

Database System Performance Measurement

Chairperson Haran Boral - MCC
Panelists Richard Brice - MCC
 David DeWitt - U of Wisconsin
 Dieter Gawlick - Amdahl
 Paula Hawthorn - Britton Lee

In the past few years we have seen in the literature a number of proposals for benchmarks to be used in measuring the performance of database management and transaction processing systems. The TP1 benchmark [Anon et al 1985] and the Wisconsin benchmark [Bitton et al 1983], [Boral and DeWitt 1984], and [Bitton and Turbyfill 1985] have been used to benchmark several systems. Other benchmarks have also been proposed.

The TP1 benchmark actually consists of three different benchmarks: Debit-Credit, Scan, and Sort. It is oriented towards transaction processing systems. Each of the benchmarks consists of a single transaction type and operates on a large database -- around 10 GBytes. The database consists of artificial data but is modeled around data maintained by a large bank.

The Debit-Credit benchmark consists of a transaction that reads and updates a small number (about 4) of random records. It imposes stringent response time and throughput requirements on the system.

The Scan benchmark consists of a COBOL program that exercises the system by executing 1,000 scan transactions each of which accesses and updates 1,000 records in a sequentially organized file.

Finally, the Sort benchmark sorts 1M records of 100 bytes each.

Each of the benchmarks stresses different aspects of the system. Each requires different amounts of CPU, communication, and I/O cycles. In addition to the diversity of system resource requirements the benchmark metho-

dology described in [Anon et al 1985] also requires that the cost of the system be calculated. Thus, the final measure one obtains from running the TP1 benchmark is \$K/TPS.

Whereas TP1 is oriented towards transaction processing the Wisconsin benchmark was conceived for the purpose of measuring the performance of relational database systems. It consists of two parts: a single user benchmark in which a suite of approximately 30 different queries are used to obtain response time measures in standalone mode (described in [Bitton et al 1983]), and a multi-user benchmark in which several queries of varying complexity are used to determine the response time and throughput behavior under a variety of conditions (one version of the multi-user benchmark is described in [Boral and DeWitt 1984] and a second version in [Bitton and Turbyfill 1985]).

The test database consists of a number of relations of varying sizes. The relations are generated according to statistical distributions and do not model any real world data. Users of the benchmark can modify the database generator routines to adapt the database characteristics so that they are more representative of their application.

It appears as though both the TP1 and the Wisconsin benchmark have the potential of becoming de facto standard benchmarks, in their respective areas, to be used in a variety of ways. For example, a vendor could use the benchmarks to stress test a system under development. Another use for a vendor is in establishing a particular rating for a system (equivalent MIPS Whetstones, etc for mainframes). Finally, a user can use a benchmark to compare several systems before purchasing one.

The purpose of this panel is to discuss the use of benchmarking for measuring the performance of transaction processing systems and database management systems in general and

the use of the TP1 and Wisconsin benchmarks in particular

The panelists have been chosen so that we have a representation of experts in the particular benchmarks (Gawlick and DeWitt), a benchmark "consumer" (Hawthorn), and a "performance expert" -- someone who understands benchmarking as a science/art (Brice)

The panelists will address the following issues (as well as others raised by the audience)

- (1) What are the strengths and weaknesses of the TP1 and Wisconsin benchmarks?
- (2) Is benchmarking a good technique for measuring the performance of data management and transaction processing systems?
- (3) What can these benchmarks tell us about a system and what can they not tell us about it?

References

- [Anon et al 1985] Anon et al , "A Measure of Transaction Processing Power," *Datamation*, April 1, 1985
- [Bitton et al 1983] Bitton D , D J DeWitt, and C Turbyfill, "Benchmarking Database Systems A Systematic Approach," *Proceedings of the 1983 VLDB*, 1983
- [Bitton and Turbyfill 1985] Bitton D and C Turbyfill, "Evaluation of a Backend Database Machine," *18th Hawaii Int'l Conf on System Sciences*, 1985
- [Boral and DeWitt 1984] Boral H and D J DeWitt, "A Methodology for Database System Performance Evaluation," *SIGMOD '84 Proceedings of the Annual Meeting*, 1984