kubernetes

Kubernetes: Demo Auto-Scaling

- Create Some Deployment/Replica-Set of some Service.
- Using Apache Web-Server.
- Generate YMLs and Execute them.
- Start HPA on Deployment.
  kubectl autoscale deployment <deployment-name> --min=2 --max=5 —cpu-percent=80
- Get Status of HPA
  kubectl get hpa
- Describe HPA Uses and Status.
  kubectl describe hpa
- Get HPA YML for reference.
  kubectl get <HPA Name> -o yaml > web-hpa.yaml

- Add Metrics API Server on Kubernetes Cluster.
- Add WebHook for kubectl
- Restart the Cluster.
- Re-create the HPA

- Launch Shell and connect to WebServer.
  kubectl run -i --tty service-test --image=busybox /bin/sh
- Connect with Service Running on Pod.
  wget http://hpa-example.default.svc.cluster.local:31010
- Run Above Command in Loop.
  while true; do wget -q -O- http://hpa-example.default.svc.cluster.local:31111; done

Will see you in Next Lecture…

See you in next lecture …