CONTINUOUS ASSESSMENTS (C.A)-3

ST_NAME   : -   **EKHLAKH AHMAD**
REG NO.   : -   **12209166**
ROLL NO. : -   **RD2215B50**
SECTION   : -   **D2215**
GROUP   : -   **2**
DATE   : -   **10/04/2023**

Q.1. implement the WordCount mapreduce in Hadoop.
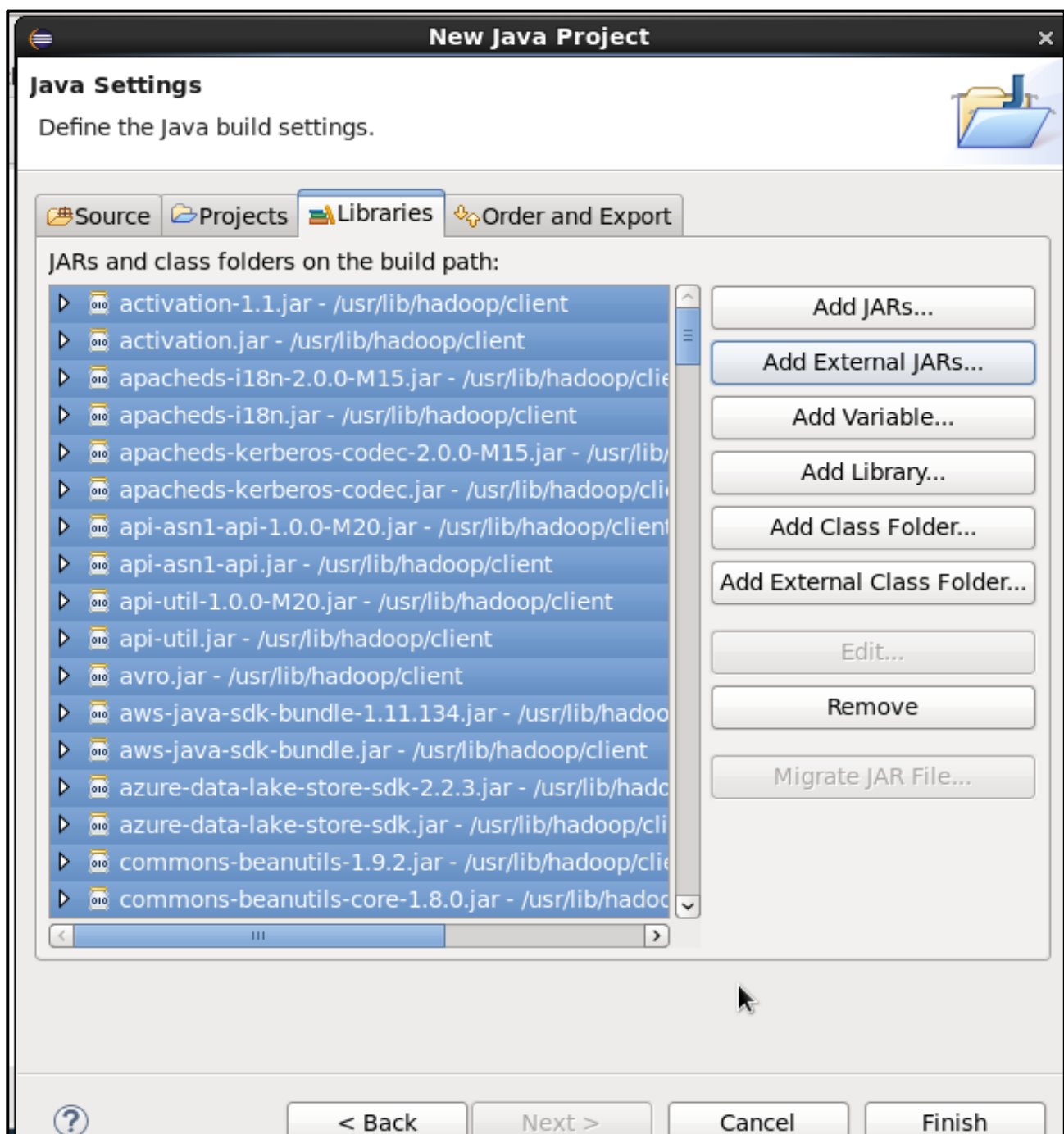
JAR Export

JAR File Specification

Define which resources should be exported into the JAR.

Select the resources to export:

▷ ⊟ 🗁 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

JAR Export

JAR File Specification
Define which resources should be exported into the JAR.

Select the resources to export:

▶ ☐ 🗁 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
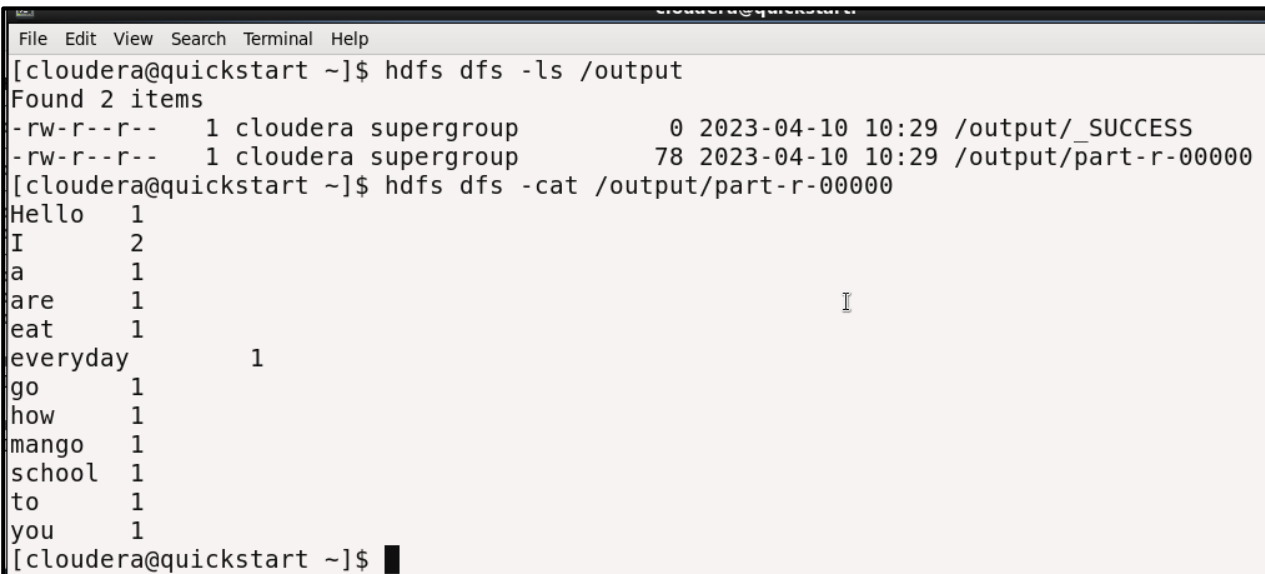☐ Overwrite existing files without warning

[< Back] [Next >] [Cancel] [Finish]

---
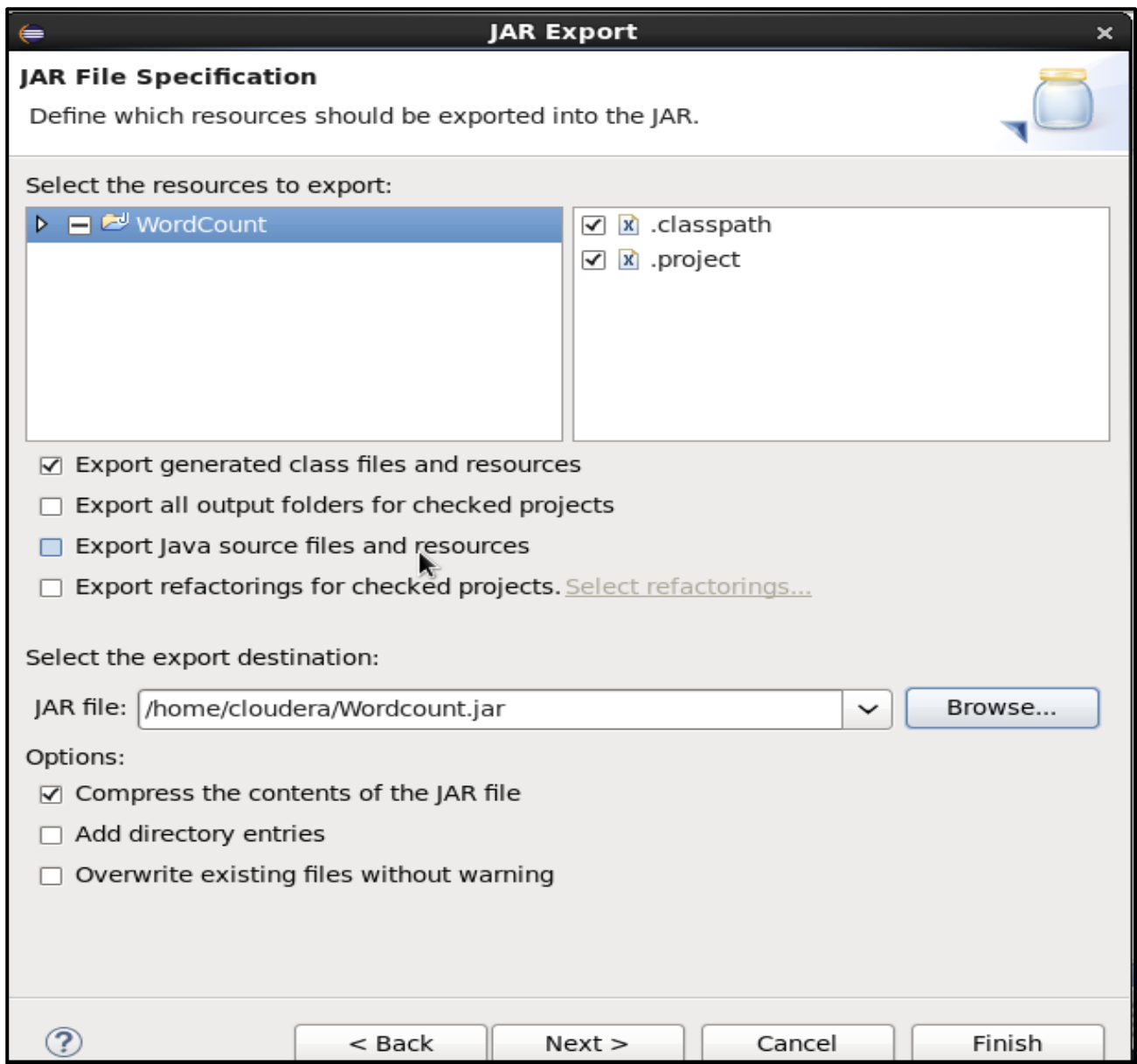
cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ gedit ekhlakh.txt
[cloudera@quickstart ~]$ cat ekhlakh.txt
Hello how are you
I eat a mango
I go to school everyday
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/ekhlakh.txt /
[cloudera@quickstart ~]$ hdfs dfs -ls /
Found 9 items
drwxrwxrwx   - hdfs      supergroup          0 2017-10-23 09:15 /benchmarks
-rw-r--r--   1 cloudera supergroup         56 2023-04-10 10:26 /ekhlakh.txt
drwxr-xr-x   - hbase     supergroup          0 2023-04-10 14:27 /hbase
drwxr-xr-x   - cloudera supergroup          0 2023-04-10 15:01 /process
drwxr-xr-x   - solr      solr                0 2017-10-23 09:18 /solr
-rw-r--r--   1 cloudera supergroup      61117 2023-04-10 06:24 /titanic_data.txt
drwxrwxrwt   - hdfs      supergroup          0 2023-02-23 13:50 /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 ~]$
```

Command for wordcount –

Hadoop jar /home/cloudera/WordCount.jar /ekhlakh.txt /output

```
File  Edit  View  Search  Terminal  Help
[cloudera@quickstart ~]$ hdfs dfs -ls /output
Found 2 items
-rw-r--r--   1 cloudera supergroup          0 2023-04-10 10:29 /output/_SUCCESS
-rw-r--r--   1 cloudera supergroup         78 2023-04-10 10:29 /output/part-r-00000
[cloudera@quickstart ~]$ hdfs dfs -cat /output/part-r-00000
Hello   1
I       2
a       1
are     1
eat     1
everyday        1
go      1
how     1
mango   1
school  1
to      1
you     1
[cloudera@quickstart ~]$
```