

IMPLEMENT HPA ON K8S

ONCE CLUSTER IS READY, INSTALL METRIC SERVER

```
kubectl apply -f https://github.com/kubernetes-sigs/metrics-server/releases/latest/download/components.yaml
```

Verify that metrics-server is running:

```
kubectl get deployment metrics-server -n kube-system
```

Deploy an Application:

```
apiVersion: apps/v1
```

```
kind: Deployment
```

```
metadata:
```

```
  name: nginx-deployment
```

```
spec:
```

```
  replicas: 1
```

```
  selector:
```

```
    matchLabels:
```

```
      app: nginx
```

```
  template:
```

```
    metadata:
```

```
      labels:
```

```
        app: nginx
```

```
    spec:
```

containers:

- name: nginx

image: nginx

ports:

- containerPort: 80

resources:

requests:

cpu: "100m"

limits:

cpu: "500m"

kubectl create -f deployment.yml

Create HPA:

apiVersion: autoscaling/v2

kind: HorizontalPodAutoscaler

metadata:

name: nginx-hpa

spec:

scaleTargetRef:

apiVersion: apps/v1

kind: Deployment

name: nginx-deployment

minReplicas: 1

maxReplicas: 5

metrics:

- type: Resource

resource:

name: cpu

target:

type: Utilization

averageUtilization: 50 # Target 50% CPU utilization

kubectl apply -f nginx-hpa.yml

kubectl get hpa

TEST HPA:

```
kubectl exec -it <nginx-pod-name> -- bash
```

```
apt-get update && apt-get install stress -y
```

```
stress --cpu 2 --timeout 300
```

OPEN ANOTHER TERMINAL and perform `kubectl get hpa -w` **command**

this will give logs

```
[root@ip-172-31-30-254 ~]# kubectl get hpa -w
```

NAME	REFERENCE	TARGETS	MINPODS	MAXPODS	REPLICAS	AGE
nginx-hpa	Deployment/nginx-deployment	164%/50%	1	5	1	2m40s
nginx-hpa	Deployment/nginx-deployment	34%/50%	1	5	4	2m45s
nginx-hpa	Deployment/nginx-deployment	158%/50%	1	5	4	3m
nginx-hpa	Deployment/nginx-deployment	125%/50%	1	5	5	3m15s
nginx-hpa	Deployment/nginx-deployment	100%/50%	1	5	5	3m45s
nginx-hpa	Deployment/nginx-deployment	99%/50%	1	5	5	4m15s
nginx-hpa	Deployment/nginx-deployment	75%/50%	1	5	5	4m30s
nginx-hpa	Deployment/nginx-deployment	0%/50%	1	5	5	4m45s
nginx-hpa	Deployment/nginx-deployment	0%/50%	1	5	5	9m31s
nginx-hpa	Deployment/nginx-deployment	0%/50%	1	5	1	9m46s

SEE THE PODS

```
[root@ip-172-31-30-254 mustafa]# kubectl get po
```

NAME	READY	STATUS	RESTARTS	AGE
nginx-deployment-96ddb4849-44jqj	1/1	Running	0	106s
nginx-deployment-96ddb4849-kjd6h	1/1	Running	0	4m33s
nginx-deployment-96ddb4849-v6856	1/1	Running	0	106s
nginx-deployment-96ddb4849-wnv46	1/1	Running	0	76s
nginx-deployment-96ddb4849-wvm24	1/1	Running	0	106s

```
[root@ip-172-31-30-254 mustafa]# kubectl get po
```

NAME	READY	STATUS	RESTARTS	AGE
nginx-deployment-96ddb4849-44jqj	1/1	Running	0	3m44s
nginx-deployment-96ddb4849-kjd6h	1/1	Running	0	6m31s
nginx-deployment-96ddb4849-v6856	1/1	Running	0	3m44s
nginx-deployment-96ddb4849-wnv46	1/1	Running	0	3m14s
nginx-deployment-96ddb4849-wvm24	1/1	Running	0	3m44s

```
[root@ip-172-31-30-254 mustafa]# kubectl get po
```

NAME	READY	STATUS	RESTARTS	AGE
nginx-deployment-96ddb4849-44jqj	1/1	Running	0	4m39s
nginx-deployment-96ddb4849-kjd6h	1/1	Running	0	7m26s
nginx-deployment-96ddb4849-v6856	1/1	Running	0	4m39s
nginx-deployment-96ddb4849-wnv46	1/1	Running	0	4m9s
nginx-deployment-96ddb4849-wvm24	1/1	Running	0	4m39s

```
[root@ip-172-31-30-254 mustafa]# kubectl get po
```

NAME	READY	STATUS	RESTARTS	AGE
nginx-deployment-96ddb4849-44jqj	1/1	Running	0	7m59s

```
[root@ip-172-31-30-254 mustafa]#
```

