

Revolution R Enterprise 6.2 README

Revolution R Enterprise 6.2 for 32-bit and 64-bit Windows and 64-bit Red Hat Enterprise Linux (RHEL 5.x and RHEL 6.x) features an updated release of Open Source R, an updated release of the bundled RevoScaleR 'big data' package, and an updated release of the optional RevoDeployR 'web deployment' package.

Installation instructions are provided in your welcome e-mail.

What's New in Revolution R Enterprise 6.2

Open Source R

R 2.15.3

- Revolution R Enterprise 6.2 provides the latest stable version of open source R: R 2.15.3

RevoScaleR

High Speed Teradata Data Connection

- With the RxTeradata data source, you can seamlessly extract data from a Teradata database using the Teradata Parallel Transporter and write it to a high performance .xdf format file, or simply analyze the data directly.
- A new manual, *RevoScaleR Teradata Data Source Getting Started Guide*, is provided to help you get up to speed quickly.

Stepwise Regression for 'Big Data' Linear Models

- The rxLinMod function contains a new 'variableSelection' argument to allow for stepwise regression.
- The helper function 'rxStepControl' provides a convenient way to control the stepwise regression. See the rxStepControl help file for more information.

Parallel Random Number Generation

- The new functions rxRngNewStream, rxRngGetStream, rxRngSetStream, and rxRngDelStream provide an R interface to the parallel random number generators supplied with the Intel MKL libraries. These allow high quality parallel random numbers to be used in distributed computations performed by RevoScaleR's rxExec function. See the help files and Section 1.6 of the RevoScaleR User's Guide for more information.

Fast Fixed Format Text Data Source

- RevoScaleR now provides a faster way to import or analyze fixed format text data. When using the rxImport function, setting the type to 'fixedFast' will utilize this faster method. Alternatively, when using an RxTextData data source, setting 'useFastRead' to TRUE (the default) will utilize this functionality.

Enhanced Control Over Interaction Terms in Model Fitting

- The rxLinMod, rxGlm, and rxLogit functions have a new 'dropMain' argument. If set to FALSE, the main terms of the interactions will be kept and interactions dropped as needed.

- The rxLinMod, rxGlm, and rxLogit functions contains a new coefLabelStyle argument. If set to “R”, the labels created are compatible with labels used in standard R’s lm function.

By-Group ‘Big Data’ Summary Statistics Written to File

- The rxCube and rxSummary functions now provide options to write by-group counts or summary statistics directly to a high performance .xdf file for further analysis. Users also now have more control over the summary statistics that are reported. See the help files for more information on the ‘outFile’ argument for rxCube and the ‘byGroupOutFile’ and ‘summaryStats’ arguments for rxSummary.

Sort, Merge, and Split Enhancements

- The rxSort and rxMerge functions now by default make better use of available memory; previously, they used a fixed default buffer size which could be unnecessarily small on computers with generous amounts of RAM. To fine tune memory usage for these functions, see the ‘bufferLimit’ argument.
- The rxSplit function now creates .xdf files by default if inData is an .xdf file. A list of data frames can be returned (given sufficient memory) by setting the baseOutFile to “”.

RevoDeployR

MongoDB replacement for MySQL

- The updated RevoDeployR server provides a new high-performance, scalable persistence infrastructure built on top of the MongoDB NoSQL database. It manages the reliable persistence of all user, project and repository data. MySQL is no longer used. A migration utility is provided to move RevoDeployR scripts, repository data, and projects from a RevoDeployR 2.0 server to a 6.2 server.

New API support for repository and script management

- Priority scheduling for asynchronous jobs
- Executing scripts found on external URLs or file paths including scripts in GIT and SVN repositories
- The lifecycle management of repository-managed files, including scripts
- Enhanced file versioning for repository-managed files, including scripts

Apache Web Server no longer required

Updated Javascript, Java, .NET client libraries

Documentation/Examples

- Updated documentation and accompanying sample applications
- A new RevoDeployR Deployment Planning guide is now available to help administrators plan the provisioning of server and grid capacity

ParallelR

NetWorkSpaces Deprecated

- NetWorkSpaces-based packages for parallel computing are deprecated and will be removed from future releases, starting with Revolution R Enterprise 7.0. This includes the following packages:
 - nws
 - nwserver
 - bootNWS
 - randomShrubberyNWS
 - sprngNWS
 - doNWS

SNOW and multicore Packages Deprecated

- The snow and multicore packages have been subsumed into the R base package parallel; the multicore package is now obsolete, per the R core team. The doMC package, which previously used multicore, has been updated to use the parallel package. The multicore package has been removed from Revolution R Enterprise 6.2, snow will be removed from subsequent releases.
- The doSNOW and doMC packages are deprecated; use doParallel instead.

Known Issues

- [Known Issues in Revolution R Enterprise 6.2](#)