

Cortana Intelligence Suite

Focus – Microsoft R for Architects

- Main page: <http://cortanaanalytics.com>
- To begin this module, you should have:
 - Basic Math and Stats skills
 - Business and Domain Awareness
 - General Computing Background

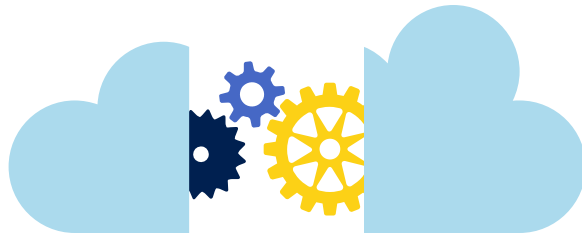
NOTE: These workbooks contain many resources to lead you through the course, and provide a rich set of references that you can use to learn much more about these topics. If the links do not resolve properly, type the link address in manually in your web browser. If the links have changed or been removed, simply enter the title of the link in a web search engine to find the new location or a corollary reference.

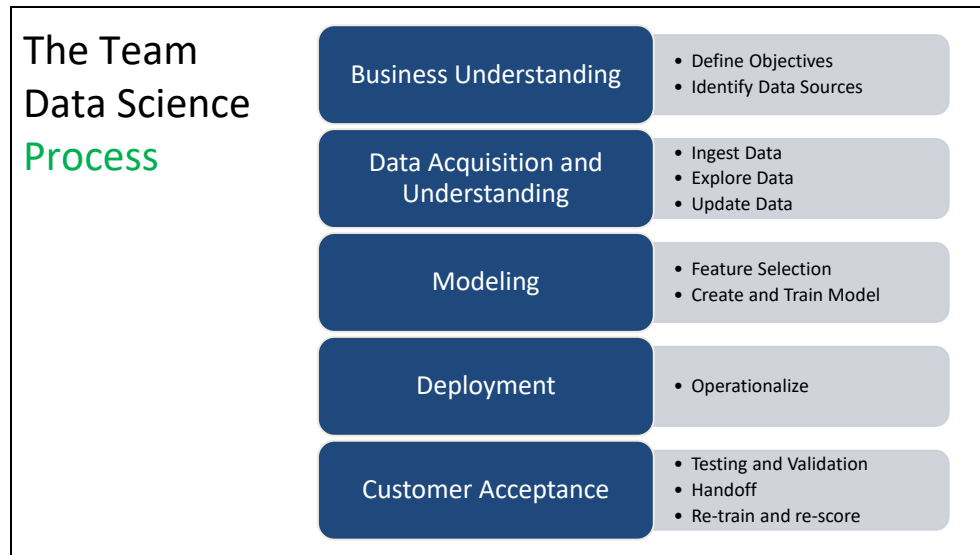
Learning Objectives

1. Understand the R Language and where it is used
2. Understand the Microsoft R Platform and its capabilities
3. Set up and use the server and various client tools for a R environment
4. Know how to operationalize a SQL Server R Services environment
5. Use the Microsoft R capabilities in a solution

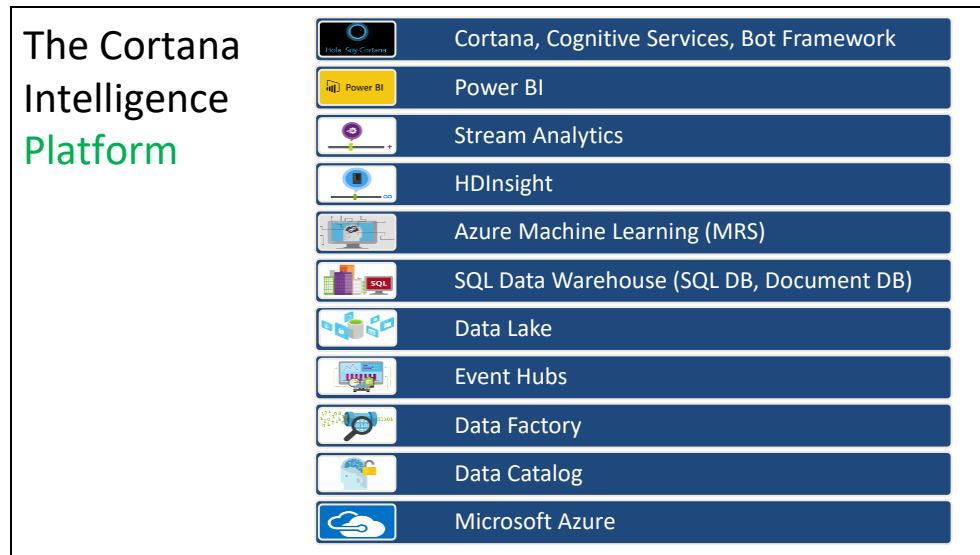
- Understand the R Language and where it is used
- Understand the Microsoft R Platform and its capabilities
- Set up and use the server and various client tools for a R environment
- Know how to operationalize a SQL Server R Services environment
- Use the Microsoft R capabilities in a solution

The Data Science Process and Platform





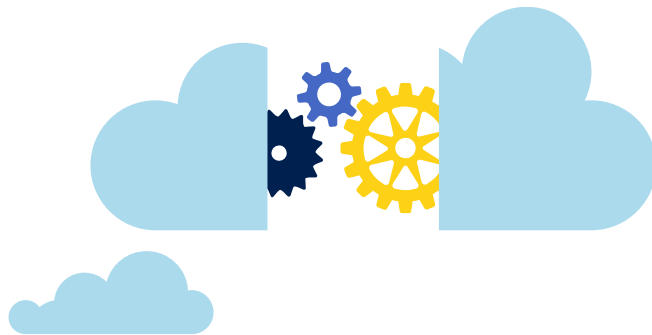
- This process largely follows the CRISP-DM model: <http://www.sv-europe.com/crisp-dm-methodology/>
- It also references the Cortana Intelligence process: <https://azure.microsoft.com/en-us/documentation/articles/data-science-process-overview/>
- A complete process diagram is here: <https://azure.microsoft.com/en-us/documentation/learning-paths/cortana-analytics-process/>
- Some walkthrough's of the various services: <https://azure.microsoft.com/en-us/documentation/articles/data-science-process-walkthroughs/>
- An integrated process and toolset allows for a more close-to-intent deployment
- Iterations are required to close in on the solution – but are harder to management and monitor



- Platform and Storage: Microsoft Azure – <http://microsoftazure.com> Storage: <https://azure.microsoft.com/en-us/documentation/services/storage/> (Host It)
- Azure Data Catalog: <http://azure.microsoft.com/en-us/services/data-catalog> (Doc It)
- Azure Data Factory: <http://azure.microsoft.com/en-us/services/data-factory/> (Move It)
- Azure Event Hubs: <http://azure.microsoft.com/en-us/services/event-hubs/> (Bring It)
- Azure Data Lake: <http://azure.microsoft.com/en-us/campaigns/data-lake/> (Store It)
- Azure DocumentDB: <https://azure.microsoft.com/en-us/services/documentdb/> , Azure SQL Data Warehouse: <http://azure.microsoft.com/en-us/services/sql-data-warehouse/> (Relate It)
- Azure Machine Learning: <http://azure.microsoft.com/en-us/services/machine-learning/> (Learn It)
- Azure HDInsight: <http://azure.microsoft.com/en-us/services/hdinsight/> (Scale It)
- Azure Stream Analytics: <http://azure.microsoft.com/en-us/services/stream-analytics/> (Stream It)
- Power BI: <https://powerbi.microsoft.com/> (See It)
- Cortana: <http://blogs.windows.com/buildingapps/2014/09/23/cortana-integration-and-speech-recognition-new-code-samples/> and <https://blogs.windows.com/buildingapps/2015/08/25/using-cortana-to-interact-with-your-customers-10-by-10/> and <https://developer.microsoft.com/en-us/Cortana> (Say It)
- Cognitive Services: <https://www.microsoft.com/cognitive-services>
- Bot Framework: <https://dev.botframework.com/>

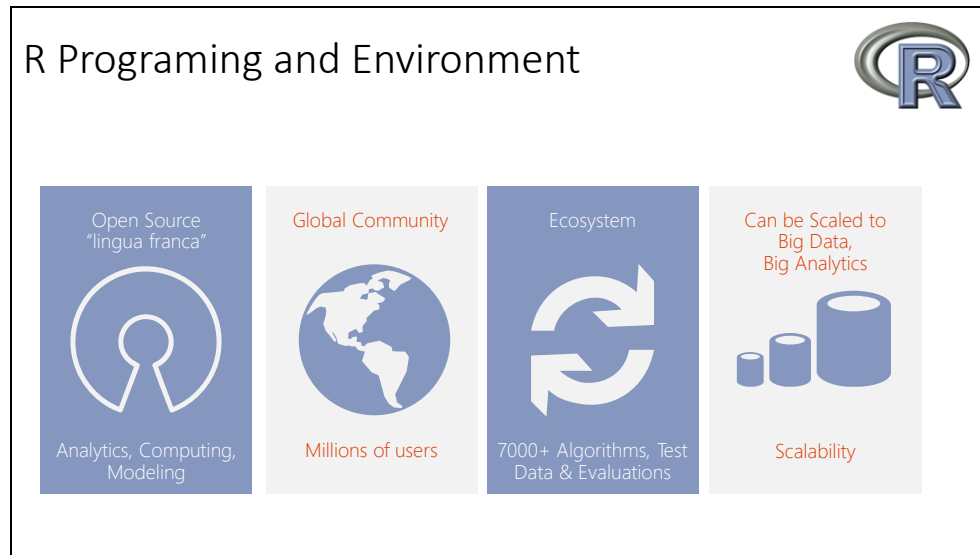
- All of the components within the suite: <https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/what-is-cortana-intelligence.aspx>
- What can I do with it? <https://gallery.cortanaintelligence.com/>
- Getting Started Quickly: <https://caqs.azure.net/#gallery>

Module 1: The R Programming Environment



6

- Video Introduction to R: <https://mran.revolutionanalytics.com/documents/what-is-r/>



- R In Youtube:
<https://www.youtube.com/user/thelearnr>
- R Links:
<http://www.datasciencecentral.com/m/discussion?id=6448529%3ATopic%3A280135>
- R resources: <https://msdn.microsoft.com/en-us/microsoft-r/microsoft-r-more-resources>

SQL and R Contrasted

SQL

1. Client/Server
2. Database Objects
3. DML, DDL
4. DCL
5. Declarative Code

R

1. Interactive Environment
2. Data Structures
3. Functions
4. Libraries (Packages)
5. Functional Code Flow

- Learn SQL: <http://www.w3schools.com/SQL/default.asp>
- Try R, with a great interface.
<http://tryr.codeschool.com/levels/1/challenges/22>
- R and Statistics Intro:
<https://www.youtube.com/watch?v=xb5P5xdcr2U&feature=youtu.be&a>
- R Online: http://www.tutorialspoint.com/r_terminal_online.php
- Using R to explore data:
<http://www.analyticsvidhya.com/blog/2015/10/cheatsheet-11-steps-data-exploration-with-codes/>
- Quick R Intro:
<http://www.datasciencecentral.com/m/blogpost?id=6448529%3ABlogPost%3A112754>

- Creating a recommender engine in R:
<http://www.analyticbridge.com/profiles/blogs/build-basic-recommendation-engine-using-r>
- Visualizations cheat-sheet in R:
<http://www.datasciencecentral.com/forum/topics/cheat-sheet-data-visualization-with-r?groupUrl=tutorials>

R Data Types

- Numeric
- Integer
- Complex
- Logical
- Character

R Data Structures

- **Vector**
 - A single-line sequence of one datatype
- **List**
 - An ordered collection of objects, allowing a variety of (possibly unrelated) objects under one name
- **Matrix**
 - A multi-line sequence of the same length and datatype
- **Array**
 - Like a Matrix, but with more dimensions

Lab:

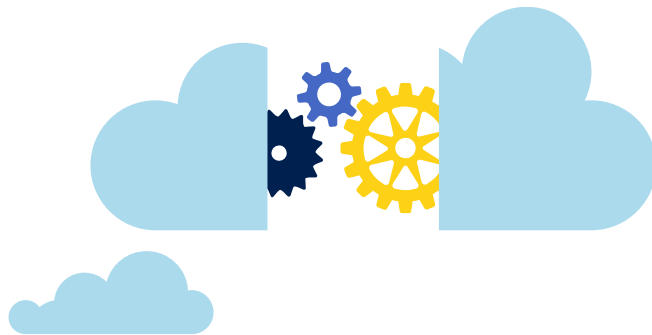
VM Setup

Step 1: Getting started

- If you do not have a Microsoft Azure account, go here: <https://azure.microsoft.com/en-us/free/> (You will need a credit card, but you will not be charged)
- Log in to the Azure Portal: <https://ms.portal.azure.com>
- Create a new Windows Data Science Virtual Machine (Size DS4_v2 or bigger): <https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-vm-do-ten-things/>
- On the DSVM, open up Visual Studio and ensure you can go through the "Getting started" lab on the MRS course at <https://rheartpython.github.io/cisw/> - you will need to install some packages with the **install.packages()** command so search the web if you don't know how to use it or ask someone.
- Optional: For those just beginning with R, Open this site, complete the lessons the instructor assigns: <http://tryr.codeschool.com/>

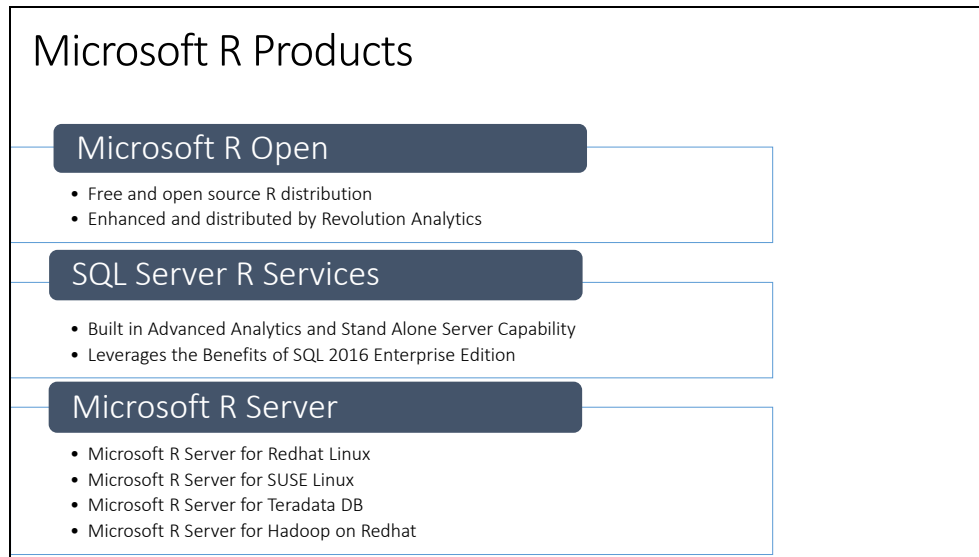
- Optional: For SQL, Open this site, complete the lessons the instructor assigns: <http://www.w3schools.com/SQL/default.asp>

Module 2: The Microsoft R Platform



12

- Primary Microsoft R Site: <https://msdn.microsoft.com/en-us/microsoft-r/index>



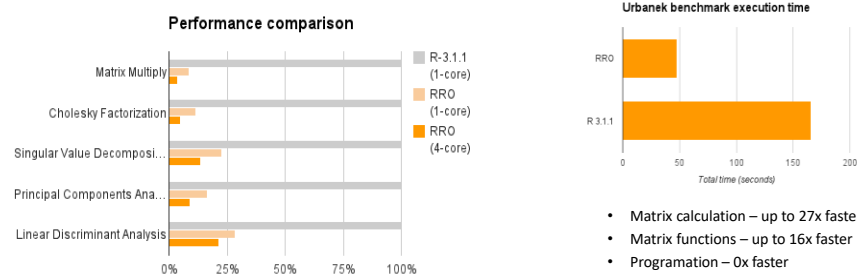
- Channel 9 videos on Microsoft R:
<https://channel9.msdn.com/Search?term=Microsoft%20R#lang-en=en&ch9Search>

Microsoft R Open

- **Enhanced Open Source R distribution**
 - Based on the latest Open Source R (3.2.4 (5))
 - Built, tested and distributed by Microsoft
 - Enhanced by Intel MKL Library to speed up linear algebra functions
- **Compatible with all R-related software**
 - CRAN packages, RStudio, third-party R integrations, ...
- **Revolutions Open-Source R packages**
 - Reproducible R Toolkit – checkpoint
- **MRAN website mrان.revolutionanalytics.com**
 - Enhanced documentation and learning resources
 - Discover 7500 free add-on R packages
- **Open source (GPLv2 license) - 100% free to download, use and share**

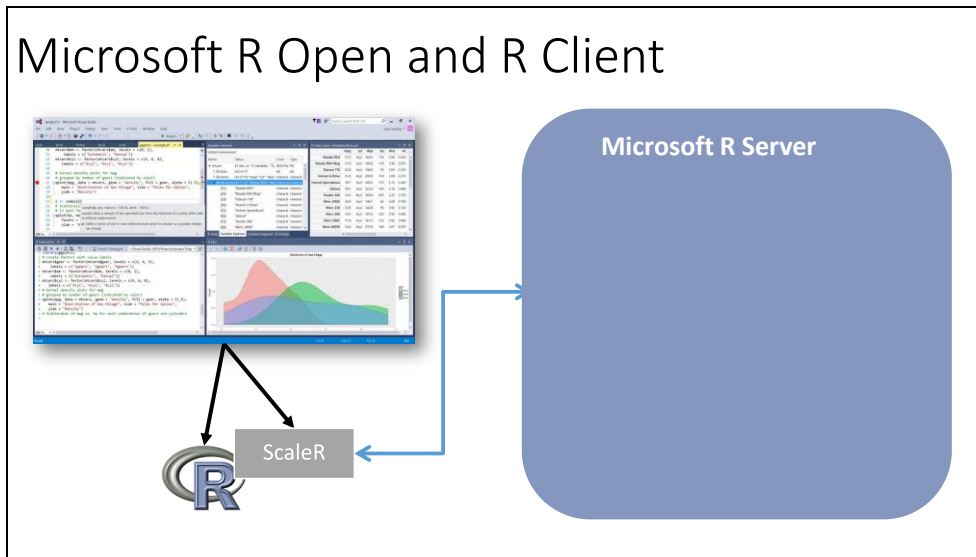
- Quick Video on R Client: <https://channel9.msdn.com/blogs/MicrosoftR/Microsoft-Introduces-new-free-Microsoft-R-Client>

CRAN R compared to Microsoft R Open



- More efficient and multi-threaded math computation.
- Benefits math intensive processing.
- No benefit to program logic and data transform

- Overview: <https://channel9.msdn.com/Series/Microsoft-R-Server-Series/Introduction-to-Microsoft-R-Server-Session-1--Overview>



- Book and Series: <http://dacrook.com/introduction-to-microsoft-r-open/>
- Microsoft R Client: <https://msdn.microsoft.com/en-us/microsoft-r/index#mrc>

Microsoft R Components

- Microsoft R Open
 - Microsoft R Client
 - Microsoft R Server
 - HDInsight SparkR / SQL Server R Services
-
- R in Azure Machine Learning

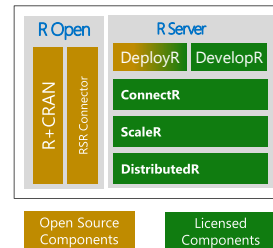
- Supported Platforms for Microsoft R Server: <https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-supported-platforms>
- Book and Series: <http://dacrook.com/introduction-to-microsoft-r-open/>
- Microsoft R Client: <https://msdn.microsoft.com/en-us/microsoft-r/index#mrc>
- Microsoft R Server: <https://msdn.microsoft.com/en-us/microsoft-r/index#mrs>
- SQL Server R Services: <https://msdn.microsoft.com/en-us/microsoft-r/index#sqlr>
- HDInsight SparkR: <https://azure.microsoft.com/en-gb/services/hdinsight/apache-spark/>

Microsoft R Server

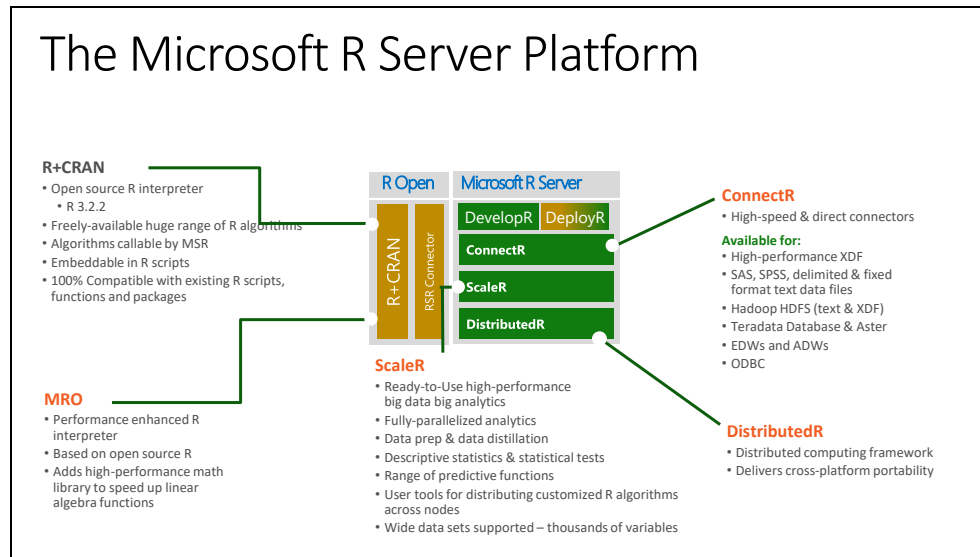
Microsoft R Server is a broadly deployable enterprise-class analytics platform based on R that is supported, scalable and secure. Supporting a variety of big data statistics, predictive modeling and machine learning capabilities, R Server supports the full range of analytics – exploration, analysis, visualization and modeling

High-performance open source R plus:

- Data source connectivity to big-data objects
- Big-data advanced analytics
- Multi-platform environment support
- Inpredictive modeling
- Development and production environment support
 - IDE for data scientist developers
 - Secure, Scalable R Deployment




Microsoft R Server: <https://msdn.microsoft.com/en-us/microsoft-r/index#mrs>



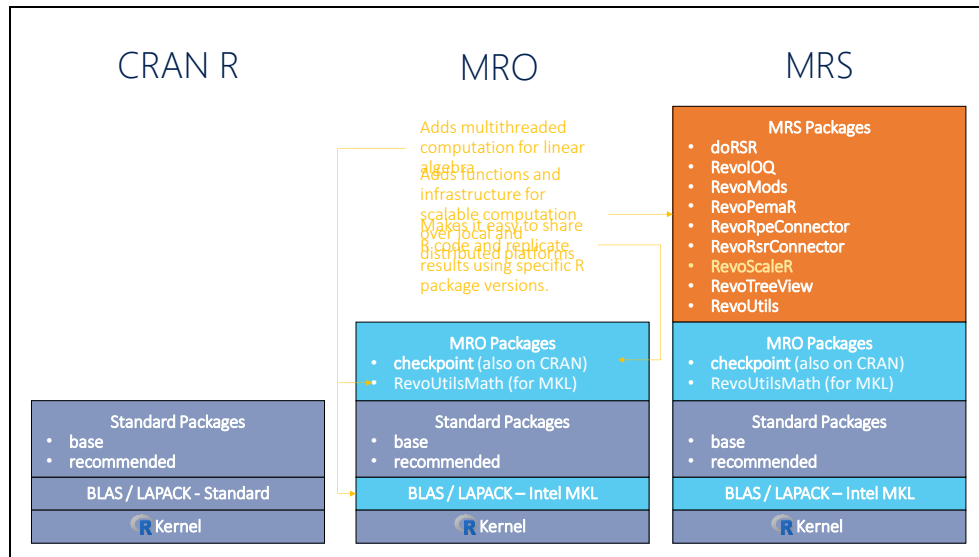
- Installing on Linux: <https://channel9.msdn.com/Series/Microsoft-R-Server/Microsoft-R-Server-Installation-Linux>

CRAN, MRO, MRS Comparison

		Microsoft R Open	Microsoft R Server
Datasize	In-memory	In-memory	In-Memory or Disk Based
Speed of Analysis	Single threaded	Multi-threaded	Multi-threaded, parallel processing 1:N servers
Support	Community	Community	Community + Commercial
Analytic Breadth & Depth	7500+ innovative analytic packages	7500+ innovative analytic packages	7500+ innovative packages + commercial parallel high-speed functions
License	Open Source	Open Source	Commercial license. Supported release with indemnity

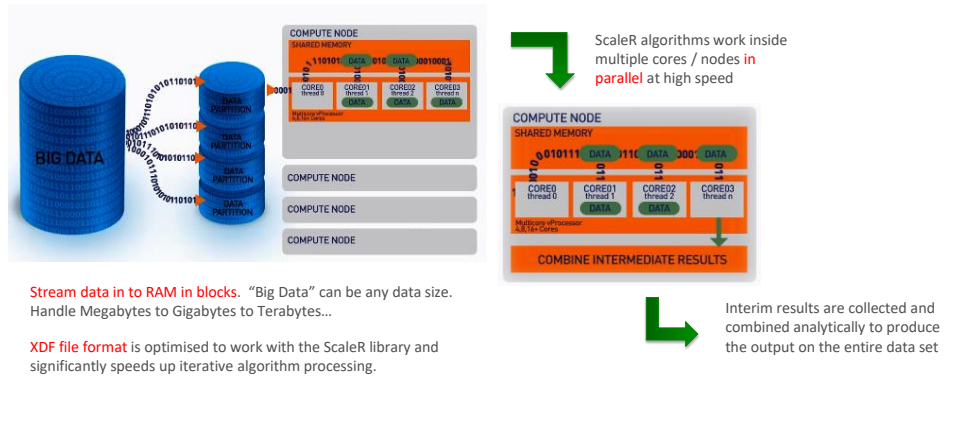
- Technology Overview: <https://channel9.msdn.com/Series/Microsoft-R-Server/Technology-Overview-for-Microsoft-R-Server-2016>

Slide 21




- Getting Started: <https://msdn.microsoft.com/en-us/microsoft-r/?f=255&MSPPError=-2147217396>

ScaleR – Parallel + “Big Data”



- Function Breakdown: <https://msdn.microsoft.com/en-us/microsoft-r/scaler/scaler>

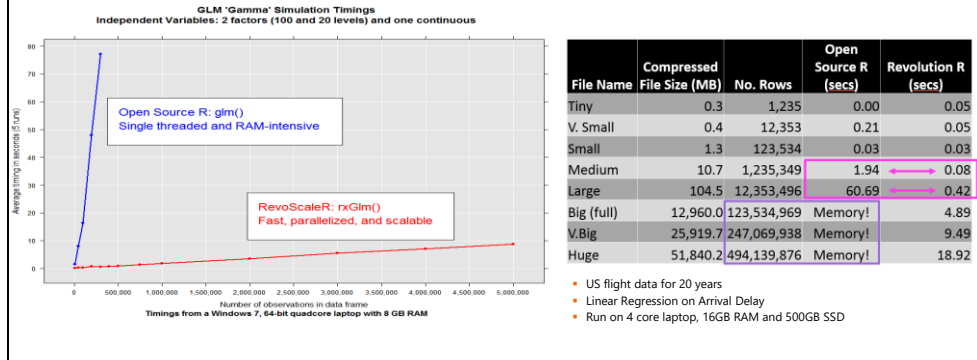
Scale R – Parallelized Algorithms & Functions

<p>Data Preparation</p> <ul style="list-style-type: none"> ▪ Data Import – Delimited, Fixed, SAS, SPSS, ODBC ▪ Variable creation & transformation ▪ Recode variables ▪ Factor variables ▪ Missing value handling ▪ Sort, Merge, Split ▪ Aggregate by category (means, sums) <p>Descriptive Statistics</p> <ul style="list-style-type: none"> ▪ Min / Max, Mean, Median (approx.) ▪ Quantiles (approx.) ▪ Standard Deviation ▪ Variance ▪ Correlation ▪ Covariance ▪ Sum of Squares (cross product matrix for set variables) ▪ Pairwise Cross tabs ▪ Risk Ratio & Odds Ratio ▪ Cross-Tabulation of Data (standard tables & long form) ▪ Marginal Summaries of Cross Tabulations 	<p>Statistical Tests</p> <ul style="list-style-type: none"> ▪ Chi Square Test ▪ Kendall Rank Correlation ▪ Fisher's Exact Test ▪ Student's t-Test <p>Sampling</p> <ul style="list-style-type: none"> ▪ Subsample (observations & variables) ▪ Random Sampling <p>Predictive Models</p> <ul style="list-style-type: none"> ▪ Sum of Squares (cross product matrix for set variables) ▪ Multiple Linear Regression ▪ Generalized Linear Models (GLM) exponential family distributions: binomial, Gaussian, inverse Gaussian, Poisson, Tweedie. Standard link functions: cauchit, identity, log, logit, probit. User defined distributions & link functions. ▪ Covariance & Correlation Matrices ▪ Logistic Regression ▪ Classification & Regression Trees ▪ Predictions/scoring for models ▪ Residuals for all models 	<p>Variable Selection</p> <ul style="list-style-type: none"> ▪ Stepwise Regression <p>Simulation</p> <ul style="list-style-type: none"> ▪ Simulation (e.g. Monte Carlo) ▪ Parallel Random Number Generation <p>Cluster Analysis</p> <ul style="list-style-type: none"> ▪ K-Means <p>Classification</p> <ul style="list-style-type: none"> ▪ Decision Trees ▪ Decision Forests ▪ Gradient Boosted Decision Trees ▪ Naive Bayes <p> Combination</p> <ul style="list-style-type: none"> ▪ rxDataStep ▪ rxExec ▪ PEMR-R API Custom Algorithms
--	--	--

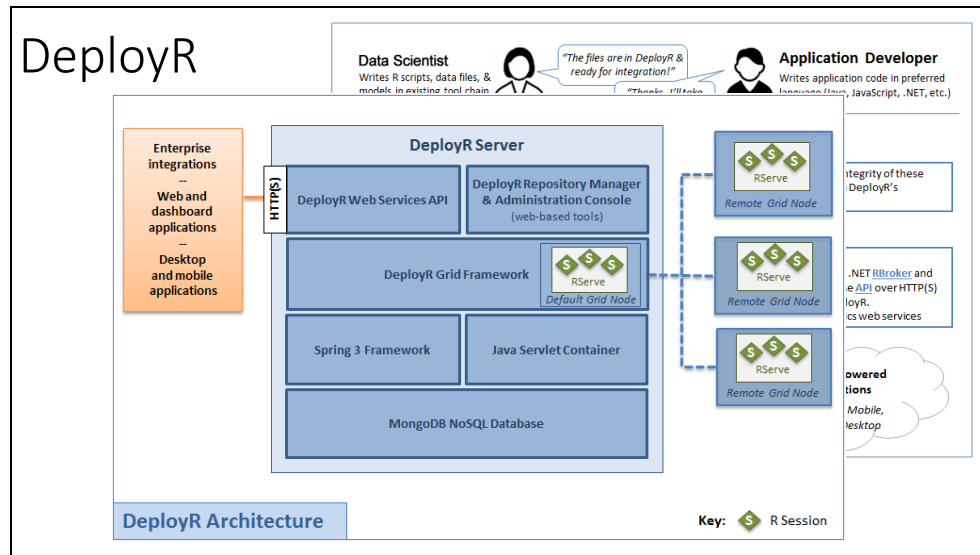
- SQL Server Implementation of ScaleR Functions: <https://msdn.microsoft.com/en-us/library/mt652103.aspx>

ScaleR - Performance comparison

Microsoft R Server has no data size limits in relation to size of available RAM. When open source R operates on data sets that exceed RAM it will fail. In contrast Microsoft R Server scales linearly well beyond RAM limits and parallel algorithms are much faster.

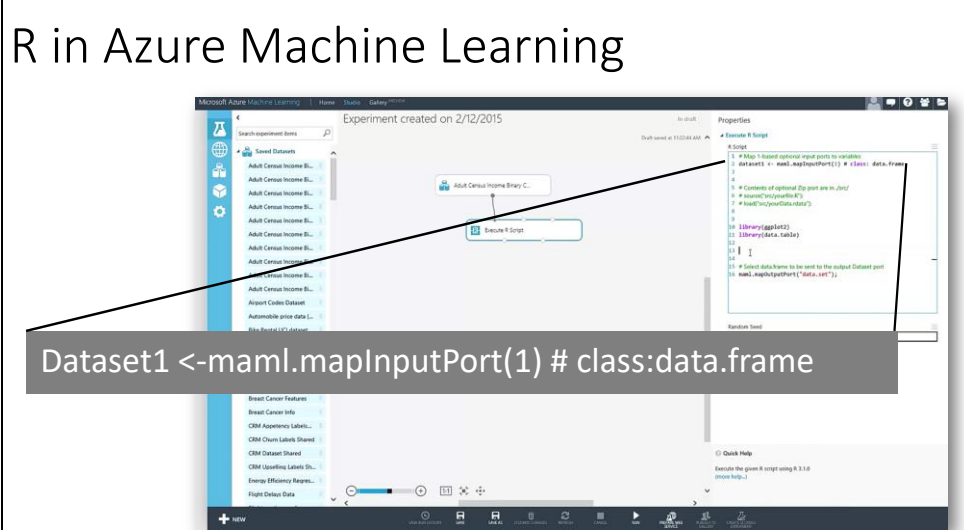


- ScaleR Functions for Working with SQL Server Data: <https://msdn.microsoft.com/en-us/library/mt732681.aspx>



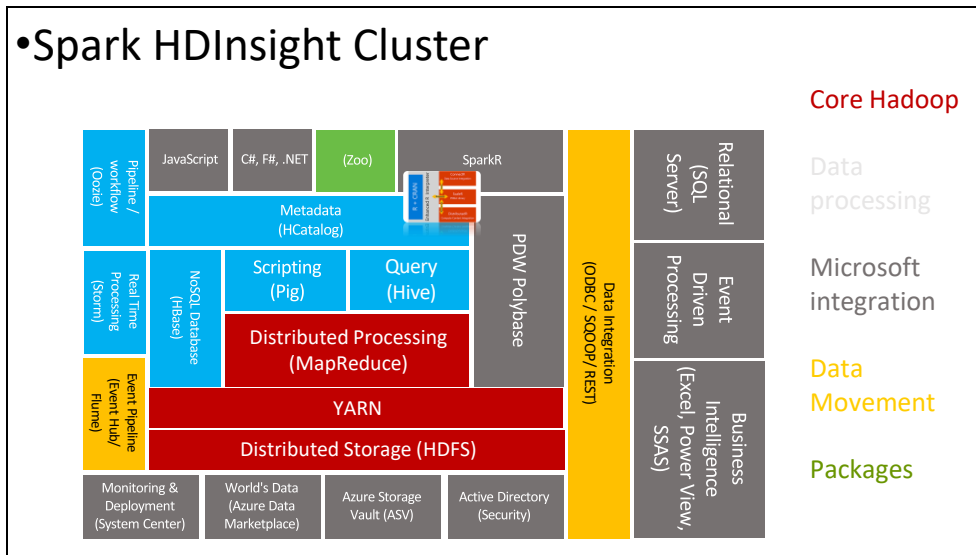
- Microsoft DeployR Documentation: <https://msdn.microsoft.com/en-us/microsoft-r/deployr-about>
- Previous Documentation: <https://deployr.revolutionanalytics.com/documents/getting-started/about/>

R in Azure Machine Learning

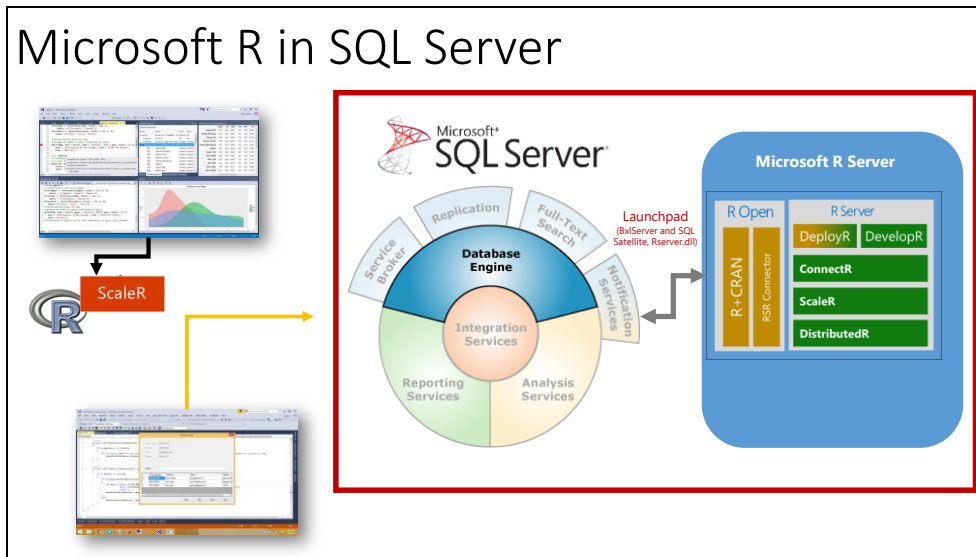


```
Dataset1 <- maml.mapInputPort(1) # class:data.frame
```

- Primary reference: <https://msdn.microsoft.com/en-us/library/dn905952.aspx>
- Using R in Azure Machine Learning: <https://azure.microsoft.com/en-us/documentation/articles/machine-learning-r-quickstart/>
- Overview Video: <https://channel9.msdn.com/Blogs/Windows-Azure/R-in-Azure-ML-Studio>
- R Packages supported: <https://msdn.microsoft.com/en-us/library/mt741980.aspx>

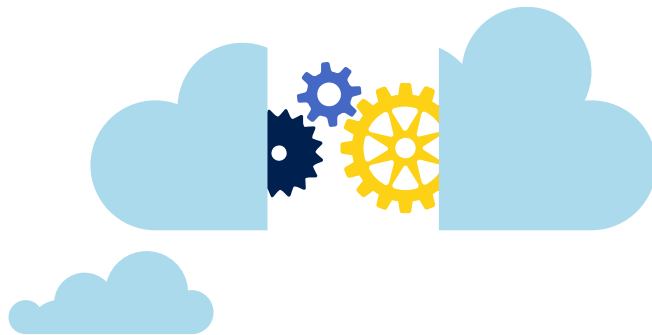


- Full training example for the local HDP Instance: <http://hortonworks.com/hadoop-tutorial/hello-world-an-introduction-to-hadoop-hcatalog-hive-and-pig/>
- More detail on the Hadoop Components: <http://www.datasciencecentral.com/profiles/blogs/hadoop-herd-when-to-use-what>



- Primary Documentation and training: <https://msdn.microsoft.com/en-us/library/mt604845.aspx>
- Great set of resources: <https://www.r-bloggers.com/r-and-sql-server-articles/amp/>

Module 3: R Client Options



29

- The Microsoft R Client: <https://msdn.microsoft.com/en-us/microsoft-r/install-r-client-windows>

Microsoft R Development Tools

- Microsoft R Client
- RStudio
- R Tools for Visual Studio (RTVS)
- SQL Server tools

- Installing Microsoft R Client on Windows: <https://msdn.microsoft.com/en-us/microsoft-r/install-r-client-windows>
- Files located at: C:\Program Files\Microsoft\R Client\R_SERVER\bin
- Learn more about RTVS and SQL Server: <https://microsoft.github.io/RTVS-docs/sqlserver.html>

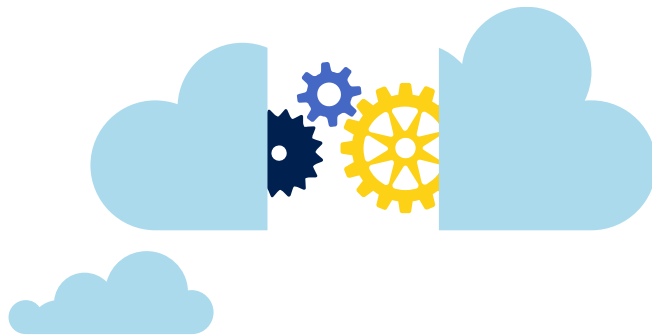
Lab:

**Install and configure
Client Environments**

Microsoft Azure

- *Optional:* Install Visual Studio (<https://www.visualstudio.com/downloads/download-visual-studio-vs>) (Select Optional, and select SQL Server Data Tools)
- *Optional:* Install RTVS (<http://microsoft.github.io/RTVS-docs/installer.html>)
- *Optional:* Install Rstudio (<https://www.rstudio.com/products/rstudio/download2/>)
- You may need to update the version of R the Visual Studio is using by going to **R Tools -> Options** (to C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\R_SERVICES)
- Connect to R in Visual Studio or Rstudio or Command line (C:\Program Files\Microsoft SQL Server\MSSQL13.MSSQLSERVER\R_SERVICES\bin>R.exe) , and Run **Revo.version** to ascertain MRS running

Module 4: Operationalize Microsoft R

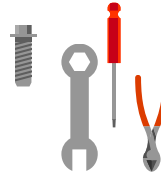


32

- Complete introduction: <https://msdn.microsoft.com/en-us/microsoft-r/microsoft-r-getting-started>
- Data Exploration and Modeling with R: <https://msdn.microsoft.com/en-us/library/mt590947.aspx>

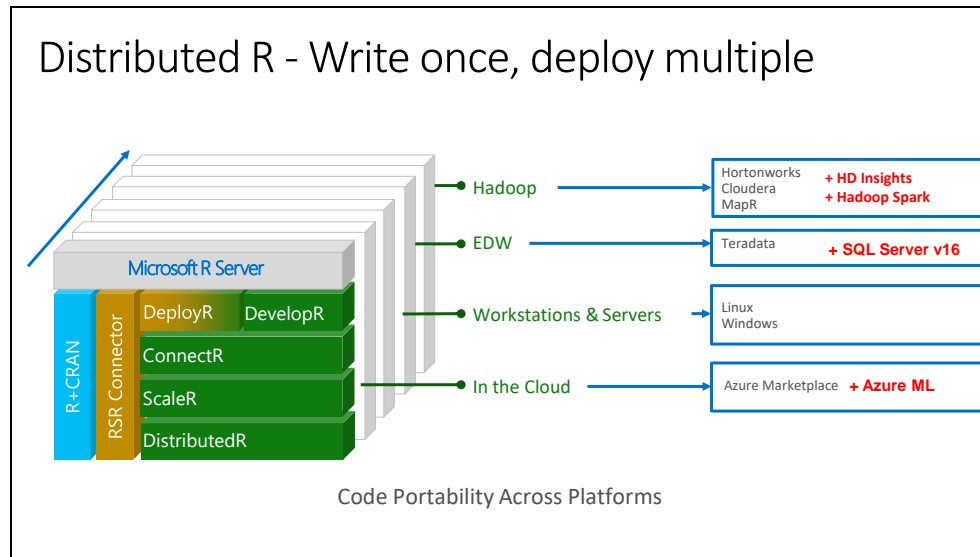
Configuration and Operation

- Planning
 - Specific Environments
- File Locations
- Services and Background Processes
- Package Management
- DeployR Planning

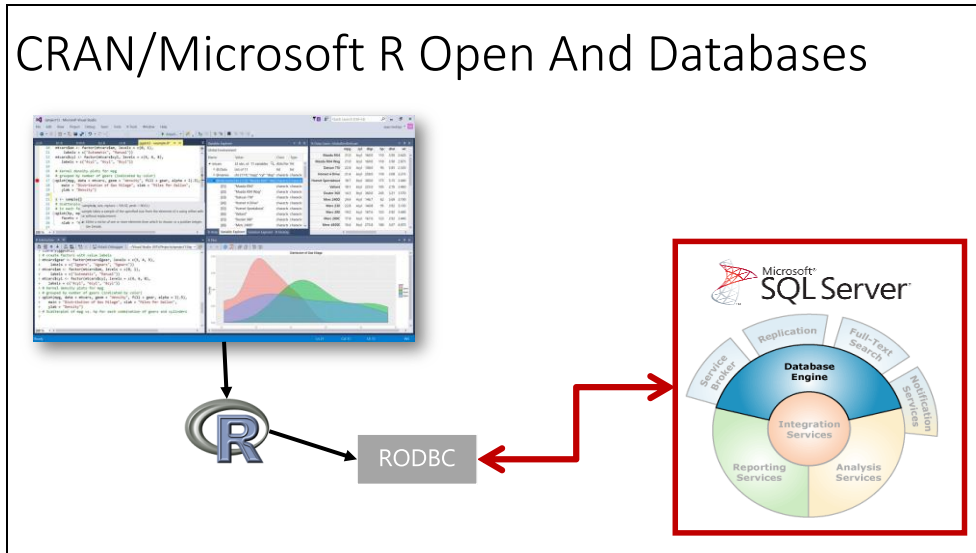


33

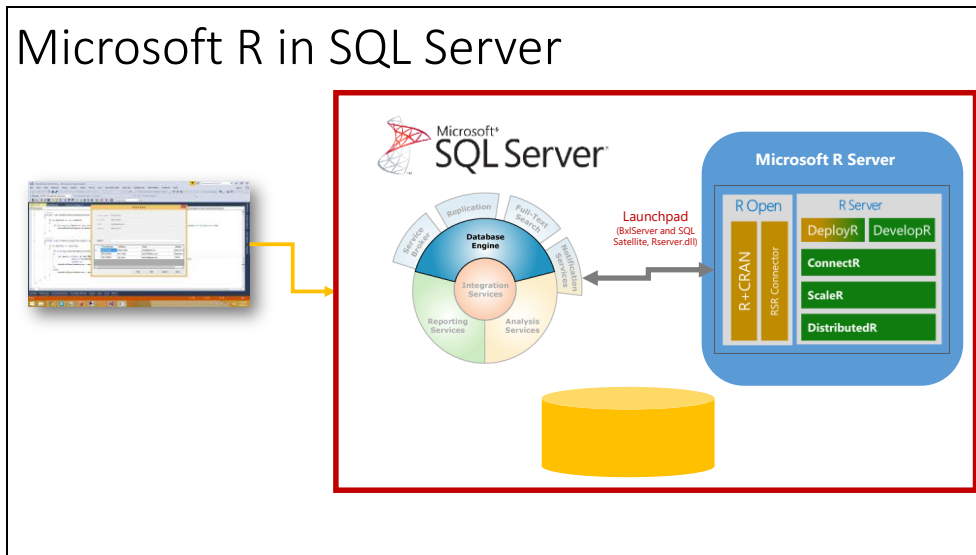
- Features and Tasks: <https://msdn.microsoft.com/en-us/library/mt590811.aspx>
- Differences in Features: <https://msdn.microsoft.com/en-us/library/mt721284.aspx>
- Installing on VM's: <https://msdn.microsoft.com/en-us/library/mt748179.aspx>
- Setting up R Services: <https://msdn.microsoft.com/en-us/library/mt696069.aspx>



- DeployR Workflow: <https://msdn.microsoft.com/en-us/microsoft-r/deployr-about>



- Book and Series: <http://dacrook.com/introduction-to-microsoft-r-open/>
- Microsoft R Client: <https://msdn.microsoft.com/en-us/microsoft-r/index#mrc>



- Primary Documentation and training: <https://msdn.microsoft.com/en-us/library/mt604845.aspx>

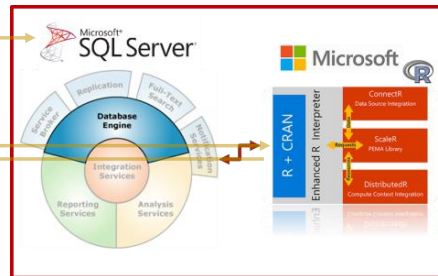
T-SQL and R Interaction

```
EXEC sp_execute_external_script
@language =N'R',

-- SQL Part (sends to @script)
@input_data_1 =N 'SELECT 1 as Installed',

-- R Part (gets @input_data_1)
@script=N'OutputDataSet<-InputDataSet'

WITH RESULT SETS
(((Installed] int not null));
GO
```

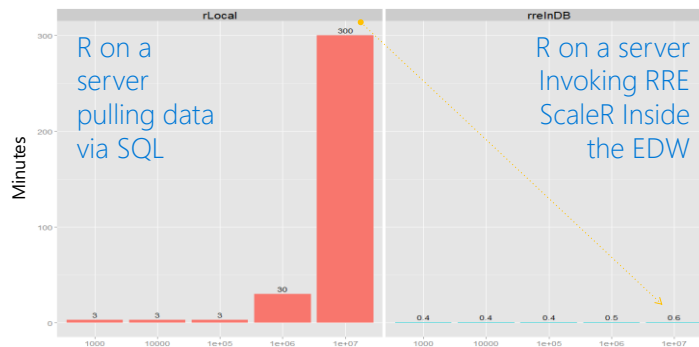


37

- Components and Architecture: <https://msdn.microsoft.com/en-us/library/mt709082.aspx> (with graphics)

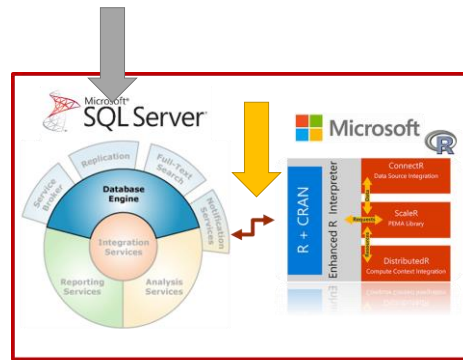
Example of In-Database Acceleration

5+ hours to 40 seconds:



T-SQL and R Interaction

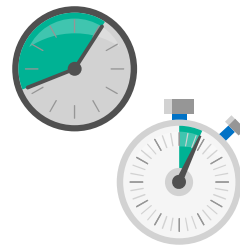
1. T-SQL Code
 1. `SELECT` data
2. `sp_execute_external_script`
 1. Launchpad (*BxlServer* and *SQL Satellite, Rserver.dll*)
3. R Data or Object Returns



- Components and Architecture: <https://msdn.microsoft.com/en-us/library/mt709082.aspx> (with graphics)

Performance and Monitoring

- Performance considerations
- Monitoring
- Tuning



- Extended Events for SQL Server R Services: <https://msdn.microsoft.com/en-us/library/mt628054.aspx>

Security and Governance

- Principals
- Securables



- Security Overview: <https://msdn.microsoft.com/en-us/library/mt709078.aspx>

Implementation Considerations



- Coordinating with the R professional
- Best Practices



42

- Managing and monitoring R Solutions for SQL Server: <https://msdn.microsoft.com/en-us/library/mt590866.aspx>
- Upgrade and Installation: <https://msdn.microsoft.com/en-us/library/mt653951.aspx>
- Considerations: <https://msdn.microsoft.com/en-us/library/mt590540.aspx>

Lab:

More labs on ScaleR

Step 2: Reading the data

Step 3: Preparing the data

1. Navigate to <https://rheartpython.github.io/cisw/> and the MRS labs and go through the Step 2 and 3 for the "*Analyzing Big Data with MRS Labs - Scenario around NYC Taxi data*" section.

Module 5: Creating a Microsoft R Solution



<https://mva.microsoft.com/ebooks/>


44

- Complete introduction: <https://msdn.microsoft.com/en-us/microsoft-r/microsoft-r-getting-started>

Lab:

Comprehensive Example using
SQL Server

- Option 1: Follow the instructions at the MRS set of labs at <https://rheartpython.github.io/cisw/intro-mrs-labs/> for “*SQL Server R Services - Scenario around Hospital Length of Stay*” or simply go to <https://microsoft.github.io/r-server-hospital-length-of-stay/> and follow the instructions for the **Typical Workflow**
- Option 2 Refer to this link: <https://gallery.cortanaintelligence.com/Tutorial/Predictive-Maintenance-Template-with-SQL-Server-R-Services-1> and work through that example.
- Demand Forecasting Template: <https://channel9.msdn.com/Blogs/Seth-Juarez/Energy-Demand-Forecasting-Template-with-SQL-Server-R-Services>
- More labs: <https://github.com/Microsoft/SQL-Server-R-Services-Samples> and <https://gallery.cortanaintelligence.com/Collection/ML-Templates-with-SQL-Server-R-Services-1>



Microsoft

1. Understand the R Language and where it is used
2. Understand the Microsoft R Platform and its capabilities
3. Set up and use the server and various client tools for a R environment
4. Know how to operationalize a SQL Server R Services environment
5. Use the Microsoft R capabilities in a solution

© 2015 Microsoft Corporation. All rights reserved.

Questions?

More resources:

<https://msdn.microsoft.com/en-us/microsoft-r/microsoft-r-more-resources>

[Revolutions Blog](#)

Microsoft R and SQL Server: <https://www.r-bloggers.com/r-and-sql-server-articles/amp/>
Blog: Joseph Sirosh, "Making R the Enterprise Standard..."

[Getting Started with Microsoft R](#)

[Diving In.. Data Analysis in Microsoft R](#)

[R Server Technology – Video](#)

[R Tools for Visual Studio Sneak Peek](#)

[R Tools for Visual Studio Overview](#)

[SQL R Services Overview – Youtube](#)

[SQL R Services Feature Overview - Youtube](#)

[SQL R Services Overview at Build](#)

[SQL R Services Tutorial](#)

Addendum:

Install R Platforms

- Read the installation page for MRS – <https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-supported-platforms>
- *As assigned:* Install Microsoft R Client – <https://msdn.microsoft.com/en-us/microsoft-r/install-r-client-windows>
- *As assigned:* Install MRS on Windows – <https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-windows?f=255&MSPPErr=-2147217396>
- *As assigned:* Install MRS on Linux – note: MSDN account required: <https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-linux-server?f=255&MSPPErr=-2147217396>
- *As assigned:* Install SQL Server 2016 and ensure you select R Services – see this link: <https://www.microsoft.com/en-us/cloud-platform/sql-server-editions-developers>

- Find out if MRS is loaded: `sessionInfo()`