Teradata Aster R Enterprise

Srujana Vallabhaneni[1,*]

1. Teradata Aster
*Contact author: Srujana.Vallabhaneni@Teradata.com

Keywords: parallel computing, high performance computing, big data, scale

Teradata Aster R Enterprise™ is next generation analytics at speed using R interface. Teradata Aster has enabled the scaled deployment of R analytics within Aster Big Data platform in several ways. First, Aster has enabled R to be executed in parallel on the Aster cluster. Score R models in-database in parallel, which are no longer bound by single node resource limits (memory, CPU).

Secondly, Aster has developed two packages for R clients namely **asterR** and **asterRExt.** These packages allow for multiple out-of-box analytics including SQL, SQL-MapReduce and R analytics using familiar R interface. R is integrated into Teradata Aster's SNAP Framework™ for Discovery (SNAP = Seamless Network Analytics Processing). SNAP Framework enables users to snap together multiple analytic engines and file stores with ease, providing them with unmatched power and speed to delve deeply into data. The SNAP framework includes the Integrated Optimizer and Executor, Unified R/SQL Interface, and Common Storage System and Services. R analysis using multiple analytics (Graph, Text, Path/Pattern, open source R etc.) is automatically optimized and executed with speed across the SQL, SQL-MapReduce, SQL-GR and open source R engines. This will increase productivity of R users through rapid iterations on a single platform that empowers all users – R user, analyst, data scientist. These different users have access to various SQL, SQL-MR, R analytics through an interface of their choice: R or SQL.

Thirdly, scaling R through MapReduce approach: Extended the R language by adding Aster-specific runner constructs that enable R programmers to write their programs in a partial and final aggregation/map-reduce framework ("split-apply-combine") for execution on the R instances installed on the Aster cluster. This scales open source R using the R-MapReduce framework. You can build and execute parallel R programs based on MapReduce approach and rapidly process large data.

The session will included demo of packages in the Teradata Aster R Extensions including database access functions (ta.connect, ta.data.frame), fast data pipe (ta.pull, ta.push), data exploration and manipulation (ta.show, ta.summary) Aster SQL-MapReduce R wrappers - (tadf.nPath, tadf.glm), push down open source R Statistics functions (ta.sd, ta.var, ta.wilcox.test). The session will show Aster specific runner constructs for in-database R execution. Demo will include running open source R engine across nodes on Aster's MPP platform and show how users can push down their R scripts into Aster. Also show how to overcome R's memory scalability problem by MapReduce approach.