Exercise 14b

Create a Kubernetes Cluster in DigitalOcean and Deploy Cassandra

Prior Knowledge

Unix Command Line Shell YAML Completion of Ex 14a

Learning Objectives

See how Cassandra replicates Introduction to Kubernetes

Software Requirements

Browser kubectl

Overview

In this exercise we are going to instantiate a Kubernetes cluster in DO, then install a Cassandra ring onto the kubernetes cluster. Finally we will do some load-testing.

- 1. If you have left the cluster running from Ex14a, go straight to **step 2**
 - a. Otherwise redo the steps to create a cluster:



b. Make sure you install the monitoring 1-click app.



d. Download the config file, then:

```
mv ~/Downloads/k8s-cass-kubeconfig.yaml ~/.kube/
```

In your terminal window:

```
export KUBECONFIG=~/.kube/k8s-cass-kubeconfig.yaml
```

(There are also other things we can do, but this works fine)

e. Check it works:

```
kubectl get all
```

You should see something like:

```
NAME TYPE CLUSTER-IP EXTERNAL-IP PORT(S) AGE service/kubernetes ClusterIP 10.245.0.1 <none> 443/TCP 45m
```

2. Now let's make a directory for our Cassandra YAMLs:

```
mkdir ~/cassandra
```

3. There are only two small YAML files required to get Cassandra running. They come from this webpage:

https://kubernetes.io/docs/tutorials/stateful-application/cassandra/

Because we need to modify one of them, let's download them:

```
cd ~/cassandra
wget https://k8s.io/examples/application/cassandra/cassandra-service.yaml
wget https://k8s.io/examples/application/cassandra/cassandra-statefulset.yaml
```

4. Take a look at the cassandra-service.yaml:

It is really simple (for YAML!):

```
apiVersion: v1
kind: Service
metadata:
    labels:
        app: cassandra
    name: cassandra
spec:
    clusterIP: None
    ports:
        - port: 9042
    selector:
        app: cassandra
```



- 5. This is "kind of" the equivalent of EXPOSE in Docker. You could compare this to the one in the "hello-kubernetes" app if you like.
- 6. Let's deploy this: kubectl apply -f cassandra-service.yaml

service/cassandra created

7. Check if it is happy: kubectl get svc

```
NAME
            TYPE
                        CLUSTER-IP
                                     EXTERNAL-IP
                                                   PORT(S)
                                                              AGE
            ClusterIP
                                                   9042/TCP
cassandra
                                                              295
                        None
                                     <none>
                                                   443/TCP
kubernetes
            ClusterIP
                        10.245.0.1
                                     <none>
                                                              28m
```

- 8. You should see the cassandra service running alongside the kubernetes master. We need this to be in place **before** we start the next part because the different pods need to be able to access each other via port 9042 for the cluster to form.
- 9. The second file is more complex. It basically defines three things:
 - a. The cassandra images and config to start a cassandra container

```
apiVersion: apps/v1
kind: StatefulSet
metadata:
   name: cassandra
   labels:
    app: cassandra
spec:
   serviceName: cassandra
replicas: 3
   selector:
    matchLabels:
    app: cassandra
template:
   metadata:
   labels:
   app: cassandra
```

(That is just the start of that bit)



 Defines the storage that will be needed by these servers in a StatefulSet

(https://cloud.google.com/kubernetes-engine/docs/concepts/statefulset)



c. Defines a StorageClass for deploying into Minikube (a kubernetes distro designed to run on developers' machines).

kind: StorageClass
apiVersion: storage.k8s.io/v1
metadata:
 name: fast
provisioner: k8s.io/minikube-hostpath
parameters:
 type: pd-ssd

- 10. This YAML will **not work** as is. That is because we need the Kubernetes cluster to request disk from DigitalOcean specifically, not from minikube.
- 11. Let's see what StorageClass is available in our DO cluster:

```
kubectl get storageclass

NAME do-block-storage (default) PROVISIONER dobs.csi.digitalocean.com Delete Unmediate ALLOWVOLUMEEXPANSION AGE 5h26i
```

This is a pre-configured storage class for DO called do-block-storage

12. We need to edit the YAML file:

code ~/cassandra/cassandra-statefulset.yaml

13. Firstly, lets delete the StorageClass section: Delete all the highlighted lines:

```
91
               requests:
                 storage: 1Gi
92
93
      kind: StorageClass
94
      apiVersion: storage.k8s.io/vl
95
96
        name: fast
97
      provisioner: k8s.io/minikube-hostpath
98
99
100
        type: pd-ssd
101
```



14. Secondly, change the storageClassName from **fast** to **do-block-storage** (which we just identified above).

```
# These are converted to volume claims by the controller
# and mounted at the paths mentioned above.
# do not use these in production until ssd GCEPersistentDisk or
volumeClaimTemplates:
- metadata:
| name: cassandra-data
| name: cassandra-data
| spec:
| accessModes: [ "ReadWriteOnce" ]
| storageClassName: do-block-storage
| resources:
| requests:
| storage: 1Gi
| 93
| 94
```

- 15. Save the file
- 16. Let's deploy this now:

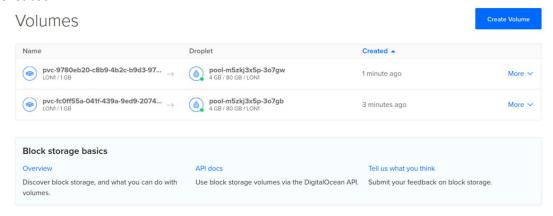
```
kubectl apply -f cassandra-statefulset.yaml
statefulset.apps/cassandra created
```

17. We need to wait a bit for this to start. Basically the system is making API requests to DigitalOcean to provision disks.

If you go to k9s, you should see the containers appearing and starting up:



19. If you go to the DigitalOcean control panel you will see **Volumes** being created:



20. After about 5 minutes the cluster should be up and running:

kubectl get all

| NAME pod/cassandra-0 pod/cassandra-1 pod/cassandra-2 | READY 1/1 1/1 1/1 | STATUS Runnin Runnin Runnin | g 0 g 0 | AGE 10m 9m21s 7m37s | | |
|---|----------------------------|--------------------------------------|----------------------------------|------------------------------|--------------------------------|-----------------------|
| NAME service/cassandra service/kubernete | | sterIP sterIP | CLUSTER-IP None 10.245.0.1 | EXTERNAL-IP <none></none> | PORT(S) 9042/TCP 443/TCP | AGE 5h14m 5h42m |
| NAME statefulset.apps/ | /cassandr | REA a 3/3 | | | | |

21. We can now execute commands in the cassandra cluster.

"kubectl exec -ti" is a bit like docker exec. This executes the command that follows -- on the cassandra-0 container instance.

kubectl exec -ti cassandra-0 -- nodetool status



22. Now let's stress test the server. We can do it over the network between our machine and the DO cluster, but be warned this isn't terribly efficient.

First, let's forward the cluster port 9042 to local port 9040. We are choosing 9040 locally because we might still be running cassandra locally on 9042 and we want to be sure we are talking to the remote cluster:

kubectl port-forward pods/cassandra-0 9040:9042

You should see:

```
Forwarding from 127.0.0.1:9040 -> 9042 Forwarding from [::1]:9040 -> 9042
```

Leave that window and **start a new terminal window**.

23. Now let's to a stress test on port 9040:

```
cassandra-stress write n=100000 -port native=9040 -rate threads=1000
```

You may see some Java exceptions in the logs (due to networking challenges). Eventually it should finish with something like:

```
Results:
                              1,745 op/s
Op rate
                                          [WRITE: 1,745 op/s]
Partition rate
                              1,745 pk/s [WRITE: 1,745 pk/s]
Row rate
                              1,745 row/s [WRITE: 1,745 row/s]
                     : 562.7 ms [WRITE: 562.7 ms]
: 504.9 ms [WRITE: 504.9 ms]
Latency mean
Latency median
Latency 95th percentile : 995.1 ms [WRITE: 995.1 ms]
Latency 99th percentile : 1297.1 ms [WRITE: 1,297.1 ms]
Latency 99.9th percentile : 1903.2 ms [WRITE: 1,903.2 ms]
                  : 2493.5 ms [WRITE: 2,493.5 ms]
Latency max
                         : 100,000 [WRITE: 100,000]
Total partitions
                                   0 [WRITE: 0]
Total errors
                        : 0
Total GC count
Total GC memory
                        : 0.000 KiB
Total GC time
                        : 0.0 seconds
Avg GC time
                             NaN ms
                        :
StdDev GC time
                             0.0 ms
Total operation time : 00:00:57
END
```

- 24. Kill the port forwarding process (Ctrl-C)
- 25. Let's now do the same test from within the cluster. To do this, we can start a new pod and get shell access:

 $kubectl\ apply\ -f\ \underline{https://raw.githubusercontent.com/pzfreo/ox-clo/master/code/cass-tools/shell.yaml$



27. This is just an Ubuntu container with cassandra tools installed. Let's check it started:

kubectl get all

| pod/casstool | 1/1 | Running | 0 | 23m |
|-----------------|-------|---------|----------|-------|
| pod/cassandra-2 | 1/1 | Running | 0 | 4h24m |
| pod/cassandra-1 | 1/1 | Running | 0 | 4h26m |
| pod/cassandra-0 | 1/1 | Running | 0 | 4h27m |
| NAME | READY | STATUS | RESTARTS | AGE |

| NAME | • | TYPE | CLUSTER-IP | |
|---------------|---------|-----------|------------|---------------|
| EXTERNAL-IP | PORT(S) | AGE | | |
| service/cassa | ndra | ClusterIP | None | <none></none> |
| 9042/TCP 9h | ı | | | |
| service/kuber | netes | ClusterIP | 10.245.0.1 | <none></none> |
| 443/TCP 9h | 1 | | | |

| NAME | READY | AGE |
|----------------------------|-------|-------|
| statefulset.apps/cassandra | 3/3 | 4h27m |

28. Inside kubernetes, the networking is different to the real world. We need to know a host IP to contact the pods on:

kubectl describe svc/cassandra

Name: cassandra Namespace: default Labels: app=cassandra

Annotations: Selector: app=cassandra

Type: ClusterIP IP: None

Port: <unset> 9042/TCP

9042/TCP TargetPort:

TargetPort: 9042/TCP Endpoints: **10.244.0.153**:9042,10.244.0.73:9042,10.244.1.109:9042

Session Affinity: None Events: <none>

Choose one of the IP addresses listed as an endpoint. Make a note (in my case 10.244.0.153)

29. We can now get a command-line in that container:

kubectl exec -it casstool -- bash



30. Now let's redo that test from within the cluster (with your IP address)

```
cassandra-stress write n=100000 -rate threads=1000 -node 10.244.0.153
```

Unless you have a massively fast connection from your machine to the DO datacentre, you should see much better performance now:

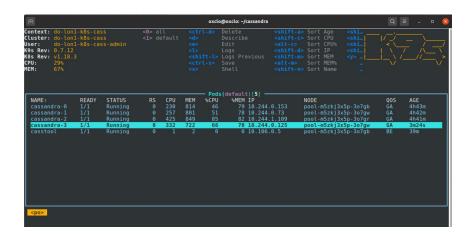
```
Results:
Op rate
                               7,527 op/s
                                            [WRITE: 7,527 op/s]
Partition rate
                               7,527 pk/s
                                            [WRITE: 7,527 pk/s]
                               7,527 row/s [WRITE: 7,527 row/s]
Row rate
Latency mean
                             123.7 ms [WRITE: 123.7 ms]
Latency median
                              87.9 ms
                                       [WRITE: 87.9 ms]
Latency 95th percentile
                             419.7 ms
                                      [WRITE: 419.7 ms]
Latency 99th percentile
                             649.6 ms [WRITE: 649.6 ms]
Latency 99.9th percentile : 1015.0 ms [WRITE: 1,015.0 ms]
Latency max
                          : 1106.2 ms [WRITE: 1,106.2 ms]
Total partitions
                               100,000 [WRITE: 100,000]
Total errors
                                     0 [WRITE: 0]
Total GC count
                          : 0
Total GC memory
                          : 0.000 KiB
Total GC time
                               0.0 seconds
                               NaN ms
Avg GC time
StdDev GC time
                               0.0 ms
Total operation time
                          : 00:00:13
END
```

31. Now let's add another node into the cluster.

Quit that container shell (Ctrl-D) and execute this command:

kubectl scale --replicas 4 statefulset/cassandra

Wait for the new instance to be live:





33. Now lets ask the status from Cassandra:

34. Rerun the stress test (Steps 29 and 30).

Unfortunately, we may not be any faster - because now at least 2 pods are on one node (4 pods, 3 nodes).

```
Results:
Op rate
                               5,695 op/s
                                            [WRITE: 5,695 op/s]
Partition rate
                               5,695 pk/s
                                            [WRITE: 5,695 pk/s]
Row rate
                               5,695 row/s [WRITE: 5,695 row/s]
Latency mean
                             164.1 ms [WRITE: 164.1 ms]
                             72.7 ms [WRITE: 72.7 ms]
Latency median
Latency 95th percentile : 590.3 ms [WRITE: 590.3 ms]
Latency 99th percentile : 788.0 ms [WRITE: 788.0 ms]
Latency 99.9th percentile :
                            976.2 ms [WRITE: 976.2 ms]
Latency max
                          : 1406.1 ms [WRITE: 1,406.1 ms]
                               100,000 [WRITE: 100,000]
Total partitions
                                     0 [WRITE: 0]
Total errors
                          : 0
Total GC count
Total GC memory
                          : 0.000 KiB
Total GC time
                               0.0 seconds
Avg GC time
                               NaN ms
StdDev GC time
                               0.0 ms
Total operation time
                          : 00:00:17
```

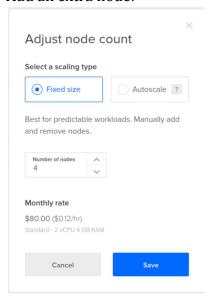
35. Let's fix that:

Go to the Digital Ocean **Kubernetes-> k8s-cass-> Nodes**. Click on the Node Pool ...

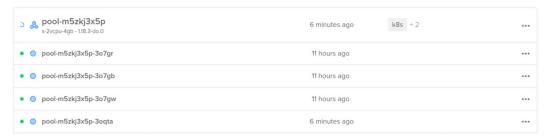




Add an extra node:



36. Wait until the new node is in place:



37. Check from kubectl that the new node is active:

kubectl get nodes

| NAME | STATUS | ROLES | AGE | VERSION |
|----------------------|--------|---------------|-------|---------|
| pool-m5zkj3x5p-3o7gb | Ready | <none></none> | 11h | v1.18.3 |
| pool-m5zkj3x5p-3o7gr | Ready | <none></none> | 11h | v1.18.3 |
| pool-m5zkj3x5p-3o7gw | Ready | <none></none> | 11h | v1.18.3 |
| pool-m5zkj3x5p-3oqta | Ready | <none></none> | 5m57s | v1.18.3 |

- 38. Use either the Kubernetes Dashboard or k9s or kubectl to see how the pods are assigned to nodes. Unfortunately we are unbalanced, because Kubernetes hasn't had any impetus to move the extra pod to the new node.
- 39. Rerun the stress test and see what the performance is like
- 40. We have two options. We could just create a few more Cassandra nodes to more evenly use the nodes. However, there is another option using a cool tool called the Kubernetes Descheduler:

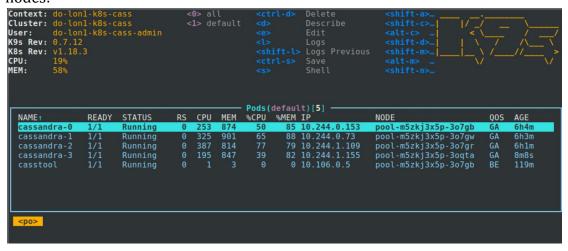
https://github.com/kubernetes-sigs/descheduler



41. In a window do this:

```
git clone
https://github.com/kubernetes-sigs/descheduler.git
cd descheduler
kubectl create -f kubernetes/rbac.yaml
kubectl create -f kubernetes/configmap.yaml
kubectl create -f kubernetes/job.yaml
```

- 42. Watch what happens using k9s
- 43. Once the pod has moved, all the cassandra pods should be on different nodes:



44. Now rerun the stress test one more time:

You can see my results were considerably better:

```
Results:
Op rate
                                           [WRITE: 9,650 op/s]
                               9,650 op/s
Partition rate
                                            [WRITE: 9,650 pk/s]
                          :
                               9,650 pk/s
Row rate
                          :
                               9,650 row/s [WRITE: 9,650 row/s]
Latency mean
                          :
                              98.2 ms [WRITE: 98.2 ms]
Latency median
                         :
                              81.1 ms [WRITE: 81.1 ms]
Latency 95th percentile : 299.9 ms [WRITE: 299.9 ms]
Latency 99th percentile : 464.0 ms [WRITE: 464.0 ms]
Latency 99.9th percentile: 724.6 ms [WRITE: 724.6 ms]
Latency max
                             801.6 ms [WRITE: 801.6 ms]
                          :
Total partitions
                               100,000 [WRITE: 100,000]
                          :
Total errors
                          :
                                     0 [WRITE: 0]
Total GC count
                          : 0
Total GC memory
                            0.000 KiB
Total GC time
                               0.0 seconds
                          :
Avg GC time
                               NaN ms
                          :
StdDev GC time
                               0.0 ms
                          :
                          : 00:00:10
Total operation time
```

45. Congratulations - we have deployed cassandra, scaled it, tested it and increased performance all in a kubernetes cluster.



- 46. Let's clean up.
- 47. Firstly, let's delete our cassandra cluster.

cd ~/cassandra
kubectl delete -f cassandra-statefulset.yaml

48. Delete our casstool pod:

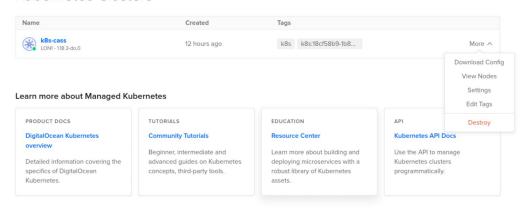
kubectl delete pod/casstool

49. Delete the volumes / volume claims:

kubectl delete persistentvolumeclaim -l app=cassandra

50. Delete the kubernetes cluster (from the DO web panel):

Kubernetes Clusters



Click **Destroy**



Then **Destroy** again.

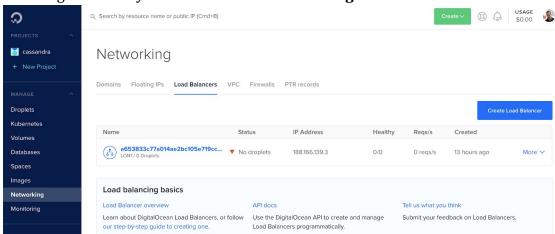


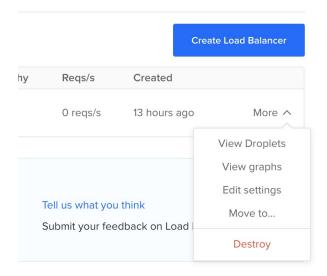
Then enter the cluster name (k8s-cass) and actually finally **Destroy**



51. If you came straight to this lab from the last one, you will also have one last remaining load balancer running.

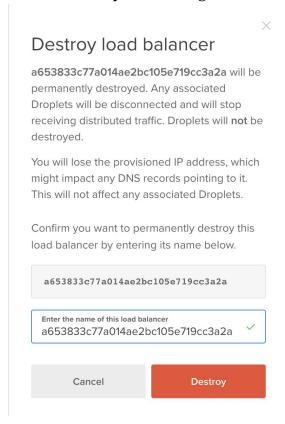
DigitalOcean will also have created a load-balancer to handle the incoming traffic for your service. Go to **Networking -> Load Balancers**







51. Click on **Destroy** and once again enter the name (copy and paste!)



52. This lab is done! Congratulations.

