DAT218x

Cleansing Data with Data Quality Services

Lab 1-1 | Installing Data Quality Services

Estimated time to complete this lab is 20 minutes

Overview

In this lab, you will install Data Quality Services, and then explore the three DQS databases.

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

Exercise 1: Connecting to the VM

Go to the next exercise if you are already connected to the lab VM.

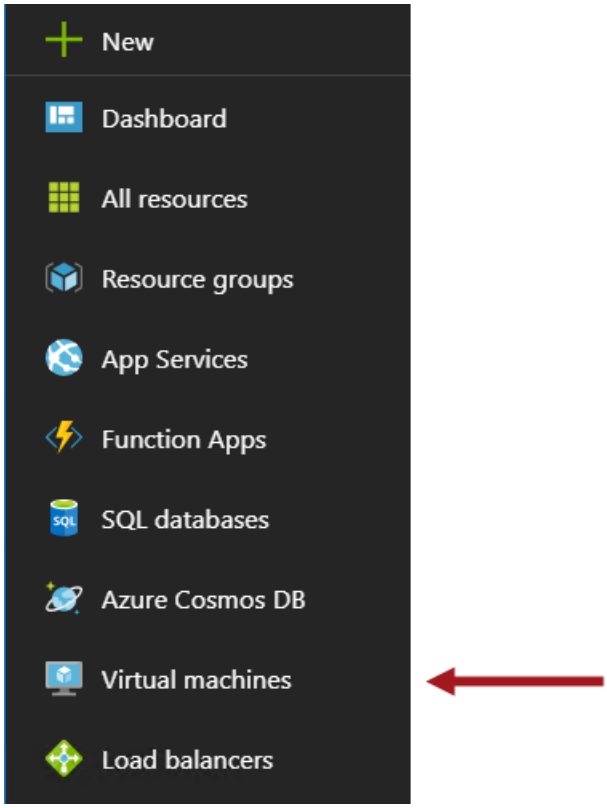
In this exercise, having signed in to the Azure Portal by using your Azure subscription, you will connect to the lab VM which you provisioned in **Lab 0-1**.

Connecting to the VM

In this task, you will sign in to the Azure Portal, and then connect to your lab VM.

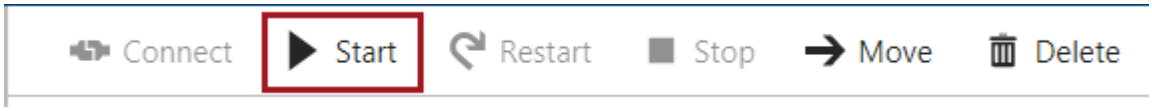
1. Sign in to the **Microsoft Azure Portal** by using your subscription.

2. In the left pane, select **Virtual Machines**.



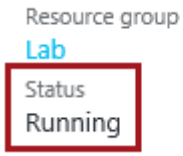
3. In the **Virtual Machines** blade, select the VM you provisioned in **Lab 0-1**.

4. In the VM blade, click **Start**.



5. Wait for the VM status to update to **Running**.

   It usually takes 1-2 minutes for the VM to start.



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

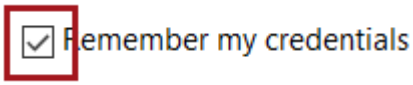
7. When prompted to open the Remote Desktop File, click **Open**.

8. 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**.*

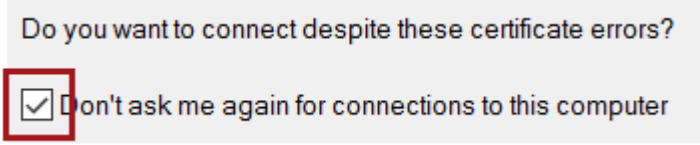
9. In the **Windows Security** window, enter the VM admin credentials used when provisioning the VM.

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



11. Click **OK**.

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



13. Click **Yes**.

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

Exercise 2: Installing Data Quality Services

In this exercise, you will install Data Quality Services, and then explore the three DQS databases.

Installing Data Quality Services

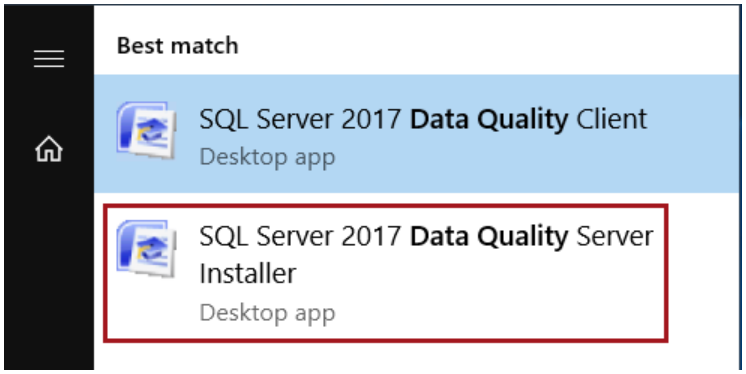
In this task, you will install Data Quality Services.

1. If you did not shutdown the VM after completing **Lab 0-1**, you must restart the VM now.

   *The VM setup you completed in **Lab 0-1** requires a reboot before additional software, including DQS, can be installed.*

2. At the bottom-left corner, click the **Windows** icon, and then commence typing **Data Quality**.

3. When the search results appear, click **SQL Server 2017 Data Quality Server Installer**.



   The two applications were installed by the SQL Server setup process used to create the Azure VM.

4. In the installer window, when prompted to enter the Database Master Key, enter **Pass@word1**, and then press **Enter**.

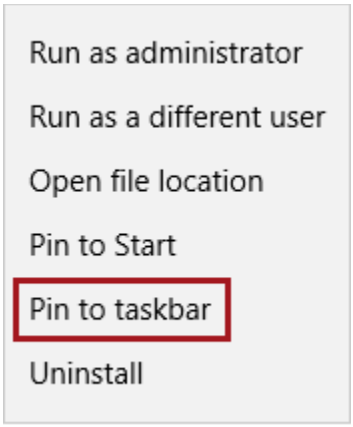
5. When prompted to re-enter the key, enter **Pass@word1** again, and then press **Enter**.

   The installation usually takes about 2-3 minutes to complete.

6. When the installation has completed, press any key to close the installer window.

7. To open Data Quality Client, click the **Windows** icon, and then commence typing **Data Quality**.

8. In the **Apps** section, when the search results appear, right-click **SQL Server 2017 Data Quality Client**, and then select **Pin to Taskbar**.

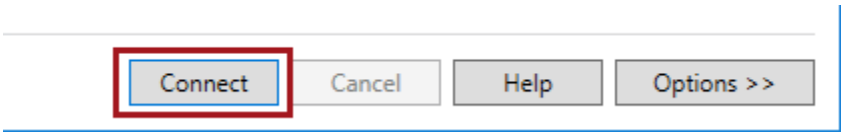


9. Return to the desktop, and then notice the **Data Quality Client** shortcut.

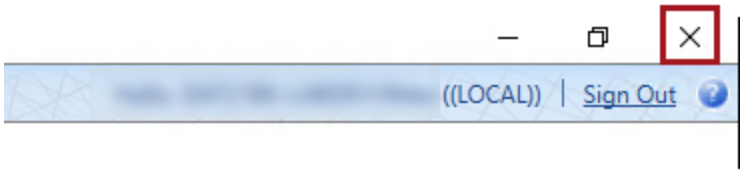
10. To open Data Quality Client, click the **Data Quality Client** taskbar shortcut.

11. In the **Connect to Server** window, in the **Server Name** dropdown list, select **(LOCAL)**.

12. Click **Connect**.



13. To close Data Quality Client, located at the top-right corner, click **X**.



You will work with Data Quality Client in all labs.

Exploring the Data Quality Services Databases

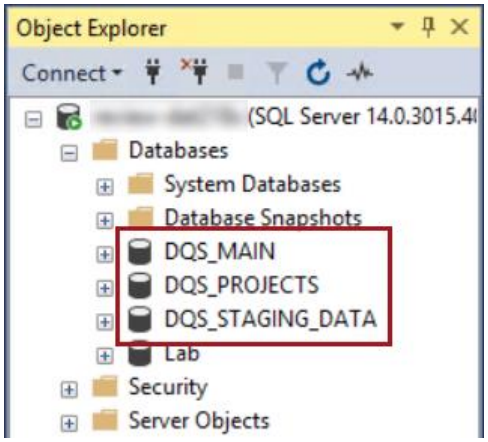
In this task, you will explore the three DQS databases.

1. To open SQL Server Management Studio, click the **SQL Server Management Studio** taskbar shortcut.

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

3. To verify that the DQS databases were created, in **Object Explorer** (located at the left), expand the **Databases** folder.

4. Verify that three DQS databases are listed.



*The Data Quality Server is primarily implemented by two SQL Server databases: **DQS_MAIN** and **DQS_PROJECTS**.*

*The **DQS_STAGING_DATA** is an empty database provisioned to support preparing or managing data for import or export by the server. You will not use this database in the labs.*

5. Expand the **DQS_MAIN** database, and then expand the **Tables** folder.

   *Ordinarily, you will not work directly in the **DQS_MAIN** database. All data quality activities are typically done in the Data Quality Client application.*

6. Notice the tables designed to store server metadata, and also tables in the **KnowledgeBase1000000** schema.

   *Each knowledge base you create will result in the creation of a dedicated schema. The first knowledge base is in fact the sample **DQS Data** knowledge base. You should not directly manipulate the data in these tables.*

7. Collapse the **Tables** folder.

8. Expand the **Programmability** folder, and then the **Assemblies** folder.

9. Notice the large list of assemblies, and specifically those prefixed **Microsoft.Ssdqs**.

   The Data Quality Server functionality is implemented by .NET CLR stored procedures.

10. Collapse the **Assemblies** folder.

11. Within the **Programmability** folder, expand the **Stored Procedures** folder.

12. Notice the large list of stored procedures that implement the Data Quality Server functionality.

13. Explore also the **DQS_PROJECTS** database also.

    This database is designed to store data in relation to Data Quality Projects. Presently, the database is almost empty.

14. Explore also the **DQS_STAGING** database.

    *The database is empty, unless you choose to create tables. In this lab, you will use the **Lab** database to store and manage data.*

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

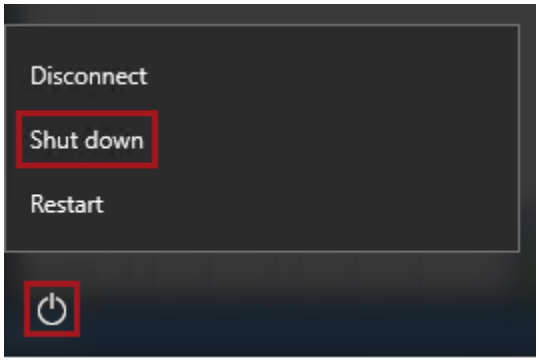
    *You have now completed the lab. If you are not commencing the next lab, you should complete the **Finishing Up** exercise to shut down and stop the VM.*

Finishing Up

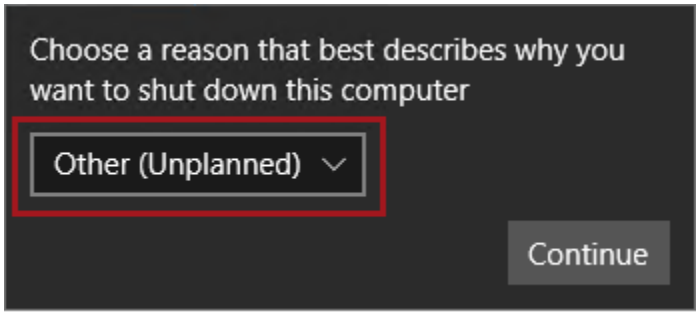
In this exercise, you will shut down and stop the VM.

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**.

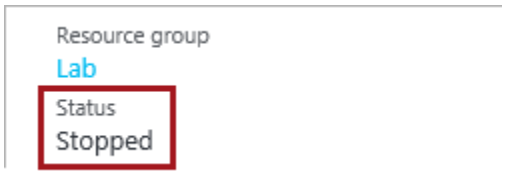


3. When prompted to choose a reason, select **Other (Unplanned)**.



4. Click **Continue**.

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

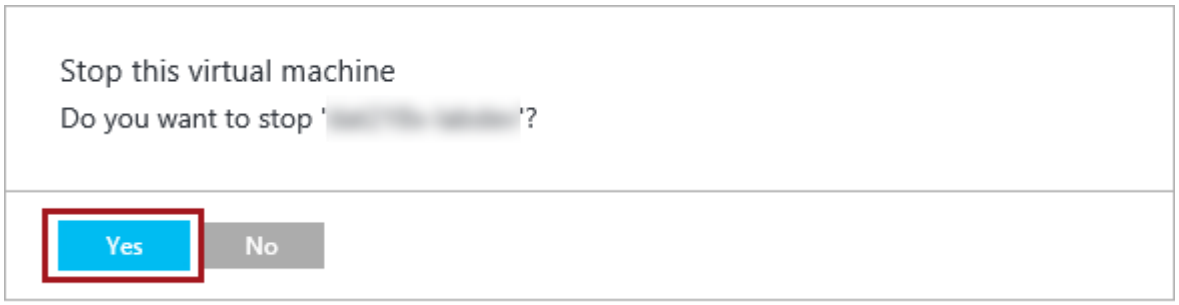


In this state, however, the VM is still billable.

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

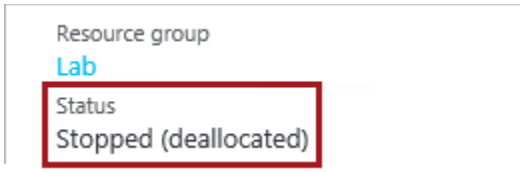
   Deallocation will take some minutes to complete, and also extends the time required to restart the VM. Consider deallocating the VM if you want to reduce costs, or if you choose to complete the next lab after an extended period.

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**.