

# Homework 4

Due: December 11th, 2015 23:59:59pm

Please typewrite your answer to the homework problems, save your file in .pdf format, name your file as “your BU email id”-homework4.pdf (e.g., jdoe-homework4.pdf), and submit it to Blackboard before deadline. Late homework will not be accepted.

**Problem 1.** (Chapter 18 Problem 19 in Cbook) (15 pts)

Gifford’s quorum consensus replication is in use at servers X, Y and Z which all hold replicas of data items A and B. The initial values of all replicas of A and B are 100 and the votes for A and B are 1 at each of X, Y and Z. Also  $R = W = 2$  for both A and B. One client reads the value of A and then writes it to B.

- (a) At the time the client performs these operations, a partition separates servers X and Y from server Z. Describe the quora obtained and the operations that take place if the client can access servers X and Y.
- (b) Describe the quora obtained and the operations that take place if the client can access only server Z.
- (c) The partition is repaired and then another partition occurs so that X and Z are separated from Y. Describe the quora obtained and the operations that take place if the client can access servers X and Z.

**Problem 2.** (10 pts)

Explain how three phase commit protocol avoids delay to participants during their “uncertain” period due to the failure of the coordinator or other participants. Assume that communication does not fail.

**Problem 3.** (10 pts)

What is a virtual node and why does Dynamo use virtual nodes in consistent hashing over physical nodes?

**Problem 4.** (10 pts)

Both node failures and concurrent updates may lead to conflicting versions of an object. How does Dynamo determine whether two updates are causally related or concurrent?

**Problem 5.** (15 pts)

Dynamo allows an application to customized its storage system to meet the desired level of performance, durability, and availability SLAs by allowing them to tune the parameters N, R, and W. Explain what these parameters are, and how they can be tuned.

**Problem 6.** (10 pts)

Dynamo uses a “sloppy” quorum mechanism. Explain what a “sloppy” quorum is, and what implication it has on Dynamo’s consistency model.

**Problem 7.** (15 pts)

Consider a 32-bit Bloom filter, and 3 hash functions  $h_1$ ,  $h_2$ , and  $h_3$ , where  $h_i(x) = ((x^2 + x^3) * i) \bmod m$ ,  $m$  is the number of bits in the Bloom filter.

- (a) (10 pts) Show the Bloom filter bits following each of the following six insertions in order: 2013, 2010, 2007, 2004, 2001, 1998.
- (b) (5 pts) For the Bloom filter obtained after part (a), find one value that is not among the six inserted values, but is a false positive.

**Problem 8.** (15 pts)

Consider a Cassandra deployment using the `RackInferring` snitch. Every data is stored at 3 replicas. Given the following IP addresses for the replicas, say in each case whether the allocation is: fault-tolerant to the failure of: i) a single server; ii) a single rack; iii) a single datacenter. (You should give 3 answers for each case.)

- (a) 123.231.111.222, 123.231.112.222, 123.231.113.222
- (b) 123.231.111.222, 123.230.1.2, 123.3.4.5
- (c) 123.231.111.222, 123.231.111.223, 123.231.111.212
- (d) 1.231.111.222, 1.231.112.222, 1.231.113.222
- (e) 1.2.3.4, 1.2.3.4, 1.2.3.5