

**ALGEBRAIX
MATHEMATICS**

INTRODUCTION

Most end-user complaints about conventional relational database management systems (RDBMSs) stem from the fact that they lack “strong data independence” -- which requires that users’ logical views of the data, or “schemas,” be completely independent of the data structures employed to manage physical data storage. To achieve strong data independence, database management systems must be capable of unambiguously mapping logical data structures to physical data structures and vice versa. Unfortunately, the classical Relational Data Model (RDM), on which all conventional RDBMSs are based, lacks a rigorous mathematics with which to perform such mappings and falls well short of what is required. For example, the classical RDM joins tables and treats the operation as if it were both commutative and associative. In fact, the classical RDM join operation is neither commutative nor associative, but what is worse is that, mathematically speaking, the classical RDM join operation is not well-defined.

In stark contrast, ALGEBRAIX technology employs a uniquely powerful and mathematically rigorous extended set algebra (ESA) which enables Algebraix Data products to achieve strong data independence. Moreover, by subsuming the functions of the classical RDM, ALGEBRAIX technology enables full backward compatibility to be achieved with conventional RDBMSs while enabling many unique and powerful new features, functions and benefits. The consequences of this fundamental breakthrough are far-reaching and profound.

Extended Set Algebra

The underpinning of ALGEBRAIX technology is an ESA composed of three related algebras denoted by **Exts(U)**, **Clans(U)** and **f(Clans(U))** whose underlying objects are sets of couplets, sets of sets of couplets and functions defined on sets of sets of couplets, respectively.

A couplet **d.v** is the set **{{d},{d,v}}** and models a datum as a coupling of a domain and a value. The set of all couplets that can be formed from a nonempty set **U** is **U x U = {d.v : d,v ∈ U}**. A subset of **U x U** is referred to as an extended set and models a generalization of the notion of a “record.” A set of extended sets is referred to as a clan and models a generalization of the notion of a “table.”

Exts(U) and **Clans(U)** have the usual set theoretic binary operations. **Exts(U)** also has a binary operation which generalizes the notion of function composition by trading on the fact that an extended set is a collection of couplets. **Clans(U)** also has two proprietary binary operations and there is a natural inclusion morphism from **Exts(U)** to **Clans(U)** which enables us to view **Exts(U)** as a subalgebra of **Clans(U)** when appropriate. The binary operations of **f(Clans(U))** include the lifts of each binary operation of **Clans(U)** and the usual composition of functions.

Among the consequences of this rigorous mathematical model are the facts that its RDM join operation is both commutative and associative, predicates can be invoked algebraically and many elements of **f(Clans(U))** can be viewed as linear operators.

MAJOR TECHNOLOGICAL ADVANCES

Three major technological advances are made possible by ESA, namely: a universal data model, a comprehensive computing universe, and algebraic equivalence.

Universal Data Model

ESA provides a universal means for simultaneously modeling disparate logical and physical structures and for mathematically mapping any one to any other. This confers a universal data model which enables the simultaneous modeling of multiple disparate logical and physical data structures together with a means of unambiguously mapping from any one to any other without programming. As a direct result, ESA enables:

- Strong data independence;
- Extensions to support multiple disparate data models, including flat files (e.g., CSV files), nested hierarchies (e.g., XML documents), relational data tables, directed graphs or any other ad hoc data structure;
- Management of both structured and unstructured data; and
- “Data virtualization” via simultaneous access to multiple disparate logical and physical data structures without programming.

Comprehensive Computing Universe

ESA provides a means for establishing a computing “universe,” comprising all the logical and physical resources within a host system, for which a complete and rigorous mathematical model can be maintained at all times. As a direct result, ESA enables:

- Time-stamped nondestructive recording of all data “insert,” “update” and “delete” functions;
- Establishment of “data immutability” ensuring that all source data enters the computing universe at discrete points in time and is forever unchanged thereafter until discarded; and
- Establishment of “temporal integrity” in which the value, structure and location of all data are held constant and related over time.

Algebraic Equivalence

In a fundamental departure from all prior data-management technologies, ESA provides a means for performing all data manipulations via execution of algebraic expressions -- i.e., sequences of algebraic operations -- and for determining the mathematical equivalence of alternative expressions irrespective of the structures of their operands. As a direct result, ESA enables:

- All data manipulations to be performed via mathematical operations versus ad hoc programmed procedures;

- Algebraic decomposition of complex expressions into mathematically equivalent subexpressions composed of basic operations which maximize parallelism and scalability;
- Continuous dynamic optimizations based on mathematical metadata which ensure that optimal techniques are applied at all times by eliminating unnecessary data manipulations and minimizing data transfers across the volatile-memory/persistent-storage performance barrier;
- Data networks to be distributed across computing networks via algebraic decomposition and recomposition of distributed data processing operations; and
- Simultaneous in situ discovery, exploration and mining of any combination of locally and globally distributed data via data virtualization.

MANY PRODUCT BENEFITS

The many unique and powerful features, functions and benefits of Algebraix Data's **Advanced Analytic Database (A²DB)** are derived directly from ALGEBRAIX technology and would not be possible without it.

Real-Time Data Updates

The unique nondestructive time-stamp data recording facilitated by ALGEBRAIX technology enables **A²DB** to ingest data inputs in real time with a negligible impact on query performance -- effectively minimizing "data latency" or lag between the time data is received or updated and when it can be queried.

Similarly, by time-stamping and permanently recording all data updates and employing mathematical operations to retrieve data at any time in its recorded history, **A²DB** enables instantaneous hot restarts, effortless point-in-time snapshots and time-series analyses.

Dynamic Self-Tuning

The strong data independence and algebraic equivalence made possible by ALGEBRAIX technology enable **A²DB** to adaptively and dynamically optimize database operations by restructuring both logical and physical data structures at runtime through a process called "adaptive data restructuring" -- a patented process in which multiple alternative logical and physical data structures are automatically and dynamically generated to minimize query-response time. **A²DB** exploits adaptive data restructuring and mathematical metadata -- which characterize and record working data sets, the operands and operations most-frequently and/or most-recently used in query processing, and the relationships between them -- to adaptively and dynamically generate, partition, refine, replicate, restructure or otherwise optimize physical data sets and queries based on changing data and query patterns. As a result, **A²DB** is completely freed from the performance optimization limitations of conventional RDBMSs, and its dynamic self-optimizations go well beyond anything possible in such systems.

Accordingly, **A²DB** dynamically monitors system usage patterns in response to changing applications, query patterns and system workloads to determine whether alternative data and query structures can be generated to improve query processing and, if so, **A²DB** adaptively and dynamically generates the appropriate alternative structures. Through such devices as generating multiple alternative logical and physical data structures, generating data aggregations as needed, manipulating algebraic relations as a substitute for manipulating the data referenced by them, and substitution of mathematically equivalent but more-efficient sequences of operations for less-efficient sequences of operations, **A²DB** dramatically improves query performance.

The inability of conventional RDBMSs to adaptively and dynamically optimize database operations, forces database administrators to attempt to optimize them manually via a variety of database design processes such as “denormalization,” generation of predefined data cubes, star schemas or other static intermediary data constructs, or static partitioning and replication of physical data sets. Of course, the static nature of manual optimizations seriously limits the kinds of optimizations that can be performed and necessarily results in databases that are tailored to specific applications and their associated queries. Changes to either may require manual database redesigns and new manual optimizations. By comparison, **A²DB** constantly stays in tune with active user queries -- with no need for database redesigns or manual optimizations.

Commodity Hardware

The dynamic self-tuning made possible by ALGEBRAIX technology enables **A²DB** to minimize data transfers across the volatile-memory/persistent-storage performance barrier as well as scatter/gather operations to and from persistent storage media. Consequently, **A²DB** achieves high-speed data streaming with inexpensive direct-attached “commodity” disk drives -- with no need for expensive disk controllers and large caches of high-speed semiconductor storage.

Similarly, the dynamic self-optimizations made possible by ALGEBRAIX technology enable **A²DB** to eliminate many unnecessary data manipulations, thereby reducing the drain on computing resources and enabling **A²DB** to run on relatively inexpensive commodity processors -- with no need for expensive “mainframe” computers.

Superior Performance and Scalability

The performance of conventional RDBMSs typically falls off geometrically as data volumes increase. 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 have proven to be problematical in the presence of changing applications and workloads.

By comparison, ALGEBRAIX technology enables **A²DB** to employ massively-parallel operations as needed to scale linearly relative to both the volume of data under management and the computer resources available. To date, **A²DB** has demonstrated linear scalability over four decimal orders of magnitude with no modifications to system architecture or forced changes to data models or query semantics.

The performance of conventional RDBMSs is also extremely sensitive to data structures, query semantics, hardware configurations and other aspects of their environment. By comparison, **A²DB** dynamically adapts to changes in its operating environment -- with no need for costly and time-consuming intervention by applications programmers or database administrators. Regardless of the external data model chosen or the particular semantics of a query, **A²DB** employs mathematical optimizations to ensure that the most efficient and highest-performing methods are used to process user queries. Indeed, performance benchmarks confirm that **A²DB** self-optimizations frequently accelerate query processing by decimal orders of magnitude while eliminating the costs and delays associated with manual database restructuring and tuning operations.

Data Virtualization

The forced manual optimizations of conventional RDBMSs frequently result in the same data being accessed via different schemas in different databases. A common result is proliferation of application-specific schemas and application-specific databases or data “silos” containing some, if not all, of the same data in different logical and physical data structures. As conventional RDBMSs lack an effective 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, data federation and integration across such silos requires costly and time-consuming “extract, transform and load” procedures, in which data are restructured to conform to different schemas.

The universal data model and algebraic equivalence made possible by ALGEBRAIX technology enable **A²DB** to algebraically relate the same data across different logical and physical data structures. This unique capability enables **A²DB** to employ multiple schemas and logical data models to access the same physical data and to employ the same schemas and logical data models to access different physical data. Consequently, **A²DB** is “schema agnostic” and can employ the same or different schemas and logical data models to simultaneously access the same or different data in the same or different physical data silos. This “schema agnosticism” greatly simplifies data federation and integration by enabling direct simultaneous cross-silo data access -- without resort to costly and time-consuming extract, transform and load procedures. Moreover, when combined with distributed data networks made possible by ALGEBRAIX technology, this schema agnosticism results in meaningful data virtualization by enabling simultaneous in situ discovery, exploration and mining of any combination of locally and globally distributed data. By eliminating the need to transform or upload and download large volumes of data, data virtualization well may be a prerequisite for the “cloud computing” networks of the future.

Ease of Installation, Operation and Use

Conventional RDBMSs require data to be manually prestructured and optimized prior to the commencement of database operations. Furthermore, conventional RDBMSs require manual checkpoint, rollback and restart procedures to recover from hardware failures. All of these manual processes require substantial ongoing efforts by applications programmers and database administrators with extensive specialized training, knowledge and experience, making conventional RDBMSs difficult and time consuming to install, operate and use.

By comparison, **A²DB** is self-configuring, self-installing and self-tuning. The strong data independence and dynamic adaptive data restructuring made possible by ALGEBRAIX technology enable **A²DB** to adaptively and dynamically optimize usage of all available system resources -- with none of the manual data prestructuring and optimization required by conventional RDBMSs. Moreover, the unique nondestructive time-stamp data recording facilitated by ALGEBRAIX technology enables **A²DB** to achieve instantaneous "hot" restarts at all times -- completely eliminating the manual checkpoint, rollback and restart procedures required by conventional RDBMSs.

By forcing users to convert all data into the relational data model and requiring that all applications use that same data model and semantics to access the same data, conventional RDBMSs are anything but flexible. Enormous investments in database design and data prestructuring by applications programmers and database administrators are required to facilitate access to conventional relational databases. Worse, any modified or new application requiring changes to the contents or semantics of a conventional 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 conventional relational database can force modifications to existing applications.

By comparison, **A²DB** enables different applications to use different data models to access the same or different data. This unique capability enables complete, correct, consistent and simultaneous access to enterprise data via multiple data models. For example, data presented to **A²DB** in CSV (comma separated values) or other textual formats are accessed directly in their native formats with no preprocessing.

SUMMARY

By rejecting the pseudomathematics of conventional RDBMS technology and employing ALGEBRAIX technology instead, **A²DB** gains many unique and powerful features, functions and benefits which, taken together, make **A²DB** vastly superior for database applications requiring advanced analytic analyses of large volumes of rapidly changing data.◇

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
2.9.2010