

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.

Section 4 Learning Objectives

1. Understand Azure ML and how experiments are created
2. Understand how MRS can be used to perform Machine Learning experiments
3. Use ADF to schedule Azure ML Activities

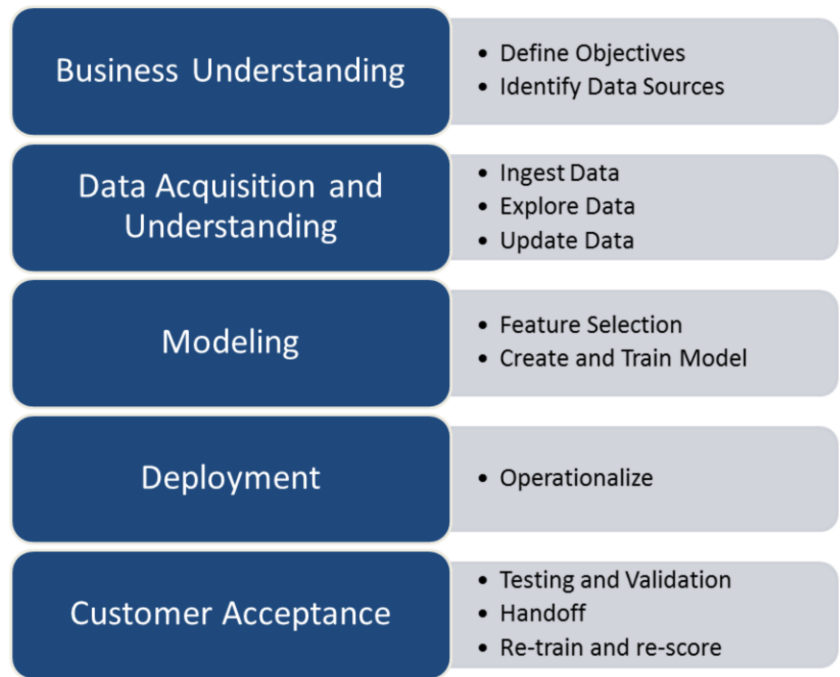


1. At the end of this Module, you will:
 1. Understand Azure ML and how experiments are created
 2. Understand how MRS can be used to perform Machine Learning experiments
 3. Use ADF to schedule Azure ML Activities

The Data Science Process and Platform














The Team Data Science Process



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/>
5. An integrated process and toolset allows for a more close-to-intent deployment

6. Iterations are required to close in on the solution – but are harder to management and monitor

The Cortana Intelligence Platform

	Cortana, Cognitive Services, Bot Framework
	Power BI
	Stream Analytics
	HDInsight
	Azure Machine Learning (MRS)
	SQL Data Warehouse (SQL DB, Document DB)
	Data Lake
	Event Hubs
	Data Factory
	Data Catalog
	Microsoft Azure

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: Azure ML



6

1. Example paths for using Azure ML:
<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-plan-sample-scenarios/>

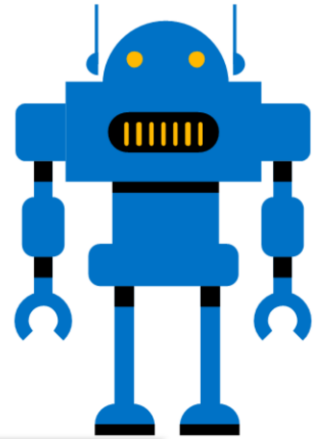
Machine Learning in 5 Minutes

The Formal one:

"A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P** if its performance at tasks in **T**, as measured by **P**, improves with experience **E**."

A Practical Example:

Look at data. Do the thing. Better? No? Look at the data. Do something different. Better? Yes? *Do that again.* (Repeat)



1. Choosing an Algorithm for Machine Learning:
<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-choice/>

Machine Learning Capabilities

Which category
(Classification)



How much/many
(Regression)



Which group
(Clustering, Recommender)



Is it odd
(Anomaly)



Which action
(Reinforcement Learning)



1. Regression: Predict a real value for each item (stock/currency value, temperature). – How much/how many?
2. Classification: Assign a category to each item (Chinese | French | Indian | Italian | Japanese restaurant). – Which Category?
3. Clustering/Recommendation: Partition items into homogeneous groups (clustering twitter posts by topic). – Which Groups?
4. Anomaly: Identify when something unexpected happens. – Is this weird?
5. Reinforcement Learning: Make an appropriate action for some new data. – Which action?

Machine Learning Algorithms

Split into two main categories:

- Supervised learning
 - Predicting the future
 - Learn from known past examples to predict future
 - Labels provided
- Unsupervised learning
 - Making sense of data
 - Understanding the past
 - Learning the structure of data
 - Labels no provided



1. Algorithm Documentation:
<https://msdn.microsoft.com/library/dn905974.aspx>
2. Exploring:
<https://azuremlsimplifieds.azurewebsites.net/simplifieds/>

The Azure ML Environment

Development Environment

- Creating Experiments
- Sharing a Workspace



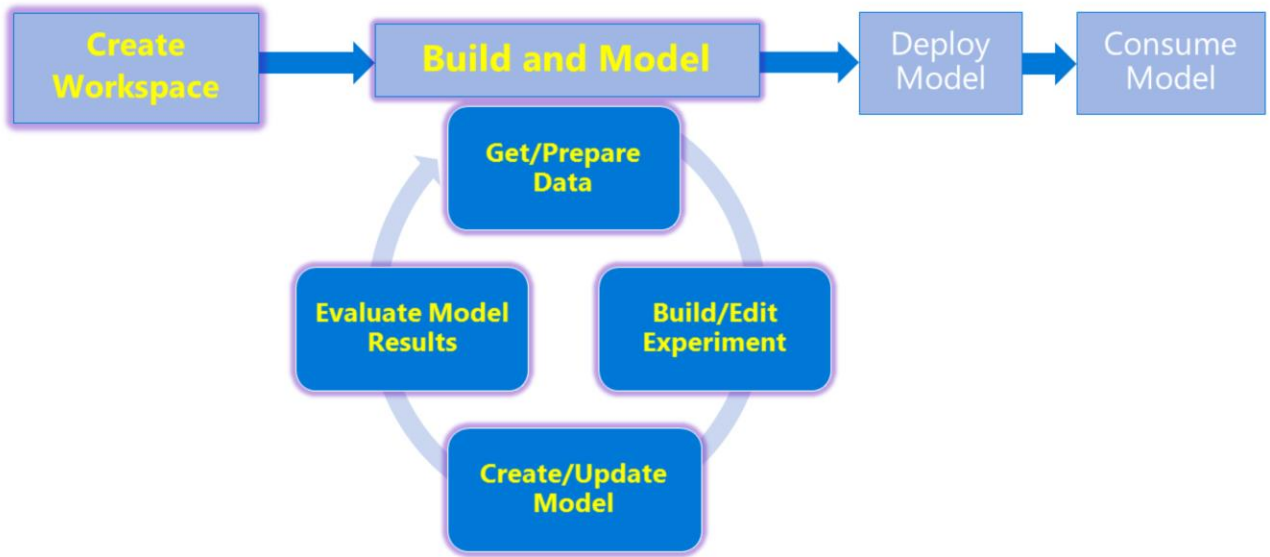
Deployment Environment

- Publishing the Model
- Using the API
- Consuming in various tools



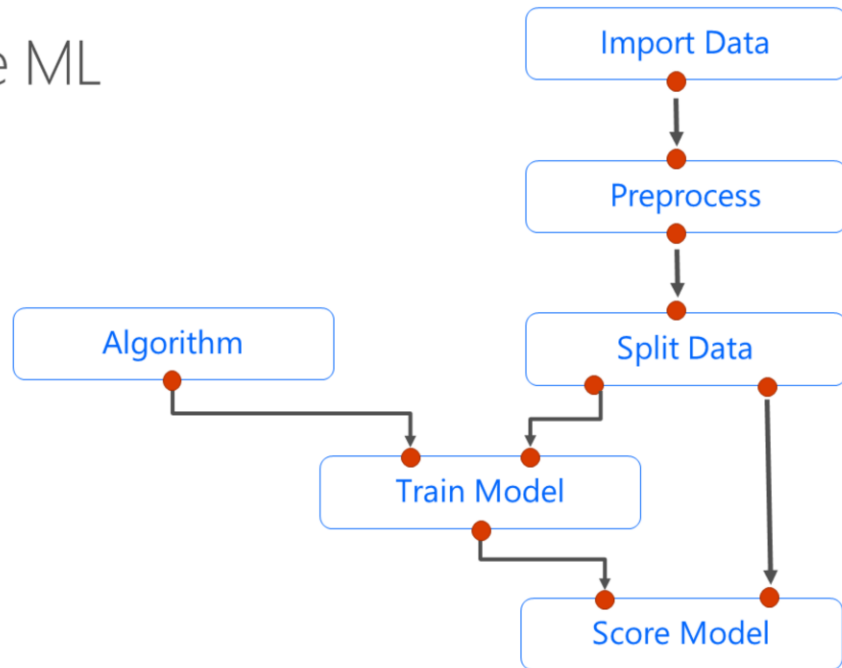
1. Guided tutorials: <https://azure.microsoft.com/en-us/documentation/services/machine-learning/>
2. Microsoft Azure Virtual Academy course: https://mva.microsoft.com/en-US/training-courses/microsoft-azure-machine-learning-jump-start-8425?l=ehQZFoKz_7904984382

Creating an Experiment



1. Beginning Series: <https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-for-beginners-the-5-questions-data-science-answers/>

Basic Azure ML Elements



1. Designing an experiment in the Studio:
<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-what-is-ml-studio/>



1. Open the **AML Student Workbook** from your \Resources folder
2. Follow the instructions you find there

Module 2: Microsoft R Server for Machine Learning



14

1. Primary documentation: <https://www.microsoft.com/en-us/server-cloud/products/r-server/>

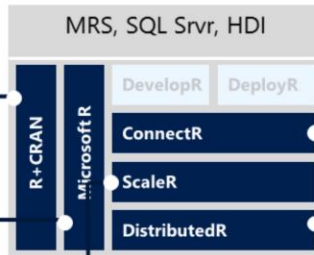
The Microsoft R Server Platform Parallelization & Data Constructs

R+CRAN

- Open source R interpreter
 - R 3.1.2
- Freely-available huge range of R algorithms
- Algorithms callable by RevoR
- Embeddable in R scripts
- 100% Compatible with existing R scripts, functions and packages

Microsoft R

- Performance enhanced R interpreter
- Based on open source R
- Adds high-performance math library to speed up linear algebra functions



ScaleR

- Ready-to-Use high-performance big data big analytics
- Fully-parallelized analytics
- Data prep & data distillation
- Descriptive statistics & statistical tests
- Range of predictive functions
- User tools for distributing customized R algorithms across nodes
- Wide data sets supported – thousands of variables

ConnectR

- High-speed & direct connectors

Available for:

- High-performance XDF
- SAS, SPSS, delimited & fixed format text data files
- Hadoop HDFS (text & XDF)
- Teradata Database & Aster
- EDWs and ADWs
- ODBC

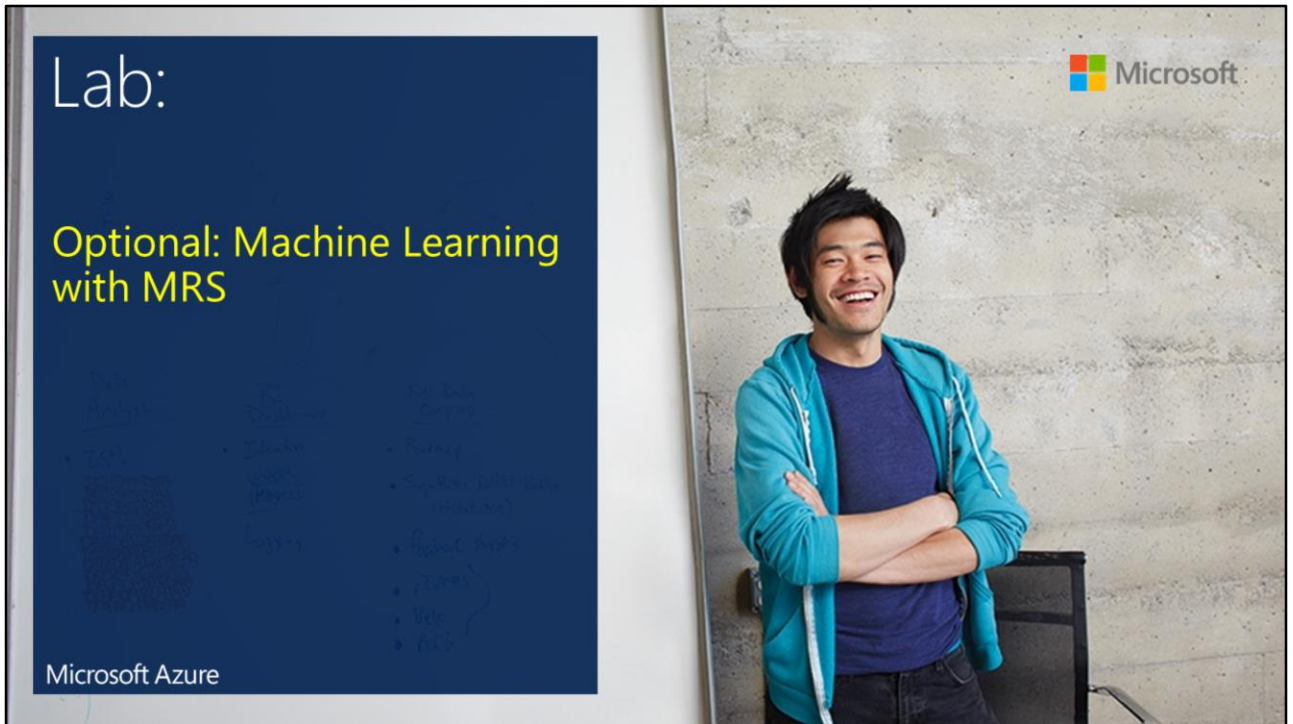
DistributedR

- Distributed computing framework
- Delivers cross-platform portability

Available on:

- Windows Servers
- Red Hat and SuSE Linux Servers
- IBM Platform LSF Linux
- Microsoft HPC Clusters
- Teradata Database
- Cloudera Hadoop
- Hortonworks Hadoop
- MapR Hadoop

1. Getting started with Microsoft R Server:
<https://ppe.msdn.microsoft.com/en-us/microsoft-r/index?branch=master&f=255&MSPPError=-2147217396>



1. Open the **MRS Student Workbook** document from your \Resources file
2. Locate the section marked “**Predictive Modeling with MRS**” and follow the instructions there

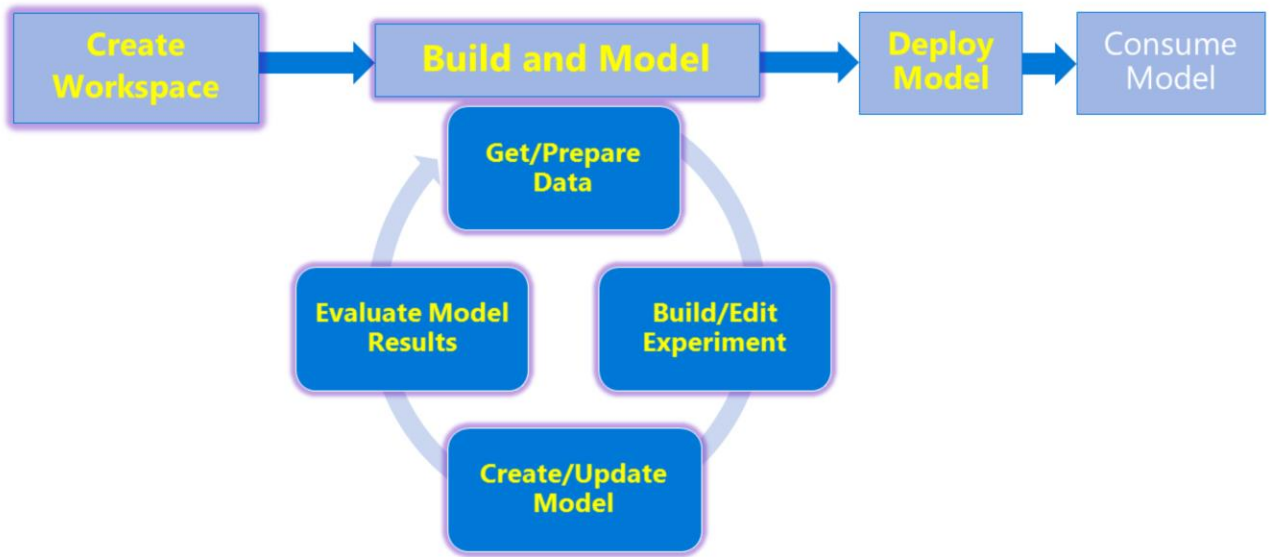
Module 3: Azure Data Factory and Azure ML



17

1. Create Predictive Pipelines using Azure ML Activities in ADF: <https://azure.microsoft.com/en-us/documentation/articles/data-factory-azure-ml-batch-execution-activity/>

Deploying the Model



1. Deploying the Azure ML Model:
<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-walkthrough-5-publish-web-service/>



1. Understand Azure ML and how experiments are created
2. Understand how MRS can be used to perform Machine Learning experiments
3. Use ADF to schedule Azure ML Activities

© 2018 Microsoft Corporation. All rights reserved.

Questions?