

ISOM3015 (001 & 002)

Assignment 1 (Due date: 14 October)

(e-copy submission before class in Moodle & hard-copy submission in class)

Q1) 60%

There is a small table called Employee with the following data:

Id	Name	Age	Gender	Salary
1201	gopal	45	Male	50,000
1202	manisha	40	Female	50,000
1203	khalil	34	Male	30,000
1204	prasanth	30	Male	30,000
1205	kiran	20	Male	40,000
1206	laxmi	25	Female	35,000
1207	bhavya	20	Female	15,000
1208	reshma	19	Female	15,000
1209	kranthi	22	Male	22,000
1210	Saish	24	Male	25,000
1211	Krishna	25	Male	25,000
1212	Arshad	28	Male	20,000
1213	lavanya	18	Female	8,000

We would like to write an application to process the input dataset to find the highest salaried employee by gender in different age groups (for example, below 20, between 21 to 30 and above 30).

Suppose the above data is saved as input.csv in the CentOS system in the “/home/hadoop/hadoopPartitioner”.

The map task accepts the key-value pairs as input from input.txt. as follows:

Input – Assuming that the key would be a pattern: “any special key + filename + line number” (example: key = @input1) and the value would be the data in that line (example: value = 1201, gopal,45,Male,50000), e.g. the key-value pair for the first line (first record) of input.txt would be:

```
<“@input1”, “1201,tgopal,45,Male,50000”>
```

- Implement a Java function map(String key, String value) to display the mapping result of a specific input record based on its corresponding key and value, e.g. it would display <xxx, yyy> while xxx indicates the key and yyy indicates the value.
- Implement a Java function getPartition(String key, String value) to return the partition number starting from 0.

- c) Implement a Java function `reduce(String key, String[] value)` to display the reducing value result for a particular key based on the string array of values for the particular key.

Sample output:

There are three output files generated:

Output in Part-00000

Female	15000
Male	40000

Output in Part-00001

Female	35000
Male	31000

Output in Part-00002

Female	55000
Male	50000

Assignment Project Exam Help

<https://powcoder.com>

Q2)

Suppose the file `input.csv` has been stored in Linux file system but you are not sure where it is. What is the command used to find out the location of the file? (5%)

Q3)

What is the command used to move `input.csv` file to `/root/data` directory suppose `/root/data` directory already exists and the current working directory contains `input.csv`. (5%)

Q4)

Inside Linux OS, what is the command if I would like to know the number of records inside `input.csv`? (5%)

Q5)

What is the command used to display every five records of `input.csv` each time? (5%)

Q6)

Suppose there are three related dataset files in my current directory; namely employee.csv, contact.csv and order.csv. What is the command to compress the three files into an archive file transaction.tar.gz in my current directory? (5%)

Q7)

I would like to de-compress the file in Q7) and store the files in a directory /opt after de-compression. What is the command? (5%)

Q8)

Suppose your current directory is /root/Desktop, you would like to display all the files with txt extension in ascending order files in the /root/Download directory and then store the results in a text file abc.txt in my current directory instead of displaying the results. What is the command? (5%)

Assignment Project Exam Help

Q9)

Suppose you are the root user and you have a shell script file (config.sh). The details of the file are shown below:

```
-rwxrwxrwx 1 root root 0 Nov  9 16:40 config.sh
```

Add WeChat powcoder

You would like to remove the write and executable mode for all other users except yourself. What is the command (just one single command)? (5%)