# Teradata Aster R


DATA WAREHOUSING

## What Could You Accomplish with a Scalable R Solution?

The R statistical software and programming language has already proven to be a powerful solution for implementing analytic business applications, everything from churn and cross-selling to credit risk analysis. With over 6500 packages and cutting edge algorithms contributed by a vibrant user community, it's obvious why R is now the favorite analytic language of data scientist and data miners. But as powerful, flexible and extensible R is, it carries distinct memory, data and processing limitations, especially when it comes to scalability, performance, and enterprise deployment.

But what could you do if these limitations were eliminated? For starters, you could perform analyses on ever larger data sets, dramatically cut the time it takes to run those analyses, allowing you to iterate faster against all your data. In addition, you can easily share data, analytics, and results with users across your enterprise.

That's why Teradata has introduced Teradata® Aster® R, a gateway to scalable R for business analytics, a solution that allows you to lift the limitations of R by leveraging scalable Teradata Aster Analytics and Aster's MPP architecture: your high performance computing platform for R.

Aster R features the following components to enable enterprise scale R:

- **Aster R package** – A library of high performance, prebuilt R functions that hides the complexity of parallel programming.
- **R Engine** – Open source R engine and packages embedded in MPP architecture to run existing R scripts at scale.
- **Aster AppCenter** – A framework to deploy your R model to end users in a web-based interface.

"Using R gives an organization access to a pipeline of research, a huge body of proven algorithms… and a broader analytic ecosystem."

– James Taylor, Chief Executive Officer

## Scalable, Reliable, Powerful

The advantages of using R in-database include performing R analytics at scale by leveraging the high speed scalable processing, integrating all your analytic data and results in a single managed environment to simplify access and collaboration. Aster R brings you all those benefits plus unmatched ease of use.

Because users currently run R analyses on a single server, they're limited to the processing power of that lone server. Instead of working with entire sets of data, they're forced to create sample sets and run their analysis on just those samples. But more and more often today you need to be able to look at *all* your data, beyond just the sample that fits in memory.

Another data reduction method is to partition data. This works effectively only if your data can be neatly partitioned into smaller segments. It does not, however, allow you to analyze data across multiple partitions. To process large volumes of data using R, you must understand and develop parallel programs in R. With the Aster R package, you simply call the prebuilt parallel functions as you would with R. The prebuilt functions will seamlessly parallelize the analytic task and return a single answer.

Aster R solves that scalability challenge three ways. First, it allows R programmers to run functions available in the Aster R package in the MPP platform by leveraging the Aster Analytic Portfolio that includes text, path, pattern, statistics, machine learning and graph analytics. Since these are Aster analytic functions, they will transparently distributes the processing across multiple servers running in parallel so that users can quickly analyze data and get results.

Secondly, the embedded R engine allows users to run their existing R scripts in the database. This is ideal for "embarrassingly parallel" functions such as scoring that processes each row independently. Now you can distribute the analytic task across multiple nodes to accelerate processing.

Third method is to use the Aster R Runners that allows advanced R users to write their own parallel programs in

TERADATA.

R based on the split/apply/combine strategy (similar to the map and reduce constructs). That leads to additional benefits because R users can perform analysis much faster, taking advantage of Aster's high-speed parallel processors. Tasks that previously took hours can be done in minutes.

## We've Leveraged *Your* Language

Aster R lets R users continue to use the R language, tools, and interfaces they're most familiar with. There's no need to learn SQL, new interface or parallel programming. Instead, programmers can continue to build R programs on their R client while the solution pushes the processing into the highly scalable platform for parallel processing for added ease and productivity.

Aster R allows the R user to interact with data within the Aster, Hadoop or Teradata as data frames. Users simplify create a virtual data frame that points to an Aster table or view in Hadoop, Teradata or any other source supported by Teradata QueryGrid. Once you establish a data frame, you can interact with the data as if it's an R object.

Some R-like solutions require you to include a proprietary distribution of R, however those solutions are not fully compatible with open source R. Other solutions require users to wrap their R programs in a SQL statement to leverage in-database processing. With Teradata Aster R, however, those programmers can simply follow similar R language, syntax, and signatures to provide a familiar, complete, and end-to-end environment.

## Why Choose Teradata?

Why should you select Teradata Aster R to complement your R solution? Because as the recognized industry leader in data and analytics, only Teradata brings you and your organization an unparalleled blend of technology, insight, and innovation. Our unique combination of industry knowledge, consulting expertise, scalable multi-genre advanced analytics expertise, world-leading software, and hardware technology offers you everything you need to scale your current R solution and deploy analytics more effectively across your business.

> "Organizations should plan on working with a commercial vendor that has a solid plan for R in terms of providing scalable implementations of the algorithms they care about."
>
> – James Taylor, Chief Executive Officer

## Aster R Built-In Benefits Include:

### Lifts the limitations of R

- Teradata Aster R lifts memory, processing and data limitations by leveraging Aster Analytics and Aster's MPP architecture that enables massively scalable processing of complex analytics across all your data. Analyst can focus on the analytics, instead of sampling, partitioning or coding around limitations.

### Ease-of-use and productivity for R users

- R analyst can continue to use familiar R client and R language to access a library of parallel functions including the rich and powerful Aster Analytics Portfolio. With Aster R, there's no new tool, language or data format to learn.

### Powerful analytics combining Aster and R

- Analyst now have the flexibility to mix and match R, Aster R and SQL functions within a single script for multi-genre analytics to solve complex business problems. Analyst can now incorporate results from a path and sentiment analysis to enhance their R churn model.

### Enterprise Ready Solution

- To improve data security, stop moving the data from a secure environment to unsecured desktops or laptops. By keeping the data with the Aster environment, you not only reduce the risk to your data, but also improve manageability of your data.

Teradata also provides a web-based framework, AppCenter, to easily deploy your R models to your end user so they can derive business value from your R models.

## For More Information

To learn more about the potential of Teradata Aster R and how it can help your enterprise take better advantage of R, contact your Teradata representative or visit **Teradata.com/Teradata-Aster-R**.

10000 Innovation Drive, Dayton, OH 45342 **Teradata.com**



01.16 EB8258

TERADATA.