Module 12 - Discussion Prompt Questions

Adam Rich

May 1, 2018

Section EN.605.202.87.SP18

1. What are some of the largest employers with which you are familiar? What do you imagine are some of the greatest challenges facing them in terms of information management?

   I believe that the largest employer in the world in the US Department of Defense, although I could be wrong. McDonald's and Wal-Mart are another two. Their greatest challenges with Information Management would be just the massive amount of data they have to track and the various locations where the data are entered and needed. An employee's record will not be needed at every location. But, it will probably need to be kept at their location of employment and also at headquarters locations.

2. What are some of the many ways that company may need to access data?

   Data entry. Making sure that someone actually works for them. Payroll. Handling leave. Tracking productivity.

3. What are the pros and cons of using a social security number as a primary key? What are some good alternatives?

   SSNs are supposed to be unique, so that can help. However, not everyone has one. Also, it is such an important identifier that you really should not use it if you do not have to. An alternative could be a randomly generated key. Or, a key that embeds some sort of important information like first and last name and then a suffix to make it unique. Then, it would be easier to find the record needed searching just on the key values.

4. How big an impact on your design considerations in general is it to need a sorted file only periodically (for a directory, for example)?

   Keeping a file sorted has a maintenance cost. So, if you don't need to keep it sorted then that opens up the possibility of being less concerned about that maintenance and more concerned with just searching the data. Re-sorting, if desired, can be done periodically.

5. How big an impact on your design considerations to sort by different fields of the data periodically (a directory is likely ordered by last name, first name, for example)?

   This is a bigger deal, depending on the size of the dataset. You have to decide whether to re-sort the data in place or to make a copy, both of which can be expensive operations. Or, if you know that you will need to sort parts of the record in different ways, maybe the keys could be designed in a way to allow them to be sorted with pointers to the unsorted data.

6. How would the nature of your data set affect your choice of collision handling scheme?

   If there is not a lot of data, you can easily pick a hashing function that does not have collisions. If there is a lot of data but a natural and easy hashing function, then maybe collision can be avoided. Whatever the hashing function is, if there is collision then you will want to minimize it. Having too many keys hashed to the same value will slow down finding things are kind of work against the idea of hashing.