DAT218x

Cleansing Data with Data Quality Services

Lab 0-1 | Getting Started

Estimated time to complete this lab is 60 minutes

Overview

In this lab, you will provision a Microsoft Azure Virtual Machine (VM) that will be used by all labs in this course. Once the VM is provisioned, you will complete the setup required to support the labs.

The labs in this course are accumulative. You cannot complete the following labs if this lab has not been successfully completed.

What You'll Need

To complete this lab, you will need the following:

- High-speed and reliable internet connectivity (for remote connections to the VM)
- A second monitor is recommended (for the Remote Desktop connection)
- A Microsoft account (such as one used for outlook.com, Hotmail, or other Microsoft services)
- A Microsoft Azure subscription
- The lab files for this course (available for download from GitHub, as described in this lab)

Creating a Free Trial Azure Subscription

If you already have an 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, refer to https://aka.ms/edx-dat218-faq-az. Note that the free trial is not available in all regions.

If you already have a Microsoft account that has not already been used to sign up for a free Microsoft 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.

After you've created a Microsoft account, browse to https://aka.ms/edx-dat218-free-trial-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:

- Enter your cellphone number and have Microsoft send you a text message to verify your identity
- Enter the verification code sent to you
- 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 expressly decide to keep it active.

Exercise 1: Provisioning an Azure VM

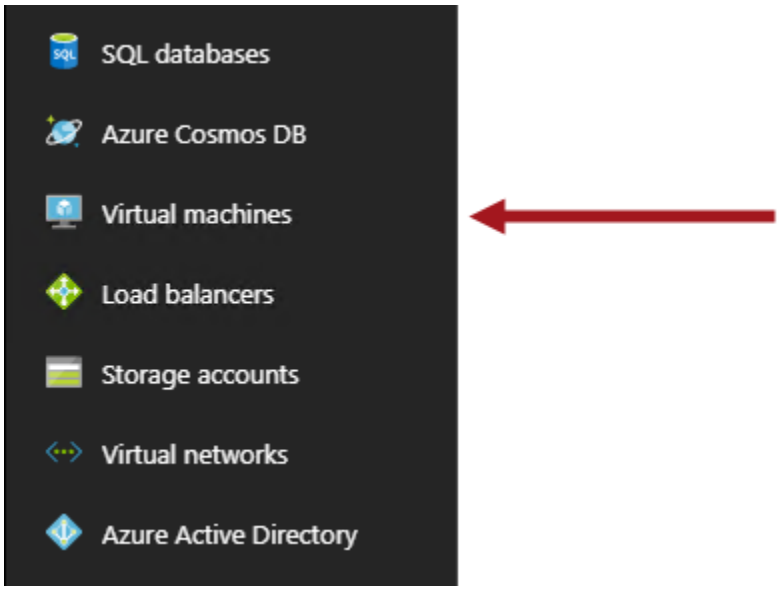
In this exercise, having signed in to the Azure Portal by using your Azure subscription, you will provision an Azure VM to support all labs for this course.

The Azure VM should be stopped when you have completed a lab so that your subscription is not charged (for free trial subscriptions, this will ensure you will have sufficient credits left to complete the labs over the duration of the course).

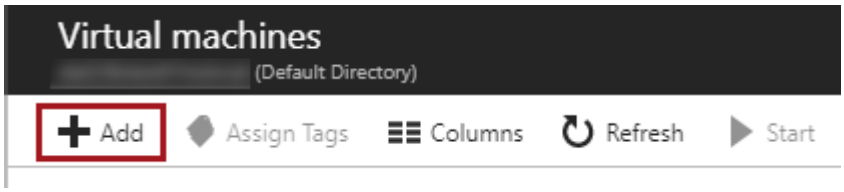
Provisioning an Azure VM

In this task, you will sign in to the Azure Portal, and then provision an Azure VM.

1. In a web browser, navigate to https://portal.azure.com.
2. Sign in to the **Azure Portal** by using your subscription.
3. In the left pane, select **Virtual Machines**.



4. In the **Virtual Machines** blade, click **Add**.



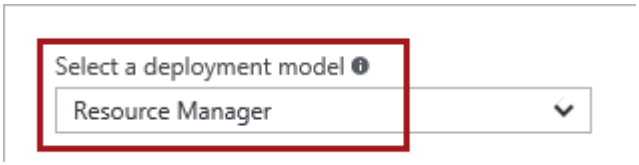
5. In the **Compute** blade, in the search box, enter the search text **Free SQL Server 2017**, and then press **Enter**.

6. Select the
Free SQL Server License: SQL Server 2017 Developer on Windows Server 2016 image.

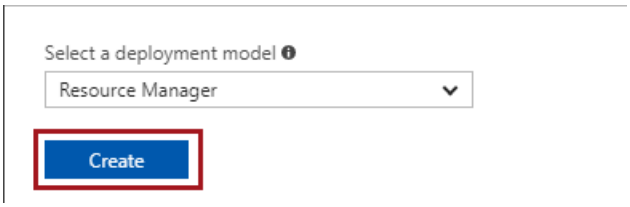
NAME	PUBLISHER	CATEGORY
Free SQL Server License: SQL Server 2017 Developer on Ubuntu Server 16.04 LTS	Microsoft	Linux based
Free SQL Server License: SQL Server 2017 Developer on Red Hat Enterprise Linux 7.4 (RHEL)	Microsoft	Linux based
Free SQL Server License: SQL Server 2017 Express on Windows Server 2016	Microsoft	Windows based
Free SQL Server License: SQL Server 2017 Developer on Windows Server 2016	Microsoft	Windows based
Free SQL Server License: SQL Server 2017 Developer on SUSE Linux Enterprise Server (SLES) 12 SP2	Microsoft	Linux based
Free SQL Server License: SQL Server 2017 Express on SUSE Linux Enterprise Server (SLES) 12 SP2	Microsoft	Linux based
Free SQL Server License: SQL Server 2017 Express on Ubuntu Server 16.04 LTS	Microsoft	Linux based
Free SQL Server License: SQL Server 2017 Express on Red Hat Enterprise Linux 7.4 (RHEL)	Microsoft	Linux based

7. In the image blade, review the text that describes the virtual machine setup.

8. In the lower section of the blade, in the **Select a Deployment Model** dropdown list, ensure that **Resource Manager** is selected.



9. To provision the virtual machine, click **Create**.



10. Notice that the **Create Virtual Machine** blade opens, and that also the **Basics** blade (step 1) opens.

11. In the **Basics** blade, in the **Name** box, enter a name for the virtual machine (this will become the name of the machine to which you will connect).

12. In the **VM Disk Type** dropdown list, select **HDD**.

13. In the **User Name** box enter **VM-Admin**.

14. In the password boxes, enter and confirm an appropriate password.

    This will become the machine administrator account password. Note that the password must be at least 12 characters in length, and must have three of the following: one lower case character, one upper case character, one number, or one special character.

    Be sure to permanently record these credentials, as you will be required to use them to sign in every time you will connect to the VM.

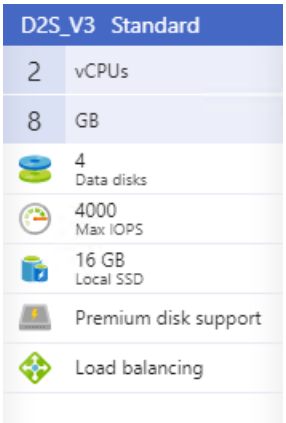
15. In the **Resource Group** box, enter **Lab**.

16. In the **Location** box, select a data center that is near you.

17. Click **OK**.

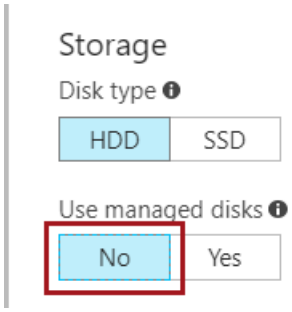
18. In the **Size** blade (step 2), scroll down to locate and select a **Standard** size VM which provides at least 2 vCPUs and 8GB RAM (like **D2S_V3**, if available).

    The labs in this course will not require excessive storage, memory or processing. Also, you will be prompted to deallocate your VM between labs, and so the monthly cost you see will only apply when the VM is running.



19. Click **Select**.

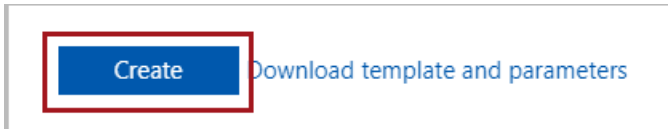
20. In the **Settings** blade (step 3), for **Use Managed Disks**, select **No**.



21. Click **OK**.

22. In the **SQL Server Settings** blade (step 4), to accept the default settings, click **OK**.

23. In the **Summary** blade, click **Create**.



*While the VM is being provisioned, you are directed to the **Azure Portal** dashboard.*

24. On the **Azure Portal** dashboard, notice the tile displaying the status of the deployment process.

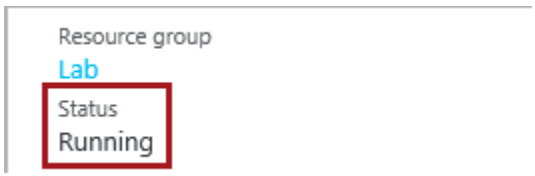
The deployment usually takes 15-20 minutes to complete, and this time depends largely on the VM size selected. The VM blade will open when the deployment completes.

You cannot proceed to the next task until the deployment completes.

Connecting to the VM

In this task, once the VM has successfully deployed, you will connect to the VM.

1. In the **Azure Portal**, notice that the VM blade automatically opens, and that the VM status is **Running**.



*You are charged when the VM status is **Running**, but you are not charged—except for a relatively smaller storage cost—when the VM status is **Stopped (Deallocated)**.*

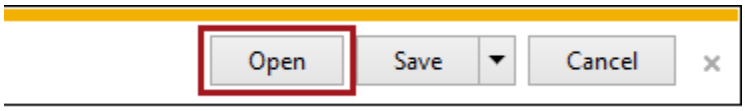
At the end of this lab, there will be instructions to guide you on how to stop and optionally deallocate the VM. You should consider doing this if you choose to commence the next lab at a much later time.

2. To connect to the VM, click **Connect**.

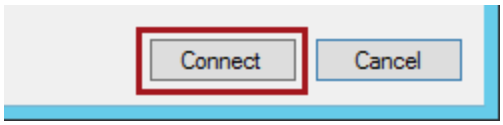
A Remote Desktop File (.rdp) file is downloaded to your computer.

This file can be used to reconnect to the remote desktop session, but note that if you deallocate the VM and later re-start the VM, it will be likely that a different IP address will be assigned.

3. If prompted by the web browser to open the Remote Desktop File, click **Open**, otherwise, locate the downloaded file, and then double-click it.



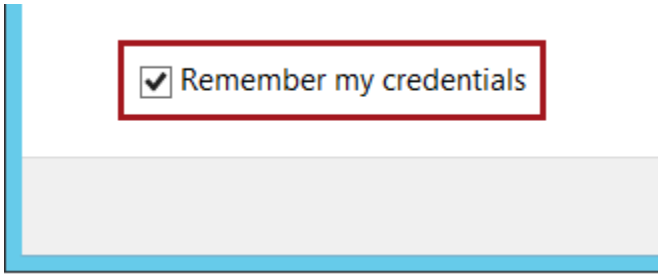
4. If prompted to connect to the unknown publisher, click **Connect**.



*You need to enter the VM administrator credentials. If the authentication window defaults to an existing account, you will need to select **More Choices**, and then select **Use a Different Account**.*

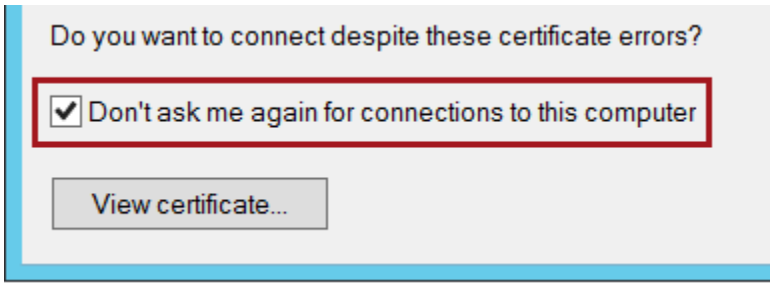
5. In the **Windows Security** window, enter the VM admin credentials (retrieved from **MySolution.txt**: VM admin user name, and VM admin password).

6. Check the **Remember My Credentials** checkbox.



7. Click **OK**.

8. In the **Remote Desktop Connection** window, check the **Don't Ask Me Again for Connections to This Computer** checkbox.



9. Click **Yes**.

10. If you have a second monitor, maximize the Remote Desktop window inside a single monitor.

Exercise 2: Setting Up the Azure VM

In this exercise, you will complete several VM setup tasks.

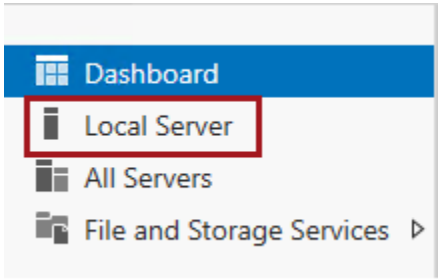
Configuring the Server

In this task, you will configure the server to support the lab experience.

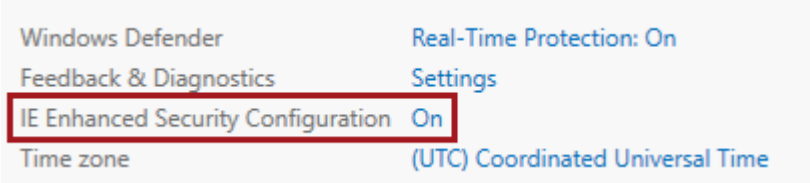
1. In the Remote Desktop window, when the **Networks** panel opens on the right, to ensure the machine is not discoverable by other machines, click **No**.

2. Wait until **Server Manager** opens (in a new VM instance, it is set to open automatically).

3. In **Server Manager**, in the left pane, select **Local Server**.

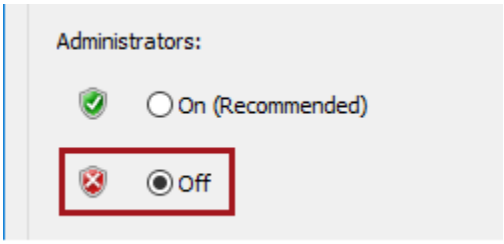


4. In the **Properties** pane, notice the **IE Enhanced Security Configuration** is set to **On**.



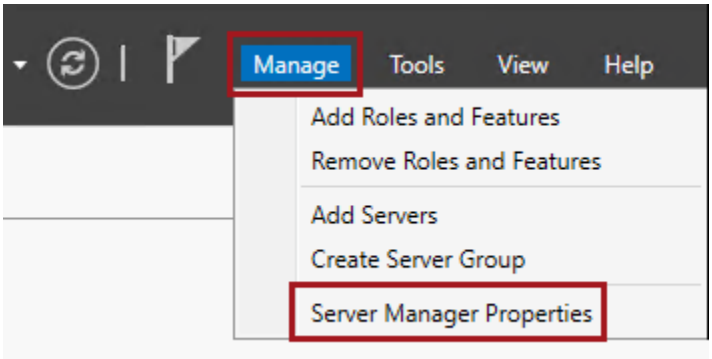
5. Click the **On** link.

6. In the window, for **Administrators**, select the **Off** option.

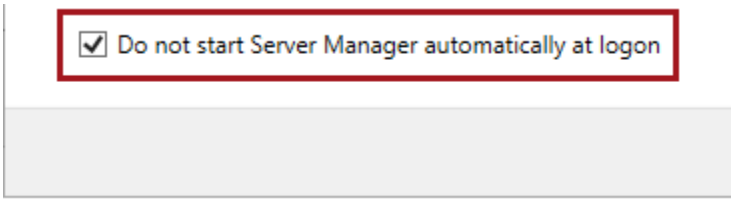


7. Click **OK**.

8. Located at the top-right corner, select **Manage**, and then select **Server Manager Properties**.

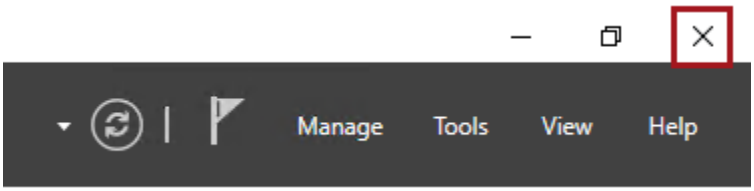


9. In the window, check the **Do Not Start Server Manager Automatically at Logon**.



10. Click **OK**.

11. To close Server Manager, located at the top-right corner, click **X**.

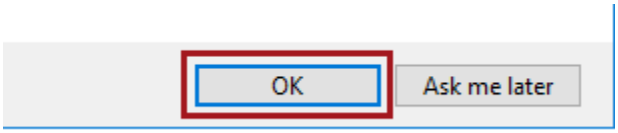


Installing the Lab Resources

In this task, you will download and extract the lab resources that support the labs.

1. To open Internet Explorer, on the taskbar, click the **Internet Explorer** shortcut.

2. In the **Internet Explorer 11** window, to accept the recommended settings, click **OK**.

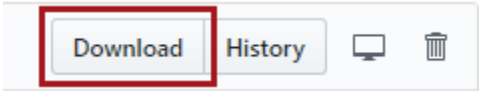


3. Maximize the Internet Explorer window.

4. In the Internet Explorer **URL** box, enter https://github.com/MicrosoftLearning/Cleansing-Data-with-DQS.

5. On the web page, click the **DAT218x-LabResources.zip** link.

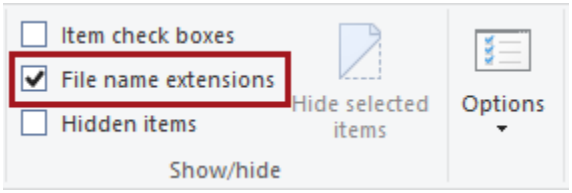
6. To download the zip file, click **Download**.



7. Download the file to **F:**.

8. When downloaded, open File Explorer.

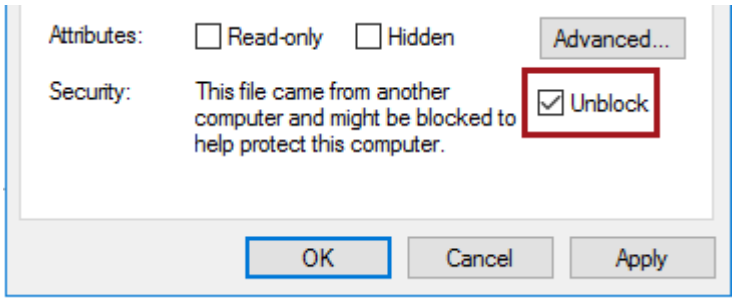
9. In the File Explorer window, on the **View** ribbon, check **File Name Extensions**.



10. Navigate to **F:**.

11. Right-click the **DAT218x-LabResources.zip** file, and then select **Properties**.

12. In the window, check **Unblock**.



13. Click **OK**.

14. To extract the file content, right-click the **DAT218x-LabResources.zip** file, and then select **Extract All**.

15. Extract the files to **F:**.

16. Verify that you have the **F:\Labs** folder.

17. Optionally, delete the **F:\DAT218x-LabResources.zip** file.

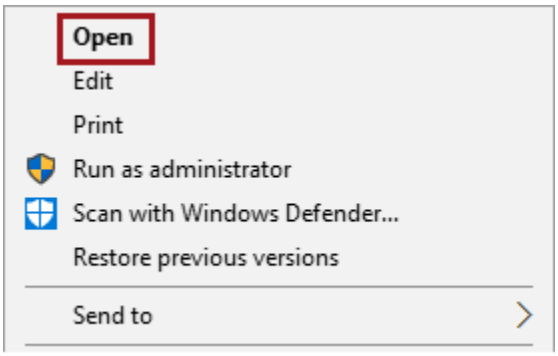
18. Close Internet Explorer (GitHub).

Creating the Lab Database

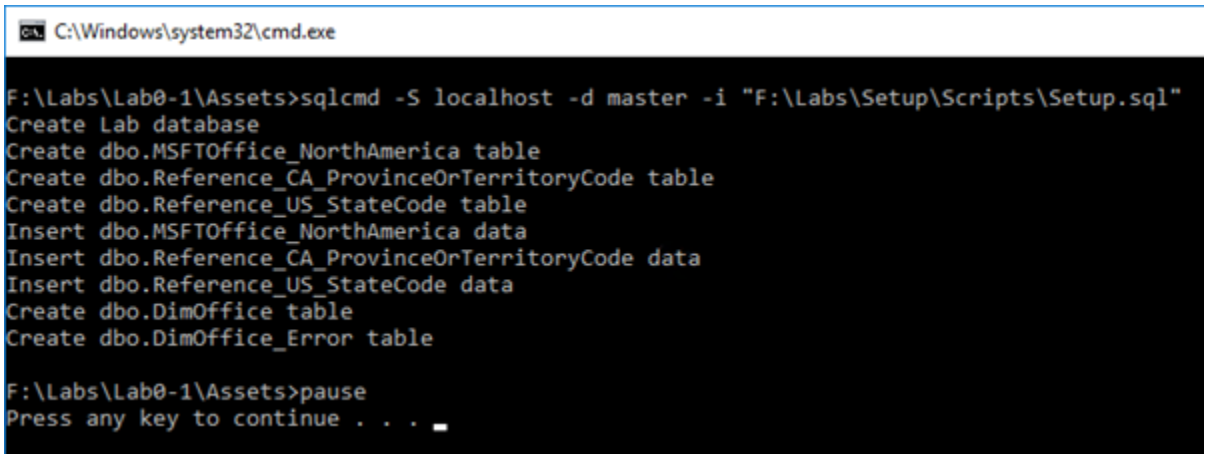
In this task, you will run a script to create the lab database.

1. Navigate to the **F:\Labs\Lab0-1\Assets** folder.

2. Right-click the **Setup.cmd** file, and then select **Open**.



*The setup creates the **Lab** database and various database objects to support the labs.*



3. When the script execution completes, verify that the you see the same feedback which describes that the **Lab** database and tables were created.

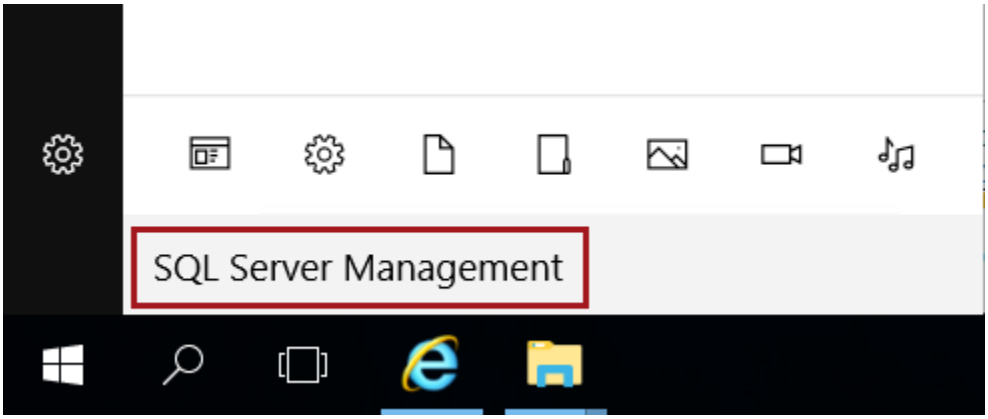
   If you do not see the same feedback as the image above, you will need to troubleshoot why the database and tables were not created. You cannot start the next lab if this database has not been successfully created.

4. Press any key to close the command window.

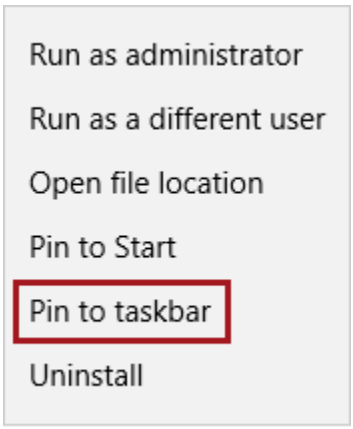
Configuring SQL Server Management Studio

In this task, you will configure SQL Server Management Studio (SSMS). This tool will be required to explore database, and also to execute scripts.

1. To add a shortcut to the taskbar, at the bottom-left corner, click the **Windows** icon, and then commence typing **SQL Server Management**.



2. In the **Apps** section, when the search result appears, right-click **Microsoft SQL Server Management Studio 17**, and then select **Pin to Taskbar**.

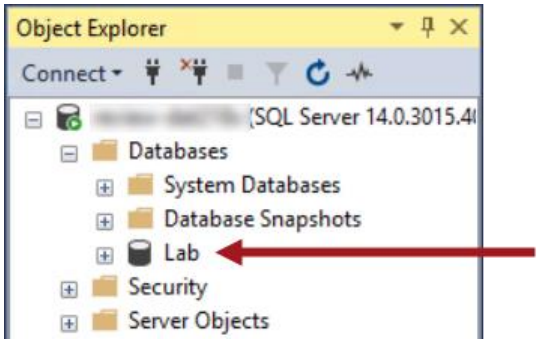


3. Return to the desktop, and then click the **SQL Server Management Studio** shortcut.

4. In the **Connect to Server** window, click **Connect**.

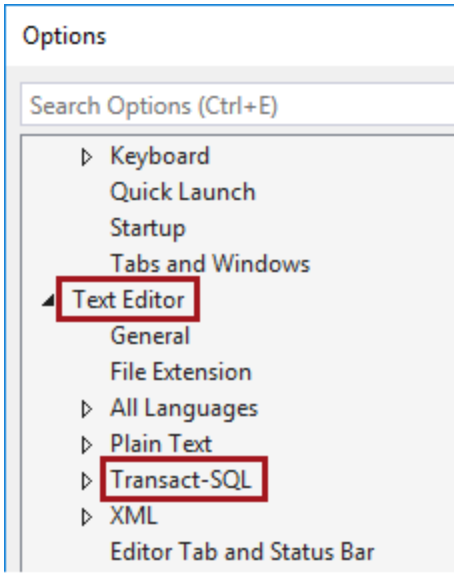
5. To verify that the **Lab** database was created, in **Object Explorer** (located at the left), expand the **Databases** folder.

6. Verify that the **Lab** database is listed.

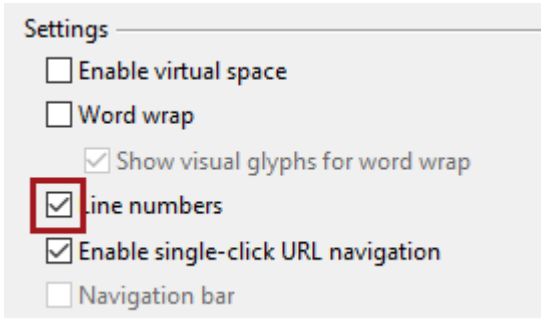


7. To configure the SSMS environment, on the **Tools** menu, select **Options**.

8. In the **Options** window, in the left pane, expand **Text Editor**, and then select **Transact-SQL**.



9. Check the **Line Numbers** checkbox.



10. Click **OK**.

11. To close SQL Server Management Studio, on the **File** menu, select **Exit**.

Installing SQL Server Tools

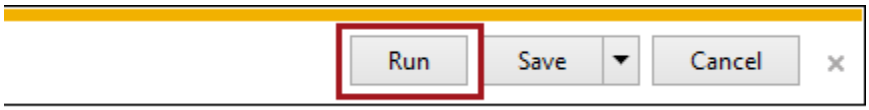
In this task, you will install SQL Server Data Tools (SSDT). This tool is required to develop and execute Integration Services (SSIS) packages.

1. In Internet Explorer, navigate to **https://aka.ms/edx-dat218-ssdt-sql**.

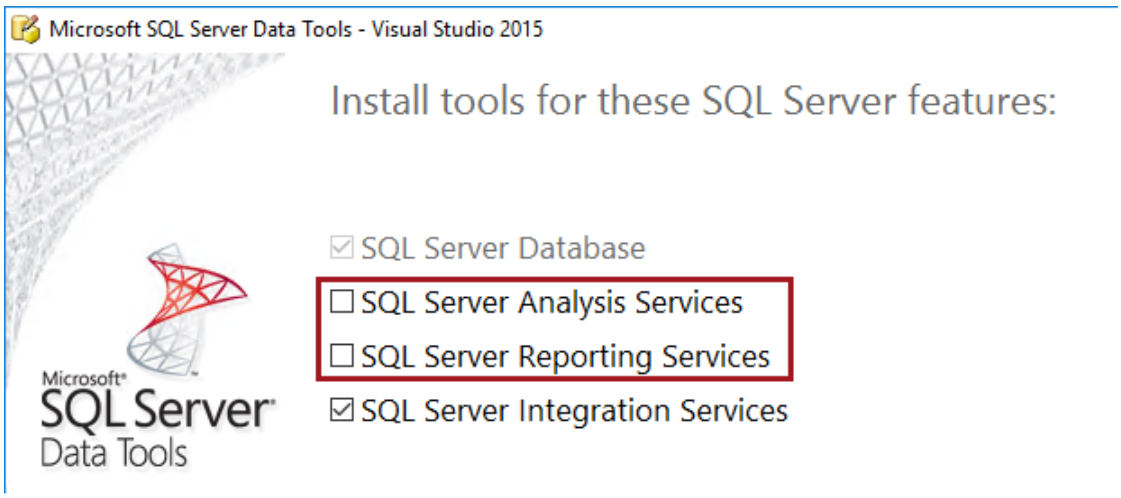
   Tip: You can copy-and-paste the URL into the Remote Desktop window.

2. Click the **Download SQL Server Data Tools** link.

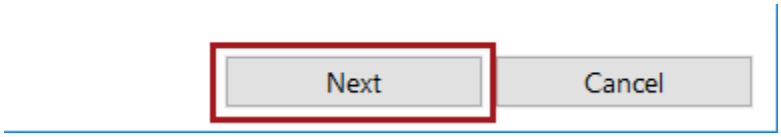
3. When prompted by Internet Explorer to run the **SSDTSetup.exe** file, click **Run**.



4. In the installation window, uncheck both the **SQL Server Analysis Services** and **SQL Server Reporting Services** checkboxes.

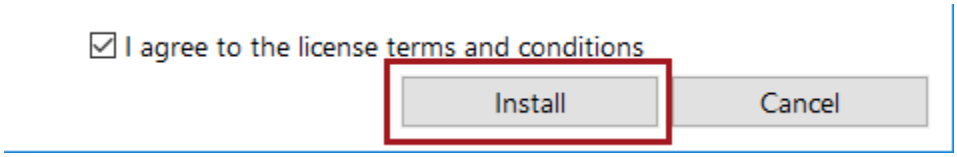


5. Click **Next**.



6. Read the license terms, and if you accept them, check the checkbox.

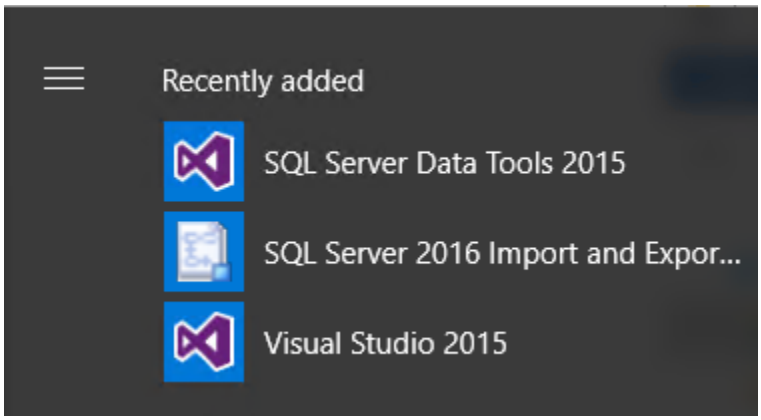
7. Click **Install**.



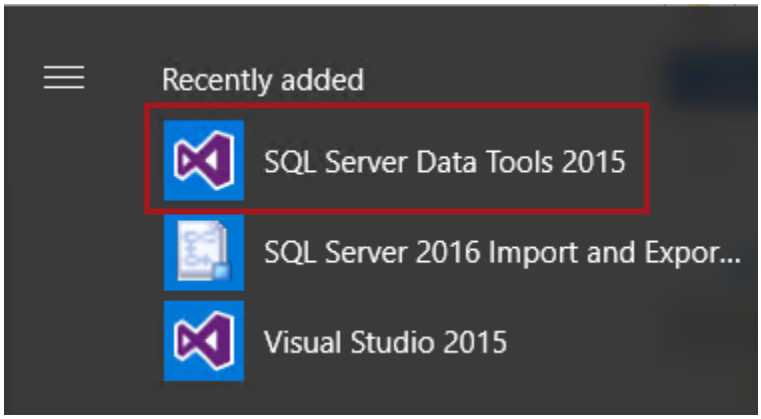
The installation usually takes 10-15 minutes to complete.

8. When the installation completes, click **Close**.

9. To launch SSDT, at the bottom-left corner, click the **Windows** icon, and notice the items in the **Recently Added** section.

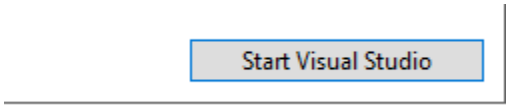


10. Select **SQL Server Data Tools 2015**.



11. In the Visual Studio getting started window, in the **Development Settings** dropdown list, select **Business Intelligence Settings**.

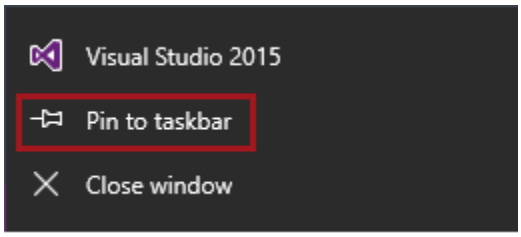
12. Click **Start Visual Studio**.



The installation usually takes 1-2 minutes to complete.

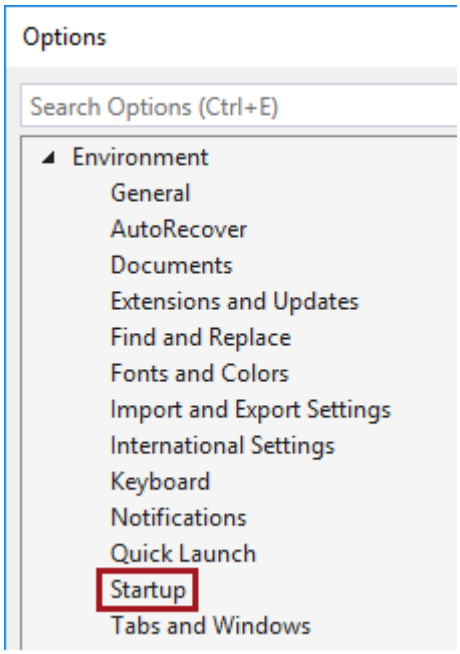
13. If you receive an error describing that Visual Studio was unable to load one or more of the requested types, click **OK**. (This will not impact on the completing the labs.)

14. To create a shortcut, on the taskbar, right-click the **Visual Studio 2015** icon, and then select **Pin to Taskbar**.

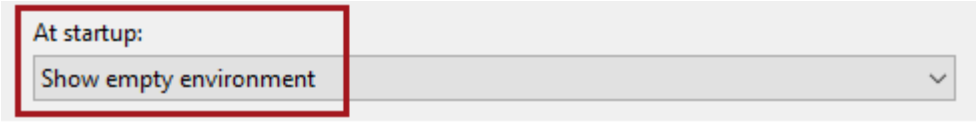


15. To configure the SSDT environment, on the **Tools** menu, select **Options**.

16. In the **Options** window, in the left pane, select the **Startup** page.



17. In the **At Startup** dropdown list, select **Show Empty Environment**.



18. Click **OK**.

19. To close SSDT, on the **File** menu, select **Exit**.

*You will work with SSDT to create Integration Services packages in **Lab 1-2**, and **Lab 2-2**.*

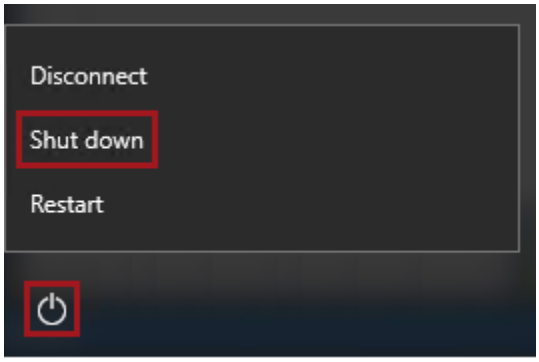
*You have now completed the lab. If you are not commencing the next lab, you should complete the **Finishing Up** exercise to shut down and stop the VM.*

Finishing Up

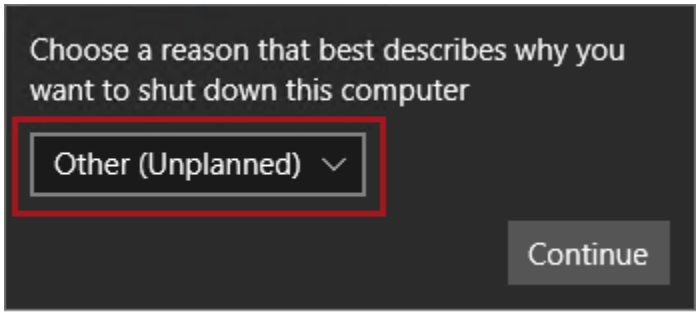
In this exercise, you will shut down and stop the VM.

1. Close all open applications.

2. Press the **Windows** key, and then in the **Start** page, located at the bottom-left, click the **Power** button, and then select **Shut Down**.

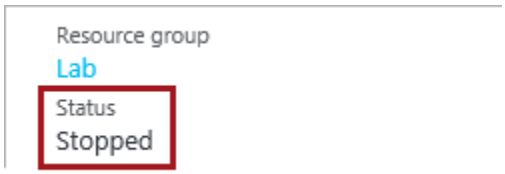


3. When prompted to choose a reason, select **Other (Unplanned)**.



4. Click **Continue**.

5. In the **Azure Portal** Web browser page, wait until the status of the VM updates to **Stopped**.

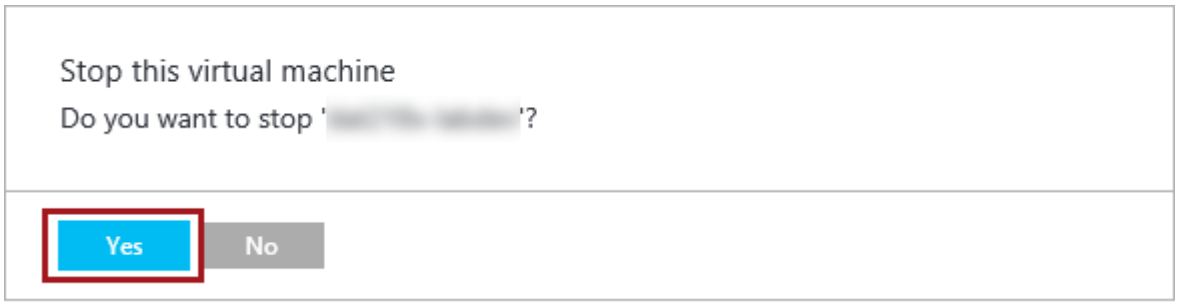


In this state, however, the VM is still billable.

6. Optionally, to deallocate the VM, click **Stop**.

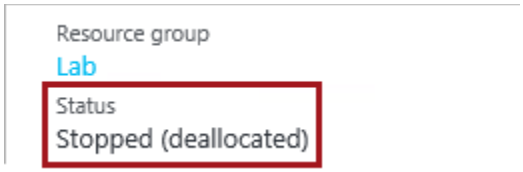
   Deallocation will take some minutes to complete, and also extends the time required to restart the VM. Consider deallocating the VM if you want to reduce costs, or if you choose to complete the next lab after an extended period.

7. When prompted to stop the VM, click **Yes**.



   The deallocation can take several minutes to complete.

8. Verify that the VM status updates to **Stopped (Deallocated)**.



   In this state, the VM is now not billable—except for a relatively smaller storage cost.

   Note that a deallocated VM will likely acquire a different IP address the next time it is started.

9. Sign out of the **Azure Portal**.