Mason Howard

01/13/2014

CS 2243 Chapter 01 Homework

Revisit, revise and resubmit: Q5 – examples. P2 & P9 -- see the grey text

I. Explain the difference between facts, data, information and knowledge. How are these terms “a progression?” How does a database work/help in this context? Reasonable start but needs s little work

Facts, or data, are just individual tidbits. Things like phone numbers, ages, or addresses, but without the knowledge that it is a phone number, an age, or an address. (can draw a reasonable conclusion, tho and Metadata is *in* a database) Metadata would give that knowledge. Information is analysis on the data. Examples would be things like the average age is such and such, or we've got a 20% budget surplus. “Knowledge,” as it's used here, could also be restated as subject understanding plus situational awareness.

The terms are a progression because each one ‘leverages’ the one before. Data is data. “Information” is the result of (need a better descriptor) ~~leveraging~~ data. “Knowledge” is the collection and accumulation of that information over time. Takes more than just collecting data to become knowledge

Databases and database management systems are a solution that provides quick access to “information,” maintains data integrity, and better security against intrusions and data loss.

II Please provide answers to these Exercises:

The designated Review Questions and Problems are located at the end of the chapter.

Identify your answer with the question or problem number.

from pages 29 - 31

* Answer Review Question (RQ) 5 . Answer, also, for Data Independence
  + Also give an example of each (different from the ones mentioned in class)
* Answer RQ 12 – consider more than $$ or monetary costs
  + don’t simply copy the words from the text, please
* Answer RQ 15
* for Figure P1.1
  + Answer Problem 2 and 4
* for Figure P1.9
  + Distinguish between attributes & entities.
  + What *entities* do you perceive in this figure?

Answer Problem 9

Review Question 5: What is structural independence, and why is it important? What is data independence and why is it important? Give an example of each.

Structural independence: Structural independence exists when you can change the file structure without affecting the application's ability to access the data. The ability to add different categories of data to a file (file X contains names, passwords, and now credit card numbers) and not affect the DBMS's ability to pull info from it is an exhibition of structural independence.

Note: Would need to change programming or write new to access the new data. Real question is how would placement of this new data affect location of “old” data and existing program’s access to that.

Data independence: Data independence exists when you can change the data storage characteristics without affecting the program's ability to access the data. For example, having some things stored in a CSV and others in a binary file. The DBMS should be able to work with both.

Wrong direction – consider an attribute rather than the entire file. Also, changing the file type would require change in programming.

Review Question 12: What are the potential costs of implementing a database system? Consider more than monetary costs.

More than the monetary costs of the hardware, software, training, support, and invariable downtime as it is installed / upgraded, implementing a database system likely involves you with the vendor for a very long time. One should weigh security risks as a valid concern. DBMS vendors frequently upgrade their products by adding new functionality, and possibly requiring hardware improvements. ✓

Review Question 15: What common problems does a collection of spreadsheets created by end users share with the typical file system?

Synchronization, authorization, authentication, difficulty leveraging data, limited data sharing, complex administration, etc. ✓ Though how much “administration” really occurs with a spreadsheet?

Figure P1.1, Questions 2 and 4:

2: You'd have to search the texts for something that looks like a city, extract it, associate it with the rows in some fashion, sort the rows by those, and then return the rows. I'd have a table with columns that define an address and a column that gives a unique identifier to each, and give the unique identifier to the column of “MANAGER\_ADDRESS” and then you can sort according to city much easier, without having to search strings and whatnot (something that probably would have to be custom done). Reconsider: there is a much simpler solution.

4: The table doesn't seem to have a unified purpose. The data's there all willy-nilly and it's got the brilliant “Holly B. Parker” as a project manager for two projects, basically wasting rows. George has got two projects, and I'm pretty sure their phone numbers don't change. Unless they got a new phone and didn't update all projects under the owner with the new phone number.

Yes, you found the problems but study the table again – it’s not a random collection of facts.

Figure P1.9, Distinguish between attributes and entities. What entities do you perceive in this figure? Answer Question 9.

Not question 9: There's buildings, rooms, teachers, and times. Attributes would be \_FNAME, \_LNAME, \_INITIAL. Those times could be improved by using a time block table, and referring to unique identifiers on it to specify times. Only 3 attributes? Consider the table as a whole – what is it (its purpose)?

9: Teachers should have a unique identifier assigned to them by their own table that describes first names, last names, and initials. Rooms should have their own unique identifier assigned in a table defining room numbers and building locations. Times should have their own table identifying where specified time is taking place, as well as when. (as will see in a few weeks, this is “too” normal and will be problematic.) It needs to be cleaned up, made pretty.

Need to identify the specific problems with current arrangement.