=============================================================

How to write a manifest yml for creating deployment resources

=============================================================

---

apiVersion: apps/v1

kind: Deployment

metadata:

name: javawebappdeploy

spec:

replicas: 2

strategy:

type: RollingUpdate

selector:

matchLabels:

app: javawebapppod

template:

metadata:

name: javaewebpod

labels:

app: javawebapppod

spec:

containers:

- name: javawebcontainer

image: vinodses/my-web-app

ports:

- containerPort: 8080

---

apiVersion: v1

kind: Service

metadata:

name: javawebsvc

spec:

type: LoadBalancer

selector:

app: javawebapppod

ports:

- port: 80

targetPort: 8080

...

$ kubectl delete all --all

$ kubectl get all

$ kubectl apply -f <yml>

$ kubectl get all

Note: Access app using LBR URL

$ kubectl scale deployment javawebdeploy --replicas 4

$ kubectl get pods

$ kubectl scale deployment javawebdeploy --replicas 10

When you are configuring HPA (horizontal pod autoscaling) then only we will get autoscaling feature

=============================================================================

Autoscaling

It is the process of increasing/decreasing the infrastructure resources based on the demand

Autoscaling mainly can done in 2 ways

-> Horizontal Scaling means increasing number of instances/servers/pods.

-> vertical Scaling means increasing capacity of single system.

HPA : Horizontal POD Autoscaling

VPA : Vertical POD Autoscaling (we don't use this)