Microsoft Data Science

Cortana Intelligence Suite Workshop – Foundations

Training Overview

Welcome to the Cortana Intelligence Suite workshop delivered by your Microsoft Data Science team. In this workshop you'll cover a series of modules that guide you from understanding an analytics workload, the Cortana Intelligence Suite Process, the foundations of data transfer and storage, data source documentation, storage and processing using various tools. You'll also learn how to work through a real-world scenario using the Cortana Intelligence Suite tools, including the Microsoft Azure Portal, PowerShell, and Visual Studio, among others.

This course is designed to take approximately one to two days, depending on what is covered and how many of the labs are done in-class. The longer course is marked *(Extended Class)* below. All materials are provided regardless of the length of the course.

Note:

If you're getting these materials without classroom instruction, simply walk through the videos and other materials on the delivery site step-by-step, making sure you read and follow the links in the handouts for each topic.

Audience

Technical professionals (Data Scientists, Database professionals, Analysts, BI Professionals) who are familiar with building solutions but not familiar with the entire CIS Platform of products.

Prerequisites

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

- A subscription to Microsoft Azure (this may be provided through your company or as part of your invitation – you **must** have this enabled *prior to class* – you will be using Azure throughout the course, for all labs, work and exercises)
    - You can use your MSDN subscription – https://azure.microsoft.com/en-us/pricing/member-offers/msdn-benefits/
    - Your employer may provide Azure resources to you, but make sure you check to see if you can deploy assets and that they know you'll be using their subscription in the class.
    - Optionally, you may receive instructions in your class invitation.
- We'll be using the Data Science Virtual Machine in Azure for the course. It has all of the tools you will need to work with the materials. Make sure you're able to use the Remote Desktop Protocol (RDP) from your system to be able to work through the labs.
- If you would also like to work with some of the tools locally (you still need an Azure subscription for this class), you can optionally obtain:

- A laptop that you can install software on
- 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 Installed (https://powerbi.microsoft.com/en-us/desktop/)A background in data technologies, such as working with Relational and Non-Relational data processing systems
- Install the Microsoft R Client: http://aka.ms/rclient/download with the R tools for Visual Studio

- It's also a good idea to have a general level of predictive and classification Statistics, and a basic understanding of Machine Learning

Syllabus

Course Session

Each Training Module guides you through a logical progression with hands-on tasks in do-verb form. Each day is broken up into 1-4 hour Modules, where you will learn and perform labs as a group and individually.

NOTE: *The workbooks you receive as part of the classes 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.*

We'll cover the following concepts and technologies in this course, and you'll have the following skills when you complete all of the labs and class activities:

Modules in the complete course

1. Process and Platform, Environment Configuration
2. Data Discovery and Ingestion
3. Data Preparation
4. Modeling for Machine Learning and Data Mining (Extended Class)
5. Business Validation and Model Evaluation (Extended Class)
6. Deploying and Accessing the Solution
7. Workshop recap (Extended Class)

Concepts delivered

1. The Data Science Process, CIS Platform components, Tools installation and overview
2. Data sourcing, Feature selection techniques, Data cataloging, Data Ingestion, Data Exploration
3. Data selection, including Features, Dimension reduction, Data processing, Data transformation and augmentation
4. Algorithm selection and application, Parameter selection and adjustment
5. Business validation of report and results, Model testing and cross-validation
6. Deploying the solution using Data Destinations, Deploying the solution using API's, Deploying the Solution using Queries and Reports
7. Mapping requirements to CIS solution elements, what to use when in CIS

Technologies covered

1. The Data Science Process, Azure Portal, ADC Interface, Visual Studio Interface (and RTVS), Power BI Interface
1. Azure Machine Learning Interface, Azure PowerShell, Azure Storage Explorer
2. Azure Data Catalog, Azure Storage, Techniques for discovery
3. Azure Data Factory, HDInsight
4. Azure Machine Learning, Microsoft R Server overview, Azure Data Factory
5. Azure Machine Learning, Microsoft R Server overview, Azure Data Factory, Business Validation, SQL DB, Azure Storage
6. Azure Data Storage, SQL DB, Azure Machine Learning API, Cognitive Services API, HIVE, Power BI
7. Cortana Intelligence Process, Cortana Intelligence Suite Platform

Skills taught

1. Understand the CIS Process (General level), Understand CIS Components (General Level), Set up and configure the development environment
2. Understand how to source and vet proper data, Understand feature selection, Understand Azure Storage Options, Use various methods to ingest data into Azure Storage, Examine data stored in Azure Storage, Use various tools to explore data
3. Understand ADF and its constructs, Implement an ADF Pipeline referencing Data Sources and with various Activities including on-demand HDInsight Clusters, Understand the HIVE language and how it is used
4. Understand how to use Azure ML and how experiments are created, Understand how MRS can be used to perform Machine Learning experiments, Use ADF to schedule Azure ML Activities
5. Understand how to evaluate the efficacy and performance of an Azure ML experiment, Understand how to evaluate the efficacy and performance of an MSR ML experiment, Access and show data from Azure Storage, Access and Query Azure SQL DB
6. Understand how to publish an Azure ML API, Understand the access methods of Azure Storage and Intelligent Processing, Understand the options to send a HIVE query to an HDI system, Use Power BI to query the results of a solution and create reports in Power BI Desktop, Power BI Service, and Power BI in Microsoft Excel
1. 7. Understand when to use each component within CIS

Live Deliveries: Follow-up Q&A with MS architects and trainers