

MapReduce programming

Unit - II





Steps for MapReduce program execution

1. Install Eclipse

<https://www.eclipse.org/downloads/packages/installer>

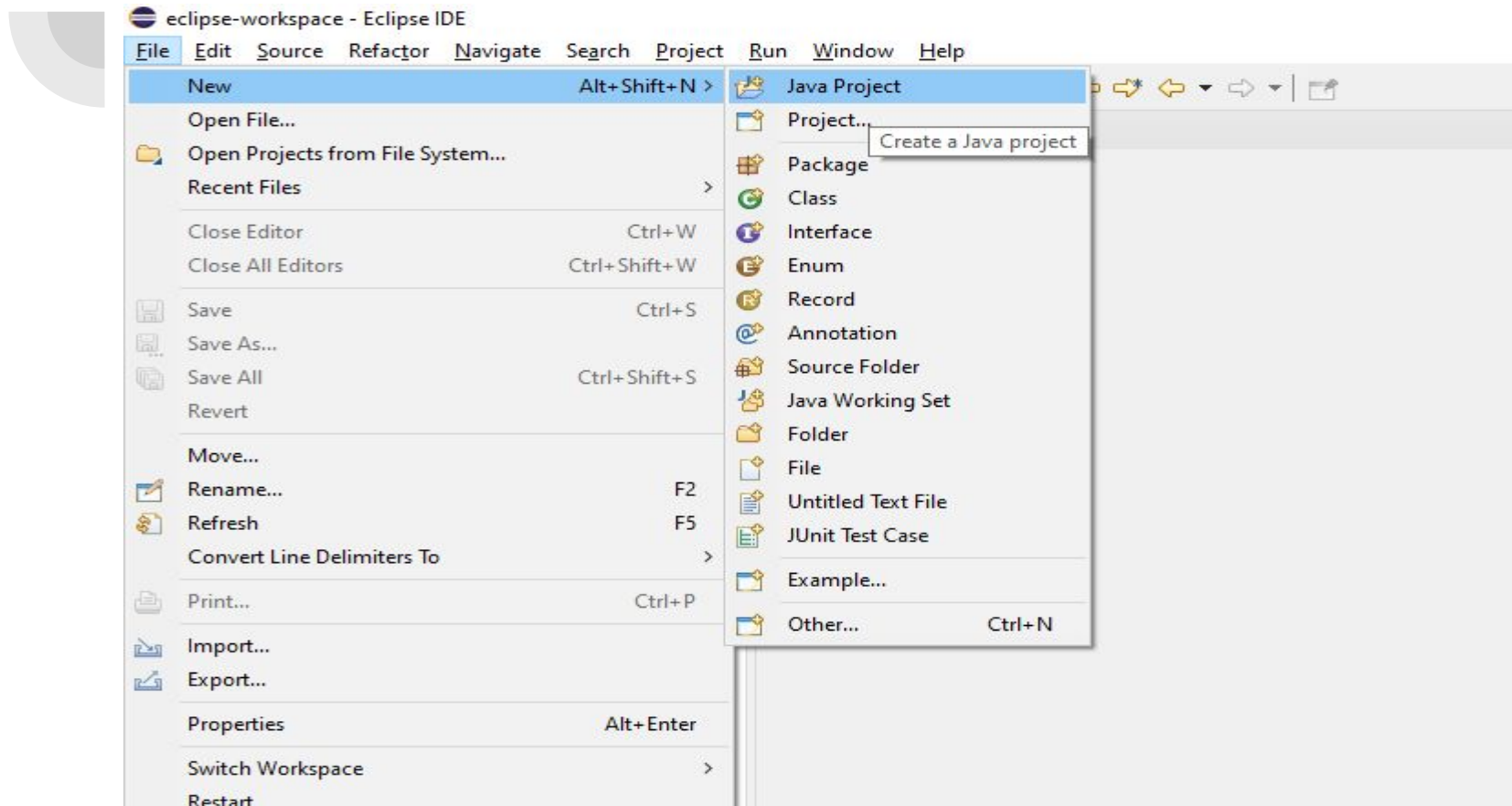
2. Install JDK

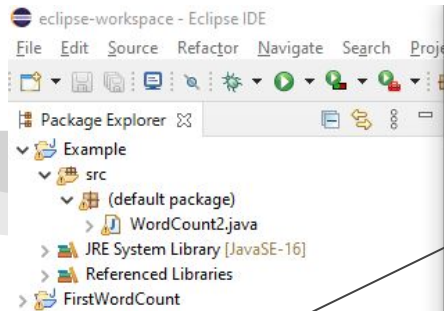
<https://www.oracle.com/in/java/technologies/javase-jdk16-downloads.html>

3. Download two Hadoop Jar files (Hadoop-common-2.7.7.jar and Hadoop-mapreduce-client-core-2.7.7.jar)

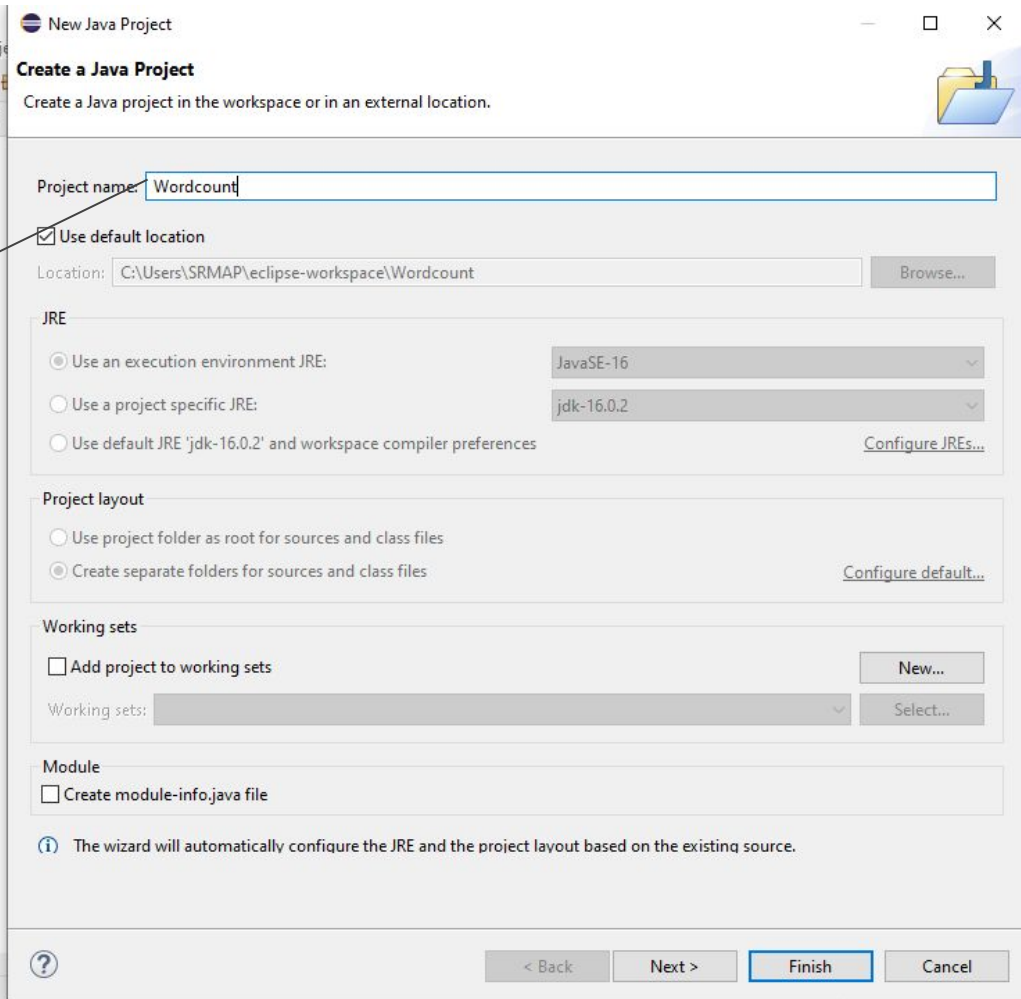
<https://github.com/SumalathaSaleti/WordCountProgram>

Create a new Java project in Eclipse

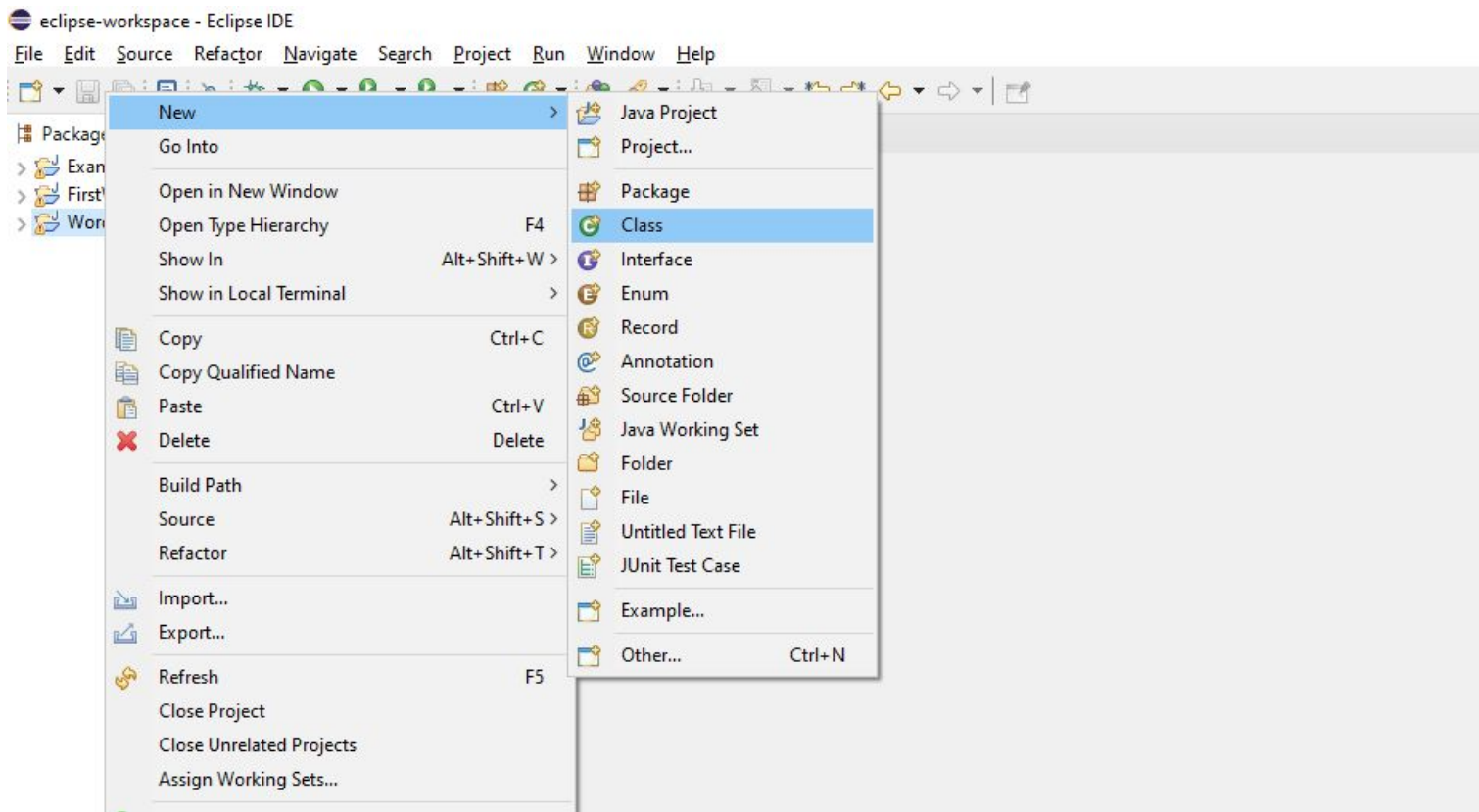




Enter the project name



Create a class





New Java Class

The use of the default package is discouraged.

Source folder:

Package:

☐ Enclosing type:

Name:

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

Superclass:

Interfaces:

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

Enter the class name
and check the
method 'main'



Write the code

Find the code in the below link

<https://github.com/SumalathaSaleti/WordCountProgram>

Check the Java compiler

The screenshot shows the Eclipse IDE interface. On the left, the Package Explorer shows a project named 'Wordcount'. The Properties view on the right shows the 'Java Compiler' settings for the 'Wordcount' project. The 'Enable project specific settings' checkbox is unchecked. An arrow points to this checkbox with the text 'Diable project specific settings'. Another arrow points to the 'Configure Workspace Settings...' link with the text 'Click here'.

Properties for Wordcount

- Resource
- Builders
- Coverage
- Java Build Path
- Java Code Style
- Java Compiler**
- Javadoc Location
- Java Editor
- Project Natures
- Project References
- Run/Debug Settings
- Task Tags
- Validation
- WikiText

Java Compiler

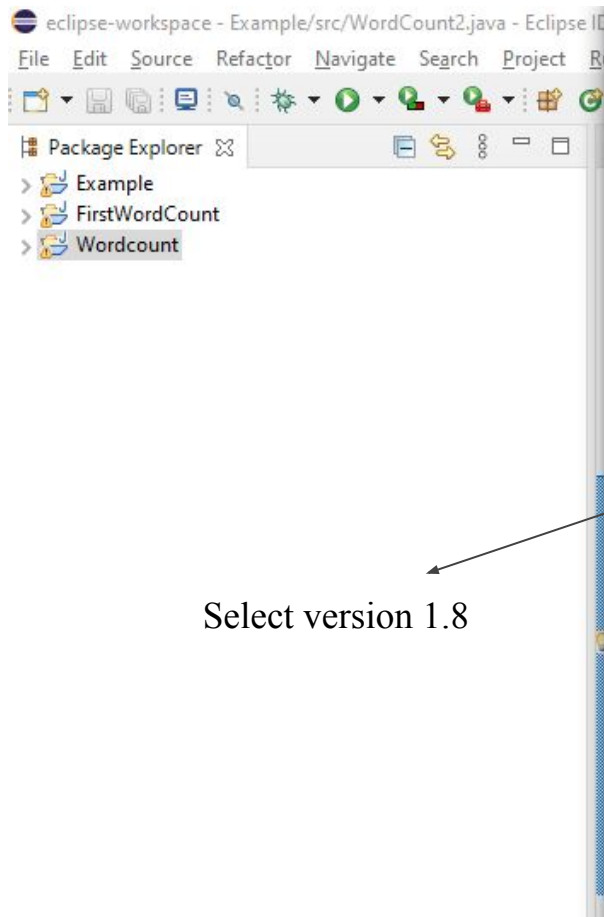
☐ Enable project specific settings [Configure Workspace Settings...](#)

JDK Compliance

- ☒ Use compliance from execution environment 'JavaSE-1.8' on the 'Java Build'
- Compiler compliance level: 1.8
- ☐ Use '--release' option
- ☒ Use default compliance settings
 - ☐ Enable preview features for Java 16
 - Preview features with severity level: Warning
- Generated .class files compatibility: 1.8
- Source compatibility: 1.8
- Disallow identifiers called 'assert': Error
- Disallow identifiers called 'enum': Error

Classfile Generation

- ☒ Add variable attributes to generated class files (used by the debugger)
- ☒ Add line number attributes to generated class files (used by the debugger)
- ☒ Add source file name to generated class file (used by the debugger)
- ☒ Preserve unused (never read) local variables
- ☒ Inline finally blocks (larger class files, but improved performance)
- ☐ Store information about method parameters (usable via reflection)



Preferences (Filtered)

type filter text

Java
 > Compiler

Compiler

JDK Compliance

Compiler compliance level: 1.8

☐ Use '--release' option

☒ Use default compliance settings

☐ Enable preview features for Java 16

Preview features with severity level:

Warning

Generated .class files compatibility:

1.8

Source compatibility:

1.8

Disallow identifiers called 'assert':

Error

Disallow identifiers called 'enum':

Error

Classfile Generation

☒ Add variable attributes to generated class files (used by the debugger)

☒ Add line number attributes to generated class files (used by the debugger)

☒ Add source file name to generated class file (used by the debugger)

☒ Preserve unused (never read) local variables

☒ Inline finally blocks (larger class files, but improved performance)

☐ Store information about method parameters (usable via reflection)

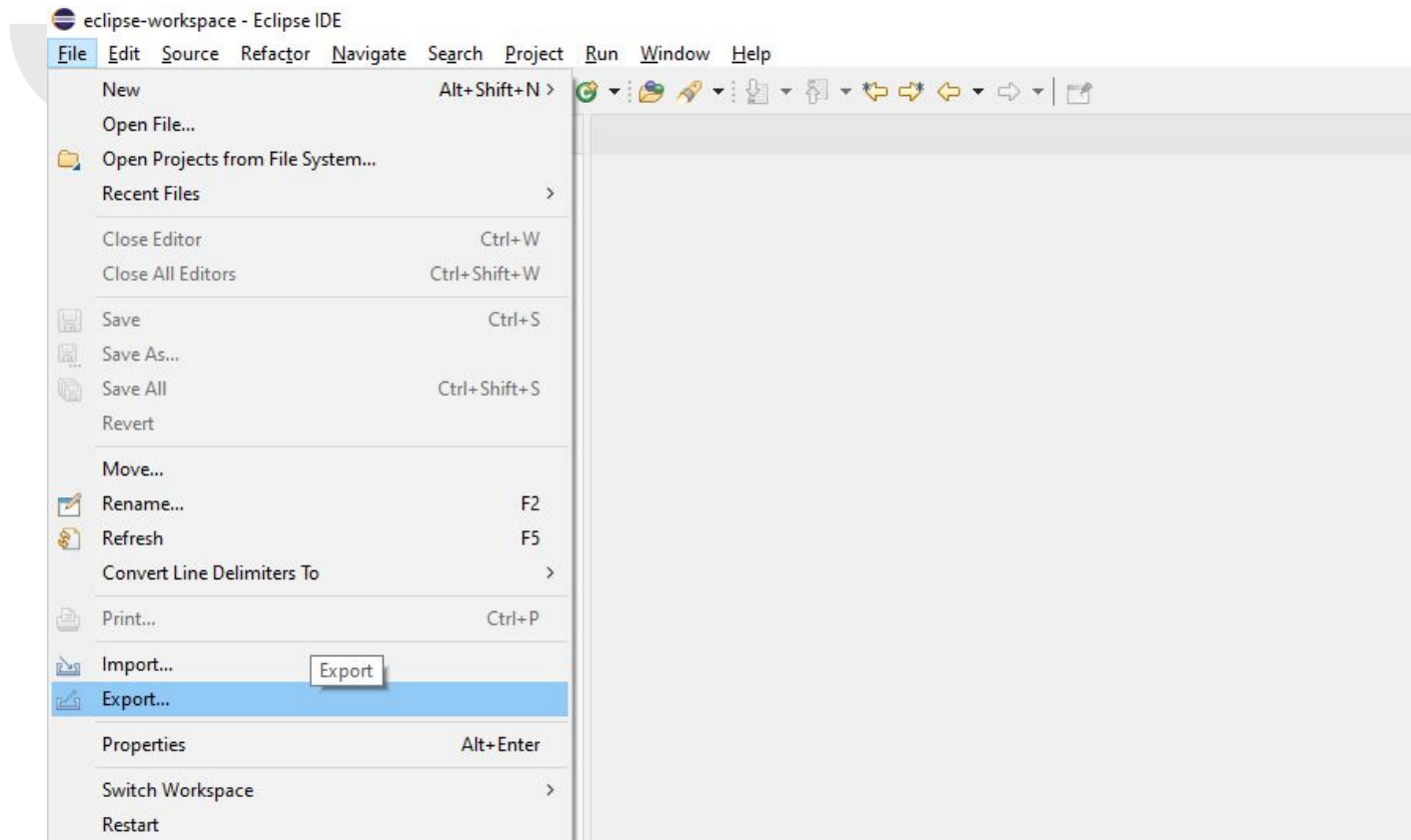
⚠ When selecting 1.8 compliance, make sure to have a compatible JRE installed and activated (currently 16). [Configure...](#)

Select version 1.8



- Include Hadoop common jar files

Export Jar file



File Edit Source Refactor Navigate Search Project Run Window



Package Explorer

- > Example
- > FirstWordCount
- > Wordcount

Export

Select

Export resources into a JAR file on the local file system.



Select an export wizard:

type filter text

- General
 - Ant Buildfiles
 - Archive File
 - File System
 - Preferences
- Install
- Java
 - JAR file
 - Javadoc
 - Runnable JAR file
- Run/Debug
- Team
- XML

Select Jar file option and
click 'next'

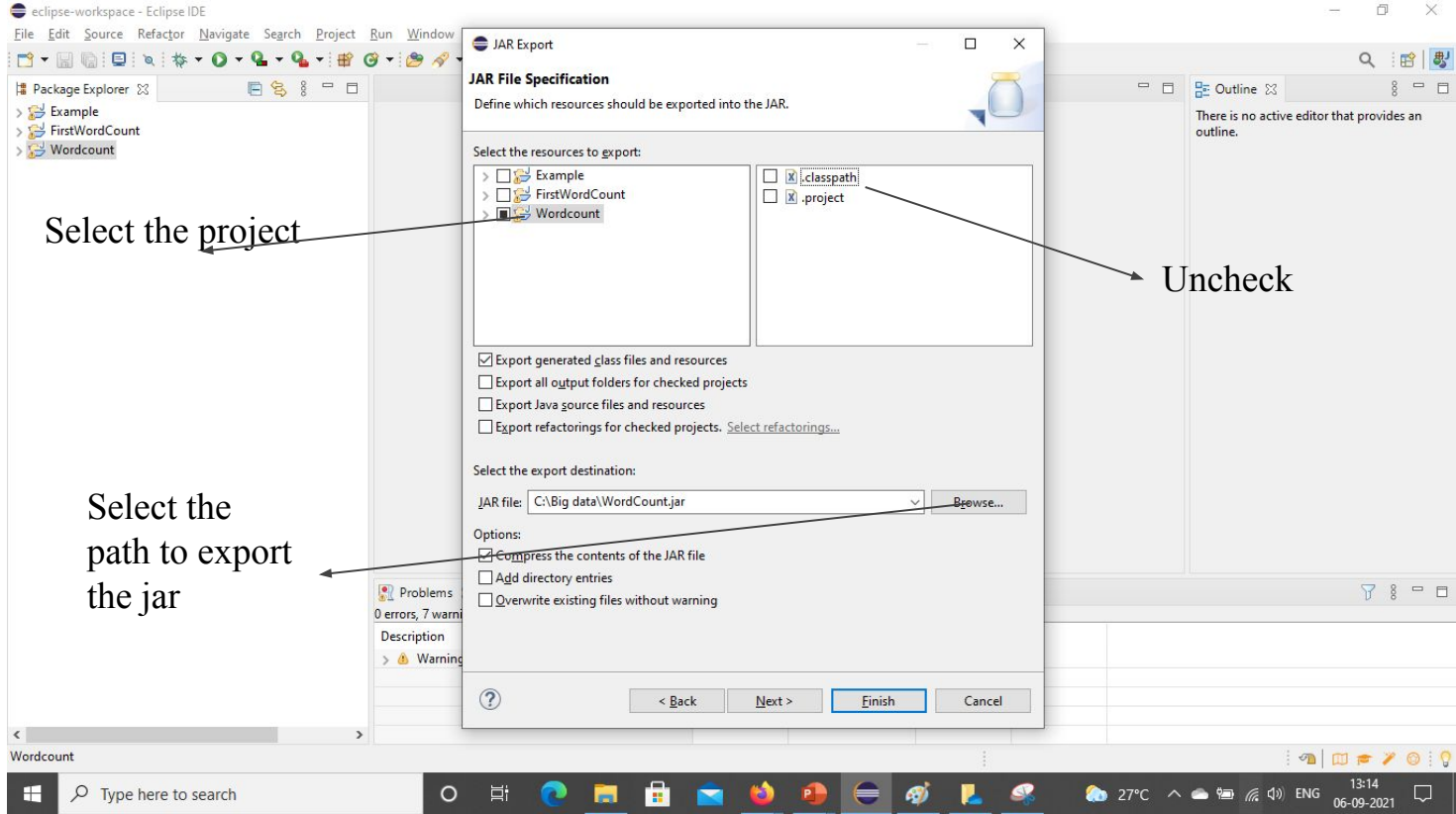


< Back

Next >

Finish

Cancel





Create an input directory in HDFS

```
sandbox-hdp login: root
root@sandbox-hdp.hortonworks.com's password:
Last login: Tue Aug 24 12:21:03 2021 from 172.18.0.2
[root@sandbox-hdp ~]# hadoop fs -mkdir /CSEBDA
[root@sandbox-hdp ~]# hadoop fs -mkdir /CSEBDA/input
```

Upload the jar file and input file into HDFS

Ambari - Sandbox — Mozilla Firefox

localhost:8080/#/main/view/FILES/auto_files_instance

Ambari Sandbox 0 ops

Dashboard Services Hosts Alerts Admin

admin

/ > CSESDA > input

Total: 0 files or folders

+ Select All New Folder Upload

Search in current directory...

Name >	Size >	Last Modified >	Owner >	Group >	Permission
--------	--------	-----------------	---------	---------	------------

Input directory

Upload the text file



/ > CSEBDA > input

Total: 0 files or folders

+ Select All

New Folder

Upload



0

Search in current directory...



Name >



Group >

Permission

Upload file to /CSEBDA/input



Drag file to upload or click to browse

Currently supports single file upload

Cancel

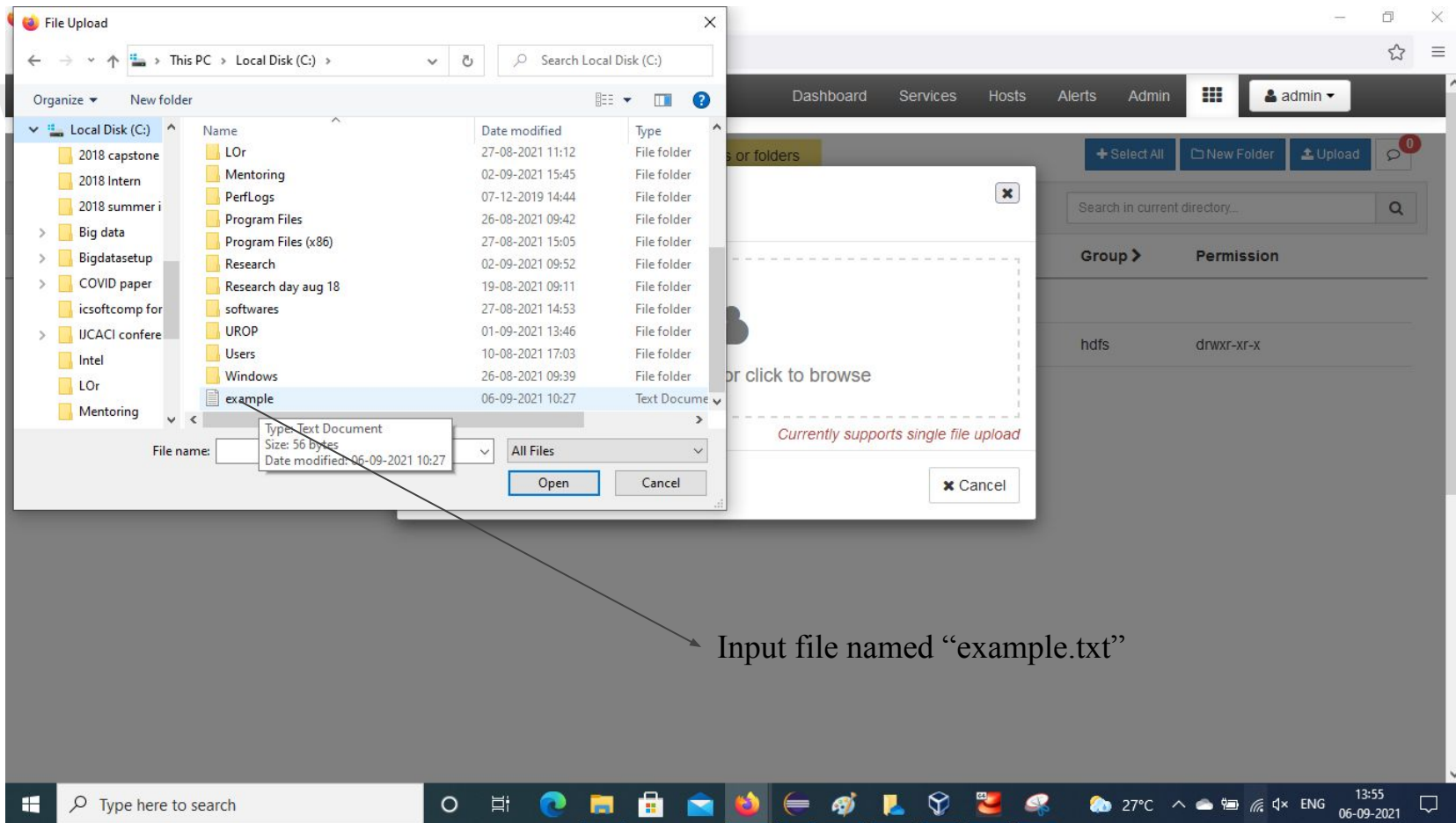


Type here to search

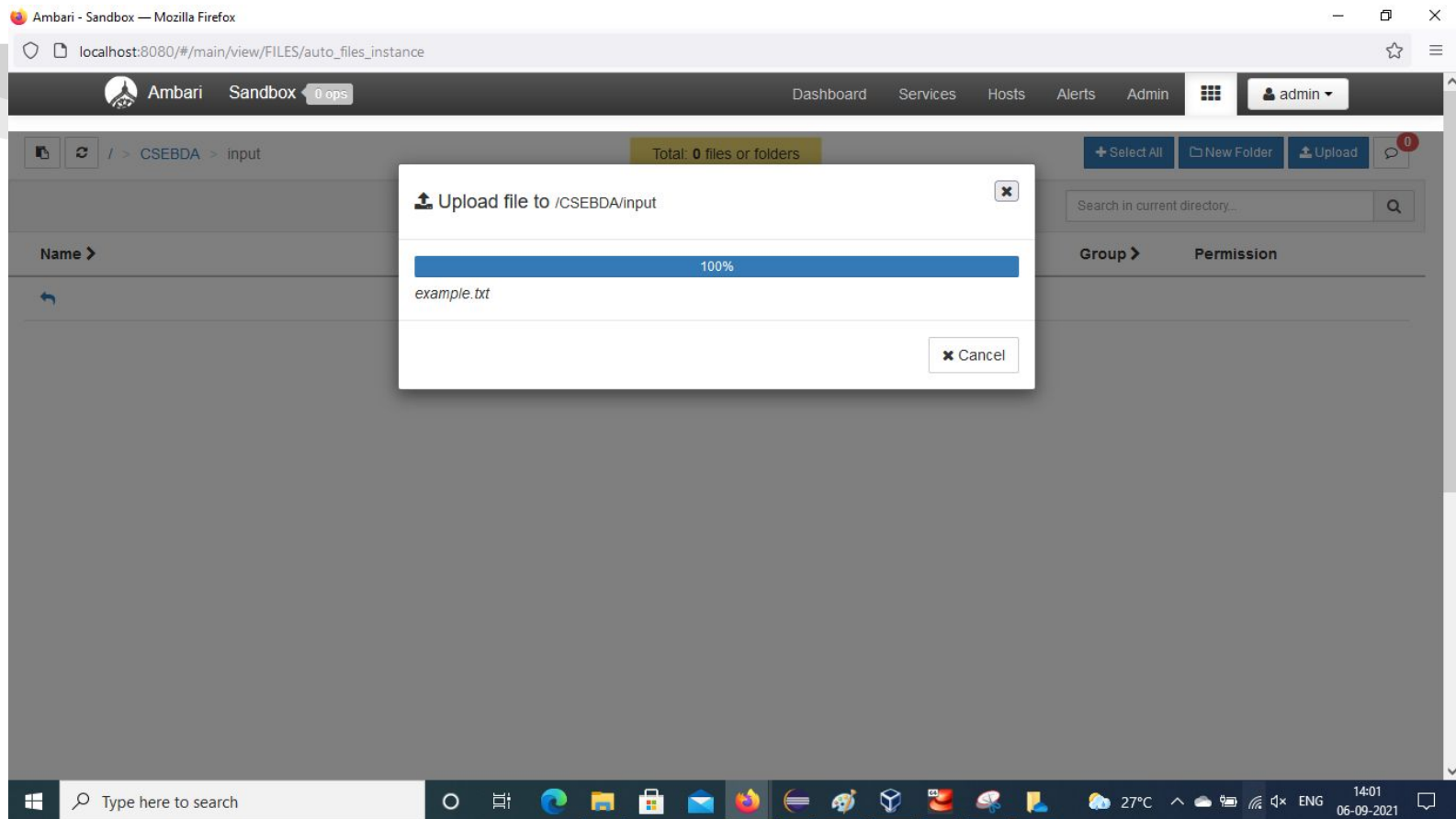


27°C

14:01
06-09-2021



Input file named "example.txt"



Ambari - Sandbox — Mozilla Firefox

localhost:8080/#/main/view/FILES/auto_files_instance

Ambari Sandbox 0 ops

Dashboard Services Hosts Alerts Admin

admin

/ > CSEBDA > input

Total: 1 files or folders

+ Select All New Folder Upload

Search in current directory...

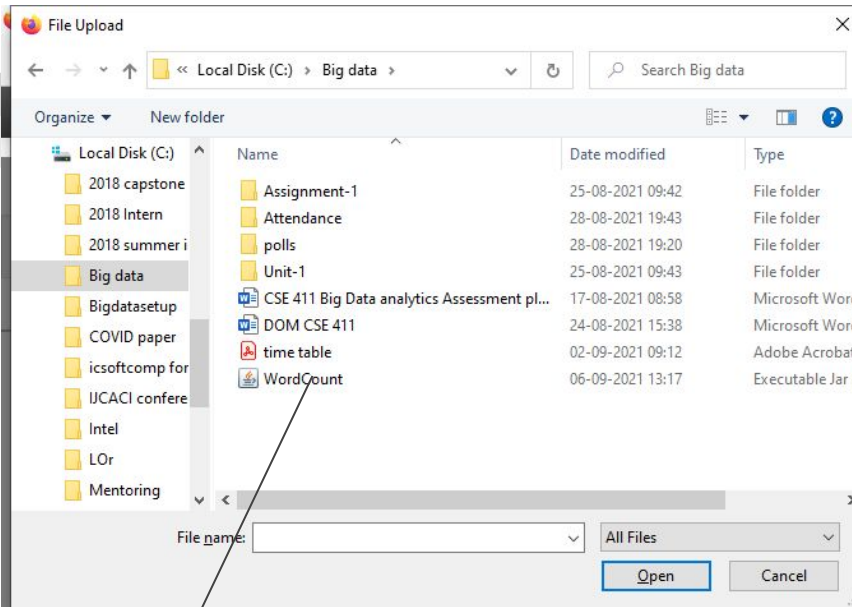
Name >	Size >	Last Modified >	Owner >	Group >	Permission
example.txt	0.1 KB	2021-08-24 19:07	admin	hdfs	-rw-r--r--

Input file is uploaded to HDFS

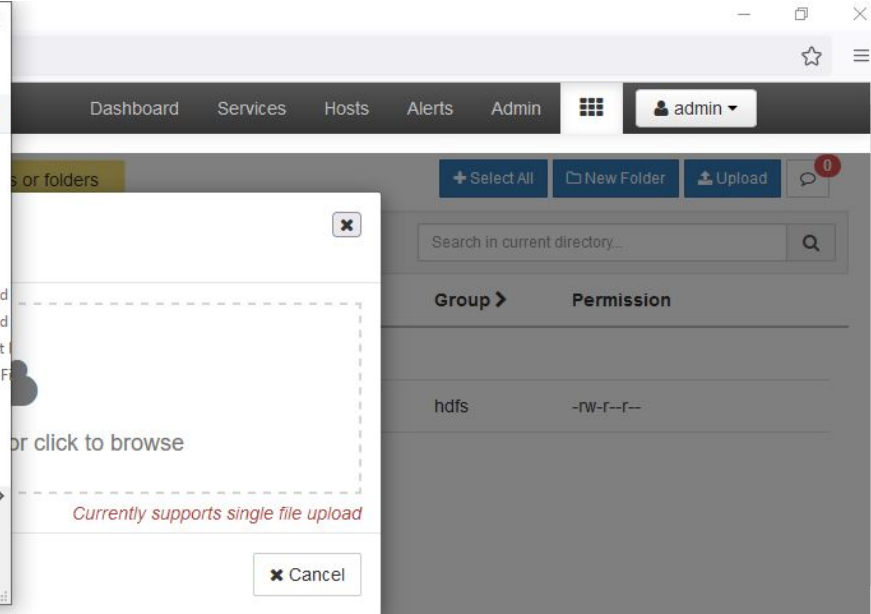
Type here to search

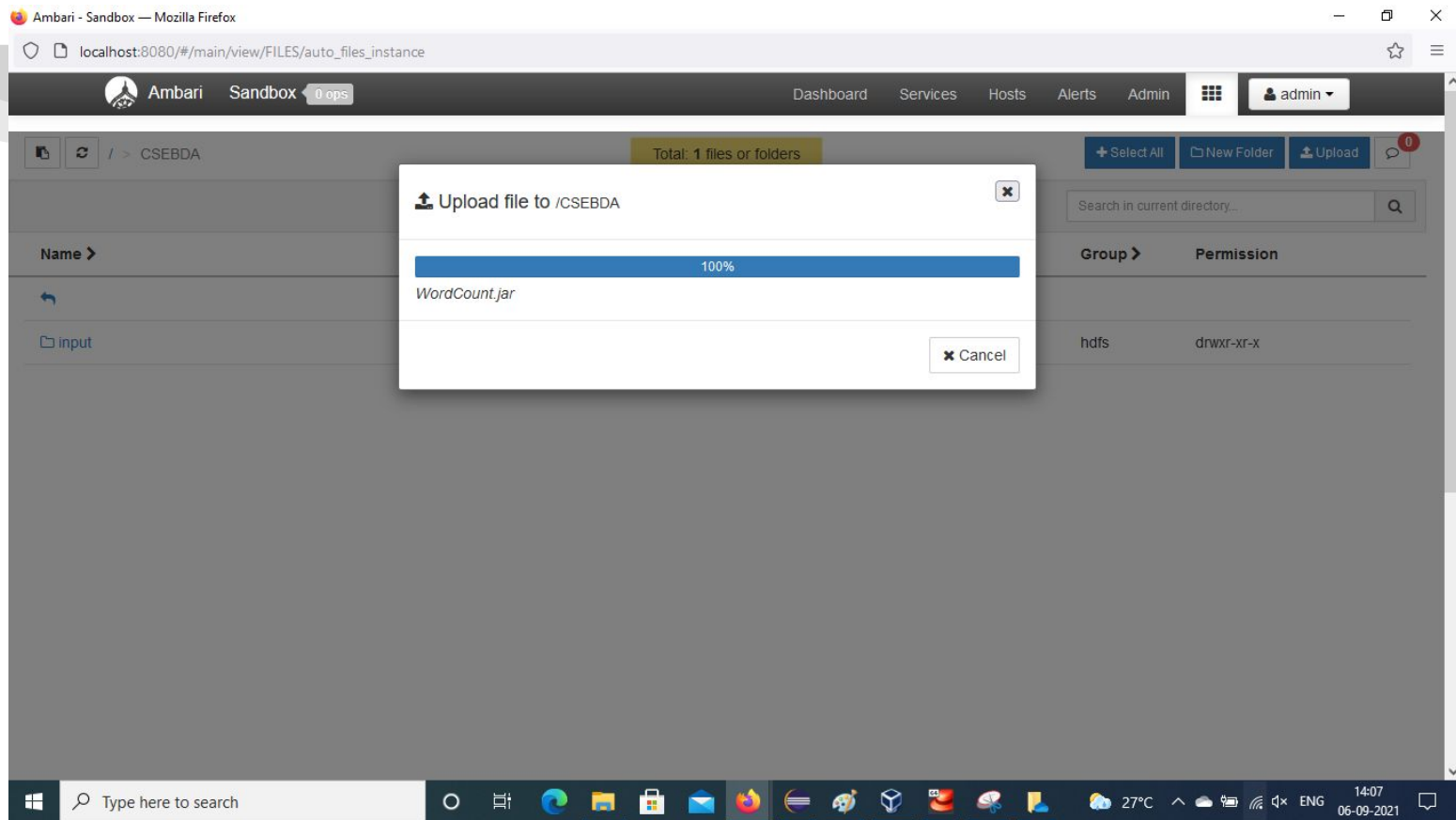
27°C

14:02 06-09-2021



Upload the jar WordCount.jar







Ambari

Sandbox

0 ops

Dashboard

Services

Hosts

Alerts

Admin



admin ▾



/ > CSEBDA

Total: 2 files or folders

+ Select All

New Folder

Upload



0

Search in current directory...



Name >	Size >	Last Modified >	Owner >	Group >	Permission
↶					
WordCount.jar	3.6 kB	2021-08-24 19:13	admin	hdfs	-rw-r--r--
input	--	2021-08-24 19:12	root	hdfs	drwxr-xr-x



Type here to search



14:08

06-09-2021




Copy the jar file into HDP sandbox

```
← → ↻ localhost:4200
[root@sandbox-hdp ~]# hadoop fs -get /CSEBDA/WordCount.jar
[root@sandbox-hdp ~]#
```



Run the Mapreduce program

```
[root@sandbox-hdp ~]# hadoop jar WordCount3.jar Wordcount /CSE411/input /CSE411/output
21/08/24 09:22:16 INFO client.RMProxy: Connecting to ResourceManager at sandbox-hdp.hortonworks.com/172.18.0.2:8032
21/08/24 09:22:19 INFO client.AHSPProxy: Connecting to Application History server at sandbox-hdp.hortonworks.com/172.18.0.2:10200
21/08/24 09:22:24 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your
  application with ToolRunner to remedy this.
21/08/24 09:22:27 INFO input.FileInputFormat: Total input paths to process : 1
21/08/24 09:22:27 INFO mapreduce.JobSubmitter: number of splits:1
21/08/24 09:22:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1629786961701_0001
21/08/24 09:22:41 INFO impl.YarnClientImpl: Submitted application application_1629786961701_0001
21/08/24 09:22:42 INFO mapreduce.Job: The url to track the job: http://sandbox-hdp.hortonworks.com:8088/proxy/application_1629786961701_0001/
21/08/24 09:22:42 INFO mapreduce.Job: Running job: job_1629786961701_0001
21/08/24 09:28:21 INFO mapreduce.Job: Job job_1629786961701_0001 running in uber mode : false
21/08/24 09:28:22 INFO mapreduce.Job:  map 0% reduce 0%
21/08/24 09:29:15 INFO mapreduce.Job:  map 100% reduce 0%
21/08/24 09:29:39 INFO mapreduce.Job:  map 100% reduce 100%
21/08/24 09:29:45 INFO mapreduce.Job: Job job_1629786961701_0001 completed successfully
21/08/24 09:29:45 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=205
        FILE: Number of bytes written=306241
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=221
```

HDFS: Number of bytes written=92
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)=45627
Total time spent by all reduces in occupied slots (ms)=21667
Total time spent by all map tasks (ms)=45627
Total time spent by all reduce tasks (ms)=21667
Total vcore-milliseconds taken by all map tasks=45627
Total vcore-milliseconds taken by all reduce tasks=21667
Total megabyte-milliseconds taken by all map tasks=11406750
Total megabyte-milliseconds taken by all reduce tasks=5416750

Map-Reduce Framework

Map input records=3
Map output records=18
Map output bytes=163
Map output materialized bytes=205
Input split bytes=129
Combine input records=0
Combine output records=0



Reduce input groups=13
Reduce shuffle bytes=205
Reduce input records=18
Reduce output records=13
Spilled Records=36
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=537
CPU time spent (ms)=2690
Physical memory (bytes) snapshot=303505408
Virtual memory (bytes) snapshot=3903258624
Total committed heap usage (bytes)=142082048

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=92

File Output Format Counters

Bytes Written=92



View the output file

```
[root@sandbox-hdp ~]# hadoop fs -cat /CSE411/output/part-r-00000
COUNT      3
EACH        1
FREQUENCY    1
HELLO        1
HI           1
IS           2
OF           1
PROGRAM      1
THF          1
THIS         1
TO           1
USED         1
WORD         3
```