GROWTH PROSPECTS IN ENGINEERING DATABASES

J. Bradley, University of Calgary

The subject of databases and database systems has enjoyed spectacular growth in the past 15 years. Databases reside on computer disks, and are the primary consumers of disk space. A measure of the importance of data bases is the similar spectacular growth in sales of large capacity disks. This growth is such that sales of large disks from manufacturers in Silicon Valley now outnumber sales of microcircuit components from that famous valley by a factor of some ten to one.

Large data bases are the single most important factor in the continued demand for large mainframe computers in business and industry. The sheer size of these data bases makes manipulation and control by cheaper mini and microcomputers impossible. The personal computer user typically obtains information from these databases via a micro-mainframe link, and then locally manipulates small files and data bases stored on the low capacity personal computer disks (less than about 50 MB).

The whole field of data bases grew out of the need in business to integrate important master files. Indeed, for the first five years of the 1970s, it is probably true to say that there was no academic interest in the subject whatsoever. Admittedly, the general hodge-podge of systems and ideas then prevailing was a factor in this neglect. That changed with the development of the basic theory of relational databases, and the demonstration of the direct applicability of predicate calculus to non procedural data base languages. This work was largely originated at IBM, but by the late 1970s had led to a large involvement by researchers in universities, mainly at computer science departments, but to some extent in electrical engineering departments.

By the mid 1980s, the discipline of data base systems, particularily relational data base systems, had become well established in universities. Correspondingly, in commerce data bases and database management systems had become an essential part of commercial computer systems.

However, in the first half of the 1980s there were signs of growing interest in databases by engineers. Large engineering projects, from space stations to ULSI chips, require the manipulation and storage of vast amounts of data, often by many designers, who are often geographically remote from each other. Data base technology appeared to offer a tool that could help. However, there was a serious limitation to what conventional data basse systems could do, as far as the design of engineering function was concerned.

Conventional data bases are largely static, or passive, and essentially provide a description of how entities, be it engineering components or whatever, interlock. In other words, they are good at storing the form of an engineering design. The other part of any engineering design is, of course, function, or how the design will behave, in terms of response in time to a stimulus. This type of data base clearly requires the involvement of time and event sequences.

The author wrote one of the first papers on the involvement of time (back in 1978) in data bases (when there was so little interest in the subject that he gave up and returned to more conventional data base pursuits). But in the 1980s, interest in this aspect of data bases, or active data bases, has snowballed, giving us the beginnings of a technology that can be used for both engineering form and function.

A parallel development has been knowledge databases and systems. These can be looked upon as extensions of passive and active data bases to include design rules and procedures. As a result we now also have the emerging area of engineering knowledge databases. The integration of this type of data base with passive and active engineering data promises to be very fruitful.

Another important area is image data bases and their interfaces with knowledge data bases. These data bases and the attendant tecnologies of pattern and image recognition, and entity perception and recognition, are clearly at the foundation of artificial vision systems for robotics. Research in this area is widespread, and there have been many developments. This type of work is heavily funded by the U.S. Department of Defense.

A futher area that is strongly in the field of engineering is database machines, or machines specially designed to obtain data swiftly from large data bases, using special memory techniques. Most of the work in this area has been carried out in Electrical Engineering departments in universities. Database machines are coming to market too, and already there are some 500 installed in the U.S. The use of database machines with supercomputers is an expanding research area, and the use of data base machines with image data bases for machine vision looks promising.

Finally, we can mention the most important engineering data base of them all, the one each of us carries in every cell in our bodies, namely DNA data base that contains the information about how we are constructed. Genetic engineering is advancing by leaps and bounds, and genetic engineering data bases, on disk, are in the beginning stages. Eventually, they will replicate the data bases that occur in living things, and will evolve to become the blueprints for modified species, although whether such data bases should ever be used is another question. At the very least, we can expect that these replicated genetic engineering data bases on disk will be an important tool in cancer research.

All these considerations lead us to the conclusion that the field of engineering data bases is going to grow, and will be growing very rapidly within 5 years. It is a fair bet that by the turn of the century, the amount of engineering data in on-line operational data bases (as distinct from text data bases, a relatively trivial kind of data base) will exceed the amount of data in conventional commercial data bases.

A great deal of the material on engineering data bases is scattered over many journals and system manuals. However, there is already one book on the subject (from Electrical Engineering Dept., Univ. of California at Berkeley). Many more books and papers can be expected in the near future.

A public seminar series at the University of Calgary would do a great deal to stimulate interest among engineers and computer scientists in this area - particularily in Canada. The tentative outline of the contents of such a course is given at the end of this report. Such a seminar could be given in the late fall of 1985 or early spring of 1986.

There will be many opportunities in coming years to develop engineering data base systems and associated software, and use these systems in advanced engineering design, and such a course should go a long way towards alerting professionals on both sides of the border about the opportunities. There should be little difficulty in getting full attendance for the course. The author is an internationally known data base expert (with many papers, two books published and more on the way). In addition, he has a publication and career background in engineering, with 7-years engineering experience in industry. He spent 1984-85 on sabbatical leave at the Electrical Engineering Department, University of Victoria.

Bibliography

Bradley, J. "Operations Databases", ACM Conf. on Very Large Databases, Berlin, 1978.

Bradley, J. File & Database Techniques, Holt, Rinehart & Winston, New York, 1982, 562 pages.

Katz, R. H. Information Management for Engineering Design, Springer-Verlag, June, 1985

U N I V E R S I T Y   O F   C A L G A R Y

A 3-day seminar, fall, 1985

E N G I N E E R I N G   D A T A B A S E S ,   A N   E M E R G I N G   T E C H N O L O G Y

Sponsored by the Faculties of Science and Engineering

Seminar leader:
J. Bradley, Ph.D.
University of Calgary

COURSE OUTLINE (Tentative)

1. FUNDAMENTALS OF DATABASE TECHNOLOGY

(a) Databases in engineering - an overview
(b) Large direct access files- primary key access
(c) Systems for non key direct access
(d) Integration of direct access computer files
(e) The essential role of data dictionaries
(f) Generalized database systems
(g) Data independence and database schemas

2. RELATIONAL ENGINEERING DATABASE STRUCTURES

(a) Theory of relations; expanded Cartesian product
(b) The normal forms; advantages of higher normal form relations
(c) Dependencies in relations; 2 and 3-D geometry applications
(d) Database relationships
(e) Hierarchical and network database structures
(f) Recursive relationships and complex structures

3. FOURTH GENERATION DATABASE LANGUAGES

(a) Procedural and non procedural languages
(b) Relational algebra
(c) Relational predicate calculus
(d) IBM's SQL, and ANS SQL
(e) SQL extensions for engineering databases

4. PASSIVE ENGINEERING DATABASES - ENGINEERING FORM

(a) Form versus function in engineering design
(b) Bill-of-materials databases (BOM)
(c) Building a BOM database with a conventional DBMS
(d) Creating new product subassemblies
(e) Determining component price impacts
(f) Use of SQL with BOM data bases
(g) Application to electronic circuit topology
(h) Application to digital systems topology

5. ACTIVE ENGINEERING DATABASES - ENGINEERING FUNCTION

(a) Inadequacy of conventional passive databases
(b) Incorporation of time and events - active databases
(c) Databases for physical engineering operations
(d) Use of SQL for tracing event sequences
(e) Application to industrial process control
(g) Application to logical and circuit design
(h) Application to complex engineering systems

6. ENGINEERING KNOWLEDGE DATABASES

(a) Integration of engineering form and function data
(b) Integration of engineering design procedures
(c) Concept of an engineering knowledge data base
(d) Knowledge database schemas
(e) Adaptation of conventional DBMSs
(f) New types of DBMS for knowledge data bases
(g) Engineering expert systems
(h) SQL and retrieval of design procedures
(i) Manipulation of engineering form and function
(k) Manipulation of engineering design rules
(h) Interface with generalized engineering object compilers
(i) Interface with conventional CAD systems
(j) Engineering decision support systems (EDSS)
(k) EDSS models and databases

7. IMAGE DATABASES AND COMPUTER VISION

(a) Image data base structures
(b) Integration of symbolic data bases
(c) Use of analysis software for symbol extractions
(d) Direct access using symbol extractions
(e) Image recognition using image data bases
(f) Image recognition using knowledge data bases
(g) Image and knowledge databases for robotics

8. DATABASE MACHINES AND SUPERCOMPUTERS

(a) The database machine concept
(b) Advantages and techniques of associative memory
(c) The engineering database machine
(d) Multiple database machine systems
(e) The Britton-Lee database machine
(f) Database machines for supercomputers
(f) Parallel methods for executing complex database functions

9 GENETIC ENGINEERING DATABASES

(a) Molecular databases
(b) DNA as a knowledge data base medium
(c) Access and control mechanisms for complex molecule construction
(d) Magnetic replication of molecular data bases
(e) Replicated data bases for genetic engineering design and research

10. OVERVIEW OF DATABASE COMMUNICATIONS

(a) Principles of on-line memory partitions
(b) Single and multi-thread systems
(c) Teleprocessing monitors
(d) Communications controllers