Orchestrating Big Data Solutions with Azure Data Factory

Setup Guide

Overview

This course describes how to orchestrate big data solutions using Microsoft Azure Data Factory. In this lab, you will build data factory pipelines to transfer and transform data.

What You'll Need

- A Microsoft account
- A Microsoft Azure subscription
- A Microsoft Windows, Apple Macintosh, or Linux computer

Create an Azure Subscription

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

If you have never had an Azure subscription, you can create a 30-day free 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.

If you have previously had a free trial subscription that has now expired, you will <u>not</u> be able to sign up for a second free trial. In this case, you must sign up for a pay-as-you-go subscription. The lab instructions in this course are designed to minimize the cost of the Azure resources required to complete the hands-on exercises.

To set up a free trial subscription:

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 https://aka.ms/edx-dat223_3x-az 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 – don't worry, 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.

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).

Install a SQL Client Tool

In some exercises, you will need to query an Azure SQL Database instance to view the data it contains. To do this, you will need a client tool. If you do not already have a SQL Server client tool installed (such as Visual Studio or SQL Server Management Studio on Windows, or Talend Open Studio for Data Integration or Navicat for SQL Server on Linux / Mac OSX), you can follow the steps below to install the cross-platform SQL Server command line interface, which is an open source tool for working with SQL Server databases from Windows, Linux or Mac OSX. You can learn more at https://www.npmjs.com/package/sql-cli.

1. Browse to https://nodejs.org/en/download/ and follow the instructions to download and install the latest version of Node.js for your operating system (Windows, OSX, or Linux) and architecture (64-bit or 32-bit).
2. Open a Node.JS command line and enter the following command line to install the SQL Server Command line interface package:

```
npm install -g sql-cli
```

Note: Depending on the security configuration of your system, you may need to run this command as an administrator. On Linux or Mac, you can do this by prefixing the command line above with the **sudo** command and entering the administrator password when prompted. On Windows, you can do this by opening a command line as Administrator.

3. Verify installation by viewing the help information for the SQL Server command line interface using the following command:

```
mssql -h
```

Install the Lab Files

To complete the labs in this course, you will need to download and extract the lab files from the course GitHub repository.

1. Download the lab files zip archive from https://aka.ms/edx-dat223.3x-labs.
2. Extract the lab files to a folder on your computer.