



From R to Machine Learning Services and Beyond

April 27, 2019

Others Talk,  
We Listen.

## 2 Introduction

- Mark Hudson

- >20 years mixing technology with data
- >10 years with CapTech
- Microsoft Certified IT Professional – Business Intelligence
- Member of the Richmond SQL Server User Group
- Email: [mhudson@captechconsulting.com](mailto:mhudson@captechconsulting.com)
- Twitter: @HMarkHudson



- CapTech

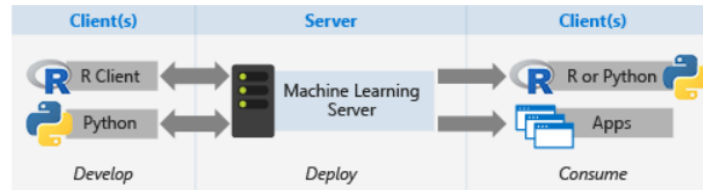
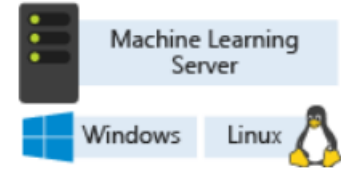
- Headquarters – Richmond
- Offices – Reston, Charlotte, Philadelphia, Columbus, Chicago, Atlanta, Denver
- Local, national, and international clients
- Microsoft Gold Partner
- Web: [www.captechconsulting.com](http://www.captechconsulting.com)
- Twitter: @CapTechListens

# Thanks to our sponsors



## 4 Agenda

- What is R?
- Microsoft's Contributions to R
- How Does It Work?
- Demo
- Questions



## 5 What is R?

- Open-source programming language for analyzing and visualizing data
  - Developed in the 1990s modeled on the S Language for Statistical Computing
  - Collection of libraries each with a collection of functions
- Runs on Windows, Unix, and MacOS platforms
- Profile data
  - head, summary, and by functions profile data
- Visualize data
  - ggplot with geom\_line and geom\_boxplot visualize data
- Analyze data
  - cor (correlate) and lm (linear model) analyze data
- Available from Comprehensive R Archive Network (CRAN)
  - <https://cran.r-project.org/>



## 6 Strengths & Weaknesses of R

- Strengths
  - Open Source ... as in free for download
  - Supports broad analysis from simple summaries to complex predictive models
  - 1,000s of packages with data transformation, analysis, and visualization functions
  - Imports data from broad collection of data sources
- Weaknesses
  - Open Source ... as in (arguably) limited support
  - Single-thread processing without separate libraries
  - Constrained by available RAM
  - Hard to productionize code

## 7 Microsoft's Contributions to R

- Microsoft acquires Revolution Analytics (2015)
- Microsoft integrates Revolution Analytics' tools (2016)
  - RevoScaleR package
  - Open Source, Client, and SQL Server offerings
- Microsoft R Open available from Microsoft R Application Network (MRAN)
  - <https://mran.microsoft.com/>
- Microsoft R Client available for local analysis
  - <https://docs.microsoft.com/en-us/machine-learning-server/r-client/what-is-microsoft-r-client>
- Machine Learning Server available for server license
  - RevoScaleR
  - revoscalepy



## 8 What is RevoScaleR?

- Microsoft SQL Server platform + Revolution Analytics functions
  - >200 data related functions
  - ODBC connections to many data sources
  - Remote computing concepts for SQL Server and Hadoop platforms
- Microsoft integrates RevoScaleR with SQL Server
  - SQL Server 2016 – R Server
  - SQL Server 2017 – Machine Learning Server
  - Azure SQL – Coming now in Preview
- Current Versions
  - Microsoft R Open – Community Support
  - Microsoft R Client – Community w/RevoScaleR
  - Machine Learning Server
    - Includes RevoScaleR and revoscalepy

```
> ls("package:RevoScaleR")
 [1] "as.gbm"           "as.glm"           "as.kmeans"
 [4] "as.lm"           "as.naiveBayes"   "as.randomForest"
 [7] "as.rpart"        "as.xtabs"         "colnames"
[10] "formula"         "prune.rxDTree"   "row.names"
[13] "rxAddInheritance" "rxAuc"           "rxBTrees"
[16] "rxCancelJob"     "rxChiSquaredTest" "rxCleanupJobs"
[19] "rxClose"         "rxCompareContexts" "rxCompressXdf"
[22] "RxComputeContext" "rxCor"           "rxCorCoef"
[25] "rxCorData"       "rxCov"           "rxCovCoef"
[28] "rxCovCor"        "rxCovData"       "rxCreateColInfo"
[31] "rxCrossTabs"     "rxCube"          "rxDataFrameToXdf"
[34] "RxDataSource"    "rxDataStep"      "rxDataStepXdf"
[37] "rxDeleteObject"  "rxDForest"       "rxDistributeJob"
[40] "rxDTree"         "rxDTreeBestCp"  "rxElemArg"
[43] "rxExec"          "rxExecBy"        "rxExecuteSQLDDL"
[46] "rxFactors"       "rxFileSystem"    "rxFindFilePath"
```



## 9 Strengths & Weaknesses of RevoScaleR

- Strengths

- Compatible with non-RevoScaleR functions
- Multi-thread processing
- Processes chunks of data – Avoids RAM constraint
- Compute context enables remote execution – “Take analytics to the data”
- Integrates R code with T-SQL
- Introduces External Data Frame (.xdf) files – Compressed, columnar data storage

- Weaknesses

- Uses external scripts
- Limited Azure SQL support – For now, in preview
- Partially converted functions – For now?
- Converted function consistency – `summary()` v. `rxSummary()`

## 10 How Machine Learning Server Works

- External Scripts
  - Executes commands through `sp_execute_external_script`
  - Requires users to be granted External Script permissions
  - Uses SQL Server Trusted Launchpad
  - External processes executed through external users
- Native
  - Fastest processing speeds using `Predict()` T-SQL function
  - Limited to RevoScaleR functions and SQL Server 2017
  - Requires trained models in binary format

## 11 R Server Demo

- Microsoft Machine Learning Server 2017
- Visual Studio 2017
- Lahman's Baseball Database
  - <http://www.seanlahman.com/baseball-archive/statistics/>

## 12 Microsoft Machine Learning Server

Machine Learning Server Documentation

<https://docs.microsoft.com/en-us/machine-learning-server/>

Machine Learning Server Tutorials

<https://docs.microsoft.com/en-us/sql/advanced-analytics/tutorials/machine-learning-services-tutorials?view=sql-server-2017>

Preview of Machine Learning Services with R support in Azure SQL Database

<https://cloudblogs.microsoft.com/sqlserver/2018/11/07/announcing-preview-of-machine-learning-services-with-r-support-in-azure-sql-database/>

Compare Machine Learning Server and Microsoft R products

<https://docs.microsoft.com/en-us/machine-learning-server/what-is-r-server-interopability>

# 13 Questions

