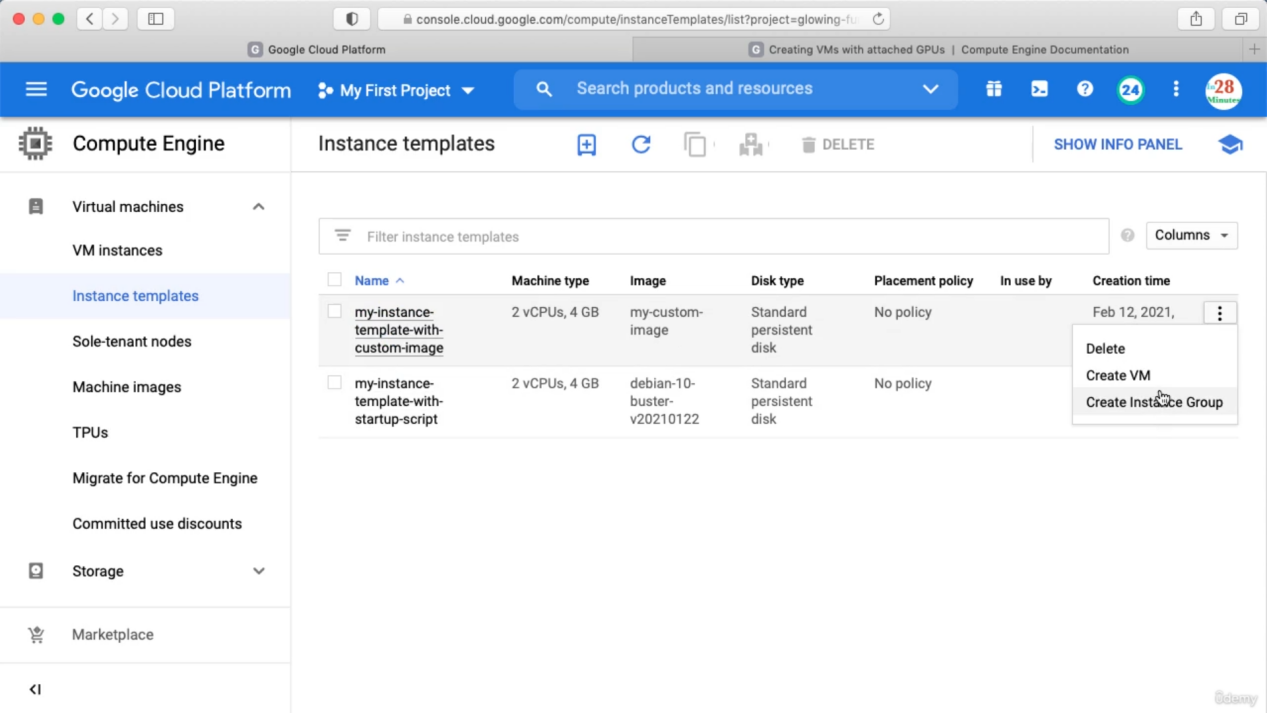
MIG => Managed instance group.

Auto Healing => you can configure a health check and if the health of a particular VM is not good then it is replaced automatically.

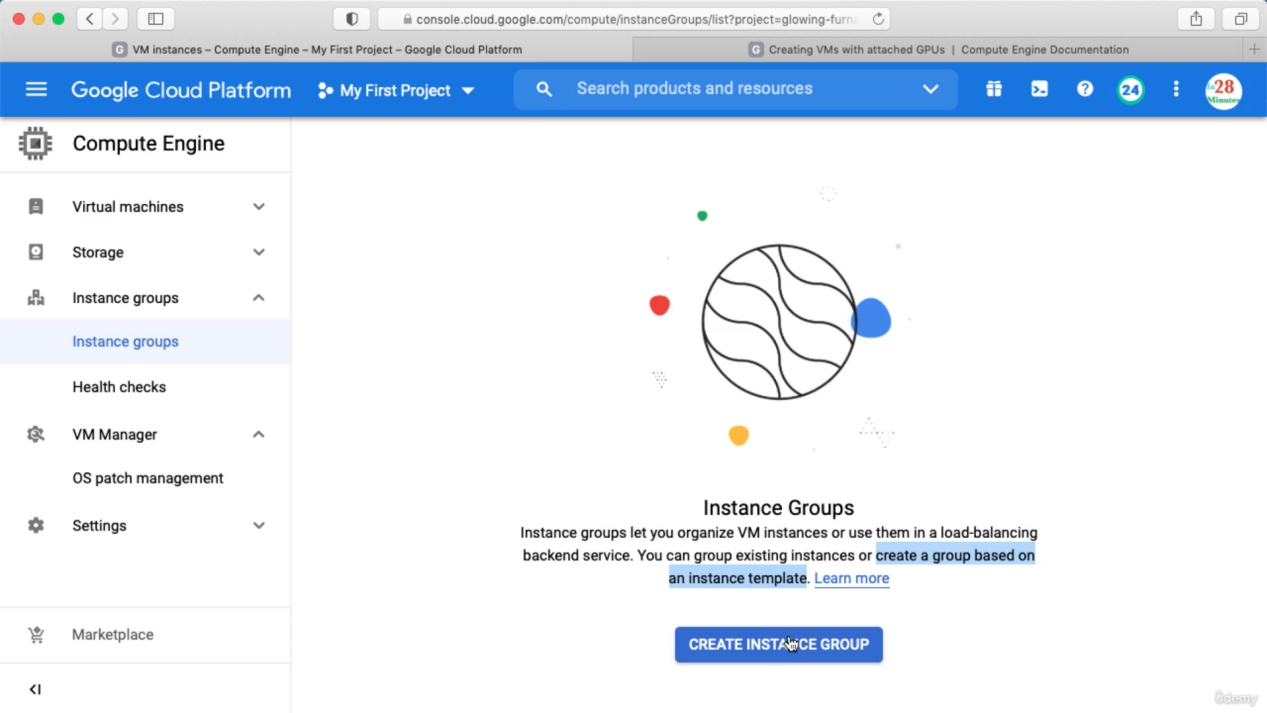
Managed releases => you can go to one version to another version without any downtime.

Stack driver is a monitoring tool provided by Google cloud.

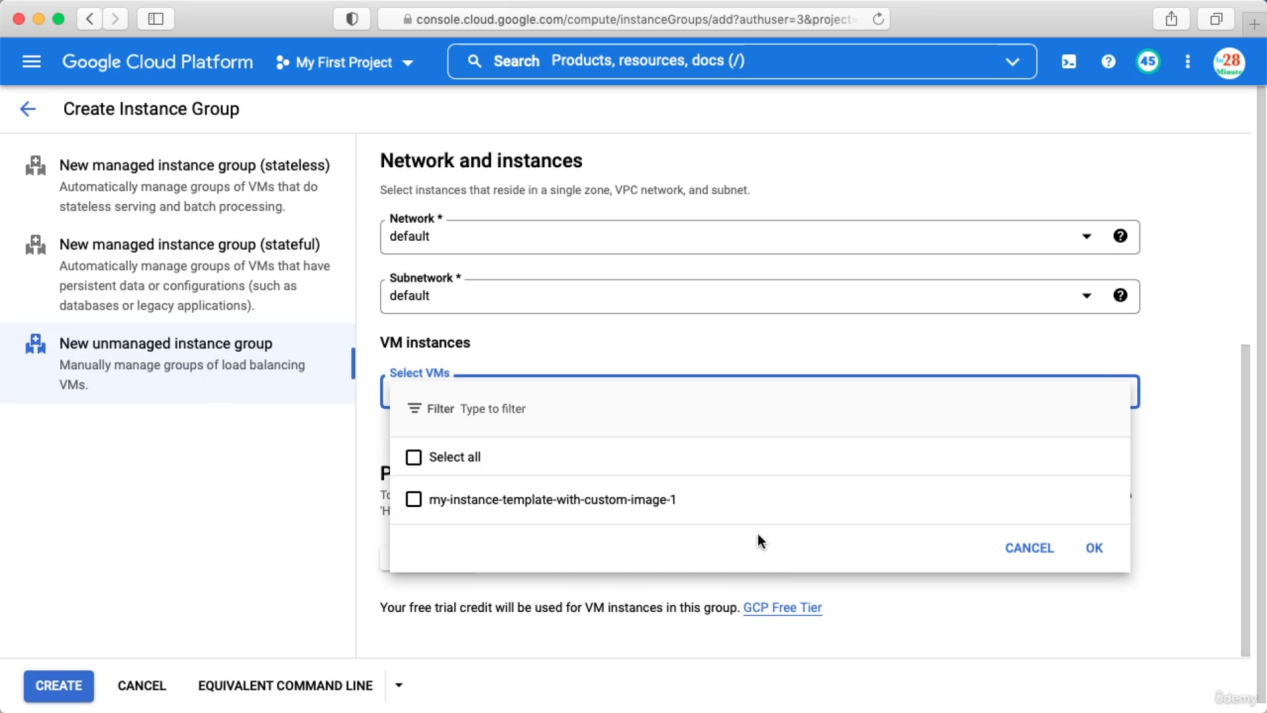
Creating MIG



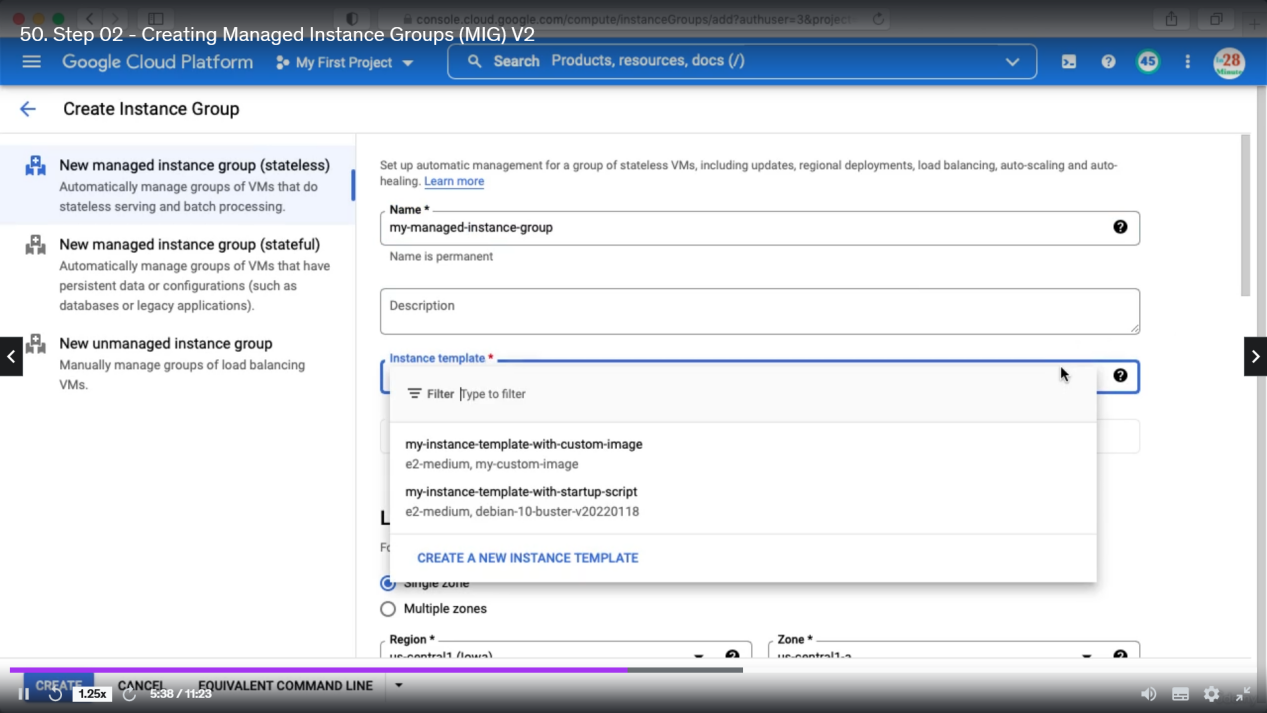
Go to instance group and click create



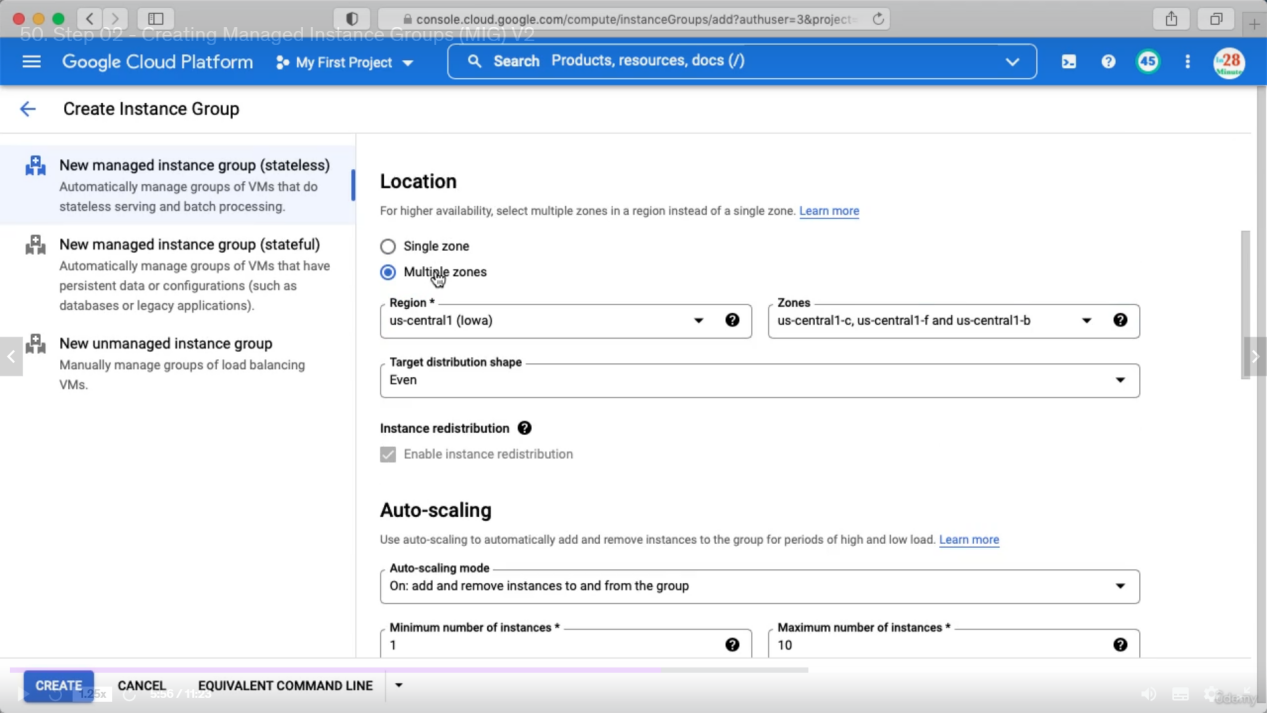
For UN-managed instance group it is allowing to select multiple VM’s as below.



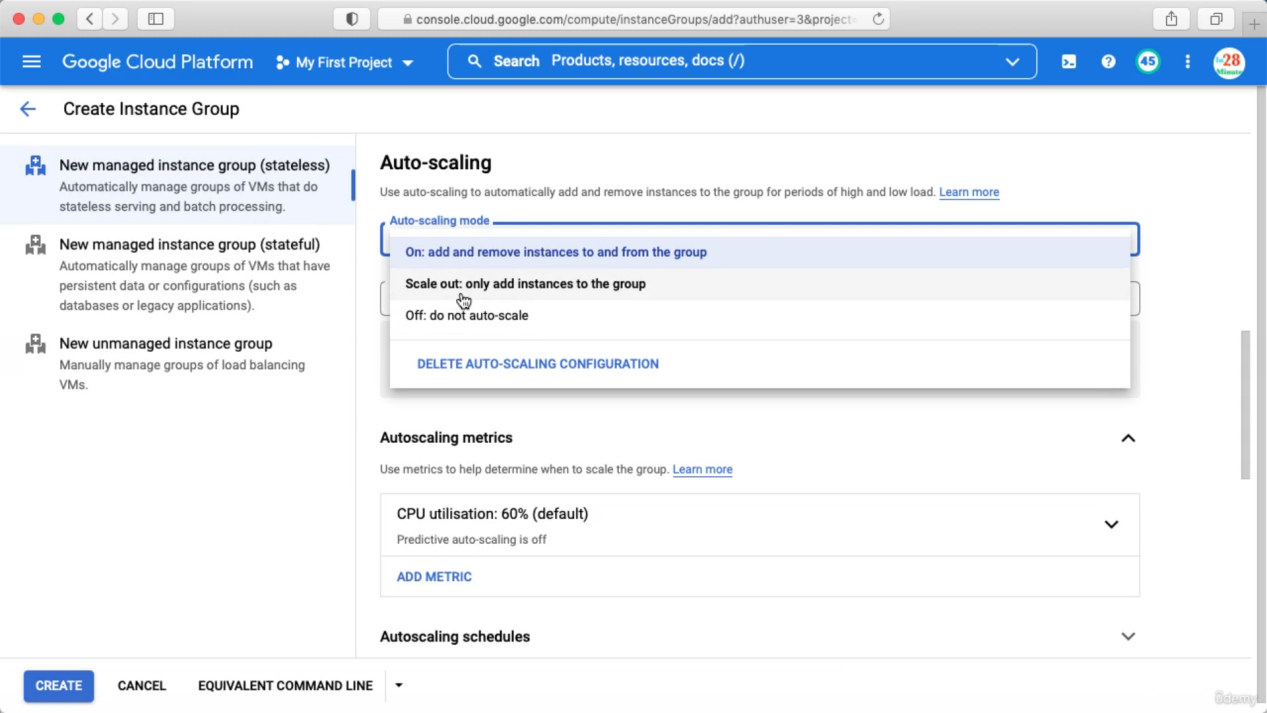
For managed stateless instance group create this as below. Select a instance template from which instance group will be created.



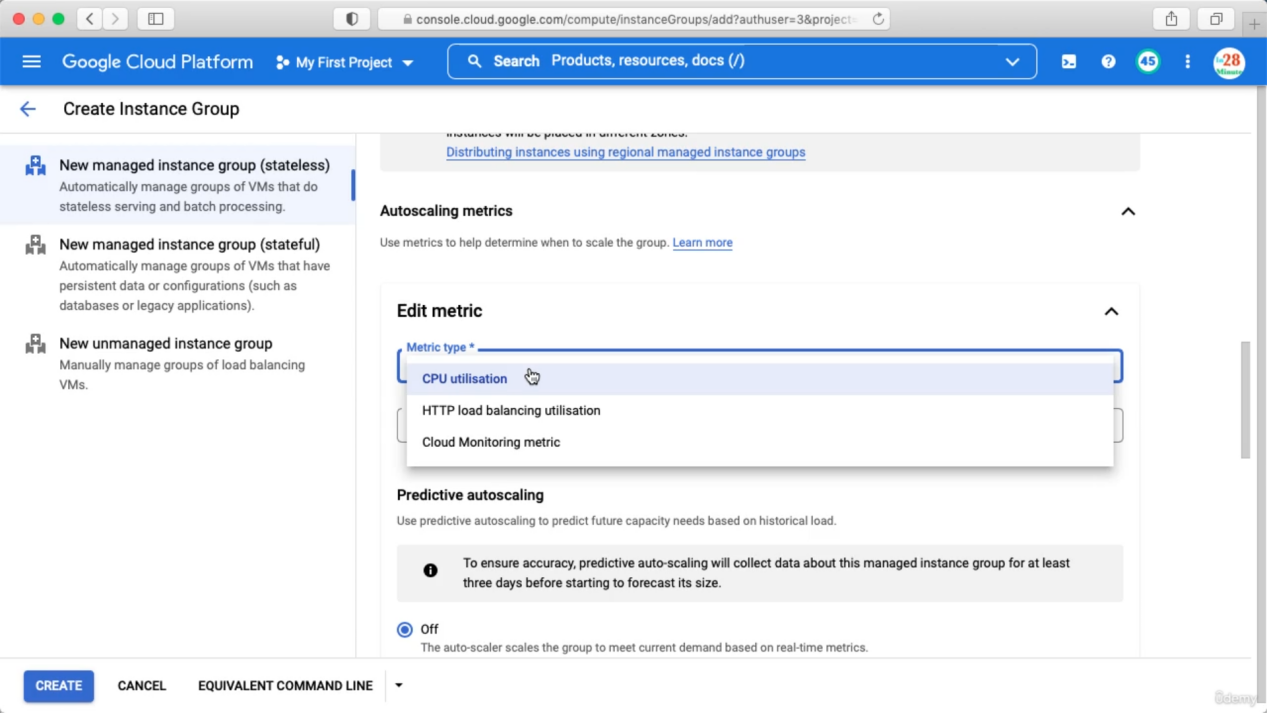
You can check multiple zones that will give you higher availability



Next you need to select Auto Scaling. Scale out means it allows only add instance and no remove. Best is to select on.

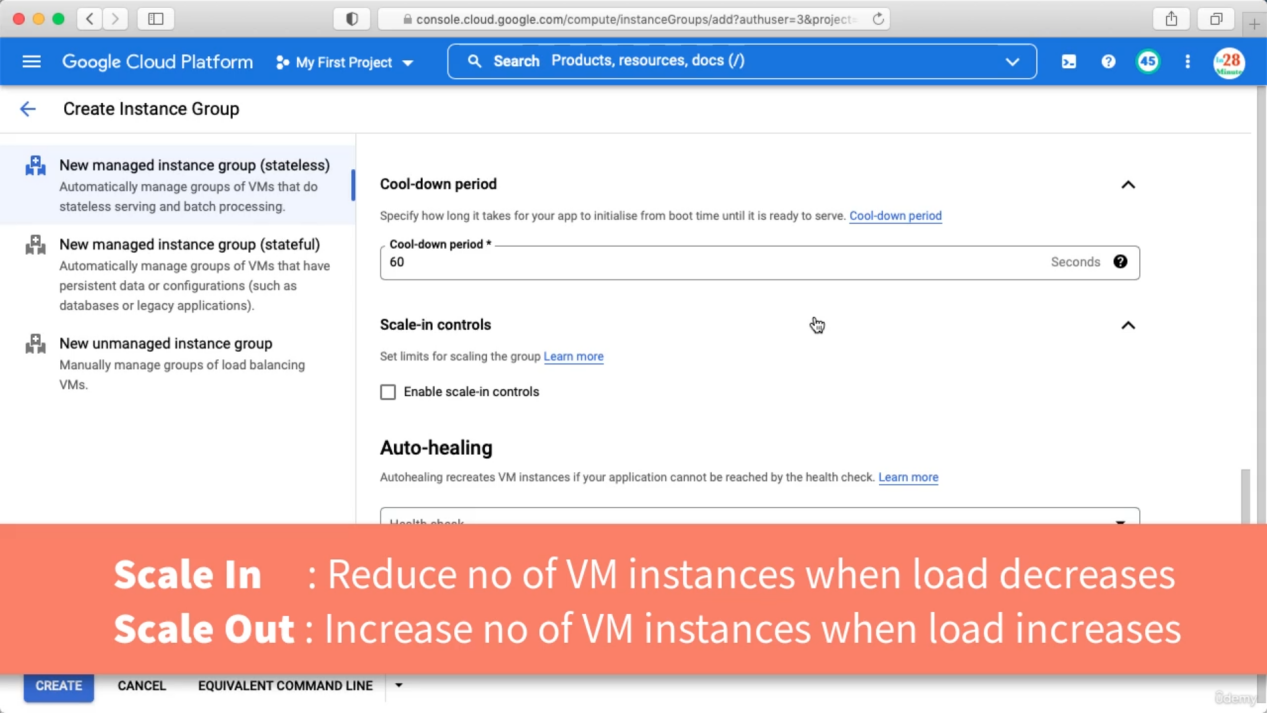


You also can choose different metrices other than cpu utilization.

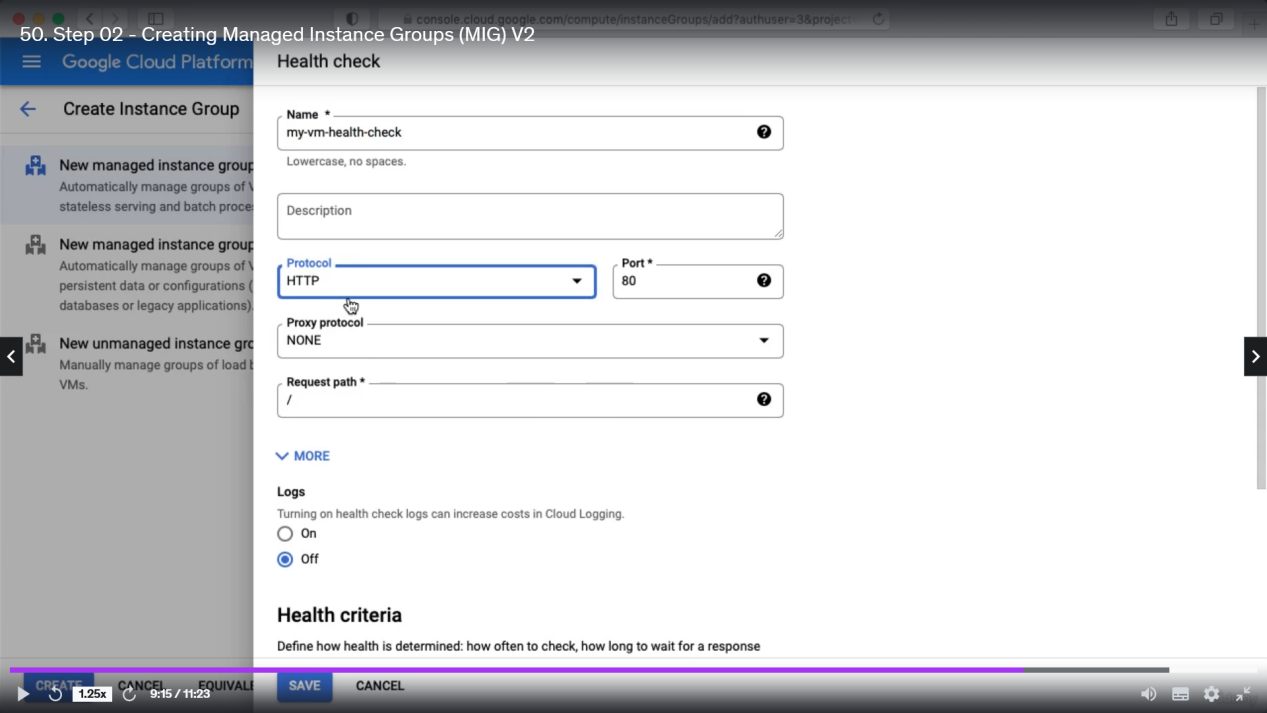


Predictive auto scaling : based on historical data Google cloud will decide whether to scale up or down.

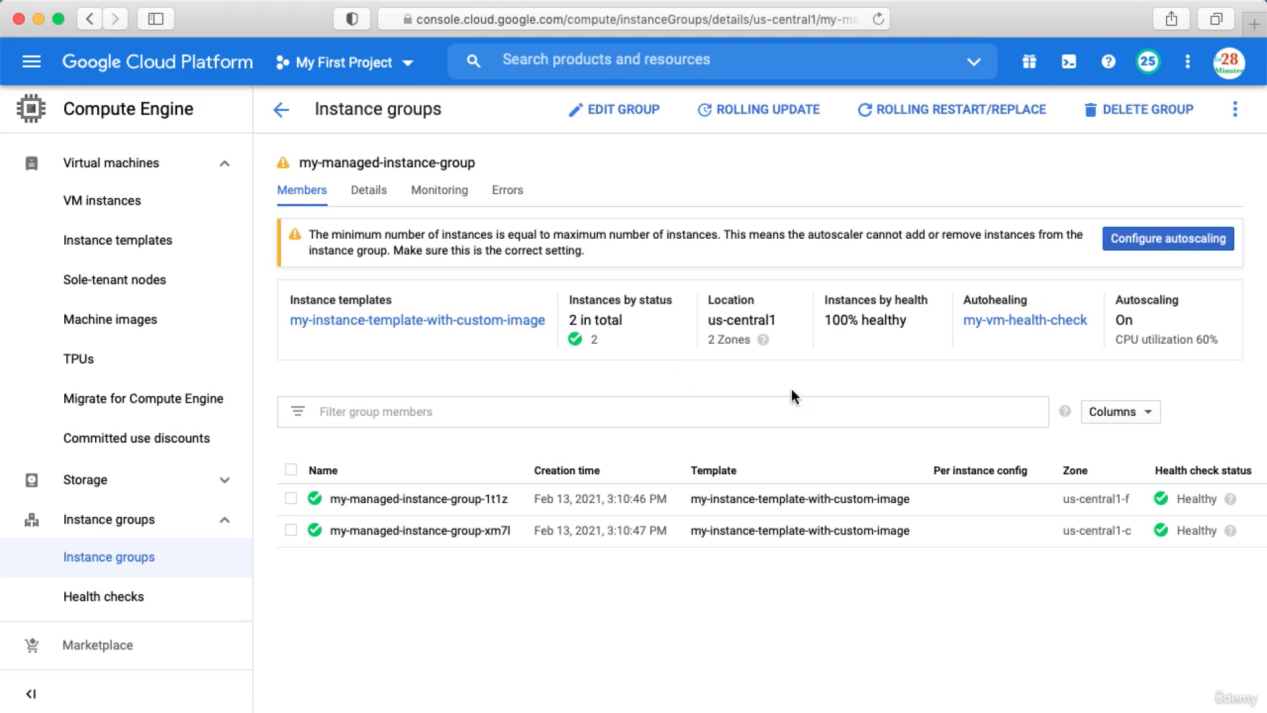
Cool-down-period : after auto scaling how long it should wait before looking to metrics.



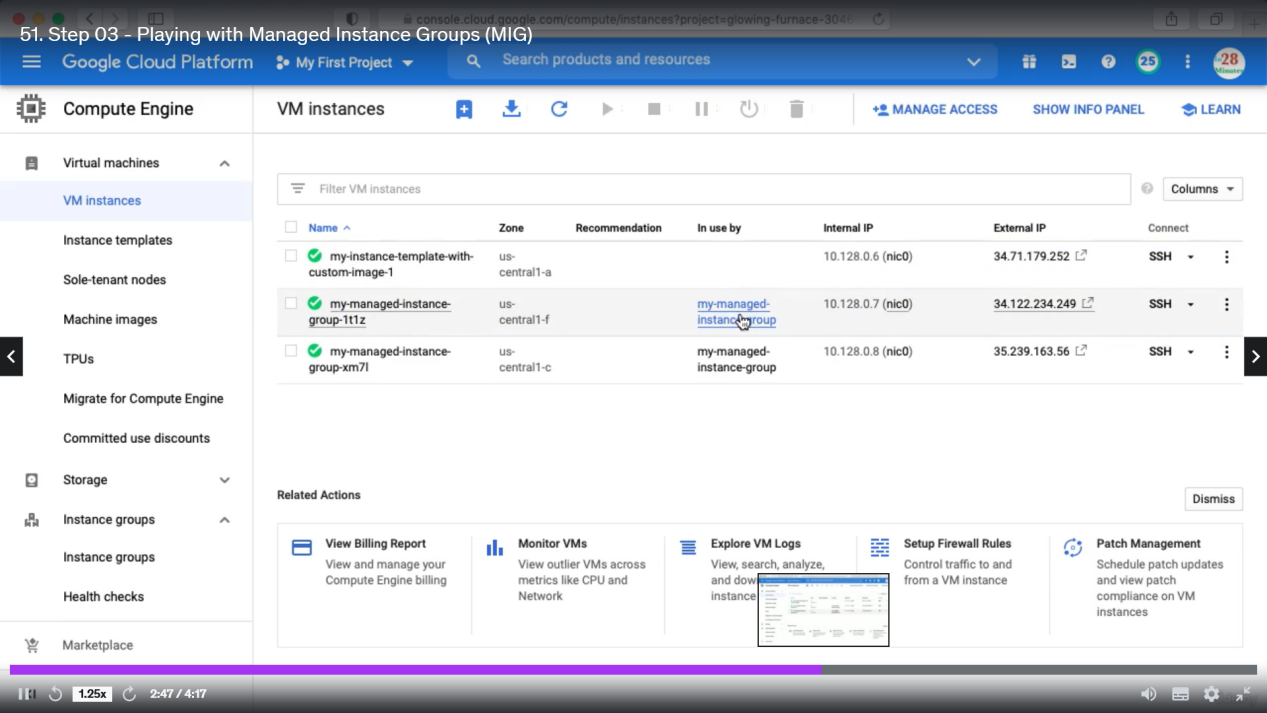
For auto healing we need to create out health check and you should keep log on. Cloud logging is a managed log service for all the services in google cloud.



If you click the created instance group you will see below details



If you click vm, then you see the 2 instances are used by instance group



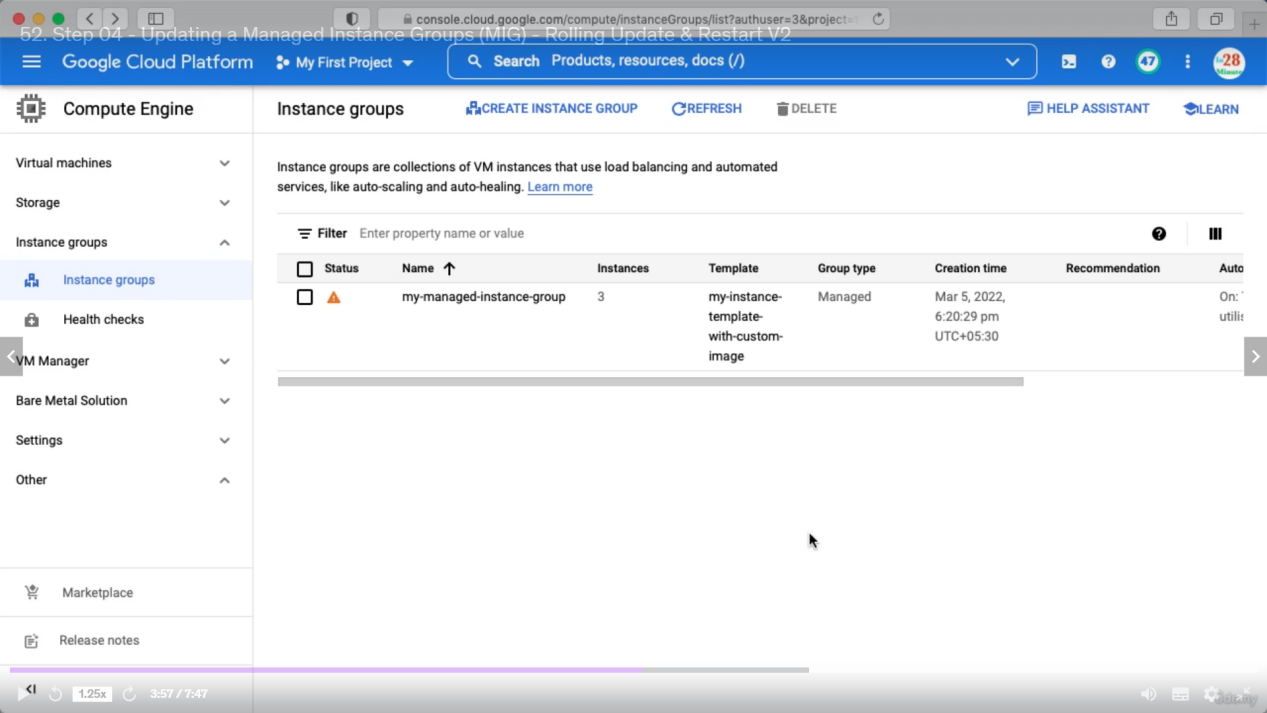
You also can edit the instance group and change the parameters like increase or decrease number of instances.

Updating Instance group

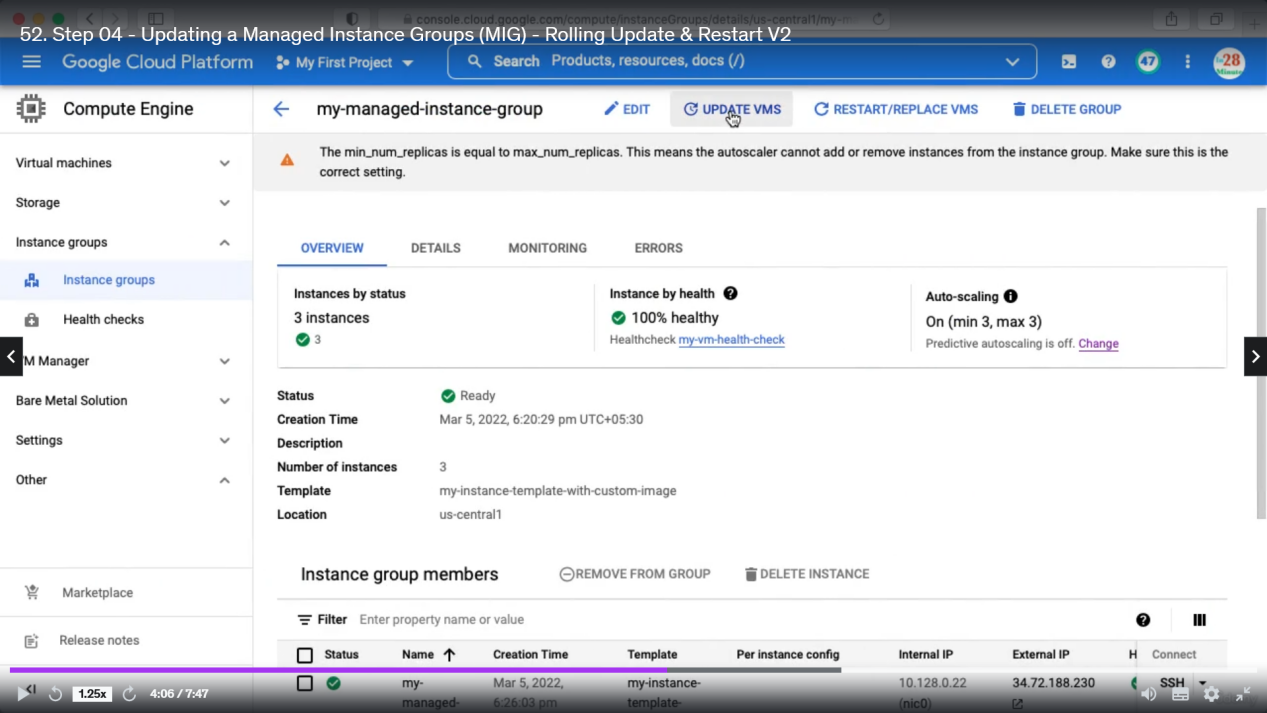
Let's say, I have 10 instances, I can say I would want to update 2 instances at a time,so I would want 2 instances first, 2 instances next, 2 instances after that.So we are updating in a group of 2 instances each.The other option is to test a set of instances first.So you would configure a new template that basically called the canary template and you would test 2 instances.Once the 2 instances are all tested, you would roll out at once to the remaining 8 instances. This approach is called a canary testing approach.

Update instance group demo

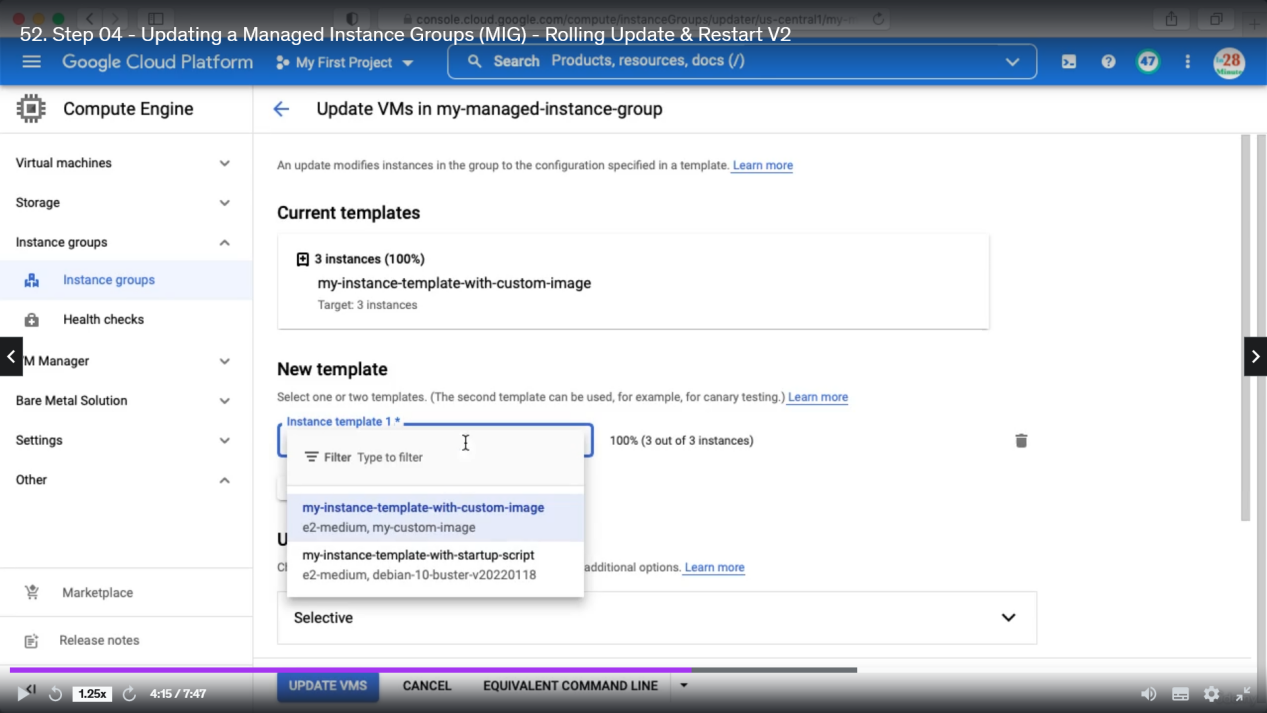
Click your instance group



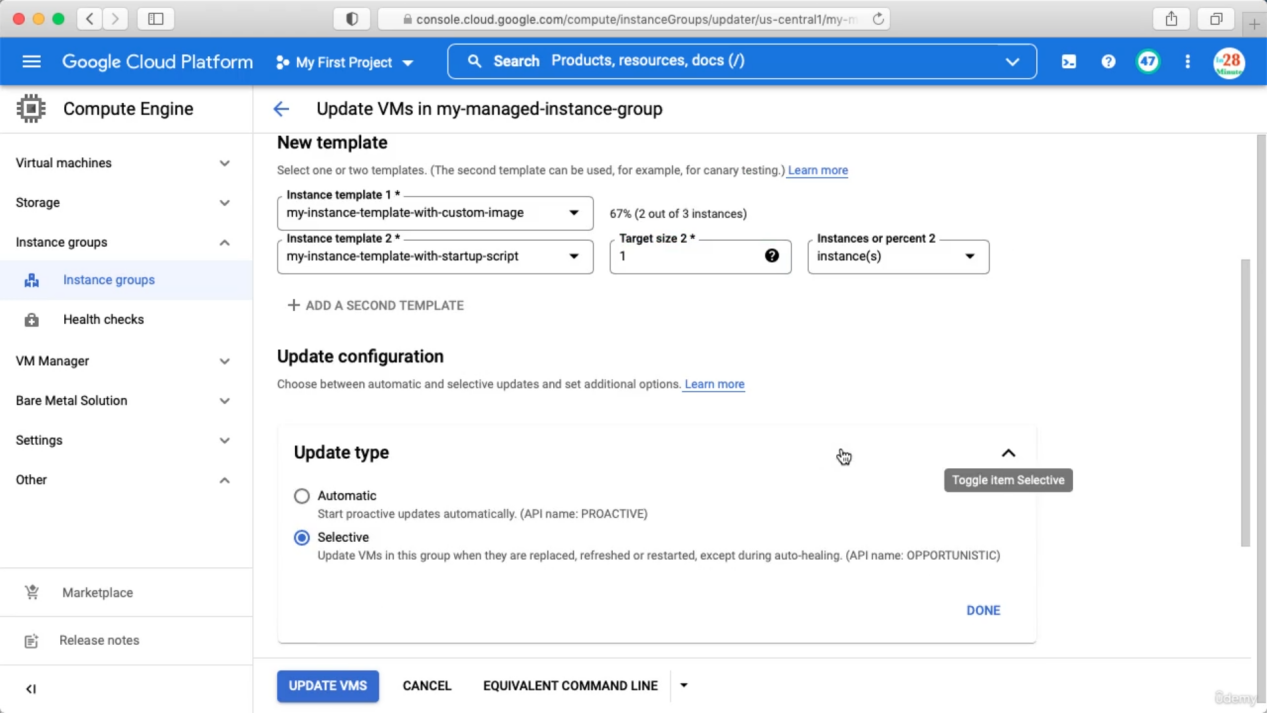
For rolling update click below



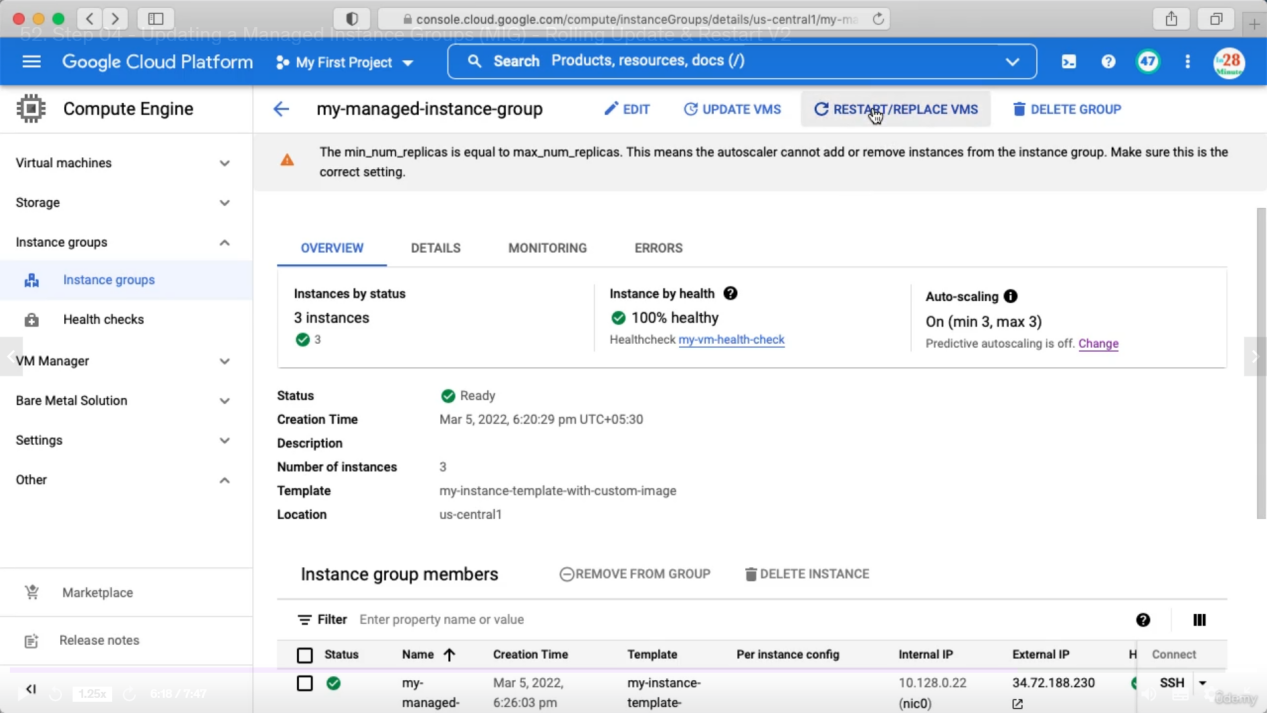
Here you can configure new instance template



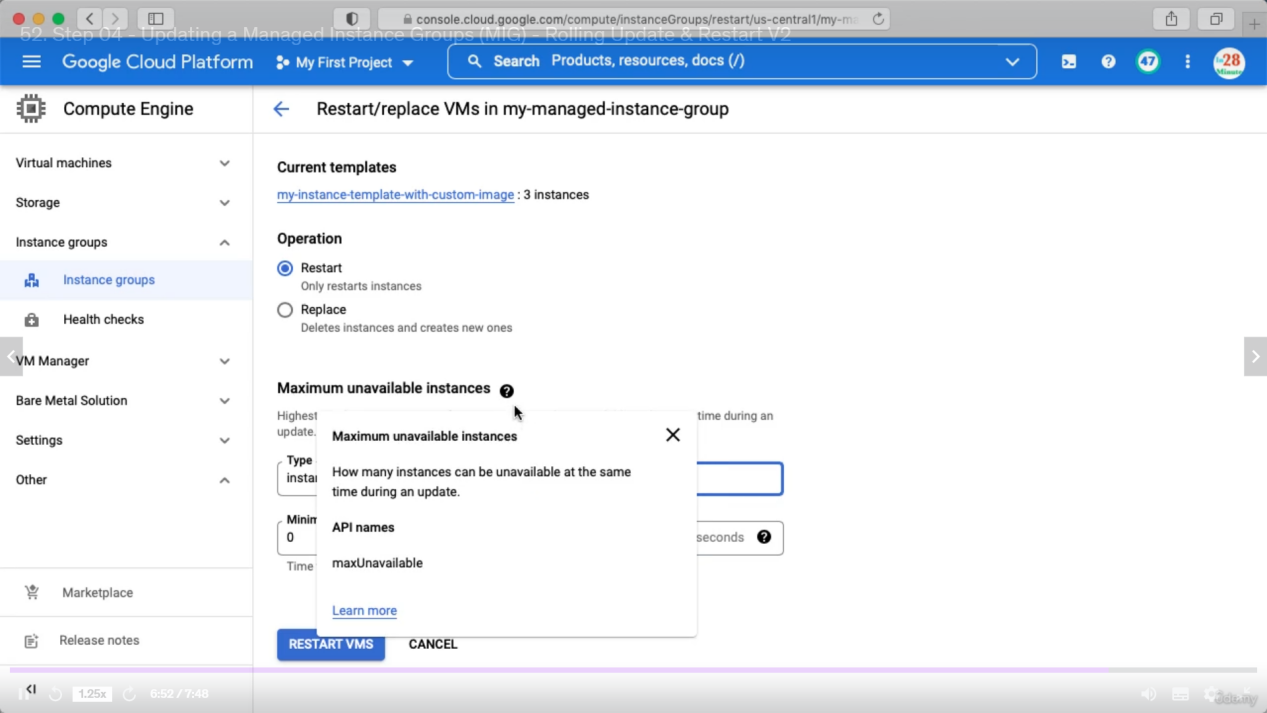
Canary update. Here in the template 2 you can specify new template and number of instances to be updated first.



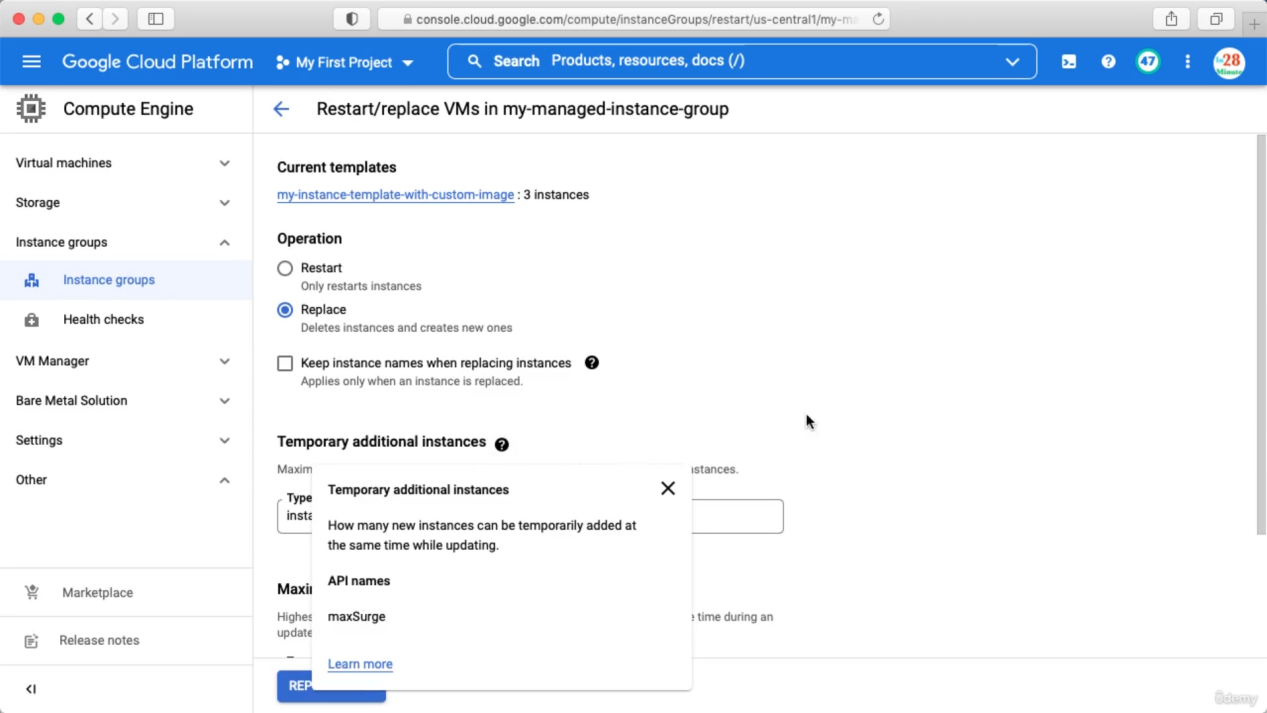
Next approach: Restart/Replace VM



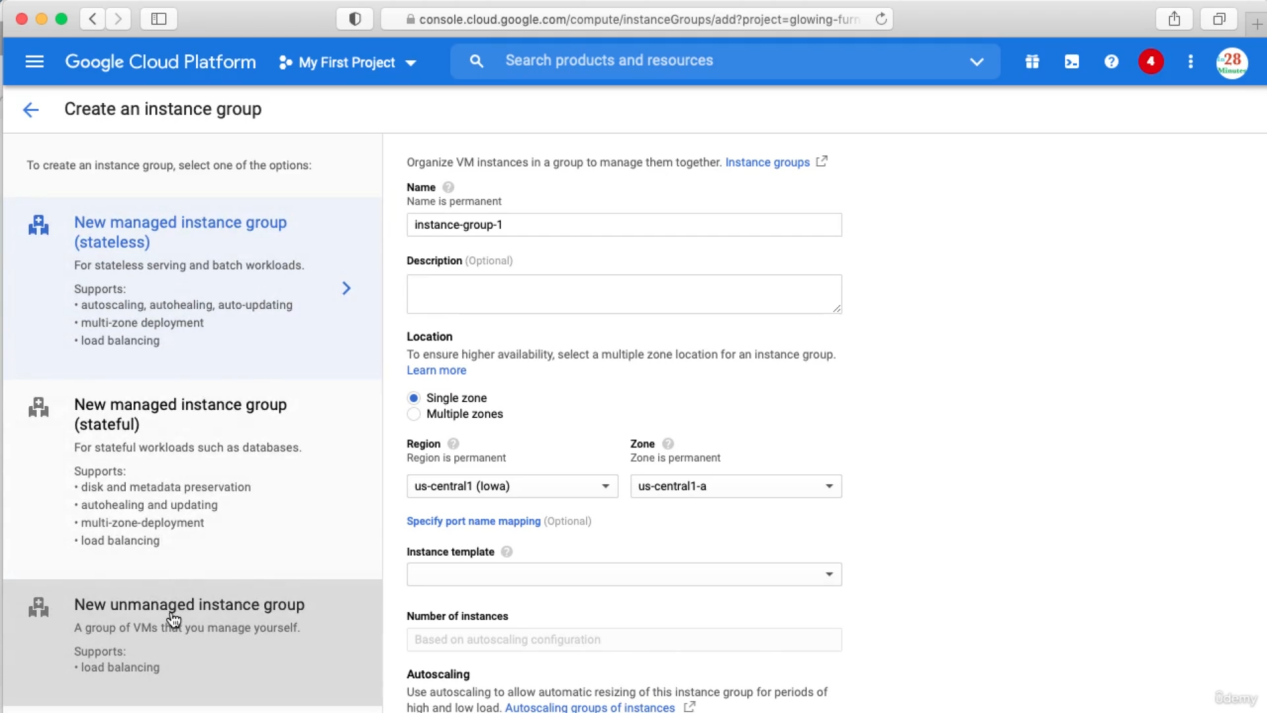
In the case of rolling restart or replace, we are not changing the instance template. All that we need to do is either restart or replace of existing VM instances using the existing template itself. So you can choose the right operation you want to do. If you want to restart then you can configure only maxUnavailable. How many instances can be unavailable during the restart? So if there are 10 instances, let's say you'd want to restart 2 instances at a time then you can configure that in here. If you'd want to replace the VMs, in that kind of situation you also have the option to configure a maxSurge. So in addition to maximum unavailable instances, you can also configure temporary additional instances. That's basically maxSurge. So I can say I'd want to create 2 new instances.before starting the replacement process. Understanding maxSurge and maxUnavailable is really, really important.



Replace. Means we want to update 2 instances first before replacement.

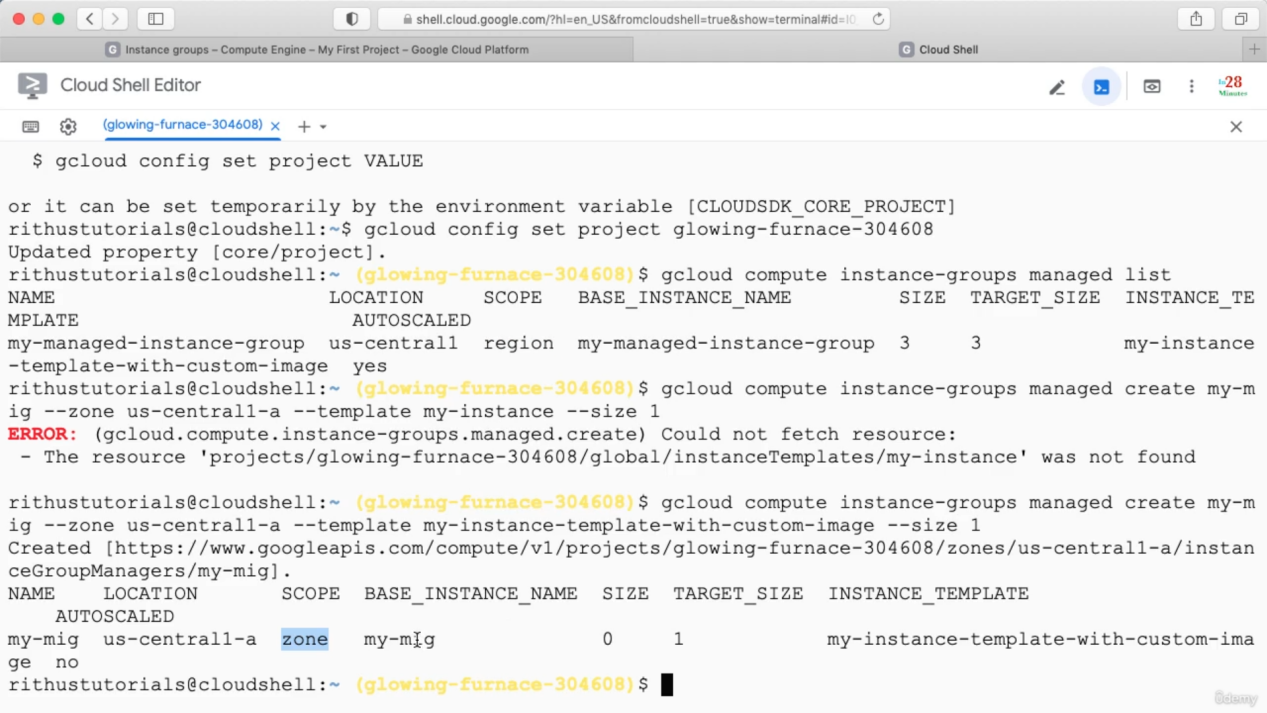


Demo: un-managed instance groups



Gcloud for managed instance group

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



--minimal-action : minimum it will replace/refresh or restart, --most-disruptive-allowed-action: maximum it will do none or replace or restart

