PRACTICAL PROBLEMS IN DATA MANAGEMENT -- A position paper

Jim Gray
Tandem Computers, Cupertino CA 95014

Database theory seems to lag behind or be orthogonal to the practical problems of database design and implementation. In my own specialty of transaction management, algorithms appear in working systems five years before they appear in the literature. The best theoretical papers explain the algorithms, expose subtle bugs in them or generalize them.

Most researchers do not view system descriptions as part of the literature. This is true both in research (they do not read system descriptions) and in publications (if an idea is well documented in some system but is not "published" in a journal, the idea is viewed as "unpublished"). An important practical problem to be solved is to get theoreticians to read the papers (manuals) written by practitioners. I beleive this would inspire abstractions, clarifications and generalizations of mechanisms not fully thought out or realized by practitioners.

Systems worth examining include IMS-FastPath, Encompass, No. 4 ESS, R*, Ingress, View 3000, Star, and Lisa.

Here are some hard problems faced by practitioners today.

Design a system that can run 10,000 transactions per second against a 1 record database. An even more difficult problem is running 10,000 transactions per second against a 2 record database. See IMS-FastPath for a hint.

The area of database design is in a mess. Design:

- a theory of normalization which can be understood by a personal computer user,
- algorithms which given a logical database design do all the physical database design (and redesign),
- powerful languages for data manipulation and display including: tabular, form, icon and graphical display.
- performance measurement and modeling tools so that users can measure and direct the physical database design.

The area of data display is much neglected. Many systems do an excellent job of organizing and manipulating data but very few (if any) have a clear model of how to display stored information and how to accept new information. The advent of:

color bitmap displays
voice input/output devices
facsimile input/output devices

will revolutionize this area (no comparable revolution is slated for the data storage or manipulation area).

Managing change is another unexplored topic. Change is a major source of faults in fault tolerant systems. How can one reorganize a database while it is operating? How can one install a new data management system while it is operating? How can one install new hardware while it is operating? The phone companies have done the most work in this area, but there is little literature which presents concepts or tools to manage change.