Processing Big Data 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 computer with the following software installed:
    - Microsoft Azure PowerShell
    - Microsoft Visual Studio with the Azure SDK
    - Microsoft Power BI Desktop
- 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. Otherwise, follow these steps to create a free trial subscription. 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 http://azure.microsoft.com/pricing/free-trial-faq/.

1. If you already have a Microsoft account that has not 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://azure.microsoft.com and click the **Free Trial** link. Then 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.

Setting Up a Client Computer

In addition to an Azure subscription, in which your HDInsight cluster will be hosted, you'll need a client computer from which you can manage the cluster and the big data processing jobs you will run on it. You can use any Internet-connected computer running Windows. The lab procedures assume you are using Windows 8.1, but they can be adapted for any of the following operating systems:

- Windows 7
- Windows 8
- Windows 8.1
- Windows 10
- Windows Server 2012
- Windows Server 2012 R2

You can download an evaluation edition of Windows 8.1 from https://www.microsoft.com/en-us/evalcenter/evaluate-windows-8-1-enterprise.

Note: If you do not have a Windows PC you can use for these labs, consider installing Windows in a virtual machine (VM), either using virtualization software on your own computer or in a virtual machine in Microsoft Azure.

Installing Microsoft Azure PowerShell and Visual Studio

The Microsoft Azure PowerShell module includes cmdlets for managing Microsoft Azure services, including Microsoft Azure HDInsight. Visual Studio, when used with the Microsoft Azure SDK provides a development environment for building Azure solutions; including HDInsight Big Data processing solutions.

Install PowerShell and Visual Studio

1. On the computer where you intend to perform the labs, start a web browser and navigate to http://azure.microsoft.com.
2. Click **Downloads**, and in the **Command-Line Tools** section, under **Windows PowerShell**, click **Install**.
3. Run the Microsoft Web Platform Installer and install the Microsoft Azure PowerShell module.
4. When PowerShell has been installed, in the Web Platform Installer, on the **Products** tab, next to **Visual Studio Community 2015 with Microsoft Azure SDK 2.7**, click **Add**. Then click **Install** and accept the license agreement to install Visual Studio with the Azure SDK. (This will take a considerable while, so now would be a good time to resume viewing the course content until installation is complete!)

   Note: If you already have Visual Studio 2013 or 2015 installed on your computer, just install the **Microsoft Azure SDK Tools** for the version of Visual Studio you have installed.

5. When the installation is complete, restart the computer and log back in.

> **Note:** If the installation is interrupted or fails for any reason, restart your computer and re-launch the Web Platform Installer at http://www.microsoft.com/web/downloads/platform.aspx.

6. On the Start screen, view all apps. Then right-click each of the following apps and click **Pin to Taskbar**:
   - **Windows PowerShell**
   - **Visual Studio 2015** (or whatever version you had previously installed)

Configure PowerShell

1. On the desktop, on the task bar, right-click the PowerShell icon, right-click **Windows PowerShell**, and click **Run as administrator**. Click **Yes** when prompted, and then in the PowerShell command line, enter the following command:

   ```
   Set-ExecutionPolicy UnRestricted -Force
   ```

2. Close the PowerShell command line window.
3. On the task bar, click the PowerShell icon Then in the PowerShell command line, enter the following command to identify the Microsoft Azure subscriptions currently associated with PowerShell (there should be none):

   ```
   Get-AzureSubscription
   ```

4. If any Microsoft Azure subscriptions are listed, enter the following command to remove each one (replacing *<subscription_name>* with the Microsoft Azure subscription name):

   ```
   Remove-AzureSubscription "<subscription_name>" -Force
   ```

5. In the PowerShell command line, enter the following command to obtain a new credentials certificate for your Microsoft Azure subscription:

   ```
   Get-AzurePublishSettingsFile
   ```

6. If you are prompted to sign into Microsoft Azure, sign in using the Microsoft account associated with your Azure subscription. Then, in the new tab that is opened, save the publish settings file to a folder on your computer – noting the file name and the path to the folder where it is saved.
7. In the PowerShell editor, in the PowerShell command line, enter the following command to associate your Microsoft Azure subscription with the PowerShell environment (replacing *<path_to_credentials.publishsettings>* with the path to the file you downloaded in the previous step):

   ```
   Import-AzurePublishSettingsFile "<path_to_credentials.publishsettings>"
   ```

8. In the PowerShell command line pane, enter the following command to verify that your Microsoft Azure subscription is now associated with PowerShell:

   ```
   Get-AzureSubscription
   ```

9. Close the Windows command line window.

Configure Visual Studio

1. On the taskbar, click the Visual Studio icon. When prompted, do not sign in.
2. Accept the default layout options, and start Visual Studio.
3. Wait for Visual Studio to start, and then close it.

Installing Microsoft Power BI Desktop

Microsoft Power BI Desktop (formerly *Power BI Designer*) is a tool for data analysis and reporting. You can use Power BI Desktop to analyze and visualize data generated by HDInsight jobs.

Install Power BI Desktop

Note: If you already have Power BI Desktop installed on your computer, skip this task.

1. In a web browser, navigate to https://powerbi.microsoft.com, and in the **Downloads** section near the bottom of the page, download **Microsoft Power BI Desktop**.
2. When the download has completed, run the downloaded installation program to install Power BI Designer.
3. When installation is finished, on the Start screen, view all apps and then right-click **Microsoft Power BI Desktop** and click **Pin to Taskbar**. This will make it easier to find the app when you want to start it.
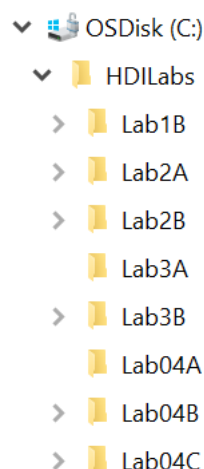
Note: Power BI Desktop is the released version of the Power BI Designer preview tool used in the demonstrations for this course. The tool has been renamed and updated, and looks cosmetically different from the preview version; but still provides the same functionality as shown in the demonstrations.

Installing the Lab Files

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

Download the Lab Files

1. On the **Course Info** page for this course at https://courses.edx.org/courses/course-v1:Microsoft+DAT202.1x+1T2016/info, download the **Lab Files** handout, saving the .zip file to a folder on your computer.
2. Extract the **HDILabs.zip** archive you downloaded to **C:** to create the following folder structure:

    ⌄ 🖥️ OSDisk (C:)
        ⌄ 📁 HDILabs
            > 📁 Lab1B
            > 📁 Lab2A
            > 📁 Lab2B
               📁 Lab3A
            > 📁 Lab3B
               📁 Lab04A
            > 📁 Lab04B
            > 📁 Lab04C

3. Right-click the **C:\HDILabs** folder and click **Properties**. Then clear the **Read-only** option for the folder and click **OK**. When prompted, apply the changes to the folder and its subfolders and files.