GCP: Cloud Run vs Cloud Function versus GKE

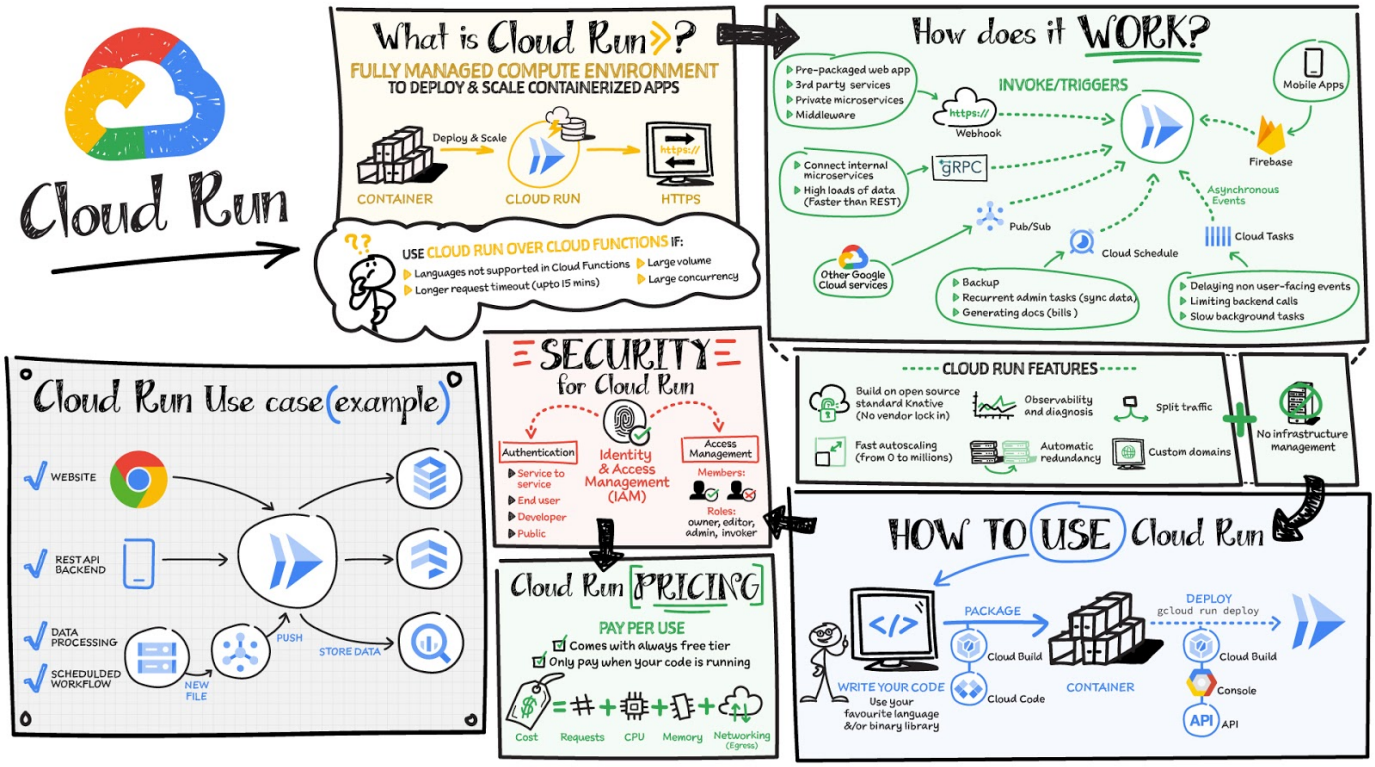
Epic	TBD
Document Type	Software Engineering White-paper
Audience	Evie Developers
Document status	DRAFT
Document owner	@ Michael Naughton @ wajahat.syed
Designer	@ Michael Naughton

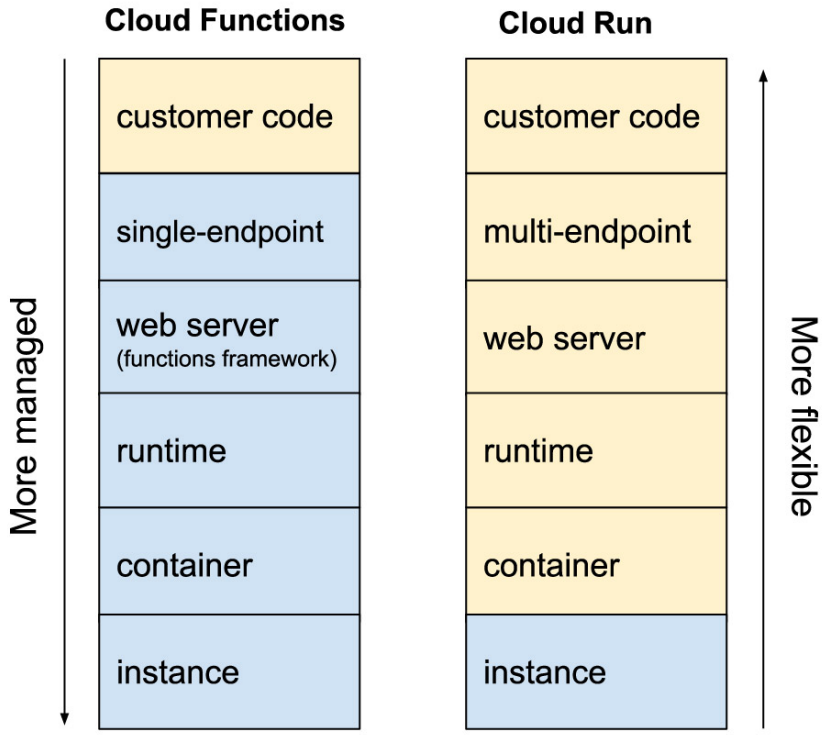
Background

In GCP we have a number of options on how to run a docker container.

https://www.youtube.com/watch?v=jh0fPT-AWwM

When to use





Deployment	What is it?	When To Use	Drawbacks	Reference
Cloud Run	A single, fully managed, stateless container which is built from source externally (i.e Docker Hub, Cloud Build) and deployed within Cloud Run. The single container is the core of the Cloud run instance architecture. Can be invoked from HTTP, gRPC, websockets, Eventarc, Pub/Sub, Cloud Scheduler, and Cloud Tasks. Can scale from 0 N instances Derives stability based on externally built versions	• For serving multiple interfaces when a longer period of time is required • Good at asynchronous, event-driven applications and microservices • Great for API's • No downtime • Simple configuration • Highly scalable • High portability • Can use any programming language • Integration with Cloud Monitoring, Cloud Logging, Cloud Trace, and Error Reporting • Only pay when code is running	• Need to build code to docker container. This step can be automated with CI /CD in future • Low cost option for light workloads but expensive as application scales up. • Questionable security due to high abstraction	https://cloud.google.com/run/docs/overview/what-is-cloud-run

Cloud Functions	A stateless, event driven application which is good for processing short lived actions, triggered from other Google services such as Cloud Storage, PubSub or EventArc. Typically Cloud Functions are used as middleware/data pipelines	• Fully managed. No need to package code as docker container • Several small units of work that make up a business service • Real time processing • Event driven architecture • Little to no configuration needed • Data pipeline automation • process short-lived, event-based actions triggered from other systems such as Cloud Storage, Eventarc, or PubSub	• Downtime between deployments • Low cost option for light workloads but expensive as application scales up.	https://cloud.google.com/functions/docs/concepts/overview
Google Kubernetes Engine	Can run in 2 configurations: 1. Autopilot (managed cluster) 2. Standard GKE	• GKE gives you complete control over every aspect of container orchestration, from networking, to storage, to how you set up observability—in addition to supporting stateful application use cases. • Good for long running jobs • Good security options due to high customisability (Cloud Armour)	• Need to deal with complexity of kubernetes. • Does not manage builds. • Additional Operational overhead. • Manage kubernetes version upgrades. • Expensive in the start but saves cost later when application scales up	https://cloud.google.com/kubernetes-engine/docs/concepts/kubernetes-engine-overview

How to use

Deploying to Cloud Run

```
REGION=us-central1

gcloud run deploy $SERVICE_NAME \
  --image gcr.io/$PROJECT_ID/$SERVICE_NAME \
  --allow-unauthenticated \
  --region $REGION
```

… our devops pipeline has this stage to deploy to GKE(under construction) and this stage

Conclusion

Generally we use Cloud Run for most loads - we currently manage the build process through DockerHub.

Cloud functions are generally used as a service boundary I/O or small utility functions.

GKE is used for stateful, large scale applications - but as for March 2023, we do not use GKE at all due to the complexity overhead.