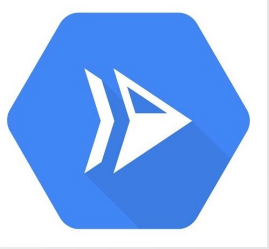
Google Cloud Run : Serverless

Presented By :

Amrit Choudhary

GCR



Serverless containers

It is fully managed serverless platform for developing and running web applications using docker containers.

Container image with a web server, combination of memory/cpu with concurrency are the only requirement, rest is taken care by cloud run like creating http endpoint, handling and routing requests, managing containers.

Cloud run is very easy compared to running docker in docker swarm or kubernetes.

No architectural restrictions like GCF or GAE.

There are actually two options when using Google Cloud Run:

a) fully managed environment

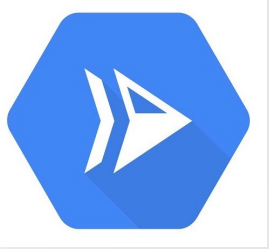
b) Google Kubernetes Engine (GKE) cluster.

GCR

Merits:

- ✓ It's going to make lives simpler for people running containerized, stateless web applications to get server less benefits.
- ✓ Easy scaling, fully managed, flexible and fine grained billing
- ✓ Extremely simple testing as everything is inside the container and we just need to use the same container everywhere we deploy.
- ✓ Portability : containers can be run anywhere , no such dependency to run only on cloud run, kubernetes, GKE etc.
- ✓ It can serve http requests only + push of pubsub events + processing of storage events.

GCR



De-Merits:

- ✓ **Beta Phase :** Things may change until the final version is rolled out.
- ✓ Applications need to run on 8080 port only.
- ✓ Not exactly Server less model as still significantly involves more code ownership.

GCR

Note : google container run time contract

- The container must be compiled for Linux 64-bit, but it can use any programming language or base image of your choice.
- The container must listen for HTTP requests on the IP address 0.0.0.0, on the port defined by the PORT environment variable (almost always 8080).
- The container instance must start an HTTP server within 4 minutes of receiving the HTTP request.
- The container's file system is an in-memory, writable file system. Any data written to the file system will not persist after the container has stopped.

Resource allocation of container :

- **CPU:** 1 vCPU (virtual CPU) for each container instance. However, the instance may run on multiple cores at the same time.
- **Memory:** By default, each container instance has 256 MB of memory. Google says this can be increased up to a maximum of 2 GB.

The free monthly quotas for Google Cloud Run are as follows:

- **CPU:** The first 180,000 vCPU-seconds
- **Memory:** The first 360,000 GB-seconds
- **Requests:** The first 2 million requests
- **Networking:** The first 1 GB egress traffic (platform-wide)
- Cloud Run on GKE uses a separate pricing model that will be announced before the service reaches general availability.