The Enterprise Database: How Does M Compare?

by Frederick L. Hiltz and Pasha Roberts

M has grown to compete with traditional solutions for enterprise computing on a large scale, yet the technology remains relatively unknown. As the Brigham and Women's Hospital installs a large M information system with a client/server architecture, observations of its database performance in daily production invite comparison with other ways of serving about 33 gigabytes of data to fifteen hundred users. [1] Informal comparisons address throughput, availability, management techniques, expansion, integration, and security.

We would like to afford the considerable effort and expense of a formal comparison such as the TPC benchmarks specified by the Transaction Processing Performance Council. [2] In its absence, however, the literature and interviews with vendors and users of comparable systems have provided some points of comparison. Taking care to recognize the "apples and oranges" nature of some comparisons, we feel that even informal measures can show the standing of a client/server M network among some popular alternatives.

Because it is the least well known and because we know it best, we present the Brigham and Women's Hospital (BWH) approach in the most detail, followed by DSM for OpenVMS (from Digital Equipment Corporation), ORACLE7 (Oracle Corporation), and SYBASE (Sybase, Inc.), chosen as typical and popular alternatives among many.

The Application

Brigham and Women's Hospital (BWH) is a 726-bed teaching hospital in Boston, closely affiliated with Harvard Medical School. In 1992, it cared for 8,896 births, handled 585,090 outpatient visits, and brought in $590 million in gross revenues. The hospital is the site of extensive research; it is the largest hospital recipient of National Institutes of Health funds.

The BWH information system has been an important factor in the hospital's success. To enable further expansion of the system, in 1991 the hospital began moving its ninety-six application modules, which support virtually the entire enterprise, to a client/server network of personal computers running DataTree M from InterSystems. The network comprises 2,500 computers; 4,000 are planned by 1995.

The application requires constant availability, that is, no scheduled shut-downs. Brief interruptions due to a hardware failure or power outage are tolerated.

Alternative Approaches

BWH—The enterprise network, shown in figure 1, is built around fifty-six local IBM token rings, allocated by geographic and load considerations. Each ring serves about forty workstations, and a bridge connects it to a backbone token ring where Novell servers provide file service, initial program load, and personal computer (PC) applications such as WordPerfect, Quattro Pro, and Paradox.

A router PC on each local ring runs DataTree M and buffers data germane to the locations wired by the ring. The routers connect to the database servers via a second backbone token ring. The servers' disks store the database, which is not replicated. When a workstation needs data, the router passes the request over the backbone ring to the server and returns the data to the workstation. McManus, Glaser, and Reder and Thrasivoulos have described the design and implementation of this architecture in more detail. [3,4,5]

Most of the two thousand workstations are diskless PCs with Intel 386- or 486-series processors. Forty routers and thirty servers contain 486-series processors; the servers support 2.3 gigabytes of disk storage each.

BWH's database has hierarchical structure internally, and programmers deal with it in that form. Many M products present a relational view of the database. Our purpose, however, is not to enter the database structure controversy, but to focus on connecting large databases to an enterprise's applications whatever the method.

DSM for OpenVM—VAX and Alpha AXP processors combine with virtually unlimited amounts of disk storage in VAXcluster systems. The nodes may be loosely coupled via Ethernet or tightly coupled through high-performance proprietary connections. One or more DSM systems run on each node and terminals (character based or X/) log on symmetrically to a node chosen by a load-balancing process.

ORACLE7—Oracle offers versions for most hardware. Some large enterprise systems are running on VAXclusters, often with terminals connected by serial lines or by LAT and

Continued on page 22

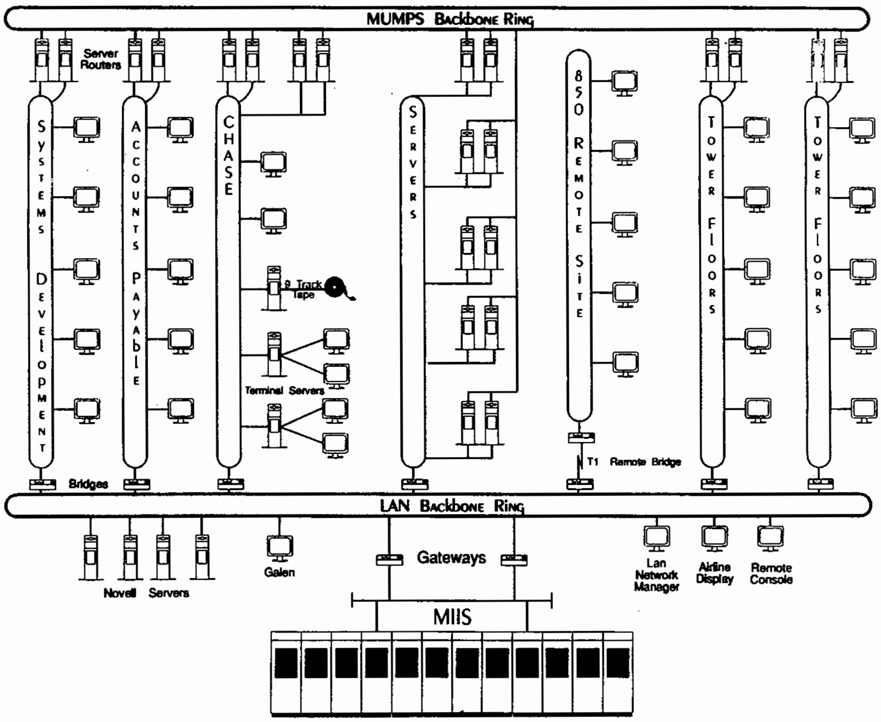

Figure 1. The architecture of rings, servers, and workstations.

Continued from page 20

Ethernet. Others run under UNIX on multiprocessor computers supporting PC or X/ terminals. ORACLE7 also supports PC, Macintosh, and UNIX clients on local networks in a client/server configuration. The ORACLE7 database is relational in structure and in appearance.

SYBASE—The SYBASE relational database is available for most computers, supporting both terminal/host and client/ server architectures including PC and Macintosh clients. We know of no SYBASE sites as large as the BWH configuration.

Throughput

Meaningful comparisons of throughput require identical tests in controlled conditions on explicitly defined hardware. The TPC benchmarks do this; we do not. We report on successful applications of each database system.

BWH—When data are averaged over two weeks, 24 hours a day, the servers performed 702 physical disk block operations per second. Assuming one logical reference per block and that 95 percent of blocks are found in a memory cache, there were 14,040 database references (gets/sets/kills) per second by M application programs. A user signed on every three seconds, on average, during the period.

Users consider the performance good, but with occasional one- to five-second delays for large blocks of data. One application, order entry, processed complex transactions involving several servers at an average rate of one every 1.3 minutes during 101 days of normal production. Its maximum rate is unknown.

DSM for OpenVMS—Large clusters typically perform 100 physical disk operations per second per node, or 20,000 to 30,000 gets/sets/kills per second total.

ORACLE7 and *SYBASE* are considered very fast, although SYBASE is untested in environments of one thousand-plus users and multiple servers. Godfrey reports SYBASE speed twelve times faster than DataTree M in one small single-user search test. [6]

Availability

BWH—The platform does not offer disk mirroring or redundant array of inexpensive disks (RAID) to survive disk failures; it uses redundant inexpensive servers to survive both disk and processor failures. A shadow server reads each server's journal (a log of database changes) and maintains an exact copy of the server's database on its own disks. When a server fails, an operator's menu choice substitutes its shadow. Workstations display an advisory message and their tasks pause for about two minutes, then the tasks resume their sessions with the new server.

Daily backup copies the shadows' disks to tape, during which a disk elsewhere on the network assumes shadowing duties.

DSM for OpenVMS—RAID is available. Mirrored disks overcome disk failures invisibly and without delay to applications. One mirror copy can be backed up to tape while the other remains on line. Incremental backup can copy a database that is changing. Database replication is available but seldom needed with mirrored disks and the high-performance interconnection to all nodes of a cluster.

Failure of one node does not affect others. Affected terminals log on to another node and start a new session. Two kinds of journaling (or backup)—before- and after-image—can assure database integrity in the event of disk and central processing unit failures even without hardware redundancy. Disaster tolerance extends a cluster across tens or hundreds of miles with fiber optic cable or microwave links.

SYBASE and ORACLE7—use the disk and backup facilities of their underlying operating systems, most of which offer disk mirroring and online backup to tape. RAID is common. Failure of the server terminates clients' tasks, which must be restarted after repairs. Both products permit journaling, and both can replicate part or all of the database, providing some resilience and substantial speedup at the cost of complexity in maintaining database integrity. ORACLE7 will soon offer shadow servers.

Management Techniques

Large systems present a real management challenge; the solutions may well be their most important features. Centralized architectures provide centralized management, but distributed databases need extra support. It is difficult to measure the management effort required of each system, but the different tools available deserve mention.

BWH—The flexibility and power of M have enabled much development of central database manager's tools, augmenting the vendor's utilities. M programs use NetWare APIs (application program interfaces) and observations of the network to gather information from the lower levels, add application statistics, and assimilate them all into a common database.

The most visible of these tools is the "airline screen" monitor, developed at BWH, which integrates exception reporting from all levels. The name refers to its display, containing an entry for every M and Novell server, just as an airport display shows the status of incoming flights. Servers run diagnostic software once a minute; error conditions cause the node to change color and sound alert tones on the monitor. The software also corrects errors when possible, making the network selfreporting and selfcleaning. Operators, programmers, and managers see the status of the system at a glance, including the number of clients on each server—a rough estimate of activity.

Many more detailed statistics are collected, tabulated, and graphed every six minutes. Most of this information is about M activity, but similar systems for Novell networks, such as LANAlert and NetWare Management, are being introduced. These welcome developments will be integrated in the future.

Supporting the database manager's tools are utilities for the network and its operating system. LAN Network Manager maintains configuration, reports activity, and signals alarms at the wire level. Novell's NetWare adds some control, providing more reports, user identification, file services, initial program load, and common PC applications. DataTree M checks and restores internal database integrity. It also reconfigures the database, adding and moving data sets or even entire servers while applications continue.

DSM for OpenVM—The system manager's utilities treat the cluster as a whole for reporting, control, and mounting and dismounting of dataset volumes. The whole computer room can be managed from one terminal, which may be remote. Third-party products augment the extensive set of management tools supplied by Digital Equipment Corporation.

ORACLE7—Beyond the system-management functions of the underlying operating systems, ORACLE7 provides online database configuration but little more. As a result, an active third-party market has arisen to offer some good tools for the tasks common to all database management systems.

SYBASE—offers the system-management functions of the underlying operating systems and some performance tuning tools, but not online database configuration. Good third-party tools are available.

Expansion

Terminal/host platforms and client/server platforms grow differently. Adding terminals to a host slices the performance pie into smaller pieces. When the pieces get unacceptably small, a bigger host restores performance. Adding clients to a server adds the clients' computing power to match their applications' demand. Servers need to grow, too, as the database load grows, by adding more or bigger servers.

BWH—The hospital information system will double in size by 1995, to four thousand workstations and 60 gigabytes of data on fifty servers. Each new workstation will run one interactive task, bringing ample computing power and buffer memory for programs and data, so response time for its user should not degrade. The token ring network and the servers' disks are lightly loaded. CPU speed and buffer capacity in the server/routers will expand easily; adding one more server to a collection of forty will cause no disruption and little incremental cost. Batch processing is not dead; a dedicated server runs background M tasks, and more will be added as necessary.

DSM for OpenVMS—Disks, nodes, and even new interconnections can be added to a cluster, which may include both VAX and the new Alpha AXP processors. Dataset volumes can be added, moved, or expanded on any node by moving users to another node during the process. Future releases will allow users to remain active on the node while volumes are mounted and dismounted.

ORACLE7—Several customers operate 100-gigabyte databases that are roughly equivalent to BWH's 33 gigabytes, given the more compact nature of M globals and the absence of replication at BWH. ORACLE7 runs on both terminal/host and client/server platforms, which determine the growth path. Private Healthcare Systems (a user we interviewed while preparing this article) expects to grow from one hundred to one thousand users and 70 gigabytes by 1994. [7]

SYBASE—As with ORACLE7, the SYBASE expansion path depends on the underlying platform. In its current version, SYBASE may not scale to 100 gigabytes very well due to some internal size limits. Programs must explicitly refer to the server on which a table resides, making database reconfiguration awkward.

Integration

M, which combines an application development language and a database, integrates them very well but often slights the outside world of other languages and applications. Dedicated database management systems like ORACLE7 and SYBASE, having always depended on other products for application development, offer easy connection to external programs, typically through the SQL (structured query language) data-manipulation language.

BWH—Many workstations run conventional PC applications in addition to the M hospital information system. Choosing a PC application halts M, runs the application from an MS-DOS batch file and then restarts M. The 30-second switch time and limited transfer of data via DOS files are real nuisances that a multitasking environment on the workstations would eliminate. Transfer of data to other MS-Windows applications via Windows services is likely in the future, but this will not help Macintosh or UNIX clients.

DSM for OpenVMS—offers an external calling interface both ways: M calling other languages and vice versa. With this foundation, M programs gain access to the VMS file system, multitasking services, and applications written in other languages. Complementing the calling interface, a data mapping facility lets M programs work with C structures and function prototypes. Open Data Base Connect (ODBC) provides a language similar to SQL for external access to M databases.

ORACLE7 and *SYBASE*—Client programs in any language that can send SQL statements over their network have full transparent access to tables on any server, and to the definitions of those tables' contents, or schema.

Security

Of the many facets of security, authorization is most important for database access. Privacy, authentication, and identification belong to other parts of the computer system. Application programs can perform authorization in all the alternatives, some of which also offer authorization as part of database management.

BWH—Users need access to M applications, NetWare services, and (for another year or two) the retiring minicomputer network. Most people would not accept three different identification/password combinations, nor is this necessary. Administrators authorize users for specific applications. Together with the passwords, authorizations originate in the old network and pass through a gateway to an M program that relays them to NetWare. As in most M implementations, the restriction of users to authorized operations is effective when running applications, but completely absent in "programmer mode."

DSM for OpenVMS—The operating system provides identification and authentication of users, and M applications enjoy very effective restriction of programmers to authorized environments only. [8] VMS categorizes users as managers, application programmers, and application users. They are separately authorized for DSM, specific M applications, and dataset volumes. Authorization may extend to read, write, and delete privileges on individual globals. VMS provides Department of Defense C2 security, which M programs can extend to applications.

ORACLE7—has very tight yet flexible security. The system administrator authorizes users for different kinds of access to tables and column individually or by "role" or group. Application programs can also be given similar authorizations independently. Roles are inheritable.

SYBASE—has an excellent authorization scheme based on a model similar to ORACLE7's, although roles are not inheritable.

Conclusion

The M database is one of the largest serving the BWH enterprise. Its performance, while adequate, is not the equal of more mature large database products. BWH's network of loosely coupled computers affords more flexibility than the alternatives, which renders it valuable for responding to changing requirements and for easily taking advantage of evolving computer hardware.

Managing this new architecture has been a do-it-yourself project, whereas most of the tools can be bought for the traditional alternatives. Management of any system this big, however, requires a substantial investment.

BWH's system has grown by a factor of one hundred in the past two years, and can grow by another factor of two to five less painfully than the alternatives can.

As with most M installations, BWH's integration with the rest of the world is poor. We are counting on the language and operating system vendors to make possible the outside database access so long enjoyed by users of SYBASE and ORACLE7. Also like most M installations, restriction of the database to authorized users is nonexistent for programmers. Application programs must manage authorization for users, a task handled by the DBMS of alternative solutions.

We have purposely avoided comparing costs, a futile exercise in a general discussion such as this. The BWH alternative, however, is probably less expensive than the others considered here, due to its use of personal computer hardware and to M Technology.

M Technology does compete in the large database market, adding valuable alternatives to the variety of available choices.

The authors gratefully acknowledge Jerry Melnick of Digital Equipment Corporation (Marlboro, MA), Richard White of Private Healthcare Systems (Lexington, MA), and John Godfrey of Cornell University (Ithaca, NY) for their advice and information about DSM for OpenVMS, ORACLE7, and SYBASE. Any shortcomings in the descriptions or interpretations of these products are the authors' alone.

Drop us a Note!

Tell us which article you think is the best in this issue.

Frederick L. Hiltz is a principal software engineer at the Center for Applied Medical Information Systems Research of Brigham and Women's Hospital, where he applies the enterprise network to hospital information requirements.

Pasha Roberts is a technology designer at the Technology Planning Division. He has developed much of the systems and network management infrastructure that supports the Brigham's network, and is constantly looking toward future applications of enterprise data.

Endnotes

1. All brand and product names are trademarks or registered trademarks of their respective companies.
2. TPC benchmarks are specified by the Transaction Processing Performance Council, c/o Shanley Public Relations, 777 N. First Street, Suite 600, San Jose, CA 95112-6311.
3. P. A. McManus et al., "Distributed Design for a Large-Scale Workstation Network," *MUG Quarterly* 19:4 (April 1992): 15–19.
4. Glaser et al., "A Very Large PC LAN as the Basis for a Hospital Information System," *Journal of Medical Systems* 15:2 (1991).
5. Reder and Jacquelyn Thrasivoulos, "Planning for the Next Millennium with M," *MUMPS Computing* 22:5 (November 1992): 7–11.
6. John Godfrey, personal communication.
7. Richard White, personal communication.
8. Ron Watt, "Security in VAX DSM Version 6.0," *MUG Quarterly* 21:3 (June 1991): 106–111.