

# Revolution R Enterprise 5.0 for Windows README

Revolution R Enterprise 5.0 for 32-bit and 64-bit Windows features an updated release of the RevoScaleR package that provides scalable data management and data analysis: the same code scales from data frames to local, high-performance .xdf files to data distributed across a Windows HPC Server cluster. It also provides open source R 2.13.2 and includes an enhanced R Productivity Environment (RPE), a full-featured integrated development environment with visual debugging capabilities. Also available as a separate product is RevoDeployR 2.0, designed to help you deliver R analytics via your corporate intranet or in the cloud.

The latest release of Revolution R Enterprise 5.0 is Revolution R Enterprise 5.0.1, a bug fix release that includes the following changes:

- The Revolution R Enterprise shortcut can now be pinned to your Start menu.
- rxKmeans with missing values was sometimes incorrectly reporting the number of valid and missing observations.
- rxMerge was erroneously returning FALSE if the outfile argument had no .xdf extension, even though the merge was successful. Also, the use of the newVarNames argument sometimes led to the renaming of additional variables.
- rxSetVarInfo now returns an RxXdfData object if the inputdata is an .xdf file. If the input data is a data frame, a data frame is returned.
- rxFactors now returns an RxXdfData object if an output file is specified. Otherwise, a data frame is returned.
- The doRSR package now provides correct results when registered with a distributed compute context with wait=FALSE.
- RxHpcServer now behaves correctly if the names of all the compute nodes in a cluster are alphabetically after "Local Client."

To get started using RevoScaleR with huge (or small) data sets, we recommend that you review the *RevoScaleR Getting Started Guide* ( found on your start-up menu after installation under the Revolution R program group - All Programs/Revolution R/Enterprise 5.0/Documentation). It provides a tutorial introduction to using the RevoScaleR functionality. Distributed computing features are described in the *RevoScaleR Distributed Computing Guide*. The *RevoScaleR User's Guide* is also available in .pdf format, and help files give detailed descriptions of the individual functions.

If you are new to Revolution R Enterprise for Windows, we recommend that you review the *R Productivity Environment Getting Started Guide* before starting to become familiar with the user interface for Windows. (It is also found on your start-up menu after installation under the Revolution R program group). It provides a tutorial introduction to using the Revolution R Enterprise Windows graphical user interface and the visual debugger.

Installation instructions are provided in your confirmation e-mail.

## What's New in Revolution R Enterprise 5.0

### RevoScaleR

#### *Distributed/parallel computing*

- Automatically distribute statistical analyses from your desktop across nodes of a cluster [Currently supported for Windows HPC Server]. Analyses include summary statistics, crosstabs, linear regression, logistic regression, covariance matrix computations for factor analysis and principal components, and k-means clustering. Binning computations for histograms are also distributed.
- Distribute R function calls, including data manipulation functions, across nodes. Easily distribute “embarassingly parallel” computations across nodes or cores of a [Microsoft HPC] cluster, or the cores of your desktop or laptop using the new *rxExec* function.
- Compute in parallel with *foreach* using RevoScaleR using the new doRSR backend.
- New *RevoScaleR Distributed Computing Guide* (choose Help/R Manuals (PDF) in the RPE).

#### *Scalable data management*

- Data import:
  - New versatile *rxImport* function for using external data with R (delimited and fixed-format text, SAS, SPSS, or ODBC). Bring smaller data sets directly into an R data frame; store larger data sets in the native .xdf file format, very efficient for storing and accessing large data sets. The *rxImport* function returns a data frame or an RxXdfData object representing the created .xdf file. Either can be used in subsequent data analysis functions.
  - Two alternative modes of Delimited Text import, and two alternative modes of ODBC import – one supported on Linux
  - Ability to keep or drop variables on import
  - Ability to specify start row and number of rows of data to import
- Data Cleaning and Manipulation
  - New versatile *rxDataStep* function allows you to perform data transformations on big data using the power and flexibility of the R language. Experiment with a small data frame, then apply the same code to a huge data set.
    - Returns data frame or RxXdfData object representing an .xdf file that can be used in subsequent scalable analyses.
    - Works with data frames or .xdf files (as input data or output), making it easy to convert from one type to another.
    - Ability to “re-block” xdf files with a user-specified number of rows.
    - Improved evaluation environments for user-defined transforms and transform functions, and new internal variable, *.rxNumRows* (containing the number of rows in the current block) for use within transformations.
  - Big data merge with the new *rxMerge* function. Merge two large data files, or merge a smaller in-memory data set into a large data file.

- Improved performance for big data sort. New general *rxSort* function to work on data frames or .xdf file
- Ability to create and recode factors in .xdf files and data frames using new *rxFactors* function
- Split an .xdf file into multiple files by number of rows, blocks, or levels of a factor variable using new *rxSplitXdf* function.
- Support for additional data types in .xdf files: ordered factors and POSIXct, and improved support for Date data type.
- New functions *rxGetVarInfo*, *rxGetInfo*, and *rxSetVarInfo* work for both data frames and .xdf files
- New examples in the *RevoScaleR User's Guide* for big data data step and import.

### **Expanded scalable statistical functionality**

- New functions utilizing output from *rxCrossTabs* objects:
  - *rxChiSquaredTest*: Chi-squared Test
  - *rxFisherTest*: Fisher's Exact Test
  - *rxKendallCor*: Kendall's Tau Rank Correlation Coefficient
  - *rxPairwiseCrossTab*: Apply a function 'FUN' to all pairwise combinations of the rows and columns of an *xtabs* object, stratifying by higher dimensions
  - *rxRiskRatio*: Calculate the relative risk ratio on a two-by-two table
  - *rxOddsRatio*: Calculate the relative odds ratio on a two-by-two table
  - *rxMultiTest*: Collects a list of tests for variable independence into a table.
- Also a new *rxResultsDF* method for *rxCrossTabs*, *rxSummary*, and *rxLinMod* for extracting a data frame from results objects
- Improved performance for scalable analysis functions operating on data frames.
- Option in *rxPredict* and *rxKmeans* to write out model variables in addition to predictions/cluster number.
- Option in *rxSummary* to remove missing values by term.
- Option in *rxLinMod* and *rxLogit* to drop first or last factor levels, and ability to set starting parameter values in *rxLogit*.
- *rxHistogram* now supports logical data and frequency weights with continuous data, and has *transforms* and related arguments.
- Prediction standard errors and confidence intervals can now be calculated using *rxPredict* with models fit using *rxLinMod* and *rxLogit*.
- New examples in the *RevoScaleR User's Guide* for factor analysis and principal components analysis.

### **Enhanced RPE**

- Support for creating and building R packages
  - R Package Project type in RPE Solution Explorer to create the directory structure for a new R package.

- Create an Rd Help file template for a user-created function from the Solution Explorer by adding a new item and specifying the function name.
- Build an R package from the Solution Explorer.
- Support for Windows HPC Server
  - Access the HPC job scheduler directly from the Windows R Productivity Environment (RPE).
  - View the status of pending jobs in the RPE Object Browser.
  - Code snippets for distributed computing with HPC Server
- Option to load last-loaded solution on startup
- New projects now starts by default in release mode instead of debug mode

## **Update to Open Source R 2.13.2**

### **Known Issues**

- [\*Known Issues in Revolution R Enterprise 5.0\*](#)