Kubernetes Practical Intro for Data Science

With a fun Google slides theme.

Learning Objectives

By the end of this presentation, you should be able to…

- Understand why people use Kubernetes
- Describe some basic buildings blocks at the high level
- Deploy a simple ML app on a local Kubernetes cluster

We will NOT be covering…

- Kubernetes internal architecture
- Cluster setup and management
- Or really anything super in depth and out of my league

What is Kubernetes?

“*Kubernetes, also known as K8s, is an open-source system for automating deployment, scaling, and management of containerized applications.*” - The K8s Website

OK? BUT WHAT DOES THAT EVEN MEAN????

Deploying and Managing A Simple Application Without K8s

To explain why we might want to use Kubernetes, let's look at a simple example of deploying and *managing* an application without it.

Here's a dramatically oversimplified workflow...

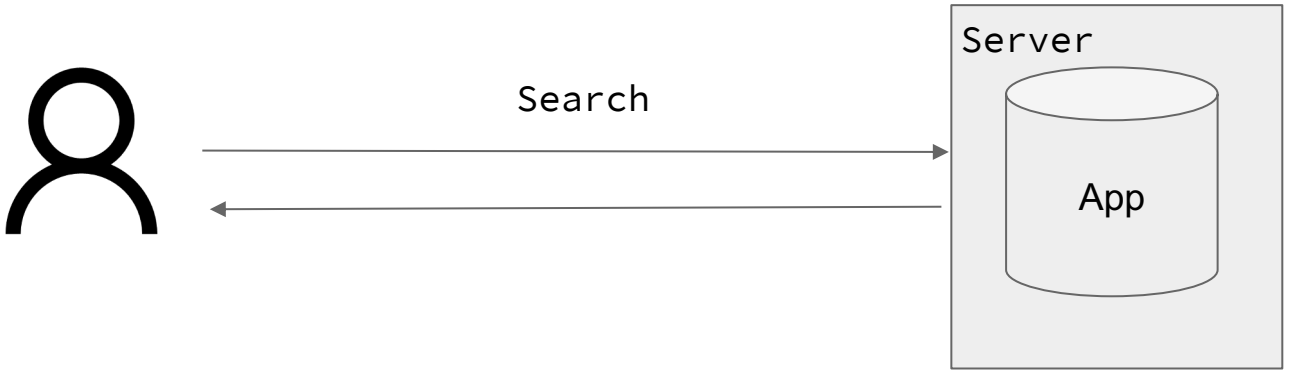
1. You build a simple search app

2. You deploy the app on a remote server

3. Users access the app to get Search results

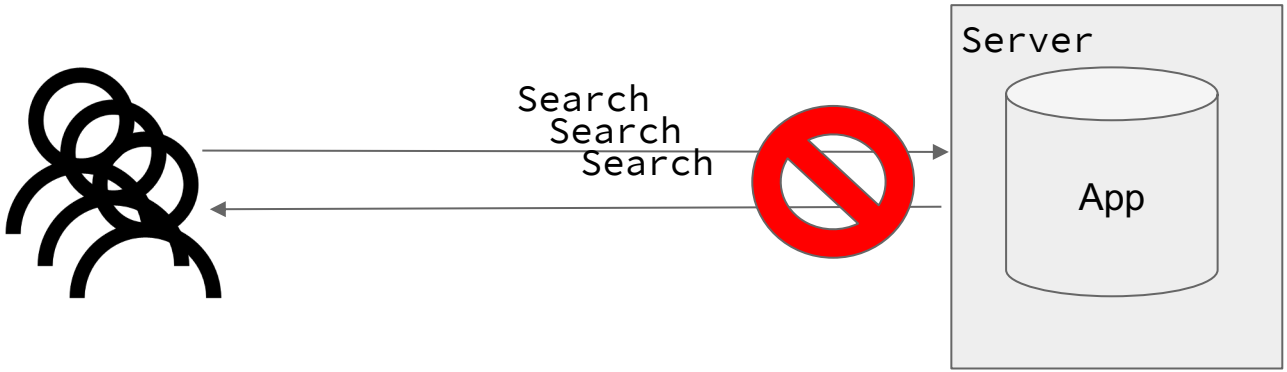
Deploying and Managing A Simple Application Without K8s

When 1 user accesses the application, things are fine.



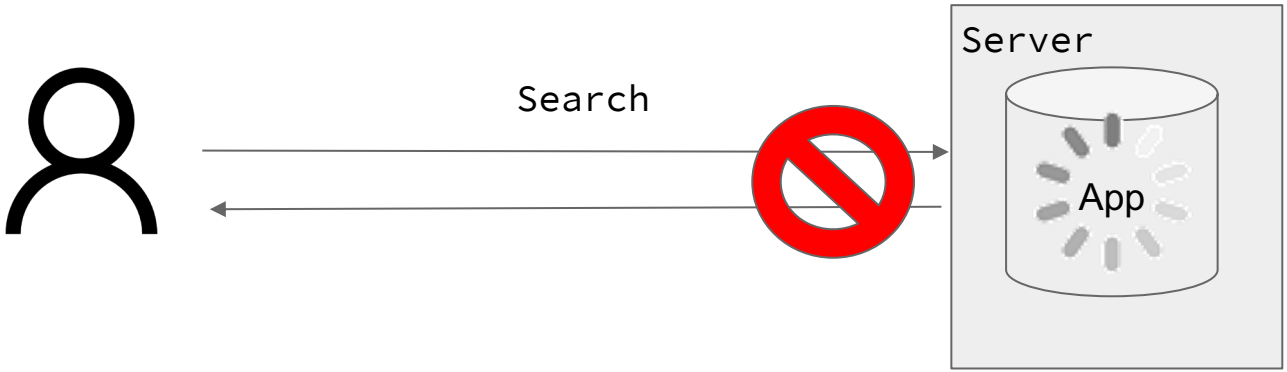
Deploying and Managing A Simple Application Without K8s

But when 50 users accesses the application, the single instance gets overloaded!



Deploying and Managing A Simple Application Without K8s

If we need to upgrade the application, it goes down making searches unavailable.



Some Reasons We Need Kubernetes

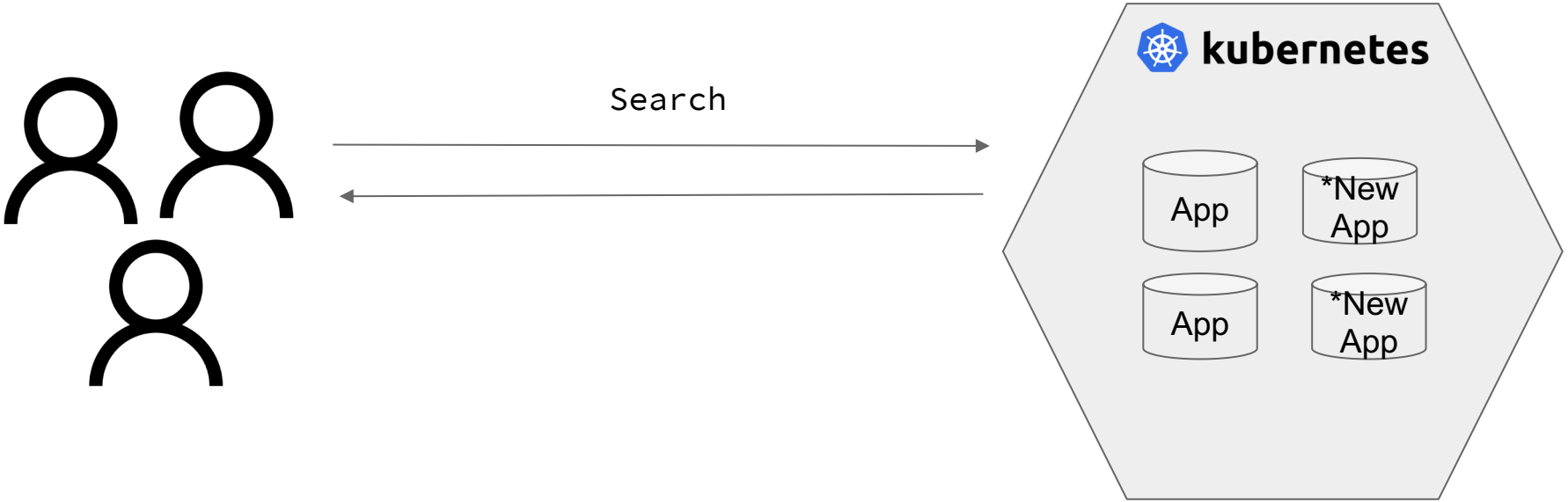
There are many more examples, but as you can see deploying an application is only a small portion of the battle. Managing an application is significantly more difficult to do automatically at scale. We need to:

- Scale applications according to workload
- Make changes without application downtime
- Distribute user traffic to multiple instances of our application

Kubernetes is **declarative**, which means we tell it what we want to happen and Kubernetes does it for us.

Deploying and Managing A Simple Application With* K8s

Kubernetes can automatically scale your app up and down, distribute traffic and roll out new versions of your app with 0 downtime across a cluster of nodes. And much more!



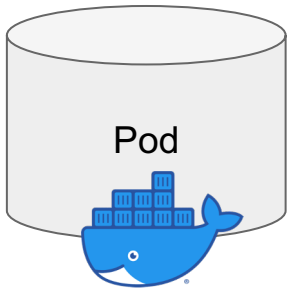
Kubernetes Objects 101

Kubernetes objects are persistent entities in the Kubernetes ecosystem. They defined how applications run on the cluster. We'll look at:

- Pods
- Deployments
- Services
- Ingresses

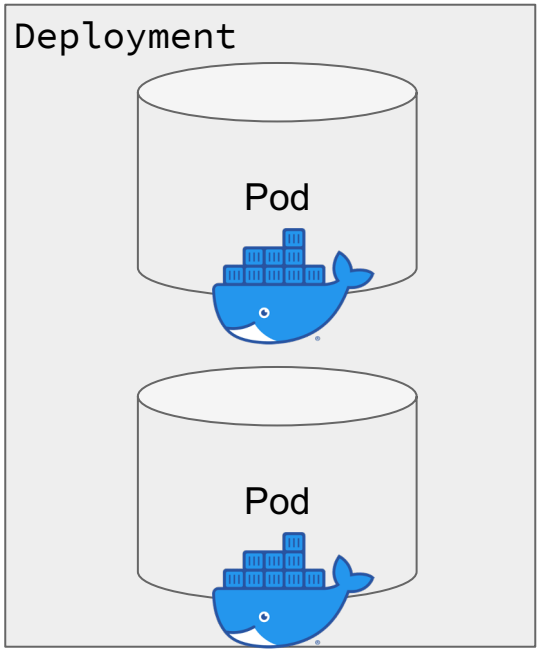
Kubernetes Objects 101 - Pods

Pods are the smallest deployable units of computing you can deploy in Kubernetes. They’re typically instances of a containerized application. Basically, running Docker containers.



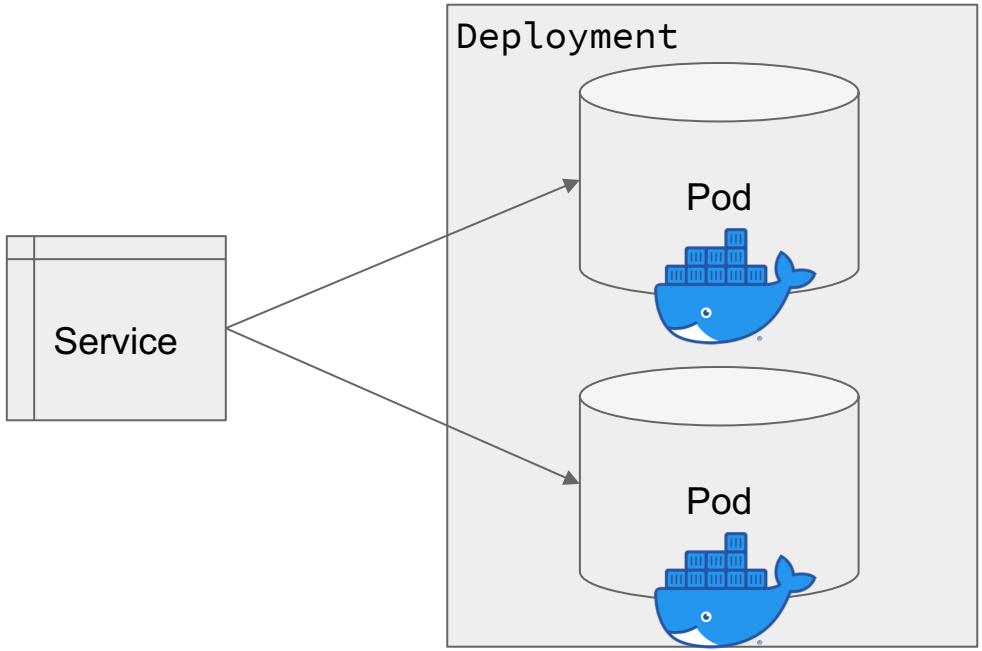
Kubernetes Objects 101 - Deployments

Deployments provide a declarative definition for 1 or pods, making it easy to scale apps.



Kubernetes Objects 101 - Services

Services provide a lookup for pods in a deployment. This is a layer of abstraction over many instances of an application.



Kubernetes Objects 101 - Ingresses

Ingresses provide entry into the cluster for requests so that you can expose applications to the world.