

THE DATA MANAGEMENT MACHINE, A CLASSIFICATION

Robert S. Rosenthal
MRI Systems Corporation

There has been much interest in the use of special purpose processors as the data base management component of data processing systems. The generic terms "backend" and "data management machine" have been applied to such devices. Examination of the literature reveals a broad cross section of host to backend functional distribution and interconnection methodology. This discussion represents an attempt to examine and classify several of these backend data base management machine configurations in terms of their operational parameters and application constraints. A formal taxonomy of such systems remains yet to be performed.

At least three distinct classes of data management machine (DMM) are evidenced in the literature; they are the large host backend, distributed network data node and smart peripheral. The intended classes of problem that the various authors envision amenable to solution by the DMM approach exhibit overlap while the performance envelope in which each DMM architecture would provide a technically acceptable, economically sound solution to a given user requirement set varies. Some of the papers used as source for this work contained no explicit mention of either the problem classes or performance constraints that the described configuration was to address; thus liberty has been taken in interpreting the implicit application goals of these authors.

LARGE HOST BACKEND

The Large Host Backend is a semi-autonomous processor which is closely coupled to a general purpose host or host network and has access to a complement of secondary and tertiary storage on which data bases reside. The function of the Large Host Backend is to provide data base management services for the host. The term "closely coupled" as used here is weaker than "tightly coupled", which is applied to a multiprocessing network in which main storage is shared between two or more processors. This term is used to denote communications between the host and its data base management components

at an effective data rate and with processor overhead in the same range as those of a present day rotating storage peripheral device.

The motivation for such an architecture has been stated to lie primarily in the expectation of reduced mainframe processing load, increased data base security, and higher dbms performance. These benefits are based on the economy of specialization in the backend hardware and control program. It has been assumed that the bulk of the dbms algorithmic and control program processing is removed from the host to the backend processor. The level of data manipulation language (DML) could be similar to that now utilized by currently implemented "navigational" (point record oriented as distinguished from set oriented) data base management systems without creating severe system thruput bottlenecks. In order to achieve the anticipated benefits, some backend infrastructural constraints must be observed. The author has covered these more fully in a previous paper [11]; however, a summary includes:

- Tightly coupled multiprocessing within the DMM
- Adequate processor address space/task address space to meet buffering and executable code requirements
- Minimal context switching overhead
- Powerful data manipulation at the character level

NETWORK NODE DATA MANAGEMENT MACHINE

The Network Node Data Management Machine (NNDMM) is a loosely coupled autonomous processor which is an element in a network comprised of both general and special purpose processors. The NNDMM has direct control over all or a portion of the total data base accessible to the network and functions as a co-equal member rather than a slave to a specific host; that is, it may request data from other nodes in the pursuit of its task. The loose coupling aspect implies a telecommunications linkage although this is not necessarily the case. The basic attribute of a loosely coupled interface is the bidirectional nature of control flow. The services provided within currently available commercial common carrier technology offer low data rates relative to channel or bus coupled backends, very high linkage processing cost per message, and slow line turnaround time. The near future holds promise of packet switching, digital satellite communications networks, and fast dial. These technologies all accentuate the long turnaround time problem. Even in the case of the most rapid, a satellite network, propagation delays of over two seconds per leg are unavoidable. Hence, the low level navigational DML becomes undesirable both in processing overhead and thruput restrictions. Distribution to the DMM node

of more sensitivity to the requesting node's procedural requirements as well as the use of a higher level DML (e.g., query language) which offers greater data reduction as discussed by Lowenthal [5] and lower message traffic count provides a workable solution to these problems.

THE SMART PERIPHERAL

The third category is that of the spectrum of smart peripherals. At the low end are the 'intelligent controllers' for conventional rotating storage, which simply move some of the access method function from the mainframe to the controller. Frequently mentioned candidate functions for migration into the controller are error detection and correction, arm position scheduling, and record content search. Some small scale data base management capability has been implemented in this manner. The high end of the range is represented by complete controller resident data base processors such as the Relational Associative Processor (RAP) [10] and CASSM [4]. These systems are based on head-per-track rotational storage and electronic memories implemented in CCD.

Although these systems are in their infancy, they promise a breakthrough in performance, offering two revolution time location and retrieval and single rotation time update. Even continuous available space consolidation can be performed during idle time.

No specific application limitation can be discussed for this class. The 'smart controller' is a direct replacement for existing controllers. It may reduce the host central processor burden somewhat, but does not offer sustainably improved performance or function. The controller resident data base processor has not yet been investigated at a level which would allow a determination of its best uses.

CONCLUSION

The distributed computing network of the future will likely contain a range of nodes which encompass a combination of the subarchitecture discussed here. The diversity of requirement imposed by large data bases, high volume batch processing, remote and local query and local data sharing will tend to shape development into a modular direction. It is evident that these dissimilar nodes, operating in diverse manners, can not all be called "backends"; thus a definitive taxonomy would be a welcome entry in the field.

BIBLIOGRAPHY

1. Canady, R. H., et al. A back-end computer for data base management, Comm. ACM 17, 10 (Oct. 1974), pp. 575-582.
2. Codespoti, D. J., Maryanski, F. A microprocessor enhanced distributed operating system: a simulation mode, Technical Report CS76-01, Computer Science Department, Kansas State University, (Jan. 1976).
3. Date, C. J. and Codd, E. F. The relational and network approaches: comparison of the application programming interfaces, (Report RJ 1401) IBM Thomas J. Watson Research Center, Yorktown Heights, New York (June 1974).
4. Lipovski, J. G., and Su, S. Y. W. On non-numeric architecture, NSF Grant Number G3-43222, Department of Electrical Engineering, University of Florida, (1975).
5. Lowenthal, E. I. The backend computer, Part I and Part II, Auerbach (Data Management) Series, 24-01-04 and 24-01-05 (1976).
6. Lowenthal, E. I. The distributed data management function, Proceedings of the National Computer Conference, 1974.
7. Maryanski, F. J., and Wallentine, V. E. A simulation model of a back-end data base management system, Technical Report CS76-12, Computer Science Department, Kansas State University, (April 1976).
8. Maryanski, F. J., Fisher, P. S., Wallentine, V. E. Evaluation of conversion to a back-end data base management system, USACSC Grant No. DAH-75-G-0137, U. S. Army Computer Systems Command, Fort Belvoir, Va. (1975).
9. Moreira, A., Pinheiro, C., D'Elia, L. Integrating data base management into operating systems - an access method approach, Proceedings of the National Computer Conference, 1974, pp. 57-62.
10. Ozkarahan, E. A., Schuster, S. A., Smith, K. C. A data base processor, Technical Report CSRG-43, Computer Systems Research Group, University of Toronto (November 1974).

11. Rosenthal, R. S. An evaluation of a backend data base management machine, Proceedings of the Annual Computer Related Information Systems Symposium, U. S. Air Force Academy, 1977.

12. Wallentine, V., et al. Technical report on the implementation of a backend data base management system, Department of Computer Science, Kansas State University, 1975.