

**CSE488: Big Data Analytics**

**[SPRING 2023]**

**Lab-6**

**MapReduce program to find the occurrences of each page, most visited page, least visited page, and frequency of pairs.**

**Submitted by:**

Student ID: 2019-3-60-046

Student Name: Mohsenul Kabir Mithun

**Creating a Hadoop project in Eclipse:**

**1.  Find the occurrences of each page (numbered as 1, 2, 3, and so on).**

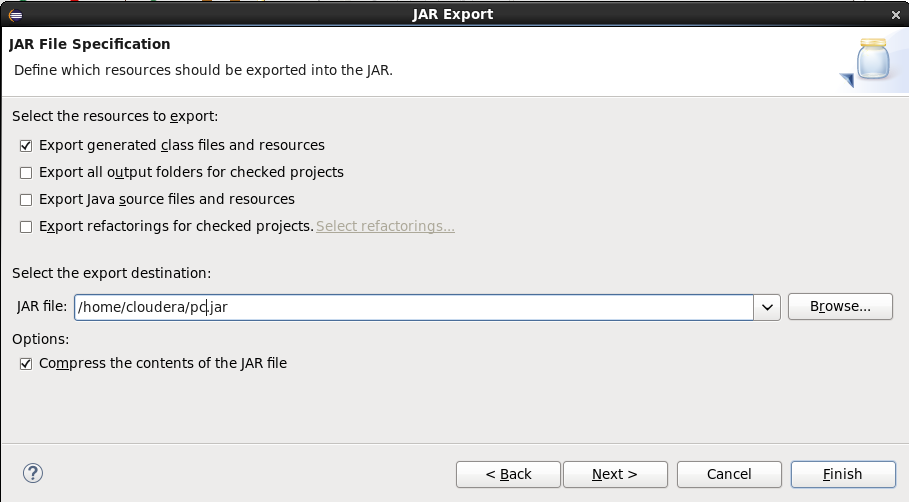
Source Code SS: (Mapper and Reducer Class)

Graphical user interface, text, application, email

Description automatically generated A picture containing text

Description automatically generated

Creating Jar file:



Creating Input file:

Graphical user interface, text, application

Description automatically generated

CMD run the program:

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Output:

Graphical user interface, text, application, email

Description automatically generated

**2.** **Find the most visited page. You may use the output file generated from Question 1 as an input file to solve this problem.**

Source Code SS: (Mapper and Reducer Class)

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generated with medium confidence

Creating Jar file:

Graphical user interface, text, application, email

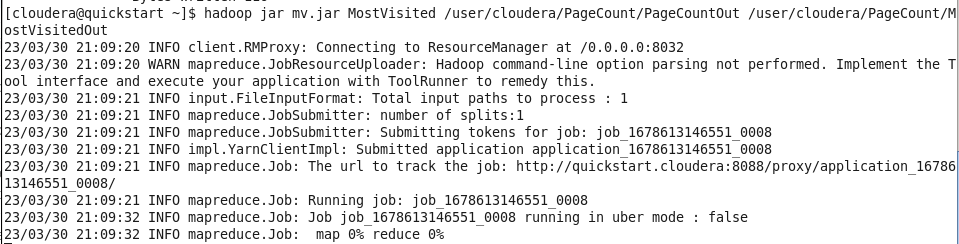
Description automatically generated

Creating Input file:

Graphical user interface, text, application

Description automatically generated

CMD run the program:



Text

Description automatically generated

Output:

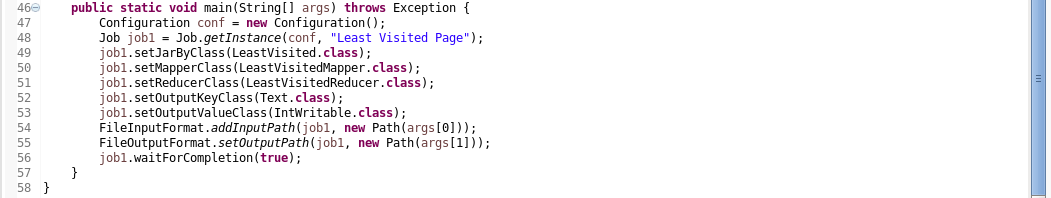
Graphical user interface, text, application, website

Description automatically generated

**3.** **Find the least visited page. You may use the output file generated from Question 1 as an input file to solve this problem.**

Source Code SS: (Mapper and Reducer Class)





Creating Jar file:

Graphical user interface, text, application, email

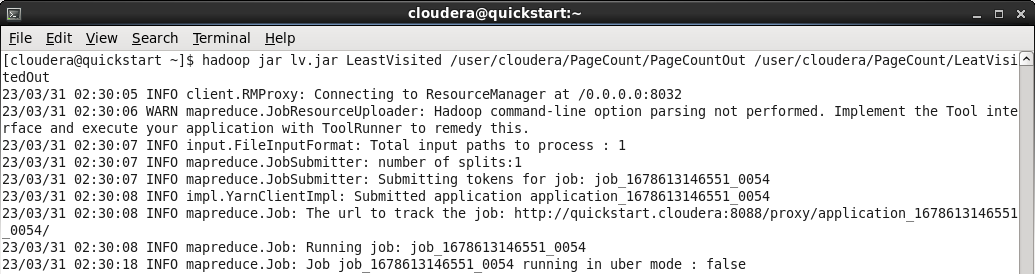
Description automatically generated

Creating Input file:

Graphical user interface, text, application

Description automatically generated

CMD run the program:



Graphical user interface, text

Description automatically generated

Output:

Graphical user interface, application, website

Description automatically generated

**4.** **Find the frequency of pairs. In this case, remove the duplicate numbers from each transaction as a data pre-processing step. Use the concept of Java set for removing duplicates.**

Source Code SS: (Mapper and Reducer Class)

Graphical user interface, text, application, email

Description automatically generated Graphical user interface

Description automatically generated with medium confidence

Creating Jar file:

Graphical user interface, text, application, email

Description automatically generated

Creating Input file:

Graphical user interface, text, application

Description automatically generated

CMD run the program:

Text

Description automatically generated

Text

Description automatically generated

Output:

