

# Résumé Module 6 - Elastic Google Cloud Infrastructure : Scaling and Automation

## Interconnexion des réseaux

### Cloud VPN

Le Cloud VPN permet de connecter de façon sécurisée notre réseau local au réseau VPC Google Cloud. C'est pratique quand on a pas énormément de données à faire passer, avec un SLA de 99,9%. Il supporte différents types de routage comme les routes statiques ou dynamiques avec Cloud Router, et utilise les protocoles IKEv1 et IKEv2 pour établir les tunnels sécurisés.

Il existe plusieurs types de VPN. Le **Classic VPN** utilise deux passerelles et deux tunnels avec une adresse IP externe régionale. Le MTU maximum est de 1460 bytes avec cette solution.

Le **HA VPN** offre une meilleure disponibilité à 99,99% en choisissant automatiquement deux adresses IP externes. Il supporte plusieurs tunnels et doit obligatoirement utiliser le routage dynamique BGP. On peut l'utiliser dans différentes topologies : vers des équipements VPN classiques, vers AWS, ou entre deux réseaux Google Cloud.

Pour le routage dynamique, Cloud Router établit une session BGP entre le VPC et la passerelle VPN locale qui doit supporter BGP. Ça permet de propager automatiquement les changements de configuration réseau.

### Cloud Interconnect et Peering

Ces solutions offrent des connexions dédiées ou partagées au réseau Google. Les connexions de couche 2 utilisent des VLAN qui donnent accès aux adresses IP internes RFC 1918, tandis que les connexions de couche 3 donnent accès aux services Google via des IP publiques.

Dedicated Interconnect fournit une connexion physique directe dans une installation de colocation Google. Partner Interconnect passe par un fournisseur de services supporté. Cross-Cloud Interconnect permet de connecter Google Cloud à un autre fournisseur cloud avec des débits de 10 ou 100 Gbps.

Pour le peering, Direct Peering connecte directement au réseau Google sans SLA, tandis que Carrier Peering passe par un partenaire supporté.

### VPC partagé et VPC Peering

Le VPC partagé permet à plusieurs projets de partager un réseau VPC commun pour communiquer avec des adresses IP internes. Le VPC Peering connecte deux réseaux VPC différents, même s'ils appartiennent à des projets ou organisations différentes, en utilisant des adresses RFC 1918.

## **Load Balancing et Autoscaling**

### **Répartiteurs de charge (Load Balancers)**

Google Cloud propose des répartiteurs globaux (HTTP(S), SSL Proxy, TCP Proxy) et régionaux (HTTP(S) régional, Internal, TCP/UDP Network). Les répartiteurs globaux utilisent une adresse IP Anycast pour diriger le trafic vers l'instance la plus proche de l'utilisateur.

### **Managed Instance Groups**

Ces groupes permettent de contrôler plusieurs VM identiques comme une seule entité en utilisant un modèle d'instance. Le gestionnaire s'assure que toutes les instances fonctionnent et peut redimensionner le groupe automatiquement. On peut les déployer dans une zone ou sur plusieurs régions.

L'autoscaling ajoute ou supprime automatiquement des instances selon la charge. Les politiques peuvent se baser sur l'utilisation CPU, la capacité du répartiteur, des métriques de monitoring, ou des plannings. Les vérifications de santé déterminent si les instances répondent correctement au trafic.

### **Load Balancer HTTP(S)**

Il offre une répartition de charge globale pour les requêtes HTTP(S) avec une adresse IP Anycast. Il supporte IPv4 et IPv6, l'autoscaling et les URL maps pour diriger le trafic selon le contenu.

Les services backend définissent les vérifications de santé (health checks), l'affinité de session optionnelle, et les paramètres de timeout. Chaque backend peut être un groupe d'instances avec un mode de répartition basé sur l'utilisation CPU ou les requêtes par seconde.

Pour HTTPS, il faut au minimum un certificat SSL installé, avec jusqu'à 15 certificats par proxy cible. Le protocole QUIC est supporté pour des connexions client plus rapides.

### **Cloud CDN**

Le réseau de diffusion de contenu utilise les points de présence globaux de Google pour mettre en cache le contenu près des utilisateurs. Trois modes de cache sont disponibles : `USE_ORIGIN_HEADERS` qui respecte les directives du serveur d'origine, `CACHE_ALL_STATIC` qui met automatiquement en cache le contenu statique, et `FORCE_CACHE_ALL` qui met tout en cache.

# Automatisation de l'infrastructure

## Terraform

Terraform utilise des templates structurés pour documenter l'infrastructure de façon lisible. Il permet un processus de déploiement répétable avec un langage déclaratif, se concentre sur l'application plutôt que l'infrastructure, et peut faire des déploiements parallèles.

L'Infrastructure as Code (IaC) permet de provisionner et supprimer rapidement l'infrastructure, créer des environnements identiques pour le développement, les tests et la production, et s'intégrer dans un pipeline CI/CD.

En dehors du code, avec Terraform on a des commandes de base que sont *terraform init* pour télécharger les plugins, *terraform plan* pour vérifier les changements prévus, et *terraform apply* pour créer l'infrastructure.

## Google Cloud Marketplace

Il permet de déployer rapidement des solutions logicielles sur Google Cloud avec une facturation unique, la gestion via Terraform, des notifications de mises à jour de sécurité, et un accès direct au support des partenaires.

## Managed Services

Ces services offrent des solutions partielles ou complètes sans avoir à gérer l'infrastructure sous-jacente.

**BigQuery** est un entrepôt de données serverless très évolutif avec une interface SQL, accessible via la console, les lignes de commande ou l'API REST.

**Dataflow** permet d'exécuter des traitements de données batch ou streaming avec autoscaling, en utilisant Apache Beam.

**Dataprep** aide à explorer et préparer visuellement les données pour l'analyse, en suggérant des transformations idéales sans écrire de code.

**Dataproc** est un service géré pour faire tourner des clusters Apache Spark et Hadoop rapidement et à faible coût, avec facturation à la seconde et instances préemptibles.

Ces services gérés sont utiles car ils peuvent remplacer la création et gestion de solutions d'infrastructure complexes, permettant de se concentrer sur la valeur métier plutôt que sur la technique.