**Lesson 4 Demo 13**

**Understanding Horizontal Pod Auto Scaling**

**Objective:** To create a Horizontal Pod Autoscaler

**Tools required:** kubeadm and kubectl

**Prerequisites:** Kubernetes cluster should be installed

**Steps to be followed:**

1. Creating HPA in the Master node
2. Checking the deployment service
3. Verifying HPA

**Step 1: Creating HPA in the Master node**

* 1. Open app-hpa.yaml using the following command:

**vi app-hpa.yaml**

* 1. Copy and paste the below code into app-hpa.yaml file for php-apache deployment:

**apiVersion: v1**

**kind: Service**

**metadata:**

**creationTimestamp: null**

**name: php-apache**

**spec:**

**ports:**

**- port: 80**

**protocol: TCP**

**targetPort: 80**

**selector:**

**run: php-apache**

**status:**

**loadBalancer: {}**

**---**

**---**

**apiVersion: apps/v1**

**kind: Deployment**

**metadata:**

**creationTimestamp: null**

**labels:**

**run: php-apache**

**name: php-apache**

**spec:**

**replicas: 1**

**selector:**

**matchLabels:**

**run: php-apache**

**strategy: {}**

**template:**

**metadata:**

**creationTimestamp: null**

**labels:**

**run: php-apache**

**spec:**

**containers:**

**- image: k8s.gcr.io/hpa-example**

**name: php-apache**

**ports:**

**- containerPort: 80**

**resources:**

**requests:**

**cpu: 200m**

**status: {}**

Text

Description automatically generated

**Step 2: Checking the deployment service by using the following commands**

* 1. Creating and checking HPA:

**kubectl create -f app-hpa.yaml**

* 1. Enter the following command to get the Pod status:

**kubectl get pods**

* 1. Deployment of HPA using the following command:

**kubectl get deployment**

* 1. Run the following command to get svc:

**kubectl get svc**

Graphical user interface, text

Description automatically generated

2.5 Copy and paste the following code in hpa.yaml to create HPA:

**apiVersion: autoscaling/v1**

**kind: HorizontalPodAutoscaler**

**metadata:**

**creationTimestamp: null**

**name: php-apache**

**spec:**

**maxReplicas: 10**

**minReplicas: 1**

**scaleTargetRef:**

**apiVersion: apps/v1**

**kind: Deployment**

**name: php-apache**

**targetCPUUtilizationPercentage: 50**

**status:**

**currentReplicas: 0**

**desiredReplicas: 0**

Text

Description automatically generated

2.6 Run the following command to create HPA:

**kubectl create -f hpa.yaml**

**Step 3: Verifying HPA**

3.1 Verify HPA by using the following command:

**kubectl get hpa**

Graphical user interface, text

Description automatically generated

3.2 Run the following command to load-generate:

**kubectl run load-generator --image=busybox -- /bin/sh -c "while sleep 0.01; do wget -q -O- http://php-apache; done"**

**kubectl get hpa php-apache –watch**

**kubectl get deployment php-apache**

* 1. Run the following command to stop the Pod:

**kubectl delete pod load-generator**

A screenshot of a computer

Description automatically generated