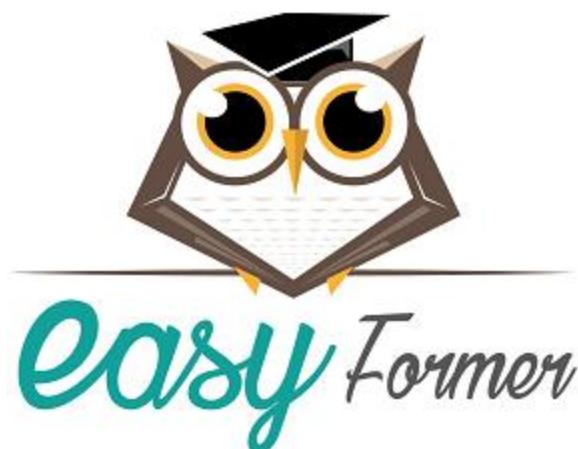


Load balancing et autoscaling GCP

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

Sommaire

page

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

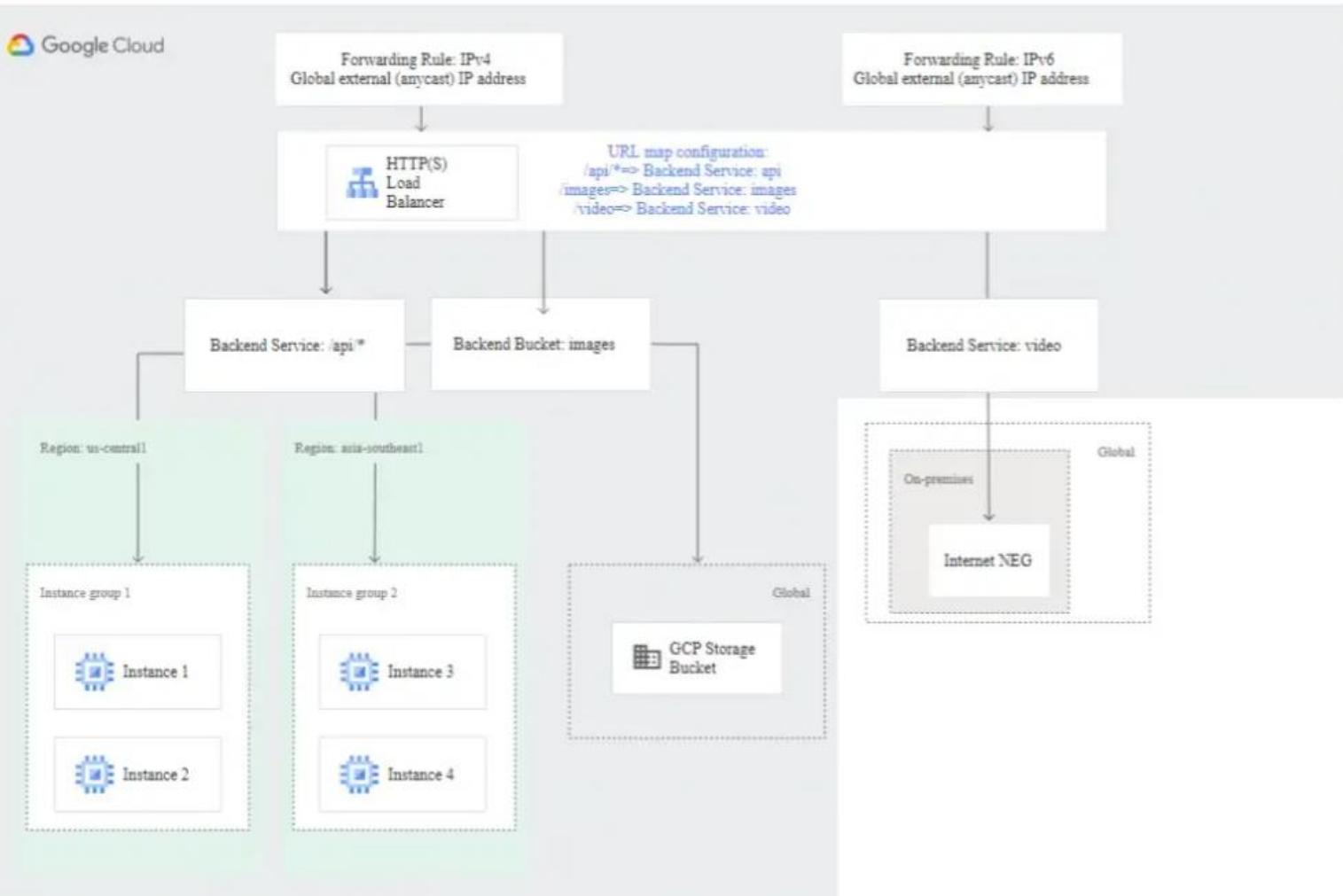


1 Introduction

Un load balancing est un équilibrer de charge, il va réparti 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 Compute Engine interface. On the left, there's a sidebar with 'Compute Engine' selected. Under 'Machines virtuelles', 'Modèles d'instances' is highlighted. The main area is titled 'Modèles d'ins...' and features a 'CRÉER UN MODÈLE D'INSTANCE' button at the top right. Below it is a search bar labeled 'Filtre Filtrer les modèles d'instances'. A table header includes columns for 'Nom', 'Type de machine', 'Image', 'Type de disque', and 'Stratégie d'emplacement'. In the center, there's a decorative graphic of a globe with colored dots (yellow, green, blue, red) around it. Below the graphic, the section title 'Modèles d'instances' is followed by a descriptive text: 'Un modèle d'instance vous permet de décrire une instance de VM. Vous pouvez ensuite créer des groupes d'instances identiques sur la base du modèle.' It also includes a 'Learn more' link and a 'ACCÉDER AU GUIDE DE DÉMARRAGE RAPIDE' button.



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 0.25 à 2 vCPU (1 cœur partagé)	Memory 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)

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

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

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



On peut ensuite créer notre groupe.

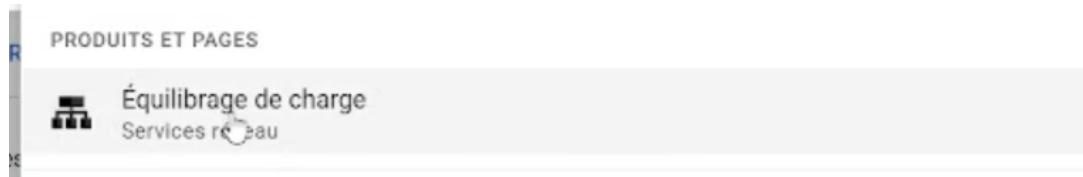
The screenshot shows the Google Cloud Compute Engine Instances page. On the left, there's a sidebar with categories like Machines virtuelles, Stockage, and Marketplace. The main area is titled "Groupes d'instances" and includes buttons for "CRÉER UN GROUPE D'INSTANCES", "ACTUALISER", and "SUPPRIMER". A search bar at the top right says "Rechercher des ressources, des documents, des produits, etc." Below the search bar is a "Filtre" section with a placeholder "Saisissez le nom ou la valeur de la propriété". A table lists an instance group named "instance-group-1" with 0 instances, created by "inst-1" on "Généré" on "nov. 16, 2022, 12:39:08 AM UTC+01:00". The table columns are: État, Nom ↑, Instances, Modèle, Type de groupe, Date/Heure de création, Recommandation, Autoscaling, and Zone. A success message at the bottom says "Le groupe d'instances 'instance-group-1' a bien été créé." with a close button.



1.3 Load balancing

1.3.1 Crédation de l'équilibrer de charge

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



On choisira ensuite un équilibrage http

[Créer un équilibrur de charge](#)

Équilibrage de charge HTTP(S)

Équilibrage de charge de niveau 7 pour les applications HTTP et HTTPS.[En savoir plus](#)

Configurer

Équilibrur de charge HTTP

Équilibrur de charge HTTPS (inclus l'équilibrur de charge HTTP/2)

Options

Web ou interne

Régional ou multirégional

[DÉMARRER LA CONFIGURATION](#)



On sélectionnera les options suivantes

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 Platform interface for managing network services. The left sidebar lists various services: Services réseau, Équilibrage de charge, Cloud DNS, Cloud CDN, Cloud NAT, Traffic Director, Annuaire des services, Cloud Domains, and Private Service Connect. The 'Équilibrage de charge' option is selected. The main panel shows a configuration for a load balancer named 'lb-1'. The 'Configuration de l'interface' section specifies 'Protocole : HTTP, adresse IP : Éphémère, port : 80'. Below this, there are four steps listed: 'Configuration du frontend' (checked), 'Configuration du backend' (checked), 'Règles de routage' (checked), and 'Vérifier et finaliser (facultatif)'. At the bottom are 'METTRE À JOUR' and 'ANNULER' buttons.



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 * verif

Minuscules, sans espaces.

Description

Région europe-west9 (Paris)

Protocole TCP Port * 80

Protocole de proxy AUCUN

Demandeur 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

ENREGISTRER ANNULER



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

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



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

The screenshot shows the 'Détails de l'équilibreur de charge' (Load Balancer details) page for a load balancer named 'lb-1'. The left sidebar lists various network services: Équilibrage de charge, Cloud DNS, Cloud CDN, Cloud NAT, Traffic Director, Annuaire des services, Cloud Domains, and Private Service Connect. The main panel displays the load balancer's interface settings. Under 'Interface', there is one entry for 'HTTP' with the IP:Port value '35.223.196.207:80'. Other tabs like 'SURVEILLANCE' and 'MISE EN CACHE' are also visible.



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 récréées automatiquement

The screenshot shows the Google Cloud Compute Engine Instances page. On the left, there's a sidebar with 'Compute Engine' selected under 'Machines virtuelles'. The main area displays a table of instances:

	État	Nom	Zone	Recommandations	Utilisé par	Adresse IP	Connexion
<input type="checkbox"/>	✓	instance-group-1-bwz	europe-west9-a	instan...	10.200.0.1	SSH	⋮
<input type="checkbox"/>	▼	instance-group-1-g49	europe-west9-a			SSH	⋮
<input type="checkbox"/>	▼	instance-group-1-t8p6	europe-west9-a			SSH	⋮

Below the table, there's a section titled 'Actions associées' with four cards:

- Découvrir la sauvegarde et la reprise après sinistre (NOUVEAU)
- Afficher le rapport de facturation
- Surveiller les VM
- Suppression des instances... (button)

A modal window titled 'Suppression des instances...' is open in the bottom right corner.



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

The screenshot shows the Google Cloud Compute Engine Instances page. The left sidebar has 'Compute Engine' selected under 'Machines virtuelles'. The main area displays a table of instances:

Etat	Nom	Zone	Utilisé par	Adresse IP	Connecter
✓	instance-group-1-bwz	europe-west9-a	instan...	10.200.0.1	SSH
✓	instance-group-1-gl49	europe-west9-a	instan...	10.200.0.1	SSH
✓	instance-group-1-p3zl	europe-west9-a	instan...	10.200.0.2	SSH
✓	instance-group-1-t8p6	europe-west9-a	instan...	10.200.0.1	SSH

