Revolution R Enterprise 7.3 README

Revolution R Enterprise 7.3 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 Hadoop and Teradata. Supporting a variety of big data statistics, predictive modeling, and machine learning capabilities, Revolution R Enterprise is also 100% R. The 7.3.0 release includes R 3.1.1 and 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 7.3

RevoR

Updated R

- The included version of R has been updated to R 3.1.1, "Sock it to Me".

Install Enhancements on Red Hat Enterprise Linux Systems Running Hadoop

- When installing Revolution R Enterprise 7.3 on a Red Hat Enterprise Linux configured for Hadoop, the installer attempts to configure Revolution R Enterprise with all environment variables necessary to execute jobs via the RxHadoopMR compute context. If successful, a new file, RevoHadoopEnvVars.site, is written to the Revolution R Enterprise install directories; this file replaces the old bash_profile_additions and rhadoop.sh files, which are now deprecated. (Users who already have successfully installed RRE on their Hadoop clusters have the option of continuing to use their existing bash_profile_additions and rhadoop.sh files.)

Enhanced Cloudera Manager Parcels

- Revolution R Enterprise Cloudera Manager parcels are now fully compliant with Cloudera guidelines, and should require significantly less post-installation configuration.

Updated RevoPemaR Package

- The RevoPemaR package now includes a PemaByGroup class framework for performing arbitrary by-group computations. Expanded examples include an enhanced gradient descent algorithm. Functionality has also been extended to work with RevoScaleR distributed compute contexts (experimental).

RevoScaleR

RxHadoopMR Compute Context

- The RxHadoopMR compute context now supports additional distributions of Hadoop, including Cloudera CDH5.1 and MapR 4.0.1.

- A new argument, scheduleOnce, to rxDForest and rxBTrees, allows parallel partial decision forests to be fit using rxExec and a single MapReduce job, which can be a significant time-saver when running those functions on small to moderate-sized data sets.

RxInTeradataCompute Context
- The RxInTeradata compute context now works with Teradata 15.00 and includes RevoScaleR memory limits. Because a Teradata cluster node typically has many worker processes (AMPs) running constantly, it is important to limit how much memory a distributed analysis consumes. Revolution R Enterprise 7.3.0 and later includes both default and configurable memory limits for master and worker processes working on a RevoScaleR job. By default, the master process memory limit is 2000000000 bytes (approximately 2GB) and the worker process memory limit is 1000000000 (approximately 1GB). This may be customized by the cluster administrator using the environment variables REVOMASTERMEMLIMIT and REVOWORKERMEMLIMIT.

Stochastic Gradient Boosting
- A new function, rxBTrees, can be used to create classification and regression decision forests by means of stochastic gradient boosting.

Add Extra Variables to rxPredict Output
- A new argument to the rxPredict function, extraVarsToWrite, can be used to write additional variables to the rxPredict output. This can be used, for example, to add customer IDs or other non-model variables to the output.

Support for PMML Output
- New functions as.randomForest and as.gbm convert appropriate RevoScaleR model objects to objects of class randomForest and gbm, respectively. The underlying structure of the output object will be a subset of that produced by an equivalent call to the corresponding function. In many cases, this method can be used to coerce an object for use with the pmml package, available from CRAN or the Revolution Analytics versioned source repository. (Currently, gbm support in PMML is available only via a commercial package available from Zementis, Inc.)

Known Issues
- ***Known Issues in Revolution R Enterprise 7.3***