

CS 437 / 537: Database Systems

Assignment 3

Assigned: Tuesday, November 17, 2015

Due: 2359hrs Thursday, December 3, 2015

Instructions

- Since this HW involves quite a bit of drawing and longer answers, you need not submit this online – feel free to drop it off in my mailbox (503) in the computer science building (AKW) anytime before the deadline. I prefer typed solutions.
- You are free to use any and all resources you can access, as long as you do the work yourself, and cite all sources. This means that you can find your definitions on Wikipedia, search for solutions to SQL issues on forums, etc. You should know this already, but we'll say it anyway: You may not get other people to do your HW for you.
- Since many students have said that the project is taking up all of their time, this will be the last HW – that's why this is longer and why you have multiple reading assignments. You have plenty of time, so I shall expect very good solutions.
- If you choose to email the assignment, send it to debayan.gupta@yale.edu. Make sure you check your attachments carefully before you hit send.

Part I: Reading (40 points)

As part of this assignment, you are expected to read and critique 2 papers. Your critique should be in 2 parts, each ~5-10 lines:

1. Summary of the paper. Highlight the primary technical innovations that the paper talks about.
2. Errors, improvements, whatever you found interesting about the paper.

The two papers for this assignment are:

1. Amazon's DynamoDB
(<http://www.allthingsdistributed.com/files/amazon-dynamo-sosp2007.pdf>; also see <http://aws.amazon.com/dynamodb/>)
2. Google's BigTable (<https://research.google.com/archive/bigtable-osdi06.pdf>)

Part II: (60 points)

1. Create a B+ tree of degree 5 and add the following 20 values to it, in order:
12, 2, 15, 4, 19, 5, 17, 8, 14, 10, 6, 1, 11, 0, 13, 7, 3, 16, 18, 9

Note: Show only the following steps – after inserting 14, 3, and 9 (13 points)

2. Delete the following values from the tree (in order): 14, 2, 17, 8, 9
Show only the following steps – after deleting 2 and 9 (10 points)
3. Explain the difference between view and conflict serializability; provide an example of each type. (10 points)
4. Your database system uses a special mechanism to store tuples using encryption and digital signatures to ensure that the data has not been tampered with. As such, it does not possess the ability to perform updates, since a tuple, once inserted, cannot be changed without setting off an alarm.

You decide to implement updates in this system by using deletion followed by insertion. What are the pros and cons of this idea? If there are any problems, suggest some possible solutions. ~5 lines would be ideal. Please do not write more than 1 page. (10)

5. Final question: This is a general question, and you'll probably get a good score unless you write something that contradicts the basic principles of databases. Or the laws of physics.

You've just taken over the world, and want to keep track of all your loyal subjects. To this end, you want to create a giant database with data about every living human. As you might guess,

this is a giant database, with many attributes, some of which are also quite large. What general principles would you follow when commanding your minions to construct this database system? (e.g. B+ trees? Binary relations? Hash tables? Two phase locking for executing transactions? What tables would you use? How many backups would you keep in order to prevent localized rebellions from destroying your data? Would it be a distributed system with a lot of parallelization, or a centralized supercomputer?)

Given what you have learned this semester, I want a general idea of your concepts, and whether or not you would make a good dictator. (17)