Microsoft

Implementing Real-Time Analysis with Hadoop in Azure HDInsight

Lab Setup Guide

Overview

This course includes optional labs in which you can try out the techniques demonstrated in the course for yourself.

What You'll Need

To complete the labs, you will need the following:

- A web browser
- A Microsoft account
- A Microsoft Azure subscription
- A Microsoft Windows, Linux, or Apple Mac OS X computer
- The lab files for this course

Creating a Free Trial Azure Subscription

If you already have a Microsoft Azure subscription, you can skip this section.

If you do not have an Azure subscription, you can sign up for the Visual Studio Dev Essentials program at https://visualstudio.com/dev-essentials. This will give you $25 of Azure credit per month for a year. Note that HDInsight clusters consume credit even when not in use, so be careful to delete your clusters after each lab if you don't intend to use them immediately; otherwise you will run out of credit before the month ends.

Alternatively, follow these steps to create a free 30-day trial subscription, which includes enough free credit in your local currency to complete the labs. You will need to provide a valid credit card number for verification, but you will not be charged for Azure services – for more information, see the frequently asked questions in the Azure sign-up page.

1. If you already have a Microsoft account that has <u>not</u> already been used to sign up for a free Azure trial subscription, you're ready to get started. If not, don't worry, just create a new Microsoft account at https://signup.live.com.
2. After you've created a Microsoft account, browse to http://aka.ms/edx-dat202.2x-az and follow the instructions to sign up for a free trial subscription to Microsoft Azure. You'll need to sign-in with your Microsoft account if you're not already signed in. Then you'll need to:

a. Enter your cellphone number and have Microsoft send you a text message to verify your identity.
b. Enter the code you have been sent to verify it.
c. Provide valid payment details. This is required for verification purposes only – your credit card won't be charged for any services you use during the trial period, and the account is automatically deactivated at the end of the trial period unless you explicitly decide to keep it active.

Configuring a Client Computer

You can use a variety of tools to work with Hadoop in HDInsight from Windows, Linux and OSx client computers.

Install Azure Storage Explorer

You will be working with Azure blob storage in this course. You can use any Azure storage client to upload and download files to Azure. If you do not already have an Azure storage client installed, you can install Azure Storage Explorer, which is available for Windows, Mac OSX, and Linux.

1. Browse to http://storageexplorer.com/ and follow the instructions to download and install the latest version of Azure Storage Explorer for your operating system (Windows, Mac OSX, or Linux).

Note: We recommend using Azure Storage Explorer to transfer files between your local computer and Azure Blob Storage. It provides an intuitive, easy-to-use interface and works on Windows, Mac OS X, and Linux. However, if you prefer you can use the Azure Command Line Interface (which you can download from https://azure.microsoft.com/en-us/downloads/) or any other Azure storage client tool, including Microsoft Visual Studio, AzCopy, and others.

Install PuTTY on Windows

HDInsight Hadoop clusters can be provisioned as Linux virtual machines in Azure. When using a Linux-based HDInsight cluster, you connect to Hadoop services using a remote SSH session. Linux and Mac OSx computers have an SSH client interface built-in, but if you plan to use a Windows client computer with a Linux-based HDInsight cluster, you must install an SSH client such as PuTTY.

Tip: Complete this procedure only if you are using a *Windows* client.

1. In a Web browser, navigate to http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html.
2. Download **putty.exe**, saving it to a suitable folder on your local file system (for example, C:\putty).
3. Create a shortcut to **putty.exe** on your desktop for convenience.

Download the Lab Files

The course materials for this course include files that are required to complete the labs.

1. Download the lab files for this course from http://aka.ms/edx-dat202.2x-labs.
2. Extract the **HDILabs.zip** archive you downloaded to a folder on your local computer.
3. Ensure that the extracted folder and all subfolders are not read-only.

Optional - Install Visual Studio and the .NET SDK for Azure

The labs in this course have been designed so that they can be completed on Windows, Linux, or Mac OS X computers, and they include some programming tasks that require you to build applications in Java. Optional steps for these tasks have been included for existing developers who prefer to program in Microsoft C#, and if you wish to follow these alternative steps you will need Visual Studio 2015 and the latest version of the Microsoft .NET SDK for Azure.

1. If you do not already have Visual Studio 2015 installed, download and run the Visual Studio 2015 Community Edition installer from http://aka.ms/edx-dat202.2x-vs. Then complete the installation, being sure to include the **SQL Server Client Tools** component (which is required for the .NET SDK for Azure).

2. Use the Microsoft Web Platform Installer to install the latest version of the .NET SDK for Azure for Visual Studio 2015. You can download and run the Web Platform Installer from http://go.microsoft.com/fwlink/?linkid=255386.