Introduction to Kubernetes

Objectives

In this lab, you will:

- Use the `kubectl` CLI
- Create a Kubernetes Pod
- Create a Kubernetes Deployment
- Create a ReplicaSet that maintains a set number of replicas
- Witness Kubernetes load balancing in action

> **Note: Kindly complete the lab in a single session without any break because the lab may go on offline mode and may cause errors. If you face any issues/errors during the lab process, please logout from the lab environment. Then clear your system cache and cookies and try to complete the lab.**

Verify the environment and command line tools

1. If a terminal is not already open, open a terminal window by using the menu in the editor: `Terminal > New Terminal`.

   > **Note: Please skip this step if the terminal already appears.**

2. Verify that `kubectl` CLI is installed.

   ```
   kubectl version
   ```

You should see the following output, although the versions may be different:

3. Change to your project folder.

   > **Note: Please skip this step if you are already on the '/home/project' directory**

   ```
   cd /home/project
   ```

4. Clone the git repository that contains the artifacts needed for this lab, if it doesn't already exist.

   ```
   [ ! -d 'CC201' ]  && git clone https://github.com/ibm-developer-skills-network/abahi-CC_201_labs.git CC201
   ```

5. Change to the directory for this lab by running the following command. `cd` will change the working/current directory to the directory with the name specified, in this case **CC201/labs/2_IntroKubernetes**.

   ```
   cd CC201/labs/2_IntroKubernetes/
   ```

6. List the contents of this directory to see the artifacts for this lab.

   ```
   ls
   ```

Use the `kubectl` CLI

Recall that Kubernetes namespaces enable you to virtualize a cluster. You already have access to one namespace in a Kubernetes cluster, and `kubectl` is already set to target that cluster and namespace.

Let's look at some basic `kubectl` commands.

1. `kubectl` requires configuration so that it targets the appropriate cluster. Get cluster information with the following command:

   ```
   kubectl config get-clusters
   ```

2. A `kubectl` context is a group of access parameters, including a cluster, a user, and a namespace. View your current context with the following command:

   ```
   kubectl config get-contexts
   ```

3. List all the Pods in your namespace. If this is a new session for you, you will not see any Pods.

   ```
   kubectl get pods
   ```

Create a Pod with an imperative command

Now it's time to create your first Pod. This Pod will run the `hello-world` image you built and pushed to IBM Cloud Container Registry in the last lab. As explained in the videos for this module, you can create a Pod imperatively or declaratively. Let's do it imperatively first.

1. Export your namespace as an environment variable so that it can be used in subsequent commands.

   ```
   export MY_NAMESPACE=sn-labs-$USERNAME
   ```

2. Click the Explorer icon (it looks like a sheet of paper) on the left side of the window, and then navigate to the directory for this lab: `CC201 > labs > 2_IntroKubernetes`. Click on `Dockerfile`. This is the file that will be used to build our image.

3. Build and push the image again, as it may have been deleted automatically since you completed the first lab.

   ```
   docker build -t us.icr.io/$MY_NAMESPACE/hello-world:1 . && docker push us.icr.io/$MY_NAMESPACE/hello-world:1
   ```

4. Run the `hello-world` image as a container in Kubernetes.

```
kubectl run hello-world --image us.icr.io/$MY_NAMESPACE/hello-world:1 --overrides='{"spec":{"imagePullSecrets":[{"name":"icr"}],"containers":[{"
```

The `--overrides` option here enables us to specify the needed credentials to pull this image from IBM Cloud Container Registry. Note that this is an imperative command, as we told Kubernetes explicitly what to do: run `hello-world`.

5. List the Pods in your namespace.

```
kubectl get pods
```

Great, the previous command indeed created a Pod for us. You can see an auto-generated name was given to this Pod.

You can also specify the wide option for the output to get more details about the resource.

```
kubectl get pods -o wide
```

6. Describe the Pod to get more details about it.

```
kubectl describe pod hello-world
```

Note: The output shows the pod parameters like **Namespace, Pod Name, IP address, the time when the pod started running** and also the container parameters like **container ID, image name & ID, running status and the memory/CPU limits.**

7. Delete the Pod.

```
kubectl delete pod hello-world
```

This command takes a while to execute the deletion of the pod. Please wait till the terminal prompt appears again.

8. List the Pods to verify that none exist.

```
kubectl get pods
```

Create a Pod with imperative object configuration

Imperative object configuration lets you create objects by specifying the action to take (e.g., create, update, delete) while using a configuration file. A configuration file, `hello-world-create.yaml`, is provided to you in this directory.

1. Use the Explorer to view and edit the configuration file. Click the Explorer icon (it looks like a sheet of paper) on the left side of the window, and then navigate to the directory for this lab: `CC201 > labs > 2_IntroKubernetes`. Click `hello-world-create.yaml` to view the configuration file.

2. Use the Explorer to edit `hello-world-create.yaml`. You need to insert your namespace where it says `<my_namespace>`. Make sure to save the file when you're done.

3. Imperatively create a Pod using the provided configuration file.

```
kubectl create -f hello-world-create.yaml
```

Note that this is indeed imperative, as you explicitly told Kubernetes to *create* the resources defined in the file.

4. List the Pods in your namespace.

```
kubectl get pods
```

5. Delete the Pod.

```
kubectl delete pod hello-world
```

This command takes a while to execute the deletion of the pod. Please wait till the terminal prompt appears again.

6. List the Pods to verify that none exist.

```
kubectl get pods
```

Create a Pod with a declarative command

The previous two ways to create a Pod were imperative – we explicitly told `kubectl` what to do. While the imperative commands are easy to understand and run, they are not ideal for a production environment. Let's look at declarative commands.

1. A sample `hello-world-apply.yaml` file is provided in this directory. Use the Explorer again to open this file. Notice the following:

- We are creating a Deployment (`kind: Deployment`).
- There will be three replica Pods for this Deployment (`replicas: 3`).
- The Pods should run the `hello-world` image (`- image: us.icr.io/<my_namespace>/hello-world:1`).

You can ignore the rest for now. We will get to a lot of those concepts in the next lab.

2. Use the Explorer to edit `hello-world-apply.yaml`. You need to insert your namespace where it says `<my_namespace>`. Make sure to save the file when you're done.

3. Use the `kubectl apply` command to set this configuration as the desired state in Kubernetes.

```
kubectl apply -f hello-world-apply.yaml
```

4. Get the Deployments to ensure that a Deployment was created.

```
kubectl get deployments
```

5. List the Pods to ensure that three replicas exist.

```
kubectl get pods
```

With declarative management, we did not tell Kubernetes which actions to perform. Instead, `kubectl` inferred that this Deployment needed to be created. If you delete a Pod now, a new one will be created in its place to maintain three replicas.

6. Note one of the Pod names from the previous step, replace the `pod_name` in the following command with the pod name that you noted and delete that Pod and list the pods. To see one pod being terminated, there by having just 2 pods, we will follow the **delete**, immediately with **get**.

```
kubectl delete pod <pod_name> && kubectl get pods
```

This command takes a while to execute the deletion of the pod. Please wait till the terminal prompt appears again.

7. List the Pods to see a new one being created.

   You may have to run this command a few times as it may take a while to create the new pod.

```
kubectl get pods
```

```
NAME                     READY   STATUS    RESTARTS   AGE
hello-world-774ddf45b5-28k7j   1/1     Running   0          36s
hello-world-774ddf45b5-9cbv2   1/1     Running   0          112s
hello-world-774ddf45b5-svpf7   1/1     Running   0          112s
```

The output should reflect three pods running.

Load balancing the application

Since there are three replicas of this application deployed in the cluster, Kubernetes will load balance requests across these three instances. Let's expose our application to the internet and see how Kubernetes load balances requests.

1. In order to access the application, we have to expose it to the internet using a Kubernetes Service.

   ```
   kubectl expose deployment/hello-world
   ```

This command creates what is called a ClusterIP Service. This creates an IP address that accessible within the cluster.

2. List Services in order to see that this service was created.

   ```
   kubectl get services
   ```

3. Open a new split terminal window by locate the split icon in the top-right corner of the terminal panel.

4. Since the cluster IP is not accessible outside of the cluster, we need to create a proxy. Note that this is not how you would make an application externally accessible in a production scenario. Run this command in the new terminal window since your environment variables need to be accessible in the original window for subsequent commands.

   ```
   kubectl proxy
   ```

This command doesn't terminate until you terminate it. Keep it running so that you can continue to access your app.

5. In the original terminal window, ping the application to get a response.

   ```
   curl -L localhost:8001/api/v1/namespaces/sn-labs-$USERNAME/services/hello-world/proxy
   ```

Notice that this output includes the Pod name.

6. Execute the following command to send ten consecutive requests to the hello-world service via the Kubernetes API proxy. As each request is forwarded, note the pod name in the response (which shows which pod handled the request).

   ```
   for i in `seq 10`; do curl -L localhost:8001/api/v1/namespaces/sn-labs-$USERNAME/services/hello-world/proxy; done
   ```

You should see more than one Pod name, and quite possibly all three Pod names, in the output. This is because Kubernetes load balances the requests across the three replicas, so each request could hit a different instance of our application.

7. Delete the Deployment and Service. This can be done in a single command by using slashes.

   ```
   kubectl delete deployment/hello-world service/hello-world
   ```

   Note: If you face any issues in typing further commands in the terminal, press Enter.

8. Return to the terminal window running the `proxy` command and kill it using `Ctrl+C`.

Congratulations! You have completed the lab for the second module of this course.