EAI 6010:

APPLICATIONS OF ARTIFICIAL INTELLIGENCE

Module 5: Dog breed Microservice Deployment On GCP

Submitted To:
Prof RJ Munthali, Faculty Lecturer

Submitted By:
Abhilash Dikshit

Academic Term: Fall 2023
Northeastern University, Vancouver, BC,Canada
Master of Professional Studies in Analytics

December 19, 2023

Title: Dog Breed Classifier Microservice Deployment on GCP

I. Service Description

The microservice developed is a Dog Breed Classifier based on the Stanford Dog Breed dataset which was used in Module 2 assignment for dog breed prediction. It utilizes transfer learning with pre-trained weights from the inception model. Additional layers are strategically added over the inception layer to enhance training efficiency and improve classification accuracy.

General Input: The service accepts images of dogs as input.

Example Input: An image of a Golden Retriever

General Output: The output is the predicted dog breed based on the input image.

Example Output: Predicted Breed > Golden Retriever

II. Service URL

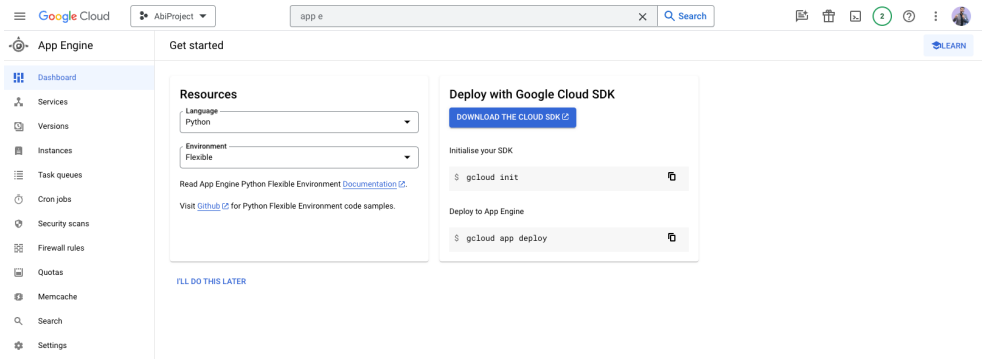
The Dog Breed Classifier microservice is deployed at the following URL:

LINK: https://abiproject-391006.uc.r.appspot.com

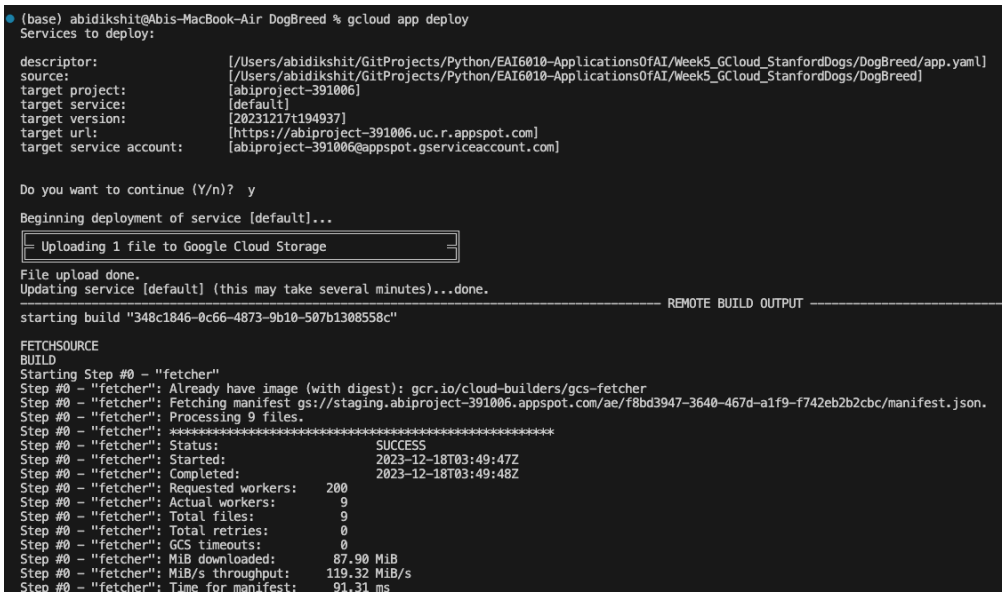
Deployment Process:

To deploy the Dog Breed application on Google Cloud Platform (GCP), we follow these steps:

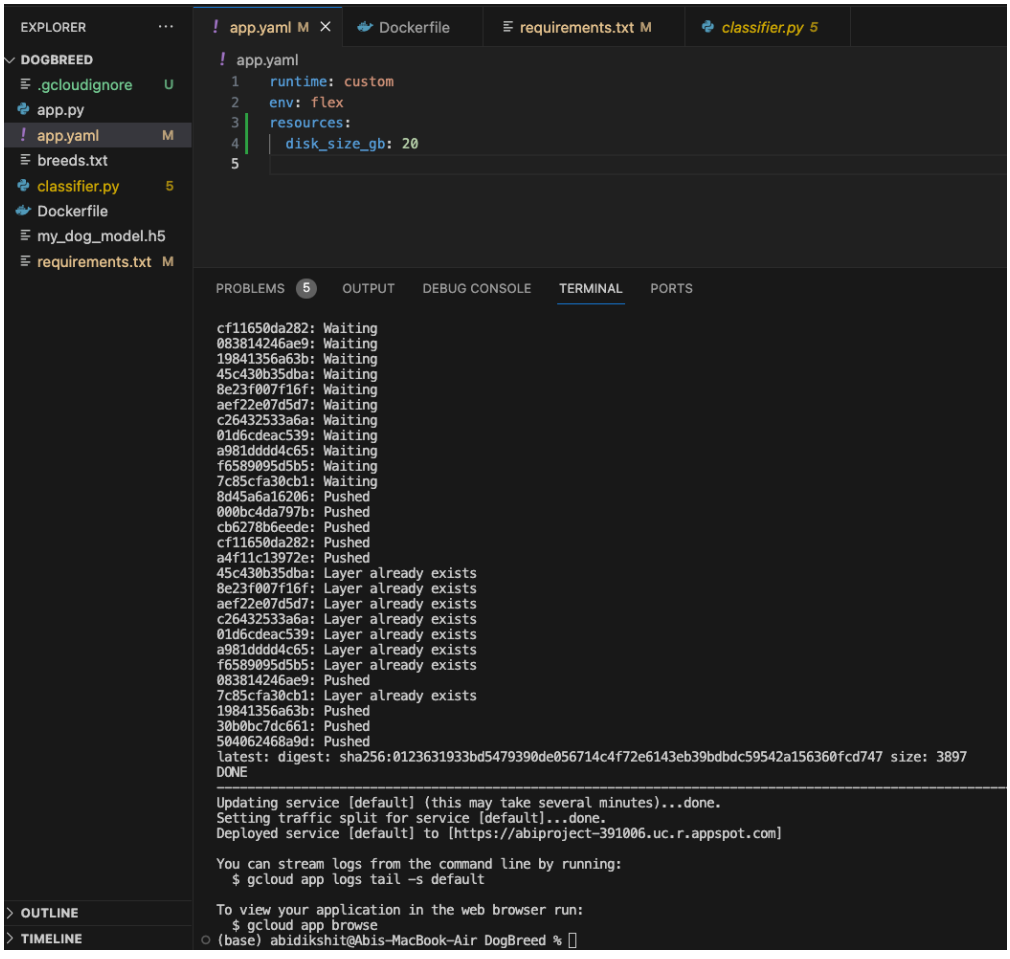
1. Created a project "abiproject-391006" in the GCP Console.
2. Enabled the below API services in "APIs & Services" > "Library" page:
    - Google Cloud Storage API
    - Google Cloud SQL API
    - Google App Engine API
3. Created a new App Engine service.



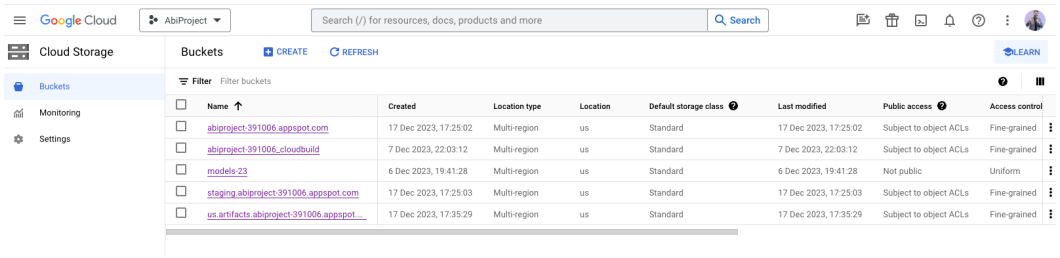
4. Deployed the code to App Engine.
    - Increased the disk size to 20 GB in “app.yaml” file as the deployment was failing due to less space.



```
(base) abidikshit@Abis-MacBook-Air DogBreed % gcloud app deploy
Services to deploy:

descriptor:                  [/Users/abidikshit/GitProjects/Python/EAI6010-ApplicationsOfAI/Week5_GCloud_StanfordDogs/DogBreed/app.yaml]
source:                      [/Users/abidikshit/GitProjects/Python/EAI6010-ApplicationsOfAI/Week5_GCloud_StanfordDogs/DogBreed]
target project:              [abiproject-391006]
target service:              [default]
target version:              [20231217t194937]
target url:                  [https://abiproject-391006.uc.r.appspot.com]
target service account:      [abiproject-391006@appspot.gserviceaccount.com]


Do you want to continue (Y/n)?  y

Beginning deployment of service [default]...
╔═ Uploading 1 file to Google Cloud Storage ═╗
File upload done.
Updating service [default] (this may take several minutes)...done.
------------------------------------------------------------------------------- REMOTE BUILD OUTPUT ------------------------------------------------------------------------------
starting build "348c1846-0c66-4873-9b10-507b1308558c"

FETCHSOURCE
BUILD
Starting Step #0 - "fetcher"
Step #0 - "fetcher": Already have image (with digest): gcr.io/cloud-builders/gcs-fetcher
Step #0 - "fetcher": Fetching manifest gs://staging.abiproject-391006.appspot.com/ae/f8bd3947-3640-467d-a1f9-f742eb2b2cbc/manifest.json.
Step #0 - "fetcher": Processing 9 files.
Step #0 - "fetcher": ******************************************************
Step #0 - "fetcher": Status:                     SUCCESS
Step #0 - "fetcher": Started:                    2023-12-18T03:49:47Z
Step #0 - "fetcher": Completed:                  2023-12-18T03:49:48Z
Step #0 - "fetcher": Requested workers:    200
Step #0 - "fetcher": Actual workers:         9
Step #0 - "fetcher": Total files:            9
Step #0 - "fetcher": Total retries:          0
Step #0 - "fetcher": GCS timeouts:           0
Step #0 - "fetcher": MiB downloaded:       87.90 MiB
Step #0 - "fetcher": MiB/s throughput:    119.32 MiB/s
Step #0 - "fetcher": Time for manifest:    91.31 ms
```



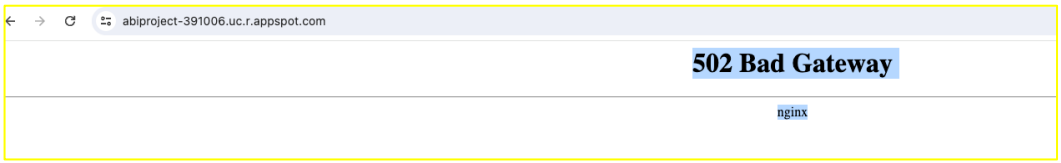
```
EXPLORER
DOGBREED
.gcloudignore U
app.py
app.yaml M
breeds.txt
classifier.py 5
Dockerfile
my_dog_model.h5
requirements.txt M
OUTLINE
TIMELINE

app.yaml M | Dockerfile | requirements.txt M | classifier.py 5

app.yaml
1  runtime: custom
2  env: flex
3  resources:
4    disk_size_gb: 20
5

PROBLEMS 5  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

cf11650da282: Waiting
083814246ae9: Waiting
19841356a63b: Waiting
45c430b35dba: Waiting
8e23f007f16f: Waiting
aef22e07d5d7: Waiting
c26432533a6a: Waiting
01d6cdeac539: Waiting
a981dddd4c65: Waiting
f6589095d5b5: Waiting
7c85cfa30cb1: Waiting
8d45a6a16206: Pushed
000bc4da797b: Pushed
cb6278b6eede: Pushed
cf11650da282: Pushed
a4f11c13972e: Pushed
45c430b35dba: Layer already exists
8e23f007f16f: Layer already exists
aef22e07d5d7: Layer already exists
c26432533a6a: Layer already exists
01d6cdeac539: Layer already exists
a981dddd4c65: Layer already exists
f6589095d5b5: Layer already exists
083814246ae9: Pushed
7c85cfa30cb1: Layer already exists
19841356a63b: Pushed
30b0bc7dc661: Pushed
504062468a9d: Pushed
latest: digest: sha256:0123631933bd5479390de056714c4f72e6143eb39bdbdc59542a156360fcd747 size: 3897
DONE
------------------------------------------------------------------------------------------------------------------------------------------------------------------------
Updating service [default] (this may take several minutes)...done.
Setting traffic split for service [default]...done.
Deployed service [default] to [https://abiproject-391006.uc.r.appspot.com]

You can stream logs from the command line by running:
  $ gcloud app logs tail -s default

To view your application in the web browser run:
  $ gcloud app browse
(base) abidikshit@Abis-MacBook-Air DogBreed %
```




```
(base) abidikshit@Abis-MacBook-Air DogBreed % gcloud app browse
Opening [https://abiproject-391006.uc.r.appspot.com] in a new tab in your default browser.
(base) abidikshit@Abis-MacBook-Air DogBreed %
```

LINK: https://abiproject-391006.uc.r.appspot.com

- Unable to get the frontend app due to 502 Bad Gateway error.



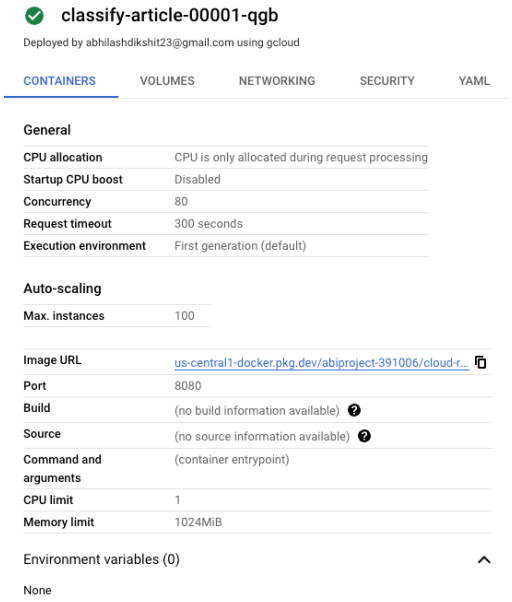
Troubleshooting:

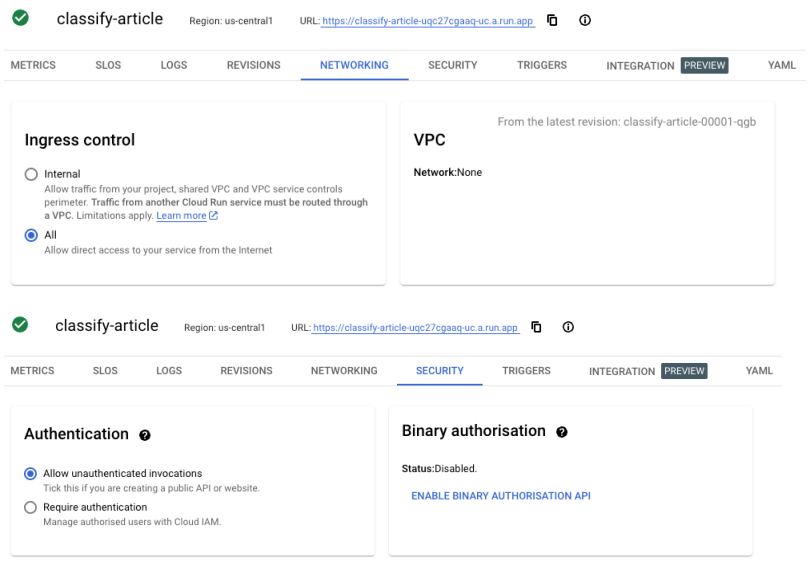
- Checked the service for "classify-article" app engine. – Set correctly
- Checked the firewall rules set. – Set correctly
- Checked the Networking Ingress and Egress setting of app engine. – Set correctly
- Checked the Security setting. – Set correctly

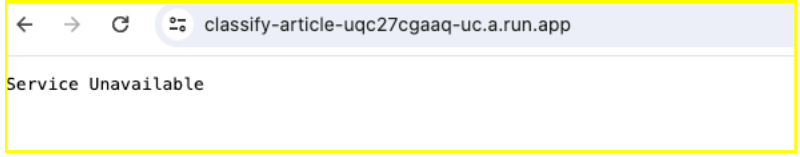
```
(base) abidikshit@Abis-MacBook-Air DogBreed % gcloud run services describe classify-article
Please specify a region:
 [1] asia-east1
 [2] asia-east2
 [3] asia-northeast1
 [4] asia-northeast2
 [5] asia-northeast3
 [6] asia-south1
 [7] asia-south2
 [8] asia-southeast1
 [9] asia-southeast2
 [10] australia-southeast1
 [11] australia-southeast2
 [12] europe-central2
```

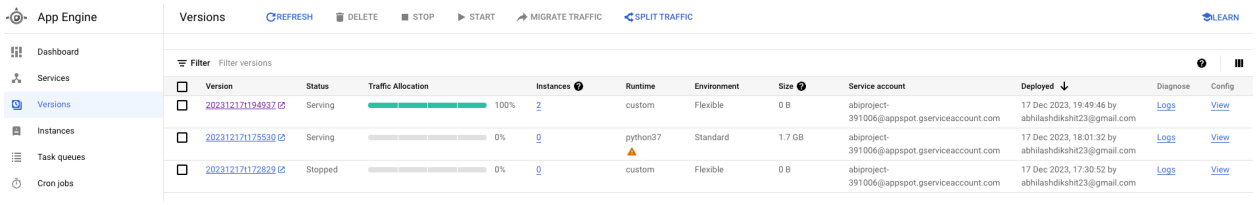
```
[38] us-west3
[39] us-west4
[40] cancel
Please enter numeric choice or text value (must exactly match list item):  31

To make this the default region, run `gcloud config set run/region us-central1`.

✓ Service classify-article in region us-central1

URL:     https://classify-article-uqc27cgaaq-uc.a.run.app
Ingress: all
Traffic:
  100% LATEST (currently classify-article-00001-qgb)

Last updated on 2023-12-11T20:14:21.649802Z by abhilashdikshit23@gmail.com:
  Revision classify-article-00001-qgb
  Container None
    Image:          us-central1-docker.pkg.dev/abiproject-391006/cloud-run-source-deploy/classify-article@sha256:a6edec5ee052d57a4df34507df50d4ec186fef530f39d1f0c9f0697a501b4bc2
    Port:           8080
    Memory:         1024Mi
    CPU:            1000m
    Startup Probe:
      TCP every 240s
      Port:         8080
      Initial delay: 0s
      Timeout:      240s
      Failure threshold: 1
      Type:         Default
  Service account:  612279551194-compute@developer.gserviceaccount.com
  Concurrency:      80
  Max Instances:    100
  Timeout:          300s
(base) abidikshit@Abis-MacBook-Air DogBreed %
```

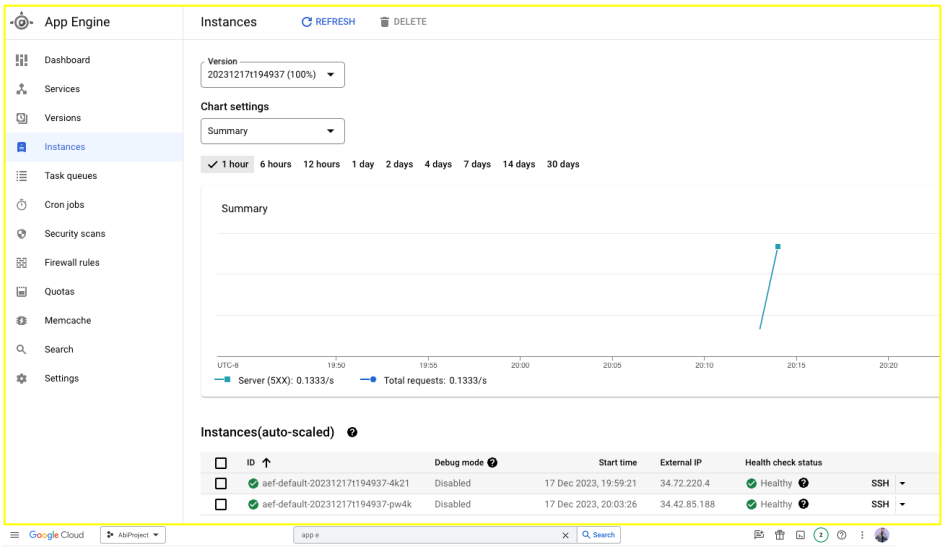
```
(base) abidikshit@Abis-MacBook-Air DogBreed % gcloud compute firewall-rules list
NAME                    NETWORK  DIRECTION  PRIORITY  ALLOW                         DENY  DISABLED
default-allow-icmp      default  INGRESS    65534     icmp                                False
default-allow-internal  default  INGRESS    65534     tcp:0-65535,udp:0-65535,icmp        False
default-allow-rdp       default  INGRESS    65534     tcp:3389                            False
default-allow-ssh       default  INGRESS    65534     tcp:22                              False

To show all fields of the firewall, please show in JSON format: --format=json
To show all fields in table format, please see the examples in --help.
```

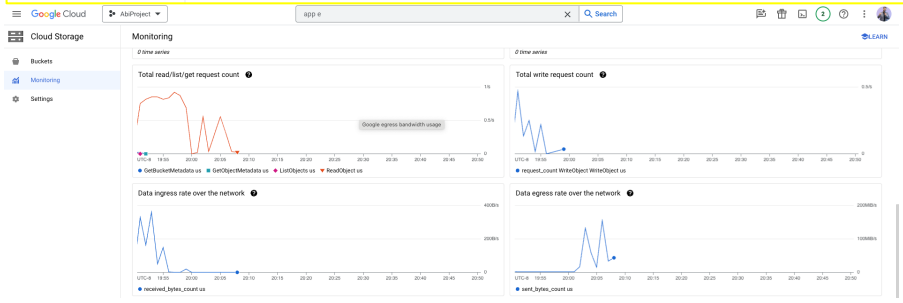


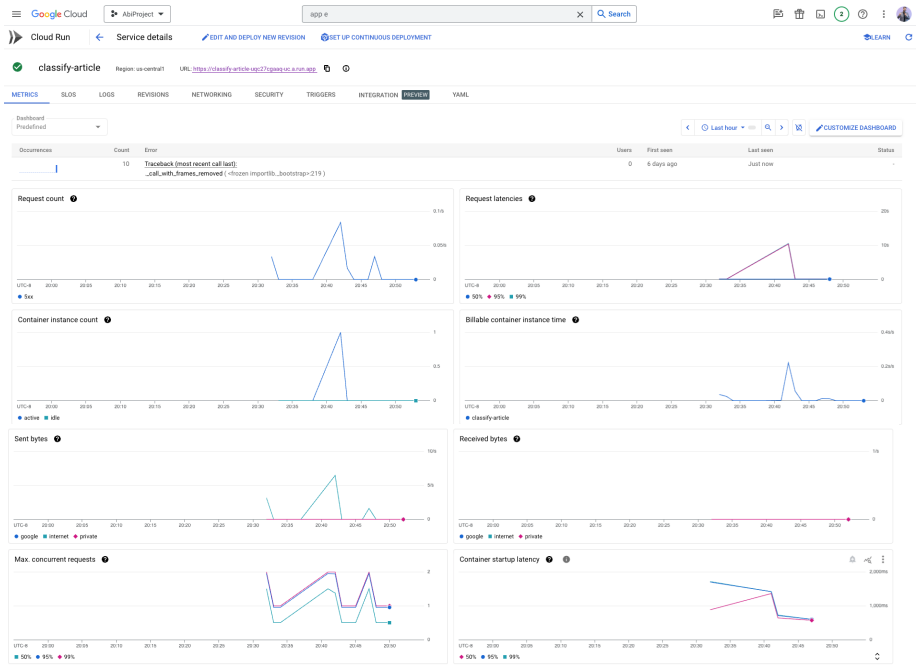





App Engine
Versions
REFRESH
DELETE
STOP
START
MIGRATE TRAFFIC
SPLIT TRAFFIC
LEARN
Dashboard
Services
Versions
Instances
Task queues
Cron jobs
Filter
Filter versions
Version
Status
Traffic Allocation
Instances
Runtime
Environment
Size
Service account
Deployed
Diagnose
Config
20231217t194937
Serving
100%
2
custom
Flexible
0 B
abiproject-391006@appspot.gserviceaccount.com
17 Dec 2023, 19:49:46 by abhilashdikshit23@gmail.com
Logs
View
20231217t175530
Serving
0%
0
python37
Standard
1.7 GB
abiproject-391006@appspot.gserviceaccount.com
17 Dec 2023, 18:01:32 by abhilashdikshit23@gmail.com
Logs
View
20231217t172829
Stopped
0%
0
custom
Flexible
0 B
abiproject-391006@appspot.gserviceaccount.com
17 Dec 2023, 17:30:52 by abhilashdikshit23@gmail.com
Logs
View


App Engine
Instances
REFRESH
DELETE
Dashboard
Services
Versions
Instances
Task queues
Cron jobs
Security scans
Firewall rules
Quotas
Memcache
Search
Settings
Version
20231217t194937 (100%)
Chart settings
Summary
1 hour
6 hours
12 hours
1 day
2 days
4 days
7 days
14 days
30 days
Summary
UTC-8
19:50
19:55
20:00
20:05
20:10
20:15
20:20
Server (5XX): 0.1333/s
Total requests: 0.1333/s
Instances(auto-scaled)
ID
Debug mode
Start time
External IP
Health check status
aef-default-20231217t194937-4k21
Disabled
17 Dec 2023, 19:59:21
34.72.220.4
Healthy
SSH
aef-default-20231217t194937-pw4k
Disabled
17 Dec 2023, 20:03:26
34.42.85.188
Healthy
SSH


Google Cloud
Cloud Storage
Monitoring
LEARN
Buckets
Monitoring
Settings
Total read/list/get request count
Total write request count
Data ingress rate over the network
Data egress rate over the network


Google Cloud
Cloud Run
Service details
EDIT AND DEPLOY NEW REVISION
SET UP CONTINUOUS DEPLOYMENT
LEARN
classify-article
METRICS
SLOS
LOGS
REVISIONS
NETWORKING
SECURITY
TRIGGERS
INTEGRATION
PREVIEW
YAML
CUSTOMIZE DASHBOARD
Request count
Request latencies
Container instance count
Billable container instance time
Sent bytes
Received bytes
Max. concurrent requests
Container startup latency

III. Service Availability

The Dog Breed Classifier service is currently experiencing technical difficulties, resulting in a 502 Bad Gateway error. Despite multiple troubleshooting attempts, I have been unable to resolve the issue promptly.

I am diligently working to identify and rectify the underlying problem as I want to upload this project on GitHub to showcase the end-to-end flow.

IV. Conclusion

In the development of the Dog Breed Classifier, the utilization of transfer learning with the inception model has proven effective. The addition of extra layers over the pre-trained model significantly reduces training time and yields satisfactory results. The web application, built on Streamlit and deployed using GCP, enhances user experience but due to technical difficulties, it cannot be accessed now using the provided URL.

The conclusion highlights the importance of additional layers for task-specific training and emphasizes the efficiency of the Inception model trained on the ImageNet dataset in achieving accelerated results. The successful deployment of the microservice will reflect my dedication to providing a seamless and accessible solution.

V. Future Actionable

- Configuring a MYSQL database in us-central1, and setting "Machine type" field to "POSTGRES_11".
- Updating the App Engine configuration to use the database:
- Testing of the application by clicking on the different breeds of dogs and viewing their details.

VI. References:

1. Stanford Dogs dataset for Fine-Grained Visual Categorization. (n.d.). http://vision.stanford.edu/aditya86/ImageNetDogs/
2. PyTorch Computer Vision - Zero to Mastery Learn PyTorch for Deep Learning. (n.d.). https://www.learnpytorch.io/03_pytorch_computer_vision/