

Load balancing et autoscaling GCP

Ce document a pour but de résumer les étapes à suivre pour mettre en place un load balancing et de l'auto-scale



Référence : EF-CLOUD-LB

Auteur(s) :

Dorian Manzanares

Destinataire(s) :

Easyformer

Date de modification : 17/11/22

Version : 1

Sommaire

page

1	INTRODUCTION	3
1.1	LE MODELE D'INSTANCE	4
1.1.1	<i>Création du modèle</i>	<i>4</i>
1.2	LE GROUPE D'INSTANCE	6
1.2.1	<i>Création du groupe.....</i>	<i>6</i>
1.3	LOAD BALANCING	8
1.3.1	<i>Création de l'équilibreur de charge</i>	<i>8</i>
1.3.2	<i>Le frontend</i>	<i>10</i>
1.3.3	<i>Backend.....</i>	<i>11</i>
1.4	TEST DU DE L'AUTO-SCALING.....	15
1.4.1	<i>Les instances de VM</i>	<i>15</i>

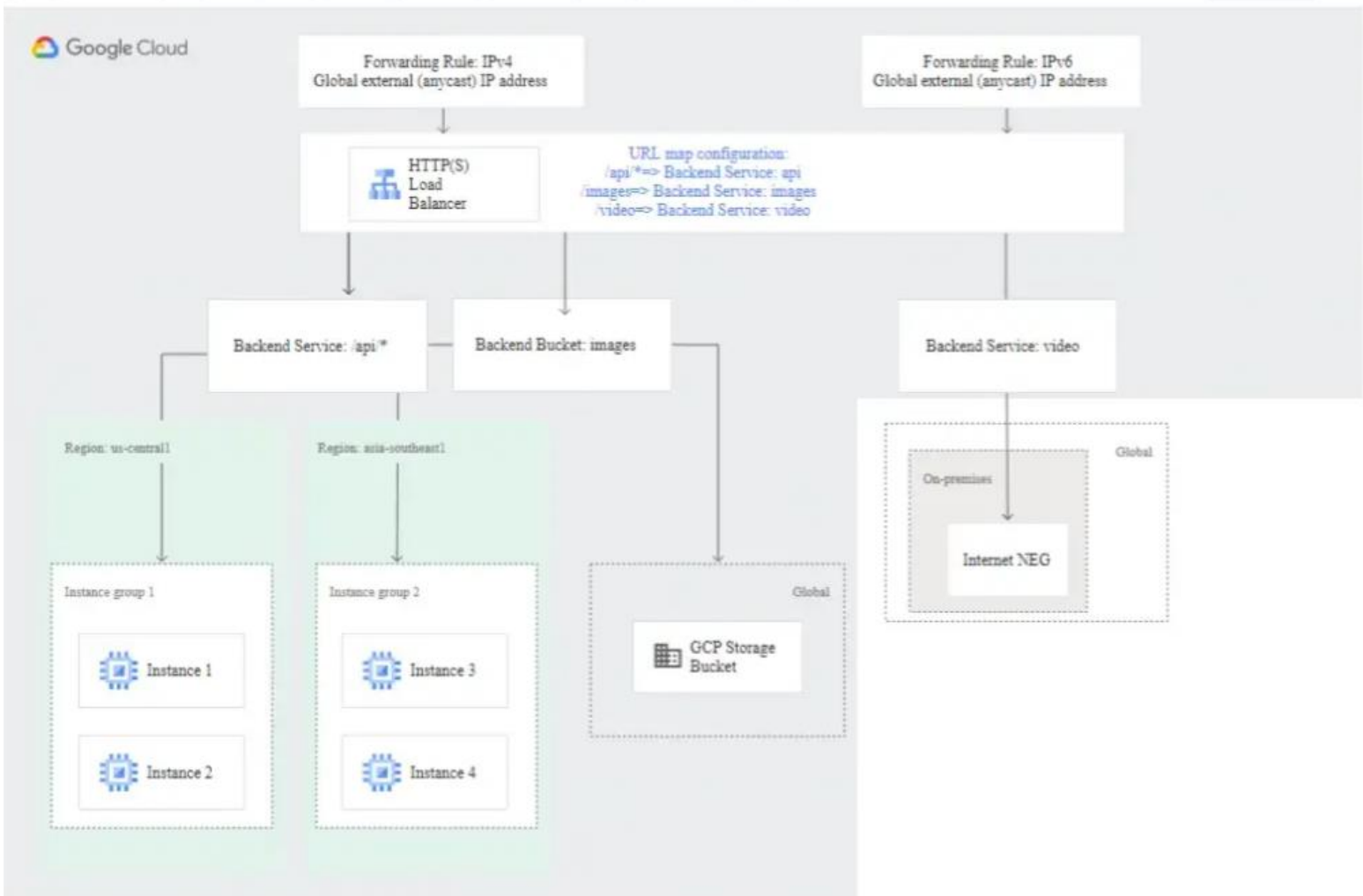


1 Introduction

Un load balancing est un équilibreur de charge, il va répartir le flux entrant sur les différentes instances disponibles.

Le principe d'auto-scale et l'adaptabilité du flux par nos instances de manière automatique.

Voici la répartition du load balancing HTTP de GCP.

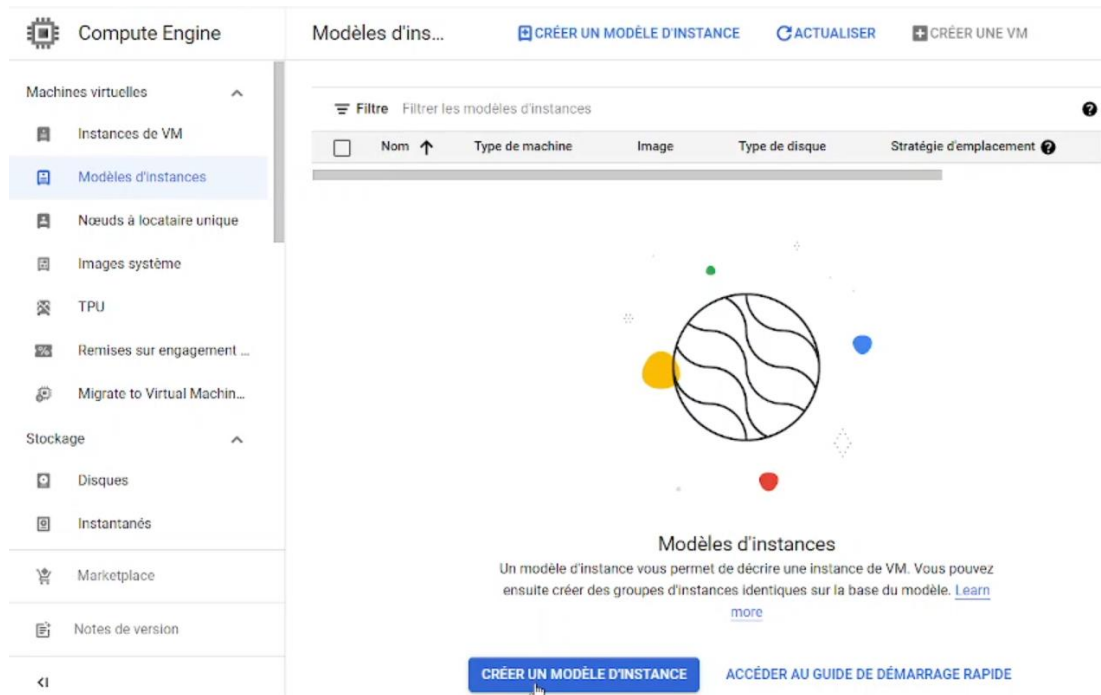


1.1 Le modèle d'instance

1.1.1 Création du modèle

Nous allons créer un modèle d'instance, concrètement, une instance qui servira d'exemple pour en déployer plusieurs identiques.

Il faudra ici créer un nouveau modèle dans « compute engine »



The screenshot shows the Google Cloud Platform console interface for the Compute Engine service. The left sidebar contains a navigation menu with categories like 'Machines virtuelles', 'Stockage', and 'Marketplace'. The main content area is titled 'Modèles d'instances' and includes a search bar, a table header with columns 'Nom', 'Type de machine', 'Image', 'Type de disque', and 'Stratégie d'emplacement', and a central graphic of a globe. Below the graphic, there is a description of instance models and two buttons: 'CRÉER UN MODÈLE D'INSTANCE' and 'ACCÉDER AU GUIDE DE DÉMARRAGE RAPIDE'.



Il faudra commencer par la nommer, le nom est libre.

Puis sélectionner le type d'instance (nous prendrons ici le type présent dans le free tier , qui ne coutera rien)

Configuration de la machine

Famille de machines

USAGE GÉNÉRAL OPTIMISÉE POUR LE CALCUL MÉMOIRE OPTIMISÉE GPU

Types de machines pour les charges de travail courantes permettant d'optimiser les coûts et la flexibilité

Série
E2

Sélection de la plate-forme de processeur en fonction de la disponibilité

Type de machine
e2-micro (2 processeurs virtuels, 1 Go de mémoire)

	vCPU	Memory
	0.25 à 2 vCPU (1 cœur partagé)	1 Go

✓ PLATE-FORME DU CPU ET GPU

On va également faire en sorte d'autoriser le trafic depuis l'extérieur, ainsi que l'autorisation des protocoles web.

(A noter ici le prix normal de l'instance qui sera déployer sans le compte gratuit)

Compute Engine

Créer un modèle d'instance

Compte de service
Compute Engine default service account

Vous devez définir le rôle "Utilisateur du compte de service" (roles/iam.serviceAccountUser) pour les utilisateurs qui souhaitent accéder aux VM avec ce compte de service. En savoir plus

Niveaux d'accès

Autoriser l'accès par défaut

Autoriser l'accès complet à l'ensemble des APIs Cloud

Définir l'accès pour chaque API

Pare-feu

Ajouter des tags et des règles de pare-feu pour autoriser un trafic réseau spécifique provenant d'Internet

Autoriser le trafic HTTP

Autoriser le trafic HTTPS

Options avancées
Mise en réseau, disques, sécurité, gestion, location unique

These are estimated costs for a VM instance created using this template:

Estimation mensuelle
7,99 \$US
Soit un coût horaire d'environ 0,01 \$US

Vous payez ce que vous consommez : facturation à la seconde, sans frais initiaux à supporter

Élément	Estimation mensuelle
2 vCPU + 1 GB memory	6,89 \$US
Disque persistant avec équilibrage, d'une taille de 10 Go	1,10 \$US
Sustained use discount	-0,00 \$US
Total	7,99 \$US

Tarifs de Compute Engine

^ LESS

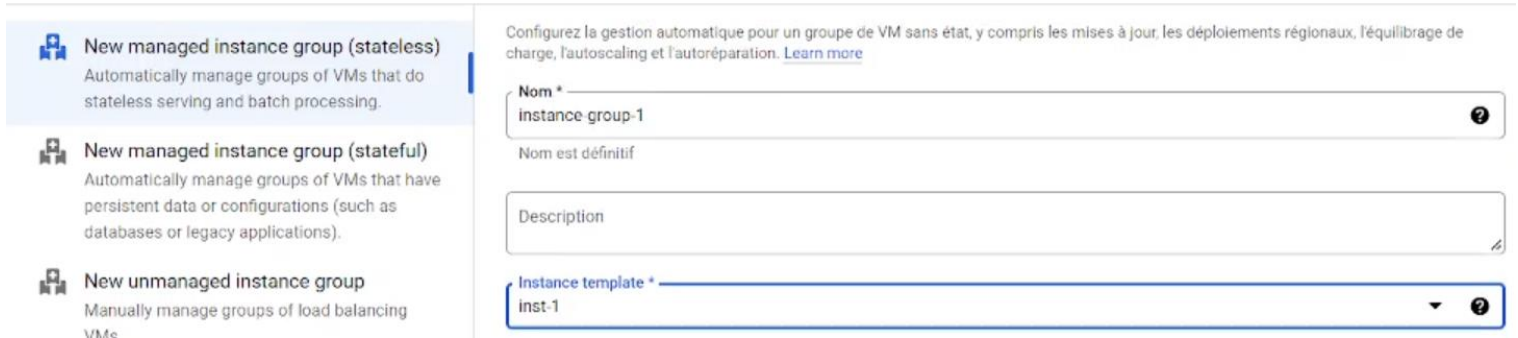
On pourra ensuite déployer le modèle.



1.2 Le groupe d'instance

1.2.1 Création du groupe

On prendra ici un nouveau groupe avec un nom libre



Configurez la gestion automatique pour un groupe de VM sans état, y compris les mises à jour, les déploiements régionaux, l'équilibrage de charge, l'autoscaling et l'autoréparation. [Learn more](#)

Nom *
instance group-1

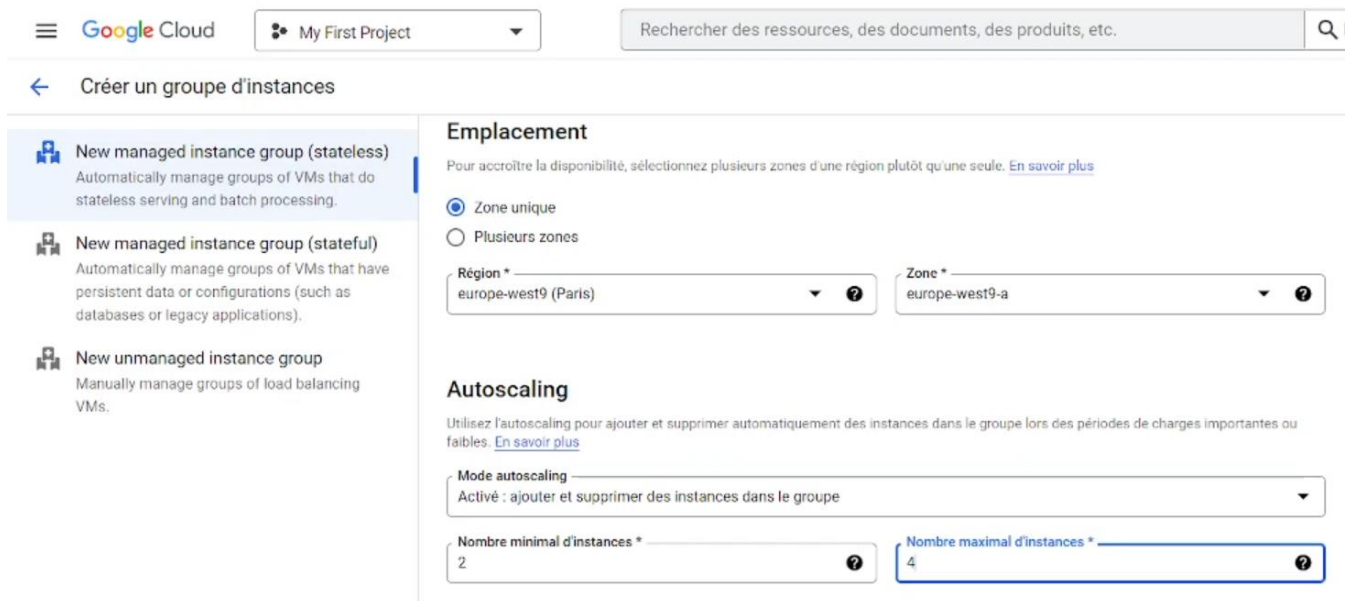
Nom est définitif

Description

Instance template *
inst-1

- New managed instance group (stateless)**
Automatically manage groups of VMs that do stateless serving and batch processing.
- New managed instance group (stateful)**
Automatically manage groups of VMs that have persistent data or configurations (such as databases or legacy applications).
- New unmanaged instance group**
Manually manage groups of load balancing VMs.

On choisira ensuite la bonne zone et le nombre d'instances minimum et maximum en cas de pannes.



Google Cloud My First Project Rechercher des ressources, des documents, des produits, etc.

← Créer un groupe d'instances

Emplacement
Pour accroître la disponibilité, sélectionnez plusieurs zones d'une région plutôt qu'une seule. [En savoir plus](#)

Zone unique
 Plusieurs zones

Région * europe-west9 (Paris) Zone * europe-west9-a

Autoscaling
Utilisez l'autoscaling pour ajouter et supprimer automatiquement des instances dans le groupe lors des périodes de charges importantes ou faibles. [En savoir plus](#)

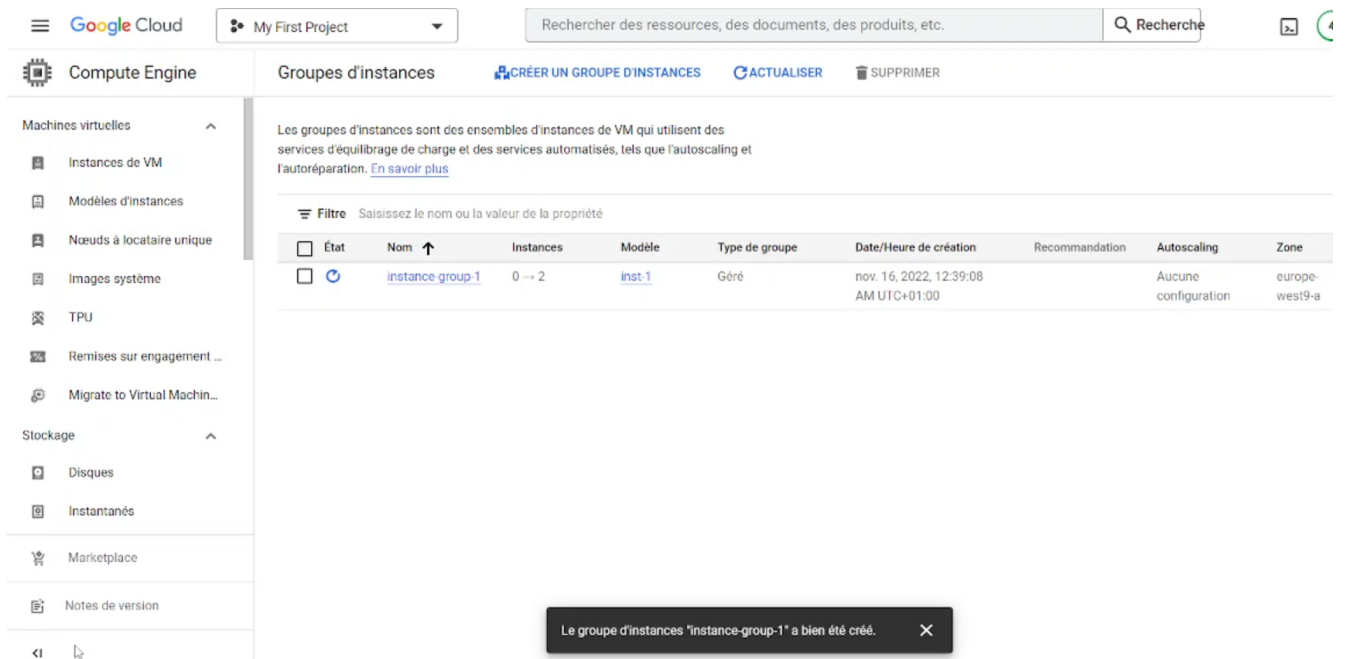
Mode autoscaling
Activé : ajouter et supprimer des instances dans le groupe

Nombre minimal d'instances * 2 Nombre maximal d'instances * 4

- New managed instance group (stateless)**
Automatically manage groups of VMs that do stateless serving and batch processing.
- New managed instance group (stateful)**
Automatically manage groups of VMs that have persistent data or configurations (such as databases or legacy applications).
- New unmanaged instance group**
Manually manage groups of load balancing VMs.



On peut ensuite créer notre groupe.



Google Cloud My First Project Rechercher des ressources, des documents, des produits, etc. Recherche

Compute Engine Groupes d'instances CRÉER UN GROUPE D'INSTANCES ACTUALISER SUPPRIMER

Machines virtuelles

- Instances de VM
- Modèles d'instances
- Nœuds à locataire unique
- Images système
- TPU
- Remises sur engagement ...
- Migrate to Virtual Machin...

Stockage

- Disques
- Instantanés

Marketplace

Notes de version

Les groupes d'instances sont des ensembles d'instances de VM qui utilisent des services d'équilibrage de charge et des services automatisés, tels que l'autoscaling et l'autoréparation. [En savoir plus](#)

Filtre Saisissez le nom ou la valeur de la propriété

État	Nom ↑	Instances	Modèle	Type de groupe	Date/Heure de création	Recommandation	Autoscaling	Zone
<input type="checkbox"/>	instance-group-1	0 → 2	inst-1	Géré	nov. 16, 2022, 12:39:08 AM UTC+01:00		Aucune configuration	europe-west9-a

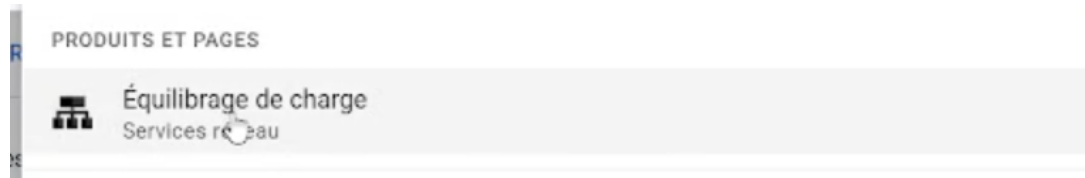
Le groupe d'instances "instance-group-1" a bien été créé.



1.3 Load balancing

1.3.1 Création de l'équilibreur de charge

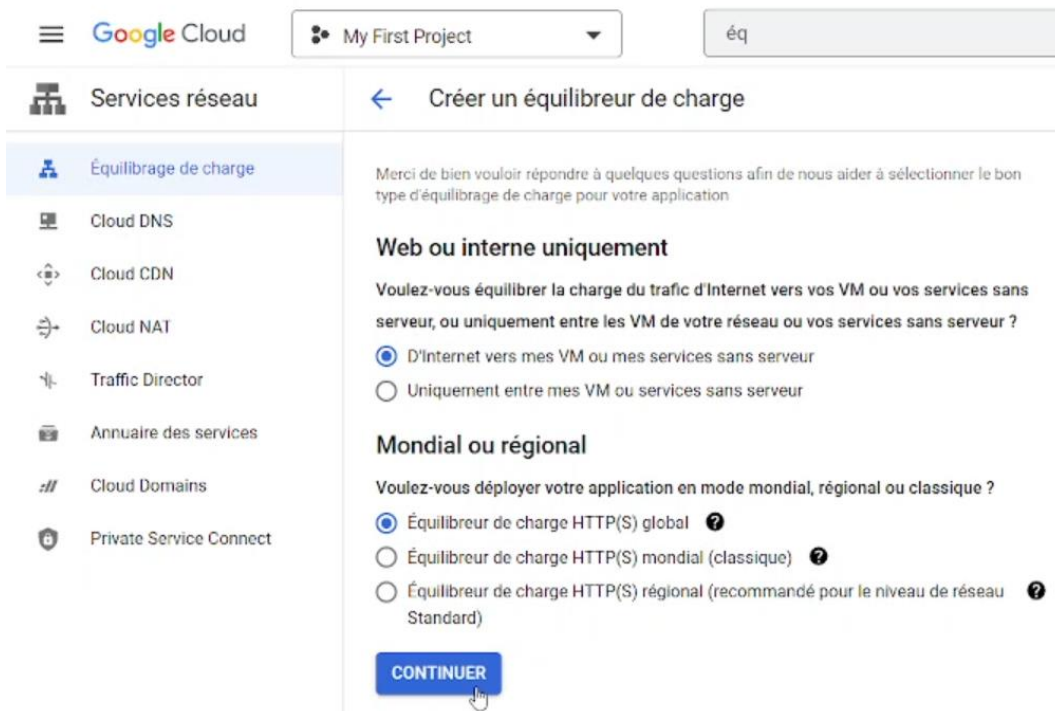
On peut se rendre dans la recherche des services et chercher l'équilibreur de charges



On choisira ensuite un équilibrage http



On sélectionnera les options suivantes



Google Cloud My First Project éq

Services réseau

- Équilibrage de charge
- Cloud DNS
- Cloud CDN
- Cloud NAT
- Traffic Director
- Annuaire des services
- Cloud Domains
- Private Service Connect

← Créer un équilibreur de charge

Merci de bien vouloir répondre à quelques questions afin de nous aider à sélectionner le bon type d'équilibrage de charge pour votre application

Web ou interne uniquement

Voulez-vous équilibrer la charge du trafic d'Internet vers vos VM ou vos services sans serveur, ou uniquement entre les VM de votre réseau ou vos services sans serveur ?

D'Internet vers mes VM ou mes services sans serveur

Uniquement entre mes VM ou services sans serveur

Mondial ou régional

Voulez-vous déployer votre application en mode mondial, régional ou classique ?

Équilibreur de charge HTTP(S) global ?

Équilibreur de charge HTTP(S) mondial (classique) ?

Équilibreur de charge HTTP(S) régional (recommandé pour le niveau de réseau Standard) ?

CONTINUER



1.3.2 Le frontend

On va créer un nouveau frontend

The screenshot shows the Google Cloud console interface for configuring a load balancer. The top navigation bar includes the Google Cloud logo, the project name 'My First Project', and a search bar. The left sidebar lists various network services, with 'Équilibrage de charge' (Load Balancing) selected. The main content area is titled 'Modifier l'équilibreur de charge HTTP(S)' and shows a list of configuration steps: 'Configuration du frontend' (checked), 'Configuration du backend' (checked), 'Règles de routage' (checked), and 'Vérifier et finaliser (facultatif)' (with an information icon). Below the list are 'METTRE À JOUR' and 'ANNULER' buttons. The right panel, 'Configuration de l'interface', provides instructions and a dropdown menu for 'Protocole : HTTP, adresse IP : Éphémère, port : 80', along with an 'AJOUTER UN IP ET PORT FRONTEND' button.



1.3.3 Backend

Puis un nouveau backend

Modifier le service de backend

Nom
back

Description

Type de backend
Groupe de points de terminaison du réseau de connectivité hybride (zonal)

Protocole
HTTP

Port nommé
http

Inactif *
30 secondes

Backends

Nouveau backend

Groupe de points de terminaison du réseau *

Mode d'équilibrage ?
Taux

Nombre maximal de RPS * RPS

Champ d'application
par point de terminaison

METTRE À JOUR ANNULER



Puis une vérification d'état

Vérification de l'état

Nom * ?
Minuscules, sans espaces.

Description

Région ?

Protocole Port * ?

Protocole de proxy

Demander ? Réponse ?

Journaux

Activé
L'activation des journaux de vérification d'état peut entraîner une augmentation des coûts dans Cloud Logging.

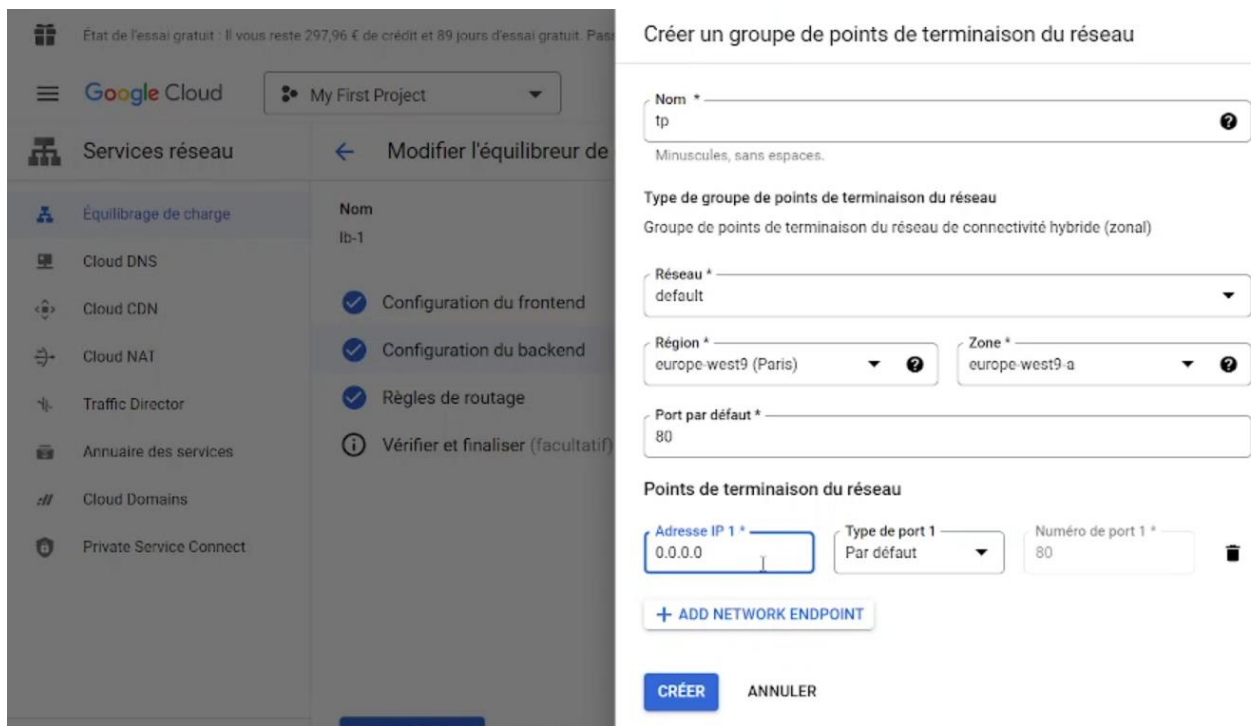
Désactivé

Critères de vérification d'état

Définir les critères qui indiquent si une instance est opérationnelle : fréquence des vérifications, délai d'attente des réponses et nombre de tentatives ayant réussi ou échoué avant l'établissement d'un résultat



Puis un point de terminaison réseau, qui sera le point sur lequel on va arriver en venant depuis le WAN

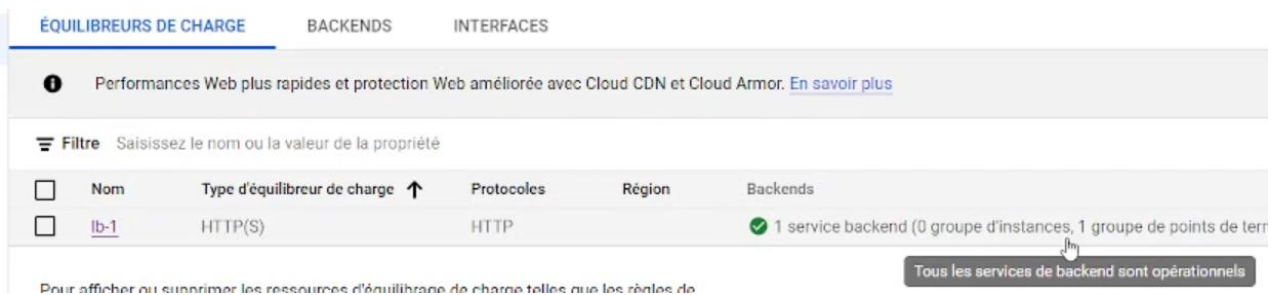


The screenshot shows the Google Cloud console interface for configuring a network endpoint group. The left sidebar shows the 'Services réseau' menu with 'Équilibrage de charge' selected. The main area is titled 'Modifier l'équilibreur de charge' and shows the configuration for 'lb-1'. The 'Configuration du frontend' and 'Configuration du backend' sections are checked. The 'Vérifier et finaliser (facultatif)' section is also visible. The right panel is titled 'Créer un groupe de points de terminaison du réseau' and contains the following fields:

- Nom *: tp
- Type de groupe de points de terminaison du réseau: Groupe de points de terminaison du réseau de connectivité hybride (zonal)
- Réseau *: default
- Région *: europe-west9 (Paris)
- Zone *: europe-west9-a
- Port par défaut *: 80
- Points de terminaison du réseau: Adresse IP 1 *: 0.0.0.0, Type de port 1: Par défaut, Numéro de port 1 *: 80

Buttons for '+ ADD NETWORK ENDPOINT', 'CRÉER', and 'ANNULER' are visible at the bottom.

On peut ensuite créer notre équilibrage et constater son bon déploiement.



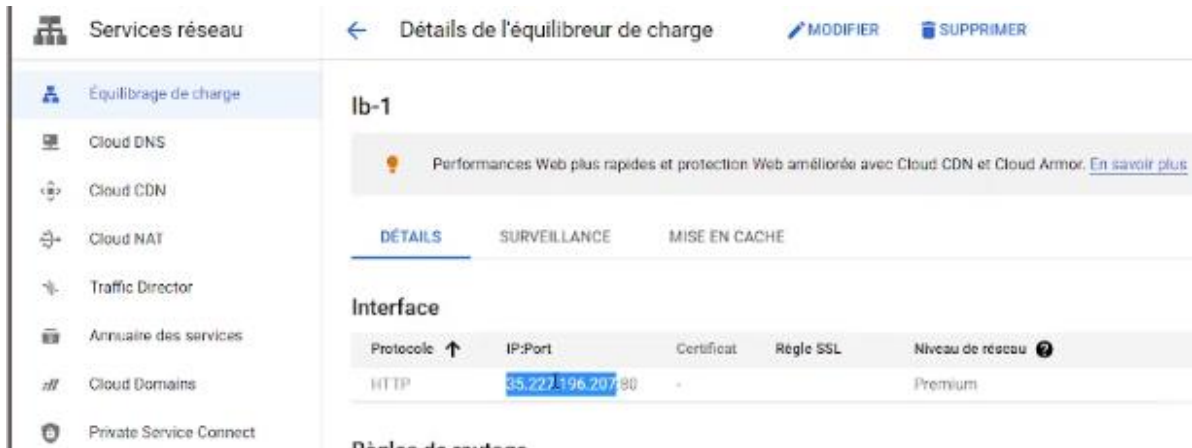
The screenshot shows the Google Cloud console interface for the 'ÉQUILIBREURS DE CHARGE' section. The 'BACKENDS' tab is selected. A notification banner at the top indicates 'Performances Web plus rapides et protection Web améliorée avec Cloud CDN et Cloud Armor'. Below the notification, there is a filter bar and a table of load balancers.

Nom	Type d'équilibreur de charge	Protocoles	Région	Backends
lb-1	HTTP(S)	HTTP		1 service backend (0 groupe d'instances, 1 groupe de points de terminaison)

A status message at the bottom right of the table indicates 'Tous les services de backend sont opérationnels'.



On dispose maintenant d'une IP externe qui sera celle de notre équilibreur



The screenshot displays the Google Cloud Platform console interface for a load balancer. On the left, a sidebar menu lists various network services, with 'Équilibrage de charge' (Load Balancing) selected. The main content area shows the 'Détails de l'équilibreur de charge' (Load Balancer Details) for 'lb-1'. At the top, there are 'MODIFIER' (Edit) and 'SUPPRIMER' (Delete) buttons. Below this, a notification banner mentions 'Performances Web plus rapides et protection Web améliorée avec Cloud CDN et Cloud Armor'. The 'DÉTAILS' tab is active, showing an 'Interface' table with the following data:

Protocole	IP:Port	Certificat	Règle SSL	Niveau de réseau
HTTP	35.227.196.207:80	-	-	Premium



1.4 Test du de l'auto-scaling

1.4.1 Les instances de VM

Afin de tester notre auto-scaling et la résistance à la charge, il est possible de supprimer des instances afin de voir si elles sont recréées automatiquement

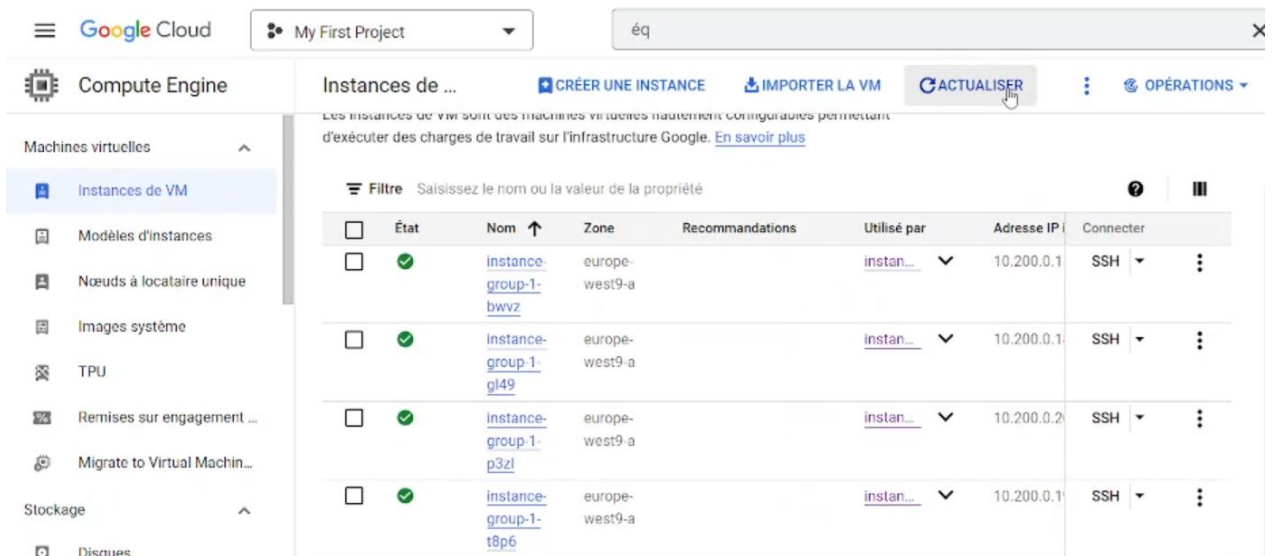
The screenshot shows the Google Cloud Console interface for a project named 'My First Project'. The 'Compute Engine' section is active, displaying a list of VM instances. The table below shows the details of three instances:

État	Nom	Zone	Recommandations	Utilisé par	Adresse IP	Connecter
✓	instance-group-1-bwvz	europa-west9-a		instan...	10.200.0.1	SSH
🔄	instance-group-1-gl49	europa-west9-a				SSH
🔄	instance-group-1-t8pé	europa-west9-a				SSH

Below the table, there are several action cards: 'Découvrir la sauvegarde et la reprise après sinistre', 'Afficher le rapport de facturation', and 'Surveiller les VM'. A 'Suppression des instances...' dialog box is overlaid on the bottom right.



Et on peut voir que même après un rechargement de l'interface de gestion de google, elles sont régénérées automatiquement.



The screenshot shows the Google Cloud Compute Engine console for a project named "My First Project". The left sidebar contains navigation options for "Machines virtuelles" (Virtual Machines), including "Instances de VM", "Modèles d'instances", "Nœuds à locataire unique", "Images système", "TPU", "Remises sur engagement...", and "Migrate to Virtual Machin...". Below this is the "Stockage" (Storage) section with "Disques" (Disks).

The main content area is titled "Instances de ..." and includes buttons for "CRÉER UNE INSTANCE", "IMPORTER LA VM", "ACTUALISER" (highlighted with a mouse cursor), and "OPÉRATIONS". Below the buttons is a descriptive text: "Les instances de VM sont des machines virtuelles hautement configurables permettant d'exécuter des charges de travail sur l'infrastructure Google. [En savoir plus](#)".

A search filter is present: "Filtre Saisissez le nom ou la valeur de la propriété". Below it is a table of VM instances:

<input type="checkbox"/>	État	Nom ↑	Zone	Recommandations	Utilisé par	Adresse IP	Connecter
<input type="checkbox"/>	✓	instance-group-1-bwvz	europa-west9-a		instan...	10.200.0.1	SSH ⌵ ⋮
<input type="checkbox"/>	✓	instance-group-1-gl49	europa-west9-a		instan...	10.200.0.1	SSH ⌵ ⋮
<input type="checkbox"/>	✓	instance-group-1-p3zl	europa-west9-a		instan...	10.200.0.2	SSH ⌵ ⋮
<input type="checkbox"/>	✓	instance-group-1-t8p6	europa-west9-a		instan...	10.200.0.1	SSH ⌵ ⋮

