

Revolution R Enterprise 6.1 README

Revolution R Enterprise 6.1 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 the RevoScaleR package that provides fast, 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, Windows HPC Server Azure Burst cluster, or IBM Platform Computing LSF cluster. RevoScaleR also allows distribution of the execution of essentially any R function across cores and nodes, delivering the results back to the user.

Installation instructions and instructions for getting started are provided in your confirmation e-mail.

What's New in Revolution R Enterprise 6.1

Big Data Decision Tree Models

- New RevoScaleR function `rxDTree` can be used to create decision tree models. It is based on a binning algorithm so that it can scale to huge data. Both classification and regression trees are supported. The model objects returned can be made to inherit from the `rpart` class of the `rpart` package, so that `plot.rpart`, `text.rpart`, and `printcp` can be used for subsequent analysis.
- Prediction for models fitted by `rxDTree` can be done using `rxPredict`.
- See Chapter 10 of the RevoScaleR User's Guide for examples on how to create decision tree models with `rxDTree`. Additional information is available in the `rxDTree` help file, seen by entering `?rxDTree` at the R command line.

Support for Compression in .xdf Files

- RevoScaleR's .xdf files can now be created using zlib compression. The `xdfCompressionLevel` argument to `rxImport` and other .xdf creating functions allows you to specify compression for individual files and data sources, while the new `rxOptions'` argument `xdfCompressionLevel` allows you to specify the compression level session-wide. The `rxCompressXdf` function allows for compression of an existing .xdf file or a directory of existing .xdf files.
- The current default for the `xdfCompressionLevel` is 0, that is, to not compress. Level 0 compression will create files that are compatible with Revolution R Enterprise 6.0.
- See Section 2.1 of the *RevoScaleR User's Guide* for information on creating compressed .xdf files when importing. Enter `?rxCompressXdf` for information on compressing existing .xdf files.

Support for Hadoop Distributed File System Data Storage

- RevoScaleR can now read delimited text data or .xdf files from the Hadoop Distributed File System (HDFS). You can set a global file system option using `rxSetFileSystem`, which

takes options “native” and “hdfs”; to specify an HDFS system, you must supply the host name and port number for your Hadoop name node.

- See Section 2.14 of the *RevoScaleR User’s Guide* for information on using data in an HDFS file system. Enter `?RxHdfsFileSystem` and `?rxSetFileSystem` for information on creating and setting HDFS file system objects.

New Fixed-Width Character Data Type

- RevoScaleR now has a data storage type for fixed-width character data. This is always used for character data in fixed-width text data files, and can be used with other data files such as delimited text and SAS data files. Specifying the width of character data in `colInfo` when using `rxImport` may speed up import considerably.

Known Issues:

- [*Known Issues in Revolution R Enterprise 6.1*](#)