A Kubernetes cluster is a collection of physical/virtual machines (nodes) that run containerized applications. A cluster is made up of 1 parent node that manages the cluster, and multiple worker nodes (pods) that run the actual application instances inside Kubernetes.

A pod represents a single running process that encapsulates a container (or many closely coupled containers). Pods can be replicated to scale applications and handle more traffic. From the perspective of a cluster, a set of replicated pods is still one application instance.

A pod (or replicated pods) is managed by a controller. Controllers handle replication, self-healing, rollout of updates, and general management of pods.

Service objects define a set of rules by which a pod (or set of replicated pods) can be accessed.