

Name: Teo Yick Fong Alex

Student Number: A0221444R

[Github Repo](#)

Task A3

Deploy a metrics-server and HorizontalPodAutoscaler

1. Creating metrics-server

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

2. Disable TLS

```
alex@homeserver: ~/Docker/ X + v
# Please edit the object below. Lines beginning with a '#' will be ignored,
# and an empty file will abort the edit. If an error occurs while saving this file will be
# reopened with the relevant failures.
#
apiVersion: apps/v1
kind: Deployment
metadata:
  annotations:
    deployment.kubernetes.io/revision: "1"
    kubectl.kubernetes.io/last-applied-configuration: |
      {"apiVersion":"apps/v1","kind":"Deployment","metadata":{"annotations":{"k8s-app":"metrics-server"},"name":"metrics-server","namespace":"kube-system"},"spec":{"maxUnavailable":0},"template":{"metadata":{"labels":{"k8s-app":"metrics-server"},"spec":{"containers":{"args":["--cert-dir=/tmp","--secure-port=4443","--kubelet-preferred-address-types=InternalIP,ExternalIP,Hostname"],"image":"k8s.gcr.io/metrics-server/metrics-server:v0.6.1","imagePullPolicy":"IfNotPresent","livenessProbe":{"failureThreshold":3,"httpGet":{"path":"/livez","port":https,"scheme":"HTTPS"},"periodSeconds":10,"successThreshold":1},"readinessProbe":{"failureThreshold":3,"httpGet":{"path":"/readyz","port":https,"scheme":"HTTPS"},"periodSeconds":10,"successThreshold":1},"resources":{"cpu":"100m","memory":"200Mi"},"securityContext":{"allowPrivilegeEscalation":false,"readOnlyRootFilesystem":true,"runAsNonRoot":true,"runAsUser":1000},"volumeMounts":[{"name":"tmp-dir"}]}}}
creationTimestamp: "2022-10-11T18:32:38Z"
generation: 1
labels:
  k8s-app: metrics-server
  name: metrics-server
  namespace: kube-system
  resourceVersion: "3914"
  uid: 38ddb02-5174-415c-a87e-9361de5a1b6b
spec:
  progressDeadlineSeconds: 600
  replicas: 1
  revisionHistoryLimit: 10
  selector:
    matchLabels:
      k8s-app: metrics-server
  strategy:
    rollingUpdate:
      maxSurge: 25%
      maxUnavailable: 0
    type: RollingUpdate
  template:
    metadata:
      creationTimestamp: null
      labels:
        k8s-app: metrics-server
    spec:
      containers:
        - args:
            - --kubelet-insecure-tls
            - --cert-dir=/tmp
            - --secure-port=4443
            - --kubelet-preferred-address-types=InternalIP,ExternalIP,Hostname
            - --kubelet-use-node-status-port
            - --metric-resolution=15s
          image: k8s.gcr.io/metrics-server/metrics-server:v0.6.1
          imagePullPolicy: IfNotPresent
          livenessProbe:
            failureThreshold: 3
            httpGet:
              path: /livez
              port: https
              scheme: HTTPS
            periodSeconds: 10
            successThreshold: 1
          resources: {}
          securityContext:
            allowPrivilegeEscalation: false
            readOnlyRootFilesystem: true
            runAsNonRoot: true
            runAsUser: 1000
          volumeMounts:
            - name: tmp-dir
              mountPath: /tmp
      dnsPolicy: ClusterFirst
      restartPolicy: Always
      schedulerName: default-scheduler
      serviceAccountName: metrics-server
      terminationGracePeriodSeconds: 30
      volumes:
        - emptyDir: {}
          name: tmp-dir
```

3. Restart Deployment

1. `kubectl -nkube-system rollout restart deploy/metrics-server`

4. Verify metrics-server

1. `kubectl get pods --all-namespaces | grep metrics-server`

Viewing Metrics-Server					
kube-system	metrics-server-6c9466598b-whsq5	1/1	Running	0	45s

5. Creating HorizontalPodAutoscaler

1. `kubectl apply -f ./manifests/hpa.yaml`

6. Verify HorizontalPodAutoscaler

1. `kubectl describe hpa`

```

Viewing HPA
Warning: autoscaling/v2beta2 HorizontalPodAutoscaler is deprecated in v1.23+, unavailable in v1.26+; use autoscaling/v2 HorizontalPodAutoscaler
Name: backend
Namespace: default
Labels: <none>
Annotations: <none>
CreationTimestamp: Wed, 12 Oct 2022 02:42:44 +0800
Reference: Deployment/backend
Metrics: ( current / target )
  resource cpu on pods  (as a percentage of request): 0% (0) / 10%
Min replicas: 1
Max replicas: 10
Deployment pods: 3 current / 3 desired
Conditions:
  Type            Status  Reason
  --            --
  AbleToScale     True    ScaleDownStabilized
  ScalingActive   True    ValidMetricFound
  ScalingLimited  False   DesiredWithinRange
Events:
  recent recommendations were higher than current one, applying the highest recent recommendation
  the HPA was able to successfully calculate a replica count from cpu resource utilization (percentage of request)
  the desired count is within the acceptable range

```

7. HPA Scaling up under load

```

NAME                                READY    STATUS    RESTARTS   AGE
backend-887ccbfbcb-75xr6            1/1     Running   0           12m
backend-887ccbfbcb-n2lbk            1/1     Running   0           12m
backend-887ccbfbcb-q9b2            1/1     Running   0           12m

```

```

NAME                                READY    STATUS    RESTARTS   AGE
backend-887ccbfbcb-75xr6            1/1     Running   0           15m
backend-887ccbfbcb-n2lbk            1/1     Running   0           15m
backend-887ccbfbcb-q9b2            1/1     Running   0           15m
backend-887ccbfbcb-mz29f            0/1     Pending   0           0s
backend-887ccbfbcb-j4xpf            0/1     Pending   0           0s
backend-887ccbfbcb-mz29f            0/1     Pending   0           0s
backend-887ccbfbcb-j4xpf            0/1     Pending   0           0s
backend-887ccbfbcb-mz29f            0/1     Pending   0           0s
backend-887ccbfbcb-j4xpf            0/1     Pending   0           0s
backend-887ccbfbcb-mz29f            0/1     ContainerCreating 0           0s
backend-887ccbfbcb-j4xpf            0/1     ContainerCreating 0           0s
backend-887ccbfbcb-j4xpf            1/1     Running   0           5s

```

Task A3.2

Deploy Docker image as zone-aware manner

1. Creating Zone Aware Deployment

```
1. kubectl apply -f ./manifests/backend-zone-aware.yaml
```

2. Create new Service

```
1. kubectl apply -f ./manifests/service-zone-aware.yaml
```

3. Viewing Pods

```
1. kubectl get po -lapp=backend-zone-aware -o wide --sort-by='.spec.nodeName'
```

```

NAME                                READY    STATUS    RESTARTS   AGE   IP            NODE              NOMINATED NODE   READINESS GATES
backend-zone-aware-5d7d6ff679-9mt44 1/1     Running   0           171m  10.244.3.6    kind-1-worker     <none>            <none>
backend-zone-aware-5d7d6ff679-5bh2g 1/1     Running   0           171m  10.244.1.10   kind-1-worker2    <none>            <none>
backend-zone-aware-5d7d6ff679-9f8b8 1/1     Running   0           172m  10.244.1.9    kind-1-worker2    <none>            <none>
backend-zone-aware-5d7d6ff679-df28b 1/1     Running   0           172m  10.244.1.8    kind-1-worker2    <none>            <none>
backend-zone-aware-5d7d6ff679-9b15x 1/1     Running   0           172m  10.244.2.10   kind-1-worker3    <none>            <none>
backend-zone-aware-5d7d6ff679-dv2ml 1/1     Running   0           171m  10.244.2.13   kind-1-worker3    <none>            <none>
backend-zone-aware-5d7d6ff679-lfhxx 1/1     Running   0           172m  10.244.2.9    kind-1-worker3    <none>            <none>
backend-zone-aware-5d7d6ff679-m4hzl 1/1     Running   0           171m  10.244.2.11   kind-1-worker3    <none>            <none>
backend-zone-aware-5d7d6ff679-tf28h 1/1     Running   0           172m  10.244.2.8    kind-1-worker3    <none>            <none>
backend-zone-aware-5d7d6ff679-v46j9 1/1     Running   0           171m  10.244.2.12   kind-1-worker3    <none>            <none>

```

Deployed Site

