

**CSE488: Big Data Analytics**

**[SPRING 2023]**

**Assignment**

**MapReduce to Calculate SUM**

**Submitted by:**

Student ID: 2019-3-60-046

Student Name: Mohsenul Kabir Mithun

**Creating a Hadoop project in Eclipse: (MapReduce used for** SumCalculator**)**

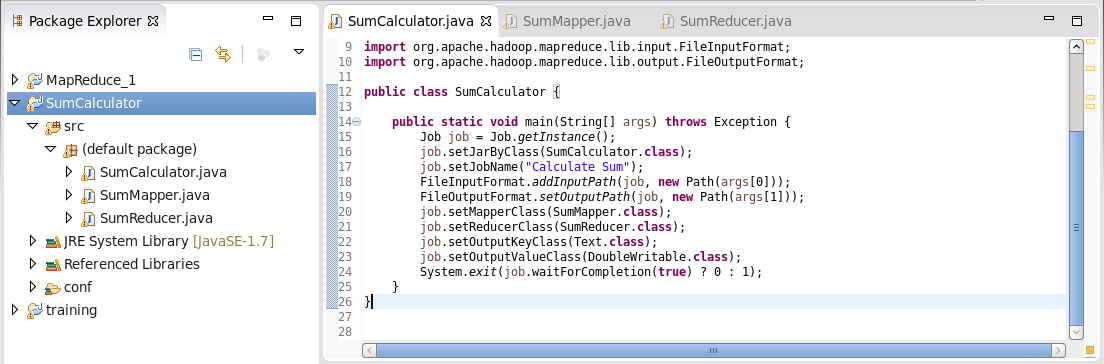
**🡪 MapReduce Program for** SumCalculator**:**

1. Open Eclipse IDE

2. Copy 'training' project and paste it on same place by name it "SumCalculator "

3. Expand SumCalculator 🡪 src 🡪 (default package)

4. Rename 4 of 3 java files are " SumCalculator.java", " SumReducer.java", " SumMapper.java" and replace the code that I provide bellow,



Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

5. Right click on ' SumCalculator ' project 🡪  export 🡪 java 🡪 select jar 🡪 Next 🡪 give a jar file name as "Sum.jar" 🡪 Next 🡪 Next 🡪 Finish

6. Avoid if any error message occurs.

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

7. Go to http://quickstart.cloudera:8888/accounts/login/?next=/

8. Login by using Username: cloudera & Password: cloudera

9. Go to file browser that you see upper right side of the browser

10. Create new directory by click on New 🡪 Directory 🡪 provide name as ‘sum’ 🡪 go that directory create new file as file1.csv

11. Go to that file1.csv 🡪 click on 'edit file' option and add some text as coma separated. Then save it.

Graphical user interface, text, application

Description automatically generated

12. open CMD and write bellow command

hadoop jar sum.jar SumCalculator  /user/cloudera/sum/file1.txt  sumout

More generic way to explain:

🡪 hadoop jar <jar file> <main class> <input path> <output path>

Text, letter

Description automatically generated

Text

Description automatically generated

13. To see output go to ‘Web Browser’ where you create a folder and add csv file. Here you also see another folder name as "sumout" In this folder you will find the output.

Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Thank you 🙂