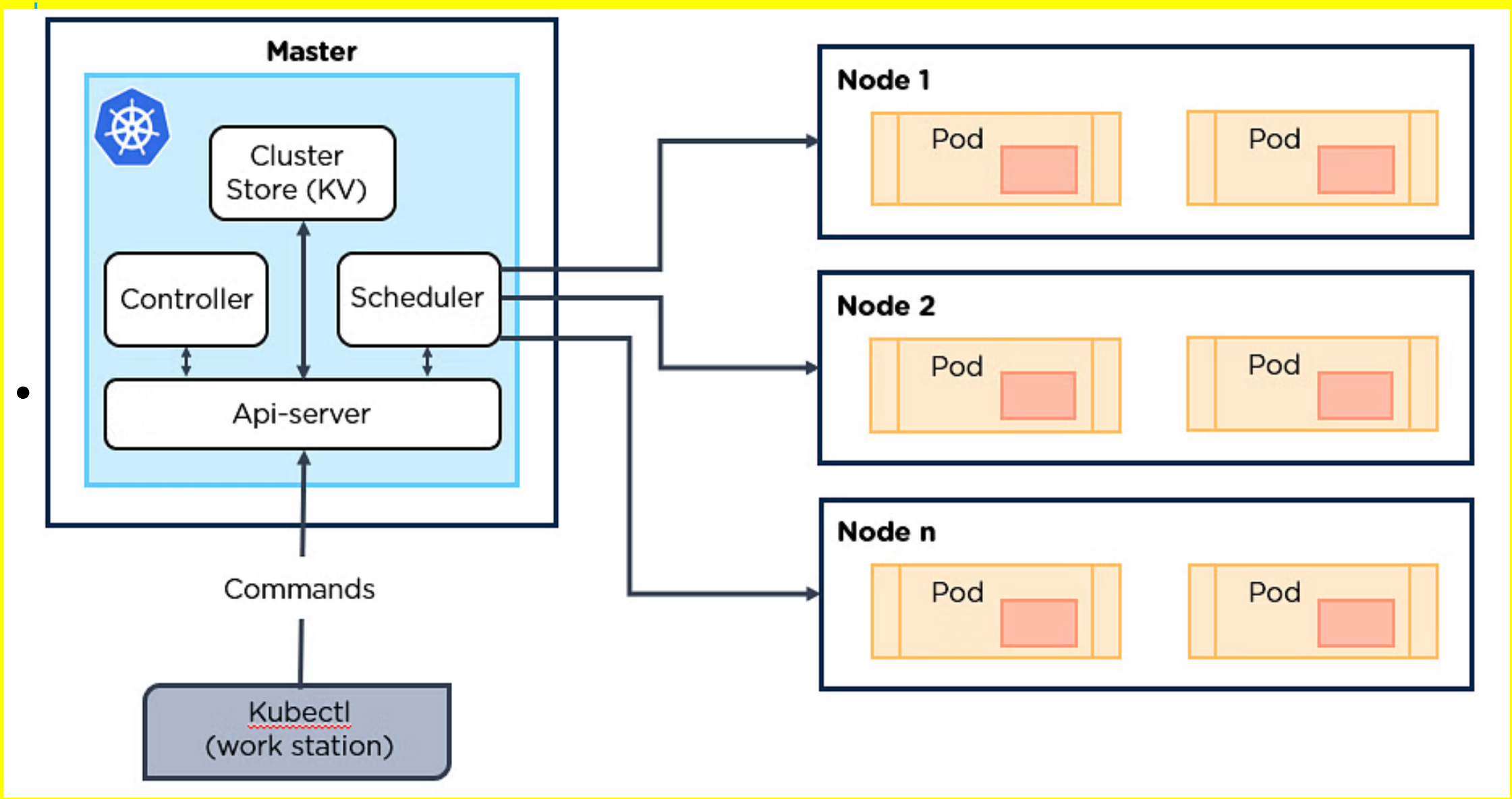
22AIE305: CLOUD COMPUTING

Kubernetes Architecture



Kubernetes cont'd

The Kubernetes architecture consists of various components that help manage clusters.

A Kubernetes cluster is a group of machines, called nodes, that are used to run containerized applications.

These nodes are managed by Kubernetes platform, which provides a way to automate the deployment, scaling, and management of the applications.

Kubernetes+Minikube

Kubernetes is a Greek word for pilot or Helmsman (navigator, guide). It is abbreviated as K8s (because there are 8 letters between the first k and last s).

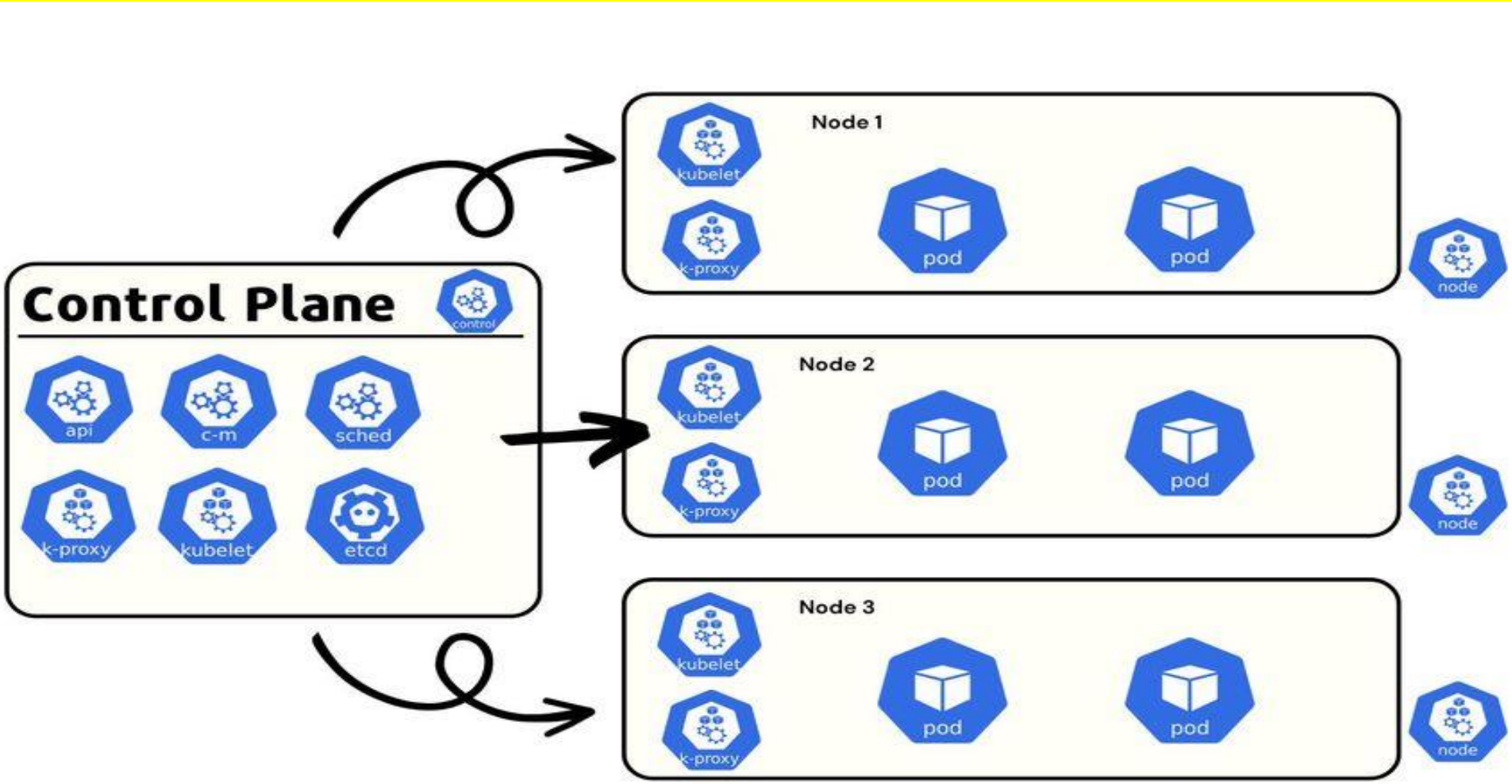
Used for automating software deployment, scaling, and management on the Cloud.

Kubernetes assembles one or more computers, either virtual machines or bare metal, into a cluster which can run workloads in containers.

Minikube is then used to manage Kubernetes clusters.

Distributed cluster machines can be created using Kubernetes+Minikube

K8s architecture



Kubernetes Control Plane Components

The control pane node contains the critical components required to run the cluster. If even one of these components does not exist in the Control Pane, you will be unable to use the cluster.

- API Server
- ETCD
- Scheduler
- Controller Manager
- Cloud Controller Manager

cloud agnostic abstraction layer

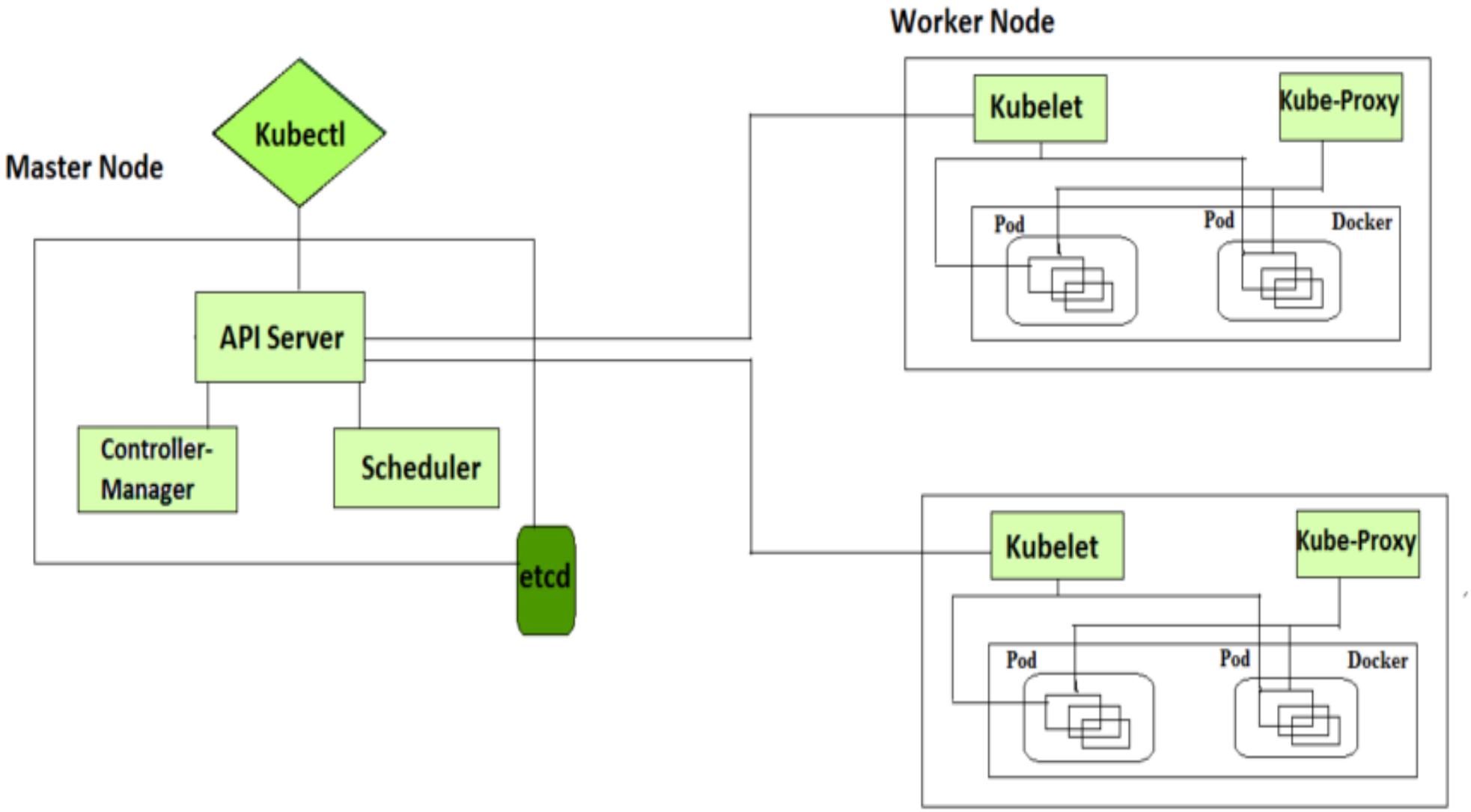
Kubernetes serves as a cloud agnostic abstraction layer because almost all Cloud providers support it.

Google Kubernetes Engine (GKE) is the most stable, performant, and reliable managed kubernetes service.

gcloud container clusters create --enable-ip-alias --enable-autoscaling --max-nodes=20 --min-nodes=1 --region=us-central1 --node-locations=us-central1-b --image-type=cos_containerd --disk-size=100 --disk-type=pd-balanced --machine-type=n2-highmem-8 --cluster-version latest --no-enable-autoupgrade --enable-network-policy --create-subnetwork="" --tags=hub-cluster --no-enable-insecure-kubelet-readonly-port <cluster-name>

API Server

- The Kubernetes API server acts as the entry point for the Kubernetes Cluster. Whenever any external request is made to the cluster, or by any of the other control plane components, the request first goes to the API server. The API server authenticate and authorizes the request, and then sends the request to the relevant K8s components.
- Whenever a kubectl command is issued, the request goes to the kube-api-server. The API Server lets you query requests and interact with the different components of the cluster. For example, if you wanted to create a deployment, you would first write a YAML manifest and use kubectl apply -f to create the object.



What is Ingress?

Ingress is a collection of routing rules that decide how the external services access the services running inside a Kubernetes cluster.

Ingress provides load balancing, SSL termination, and name-based virtual hosting.

Docker swarm vs Kubernetes

Category	Docker Swarm	Kubernetes
Scaling	No Auto Scaling	Auto Scaling
Load Balancing	Autoload Balancing	Manually configures load balancing
Installation	Easy and fast	Long and time-consuming
Scalability	Cluster strength is weak when compared to Kubernetes	Cluster strength is strong
Storage Volume Sharing	Shares storage volumes with any other container	Shares storage volumes between multiple containers inside the same pod
GUI	Not available	Available

Kubernetes does not work with individual nodes; it works with the cluster as a whole. Kubernetes clusters make up the master and slave node and manage it as a whole. There can be more than one cluster in Kubernetes.