

## Weekly Assignment 4

Endpoint: 34.134.197.8

Repository: <https://github.com/PutraAlFarizi15/CI-CD-GKE-Deploy>

### 1. Overview of the Project

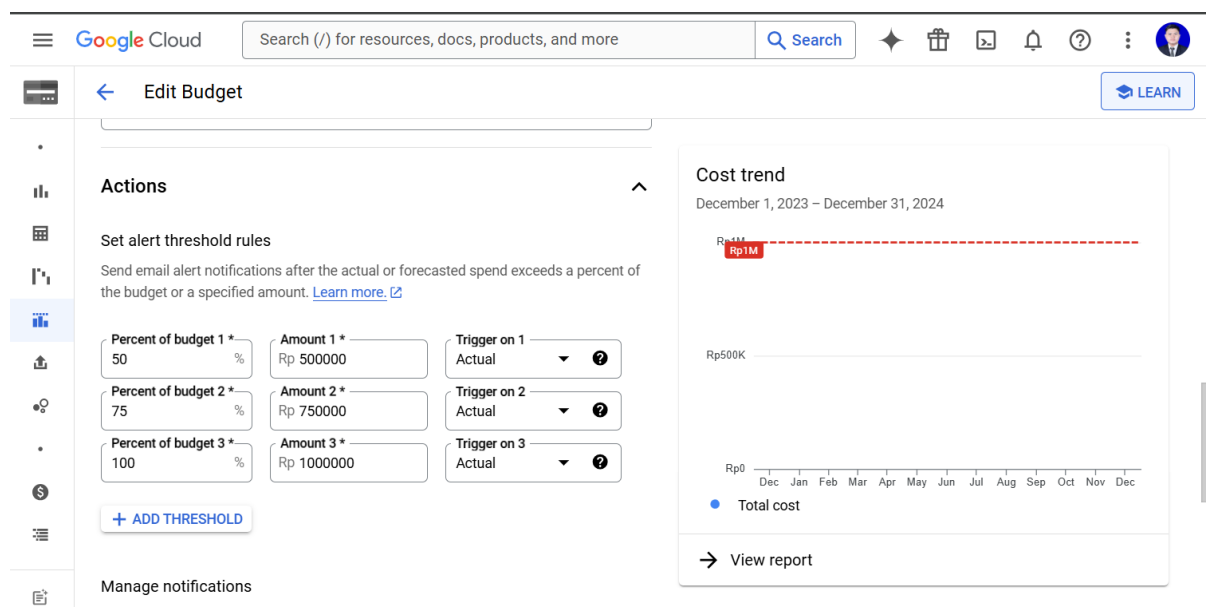
This project focuses on deploying a text generation model based on GPT-2 to Google Kubernetes Engine (GKE) using an automated CI/CD pipeline. The GPT-2 model generates coherent text based on user prompts and is exposed as an API endpoint. The deployment ensures scalability and availability by leveraging Kubernetes and Cloud Build for automation. This project also includes a cost management feature via GCP budget alerts.

### 2. Configure the Budget Alert

Setting up a budget alert with a maximum limit of one million Rupiah. Three notification levels are configured based on the percentage of budget usage:

- **50%:** The first notification is triggered when the budget usage reaches half of the total.
- **75%:** The second notification serves as an additional warning before approaching the limit.
- **100%:** The final notification is sent when the budget is fully utilized.

This configuration aims to effectively monitor and manage cloud budget usage.



### 3. Kubernetes Clusters

Google Cloud

contai

Search

◆

📦

📄

🔔

?

⋮

👤

Kubernetes clusters


CREATE


DEPLOY


REFRESH


ONBOARDING


LEARN




































































#### easier at enterprise scale

GKE Enterprise combines multi-cluster and multi-team operations with fully managed security, governance, and service networking components. Enjoy all the benefits of GKE Standard along with the tools that secure workloads, enforce compliance policies, and provide application visibility with actionable insights and an application-aware network for resiliency.

When you're ready to scale beyond a single team or cluster, GKE Enterprise delivers an integrated and consistent way to configure, secure, protect, and monitor container workloads.

LEARN AND ENABLE

OVERVIEW

OBSERVABILITY

COST OPTIMIZATION

Filter

Enter property name or value

?

⋮

☐

Status

☐

✓

ci-cd-cluster

us-central1-f

Standard

2

4

8 GB

—

⋮

Google Cloud

contai

Search

◆

📦

📄

🔔

?

⋮

👤

Clusters

EDIT

DELETE

ADD NODE POOL

DEPLOY

CONNECT

DUPLICATE

DETAILS

NODES

STORAGE

OBSERVABILITY

LOGS

APP ERRORS (2)

You can find your general cluster logs and your cluster's Autoscaler logs below.

CLUSTER LOGS

AUTOSCALER LOGS

Suggested filters (2)


⌵

Description

Occurrences (last 1 day)

Actions

Google Kubernetes cluster pod creation and deletion for a specific cluster




54

Apply

⋮

GKE audit logs



130,059

Apply

⋮

Severity

Default

Filter

Search all fields and values

?

🔄

📄

SEVERITY

TIMESTAMP

SUMMARY

>

2024-12-01 09:37:15.314 ICT

Kubernetes Apiservice Requests update kube-system:snapshot-controller-leader system:snapshot-controller {#type: type.googleapis...

>

2024-12-01 09:37:15.334 ICT

Kubernetes Apiservice Requests update kube-system:service-steering.networki... system:service-steering {#type: type.googleapis...

>

2024-12-01 09:37:15.337 ICT

Kubernetes Apiservice Requests update kube-system:external-snapshotter-lead... system:pdcsi-controller {#type: type.googleapis...

### 4. Triggers

Google Cloud

Kubernetes-Project

trigger

Search

◆

📦

📄

🔔

?

⋮

👤

Triggers

Create trigger

Connect repository

Manage repositories

Learn

Region

global (non-regional)

Filter

Enter property name or value

?

⋮

Name

Description

Repository

Event

Build configuration

Status

GKE-Deploy-Week-4

This GKE Deploy for Model

PutraAlFarizi15/Ci-CD-GKE-Deploy

Push to branch

cloudbuild.yaml

Enabled

Run

⋮

<div> <div>Google Cloud</div> <div>Kubernetes-Project</div> <div>trigger</div> <div>Search</div> </div>																																																						
<div> <div>Build history</div> <div>Stop streaming builds</div> <div>Learn</div> </div>																																																						
<div> <div>Region</div> <div>global (non-regional)</div> </div>																																																						
<div> <div>Filter</div> <div>Enter property name or value</div> </div>																																																						
<table> <tr> <th></th><th>Status</th><th>Build</th><th>Source</th><th>Ref</th><th>Commit</th><th>Trigger Name</th><th>Created</th><th>Duration</th><th>Security Insights</th><th></th></tr> <tr> <td><input type="checkbox"/></td><td>✓</td><td>9d0a7cac</td><td>PutraAlFarizi15/Cl...</td><td>master</td><td>085fcc6</td><td>GKE-Deploy...</td><td>12/1/24, 9:26 AM</td><td>9 min 30 sec</td><td>View</td><td>⋮</td></tr> <tr> <td><input type="checkbox"/></td><td>✓</td><td>8647dd1d</td><td>PutraAlFarizi15/Cl...</td><td>master</td><td>f853fc5</td><td>GKE-Deploy...</td><td>12/1/24, 6:48 AM</td><td>9 min 35 sec</td><td>View</td><td>⋮</td></tr> <tr> <td><input type="checkbox"/></td><td>✗</td><td>8aaacc19</td><td>PutraAlFarizi15/Cl...</td><td>master</td><td>fb1a70d</td><td>GKE-Deploy...</td><td>12/1/24, 6:35 AM</td><td>10 min</td><td>-</td><td>⋮</td></tr> </table>												Status	Build	Source	Ref	Commit	Trigger Name	Created	Duration	Security Insights		<input type="checkbox"/>	✓	9d0a7cac	PutraAlFarizi15/Cl...	master	085fcc6	GKE-Deploy...	12/1/24, 9:26 AM	9 min 30 sec	View	⋮	<input type="checkbox"/>	✓	8647dd1d	PutraAlFarizi15/Cl...	master	f853fc5	GKE-Deploy...	12/1/24, 6:48 AM	9 min 35 sec	View	⋮	<input type="checkbox"/>	✗	8aaacc19	PutraAlFarizi15/Cl...	master	fb1a70d	GKE-Deploy...	12/1/24, 6:35 AM	10 min	-	⋮
	Status	Build	Source	Ref	Commit	Trigger Name	Created	Duration	Security Insights																																													
<input type="checkbox"/>	✓	9d0a7cac	PutraAlFarizi15/Cl...	master	085fcc6	GKE-Deploy...	12/1/24, 9:26 AM	9 min 30 sec	View	⋮																																												
<input type="checkbox"/>	✓	8647dd1d	PutraAlFarizi15/Cl...	master	f853fc5	GKE-Deploy...	12/1/24, 6:48 AM	9 min 35 sec	View	⋮																																												
<input type="checkbox"/>	✗	8aaacc19	PutraAlFarizi15/Cl...	master	fb1a70d	GKE-Deploy...	12/1/24, 6:35 AM	10 min	-	⋮																																												

Google Cloud

Kubernetes-Project

trigger

×

Search

Build details

Rebuild

Copy url

Learn

Successful: 9d0a7cac-d781-4146-b648-eea05d536f33

Started on Dec 1, 2024, 9:26:07 AM

Trigger

GKE-Deploy-Week-4

Source

PutraAlFarizi15/CI-CD-GKE-Deploy

Branch

master

Commit

085fcc6

Steps

Duration

Build Summary

4 Steps

0: gcr.io/cloud-builders/docker build -t gcr.io/corded-terrain-4413...

00:03:45

1: gcr.io/cloud-builders/docker push gcr.io/corded-terrain-441302...

00:04:23

2: gcr.io/google.com/cloudsdktool... bash -c gcloud container clusters ...

00:01:15

3: gcr.io/google.com/cloudsdktool... bash -c kubectl apply -f kubernet...

00:00:02

Build Summary

Build log

Execution details

Build artifacts

Filter

Search all fields and values

SEVERITY

TIMESTAMP

SUMMARY

> i 2024-12-01 09:35:37.557 ICT b6ca42156b9f: Layer already exists

> i 2024-12-01 09:35:37.559 ICT 00547dd240c4: Layer already exists

> i 2024-12-01 09:35:37.564 ICT 1354cd991b1c: Layer already exists

> i 2024-12-01 09:35:37.595 ICT 24b5ce0f1e07: Layer already exists

> i 2024-12-01 09:35:37.743 ICT latest: digest: sha256:dec49ee48637f135788aeebe1dd6fb9d0d282af6c79d5e377349f1c84be63...

> i 2024-12-01 09:35:37.936 ICT DONE

## 5. Endpoint Acces

To access the API, use the following endpoint: <http://34.134.197.8>.

The API interface includes two main features:

- Enter Your Prompt:** A field where you can input your prompt or initial text as the basis for text generation.
- Generate Text:** A button to initiate the text generation process based on the provided prompt.

Simply enter your prompt in the designated field, then click the **Generate** button to receive the generated text from the model.

.



## GPT-2 Text Generation

Enter your prompt:

what is cat and dog?

Generate Text

### Generated Text:

what is cat and dog? The cat is a very important part of the family. It is the most important thing in the world. The dog is very much a part. . . . (The dog) is not a dog. He is an animal. And he is part, not partless. But he has a lot of things to do with the whole family, and he does not have a home. So he can't be a housewife. (He is) a cat. A dog, a man, an ox, or a donkey. They are all part and not separate. There is no separation. You can have one dog and one man and you can live together. That is what the dog does. If you have two dogs, you will have the same dog; if you don't have both, then you are not going to live with one. In the end, the cat has to be part in all of this. I think that is why