


Tahapan Deploy Model (Docker) Kedalam Cloud Run

Sebelum melakukan deployment kedalam cloud run ada beberapa hal yang harus dipastikan :

1. Sudah memiliki Akun Docker Hub
2. Sudah membuat API yang terintegrasi dengan Docker
3. Sudah melakukan Build Image docker dan push docker kedalam Docker Hub
4. Menyalin username Docker hub beserta imagenya.


Berikut adalah tahapan ketika set up di *cloud*:

1. Enable API

 Google Cloud

TugasAkhir ▼

← Product details





Cloud Run Admin API

[Google Enterprise API](#)

Serverless agility for containerized apps

MANAGE

TRY THIS API 

 API Enabled

2. *Create Service* di Cloud Run

☰

Google Cloud

TugasAkhir ▼

Search (/)

» Cloud Run

← Create service

A service exposes a unique endpoint and automatically scales the underlying infrastructure to handle incoming requests. Service name and region cannot be changed later.

Artifact Registry

Docker Hub

☒ Deploy one revision from an existing container image

☐ Continuously deploy from a repository

GitHub

☐ Continuously deploy from a repository

Pada bagian ini, peneliti memilih *deploy* di *container*. Kemudian mengisi *container image URL*.

Container image URL [SELECT](#)

TEST WITH A SAMPLE CONTAINER

Should listen for HTTP requests on \$PORT and not rely on local state. [How to build a container?](#)

Selanjutnya, memasukkan nama *service name* dan memilih *region* apa yang ingin dipilih.

Configure

Service name *

Region *

[How to pick a region?](#)

Pada bagian autentikasi, peneliti memilih 'allow unauthenticated invocations' agar dapat digunakan untuk semua orang.

Authentication *

- ☒ Allow unauthenticated invocations
Check this if you are creating a public API or website.
- ☐ Require authentication
Manage authorized users with Cloud IAM.

Pada bagian ini, peneliti memilih CPU dan ingress secara default dan untuk *service autoscaling* peneliti memilih 1 agar dapat menghindari isu *cold start*.

CPU allocation and pricing ?

- ☒ CPU is only allocated during request processing
You are charged per request and only when the container instance processes a request.
- ☐ CPU is always allocated
You are charged for the entire lifecycle of the container instance.

Service autoscaling ?

Minimum number of instances

1

Ingress control ?

- ☐ Internal
Allow traffic from your project, shared VPC, and VPC service controls perimeter. Traffic from another Cloud Run service must be routed through a VPC. Limitations apply. [Learn more](#)
- ☒ All
Allow direct access to your service from the internet

Pada bagian ini, peneliti menyesuaikan port berapa *container* akan berjalan dan peneliti menggunakan port bawaan dari Google Cloud Platform.

Container image URL

rolaspane/api_tugas_akhir_fix

Container port

8080

Requests will be sent to the container on this port. We recommend listening on \$PORT instead of this specific number.

Peneliti menggunakan memory sebanyak 4 GiB dan 2 CPU sesuai kebutuhan.

Resources

Memory 4 GiB ▼	CPU 2 ▼
Memory to allocate to each instance of this container.	Number of vCPUs allocated to each instance of this container.

Pada bagian ini, peneliti menerapkan *request timeout* dan juga maksimum *requests per instance* sesuai bawaan dari Google Cloud Platform.

Requests

Request timeout 300 seconds
Time within which a response must be returned (maximum 3600 seconds).
Maximum concurrent requests per instance 80
The maximum number of concurrent requests that can reach each instance. What is concurrency? ↗

Pada bagian ini, peneliti mengisi 1 pada bagian minimum dan 5 pada bagian maksimum untuk memaksimalkan *autoscaling*. Kemudian klik **Create** untuk membuat *service* di Cloud Run.

Revision autoscaling

Minimum and maximum numbers of instances for the new revision.

Minimum number of instances * 1	Maximum number of instances * 5
The service minimum instances is preferable for most use-cases. Only use this setting if you specifically require per-revision settings.	

- ☒ **Startup CPU boost**
Start containers faster by allocating more CPU during startup time. [Learn more](#) [↗](#)

Berikut merupakan tampilan ketika berhasil membuat *service* di Cloud Run dan berhasil men-*deploy container* di Cloud Run.

Google Cloud

Tugas Akhir

Search (/) for resources, docs, products, and more

Search

Cloud Run

Service details

EDIT & DEPLOY NEW REVISION

SET UP CONTINUOUS DEPLOYMENT

tugasakhir-16

Region: asia-southeast2

URL: <https://tugasakhir-16-wujlbqzeq-et.a.run.app>

Min instances: 1

METRICS

SLOS

LOGS

REVISIONS

NETWORKING

SECURITY

TRIGGERS

INTEGRATIONS

PREVIEW

YAML

Revisions

MANAGE TRAFFIC

Filter

Filter revisions

Name	Traffic	Deployed	Revision URLs (tags)	Actions
tugasakhir-16-00001-h62	100% (to latest)	Just now	+	⋮

tugasakhir-16-00001-h62

Deployed by taniafany24@gmail.com using Cloud Console

CONTAINERS

VOLUMES

NETWORKING

SECURITY

YAML

General

CPU allocation

CPU is only allocated during request processing

Startup CPU boost

Enabled

Concurrency

80

Request timeout

300 seconds

Execution environment

Default

Autoscaling

Min instances

1

Max instances

5

Image URL

mirror.gcr.io/rolaspane/api_tugas_akhir_fix@sha256-...

Port

8080

Revisions ...sakhir-16 – Cloud R

tugasakhir-16-wujlbqzeq-et.a.i

←

→

↺

tugasakhir-16-wujlbqzeq-et.a.run.app

Pretty-print

```
{"detail": "Method Not Allowed"}
```