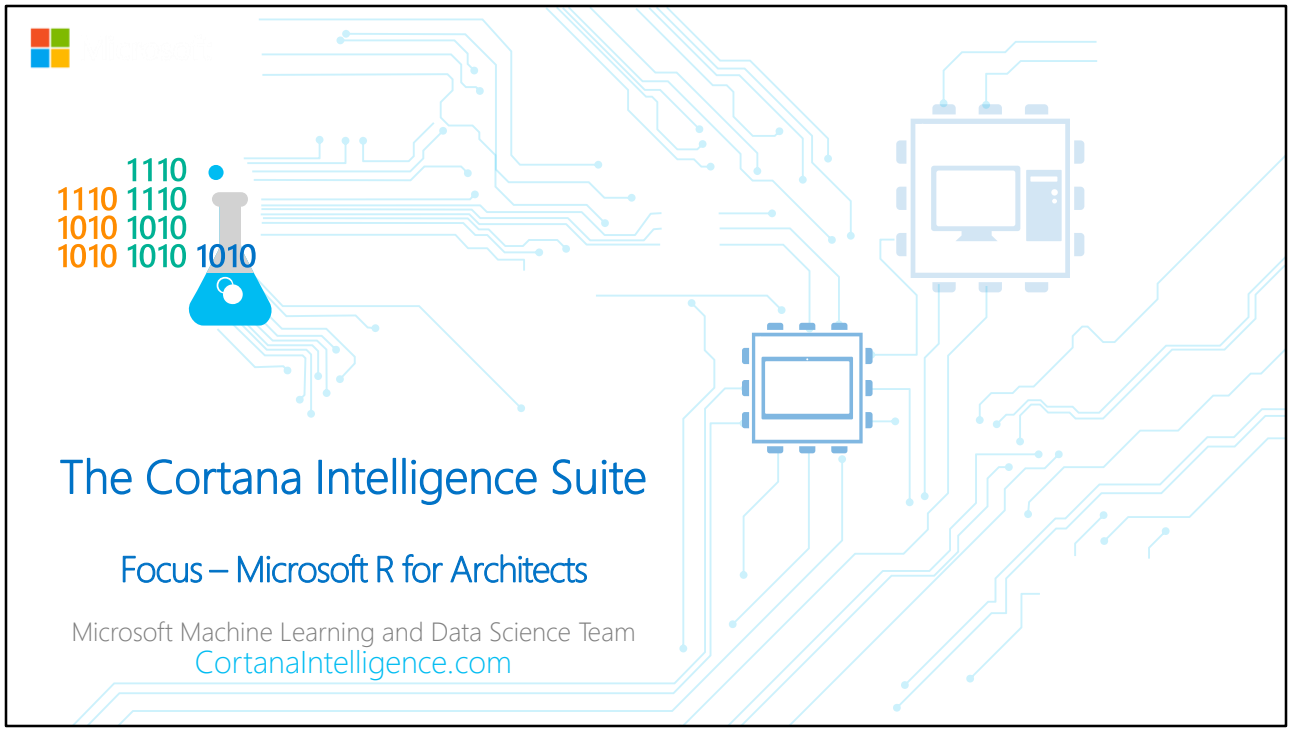


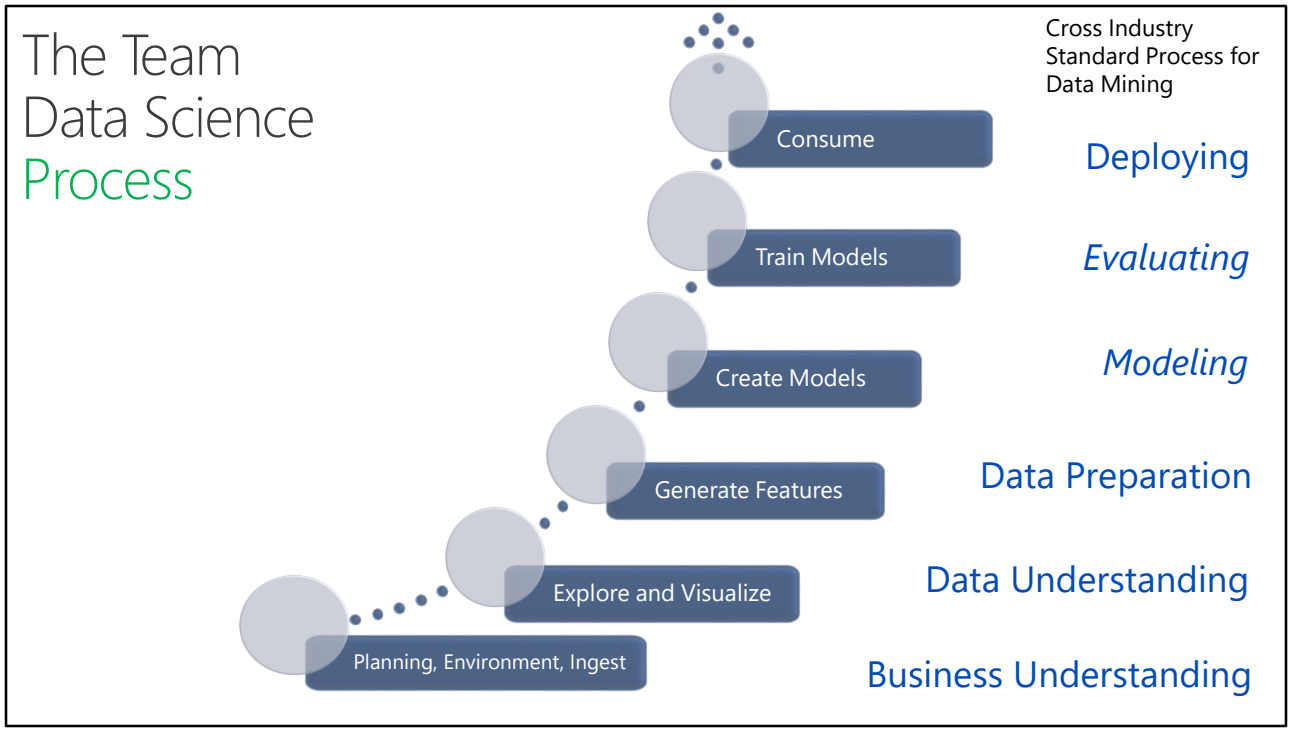
1. Main page: http://cortanaanalytics.com
2. To begin this module, you should have:
    1. Basic Math and Stats skills
    2. Business and Domain Awareness
    3. General Computing Background

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

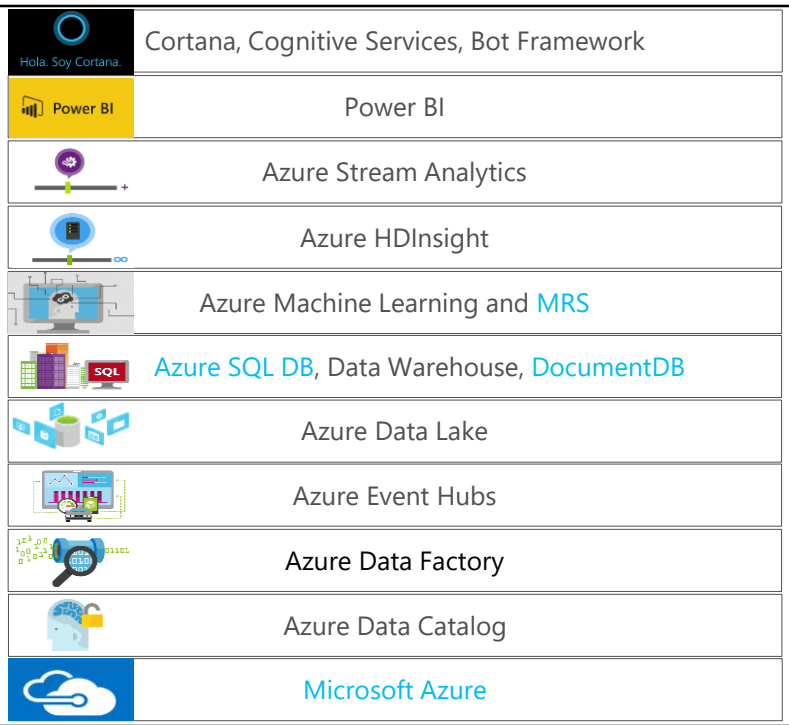
1. Understand the R Language and where it is used
2. Understand the Microsoft R Platform and its capabilities
3. Set up and use the server and various client tools for a R environment
4. Know how to operationalize a SQL Server R Services environment
5. Use the Microsoft R capabilities in a solution

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



1. This process largely follows the CRISP-DM model: http://www.sv-europe.com/crisp-dm-methodology/
2. It also references the Cortana Intelligence process: https://azure.microsoft.com/en-us/documentation/articles/data-science-process-overview/
3. A complete process diagram is here: https://azure.microsoft.com/en-us/documentation/learning-paths/cortana-analytics-process/
4. Some walkthrough's of the various services: https://azure.microsoft.com/en-us/documentation/articles/data-science-process-walkthroughs/

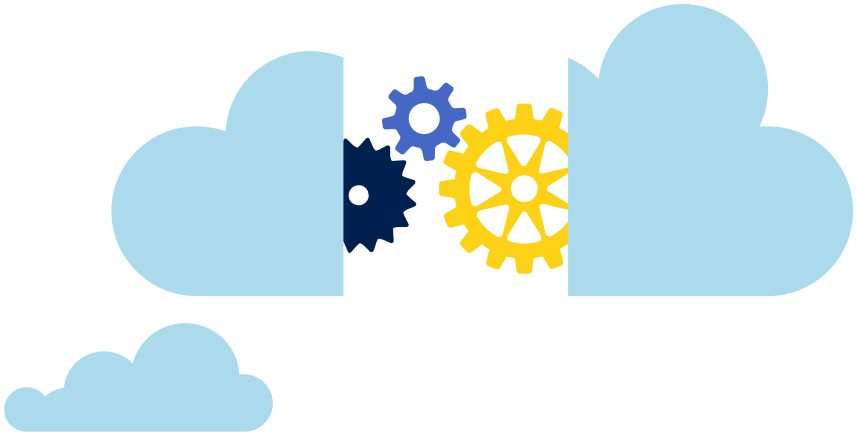
The Cortana Intelligence Platform



1. Platform and Storage: Microsoft Azure – http://microsoftazure.com Storage: https://azure.microsoft.com/en-us/documentation/services/storage/ (Host It)
2. Azure Data Catalog: http://azure.microsoft.com/en-us/services/data-catalog (Doc It)
3. Azure Data Factory: http://azure.microsoft.com/en-us/services/data-factory/ (Move It)
4. Azure Event Hubs: http://azure.microsoft.com/en-us/services/event-hubs/ (Bring It)
5. Azure Data Lake: http://azure.microsoft.com/en-us/campaigns/data-lake/ (Store It)
6. 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)
7. Azure Machine Learning: http://azure.microsoft.com/en-us/services/machine-learning/ (Learn It)
8. Azure HDInsight: http://azure.microsoft.com/en-us/services/hdinsight/ (Scale It)
9. Azure Stream Analytics: http://azure.microsoft.com/en-us/services/stream-analytics/ (Stream It)
10. Power BI: https://powerbi.microsoft.com/ (See It)
11. 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)
12. Cognitive Services: https://www.microsoft.com/cognitive-services
13. Bot Framework: https://dev.botframework.com/
14. All of the components within the suite: https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/what-is-cortana-intelligence.aspx
15. What can I do with it? https://gallery.cortanaintelligence.com/

16. Getting Started Quickly: https://caqs.azure.net/#gallery

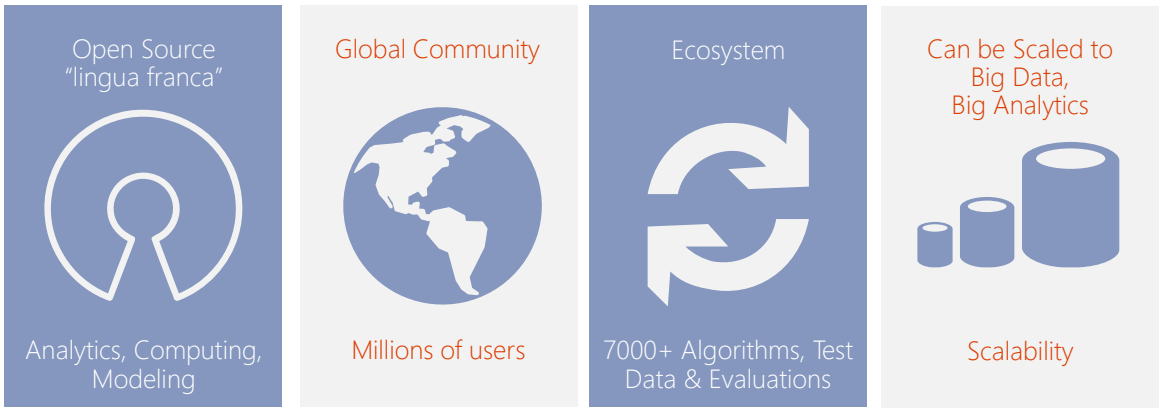
Module 1:
The R Programming Environment



5

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

R Programing and Environment



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

SQL and R Contrasted

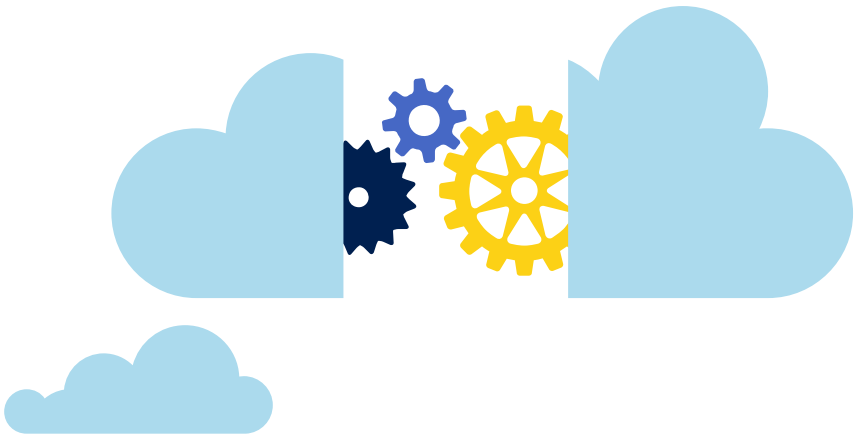
SQL	R
1. Client/Server	1. Interactive Environment
2. Database Objects	2. Data Structures
3. DML, DDL	3. Functions
4. DCL	4. Libraries (Packages)
5. Declarative Code	5. Functional Code Flow

1. Learn SQL: http://www.w3schools.com/SQl/default.asp
2. Try R, with a great interface. http://tryr.codeschool.com/levels/1/challenges/22
3. R and Statistics Intro: https://www.youtube.com/watch?v=xb5P5xdcr2U&feature=youtu.be&a
4. R Online: http://www.tutorialspoint.com/r_terminal_online.php
5. Using R to explore data: http://www.analyticsvidhya.com/blog/2015/10/cheatsheet-11-steps-data-exploration-with-codes/
6. Quick R Intro: http://www.datasciencecentral.com/m/blogpost?id=6448529%3ABlogPost%3A112754
7. Creating a recommender engine in R: http://www.analyticbridge.com/profiles/blogs/build-basic-recommendation-engine-using-r
8. Visualizations cheat-sheet in R: http://www.datasciencecentral.com/forum/topics/cheat-sheet-

data-visualization-with-r?groupUrl=tutorials

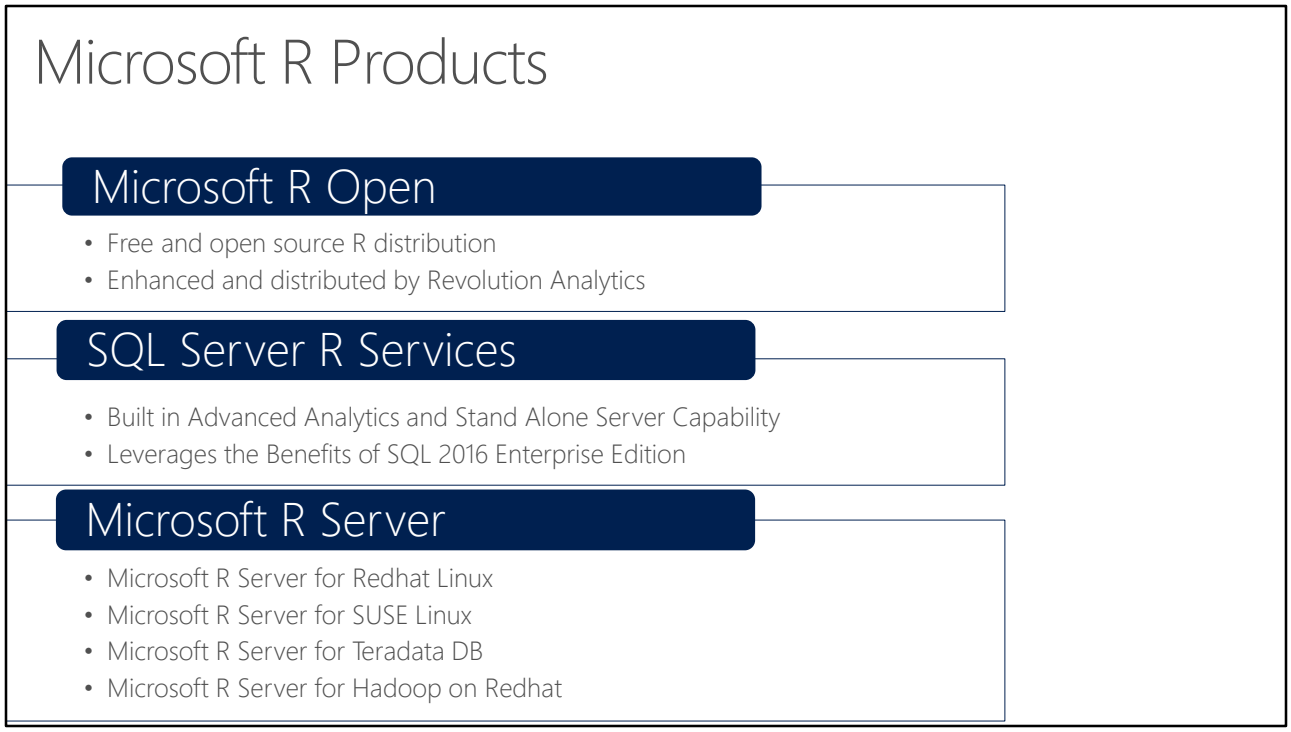
1. If you do not have a Microsoft Azure account, go here: https://azure.microsoft.com/en-us/free/ (You will need a credit card, but you will not be charged)
2. Log in to the Azure Portal: https://ms.portal.azure.com
3. Create a new Windows Data Science Virtual Machine (Size DS1_V1): https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-vm-do-ten-things/
4. For R, Open this site, complete the lessons the instructor assigns: http://tryr.codeschool.com/
5. For SQL, Open this site, complete the lessons the instructor assigns: http://www.w3schools.com/SQl/default.asp

Module 2:
The Microsoft R Platform



9

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



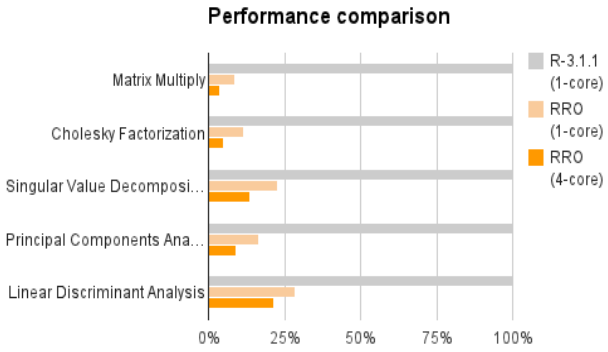
- Channel 9 videos on Microsoft R: https://channel9.msdn.com/Search?term=Microsoft%20R#lang-en=en&ch9Search

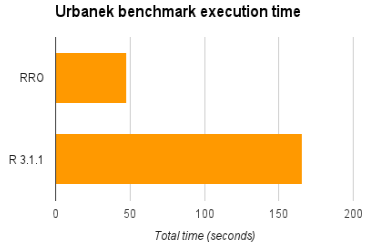
Microsoft R Open

- 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 mran.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



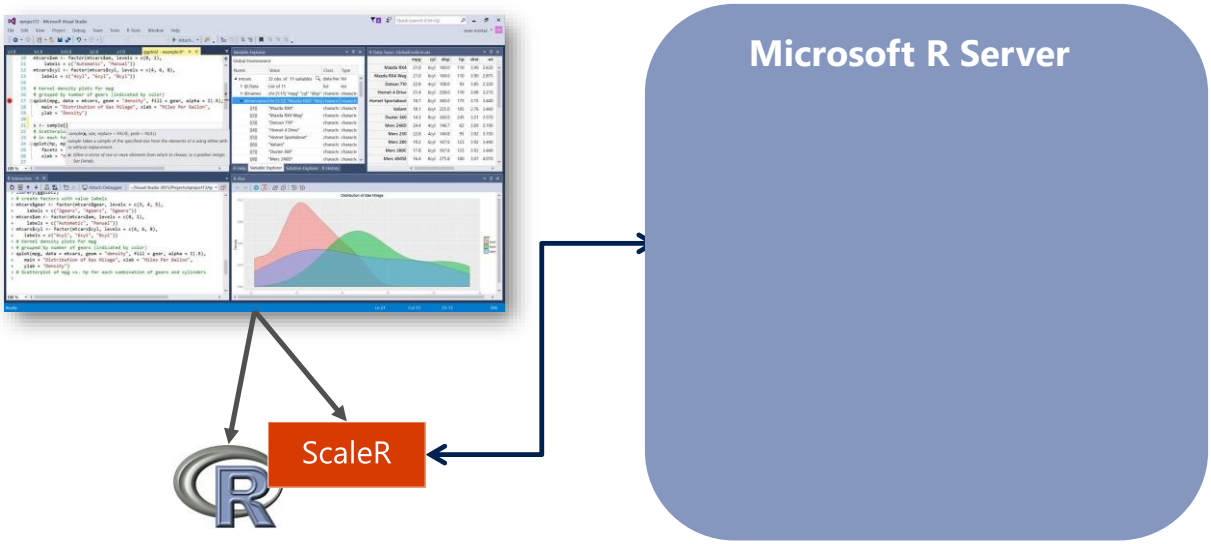


- Matrix calculation – up to 27x faster
- Matrix functions – up to 16x faster
- Programation – 0x faster

- 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

Microsoft R Open and R Client



1. Book and Series: http://dacrook.com/introduction-to-microsoft-r-open/
2. 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

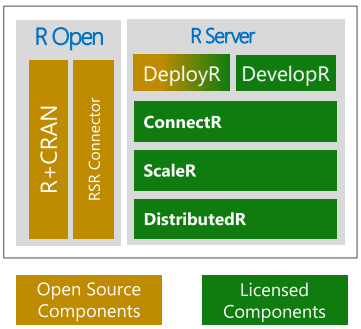
1. Supported Platforms for Microsoft R Server: https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-supported-platforms
2. Book and Series: http://dacrook.com/introduction-to-microsoft-r-open/
3. Microsoft R Client: https://msdn.microsoft.com/en-us/microsoft-r/index#mrc
4. Microsoft R Server: https://msdn.microsoft.com/en-us/microsoft-r/index#mrs
5. SQL Server R Services: https://msdn.microsoft.com/en-us/microsoft-r/index#sqlr
6. 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

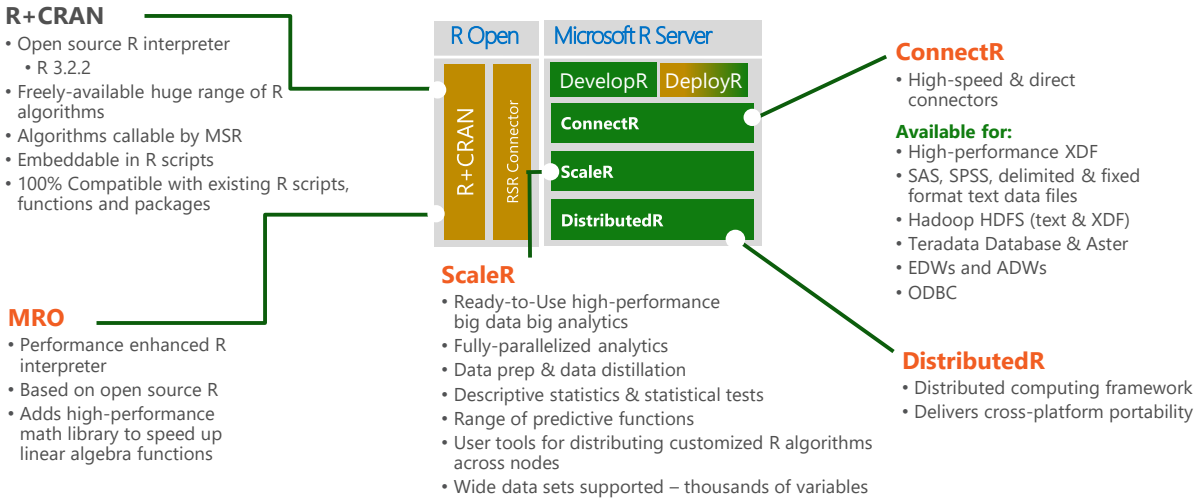
High-performance open source R plus:

- 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

The Microsoft R Server Platform

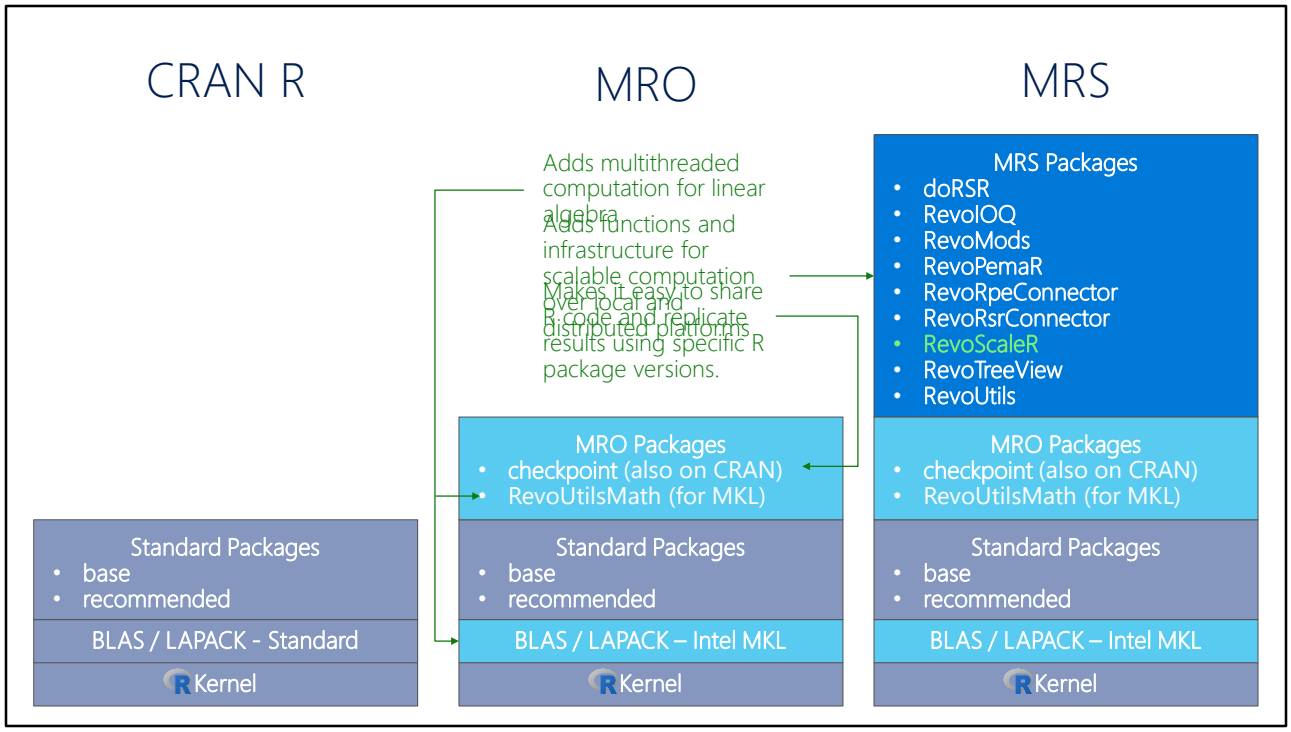


- Installing on Linux: https://channel9.msdn.com/Series/Microsoft-R-Server/Microsoft-R-Server-Installation-Linux

CRAN, MRO, MRS Comparison

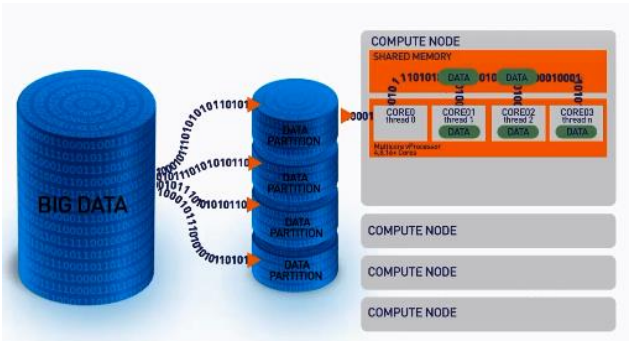
		Microsoft R Open	Microsoft R Server
Datasize	In-memory	In-memory	**In-Memory or Disk Based**
Speed of Analysis	Single threaded	Multi-threaded	**Multi-threaded, parallel processing 1:N servers**
Support	Community	Community	**Community + Commercial**
Analytic Breadth & Depth	7500+ innovative analytic packages	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



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

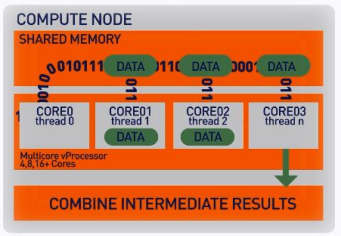
ScaleR – Parallel + "Big Data"



Stream data in to RAM in blocks. "Big Data" can be any data size. Handle Megabytes to Gigabytes to Terabytes…

XDF file format is optimised to work with the ScaleR library and significantly speeds up iterative algorithm processing.

ScaleR algorithms work inside multiple cores / nodes in parallel at high speed



Interim results are collected and combined analytically to produce the output on the entire data set

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

Scale R – Parallelized Algorithms & Functions

Data Preparation

- Data import – Delimited, Fixed, SAS, SPSS, OBDC
- Variable creation & transformation
- Recode variables
- Factor variables
- Missing value handling
- Sort, Merge, Split
- Aggregate by category (means, sums)

Descriptive Statistics

- 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

Statistical Tests

- Chi Square Test
- Kendall Rank Correlation
- Fisher's Exact Test
- Student's t-Test

Sampling

- Subsample (observations & variables)
- Random Sampling

Predictive Models

- 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

Variable Selection

- Stepwise Regression

Simulation

- Simulation (e.g. Monte Carlo)
- Parallel Random Number Generation

Cluster Analysis

- K-Means

Classification

- Decision Trees
- Decision Forests
- Gradient Boosted Decision Trees
- Naïve Bayes

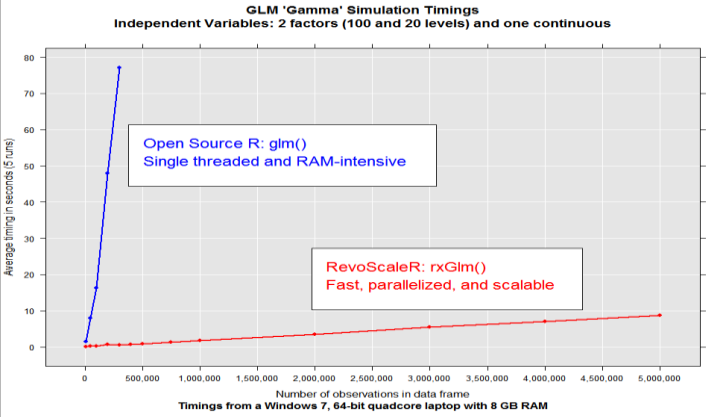
Combination

- rxDataStep
- rxExec
- PEMA-R API Custom Algorithms

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

ScaleR - Performance comparison

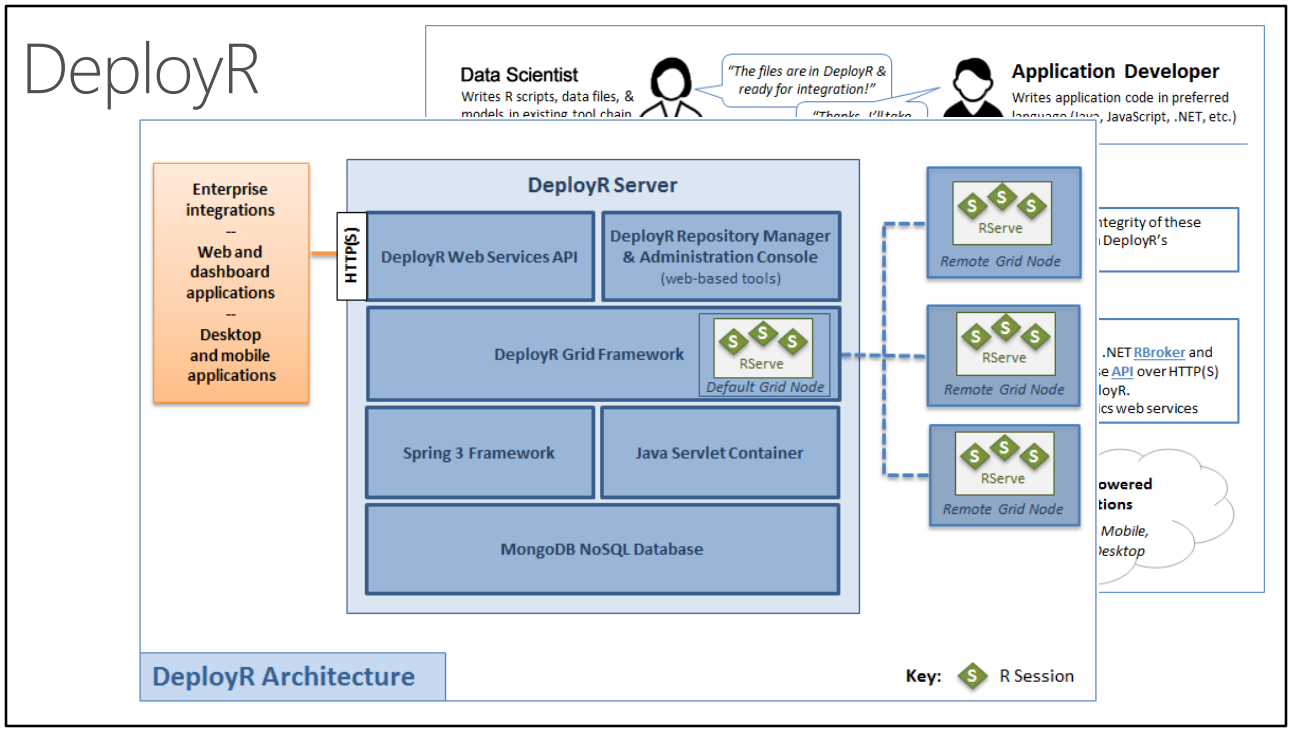
Microsoft R Server has no data size limits in relation to size of available RAM. When open source R operates on data sets that exceed RAM it will fail. In contrast Microsoft R Server scales linearly well beyond RAM limits and parallel algorithms are much faster.



File Name	Compressed File Size (MB)	No. Rows	Open Source R (secs)	Revolution R (secs)
Tiny	0.3	1,235	0.00	0.05
V. Small	0.4	12,353	0.21	0.05
Small	1.3	123,534	0.03	0.03
Medium	10.7	1,235,349	1.94	0.08
Large	104.5	12,353,496	60.69	0.42
Big (full)	12,960.0	123,534,969	Memory!	4.89
V.Big	25,919.7	247,069,938	Memory!	9.49
Huge	51,840.2	494,139,876	Memory!	18.92

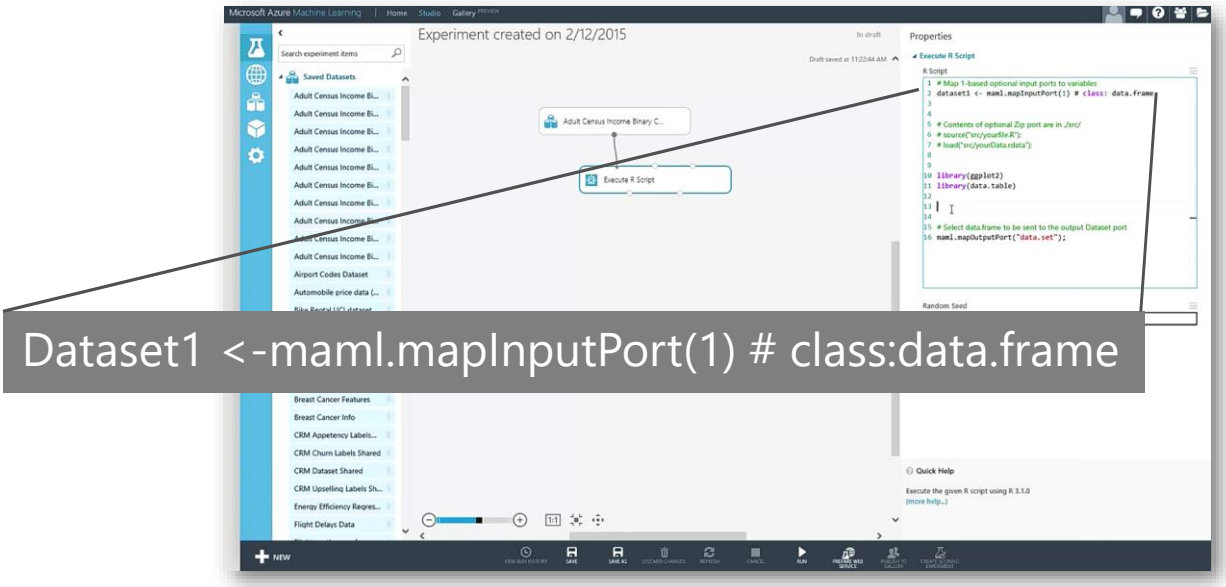
- US flight data for 20 years
- Linear Regression on Arrival Delay
- Run on 4 core laptop, 16GB RAM and 500GB SSD

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

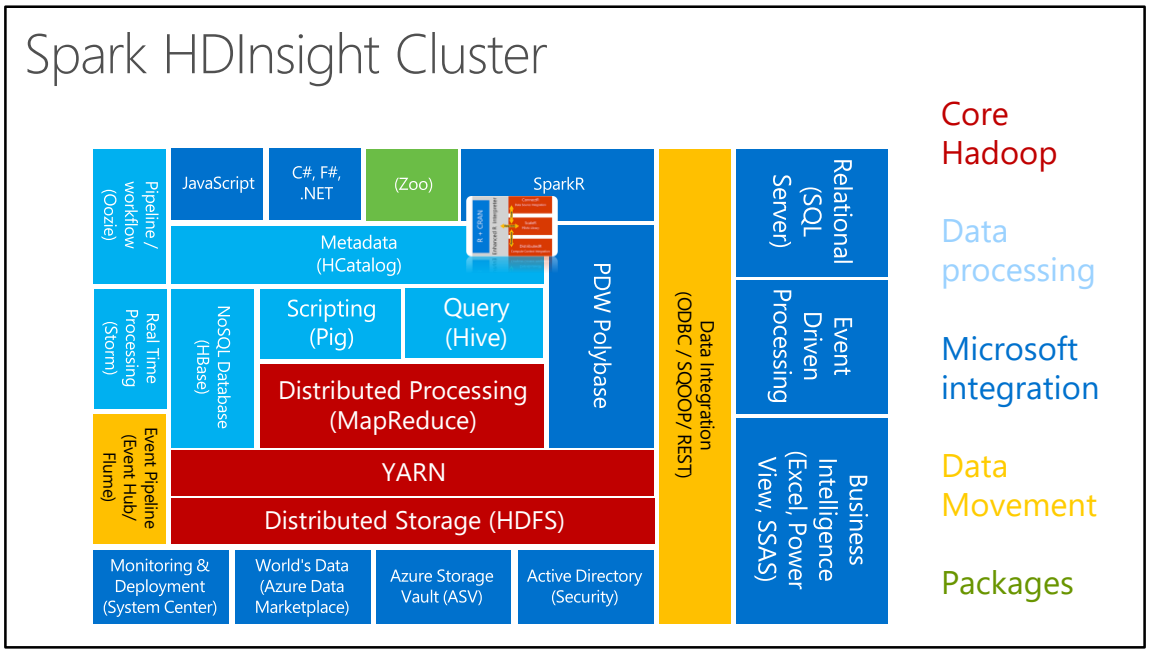


1. Microsoft DeployR Documentation: https://msdn.microsoft.com/en-us/microsoft-r/deployr-about
2. Previous Documentation: https://deployr.revolutionanalytics.com/documents/getting-started/about/

R in Azure Machine Learning

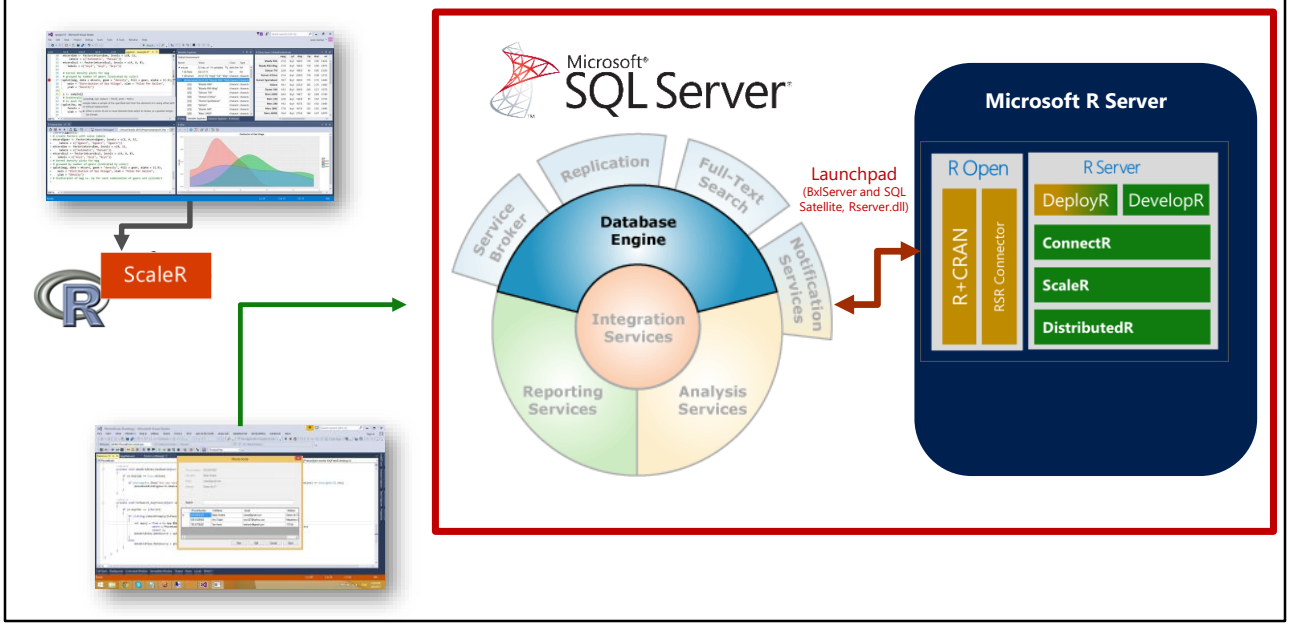


1. Primary reference: https://msdn.microsoft.com/en-us/library/dn905952.aspx
2. Using R in Azure Machine Learning: https://azure.microsoft.com/en-us/documentation/articles/machine-learning-r-quickstart/
3. Overview Video: https://channel9.msdn.com/Blogs/Windows-Azure/R-in-Azure-ML-Studio
4. R Packages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

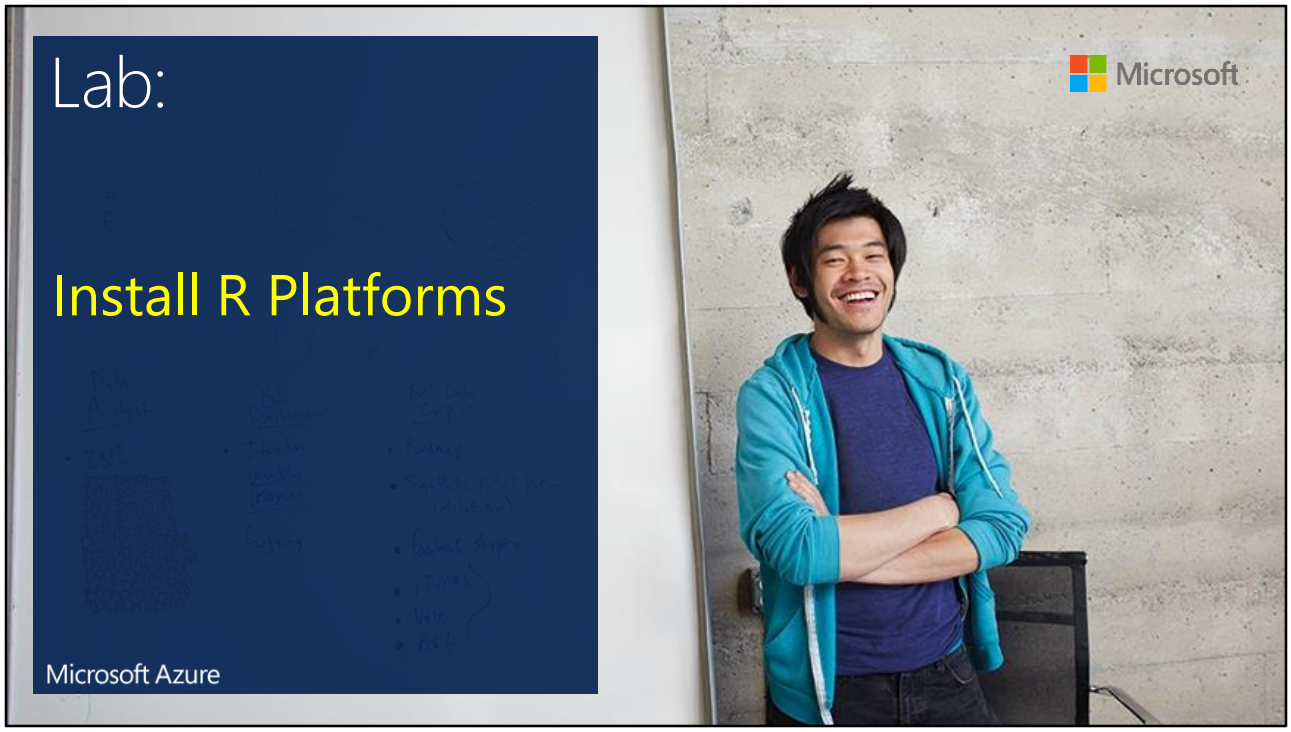


1. Full training example for the local HDP Instance: http://hortonworks.com/hadoop-tutorial/hello-world-an-introduction-to-hadoop-hcatalog-hive-and-pig/
2. More detail on the Hadoop Components: http://www.datasciencecentral.com/profiles/blogs/hadoop-herd-when-to-use-what

Microsoft R in SQL Server

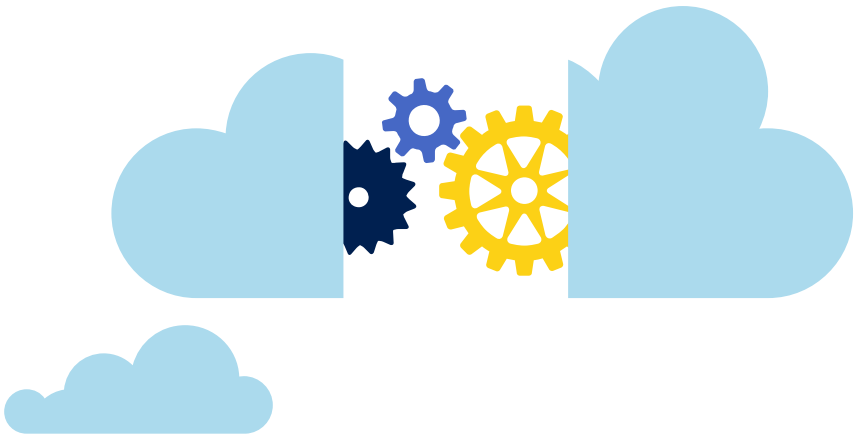


1. Primary Documentation and training: https://msdn.microsoft.com/en-us/library/mt604845.aspx



1. Read the installation page for MRS – https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-supported-platforms
2. *As assigned:* Install Microsoft R Client - https://msdn.microsoft.com/en-us/microsoft-r/install-r-client-windows
3. *As assigned:* Install MRS on Windows - https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-windows?f=255&MSPPError=-2147217396
4. *As assigned:* Install MRS on Linux – note: MSDN account required: https://msdn.microsoft.com/en-us/microsoft-r/rserver-install-linux-server?f=255&MSPPError=-2147217396
5. *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
6. Find out if MRS is loaded: **sessionInfo()**

Module 3:
R Client Options



27

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

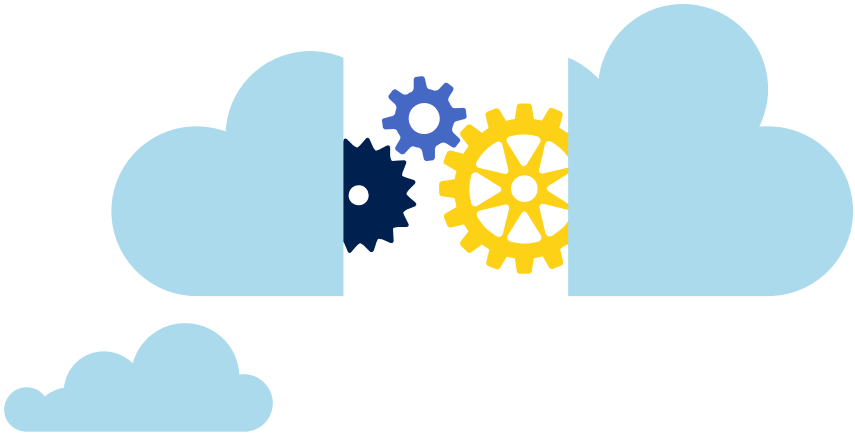
Microsoft R Development Tools

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

1. Installing Microsoft R Client on Windows: https://msdn.microsoft.com/en-us/microsoft-r/install-r-client-windows
2. Files located at: C:\Program Files\Microsoft\R Client\R_SERVER\bin

1. *Optional:* Install Visual Studio (https://www.visualstudio.com/downloads/download-visual-studio-vs) (Select Optional, and select SQL Server Data Tools)
2. *Optional:* Install RTVS (http://microsoft.github.io/RTVS-docs/installer.html)
3. *Optional:* Install Rstudio (https://www.rstudio.com/products/rstudio/download2/)
4. 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

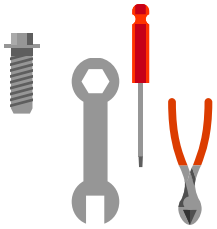
Module 4:
Operationalize Microsoft R



30

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

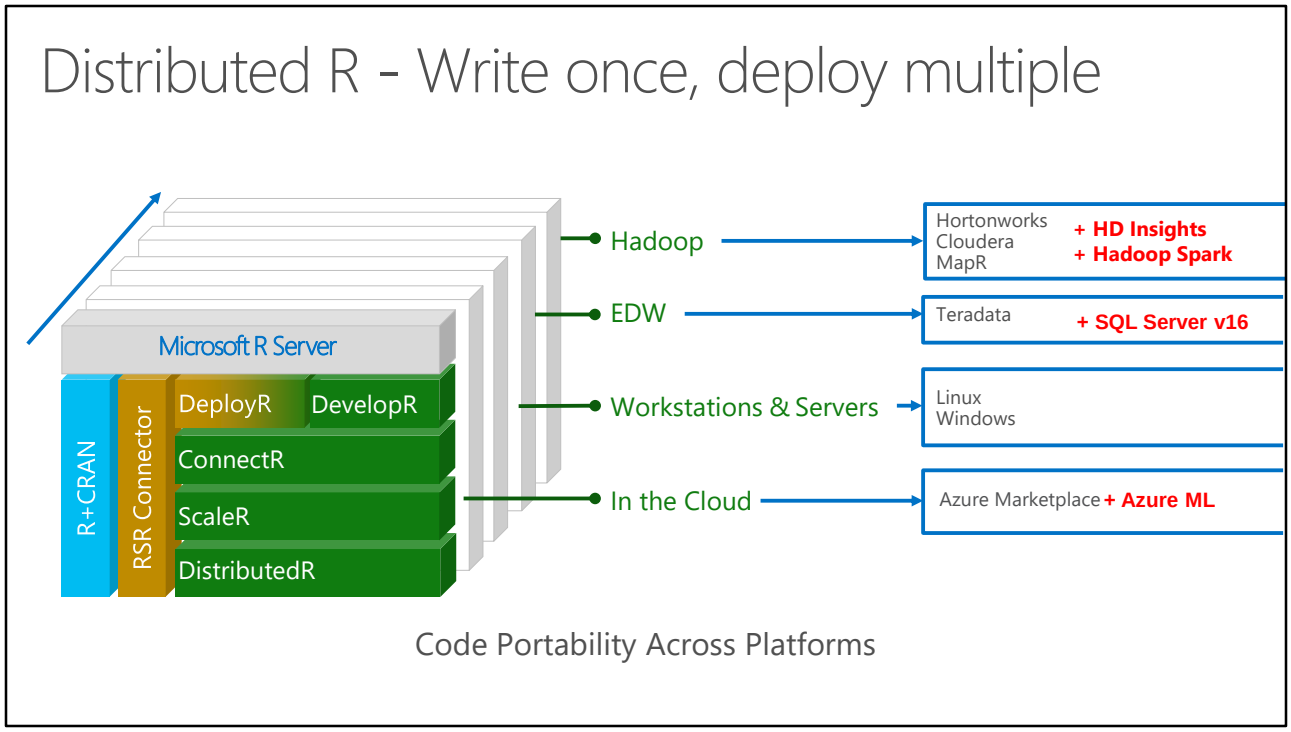
Configuration and Operation



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

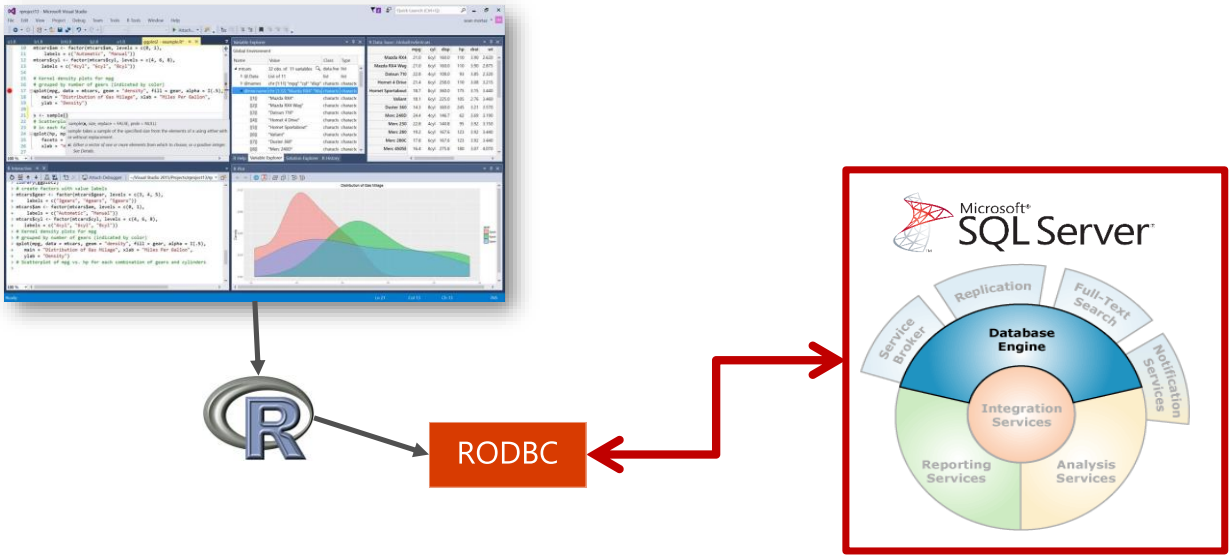
31

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



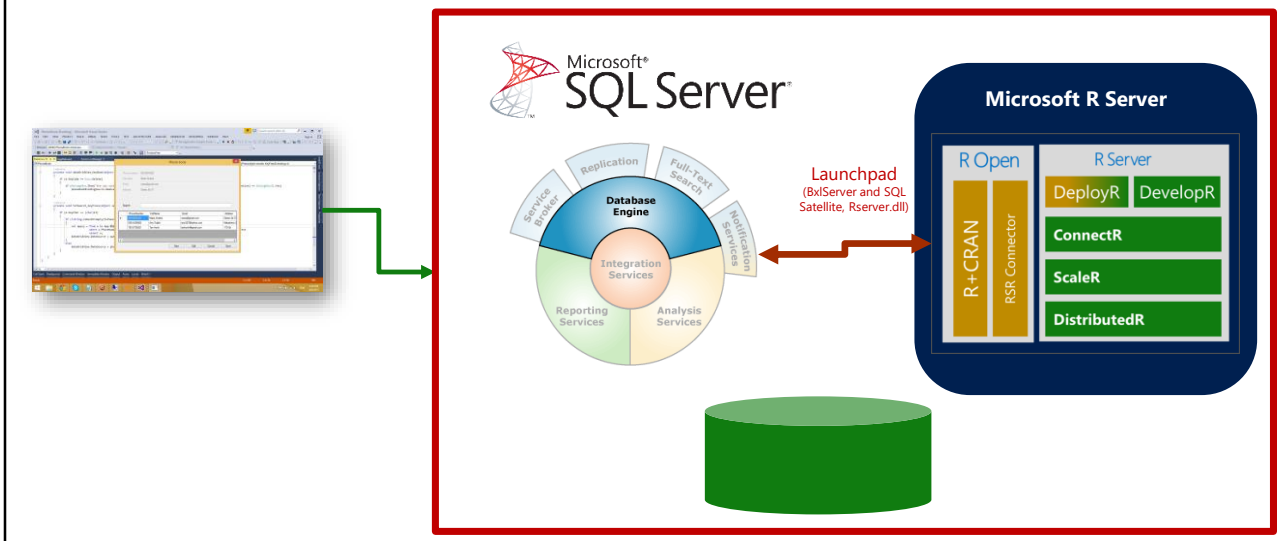
- DeployR Workflow: https://msdn.microsoft.com/en-us/microsoft-r/deployr-about

CRAN/Microsoft R Open And Databases



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

Microsoft R in SQL Server



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

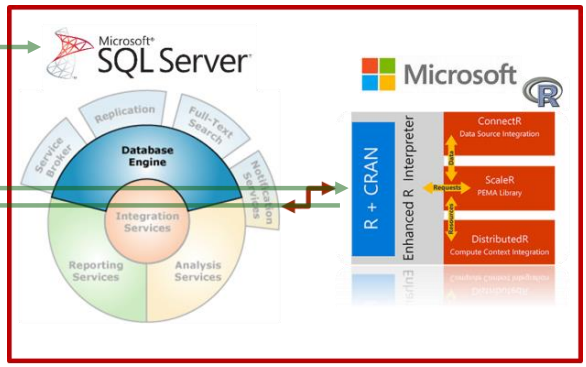
T-SQL and R Interaction

```
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
```

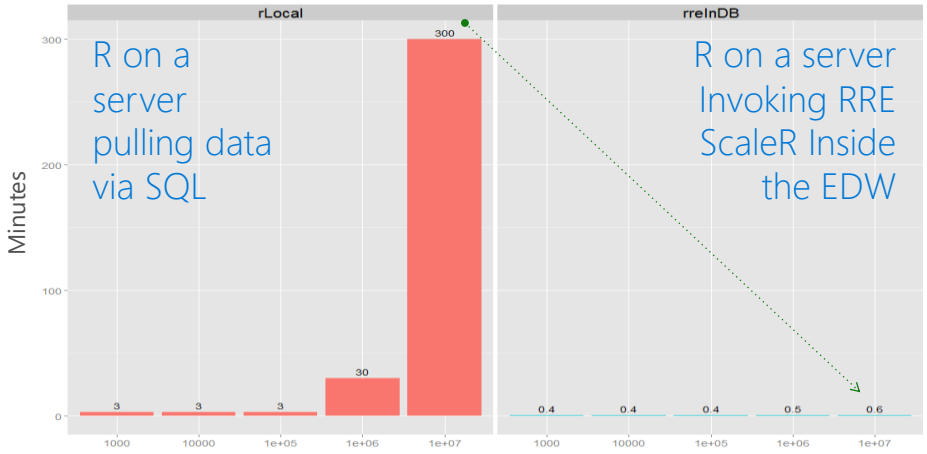


35

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

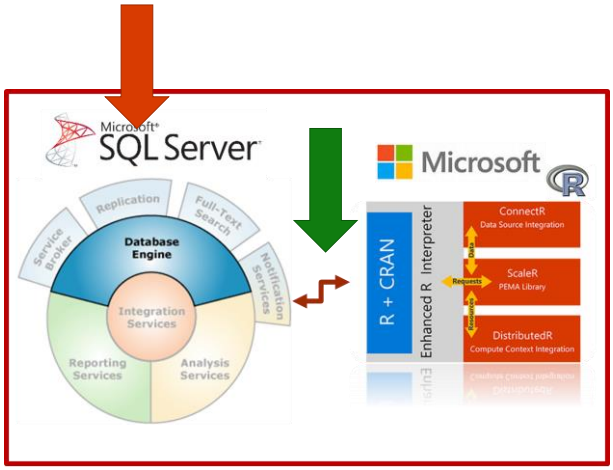
Example of In-Database Acceleration

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

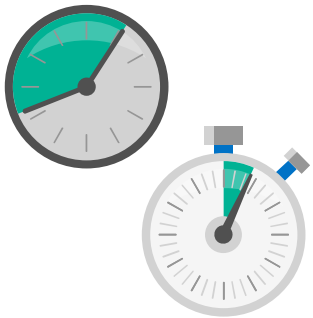


37

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

Performance and Monitoring

- Performance considerations
- Monitoring
- Tuning



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

Security and Governance

- Principals

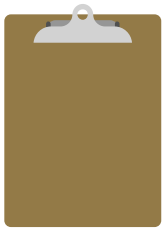
- Securables

39

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

Implementation Considerations

- Coordinating with the R professional
- Best Practices



40

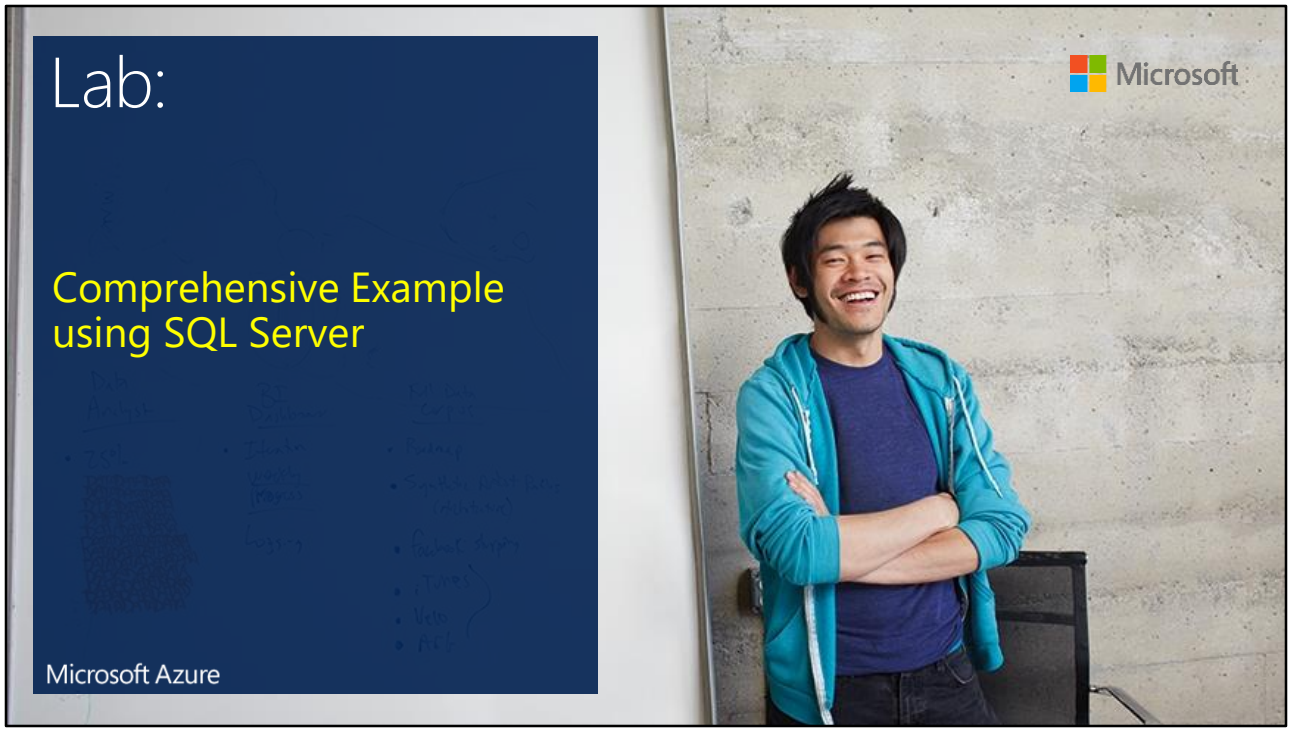
1. Managing and monitoring R Solutions for SQL Server: https://msdn.microsoft.com/en-us/library/mt590866.aspx
2. Upgrade and Installation: https://msdn.microsoft.com/en-us/library/mt653951.aspx
3. Considerations: https://msdn.microsoft.com/en-us/library/mt590540.aspx

Module 5:
Creating a Microsoft R Solution

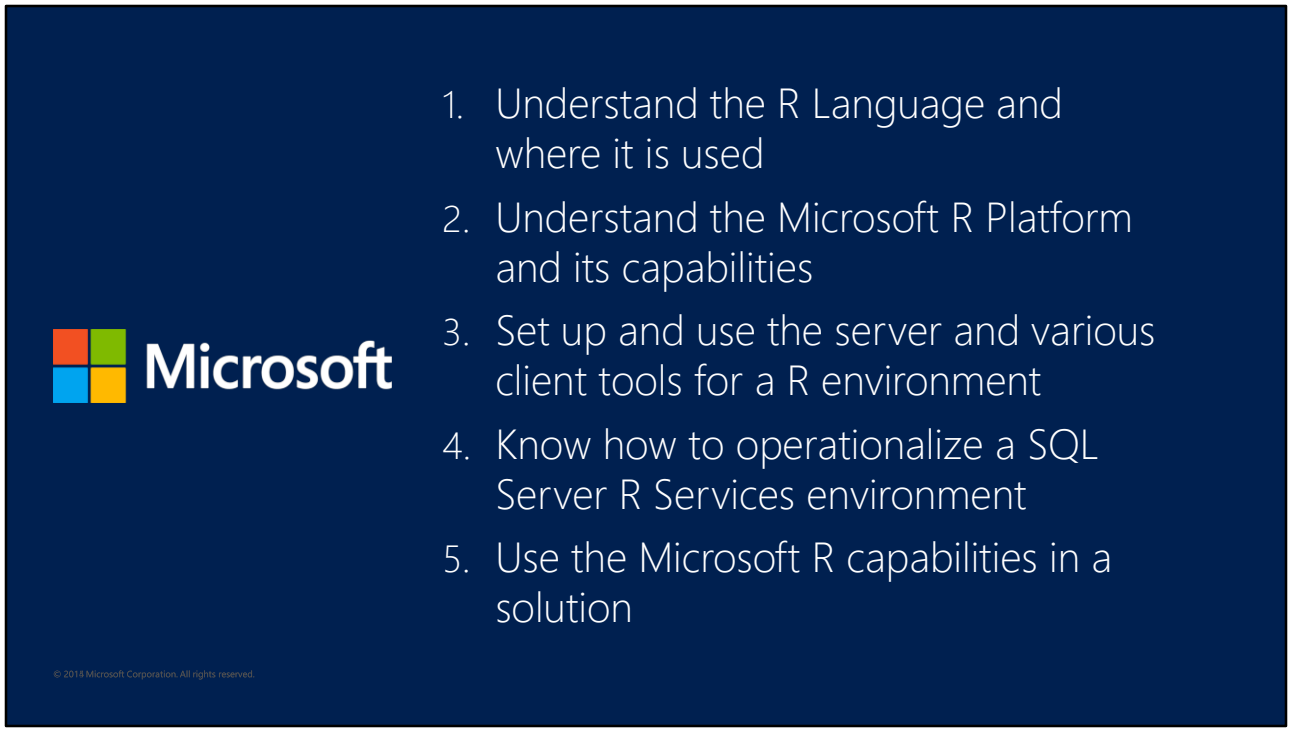
https://mva.microsoft.com/ebooks/

41

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



1. Option 1: Open the **SQL Server R Services Lab** from your \Resources folder, follow the instructions you find there. Source Materials are at: https://github.com/Microsoft/sql-server-samples/tree/master/samples/features/r-services/Telco%20Customer%20Churn
2. 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.
3. Demand Forecasting Template: https://channel9.msdn.com/Blogs/Seth-Juarez/Energy-Demand-Forecasting-Template-with-SQL-Server-R-Services
4. 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



Questions?

More resources:
https://msdn.microsoft.com/en-us/microsoft-r/microsoft-r-more-resources
Revolutions Blog
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