

# Revolution R Enterprise 7.4 README

Revolution R Enterprise 7.4 for 64-bit Windows and Linux (Red Hat Enterprise Linux 5.x and 6.x and SUSE Linux Enterprise Server 11.x) is the fastest, most cost effective, enterprise-class, big data big analytics software available for workstations, servers, and distributed platforms such as Microsoft HPC Server, Hadoop, and Teradata. Supporting a variety of big data statistics, predictive modeling, and machine learning capabilities, Revolution R Enterprise is also 100% R. The 7.4.1 release extends the functionality of R 3.1.3 and includes updated releases of the RevoScaleR package for scalable, fast (multicore), and extensible data analysis with R and the RevoPemaR package for developing your own parallel external memory algorithms in R.

Download and installation instructions are provided in your welcome e-mail.

## What's New in Revolution R Enterprise 7.4

### RevoR

#### *Installer Enhancements*

- Revolution R Enterprise 7.4.1 and Revolution R Connector 7.4.1 now install by default into the Revolution R Open 8.0.3 distribution of R 3.1.3. It is highly recommended that users of Revolution R Enterprise 7.4.0 uninstall 7.4.0 before attempting to install Revolution R Enterprise 7.4.1.

#### **New and Removed Packages**

- The Revobase and RevobaseEnt packages have been removed.
- The new RevoUtils package contains basic utility functions such as 'readNews'.
- The new RevoUtilsMath package contains functions to get and set the MKL threads.
- The new RevoIOQ package contains the RevoIOQ function for running an installation and operational qualification test suite on a single machine.

### RevoScaleR

#### *Teradata Support*

- RevoScaleR version 7.4.1 supports in-database operation using the Teradata Enterprise Data Warehouse, versions 14.10 and 15.00. Server installation documentation has been expanded in the [Revolution R Enterprise 7 Server Installation Guide for Teradata](#); client installation and configuration documentation, which was formerly part of the *RevoScaleR Teradata Getting Started Guide*, can now be found in the new [Revolution R Enterprise 7 Client Installation Guide for Teradata](#).
- In-Teradata memory limits are now set using the Teradata stored procedures SetMasterMemoryLimitMB() and SetWorkerMemoryLimitMB(), rather than using

environment variables as in Revolution R Enterprise 7.3.0. See the [Revolution R Enterprise 7 Server Installation Guide for Teradata](#) for details.

#### ***HPC Server Pack 2012 Support***

- RevoScaleR version 7.4.1 supports High-Performance Computing using clusters managed by HPC Server Pack 2012. (Clusters using HPC Server 2008R2 are also still supported, but are now deprecated.)

#### ***Support for Reading and Writing Folders of CSV Files [Experimental]***

- RevoScaleR version 7.4.1 allows you to import from and export to folders of CSV-format files in both local compute contexts and in the Hadoop Distributed File System (hdfs). This enables ready conversion of data used in Hadoop computations in the distributed xdf format to CSV for sharing with other users.

#### ***Two-process Architecture***

- RevoScaleR version 7.4 now features a two-process architecture—R itself runs in one process, while RevoScaleR's underlying computational engine runs in a separate process. For RRE 7.4, this is primarily an infrastructure investment, but we expect to take advantage of the new architecture in future releases by providing exciting new interfaces to the computational engine.

#### ***Naïve Bayes Classifier***

- The new function rxNaiveBayes performs classification using Bayes Theorem to determine the probability that an observation belongs to a certain class

#### ***Stepwise Fit Improvements***

- A new argument, keepStepCoefs, has been added to the rxStepControl function. If TRUE, a data frame stepCoefs will be returned with the fitted model with rows corresponding to the coefficients and columns corresponding to the iterations. Additional computation may be required to generate the coefficients at each step. Those stepwise coefficients can be visualized by plotting the fitted model with rxStepPlot.

#### ***Efficiency Improvements***

- Import of wide data is now faster; up to 180 times faster on sample data sets with 15 thousand variables.
- Fitting of tree-based models is now faster in most cases because splits are found in parallel. (In the relatively rare case where you are building shallow trees on data sets with few variables, it can be faster to set the findSplitsInParallel flag to FALSE.)
- Prediction on rxDForest and rxBTrees objects has been speeded up by performing the prediction in a single pass over the data.

#### **Deprecated and Obsolete Functionality**

Copyright © 2015 Revolution Analytics. All rights reserved. Revolution R, Revolution R Enterprise, RPE, RevoScaleR, DeployR, RevoTreeView, and Revolution Analytics are trademarks of Revolution Analytics. All other trademarks are the property of their respective owners

- The rxGetVarInfoXdf and rxGetInfoXdf functions in RevoScaleR are now deprecated. Use rxGetVarInfo and rxGetInfo.
- Windows HPC Server 2008 is still supported, but is now deprecated.

### **Defunct Functionality**

- Netezza databases (also known as IBM PureData databases) are no longer supported.
- Windows XP and Windows Server 2003 are no longer supported.

### **Known Issues**

- [\*Known Issues in Revolution R Enterprise 7.4\*](#)