Tor Metrics - Rserve

by Kevin Berry <kevin.berry@villanova.edu>

August 19, 2010

1 Overview

Rserve is a TCP/IP interface which allows other tools and languages to use the facilities of the R language. In our case, Java, Tomcat, Postgres and other parts of Metrics work with Rserve to generate graphs on-demand. For more information about the Metrics website, see the sister repository *metrics-web*. Here, we will cover how to install R, Rserve, the R Postgres driver, and ggplot2. For more information about the Postgres setup, see manual.pdf. The database should be set up before continuing this.

2 Architecture

See the *metrics-db/rserve* directory for the start/stop scripts and config. The graphs are all pre-loaded when Rserve starts, so, if any changes are made to the graph code, Rserve must be restarted. The database name, user, and password can be configured in *rserve-init.R*, as well as the pre-loaded libraries. Rserve forks itself upon connection, so R code can be pre-loaded to speed things up.

3 Setup

3.1 Installing R

Before we get started, we need to have R installed. We need to have the R dev package installed so we can use the add-ons.

```
$ sudo apt-get install r-base-dev
```

3.2 Installing and testing Rserve

There are a few different ways to install Rserve. However, the easiest and most direct way to install it is through R's built-in package manager and package network, CRAN (*The Comprehensive R Archive Network - http://cran.r-project.org*). Unfortunately, Rserve isn't packaged currently for many Linux

distributions, so it requires a bit manual configuration and administration.

R needs to be started as root so its build-in package manager can access the file system to install its own packages. Select the mirror through the Tcl/tk or command line interface and it should install.

```
$ sudo R
> install.packages("Rserve")
```

We want to start a bare server and see if it works correctly.

```
$ R CMD Rserve
```

Now, test it with the built-in R connector.

```
$ R
> library(Rserve)
> c <- RSconnect()
> RSshutdown(c)
```

If this worked, the server is listening, so it installed and started correctly.

3.3 Installing ggplot2

ggplot2 is the second necessary R package that we need for Metrics.

```
$ R
> install.packages("ggplot2")
```

3.4 Installing and testing the R PostgresSQL driver

The Postgres driver installs similarly to the other R packages.

```
$ R
> install.packages("RPostgreSQL")
```

First, make sure Postgres is started and configured correctly according to the Metrics specifications (see manual.pdf).

Start the R console, load the driver, and connect to the database. The database user, password may need to be changed.

```
$ R
> library(RPostgreSQL)
> drv <- dbDriver("PostgreSQL")
> con <- dbConnect(drv, user="ernie", password="", dbname="tordir")
> dbDisconnect(con)
> dbUnloadDriver(drv)
```

4 Administrating Rserve

Since Rserve is not standardly packaged, a few things must be done to ensure it runs smoothly and securely. We need to adjust permissions, add users, and modify groups so it works nicely with Tomcat. Feel free to do this differenty according to your system's requirements.

4.1 Adding users and groups

We'll add the user 'rserve' without a shell and no home directory.

```
$ useradd rserve -s /bin/false -U
```

Now, find the user id and group id of the rserve user, and edit them in rserve/Rserv.conf. This is so Rserve properly forks itself and runs as the correct user when it is started.

```
$ id rserve
uid=1011(rserve) gid=1012(rserve) groups=1012(rserve)
```

Next, we need to add the rserve user to the 'apache' group (The default user for Tomcat), so it can communicate correctly with Tomcat and have the necessary permissions for writing graphs. In this case, we will add apache to the rserve group.

```
$ usermod -a -G rserve apache
```

5 Start Rserve

Now we are ready to start Rserve! Run the script rserve/start.sh as ROOT (or else Rserve will not fork itself properly). Then, check the rserve log file (*rserve.log*). The path to this log file can be changed by modifying the script. Do a "ps -ef — grep rserve" to see if it has started. Now, with Rserve installed and Postgres is running, Metrics is almost ready to start generating some graphs!