



CSE488: Big Data Analytics [SPRING 2023]

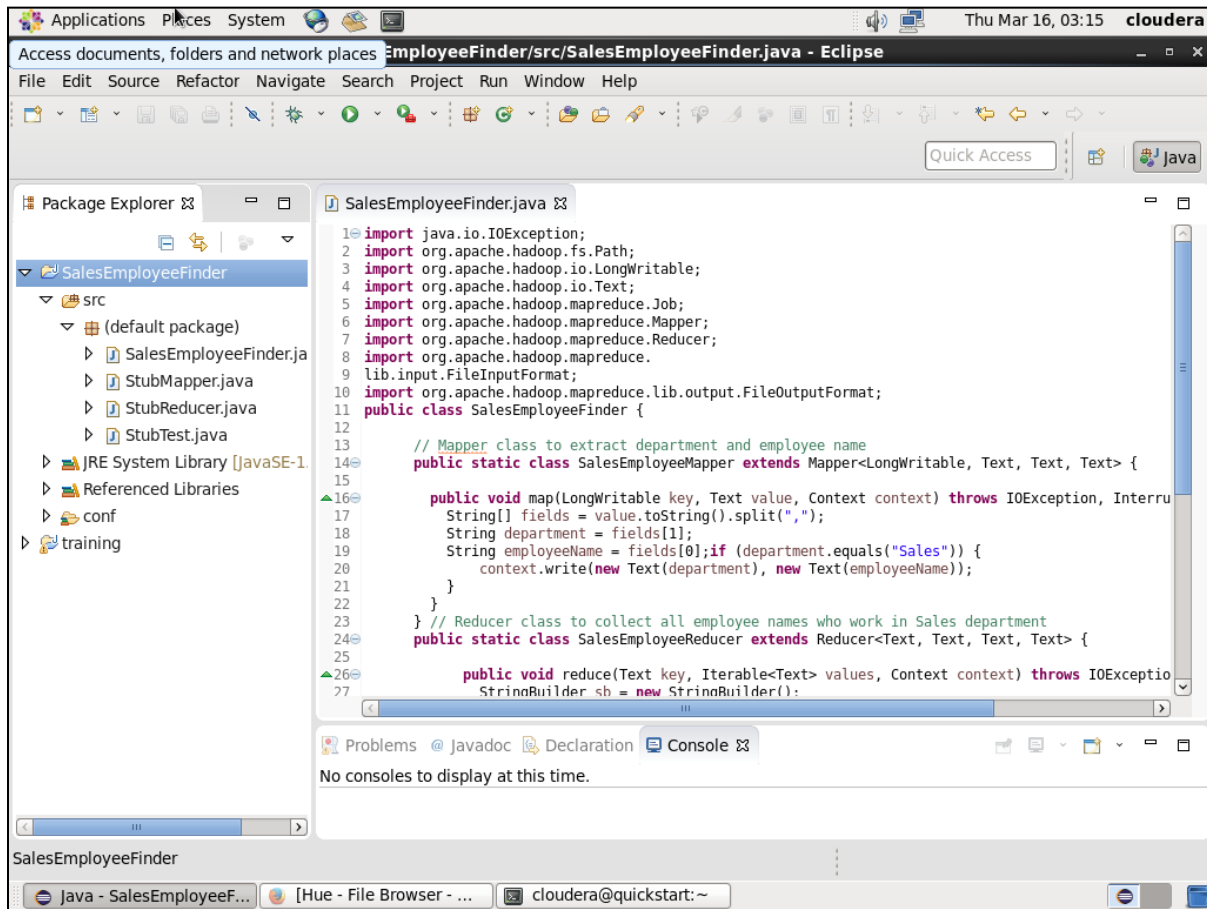
Assignment 4 Hadoop MapReduce

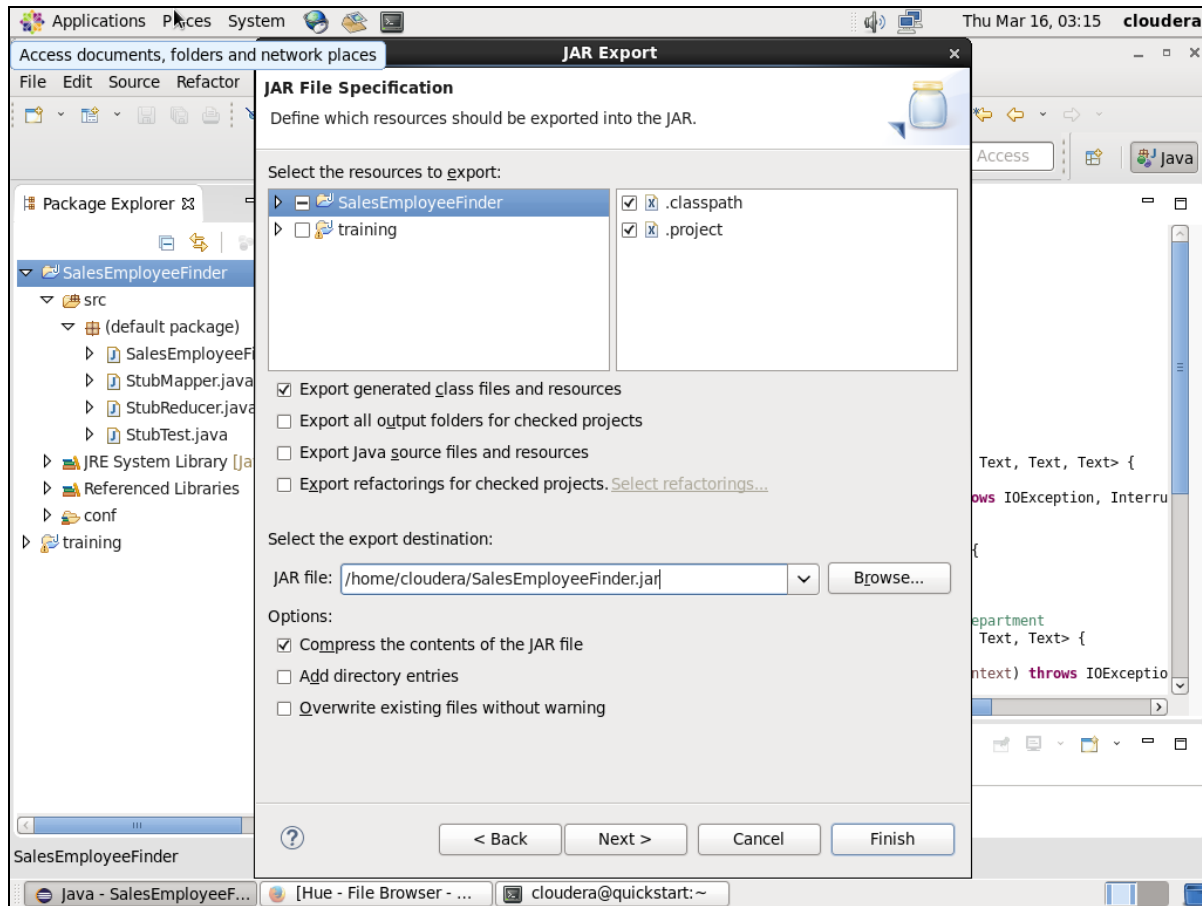
Submitted by:

Student ID: 2020-1-60-215

Student Name: Md. Abdul Ahad Rifat

Problem 1





Applications Places System Thu Mar 16, 03:15 cloudera

Browse and run installed applications cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/SalesEmployeeFinder.jar SalesEmployeeFinder /user/cloudera/SalesEmployeeFinder/sample_dataset_1.csv /user/cloudera/SalesEmployeeFinder/output.csv
23/03/16 03:13:05 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/16 03:13:06 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
23/03/16 03:13:06 INFO input.FileInputFormat: Total input paths to process : 1
23/03/16 03:13:06 INFO mapreduce.JobSubmitter: number of splits:1
23/03/16 03:13:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1678959848727_0002
23/03/16 03:13:07 INFO impl.YarnClientImpl: Submitted application application_1678959848727_0002
23/03/16 03:13:07 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1678959848727_0002/
23/03/16 03:13:07 INFO mapreduce.Job: Running job: job_1678959848727_0002
23/03/16 03:13:13 INFO mapreduce.Job: Job job_1678959848727_0002 running in uber mode : false
23/03/16 03:13:13 INFO mapreduce.Job: map 0% reduce 0%
23/03/16 03:13:23 INFO mapreduce.Job: map 100% reduce 0%
23/03/16 03:13:32 INFO mapreduce.Job: map 100% reduce 100%
23/03/16 03:13:32 INFO mapreduce.Job: Job job_1678959848727_0002 completed successfully
23/03/16 03:13:32 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=27
        FILE: Number of bytes written=286603
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=774
        HDFS: Number of bytes written=20
        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
```

Java - SalesEmployeeF... [Hue - File Browser - ... cloudera@quickstart:~

Applications Places System Thu Mar 16, 03:16 cloudera

Browse and run installed applications Hue - File Browser - Mozilla Firefox

Hue - File Browser x

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/Sales

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

Query Search data and saved document Jobs cloudera

default (0) Tables The database has no tables

View as binary

Edit file

Download

View file location

Refresh

Last modified 03/16/2023 10:13 AM

Page 1 to 1 of 1

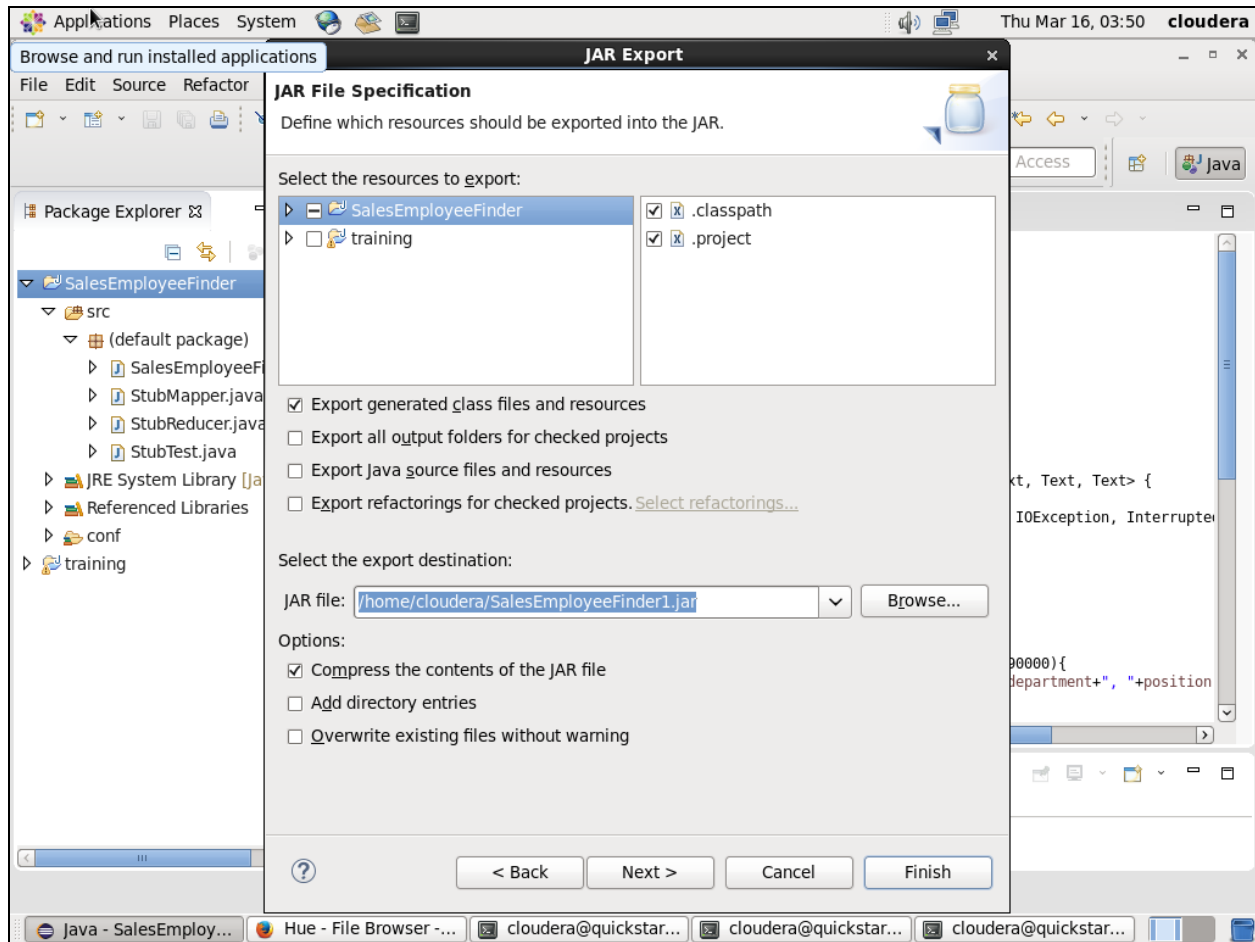
Home

/ user / cloudera / SalesEmployeeFinder / output.csv / part-r-00000

Sales Jasmine Wong,

java - SalesEmployeeF... Hue - File Browser - M... cloudera@quickstart:~

Problem 2



Applications Places System Thu Mar 16, 03:50 cloudera

Browse and run installed applications cloudera@quickstart:~

File Edit View Search Terminal Help

```
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/SalesEmployeeFinder1.jar SalesEmployeeFinder /user/cloudera/SalesEmployeeFinder/sample_dataset_1.csv /user/cloudera/SalesEmployeeFinder/salary.csv
23/03/16 03:49:31 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/03/16 03:49:31 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool in
terface and execute your application with ToolRunner to remedy this.
23/03/16 03:49:31 INFO input.FileInputFormat: Total input paths to process : 1
23/03/16 03:49:31 INFO mapreduce.JobSubmitter: number of splits:1
23/03/16 03:49:32 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1678959848727_0004
23/03/16 03:49:32 INFO impl.YarnClientImpl: Submitted application application_1678959848727_0004
23/03/16 03:49:32 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_16789598487
27_0004/
23/03/16 03:49:32 INFO mapreduce.Job: Running job: job_1678959848727_0004
23/03/16 03:49:38 INFO mapreduce.Job: Job job_1678959848727_0004 running in uber mode : false
23/03/16 03:49:38 INFO mapreduce.Job: map 0% reduce 0%
23/03/16 03:49:44 INFO mapreduce.Job: map 100% reduce 0%
23/03/16 03:49:52 INFO mapreduce.Job: map 100% reduce 100%
23/03/16 03:49:53 INFO mapreduce.Job: Job job_1678959848727_0004 completed successfully
23/03/16 03:49:53 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=260
        FILE: Number of bytes written=287069
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=774
        HDFS: Number of bytes written=217
        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)=2950
        Total time spent by all reduces in occupied slots (ms)=5565
        Total time spent by all map tasks (ms)=2950
        Total time spent by all reduce tasks (ms)=5565
        Total vcore-milliseconds taken by all map tasks=2950
        Total vcore-milliseconds taken by all reduce tasks=5565
```

Java - SalesEmploy... Hue - File Browser ... cloudera@quickstar... cloudera@quickstar... cloudera@quickstar...

```
Applications Places System Thu Mar 16, 03:52 cloudera
Access documents, folders and network places cloudera@quickstart:~
File Edit View Search Terminal Help
23/03/16 03:49:52 INFO mapreduce.Job: map 100% reduce 100%
23/03/16 03:49:53 INFO mapreduce.Job: Job job_1678959848727_0004 completed successfully
23/03/16 03:49:53 INFO mapreduce.Job: Counters: 49
  File System Counters
    FILE: Number of bytes read=260
    FILE: Number of bytes written=287069
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=774
    HDFS: Number of bytes written=217
    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)=2950
    Total time spent by all reduces in occupied slots (ms)=5565
    Total time spent by all map tasks (ms)=2950
    Total time spent by all reduce tasks (ms)=5565
    Total vcore-milliseconds taken by all map tasks=2950
    Total vcore-milliseconds taken by all reduce tasks=5565
    Total megabyte-milliseconds taken by all map tasks=3020800
    Total megabyte-milliseconds taken by all reduce tasks=5698560
  Map-Reduce Framework
    Map input records=11
    Map output records=5
    Map output bytes=244
    Map output materialized bytes=260
    Input split bytes=151
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=260
    Reduce input records=5
    Reduce output records=1
    Spilled Records=10
```


ApplicationsPlacesSystem

Thu Mar 16, 04:00cloudera

Access documents, folders and network placesHue - File Browser - Mozilla Firefox

Hue - File Browser

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/Sales

Search

ClouderaHueHadoopHBaseImpalaSparkSolrOozieCloudera ManagerGetting Started

Query

Search data and saved document

Jobs

cloudera

default

Tables

(0)

The database has no tables

Search for file name

Actions

Move to trash

Home

/ user / cloudera / SalesEmployeeFinder

Trash

	Name	Size	User	Group	Permissions	Date
<input type="checkbox"/>			cloudera	cloudera	drwxr-xr-x	March 16, 2016
<input type="checkbox"/>			cloudera	cloudera	drwxr-xr-x	March 16, 2016
<input type="checkbox"/>	output.csv		cloudera	cloudera	drwxr-xr-x	March 16, 2016
<input type="checkbox"/>	salary.csv		cloudera	cloudera	drwxr-xr-x	March 16, 2016
<input type="checkbox"/>	sample_dataset_1.csv	623 bytes	cloudera	cloudera	-rw-r--r--	March 16, 2016

Show

45

of 3 items

Page

1

of 1

Java - SalesEmploy...Hue - File Browser - ...cloudera@quickstar...cloudera@quickstar...cloudera@quickstar...

ApplicationsPlacesSystem

Thu Mar 16, 03:56cloudera

Access documents, folders and network placesHue - File Browser - Mozilla Firefox

Hue - File Browser

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/S

Search

ClouderaHueHadoopHBaseImpalaSparkSolrOozieCloudera ManagerGetting Started

HUE

Query

Search data and saved document

Jobs

cloudera

Search for file name

Actions

Move to trash

/ user / cloudera / SalesEmployeeFinder / salary.csv

Trash

	Name	Size	User	Group	Permissions	Date
			cloudera	cloudera	drwxr-xr-x	March 16, 2023 0
	.		cloudera	cloudera	drwxr-xr-x	March 16, 2023 0
	_SUCCESS	0 bytes	cloudera	cloudera	-rw-r--r--	March 16, 2023 0
	part-r-00000	222 bytes	cloudera	cloudera	-rw-r--r--	March 16, 2023 0

Show45of 2 itemsPage1of 1

default

Tables

(0)

The database has no tables

Java - SalesEmploy...

Hue - File Browser - ...

cloudera@quickstar...

cloudera@quickstar...

cloudera@quickstar...

Applications Places System Thu Mar 16, 03:57 cloudera

Browse and run installed applications Hue - File Browser - Mozilla Firefox

Hue - File Browser x

quickstart.cloudera:8888/hue/filebrowser/view=/user/cloudera/Sales

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

HUE Query Search data and saved document Jobs cloudera

default (0) Tables The database has no tables

View as binary Home Edit file / user / cloudera / SalesEmployeeFinder / salary.csv / part-r-00000 Download View file location Refresh

Page 1 to 1 of 1

Result Ava Martinez, Marketing, Marketing Manager
Jasmine Wong, Sales, Sales Representative
Sarah Chen, Human Resources, HR Manager
Michael Lee, Finance, Financial Analyst
John Smith, Engineering, Software Developer

Last modified 03/16/2023 10:56 AM

Java - SalesEmploy... Hue - File Browser ... cloudera@quickstar... cloudera@quickstar... cloudera@quickstar...