

Implementing Predictive Analytics with Spark 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

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 the frequently asked questions in the Azure sign-up page.

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://aka.ms/edx-dat202.3x-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 the Azure Cross-Platform Command-line Interface (CLI)

The Azure CLI is a command line interface for working with Azure services. There are versions available for Windows, Linux, and Mac OS X.

1. In a web browser, navigate to <https://azure.microsoft.com/downloads>.
2. In the **Command-line Tools** section, under **Azure command-line interface**, click **Documentation** and follow the instructions to install the Azure CLI on your client operating system.

Note: In up-to-date Windows and Mac OS X systems on which Node.js is already installed, you should simply download and run the installer package, and on Linux you may be able to use a package manager tool such as **npm** from the command line. On some Mac OS X systems and Linux distributions, you may need to install Node.js before installing the Azure CLI. Note that the specific steps to install these packages may vary depending on your Linux distribution. For more information about installing Node.js, see <https://nodejs.org>.

3. Restart your computer after installing the Azure CLI.
4. Open a command window (for example, Windows command prompt, Bash, or Terminal) and enter the following command to verify that the Azure CLI is installed correctly:

```
azure help
```

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 HDInsight, 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.