Teradata and Revolution Analytics:
For the Big Data Era, An Analytics Revolution

Enterprise-Class R for Massively Parallel Analytics in Teradata Database

TERADATA.

Table of Contents

Suddenly, More is Possible

Companies in many industries have invested in R, because they have understood its many advantages. Whenever they had a business problem they needed to solve, however, they would have to choose an algorithm and run it serially across individual workstations. While workable, this is a flawed solution, especially as data volumes and analytical complexity expand. Because it cannot deliver results as quickly as a truly parallel approach, and because the results are less consistent, this serial solution is also less reliable.

By contrast, a truly parallel platform takes full advantage of the strengths of a massively parallel database system. RRE is the first R solution to achieve this goal.

Abstract

Despite R being one of the most powerful and widely used types of data analysis software in the world, until now there has not been an out-of-the box R solution that runs in parallel and in database, thus limiting the analytic value of R for driving competitive advantage within large massively parallel processing (MPP) platforms. Now, a partnership between Teradata and Revolution Analytics brings enterprise-class R in database for massively parallel analytical processing inside the Teradata platform. Known as Revolution R Enterprise™ (RRE), this enterprise-class R enables companies that use R to uncover trustworthy insights from huge data assets in near real time.

Introduction

Few industry experts question the power of in-database analytics to deliver the kinds of insights that are relevant enough to alter a competitive landscape. The timeliness, analytical power, scalability, and cost-effectiveness of running analytics in parallel and in database are the keys to repeatedly finding so-called golden needles in massive haystacks and quickly operationalizing those insights.

Yet until now, massively parallel analytical processing in database has only been available using SQL. This has hampered the ability of organizations invested in non-SQL analytics to glean the most powerful competitive advantage possible from their investments in this era of big data.

It has been especially frustrating in the case of R, which is among the most widely used types of data analysis software in the world. Free and open source, R's vibrant community of approximately 2 million users and contributors has created a huge library of algorithms for data access, manipulation, analysis and visualization. This, along with R's proven power and reliability, has made it an attractive option for any enterprise that understands why big data exploration and predictive analytics are essential components in modern business success.

Revolution Analytics and Teradata recognized this and have created an enterprise-class R, known as RRE, which runs in database for massively parallel analytical processing inside the Teradata platform.

TERADATA

How Revolution R Enterprise and Teradata Work Together

Success with predictive analytics and modeling, machine learning, and big data exploration all depend on user-friendly technology that gives staff the tools and performance they need to succeed.

By bringing parallelized R algorithms and a platform to host them in-database, not only is analytics performance greatly enhanced, but latencies previously required to extract data and deliver it to other analytical platforms is eliminated. The result can amount to multiple orders of magnitude of performance improvement.

As noted, RRE supplies an R execution platform and parallelized analytical algorithms that run in-database, in parallel, in Teradata. Programmers use the Revolution Analytics integrated development environment (IDE) on a workstation or laptop to build R scripts, and then transparently execute them in Teradata Database from their workstations or laptops. Revolution R Enterprise runs standard R, preserving portability and enabling new users of R on Teradata Database to be immediately productive while capitalizing on the scale of analytics made possible by running R in-database in Teradata Database.

In addition, RRE for Teradata includes a library of Parallel External Memory Algorithms (PEMAs) that help achieve the highest possible performance. PEMAs are pre-built, extended-memory, parallelized versions of the most common statistical and predictive analytics algorithms and run directly in parallel on the Teradata nodes. They:

- Make Teradata's performance and capacity accessible to all R programmers, by leveraging the Teradata Database's memory, disk, and processors transparently, requiring no change to the R program that called the library.

- Enable R users to tackle larger data analytics problems by combining memory-based and disk-based storage to eliminate open source R's memory limits.

- Accelerate model development and execution by parallelizing execution to spread computation across all nodes in Teradata.

Users of RRE also have access to a huge array of contributed open-source R algorithms made available through the Comprehensive R Archive Network (CRAN) repository. Developers can include these algorithms in their R scripts and run them on multiple Teradata nodes.

Moreover, installation and administration only require standard Teradata tools. Teradata workload manager provides load balancing and task management, thus ensuring no overloading of the system. Furthermore, the Teradata Data Lab provides analytic sand boxes that enable developers to experiment without having to worry about the effect on day-to-day operational use. Access to RRE is authorized and authenticated using standard security capabilities in the Teradata Database, thereby assuring information protection while minimizing disruption for system administrators.

Finally, the Teradata and Revolution partnership complements other analytical applications by providing a lower cost solution for exploratory analytics, predictive modeling, and data scoring that ensures scalability, and broadly deploys analytical results via business intelligence tools, visualization tools, and custom applications to users throughout an organization.

Why This Matters

The partnership and implementation of RRE for Teradata offers four essential benefits to a business:

Faster Results
Operating in parallel, in-database analytics eliminate the need to move data to a middle tier for analysis, thus dispensing with more than half of the typical analytics cycle. Furthermore, in-database operation brings computational parallelism, which dramatically accelerates the delivery of results from organizational data.

Enhanced Analytical Power
The computational performance of RRE scales linearly with system size, enabling developers to run more—and more sophisticated—analytic models on larger sets of a company's data. This facilitates deeper and more reliable data exploration, faster model execution, and finer segmentation, which result in considerably faster and more

TERADATA

predictive analytics. Perhaps most importantly, transparent in-database execution reduces dependence on IT staff to move data, freeing developers to pursue more forward-thinking projects, including developing applications that enhance big data discovery.

Reduced Costs and Risks

Based on open source R, RRE typically costs less than half as much as other solutions. Moreover, accelerating analytics with in-database processing also reduces the costs and labor associated with the need for additional data marts. R is also widely known and taught, broadening the available talent pool, reducing training burdens, shortening ramp-up time, and cutting project cost. An added bonus: commercial support by Revolution Analytics' R experts reduces project risk.

Coherent Solution for Unified Data Architecture™ (UDA)

RRE for Teradata provides data integration with third-party platforms like open source Hadoop and the Teradata Appliance for Hadoop, unifying information and analytics across the organization to increase the value of Teradata's UDA solution.

A Powerful Partnership

Teradata helps companies get more value from data than any other company. Our big data analytic solutions, integrated marketing applications, and team of experts can help your company gain a sustainable competitive advantage with data. This is because the Teradata UDA is a truly integrated analytics solution that unifies multiple technologies into a cohesive and transparent architecture by leveraging the best-of-breed and complementary values of Teradata, Teradata Aster® for big data discovery, and

open source Hadoop. A team of data experts with deep industry experience supports the UDA.

Revolution Analytics delivers advanced analytics software at significantly less cost than existing solutions. It builds upon open source R and innovates in big data analysis, integration, and user experience to meet the demands and requirements of modern data-driven businesses.

The company has taken the R language to unprecedented levels of scale, capacity, and performance, enabling statistical analysis of very large data sets in a fraction of the time of legacy products—all without requiring expensive or specialized hardware. For enterprises that want to deploy their analytics solutions to large user populations, RRE brings structure, control, governance and support to open source R. RRE also features a visual development environment that enables faster, more accurate R programming for developers.

Flexibility is a Business Necessity

As data-driven companies seek to compete and win in the era of big data, they must find cost-effective solutions that don't compromise their chances for success. Increasingly, many companies have begun to understand the potential of R to deliver the analytics they need, but have been waiting for an enterprise type solution with parallel, in-database capabilities.

The Teradata and Revolution Analytics partnership delivers that solution—and more. Users get their enterprise-class R analytics, but also get Teradata Corporation, which offers the finest massively parallel processing platform in the world and a UDA that takes advantage of big data in a way that no other technology can match.

TERADATA.