

A²DB

Advanced Analytic Database

Executive Summary

What Are Analytics?

Analytics are the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. Analytics may drive fully automated decisions or may be input for human decisions. Analytics belong to the larger category of business intelligence or BI. BI includes both data access and reporting, as well as analytics. The questions that analytics are able to answer represent the higher-value end of the spectrum – the decisions that impact the bottom line. Analytics don't by themselves constitute a strategy, but using them to optimize a distinctive business capability most definitely does.

Why Companies that Compete With Analytics Win

Companies that compete with analytics win. Statistical analysis of customer behavior and buying power enables companies to track, evaluate and personalize their offerings in a way that companies with little access to data could not dream of doing. These companies gain access to the kinds of data that will shape the future of tomorrow's decision-making. Companies used to make decisions based on hunches or "gut feeling." Companies like Netflix have built empires based on analytics, and most of the Fortune 100, including Anheuser Busch, Barclays Bank, AstraZeneca, FedEx, Amazon.com and Google all use analytics. At a time when many companies are competing with similar products, high-performance analytics offer the last battleground for differentiation. Organizations that have selected a few distinctive capabilities upon which to base their strategies, and have applied extensive data, statistical and quantitative analyses and fact-based decision-making to support their strategies, emerge as the winners in their categories.

The Problem with Conventional RDBMS Technology

Relational database management systems (RDBMSs) have been the most popular database format since the 1980s, about ten years after E. F. Codd first introduced the concept in his landmark paper "A Relational Model of Data for Large Shared Data Banks." However, the volume and complexity of data in the enterprise has grown to unprecedented size during the past decade, making antiquated row and column-based systems increasingly challenging. Additionally, the "row and column" format for storing and managing data is not well suited for modern analytics.

It wasn't until the early 2000s that analytics became a multibillion dollar IT market and began to earn its place as the enterprise staple it is today. As the "status quo" database methodology, conventional RDBMSs have been adopted for business analytics, a purpose for which they were not originally built and are not well suited.

Information contained in conventional RDBMSs is not inherently "searchable." This creates delays in data availability, poor performance and limited functionality for analytics applications, because of the onerous need to constantly restructure and index source data to make it searchable. As a result, vendors have relied on various techniques and work-arounds to overcome some of the limitations of conventional RDBMSs, including proprietary techniques which require special programming. For example:

- Some, like **Netezza**, **Oracle** and **Teradata**, employ proprietary hardware.
- Others, like **Aster Data**, **Greenplum**, **IBM**, **Microsoft**, **Oracle** and **Vertica**, employ massively-parallel clusters of computers or so-called "column-oriented" RDBMSs in attempts to increase parallelism and performance.

While these techniques represent incremental improvements, they all rely on decades-old conventional RDBMS technology. Moreover, they dramatically increase the cost of deploying RDBMSs due to the expense in hardware, programming and time; often with limited results.

How ALGEBRAIX Technology Works

ALGEBRAIX™ technology employs advanced mathematics to analyze data relationships independently of their data structures. This approach increases query performance up to 300X and eliminates the need for manual tuning operations, while reducing the costs of hardware and support.

The consequent **boost in simplicity, performance and value** represents a breakthrough in information technology not seen since the invention of the relational data model nearly four decades ago.

Data Management Challenges

Today IT organizations are facing exponential growth in data volumes. This presents a challenge to store, manage and make sense of these data as well as an opportunity to gain a competitive advantage -- provided the data can be used effectively. As the volume of data grows exponentially, so do the performance challenges of conventional RDBMSs.

The fundamental drawbacks of conventional RDBMSs are the source of the most daunting data-management challenges enterprises face today. In just a few more years, these problems will become utterly crippling for IT.

Tables: A Full-Time Job

With conventional RDBMSs, an enormous amount of time is spent managing relational data tables -- creating them, loading them, joining them, reading them into memory, scanning them, sorting them, storing them, reorganizing them ...

It is from these tables that data gains its meaning and context in a relational database. Since the tables are created and managed manually, each table inherently produces a corresponding growth in database administrator (DBA) workloads.

The inefficiency, overhead and excessive I/O associated with all of this table housekeeping is the fundamental reason that conventional RDBMSs are failing to keep up with new requirements for analytics.

Conventional RDBMSs Require Repeated Manual Performance Optimizations

The performance of conventional RDBMSs is highly dependent on how relational data tables are structured. Yet, someone must decide in advance how to arrange data into tables.

Consequently, conventional RDBMSs are application-specific. Their design, implementation and use are dependent on the specialized schemas, table structures, table relationships and indexing schemes required to support individual applications.

The manual intervention required to optimize RDBMSs prevents effective optimization over time, as applications and their use change.

Conventional RDBMSs require:

- Data to be manually prestructured, partitioned and loaded into relational data tables prior to the commencement of database operations;
- Data to be manually tuned to the hardware platforms on which they operate, as well as to the specific data structures and queries they process;
- Complex analyses and database redesign in the form of data repartitioning and restructuring as well as applications program changes; and
- Manual checkpoint, rollback and restart procedures to recover from hardware failures.

Conventional RDBMS are:

Unable to Optimize Physical Storage. Since data stored in tables is not mathematically mapped to physical storage, conventional RDBMSs are virtually unable to optimize the way the data is stored on physical disk drives.

Unable to Dynamically Adapt to Changing Requirements. How do you create an environment that performs well when you do not know in advance how the data will be accessed? As computer systems are reconfigured to accommodate growth in transactions and data, predetermined data tables set up at the outset by DBAs become less and less optimal.

Unable to Accommodate Real-Time Data Updates. Increasingly, enterprises want to obtain a competitive advantage by analyzing and reacting to business data in real time. For example:

- Financial institutions need real-time event processing to optimize trading, manage risk and detect fraud.
- Telecommunications companies seek to optimize their rate plans in the light of up-to-the-minute analysis of actual usage of their networks.
- Intelligence and law enforcement agencies are charged with “finding the needle in the haystack” in massive real-time data flows gathered around the clock.

The indexing and loading process for new data in conventional RDBMSs creates a substantial delay between when new data is created and when it becomes available for analysis or security functions.

ALGEBRAIX Technology

ALGEBRAIX technology is disrupting the entire BI complex. Algebraix Data’s Advanced Analytic Database (**A²DB**) enables real-time data access, eliminates manual performance tuning, and runs on affordable commodity hardware.

Performance

Because data is not locked into static table structures, **A²DB** can perform “adaptive data restructuring,” a patented process in which multiple alternative logical and physical data structures are automatically and dynamically generated to reduce query-response time.

A²DB automatically and continually:

- Adapts to changes in the operating environment to dynamically optimize use of available system resources;
- Generates, refines, restructures or otherwise optimizes physical data sets and queries based on recent data and query patterns;
- Compiles and analyzes metadata characterizing data sets, relationships and operations to reduce query time; and
- Partitions and replicates data as required to accelerate query processing; thereby eliminating the need for static manual tuning operations.

A²DB dynamically monitors system usage patterns to determine whether alternative data structures can be created to improve query processing and, if so, they automatically and dynamically generate the alternative data structures.

Additionally, **A²DB** uses mathematical operations to store and retrieve data at any time in its recorded history, achieving instantaneous “hot” restarts at all times and eliminating the manual checkpoint, rollback and restart procedures required by conventional RDBMSs.

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 -- completely eliminating “data latency” or lag between the time data is received or updated and when it can be queried.

Rather than force time-consuming and expensive “extract, transform and load” (ETL) operations, ALGEBRAIX technology accepts data in its native schema and adaptively restructures it as required to improve performance during use. By employing an “it is what it is” approach to working with data, as well as eliminating the initial load and index process, ALGEBRAIX technology makes data immediately available for use while greatly reducing the work required to access the data.

Flexibility

By providing a means to map between different logical and physical data models, ALGEBRAIX technology enables different applications to use different schemas to access the same data or the same schema to access different data.

This greatly simplifies data federation and integration by enabling direct simultaneous cross-silo data access -- without resorting to costly and time-consuming ETL procedures. Also, since no conversion of data is required either as part of the ETL or load process, data can be available immediately for query processing.

Consequently, **A²DB** is “schema agnostic.” This unique capability enables complete, correct, consistent and simultaneous access to enterprise data via multiple data models.

A²DB employs 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.

Use Cases

Although the applications of the **A²DB** are numerous, several use cases are commonly encountered:

- Short-term ad hoc strategic data mining
- Routine enterprise reporting
- Ad hoc tactical business support
- Real-time data warehousing
- Rapid prototyping

By providing a means to quickly access data without the need for complex ETL procedures and database tuning for specific applications, it is now possible to capture data and perform ad hoc tactical or strategic data mining on demand.

By eliminating manual tuning operations and adaptively restructuring the data, reporting systems using **A²DB** will always operate at optimal efficiency. Changes in data patterns and the introduction of new information or alterations to reports that would normally take days or even weeks to perfect database modifications can be handled automatically by **A²DB**.

Data warehouses built on ALGEBRAIX technology also benefit. As the data structures are managed through algebraic equivalence, it is no longer necessary to do periodic rebuilds of “information cubes” or to suffer the delays such artificial constructs entail. ALGEBRAIX technology enables real-time data warehouses by enabling updates to be made at whatever interval suites the enterprise without the need to recalculate preaggregate structures or suffer the delays created by such activities. Data warehouses built on ALGEBRAIX technology are always current and available with exceptional performance.

Conclusions

Conventional RDBMSs have served well in transaction applications for decades. But since conventional RDBMSs do not scale well for massive volumes of data or perform well for analytics, it's time for the industry to take a different approach.

Using advanced mathematics, Algebraix Data has developed a new generation of data-management systems which enable unlimited scalability, improved performance and lower costs. **A²DB** uses mathematics to define the relationships between data instead of using tables. This approach eliminates time-consuming table maintenance and the performance problems related to indexing, importing, and cataloging data that's required to be in a specific format. Additionally it creates a system that can self-optimize with mathematical precision and automatically store all the historical information of data over time.

It's about time we found a compelling alternative to free us from the shackles of tables and Algebraix Data has created just that. ♦

About Algebraix Data

Algebraix Data Corporation's patented software is disruptive to the entire BI complex. Its Advanced Analytic Database (**A²DB**) 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.