



# **CSE488: Big Data Analytics [SPRING 2023]**

## **Assignment 3 Hadoop MapReduce**

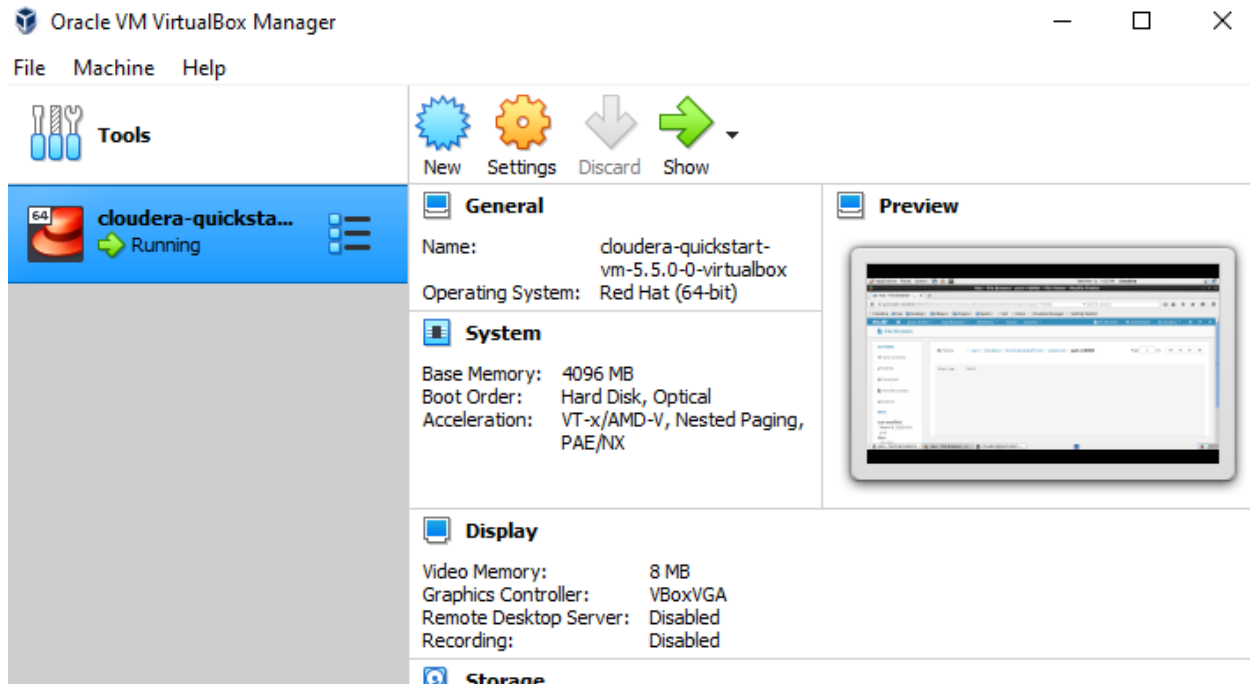
**Submitted by:**

Student ID: 2020-1-60-215

Student Name: Md. Abdul Ahad Rifat

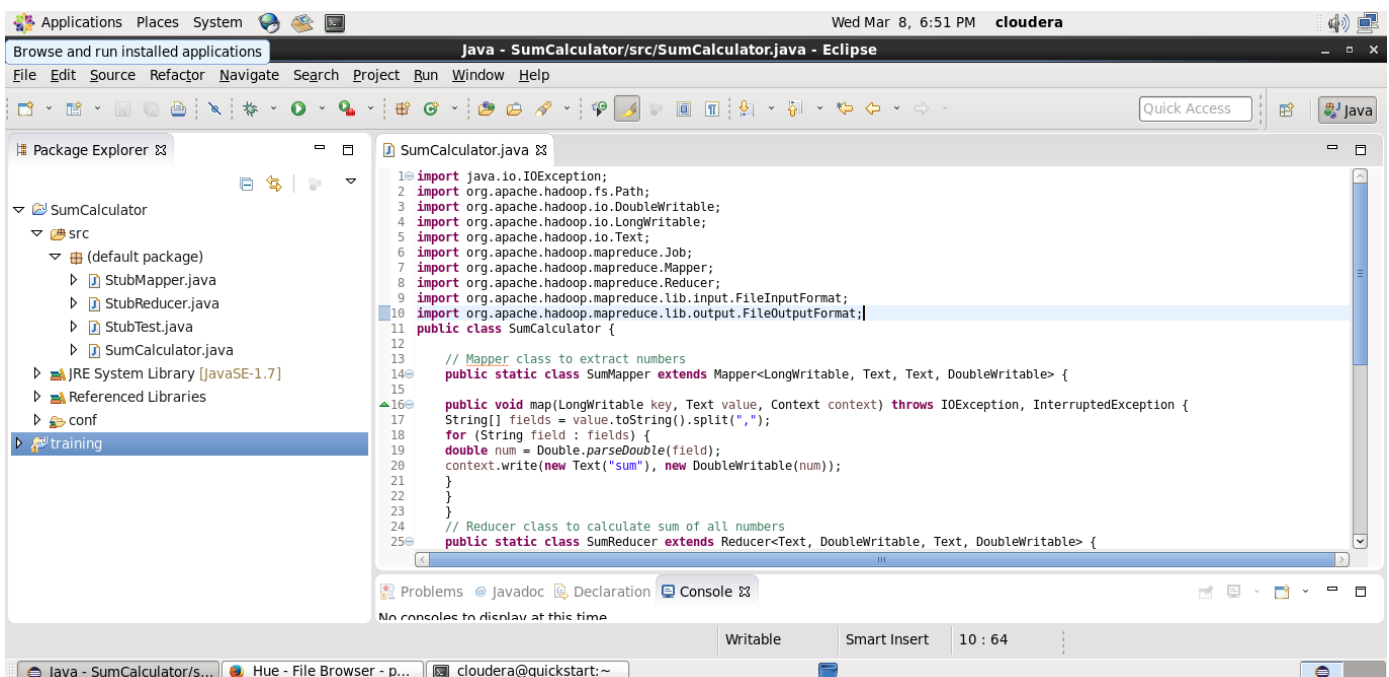
## Step 1:

Open Virtual machine and run cloudera OS



## Step 2:

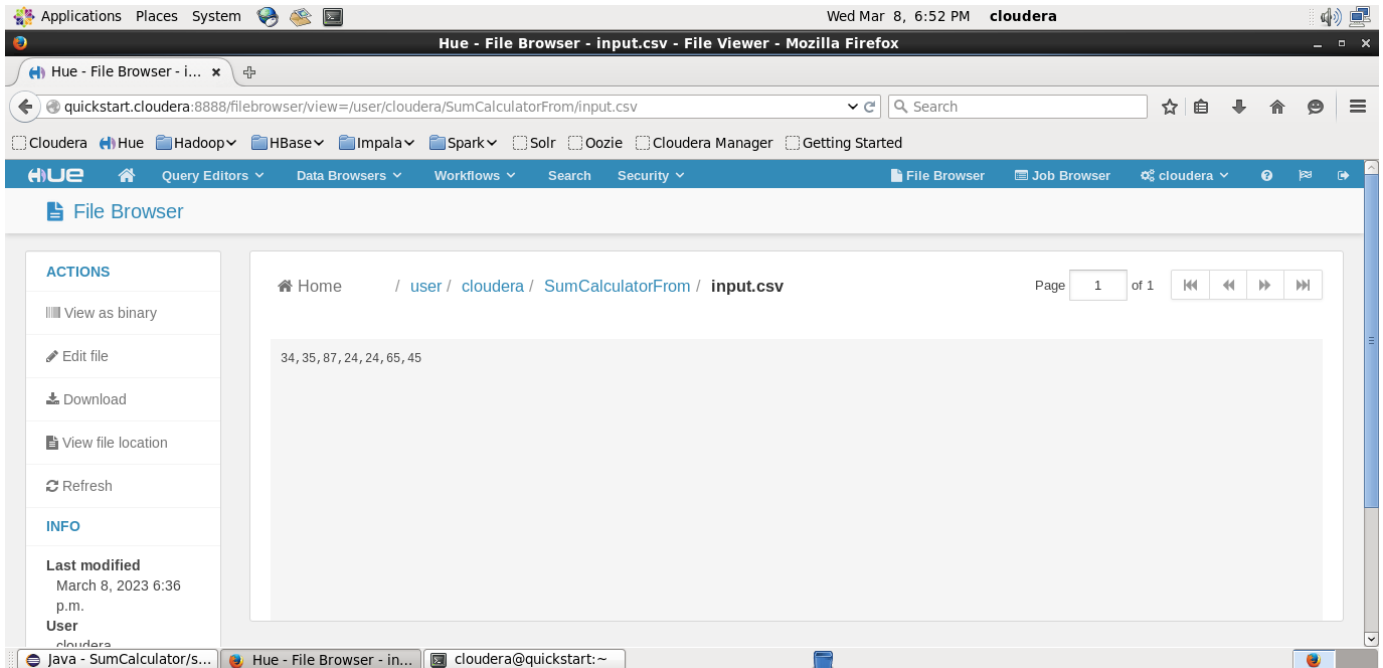
Here we copied training the project and rename it to SumCalculator and paste the code Then create one driver class same name as project (basically rename it) To change file name press F2



### Step 3:

Open the browser and goto the hue section and create new directory(SumCalculatorFrom) then under the directory create one csv file where the input is 34,35,87,24,24,65,45

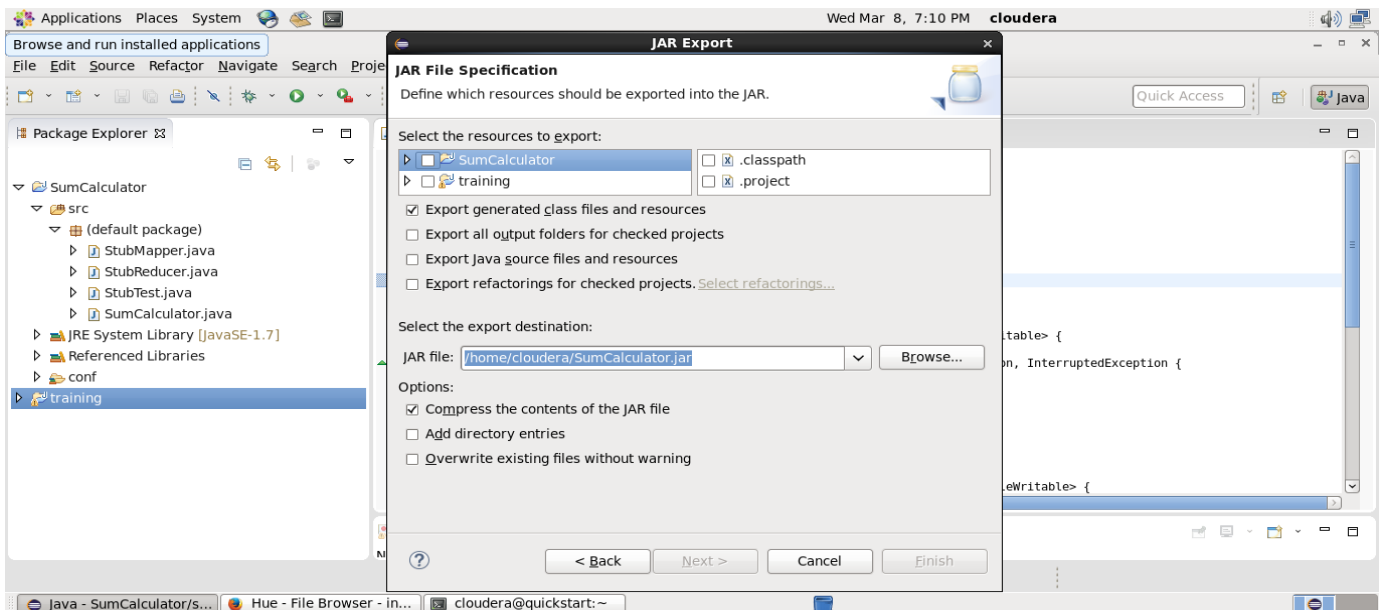
[/user/cloudera/SumCalculatorFrom/input.csv](#)



### Step 4:

Here we create one jar file under SumCalculator Project

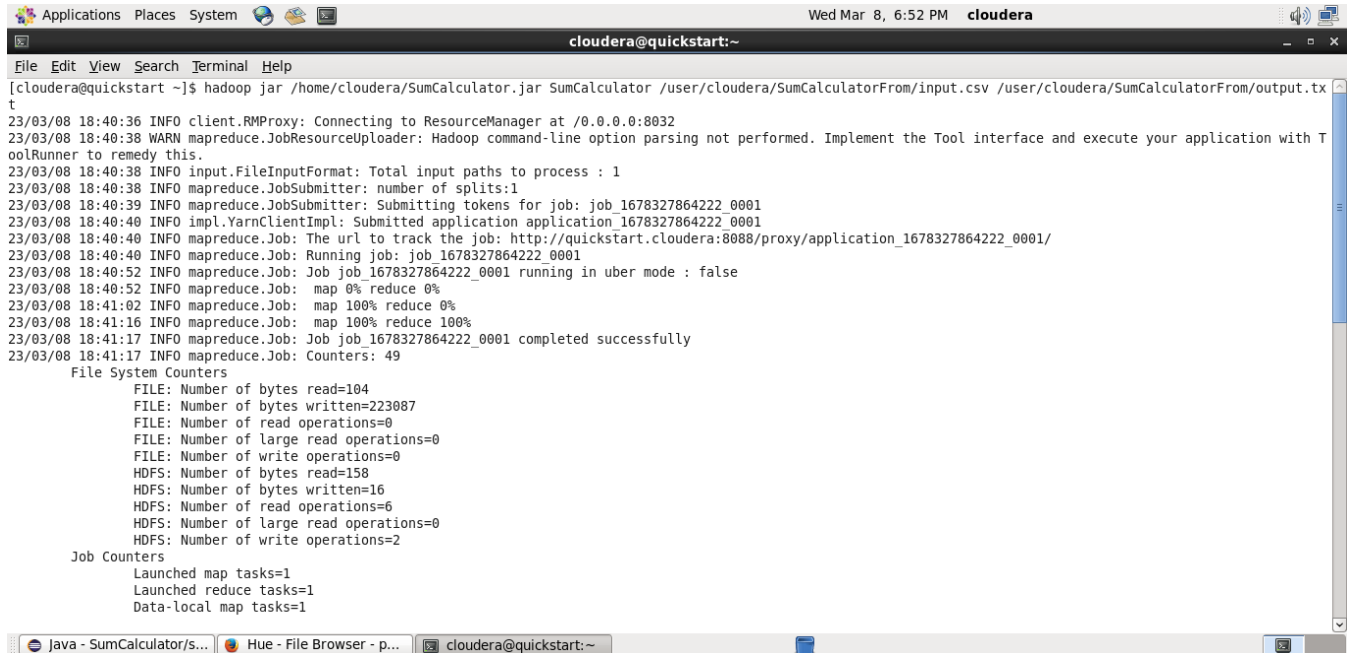
Jar file location [/home/cloudera/SumCalculator.jar](#)



## Step 5:

In this section we opening the terminal and execute the command (**hadoop jar**)

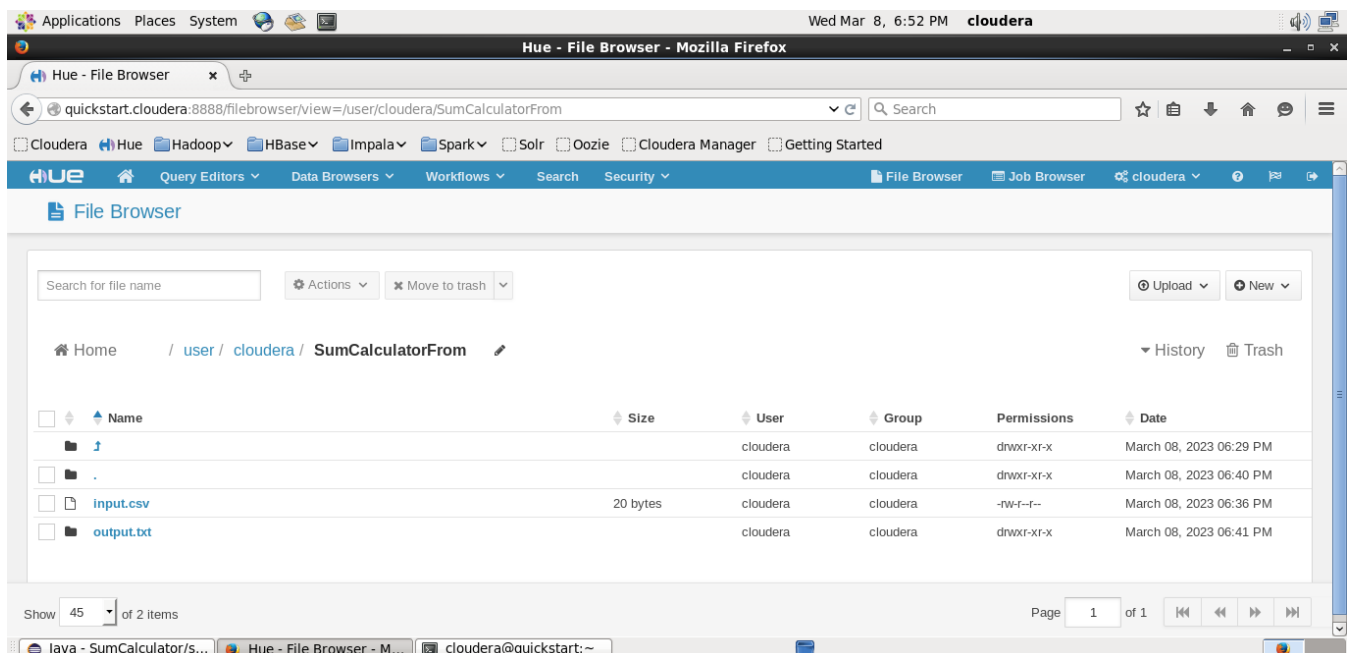
```
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/SumCalculator.jar SumCalculator  
/user/cloudera/SumCalculatorFrom/input.csv /user/cloudera/SumCalculatorFrom/output.txt
```



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hadoop jar /home/cloudera/SumCalculator.jar SumCalculator /user/cloudera/SumCalculatorFrom/input.csv /user/cloudera/SumCalculatorFrom/output.txt  
23/03/08 18:40:36 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032  
23/03/08 18:40:38 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.  
23/03/08 18:40:38 INFO input.FileInputFormat: Total input paths to process : 1  
23/03/08 18:40:38 INFO mapreduce.JobSubmitter: number of splits:1  
23/03/08 18:40:39 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1678327864222_0001  
23/03/08 18:40:40 INFO impl.YarnClientImpl: Submitted application application_1678327864222_0001  
23/03/08 18:40:40 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1678327864222_0001/  
23/03/08 18:40:40 INFO mapreduce.Job: Running job: job_1678327864222_0001  
23/03/08 18:40:52 INFO mapreduce.Job: Job job_1678327864222_0001 running in uber mode : false  
23/03/08 18:40:52 INFO mapreduce.Job: map 0% reduce 0%  
23/03/08 18:41:02 INFO mapreduce.Job: map 100% reduce 0%  
23/03/08 18:41:16 INFO mapreduce.Job: map 100% reduce 100%  
23/03/08 18:41:17 INFO mapreduce.Job: Job job_1678327864222_0001 completed successfully  
23/03/08 18:41:17 INFO mapreduce.Job: Counters: 49  
File System Counters  
FILE: Number of bytes read=104  
FILE: Number of bytes written=223087  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=158  
HDFS: Number of bytes written=16  
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
```

## Step 6:

Output file created under SumCalculator directory



Hue - File Browser - Mozilla Firefox

quickstart.cloudera:8888/filebrowser/view=/user/cloudera/SumCalculatorFrom

Cloudera Hue HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

File Browser

Search for file name Actions Move to trash Upload New

Home / user / cloudera / SumCalculatorFrom History Trash

Name	Size	User	Group	Permissions	Date
.		cloudera	cloudera	drwxr-xr-x	March 08, 2023 06:29 PM
input.csv	20 bytes	cloudera	cloudera	-rw-r--r--	March 08, 2023 06:36 PM
output.txt		cloudera	cloudera	drwxr-xr-x	March 08, 2023 06:41 PM

Show 45 of 2 items Page 1 of 1

## Step 7:

**Part-r-00000**

**Output file is the txt format**

**Output ( Total Sum 314.0)**

