1. Deploy an app

Like minikube, kubectl comes installed in the online terminal. Type kubectl in the terminal to see its usage. The common format of a kubectl command is: kubectl action resource. This performs the specified action (like create, describe) on the specified resource (like node, container). You can use --help after the command to get additional info about possible parameters (kubectl get nodes --help).

Check that kubectl is configured to talk to your cluster, by running the kubectl version command:

$

$ sleep 1; launch.sh

Starting Kubernetes...

Kubernetes Started

$ kubectl version

Client Version: version.Info{Major:"1", Minor:"11", GitVersion:"v1.11.0", GitCommit:"91e7b4fd31fcd3d5f436da26c980becec37ceefe", GitTreeState:"clean", BuildDate:"2018-06-27T20:17:28Z", GoVersion:"go1.10.2", Compiler:"gc", Platform:"linux/amd64"}

Server Version: version.Info{Major:"1", Minor:"10", GitVersion:"v1.10.0", GitCommit:"fc32d2f3698e36b93322a3465f63a14e9f0eaead", GitTreeState:"clean", BuildDate:"2018-04-10T12:46:31Z", GoVersion:"go1.9.4", Compiler:"gc", Platform:"linux/amd64"}

--

To view the nodes in the cluster, run the kubectl get nodes command:

kubectl get nodes

NAME STATUS ROLES AGE VERSION

minikube Ready <none> 1m v1.10.0

Here we see the available nodes (1 in our case). Kubernetes will choose where to deploy our application based on Node available resources.

Continue…………………………..

performed a few things for you:

* searched for a suitable node where an instance of the application could be run (we have only 1 available node)
* scheduled the application to run on that Node
* configured the cluster to reschedule the instance on a new Node when needed

To list your deployments use the get deployments command:

Pods that are running inside Kubernetes are running on a private, isolated network. By default they are visible from other pods and services within the same kubernetes cluster, but not outside that network. When we use kubectl, we're interacting through an API endpoint to communicate with our application.

We will cover other options on how to expose your application outside the kubernetes cluster in Module 4.

The kubectl command can create a proxy that will forward communications into the cluster-wide, private network. The proxy can be terminated by pressing control-C and won't show any output while its running.

We will open a second terminal window to run the proxy.

kubectl proxy

We now have a connection between our host (the online terminal) and the Kubernetes cluster. The proxy enables direct access to the API from these terminals.

You can see all those APIs hosted through the proxy endpoint, now available at through [http://localhost:8001](http://localhost:8001/). For example, we can query the version directly through the API using the curl command:

curl http://localhost:8001/version

The API server will automatically create an endpoint for each pod, based on the pod name, that is also accessible through the proxy.

First we need to get the Pod name, and we'll store in the environment variable POD\_NAME:

export POD\_NAME=$(kubectl get pods -o go-template --template '{{range .items}}{{.metadata.name}}{{"\n"}}{{end}}') echo Name of the Pod: $POD\_NAME

Now we can make an HTTP request to the application running in that pod:

curl http://localhost:8001/api/v1/namespaces/default/pods/$POD\_NAME/proxy/

The url is the route to the API of the Pod.

*Note: Check the top of the terminal. The proxy was run in a new tab (Terminal 2), and the recent commands were executed the original tab (Terminal 1). The proxy still runs in the second tab, and this allowed our curl command to work using localhost:8001.*

Expose your app -🡪interactive tutorial – exploring your app

Let’s verify that the application we deployed in the previous scenario is running. We’ll use the kubectl get command and look for existing Pods:

kubectl get pods

If no pods are running, list the Pods again.

Next, to view what containers are inside that Pod and what images are used to build those containers we run the describe pods command:

kubectl describe pods

We see here details about the Pod’s container: IP address, the ports used and a list of events related to the lifecycle of the Pod.

The output of the describe command is extensive and covers some concepts that we didn’t explain yet, but don’t worry, they will become familiar by the end of this bootcamp.

*Note: the describe command can be used to get detailed information about most of the kubernetes primitives: node, pods, deployments. The describe output is designed to be human readable, not to be scripted against.*

**Step 2 Show the app in the terminal**

Recall that Pods are running in an isolated, private network - so we need to proxy access to them so we can debug and interact with them. To do this, we'll use the kubectl proxy command to run a proxy in a second terminal window. Click on the command below to automatically open a new terminal and run the proxy:

kubectl proxy

Now again, we'll get the Pod name and query that pod directly through the proxy. To get the Pod name and store it in the POD\_NAME environment variable:

export POD\_NAME=$(kubectl get pods -o go-template --template '{{range .items}}{{.metadata.name}}{{"\n"}}{{end}}') echo Name of the Pod: $POD\_NAME

To see the output of our application, run a curl request.

curl http://localhost:8001/api/v1/namespaces/default/pods/$POD\_NAME/proxy/

The url is the route to the API of the Pod.

#### Step 3 View the container logs

Anything that the application would normally send to STDOUT becomes logs for the container within the Pod. We can retrieve these logs using the kubectl logs command:

kubectl logs $POD\_NAME

Note: We don’t need to specify the container name, because we only have one container inside the pod.

#### Step 4 Executing command on the container

We can execute commands directly on the container once the Pod is up and running. For this, we use the exec command and use the name of the Pod as a parameter. Let’s list the environment variables:

kubectl exec $POD\_NAME env

Again, worth mentioning that the name of the container itself can be omitted since we only have a single container in the Pod.

Next let’s start a bash session in the Pod’s container:

kubectl exec -ti $POD\_NAME bash

We have now an open console on the container where we run our NodeJS application. The source code of the app is in the server.js file:

cat server.js

You can check that the application is up by running a curl command:

curl localhost:8080

Note: here we used localhost because we executed the command inside the NodeJS container

To close your container connection type exit.

----continue

Expose your app publicly – interactive tutorial – exposing your app

Let’s verify that our application is running. We’ll use the kubectl get command and look for existing Pods:

kubectl get pods

Next let’s list the current Services from our cluster:

kubectl get services

We have a Service called kubernetes that is created by default when minikube starts the cluster. To create a new service and expose it to external traffic we’ll use the expose command with NodePort as parameter (minikube does not support the LoadBalancer option yet).

kubectl expose deployment/kubernetes-bootcamp --type="NodePort" --port 8080

Let’s run again the get services command:

kubectl get services

We have now a running Service called kubernetes-bootcamp. Here we see that the Service received a unique cluster-IP, an internal port and an external-IP (the IP of the Node).

To find out what port was opened externally (by the NodePort option) we’ll run the describe service command:

kubectl describe services/kubernetes-bootcamp

Create an environment variable called NODE\_PORT that has the value of the Node port assigned:

export NODE\_PORT=$(kubectl get services/kubernetes-bootcamp -o go-template='{{(index .spec.ports 0).nodePort}}') echo NODE\_PORT=$NODE\_PORT

Now we can test that the app is exposed outside of the cluster using curl, the IP of the Node and the externally exposed port:

curl $(minikube ip):$NODE\_PORT

And we get a response from the server. The Service is exposed.

#### Step 2: Using labels

The Deployment created automatically a label for our Pod. With describe deployment command you can see the name of the label:

kubectl describe deployment

Let’s use this label to query our list of Pods. We’ll use the kubectl get pods command with -l as a parameter, followed by the label values:

kubectl get pods -l run=kubernetes-bootcamp

You can do the same to list the existing services:

kubectl get services -l run=kubernetes-bootcamp

Get the name of the Pod and store it in the POD\_NAME environment variable:

export POD\_NAME=$(kubectl get pods -o go-template --template '{{range .items}}{{.metadata.name}}{{"\n"}}{{end}}') echo Name of the Pod: $POD\_NAME

To apply a new label we use the label command followed by the object type, object name and the new label:

kubectl label pod $POD\_NAME app=v1

This will apply a new label to our Pod (we pinned the application version to the Pod), and we can check it with the describe pod command:

kubectl describe pods $POD\_NAME

We see here that the label is attached now to our Pod. And we can query now the list of pods using the new label:

kubectl get pods -l app=v1

And we see the Pod.

#### Step 3 Deleting a service

To delete Services you can use the delete service command. Labels can be used also here:

kubectl delete service -l run=kubernetes-bootcamp

Confirm that the service is gone:

kubectl get services

This confirms that our Service was removed. To confirm that route is not exposed anymore you can curl the previously exposed IP and port:

curl $(minikube ip):$NODE\_PORT

This proves that the app is not reachable anymore from outside of the cluster. You can confirm that the app is still running with a curl inside the pod:

kubectl exec -ti $POD\_NAME curl localhost:8080

We see here that the application is up.

Scale Your App

We should have 1 Pod. If not, run the command again. This shows:

The DESIRED state is showing the configured number of replicas

The CURRENT state show how many replicas are running now

The UP-TO-DATE is the number of replicas that were updated to match the desired (configured) state

The AVAILABLE state shows how many replicas are actually AVAILABLE to the users

Next, let’s scale the Deployment to 4 replicas. We’ll use the kubectl scale command, followed by the deployment type, name and desired number of instances:

kubectl scale deployments/kubernetes-bootcamp --replicas=4

To list your Deployments once again, use get deployments:

kubectl get deployments

The change was applied, and we have 4 instances of the application available. Next, let’s check if the number of Pods changed:

kubectl get pods -o wide

There are 4 Pods now, with different IP addresses. The change was registered in the Deployment events log. To check that, use the describe command:

kubectl describe deployments/kubernetes-bootcamp

You can also view in the output of this command that there are 4 replicas now.

#### Step 2: Load Balancing

Let’s check that the Service is load-balancing the traffic. To find out the exposed IP and Port we can use the describe service as we learned in the previously Module:

kubectl describe services/kubernetes-bootcamp

Create an environment variable called NODE\_PORT that has a value as the Node port:

export NODE\_PORT=$(kubectl get services/kubernetes-bootcamp -o go-template='{{(index .spec.ports 0).nodePort}}') echo NODE\_PORT=$NODE\_PORT

Next, we’ll do a curl to the exposed IP and port. Execute the command multiple times:

curl $(minikube ip):$NODE\_PORT

We hit a different Pod with every request. This demonstrates that the load-balancing is working

#### Step 3: Scale Down

To scale down the Service to 2 replicas, run again the scale command:

kubectl scale deployments/kubernetes-bootcamp --replicas=2

List the Deployments to check if the change was applied with the get deployments command:

kubectl get deployments

The number of replicas decreased to 2. List the number of Pods, with get pods:

kubectl get pods -o wide

This confirms that 2 Pods were terminated.

Update the app

#### Step 1: Update the version of the app

To list your deployments use the get deployments command: kubectl get deployments

To list the running Pods use the get pods command:

kubectl get pods

To view the current image version of the app, run a describe command against the Pods (look at the Image field):

kubectl describe pods

To update the image of the application to version 2, use the set image command, followed by the deployment name and the new image version:

kubectl set image deployments/kubernetes-bootcamp kubernetes-bootcamp=jocatalin/kubernetes-bootcamp:v2

The command notified the Deployment to use a different image for your app and initiated a rolling update. Check the status of the new Pods, and view the old one terminating with the get pods command:

kubectl get pods

#### Step 2: Verify an update

First, let’s check that the App is running. To find out the exposed IP and Port we can use describe service:

kubectl describe services/kubernetes-bootcamp

Create an environment variable called NODE\_PORT that has the value of the Node port assigned:

export NODE\_PORT=$(kubectl get services/kubernetes-bootcamp -o go-template='{{(index .spec.ports 0).nodePort}}') echo NODE\_PORT=$NODE\_PORT

Next, we’ll do a curl to the the exposed IP and port:

curl $(minikube ip):$NODE\_PORT

We hit a different Pod with every request and we see that all Pods are running the latest version (v2).

The update can be confirmed also by running a rollout status command:

kubectl rollout status deployments/kubernetes-bootcamp

To view the current image version of the app, run a describe command against the Pods:

kubectl describe pods

We run now version 2 of the app (look at the Image field)

#### Step 3: Rollback an update

Let’s perform another update, and deploy image tagged as v10 :

kubectl set image deployments/kubernetes-bootcamp kubernetes-bootcamp=gcr.io/google-samples/kubernetes-bootcamp:v10

Use get deployments to see the status of the deployment:

kubectl get deployments

And something is wrong… We do not have the desired number of Pods available. List the Pods again:

kubectl get pods

A describe command on the Pods should give more insights:

kubectl describe pods

There is no image called v10 in the repository. Let’s roll back to our previously working version. We’ll use the rollout undo command:

kubectl rollout undo deployments/kubernetes-bootcamp

The rollout command reverted the deployment to the previous known state (v2 of the image). Updates are versioned and you can revert to any previously know state of a Deployment. List again the Pods:

kubectl get pods

Four Pods are running. Check again the image deployed on the them:

kubectl describe pods

We see that the deployment is using a stable version of the app (v2). The Rollback was successful.