Exercise 16.1: High Availability Steps

Overview

In this lab we will add two more control planes to our cluster, change taints and deploy an application to a particular node, and test that we can access it from outside the cluster. The nodes will handle various infrastructure services and the **etcd** database and should be sized accordingly.

The steps are presented in two ways. First the general steps for those interested in more of a challenge. Following that will be the detailed steps found in previous labs.

Very Important

If using **cri-o** the way the server is kept track of does not allow adding nodes. You will need to create a new cluster and pass different options to **kubeadm init**. More information can be found here: https://kubernetes.io/docs/setup/production-environment/tools/kubeadm/high-availability/#first-steps-for-both-methods

You will need three more nodes. One to act as a load balancer, the other two will act as cp nodes for quorum. Log into each and use the **ip** command to fill in the table with the IP addresses of the primary interface of each node. If using **GCE** nodes it would be `ens4`, yours may be different. You may need to install software such an editor on the nodes.

Proxy Node	
Second Control Plane	
Third Control Plane	

As the prompts may look similar you may want to change the terminal color or other characteristics to make it easier to keep them distinct. You can also change the prompt using something like: **PS1=”ha-proxy$ ”**, which may help to keep the terminals distinct.

1. Deploy a load balancer configured to pass through traffic. HAProxy is easy to deploy. Start with just the working cp.
2. Install the Kubernetes software on the second and third cp.
3. Join the second cp to the cluster using the node name. You will need an additional hash and flag from the **kubeadm join** you used to add a worker node.
4. Join the third cp to the cluster using the node name.
5. Update the proxy to use all three cps
6. Temporarily shut down the first cp and monitor traffic