HANDS ON WORKSHOP

Deploy Your Application at Scale in Kubernetes

Gineesh Madapparambath





Kubernetes Community Days Kerala 2024

What we'll discuss today

- Deployment Imperative and Declarative
- Fundamental of scaling
- Metrics and monitoring
- Auto scaling based on metrics
- Quick Demo
- QnA

Target Audience

• Beginners to Kubernetes

Takeaways for Attendees

• Basic understanding of deploying and scaling applications on Kubernetes.

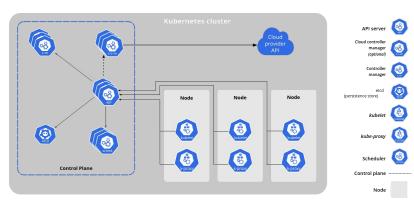
\$ whoami

- Focusing on Ansible Automation,
 Containerization (OpenShift & Kubernetes)
 and Infrastructure as Code (Terraform)
- Published a book,
 Ansible for Real-Life Automation (ansiblehandbook.com)
- Write at techbeatly and Red Hat #EnableSysadmin



What is Kubernetes?

- **Kubernetes**, often abbreviated as K8s, is an open-source container orchestration platform.
- It automates the deployment, scaling, and management of containerized applications, making it easier to manage containerized workloads.



What is a Pod?

- In Kubernetes, a Pod is the smallest and simplest unit.
- It represents a single instance of a running process in a cluster.
- Pods are used to deploy and manage containers, sharing the same network namespace and storage, enabling easy communication.



Deployment vs ReplicaSet vs Replication Controller

Deployment

- Manages the deployment and scaling of a set of Pods.
- Allows declarative updates to applications.
- Provides rolling updates and rollbacks.

ReplicaSet

- Ensures a specified number of replicas of a Pod are running at all times.
- Often managed by Deployments.
- Provides high availability and scalability.

Replication Controller

- Legacy concept, mostly replaced by ReplicaSets and Deployments.
- Ensures a specified number of replicas of a Pod are running.
- Does not support rolling updates and other advanced deployment strategies.

The Deployment Methods

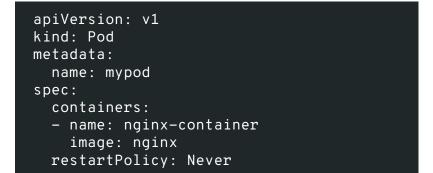
```
$ kubectl run mypod --image=nginx \
   --restart=Never
```



Imperative

Fast but not useful when repeat

Chance for error



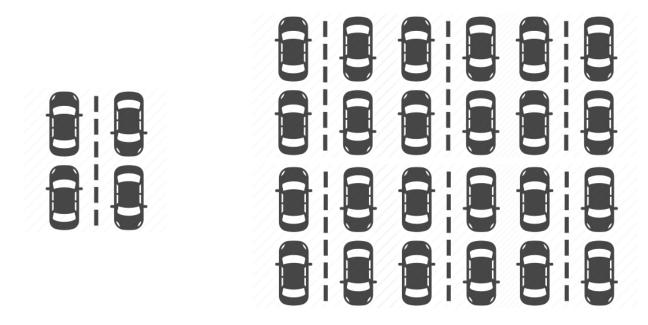
Declarative

Prepare in advance and apply

Easy to recreate



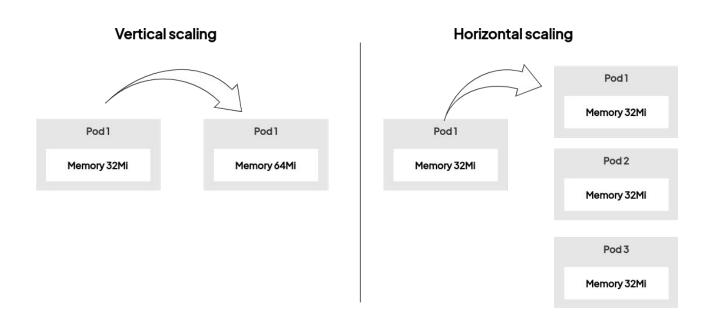
The need for Scaling Application



Expected Traffic

Planned Scaling

Vertical Scaling vs. Horizontal Scaling



How do we know the traffic and load?

- Metrics for nodes
- Metrics for pods
- Metrics for applications



Autoscaling Pods





















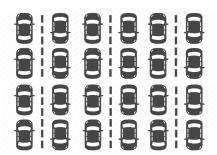
























Prerequisites

- A working Kubernetes Cluster (single node minikube for the demonstration).
- **kubectl** installed and configured.
- Access to the demo repository
 (github.com/iamgini/workshops-demos/tree/main/Kubernetes-Deploy-Your-Application-at-Scale)



References

Workshop instructions

workshops.techbeatly.com/docs/containers/Kubernetes-Deploy-Your-Application-at-Scale-in-K8S

Sample repo for deployment YAML

github.com/iamgini/workshops-demos/tree/main/Kubernetes-Deploy-Your-Application-at-Scale

How to practice Kubernetes

techbeatly.com/practicing-kubernetes

Top 15 Free Kubernetes Courses

techbeatly.com/kubernetes-free-courses

Thank You

- in linkedin.com/in/gineesh
- iamgini.com
- techbeatly.com
- t.me/telegram

kcdkerala.in