Take Home Quiz

1. [2 points] Describe how Map-Reduce environment takes care of the following items:
   a. machine failures,
   b. inter-machine communication

   a. Map Worker Failure: Map tasks completed or in-progress at worker are reset to idle. Reduce workers are notified when 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
   b. Performs status checking and handles machine failures. When the map task completes, send the master the location and size of intermediate files, one for each reducer. Master pushes this info to the reducers. Task Pipelining.
2. [1 point] Explain lazy and eager operations in Spark with examples
   Lazy operations(transformations) are not executed immediately and it's execution is deferred until an action is performed. Examples map(), flatmap() [0.25 + 0.25 points]
   Eager operations(actions) are executed immediately. Examples take(), count() [0.25 + 0.25 points]
3. [3 point] Consider multiplying two matrices A (2X3) and B (3X2). Consider the **two-stage** approach to matrix multiplication(AXB) as discussed in class.

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

   B= [1 0]
   [3 1]
   [1 2]

   Show a detailed computation performing the matrix multiplication.

   1st map function: [1 point]
   for each matrix element A[i][j]: emit(j, (A,i,A[i,j])
   for each matrix element B[j][k]: emit(j, (B,k,B[j,k])

   For A:
   A[1][1]: emit(1, (A,1,1))
   A[1][2]: emit(1, (A,2,2))
   A[1][3]: emit(1, (A,3,1))
   A[2][1]: emit(2, (A,1,2))
   A[2][2]: emit(2, (A,2,1))
   A[2][3]: emit(2, (A,3,2))

For B:
emit(j, (B,k,B[j,k])
B[1][1]:emit(1, (B,1,1))
B[1][2]:emit(1, (B,2,0))
B[2][1]:emit(2, (B,1,3))
B[2][2]:emit(2, (B,2,1))
B[3][1]:emit(3, (B,1,1))
B[3][2]:emit(3, (B,2,2))

1st reduce function: [1 point]
For each value of (i, k) which comes from A and B, i.e.(A, i, A[ij]) and (B, k, B[jk]):
Emit(i, k), (A[ij]*B[jk])

emit((1,1),1)
emit((1,1),6)
emit((1,1),1)

emit((1,2),2)
emit((1,2),2)

emit((2,1),2)
emit((2,1),3)
emit((2,1),2)

emit((2,2),1)
emit((2,2),4)

2nd map function: [0.5 points]
pass through input

2nd reduce function:
Add up all the values
emit((1,1),(1+6+1))
emit((1,2),(2+2))
emit((2,1),(1+4))
emit((2,2),(1+4))

Final Answer - [0.5 points]
[8 4]
[7 5]

4. [2 points] Write a map and reduce tasks and their outputs for joining these 2 tables.

Order(orderId, account, date)
1, aaa, d1
2, aaa, d2
3, bbb, d3

LineItem(orderId, itemId, quantity)
1, 10, 1
1, 20, 3
2, 10, 5
2, 50, 100
3, 20, 1

Map(key, value) [0.75 points]
  If table is "Order":
    emit(orderId, ["Order", (orderId, account, date)])
  If table is "LineItem":
    emit(orderId, ["LineItem", (orderId, itemId, quantity)])

Reduce(key, value) [0.75 points]
  emit(orderId, Order JOINS LineItem for each key-value pair(exclude table name from results))

The output of the map step for the example will be:
(1, ["Order", (1, aaa, d1)])
(2, ["Order", (2, aaa, d2)])
(3, ["Order", (3, bbb, d3)])
(1, ["LineItem", (1, 10, 1)])
(1, ["LineItem", (1, 20, 3)])
(2, ["LineItem", (2, 10, 5)])
(2, ["LineItem", (2, 50, 100)])
(3, ["LineItem", (3, 20, 1)])

The output of the reduce step for the example will be:
(1, aaa, d1, 1, 10, 1)
(1, aaa, d1, 1, 20, 3)
(2, aaa, d2, 2, 10, 5)
(2, aaa, d2, 2, 50, 100)
(3, bbb, d3, 3, 20, 1)
Final output [0.5 points]

5. [1 point] Consider **one-stage** matrix multiplication using map-reduce. Let A and B be two matrices of shape 5X5. What will be **emitted by Map task** for each element of (i,j) of A and (j,k) of B respectively?
    a. emit ((i,k), ('A', i, j, A[i,j]) for k in 1..5, emit ((i,k), ('B', i, j, B[j,k])) for i in 1..5

b. emit(j , (‘A’, i, A[i,j])), emit(j , (‘B’, k, B[j,k]))
c. emit ((i,k), (‘A’, i, A[i,j]) for k in 1..5, emit ((i,k), (‘B’, k, B[j,k])) for i in 1..5
d. emit(i , (‘A’, j, A[i,j])), emit(k , (‘B’, j, B[j,k]))

6. [1 point] Consider a Map-Reduce program that computes the **largest integer** in a large set of integers. Suppose one of its Mapper outputs a key-value pair : (8,1). Which of the following is/are **unlikely** to be the input of Mapper?
   a. (“8”, [3,5,1,8])
   b. (“1”, [1,8,5,9])
   c. (“8”, [10,9,8,11])
   d. (“1”,[7,8,8,6])