02 K8s Architecture

Node Process

Each node has multiple pods on it, the way k8s does is using 3 processes that must be installed on every Node that are used to manage and schedule those pods.

Nodes are the cluster servers that do the actual work, so they are also called the worker nodes.

1. Container Runtime → Eg: Docker
2. Kubelet → Schedules those pods and the containers underneath. It is an interface between the container runtime and the machine. Kubelet is responsible to take the configuration and running the pod with a container inside and assigning resources (cpu, ram, storage) from the node to the container.
3. Kube proxy → responsible for forwarding requests from services to pods. It has intelligent forwarding logic that makes sure that the communication also works in a performant way with low overhead.

!Communication between the pods works using services which is sort of a load balancer that catches the request directed to the pod/application and then forwards it to the respective pods.

Master Process

It manages the worker/node processes, i.e., it new node joins the cluster, if one node dies, schedule pod, monitor pods, re-schedule/re-start pod, etc

There are 4 processes that run on every master node that control the cluster state and the worker nodes.

1. API Server → is like a cluster gateway which gets the initial request of any updates into the cluster or even the queries from the cluster and also acts as a gatekeeper for authentication and validating request.
2. Scheduler → The Api Server hands over the request to the scheduler which decides where the application pod will start on one of the worker nodes. It doesn't randomly put the pod, but intelligently decides where to put the pod. The Kubelet gets the request from the scheduler and executes the request.
3. Contoller Manager → When a pod dies on any node, there must be a way to detect that the nodes died and then re-schedule thode pods as soon as possible. So controller manager detects the state changes like crashing of pods and tries to recover the cluster state as soon as possible. And for that, it makes a request to the scheduler.
4. etcd → It is a key-value store of a cluster state that is like a cluster brain, which means that every time there is any change in the cluster(new pod, pod dies, etc) get updated to the key-value store.

In prctice, K8s cluster is made up of multiple masters.