Kubernetes Architecture

Kubernetes is an open-source container orchestration platform. It automates the deployment (keeps pods running), scaling, and management of containerized applications.

Kubernetes cluster is a set of machines called nodes that is used to run containerized applications

Kubernetes supports enterprise-level standards, like auto-scaling, multiple hosts, auto-healing, load balancing

the api-server is the primary interface between the control plane and the rest of the cluster. It exposes a RESTful API that allows clients to interact with the control plane and submit requests to manage the cluster. the api-server exposes Kubernetes to the external world

cloud controller manager translates the provision of cloud resources. this utility is not required for on-premises procurement.

the controller manager is responsible for running controllers that manage the state of the cluster. eg. replica controller, and deployment controller.

the scheduler is responsible for scheduling pods onto the worker nodes in the cluster. it uses information about the resources required by the pods and the available resources on the worker nodes to make placement decisions.

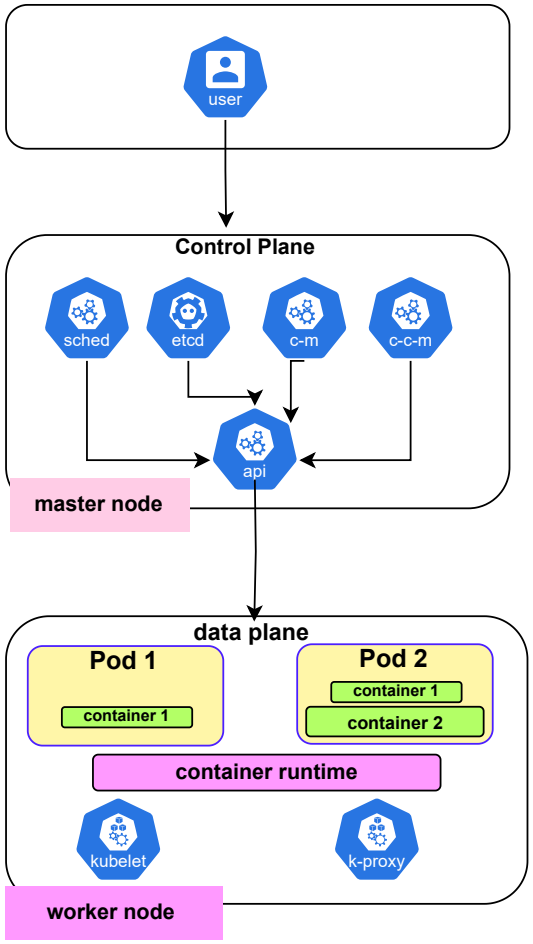
etcd is a distributed key-value store. It stores cluster's persistent state. It is used by the API - server and other components of the control plane to store and retrieve information about the cluster

a pod is a wrapper around the container with advanced capabilities. the pod is the smallest deployable unit in kubernetes. a pod provides shared storage and networking for containers inside the pod.

kube-proxy is a networking proxy that runs in each worker node. it is responsible for routing traffic to the correct pods. it also provides load balancing and ensures that traffic is distributed evenly across the pods.

kubelet is a deamon that runs on each worker node. it is responsible for communicating with the control plane about which pods run on the node and ensures that the desired state of the pod is maintained.

cluster



the control plane controls the action, the data plane executes the actions

container runtime

the container runtime runs the containers on the worker nodes. it is responsible for pulling the container images from a registry, starting and stopping the containers and managing container resources.

key notes

when you deploy kubernetes, you get a cluster. a cluster is a set of machines called nodes. cluster is made up of master and worker nodes. nodes contain pods. pods contain container(s). Kubernetes is scalable and highly available. it provides features like self-healing automatic roll-backs and horizontal scaling. It allows us to respond to changes in demand quickly. Kubernetes is portable. it helps us deploy and manage applications in a consistent and reliable way regardless of the underlying infrastructure.

managed Kubernetes services:
-Amazon EKS
-Google Kubernetes Engine
-Azure Kubernetes Services