Microsoft Data Science

Cortana Analytics Workshop

Workshop Overview

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

After the initial Cortana Analytics Workshop, this week will also include two in-depth modules on days three and four: Microsoft R Server and the Microsoft HDInsight One-Day Course.

Prerequisites

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)
- 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/)

Syllabus and Timeline

Course Modules

Each Training Module guides you through a logical progression with hands-on tasks in do-verb form. Each day is broken up into 1-3 hour Modules, where you will learn and perform labs as a group.

The modules are as follows:

Business Case and the Data Science Process

- *Data Science Overview*
- *Process Visualization*
- *Domain knowledge importance*
- *High level architecture design*

The Microsoft Azure Platform

- *IaaS, PaaS and SaaS for the Data Analyst*
- *Data Security*
- *Account Setup and Billing Overview*

Cortana Analytics

- *Components overview*
- *CAP overview and application*
- *Architecture selections*
- *Finalizing the architecture based on factors such as cost, location, team familiarity and other considerations*

Solution Review

- *Reference solution review of selection of components based on requirements and constraints*

Data Sourcing and Identification

- *Working with on-prem and cloud data sources*
- *Data documentation using Azure Data Catalog*

Loading Data to Azure

- *Using the Storage Explorer*
- *Using Visual Studio*
- *AZCOPY*
- *Disk-ship*

Orchestrating Data Ingestion and the Project Flow

- *Azure Data Factory*

Using the Hadoop Ecostructure to process data

- *Prepping and conditioning data*
- *HDInsight*
- *HIVE Queries for data selection*

Feature selection and implementation in Azure ML

- *Feature Selection Options*
- *Using Azure ML to choose Features*

Developing and Selecting Machine Learning Models in Azure ML

- *Model creation*
- *Model testing*
- *Model evaluating*
- *Sharing the experiment among a team*
- *Selecting a model in Azure ML*
- *Landing the results in Azure SQL DB using Azure Data Factory*

Deploying and Using the Azure ML Model

- *Publishing the model in Azure ML*

- *Scheduling the Model in Azure Data Factory as a process*
- *Using Power BI to examine results*

Microsoft R Server – Day Three

- *Introduction*
- *Data Exploration*
- *Data Manipulation*
- *Data Analysis*

Microsoft HDInsight One-Day Course – Day Four

- *Envisioning Big Data Solutions*
- *HDInsight Tools Overview*
- *Map Reduce and Hive Overview*
- *Pig and Sqoop Overview*
- *HBase and Storm Overview*
- *Process Automation Overview*
- *Connecting Power BI with Big Data Sources*
- *Architecture Overview*