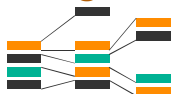**Microsoft** Data Science

Cortana Intelligence Suite Workshop – HDInsight and R

Training Overview

Welcome to the Cortana Intelligence Suite Workshop delivered by your Microsoft team. In these sessions 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 HDInsight (Hadoop) and other techniques. 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.

After the initial Cortana Analytics Bootcamp, this week will also include an in-depth module on Microsoft R Server spanning the afternoon of the second day and morning of the first.

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 may also use your MSDN account if you wish)
- Install the 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
- Optional, *but highly recommended*: Visual Studio installed – the Community Edition (free) is acceptable – Version 2015 works best (https://www.visualstudio.com/en-us/products/visual-studio-community-vs.aspx)
- Optional, *but highly recommended*: Azure SDK and Command-line Tools installed (https://azure.microsoft.com/en-us/downloads/)

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 1-4 hour Modules, where you will learn and perform labs as a group.

The modules, broken up into a general agenda are as follows. The specific modules may bleed across sessions depending on engagement of the audience.

Day 1 Morning Session
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 Intelligence Suite

- *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*

Day 1 Afternoon Session
Data Sourcing and Identification

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

Transferring Data to Azure

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

Day 2 – Morning Session
Using HDInsight to process data

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

Deploying and Using the Solution

- *Publishing and accessing the HDInsight Solution*
- *Using Power BI to examine results*

Day 2 – Afternoon Session

Microsoft R Server

- *Foundations of R*
- *Data Exploration and Manipulation*

Day 3 Morning Session

Microsoft R Server

- *Foundations of R Server*
- *Modeling and Prediction*

Day 3 – Afternoon Session (Optional)

Follow-up Q&A with MS architects, TSPs, and trainers