GCP

Mo D Jabeen

September 20, 2022

1 General

GCP is Google cloud platform, which includes a variety of products from VM instance creation, DB hosting servers and container orchestration.

2 VMs (Compute Engine)

VMs can run standard linux or windows OS. Private images can also be used or containers which are uploaded onto the container optimized image.

Each instance belongs to a **console project** which can have one or more instances. The zone and OS and machine type is specified for each instance.

Each instance has a default boot persistent disk of 10 GB, any additional storage options can be added. Each instance also has a network interface associated with a unique VPC network.

Tools to manage the VM: Google cloud console, gcloud CML tool, and REST API. Manage Access: OS login or SSH keys.

2.0.1 What are instance groups ?

A collection of VM instances you can manage as a single entity.
Two types:
Managed (MIG): Operate apps on multiple identical VMs, make VMs scalable and highly available by using automated managed services, autoscaling, autohealing, deployment and updating (PaaS).
Unmanged (UIG): Only let you load balance across a fleet otherwise managed by you (IaaS).

2.1 Managed Instance Groups

Maintained based on config specified in an instance template and optional stateful config.
Pros:

- High Availability (Keep VMs running despite crashes)
- Autohealing unexpected VMs states
- Regional coverage (switch zones in the same region)
- Load balance within the group (traffic or compute)

- Autoupdates: Safely deploy new software versions. Flexible scenarios available ie rolling updates canary udpates, speed, scope and disruption of services can be tuned.
- **stateful MIGs** preserve each instances unique state on machine restart and update.
- Load balance health check: divert traffic from non healthy VMs to healthy ones.
- MIG health checks will delete and recreate unhealthy instances.

The majority of scenarios checks for load balance healing should be more aggressive whereas the MIG checks should be more conservative.

Two types of MIGs zonal or regional, to use regional MIG autoscaling need Pub/Sub (Queue based workload).
Preemtible instances can be used of speed of execution is less important than costs.
GCP will auto use a default VPC is not configured.

2.2 Machine Families

Broken into series and then machines types ie general purpose family, series: NZ and type: n2-standard-4.

2.2.1 What are the families ?

- General purpose: Best performance price ratio (x86 or Arm architecture)
- Compute optimized: Highest performance per core, optimized for compute intensive workloads. (Intel scalable processor or AMD EPYC Milan platform)
- Memory optimised machine: Ideal for OLAP and OLTP SAP (on line processing) workloads, genomic modelling and memory intensive workloads.
- Accelerator optimised: Massively parallelized compute unified architecture ML or HPC. Allow GPU usage.

2.2.2 General purpose machines

Series:

- E2 - Cost optimized
- N2,N2D,N1 - Balance price/performance
- T2D, T2A - Scale out optimized

E2 does not support GPUs, local SSD, sole tenant nodes (allocated physical server) or nested virtualization. N2 has higher memory to core ratios.

2.2.3 Other

Rightsizing recommendations: used to optimise based on workload.
CPU bursting: short periods of increased CPU coverage when required.
Nested VMs require special hypervisor instructions to allow running additional VMs.

2.3 SSH connections

Key based authentication, passwords are not configured by default. SSH key and username need to be saved to metadata before connection google console and gcloud CLI do this automatically. Either custom project or instance meta data is used.
An extra layer of security can be added by using guest attributes.

2.4 Web protection

Methods:

- Firewalls
- HTTPS and SSL
- Port forwarding over SSH
- SOCKS proxy over SSH (create local names for each server)

Access to VMs without external IPs:

- Other VMs on the network or Bastion host
- Proxy TCP forward
- Metadata server
- GCP SDK
- VPN Gateway

2.5 Transfer files to VM

Cloud buckets, SSH in browser or gcloud has built in SCP.

2.5.1 Cloud storage buckets

Easy to access from multiple, upload any data object affordably.

2.6 Storage options

Can choose between zonal or regional.

- Standard Persistent disk (PD): Efficient reliable block storage
- Balanced PD: Cost effective and reliable block storage
- SSD PD: Fast and reliable block storage
- Extreme PD: Highest performance persistent block storage
- Local SSD: High performance local block storage
- Buckets: Affordable object storage

2.7 Data protection options

- Standard snapshots: Capture the state of the disk, good for long term storage.
- Archive snapshots: Lower cost than standard snapshots, rarely accessed.
- Machine images: Stores all config, meta data, permissions and data from one or more disks. can be used to create more VMs.
- Regional persistent disks: Replicate data synchronously across two zones.
- Disk clone: Create a live attachable, fully provisioned disk with data from source disk allowing you to stage environments, backup verification, export net disaster recovery.
- Images: Contain the set of programs and files required to boot OS on a VM instance.

2.8 Other

Start up and shutdown scripts allow commands to be executed at those VM stages.

Compute engine uses Network Time Protocol (NTP). Smears the leap second over 24 hours.

Vitrio Random number generator is used in GCP.

3 Cloud Storage

Create single purpose, standalone, functions that respond to cloud events without needing to manage a server (serverless). Great for data collection into Big Query.

4 Anthos

Build, deploy and optimize VMs (CI/CD).

5 Cloud run

Fully managed container orchestration, for partial managed use Google Kubernetes.