

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.



There's a lot of information here. You'll get hands-on experience in class, with a chance to ask an instructor any questions you might have. However, this forms only a small part of learning so many technologies and processes.

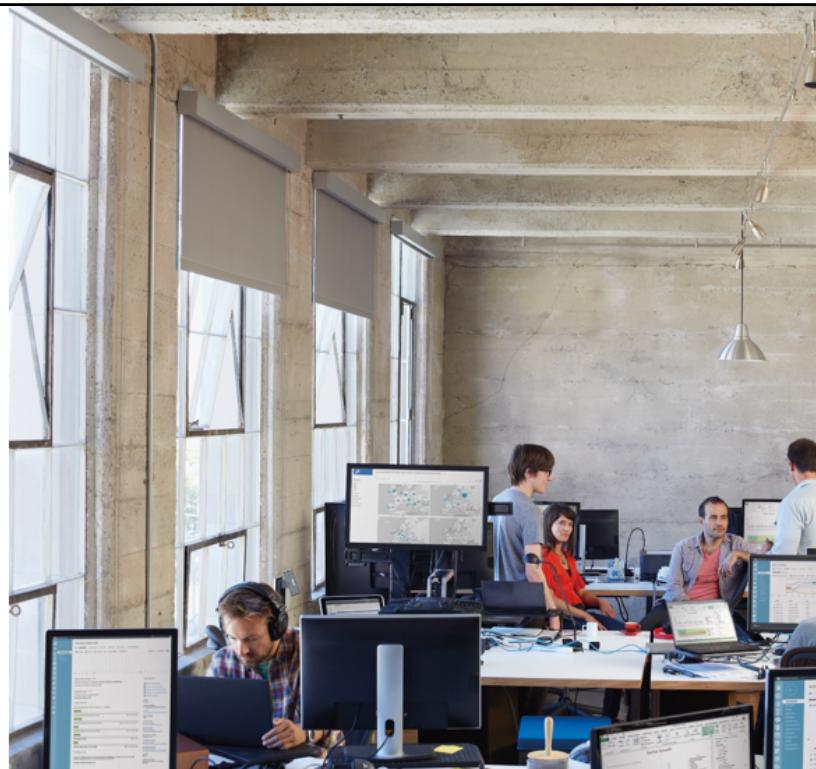
The next largest section of your learning is to read each and every reference in the student handouts. Many point to books, websites and entire courses you can take in each topic. You won't be expected to learn all of them at once – it's more common to pull out the sections you need when you do a project, and learn about that technology before you implement it.

You'll learn the most by actually creating and implementing solutions for your customers. You will do a quick design at the

end of this course, but that is a small sample of working with the technologies in a hands-on fashion. The good news is that you'll find everything you need to be successful in this course and these materials.

# Introduction

- Class hours
- Facilities
- Meals and Breaks
- Internet and Azure Access
- Labs
- Feedback



There will be a general feedback form, but after each module you are asked if you can complete the objectives. If you cannot, let the instructor know.

There are a few things you need prior to coming to class:

- A background in data technologies, such as working with Relational and Non-Relational data processing systems
- A general level of predictive and classification Statistics
- A general understanding of Machine Learning
- A subscription to Microsoft Azure (this may be provided through your company or as part of your invitation)

We'll be using the Windows Data Science Virtual Machine

(<https://azure.microsoft.com/en-us/documentation/articles/machine-learning-data-science-provision-vm/#tools-installed-on-the-microsoft-data-science-virtual-machine>) but you can also install things locally if you wish.

- A laptop with Visual Studio installed – the Community Edition (free) is acceptable  
– Version 2015 preferable (<https://www.visualstudio.com/en-us/products/visual-studio-community-vs.aspx>)
- Azure SDK and Command-line Tools installed (<https://azure.microsoft.com/en-us/downloads/>)
- Azure Storage Explorer  
(<http://go.microsoft.com/fwlink/?LinkId=698844&clcid=0x409>)
- Power BI Desktop (<https://powerbi.microsoft.com/en-us/desktop/?gated=0&number=1>)

Session	Concepts	Technologies
Process and Platform Environment Configuration	The CIS Process, CIS Platform components, Tools installation and overview	CRISP-DM, CIS, Azure Portal, ADC Interface, Visual Studio Interface (and RTVS), Power BI Interface, Azure Machine Learning Interface, Azure PowerShell, Azure Storage Explorer
Data Discovery and Ingestion	Data sourcing, Feature selection techniques, Data cataloging, Data Ingestion, Data Exploration	Azure Data Catalog, Azure Storage, Techniques for discovery
Data Preparation	Data selection, including Features, Dimension reduction, Data processing, Data transformation and augmentation	Azure Data Factory, HDInsight
Modeling for Machine Learning and Data Mining	Algorithm selection and application, Parameter selection and adjustment	Azure Machine Learning, Microsoft R Server overview, Azure Data Factory
Business Validation and Model Evaluation	Business validation of report and results, Model testing and cross-validation	Azure Machine Learning, Microsoft R Server overview, Azure Data Factory, Business Validation, SQL DB, Azure Storage
Deploying and Accessing the Solution	Deploying the solution using Data Destinations, Deploying the solution using API's, Deploying the Solution using Queries and Reports	Azure Data Storage, SQL DB, Azure Machine Learning API, Cognitive Services API, HIVE, Power BI
Workshop recap	Mapping requirements to solution elements, What to use When	Understand when to use each component within CIS

## 1. Components and capabilities of Cortana Intelligence:

<https://www.microsoft.com/en-us/cloud-platform/what-is-cortana-intelligence-suite>

# Section 1 Learning Objectives

1. Understand the CIS Process
2. Understand the CIS Platform
3. Set up and Configure your Development Environments



1. At the end of this Module, you will:
  1. Understand the CIS Process
  2. Understand the CIS Components
  3. Set up and Configure your Development Environments

## Module 1: Understanding Cortana Intelligence



7

1. The official documentation for the Cortana Intelligence Suite:  
<https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/>
2. The Advanced Analytics Learning Portal:  
<http://learnanalytics.microsoft.com/>

## Cortana Intelligence in a Sentence:

Cortana Intelligence is a **Platform** and a **Process** to perform advanced analytics from start to finish

1. What you can do with CIS: <https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/why-cortana-intelligence.aspx>
2. More about the process:  
<https://channel9.msdn.com/Blogs/Seth-Juarez/Understanding-Data-Science-for-building-Predictive-Analytics-Solutions-by-Francesca-Lazzeri>

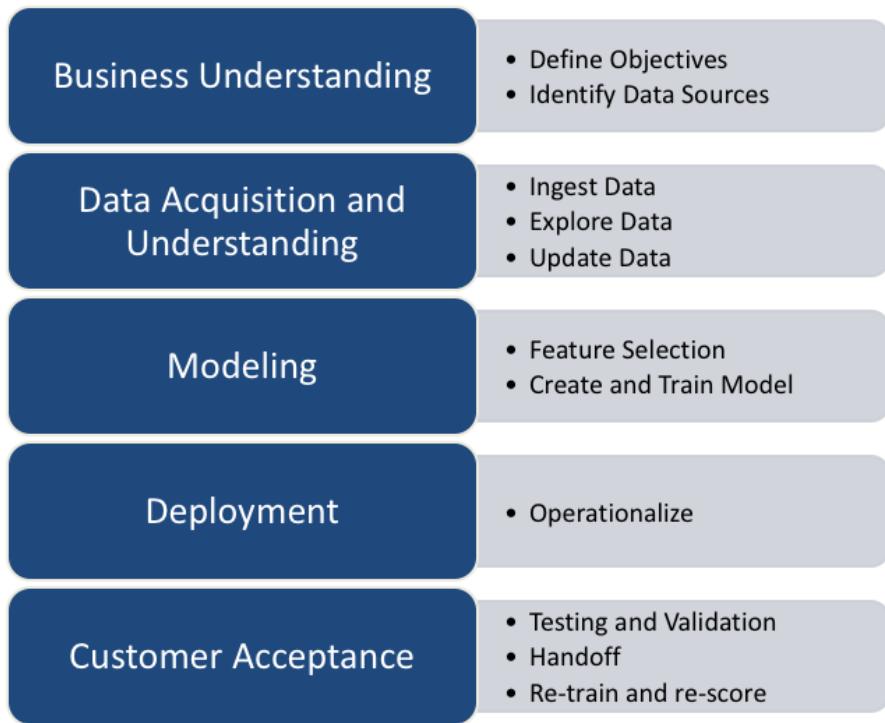
## The Data Science Process and Platform



Great overview of the process and the platform:

<https://blogs.msdn.microsoft.com/azuredev/2017/02/19/the-data-science-process-with-azure-machine-learning/>

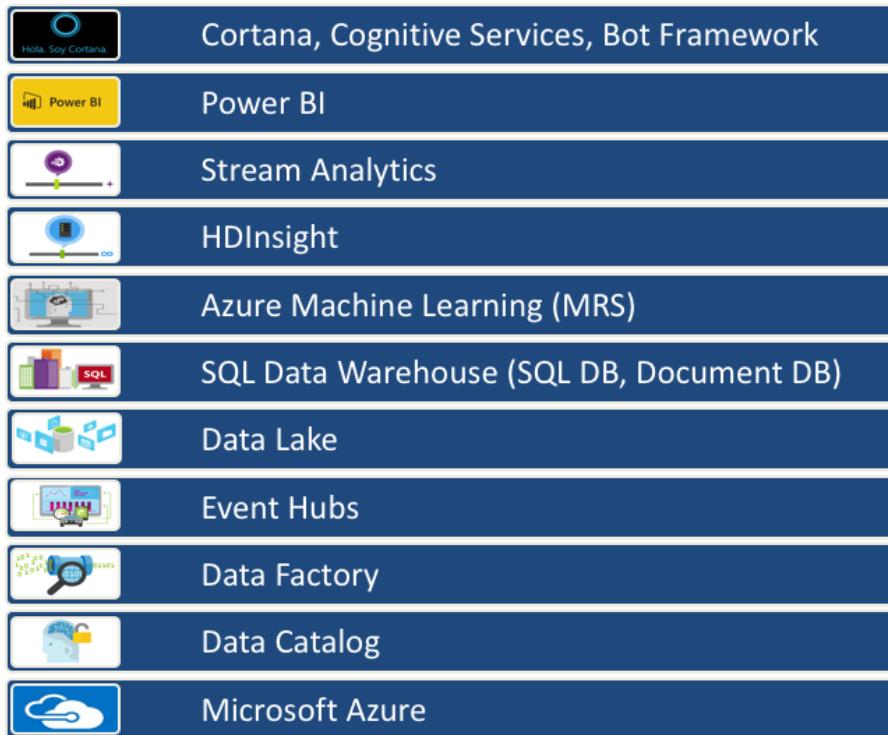
# The Team Data Science Process



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



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 2: The Cortana Intelligence Suite



12

1. All of the components within the suite:

<https://www.microsoft.com/en-us/server-cloud/cortana-intelligence-suite/what-is-cortana-intelligence.aspx>



# Microsoft Azure

What it is:

Microsoft's Cloud Platform including IaaS, PaaS and SaaS

When to use it:

- Storage and Data
- Networking
- Security
- Services
- Virtual Machines
- On-demand Resources and Services

1. What you can do with it: <https://azure.microsoft.com/en-us/overview/what-is-azure/>
2. Platform: <http://microsoftazure.com>
3. Storage: <https://azure.microsoft.com/en-us/documentation/services/storage/>
4. Networking: <https://azure.microsoft.com/en-us/documentation/services/virtual-network/>
5. Security: <https://azure.microsoft.com/en-us/documentation/services/active-directory/>
6. Services: <https://azure.microsoft.com/en-us/documentation/articles/best-practices-scalability-checklist/>
7. Virtual Machines: <https://azure.microsoft.com/en-us/documentation/services/virtual-machines/windows/> and <https://azure.microsoft.com/en-us/documentation/services/virtual-machines/linux/>
8. PaaS: <https://azure.microsoft.com/en->

[us/documentation/services/app-service/](https://us documentation/services/app-service/)



1. Activate your Azure Subscription (Learn more here: <https://docs.microsoft.com/en-us/azure/billing/billing-buy-sign-up-azure-subscription>)
  1. Open the Azure Portal
  2. Create one empty Resource Group (More here: <https://docs.microsoft.com/en-us/azure/azure-resource-manager/resource-group-overview>)
2. Explain a situation where data was used in a new and unexpected way in a business or an organization
3. List three advantages to using a cloud or hybrid architecture
4. List three objections to hosting an application in the cloud, and three responses to those objections



## Azure Data Catalog

### What it is:

On-Line Catalog of Meta-Data about your Data Sources, with easy tagging and searching

### When to use it:

- Sourcing data
- Data discovery
- Data vetting

1. Azure Data Catalog: <http://azure.microsoft.com/en-us/services/data-catalog>



# Azure Data Factory

## What it is:

A pipeline system to move data in, perform activities on data, move data around, and move data out

## When to use it:

- Create solutions using multiple tools as a single process
- Orchestrate processes - Scheduling
- Monitor and manage pipelines
- Call and re-train Azure ML models

1. Azure Data Factory: <http://azure.microsoft.com/en-us/services/data-factory/>



# Event Hubs

## What it is:

A system to ingest data from the web, IoT, and apps at scale

## When to use it:

- To stream in large amounts of data
- With IoT workloads
- Use with variable or unpredictable large data loads
- Similar to Kafka

1. Azure Event Hubs: <http://azure.microsoft.com/en-us/services/event-hubs/>



# Data Lake

## What it is:

Data storage (Web-HDFS) and Distributed Data Processing (HIVE, Spark, HBase, Storm, U-SQL) Engines

## When to use it:

- Low-cost, high-throughput data store
- Non-relational data
- Larger storage limits than Blobs

1. Azure Data Lake: <http://azure.microsoft.com/en-us/campaigns/data-lake/>

 DocumentDB

## What it is:

An automatically-indexed, schema-agnostic JSON database

## When to use it:

- Query non-relational data
- Schema defined per object
- Document (JSON) – Oriented database
- Ad-hoc queries
- Stored Procedures

1. Azure DocumentDB: [https://azure.microsoft.com/en-us/services/documentdb/?WT.srch=1&WT.mc\\_ID=SEM\\_JQ3fO8dU](https://azure.microsoft.com/en-us/services/documentdb/?WT.srch=1&WT.mc_ID=SEM_JQ3fO8dU)



# SQL Database

What it is:

A SQL Server Database Service in the Cloud

When to use it:

- When you need a relational store
- When you need full transactional support
- When you have familiarity with SQL and T-SQL and SQL Server Objects
- When you need lots of flexible indexing
- When you do not want to manage a SQL Server
- When you have multitenant databases needed

1. Azure SQL DB: <https://azure.microsoft.com/en-us/services/sql-database/?b=16.18>



# SQL Data Warehouse

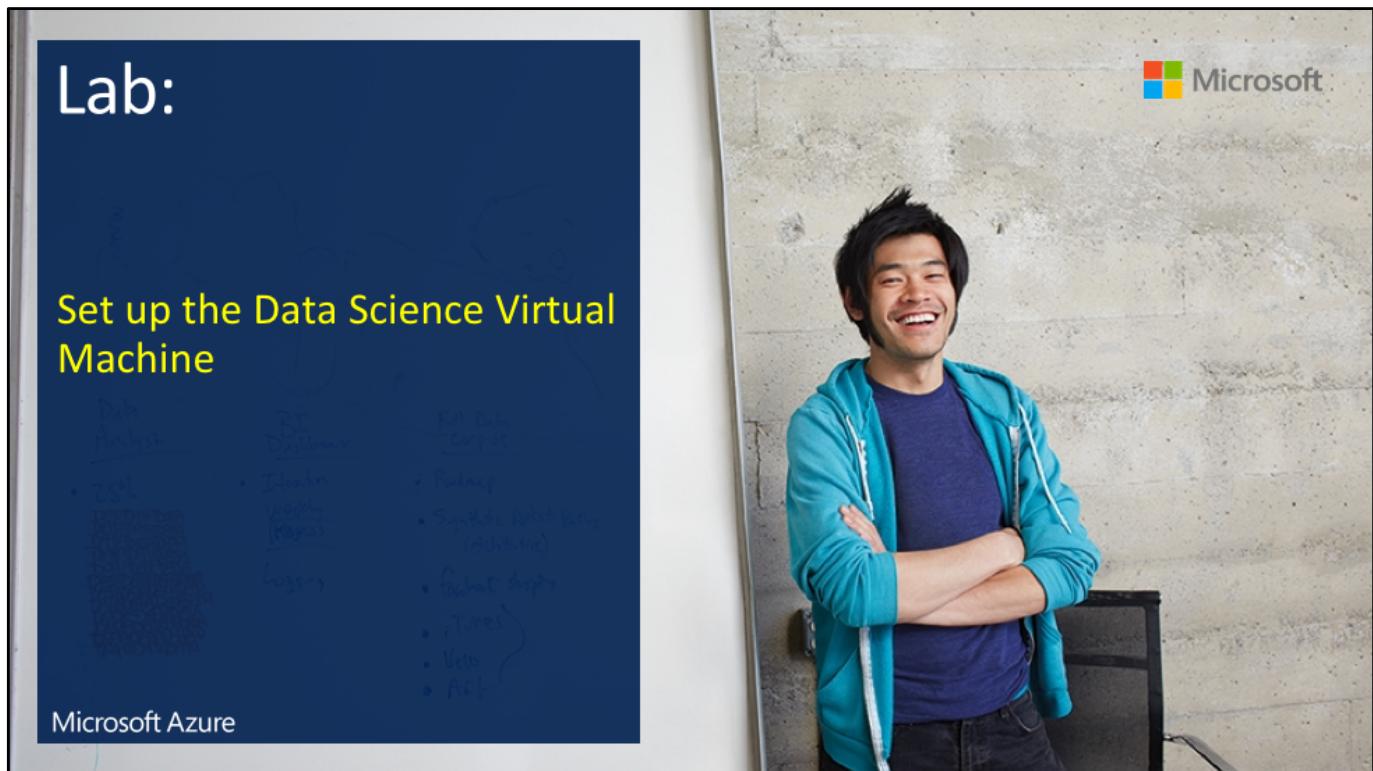
What it is:

A Scaling Data Warehouse Service in the Cloud

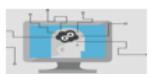
When to use it:

- When you need a large-data BI solution in the cloud
- When you are using lots of relational data
- When you need lower cost relational storage than Blobs
- When you need pause-able scaled compute

1. Azure SQL Data Warehouse: <http://azure.microsoft.com/en-us/services/sql-data-warehouse/>



1. Log in to the Azure Portal
2. Deploy one Windows Data Science Virtual Machine (DSVM) – note your admin name and password
  1. Need help? Check here – except make sure you pick the Windows Data Science Machine! <https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-windows-hero-tutorial/>
3. Start the DVSM
4. Connect to the DSVM and begin updating the Power BI, Visual Studio, and Windows environments
  1. Need help? Check here: <https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-windows-connect-logon/>



## Azure ML

### What it is:

A multi-platform environment and engine to create and deploy Machine Learning models and API's

### When to use it:

- When you need to create predictive analytics
- When you need to share Data Science experiments across teams
- When you need to create call-able API's for ML functions
- When you also have R and Python experience on your Data Science team

1. Azure Machine Learning: <http://azure.microsoft.com/en-us/services/machine-learning/>



# Microsoft R Server (MRS)

## What it is:

A scalable, high-performance R engine used in on-prem, in-cloud, and in-service areas

## When to use it:

- When you need to use the R language and environment for data processing at scale

1. Microsoft R Server: <https://www.microsoft.com/en-us/server-cloud/products/r-server/>



# HDInsight

## What it is:

Microsoft's implementation of apache Hadoop (as a service) that uses Blob or Azure Data Lake for persistent storage

## When to use it:

- When you need to process large scale data (PB+)
- When you want to use Hadoop or Spark as a service
- When you want to compute data and retire the servers, but retain the results
- When your team is familiar with the Hadoop Zoo

1. Azure HDInsight: <http://azure.microsoft.com/en-us/services/hdinsight/>



## Stream Analytics

What it is:

Real-time cloud-based stream processing

When to use it:

- For complex event processing
- IoT, streaming workloads
- When you need to ingest millions of records per second
- When you need JSON, Delimited, and Avro data processing
- Similar to Apache Storm

1. Azure Stream Analytics: <http://azure.microsoft.com/en-us/services/stream-analytics/>



# Power BI

## What it is:

Interactive Report and Visualization creation for computing and mobile platforms

## When to use it:

- When you need to create and view interactive reports that combine multiple datasets
- When you need to embed reporting into an application
- When you need customizable visualizations
- When you need to create shared datasets, reports, and dashboards that you publish to your team

## 1. Power BI: <https://powerbi.microsoft.com/>



## Cortana and Cognitive Services, Bot Framework

### What it is:

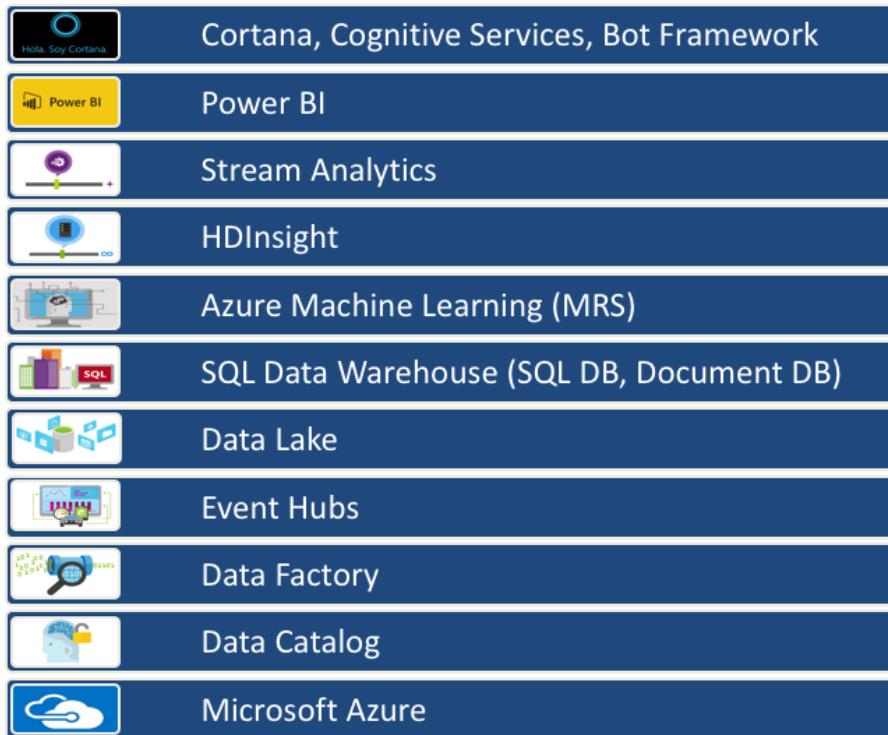
Intelligent assistant available in computing and mobile platforms, integrated into user's ecostructure, speech and vision interaction

### When to use it:

- When you want your users to interact with your solution in a natural language format
- When you have an application of your solution that lends itself to the user's connected ecostructure

1. Cortana: <http://windows.microsoft.com/en-us/windows-10/getstarted-what-is-Cortana>

# 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 3: Setting up Your Development Environment



30

1. Using Azure for Dev/Test: <https://azure.microsoft.com/en-us/solutions/dev-test/>

# Primary Development Tools:

- The Azure Portal
- Azure SDK
- Azure PowerShell and ARM Templates
- Azure Data Catalog
- Azure ML Interface
- Visual Studio (and RTVS)
- Storage Explorer

1. The Azure Portal: <https://portal.azure.com/> and OMS:  
<https://azure.microsoft.com/en-us/documentation/articles/operations-management-suite-overview/> - also, customizations:  
<https://www.youtube.com/playlist?list=PLFuGXEPUdlxLwxsfkvpdvAGInsLDfgvvC>
2. Azure SDK: <https://go.microsoft.com/fwlink/?LinkId=518003&clcid=0x409>
3. Azure Powershell: <https://azure.microsoft.com/en-us/documentation/articles/powershell-install-configure/#what-is-azure-powershell> and ARM Templates: <https://azure.microsoft.com/en-us/documentation/articles/resource-group-overview/>
4. Azure Data Catalog (use an inPrivate browser tab):  
<http://azuredatacatalog.com>
5. Azure Machine Learning: <http://studio.azureml.net>
6. Visual Studio Interface: <https://msdn.microsoft.com/en-us/library/dn762121.aspx>
7. Installing R Tools for Visual Studio (RTVS): <https://www.visualstudio.com/en-us/features/rtvs-vs.aspx>
8. Azure PowerShell: <http://aka.ms/webpi-azps>
9. Storage Explorer:

<https://go.microsoft.com/fwlink/?linkid=698844&clcid=0x409>



1. Create a Storage Account in the region closest to the class location – note the name and access keys
2. Connect to the Azure Data Catalog as described in the classroom login information
3. Connect to <http://studio.azureml.net> and create a free account for the class

Skills check – you should now be able to:



1. Understand the CIS Process
2. Understand the CIS Platform
3. Set up and Configure your Development Environments

© 2014 Microsoft Corporation. All rights reserved.

Questions?