Lavastorm Analytics Library
R Node (Version 5.1.1.0)
Release Notes

Copyright

Confidentiality

Disclaimer

Contact Details

For product demonstrations, enhancement requests or technical questions regarding the use of any Lavastorm product, contact us as follows:

HQ Address: Lavastorm, 321 Summer Street, 5th Floor, Boston, MA 02210 USA
Telephone: +1 617 345 5422 ext. 244
Fax: +1 617 345 5475
Email: info@lavastorm.com
Internet: www.lavastorm.com

Contents

R 5.1.1.0

The 5.1.1.0 release of R brings the R functionality into the LAE 5.1 application.

Please note this release is ONLY compatible with LAE 5.1.x releases, for 5.0.x and 4.6.x customers please install the corresponding R installer (currently 5.0.1.0 for 5.0.x and 1.1.0 for 4.6.x)

R 1.1.1 Update

The 1.1.1 update to R reduces the processing time required when processing large volumes of data. Issue 5394

Version 1.1.1 of the R Node is compatible with version 4.6.0 of LAE and later LAE versions.

Please note that Rserve 1.8-0 has been compiled against java 1.7, therefore in order for LAE to use the r functionality you must ensure that your LAE instance also uses java 1.7.

R 1.1 Update

The 1.1 update to R introduces the ability to output images in addition to data frames. To assign an R plot to an output pin, the user should call createImage("pinName"). This function opens a new graphical device and makes it the current device. The user can then plot information and display text using standard R calls to the open device. R will close the device once the node is finished running and output the image to the designated pin. The node will only output images in the JPEG format, and, at this time, the Cairo engine is not supported. More information on outputting images can be found in the updated node help.

Version 1.1.0 of the R Node is compatible with version 4.6.0 of LAE and later LAE versions.

Overview

R is a popular, powerful, and extensible statistical computing platform that enables data analysts to create complex computations and apply them to solve big data problems. By leveraging the power of R, the Lavastorm Analytics Engine can be extended beyond its native analytic capabilities to address a larger problem set while retaining the LAE's other facilities for acquiring, federating, and publishing data.

The R node provides integration with the R processing engine for advanced statistical functionality. The R node can take a variable number of inputs from the LAE and make them easily available in R as native data frames. It can also extract data from R in the form of data frames or plot images and make them available in the LAE. This creates a seamless integration between the two engines, allowing the user to execute R scripts on data alongside traditional LAE nodes.

Version 1.1.0 of the R Node is compatible with version 4.6.0 of LAE and later LAE versions.

Additional Required Software

In order to use the R Node, you must first obtain and set up two additional pieces of software:

- R, available from *r-project.org*
- The RServe Client, available from RServe's Web site

Lavastorm does not distribute these software programs. Lavastorm distributes only the R node that communicates with these programs in order to execute R scripts.

Setting up R for the Lavastorm Desktop

Instructions for Setting Up R

These directions assume that both the LAE server and R are on the same Windows machine. This setup is probably appropriate for most Lavastorm Desktop installations. However, R could be located on a separate server if high-performance is desired. In this case, skip to step 8 for the LAE client-side installation instructions and read the section entitled "Setting up R for LAE Enterprise Environments" for directions on installing R on Unix.

1. Install R from *r-project.org*, following instructions on that Web site. The current version of the R node was tested against 3.1.0.

2. Start up R, using the Start menu created by the R installer. This will launch an R window.

3. Install the additional packages that the R node needs. In order to do this, run the following commands inside your R window:

   ```
   install.packages("Rserve")
   install.packages("jpeg")
   install.packages("png")
   ```

The current version of the R node was tested against Rserve release 1.8-0, png release 0.1-7, jpeg release 0.1-8; later versions of these libraries may not be compatible with the R node.

4. Install the Rserve client into your existing LAE server.
   a. Navigate to Rserve's Web site (currently, http://rforge.net/Rserve/).
   b. Click the "Download/Files" link.
   c. Download the Rserve_1.8-0.tar.gz file or Rserve_1.8-0.zip for windows binary. Note that later versions of Rserve may not be compatible with the R Node.
   d. Unzip the downloaded file, if you are using the tar.gz file then you will also then need to unzip (untar) the inner file. Windows cannot do this tar unzipping by default, but programs like WinZip, jZip, and 7-Zip can.
   e. Inside the files that were unzipped and untarred, navigate to *Rserve\inst\java*.
   f. Copy the files *REngine.jar* and *Rserve.jar* to the directory *lib\java\ext* in your LAE installation. For a standard installation of Lavastorm Desktop 5.1, this location would be *C:\Program Files (x86)\Lavastorm\LAE5.1\lib\java\ext*. If the "ext" directory does not exist, you can create it yourself.

5. Install the R Node using the R Node installer *Lavastorm_Analytics_R_Node-5_1_1_0_3-windows-x86-32.exe.*

Running the R Node

In order to run the R node, you need to start up the Rserve Server first. To do this:

1. If you do not have an R window open, start up R, using the Start menu icon created by the R installer.

2. Run the following commands in the R window:

   library(Rserve)
   Rserve()

The Rserve Server will now be running. At this point, you can close the R window and the Rserve Server will continue to run normally.

In order to use the R node in BRE, simply click on Graph->Edit Base Libraries, and add a new library named "r". The R node will then appear in the Interfaces and Adapters category in the BRE node palette.

Note: there is no Startup menu item to shut down the RServe server. In order to close Rserve, you have to open up Windows Task Manager and end the process named "Rserve.exe".

Note on Location of R Server

If you have R on the same machine as the LAE Server, you can leave the RUsername, RPassword, RHost, and RPort parameters blank in the R Node.

If you instead install R on a different machine from the LAE Server:

1. You should use the RHost and RPort parameters to tell the R Node where R is located.

2. You should consider securing R to prevent unauthorized use (see instructions in the UNIX instructions below, in section 4). If you do this, you should provide RUsername and RPassword parameters in the R Node.

Setting up R for LAE Enterprise Environments

Instructions for Setting Up R

These installation directions assume that LAE Server and R are on two separate machines. However, you can install them each on the same machine if so desired.

1. Install R from *r-project.org*, following instructions on that Web site. The current version of the R node was tested against 3.1.0. If you choose to compile R from source, be sure to configure it with the following options:

    - --enable-R-shlib
    - --with-libpng
    - --with-jpeglib

    These options ensure that the supporting binaries for the R node are available.

2. Start up R. To do this, navigate to the R installation's *bin* directory and execute the "R" program. This will bring up an R prompt, allowing you to enter R commands.

3. Install the additional packages that the R node needs by typing the following commands at the R prompt:

    ```
    install.packages("Rserve")
    install.packages("jpeg")
    install.packages("png")
    ```

    The current version of the R node was tested against release RServe 1.8-0, png release 0.1-7, and jpeg release 0.1-8; later versions of these libraries may not be compatible with the R node.

    At this point, you can exit R by typing the command "q()".

4. Start the X server on this machine, and ensure the identity that runs R has access to X. For most systems, this will mean running *xinit* as root and then ensuring that the *DISPLAY* environment variable is set to ":0" for the identity that starts Rserve. The identity that starts RServe does not need to be in a graphical environment. As long as X is started and the *DISPLAY* variable is set, all will be well.

5. Confirm the location of the R temporary directory. The R node will use this directory to store image files as they are in transit from the R server to the LAE server. While the size of each image should be relatively small, they can add up over time, especially on systems with lots of users. The temporary directory can be set up via the standard R mechanisms (see R's tempdir() help).   At the time of this writing, they can be set via the TMP, TEMP, and TMPDIR environment variables, and default to "/tmp" if none is defined.

6. Configure RServe as described in http://rforge.net/Rserve/doc.html#conf. This is a three-step process:
    a. Create a custom *Rserv.conf* configuration file. See the documentation at the link above for more information.
    b. Consider the security issues discussed in the documentation at the link above in the section that begins "A note about security". The security steps recommended there are important because they can help prevent unauthorized users from executing arbitrary R code on your R Server.
    c. Ensure that the following options are in your *Rserv.conf* file:

        encoding utf8

interactive no

remote enable

7. Navigate to the R *bin* directory and then start the Rserve server by running the following command:
    > R CMD Rserve --RS-conf *<Rserv.conf>*

    Where *<Rserv.conf>* is a reference to your custom *<Rserv.conf>* file.

8. Install the Rserve client on the LAE Enterprise Server.
    a. Navigate to Rserve's Web site (currently, http://rforge.net/Rserve/).
    b. Click the "Download/Files link".
    c. Download the Rserve_1.8-0.tar.gz file. Note that later versions of Rserve may not be compatible with the R Node.
    d. Unzip and untar the downloaded file.
    e. Navigate inside the unzipped and untared files to *Rserve/inst/java*.
    f. Copy the files *REngine.jar* and *Rserve.jar* to the directory *lib/java/ext* in your LAE installation. If the "ext" directory does not exist, you can create it yourself.

9. Install the R Node on the user BRE machines, using the R Node installer *Lavastorm_Analytics_R_Node-5_1_1_0_3-windows-x86-32.exe.*

10. Give your R Node users their connection information so they can log in.
    - Assuming that you set up R to require authentication, as recommended above, provide your users with the appropriate username and password.
    - If the LAE and R Servers are on different machines, you should provide the R Node users with a hostname. You can also provide a port if you used a custom port for Rserve.

Running the R Node

In order to run the R node, you need to start up the Rserve Server first. On many systems, an administrator may have already launched the Rserve Server, so it may already be available to you. If it is not, follow the instructions above in item 7 above to start the Rserve Server.

In order to use the R node in BRE, you can simply click on Graph->Edit Base Libraries, and add a new library named "r". The R node will then appear in the Interfaces and Adapters category in the BRE node palette.

Adding the R Node as a default base library

You can also add the R Node as a default base library for all new graphs, so you can avoid the above step for each new graph. To do this, use the following steps:

1. In the BRE client, select *Tools->Preferences* from the top level menu.
2. Click the *Libraries* tab, which will take you to a pane that has two sides, the right table is labeled "Default Base Libraries"
3. Add a new entry to the "Default Base Libraries" list, by clicking the "New" button to the right.
4. Type "r" (without quotes) in the new box. (Please note that you also need to add core, inFlow and lal1 here if you wish to keep those base libraries)
5. Click the *OK* button to close this dialog.
6. All new graphs will now contain the R Node as a base library.

Some Interesting Settings for Shared R Servers

You can increase the throughput of R scripts by increasing the value for either (or both of) the *maxinbuf* Rserve parameter or the *ls.brain.node.r.transferRowLimit* LAE server parameter. However, this throughput increase comes at the expense of lowering the number of concurrent users on the R system.

- Increasing the *maxinbuf* Rserve parameter will decrease the number of network transactions that Rserve makes in transferring input data. However, each connection would likely require a larger input buffer, and this might increase the chances that the Rserve Server runs out of memory if many R users are using the same Rserve Server at the same time.

- Increasing the *transferRowLimit* parameter allows LAE to load larger chunks of input data in memory at once as it transports the data from LAE to R. However, this may increase the chance that the LAE server runs out of memory if many R users are using the same LAE server at the same time.

Known Issues

There is one known issue with the R Node, when used with version 2.15 of R. In this configuration, it appears that all NULL Boolean values are imported into R as logical True values. If this causes a problem for your graphs, Lavastorm recommends that you use version 3.1.0 of R. If that is not feasible, there is also a workaround: cast the Boolean values to String in a filter node in front of the R Node, and then cast the imported string values to logical values in R.