IDC Analyze the Future

I D C V E N D O R S P O T L I G H T

Object Databases: Worth Another Look

March 2011

Adapted from *The Third Generation of Database Technology: Vendors and Products That Are Shaking Up the Market* by Carl Olofson, IDC #221939

Sponsored by Versant

As the nature of business evolves, enterprise data and data analysis evolve along with it. Long a staple of enterprise applications, relational databases are falling short when it comes to data that has a very complex structure and unstructured, dynamic data, especially that which is associated with multimedia and social networking. Enterprises are beginning to supplement their traditional approaches to database management with nonstructured databases and object-oriented databases. This Vendor Spotlight provides an overview of database technology's evolution to a third generation that can provide not only linear storage and analysis but also nonlinear storage for more behavioral ("social") analysis. In addition, this document discusses how cloud computing and open source software (OSS) are driving the adoption of nonschematic ("NoSQL") and other nonrelational databases, such as object-oriented databases. Finally, this Vendor Spotlight looks at Versant, a provider of object-oriented database technology and tools.

The Changing Nature of DBMSs in the Enterprise

Database technology has been on a course of continuous evolution since its inception, driven by a combination of user need and computer technology improvements that have made various memory, computational, and storage resources better, faster, and cheaper. New products continue to emerge, and existing products continue to evolve, driven by both economic and market opportunity pressures.

The latest wave of innovation has been the result, in part, of such developments as cheap memory, affordable systems of multicore multiple processor configurations, grid computing, tiered storage, more stable flash memory, open systems, and cloud computing. This has given rise to a third generation of database management systems (DBMSs).

The first generation emerged in the early 1960s to serve two distinct purposes: to enable disparate but related applications to share data and to provide a platform for independent data query and reporting that did not require custom program code. These DBMSs were limited because they were locked in to an application or the underlying code.

The second generation was based on attributes of data and their relationships to each other and evolved to the point where SQL became a de facto approach to DBMS development; it was simple and easily comprehensible, despite its limitations. Over time, relational databases were developed for open systems, eliminating the challenges caused by being locked in at the operating system level. More recent relational databases offered clustering support for sharing resources on network-attached storage (NAS) or storage area networks (SANs). Other advances have included caching techniques, 64-bit buffers, automated database and query optimization, multilevel partitioning, and some degree of parallel query processing — all designed to deliver incremental benefits over the basic product.

But these relational databases are still based on the same core design principles that drove their early development. As a consequence, design of databases and laying out data on disk for optimal performance require database administrators to scatter the data across volumes, partition data, and

choose query optimization options that take full advantage of the way the data is organized on storage. The systems are still essentially focused on so-called spinning disks, with limitations in scalability and the bottlenecks represented by the disk I/O. Even though disks are cheaper and faster than ever before, reading and writing records on disk represent a considerable drag on overall system performance.

As a result, third-generation databases have started to emerge, driven by increased use of clustering and virtualization in cloud computing, by a rising tide of open source DBMS usage, and by other new technologies. As a result, IDC sees the following trends affecting the DBMS market:

- Users will look to diversify their DBMS portfolios, applying specialized DBMS technologies to different workloads rather than depending on one product or product line for all data management.
- Smaller, faster, cheaper processors and cheap memory will favor DBMSs that are memory based (data management and layout are optimized for in-memory operations, even if disk is used as expansion space or as a persistence/recovery platform) rather than disk based (data management and layout are optimized for disk; specifically, minimizing I/O wait times).
- Flexible, scalable nonschematic DBMSs and powerful object-oriented DBMSs (OODBMSs) deployed as cloud-based services will serve new workloads based on the collection and searching of data having variable formats from heterogeneous sources, and often, in its raw form, of very great size (hundreds of terabytes).

The Data Explosion and the Rise of Cloud Computing

It's no surprise that enterprises today face an overload of information. New types of data — in the form of social networking, multimedia, machine-to-machine (M2M) communications — and the need to manage multidimensional relationships have put a strain on enterprise databases. Therefore, organizations need advanced relationship analysis, including predictive analytics that take advantage of traditional and nontraditional data relationships.

For example, most analytic applications can provide information only on direct relational correlations, such as who bought which product or which transaction happened when. Enterprises need to deepen their understanding of data by examining context to add information on why events happened as they did and what patterns of events generate what outcomes. Faster business execution can be achieved by automating decision management and by conditioning application behavior with real-time analytics; such analytics may involve relationship analysis and deep mining of streaming data and social media content; such capability is well beyond the scope of conventional relational DBMSs.

Cloud computing is adding to the challenge. Cloud services are based on the principles of utility computing, resource virtualization, and an approach to shared resource management that enables a service to offer users the illusion of limitlessly expandable resources (memory, processing power, etc.). To fit into such a framework, database technology must also offer virtualization, horizontal scalability, and multitenancy.

IDC believes that cloud computing will further two of the following trends in the DBMS market:

- First, companies will use more emerging nonrelational database technologies, many of which are commonly called "NoSQL" and most of which are OSS components, to address data management challenges posed by cloud computing. This demands management tools as well as governance, quality, and security approaches to protect against potential software vulnerabilities and enable change management and compliance. Application life-cycle management adoption for OSS will become essential, as mission-critical systems increasingly contain significant, nontrivial percentages of OSS code.

- Second, software delivery and licensing models will diversify to include more SaaS and appliances. Buyers will expect choices in delivery method and also licensing model. Some buyers will demand the ability to switch between multiple software delivery modes, depending on specific projects or applications.

Relational Database Alternatives: "NoSQL" and the Reemergence of Object-Oriented Databases

While most DBMS products can be used as discrete, bounded entities in a cloud deployment, only a few feature dynamic, virtualized hardware and software resource management for scaling processing power or storage up or down without manual reconfiguration. Relational DBMSs are not designed to discover relationships; they have to be known beforehand. Relational databases are very good for processing data that is already well defined, but they can't be used to discover the structure of data that isn't.

However, content and data are completely separate things. Data must conform to a preconceived schema before it can be collected in a place where it can be queried and processed, and databases have fixed designs that are difficult to adjust and expensive to expand. Cloud computing has changed the parameters that have enforced these restrictions because it delivers dynamic resource allocation, virtualization, and dramatic economies of scale for managing large amounts of data. In response, newer DBMSs require no predefined schema, can accept large amounts of loosely defined data, and can blend structured data with content. IDC calls this new class of DBMS "nonschematic," while others use the term "NoSQL."

Currently, nonschematic DBMSs are mainly used for fairly simple data management problems where flexibility and dynamism are primary attributes. In the future, however, this technology will be used in applications requiring a mix of structured and unstructured data management and large-scale enterprise data indexing and cataloging projects.

In addition, OODBMS technology can address a range of enterprise data management problems that are posed by this new operating paradigm and that cannot be addressed by relational DBMS technology. A relational DBMS can manage data only in flat tables, which may have values that refer to the values of keys in other tables. This makes it good at managing data that occurs in regular, predictable lists. In dealing with social media data involving networks of relationships, and in dealing with online content and a host of other complex data constructs, the relational DBMS is nearly useless. By contrast, an OODBMS can handle nested objects, recursive relationships, collections of data objects having different structures, and data objects of arbitrary length and having nonstandard formats.

It is unlikely that either nonschematic DBMSs or OODBMSs will supplant relational DBMSs. It's more likely that the primary application of nonschematic DBMSs will be in doing jobs that are not being done (or not being done well) by schematic DBMSs, acting as enriching adjuncts to a data management environment that includes, and will continue to include, a schematic database at its core.

Object-Oriented DBMSs Making a Comeback

Because information today is in the form of data, video, audio, graphs, photos, and other nontraditional complex data types, programmers are treating these data types as "objects," with a set of attributes. These objects are treated in the same way by the application and by an object-oriented database — the main benefits being speed, flexibility, and less development effort (since most new applications are coded using object-oriented languages, including Java and C#). OODBMSs are faster than relational DBMSs because data is stored not in relational rows and columns but as objects, which are linked by pointers to establish multiple levels of relationships. In addition, because objects are the same in the application and database, there is no mapping layer to add overhead.

One criticism of object-oriented databases is that data is more hidden and available only through a published set of interface methods, unlike relational database technology. But this supports IDC's research that shows enterprises starting to use more than one type of DBMS. Traditional relational databases and analysis provide declarative and attribute-based information, but object-oriented databases provide more of a behavioral analysis of information. This behavioral analysis is just as important, if not more important, in today's increasingly social media–driven marketplace and also can lead to rapid feedback to operational applications, enabling "smart" business systems.

There is a widespread perception that OODBMS is "old" or even "failed" technology because it grew to a certain point and stagnated. This happened, in part, because the technology was oversold, targeting business data management problems that RDBMS technology could deal with quite well. It also happened because outside of the engineering, scientific, and real-time applications realm, most business computing problems did not require very sophisticated object-oriented data management. These conditions have changed, however. In the new world of big data and cloud computing, the requirements for data management have far outstripped the capabilities of RDBMSs in many areas but are well aligned with the capabilities of OODBMSs.

Further, object-oriented database providers are addressing past market concerns. Updated and new object-based offerings can be linked to industry-standard connectivity, reporting and OLAP tools, and backup and recovery. Some object-oriented databases also fully support SQL. In addition, most cloud computing development is being done in Java, C#, and C++, which are object-oriented languages. OODBMSs can manage the data for such applications transparently, with no need for object-to-relational mapping.

Considering Versant

Redwood City, California–based Versant is a well-established provider of object database management software. The company, listed on NASDAQ (VSNT), merged with POET Software, a Europe-based object database vendor, in 2004. In 2008, Versant acquired object database technology from db4objects Inc. to expand its portfolio in the open source arena. Versant boasts of more than 150,000 installations worldwide in many industries, including telecommunications, defense, simulation, and life sciences. The company's client list includes Alcatel-Lucent, Boeing, EADS, Ericsson, ESA, The Financial Times, Siemens, and Verizon.

The company's flagship product, Versant Object Database, is an enterprise DBMS for developers. Application design and development is accelerated and simplified by integrating caching, mapping, and data management into an "objects end to end" solution. The technology substitutes conventional and CPU-intensive object-to-relational mapping processes with native object storage, providing a platform for higher performance and scalability. This enables, for example, quick navigation in graph-oriented real-time data. Further separating the database from associated applications, Versant also enables true distributed data management for cloud deployments. The Versant technology can integrate thousands of dynamic database nodes across multiple locations and platforms.

The company also offers a set of enterprise database system management tools as part of the Versant Object Database family:

- V/Management Center offers real-time views of performance data and analytical information and provides remote control of a deployed Versant Object Database environment.
- V/Compact is a database reorganization tool for applications that frequently delete large numbers of objects, reclaiming unused disk space to improve performance with no interruption of operations.
- V/FTS provides automatic server failover and error recovery from hardware or software failures.

- V/Async Server supports master-slave and peer-to-peer asynchronous replication between multiple Versant Object Database servers for data recovery or to replicate data between multiple servers to increase performance and reliability.
- V/HA Backup supports online backup of very large data volumes by using disk mirroring features of leading enterprise storage systems without affecting data availability.
- V/SQL provides a SQL interface for data residing in a Versant Object Database.
- dRS supports the synchronization of disconnected mobile db4o databases with a Versant Object Database server.

Challenges

However, Versant does face some challenges. First, there are those who claim that object-oriented databases had their day and failed as a mainstream technology. Versant must continue to stress its technology as ideally suited for highly complex data structures that defy the simple limits of two-dimensional tables as well as for today's distributed and more "smart" or "social" business environments. Data analytics now involves understanding sometimes unpredictable behavior, and object technology is more suited for that.

Second, as the values of object-oriented databases are "relearned," competition in this field will grow, as companies reemphasize their object offerings. Versant must stress its industry and technology leadership.

Finally, as the economy continues its slow recovery, many enterprises are still hesitant to dramatically change the way they do business. As a result, Versant must take a "right tool for the job" approach, demonstrating that its object-oriented database products are not replacement technology but are for enterprise tasks that are not and cannot be supported by traditional relational databases.

Conclusion

There is intense pressure in business today for better database management. The desire to acquire and use business analytics in moment-by-moment decision making, accumulate and reconcile large amounts of enterprise data for reporting and analysis, and streamline operations with better and more precise data collection and movement results in demand for DBMS technologies that deliver orders of magnitude better performance, with much higher scalability, than the leading conventional products. With dynamic provisioning, flexibility, and virtualization capabilities required for cloud computing, DBMSs must take fundamentally different approaches to data definition, storage, manipulation, and delivery.

Enterprises must think more creatively about managing complex data. This will involve not only existing traditional relational databases for analyzing information linearly but also nonschematic DBMSs used for storage and analysis of data and information that cannot be predicted and OODBMSs for data with structures that defy the convention of rows and columns. As a result, enterprises are no longer insisting that one DBMS handle all workloads. Requirements for data analysis will vary and include informal data collections for ad hoc analysis, complex data collections of ongoing analysis, data sharing for clusters of application servers, very large databases, and extreme transaction throughput.

The growth of cloud computing will increase demands for limitless data space and much more complex collections of data in near real time. Today, data that needs to be analyzed is dynamically gathered, and it is difficult to build a schema or map out storage space before data is loaded, creating a demand for nonschematic ("NoSQL") databases. In addition, new kinds of data relationships introduced by Web site constructs, social media, and real-time analytical applications will not be supported by conventional DBMSs but can be handled well by OODBMSs.

OODBMSs are making a comeback because they are designed for dynamic gathering and analysis of data. These databases are ideally suited to help determine behavioral trends that are rarely linear. Market segments ideally suited for this type of database include financial services (including financial management, banking, and insurance), healthcare, pharmaceuticals, telecommunications, logistics, government (especially law enforcement including homeland security), discrete manufacturing, and retail (both online and "brick and mortar").

OODBMSs will not replace relational databases but will become an adjunct tool for the capture and analysis of critical information based on relationships that cannot be predicted. To better compete, enterprises must understand both the linear and the social, and therefore they need a more varied set of tools. To the extent that Versant can meet its challenges, the company has significant opportunity in the enterprise market.
