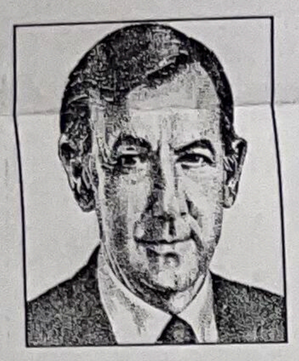
APPLIED INTELLIGENCE

Centralized Data Access Will Be Essential for Corporations



JAMES MARTIN

This is the third in a series of articles on key industry trends affecting computer hardware, software, database environments and communication environments.

A major trend in the 1990s will be the computerization of large corporations.

As shown in the figure, strategic systems required to run a corporation will need access to data that may be distributed throughout the world on a wide variety of platforms. Centralized decision-making applications will require this data to be drawn from across the organization into a centralized repository. The information in the central repository may be used to support corporationwide financial-planning applications, production-management systems and other strategically important applications.

To build these large, strategic systems, corporations are increasingly taking an organizationwide view of database resources and information modeling. They view information as a key business resource that must be managed properly to gain strategic advantage; however, to gain access to this information, it is necessary to provide an integrated data store and data model.

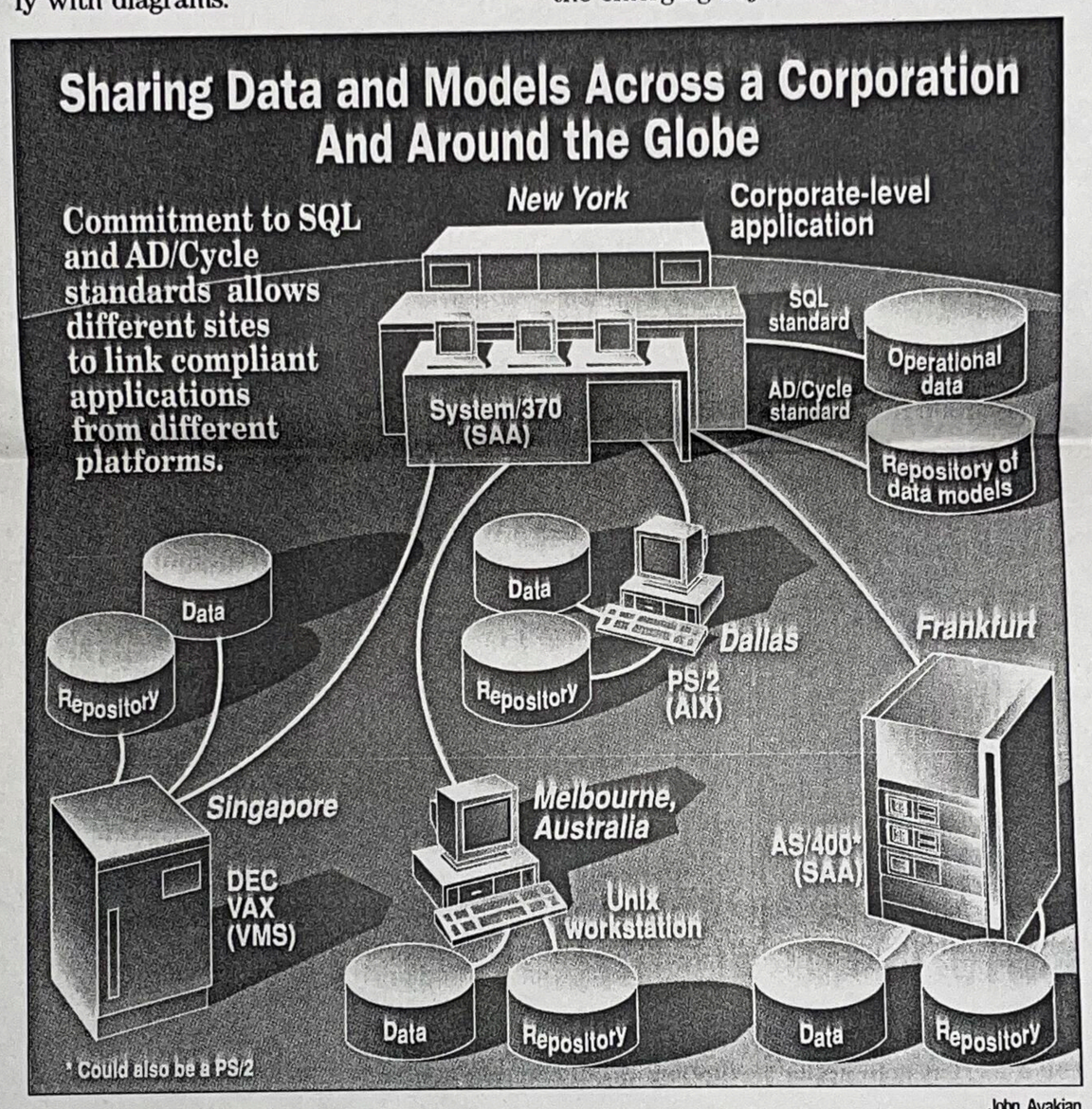
Corporationwide Data Access

The figure illustrates a problem facing many organizations attempting to implement an integrated data store. Through the years, the various divisions of a corporation may have committed to a wide range of hardware and software environments to support their different operational requirements. Manufacturing divisions may be using Digital Equipment Corp. VAX computers in the VMS environment for shop-floor control operations. R-and-D facilities may be using engineering workstations and RISC machines running under a variety of Unix operating systems. Business-oriented applications may be operating on IBM PCs, RTs, AS/400s and S/370 mainframes.

Each of these hardware and software environments may use a different database-management system. All of these systems may have been selected as an optimal solution for a local processing problem; however, the variety of incompatible systems operating within a large organization makes it extremely difficult to implement corporationwide applications.

Corporationwide database access has been greatly simplified through the use of the relational data model and the Structured Query Language (SQL) for data access and retrieval. Unfortunately, SQL is not appropriate for direct use by end users. SQL syntax is arcane, and even logically simple queries can result in a set of convoluted statements. The remedy is not to abandon SQL, but to introduce a better interface to the language.

Natural-language interfaces, simple verbs and graphical front ends are being used as interfaces to SQL. With a natural-language interface, queries are specified in a conversational language, such as English, and then automatically translated into the corresponding SQL statements. Likewise, the verbs used for data access in the Common Programming Interface (CPI) within IBM's Systems Application Architecture (SAA) translate into SQL statements. With a graphical front end, such as that of a computer-aided software egineering (CASE) tool, the user depicts the data elements and summarizations graphically with diagrams.



John Avakian

Corporations view information as a key business resource. Providing an integrated data store and data model is necessary for corporationwide access to this information.

Data models implemented across multivendor hardware platforms and supporting server architectures will be increasingly important. Products such as Oracle, Ingres and Sybase can be used to write applications for widely diverse hardware environments. From a corporate vantage point, it is important to maintain commonality among these diverse databases at the SQL level through the use of ANSI-standard implementations of database-access products.

Integrated computing architectures, such as IBM's SAA and Digital's Network Architecture Support (NAS) provide comprehensive support for data access from a distributed network of diverse hardware platforms. Within SAA, remote data access between networked IBM computers is handled via automatic generation of communication requests using the LU 6.2 peer-to-peer communications protocol. Access of data from non-IBM machines is handled through the automatic generation of database-access requests using OSI protocols such as the X.25 packet-switching protocol. These mechanisms enable data to be distributed and accessed from a network of diverse systems, all of which obey an SQL language standard.

Akin to the relational data model is the emerging object-oriented data model, in which attributes of the data are stored along with each data element. Object-oriented database-management systems will become increasingly important in the development of advanced artificial-intelligence, expert-system and CASE products.

Another problem facing organizations striving to implement global information systems is the maintenance of corporationwide data models. Each division within a corporation should build applications using a common data model to define entities, entity relationships, process models and so on. In this way, applications built on Digital VAXes, engineering workstations or personal computers will share a common pool of design information using a standard repository.

One way to ensure compatibility of data models across the corporation is to provide a central repository of design information, which can be accessed by application-development tools on machines anywhere within the network. For example, a CASE tool operating on an engineering workstation may extract and download a portion of the centralized data model that is required for the development of a specific application. During the development process, additional entities or changes to entity relationships can be defined. Using strict data-administration control procedures, these changes to the local data model can be incorporated back into the central data model.

The AD/Cycle Standard

Among a network of IBM machines operating under SAA, the AD/Cycle repository standard can be used to maintain consistency of data models across platforms. This standard defines a data metamodel that describes the information stored in the repository. For a CASE tool, for example, it provides a standard definition of the information content of a dataflow diagram, decomposition diagram or entity-relationship diagram. If AD/Cycle-compliant CASE tools are used on all platforms, then these platforms will share a common data model for application development.

Currently, AD/Cycle is still a dream. No vendor has yet introduced an AD/Cycle-compatible CASE product; however, a large number of leading CASE vendors have committed to the provision of AD/Cycle-compatible products that can be used to share design information, as defined by the data metamodel.

AD/Cycle-compliant CASE tools are likely to be available for a wide variety of development platforms, including the IBM PS/2 and RS/6000, Apollo and Sun workstations and Digital VAXstations.

Since no CASE vendor has yet implemented AD/Cycle-compliant repositories, it is unclear whether compliance with AD/Cycle will permit the integration of CASE tools from multiple vendors. If application development cannot be coordinated through AD/Cycle, then other mechanisms will have to be considered, such as multiple proprietary repositories linked by data interchange languages, such as CDIF (CASE Data Interchange Format) or ESF (External Source Format).

Next week, I will look at major trends in communications software. ■

The concepts embodied in this article are described in the High-Productivity Technology volume in The James Martin Report Series. For more information on this volume, call (617) 639-1958. For information on seminars, contact (in the United States and Canada) Technology Transfer Institute, 741 10th St., Santa Monica, Calif. 90402 (213) 394-8305. In Europe, contact Savant, 2 New St., Carnforth, Lancs., LA5 9BX United Kingdom (0524) 734 505.