An instance group is a collection of instances that are managed as a single entity.

Now we have two types of instance groups in Google Cloud platform.

There are **managed instance groups and unmanaged instance groups.**

A **managed instance group,** often abbreviated as MIG or mig, is a set of identical VMs or Virtual Machines and their configuration is defined in something called an instance template. Now, there are many useful features about managed instance groups

including they enable auto scaling. So we can add and remove virtual machines to an instance group based on workload. Instance groups support auto healing.

So if, for example, we do a health check on an instance and we don't get an indication

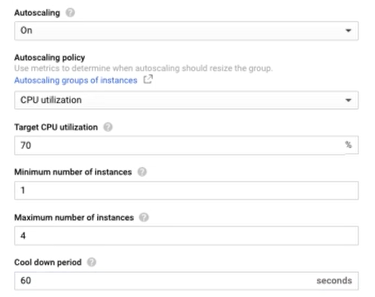
back that it's healthy the managed instance group will restart that Virtual Machine for us. We can enable or use multi-zone deployments which helps improve high availability, and we can also enable auto updating so that our operating systems and software are patched.

The second type of instance group is called an **unmanaged instance group.** Now, we don't really use these too often but basically they're used whenever you have a set of servers or a set of heterogeneous Virtual Machines that are used together and are used behind a load balancer. So if for example, you have different for some reason, you have different configurations within your load balancing group you can use an unmanaged instance group. Now, in general, it's recommended just for legacy clusters only. So you might have had something in your on premise data center that required a heterogeneous cluster but it was load, load balanced. If you need to migrate that to Google Cloud that's when you would use an unmanaged instance group.

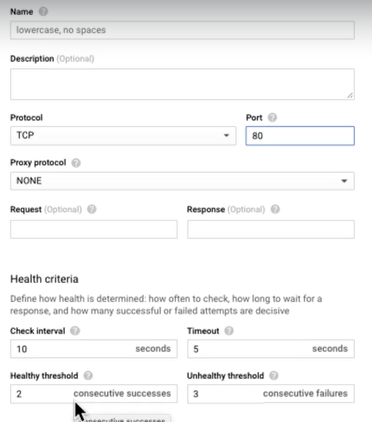
Now, unmanaged instance groups do not support a lot of the features that are really valuable in an instance group like auto scaling, auto healing, and auto updating.

Creating instance group :

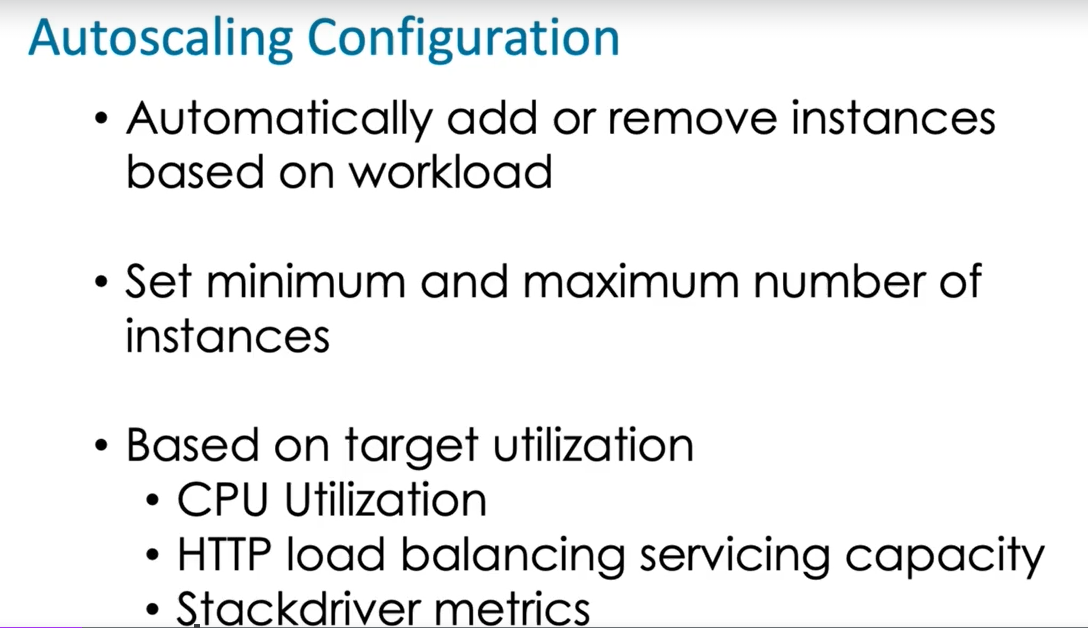
Compute engine > instance groups > name , description , location (single zone / multiple zone), region , instance template , autoscaling , autoscaling policy (metric eg:cpu utilization )

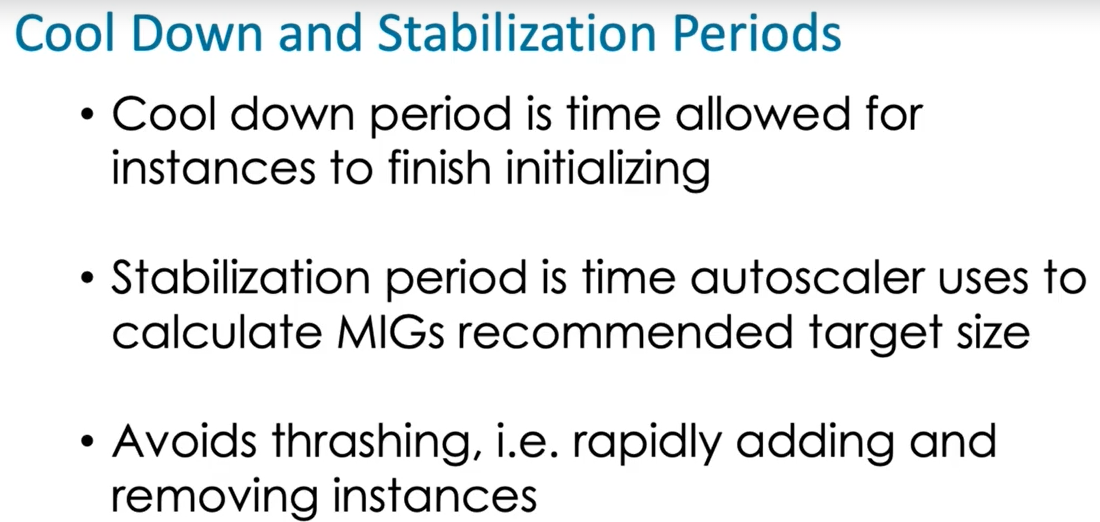


We can create health checks .



Autoscaling with managed instance group :





**Exercise: Create an Instance Group**

For this exercise:

1. Create an instance group named ace-cloud-engineer-instance-group
2. Use n1-standard-1 machine types
3. Use multiple zones
4. Set the operating system to one of your choices
5. Do not set a health check
6. Set the minimum number of instances to 1 and the maximum to 5

An unmanaged instance group is used when the instances in the group are not identically configured. Managed instance groups are used for identically configured instances. Heterogeneous and hybrid are not types of instance groups.

Only a managed instance group can load balance workloads across instances; unmanaged instance groups do not support load balancing.

During the cool down period, the autoscaler does not take a VMs metrics into account when making scale down decisions so increasing this period will give instances more time to complete initialization and start taking on load. Increasing the minimum or decreasing the maximum number of instances will not stop thrashing. Changing the metric used might avoid thrashing but a metric that avoids thrashing may not be measuring a metric useful for scaling decisions.

Heterogeneous clusters can be run on unmanaged instance groups but not managed instance groups.

An instance template is needed to enable Compute Engine to automatically add instances to a managed instance group.

Target pools must have a health check to function properly. Nodes can be in different zones but must be in the same region. Cloud Monitoring and Cloud Logging are useful but they are not required for the target pool to function properly. Nodes in a pool have the same configuration.

—-----------------------------------------------------------------------------------------------------------------------------

Create instance group

Edit managed instance group

Updating a Managed Instance Groups (MIG) - Rolling Update & Restart V2

For rolling update :

Instance group > Update vm > Change instance template > Configure new template .

Canary testing :

Add 1 old teampate in the instance group and add a new template also . set target size to do testing .

Select update type : automatic / selective

In automatic we have options : restart , refresh , restart and refresh etc for our testing instances .

The other thing you can configure here is temporary additional instances.

This is also called maxSurge.

How many new instances should be first created before the old instances are deleted .

you can configure maxUnavailable. How many available, how many instances can be unavailable during the update .

This is how we configure rolling updates .

Restart and replace vms :

On restart we can set max unavailable vm at tym of restart .

* gcloud compute instances create my-test-vm --source-instance-template=my-instance-template-with-custom-image
* gcloud compute instance-groups managed list
* gcloud compute instance-groups managed delete my-managed-instance-group
* gcloud compute instance-groups managed create my-mig --zone us-central1-a --template my-instance-template-with-custom-image --size 1
* gcloud compute instance-groups managed set-autoscaling my-mig --max-num-replicas=2 --zone us-central1-a
* gcloud compute instance-groups managed stop-autoscaling my-mig --zone us-central1-a
* gcloud compute instance-groups managed resize my-mig --size=1 --zone=us-central1-a
* gcloud compute instance-groups managed recreate-instances my-mig --instances=my-mig-85fb --zone us-central1-a
* gcloud compute instance-groups managed delete my-managed-instance-group --region=us-central1