Deploy and autoscale nginx app

3 PODS = running = 3000 users = 3 months

5 PODS = running = 5000 users = 4 months

Step 1 : Create a deployment file

vim deployment.yaml

#note: use below code in the yaml file

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:latest

        ports:

        - containerPort: 80

step 2 :deploy the app with 3 pods

kubectl create -f deployment.yaml

step 3 : create service yaml file

vim service.yaml

and use below code in the yaml file

apiVersion: v1

kind: Service

metadata:

  name: nginx-service

spec:

  selector:

    app: nginx

  ports:

    - protocol: TCP

      port: 80

      targetPort: 80

  type: LoadBalancer

step 4 : deploy the service

kubectl create -f service.yaml

Implementing Horizontal Pod Autoscaler (HPA):

After the deployment is set up, you can set up HPA to autoscale your pods. Here's a basic HPA configuration:

Step 5 : Create a horizontal pod autoscaler

vim hpa.yaml

# hpa.yaml

apiVersion: autoscaling/v1

kind: HorizontalPodAutoscaler

metadata:

  name: nginx-hpa

spec:

  scaleTargetRef:

    apiVersion: apps/v1

    kind: Deployment

    name: nginx-deployment

  minReplicas: 3

  maxReplicas: 10

  targetCPUUtilizationPercentage: 80

step 6 : list hpa with desired number of pods

kubectl create -f hpa.yaml

kubectl get hpa

====================== lab completes ========================