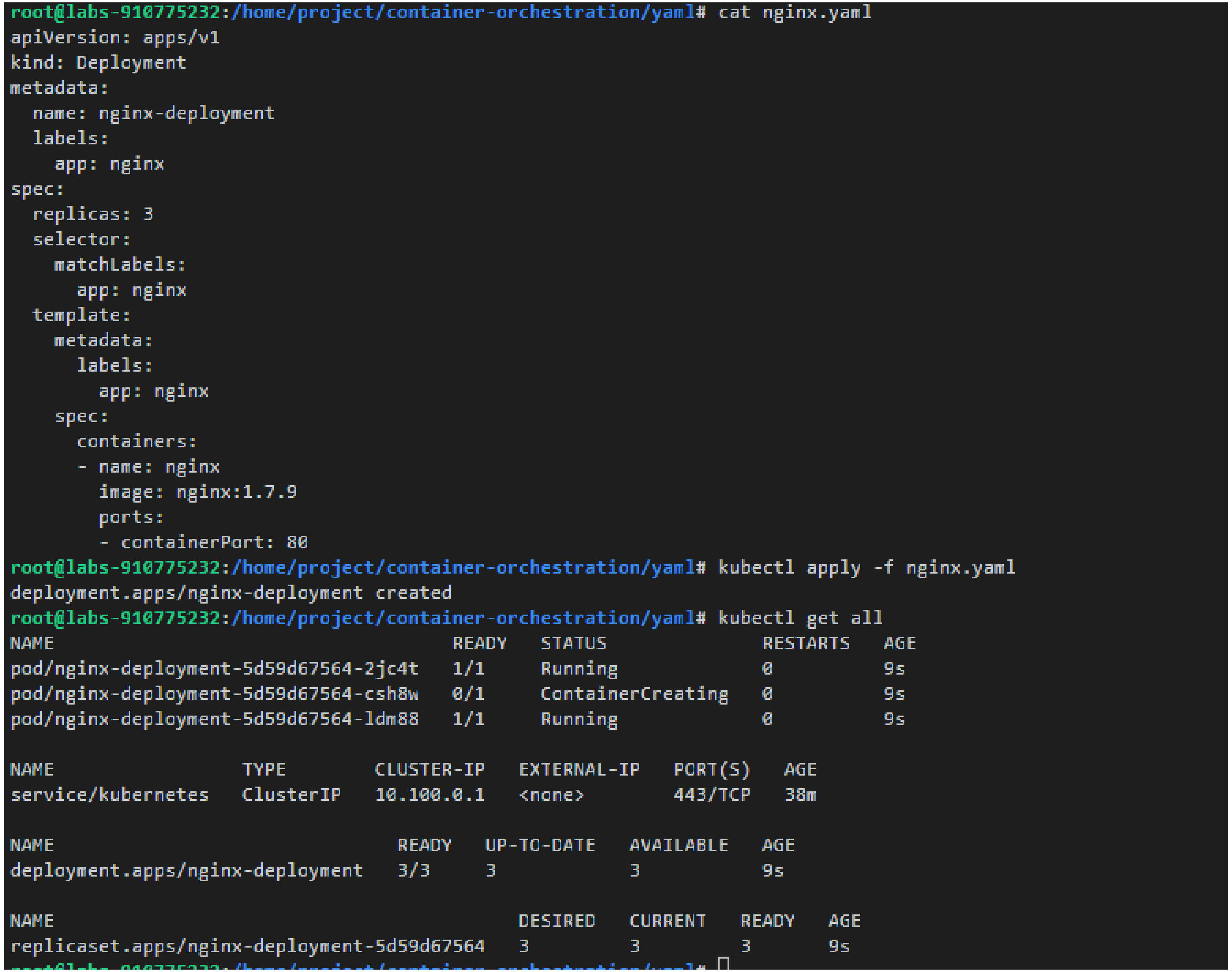
nginx.yaml 을 보면 기본 replica는 3개로 설정되어 있음.

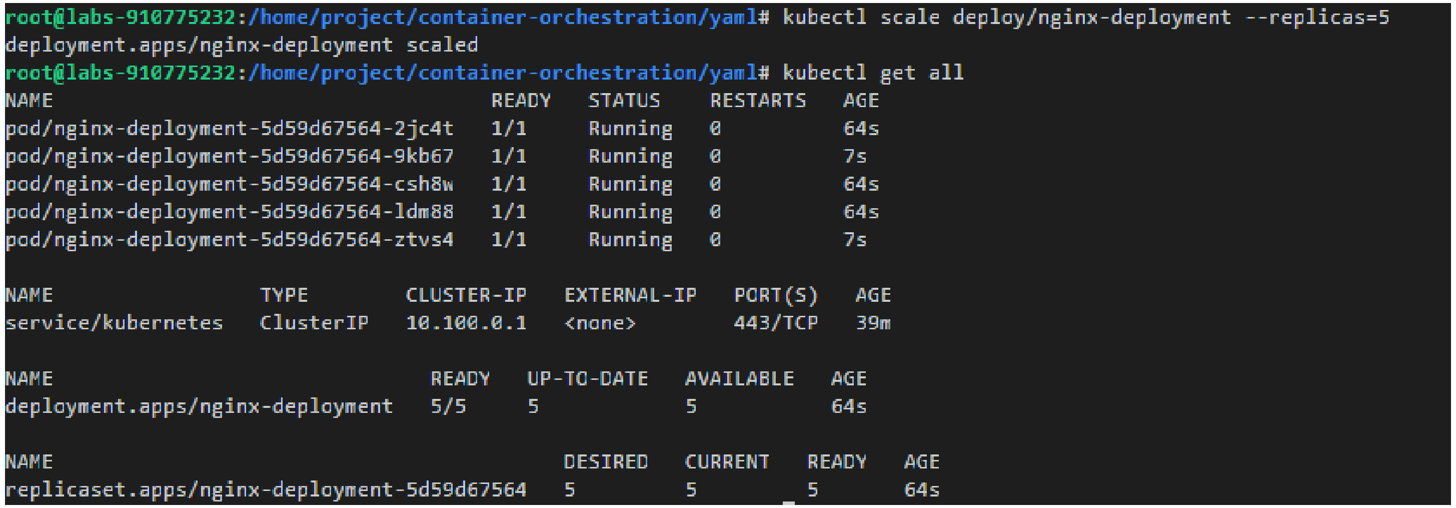


```
root@labs-910775232:/home/project/container-orchestration/yaml# cat nginx.yaml
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx-deployment
  labels:
    app: nginx
spec:
  replicas: 3
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
    spec:
      containers:
      - name: nginx
        image: nginx:1.7.9
        ports:
        - containerPort: 80
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl apply -f nginx.yaml
deployment.apps/nginx-deployment created
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get all
NAME                                    READY   STATUS              RESTARTS   AGE
pod/nginx-deployment-5d59d67564-2jc4t   1/1     Running             0          9s
pod/nginx-deployment-5d59d67564-csh8w   0/1     ContainerCreating   0          9s
pod/nginx-deployment-5d59d67564-ldm88   1/1     Running             0          9s

NAME                 TYPE        CLUSTER-IP   EXTERNAL-IP   PORT(S)   AGE
service/kubernetes   ClusterIP   10.100.0.1   <none>        443/TCP   38m

NAME                               READY   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/nginx-deployment   3/3     3            3           9s

NAME                                          DESIRED   CURRENT   READY   AGE
replicaset.apps/nginx-deployment-5d59d67564   3         3         3       9s
root@labs-910775232:/home/project/container-orchestration/yaml# 
```

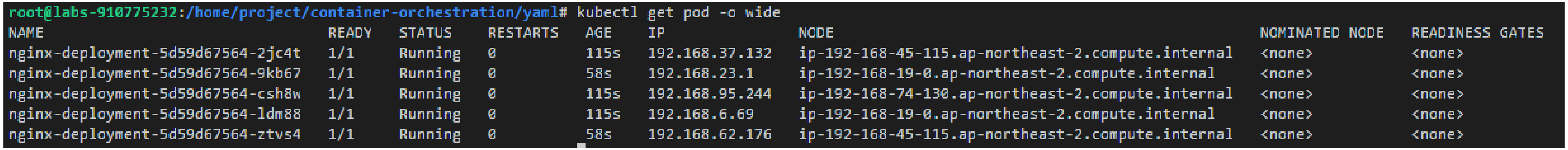
Yaml 파일에 작성한 대로 3개의 pod 생성 완료

Pod Scaling

```
kubectl scale deploy/nginx-deployment --replicas=5
kubectl get all
```



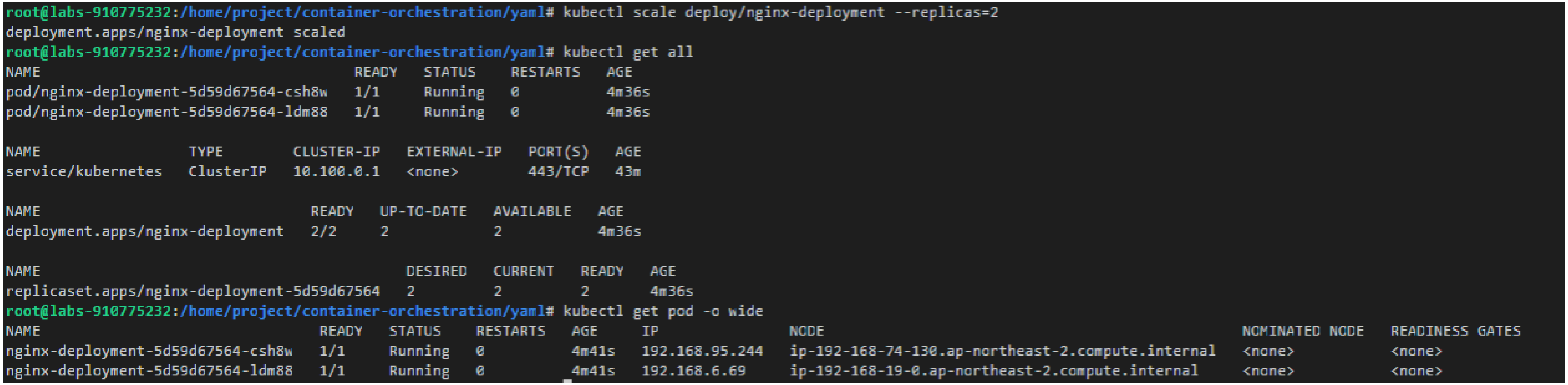
```
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl scale deploy/nginx-deployment --replicas=5
deployment.apps/nginx-deployment scaled
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get all
NAME                                    READY   STATUS    RESTARTS   AGE
pod/nginx-deployment-5d59d67564-2jc4t   1/1     Running   0          64s
pod/nginx-deployment-5d59d67564-9kb67   1/1     Running   0          7s
pod/nginx-deployment-5d59d67564-csh8w   1/1     Running   0          64s
pod/nginx-deployment-5d59d67564-ldm88   1/1     Running   0          64s
pod/nginx-deployment-5d59d67564-ztvs4   1/1     Running   0          7s

NAME                 TYPE        CLUSTER-IP   EXTERNAL-IP   PORT(S)   AGE
service/kubernetes   ClusterIP   10.100.0.1   <none>        443/TCP   39m

NAME                               READY   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/nginx-deployment   5/5     5            5           64s

NAME                                          DESIRED   CURRENT   READY   AGE
replicaset.apps/nginx-deployment-5d59d67564   5         5         5       64s
```



```
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get pod -o wide
NAME                                READY   STATUS    RESTARTS   AGE    IP               NODE                                                NOMINATED NODE   READINESS GATES
nginx-deployment-5d59d67564-2jc4t   1/1     Running   0          115s   192.168.37.132   ip-192-168-45-115.ap-northeast-2.compute.internal   <none>           <none>
nginx-deployment-5d59d67564-9kb67   1/1     Running   0          58s    192.168.23.1     ip-192-168-19-0.ap-northeast-2.compute.internal     <none>           <none>
nginx-deployment-5d59d67564-csh8w   1/1     Running   0          115s   192.168.95.244   ip-192-168-74-130.ap-northeast-2.compute.internal   <none>           <none>
nginx-deployment-5d59d67564-ldm88   1/1     Running   0          115s   192.168.6.69     ip-192-168-19-0.ap-northeast-2.compute.internal     <none>           <none>
nginx-deployment-5d59d67564-ztvs4   1/1     Running   0          58s    192.168.62.176   ip-192-168-45-115.ap-northeast-2.compute.internal   <none>           <none>
```

Pod Scaling

```
kubectl scale deploy/nginx-deployment --replicas=2
kubectl get all
```

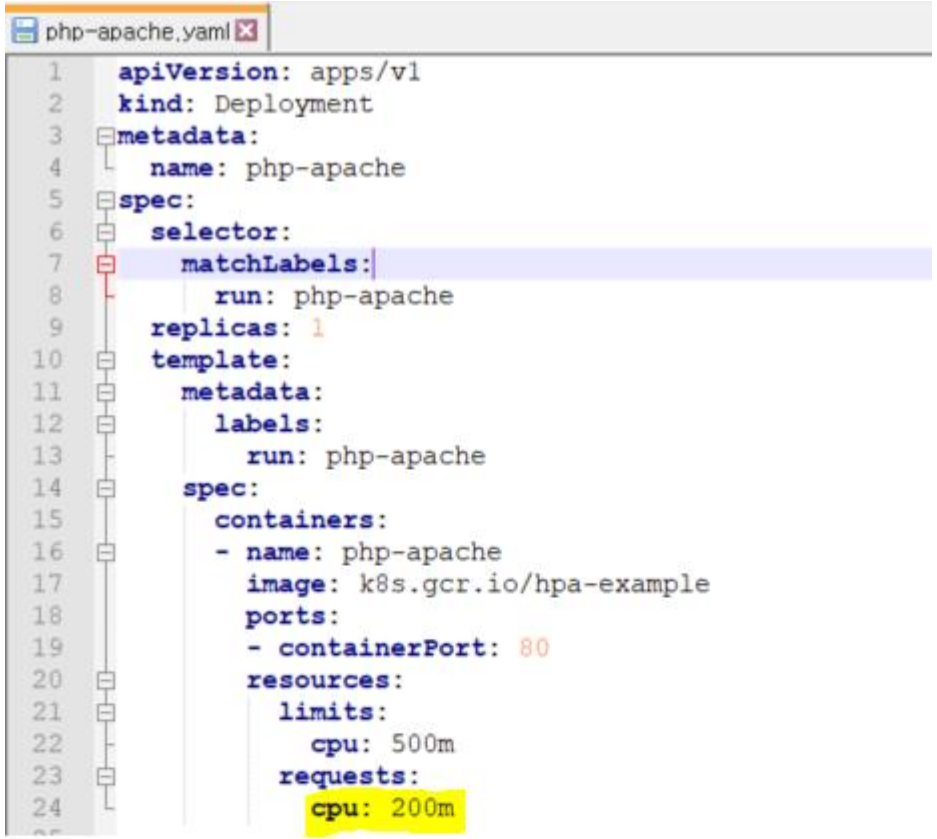


```
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl scale deploy/nginx-deployment --replicas=2
deployment.apps/nginx-deployment scaled
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get all
NAME                                    READY   STATUS    RESTARTS   AGE
pod/nginx-deployment-5d59d67564-csh8w   1/1     Running   0          4m36s
pod/nginx-deployment-5d59d67564-ldm88   1/1     Running   0          4m36s

NAME                 TYPE        CLUSTER-IP   EXTERNAL-IP   PORT(S)   AGE
service/kubernetes   ClusterIP   10.100.0.1   <none>        443/TCP   43m

NAME                               READY   UP-TO-DATE   AVAILABLE   AGE
deployment.apps/nginx-deployment   2/2     2            2           4m36s

NAME                                          DESIRED   CURRENT   READY   AGE
replicaset.apps/nginx-deployment-5d59d67564   2         2         2       4m36s
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get pod -o wide
NAME                                READY   STATUS    RESTARTS   AGE     IP               NODE                                                NOMINATED NODE   READINESS GATES
nginx-deployment-5d59d67564-csh8w   1/1     Running   0          4m41s   192.168.95.244   ip-192-168-74-130.ap-northeast-2.compute.internal   <none>           <none>
nginx-deployment-5d59d67564-ldm88   1/1     Running   0          4m41s   192.168.6.69     ip-192-168-19-0.ap-northeast-2.compute.internal     <none>           <none>
```

오토 스케일링 설정, hpa: HorizontalPodAutoscaler

```
kubectl autoscale deployment php-apache --cpu-percent=50 --min=1 --max=10
```

최소 1개에서 10개까지 Pod 의 개수 조정가능



php-apache.yaml

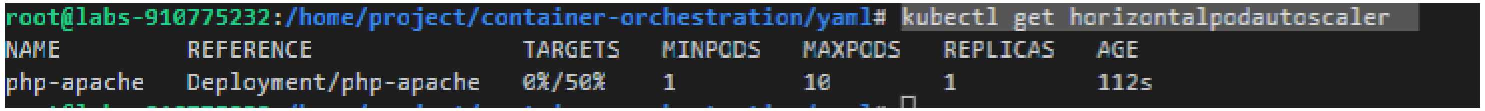
```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: php-apache
spec:
  selector:
    matchLabels:
      run: php-apache
  replicas: 1
  template:
    metadata:
      labels:
        run: php-apache
    spec:
      containers:
      - name: php-apache
        image: k8s.gcr.io/hpa-example
        ports:
        - containerPort: 80
        resources:
          limits:
            cpu: 500m
          requests:
            cpu: 200m
```

(Pod의 평균 CPU 사용율이 100 milli-cores(50%)를 넘게되면 HPA 발생)

Yaml 기본 설정이 200milli-core로 설정되어있음

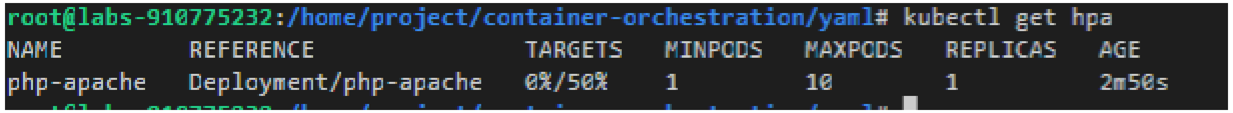
오토 스케일러 확인하기

```
kubectl get horizontalpodautoscaler
```



```
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get horizontalpodautoscaler
NAME         REFERENCE               TARGETS   MINPODS   MAXPODS   REPLICAS   AGE
php-apache   Deployment/php-apache   0%/50%    1         10        1          112s
```

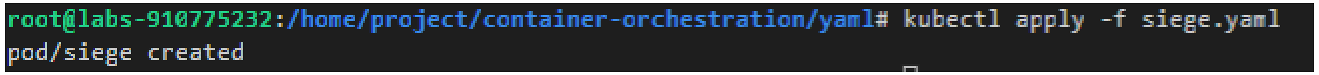
kubectl get hpa ➔ 약어로 hpa라고 씀.



```
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl get hpa
NAME         REFERENCE               TARGETS   MINPODS   MAXPODS   REPLICAS   AGE
php-apache   Deployment/php-apache   0%/50%    1         10        1          2m50s
```

로드제네레이터 설치

kubectl apply -f siege.yaml



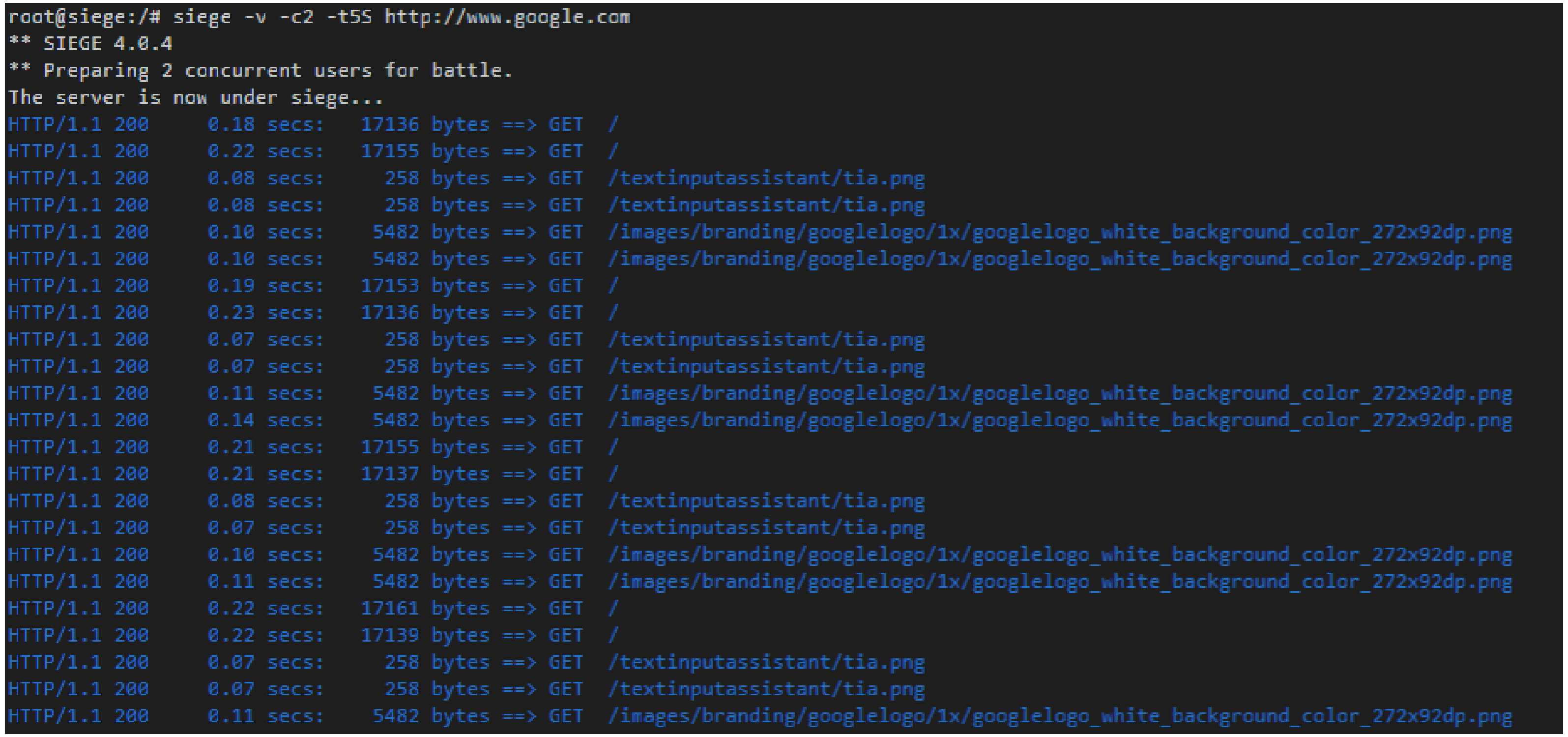
```
root@labs-910775232:/home/project/container-orchestration/yaml# kubectl apply -f siege.yaml
pod/siege created
```

root@siege:/# 접속

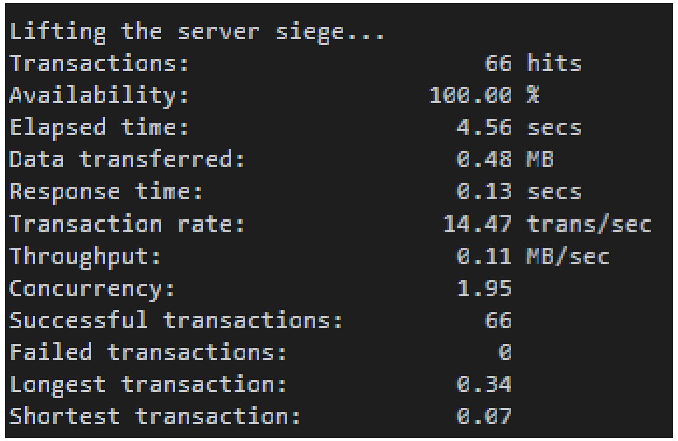
kubectl exec -it siege -- /bin/bash

부하 테스트 예제

siege -v -c2 -t5S http://www.google.com 5초 동안 타켓에 동접자 2명으로 스레드 생성하는 것을 보여주겠다.



```
root@siege:/# siege -v -c2 -t5S http://www.google.com
** SIEGE 4.0.4
** Preparing 2 concurrent users for battle.
The server is now under siege...
HTTP/1.1 200     0.18 secs:   17136 bytes ==> GET  /
HTTP/1.1 200     0.22 secs:   17155 bytes ==> GET  /
HTTP/1.1 200     0.08 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.08 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.10 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
HTTP/1.1 200     0.10 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
HTTP/1.1 200     0.19 secs:   17153 bytes ==> GET  /
HTTP/1.1 200     0.23 secs:   17136 bytes ==> GET  /
HTTP/1.1 200     0.07 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.07 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.11 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
HTTP/1.1 200     0.14 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
HTTP/1.1 200     0.21 secs:   17155 bytes ==> GET  /
HTTP/1.1 200     0.21 secs:   17137 bytes ==> GET  /
HTTP/1.1 200     0.08 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.07 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.10 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
HTTP/1.1 200     0.11 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
HTTP/1.1 200     0.22 secs:   17161 bytes ==> GET  /
HTTP/1.1 200     0.22 secs:   17139 bytes ==> GET  /
HTTP/1.1 200     0.07 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.07 secs:     258 bytes ==> GET  /textinputassistant/tia.png
HTTP/1.1 200     0.11 secs:    5482 bytes ==> GET  /images/branding/googlelogo/1x/googlelogo_white_background_color_272x92dp.png
```

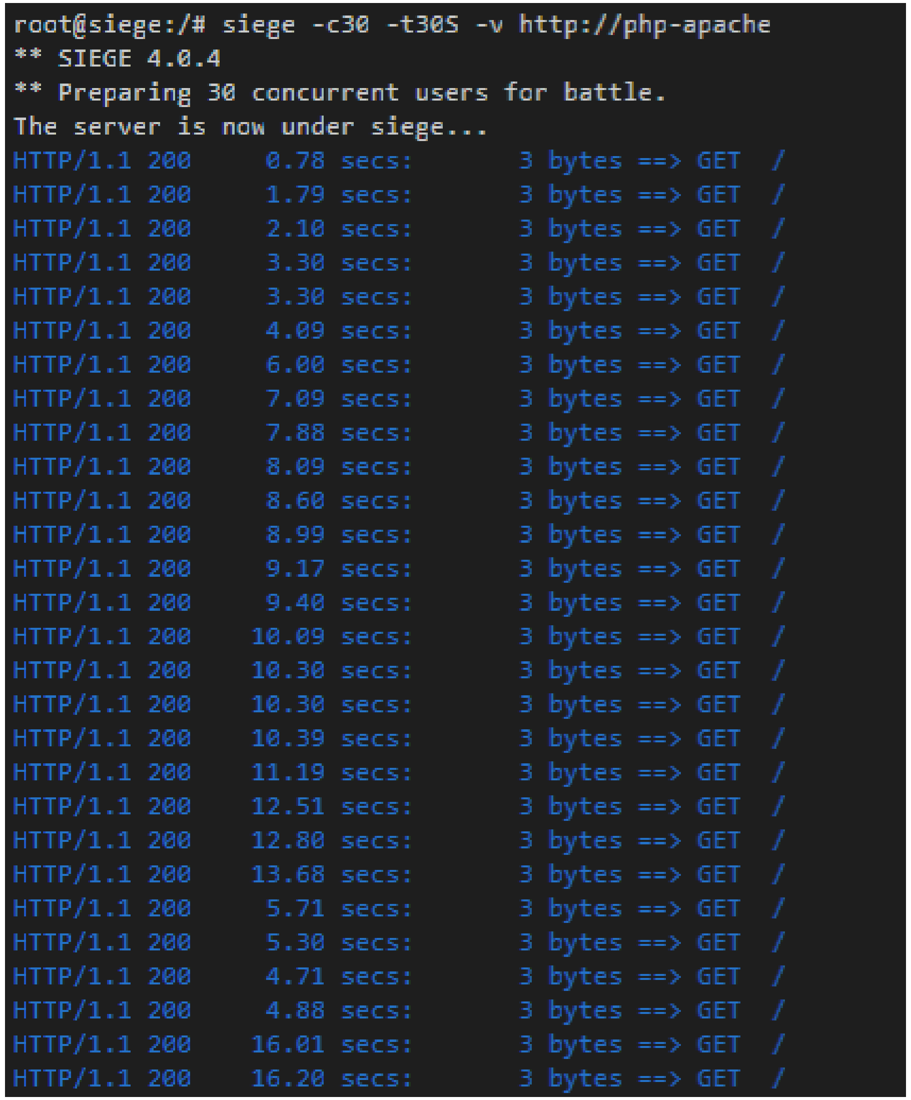


```
Lifting the server siege...
Transactions:                     66 hits
Availability:                 100.00 %
Elapsed time:                   4.56 secs
Data transferred:               0.48 MB
Response time:                  0.13 secs
Transaction rate:              14.47 trans/sec
Throughput:                     0.11 MB/sec
Concurrency:                    1.95
Successful transactions:          66
Failed transactions:               0
Longest transaction:            0.34
Shortest transaction:           0.07
```

부하 테스트

```
siege
siege -c30 -t30S -v http://php-apache
```

부하 발생 테스트



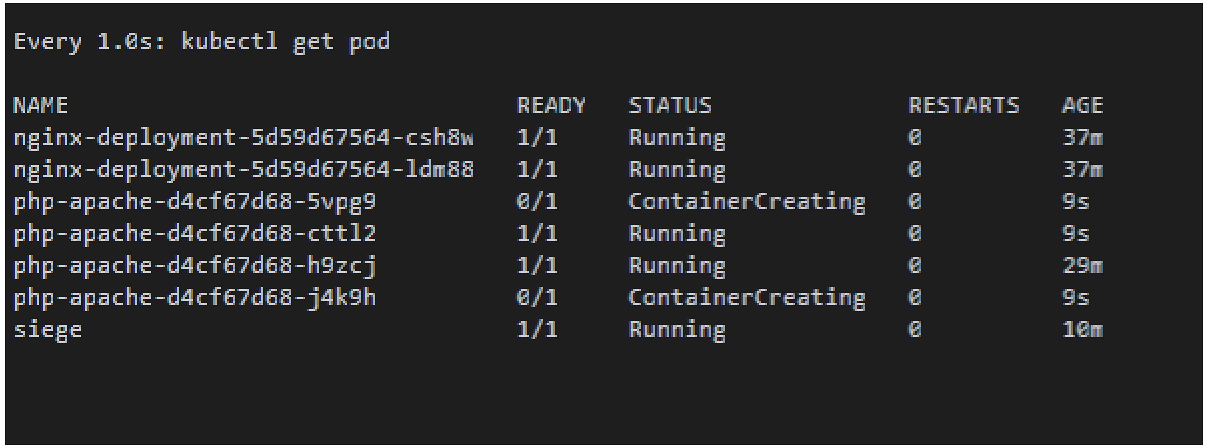
```
root@siege:/# siege -c30 -t30S -v http://php-apache
** SIEGE 4.0.4
** Preparing 30 concurrent users for battle.
The server is now under siege...
HTTP/1.1 200     0.78 secs:       3 bytes ==> GET  /
HTTP/1.1 200     1.79 secs:       3 bytes ==> GET  /
HTTP/1.1 200     2.10 secs:       3 bytes ==> GET  /
HTTP/1.1 200     3.30 secs:       3 bytes ==> GET  /
HTTP/1.1 200     3.30 secs:       3 bytes ==> GET  /
HTTP/1.1 200     4.09 secs:       3 bytes ==> GET  /
HTTP/1.1 200     6.00 secs:       3 bytes ==> GET  /
HTTP/1.1 200     7.09 secs:       3 bytes ==> GET  /
HTTP/1.1 200     7.88 secs:       3 bytes ==> GET  /
HTTP/1.1 200     8.09 secs:       3 bytes ==> GET  /
HTTP/1.1 200     8.60 secs:       3 bytes ==> GET  /
HTTP/1.1 200     8.99 secs:       3 bytes ==> GET  /
HTTP/1.1 200     9.17 secs:       3 bytes ==> GET  /
HTTP/1.1 200     9.40 secs:       3 bytes ==> GET  /
HTTP/1.1 200    10.09 secs:       3 bytes ==> GET  /
HTTP/1.1 200    10.30 secs:       3 bytes ==> GET  /
HTTP/1.1 200    10.30 secs:       3 bytes ==> GET  /
HTTP/1.1 200    10.39 secs:       3 bytes ==> GET  /
HTTP/1.1 200    11.19 secs:       3 bytes ==> GET  /
HTTP/1.1 200    12.51 secs:       3 bytes ==> GET  /
HTTP/1.1 200    12.80 secs:       3 bytes ==> GET  /
HTTP/1.1 200    13.68 secs:       3 bytes ==> GET  /
HTTP/1.1 200     5.71 secs:       3 bytes ==> GET  /
HTTP/1.1 200     5.30 secs:       3 bytes ==> GET  /
HTTP/1.1 200     4.71 secs:       3 bytes ==> GET  /
HTTP/1.1 200     4.88 secs:       3 bytes ==> GET  /
HTTP/1.1 200    16.01 secs:       3 bytes ==> GET  /
HTTP/1.1 200    16.20 secs:       3 bytes ==> GET  /
```

새로운 터미널 창 Open하여 실시간 모니터링

```
watch -n 1 kubectl get pod
```

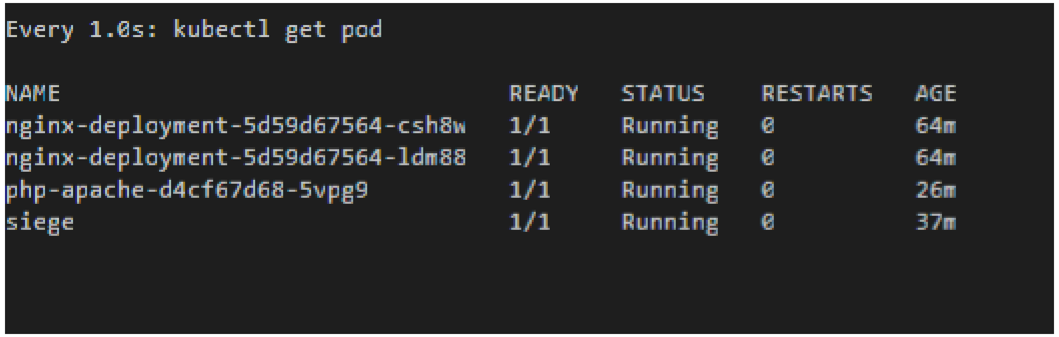
1초 주기로 pod의 상태 모니터링

설정한 임계치를 넘으면, pod가 생성되는 것을 볼 수 있음



```
Every 1.0s: kubectl get pod

NAME                                READY   STATUS              RESTARTS   AGE
nginx-deployment-5d59d67564-csh8w   1/1     Running             0          37m
nginx-deployment-5d59d67564-ldm88   1/1     Running             0          37m
php-apache-d4cf67d68-5vpg9          0/1     ContainerCreating   0          9s
php-apache-d4cf67d68-cttl2          1/1     Running             0          9s
php-apache-d4cf67d68-h9zcj          1/1     Running             0          29m
php-apache-d4cf67d68-j4k9h          0/1     ContainerCreating   0          9s
siege                               1/1     Running             0          10m
```

부하가 끝나면 기본 세팅 1개로 돌아오는 것을 확인



```
Every 1.0s: kubectl get pod

NAME                                READY   STATUS    RESTARTS   AGE
nginx-deployment-5d59d67564-csh8w   1/1     Running   0          64m
nginx-deployment-5d59d67564-ldm88   1/1     Running   0          64m
php-apache-d4cf67d68-5vpg9          1/1     Running   0          26m
siege                               1/1     Running   0          37m
```