NO→LATENCY

Cloud Run Deployments

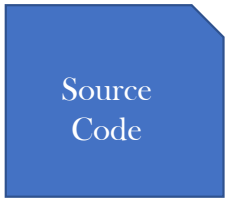
Step1 →

{docker push...}

Google Container Registry

{gcloud run deploy from image}

Step2 →



{gcloud run deploy from source}

Cloud Run Deployment - Parameters

-- **concurrency** : maximum number of concurrent requests allowed per container instance.
Default = 80 requests , max = 1000

-- **cpu**: CPU (Values : 1,2 and 4)

-- **min-instances** : minimum number of instances that will be spun up

-- **max-instances** : maximum number of instances that will be spun up

--**platform** : Possible values : managed , gke and kubernetes

-- **timeout** : Timeout in seconds

Cloud Run Deployment – Important Points

- Under normal circumstances, cloud run apps scale out by creating new instances to handle incoming traffic load.
But when you set a maximum instances limit, in some scenarios there will be insufficient instances to
meet that traffic load. In that case, incoming requests can be queued for up to 10 seconds.
During this time window, if an instance finishes processing requests, it becomes available to process queued requests.
If no instances become available during the window, the request fails with a 429 error code.

- Cloud Run does not immediately shut down instances once they have handled all requests.
To minimize the impact of cold starts, Cloud Run may keep some instances idle for a maximum of 15 minutes.
These instances are ready to handle requests in case of a sudden traffic spike.

Cloud Build

- Cloud build is a service that is used to automate the application build and deployment process
- It's a fully managed CD platform that is used to build , test and deploy the application to a range of cloud services
- Cloud build can import the source code from a range of repositories and produce artifacts including docker builds and java archives
- Cloud build executes the build as a series of steps in a sequence .

NO→LATENCY

Cloud Build - Steps

- Create a **cloudbuild.yaml** file containing the series of steps including :
    1. Building the docker image
    2. Pushing the image to container registry
    3. Deploy the image to cloud run

- Submit the build file to execute the steps in series

- **Continuos Deployment :**
    - Create a git repo
    - Upload the source code along with cloudbuild.yaml file
    - Create build triggers on GCP
    - Automatically detect code changes and deploy new application to cloud run