Data base management

Data base management is in a state of ferment due to the emergence within the past few years of many new requirements. Amongst these requirements are the need to make application programs and terminal activities much more independent of the internal representation of data in storage, and the need to support:

1. many different kinds of end users at terminals (some interactions being of unpredictable scope and complexity);
2. greatly enhanced data security and privacy;
3. increased dynamic sharing of data (including concurrent update and enquiry);
4. networks of mutually remote data bases (including very high level data sublanguages for low bandwidth communication of requests).

In the data base management sessions of this year's conference there is a strong emphasis on relational data base management. There are three reasons for this: first, this approach appears to be the most advanced in attempting to meet all these new requirements; second, the overwhelming majority of data management papers submitted to NCC 75 were concerned with implementing this approach; third, there has been almost no exposure of this approach before in this forum, even though the ideas have gained a solid acceptance in Europe and have spawned a pronounced surge in data base oriented research in numerous universities in North America (notably M.I.T., Toronto, Berkeley, Florida, and Utah).

What then are the distinctive features of this approach? This topic will be covered by Christopher J. Date in a tutorial to be presented in the first half of session L1. Mr. Date is the author of "An Introduction to Database Systems" just published by Addison Wesley. This book contains a thoughtful and very clear comparison of the hierarchic, network, and relational approaches. The second half of session L1 consists of a panel discussion on the two questions:

1. What are the major problems in implementing relational data base management systems?
2. Is there any necessary loss of performance if performance-oriented access paths are known to the system but not to the application programmer?

Each panel member has been directly involved in implementing a relational data base management system.

The remaining sessions L2 through L7 are based entirely upon submitted papers. Session L2 provides a striking contrast between two data base machines: one based upon hierarchic data structures, the other based upon non-hierarchic relations. Three more relational implementations are described in session L3. These implementations differ markedly in scope and style.

Session L4 deals with relational data base technology. A data base management system based upon the relational model must support a user view that is devoid of performance considerations. Since the problem of selecting efficient retrieval algorithms is removed from the user, this burden must fall upon the system itself. One of the papers in session L4 describes an unusual technique for efficient interpretation of data selection expressions which involve inter-entry relationships. A second paper in L4 introduces an important unifying mechanism for the services of concurrency locking, authorization, and support of multiple tabular views of data.

Session L5 is concerned with the human factors aspects of query languages and with attempts to develop objective experiments for evaluating these languages. The first paper in this session introduces a novel approach to querying a relational data base using a terminal display and employing a technique of specification by example. Then follows a psychological study in which human subjects (all non-programmers) were taught this technique and tested for speed and accuracy in formulating sample queries of various complexities given informally in English. The third paper reports on a human factors experiment designed to evaluate and compare two high level data base query languages in use by a sample of programmers and a sample of non-programmers. These research efforts can be expected to trigger many similar investigations in the future.

Session L6 deals with more traditional topics in data management. The three papers have as their topics data compression (a useful survey is provided), binary search trees (which are important for directories and certain kinds of indexes), and performance evaluation (a large scale simulation model is described).

The final session L7 deals with distributed data bases and two important application areas—the medical field and urban management. The paper on distributed data bases provides a framework in which to tackle the problem of allocating files and programs to the nodes of the network. A second paper describes a clinical data base system that attempts to cope with the multiple user and data entry problems. The urban management paper suggests that totally integrated on-line data bases for urban problem solving and decision making are not yet practical, but that the employment of data files extracted from a common base for various specialized uses represents a feasible approach.