Analyse Deployment

Team Groenpunt
Integratieproject I
2022

KdG
Karel de Grote
Hogeschool

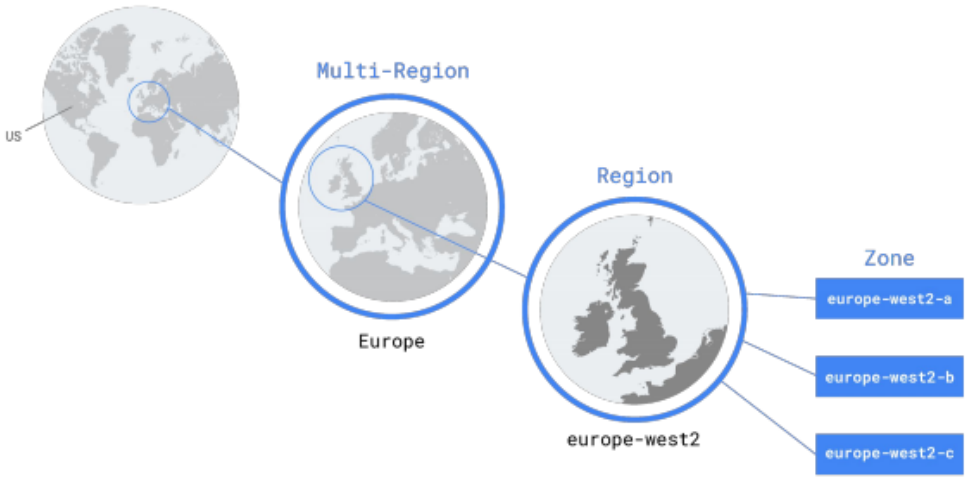
1 Onderzoek componenten

Het eerste deel van de analyse is een onderzoek naar de verschillende GCP (Google Cloud Platform) componenten die nodig zullen zijn voor dit project. Het onderzoek houdt zich uitsluitend bezig met het theoretisch aspect van deze componenten. Het al dan niet includeren van bepaalde informatie werd geanticipeerd op de benodigdheden van het project. Indien u bekend met met alle Google Cloud concepten is dit stuk overbodig.

1.1 Regions & Zones

De GCP infrastructuur bestaat uit datacenters die gesitueerd zijn over de hele wereld. De datacenters zijn onderverdeeld in verschillende regions. Een region is een geografische locatie waar resources gehost kunnen worden. Voorbeelden van regions zijn *europe-west1* en *us-central2*.

Een zone is een onderdeel van een region. Een region bestaat uit minstens drie zones. Zones binnen dezelfde region zijn verbonden met low-latency links. Voorbeelden van zones zijn *europe-west1-b* en *us-central2-c*.



1.1.1 Scope van een resource

Resources hebben een bepaalde scope. Deze scope bepaalt voor welke andere resources de resource beschikbaar is.

Scope	Uitleg
Global	Beschikbaar voor elke resource in elke zone binnen hetzelfde project.
Regional	Beschikbaar voor elke resource in dezelfde region binnen hetzelfde project.
Zonal	Beschikbaar voor elke resource in dezelfde zone binnen hetzelfde project.

1.2 Virtual private cloud

Virtual private cloud (VPC) biedt netwerk functionaliteiten aan binnen een GCP project. Het laat toe dat verschillende resources met elkaar kunnen communiceren zonder gebruik te maken van het publieke internet. Het is virtueel omdat de netwerken software gebaseerd zijn. Het zijn dus geen fysieke netwerken. Het is private omdat het intern is binnen het GCP.

1.2.1 VPC netwerk

Een VPC netwerk is een globaal netwerk binnen een GCP project. Het netwerk kan worden opgedeeld in subnetten. Elk subnet behoort tot een region. Meerdere subnetten kunnen tot dezelfde region behoren. Elk GCP project komt met een default netwerk. Dit netwerk bevat subnetten voor alle regions.

1.2.2 IP adressen

IP adressen binnen GCP hebben verschillende eigenschappen. Elke paragraaf hieronder beschrijft een eigenschap.

Een IP adres kan **private** of **public** zijn. Een private IP adres behoort tot een range van private IP adressen. deze verschillende ranges zijn algemeen vastgelegd. Publieke IP adressen zijn die adressen die niet tot een private range behoren.

Een IP adres kan **internal** of **external** zijn. Een external IP adres is bereikbaar via het publieke internet. Een internal IP adres is een IP adres binnen de VPC. Een internal IP adres moet altijd behoren tot een subnet. Het kan niet behoren tot een VPC netwerk.

Een IP adres kan **ephemeral** of **static** zijn. Een ephemeral IP adres heeft dezelfde levensduur als de resource waartoe het behoort. Het wordt dus vrijgegeven wanneer de resource waartoe het behoort wordt verwijderd. Een static IP adres is een gereserveerd IP adres dat kan toegewezen worden aan een bepaalde resource. Wanneer de resource wordt verwijderd blijft het IP adres bestaan.

1.2.3 Forwarding rules

Forwarding rules sturen internetverkeer naar een bepaalde resource binnen GCP op basis van een IP adres, een poort en een protocol. Een external forwarding rule stuurt internetverkeer van buiten het GCP sturen naar resources binnen het GCP. Een internal forwarding rule doet dit binnen in het GCP. Forwarding rules kunnen verschillende bestemmingen hebben zoals een virtuele machine of een target proxy van een load balancer.

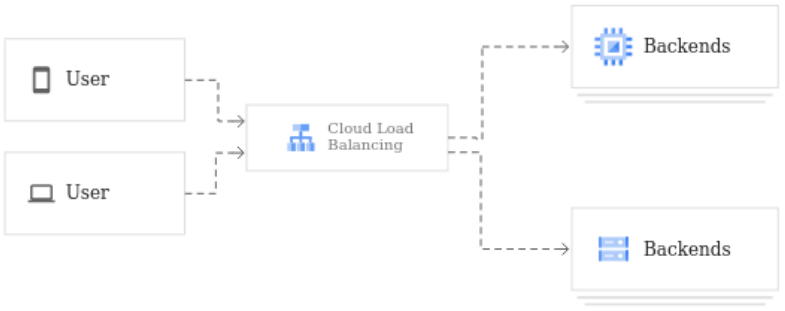
1.2.4 Firewall rules

Binnen een VPC netwerk kunnen firewall rules opgesteld worden. Ze laten toe om bepaalde connecties te accepteren of te weigeren. Firewall rules kunnen toegepast worden tussen resources en andere netwerken en ook tussen verschillende resources binnen hetzelfde netwerk. Een firewall rule kan behoren tot het netwerk of een resource in dat netwerk met behulp van een tag. Een resource kan één of meerdere tags hebben. Een tag kan tot één of meerdere resources behoren. Elke firewall rule heeft verschillende eigenschappen:

Eigenschap	**Mogelijke waarden**	**Uitleg**
Direction of connection	ingress | egress	Bepaalt of de firewall rule bestemd is voor ingaand verkeer of uitgaand verkeer. Dit wordt altijd bekeken vanuit het perspectief van de resource waaraan de firewall rule gekoppeld wordt.
Priority	0 - 65535	De prioriteit van de firewall rule. Een lager nummer betekent een hogere prioriteit. Firewall rules met een hogere prioriteit overschrijven andere firewall rules met een lagere prioriteit gedeeltelijk of volledig.
Action on match	allow | deny	Bepaalt of de firewall rule dient om de bepaalde connecties toe te laten of te weigeren.
Enforcement status	enabled | disabled	Bepaalt of de firewall rule aan of uit staat.
Target	ip adres (range) | tag | service accounts	Voor inkomend verkeer bepaalt de target parameter de destination van het verkeer. Voor uitgaand verkeer bepaalt de target parameter de source van het verkeer. De target parameter bepaalt de resource in het GCP waarvoor de firewall rule geldt.
Source	ip adres (range) | tag | service accounts	Enkel toepasbaar bij inkomend verkeer. Bepaalt vanwaar het inkomend verkeer kan komen.
Destination	ip adres (range) | tag | service accounts	Enkel toepasbaar bij uitgaand verkeer. Bepaalt naar waar het uitgaand verkeer gestuurd kan worden.
Protocol	tcp | udp | icmp | esp | ah | sctp | ipip + port	Bepaalt het protocol en poort waarvoor deze firewall rule toepasbaar is.

1.3 Cloud load balancer

Een load balancer is een service die internetverkeer verdeeld over verschillende instances van één of meerdere applicaties. De applicaties worden backend services genoemd. Load balancers helpen bij het voorkomen van problemen in verband met de performantie van de applicaties die het als backend services heeft.



1.3.1 Features

Single anycast IP address

Bij cloud load balancing kan één anycast IP adres gebruikt worden als frontend voor alle backend services over alle regions. Het any IP adres verwijst dan naar verschillende servers over de hele wereld die dan het user traffic verdelen. Het user traffic kan dan ook herleid worden naar andere regions.

Software gebaseerd

Cloud load balancing is software gebaseerd. Het is dus niet gebonden aan een fysische machine. Hierdoor zal autoscaling geen verandering van fysische machines vereisen.

Autoscaling

Cloud load balancing kan automatisch scalen naarmate het user traffic stijgt. Hiervoor dienen geen configuraties gemaakt te worden bij het aanmaken van de load balancer.

Global vs regional load

Load balancers kunnen zowel global als regional zijn. Een regional load balancer verdeeld traffic over instances in dezelfde regio. Een global load balancer verdeeld traffic over instances over de hele wereld.

Internal vs external

Een external load balancer is een load balancer die bereikbaar is via een publiek ip adres. Een internal load balancer is een load balancer binnen een VPC netwerk.

HTTP vs TCP

Een load balancer kan HTTP(S) traffic verdelen. Zo een load balancer werkt op layer 7. Een load balancer kan ook TCP traffic verdelen. Zo een load balancer werkt op layer 4.

1.3.2 Architectuur

Target proxy

Een target proxy beëindigd de HTTPS connectie van de client. Als een request binnenkomt vanuit een forwarding rule wordt de URL map aangesproken en aan de hand daarvan wordt het request doorgestuurd naar de gepaste backend service.

De request worden doorgestuurd naar een backend service vanuit een proxy-only subnet. Dit is een subnet in het VPC netwerk dat de load balancer gebruikt om connecties te maken met de backend services. Er zijn dus eigenlijk twee connecties. Eén van de originele client naar de load balancer en één van de load balancer naar de backend service.

Een target proxy kan ook voorzien worden van een SSL certificaat. Dit maakt het mogelijk om HTTPS requests te ontvangen en versturen.

URL map

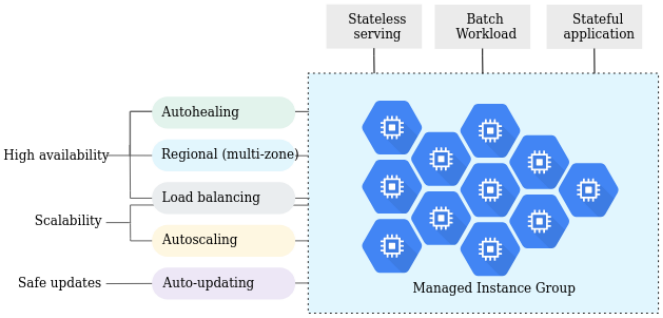
Een URL map bestaat uit verschillende regels die specificeren naar welke backend services bepaalde requests doorgestuurd moeten worden. Het is mogelijk om een default service te benoemen voor het geval dat het request geen match heeft.

Backend service

Voor elke applicatie waarvoor de load balancer gebruikt wordt moet er een backend service zijn. De backend service staat in voor de health checks van hun specifieke backend service. Een bepaalt of een backend service *ready* is om requests te ontvangen aan de hand van enkele regels.

1.4 Managed instance group

Een managed instance group (MIG) is een verzameling van identieke virtuele machines. Elke virtuele machine is opgebouwd vanuit hetzelfde instance template. MIGs bieden bepaalde features out of the box aan zoals autohealing, autoscaling en load balancing.



1.4.1 Features

Autohealing

Autohealing is het proces dat GCP gebruikt om altijd instances van een MIG's beschikbaar te hebben. Wanneer een instance stopt met draaien dan wordt er automatisch een nieuwe opgestart. Dit is autohealing op basis van de status van de virtuele machine.

Er bestaat ook autohealing op basis van de status van de applicatie die op de virtuele machine draait. Dit heet **health checking**. Wanneer de applicatie bijvoorbeeld crasht, dan wordt er een nieuwe instance aangemaakt met een werkende applicatie.

Regional or zonal groups

Er zijn twee types van MIGs. Een zonal MIG heeft instances die slechts draaien in één bepaalde zone. Regional MIGs hebben instances draaien over meerdere zones binnen dezelfde region. Regional MIGs hebben het voordeel een hogere availability te hebben doordat de workload over meerdere zones verdeeld wordt. Regionals MIGs zijn ook beschermd tegen zonal failures.

Load Balancing

Een GCP Load Balancer kan internetverkeer verdelen over de verschillende instances van een MIG. Die MIG dient dan als backend service van de load balancer.

Autoscaling

MIGs bieden out of the box autoscaling aan. Instances van een MIG kunnen toegevoegd of verwijderd worden afhankelijk van een policy. De autoscaling policy bevat signalen die gebaseerd zijn op cpu verbruik en de load balancing capaciteit.

1.4.2 Instance Templates

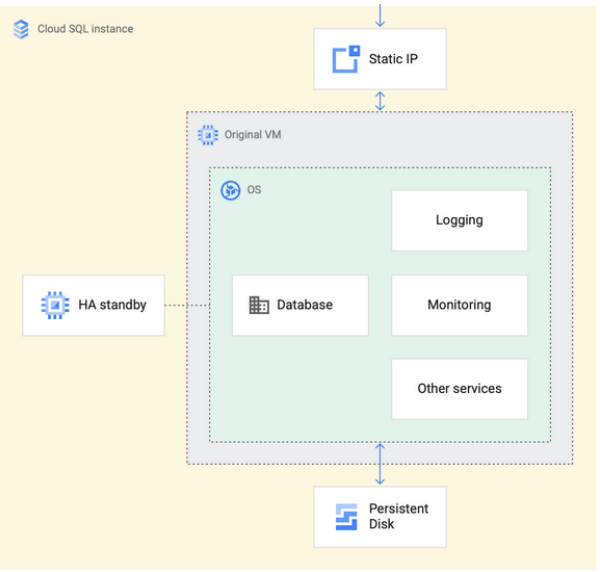
Een instance template is een blueprint waaruit een virtuele machine is opgebouwd. Een MIG gebruikt een instance template om de meerdere identieke virtuele machines te maken. Een instance template is een global resource, tenzij het gebruikt maakt van een zonal of region resource en zal dan dus gebonden zijn aan die zone of region. Een instance template definieert het machine type, het OS, labels en een startup script.

1.4.3 Named port

Een named port is een key-value pair die de naam en nummer van een poort beschrijft. De naam representeert de naam van de service waarvoor de named poort dient. Het nummer van de poort is de poort waarop de applicatie op de backend service draait.

1.5 Cloud SQL

Cloud SQL is een service die relationele databases aanbiedt in de cloud. Het helpt met het onderhouden, managen en administreren van de databases. Cloud SQL biedt drie verschillende databases aan: MySQL, PostgreSQL en SQL Server. Voor dit project zal MySQL gebruikt worden.



1.5.1 Architectuur

Primary instance

De primary instance is de werkelijke virtuele machine waarop de databank draait. Op de virtuele machine draait de databank software en andere service agents die functionaliteiten zoals logging en monitoring voorzien.

Persistance disk

Een harde schijf die verbonden is aan de virtuele machine. Op deze harde schijf wordt de werkelijke data opgeslagen.

Standby instance

Cloud SQL voorziet een high availability option bij het aanmaken van een databank die er voor zorgt dat er altijd een standby instance is van de databank. Deze standby instance bevindt zich in een andere zone binnen dezelfde region. De configuratie van de standby instance is identiek aan die van de primary instance.

1.5.2 Connectiviteit en access control

Om te connecteren naar een cloud SQL instance moet aan drie zaken gedacht worden. Ten eerste naar de connectiviteit: op welke manier wordt er met de instance een connectie opgesteld. Ten tweede autorisatie: welke applicaties mogen connecteren naar de instance. En ten derde authenticatie: op welke manier wordt er ingelogd op een databank binnen die instance.

Connectivity

Connecteren naar een cloud sql instance gebeurt via een IP adres. Dit IP adres kan een intern, private IP adres zijn binnen een VPC netwerk. Deze manier is het best om te connecteren vanuit andere resources binnen GCP zoals Compute Engine en Cloud Run omdat de connectie dan veiliger is en ook sneller. Google cloud voorziet een een service genaamd private service access. De private service acces kan een range alloceren van het VPC netwerk waar bepaalde resources gehost kunnen worden met een private IP.

Het IP adres om te connecteren naar een Cloud SQL instance kan ook een publiek IP adres zijn. Dit wil zeggen dat de connectie loopt via het publieke internet. Aangezien een publiek IP adres veel risico's met zich meebrengt qua veiligheid, moet er dan altijd geautoriseerd worden via een cloud SQL auth proxy of via geautoriseerde netwerken, dus niet met een SSL certificaat.

Authorization

Eens er een connectie gemaakt is met een cloud SQL instance moet er ook geautoriseerd worden. Dit zorgt ervoor dat niet zomaar elke applicatie kan verbinden met de instance. Autoriseren kan op drie verschillende manieren.

Ten eerste kan autoriseren via een Cloud SQL Auth proxy. Een auth proxy is een aparte instantie die instaat voor het autoriseren. Deze autorisatie gebeurt met behulp van een IAM account.

De tweede manier om te autoriseren is via een SSL certificaat. Deze zorgen ervoor dat de data geëncrypteerd wordt. Wanneer er geen gebruik gemaakt wordt van een SQL auth proxy is het zeer aangewezen om gebruik te maken van SSL.

De laatste manier om te autoriseren is via een geautoriseerd netwerk. Er wordt een lijst bijgehouden van welke netwerken allemaal mogen connecteren met de database. Alle IP adressen in de range van dat netwerk hebben dan toegang tot de database.

Authentication

Ten slotte moet er authenticatie geleverd worden bij het connecteren met een databank is een cloud SQL instance. Deze is afhankelijke van de gekozen relationele database en geeft toegang tot een specifieke database in de cloud SQL instance. Meestal is dit aan de hand van een gebruiker met paswoord.

1.6 Redis memorystore

Redis memorystore is een key value database die voornamelijk gebruikt wordt om data te cachen voor een bepaalde korte duur. Een memory store kan ook dienen als een gemeenschappelijk geheugen over verschillende virtuele machines heen.

1.6.1 Connectiviteit

Redis memorystore kan enkel intern opgezet worden binnen het GCP. Dat wil zeggen dat het geen publiek IP adres kan hebben. De enige manier om te verbinden is dus intern via een VPC netwerk. Dit kan op twee manieren gebeuren.

Direct Peering

Bij direct peering wordt er een VPC peering gemaakt tussen het VPC netwerk van de gebruiker en het VPC netwerk van Google. Dit is de default optie.

Private Service Access

Bij private service acces kan een Redis instantie opgezet worden in een Google service netwerk. Er kan dan door resources binnen hetzelfde netwerk geconnecteerd worden via een private access point.

1.6.2 Autorisatie

GCP biedt ook de mogelijkheid om autorisatie toe te voegen aan een Redis memorystore instance. Dit gebeurt door middel van een authenticatie string. Dit is een string gegenereed door de Redis instance en moet meegegeven worden wanneer een resource wil connecteren met de Redis instantie.

1.7 Cloud storage

Cloud storage is een service die het toelaat om objecten op te slaan in de cloud. Een object is een immutable file van eender welk formaat. De logische ruimte waarin objecten opgeslagen kunnen worden noemt een bucket. Een bucket behoort tot een GCP project.

1.7.1 Terminologie

Bucket

Een bucket is een container waarin object worden opgeslagen. Elk object moet in een bucket zitten. Een bucket kan je organiseren met behulp van folder en acces control. Buckets kunnen geen onderdeel zijn van een andere bucket.

Bucket labels zijn key-value pairs die informatie bijhouden over een bucket. Op die manier kan je buckets groeperen met elkaar of met andere resources zoals virtuele machines of persistent disks.

Object

Een object is een individueel stuk data dat als één geheel gezien kan worden. Bijvoorbeeld een pdf of een afbeelding. Een object moet behoren tot een bucket en er kunnen oneindig veel objects in een bucket.

Een object bestaat uit twee componenten: object data en object metadata. Object data is de werkelijke data die je wilt storen. Object metadata is een collectie van key-value pairs die informatie bijhoudt over het object.

Object names

De naam van een object is een onderdeel van de object metadata. De objectnaam moet uniek zijn en mag een maximum lengte hebben van 1024 bytes. Een bucket bekijkt objecten is een *flat namespace*. Dat wil zeggen dat er geen hiërarchie in zit zoals folders. Maar door een slash in de naam te gebruiken kunnen folder wel volgens een bepaalde structuur voorgesteld worden.

1.7.2 Autorisatie

Autorisatie op een bucket gebeurt met IAM (Identity and Access Management). Wanneer een resource een connectie wil maken met een bucket, dan moet er voor die resource een service account worden aangemaakt. Dit service account kan dan aan de bucket worden toegevoegd zodat die resource autorisatie rechten heeft. Het is ook mogelijk om bepaalde data publiek te stellen in een bucket. Deze data is dan voor iedereen op het publieke internet beschikbaar.

Uniform autorisatie

Bij een *uniform* autorisatie wordt het niveau van de autorisatie op het level van de bucket geplaatst. Dit betekent dus dat een resource zich moet autoriseren voor de hele bucket. Dit is de aangeraden manier om buckets te beveiligen.

Fine-Grained autorisatie

Fine-grained access laat toe om autorisatie te bepalen per object. Deze manier is eerder aangeraden indien dit specifiek nodig is voor een project.

1.7.3 Connectiviteit

Connecteren met een gebeurt op basis van de naam van de bucket. Elke bucket heeft een identieke naam. Er zijn verschillende mogelijkheden om te connecteren met een bucket.

1. Google Cloud Console

   Via de Google Cloud Console kan je objecten uploaden en downloaden naar of van een bucket. De autorisatie gebeurt hier automatisch indien het Google account toegang heeft tot de bucket.

2. Command line

   Met behulp van de gsutil command line tool is het ook mogelijk om objecten te uploaden of downloaden naar en van een bucket. Indien de gebruiker in de gcloud configuratie toegang heeft tot de bucket, is er aan de autorisatie voldaan.

3. Client libraries

   Er bestaan libraries in een divers aanbod van programmeertalen om te connecteren met een bucket. Autorisatie gebeurt ook met behulp van de library. Applicaties die draaien op een resource in het GCP hebben automatisch toegang tot de bucket indien de resource een service account heeft dat geautoriseerd is voor de bucket.

4. REST APIs

   Het is ook mogelijk om objecten te downloaden en uploaden van en naar een bucket door middel van HTTP requests. Elke bucket heeft bepaalde endpoints die dienen om specifieke opdrachten uit te voeren op de bucket.

2 Uitwerking architectuur

Ter voorbereiding op het deployen van het project is er een schema gemaakt waarop alle benodigde componenten en de connecties tussen die componenten te zien zijn. Op deze link is het schema te vinden in pdf formaat. Daarnaast is het schema ook toegevoegd op de laatste bladzijde van die verslag. Hieronder worden de componenten en hun connecties één voor één uitgelegd.

2.1 Region

Alle resources zullen gehost worden in dezelfde region, namelijk europe-west1. De datacenters van deze region bevinden zich in Saint Ghislain in België. De applicatie zal bijna uitsluitend gebruikt worden in België. De region beschikt over drie zones. Resources met een high availability optie kunnen dus verdeeld worden over meerdere zones.

2.2 VPC Netwerk

Er zal gebruik gemaakt worden van één VPC netwerk waarin alle resources gehost zullen worden. In het netwerk zullen twee subnetten zijn. Het eerste subnet is een proxy-only subnet dat gebruikt zal worden door de load balancer om connecties te maken met de backend services. Het tweede subnet is een standaard subnet dat door de managed instance group gebruikt wordt om de virtuele machines te hosten. Daarnaast wordt een een een ip range gealloceerd voor de private access service. De database en de memorystore zullen daaruit een IP adres krijgen. De private service access range is te bereiken via een private connection point.

2.3 Forwarding rule

Er zal gebruik gemaakt worden van een external forwarding rule om de applicatie via het publieke internet te bereiken. Het IP adres in de forwarding rule zal een external public reserved IP adres zijn. Dit IP adres zal gebruikt moeten worden in de DNS. De forwarding rule stuurt inkomende connecties door naar de target proxy van de load balancer.

2.4 Load balancer

Om het netwerkverkeer van de applicatie te verdelen over de verschillende instanties waar de applicatie op draait zal een external regional HTTPS load balancer gebruikt worden. Aangezien er gebruikt gemaakt zal worden van HTTPS moet de load balancer beschikken over een SSL certificaat. Er is slechts één backend service, namelijk de managed instance group waar de applicatie op draait. Alle connecties worden aan de hand van de named poort doorgestuurd naar de managed instance group.

2.5 Managed instance group (MIG)

2.6 Om de dotnet applicatie te draaien zal gebruik gemaakt worden van een managed instance group. Alle virtuele machines zijn gebaseerd op hetzelfde instance template dat de dotnet applicatie opstart. Alle connecties komen vanuit het proxy-only subnet van de load balancer door middel van de named poort.

2.7 MySQL database

Om data te persisteren zal er een MySQL database gebruikt worden. De database instantie kan eventueel ondersteund worden door een high availability instantie.

Om te connecteren met de database zal een private IP gebruikt worden. De database zal een IP adres krijgen uit de gereserveerde range van IP adressen voor de private service access. Op die manier kan de dotnet applicatie makkelijk verbinding maken met de database. De connectie zal dan ook intern zijn en dus niet via het publieke internet gaan.

2.8 Redis memorystore

Om bepaalde gegevens bij te houden over de verschillende virtuele machines zal een Redis memorystore gebruikt worden. Op die manier moeten inkomende connecties niet telkens naar dezelfde virtuele machine gestuurd worden. De virtuele machines worden zo helemaal stateless.

Om te connecteren met de memorystore zal ook gebruik gemaakt worden van een private IP adres in de private service acces. Dit is net hetzelfde als de database.

2.9 Cloud storage

Om bepaalde objecten zoals afbeeldingen op te slaan zal er gebruik gemaakt worden van één of meerdere buckets. De buckets zullen ten eerste bereikbaar zijn door de virtuele machines aan de hand van een service account. Daarnaast kunnen ook objecten gedownload worden via publieke URL’s. De publieke URL’s dienen niet om objecten toe te voegen.

3 Risico analyse

Hieronder volgt een lijst met bepaalde risico's die de deployment inhoudt. Alle componenten zijn belangrijk en noodzakelijk om aan de vereisten van het project toe voldoen. De risico's duiden aan waar er het meeste fout kan gaan. Ze zijn geordend van hoog naar laag. De bovenste hebben dus het meeste risico.

Connectie naar de MIG

Als de load balancer geen connecties kan sturen naar de MIG is er eigenlijk helemaal geen applicatie. Dit is de meeste essentiële connectie binnen het VPC netwerk.

Connectie naar de database

De connectie naar de database is noodzakelijk om data te kunnen persisteren. Als er geen data kan gepersisteerd worden is de applicatie zo goed als waardeloos.

Connectie naar de memorystore

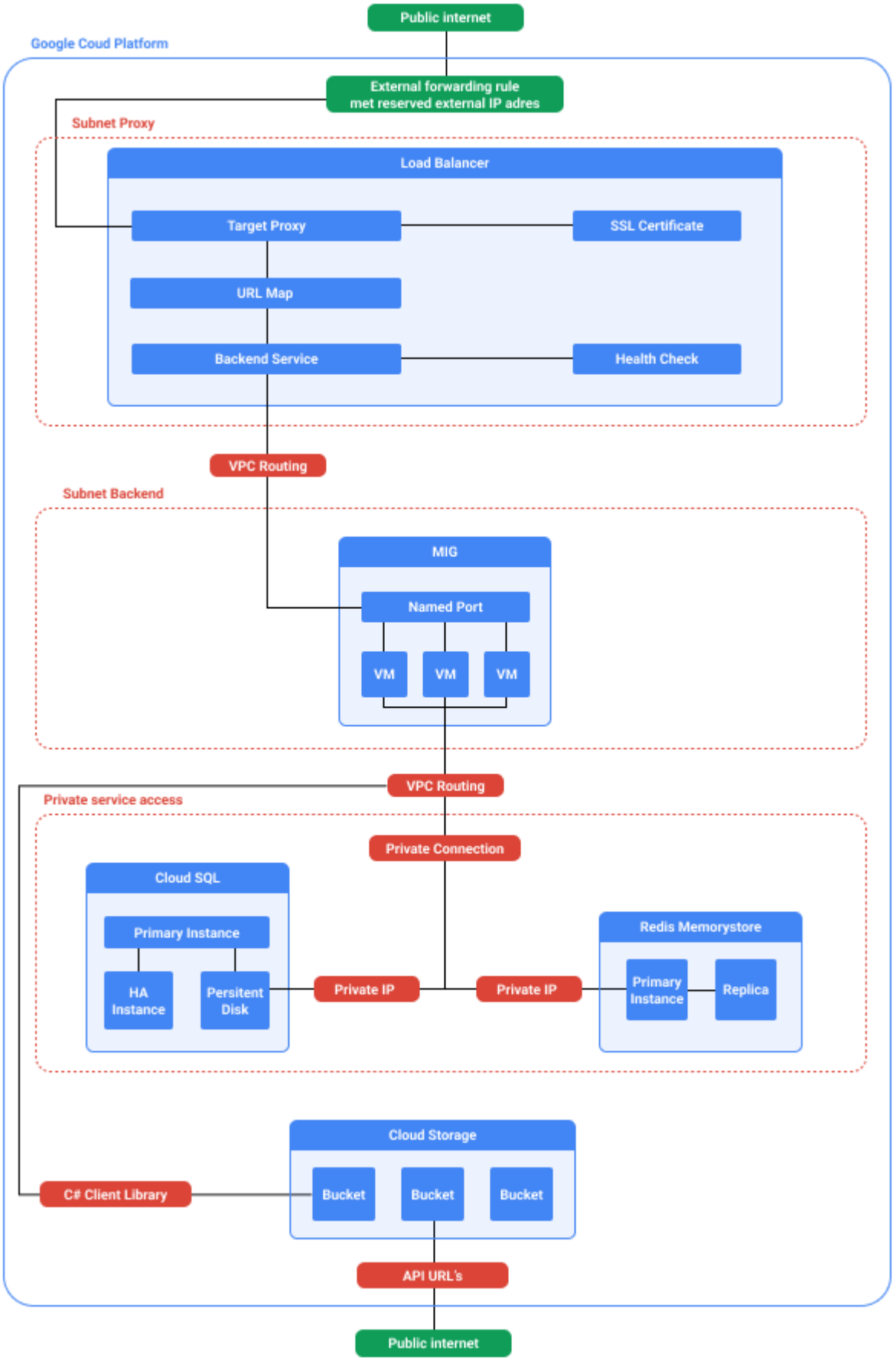
De connectie naar de memorystore is zeer belangrijk omdat het een shared memory is tussen de verschillende virtuele machines van de managed instance group. De memorystore zorgt ervoor dat clients niet telkens naar dezelfde virtuele machine gestuurd moeten worden. Dit geeft ook vrijheid in het downscalen van de managed instance group.

Opzetten van private service acces

Private service access laat toe om interne connecties tussen resources op te zetten. Dit geeft een enorme boost aan de veiligheid van het verkeer tussen de virtuele machines en de database en memorystore.

Connectie naar de cloud storage

Zonder cloud storage is het niet mogelijk om afbeeldingen te uploaden of te downloaden. Dit is wel een zeer belangrijke vereiste van het project.


Public internet
Google Coud Platform
External forwarding rule
met reserved external IP adres
Subnet Proxy
Load Balancer
Target Proxy
SSL Certificate
URL Map
Backend Service
Health Check
VPC Routing
Subnet Backend
MIG
Named Port
VM
VM
VM
VPC Routing
Private service access
Private Connection
Cloud SQL
Primary Instance
HA Instance
Persitent Disk
Private IP
Private IP
Redis Memorystore
Primary Instance
Replica
Cloud Storage
C# Client Library
Bucket
Bucket
Bucket
API URL's
Public internet