



# Server Autoscaling in Google Cloud Platform

Avi Technical Reference (v18.1)

Copyright © 2018

# Server Autoscaling in Google Cloud Platform

[view online](#)

## Overview

Autoscaling in Google Cloud platform is a feature of the managed instance groups. A managed instance group is a pool of homogenous instances, created from a common instance template.

For more information on managed instance groups, refer to [Instance Groups](#).

For Avi Vantage, the simplest form of autoscaling is to scale based on the CPU utilization of a group of virtual machine instances.

Autoscale groups can be created for both single zone (with multi zone support) and multiple zone (regional). Multi zone or regional managed instance groups improve your application availability by distributing the instances across multiple zones within a single region.

For regions that contain more than three zones, the regional managed instance group will choose three zones to create instances in. You can also selectively choose the zones to create instances or create instances in regions with less than three zones.

Note: For regions, only the region specified in the IPAM profile is supported.

## Autoscaling Groups in GCP using Instance Templates

An instance template is an API resource that you can use to create VM instances and managed instance groups. Instance templates define the machine type, boot disk image or container image, zone, labels, and other instance properties. You can use an instance template to create a managed instance group or to create individual VM instances.

Managed instance groups allow autoscaling by automatically adding or removing instances based on the increase or decrease in the load.

### Creating Instance Group with Instance Template

To create an instance group coupled with an instance template, you would require the following data: \* Name for the instance group \* Region \* Location \* Choice of the region as per the user \* Instance template \* Autoscale option \* Choosing CPU as the option and the target CPU usage \* Maximum and minimum instances to spin up using the instance templates \* Cool down period (The period after which the instance boots up and the information is collected) \* Health check for autohealing \* Delay

For complete information refer to Google documentation at [Creating Groups of Managed Instances](#).

### Creating Instance Templates

To create an instance template, you would require the following data: \* Name for the instance template \* Machine type \* Start-up script \* Networking details to include the network in which the instances will spin up \* SSH keys needed for logging in (*Optional*) \* IP forwarding (Enabled- ON)

For complete information refer to Google documentation at [Creating Instance Templates](#).

## Triggers in Instance Templates

Autoscaling is based on the following trigger types in instance templates: \* CPU usage \* Stackdriver monitoring metric \* Multiple metrics

The autoscaler is created by specifying the autoscaling policy and a target utilization level. These are used by the autoscaler to scale the group. The autoscaler will collect information based on the policy configured on Avi Vantage. This information is compared with the desired target utilization and scaling is triggered accordingly.

Use the Google API for CPU load balancing to call for scaling in and scaling out instances. Refer to [Scaling Based on CPU or Load Balancing Serving Capacity](#) for complete information on the related API.

## Rolling Update, Restart, and Replace (Optional)

Canary upgrade, rolling restart and replace are supported. This can be used for restarting or replacing existing instances.

In case of a canary upgrade, you can use another template with different conditions or set parameters for autoscaling. Ensure that the template has the same network selected as that of the instance group.

For complete information refer to Google documentation at [Updating Managed Instance Groups](#).

## Best Practices

Specify a custom image and relevant startup scripts for the instance templates. The necessary applications are installed whenever an instance is created with access to all the required data.

## Limitations

- You can only create autoscaling groups inside your SE project ID.
- Each managed instance group can support only up to 1000 VM instances.
- More than 1000 instances can not be specified in a single request while updating a managed instance group.

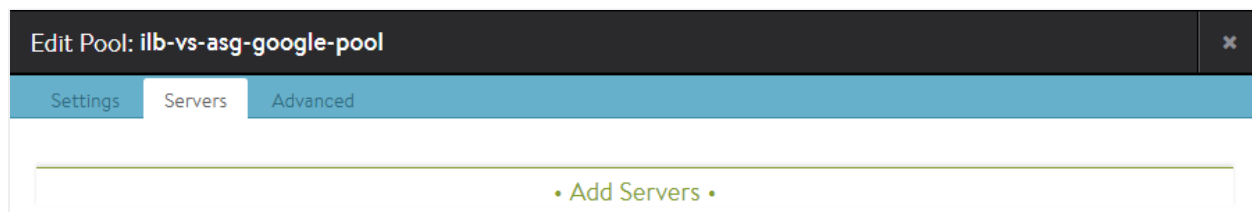
## Enabling External Autoscaling on Avi Vantage

Note: This configuration is supported only on CLI as of now.

1. On Avi shell, configure external autoscaling groups using the `external_autoscale_groups` command under the designated pool.

```
[admin:10-1-1-1]: > configure pool ilb-vs-asg-google-pool
[admin:10-1-1-1]: pool > external_autoscale_groups user-asg-ig
[admin:10-1-1-1]: pool > save
```

On UI, navigate to Applications > Pools and select the relevant pool. Navigate to the Servers tab to view the configured external autoscale group under Auto Scaling Groups field.



Select Servers ⓘ

IP Address, Range, or DNS Name | IP Group

Auto Scaling Groups ⓘ

-asg-ig ✕

• Servers •

Q

Displaying 1 item(s)

<input type="checkbox"/> ✓	Status	IP Address
<input type="checkbox"/>	Enabled	10.146.3.20

Cancel | Save

2. Create a virtual service for the cloud using the IPAM associated with it. On initiating traffic, you will notice instances attached to the pool, as they start processing traffic based on conditions defined in the instance group.