



Enhanced Image Metadata Extractor

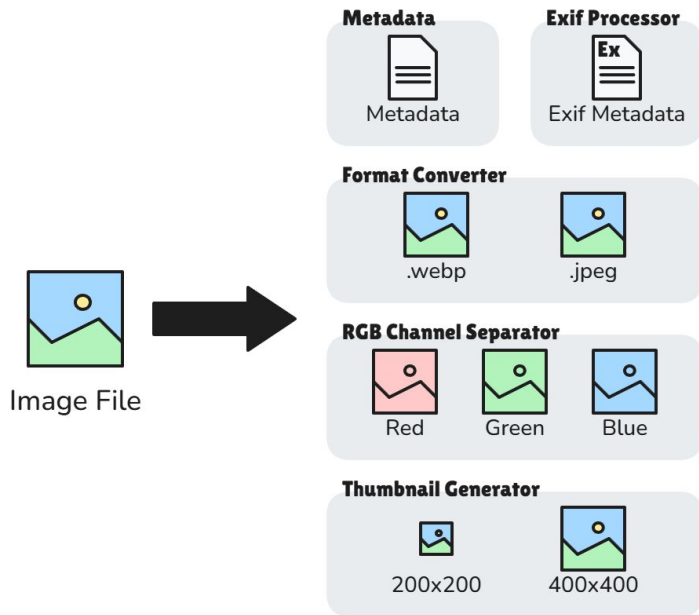
Merlin Volkmer

Bence Balázs

Florian Johann Hafner

Pietro Marchiol

Enhanced Image Metadata Extractor

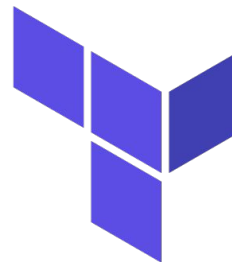




Infrastructure Overview



Google Cloud



HashiCorp

Terraform

Infrastructure Overview



Cloud IAM



Cloud
Pub/Sub



Cloud
Storage



Monitoring



Cloud
Functions



Google Cloud

Functions Framework for Python

pypi package **3.8.2**

Python Unit CI **failing** Python Lint CI **passing** Python Conformance CI **passing** openssf scorecard **7.2**

An open source FaaS (Function as a service) framework for writing portable Python functions -- brought to you by the Google Cloud Functions team.

The Functions Framework lets you write lightweight functions that run in many different environments, including:

- [Google Cloud Run Functions](#)
- Your local development machine
- [Knative](#)-based environments

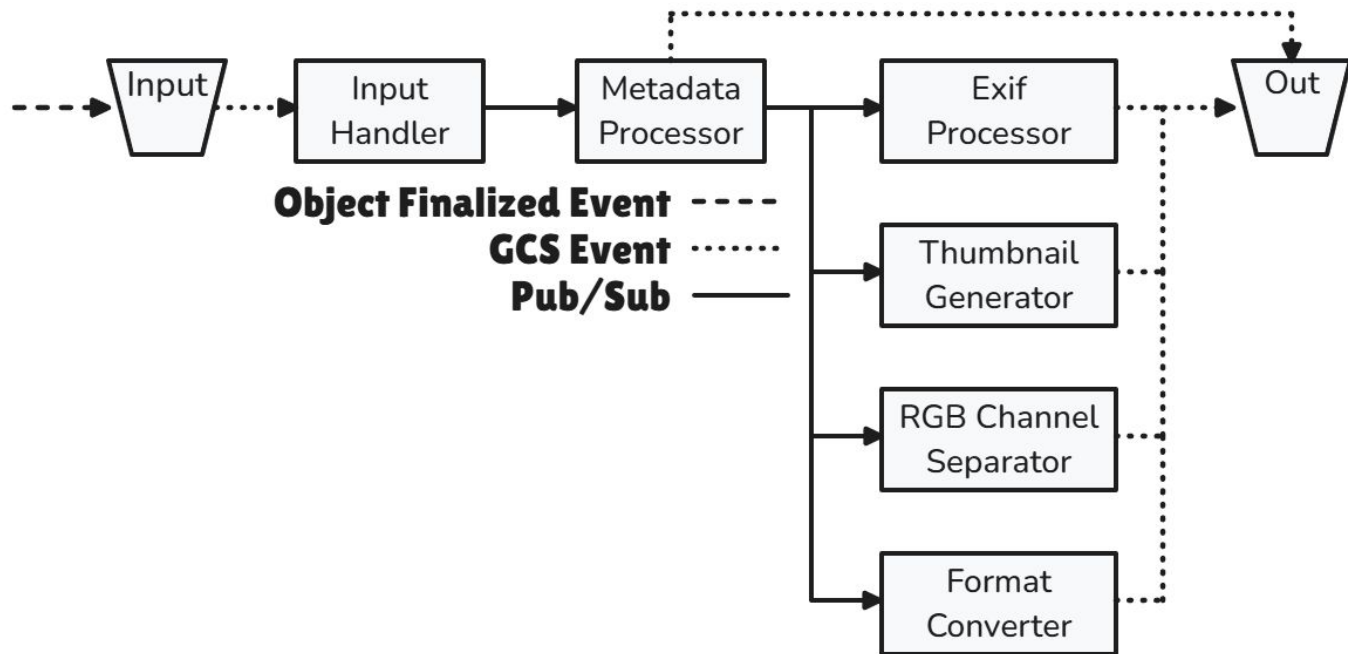
Workflow

metadata/{FILENAME}/
{FILENAME}.json

channels/{FILENAME}/
metadata.json
{FILENAME}_red.png
{FILENAME}_green.png
{FILENAME}_blue.png

converted/{FILENAME}/
metadata.json
{FILENAME}.jpg
{FILENAME}.webp

thumbnails/{FILENAME}/
metadata.json
{FILENAME}_200x200.png
{FILENAME}_400x400.png



Challenges

Setting right IAM permissions!

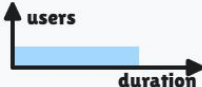







Test Timing Considerations



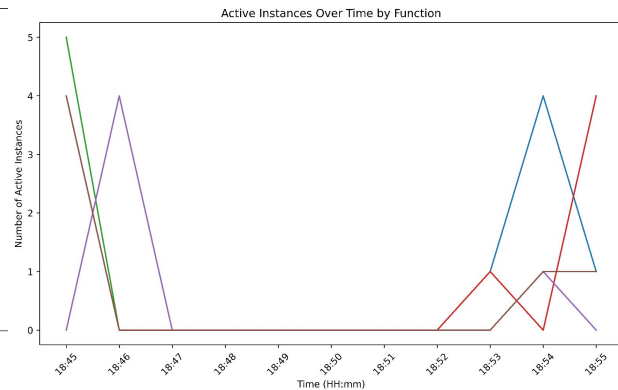
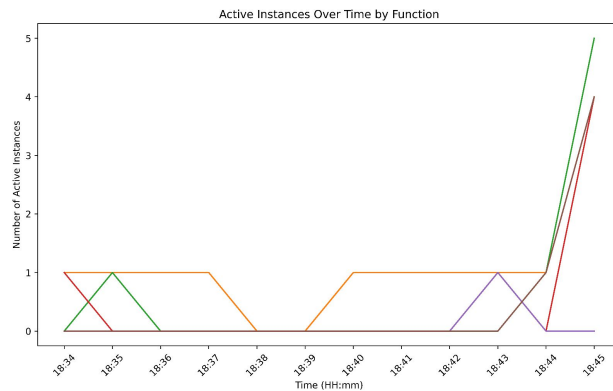
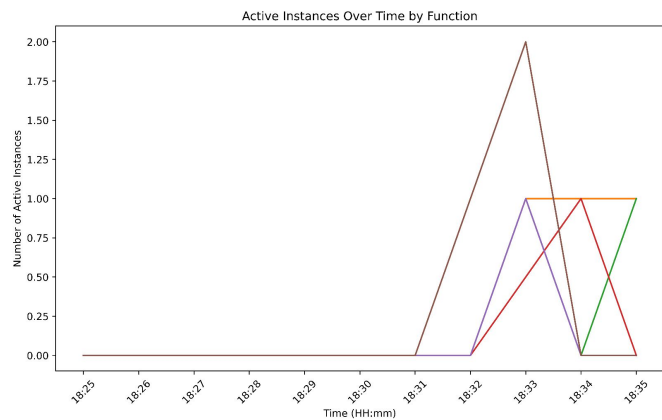
Test Scenarios

```
production_test_scenarios = [  
    {'concurrent': 5, 'duration': 60}, # Baseline  
    {'concurrent': 10, 'duration': 100}, # Medium load  
    {'concurrent': 100, 'duration': 3}, # Stress test  
]
```

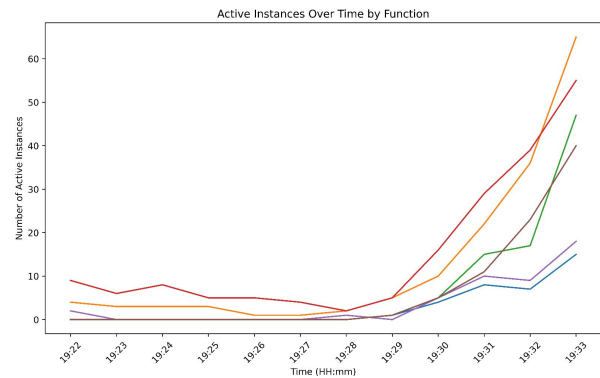
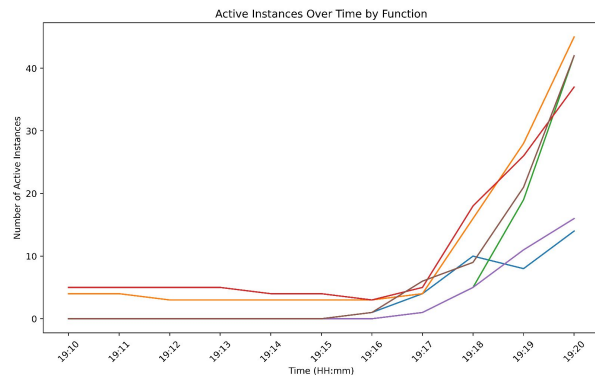
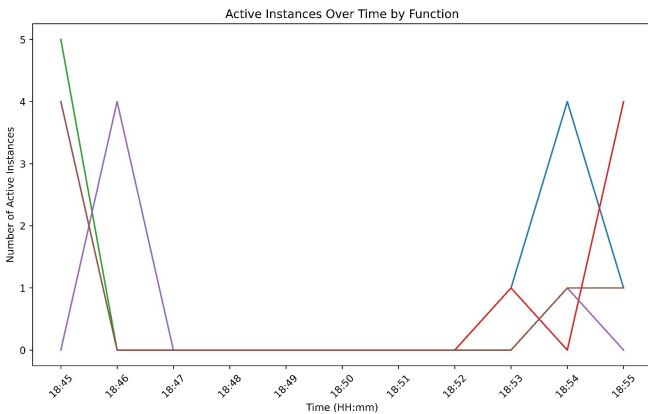
```
simple_test_scenarios = [  
    {'concurrent': 5, 'duration': 10}, # simple  
]
```

Scenario	Memory		
	low	128Mi	mid 256Mi high 512Mi
Simple 	Scenario 1 (S1) 	Scenario 2 (S2) 	Scenario 3 (S3) 
Production 	Scenario 4 (S4) 	Scenario 5 (S5) 	Scenario 6 (S6) 

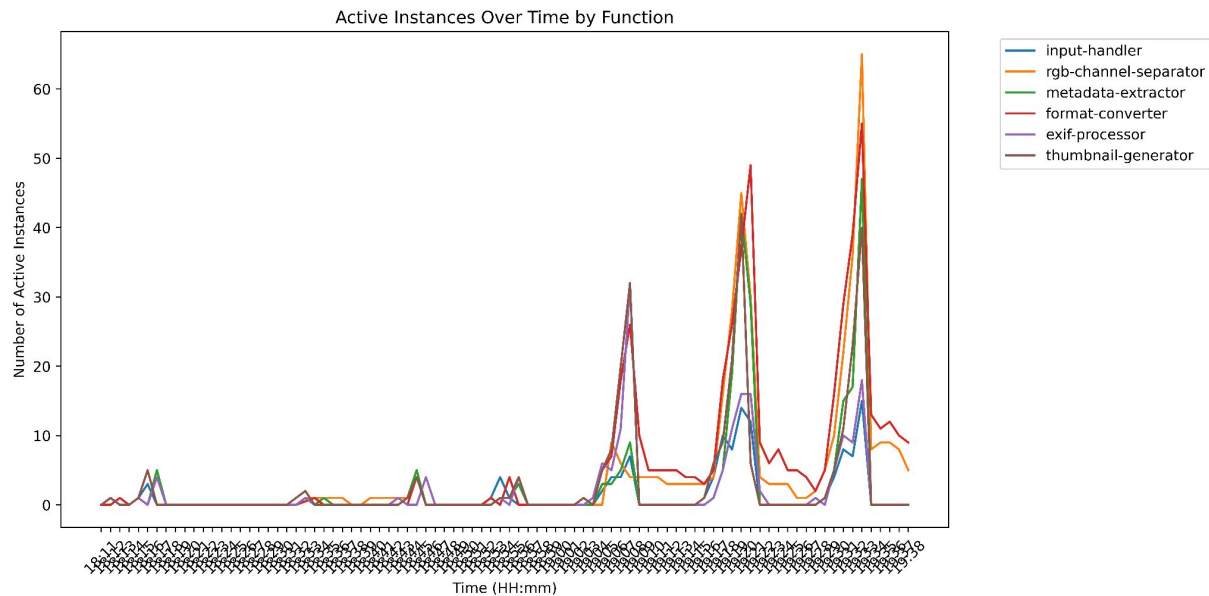
Simple Scenario Results



Production Scenario Results



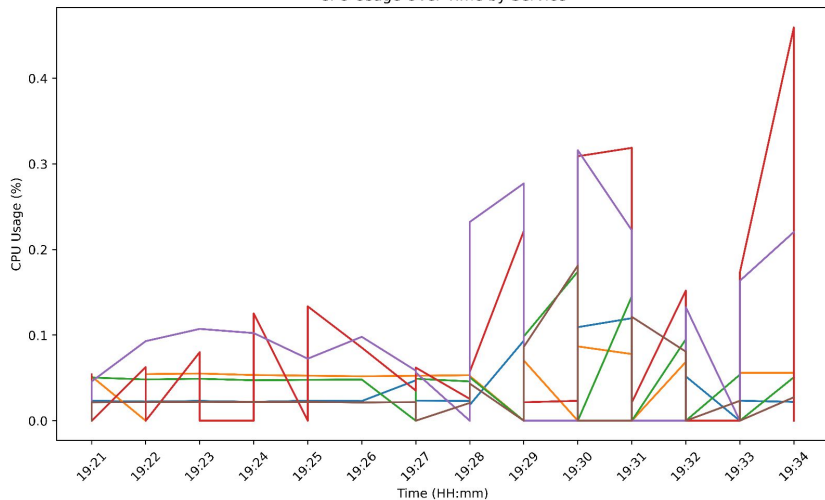
All runs



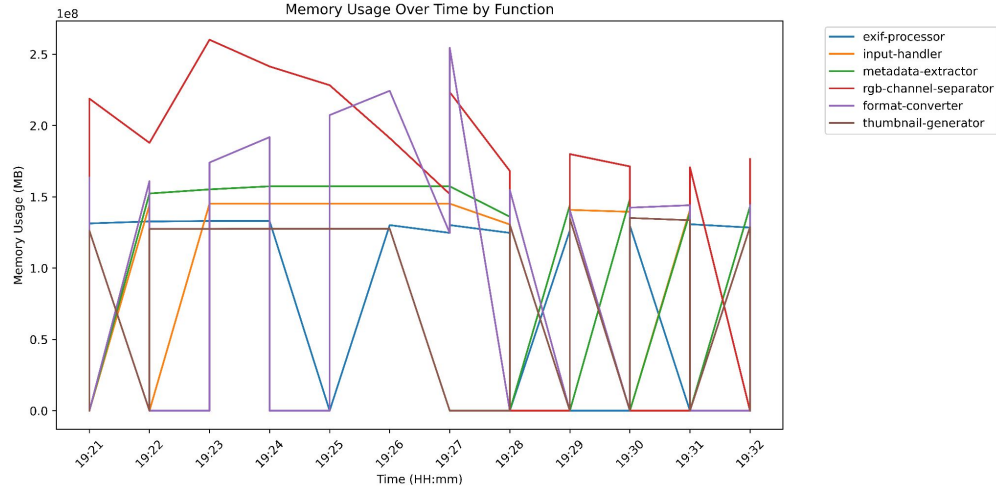
Observations

rgb-channel-seperator and format-converter need the most cpu and memory

CPU Usage Over Time by Service



Memory Usage Over Time by Function





Observations

functions need more than 128MiB

382

'Memory limit of 128 MiB exceeded with 139 MiB used. Consider increasing the memory limit, see <https://cloud.google.com/functions/docs/configuring/memory>'

```
....., retrying... (Attempt 15/15)
Verification failed, retrying... (Attempt 15/15)
Verification failed, retrying... (Attempt 15/15)
Max retries reached. Verification partially incomplete.
Max retries reached. Verification partially incomplete.
Max retries reached. Verification partially incomplete.
Max retries reached. Verification partially incomplete.
```

Scenario Summary:
Total uploads: 5
Successful uploads: 5
Failed uploads: 0
Average duration: 26.95 seconds

Verification rates:
metadata: 20.0%



Recommendations

- Read google cheat sheet documentation beforehand to understand which resources are needed
- everything except `rgb-channel-seperator` and `format-converter` needs less resources but at least ~130Mb