Full Name:
Andrew ID:

CS 15-440: Distributed Systems
Mock Quiz 2
November 11, 2019

Total Time: 25 minutes

Instructions:

- Write your answers in the spaces provided below each problem. If you make a mess, clearly indicate your final answers.
- The quiz has a maximum score of 20 points.
- Keep up with time.

Good Luck!

Question No	Max. Points	Earned Points
1	7.5	
2	4	
3	5	
4	3.5	
Total	**20**	

Problem I: True and False Questions (7.5 Points)

1) The decentralized mutual exclusion algorithm by Lin et *al.* we studied in class is a special case of the Gifford's scheme with equal read and write quorums, but not necessarily with quorums of a size greater than the number of replicas. **(True/False)**
2) The number of messages required for a process to access and release a shared resource using the token mutual exclusion algorithm could be at times less than the corresponding number of messages of a centralized permission-based mutual exclusion algorithm. **(True/False)**
3) The bully election algorithm requires 2*n* messages for electing a coordinator, where *n* is the number of processes in the underlying distributed system. **(True/False)**
4) The shared-memory programming model allows processes to interact via sending and receiving messages. **(True/False)**
5) In principle, the larger the size of the output of the map phase the less the number of map waves are recommended to improve performance. **(True/False)**
6) Overlapping communication with computation will typically degrade performance. **(True/False)**
7) In Hadoop MapReduce, the shuffle stage in the reduce phase is only started after the map phase is done. **(True/False)**
8) Pregel is a graph-parallel, asynchronous, and in-memory system. **(True/False)**
9) MapReduce and Pregel will potentially perform comparably with applications involving a single iteration. **(True/False)**
10) HDFS applies content-based partitioning. **(True/False)**
11) You can use MPI to invent synchronous and asynchronous distributed analytics systems. **(True/False)**
12) HDFS uses a master-slave architecture, but Hadoop MapReduce and Pregel use peer-to-peer ones. **(True/False)**
13) Hadoop MapReduce uses a pull-based scheduling model, thus becoming prone to over-utilizing TaskTrackers. **(True/False)**
14) Although PageRank can be solved using a system of linear equations, it is typically solved using a matrix formulation for scalability reasons. **(True/False)**
15) PageRank can run faster on Pregel than on MapReduce. **(True/False)**

Problem II: System Configurations (4 Points)

1) The best performing configuration for a write-heavy, Gifford-based distributed system with 7 replicas is:
    a. Write quorum = 4 and read quorum = 1
    b. Write quorum = 3 and read quorum = 4
    c. Write quorum = 1 and read quorum = 7
    d. Write quorum = 7 and read quorum = 7
    e. Write quorum = 4 and read quorum = 4
    f. Write quorum = 6 and read quorum = 2

2) The best configuration for a Map-only image processing Hadoop application with an underlying dataset of 2048MB is (*Note: machines in this question are assumed to be powerful enough to host up to 4 map slots independent of the size of the HDFS block*):
    a. An HDFS block size of 64MB and 4 machines, each with 2 map slots
    b. An HDFS block size of 128MB and 4 machines, each with 2 map slots
    c. An HDFS block size of 256MB and 8 machines, each with 2 map slots
    d. An HDFS block size of 128MB and 4 machines, each with 4 map slots

Problem III: Synchronous vs Asynchronous Functions and Systems (5 Points)

1) What is the difference between synchronous and asynchronous functions? Is MPI_Isend(.) a synchronous or an asynchronous routine?

2) What is the difference between synchronous and asynchronous systems? Is MapReduce a synchronous or asynchronous system?

Problem IV: Types of MapReduce Applications (3.5 Points)

List the four types of MapReduce applications discussed in class. In light of this list, what are the types of the WordCount and Sort MapReduce applications?