ALGEBRAIX DATA

TECHNOLOGY OVERVIEW

INTRODUCTION

The 21st-Century Data-Management Challenge

In just the last three years, there has been an eight-fold increase in the amount of digital information generated worldwide, and the end of this information explosion is nowhere in sight. Competitive pressures are compounding the problem as information systems executives struggle to satisfy escalating demand for "operational business intelligence" -- i.e., information needed to make better business decisions based on real-time data feeds. Yet, conventional data-management systems can't cost-effectively access and integrate 85% or more of enterprise data due to the specialized programming required to support disparate data formats. Inevitably, extremely rapid data growth rates, escalating demand for real-time information and the proliferation of new data formats will force a ten-fold increase in enterprise data-management costs unless more cost-effective means are found to manage enterprise data in multiple disparate data formats.

The ALGEBRAIX Solution

The ALGEBRAIX™ solution to these problems is a revolutionary new generation of data-management products that can accommodate databases of unprecedented size, real-time data feeds and disparate data formats, while offering the following additional benefits relative to conventional data-management systems.

- Full backward compatibility with existing industry standards
- Ten times better cost-performance
- Dramatically greater ease of installation, integration, operation and use
- Greater distributability, flexibility, functionality, modularity, parallelism, performance, portability, reliability and scalability

Algebraix Data products accomplish these feats by exploiting a new and uniquely-powerful patented ALGEBRAIX™ data-management technology based on "extended set processing" or XSP. It is a direct consequence of the power and uniqueness of ALGEBRAIX technology that Algebraix Data products are the only data-management products which can provide real-time access to distributed enterprise data in any format without extensive programming yet can be installed and configured in minutes on most hardware platforms.

WHY NEW TECHNOLOGY IS NEEDED

Old Limitations

The vast majority of enterprise database management systems in use today are based on the Relational Data Model and the associated SQL query language developed by IBM nearly 40 years ago. A serious limitation of the Relational Data Model, and the conventional relational database management systems

(RDBMSs) based on it, is that mappings of logical data models to physical data models, and of physical data models to machine operations and storage, must be predefined by applications programmers and database administrators (DBAs). Consequently, conventional relational databases are inherently application-specific as their designs and implementations are dependent on the specialized schemas, table structures, table relationships and indexing schemes required to support the data relationships and operations required for particular applications.

Conventional RDBMSs have performed reasonably well for decades in online transaction-processing (OLTP) applications requiring access to relatively small amounts of data using relatively-homogeneous data structures and queries. However, over the last two decades, whole new classes of enterprise data-management applications have emerged. In particular, business-intelligence and data-mining applications are now driving much of the demand for new enterprise data-management systems. These applications -- e.g., corporate governance and regulatory compliance, customer-acquisition analytics, customer-relationship management, enterprise resource planning, fraud detection, market analytics, operational analytics, revenue assurance, risk management, supply-chain management, etc. -- span all departments within the enterprise and require access to relatively large amounts of data using widely-diverse data structures and queries.

Increasingly, modern information processing systems employ large databases containing data in multiple formats and supporting multiple applications with widely-varying query requirements. Database queries in such systems frequently require multiple inner and outer join operations, both to satisfy query conditions and to aggregate the requested data, and these operations typically result in unacceptably-long query-response times. Unfortunately, when databases serve both OLTP and business-intelligence applications, they grow in complexity and size, and data structures and queries become increasingly diverse. Inevitably, indexes become more complex and costly in terms of memory, storage and processing resources, and their use significantly slows the ability to ingest data updates. Moreover, as increasing amounts of data are moved across the volatile-memory/persistent-storage performance barrier in such systems, query-response times become exponentially longer.

Attempts have been made to replicate data sets to improve query performance in such systems through parallel processing by replicating the supporting hardware (database servers, storage systems, I/O systems, networks, etc.) and ad hoc partitioning of the databases. While this approach can improve query response, it produces proportionate increases in system complexity and data-management costs. In addition, static manual hardware and system partitioning has proven to be problematical in the presence of changing applications and workloads. Clearly, an alternative approach is needed -- e.g., one which supports adaptive, scalable and self-optimizing databases that can dynamically accommodate changes in system resource availability and requirements.

Additionally, the Relational Data Model requires data to be prestructured into tabular arrays of row and columns -- i.e., relational data tables -- for processing and, consequently, conventional RDBMSs have difficulty accommodating new data formats. This limitation and the application-specific nature of conventional relational databases limit their usefulness in high-ingest enterprise-wide

business-intelligence and data-mining applications requiring access to large amounts of data in a multiplicity of different data formats. Again, an alternative approach is needed -- e.g., one which supports a multiplicity of different data formats and does not require data to be prestructured into relational data tables.

New Requirements

A survey of the perceived shortcomings of conventional RDBMSs reveals that there are many and that they fall into the following general categories.

- Inadequate data models
- Inadequate data integration
- Inadequate ease of installation, operation and use
- Inadequate flexibility
- Inadequate performance
- Inadequate scalability

While most of these shortcomings apply equally well to object, relational and hybrid object-relational data-management systems, this discussion is limited to relational data-management systems as they comprise the vast majority of enterprise data-management systems in use today.

Recognizing that the Relational Data Model was developed nearly 40 years ago -- at a time when no one could have anticipated that computer systems would evolve from the single-application single-user offline batch-processing systems of that time to the multiapplication multiuser online network-computing systems of today, it should surprise no one that conventional RDBMSs leave much to be desired in this new computing environment. Indeed, the proliferation of disparate data and document formats spawned by personal computers and the Internet has completely changed the focus of enterprise information systems from the relatively-static and relatively-structured business accounting data of the 1970s to the relatively-dynamic and relatively-unstructured business documents of today.

The requirements for today's enterprise data-management systems transcend anything that could have been contemplated when the Relational Data Model was conceived and, while it has been the mainstay of enterprise data management for more than 30 years, the Relational Data Model falls far short of what is now needed to cope with the extremely rapid data growth rates, escalating demand for real-time information and the proliferation of new data formats driving 21st-century information systems.

Clearly then, 21st-century information systems require something more than the Relational Data Model which forces all data to be prestructured into relational data tables for processing. Indeed, the frustration of attempting to force XML documents into relational data tables proved so great that IBM spent seven years and hundreds of millions of dollars to develop a second database for managing XML documents separately from relational data. Unfortunately, that brute-force solution did nothing to mitigate the costly and time-consuming programming

efforts required to convert relational data tables to XML documents and XML documents to relational data tables as required by such systems.

ALGEBRAIX TECHNOLOGY

ALGEBRAIX technology has a rigorous mathematical foundation and is a fundamental departure from conventional data-management and document-management technologies. The mathematical foundations for ALGEBRAIX technology are a universal data model (UDM) and a proprietary XSP algebra drawn from classical set theory. The UDM enables the modeling of both the contents and structures of internal and external information, and the XSP algebra supports XSP operations that access and manipulate data directly in secondary storage. Unlike the Relational Data Model, the UDM supports mathematical mappings of logical data models to physical data models, and physical data models to machine operations and storage. Taken together, the UDM and the XSP algebra enable all data manipulations to be performed via XSP operations in a computing universe for which a complete and rigorous mathematical model is maintained at all times.

The many functional and cost-performance advantages of Algebraix Data products flow directly from the key innovations of ALGEBRAIX technology. In addition to the UDM and the XSP algebra, key innovations include the following.

- Temporal Invariance
- Adaptive Data Restructuring
- Algebraic Integrity

Temporal Invariance

Temporal Invariance is a characteristic of XSP in which the value, structure and location of information are held constant. Temporal Invariance eliminates the need for “insert,” “update” and “delete” functions and provides the ability to examine the information contained in the XSP data universe as it existed at any time in its recorded history. Given the finite size of available storage, the XSP data universe must be moved forward in time, in a process called garbage collection, when storage limits are reached. This results in a data universe that slides forward in time, maintaining data immutability.

Adaptive Data Restructuring

Adaptive Data Restructuring (ADR) is the process of adaptively altering the logical and physical structures of information while at all times maintaining rigorous mathematical mappings between them. As XSP operations are independent of the structures on which they operate, Algebraix Data products make frequent use of ADR to minimize data transfers across the volatile-memory/persistent-storage performance barrier. Equally important, it is the information-structure independence of XSP operations that facilitates integration of information drawn from disparate native structures -- e.g., relational data tables and XML documents -- without programming.

Another consequence of the information-structure independence of XSP operations is that data-access intelligence is built into sequences of XSP operations instead of contrived data structures and pointer mechanisms. Thus, Algebraix Data products have no need for rigid prestructuring of data and the storage-consuming contrived data structures, complex indexes, B-trees and other ad hoc pointer mechanisms required by conventional data-management and document-management systems. By completely eliminating the contrived data structures and pointer mechanisms required by conventional RDBMSs, Algebraix Data products achieve dramatic reductions in storage requirements and processing time while facilitating equally dramatic reductions in applications programming and database administration costs.

Algebraic Integrity

Algebraic Integrity is the process of maintaining at all times a mathematically rigorous algebraic data model of the information contained in the XSP computing universe and its interrelationships. Algebraic Integrity enables manipulation of algebraic relations to be substituted for manipulation of the data referenced by the relations. Similarly, it enables substitutions of mathematically equivalent but more-efficient sequences of XSP operations for less-efficient sequences of XSP operations. By performing only mathematically optimized sequences of XSP operations, Algebraix Data products eliminate unnecessary data manipulations and minimize data transfers across the volatile-memory/persistent-storage performance barrier. These algebraic optimizations yield substantial reductions in query-response time relative to conventional data-management and document-management systems.

Because the Algebraix Data UDM subsumes the Relational Data Model as a trivial subset, it provides many unique and powerful data-management and document-management functions that are not possible in conventional relational data-management systems. The unique capabilities of ALGEBRAIX technology include, but are not limited to: the ability to algebraically manipulate arbitrary data structures, the ability to simultaneously execute extended-set operations across multiple types of data structures, and the ability to generate and manipulate algebraic metadata which enables powerful new data-management and document-management functions and performance optimizations.

The cumulative effect of these technological advances is to shift much of the data-management burden from I/O-intensive operations to computationally intensive operations. This shift effectively exploits the exponentially spiraling increases in the computational power of modern computer systems and vastly increases the parallelism of database operations. This fundamental difference provides even greater advantage when implemented on systems employing coprocessors that can increase parallelism by orders of magnitude.

ALGEBRAIX PRODUCT ARCHITECTURE

As all Algebraix Data products require the same basic mechanisms to manipulate, interpret and evaluate algebraic XSP expressions, they are based on a common architecture. At the heart of each is an algebraic system comprised of a query optimizer, an ADR optimizer and an XSP set processor as depicted in Figure 1.

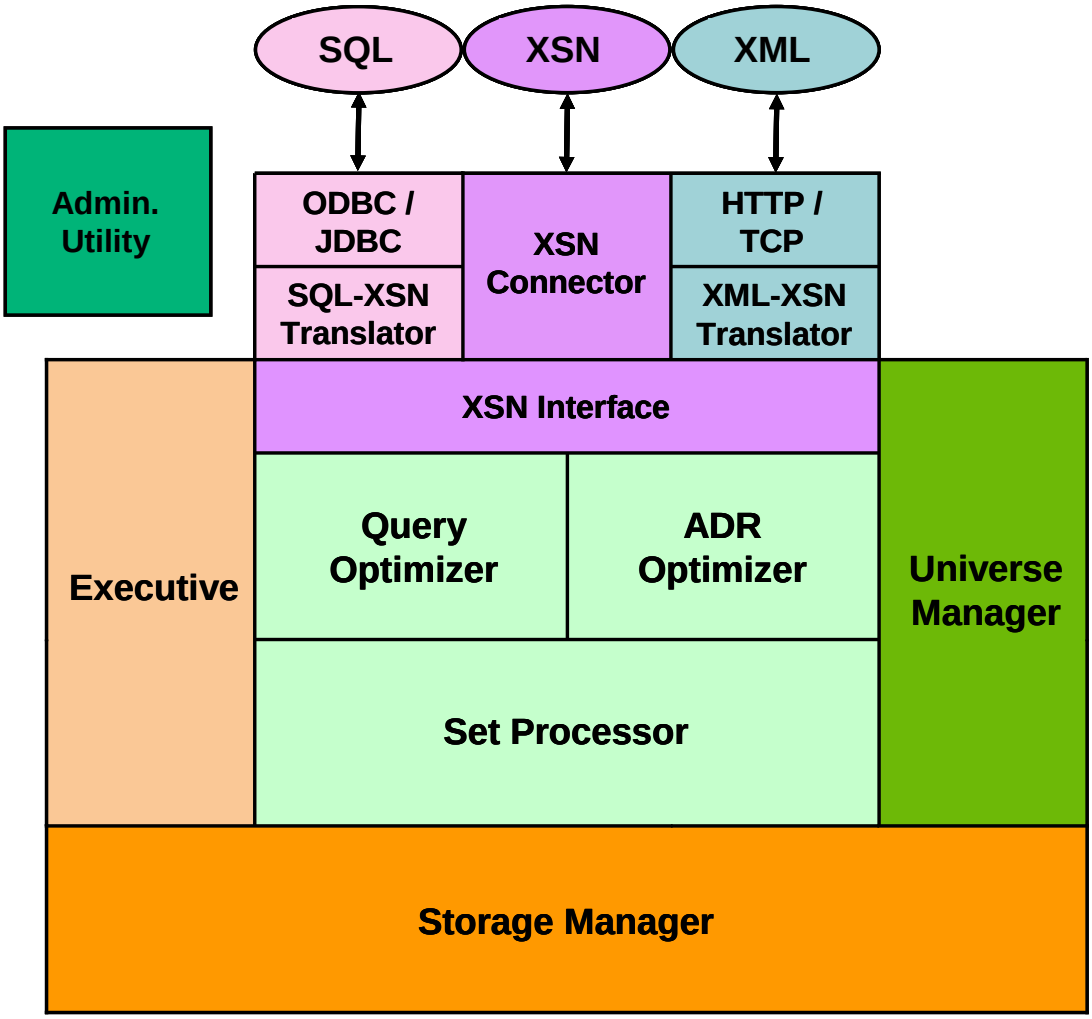

Figure 1: ALGEBRAIX Product Architecture

This algebraic system provides all the functions needed to establish and maintain a rigorous mathematical model of its entire hardware and software universe while performing the algebraic transformations and computations required to analyze, optimize and execute queries made to data in its universe.

ALGEBRAIX DATA PRODUCT ADVANTAGES

Data Models

> ... not all information is able to fit within the confines of the relational data model. ... To attain the sophistication of query optimization that we currently have with SQL, an XML query language's underlying mathematics must be well understood.
>
> Evan Lenz, XYZFind™ Corporation, in "XQuery: Reinventing the Wheel?"

The relational data model is well suited to managing data that can be maintained in tabular arrays of row and columns. However, the vast majority of enterprise data in use today cannot be easily mapped into such relational data tables. Moreover, conventional RDBMSs have no means of accessing information presented in a different data model, such as a directed graph, a nested hierarchy or other ad hoc data structure. For example, conventional RDBMSs access XML data simply as text fields, forcing applications programs to analyze and interpret the internal contents and structures of those fields. The forced mapping of all data into relational data tables results in significant increases in data redundancy and application complexity. In stark contrast, the Algebraix Data UDM provides a rigorous mathematical means to map any logical data model to any other logical data model and to map any logical data model to any physical data model. The mathematical mapping of one data model to another enables simultaneous access to all information through a preferred data model regardless of the original structure of the information. This unprecedented capability enables selection of the data models best suited to the generation, transmission and management of information contained in arbitrary ad hoc data structures.

Data Integration

> Integration dominates the agenda of today's IS departments because it is crucial to most business strategies and their associated application development projects.
>
> Gartner Group, Web Services and Application Integration Conference, May 2003

Inadequate data integration creates a large number of user pain points that are not addressed by conventional RDBMSs. These pain points include the need for application-specific databases, significant manual, time-consuming and error-prone data conversions, increased complexity and data redundancy, and a plethora of other problems, all of which drive up the costs of data integration. Algebraix Data products solve the data-integration problem through the ability to algebraically relate equivalent information in different data structures. This capability enables Algebraix Data products to synthesize and present a variety of different data structures to support a multitude of applications from the same data repository. The creation and maintenance of alternative data models of the same information is also maintained through algebraic structures, ensuring that mappings are complete, consistent and accurate. By enabling multiple data models to be presented from a single repository, data integration is vastly simplified.

Ease of Installation, Operation and Use

Conventional RDBMSs are difficult and time consuming to install, operate and maintain. Individual instances must be tuned to the hardware platforms on which they operate, as well as to the specific data structures and queries they process. Such manual tuning requires substantial ongoing efforts by applications programmers and DBAs with extensive specialized training, knowledge and experience. Algebraix Data products do not require any manual tuning. Through application of adaptive data restructuring and algebraic equivalence, Algebraix Data products automatically and continuously improve query response times by dynamically optimizing usage of all available system resources. Rather than requiring complex analyses of queries and data patterns, followed by expensive and time-consuming database restructuring and application changes, Algebraix Data products automatically and continually adapt through normal use to provide optimal query processing and resource management.

> ... many-fold increases in the number of Internet users and devices in the coming years will tax our data manipulation abilities to the limit. The answer? Simplify, simplify, simplify, said Jarvis [Mark Jarvis of Dell]. … "We could breed more people to staff a growing number of IT departments … Or we can simplify IT technology so it's easier to use and doesn't cost as much."
>
> As reported in "AMD'S Ruiz: Innovate or Die," by Damon Poeter, CMP Channel, 12 November 2007

Flexibility

As new business processes are created and old ones are modified to respond to changing environments, enterprise data and information systems must evolve to accommodate those changes. By forcing users to convert all information into the relational data model and requiring that all applications use that same data model and semantics to access the same information, conventional RDBMSs are anything but flexible. Enormous investments in database design and data prestructuring by applications programmers and DBAs are required to facilitate access to relational databases. Worse, any modified or new application requiring changes to the contents or semantics of a relational database can force a complete redesign of the database and consequent restructuring of the data. Similarly, any change to the contents or semantics of a relational database can force modifications to existing applications. In stark contrast, the data-structure independence of Algebraix Data products, which results from the algebraic equivalence of information independent of its data model, enables different applications to use different data models to access the same information. This unique capability enables complete, correct, consistent and simultaneous access to enterprise data via multiple data models. Consequently, modified or new applications using different data models and semantics can be integrated with legacy applications "as is" with no need for costly and time-consuming modifications to either, providing unprecedented flexibility and cost savings relative to conventional RDBMSs.

Performance

> The enterprise is really looking for something that has one or two magnitudes [decimal orders of magnitude] better performance and is very easy to insert in an IT infrastructure.
>
> Ted Dintersmith, Partner, Charles River Ventures, as reported in "Thinking Inside the Box," by Martin LaMonica, CNET NEWS.COM, 28 March 2005

The performance of conventional RDBMSs is extremely sensitive to data structures, query semantics, operational tuning, hardware configurations and other aspects of the environment. Further, conventional RDBMSs generally do not make use of the full resources of the computers on which they are running. In contrast, Algebraix Data products automatically and continuously adapt to changes in their operating environments. Regardless of the external data model chosen or the particular semantics of a query, Algebraix Data products employ adaptive data restructuring and algebraic optimizations to ensure that the most efficient and highest performing methods are used to process user queries. Algebraix Data products dynamically monitor system usage patterns to determine whether additional data structures can be created to improve query processing and, if so, they automatically generate the additional data structures with no need for costly and time-consuming intervention by applications programmers or DBAs. The result of these dynamic optimizations is a 10:1 cost-performance advantage over conventional RDBMSs.

Scalability

> "The overall goal is to make storage more efficient, reliable, secure and easier to manage in systems with tens or hundreds of petabytes of data spread across tens of thousands of disk drives, possibly used by tens of thousands of clients," says Ethan Miller, a computer science professor at the University of California, Santa Cruz.
>
> As reported in "AMD'S Ruiz: Innovate or Die," by Damon Poeter, CMP Channel, 12 November 2007

As enterprise data volumes continue to grow exponentially, it is essential that the systems used to manage enterprise data scale easily and cost-effectively. This is best achieved if the systems can be scaled homogeneously and incrementally without qualitative modification and in direct proportion to the additional data that must be managed. Unlike conventional RDBMSs, Algebraix Data products scale incrementally relative to both the volume of data under management and the computer resources available for query processing. Algebraix Data products have demonstrated linear scalability over three decimal orders of magnitude without modifications to system architecture or forced changes to data models or query semantics. This unprecedented scalability is achieved through application of the distributive properties of the extended set algebra and the massively parallel processing architecture of Algebraix Data products.◊

Algebraix Data Corporation's patented software is disrupting the entire BI complex. Its **A²DB** advanced analytic database enables real-time data access, eliminates performance tuning, and runs on affordable commodity hardware.

For more information about Algebraix Data Corporation, call 858.200.7215 or visit our Website at www.algebraixdata.com.

4350 Executive Drive, Suite 308
San Diego, CA 92121-2118
858.200.7215
www.algebraixdata.com
2.8.2010