Getting Started with R Services

Paul Ferrill

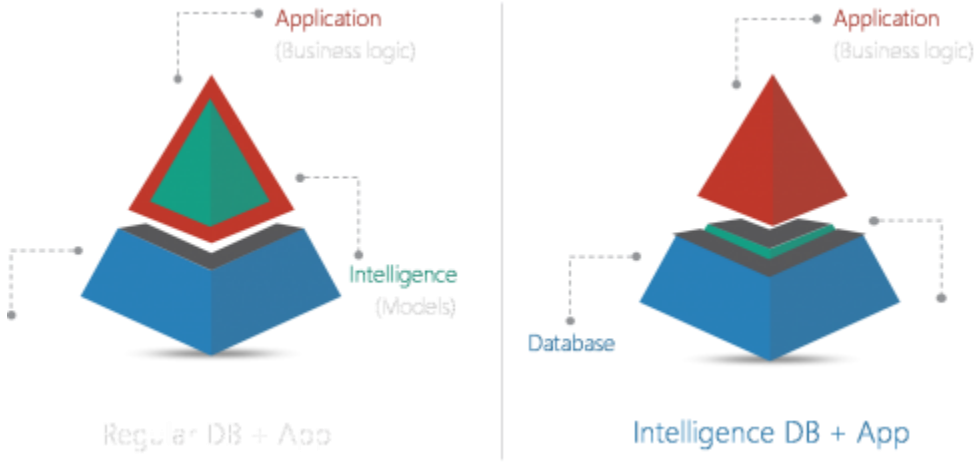
Microsoft MVP - Cloud and Data Center

About me

- BS (U of AL) / MS (U of FL) EE
- Cloud and Datacenter MVP 3 years
- Freelance writer
  - Codeproject.com, CodeGuru.com, RedmondMag.com, InfoWorld.com, Network World, PC Magazine, ServerWatch.com, IBM/developerworks
  - Pro Android Python with SL4A – Apress
  - Developing Sharepoint Apps with LightSwitch – O'Reilly
  - Upcoming Microsoft sponsored article on R Services
- Python fan

Why R in SQL Server?

SQL Server R Services is an in-database analytics feature that tightly integrates R with SQL Server. With this feature, we want to provide a data intelligence platform that moves intelligence capabilities provided with R closer to the data. So why is that a good thing? The short answer is that a platform like this makes it much easier to consume and manage R securely and at scale in applications in production.



The way we make SQL Server and R work together is by using a framework we call the extensibility architecture. Previously, CLR or extended stored procedures would enable you to run code outside the constructs of SQL Server, but in those cases, the code still runs inside the SQL Server process space. Having external code running inside the SQL Server process space can cause disruption and it is also not possible to legally embed runtimes that are not owned by Microsoft.

https://blogs.msdn.microsoft.com/sqlserverstorageengine/2017/01/10/sql-server-r-services-why-did-we-build-it

Azure

- Great learning platform for data science
- Free tier has a number of services
- https://azure.microsoft.com/en-us/free/pricing-offers/
- 30-day trial with $200 usage credit
- Continue with all free services

Azure Data Science VMs

- Windows and Linux versions
- You can download the VHD and use locally
- https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-data-science-linux-dsvm-intro
- Demo

Free Azure Test Drive

Microsoft Azure | Test Drive Preview UI

Paul Ferrill
paul@ferrill.net

My Test Drives > Linux Data Science Virtual Machine

Linux Data Science Virtual Machine

by Microsoft

▷ Start Free Test Drive

Details

The Linux Data Science Virtual Machine is a custom Azure virtual machine image purposely built for data science. It contains many of the popular data science tools pre-installed and pre-configured to jump-start advanced analytics. It also has several Azure tools and libraries installed to allow working with various Azure data and analytics products in the cloud. This virtual machine improves data scientist productivity and enables users to try our products, run analytics modeling workloads, and replace their analytics desktop with a cloud-hosted data-science machine for a significant part of their work.

Test Drive Instructions

TEST DRIVE DURATION
8 hours

ESTIMATED DEPLOYMENT DURATION
1 minute

↗ See details on Azure marketplace

↗ Add this to your Azure account

Containers

- Docker on Windows 10
- Requires Hyper-V to be installed
- https://docs.microsoft.com/en-us/virtualization/windowscontainers/quick-start/quick-start-windows-10
- Allows you to run both Linux and Windows containers
- SQL Server Express and full as a container

docker pull microsoft/mssql-server-windows

R in a Container

- To run the jupyter notebook container

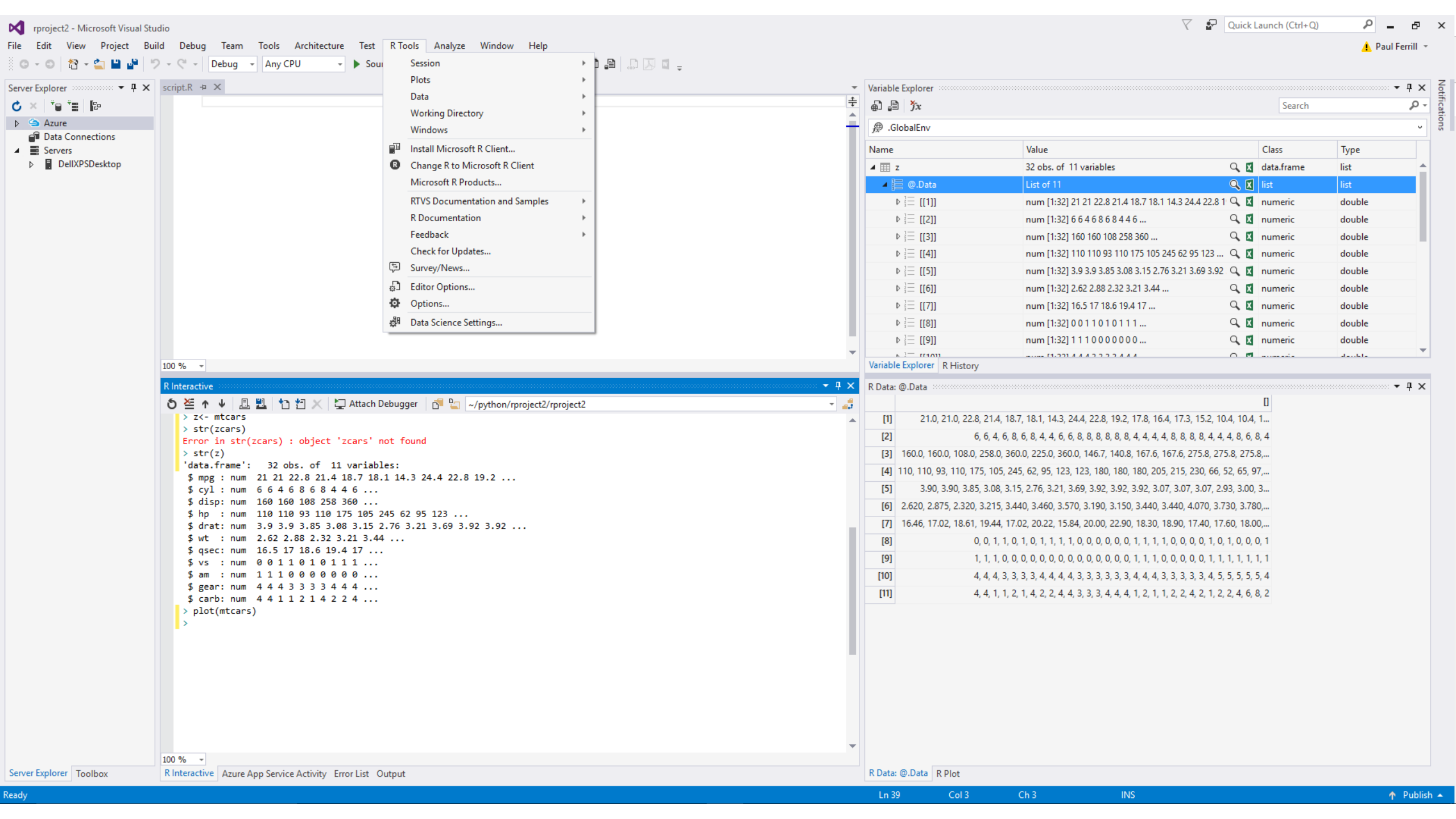
```
docker run -it --rm -p 8888:8888 jupyter/r-notebook
```

- To run the iPython r bridge container

```
docker run -it --rm -p 8888:8888 rpy2/rpy2:2.8.x ipython
```

R Tools for Visual Studio

- Available as a separate download
- https://www.visualstudio.com/vs/rtvs/
- Currently requires Visual Studio 2015 but does work with community edition
- DEMO


rproject2 - Microsoft Visual Studio
Quick Launch (Ctrl+Q)
File Edit View Project Build Debug Team Tools Architecture Test R Tools Analyze Window Help
Paul Ferrill
Debug
Any CPU
Session
Plots
Data
Working Directory
Windows
Install Microsoft R Client...
Change R to Microsoft R Client
Microsoft R Products...
RTVS Documentation and Samples
R Documentation
Feedback
Check for Updates...
Survey/News...
Editor Options...
Options...
Data Science Settings...
Server Explorer
Azure
Data Connections
Servers
DellXPSDesktop
script.R
100 %
Variable Explorer
Search
.GlobalEnv
Name
Value
Class
Type
z
32 obs. of 11 variables
data.frame
list
@.Data
List of 11
list
list
[[1]]
num [1:32] 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 1
numeric
double
[[2]]
num [1:32] 6 6 4 6 8 6 8 4 4 6 ...
numeric
double
[[3]]
num [1:32] 160 160 108 258 360 ...
numeric
double
[[4]]
num [1:32] 110 110 93 110 175 105 245 62 95 123 ...
numeric
double
[[5]]
num [1:32] 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92
numeric
double
[[6]]
num [1:32] 2.62 2.88 2.32 3.21 3.44 ...
numeric
double
[[7]]
num [1:32] 16.5 17 18.6 19.4 17 ...
numeric
double
[[8]]
num [1:32] 0 0 1 1 0 1 0 1 1 1 ...
numeric
double
[[9]]
num [1:32] 1 1 1 0 0 0 0 0 0 0 ...
numeric
double
Variable Explorer
R History
R Interactive
Attach Debugger
~/python/rproject2/rproject2
> z<- mtcars
> str(zcars)
Error in str(zcars) : object 'zcars' not found
> str(z)
'data.frame': 32 obs. of 11 variables:
$ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
$ cyl : num 6 6 4 6 8 6 8 4 4 6 ...
$ disp: num 160 160 108 258 360 ...
$ hp : num 110 110 93 110 175 105 245 62 95 123 ...
$ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
$ wt : num 2.62 2.88 2.32 3.21 3.44 ...
$ qsec: num 16.5 17 18.6 19.4 17 ...
$ vs : num 0 0 1 1 0 1 0 1 1 1 ...
$ am : num 1 1 1 0 0 0 0 0 0 0 ...
$ gear: num 4 4 4 3 3 3 3 4 4 4 ...
$ carb: num 4 4 1 1 2 1 4 2 2 4 ...
> plot(mtcars)
>
100 %
R Data: @.Data
[1] 21.0, 21.0, 22.8, 21.4, 18.7, 18.1, 14.3, 24.4, 22.8, 19.2, 17.8, 16.4, 17.3, 15.2, 10.4, 10.4, 1...
[2] 6, 6, 4, 6, 8, 6, 8, 4, 4, 6, 6, 8, 8, 8, 8, 8, 4, 4, 4, 4, 8, 8, 8, 8, 4, 4, 4, 8, 6, 8, 4
[3] 160.0, 160.0, 108.0, 258.0, 360.0, 225.0, 360.0, 146.7, 140.8, 167.6, 167.6, 275.8, 275.8, 275.8,...
[4] 110, 110, 93, 110, 175, 105, 245, 62, 95, 123, 123, 180, 180, 180, 205, 215, 230, 66, 52, 65, 97,...
[5] 3.90, 3.90, 3.85, 3.08, 3.15, 2.76, 3.21, 3.69, 3.92, 3.92, 3.92, 3.07, 3.07, 3.07, 2.93, 3.00, 3...
[6] 2.620, 2.875, 2.320, 3.215, 3.440, 3.460, 3.570, 3.190, 3.150, 3.440, 3.440, 4.070, 3.730, 3.780,...
[7] 16.46, 17.02, 18.61, 19.44, 17.02, 20.22, 15.84, 20.00, 22.90, 18.30, 18.90, 17.40, 17.60, 18.00,...
[8] 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1
[9] 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1
[10] 4, 4, 4, 3, 3, 3, 3, 3, 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 4, 4, 4, 3, 3, 3, 3, 3, 4, 5, 5, 5, 5, 5, 4
[11] 4, 4, 1, 1, 2, 1, 4, 2, 2, 4, 4, 3, 3, 3, 4, 4, 4, 1, 2, 1, 1, 2, 2, 4, 2, 1, 2, 2, 4, 6, 8, 2
R Data: @.Data
R Plot
Server Explorer
Toolbox
R Interactive
Azure App Service Activity
Error List
Output
Ready
Ln 39
Col 3
Ch 3
INS
Publish
Notifications

Jupyter Notebook

- https://www.continuum.io/blog/developer/jupyter-and-conda-r
- conda install -c r r-essentials
- Jupyter notebook
- Demo

Links

- Starting https://msdn.microsoft.com/en-us/library/mt604885.aspx
- R Studio - https://www.rstudio.com/products/rstudio/
- Samples- https://github.com/Microsoft/SQL-Server-R-Services-Samples
- Analytics - http://learnanalytics.microsoft.com/
- R Tools for VS - http://microsoft.github.io/RTVS-docs/
- Docker - https://github.com/jupyter/docker-stacks/tree/master/r-notebook
- Data Science VM – http://aka.ms/dsvmtenthings
- R Succinctly by Barton Poulson from Syncfusion