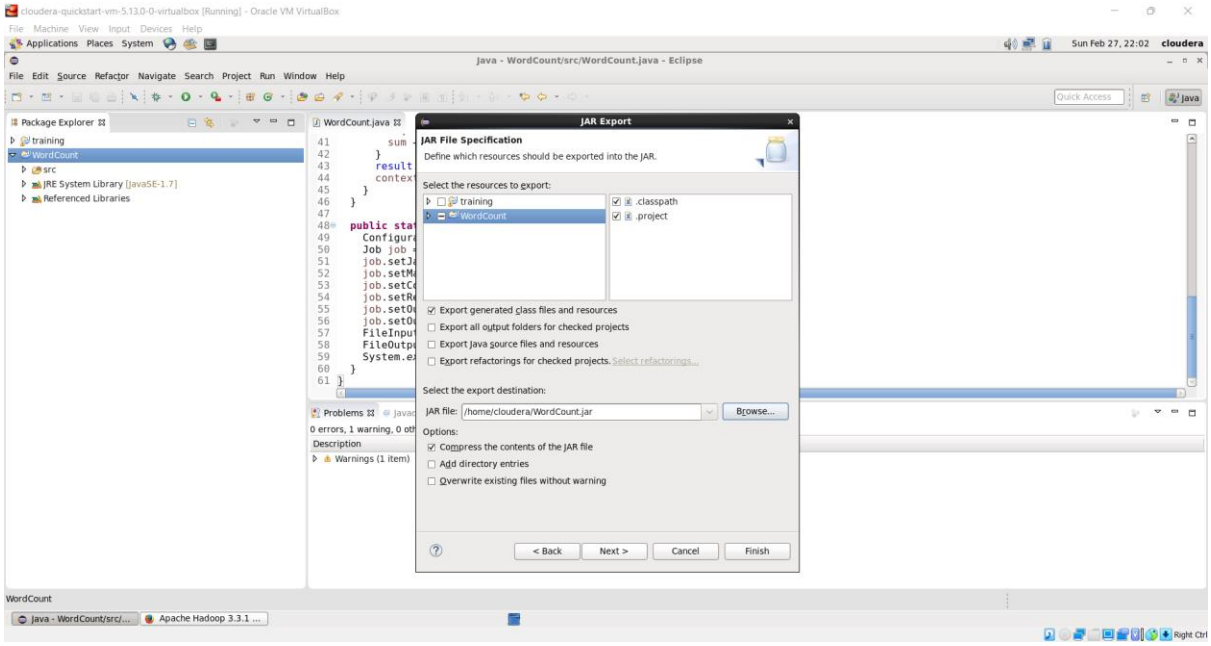

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 21:59 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
(default package)
WordCount.java
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
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 {
StringTokenizer itr = new StringTokenizer(value.toString());
while (itr.hasMoreTokens()) {
word.set(itr.nextToken());
context.write(word, one);
}
}
}
public static class IntSumReducer
extends Reducer<Text,IntWritable,Text,IntWritable> {
private IntWritable result = new IntWritable();
public void reduce(Text key, Iterable<IntWritable> values,
Context context
) throws IOException, InterruptedException {
int sum = 0;
for (IntWritable val : values) {
sum += val.get();
}
result.set(sum);
context.write(key, result);
}
}
Problems Javadoc Declaration Console
No consoles to display at this time.
Java - WordCount/src/... Apache Hadoop 3.3.1 ...
Right Ctrl

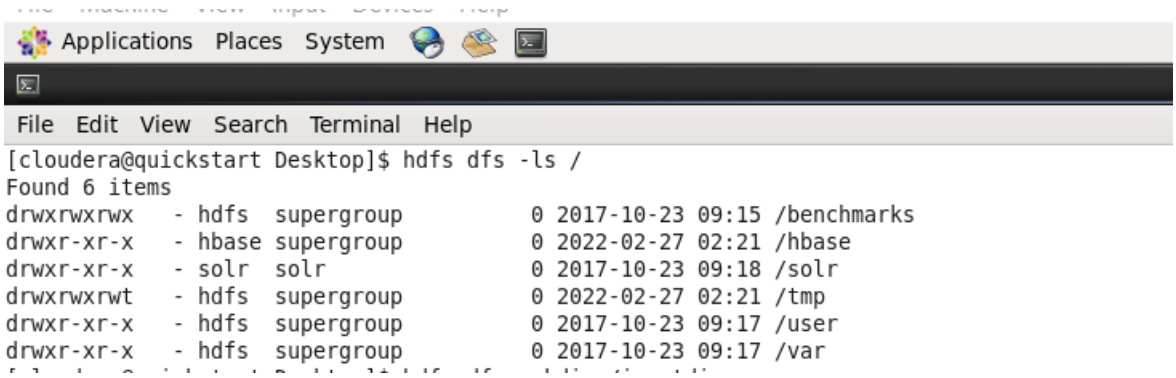

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
(default package)
WordCount.java
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
}
public static class IntSumReducer
extends Reducer<Text,IntWritable,Text,IntWritable> {
private IntWritable result = new IntWritable();
public void reduce(Text key, Iterable<IntWritable> values,
Context context
) throws IOException, InterruptedException {
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(WordCount.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);
}
}
Problems
Javadoc
Declaration
0 errors, 1 warning, 0 others
Description
Resource
Path
Location
Type
Writable
Smart Insert
14 : 25
Apache Hadoop 3.3.1 ...


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
JRE System Library
Referenced Libraries
New
Go Into
Open in New Window
Open Type Hierarchy F4
Show In Shift+Alt+W
Copy Ctrl+C
Copy Qualified Name
Paste Ctrl+V
Delete Delete
Remove from Context Shift+Ctrl+Alt+Down
Build Path
Source Shift+Alt+S
Refactor Shift+Alt+T
Import...
Export...
Refresh F5
Close Project
Close Unrelated Projects
Assign Working Sets...
Debug As
Run As
Validate
Team
Compare With
Restore from Local History...
Configure
Properties Alt+Enter
WordCount
Apache Hadoop 3.3.1 ...


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 22:01 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
Export
Select
Export resources into a JAR file on the local file system.
Select an export destination:
type filter text
General
Install
Java
JAR file
Javadoc
Runnable JAR file
Run/Debug
Tasks
Team
XML
< Back
Next >
Cancel
Finish
Problems
Description
Warnings (1 item)
WordCount
Apache Hadoop 3.3.1 ...


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 22:02 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Name:
WordCount.jar
Save in folder:
cloudera
Create Folder
Places
Search
Recently Used
cloudera
Desktop
File System
Documents
Music
Pictures
Videos
Downloads
Name
Size
Modified
Desktop 10/23/2017
Documents 10/23/2017
Downloads 21:34
eclipse 21:47
lib 10/23/2017
Music 21:34
Pictures 21:34
Public 21:34
Templates 21:34
Videos 21:34
workspace 21:49
.jar;.zip
Cancel
OK
Problems
Description
WordCount
Apache Hadoop 3.3.1 ...


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 22:02 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Select the resources to export:
training
WordCount
.classpath
.project
Export generated class files and resources
Export all output folders for checked projects
Export Java source files and resources
Export refactorings for checked projects. Select refactorings...
Select the export destination:
JAR file: /home/cloudera/WordCount.jar
Browse...
Options:
Compress the contents of the JAR file
Add directory entries
Overwrite existing files without warning
< Back
Next >
Cancel
Finish
Problems
0 errors, 1 warning, 0 oth
Description
Warnings (1 item)
WordCount
Java - WordCount/src/...
Apache Hadoop 3.3.1 ...


Applications Places System
File Edit View Search Terminal Help
[cloudera@quickstart Desktop]$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2022-02-27 02:21 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2022-02-27 02:21 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var

```
[cloudera@quickstart Desktop]$ hdfs dfs -mkdir /inputdir
[cloudera@quickstart Desktop]$ hdfs dfs -ls /
Found 7 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2022-02-27 02:21 /hbase
drwxr-xr-x   - cloudera supergroup          0 2022-02-27 22:02 /inputdir
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2022-02-27 02:21 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var
```

```
[cloudera@quickstart Desktop]$ hdfs dfs -ls /inputdir
Found 1 items
-rw-r--r--   1 cloudera supergroup        114 2022-02-27 22:08 /inputdir/abc
[cloudera@quickstart Desktop]$ hdfs dfs -cat /inputdir/abc
every night in my dreams i see you , i feel you
this is how i know you go on far across the distance and speces
[cloudera@quickstart Desktop]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /outputdir
Exception in thread "main" java.lang.ClassNotFoundException: WordCount
        at java.net.URLClassLoader$1.run(URLClassLoader.java:366)
        at java.net.URLClassLoader$1.run(URLClassLoader.java:355)
        at java.security.AccessController.doPrivileged(Native Method)
        at java.net.URLClassLoader.findClass(URLClassLoader.java:354)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:425)
        at java.lang.ClassLoader.loadClass(ClassLoader.java:358)
        at java.lang.Class.forName0(Native Method)
        at java.lang.Class.forName(Class.java:270)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
[cloudera@quickstart Desktop]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /outputdir
22/02/27 22:24:21 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/27 22:24:23 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/02/27 22:24:24 INFO input.FileInputFormat: Total input paths to process : 1
22/02/27 22:24:24 INFO mapreduce.JobSubmitter: number of splits:1
22/02/27 22:24:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1645957104446_0001
22/02/27 22:24:28 INFO impl.YarnClientImpl: Submitted application application_1645957104446_0001
22/02/27 22:24:28 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1645957104446_0001/
```

```
22/02/27 22:24:28 INFO mapreduce.Job: Running job: job_1645957104446_0001
22/02/27 22:25:06 INFO mapreduce.Job: Job job_1645957104446_0001 running in uber mode : false
22/02/27 22:25:06 INFO mapreduce.Job:  map 0% reduce 0%
22/02/27 22:25:38 INFO mapreduce.Job:  map 100% reduce 0%
22/02/27 22:26:05 INFO mapreduce.Job:  map 100% reduce 100%
22/02/27 22:26:05 INFO mapreduce.Job: Job job_1645957104446_0001 completed successfully
22/02/27 22:26:06 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=238
                FILE: Number of bytes written=287155
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=223
                HDFS: Number of bytes written=144
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=28216
                Total time spent by all reduces in occupied slots (ms)=23102
                Total time spent by all map tasks (ms)=28216
                Total time spent by all reduce tasks (ms)=23102
                Total vcore-milliseconds taken by all map tasks=28216
                Total vcore-milliseconds taken by all reduce tasks=23102
                Total megabyte-milliseconds taken by all map tasks=28893184
                Total megabyte-milliseconds taken by all reduce tasks=23656448
        Map-Reduce Framework
                Map input records=2
                Map output records=26
                Map output bytes=216
                Map output materialized bytes=238
                Input split bytes=109
                Combine input records=26
                Combine output records=22
                Reduce input groups=22
                Reduce shuffle bytes=238
                Reduce input records=22
                Reduce output records=22
                Spilled Records=44
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=641
                CPU time spent (ms)=4690
Browsing HDFS - Mozilla Firefox
```

```
                Physical memory (bytes) snapshot=338911232
                Virtual memory (bytes) snapshot=3015090176
                Total committed heap usage (bytes)=226365440
        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=114
        File Output Format Counters
                Bytes Written=144
```

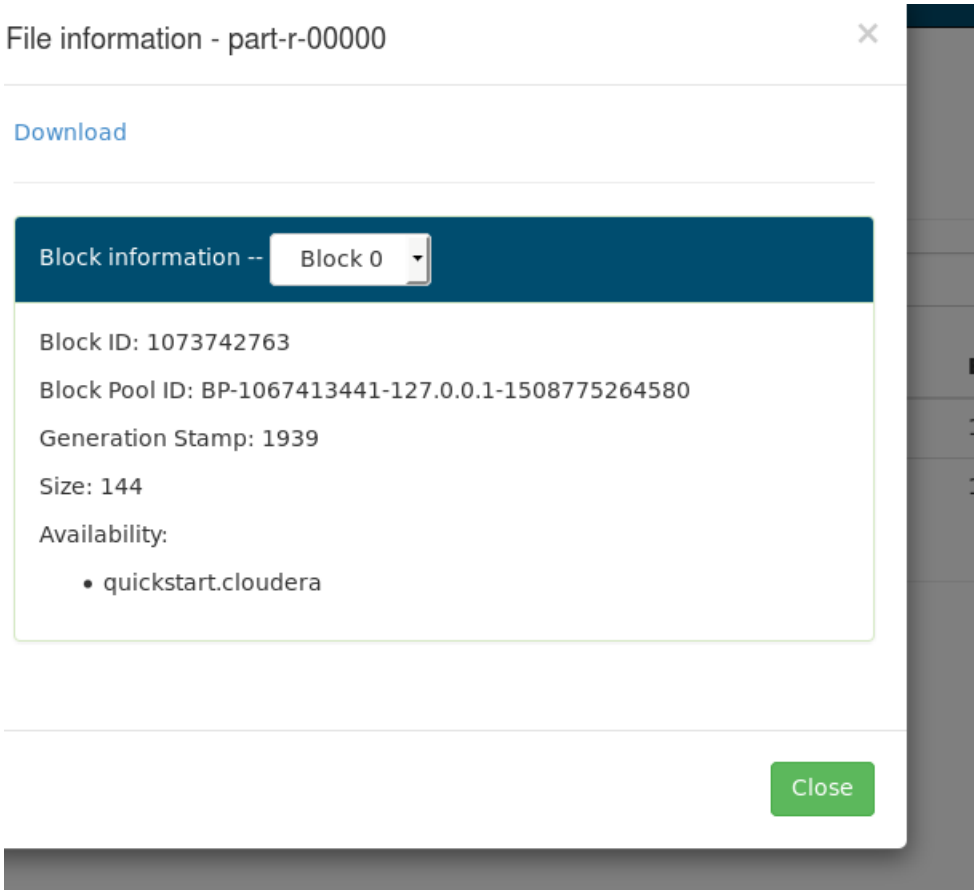
```
                Bytes Written=144
[cloudera@quickstart Desktop]$ hdfs dfs -ls /
Found 8 items
drwxrwxrwx   - hdfs     supergroup          0 2017-10-23 09:15 /benchmarks
drwxr-xr-x   - hbase    supergroup          0 2022-02-27 02:21 /hbase
drwxr-xr-x   - cloudera supergroup          0 2022-02-27 22:08 /inputdir
drwxr-xr-x   - cloudera supergroup          0 2022-02-27 22:26 /outputdir
drwxr-xr-x   - solr     solr                0 2017-10-23 09:18 /solr
drwxrwxrwt   - hdfs     supergroup          0 2022-02-27 02:21 /tmp
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /user
drwxr-xr-x   - hdfs     supergroup          0 2017-10-23 09:17 /var
[cloudera@quickstart Desktop]$ hdfs dfs -ls /outputdir
Found 2 items
-rw-r--r--   1 cloudera supergroup          0 2022-02-27 22:26 /outputdir/_SUCCESS
-rw-r--r--   1 cloudera supergroup        144 2022-02-27 22:26 /outputdir/part-r-00000
[cloudera@quickstart Desktop]$ hdfs dfs -ls /outputdir/part-r-00000
-rw-r--r--   1 cloudera supergroup        144 2022-02-27 22:26 /outputdir/part-r-00000
```

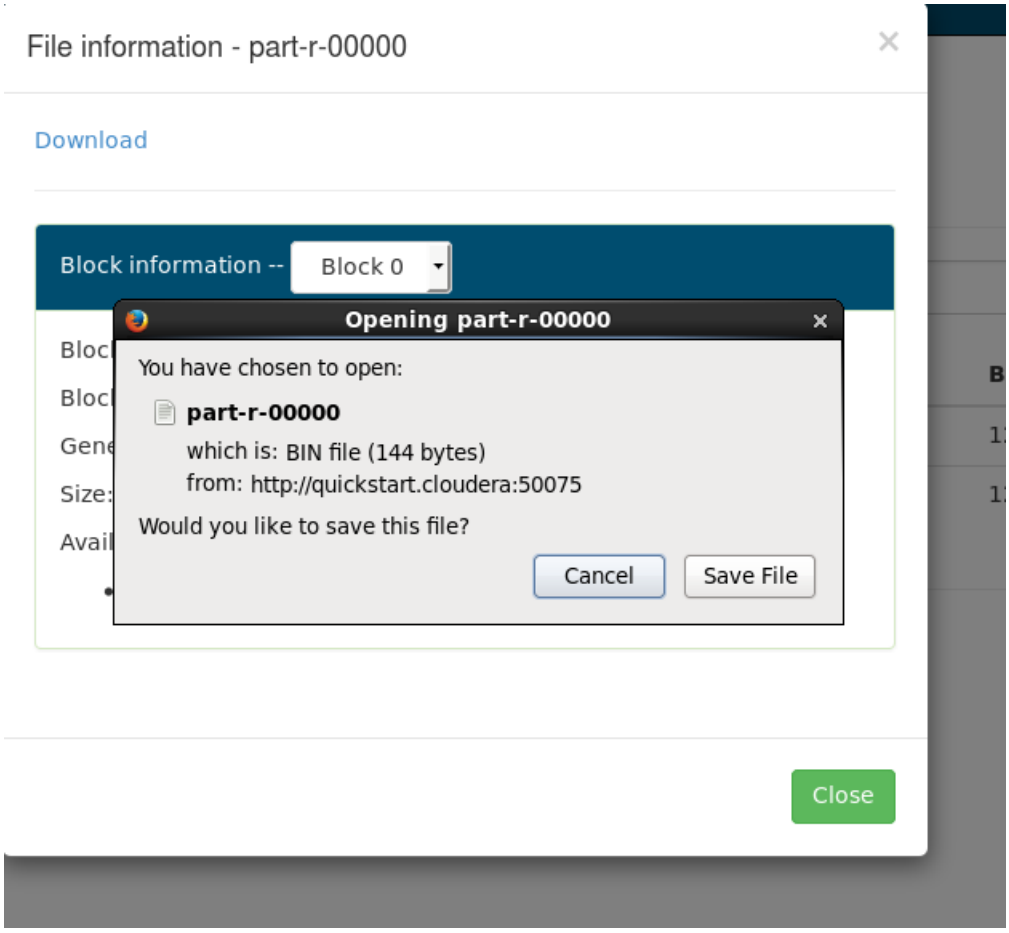
```
cloudera@quickstart Desktop]$ hdfs dfs -cat /outputdir/part-r-00000
       1
cross  1
nd     1
istance        1
reams  1
very   1
ar     1
eel    1
o      1
ow     1
       3
n      1
s      1
now    1
y      1
ight   1
n      1
ee     1
peces  1
he     1
his    1
ou     3
```

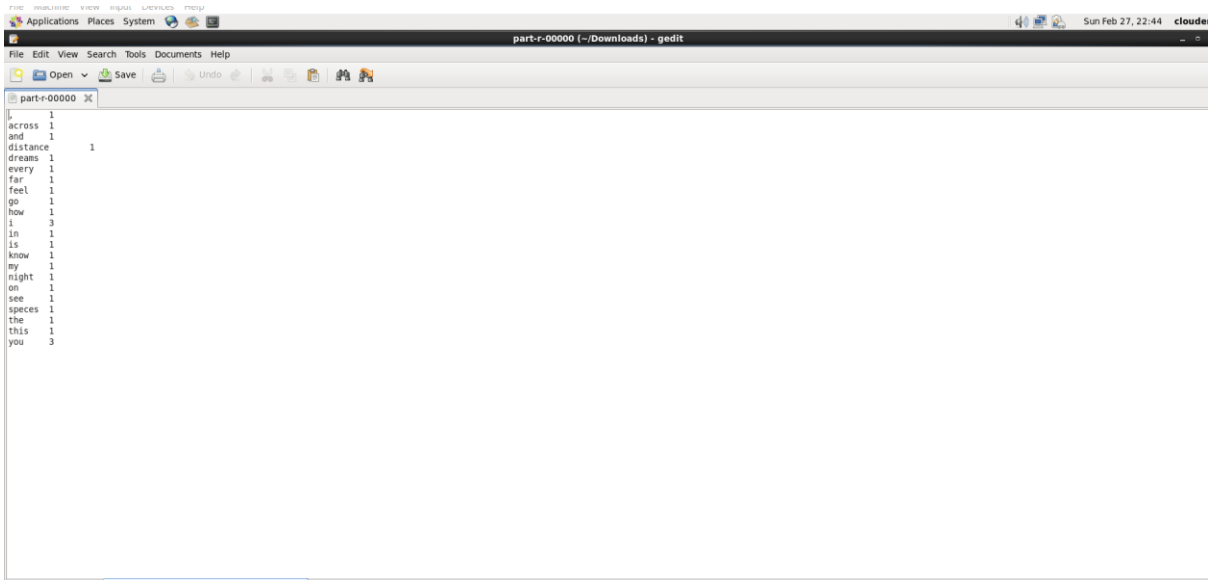
/outputdir Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	cloudera	supergroup	0 B	Sun Feb 27 22:26:03 -0800 2022	1	128 MB	_SUCCESS
-rw-r--r--	cloudera	supergroup	144 B	Sun Feb 27 22:26:02 -0800 2022	1	128 MB	part-r-00000

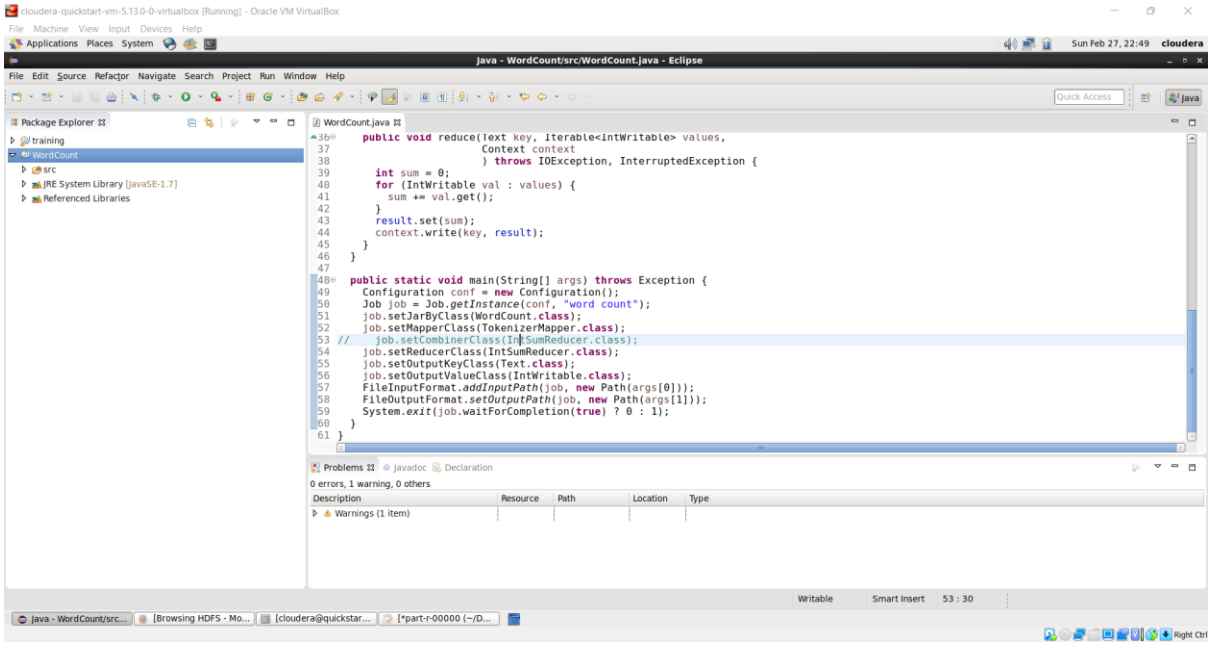
Hadoop, 2017.

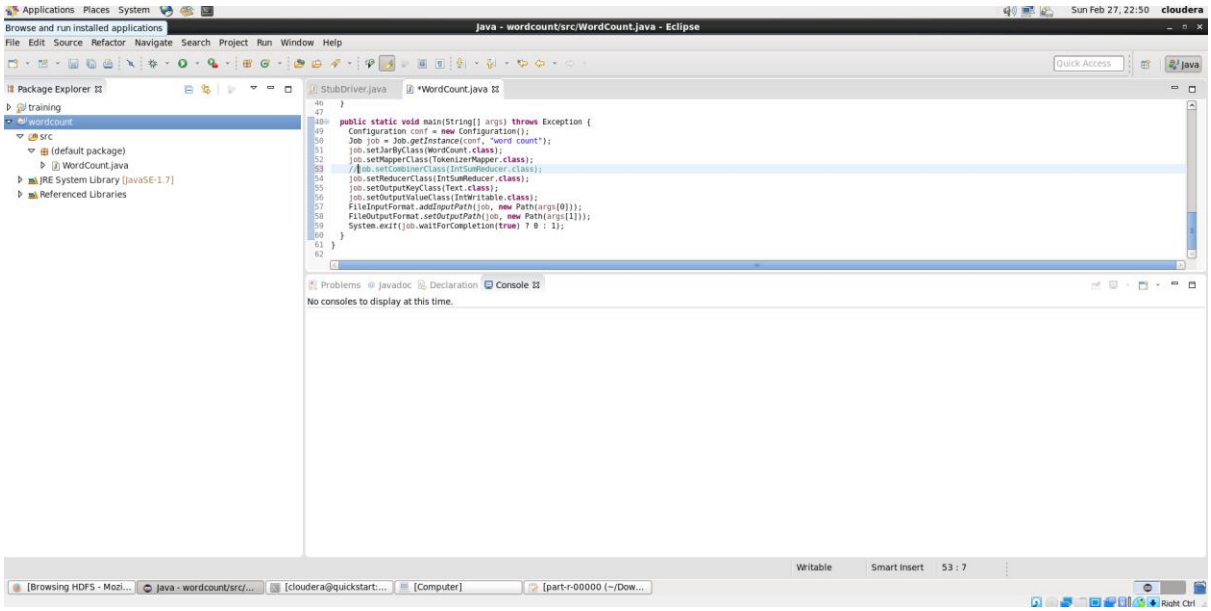

File information - part-r-00000
Download
Block information --
Block 0
Block ID: 1073742763
Block Pool ID: BP-1067413441-127.0.0.1-1508775264580
Generation Stamp: 1939
Size: 144
Availability:
quickstart.cloudera
Close

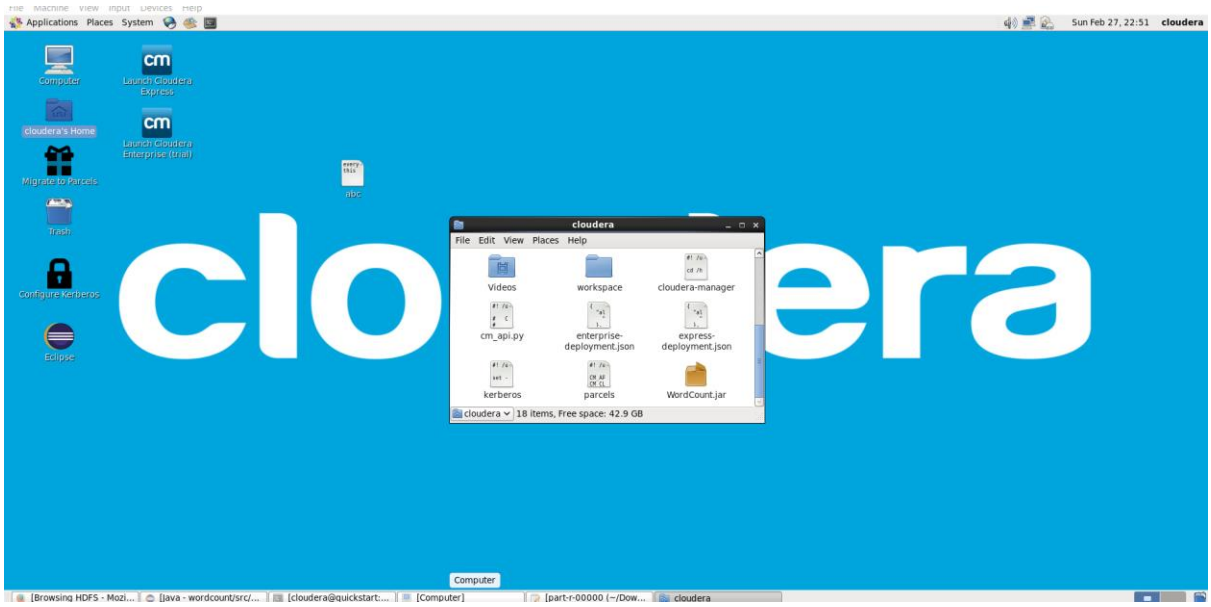


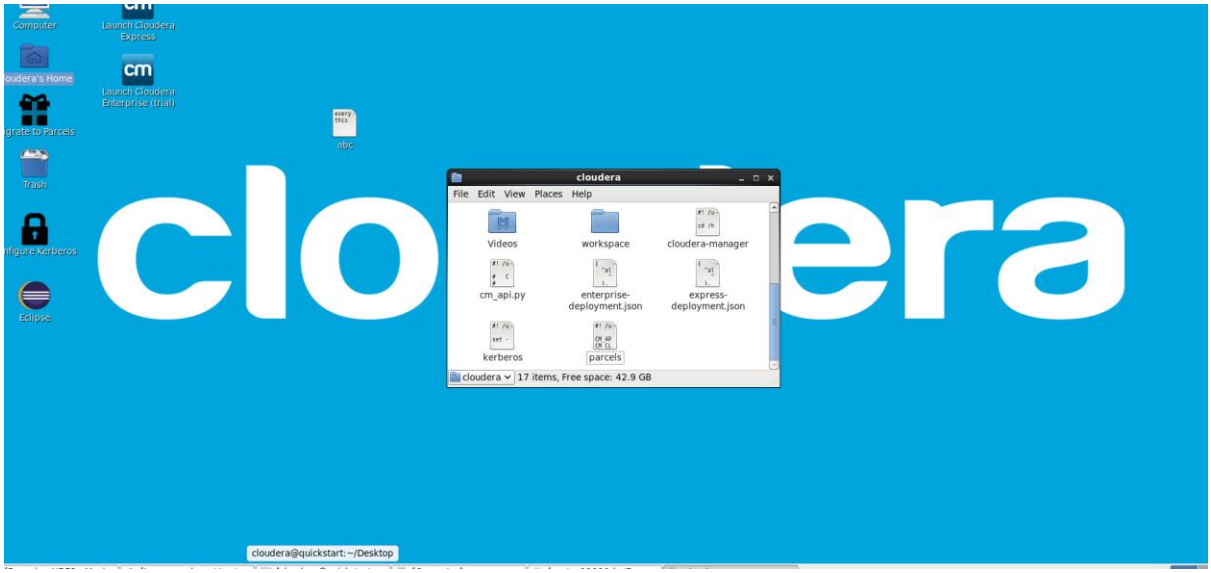


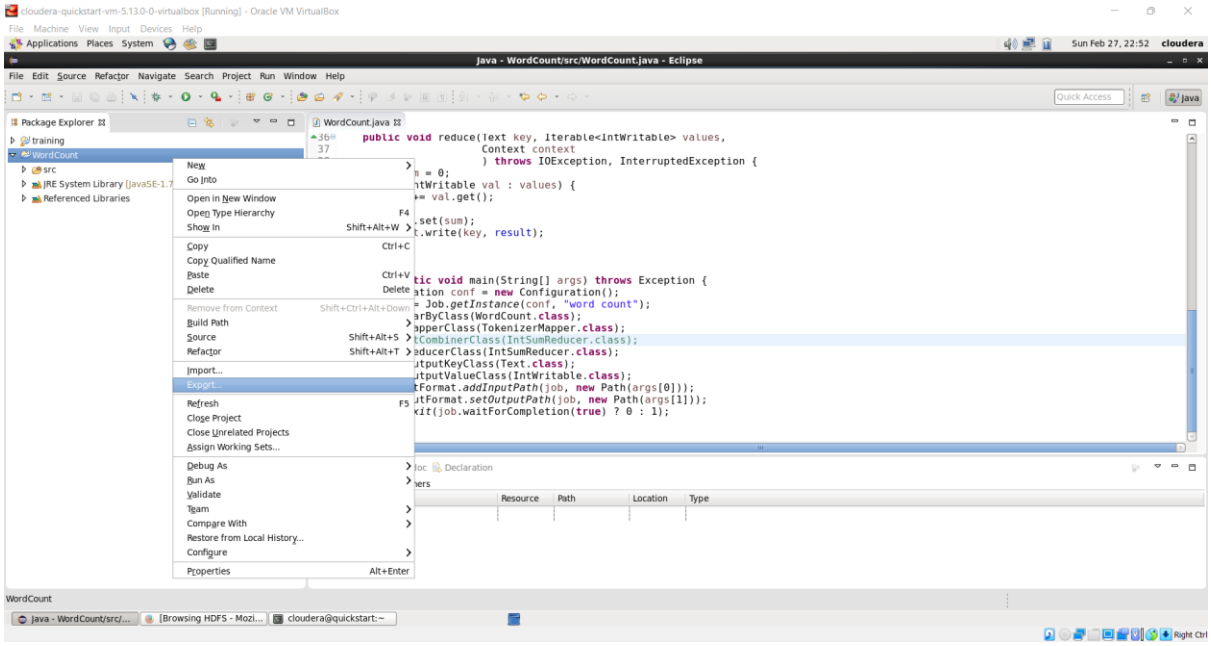
Implementation of WordCount problem using Hadoop MapReduce (With Combiner) in Eclipse:

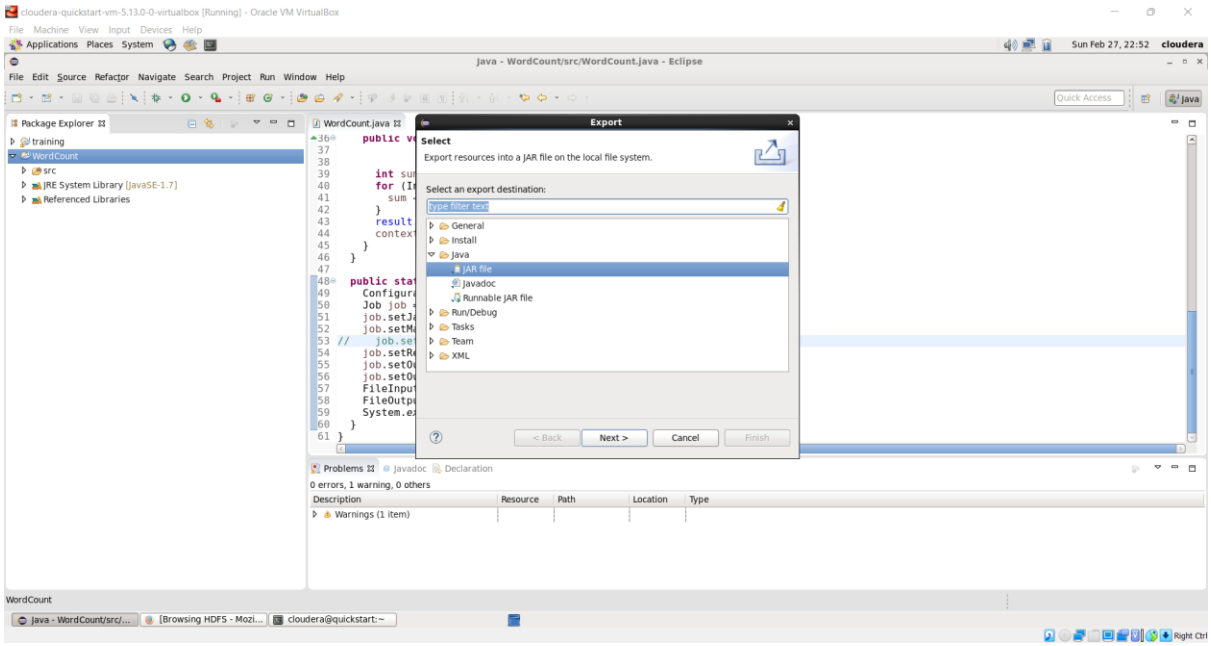

Java - WordCount/src/WordCount.java - Eclipse
public void reduce(Text key, Iterable<IntWritable> values,
Context context
) throws IOException, InterruptedException {
int sum = 0;
for (IntWritable val : values) {
sum += val.get();
}
result.set(sum);
context.write(key, result);
}
}
public static void main(String[] args) throws Exception {
Configuration conf = new Configuration();
Job job = Job.getInstance(conf, "word count");
job.setJarByClass(WordCount.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);
}
}

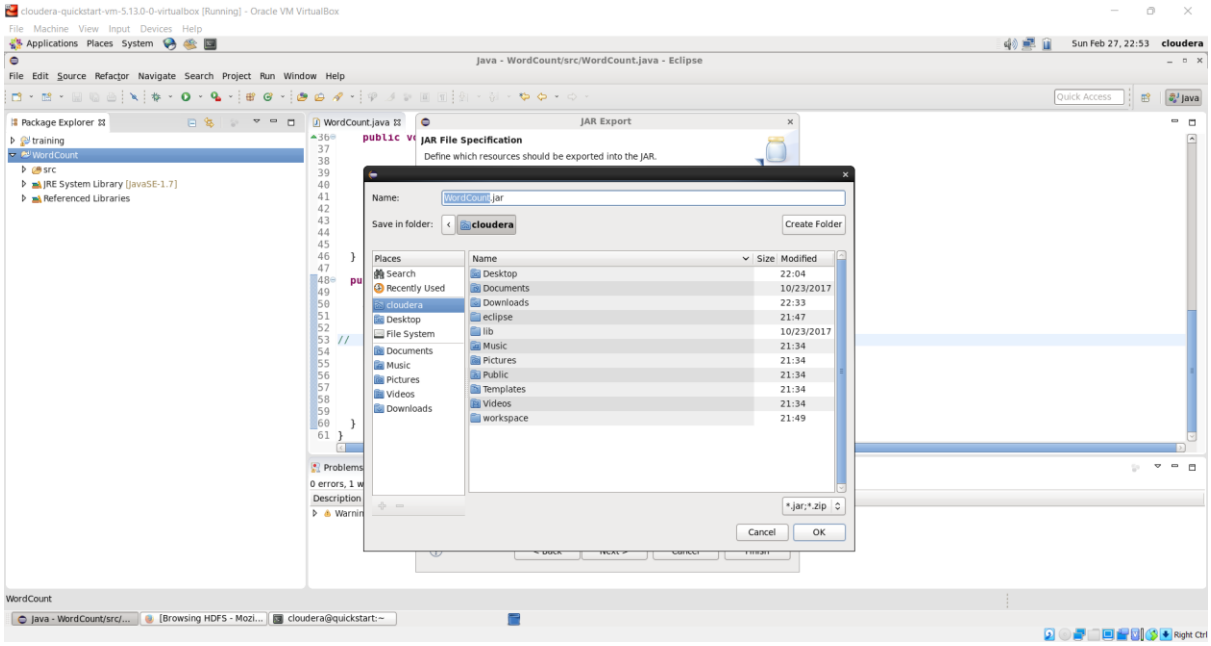

Java - wordcount/src/WordCount.java - Eclipse
public static void main(String[] args) throws Exception {
Configuration conf = new Configuration();
Job job = Job.getInstance(conf, "word count");
job.setJarByClass(WordCount.class);
job.setMapperClass(TokenizerMapper.class);
// job.setCombinerClass(IntSumReducer.class);
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);
}
}
No consoles to display at this time.


Applications Places System
Sun Feb 27, 22:51 cloudera
Computer
Launch Cloudera Express
cloudera's Home
Launch Cloudera Enterprise (trial)
Migrate to Parcels
Trash
Configure Kerberos
Eclipse
abc
cloudera
File Edit View Places Help
Videos
workspace
cloudera-manager
cm_api.py
enterprise-deployment.json
express-deployment.json
kerberos
parcels
WordCount.jar
cloudera 18 items, Free space: 42.9 GB
Computer

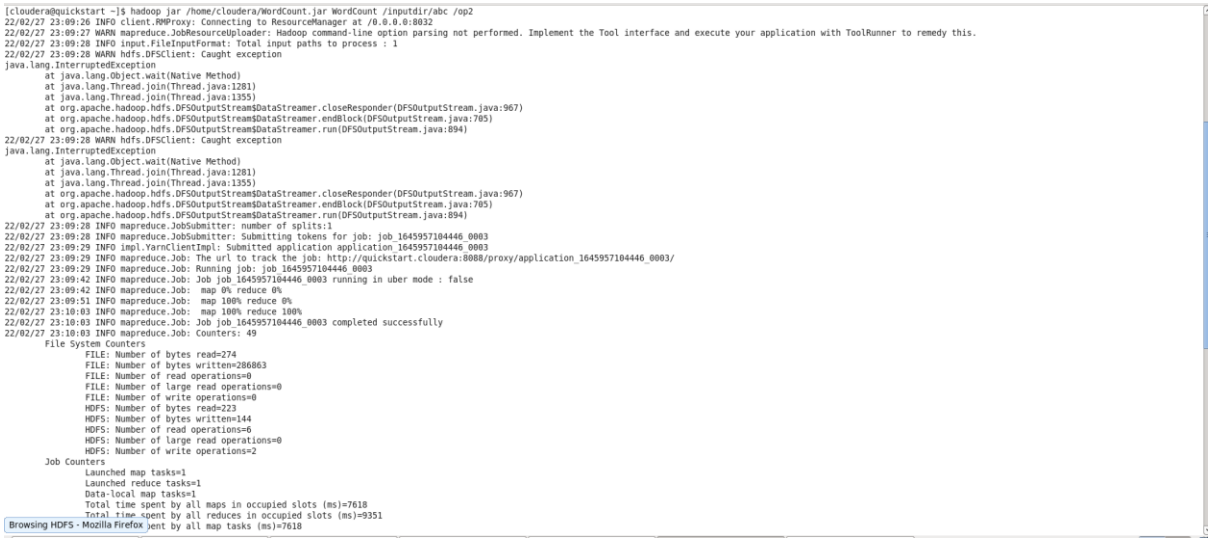

Computer
Launch Cloudera Express
Launch Cloudera Enterprise (trial)
Trash
Eclipse
abc
cloudera
File Edit View Places Help
Videos
workspace
cloudera-manager
cm_api.py
enterprise-deployment.json
express-deployment.json
kerberos
parcels
cloudera 17 items, Free space: 42.9 GB
cloudera@quickstart:~/Desktop


cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 22:52 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
public void reduce(Text key, Iterable<IntWritable> values,
Context context
) throws IOException, InterruptedException {
New
Go Into
Open in New Window
Open Type Hierarchy F4
Show In Shift+Alt+W
Copy Ctrl+C
Copy Qualified Name
Paste Ctrl+V
Delete Delete
Remove from Context Shift+Ctrl+Alt+Down
Build Path
Source Shift+Alt+S
Refactor Shift+Alt+T
Import...
Export...
Refresh F5
Close Project
Close Unrelated Projects
Assign Working Sets...
Debug As
Run As
Validate
Team
Compare With
Restore from Local History...
Configure
Properties Alt+Enter
Declaration
Resource Path Location Type
WordCount
Java - WordCount/src/... [Browsing HDFS - Mozi... cloudera@quickstart:~

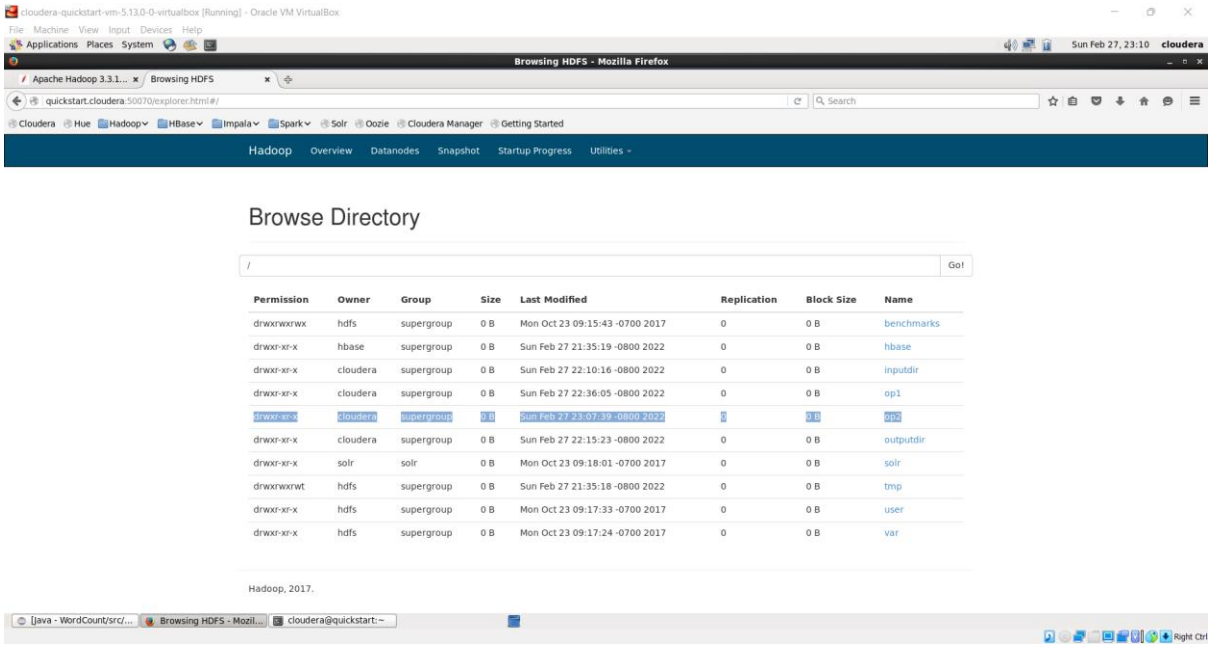

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Applications Places System
Sun Feb 27, 22:52 cloudera
Java - WordCount/src/WordCount.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help
Package Explorer
training
WordCount
src
JRE System Library [JavaSE-1.7]
Referenced Libraries
WordCount.java
Export
Select
Export resources into a JAR file on the local file system.
Select an export destination:
type filter text
General
Install
Java
JAR file
Javadoc
Runnable JAR file
Run/Debug
Tasks
Team
XML
< Back
Next >
Cancel
Finish
Problems
Javadoc
Declaration
0 errors, 1 warning, 0 others
Description Resource Path Location Type
Warnings (1 item)
WordCount
Java - WordCount/src/... [Browsing HDFS - Mozi... cloudera@quickstart:~

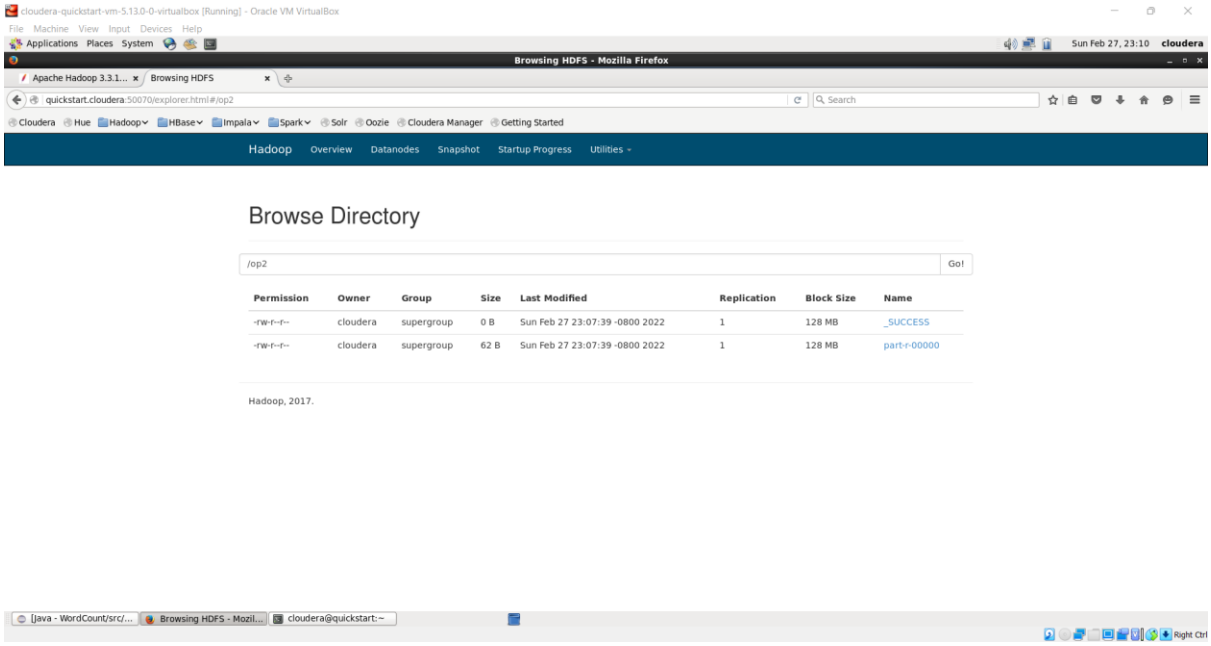

cloudera-quickstart-vm-5.13.0-0-virtualbox [Running] - Oracle VM VirtualBox
Java - WordCount/src/WordCount.java - Eclipse
JAR Export
JAR File Specification
Define which resources should be exported into the JAR.
Name: WordCount.jar
Save in folder: cloudera
Create Folder
Cancel
OK

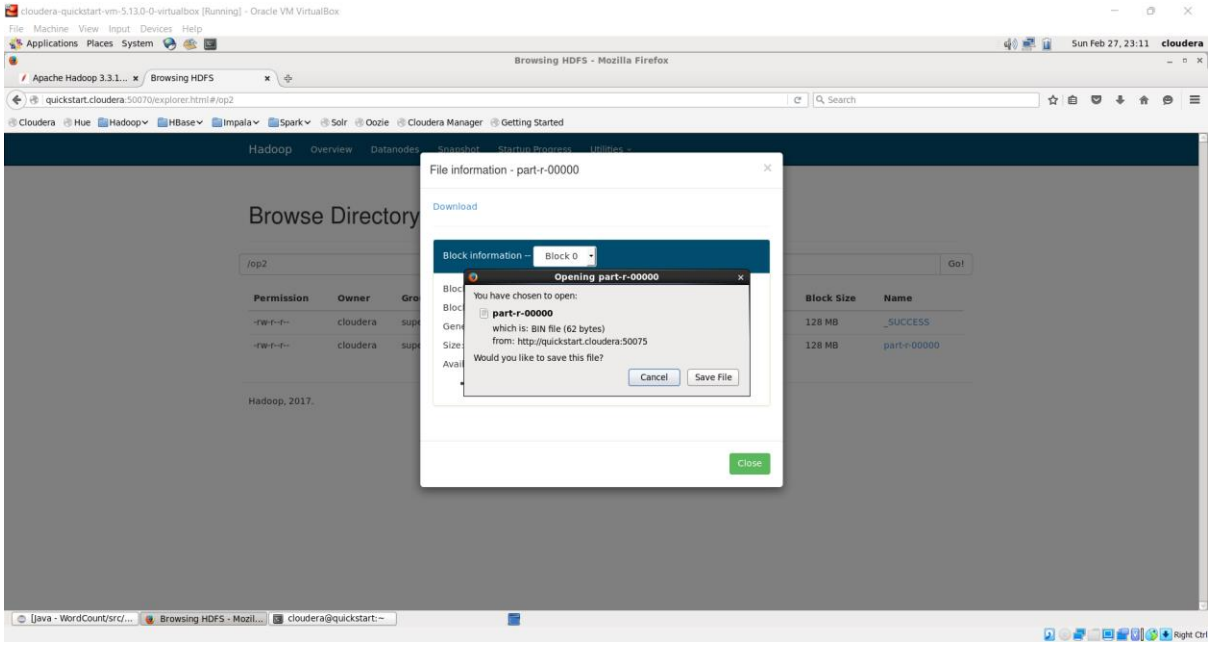
cloudera@quickstart:~
[cloudera@quickstart ~]$ ls
cloudera-manager cm_api.py Desktop Documents Downloads eclipse enterprise-deployment.json express-deployment.json kerberos lib Music parcels Pictures Public Templates Videos WordCount.jar workspace
[cloudera@quickstart ~]$

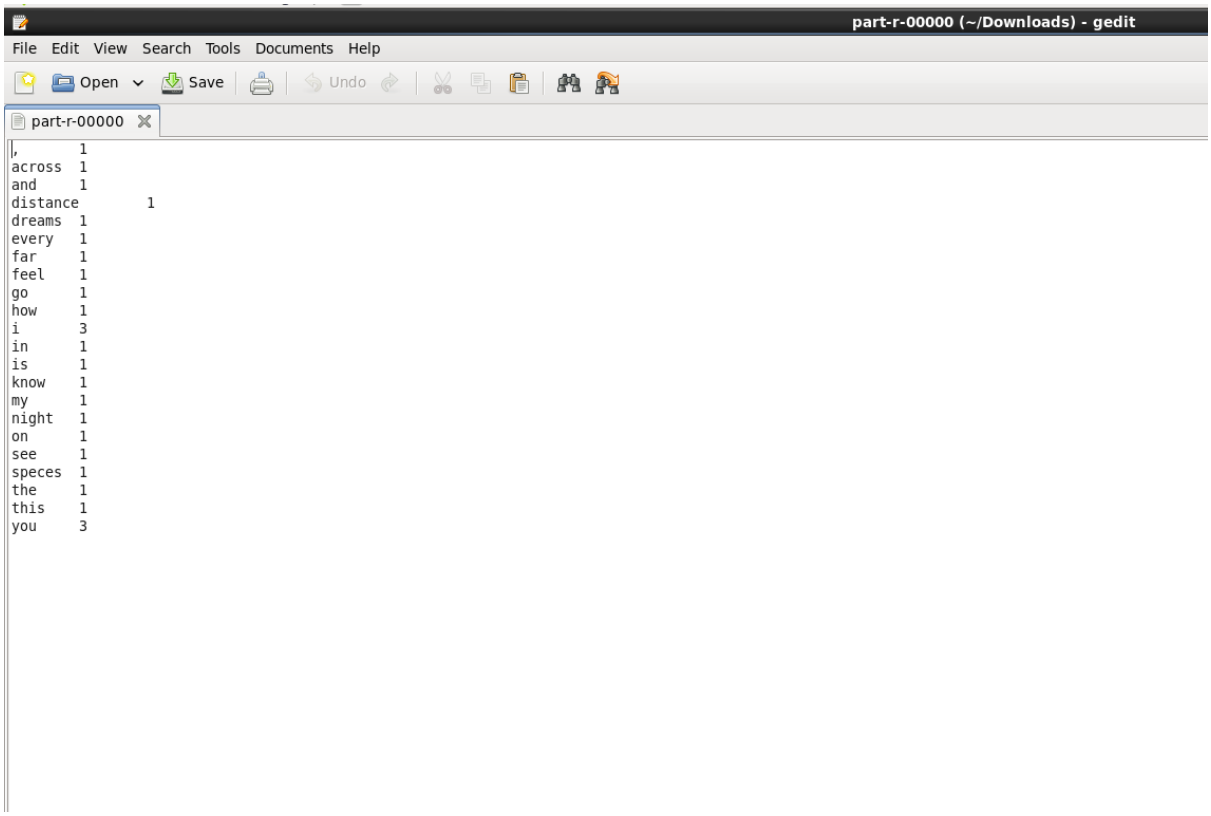

[cloudera@quickstart ~]$ hadoop jar /home/cloudera/WordCount.jar WordCount /inputdir/abc /op2
22/02/27 23:09:26 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
22/02/27 23:09:27 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/02/27 23:09:28 INFO input.FileInputFormat: Total input paths to process : 1
22/02/27 23:09:28 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/02/27 23:09:28 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
at java.lang.Object.wait(Native Method)
at java.lang.Thread.join(Thread.java:1281)
at java.lang.Thread.join(Thread.java:1355)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
22/02/27 23:09:28 INFO mapreduce.JobSubmitter: number of splits:1
22/02/27 23:09:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1645957104446_0003
22/02/27 23:09:29 INFO impl.YarnclientImpl: Submitted application application_1645957104446_0003
22/02/27 23:09:29 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1645957104446_0003/
22/02/27 23:09:29 INFO mapreduce.Job: Running job: job_1645957104446_0003
22/02/27 23:09:42 INFO mapreduce.Job: Job job_1645957104446_0003 running in uber mode : false
22/02/27 23:09:42 INFO mapreduce.Job: map 0% reduce 0%
22/02/27 23:09:51 INFO mapreduce.Job: map 100% reduce 0%
22/02/27 23:10:03 INFO mapreduce.Job: map 100% reduce 100%
22/02/27 23:10:03 INFO mapreduce.Job: Job job_1645957104446_0003 completed successfully
22/02/27 23:10:03 INFO mapreduce.Job: Counters: 49
File System Counters
FILE: Number of bytes read=274
FILE: Number of bytes written=286863
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=223
HDFS: Number of bytes written=144
HDFS: Number of read operations=6
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
Launched map tasks=1
Launched reduce tasks=1
Data-local map tasks=1
Total time spent by all maps in occupied slots (ms)=7618
Total time spent by all reduces in occupied slots (ms)=9351
Browsing HDFS - Mozilla Firefox

```
                HDFS: Number of bytes written=144
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=7618
                Total time spent by all reduces in occupied slots (ms)=9351
                Total time spent by all map tasks (ms)=7618
                Total time spent by all reduce tasks (ms)=9351
                Total vcore-milliseconds taken by all map tasks=7618
                Total vcore-milliseconds taken by all reduce tasks=9351
                Total megabyte-milliseconds taken by all map tasks=7800832
                Total megabyte-milliseconds taken by all reduce tasks=9575424
        Map-Reduce Framework
                Map input records=2
                Map output records=26
                Map output bytes=216
                Map output materialized bytes=274
                Input split bytes=109
                Combine input records=0
                Combine output records=0
                Reduce input groups=22
                Reduce shuffle bytes=274
                Reduce input records=26
                Reduce output records=22
                Spilled Records=52
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=245
                CPU time spent (ms)=2050
                Physical memory (bytes) snapshot=345268224
                Virtual memory (bytes) snapshot=3015094272
                Total committed heap usage (bytes)=226365440
        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=114
        File Output Format Counters
                Bytes Written=144
[cloudera@quickstart ~]$ s
```




Browsing HDFS - Mozilla Firefox
Hadoop Overview Datanodes Snapshot Startup Progress Utilities
Browse Directory
/op2
Go!
Permission Owner Group Size Last Modified Replication Block Size Name
-rw-r--r-- cloudera supergroup 0 B Sun Feb 27 23:07:39 -0800 2022 1 128 MB _SUCCESS
-rw-r--r-- cloudera supergroup 62 B Sun Feb 27 23:07:39 -0800 2022 1 128 MB part-r-00000
Hadoop, 2017.


Browsing HDFS - Mozilla Firefox
File information - part-r-00000
Download
Block information -- Block 0
Opening part-r-00000
You have chosen to open:
part-r-00000
which is: BIN file (62 bytes)
from: http://quickstart.cloudera:50075
Would you like to save this file?
Cancel
Save File
Close




```
,       1
across  1
and     1
distance        1
dreams  1
every   1
far     1
feel    1
go      1
how     1
i       3
in      1
is      1
know    1
my      1
night   1
on      1
see     1
speces  1
the     1
this    1
you     3
```