Quiz 2 - Thursday Batch - Rubrics

1. [2 points] What are the different nodes at which failure can occur in a Map-Reduce system? How is the failure dealt with?
   Map worker failure:
   Map tasks completed or in-progress at worker are reset to idle; Reduce workers are notified when the task is rescheduled on another worker.

   Reduce Worker failure:
   Only in-progress tasks are reset to idle
   Reduce task is restarted

   Master failure:
   MapReduce task is aborted and client is notified.

   Rubrics - All 3 correct - 2 points
   Any 1 wrong - 1 point
   Any 2 wrong - 0.5 points

2. [1 point] In MapReduce, which of the following is **max of I/O cost along any path**?
   a. Elapsed Computation cost
   b. Communication cost
   c. Elapsed Communication cost
   d. Total Payable cost

3. [1 point] Which of the following is a transformation operation?
   a. Map
   b. Count
   c. Filter
   d. distinct

   Answer- Only count is an action, rest all are transformations

4. [1 point] Consider multiplying two matrices A (3X3) and B (3X2). Consider the **one-stage** approach to matrix multiplication (AXB) as discussed in class.

   A = [1 1 1]
   [2 1 2]
   [1 2 1]

   B= [1 0]
   [0 1]
   [1 2]

If the Mapper takes as the input the element B[2,2], which of the following key-value pairs will be in its output?

a. ((1,2),(B,2,2,B[2,2]))
b. ((2,2),(B,2,2,B[2,2]))
c. ((2,1),(B,2,1,B[2,2]))
d. ((3,2),(B,2,1,B[2,2]))

Answer - emit ((i,k), ('B', j, k, B[j,k])) for i in 1..3
Where j and k are 2 and i can be 1...3
Or
Only ((2,2),(B,2,2,B[2,2])) (this answer is only given points due to misprint in the ppt)

5. [1 point] Consider multiplying two matrices A (3X3) and B (3X2). Consider the **two-stage** approach to matrix multiplication(AXB) as discussed in class.

A = [1 1 1]
[2 1 2]
[1 2 1]

B= [1 0]
[0 1]
[1 2]

If the Mapper in **stage 1** takes as the input the element A[3,2], which of the following key-value pairs will be in its output?

a. (2, (A,3,A[3,2]))
b. (3,(A,3,A[3,2]))
c. (2,(A,2,A[3,2]))
d. (3,(A,2,A[3,2]))

Answer - A[ij] : emit(j , (A, i, A[i,j])
Where i is 3 and j is 2.

6. [2 points] Write the map reduce solution for Distributed Sort. We would like to sort a very large list of (firstName, lastName) pairs by lastName followed by firstName
Examples of outputs:
Smith Anne
Smith John
Smith Ken
- Map Task:
emit(lastName, firstName)
- Group By Keys
- Reduce Task:

For each lastName key, if there are multiple firstName values, emit(lastName, firstName) in alphabetical order. Merge output from all reduce tasks.

Rubrics - 1 point for Map and group by step
1 point for Reduce step

7. [2 points] You are given input as a list of housing data where each input record contains information about a single house: (address, city, state, zip, value). The output should be the average house value in each zip code. Complete the following

```
map(key, value):
emit(zip, (value, 1))

Combiner : Sum the value and number of houses in each map task
emit(zip, (sum(value), n)
reduce(key, values):
sum = 0
count = 0
For val, num in values:
        sum += val
        count += num
average = sum / count
emit(zip, average)
```

Rubrics - 1 point for correct Map step (0 is given if sum or count is taken in the map step as it defeats the purpose of map-reduce paradigm)
1 point for correct reduce step