Server Autoscale

Overview

Server autoscale dynamically scales a server pool, by adding new pool members (scaling out) or removing them (scaling in). This article explains the steps to configure server autoscale via the Avi UI.

Creating an AutoScale Policy

An existing pool can be edited to create an autoscale policy. Alternatively, the autoscale policy can also be created during pool creation also.

To create an autoscale policy by editing an existing pool, 1. Click on the edit icon to open the Edit Pool screen.

2. Click on the AutoScale Policy drop-down menu. 3. Click on the Create AutoScale Policy button to open the New AutoScale Policy screen.

4. Enter the name of the autoscale policy. 5. Set the minimum and maximum number of servers (Instances) allowed in the pool. 6. Select Intelligent (Machine Learning) to select Avi Vantage’s intelligent machine learning algorithm that bases autoscaling decisions on a comparison of the load on the pool and the estimated capacity of member servers. Note: On enabling the option Intelligent (Machine Learning), the option Use Predicted Load is visible. Select this option to make use of predicted load instead of current load. 7. Set the Scale Out and Scale In parameters as follows:
Select the type of <b>Alerts</b> that will trigger a scale out or a scale in.

Select the minimum amount of time between two consecutive scale outs or scale ins as the <b>Cooldown Period</b>.

Set the number of server instances to add or remove with each scale out or scale in operation as the <b>Adjustment Step</b>.

<i>Note:</i> The adjustment step should not exceed the minimum or maximum number of allowed server instances.

Select the minimum extra capacity as percentage of load used by the intelligent machine learning algorithm.

The <b>New AutoScale Policy</b> screen is as shown below:

8. Click on Save.

The autoscale policy with the defined configurations is created and attached to the pool.