Lee Yan Cheng A0199141B

GitHub link: https://github.com/yanchenglee98/OTOT-A2-A3

Instructions:

1) install metrics server: kubectl apply -f https://github.com/kubernetes-sigs/metrics-server/releases/latest/download/components.yaml
2) manually edit Deployment manifest to ensure tls is working:

kubectl -nkube-system edit deploy/metrics-server

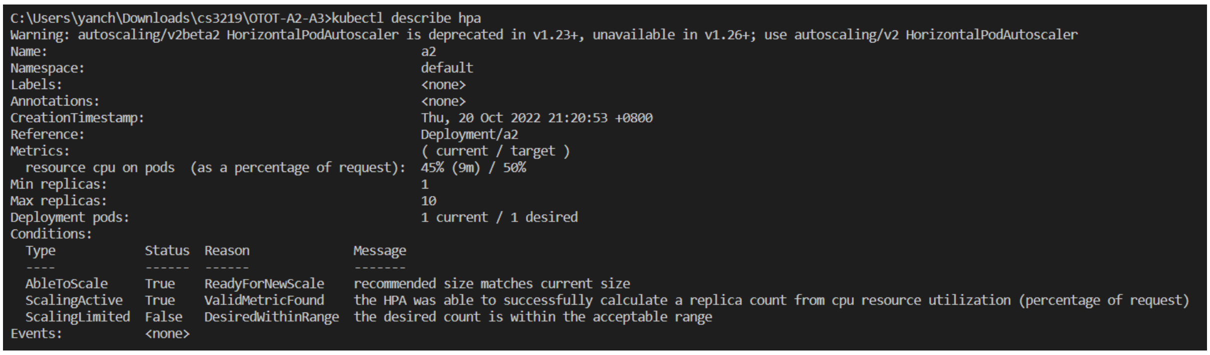
add a flag `--kubelet-insecure-tls` to `deployment.spec.containers[].args[]`

restart the Deployment using `kubectl -nkube-system rollout restart deploy/metrics-server`

3) apply hpa: kubectl apply -f ./demo/a3/hpa.yml
4) spam refresh at http://localhost/app/
5) check hpa event log: kubectl describe hpa
6) apply zone aware: kubectl apply -f ./demo/a3/deployment-zone-aware.yml
7) check zone aware allocation: kubectl get po -lapp=a2-zone-aware -owide

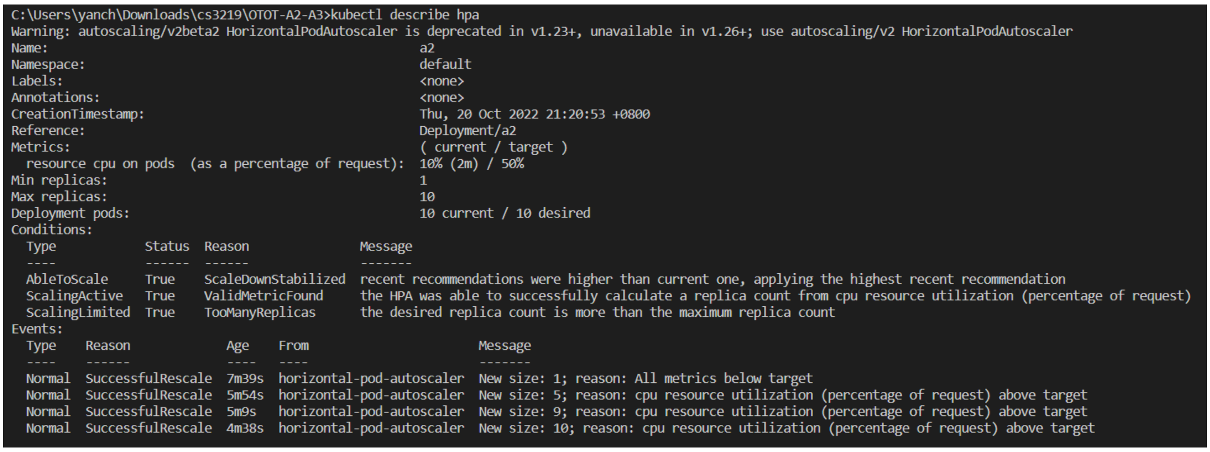
Screenshot proof:

Before scale:



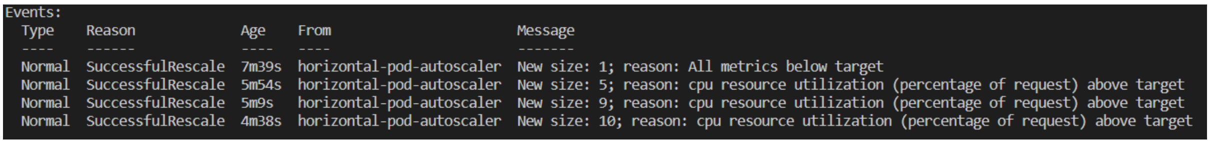
```
C:\Users\yanch\Downloads\cs3219\OTOT-A2-A3>kubectl describe hpa
Warning: autoscaling/v2beta2 HorizontalPodAutoscaler is deprecated in v1.23+, unavailable in v1.26+; use autoscaling/v2 HorizontalPodAutoscaler
Name:                                                  a2
Namespace:                                             default
Labels:                                                <none>
Annotations:                                           <none>
CreationTimestamp:                                     Thu, 20 Oct 2022 21:20:53 +0800
Reference:                                             Deployment/a2
Metrics:                                               ( current / target )
  resource cpu on pods  (as a percentage of request):  45% (9m) / 50%
Min replicas:                                          1
Max replicas:                                          10
Deployment pods:                                       1 current / 1 desired
Conditions:
  Type            Status  Reason              Message
  ----            ------  ------              -------
  AbleToScale     True    ReadyForNewScale    recommended size matches current size
  ScalingActive   True    ValidMetricFound    the HPA was able to successfully calculate a replica count from cpu resource utilization (percentage of request)
  ScalingLimited  False   DesiredWithinRange  the desired count is within the acceptable range
Events:           <none>
```

After scale:



```
C:\Users\yanch\Downloads\cs3219\OTOT-A2-A3>kubectl describe hpa
Warning: autoscaling/v2beta2 HorizontalPodAutoscaler is deprecated in v1.23+, unavailable in v1.26+; use autoscaling/v2 HorizontalPodAutoscaler
Name:                                                  a2
Namespace:                                             default
Labels:                                                <none>
Annotations:                                           <none>
CreationTimestamp:                                     Thu, 20 Oct 2022 21:20:53 +0800
Reference:                                             Deployment/a2
Metrics:                                               ( current / target )
  resource cpu on pods  (as a percentage of request):  10% (2m) / 50%
Min replicas:                                          1
Max replicas:                                          10
Deployment pods:                                       10 current / 10 desired
Conditions:
  Type            Status  Reason               Message
  ----            ------  ------               -------
  AbleToScale     True    ScaleDownStabilized  recent recommendations were higher than current one, applying the highest recent recommendation
  ScalingActive   True    ValidMetricFound     the HPA was able to successfully calculate a replica count from cpu resource utilization (percentage of request)
  ScalingLimited  True    TooManyReplicas      the desired replica count is more than the maximum replica count
Events:
  Type    Reason             Age    From                       Message
  ----    ------             ----   ----                       -------
  Normal  SuccessfulRescale  7m39s  horizontal-pod-autoscaler  New size: 1; reason: All metrics below target
  Normal  SuccessfulRescale  5m54s  horizontal-pod-autoscaler  New size: 5; reason: cpu resource utilization (percentage of request) above target
  Normal  SuccessfulRescale  5m9s   horizontal-pod-autoscaler  New size: 9; reason: cpu resource utilization (percentage of request) above target
  Normal  SuccessfulRescale  4m38s  horizontal-pod-autoscaler  New size: 10; reason: cpu resource utilization (percentage of request) above target
```

HPA event output:



```
Events:
  Type    Reason             Age    From                       Message
  ----    ------             ----   ----                       -------
  Normal  SuccessfulRescale  7m39s  horizontal-pod-autoscaler  New size: 1; reason: All metrics below target
  Normal  SuccessfulRescale  5m54s  horizontal-pod-autoscaler  New size: 5; reason: cpu resource utilization (percentage of request) above target
  Normal  SuccessfulRescale  5m9s   horizontal-pod-autoscaler  New size: 9; reason: cpu resource utilization (percentage of request) above target
  Normal  SuccessfulRescale  4m38s  horizontal-pod-autoscaler  New size: 10; reason: cpu resource utilization (percentage of request) above target
```

Video demo link:
https://drive.google.com/file/d/1LI9efI9_peSOteEwQRwC1cssJ45FLNnK/view?usp=sharing

Demo does not show scaling as it takes a while to scale