

Known Issues and Additional Information for Revolution R Enterprise 7.0

RevoScaleR

Distributed Computing:

Known Issues

- The Linux installation manual `instman.pdf` that was shipped with the product and is available in the installer directory after unpacking the installer tarball contains some obsolete information about Hadoop-related installer flags. In particular, the command given on page 28 for installing Revolution R Enterprise on a Hadoop node incorrectly includes the flag “-E”; this flag has been removed from the installer. For a corrected version of the `instman.pdf`, [click here](#).
- The file `rhadoop.sh` shipped with the 7.0.0 release has DOS line endings instead of UNIX line endings. If you have installed this file to the recommended location of `/etc/profile.d`, you may see a tell-tale error on login “: command not found”. The error also causes `rxExec` jobs to fail. To correct the error, open the file as root (or as a sudo user) using `vim` (this is the usual version of `vi` on supported Linux systems). Type a colon (:), use the command “`set ff=unix`” at the colon prompt, then save the file.
- The `rxGetJobs` function is not currently supported when using the `RxLsfCluster` compute context from Windows clients on LSF clusters.
- The ‘`dataPath`’ and ‘`outDataPath`’ arguments for the `RxHadoopMR` compute context are not yet supported.
- The ‘`rxSetVarInfo`’ function is not supported when accessing `xdf` files with the `RxHadoopMR` compute context.
- The `rxFactors` function does not give an error that it cannot be used in a distributed context.
- When running in the RPE in the Debug solution configuration, instrumented expressions fail in ‘`foreach`’ when run under a distributed compute context with `doRSR`.

Additional Information

- Handling Data in Distributed Computing Platforms

The strategy for importing and manipulating data will vary across compute platforms and file systems. For example:

- When analyzing data on an LSF or HPC cluster, the user is responsible for distributing the data. This might include copying files to nodes of the cluster and then distributing a call to `rxImport` across nodes using `rxExec`.
- When analyzing data in Hadoop with the Hadoop Distributed File System (HDFS), HDFS handles the distribution of the data. A single call to `rxImport` or `rxDataStep` will automatically access data stored across nodes of the cluster in a way that is more similar to importing data on a workstation. Note, however, that the HDFS file system does not

support modification of files once they are created (the 'append' argument is not supported for rxImport or rxDataStep). rxDataStep has these additional limitations when run within an RxHadoopMR compute context:

- 'inData' cannot be a data frame
- 'returnTransformObjects' is not supported
- The internal transformation function variable .rxStartRow is not supported
- For examples of handling data in distributed computing contexts, see the relevant Getting Started Guide:
 - RevoScaleR Getting Started with Hadoop
 - RevoScaleR Getting Started with LSF Clusters
 - RevoScaleR Getting Started with HPC Server
- rxRoc is not supported in a distributed computing framework.
- The 'outFile' argument in 'rxCube' is not supported in a distributed computing context. Similarly, the 'byGroupOutFile' argument in 'rxSummary' is not supported in a distributed computing context. (RA-10104)
- Distributed computing functionality is not supported for 32-bit Windows HPC Server node installations of Revolution R Enterprise.
- In extremely long sessions with frequent distributed computing calls to Windows HPC Server (on the order of tens of thousands of calls), a memory leak may occur.
- On Windows XP systems, rxExec and rxGetJobResults may take a long time to return if autoCleanup=TRUE.
- Distributed computing jobs can appear to be inactive if a hidden authentication window is waiting for input by the user.

Data Import and Manipulation

Known Issues

- Using a transform to change a variable's data type is not supported when repeated calls to rxImport or rxTextToXdf are used to import and append rows, combining multiple input files into a single .xdf file.
- When importing data from the Hadoop Distributed File System, attempting to interrupt the computation may result in exiting Revolution R Enterprise.
- The rxLocateFile function is not supported when the fileSystem option is globally set to HDFS.

Additional Information

- Revision 0.92 of the SQLite ODBC driver is incompatible with RevoScaleR; revisions 0.88-0.91 and 0.93 and later are known to be compatible.
- If you use data sets with extremely large numbers of variables (e.g., over 40,000), you should set the max-ppsize flag when starting R in order to use functions such as rxGetVarInfo . For example, using R directly (i.e., rgui.exe or rterm.exe), set the max-ppsize=500000 when you type

R: R --max-ppsize=500000. On Windows with the RPE, set the max-ppsize flag in the call to the RevoIDE executable: RevoIDE.exe /RCommandLine --max-ppsize=500000.

Analysis Functions

Known Issues

- The rxDTree function does not currently support in-formula transformations; in particular, using the F() syntax for creating factors on the fly is not supported. However, numeric data will be automatically binned.
- The rxDTree function cannot be used directly with an RxTeradata data source. The data must be imported to an .xdf file first.
- The rxLogit function removes missings from all variables that are read from the data source before doing any transformations or computations.
- Ordered factors are treated the same as factors in all RevoScaleR analysis functions except rxDTree.

The R Productivity Environment

Additional Information

- You cannot debug into code that has been loaded using the source() function within a script. However, if you load the code with the revoSource() function, the code will be instrumented for debugging and can be entered by the debugger.
- Installation on 32-bit Windows Server 2003 SP2 sometimes results in the warning: No template information found. To resolve the issue, run the following as an admin at the command prompt in the IDE32 directory of your installation: REvoIDE.exe /installvstemplates
- The first time you launch the help system, it may set itself to use online help before local help. If this occurs, you may not see the correct help for RPE dialogs. To be sure you have the correct help available, choose Options... from the Tool menu, then expand Environment, then expand Help, and click Online. Under "When loading help content," select "Try local first, then online."

R Base and Recommended Packages

Known Issues

In the nls function, use of the "port" algorithm occasionally causes the R front-end to stop unexpectedly. The nls help file advises caution when using this algorithm; Revolution recommends avoiding it altogether and using either the default Gauss-Newton or plinear algorithms.