

Week 8

Learnings

Resources for Pods and Containers

What's the difference between resource limits vs requests?

When you specify a limit, the kubelet enforces these limits. A request is the amount you would like the kubelet to maintain on average for your pod.

Types

CPU

Represents compute processing and is specified in units of "Kubernetes CPUs".

Memory

Specified in units of bytes.

Additional

What is a kubelet as mentioned above?

A kubelet is defined as an "agent" that runs on each node in a cluster. It controls the way pods do things. For example, it can request more resources,

Problems

Pods wait for a long time before they scale

Based on [this API reference](#) it looks like the normal behavior is to check the scale up/scale down metrics every 300 seconds (5 minutes).

Pods are taking too long to start up

At first, I thought this was because minikube was not loading the correct image. Previously I made an image that takes a very long time to start up so that I could show health checks properly working. It turns out that the container was just taking a really long time because it did not have enough resources to start up on time.

Horizontal Pod Autoscaler is not scaling down