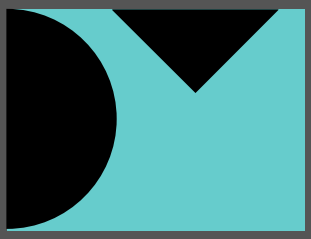


DECISION MANAGEMENT SOLUTIONS

James Taylor
CEO

Sponsored by

TERADATA.

FUZZY LOGIX

Getting Analytics Closer To Your Data

Addressing Analytic Challenges with Teradata and Fuzzy Logix

As organizations push to become more data-driven and use their data to make better decisions, advanced analytics are becoming increasingly important. The evidence that advanced analytics work, that an organization's data holds the secrets to better decisions, is compelling. More advanced analytical techniques are helping organizations address challenges and deliver business value in banking, insurance, media, telecom, retail, manufacturing, and healthcare. Advanced analytics are helping organizations with business problems that range from risk management to fraud detection, from preventing customer churn to micro segmentation and next best action, and from forecasting to simulation.

Organizations adopting advanced analytics in today's big data environment face agility and scalability challenges, even as they strive to apply an ever-wider variety of analytic techniques. While there are several ways to address these challenges, one of the most compelling is to move analytics closer to data using an in-database analytic solution, such as one from Teradata and Fuzzy Logix.

This executive brief discusses key challenges in analytics and reviews the value of a solution from Teradata Corporation and Fuzzy Logix in addressing them.

Challenges in Analytics

The big data environment that most organizations face today offers particular challenges for advanced analytics in terms of agility, scalability, and analytic variety.

Agility: From idea to deployment

Those working on advanced analytics used to have plenty of time to come up with their models: data typically did not change that rapidly, there were only a few models to develop, and problem domains were stable and well-defined. Today's fast-moving environment, however, puts a high premium on analytic agility:

- An organization's data changes faster, so being able to analyze data that is completely current is both harder and more necessary.
- Analytic teams need to be able to quickly integrate new data sources with corporate data to evaluate potential approaches.
- As business problems change and evolve, rapid iteration is critical.
- Broader adoption means more models must be developed rapidly.

An increasingly agile analytic environment is required so that less time is spent moving data around and more is spent conducting analysis. This analysis must be done faster, even as the data being analyzed is increasingly large and varied. Finally, the results must be deployed quickly so that analytics are as current as possible. When it comes to analytics, time is accuracy, and accuracy is money.

Data scalability

An analytic solution must be able to scale up to an organization's data volume. For advanced analytics, data scalability means handling both *breadth* and *depth*: breadth in terms of the number of attributes and depth in terms of the number of records.

Breadth

Today's increasing data volumes are largely the result of additional data sources, both internal and external. Integrating multiple, complex data sources can result in huge numbers of attributes. Any or all of these attributes might be the key to a powerful analytic insight. Developing advanced analytics typically adds large numbers of additional derived attributes, such as the number of times in the last 90 days that a customer has purchased a product. All these attributes, sourced or derived, need to be available when developing advanced analytics.

In a warranty analysis case, over 1,200 variables must be analyzed for hundreds of thousands of vehicles

Depth

There has been a rapid growth in the number of records available for analysis in recent years. More data is being produced, more data is being stored, and more history is readily available. Advanced analytic techniques generally work best with data from a longer period so that, for instance, rare corner cases or seasonal variations can be analyzed. Analytic modelers have historically coped with excessive volume by analyzing only samples of their data. Sampling can be fine, if it is done right, but it's one more relatively complicated step that slows analytics, constrains the number of iterations, and potentially introduces a source of error.

Market basket analysis can involve a lot of data. Take 52 weeks of data on 80,000 products and 20M customers, and the result is 486 billion rows of data that must be analyzed.

Advanced analytic teams increasingly want to work with all their data, which means their data and analytic infrastructure must scale to handle very broad, very deep data sets.

Analytic variety

It is not enough that this scale and agility is delivered only for a few core techniques. As organizations become more experienced with advanced analytics, they generally adopt an increasingly wide variety of techniques for analyzing their data. As different analytic professionals work on different projects, they need different techniques to solve their problems and analyze their data. Different delivery vehicles for analytics, from supporting human decision makers with visualizations to driving automated

decision-making in line of business systems, further increase the range of techniques required. Today's business environment is increasingly complex, and a broad array of analytics is required.

Organizations, then, need a broad spectrum of highly agile analytics that scale to very large data volumes.

Today's Available Options

Organizations adopting advanced analytics have an increasingly wide range of approaches and technologies available:

- Traditional analytic scripting languages have been supplemented by open source R with thousands of unique packages to help address specific problems.
- Modern workflow-based analytic tools deliver increased automation of the analytic process, along with improved analytic model management.
- Modern parallel architectures and multi-core hardware scale to huge volumes, while in-memory and grid solutions provide tremendous performance gains.
- In-database analytic engines allow analytics to be built on data in place, avoiding the need to move data and ensuring that absolutely current data is being used.

In-database analytic solutions, such as that from Teradata and Fuzzy Logix, are particularly well suited for meeting the challenges of scalability and agility.

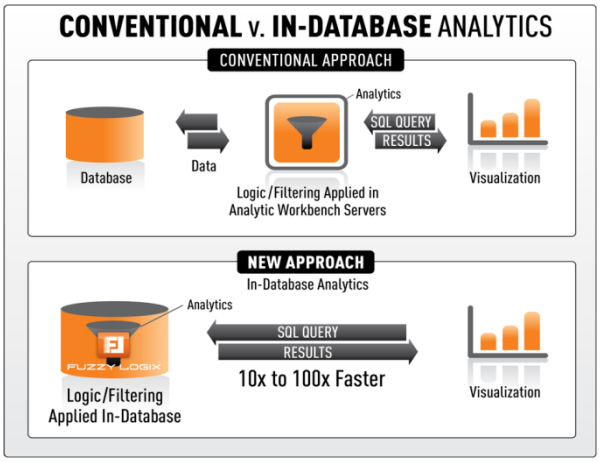
"Without the ability to bring the analytics to the data, companies cannot meet their need for performing deep analytics and delivering the resulting insights to the point in their business where decisions are made"

Scott Gnau, President Teradata Labs

Teradata and Fuzzy Logix

Teradata has partnered with Fuzzy Logix to embed the DB Lytix analytic engine directly into the Teradata database. DB Lytix is a high performance, massively parallel analytic engine. Fuzzy Logix runs in the database so that organizations can move their advanced analytics to their data.

Embedded in the Teradata database, the Fuzzy Logix algorithms can directly access data in a Teradata data warehouse. Running the algorithms on the data, where the data is stored, eliminates data movement, allowing for more rapid, flexible analysis of wide and deep data. The rapid iteration this allows is



extended with SQL-based deployment of analytic models so that analytics can be developed, refined, deployed, and used rapidly and effectively across the organization, allowing near real-time analytic processing of business events.

This speed can be combined with in-database "sandboxes"—Teradata Data Lab—so that analytic teams can evaluate multiple algorithms against multiple data sources with maximum agility. Teradata Data Lab allows analysts to self-provision space in the database to move "dirty" or "untested" data into their data lab and join with production data without replication.

Fuzzy Logix supports very large numbers of independent variables thanks to its ability to "pivot" data. Wide data sets are pivoted into an internal columnar representation that can be processed by the Fuzzy Logix algorithms. Data can be transformed as it is pivoted—turning categorical variables to numeric variables, eliminating over-correlated variables, or filtering records with a WHERE clause, for instance—and columns can be combined from multiple Teradata tables, even though the total number of columns exceed Teradata's maximum for a single table.

Fuzzy Logix algorithms can be run against extremely wide data sets as a result.

Because Fuzzy Logix runs in-database, it has direct data access. Embedded in the Teradata infrastructure, the Fuzzy Logix algorithms can take advantage of the massively parallel, high performance infrastructure that Teradata customers are already using. Even when millions of rows are being processed, this scalability eliminates the need to sample data or limit the scope of analytic projects.

In a pharmaceutical development example, this scalability allowed a 1.6 million row analytic process that had never previously run to completion to finish in just five minutes, while the duration of a second analytic process went from five hours to just three minutes.

Fuzzy Logix makes over 600 algorithms, based on over a decade of development, available. These algorithms cover a broad array of domains, from mathematical and statistical functions to probability, data mining, and specialist financial algorithms. All the algorithms are certified on the Teradata Integrated Big Data Platform, the Teradata Data Warehouse appliance, and the Teradata Active Enterprise Data Warehouse, with support for the Teradata Aster Big Analytics Appliance planned.

Fuzzy Logix is SQL-based so that organizations can use a variety of analytic tools and still take advantage of in-database performance and scalability. Analytic scripting languages, modern workbenches, and visualization, reporting, and dashboard tools can access the Fuzzy Logix functions under the covers using simple SQL statements. Using SQL against both the warehouse and the analytic algorithms eliminates the need for intermediate servers and data movement.

Contact Us

Email : info@decisionmanagementsolutions.com

Web : www.decisionmanagementsolutions.com

Teradata Catalog: **EB8385**