

Kubernetes Deployment Configuration Options

At a high level, you have four main deployment configurations:

Single Node

With a single-node deployment, all the components run on the same server. This is great for testing, learning, and developing around Kubernetes.

Single head node, multiple workers

Adding more workers, a single head node and multiple workers typically will consist of a single node etcd instance running on the head node with the API, the scheduler, and the controller-manager.

Multiple head nodes with HA, multiple workers

Multiple head nodes in an HA configuration and multiple workers add more durability to the cluster. The API server will be fronted by a load balancer, the scheduler and the controller-manager will elect a leader (which is configured via flags). The etcd setup can still be single node.

HA etcd, HA head nodes, multiple workers

The most advanced and resilient setup would be an HA etcd cluster, with HA head nodes and multiple workers. Also, etcd would run as a true cluster, which would provide HA and would run on nodes separate from the Kubernetes head nodes.

Which of the four you will use will depend on how advanced you are in your Kubernetes journey, but also on what your goals are.

A Note About Kubernetes Federation

The use of Kubernetes Federation also offers high availability. Multiple clusters are joined together with a common control plane allowing movement of resources from one cluster to another administratively or after failure. While Federation has some issues, there is hope v2 will be a stronger product.