

Name (Last, First): _____

This exam consists of 5 questions on 7 pages; be sure you have the entire exam before starting. The point value of each question is indicated at its beginning; the entire exam is worth 100 points. Individual parts of a multi-part question are generally assigned approximately the same point value: exceptions are noted. You are required to have hard copies of all the papers we have read in this class prior to today. You may not share your paper copies with other students during the exam.

Be concise and clearly indicate your answer. Presentation and simplicity of your answers may affect your grade. **If I cannot read your answer easily you will not get credit.** Answer each question in the space following the question. If you find it necessary to continue an answer elsewhere, clearly indicate the location of its continuation and label its continuation with the question number and subpart if appropriate.

You should read through all the questions first, then pace yourself.

The questions begin on the next page.

Problem	Possible	Score
1	20	
2	20	
3	20	
4	20	
5	20	
Total	100	

1. (_____/20 points)

Short Answer

(a) What are the two basic ways we can achieve reliability in a distributed system? For example, how does MapReduce achieve reliability differently than GFS?

(b) Explain why coroutines (Gevent and Greenlets) can be very efficient in network-based distributed programs?

(c) What is distributed consensus?

(d) Why is distributed consensus a hard problem?

2. (_____/20 points)

MapReduce Recall the two papers we read on MapReduce and your MapReduce implementation for Project 2. Answer the following questions

(a) (10 points) In the shuffle phase of MapReduce we can sort the k2, v2 values on the Map worker or on the Reduce worker. Explain the pros and cons of each approach.

(a) (10 points) For a particular MapReduce program with a particular input, propose a way to find out the optimal number of Reduce tasks. Note one Reduce task will be the slowest, but too many reduce tasks may also be slow due to overhead.

3. (_____/20 points)

Leader Election

Recall the paper *Elections in a Distributed Computing System* by H. Garcia-Molina.

In the Bully Algorithm, it is possible for a leader to become a follower even if it does not crash. Think of ways to change the Bully Algorithm to allow a leader to remain a leader over time even when new processes with higher id values enter the system. Try to retain as much of the Bully Algorithm as possible. Your solution does not have to force the leader to remain the leader, but rather have it stay as the leader with high probability. Be creative.

4. (_____/20 points)

GFS

Recall the paper *The Google File System* by Ghemawat, Gobioff, and Leung. Answer the following questions. Provide references sections and paragraphs in the paper to support your answers.

(a) (10 points) Explain how pipelining during chunk replication improves replication performance? What are slower, less efficient alternatives.

(b) (10 points) Hadoop's HDFS does not support small random writes to files. In fact, until recently, Hadoop did not support file append. Explain why small random writes would be inefficient in GFS. Think about what is needed to support a update rather than just a streaming write to a file.

5. (_____/20 points)

Raft

Recall the paper *In Search of an Understandable Consensus Algorithm* by D. Ongaro and J. Ousterhout. Answer the following questions. Provide references sections and paragraphs in the paper to support your answers.

(a) (5 points) Based on the description in the paper, how is the Raft algorithm different from Paxos?

(b) (5 points) In Raft, how do randomized timeouts help make leader election more efficient?

(c) (10 points) Describe what happens when a followers log becomes inconsistent with the leader and the follower must duplicate the leaders log. Consider the following situation:

log index	1	2	3	4	5	6
leader (term)	1	1	1	2	2	3
follower (term)	1	1	4	4	4	

Explain how the Raft algorithm forces the recovered follower to duplicate the leader's log. What steps are taken?

Continue your answers here if necessary.