Containerization and Big Data: Exercise 4-2

Brandon Hosley

July 15, 2020

4 Building an Apache Spark Cluster Using Kubernetes

4.4 Configuring Minikube

```
bhosl2@us2004lts:~/mySparkCluster

bhosl2@us2004lts:~$ mkdir mySparkCluster

bhosl2@us2004lts:~$ cd mySparkCluster$

bhosl2@us2004lts:~/mySparkCluster$ minikube start --cpus 6 --memory 14336

* minikube v1.11.0 on Ubuntu 20.04

* minikube 1.12.0 is available! Download it: https://github.com/kubernetes/minik

ube/releases/tag/v1.12.0

* To disable this notice, run: 'minikube config set WantUpdateNotification false

* Automatically selected the docker driver

* Starting control plane node minikube in cluster minikube

* Creating docker container (CPUs=6, Memory=14336MB) ...

* Preparing Kubernetes v1.18.3 on Docker 19.03.2 ...

- kubeadm.pod-network-cidr=10.244.0.0/16

* Verifying Kubernetes components...

* Enabled addons: default-storageclass, storage-provisioner

* Done! kubectl is now configured to use "minikube"

bhosl2@us2004lts:~/mySparkCluster$
```

4.5 Creating Spark Docker Images

```
bhosl2@us2004lts: ~/mySparkCluster/spark-2.4.6-bin-hadoop2.7
bhos12@us20041ts:~/mySparkCluster/spark-2.4.6-bin-hadoop2.7$ ls
                    RELEASE conf examples kubernetes python
bin data jars licenses sbin
LICENSE R
       README.md bin
phos12@us2004lts:~/mySparkCluster/spark-2.4.6-bin-hadoop2.7$ ls bin
peeline
                     pyspark
                                      spark-class2.cmd spark-submit
                                        spark-shell
                                                           spark-submit.cmd
peeline.cmd
                      pyspark.cmd
docker-image-tool.sh pyspark2.cmd
                                                           spark-submit2.cmd
                                        spark-shell.cmd
ind-spark-home
                      run-example
                                        spark-shell2.cmd
                                                           sparkR
find-spark-home.cmd
                      run-example.cmd spark-sql
oad-spark-env.cmd
                                        spark-sql.cmd
                                                           sparkR2.cmd
oad-spark-env.sh
                      spark-class.cmd
                                       spark-sql2.cmd
bhos12@us20041ts:~/mySparkCluster/spark-2.4.6-bin-hadoop2.7$
```

```
### bhosl2@us2004lts:~/mySparkCluster/spark-2.4.6-bin-hadcop2.7$

bhosl2@us2004lts:~/mySparkCluster/spark-2.4.6-bin-hadcop2.7$

docker image 1s

REPOSITORY

TAG

IMAGE ID

CREATED

SIZE

$ spark-r

$ v1.0.uis

$ ee40adeda556

$ 3 minutes ago

$ 1.11GB

$ spark-py

$ v1.0.uis

$ 9841553098a2

$ 4 minutes ago

$ 1.06GB

$ spark

$ v1.0.uis

$ b1c210a4fefd

$ 5 minutes ago

$ 1.7MB

$ k8s.gcr.io/kube-proxy

$ v1.18.3

$ 3439b754629

$ weeks ago

$ 117MB

$ k8s.gcr.io/kube-controller-manager

$ v1.18.3

$ 762863876bd

$ weeks ago

$ 162MB

$ k8s.gcr.io/kube-scheduler

$ v1.18.3

$ 76216634ed0c

$ weeks ago

$ 95.3MB

$ k8s.gcr.io/kube-scheduler

$ v1.18.3

$ 76216634ed0c

$ weeks ago

$ 95.3MB

$ webers ago

$ 173MB

$ 88.gcr.io/kube-scheduler

$ v1.18.3

$ 76216634ed0c

$ weeks ago

$ 95.3MB

$ webers ago

$ 173MB

$ 88.gcr.io/kube-scheduler

$ v1.18.3

$ 76216634ed0c

$ weeks ago

$ 95.3MB

$ webers ago

$ 173MB

$ 88.gcr.io/coredns

$ 1.6.7

$ 674a37a3836

$ 5 months ago

$ 284MB

$ k8s.gcr.io/coredns

$ 1.6.7

$ 674a37a38360

$ 5 months ago

$ 43.8MB

$ k8s.gcr.io/coredns

$ 1.6.7

$ 674a37a38360

$ 5 months ago

$ 288MB

$ k8s.gcr.io/coredns

$ 1.6.7

$ 674a37a38360

$ 5 months ago

$ 288MB

$ k8s.gcr.io/coredns

$ 1.6.7

$ 674a37a38360

$ 5 months ago

$ 8038B

$ webers ago

$ 173MB

$ 1.6.7

$ 674a37a38360

$ 5 months ago

$ 288MB

$ 889.gcr.io/coredns

$ 1.6.7

$ 674a37a38360

$ 5 months ago

$ 288MB

$ 889.gcr.io/k8s-minikube/storage-provisioner

$ v1.0.2

$ 300866160379

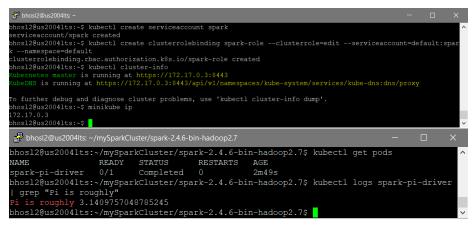
$ 8 months ago

$ 80.8MB

$ webrentestorage

$ 80.8MB
```

4.6 Running Spark App



4.7 Minikube Dashboard

