

TECHNICAL CONTRIBUTIONS

Teaching data base systems using Date and Computing Surveys

Ehud Gudes, Computer Science Department, Pennsylvania State University

Data base systems is a "hot" topic in research. However, teaching a course in data base systems causes several problems to arise. First, there is not yet one good text book which covers all the material required for such a course at the advanced undergraduate/early graduate level. Second, if a data base system (such as INGRES or EDBS) is not available for students to experiment with, it is hard to design meaningful assignments. In the following paragraphs we report how we solved these problems in a course on data base systems given last fall term at the Computer Science Department at The Pennsylvania State University.

The course (CMPSC 440) is equivalent to a 3 credit semester course and was given to a mixed class of 1/3 graduate students and 2/3 advanced undergraduate students. All students had a course in data structures and most of them had a course in operating systems before.

As textbooks we used Date [1] and Computing Surveys (CS) (see [2]). In addition, Martin's book [3] was used as a reference to several of the topics. In general, the combination of Date and CS proved very successful. Using both of them, one can cover the three data base models in great detail and compare between them. In fact, the two texts complement each other in some respects. Date is very good in the relational model, but vague and unclear in the hierarchical and network models. CS is not very detailed in relational but clear and detailed enough in the hierarchical and network. Actually, once one reads the articles on hierarchical and network in CS, Date's presentation becomes much clearer. Using two texts instead of one causes some "terminology problems". Since this is a common problem in the data base area, it was handled by the instructor, and was not considered a major obstacle. What is missed from both texts is a good discussion of storage (file) structures. Date's chapter 2 is definitely not enough. The topic of storage structures presents further complication because some of it might have been taught in other courses (e.g. in a data structures course). In our course we used IBM manual [4] and Martin [3] and for basic storage structures and Knuth Vol 3 [5] for advanced storage structures.

Following is an outline of the course plus comments on the use of textbooks for each section.

- 1) Introduction. Basic definitions. Historical survey. - Sybley and Fry's article in CS complements nicely Date's chapters 1 and 3.
- 2) Basic storage structures - Sequential, Index sequential, Direct (hashing) organizations are studied from IBM [4]. Inverted files and multi-list are defined but not analyzed in detail (using Martin [3]).

- 3) The relational model - Date is very good. Chapter 8 is complemented with more recent material on implementation of relational systems (see e.g. [6]).
- 4) The hierarchical model - Tschritizis and Lochovsky's article in CS is very good. It discusses general selection languages which Date ignores and does not stress IMS as heavily as Date. Date is used mainly to complement CS for more details on some of the IMS concepts (e.g. IMS storage structures - Date Chapter 13).
- 5) The network model - Taylor and Frank's article in CS is excellent. Date's chapter 15 is used first and then CS is used with occasional references to Date for more examples.
- 6) Security and Integrity - Date's chapters 19 and 20 are used plus some of the more recent papers on this subject.
- 7) Comparison between the models - Michaels et al article in CS is used.
- 8) Advanced storage structures -
 - a) primary key indexing - tree directories and B-trees are discussed in some detail. Knuth [5] can be used although there is not enough time to cover it at the depth of Knuth's coverage.
 - b) Secondary key indexing - Inverted files, cellular multi-list are discussed in more detail. Hsiao and Harary's generalized file structure [7] is used.
- 9) Date base design and the role of the DBA - This is used as a summary for the whole course. A topic which is hard to find good technical literature on. Several recent papers are referenced (see e.g. [8]).

Admittedly, this is a lot of material to cover and one may need to skip some of it. We covered most of it, although topics 6, 7, 8 and 9 were discussed in less detail than desirable.

The assignments for this course included homeworks (some of Date's exercises are very good), and two major projects: a) a term paper based on recent literature; b) implementation of a mini relational data base system. Date's data model of Parts and Suppliers was used (i.e. only three relations). The students had to read in a DMD and to print out the data base using that DMD. Then they had to read a query in a simplified version of relational calculus and to execute it. For the second part a simple algorithm to convert from a restricted relational calculus (only existential quantifiers) to relational algebra was used. (Because of the restriction above, the algorithm is much simpler than in [9] and

therefore can be implemented in a class project).

The students evaluation comments centered around three topics. First, they felt that too much material has been covered. Most of them preferred less detailed discussion of the three models and more practical aspects (e.g. topic 9). Second, they felt that too much assignments work was required and suggested to choose between the lab and the term paper. We will probably adopt the first suggestion in the future. Giving up the lab will seem to hurt the course unless an alternative such as existing data base system is available. Finally, a common request of all students was: one good textbook!

References

- [1] Date, C.J. An Introduction to Database Systems Addison-Wesley, 1975.
- [2] ACM Computing Surveys, Vol 8, No. 1, March, 1976.
- [3] Martin, J. Computer Data Base Organization, Prentice-Hall, 1975.
- [4] IBM, Introduction to IBM Direct Access Storage Devices and Organization Methods, GC20-1649-9, 1976.
- [5] Knuth, The Art of Computer Programming, Vol 3, Addison-Wesley, 1973.
- [6] "Implementation of Relational Data Base Management Systems", FDT Vol 7, No. 3-4, 1975.
- [7] Hsiao, D and Harary, F., "A formal system from information retrieval from files", CACM Vol 13, No. 2, February, 1970.
- [8] Kahn, B.K., "A method for describing information required by the database design process", Proceedings of SIGMOD 1976.
- [9] Codd, E.F., "Relational completeness of data base sublanguages", Data Base Systems, Prentice-Hall, 1971.