Lab - Implement and monitor autoscaling of session hosts

Student lab manual

Lab dependencies

- An Azure subscription you will be using in this lab.
- A Microsoft Entra user account with the Owner or Contributor role in the Azure subscription you will be using in this lab and with the permissions sufficient to join devices to the Entra tenant associated with that Azure subscription.
- The lab *Deploy host pools and session hosts by using the Azure portal (Entra ID)* completed
- The lab *Manage host pools and session hosts by using the Azure portal (Entra ID)* completed
- The lab *Connect to session hosts (Entra ID)* completed

Estimated Time

45 minutes

Lab scenario

You have an Azure Virtual Desktop environment which usage changes on regular basis. You want to minimize the cost by leveraging the autoscale scaling plans functionality.

Objectives

After completing this lab, you will be able to:

- implement and evaluate Azure Virtual Desktop autoscaling

Lab files

- None

Instructions

Exercise 1: Implement Azure Virtual Desktop autoscale scaling plans

The main tasks for this exercise are as follows:

1. Assign the required RBAC role to an Azure Virtual Desktop service principal
2. Stop and deallocate all session hosts
3. Adjust the host pool settings
4. Create a scaling plan
5. Evaluate the autoscaling functionality
6. Disable host pool autoscaling

Task 1: Assign the required RBAC role to an Azure Virtual Desktop service principal

> **Note**: For autoscale plans to work, you need to grant the Azure Virtual Desktop service principal the permissions to manage the power state of the session host VMs. These permissions can be granted by using the built-in **Desktop Virtualization Power On Off Contributor** RBAC role. It is important to keep in mind that the role assignment must be performed at the subscription scope. Assigning this role at any level lower than your subscription, such as the resource group, host pool, or VM, will prevent autoscale from working properly.

> **Note**: This role is different from the one (**Desktop Virtualization Power On Contributor**) used in the lab *Manage host pools and session hosts by using the Azure portal (Entra ID)*, which was required to support the *Start VM on Connect* functionality.

1. If needed, from the lab computer, start a web browser, navigate to the Azure portal and sign in by providing credentials of a user account with the Owner role in the subscription you will be using in this lab.

   > **Note**: Use the credentials of the `User1-` account listed on the Resources tab on the right side of the lab session window.

2. From the lab computer, in the web browser displaying the Azure portal, start a PowerShell session in the Azure Cloud Shell.

   > **Note**: If prompted, in the **Getting started** pane, in the **Subscription** drop-down list, select the name of the Azure subscription you are using in this lab and then select **Apply**.

3. In the PowerShell session in the Azure Cloud Shell pane, run the following command to retrieve the value of the Id property of the Azure subscription you are using in this lab and store it in a variable `$subId`:

   ```
   $subId = (Get-AzSubscription).Id
   ```

4. Run the following command to create a $parameters variable, which stores a hash table that contains the values of the RBAC role definition name, Microsoft Entra application representing the **Azure Virtual Desktop** service principal, and the subscription scope:

   ```
   $parameters = @{
       RoleDefinitionName = "Desktop Virtualization Power On Off Contributor"
       ApplicationId = "9cdead84-a844-4324-93f2-b2e6bb768d07"
       Scope = "/subscriptions/$subId"
   }
   ```

5. Run the following command to create the RBAC role assignment:

   ```
   New-AzRoleAssignment @parameters
   ```

6. Close the Cloud Shell pane.

Task 2: Stop and deallocate all session hosts

> **Note**: To evaluate the autoscaling functionality, you will stop and deallocate all of the session hosts in the Azure Virtual Desktop environment.

1. From the lab computer, in the web browser displaying the Azure portal, search for and select **Azure Virtual Desktop** and, on the **Azure Virtual Desktop** page, in the vertical menu bar, in the **Manage** section, select **Host pools**.

2. On the **Azure Virtual Desktop | Host pools** page, in the list of host pools, select **az140-21-hp1**.

3. On the **az140-21-hp1** page, in the in the vertical menu bar, in the **Manage** section, select **Session hosts**.

4. On the **az140-21-hp1 | Session hosts** page, select the checkbox next to each session host name and then select **Stop**.

5. When prompted to confirm, in the **Stop session hosts** pop-up window, select **Stop**.

   > **Note**: You might need to select the ellipsis (**...**) icon in the toolbar to display the **Stop** button.

   > **Note**: Do not wait until the session hosts are stopped and deallocated, but instead proceed to the next task. Stopping and deallocating session hosts might take about 2 minutes.

Task 3: Adjust the host pool settings

> **Note**: When using autoscale for pooled host pools, you must have a configured MaxSessionLimit parameter for that host pool. In this lab, you will set it artificially low in order to facilitate illustrating the autoscaling functionality.

1. From the lab computer, in the web browser displaying the Azure portal, on the **az140-21-hp1** page, in the **Settings** section, select **Properties**.
2. On the **az140-21-hp1|Properties** page, in the **Max session limit** text box, enter **1**.
3. On the **az140-21-hp1|Properties** page, select **Save**.

Task 4: Create a scaling plan

1. From the lab computer, in the web browser displaying the Azure portal, search for and select **Azure Virtual Desktop**, on the **Azure Virtual Desktop** page, in the **Manage** section of the vertical navigation menu, select **Scaling plans**.

2. On the **Azure Virtual Desktop | Scaling plans** page, select **+ Create**.

3. On the **Basics** tab of the **Create a scaling plan** page, specify the following settings and select **Next: Schedules**:

Setting	Value
Subscription	the name of the Azure subscription you are using in this lab

Setting	Value
Resource group	the name of a new resource group **az140-412e-RG**
Scaling plan name	**az140-scalingplan412e**
Region	the name of the Azure region where you deployed the Azure Virtual Desktop environment
Friendly name	**az140-scalingplan412e**
Time zone	the local time zone of the Azure region where you deployed the Azure Virtual Desktop environment
Host pool type	**Pooled**

> **Note**: Leave the **Exclusion tag** property not set. In general, you can use this feature to exclude Azure VMs with arbitrarily set tags from autoscaling.

4. On the **Schedule** tab, select **+ Add schedule**.

> **Note**: Schedules allow you to define ramp-up hours, peak hours, ramp-down hours, and off-peak hours for week days and specify autoscaling triggers. Scaling plan must include an associated schedule for at least one day of the week.

5. On the **General** tab of the **Add a schedule** pane, adjust the default configuration to match the following settings and then select **Next**:

Setting	Value
Time zone	the local time zone of your Azure Virtual Desktop environment (based on the region you selected earlier in this task)
Schedule name	**week_schedule**
Repeat on	**7 selected** (select all days of the week)

> **Note**: The schedule effectively covers every day of the week, which will facilitate evaluating outcome of autoscaling.

6. On the **Ramp-up** tab of the **Add a schedule** pane, adjust the default configuration to match the following settings and then select **Next**:

Setting	Value
Start time (12 hour system)	your current time minus 1 hour
Load balancing algorithm	**Breadth-first**
Minimum percentage of hosts (%)	**30**

Setting	Value
Capacity threshold (%)	**60**

> **Note**: For pooled host pools, autoscale ignores existing load-balancing algorithms in your host pool settings, and instead applies load balancing based on your schedule configuration.

> **Note**: The **Minimum percentage of hosts** setting designates the minimum percentage of session host virtual machines to start for ramp-up and peak hours. For example, if **Minimum percentage of hosts** is specified as 30% and total number of session hosts in your host pool is 3, autoscale will ensure a minimum of 1 session host is available to accept user connections.

> **Note**: Autoscale rounds up to the nearest whole number.

> **Note**: The **Capacity threshold** setting is the percentage of used host pool capacity that will be considered to evaluate whether to turn on/off virtual machines during the ramp-up and peak hours. For example, if capacity threshold is specified as 60% and your host pool capacity is 1 session (with one host running), autoscale will turn on additional session hosts once the load of the host pool exceeds 60% (it's 100% in this case).

7. On the **Peak hours** tab of the **Add a schedule** pane, adjust the default configuration to match the following settings and then select **Next**:

Setting	Value
Start time (12 hour system)	your current time plus 1 hour
Load balancing algorithm	**Depth-first**
Capacity threshold (%)	**60**

> **Note**: The **Capacity threshold (%)** setting is shared between the **Ramp-up** and **Peak hours** settings.

8. On the **Ramp-down** tab of the **Add a schedule** pane, adjust the default configuration to match the following settings and then select **Next**:

Setting	Value
Start time (12 hour system)	your current time plus 2 hours
Load balancing algorithm	**Depth-first**
Minimum percentage of active hosts (%)	**10**
Capacity threshold (%)	**80**
Force logoff users	**No**
Stop VMs when	**VMs have no active or disconnected sessions**

> **Note**: The **Minimum percentage of active hosts (%)** setting designates the minimum percentage of session host virtual machines that you would like to get to for ramp-down and off-peak hours. For example, if **Minimum percentage of active hosts (%)** is set to 10% and

> total number of session hosts in your host pool is 3, autoscale will ensure a minimum of 1 session host is available to take user connections.

> **Note**: The **Capacity threshold (%)** setting designates the percentage of used host pool capacity that will be considered to evaluate whether to turn off virtual machines during the ramp-down and off-peak hours. For example, with 1 user connection and 3 hosts running, if capacity threshold is specified as 80%, autoscale would turn off 1 host (resulting in 50% used host pool capacity).

> **Note**: In general, autoscale will stop and deallocate session hosts according to the following rules:

- The used host pool capacity is below the capacity threshold.
- Turning off session hosts will not result in exceeding the capacity threshold.
- Autoscale only turns off session hosts with no user sessions on them, unless the scaling plan is in ramp-down phase and you've enabled the setting to force user logoff.
- Pooled autoscale will not turn off session hosts in the ramp-up phase to ensure that user experience is not impacted.

9. On the **Off-peak hours** tab of the **Add a schedule** pane, adjust the default configuration to match the following settings and then select **Add**:

Setting	Value
Start time (12 hour system)	your current time plus 3 hours
Load balancing algorithm	**Depth-first**
Capacity threshold (%)	**80**

> **Note**: The **Capacity threshold** setting is shared between the **Ramp-down** and **Off-peak hours** settings.

10. Back on the **Schedule** tab of the **Create a scaling plan** page, select **Next: Host pool assignments**.

11. On the **Host pool assignments** tab, in the **Select host pool** drop-down list, select **az140-21-hp1**, ensure that **Enable autoscale** checkbox is selected, and then select **Review + create**.

12. On the **Review + Create** page, select **Create**.

> **Note**: Wait for autoscale configuration to complete. This typically takes just a few seconds.

Task 5: Evaluate the autoscaling functionality

> **Note**: You will start by evaluating the **Ramp-up** settings.

1. From the lab computer, in the web browser displaying the Azure portal, search for and select **Azure Virtual Desktop** and, on the **Azure Virtual Desktop** page, in the vertical menu bar, in the **Manage** section, select **Host pools**.

2. On the **Azure Virtual Desktop | Host pools** page, in the list of host pools, select **az140-21-hp1**.

3. On the **az140-21-hp1** page, in the in the vertical menu bar, in the **Manage** section, select **Session hosts**.

4. On the **az140-21-hp1 | Session hosts** page, review the values of the **Power state** setting of session hosts and verify that one of them is listed as **Running**.

> **Note**: You might need to wait a couple of minutes before the first session hosts reaches the **Running** state.

> **Note**: This is expected, since according to the **Ramp-up** settings of the newly created scaling plan, at least one session host should be always online. At this point, the host pool capacity is 1 (since there is only 1 host running), but the used host pool capacity is 0%, since there are no user connections.

> **Note**: Next, you will evaluate the **Ramp-up** and **Peek hours** capacity threshold setting by launching a single user session. We can evaluate this even outside of **Peek hours** window, since the two stages share the same capacity threshold.

5. From the lab computer, start the Microsoft Remote Desktop client.

6. On the lab computer, in the **Remote Desktop** client window, select **Subscribe** and, when prompted, sign in with the credentials of the User2 Entra ID user account which you can locate on the **Resources** tab in the right pane of the lab interface window.

7. Ensure that the **Remote Desktop** page displays four icons, including Microsoft Word, Microsoft Excel, Microsoft PowerPoint, and Command Prompt.

8. Double-click the Command Prompt icon.

9. When prompted to sign in, in the **Windows Security** dialog box, enter the password of the Microsoft Entra user account you used to connect to the target Azure Virtual Desktop environment.

10. Verify that a **Command Prompt** window appears shortly afterwards.

> **Note**: At this point, the used host pool capacity is 100%, which is greater than the capacity threshold (%60). This should result in autoscaling turning on another host, which will bring the used host pool capacity to 50%. Since this is below the capacity threshold, the third host will remain stopped/deallocated. You will verify this next.

11. From the lab computer, switch to the web browser displaying the Azure portal.

12. On the **az140-21-hp1 | Session hosts** page, select **Refresh**, review the values of the **Power state** setting of session hosts, and verify that now two of them are listed as **Running**.

> **Note**: Next, you will evaluate the **Ramp-down** capacity threshold setting by adjusting its time window.

13. On the **az140-21-hp1 | Session hosts** page, in the vertical navigation menu, in the **Settings** section, select **Scaling plans** and then, on the **Scaling plans** page, select **az140-scalingplan412e**.

14. On the **az140-scalingplan412e** page, in the vertical navigation menu, in the **Settings** section, select **Schedules** and then select **week_schedule**.

15. In the **week_schedule** pane, navigate to the **Ramp-down** tab and adjust the value of the **Start time (12 hour system)** setting to any time between the **Start time (12 hour system)** of the **Peek hours** phase and your current time.

    > **Note**: You might need to adjust the value of **Start time (12 hour system)** of the **Peek hours** phase.

16. In the **week_schedule** pane, navigate to the **Off-peek hours** tab and select **Save**.

17. Switch to the **Command Prompt** window representing the only RDP session to the host pool and, at the Command Prompt, enter the following and press the **Enter** key:

    ```
    logoff
    ```

18. From the lab computer, in the web browser displaying the Azure portal, search for and select **Azure Virtual Desktop** and, on the **Azure Virtual Desktop** page, in the vertical menu bar, in the **Manage** section, select **Host pools**.

19. On the **Azure Virtual Desktop | Host pools** page, in the list of host pools, select **az140-21-hp1**.

20. On the **az140-21-hp1** page, in the in the vertical menu bar, in the **Manage** section, select **Session hosts**.

21. On the **az140-21-hp1 | Session hosts** page, review the values of the **Power state** setting of session hosts and verify that now only one of them is listed as **Running**.

Task 6: Disable host pool autoscaling

> **Note**: To ensure that the autoscaling configuration will not affect other labs, you will remove the host pool assignment of the scaling plan you implemented in this lab.

1. From the lab computer, in the web browser displaying the Azure portal, search for and select **Azure Virtual Desktop**, on the **Azure Virtual Desktop** page, in the **Manage** section of the vertical navigation menu, select **Scaling plans**, and then, on the **Scaling plans** page, select **az140-scalingplan412e**.
2. On the **az140-scalingplan412e** page, in the **Manage** section, select **Host pool assignments**.
3. On the **az140-scalingplan412e | Host pool assignments** page, select **az140-21-hp1**, then select **Unassign** and, when prompted to confirm, in the **Unassign host pool** dialog box, select **Unassign**.