

Quiz #1

Due Apr 14 at 4pm

Points 100

Questions 6

Available after Apr 9 at 12am

Time Limit None

Instructions

Census Count MapReduce Program.

This is a warm-up assignment.

I will discuss this assignment in details on Thursday, April 10, 2020

Input Data Format:

```
<age><,><gender><,><city><,><state><,><country><,><salary>
```

Input Example To Be Used for Assignment, you
can NOT use any other input for this assignment

```
20,Male,Ames,IA,USA,25000
24,Male,Ames,IA,USA,29000
24,Female,Sunnyvale,CA,USA,57000
40,Male,Ames,IA,USA,43000
40,Male,Boone,IA,USA,44000
40,Female,Cupertino,CA,USA,48000
20,Male,Boone,IA,USA,24000
20,Female,Ames,IA,USA,26000
12,Male,Ames, IA,USA,1000
24,Male,Boone,IA,USA,34000
```

NOTE-1: If an input record does not follow/comply the
given input format, then that record is filtered out

(to be excluded by your MapReduce program).

NOTE-2: Age group of under 15 will be filtered out

Expected output:

Key	Value
=====	=====
<age>	<average-salary>
"unique-cities"	<list-of-unique-cities>
"number-of-males"	<number-of-males>"
"number-of-females"	<number-of-females>"

Your MapReduce Algorithm:

1. For this assignment, your MapReduce program is comprised of two functions:

map()
reduce()

2. Your MapReduce Solution must be a generic solution and work for any size data.

Attempt History

Attempt	Time	Score
---------	------	-------

	Attempt	Time	Score
LATEST	<u>Attempt 1</u>	117 minutes	100 out of 100

Score for this quiz: **100** out of 100

Submitted Apr 14 at 2:12am

This attempt took 117 minutes.

Question 1

20 / 20 pts

Write a map() function to handle the given input.

Input for your map() will be a a pair of (key, value) where key is the recored number of input and value is a single record of input; for example the first record will be passed to a map() function as (1, "20,Male,Ames,IA,USA,25000"), second record as: (2, "24,Male,Ames,IA,USA,29000"), and so on...

Your Answer:

```
map(key, value) {
```

```
    #tokenize the input
```

```
    tokens = value.split(",")
```

```
    #Identify if the record is in proper format or age is less than 15
```

```
    if( (len(tokens) != 6) & (tokens[0] < 15) ) {
```

```
#record is not in proper format, so ignore it.  
  
return  
  
}
```

```
#read the values in variables for better readability
```

```
age = tokens[0]  
salary = tokens[5]  
gender = tokens[1]  
city = tokens[2]
```

```
#generate required (key, value) pairs
```

```
emit(age, salary)  
emit(gender, 1)  
emit("unique-cities", city)
```

```
} #map function ends
```

Question 2

20 / 20 pts

Show output of all mappers: show your work as detail as possible

Your Answer:

Input is:

```
1, "20,Male,Ames,IA,USA,25000"  
2, "24,Male,Ames,IA,USA,29000"  
3, "24,Female,Sunnyvale,CA,USA,57000"  
4, "40,Male,Ames,IA,USA,43000"  
5, "40,Male,Boone,IA,USA,44000"  
6, "40,Female,Cupertino,CA,USA,48000"  
7, "20,Male,Boone,IA,USA,24000"  
8, "20,Female,Ames,IA,USA,26000"  
9, "12,Male,Ames, IA,USA,1000"  
10, "24,Male,Boone,IA,USA,34000"
```

Outputs for each record are:

```
1 =>  
(20,25000)  
(Male,1)  
(unique-cities,Ames)  
  
2 =>  
(24,29000)  
(Male,1)  
(unique-cities,Ames)  
  
3 =>  
(24,57000)  
(Female,1)  
(unique-cities,Sunnyvale)  
  
4 =>  
(40,43000)  
(Male,1)  
(unique-cities,Ames)  
  
5 =>
```

```
(40,44000)
(Male,1)
(unique-cities,Boone)

6 =>
(40,48000)
(Female,1)
(unique-cities,Cupertino)

7 =>
(20,24000)
(Male,1)
(unique-cities,Boone)

8 =>
(20,26000)
(Female,1)
(unique-cities,Ames)

9 =>
Ignored

10 =>
(24,34000)
(Male,1)
(unique-cities,Boone)
```

Question 3

10 / 10 pts

Show all of the input to "Sort & Shuffle" phase.

Your Answer:

"Sort and Shuffle" phase takes all the outputs of the Mappers as inputs:

```
1 =>
(20,25000)
(Male,1)
(unique-cities,Ames)

2 =>
(24,29000)
(Male,1)
(unique-cities,Ames)

3 =>
(24,57000)
(Female,1)
(unique-cities,Sunnyvale)

4 =>
(40,43000)
(Male,1)
(unique-cities,Ames)

5 =>
(40,44000)
(Male,1)
(unique-cities,Boone)

6 =>
(40,48000)
(Female,1)
(unique-cities,Cupertino)

7 =>
(20,24000)
(Male,1)
(unique-cities,Boone)

8 =>
(20,26000)
```

```
(Female,1)
(unique-cities,Ames)
```

```
9 =>
Ignored
```

```
10 =>
(24,34000)
(Male,1)
(unique-cities,Boone)
```

Question 4

20 / 20 pts

Show output of the "Sort & Shuffle" phase.

Your Answer:

The output of Sort and Shuffle phase:

```
(20, [24000, 25000, 26000])
(24, [29000, 34000, 57000])
(40, [43000, 44000, 48000])
(Female, [1,1,1])
(Male, [1,1,1,1,1,1])
(unique-cities, [Ames, Ames, Ames, Ames, Boone, Boone, Boone, Cupertino, Sunnyvale])
```

Question 5

20 / 20 pts

Write a reduce() function to generate the desired output.

Your Answer:

```
reduce(Key, values) {  
  
    #if Key is 'Unique-Cities', add all the unique cities present to a list  
  
    if (Key == "unique-cities") {  
  
        list_of_cities = [] #create an empty list  
  
        #to access each of the values for this key use a loop  
  
        for (v in values) {  
  
            #if the value is not already present in list, then add that value to list.  
  
            if (v not in list_of_cities) {  
  
                list_of_cities.append(v)  
  
            } # inner if closes  
  
        } #loop closes  
  
        #emit the output in desired format  
  
        emit(Key, list_of_cities)  
  
    } #outer if closes  
  
  
    #if Key is 'Male', count the number of occurrences
```

```
else if (Key == Male) {  
    count = 0 #initialize a counter  
    for ( v in values ) {  
        count += 1  
    } # loop ends  
    #emit the output in desired format  
    emit("number-of-" + Key + "s", count)  
} #outer else if closes
```

#if Key is 'Female', count the number of occurrences

```
else if (Key == Female) {  
    count = 0 #initialize a counter  
    for ( v in values ) {  
        count += 1  
    } # loop ends  
    #emit the output in desired format  
    emit("number-of-" + Key + "s", count)  
} #outer else if closes
```

#if Key is <age> calculate average of salary for that age group

```
else {  
    sum = 0 #initialize an accumulator  
    count = 0 #initialize a counter  
    average = 0.0 #initialize a variable to store average  
    for ( v in values ) {  
        sum += v  
        count += 1  
    } #loop ends  
    average = sum/count  
    #emit the output in desired format  
    emit(Key, average)  
} #outer else closes  
} #reduce function closes
```

Question 6

10 / 10 pts

Show output of all reducers.

Your Answer:

'Reducers' takes input from Sort and Shuffle Phase and generates the following output:

```
(20, 25000)
(24, 40000)
(40, 45000)
(number-of-Females, 3)
(number-of-Males, 6)
(unique-cities, (Ames, Boone, Cupertino, Sunnyvale))
```

Quiz Score: **100** out of 100