

Kubernetes





Problems:

- The need to scale up/down quickly in response to the workload
- Combining multiple services into one cluster
- Continuous integration and deployment
- High availability and Fault tolerance





Potential solutions:

- Manually handle everything
- Implement your own system:)
- Use an existing one

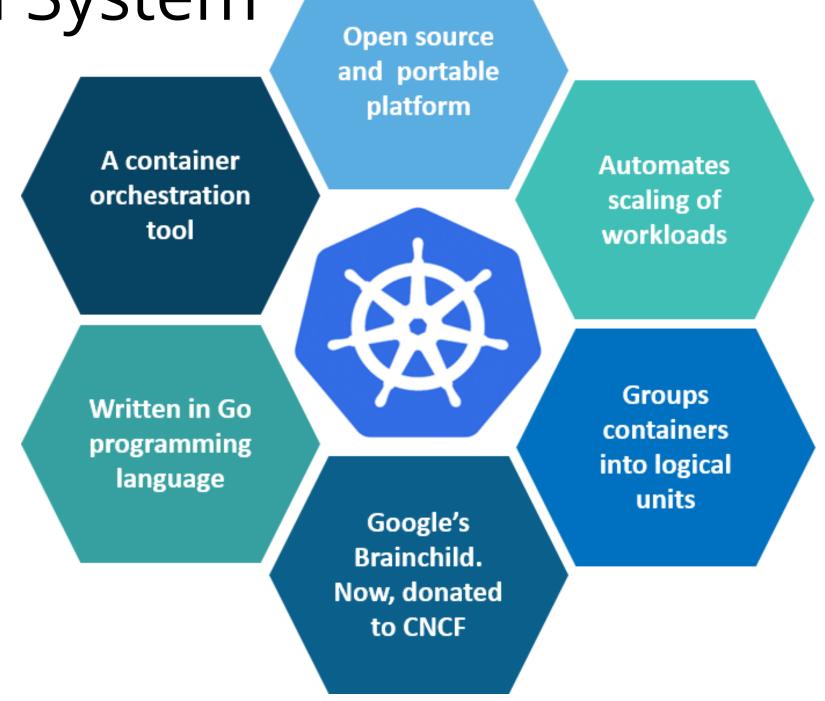




Container Orchestration System

Providing:

- Auto Software Deployment
- Auto Scaling
- System Resilience
- Service Discovery
- System Management







The Premise





The ability to enforce the *Desired State*.

"I promise that all the containers running across the cluster are always in the desired state."

-Kubernetes



The Desired State

- represents a state of the system that we aim to achieve.
- should be retained at all times, even in case a member of a cluster fails.
- can be defined through a declarative or imperative API.



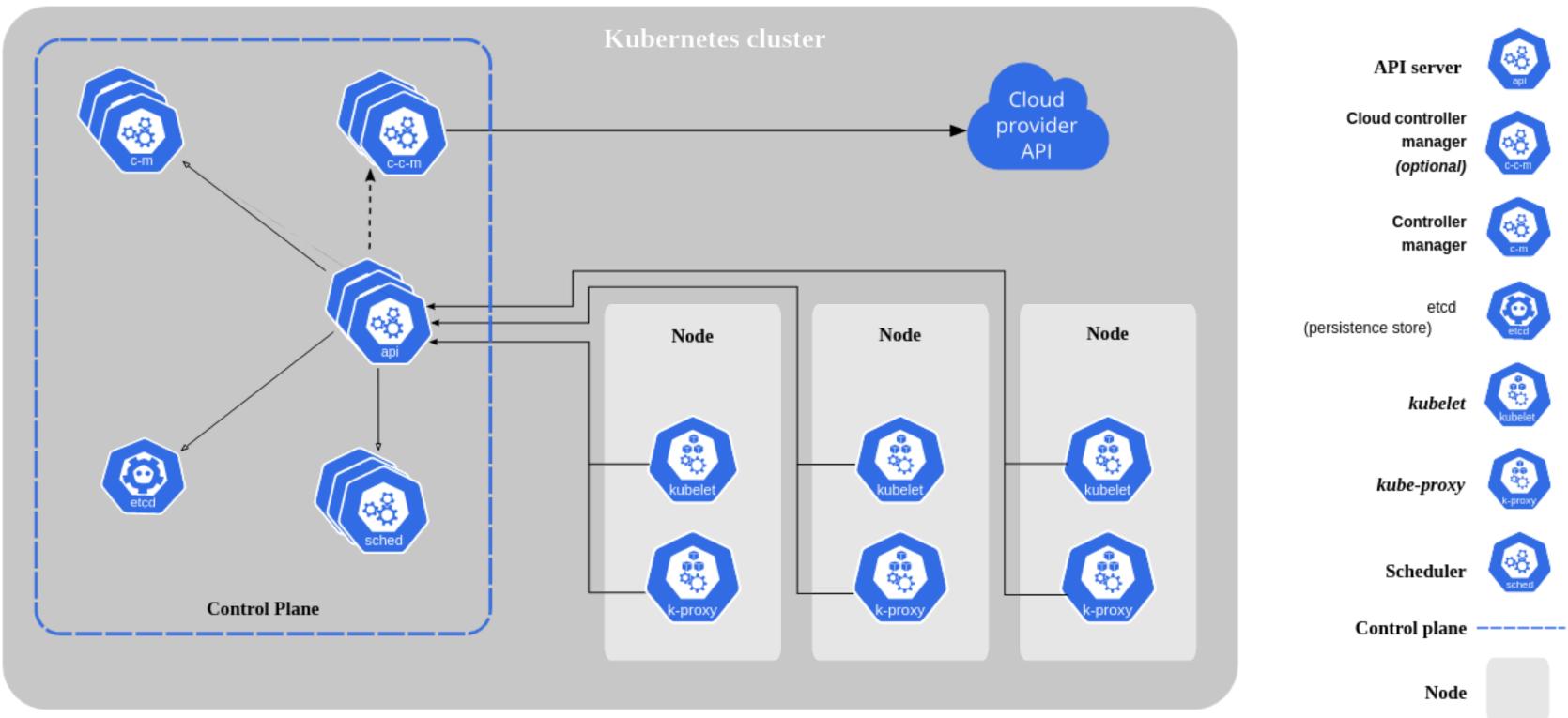


The Architecture





The Architecture







Distributions



Distributions

- Vanilla Kubernetes configure everything on your own
- Managed Kubernetes comes pre-compiled and pre-configured
- Managed Kubernetes versions are also known as Kubernetes distributions.





Key Elements of a Kubernetes Distribution

- Container Runtime
- helps create and manage containers on the physical/virtual machine in which they are hosted.
 - Storage
- offers a way to persist data, dynamically provision volumes, and manage them.
 - Networking
- allows for seamless communication and interaction between different containerized components.



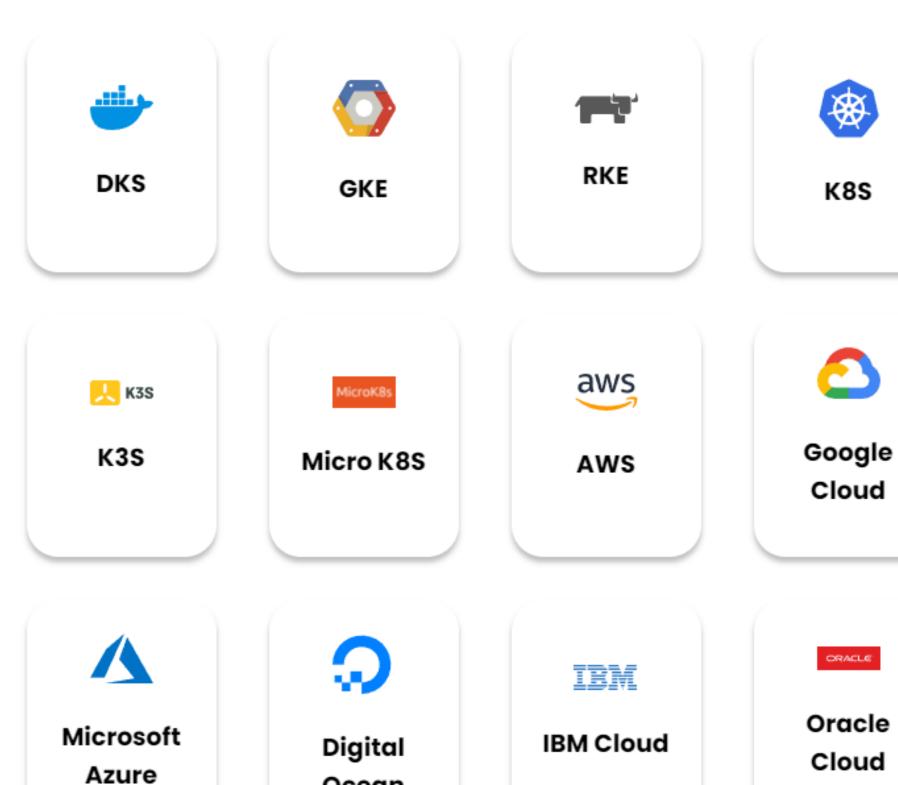


Key Elements of a Kubernetes Distribution

- Container Runtime
 - Docker, Apache Mesos, CoreOS, rkt, Canonical LXC, containerd...
- Storage
 - EBS, GlusterFS, Portworx, Rook, OpenEBS, Longhorn...
- Networking
 - Flannel, Weave Net, Calico, Canal...







Ocean

K8S

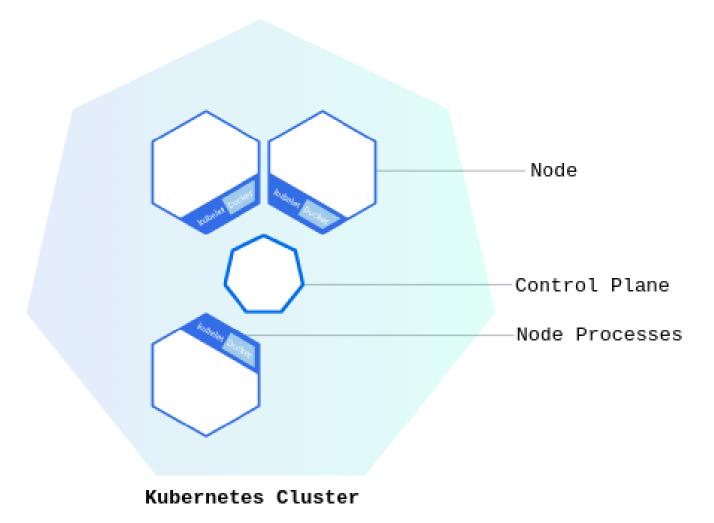
ORACLE





Core Concepts

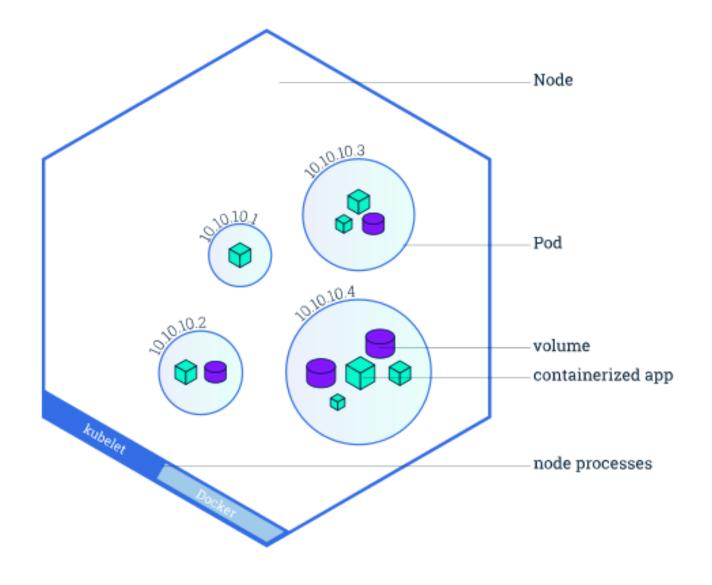




Control Plane - is responsible for managing the cluster.

Node - a VM or a physical computer that serves as a worker machine in a Kubernetes cluster.

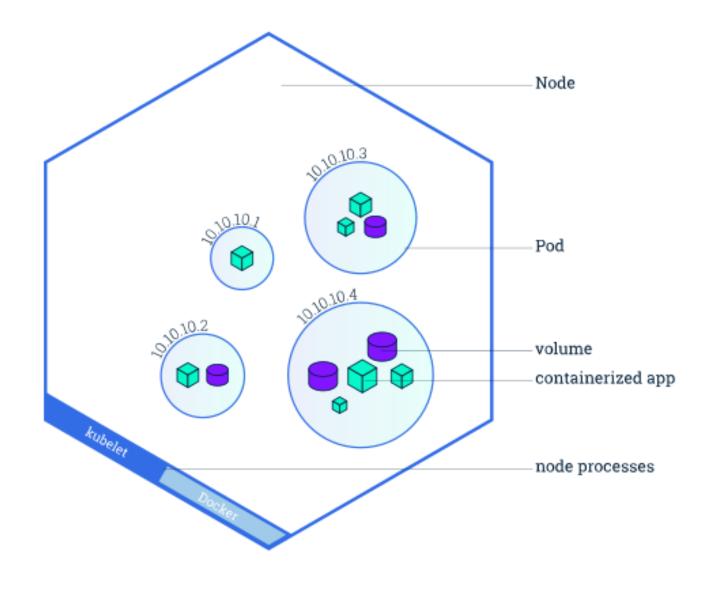




Pod - models an application-specific "logical host" and can contain different application containers which are relatively tightly coupled.

Volume - a directory, possibly with some data in it, which is accessible to the containers in a pod.





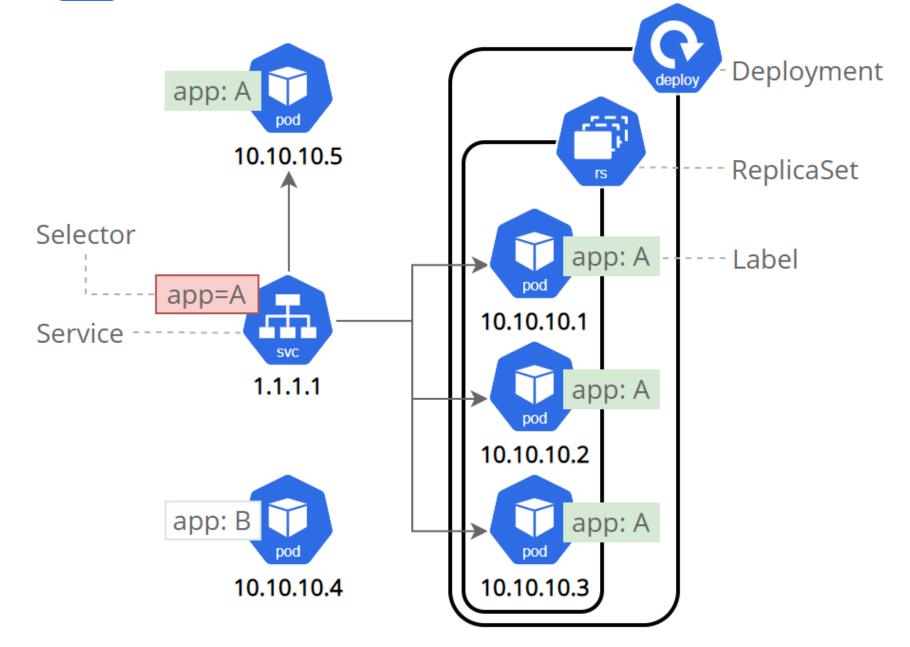
Ephemeral Volume - a volume that has the lifetime of a pod.

Persistent Volume - a volume that exists beyond the lifetime of a pod.

It can be provisioned *statically* or *dynamically* (using the *PVC*).

Persistent Volume Claim - a request for the resources that PV needs, and also acts as a claim check to the resource.





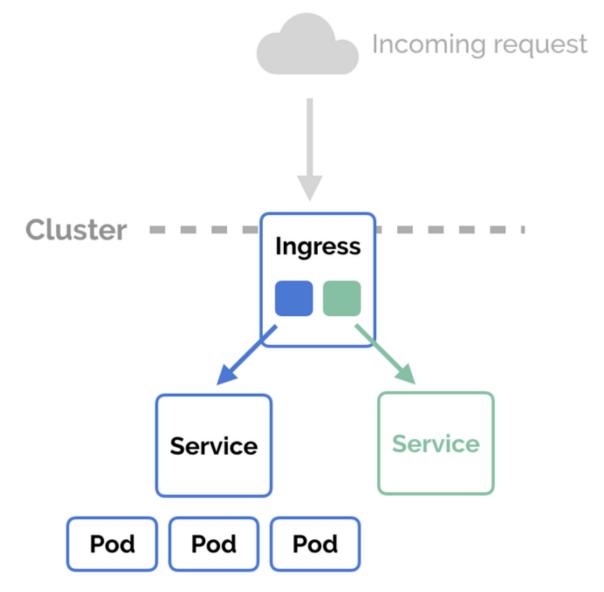
Service - an abstraction that defines a logical set of Pods and a policy by which to access them.

It can be of type *ClusterIP* (default), *LoadBalancer*, *NodePort* and *ExternalName*.

Deployment - instructs Kubernetes on how to create and update instances of your application.

ReplicaSet - maintains a stable set of replica Pods running at any given time.





Ingress - an API object that manages external access to the services in a cluster and exposes HTTP and HTTPS routes from outside the cluster to services within the cluster.





Creating a Cluster



Creating a Cluster

- Install kubeadm
- Init control plane node kubeadm init <args>
- Join a cluster -

kubeadm join <control-plane-host> --token <token> -discovery-token-ca-cert-hash sha256:<hash>



Creating a Cluster

- Install microk8s snap install microk8s --classic
- Optionally check status microk8s status --wait-ready
- Join a cluster microk8s join <cluster-host>/<hash> --worker



Creating a Cluster minikube

- Install minikube
- Start the cluster minikube start



Creating a Cluster

- Run curl -sfL https://get.k3s.io | sh -
- :)
- Join a cluster -

k3s agent --server <server-url> --token \${NODE_TOKEN}





Interacting with a Cluster



Interacting with a Cluster

List nodes in a cluster:

kubectl get node

List resources in a cluster:

- kubectl get <type>
- kubectl get pod
- kubectl get service -n custom-namespace
- kubectl get deployment -A

Get a specific resource:

- kubectl get <type>/<resouce-name>
- kubectl get pod/test-01hss -o yaml

Edit a specific resource:

- kubectl edit <type>/<resouce-name>



Describe a specific resource:

- kubectl describe <type>/<resouce-name>
- kubectl describe service/nginx-svc

Get logs of a pod:

- kubectl logs pod test-01hss
- kubectl logs deployment test

Forward a port to the cluster:

- kubectl port-forward <type>/<r-name> <lport>:<tport>
- Scale deployment:
 - kubectl autoscale deployment <r-name> --min=2 --max=10
 - kubectl scale --replicas=3 deployment <r-name>



Interacting with a Cluster

Get resource consumption:

- kubectl top node
- kubectl top pod

Execute a command in a pod:

- kubectl exec -it <pod-name> -- <command>
- Apply yaml file:
 - kubectl apply -f <file-path>

CRUD:

- kubectl create <type> <args>
- kubectl get <type> <args>
- kubectl patch <type> <args>
- kubectl delete <type> <args>





Deploying an App



Imperative

Start a pod:

- kubectl run nginx --image=nginx -n test

Expose a pod:

kubectl expose deployment nginx --port=80 --target-port=80

Declarative

Define desired state in yaml files Apply:

- kubectl apply -f <file-path>





What's next?



What's next?

- Explore Kubernetes resources
 (ConfigMaps, Secrets, Certificates, StatefulSets, Jobs...)
- Use Helm Kubernetes package manager
- Learn Kubernetes YAML file syntax
- Advanced zero-downtime deployments
- Play around :)





Thank you!

