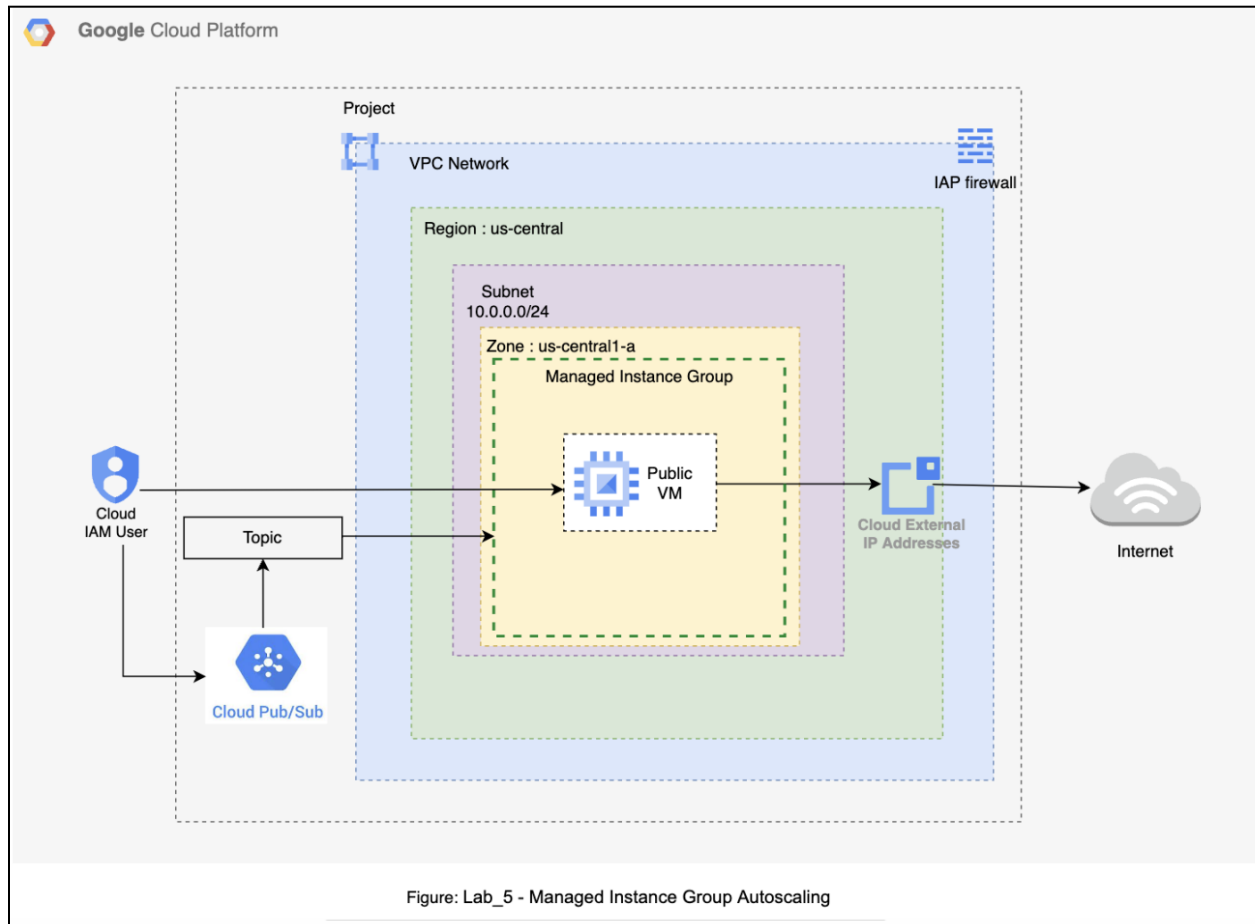


Architecture Diagram



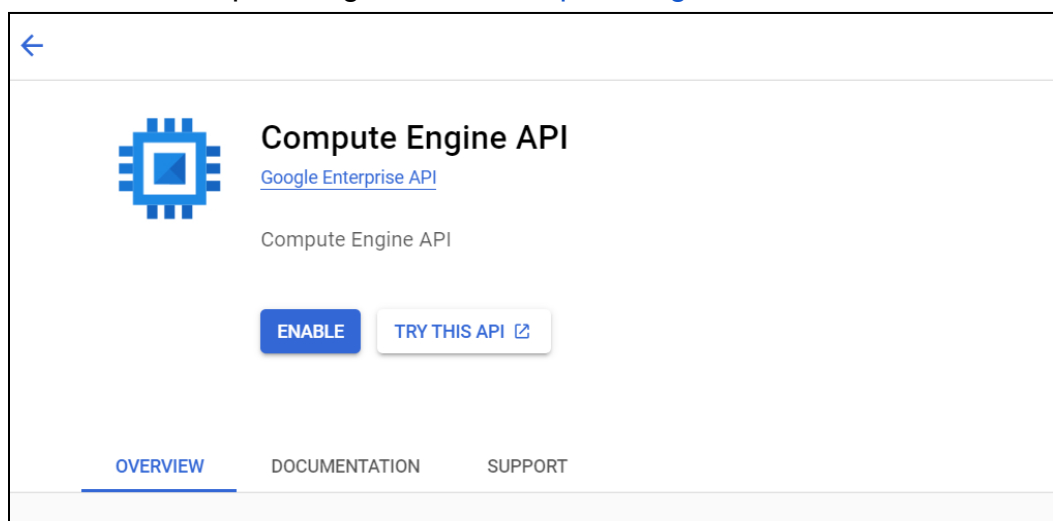
Autoscaling VM based CPU Utilization & pub/sub triggers to simulate load and autoscale the nodes

Prerequisites

Create and manage Google Cloud resources and services directly on the command line.

To create a private virtual machine in GCP following resources such as custom VPC, custom subnets, NAT gateway and IAP firewall rule is required to be provisioned.

Enable the Compute Engine API - [Compute Engine API](#)



Note: If you have followed Lab 1 for creation of VPC, subnets, NAT gateway, Firewall and Service account, Below prerequisites 4 steps are not required as you may have the resources created previously.


For someone who has not followed the previous lab follow all the prerequisites steps

1. Create a **Virtual Private Cloud (VPC) & Subnet**

VPC provides networking for your cloud-based resources and services that are global, scalable, and flexible.

Deploying Virtual Private Connect & Subnet using Google Cloud Console

Metadata - VPC & Subnet	Naming Convention
Name	labs-vpc
Subnet Creation Mode	Custom
New Subnet Name Region IP Stack type IPv4 Range Dynamic Routing	labs-subnet us-central1 IPv4 10.0.0.0/24 Regional

- In the Cloud Console, on the **Navigation menu** () , click **VPC network > VPC networks**.
This may take a minute to initialize for the first time.
- To create a new VPC, click **CREATE VPC NETWORK**.

There are many parameters you can configure when creating a **new VPC**.

- Enter VPC **Name** as it is a mandatory field.
- Description can be kept blank as it is an optional field.
- Select **Custom** checkbox for Subnet to create custom subnet.

VPC network

← Create a VPC network

VPC networks

IP addresses

Bring your own IP

Firewall

Routes

VPC network peering

Shared VPC

Serverless VPC access

Packet mirroring

<

Name *

labs-vpc

Lowercase letters, numbers, hyphens allowed

Description

VPC network ULA internal IPv6 range ?

Enabling this feature will assign a /48 from Google defined ULA prefix fd20::/20.

☐ Enabled
 ☒ Disabled

Subnets

Subnets let you create your own private cloud topology within Google Cloud. Click Automatic to create a subnet in each region, or click Custom to manually define the subnets. [Learn more](#)

Subnet creation mode ?

☒ Custom
 ☐ Automatic

- Click on **Add Subnet**.
- Enter Subnet **Name** as it is a mandatory field.
- Description can be kept blank as it is an optional field.
- Region here selected is us-central1. Region can be chosen upon end user's requirements to obtain low latency of resources.
- For more information about regions, see the Compute Engine guide, [Regions and Zones](#).
- IP Stack type can be left default with IPv4 checkbox selected.
- Enter **IPv4 range** for Subnet depending upon requirement.
- Rest of the fields can be left with default values.
- Click the **CREATE** button to create a new VPC.

New subnet ^

Name *
labs-subnet ?

Lowercase letters, numbers, hyphens allowed

Description

Region *
us-central1 ▼ ?

IP stack type

☒ IPv4 (single-stack)

☐ IPv4 and IPv6 (dual-stack) ?

IPv4 range *
10.0.0.0/24 ?

E.g. 10.0.0.0/24

[CREATE SECONDARY IPV4 RANGE](#)

Dynamic routing mode ?

☒ **Regional**
Cloud Routers will learn routes only in the region in which they were created

☐ **Global**
Global routing lets you dynamically learn routes to and from all regions with a single VPN or interconnect and Cloud Router

DNS server policy
No server policy ▼ ?

Maximum transmission unit (MTU)
1460 ▼

[CREATE](#) [CANCEL](#)

[EQUIVALENT COMMAND LINE](#) ▼

- Following resources are created
- VPC named **labs-vpc** and subnet named **labs-subnet**.

▼ labs-vpc	1	1460	Custom	None
us-central1	labs-subnet	10.0.0.0/24	None	

2. Create **NAT Gateway**

Cloud NAT ([network address translation](#)) lets certain resources without external IP addresses create outbound connections to the internet.

Deploying NAT Gateway using Google Cloud Console

Metadata - NAT Gateway	Naming Convention
Name	labs-nat-gw
Select Cloud Router Network Region	labs-vpc us-central1
Create a Router Name	labs-nat-router
Cloud NAT Mapping Resource (Internal)	Primary and Secondary Ranges for all subnets

- In the Cloud Console, on the **Navigation menu** (≡), click **Network services > Cloud NAT**.
- To create a new NAT gateway, click **CREATE CLOUD NAT GATEWAY**.

There are many parameters you can configure when creating a new NAT gateway

- Enter NAT **Gateway name** as it is a mandatory field.
- In Select Cloud Router, select the VPC Network created previously i.e. labs-vpc
- Select the region associated with the subnet.

Network services

Load balancing

Cloud DNS

Cloud CDN

Cloud NAT

Traffic Director

Service Directory

Cloud Domains

Private Service Connect

Marketplace

Release Notes

Create Cloud NAT gateway

Cloud NAT lets your VM instances and container pods communicate with the internet using a shared, public IP address.

Cloud NAT uses Cloud NAT gateway to manage those connections. Cloud NAT gateway is region and VPC network specific. If you have VM instances in multiple regions, you'll need to create a Cloud NAT gateway for each region. [Learn more](#)

Gateway name *

labs-nat-gw

Lowercase letters, numbers, hyphens allowed

Select Cloud Router

Network *

labs-vpc

Region *

us-central1 (Iowa)

One subnet.

Cloud Router *

- In Cloud Router, click on create new router.

Network services

Load balancing

Cloud DNS

Cloud CDN

Cloud NAT

Traffic Director

Service Directory

Cloud Domains

Private Service Connect

Marketplace

Release Notes

Create Cloud NAT gateway

Region *

us-central1 (Iowa)

One subnet.

Cloud Router *

Filter |Type to filter

Create new router

Primary and secondary ranges for all subnets

Cloud NAT IP addresses

Automatic (recommended)

Destination (external)

Internet

ADVANCED CONFIGURATIONS

- Enter the Cloud Router name as it is a mandatory field. Rest of the fields can be left with default values.
- Click on **CREATE** to create a Cloud router.

Create a router

Google Cloud Router dynamically exchanges routes between your Virtual Private Cloud (VPC) and on-premises networks by using Border Gateway Protocol (BGP)

Name *
labs-nat-router ?
Lowercase letters, numbers, hyphens allowed

Description

Network *
labs-vpc ?

Region *
us-central1 (Iowa) ?

BGP peer keepalive interval seconds ?

CREATE CANCEL

- The router's name is populated in the Cloud Router field.
- Rest of the fields can be left with default values.

Create Cloud NAT gateway

us-central1 (Iowa) ?

One subnet.

Cloud Router *
labs-nat-router ?

Cloud NAT mapping ?

Source (internal) ?
Primary and secondary ranges for all subnets ?
Select which subnets to map to the Cloud NAT gateway. Primary IP addresses are used by VM instances and secondary IP addresses are used by container pods. [Learn more](#)

Cloud NAT IP addresses
Automatic (recommended) ?

Destination (external)
Internet

▼ **ADVANCED CONFIGURATIONS**

CREATE CANCEL

- Cloud NAT gateway named **labs-nat-gw** is created.

<input type="checkbox"/>	labs-nat-gw	us-central1	labs-nat-router	✓ Running	⋮
--------------------------	-----------------------------	-------------	---------------------------------	-----------	---


3. Create IAP Firewall Rule.

To allow IAP to connect to your VM instances, create a firewall rule that:

- applies to all VM instances that you want to be accessible by using IAP.
- allows ingress traffic from the IP range **35.235.240.0/20**. This range contains all IP addresses that IAP uses for TCP forwarding.
- allows connections to all ports that you want to be accessible by using IAP TCP forwarding, for example, port 22 for SSH and port 3389 for RDP.

Deploying IAP Firewall using Google Cloud Console

Metadata - Firewall	Naming Convention
Name	labs-iap-fw-rule
Network	labs-vpc
Direction of Traffic Action on Match	Ingress Allow
Target	All Instance in Network
Source Filter	IPv4 Ranges
Source IPv4 Ranges	35.235.240.0/20
Protocol & Ports TCP	Specified Protocols & Ports 22, 3389

- In the Cloud Console, on the **Navigation menu** () , click **VPC network > Firewall**.
- To create a new IAP Firewall rule, click **CREATE FIREWALL RULE**.

There are many parameters you can configure when creating a new firewall rule

- Enter firewall **Name** as it is a mandatory field.
- Description can be left blank as it is an optional field.
- Select the VPC **Network** created previously i.e. **labs-vpc** to attach the firewall to that network.

- Click on **CREATE** to create an IAP firewall.

Google Cloud | search-playground-08032... | Search for resources, docs, products, and more

VPC network

- VPC networks
- IP addresses
- Bring your own IP
- Firewall**
- Routes
- VPC network peering
- Shared VPC
- Serverless VPC access
- Packet mirroring

Create a firewall rule

Protocols and ports ?

☐ Allow all

☒ Specified protocols and ports

☒ TCP

Ports: 22,3389

E.g. 20, 50-60

☐ UDP

Ports:

E.g. all

☐ Other

Protocols:

Separate multiple protocols by commas, e.g. ah, sctp

▼ DISABLE RULE

CREATE CANCEL

EQUIVALENT COMMAND LINE ▼

- The IAP firewall named **labs-iap-fw-rule** is created.

Filter	Name	Type	Targets	Filters	Protocols / ports	Action	Priority	Network	Logs	Hit count
labs-iap-fw-rule	labs-iap-fw-rule	Ingress	Apply to all	IP ranges: 35.235.240.0/20	tcp:22, 3389	Allow	1000	labs-vpc	Off	

4. Create Service Account for Compute Engine

A service account is identified by its email address, which is unique to the account. Before creating the service account following the below steps

- Enable the IAM API - [IAM API](#)

Required roles for your IAM account.

To get the permissions that you need to manage service accounts, grant the following IAM roles on the project:

To view and create service accounts:

Create Service Accounts (roles/iam.serviceAccountCreator)

Deploying Custom Service Account using Google Cloud Console

Metadata - Service Account	Naming Convention
Service Account Details Service Account Name	labs-compute-sa
Add Principals New Principals	labs-compute-sa
Assign roles	Storage Admin

- In the Cloud Console, on the **Navigation menu** (≡), click **IAM & Admin > Service Account**.
- To create a new custom service account, click **CREATE SERVICE ACCOUNT**.

There are many parameters you can configure when creating a new firewall rule

- Enter **Service Account Name** as it is a mandatory field.
- Service account ID is auto populated with service account name.
- Description can be kept blank as it is an optional field.
- Click **CREATE AND CONTINUE** to create the custom service account.

IAM & Admin

← Create service account

1 Service account details

Service account name
labs-compute-sa
Display name for this service account

Service account ID *
labs-compute-sa
Email address: labs-compute-sa@[redacted].iam.gserviceaccount.com

Service account description
Describe what this service account will do

CREATE AND CONTINUE

2 Grant this service account access to project (optional)

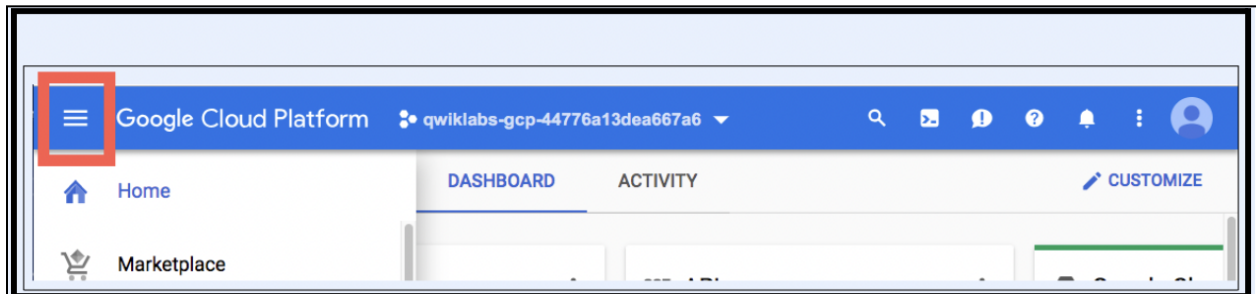
3 Grant users access to this service account (optional)

- Other fields are optional and can be skipped. The custom service account name **labs-compute-sa@project-id.iam.gserviceaccount.com** is created.

Filter labs-compute-sa Enter property name or value						
<input type="checkbox"/>	Email	Status	Name ↑	Key ID	Key creation date	Actions
<input type="checkbox"/>	labs-compute-sa@[REDACTED].iam.gserviceaccount.com	✓	labs-compute-sa	No keys		⋮

- Navigate to IAM, click **Grant ACCESS**.
- Enter the created service account name in **New Principals**.
- In Role, Select **Storage Admin** from the dropdown.

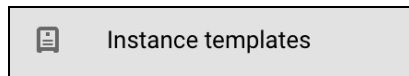
The screenshot shows the 'IAM & Admin' console. The left sidebar lists various IAM-related tools. The main area is titled 'Permissions for project "search-engine"' and shows a table of principals. A 'Filter' bar is at the top of the table. The table has columns for 'Type', 'Principal', and 'Role'. One principal is listed: 'labs-compute-sa@[REDACTED].iam.gserviceaccount.com' with the role 'Storage Admin'. The right sidebar shows the 'Add principals' section with a search bar containing the service account email. Below that, the 'Assign roles' section shows a dropdown menu with 'Storage Admin' selected. The bottom of the console has 'SAVE' and 'CANCEL' buttons.



Note: You can view the menu with a list of **Google Cloud Products** and **Services** by clicking the **Navigation menu** at the top-left.

Create Instance templates for Manage instance groups.

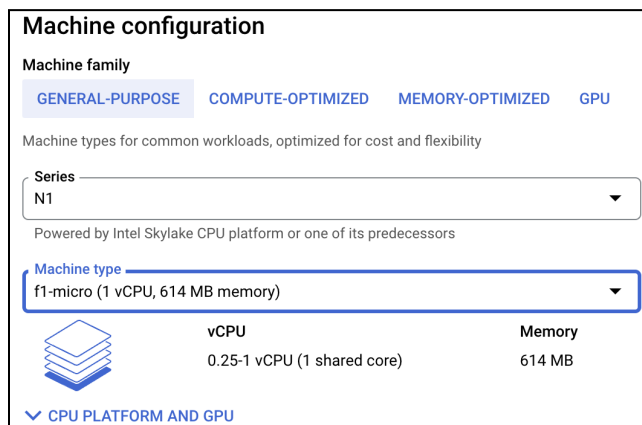
- Follow the given step to create instance template :
 - Click Compute Engine in the left navigation panel.
 - Click on instance templates.



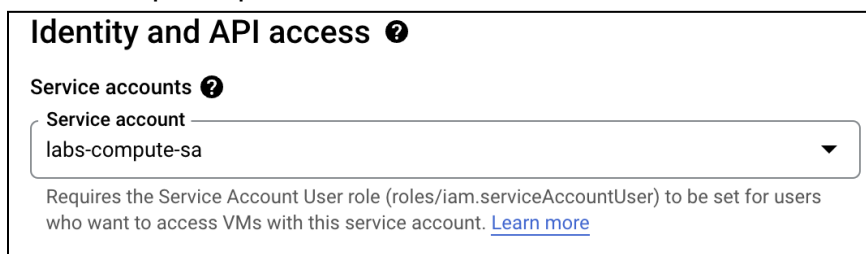
- Click on CREATE INSTANCE TEMPLATES.
- Enter the details > Name : <Instance template name>

A text input field with a light gray border. Inside the field, the text 'Name *' is at the top left, and 'instance-template-4' is entered below it.

- Machine Type >
 - series : n1
 - machine type : f1-micro

A form section titled 'Machine configuration'. It has a sub-header 'Machine family' with four tabs: 'GENERAL-PURPOSE' (selected), 'COMPUTE-OPTIMIZED', 'MEMORY-OPTIMIZED', and 'GPU'. Below the tabs is a note: 'Machine types for common workloads, optimized for cost and flexibility'. There are two dropdown menus: 'Series' with 'N1' selected, and 'Machine type' with 'f1-micro (1 vCPU, 614 MB memory)' selected. Below the 'Machine type' dropdown is a table with three columns: 'vCPU' (0.25-1 vCPU (1 shared core)), 'Memory' (614 MB), and an icon of stacked disks. At the bottom, there is a link 'CPU PLATFORM AND GPU'.

- Identity and API access > Service accounts : <Add service account created in prerequisites>

A form section titled 'Identity and API access' with a help icon. It has a sub-header 'Service accounts' with a help icon. Below it is a dropdown menu for 'Service account' with 'labs-compute-sa' selected. At the bottom, there is a note: 'Requires the Service Account User role (roles/iam.serviceAccountUser) to be set for users who want to access VMs with this service account. [Learn more](#)'.

- Advanced options > Network interfaces > Select : Networks in this project
network : <select vpc name created in prerequisites>
Subnetwork : <select subnetwork created in prerequisites>
click on Done. External IPv4 address should be None.

Network interfaces ?

Network interface is permanent

Edit network interface ⬆

☒ Networks in this project

☐ Networks shared with me (from host project: "searce-playground-host-project")

Network *

labs-vpc ▼ ?

Subnetwork *

labs-subnet (us-central1) IPv4 (10.0.0.0/24) ▼ ?

IP stack type

☒ IPv4 (single-stack)

☐ IPv4 and IPv6 (dual-stack)

Alias IP ranges

+ ADD IP RANGE

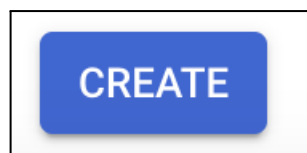
External IPv4 address

None ▼ ?

- add startup script in management > startup script :

```
#!/bin/bash
apt update
apt -y install apache2
echo "Hello world from $(hostname) $(hostname -I)" >
/var/www/html/index.html
```

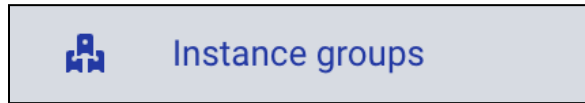
- Click on CREATE.



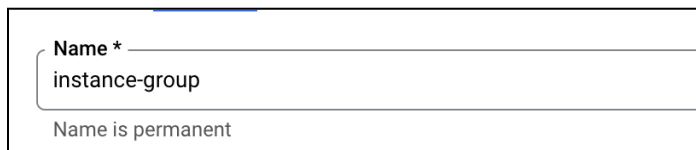
Configure Managed instance groups

- To Configure managed instance groups these are the following steps:

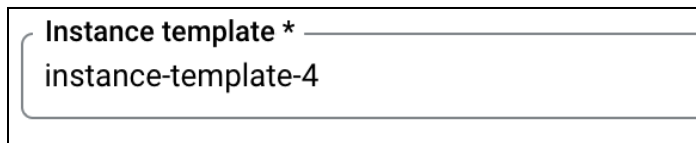
- Click Compute Engine in the left navigation panel.
- Click on Instance groups.



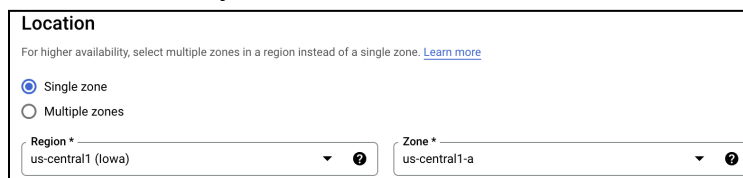
- Click on CREATE INSTANCE GROUP.
- Select > New managed instance group (stateless)

A screenshot of a form field labeled 'Name *'. The text 'instance-group' is entered. Below the field, it says 'Name is permanent'.

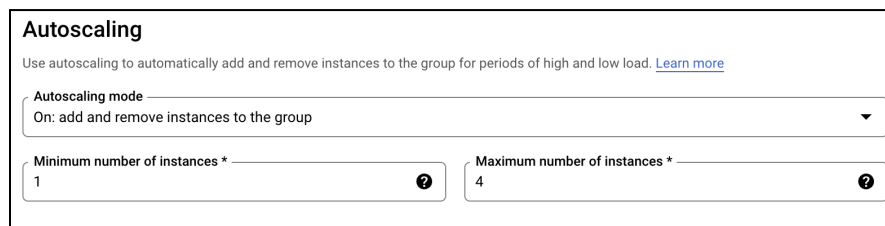
- ENTER the following details > Name : <Enter the instance groups name>

A screenshot of a form field labeled 'Instance template *'. The text 'instance-template-4' is entered.

- Instance templates : <Select the instance template create in above step>
Location > Select : Single Zone Region: <Select your region>
Zone: <Select you zone>

A screenshot of the 'Location' section. It includes a note: 'For higher availability, select multiple zones in a region instead of a single zone. [Learn more](#)'. There are two radio buttons: 'Single zone' (selected) and 'Multiple zones'. Below are two dropdown menus: 'Region *' with 'us-central1 (Iowa)' selected, and 'Zone *' with 'us-central1-a' selected. Both dropdowns have a question mark icon.




- Autoscaling >
Auto Scaling mode > select : On: add and remove instances to the group
Minimum number of instances : <Enter 1 >
Maximum number of instances : <Enter 4 >




A screenshot of the 'Autoscaling' section. It includes a note: 'Use autoscaling to automatically add and remove instances to the group for periods of high and low load. [Learn more](#)'. There is a dropdown menu for 'Autoscaling mode' with 'On: add and remove instances to the group' selected. Below are two input fields: 'Minimum number of instances *' with '1' entered, and 'Maximum number of instances *' with '4' entered. Both input fields have a question mark icon.

- Click on CREATE.

Result : MIG Auto scaling

Use following Command : To scale the CPU : `ab -n 10000 -c 100 'http://host_ip/'`
 Make sure your IP Address is added in the firewall which we create in the above step.

OVERVIEW	DETAILS	MONITORING	ERRORS
Instances by status 2 instances  2		Instance b  100% Healthcheck	
Status		 Updating	
Creation Time		Nov 11, 2022, 4:26:51 PM UTC+05:30	

OVERVIEW	DETAILS	MONITORING	ERRORS
<div>Instances by status</div> <div>2 instances</div> <div> 2</div>		<div>Instance by health </div> <div>Not configured</div> <div>Autohealing off. Configure</div>	
Status		 Ready	
Creation Time		Nov 16, 2022, 4:55:00 PM UTC+05:30	
Description			
Number of instances		2	
Template		instance-template-labs	
Location		us-central1-a	

Configure the PUB/SUB

To configure PUB/SUB below are the following step :

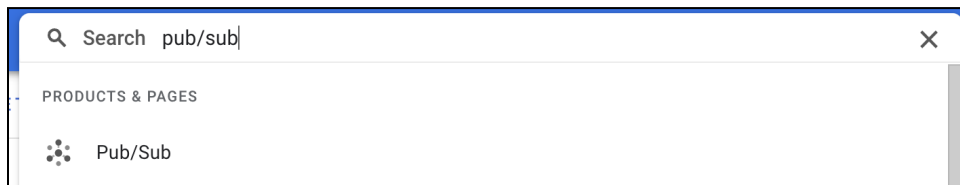
To set up Auto Scaling via Pub/Sub. First of all, create a Topic and add that topic into Subscriptions.

- Create TOPIC

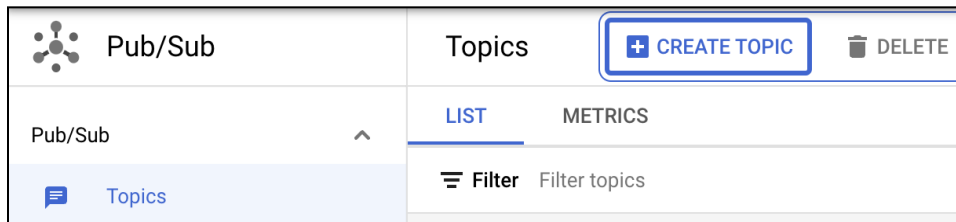
- In google console search pub/sub in the search tab.



- Click on Pub/Sub.



- Click on Pub/sub > Topics > CREATE TOPIC



- Enter the Topic name and leave the rest of the fields default.

Create a topic

A topic forwards messages from publishers to subscribers.

Topic ID *

lab-topic

?

Topic name: projects/searce-playground-v1/topics/lab-topic

☒ Add a default subscription ?

☐ Use a schema ?

☐ Set message retention duration (not free) ?

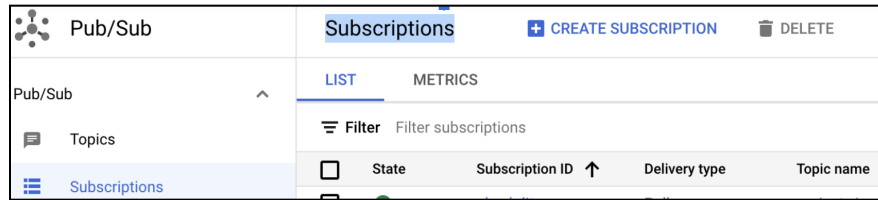
Encryption

☒ Google-managed encryption key
No configuration required

☐ Customer-managed encryption key (CMEK)
Manage via Google Cloud Key Management Service

CANCEL CREATE TOPIC

- Click on CREATE TOPIC.
- Create Subscriptions.
 - Click on the Subscriptions > CREATE SUBSCRIPTION.



- Enter the following details >

Subscriptions : <Enter the name of Subscription>

Select a Cloud Pub/Sub topic : <Select the topic created in above step>

A subscription directs messages on a topic to subscribers. Messages can be pushed to subscribers immediately, or subscribers can pull messages as needed.

Subscription ID * ?

Subscription name: projects/searce-playground-v1/subscriptions/lab-subscription

Select a Cloud Pub/Sub topic * ▼

Delivery type ?

☒ Pull

☐ Push

☐ Write to BigQuery

Message retention duration ?

Duration is from 10 minutes to 7 days

Days ▼ Hours ▼ Minutes ▼

☐ Retain acknowledged messages ?

When enabled, acknowledged messages are retained for the message retention duration specified above. This increases message storage fees. [Learn more](#)

- Leave the rest of the field default and Click on the CREATE.

To configure auto scaling based on unacknowledged messages in a Pub/Sub subscription, use the [subscription/num_undelivered_messages](#) metric provided by pubsub and filter by the subscription ID.

The subscription/num_undelivered_messages metric exports the total number of messages in the subscription, including messages that are currently being processed but that are not yet acknowledged.

Run the below Command on Cloud shell editor.

```
gcloud compute instance-groups managed set-autoscaling MIG_NAME \
--max-num-replicas=MAX_INSTANCES \
--min-num-replicas=MIN_INSTANCES \
--update-stackdriver-metric=pubsub.googleapis.com/subscription/num_undelivered_m
essages \
--stackdriver-metric-filter="resource.type=\"pubsub_subscription\" AND
resource.labels.subscription_id=\"SUBSCRIPTION_ID\"" \
--stackdriver-metric-single-instance-assignment=NUMBER_OF_MESSAGES_TO_ASSI
GN_TO_EACH_VM
```

```
Example : gcloud compute instance-groups managed set-autoscaling
instance-group-mig \
--max-num-replicas=4 \
--min-num-replicas=1 \
--update-stackdriver-metric=pubsub.googleapis.com/subscription/num_undelivered_
messages \
--stackdriver-metric-filter="resource.type=\"pubsub_subscription\" AND
resource.labels.subscription_id=\"lab-subscription\"" \
--stackdriver-metric-single-instance-assignment=2
```

Now Setup is complete for pub/sub triggers to simulate load and autoscaling of nodes.

- Check the above setup using following step :
 - In google console search pub/sub in the search tab.
 - Click on Pub/sub > Topic > your topic name

Pub/Sub		LIST	METRICS
Topics		Filter lab Filter topics	
Subscriptions		Topic ID ↑	Encryption key
		lab-topic	Google-managed

- Click on Your topic name.

- Click on MESSAGE > PUBLISH MESSAGE.

SUBSCRIPTIONS

SNAPSHOTS

MESSAGES

METRICS

DETAILS

PUBLISH MESSAGE

Manually publish a message containing data and attributes with metadata to this topic.

- Enter the following details :
Number of messages : <Enter you desire number to publish your message>
Message interval (Seconds) : <Enter you desire number So, It publish your message in given interval>
Message body > Message : <Enter your contain that you what to publish>

Publish message

Topic name
projects/searce-playground-v1/topics/lab-topic

Publish count

You can publish the given message once or multiple times in an interval. This can be useful for getting messages in new subscriptions and testing. For a more robust way to publish messages multiple times, consider using Cloud Scheduler.

Number of messages *
5
Enter an amount between 1-100.

Message interval (seconds) *
1
How long to wait before publishing the next message

Message body

The message you want to publish to this topic. Either message or attribute will be required to publish.

Message *
my first message.

PUBLISH

CANCEL

- Click on PUBLISH to publish your message.

The publish job has completed.

Pending	0
Success	4
Error	0

- Click on Subscription <your Subscription ID>

SUBSCRIPTIONS

SNAPSHOTS

Only subscriptions attached to this topic
creating a subscription from a Cloud Data

CREATE SUBSCRIPTION

EXPLORE

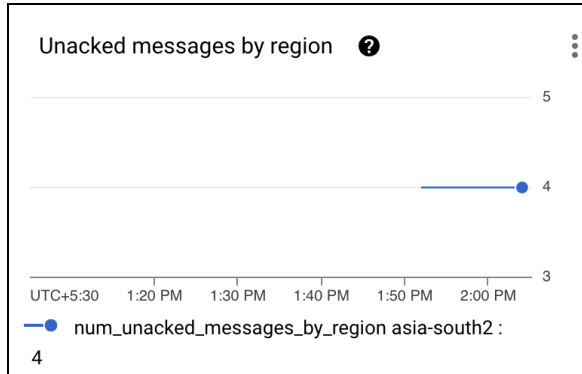
Filter

Filter subscriptions

Subscription ID	Subscription name
lab-subscription	projects/searce-p

Now check the METRICS of your subscription, <Unacked messages by region>, here It shows Unacked message Four. Now check your Manage instance groups (MIG) METRICS Via click on MONITORING. In MIG graphs It shows the scaled instance and Unacked message with the threshold value line.





- To scale down the instance just click on ACK.
 - Click Pull to view messages and temporarily delay message delivery to other subscribers. Select Enable ACK messages and then click ACK next to the message to permanently prevent message delivery to other subscribers. Note : ACK will show as a deadline exceed in some second so complete the ACK step ASAP

MESSAGES					
Click Pull to view messages and temporarily delay message delivery to other subscribers. Select Enable ACK messages and then click ACK next to the message to permanently prevent message delivery to other subscribers.					
<div>PULL <input checked="" type="checkbox"/> Enable ack messages</div>					
<div>Filter Filter messages</div>					
Publish time	Attribute keys	Message body	Ordering key	Ack ↑	
Nov 16, 2022, 1:50:36 PM	—	MY first messga	—	ACK	
Nov 16, 2022, 1:50:37 PM	—	MY first messga	—	ACK	
Nov 16, 2022, 1:50:38 PM	—	MY first messga	—	ACK	
Nov 16, 2022, 1:50:39 PM	—	MY first messga	—	ACK	

RESULTS : It takes 5 to 10 mint for showing the Scale down result. It will look like the images below.

OVERVIEW

DETAILS

MONITORING

ERRORS

Instances by status

2 instances

2

Instance by health

100% healthy

Healthcheck

Status	Updating
Creation Time	Nov 11, 2022, 4:26:51 PM UTC+05:30
Description	
Number of instances	1
Template	instance-template-anil
Location	asia-south1-c
In use by	mig-anil-autoscaling

Instance Group Members

REMOVE FROM GROUP

Filter

Enter property name or value

<input type="checkbox"/>	Status	Name	Creation Time	Template
<input type="checkbox"/>		instance-group-	Nov 16, 2022, 1:46:46 PM	-

