# DataSci Tutorial 1: R Setup for Zeppelin and RStudio Server

FINISHED

This tutorial was built for BDCS-CE version 17.4.1 as part of the Data Science Acceleration User Journey: here (https://oracle.github.io/learning-library/workshops/journey3-data-science/). Questions and feedback about the tutorial: david.bayard@oracle.com (mailto:david.bayard@oracle.com)

```
These DataSci tutorials build on top of the New Data Lake tutorial.  Please start with the New Data Lake User Journey.  In particular, be sure you previously ran the
New Data Lake Tutorials: "Citi Bike New York Demo Introduction and Setup" and "Working with Hive".
```

This notebook provides shell scripts to configure R for Zeppelin and to install RStudio Server, which is an optional third-party tool to help with R coding.

Took 0 sec. Last updated by anonymous at November 16 2017, 2:51:42 PM.

### Script to setup R for use with Zeppelin (will take about 15 minutes)

FINISHED

```
%sh

echo "Starting script to setup R for use with Zeppelin. Please be patient.  You will not see any more output until it is finished."

sudo /usr/bin/enable-zeppelin-r.sh > /tmp/install-oracle-r.log 2>&1

echo ".."
echo ".."
echo "First 50 lines..."
head -50 /tmp/install-oracle-r.log

echo ".."
echo ".."
echo "Last 50 lines..."
tail -50 /tmp/install-oracle-r.log

echo ".."
echo ".."
echo "done"

echo "done"
```

```
Starting script to setup R for use with Zeppelin. Please be patient.  You will not see any more output until it is finished.
..
..
First 50 lines...
Oracle Distribution of R version 3.2.0  (--) -- "Full of Ingredients"
Copyright (C)  The R Foundation for Statistical Computing
Platform: x86_64-unknown-linux-gnu (64-bit)
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
  Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
You are using Oracle's distribution of R. Please contact
Oracle Support for any problems you encounter with this
distribution.
> install.packages('evaluate', repos='http://cran.us.r-project.org');
Installing package into ‘/usr/lib64/R/library’
(as ‘lib’ is unspecified)
trying URL 'http://cran.us.r-project.org/src/contrib/evaluate_0.10.1.tar.gz'
Content type 'application/x-gzip' length 22177 bytes (21 KB)
==================================================
downloaded 21 KB
* installing *source* package ‘evaluate’ ...
** package ‘evaluate’ successfully unpacked and MD5 sums checked
** R
** preparing package for lazy loading
** help
*** installing help indices
  converting help for package ‘evaluate’
    finding HTML links ... done
    create_traceback                        html
Rd warning: /tmp/RtmphVbCUg/R.INSTALL41a5fcbd4f0/evaluate/man/create_traceback.Rd:11: missing file link ‘sys.calls’
    evaluate                                html
    flush_console                           html
    inject_funs                             html
    is.message                              html
    line_prompt                             html
    new_output_handler                      html
    parse_all                               html
    replay                                  html
    set_hooks                               html
..
..
Last 50 lines...
    renderGvis                              html
** building package indices
** installing vignettes
** testing if installed package can be loaded
Creating a generic function for ‘toJSON’ from package ‘jsonlite’ in package ‘googleVis’
* DONE (googleVis)
Making 'packages.html' ... done
* installing *source* package ‘mplot’ ...
** package ‘mplot’ successfully unpacked and MD5 sums checked
** R
** data
*** moving datasets to lazyload DB
** inst
```

```
** preparing package for lazy loading
** help
*** installing help indices
  converting help for package ‘mplot’
    finding HTML links ... done
    af                                      html
    artificialeg                            html
    bglmnet                                 html
    bodyfat                                 html
    diabetes                                html
    fev                                     html
    mplot-package                           html
    mplot                                   html
    plot.af                                 html
    plot.bglmnet                            html
    plot.vis                                html
    print.af                                html
```

Took 1 min 33 sec. Last updated by anonymous at November 16 2017, 2:41:33 PM.

## Install RStudio Server (will take about a minute)

FINISHED

```
%sh
sudo -- sh -c 'wget -nc -nv https://download2.rstudio.org/rstudio-server-rhel-1.0.143-x86_64.rpm; yum -y install --nogpgcheck rstudio-server-rhel-1.0.143-x86_64.rpm'
```

```
2017-11-16 19:11:27 URL:https://download2.rstudio.org/rstudio-server-rhel-1.0.143-x86_64.rpm [41442544/41442544] -> "rstudio-server-rhel-1.0.143-x86_64.rpm" [1]
Loaded plugins: security
Setting up Install Process
Examining rstudio-server-rhel-1.0.143-x86_64.rpm: rstudio-server-1.0.143-1.x86_64
Marking rstudio-server-rhel-1.0.143-x86_64.rpm to be installed
Resolving Dependencies
--> Running transaction check
---> Package rstudio-server.x86_64 0:1.0.143-1 will be installed
--> Finished Dependency Resolution
Dependencies Resolved
================================================================================
 Package         Arch    Version     Repository                            Size
================================================================================
Installing:
 rstudio-server  x86_64  1.0.143-1   /rstudio-server-rhel-1.0.143-x86_64  311 M
Transaction Summary
================================================================================
Install       1 Package(s)
Total size: 311 M
Installed size: 311 M
Downloading Packages:
Running rpm_check_debug
Running Transaction Test
Transaction Test Succeeded
Running Transaction
  Installing : rstudio-server-1.0.143-1.x86_64                             1/1
groupadd: group 'rstudio-server' already exists
rsession: no process killed
rstudio-server start/running, process 25857
  Verifying  : rstudio-server-1.0.143-1.x86_64                             1/1
```

```
Installed:
  rstudio-server.x86_64 0:1.0.143-1
Complete!
```

Took 1 min 39 sec. Last updated by anonymous at November 16 2017, 2:13:03 PM. (outdated)

### Create a user for RStudio Server

FINISHED

```
%sh
cat <<EOF > /tmp/r_user.sh
    useradd ruser
    echo -e "welcome1\nwelcome1" | passwd ruser
    echo "finsihed adduser"
    usermod -a -G hive,hadoop ruser
    echo "finsihed groups"
    sudo -u hdfs hadoop fs -mkdir /user/ruser
    sudo -u hdfs hadoop fs -chown ruser /user/ruser
    echo "finsihed hdoop fs commands"
EOF

chmod a+x /tmp/r_user.sh
sudo /tmp/r_user.sh

echo "done"
```

```
New password: BAD PASSWORD: it is based on a dictionary word
Retype new password: Changing password for user ruser.
passwd: all authentication tokens updated successfully.
finsihed adduser
finsihed groups
finsihed hdoop fs commands
done
```

Took 6 sec. Last updated by anonymous at November 16 2017, 2:13:18 PM. (outdated)

# Restart the Zeppelin notebook

FINISHED

Now that we've installed some additional R packages for Zeppelin, you may need to restart the Zeppelin Notebook for R to work. To do so,

- Go to the Setting tab
- Click on Notebook
- Click on Restart

Took 1 sec. Last updated by anonymous at November 16 2017, 3:00:01 PM. (outdated)

### Test that R is working (will take a minute for initial startup)

FINISHED

```
%r
3+3
```

```
[1] 6
```

Took 43 sec. Last updated by anonymous at November 16 2017, 2:58:19 PM.

## List of R packages now installed

FINISHED

```
%r
ip <- as.data.frame(installed.packages()[,c(1,3:4)])
rownames(ip) <- NULL
ip <- ip[is.na(ip$Priority),1:2,drop=FALSE]
print(ip, row.names=FALSE)
```

```
    Package  Version
     SparkR    2.1.0
 assertthat    0.2.0
    bestglm     0.36
         BH 1.65.0-1
      bindr      0.1
   bindrcpp      0.2
     bitops    1.0-6
 colorspace    1.3-2
       curl      3.0
   devtools   1.13.4
  dichromat    2.0-0
     digest   0.6.12
 doParallel   1.0.11
      doRNG    1.6.6
      dplyr    0.7.4
   evaluate   0.10.1
    foreach    1.4.3
    ggplot2    2.2.1
      git2r   0.19.0
     glmnet   2.0-13
       glue    1.2.0
  googleVis    0.6.2
     grpreg    3.1-2
     gtable    0.2.0
      highr      0.6
  htmltools    0.3.6
     httpuv    1.3.5
       httr    1.3.1
  iterators    1.0.8
   jsonlite      1.5
      knitr     1.17
   labeling      0.3
   lazyeval    0.2.1
      leaps      3.0
   magrittr      1.5
   markdown      0.8
    memoise    1.1.0
       mime      0.5
      mplot    0.8.0
    munsell    0.4.3
    openssl    0.9.9
  pkgconfig    2.0.1
   pkgmaker     0.22
      plogr    0.1-1
       plyr    1.8.4
         R6    2.2.2
    rCharts    0.4.5
```

```
RColorBrewer    1.1-2
          Rcpp  0.12.13
         RCurl 1.95-4.8
      registry      0.3
      reshape2    1.4.2
           RFO   4.6-10
       RJSONIO    1.3-0
         rlang    0.1.4
      rngtools    1.2.4
    rstudioapi      0.7
        scales    0.5.0
         shiny    1.0.5
 shinydashboard    0.6.1
   sourcetools    0.1.6
       stringi    1.1.5
       stringr    1.2.0
        tibble    1.3.4
   viridisLite    0.2.0
       whisker    0.3-2
         withr    2.1.0
        xtable    1.8-2
          yaml   2.1.14
```

Took 0 sec. Last updated by anonymous at November 16 2017, 2:58:27 PM. (outdated)

# Next Steps

READY

Your BDCS-CE environment is now setup with R, SparkR, RStudio Server, and many common R packages.

Proceed to the next Tutorial to see R and SparkR in action with the Zeppelin Notebook.

## Change Log

FINISHED

November 16, 2017 - Updated for 17.4.1. R is now installed, but need to still add some packages like knitr to work with Zeppelin
September 12, 2017 - Confirmed it works with BDCSCE 17.3.5
August 13, 2017 - Confirmed it works with BDCS-CE 17.3.3-20
August 11, 2017 - First version

Took 0 sec. Last updated by anonymous at November 16 2017, 2:24:12 PM.

```
%sh
```

READY