Práctica 6

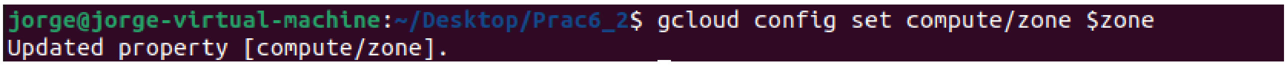
Autor: Jorge Calvo Soria

Creación del Cluster

En esta primera parte se configura un cluster de k8s, que estará sirviendo a un servicio web apache.

1. Primero realizamos gcloud init y elegimos el proyecto.

2. Establecemos la zona del proyecto.

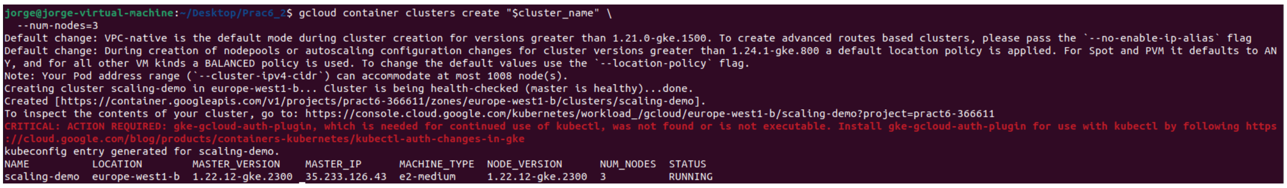


```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ gcloud config set compute/zone $zone
Updated property [compute/zone].
```

3. Previamente hemos establecido una configuración previa en un archivo "config.ini". Tras dicha configuración

```
zone="europe-west1-b"
cluster_name="scaling-demo"
php_deployment="php-apache"
php_manifest="php-apache.yaml"
min_replicas=1
max_replicas=10
cpu_threshold=50
min_nodes=1
max_nodes=5
```

4. Se inicia el cluster de kubernetes en Google con el siguiente comando

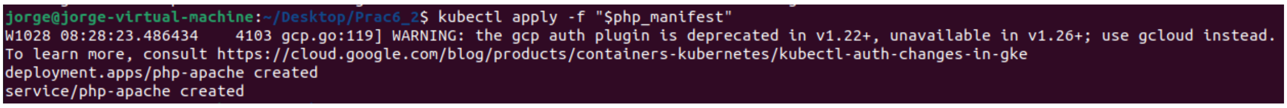


```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ gcloud container clusters create "$cluster_name" \
  --num-nodes=3
Default change: VPC-native is the default mode during cluster creation for versions greater than 1.21.0-gke.1500. To create advanced routes based clusters, please pass the `--no-enable-ip-alias` flag
Default change: During creation of nodepools or autoscaling configuration changes for cluster versions greater than 1.24.1-gke.800 a default location policy is applied. For Spot and PVM it defaults to AN
Y, and for all other VM kinds a BALANCED policy is used. To change the default values use the `--location-policy` flag.
Note: Your Pod address range (`--cluster-ipv4-cidr`) can accommodate at most 1008 node(s).
Creating cluster scaling-demo in europe-west1-b... Cluster is being health-checked (master is healthy)...done.
Created [https://container.googleapis.com/v1/projects/pract6-366611/zones/europe-west1-b/clusters/scaling-demo].
To inspect the contents of your cluster, go to: https://console.cloud.google.com/kubernetes/workload_/gcloud/europe-west1-b/scaling-demo?project=pract6-366611
CRITICAL: ACTION REQUIRED: gke-gcloud-auth-plugin, which is needed for continued use of kubectl, was not found or is not executable. Install gke-gcloud-auth-plugin for use with kubectl by following https
://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
kubeconfig entry generated for scaling-demo.
NAME          LOCATION        MASTER_VERSION    MASTER_IP      MACHINE_TYPE  NODE_VERSION      NUM_NODES  STATUS
scaling-demo  europe-west1-b  1.22.12-gke.2300  35.233.126.43  e2-medium     1.22.12-gke.2300  3          RUNNING
```

5. Para demostrar el autoescalado horizontal de pods vamos a desplegar una imagen de docker basada en php-apache.Para esto, vamos a hacer uso de un manifiesto de despliegue, php-apache.yaml, que contiene la siguiente configuración:

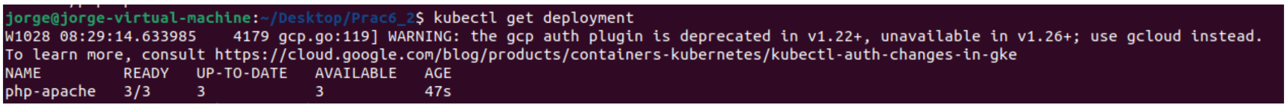
```
 apiVersion: apps/v1
kind: Deployment
metadata:
  name: php-apache
spec:
  selector:
    matchLabels:
      run: php-apache
  replicas: 3
  template:
    metadata:
      labels:
        run: php-apache
    spec:
      containers:
      - name: php-apache
        image: k8s.gcr.io/hpa-example
        ports:
        - containerPort: 80
        resources:
          limits:
            cpu: 500m
          requests:
            cpu: 200m
---
apiVersion: v1
kind: Service
metadata:
  name: php-apache
  labels:
    run: php-apache
spec:
  ports:
  - port: 80
  selector:
    run: php-apache
```

6. Aplicamos el manifiesto



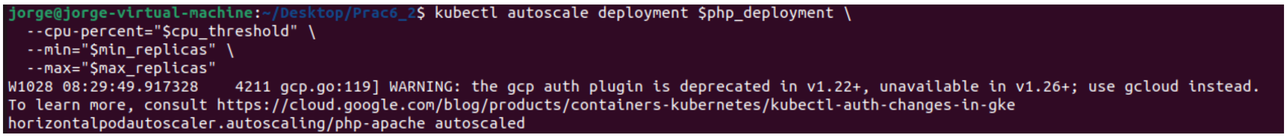
```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ kubectl apply -f "$php_manifest"
W1028 08:28:23.486434    4103 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
deployment.apps/php-apache created
service/php-apache created
```

7. Mostramos el comportamiento del clusterinspeccionando los despliegues que se encuentran en funcionamiento en nuestro cluster



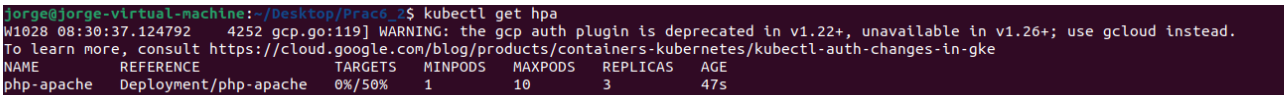
```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ kubectl get deployment
W1028 08:29:14.633985    4179 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
NAME         READY   UP-TO-DATE   AVAILABLE   AGE
php-apache   3/3     3            3           47s
```

8. Aplicamos la configuración del HPA



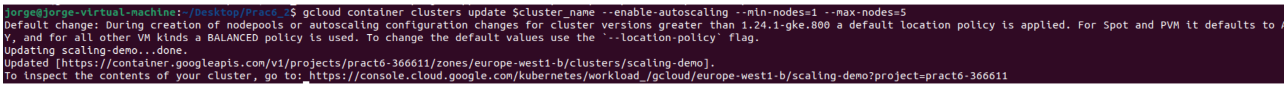
```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ kubectl autoscale deployment $php_deployment \
  --cpu-percent="$cpu_threshold" \
  --min="$min_replicas" \
  --max="$max_replicas"
W1028 08:29:49.917328    4211 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
horizontalpodautoscaler.autoscaling/php-apache autoscaled
```

9. Comprobamos el estado del HPA



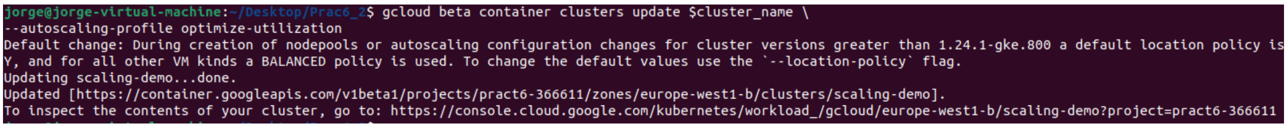
```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ kubectl get hpa
W1028 08:30:37.124792    4252 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
NAME         REFERENCE               TARGETS   MINPODS   MAXPODS   REPLICAS   AGE
php-apache   Deployment/php-apache   0%/50%    1         10        3          47s
```

10. Activamos el autoescalado horizontal del cluster



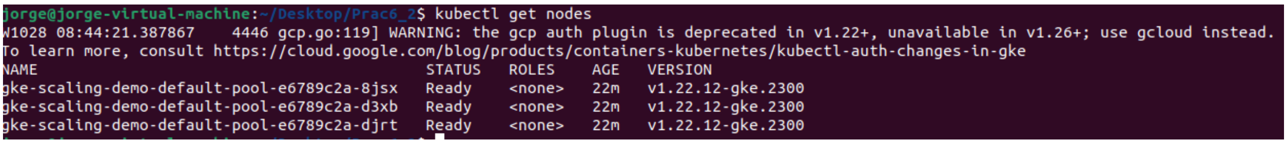
```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ gcloud container clusters update $cluster_name --enable-autoscaling --min-nodes=1 --max-nodes=5
Default change: During creation of nodepools or autoscaling configuration changes for cluster versions greater than 1.24.1-gke.800 a default location policy is applied. For Spot and PVM it defaults to A
Y, and for all other VM kinds a BALANCED policy is used. To change the default values use the `--location-policy` flag.
Updating scaling-demo...done.
Updated [https://container.googleapis.com/v1/projects/pract6-366611/zones/europe-west1-b/clusters/scaling-demo].
To inspect the contents of your cluster, go to: https://console.cloud.google.com/kubernetes/workload_/gcloud/europe-west1-b/scaling-demo?project=pract6-366611
```

11. cambiamos el perfil del autoescalado introduciendo "optimaze-utilization"



```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ gcloud beta container clusters update $cluster_name \
--autoscaling-profile optimize-utilization
Default change: During creation of nodepools or autoscaling configuration changes for cluster versions greater than 1.24.1-gke.800 a default location policy is
Y, and for all other VM kinds a BALANCED policy is used. To change the default values use the `--location-policy` flag.
Updating scaling-demo...done.
Updated [https://container.googleapis.com/v1beta1/projects/pract6-366611/zones/europe-west1-b/clusters/scaling-demo].
To inspect the contents of your cluster, go to: https://console.cloud.google.com/kubernetes/workload_/gcloud/europe-west1-b/scaling-demo?project=pract6-366611
```

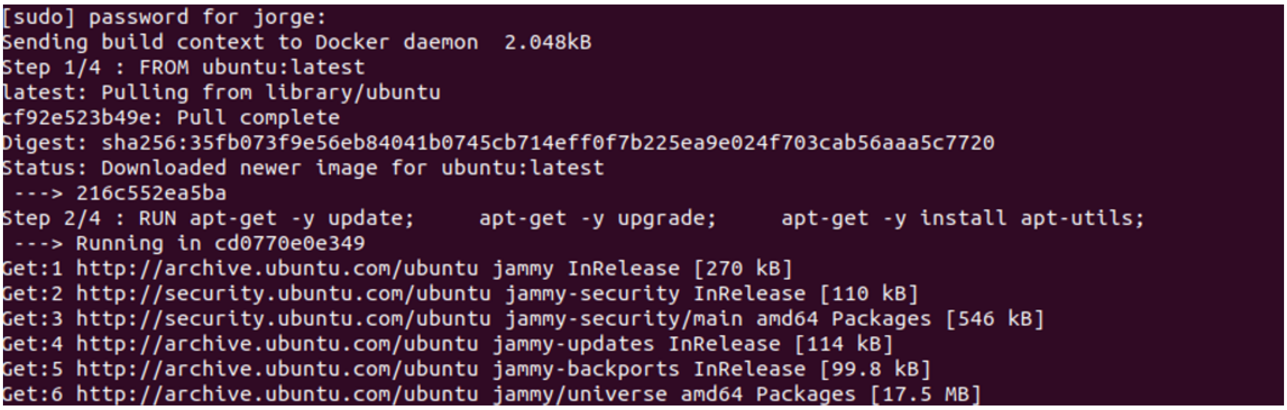
12. Nodos disponibles que podemos usar:



```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ kubectl get nodes
W1028 08:44:21.387867    4446 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
NAME                                         STATUS   ROLES    AGE   VERSION
gke-scaling-demo-default-pool-e6789c2a-8jsx   Ready    <none>   22m   v1.22.12-gke.2300
gke-scaling-demo-default-pool-e6789c2a-d3xb   Ready    <none>   22m   v1.22.12-gke.2300
gke-scaling-demo-default-pool-e6789c2a-djrt   Ready    <none>   22m   v1.22.12-gke.2300
```

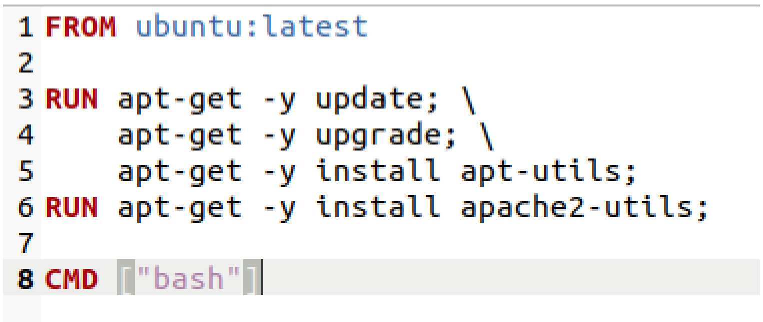
Entrega 2

1. En el directorio donde se encuentra el DockerFile creamos la imagen. Usamos el comando -> “sudo docker build -t gcr.io/pract6-366611/ab:v0.0.1 .”



```
[sudo] password for jorge:
Sending build context to Docker daemon  2.048kB
Step 1/4 : FROM ubuntu:latest
latest: Pulling from library/ubuntu
cf92e523b49e: Pull complete
Digest: sha256:35fb073f9e56eb84041b0745cb714eff0f7b225ea9e024f703cab56aaa5c7720
Status: Downloaded newer image for ubuntu:latest
 ---> 216c552ea5ba
Step 2/4 : RUN apt-get -y update;     apt-get -y upgrade;     apt-get -y install apt-utils;
 ---> Running in cd0770e0e349
Get:1 http://archive.ubuntu.com/ubuntu jammy InRelease [270 kB]
Get:2 http://security.ubuntu.com/ubuntu jammy-security InRelease [110 kB]
Get:3 http://security.ubuntu.com/ubuntu jammy-security/main amd64 Packages [546 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy-updates InRelease [114 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy-backports InRelease [99.8 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy/universe amd64 Packages [17.5 MB]
```

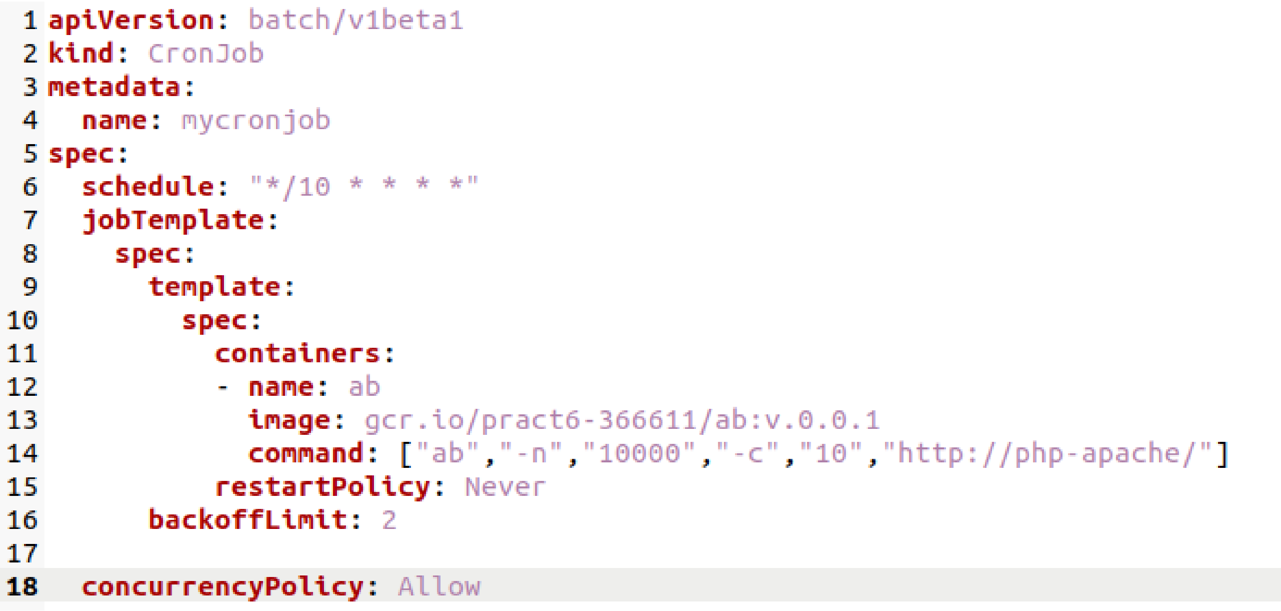
DockerFile



```
FROM ubuntu:latest

RUN apt-get -y update; \
    apt-get -y upgrade; \
    apt-get -y install apt-utils;
RUN apt-get -y install apache2-utils;

CMD ["bash"]
```

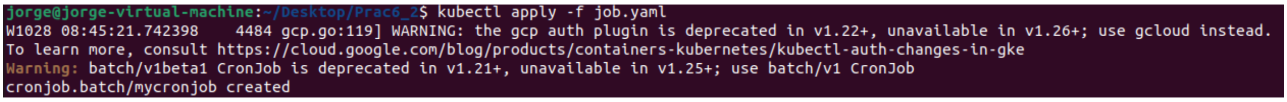
2. Pusheamos a Google la imagen con el comando “gcloud builds submit --tag gcr.io/pract6-366611/ab:v0.0.1 .”

3. Creamos el job.yaml



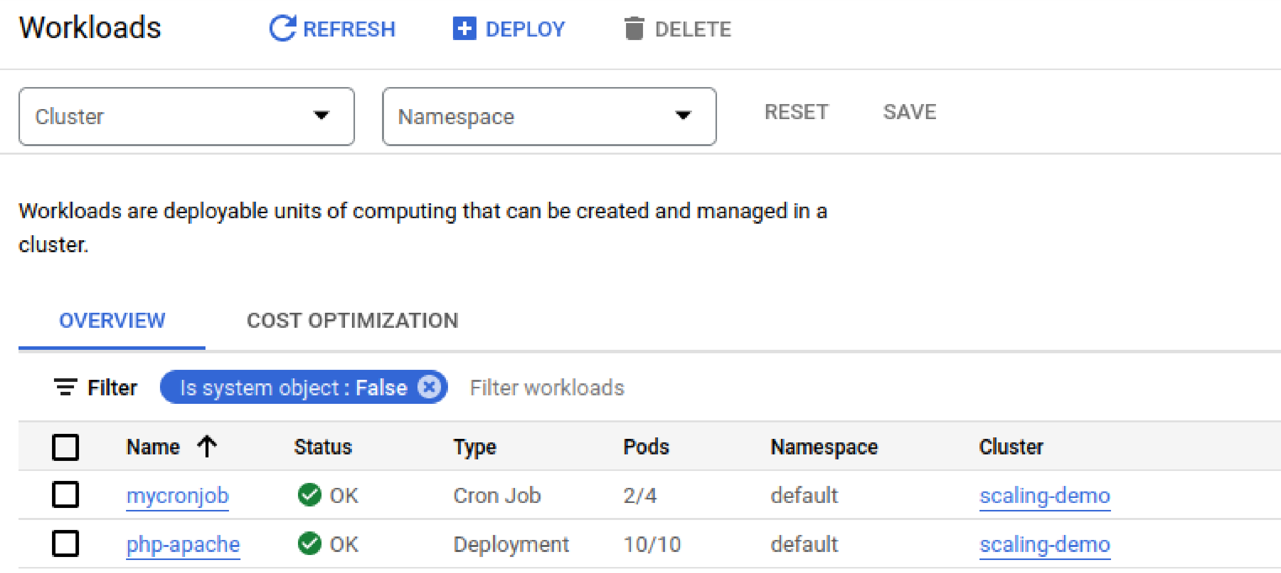
```yaml
apiVersion: batch/v1beta1
kind: CronJob
metadata:
  name: mycronjob
spec:
  schedule: "*/10 * * * *"
  jobTemplate:
    spec:
      template:
        spec:
          containers:
          - name: ab
            image: gcr.io/pract6-366611/ab:v.0.0.1
            command: ["ab","-n","10000","-c","10","http://php-apache/"]
          restartPolicy: Never
      backoffLimit: 2

  concurrencyPolicy: Allow
```

4. Comprobamos que funciona:



```
jorge@jorge-virtual-machine:~/Desktop/Prac6_2$ kubectl apply -f job.yaml
W1028 08:45:21.742398    4484 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
Warning: batch/v1beta1 CronJob is deprecated in v1.21+, unavailable in v1.25+; use batch/v1 CronJob
cronjob.batch/mycronjob created
```

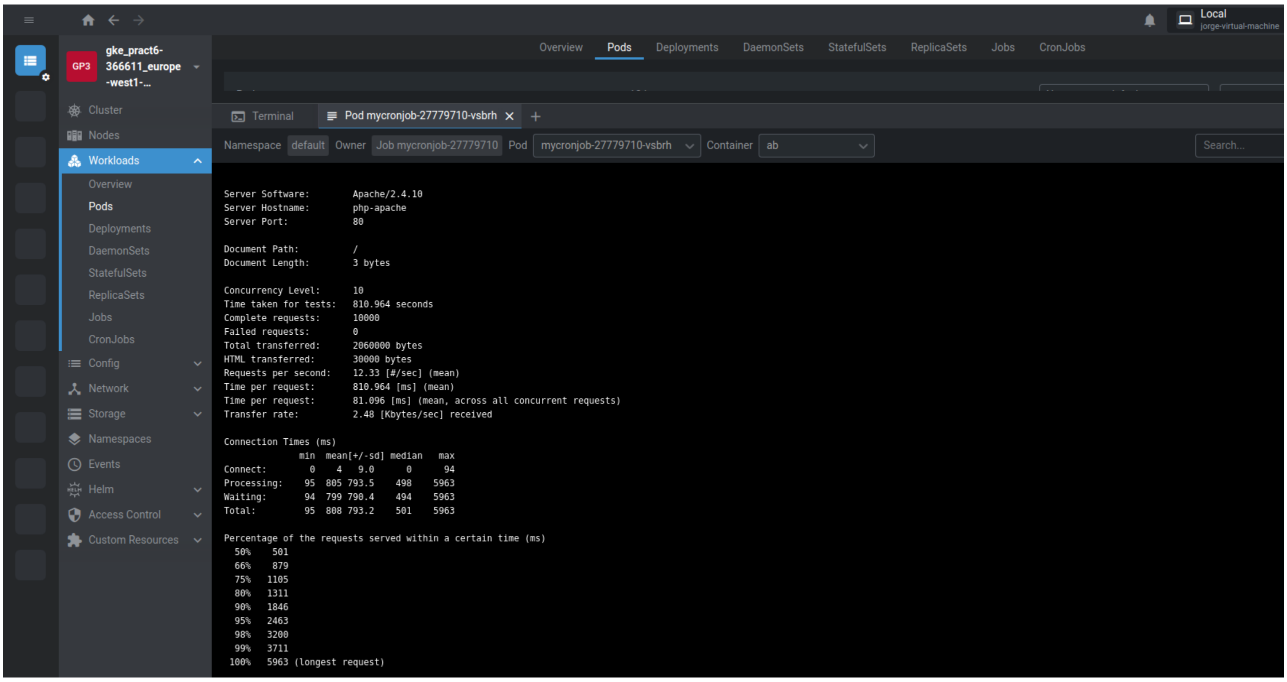
5. Cluster creado en Google:



Workloads are deployable units of computing that can be created and managed in a cluster.

□	Name ↑	Status	Type	Pods	Namespace	Cluster
□	mycronjob	✓ OK	Cron Job	2/4	default	scaling-demo
□	php-apache	✓ OK	Deployment	10/10	default	scaling-demo

6. Usando Lens sacamos información relevante sobre el cluster:



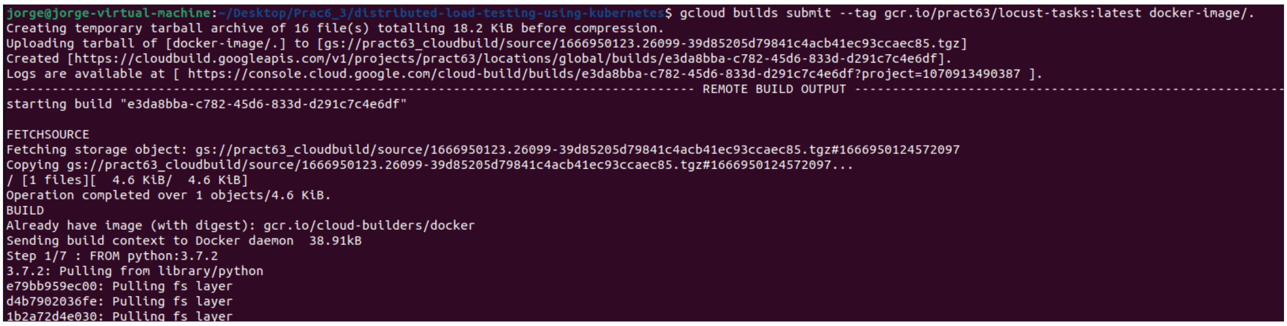
```
Server Software:        Apache/2.4.10
Server Hostname:        php-apache
Server Port:            80

Document Path:          /
Document Length:        3 bytes

Concurrency Level:      10
Time taken for tests:   810.964 seconds
Complete requests:      10000
Failed requests:        0
Total transferred:      2060000 bytes
HTML transferred:       30000 bytes
Requests per second:    12.33 [#/sec] (mean)
Time per request:       810.964 [ms] (mean)
Time per request:       81.096 [ms] (mean, across all concurrent requests)
Transfer rate:          2.48 [Kbytes/sec] received

Connection Times (ms)
              min  mean[+/-sd] median   max
Connect:        0    4   9.0      0      94
Processing:    95  805 793.5    498    5963
Waiting:       94  799 790.4    494    5963
Total:         95  808 793.2    501    5963

Percentage of the requests served within a certain time (ms)
  50%    501
  66%    879
  75%   1105
  80%   1311
  90%   1846
  95%   2463
  98%   3200
  99%   3711
 100%   5963 (longest request)
```

Entrega 3

1. Se crea el cluster igual que en la entrega 2 para un nuevo proyecto

Previamente a empezar con esta entrega hay que habilitar Cloud Build Kubernetes Engine Google App Engine Admin API Cloud Storage.

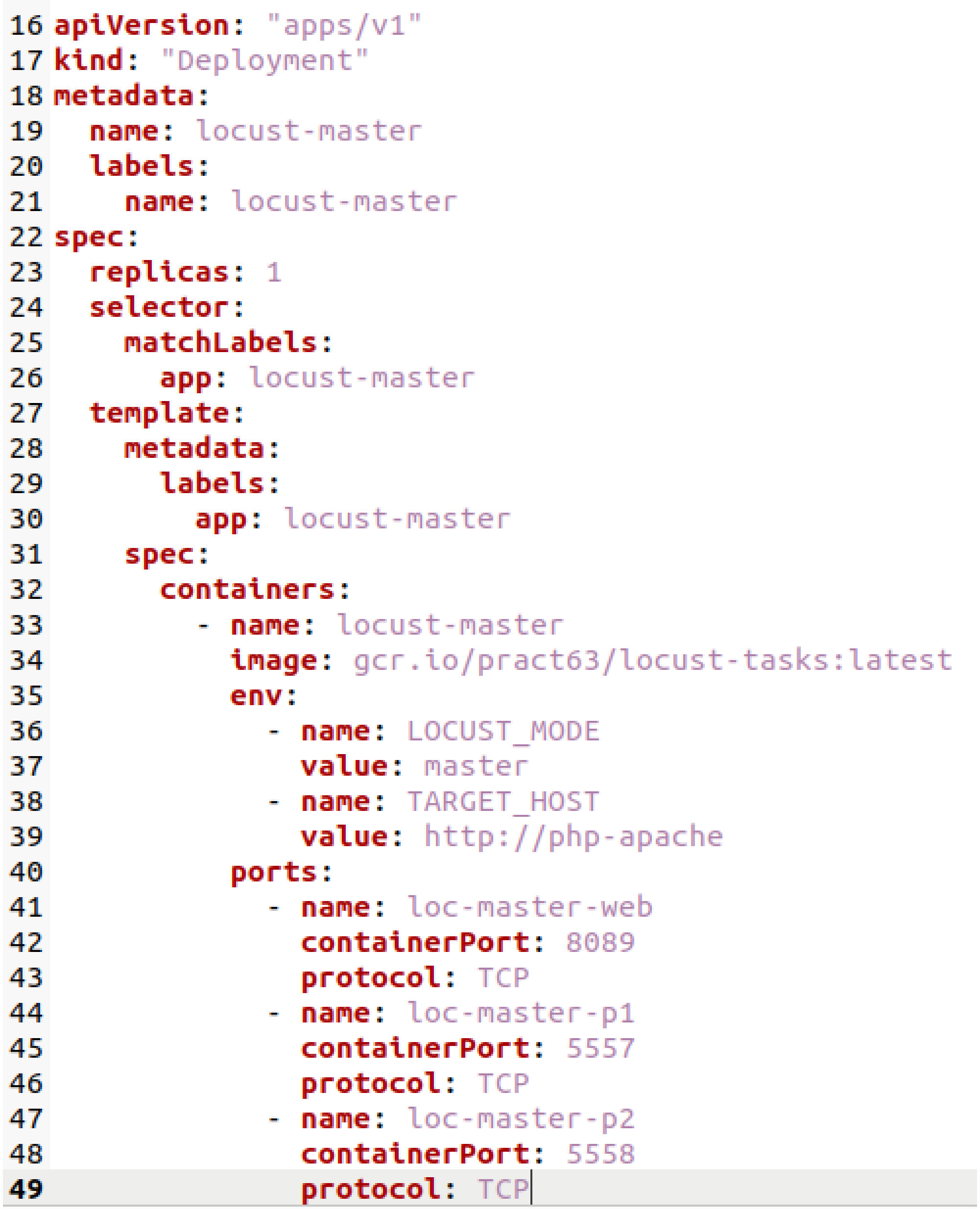
2. Construimos y subimos la imagen de Docker a gcloud



```
jorge@jorge-virtual-machine:~/Desktop/Prac6_3/distributed-load-testing-using-kubernetes$ gcloud builds submit --tag gcr.io/pract63/locust-tasks:latest docker-image/.
Creating temporary tarball archive of 16 file(s) totalling 18.2 KiB before compression.
Uploading tarball of [docker-image/.] to [gs://pract63_cloudbuild/source/1666950123.26099-39d85205d79841c4acb41ec93ccaec85.tgz]
Created [https://cloudbuild.googleapis.com/v1/projects/pract63/locations/global/builds/e3da8bba-c782-45d6-833d-d291c7c4e6df].
Logs are available at [ https://console.cloud.google.com/cloud-build/builds/e3da8bba-c782-45d6-833d-d291c7c4e6df?project=1070913490387 ].
------------------------------------------------------------------------------------------ REMOTE BUILD OUTPUT ---------------------------------------------------------------------
starting build "e3da8bba-c782-45d6-833d-d291c7c4e6df"

FETCHSOURCE
Fetching storage object: gs://pract63_cloudbuild/source/1666950123.26099-39d85205d79841c4acb41ec93ccaec85.tgz#1666950124572097
Copying gs://pract63_cloudbuild/source/1666950123.26099-39d85205d79841c4acb41ec93ccaec85.tgz#1666950124572097...
/ [1 files][  4.6 KiB/  4.6 KiB]
Operation completed over 1 objects/4.6 KiB.
BUILD
Already have image (with digest): gcr.io/cloud-builders/docker
Sending build context to Docker daemon  38.91kB
Step 1/7 : FROM python:3.7.2
3.7.2: Pulling from library/python
e79bb959ec00: Pulling fs layer
d4b7902036fe: Pulling fs layer
1b2a72d4e030: Pulling fs layer
```

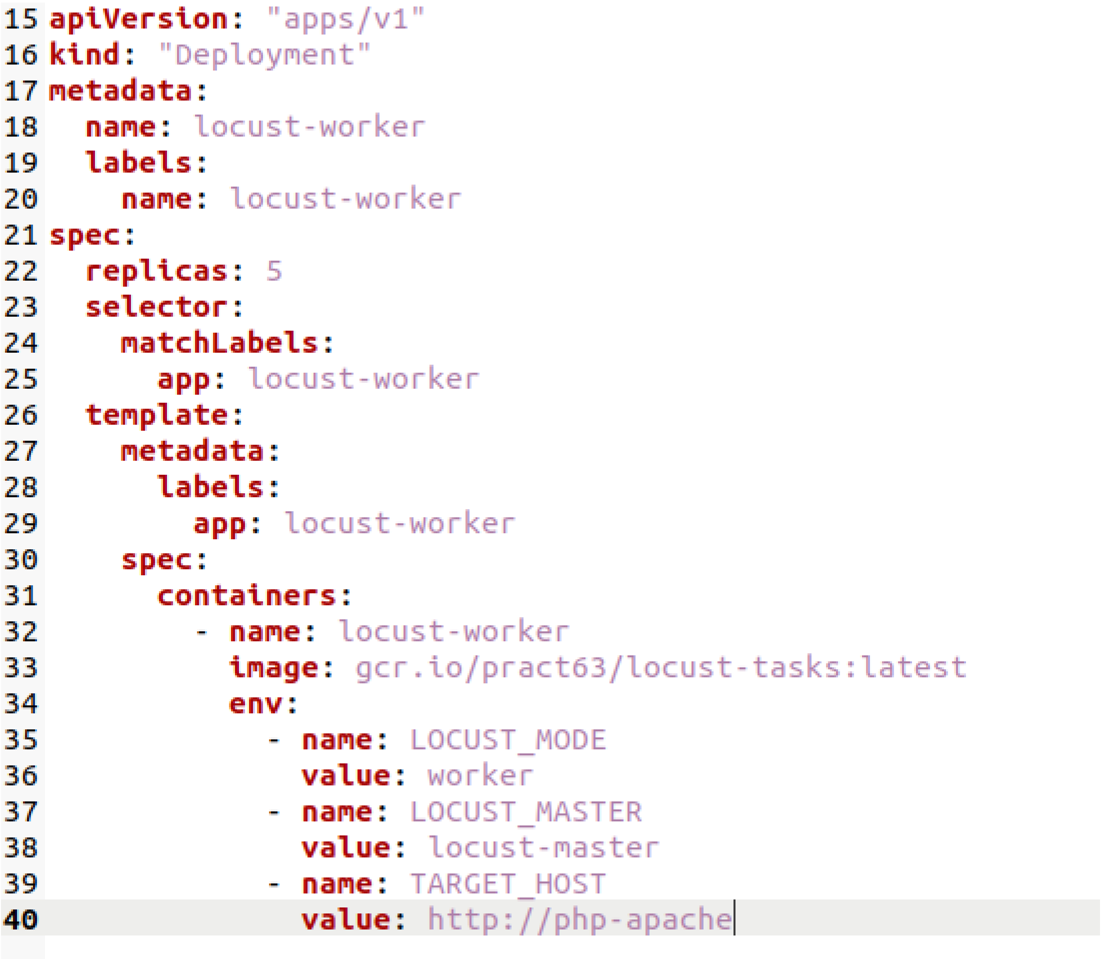
3. Cambiamos el [Target Host] -> http://php-apache y el nombre del proyecto id del proyecto de los archivos locust-master-controller.yaml y locust-worker-controller.yaml

   Master Controler



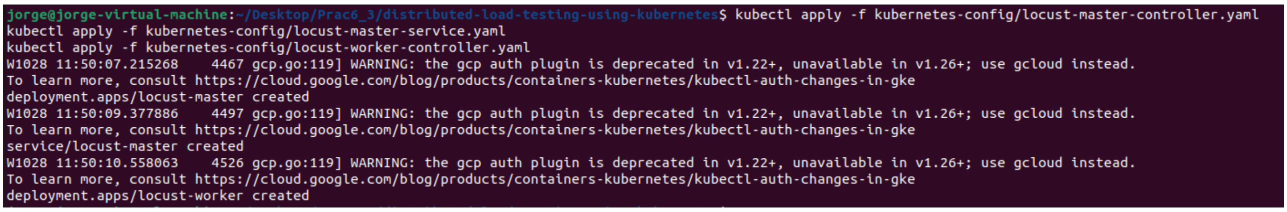
```yaml
apiVersion: "apps/v1"
kind: "Deployment"
metadata:
  name: locust-master
  labels:
    name: locust-master
spec:
  replicas: 1
  selector:
    matchLabels:
      app: locust-master
  template:
    metadata:
      labels:
        app: locust-master
    spec:
      containers:
        - name: locust-master
          image: gcr.io/pract63/locust-tasks:latest
          env:
            - name: LOCUST_MODE
              value: master
            - name: TARGET_HOST
              value: http://php-apache
          ports:
            - name: loc-master-web
              containerPort: 8089
              protocol: TCP
            - name: loc-master-p1
              containerPort: 5557
              protocol: TCP
            - name: loc-master-p2
              containerPort: 5558
              protocol: TCP
```

Worker controller



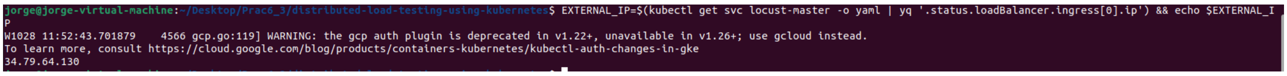
```yaml
apiVersion: "apps/v1"
kind: "Deployment"
metadata:
  name: locust-worker
  labels:
    name: locust-worker
spec:
  replicas: 5
  selector:
    matchLabels:
      app: locust-worker
  template:
    metadata:
      labels:
        app: locust-worker
    spec:
      containers:
        - name: locust-worker
          image: gcr.io/pract63/locust-tasks:latest
          env:
            - name: LOCUST_MODE
              value: worker
            - name: LOCUST_MASTER
              value: locust-master
            - name: TARGET_HOST
              value: http://php-apache
```

4. Levantamos los nodos de locust master y locust worker

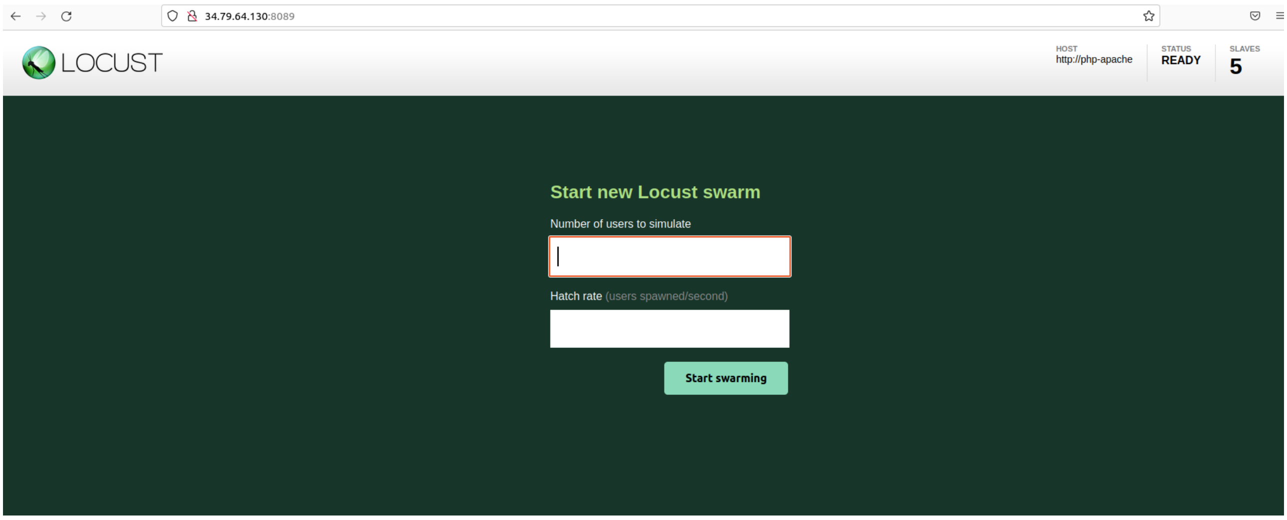


```
jorge@jorge-virtual-machine:~/Desktop/Prac6_3/distributed-load-testing-using-kubernetes$ kubectl apply -f kubernetes-config/locust-master-controller.yaml
kubectl apply -f kubernetes-config/locust-master-service.yaml
kubectl apply -f kubernetes-config/locust-worker-controller.yaml
W1028 11:50:07.215268    4467 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
deployment.apps/locust-master created
W1028 11:50:09.377886    4497 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
service/locust-master created
W1028 11:50:10.558063    4526 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
deployment.apps/locust-worker created
```

5. Obtenemos la dirección IP externa del servicio locust master y la introducimos en el buscador.



```
jorge@jorge-virtual-machine:~/Desktop/Prac6_3/distributed-load-testing-using-kubernetes$ EXTERNAL_IP=$(kubectl get svc locust-master -o yaml | yq '.status.loadBalancer.ingress[0].ip') && echo $EXTERNAL_I
P
W1028 11:52:43.701879    4566 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
34.79.64.130
```



34.79.64.130:8089

HOST http://php-apache

STATUS READY

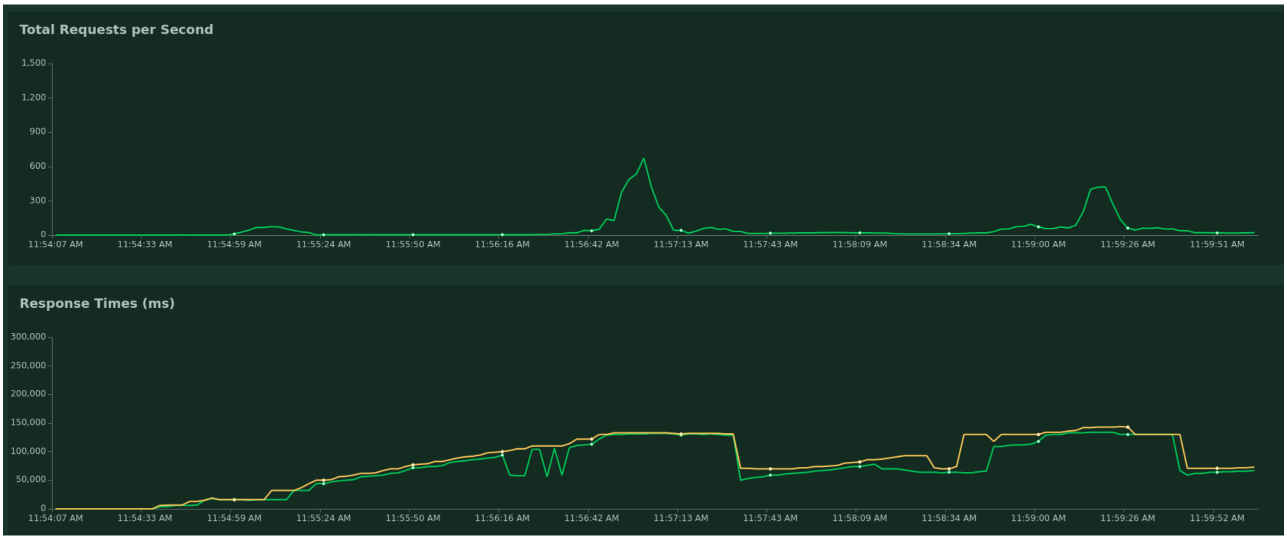
SLAVES 5

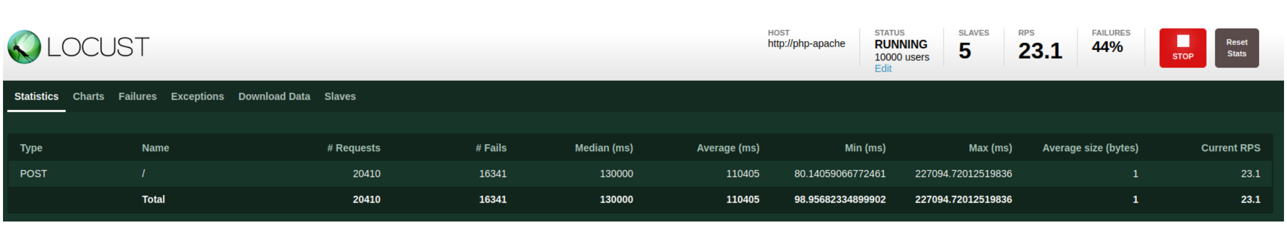
Start new Locust swarm

Number of users to simulate

Hatch rate (users spawned/second)

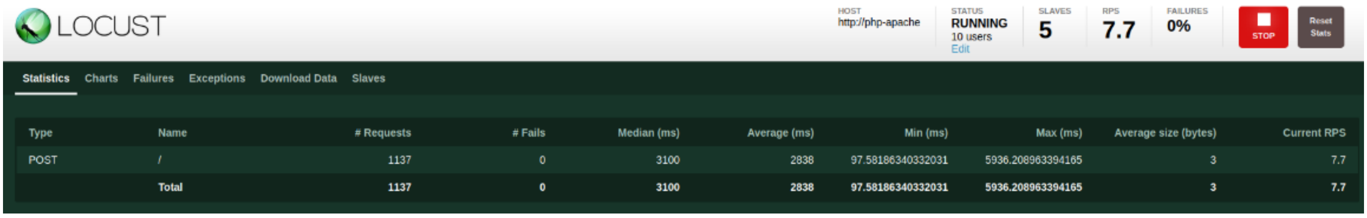
En un primer caso para 10 pods máximos establezco 10000 users y un Hatch rate de 1000.



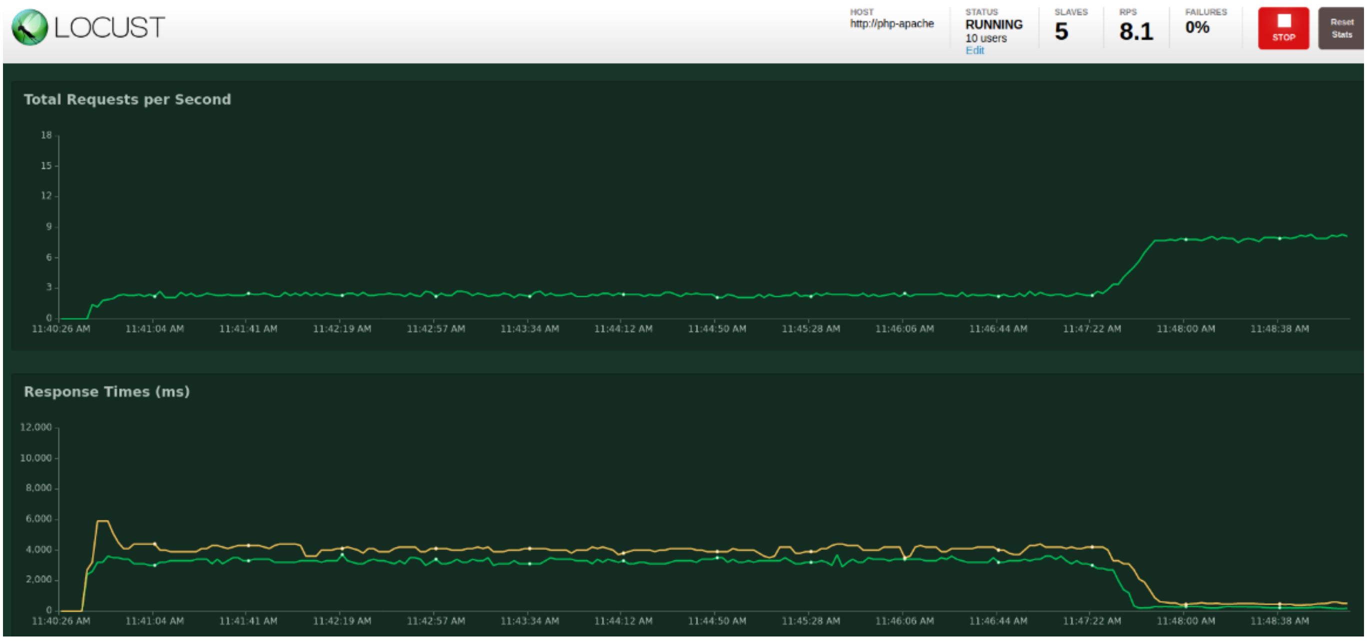


Type	Name	# Requests	# Fails	Median (ms)	Average (ms)	Min (ms)	Max (ms)	Average size (bytes)	Current RPS
POST	/	20410	16341	130000	110405	80.14059066772461	227094.72012519836	1	23.1
	Total	**20410**	**16341**	**130000**	**110405**	**98.95682334899902**	**227094.72012519836**	**1**	**23.1**

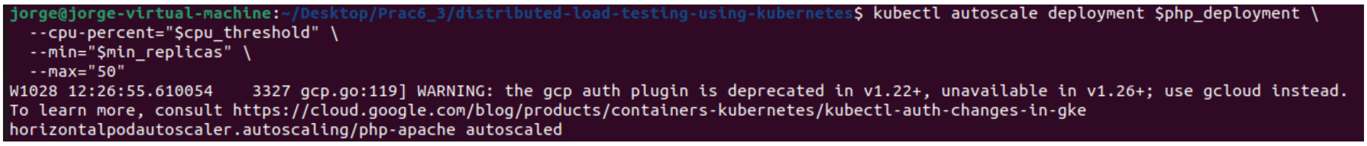
Para otro caso de 10 pods máximos establezco 10 users y un Hatch rate de 2.



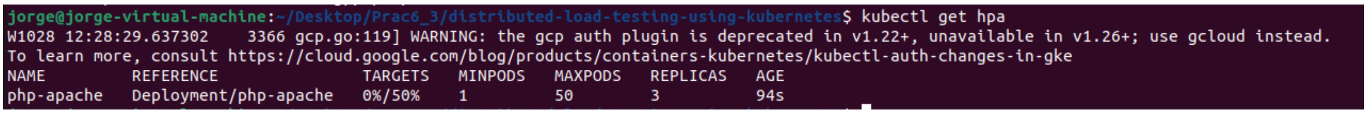
Type	Name	# Requests	# Fails	Median (ms)	Average (ms)	Min (ms)	Max (ms)	Average size (bytes)	Current RPS
POST	/	1137	0	3100	2838	97.58186340332031	5936.208963394165	3	7.7
	Total	**1137**	**0**	**3100**	**2838**	**97.58186340332031**	**5936.208963394165**	**3**	**7.7**



Ahora levanto otra vez el cluster y cambio el número máximo de réplicas

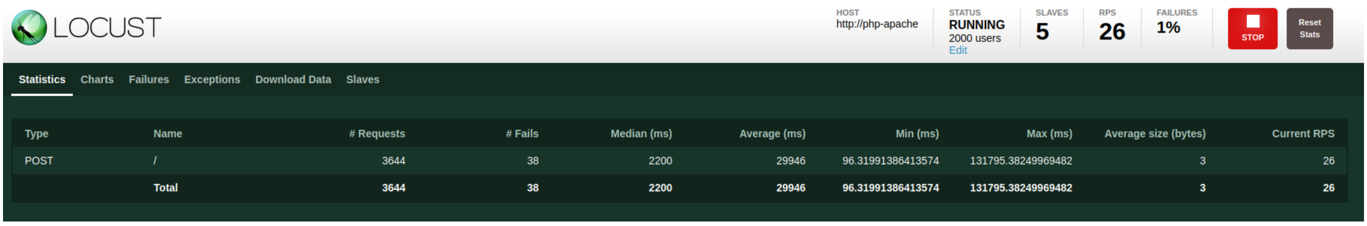


```
jorge@jorge-virtual-machine:~/Desktop/Prac6_3/distributed-load-testing-using-kubernetes$ kubectl autoscale deployment $php_deployment \
   --cpu-percent="$cpu_threshold" \
   --min="$min_replicas" \
   --max="50"
W1028 12:26:55.610054    3327 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
horizontalpodautoscaler.autoscaling/php-apache autoscaled
```

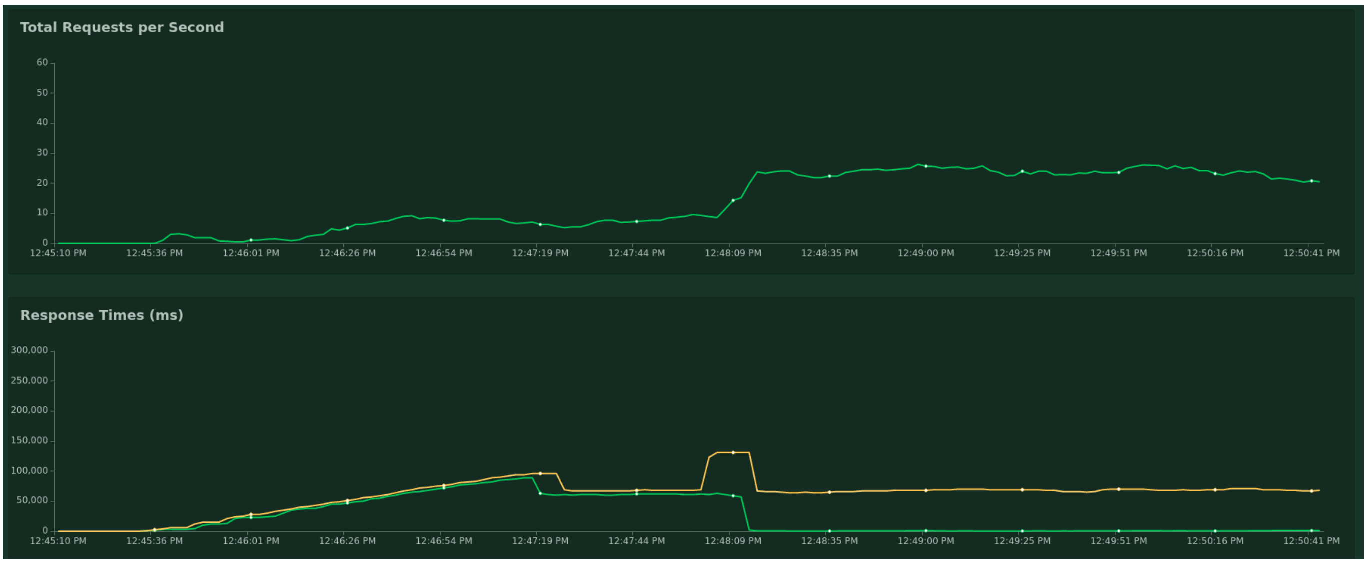


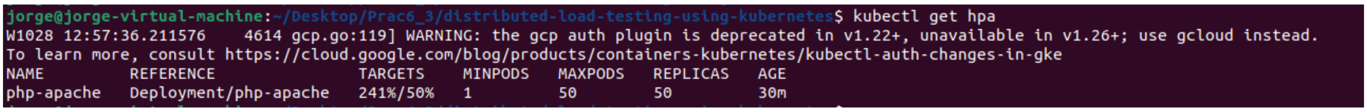
```
jorge@jorge-virtual-machine:~/Desktop/Prac6_3/distributed-load-testing-using-kubernetes$ kubectl get hpa
W1028 12:28:29.637302    3366 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
NAME         REFERENCE               TARGETS   MINPODS   MAXPODS   REPLICAS   AGE
php-apache   Deployment/php-apache   0%/50%    1         50        3          94s
```

Introducimos 2000 ususarios y Hatch rate de 100 con un máximo de 50 pods



Type	Name	# Requests	# Fails	Median (ms)	Average (ms)	Min (ms)	Max (ms)	Average size (bytes)	Current RPS
POST	/	3644	38	2200	29946	96.31991386413574	131795.38249969482	3	26
	Total	**3644**	**38**	**2200**	**29946**	**96.31991386413574**	**131795.38249969482**	**3**	**26**





```
jorge@jorge-virtual-machine:~/Desktop/Prac6_3/distributed-load-testing-using-kubernetes$ kubectl get hpa
W1028 12:57:36.211576    4614 gcp.go:119] WARNING: the gcp auth plugin is deprecated in v1.22+, unavailable in v1.26+; use gcloud instead.
To learn more, consult https://cloud.google.com/blog/products/containers-kubernetes/kubectl-auth-changes-in-gke
NAME         REFERENCE               TARGETS    MINPODS   MAXPODS   REPLICAS   AGE
php-apache   Deployment/php-apache   241%/50%   1         50        50         30m
```

Se llega a la conclusión de que a cuantos más pods se usan más request por segundo (RPS).

Además, también disminuye el porcentaje de error cuando el número de pods es más alto.