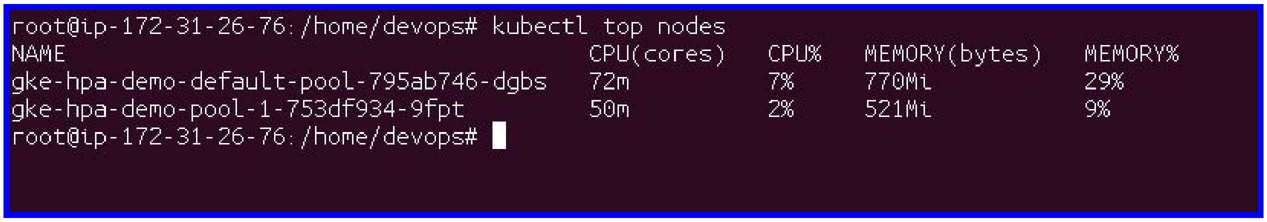
Horizontal Pod Autoscaler

Replace **<your-name>** with your **name** throughout the Lab.

1. SSH to the AWS Workstation

```
$ sudo su
# mkdir /home/devops/hpa
# cd /home/devops/hpa
# kubectl top nodes
```

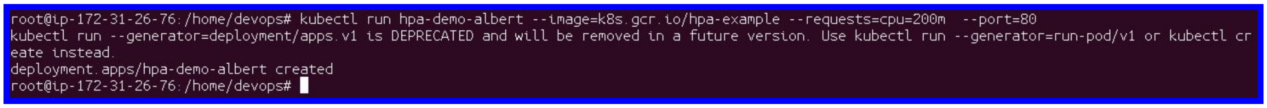


```
root@ip-172-31-26-76:/home/devops# kubectl top nodes
NAME                                  CPU(cores)   CPU%   MEMORY(bytes)   MEMORY%
gke-hpa-demo-default-pool-795ab746-dgbs   72m          7%     770Mi           29%
gke-hpa-demo-pool-1-753df934-9fpt     50m          2%     521Mi           9%
root@ip-172-31-26-76:/home/devops#
```

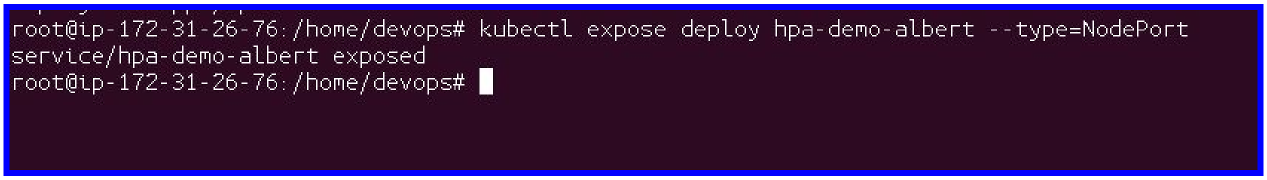
2. Run & expose the Application

Create a new deployment with the below command.

```
# kubectl run hpa-demo-<your-name> --image=k8s.gcr.io/hpa-example
--requests=cpu=200m  --port=80
# kubectl expose deploy hpa-demo-<your-name> --type=NodePort
```



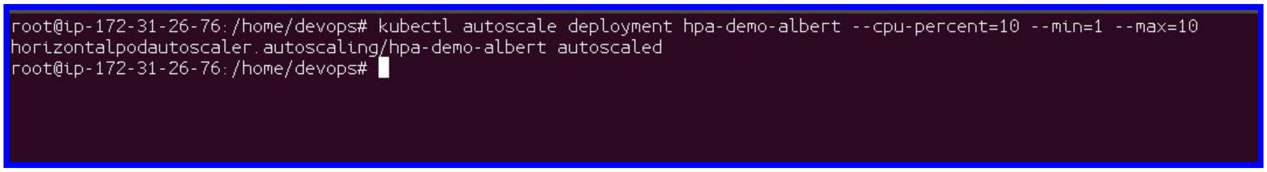
```
root@ip-172-31-26-76:/home/devops# kubectl run hpa-demo-albert --image=k8s.gcr.io/hpa-example --requests=cpu=200m  --port=80
kubectl run --generator=deployment/apps.v1 is DEPRECATED and will be removed in a future version. Use kubectl run --generator=run-pod/v1 or kubectl cr
eate instead.
deployment.apps/hpa-demo-albert created
root@ip-172-31-26-76:/home/devops#
```



```
root@ip-172-31-26-76:/home/devops# kubectl expose deploy hpa-demo-albert --type=NodePort
service/hpa-demo-albert exposed
root@ip-172-31-26-76:/home/devops#
```

3. Create Horizontal Pod Autoscaler

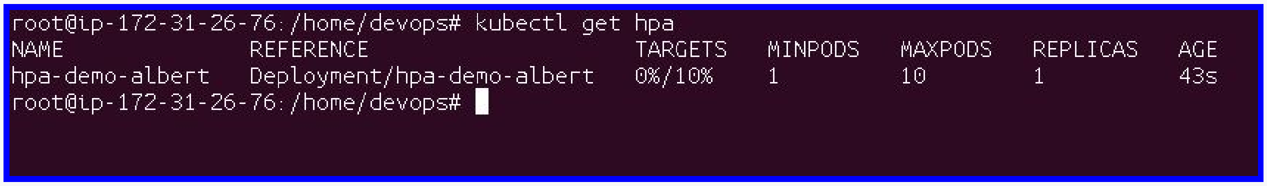
```
# kubectl autoscale deployment hpa-demo-<your-name> --cpu-percent=10
--min=1 --max=10
```



```
root@ip-172-31-26-76:/home/devops# kubectl autoscale deployment hpa-demo-albert --cpu-percent=10 --min=1 --max=10
horizontalpodautoscaler.autoscaling/hpa-demo-albert autoscaled
root@ip-172-31-26-76:/home/devops#
```

4. Please wait for 2-3 minutes before running the below command

```
# kubectl get hpa -w
```

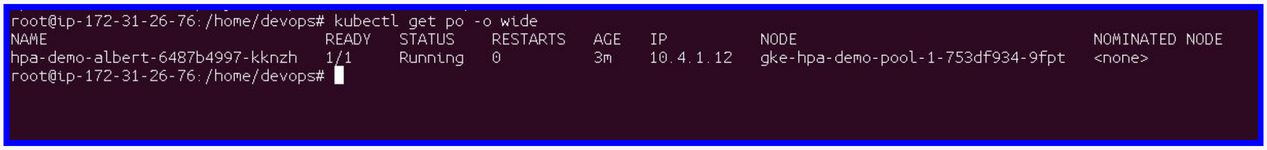


```
root@ip-172-31-26-76:/home/devops# kubectl get hpa
NAME              REFERENCE                   TARGETS   MINPODS   MAXPODS   REPLICAS   AGE
hpa-demo-albert   Deployment/hpa-demo-albert  0%/10%    1         10        1          43s
root@ip-172-31-26-76:/home/devops# 
```

Press Ctrl+c to exit.

7. Check the NODE where the HPA App has been deployed.

```
# kubectl get po -o wide
```

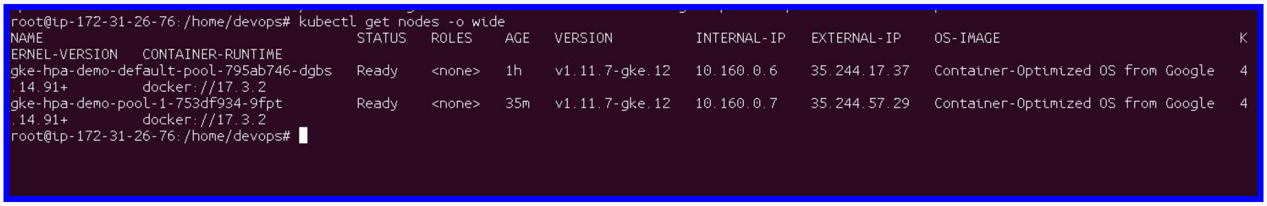


```
root@ip-172-31-26-76:/home/devops# kubectl get po -o wide
NAME                              READY   STATUS    RESTARTS   AGE   IP          NODE                                NOMINATED NODE
hpa-demo-albert-6487b4997-kknzh   1/1     Running   0          3m    10.4.1.12   gke-hpa-demo-pool-1-753df934-9fpt   <none>
root@ip-172-31-26-76:/home/devops# 
```

In this demo the POD is running on **gke-hpa-demo-pool-1-753df934-9fpt NODE.**

8. Run the below command to get the PUBLIC IP of the NODE where the POD is running.

```
# kubectl get nodes -o wide
```

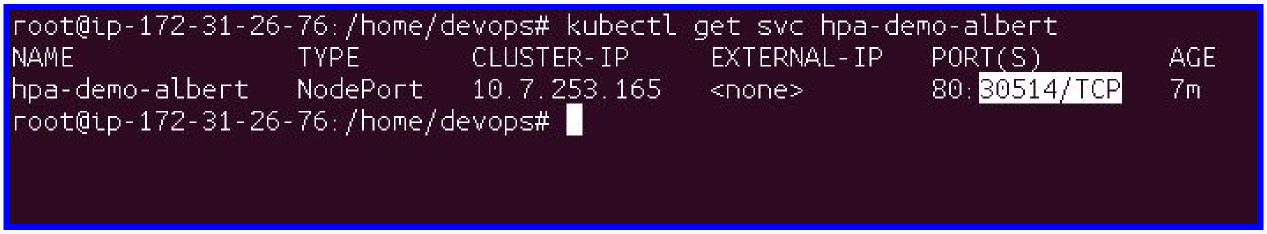


```
root@ip-172-31-26-76:/home/devops# kubectl get nodes -o wide
NAME                                     STATUS   ROLES    AGE   VERSION          INTERNAL-IP   EXTERNAL-IP    OS-IMAGE                             K
ERNEL-VERSION   CONTAINER-RUNTIME
gke-hpa-demo-default-pool-795ab746-dgbs  Ready    <none>   1h    v1.11.7-gke.12   10.160.0.6    35.244.17.37   Container-Optimized OS from Google   4
.14.91+         docker://17.3.2
gke-hpa-demo-pool-1-753df934-9fpt        Ready    <none>   35m   v1.11.7-gke.12   10.160.0.7    35.244.57.29   Container-Optimized OS from Google   4
.14.91+         docker://17.3.2
root@ip-172-31-26-76:/home/devops# 
```

Public IP for **gke-hpa-demo-pool-1-753df934-9fpt is 35.244.57.29.**

9. Check the NODEPORT on which the service is exposed.

```
# kubectl get svc hpa-demo-<your-name>
```



```
root@ip-172-31-26-76:/home/devops# kubectl get svc hpa-demo-albert
NAME              TYPE       CLUSTER-IP     EXTERNAL-IP   PORT(S)        AGE
hpa-demo-albert   NodePort   10.7.253.165   <none>        80:30514/TCP   7m
root@ip-172-31-26-76:/home/devops# 
```

In this Example the hpa app is exposed is exposed on NODEPORT 30514

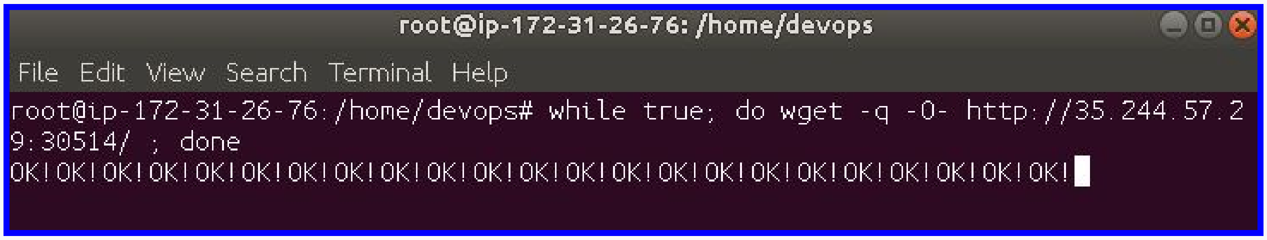
10. Run the below command to increase the Load.

```
while true; do wget -q -O- http://<NODE-Public-IP>:<NodePort>/ ; done
```

Where, NODE-Public-IP is the Public IP of the Node where the hpa-demo POD is deployed and NodePort is the Port on which the hpa-demo POD is exposed to, in this Example 35.244.57.29 is the Public IP of the NODE and 30514 is the NodePort

Example

```
while true; do wget -q -O- http://35.244.57.29:30514/ ; done
```


root@ip-172-31-26-76: /home/devops
File Edit View Search Terminal Help

```
root@ip-172-31-26-76:/home/devops# while true; do wget -q -O- http://35.244.57.2
9:30514/ ; done
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
```

12. Launch TWO more terminals and SSH to your AWS Workstation from both the terminals.

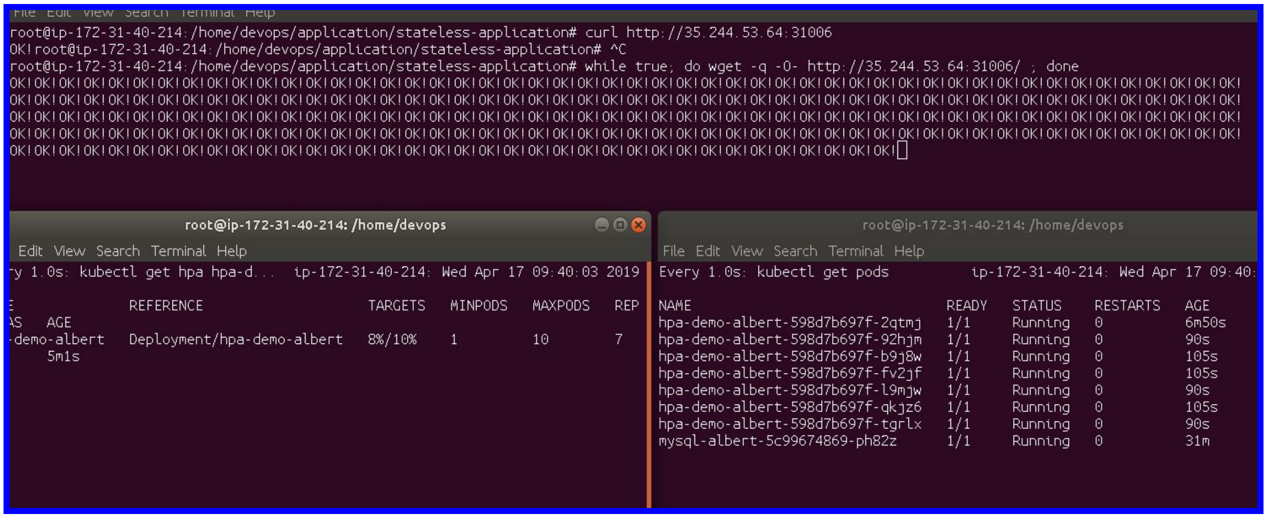
13. On terminal two run the below command

```
$ sudo su
# watch -n 1 kubectl get hpa hpa-demo-<your-name>
```

14. On terminal three run the below command

```
$ sudo su
# watch -n 1 kubectl get pods
```

We can observe that the PODS have been horizontally scaled up due to the increased load on the third terminal, as shown in the below screenshot.


File Edit View Search Terminal Help
root@ip-172-31-40-214: /home/devops
Edit View Search Terminal Help
root@ip-172-31-40-214: /home/devops
File Edit View Search Terminal Help

```
root@ip-172-31-40-214:/home/devops/application/stateless-application# curl http://35.244.53.64:31006
OK!root@ip-172-31-40-214:/home/devops/application/stateless-application# ^C
root@ip-172-31-40-214:/home/devops/application/stateless-application# while true; do wget -q -O- http://35.244.53.64:31006/ ; done
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
```

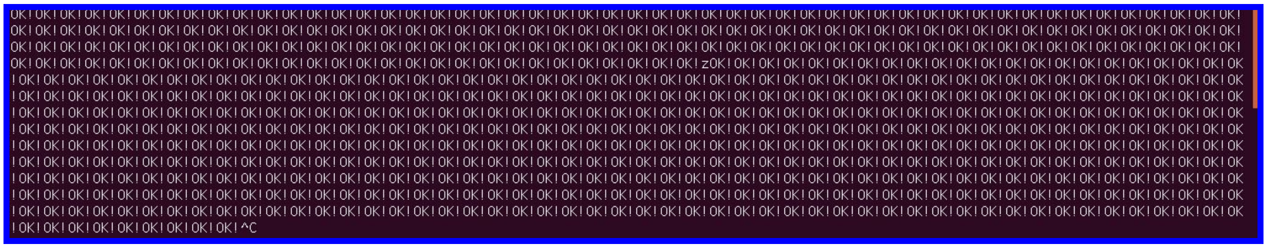
```
ry 1.0s: kubectl get hpa hpa-d...  ip-172-31-40-214: Wed Apr 17 09:40:03 2019

E              REFERENCE                    TARGETS    MINPODS    MAXPODS    REP
AS   AGE
-demo-albert   Deployment/hpa-demo-albert   8%/10%     1          10         7
     5m1s
```

```
Every 1.0s: kubectl get pods        ip-172-31-40-214: Wed Apr 17 09:40:

NAME                                 READY    STATUS     RESTARTS    AGE
hpa-demo-albert-598d7b697f-2qtmj     1/1      Running    0           6m50s
hpa-demo-albert-598d7b697f-92hjm     1/1      Running    0           90s
hpa-demo-albert-598d7b697f-b9j8w     1/1      Running    0           105s
hpa-demo-albert-598d7b697f-fv2jf     1/1      Running    0           105s
hpa-demo-albert-598d7b697f-l9mjw     1/1      Running    0           90s
hpa-demo-albert-598d7b697f-qkjz6     1/1      Running    0           105s
hpa-demo-albert-598d7b697f-tgrlx     1/1      Running    0           90s
mysql-albert-5c99674869-ph82z        1/1      Running    0           31m
```

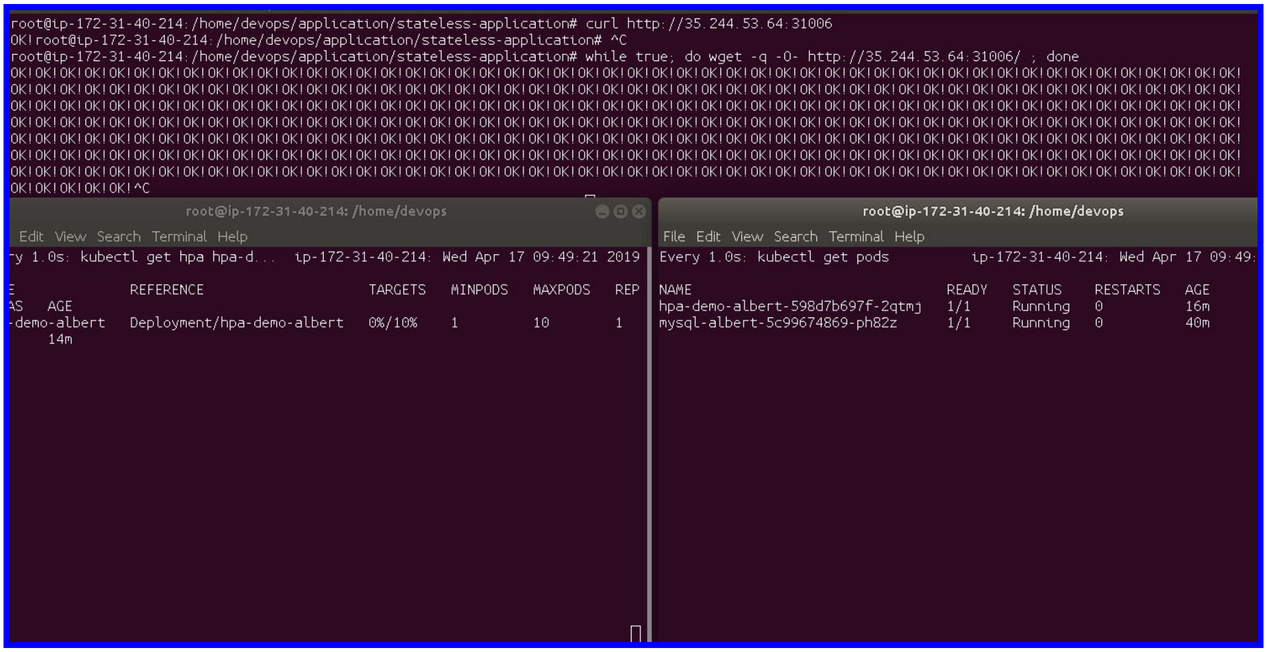
15. To Decrease the Load > Go back to the terminal ONE where we ran the command to increase the load and press CTRL+C to stop the load.



```
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!zOK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK
!OK!OK!OK!OK!OK!OK!OK!OK!OK!^C
```

16. Wait for 5-7 minutes and go back to terminal TWO to check the output of the

```
# kubectl get pods
```


root@ip-172-31-40-214: /home/devops
Edit View Search Terminal Help
root@ip-172-31-40-214: /home/devops
File Edit View Search Terminal Help

```
root@ip-172-31-40-214:/home/devops/application/stateless-application# curl http://35.244.53.64:31006
OK!root@ip-172-31-40-214:/home/devops/application/stateless-application# ^C
root@ip-172-31-40-214:/home/devops/application/stateless-application# while true; do wget -q -O- http://35.244.53.64:31006/ ; done
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!OK!
OK!OK!OK!OK!OK!^C
```

```
ry 1.0s: kubectl get hpa hpa-d...  ip-172-31-40-214: Wed Apr 17 09:49:21 2019

E              REFERENCE                    TARGETS    MINPODS    MAXPODS    REP
AS   AGE
-demo-albert   Deployment/hpa-demo-albert   0%/10%     1          10         1
     14m
```

```
Every 1.0s: kubectl get pods        ip-172-31-40-214: Wed Apr 17 09:49:

NAME                                 READY    STATUS     RESTARTS    AGE
hpa-demo-albert-598d7b697f-2qtmj     1/1      Running    0           16m
mysql-albert-5c99674869-ph82z        1/1      Running    0           40m
```

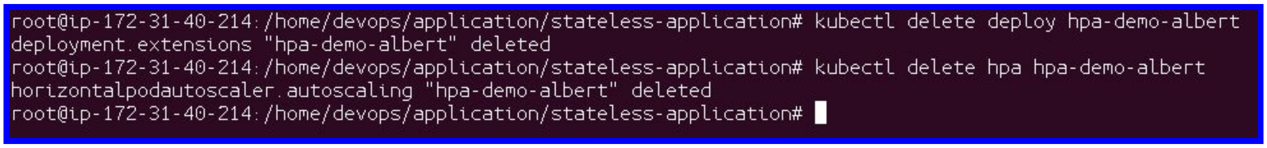
We can observe that the hpa-demo-albert app has scaled down to one again after the load is decreased.

```
# kubectl get hpa
```

Also, we can observe that the load has decreased to 0%.

19. On Terminal-1, Run the below commands to delete the Deployments and HPA

```
# kubectl delete deploy hpa-demo-<your-name>
# kubectl delete hpa hpa-demo-<your-name>
```



```
root@ip-172-31-40-214:/home/devops/application/stateless-application# kubectl delete deploy hpa-demo-albert
deployment.extensions "hpa-demo-albert" deleted
root@ip-172-31-40-214:/home/devops/application/stateless-application# kubectl delete hpa hpa-demo-albert
horizontalpodautoscaler.autoscaling "hpa-demo-albert" deleted
root@ip-172-31-40-214:/home/devops/application/stateless-application#
```