



Avi Integration with Azure VM Scale Sets

Avi Technical Reference (v20.1)

Copyright © 2020

Avi Integration with Azure VM Scale Sets

[view online](#)

This article describes how Avi Vantage platform is integrated with Azure virtual machine (VM) scale sets. The integration primarily involves the following two sections: * Configuring pools * Configuring virtual service

The pool and virtual service settings are customized as explained in the below sections for Avi Vantage to implement the auto scaling policies. An exception to this is where Azure drives the auto scaling using inbuilt policies and is covered under the section [Auto Scaling using Azure Policies](#).

Overview

An Avi pool is a group of back-end servers having similar characteristics or hosting similar applications. In an Avi-Azure integration, an Avi pool is scaled in or scaled out, to reflect the actions taken by Azure to manage the specific scale set. These actions are governed by Azure's preconfigured policies and criteria. For more information on Azure virtual machine scale sets, refer to [What are virtual machine scale sets in Azure?](#)

Background

Avi Vantage supports Azure scale sets in configuring Avi pools for a virtual service. The Avi Azure cloud connector periodically polls Azure scale set information and updates the corresponding Avi pool server membership in case of any changes.

For example, if a new server (virtual machine) is added to an Azure scale set that is used as an Avi pool, Avi Vantage automatically updates the pool membership to include the newly provisioned server. Conversely, upon deletion of a server (virtual machine) from the Azure scale set, Avi Vantage deletes this server from its pool membership. This enables seamless, elastic, and automated backend server resource management without any operator intervention or configuration updates.

Prerequisites

- Avi Vantage release 17.2.4 and later.
- The Azure credential asset needs to have access to the scale sets. The credential asset could either be a service principle object, as in case of an application or an username/password credential set. For more details on Azure role definitions, refer to [Role Setup for Installation in Microsoft Azure](#).
- Scale sets are configured in Azure.

Configuring Pool Settings

From the Navigation pane, choose Applications, and click on Pools. Click on Create Pool and choose Azure cloud, when prompted.

Enter the pool name and suitable values for Default Server Port, Graceful Disable Timeout, Load Balance, and Persistence fields.

The screenshot displays the configuration page for a new pool named 'Azure_test_pool'. The interface is divided into four steps: Step 1: Settings, Step 2: Servers, Step 3: Advanced, and Step 4: Review. The 'Advanced' step is currently selected and active. The configuration fields are as follows:

- Name:** Azure_test_pool
- Enabled:** Checked (indicated by a green checkmark and a blue toggle switch).
- Health Monitors:** Includes a checked 'Passive Health Monitor' and an '+ Add Active Monitor' button. A message states: 'No active health monitors have been added to this pool.'
- Default Server Port:** 80
- Graceful Disable Timeout (Minutes):** 1
- Load Balance:** Least Connections
- Persistence:** None
- AutoScale Policy:** None
- Other options:** 'Rewrite Host Header to Server Name' and 'SSL to Backend Servers' are unchecked.

At the bottom of the form, there are 'Cancel' and 'Next' buttons.

AutoScale Policy

To integrate Avi Vantage with Azure VM scale sets, configure the autoscale policy by defining conditions and alerts for scale in and scale out.

To create a new policy, click on the drop-down list for AutoScale Policy tab and choose Create AutoScale Policy.

New Pool: Azure_test_pool

➤ Step 1: Settings ✔ Step 2: Servers

Name* ?

Enabled ?

Default Server Port ?

Graceful Disable Timeout (Minutes) ?

Load Balance ?

Least Connections
▼

Persistence ?

None
▼

AutoScale Policy ?

None
▼

-- no valid entries --

Create AutoScale Policy

Create AutoScale Policy

- Name ? Enter a relevant name for the policy.
- Instances ? Enter the minimum and maximum number of servers for the scaling process.
- Scale Out
 - Alerts ? Click on the dropdown list to choose a pre-defined alert. To create a new alert, refer to the Create Alerts section below.
 - Cooldown Period ? Enter the time duration in seconds during which no new scaleout will be triggered in spite an alert.
 - Adjustment Step ? Enter the number of servers to scaleout simultaneously.
- Scale In
 - Alerts ? Click on the dropdown list to choose a pre-defined alert. To create a new alert, refer to the Create Alerts section below.
 - Cooldown Period ? Enter the time duration in seconds during which no new scalein will be triggered in spite an alert.
 - Adjustment Step ? Enter the number of servers to scalein simultaneously.

New AutoScale Policy: Autoscale_policy_test
✕

Name*

Instances

2	min	8	max
---	-----	---	-----

Intelligent (Machine Learning) - Beta ?

Scale Out

Alerts ?

Syslog-Config-Events ✕
▼ ✎

Cooldown Period ?	sec	Adjustment Step ?
300		1

Scale In

Alerts ?

Syslog-Config-Events ✕
▼ ✎

Cooldown Period ?	sec	Adjustment Step ?
300		1

Save

In the example shown, cooldown period is configured for 300 seconds and the adjustment step is set to 1. So, after a scale out or scale in event is triggered, the system waits for 300 seconds before triggering a new event. With the adjustment step set to one, the servers will be incremented or decremented by one for every scale out or scale in event respectively.

Create Alerts

To create a custom alert for Scale out or Scale in, click on the dropdown list for Alerts and choose Create.

New AutoScale Policy: Autoscale_policy_test

Name*

Instances min max

Intelligent (Machine Learning) - Beta ?

Scale Out

Alerts ?

None

Search

- Syslog-Config-Events
- Syslog-System-Events
- System-CC-Alert
- System-Controller-Alert
- System-SE-Alert

Create

Cooldown Period ? sec

Adjustment Step ?

Save

Basics

- Name ? Enter a name for the alert configuration.
- Click on the Enable Event Trigger checkbox to enable alerts for the configured alert.
- Description ? (Optional) Enter the description for the configured alert.

Basics

Name * ?

Enable Alert Trigger ?

Description ?

Conditions

- **Throttle Alert ?** Enter the time in seconds for which this alert will be suppressed since the last trigger.
- **Source ?** Choose the Metrics option to configure suitable metric parameters in the subsequent sections.
- **Object ?** Choose one of the options provided: Virtual Service, Service Engine, Pool. In this instance, none are chosen.
- Click on AutoScale Alert to enable this alert for autoscaling.
- **Number of Occurrences ?** Enter the number of events that should be met or exceeded within the specified time frame for the alert to be created. The allowed range is from 1 to 65536.
- **Category ?** Choose Rolling Window to define metric parameters with a time window. Only if the number of events meets or exceeds the defined limit within this time window, will an alert be generated. Choose Real-time to define metric parameters without a time window. This checks for metric triggers in real time.
- **Metric Occurs ?** Click on the dropdown list to choose from the list of event IDs. Enter relevant values in the Comparator, Value, and Duration fields to complete the metric configuration. The instance in the screenshot configures a metric for I4_server.max_open_conns event which refers to the maximum open Layer 4 connections allowed. In this case, an alert is triggered if the Layer 4 open connections exceed or is equal to 500 for a time duration of 300 seconds.
- Click on +Add New Metric to create more event triggers. You can configure the new metric using And and Or conditions with the metric that has already been configured.

Conditions

Throttle Alert ?

sec

Source

Event Metrics

Object

Virtual Service Service Engine Pool

AutoScale Alert ?

Number of Occurrences ?

Category ?

Real-time Rolling Window

Metric Occurs* ?

v

Comparator	Value	metric count	Duration
>= v	500	metric count	<input style="width: 80%; border: none;" type="text" value="300"/> sec

+ Add New Metric

Actions

- **Alert Action ?** Click on the dropdown list to choose a pre-defined alert action or click on Create Alert Action to create a new action. Click on the edit icon to customize the action parameters.
- **Alert Expiry Time ?** Enter the expiry time in seconds for the configured alert.

Actions

Alert Action * ⓘ
Syslog-System

Alert Expiry Time ⓘ
86400 sec

Save

AutoScale Launch Config

Click on the dropdown list and choose default-autoscalelaunchconfig which indicates that the servers for auto scaling are hosted by an external ecosystem, such as, Azure cloud in this case.

New Pool: Azure_test_pool

Step 1: Settings Step 2: Servers

Name * ⓘ Azure_test_pool Enabled ⓘ

Default Server Port ⓘ 80

Graceful Disable Timeout (Minutes) ⓘ 1

Load Balance ⓘ Least Connections

Persistence ⓘ System-Persistence-Http-Cookie

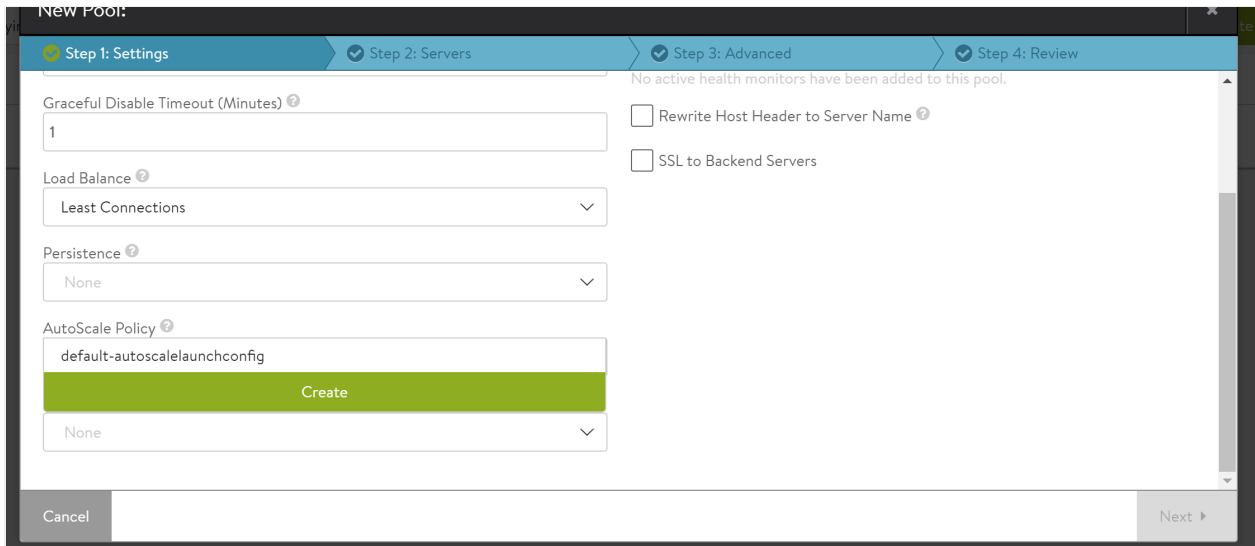
AutoScale Policy ⓘ Autoscale_policy_test

AutoScale Launch Config ⓘ
None
default-autoscalelaunchconfig

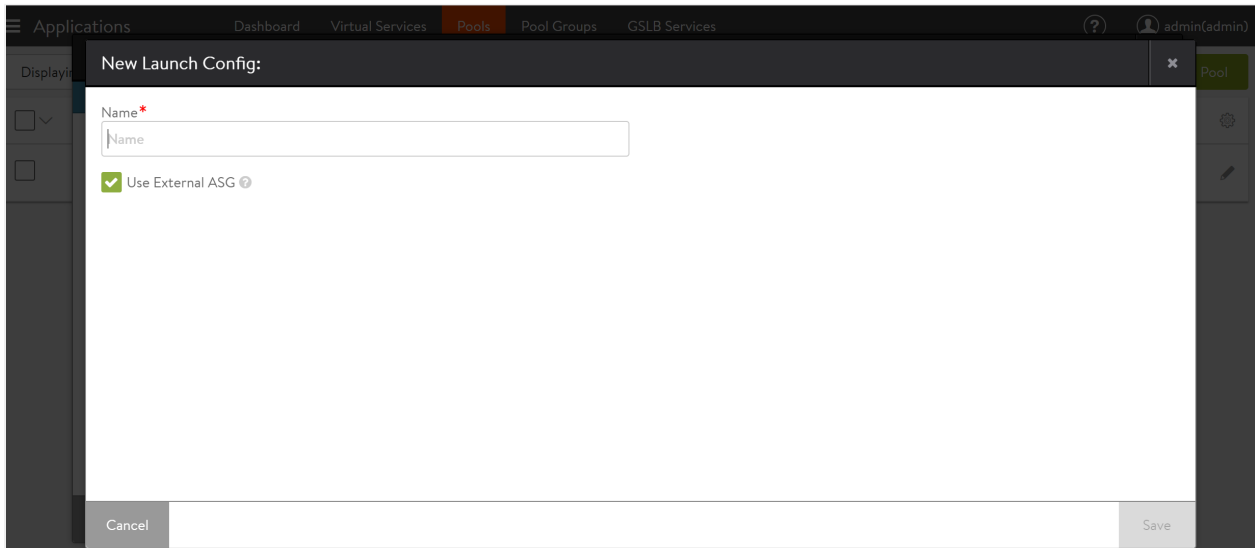
Create

Click on Next.

To create a new autoscale launch configuration, click on Create.

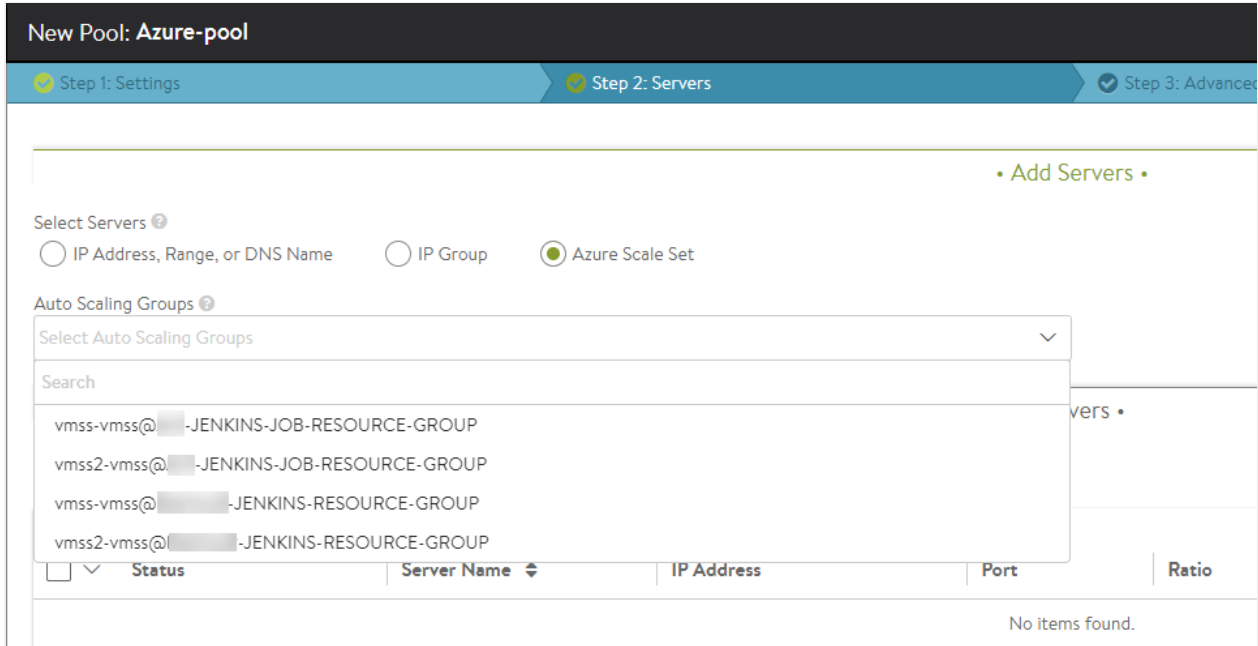


Provide the name and enable the checkbox for Use External ASG.

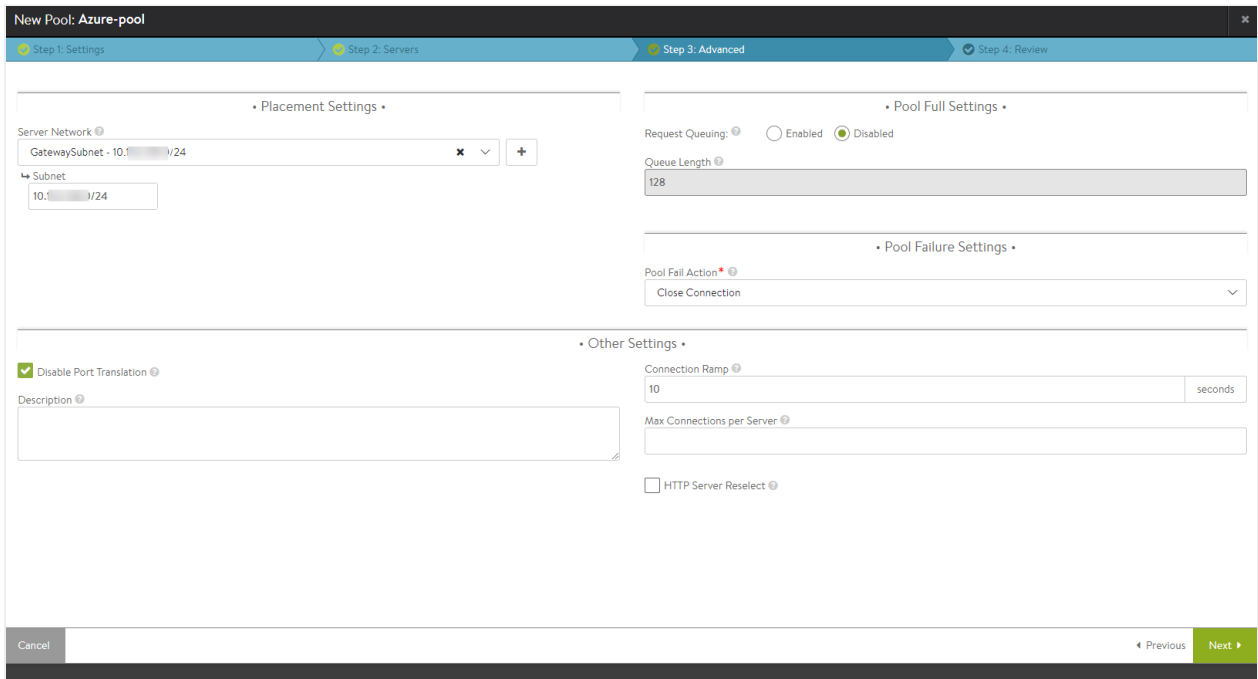


Configuring Pool Servers

Under Select Servers, choose Azure Scale Set option. Click on the dropdown for Auto Scaling Groups which will list the scale sets configured on Azure for the specified cloud.



Click on Next, to configure Placement Settings, Pool Full Settings, and Pool Failure Settings.



Click on Next to review the configuration and click on Save to complete the configuration.

This concludes the pool configuration for integrating Azure VM scale sets.

Configuring Virtual Service

Navigate to Applications > Virtual Service and click on Create Virtual Service. You can choose either Basic or Advanced setup to map the configured virtual service with Azure scale sets.

Basic Setup

Configuring under Basic Setup provides an option to select the Azure specific resource group. Under Add Servers, select Azure Scale Set and choose the resource group from the drop-down list under Auto-Scaling groups. On selecting a specific group, the configured servers will be displayed under the Servers section.

The screenshot shows the configuration interface for a new virtual service. The title is "New Virtual Service: Azure_virtual_service_test".

- Name:** Azure_virtual_service_test
- Application Type:** HTTP, HTTPS, L4, L4 SSL/TLS
- Service:** 80
- Assign Public IP for External Client Access
- Network for VIP Address Allocation:** GatewaySubnet - 10.10.10.0/24
- Add Servers:**
 - IP Address, Range, or DNS Name
 - IP Group
 - Azure Scale Set
- Auto Scaling groups:** vmss-vmss@JENKINS-JOB-RESOURCE-GROUP
- Servers:**

Server Name	IP Address	Network
vmss-vmss_10	10.10.10.10	
vmss-vmss_12	10.10.10.10	
vmss-vmss_13	10.10.10.10	
vmss-vmss_15	10.10.10.10	

Advanced Setup

Configuring under Advanced Setup provides an option to allocate the Azure pool configured for a specific virtual service. To choose the Azure pool, select Pool under the Pool section and from the drop-down list select the Azure pool configured.

• Pool •

Pool Pool Group

Pool ⓘ

Azure_test_pool x v ✎

Ignore network reachability constraints for the server pool ⓘ

Next ▶

Click on Next to navigate to the rest of the virtual service configuration and click on Save to complete the configuration.

This concludes the virtual service configuration for integrating Azure VM scale sets.

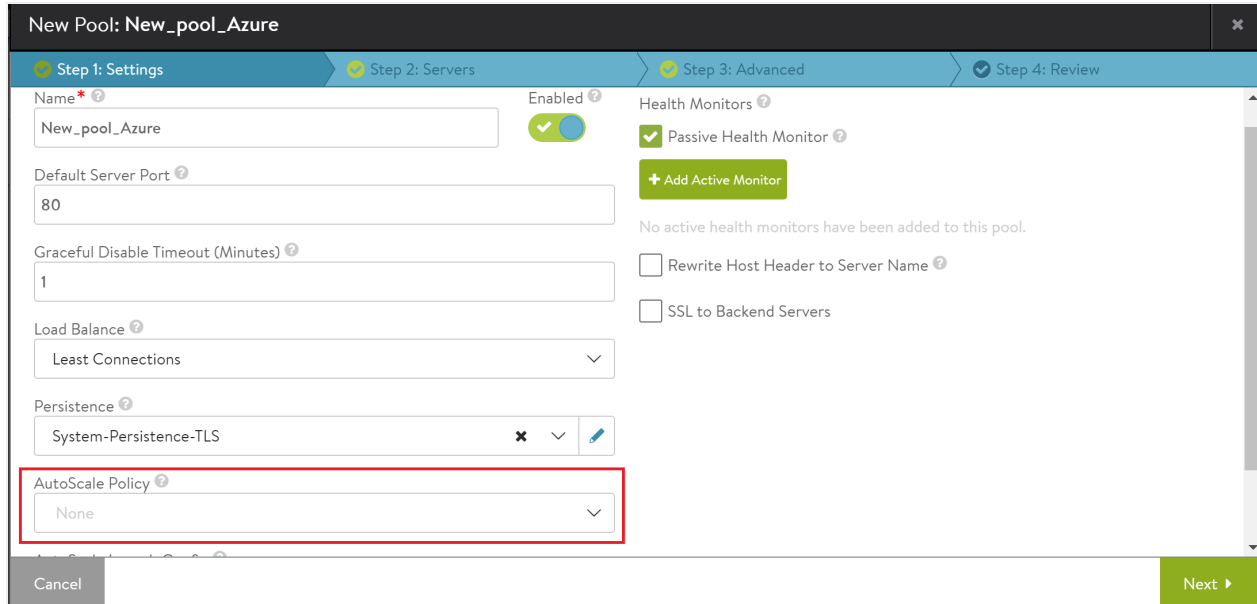
Auto Scaling using Azure Policies

The auto scaling can be driven using Azure policies instead of AutoScale Policy defined in Avi Vantage.

From the Navigation pane, choose Applications, and click on Pools. Click on Create Pool and choose Azure cloud, when prompted.

Enter the pool name and suitable values for Default Server Port, Graceful Disable Timeout, Load Balance, and Persistence fields.

Leave the AutoScale Policy field unselected and click on Next to configure pool servers as explained in [Configuring Pool Servers](#) section.



As no AutoScale policy has been defined in Avi Vantage, auto scaling is driven by the policies configured on Azure.

Configuring Autoscale Polling Interval

Starting with Avi Vantage release 18.2.2, configuration of autoscale polling interval option is supported for Azure cloud.

To enable autoscale polling interval, login to Avi CLI and execute the `autoscale_polling_interval <duration in seconds>` command under the `configure cloud Default-Cloud` mode.

```
[admin:10-1-1-1]: > configure cloud Default-Cloud
[admin:10-1-1-1]: cloud > autoscale_polling_interval 120
[admin:10-1-1-1]: cloud > save
```

120 is the duration in seconds of polling interval.