



# kubernetes

---

*Kubernetes: High Availability Deployment*

## *KUBERNETES : Administration*

---

- Creating a **HA Kubernetes Cluster**, means we want to have multiple Kubernetes Master and Workers, running across multiple zones.
- **Kubernetes High-Availability** is about setting up Kubernetes, along with its supporting components in a way that there is no single point of failure.
- A **single master cluster** can easily fail, while a multi-master cluster uses multiple master nodes, each of which has access to same worker nodes.
- In a single master cluster the important component like API server, controller manager lies only on the single master node and if it fails you cannot create more services, pods etc. However, in case of Kubernetes HA environment, these important components are replicated on multiple masters (usually three masters) and if any of the masters fail, the other masters keep the cluster up and running.

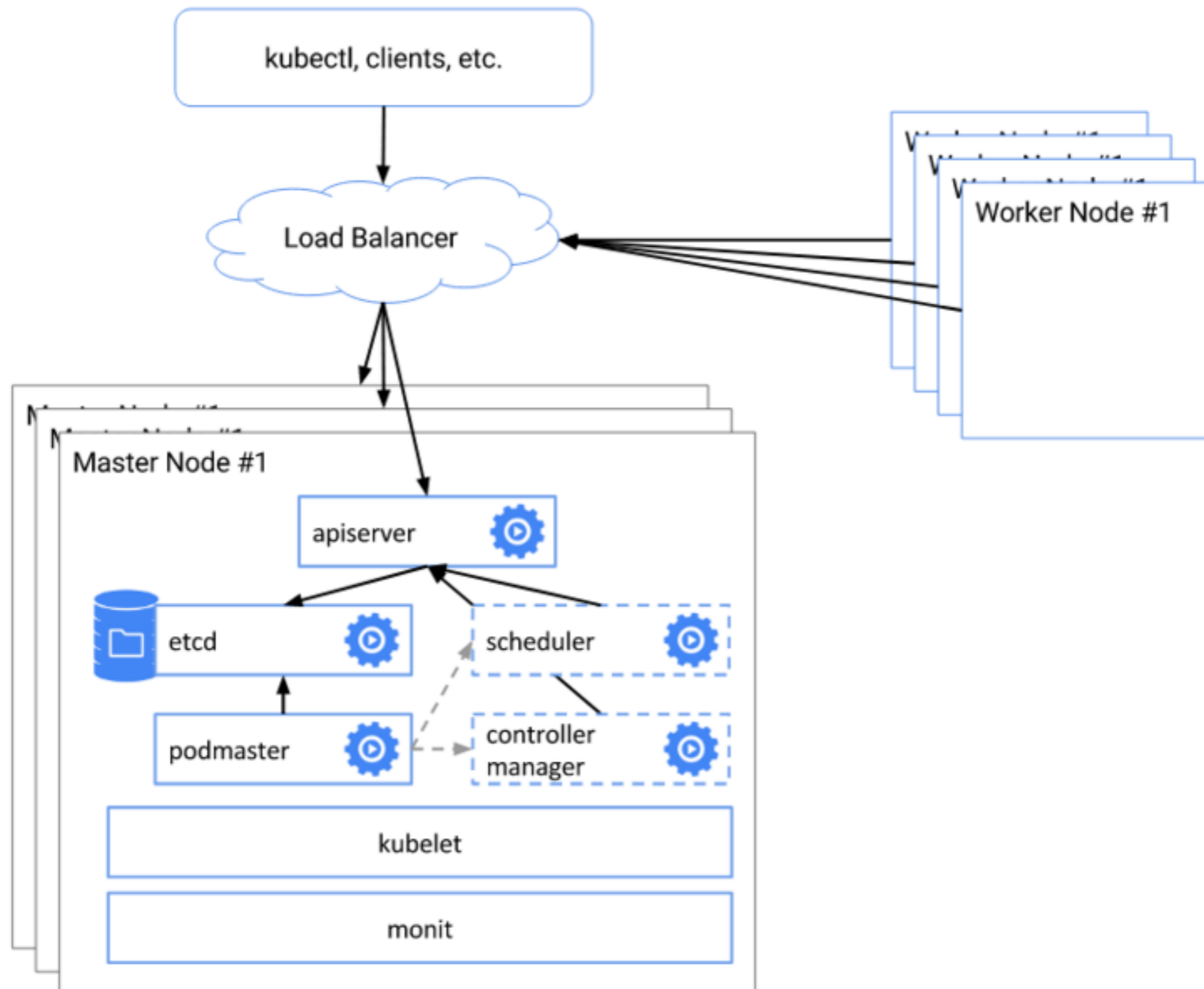
## *KUBERNETES : Administration*

---

- However, in case of **Kubernetes HA environment**, these important components are replicated on multiple masters (usually three masters) and if any of the masters fail, the other masters keep the cluster up and running.
- A **multi-master setup** protects against a wide range of failure modes, from a loss of single worker node to the failure of the master node's etcd service.
- By providing redundancy, a multi-master cluster serves a highly available system for your end users.

# KUBERNETES : Administration

## Architecture of Kubernetes HA Deployment



# *KUBERNETES : Administration*

---

## Kubernetes HA Overview

- Each of master replicas will run the following components in the following mode:
- **etcd instance:** All instances will be clustered together using consensus.
- **API server:** Each server will talk to **local etcd** - all API servers in the cluster will be available.
- **Controllers, Scheduler, and Cluster Auto-Scaler:** will use lease mechanism - only one instance of each of them will be active in the cluster.
- **Add-on Manager:** Each manager will work independently trying to keep add-ons in sync.

# *KUBERNETES : Administration*

---

## Kubernetes HA Overview

- **Load Balancing:** Kubernetes Master Nodes binds with LoadBalancer and an connect with Load Balancer with an Floating Virtual IP.
- All Worker Nodes Communicate to Master Node with Load Balancer and All Master will communicate each other with Load Balancer as well.
- Cloud Providers have their own Load Balancer Implementation and mechanism.
- Some Cloud Provider Need LoadBalance IP before Creating the HA Environment, some Setup it default and manage own.

*Will see you in Next Lecture...*

---

*Thank you!*

A close-up photograph of a hand holding a black marker, writing the words 'Thank you!' in a cursive script on a white surface. The hand is positioned on the right side of the frame, with the marker tip touching the paper. The background is plain white.

*See you in next lecture ...*