



KPLABS Course

Certified Kubernetes Application Developer 2022

Workloads & Scheduling

ISSUED BY

Zeal Vora

REPRESENTATIVE

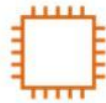
instructors@kplabs.in



Module 1: Labels & Selector

1.1 Labels

Labels are key/value pairs that are attached to objects, such as pods



Server



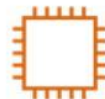
Database



Load Balancer



Load Balancer



Server



Database

1.2 Selectors

Selectors allow us to filter objects based on labels.

Example:

Show me all the objects which have a label where env: prod



name: kplabs-gateway
env: production



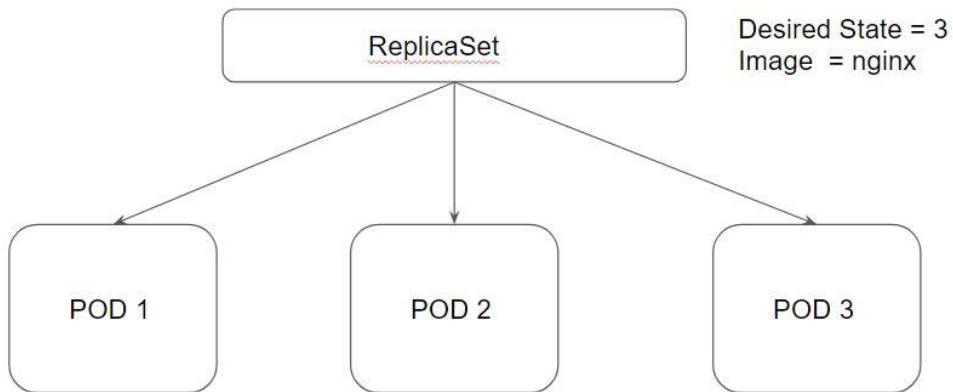
name: kplabs-db
env: production



name: kplabs-elb
env: production

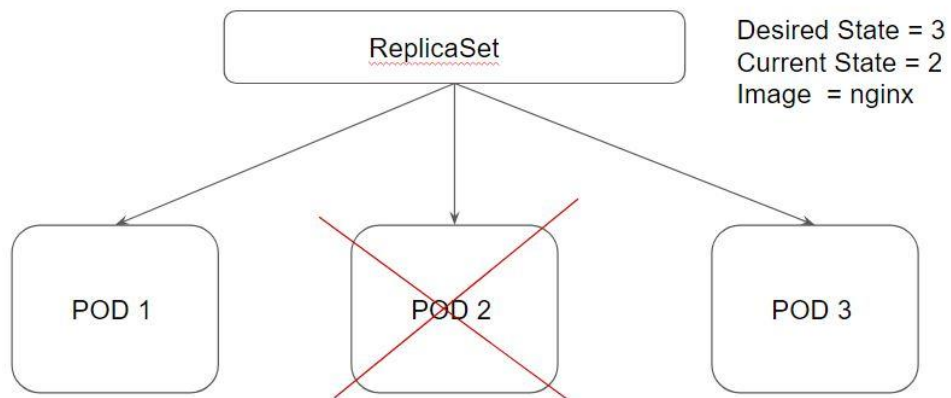
Module 2: ReplicaSets

A ReplicaSet purpose is to maintain a stable set of replica Pods running at any given time.



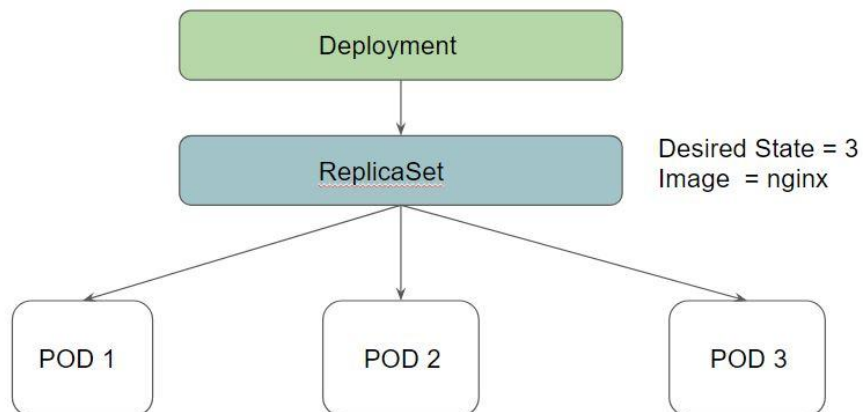
Desired State - The state of pods which is desired.

Current State - The actual state of pods that are running.



Module 3: Deployments

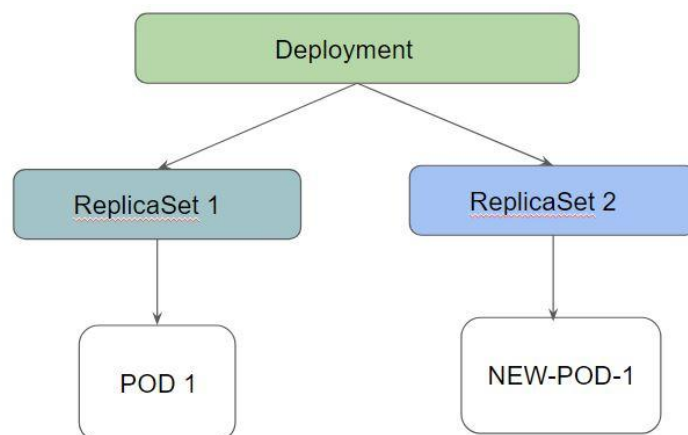
Deployments provide replication functionality with the help of ReplicaSets, along with various additional capability like rolling out of changes, rollback changes if required.



3.1 Benefits of Deployment - Rollout Changes

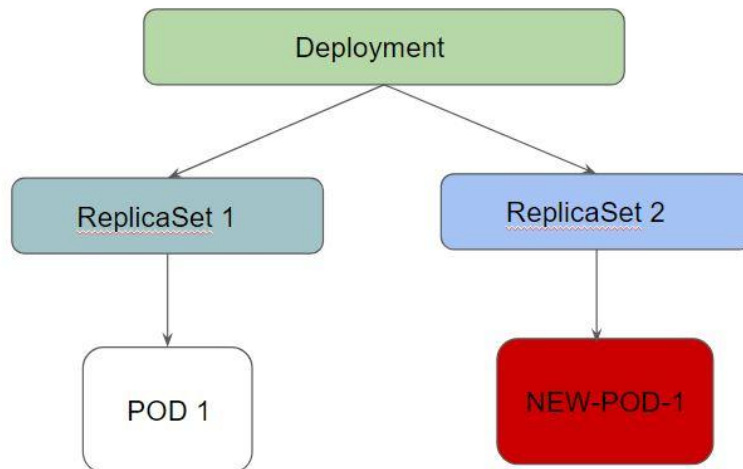
We can easily roll out new updates to our application using deployments.

Deployments will perform an update in a rollout manner to ensure that your app is not down.



3.2 Benefits of Deployment - Rollback Changes

Sometimes, you may want to rollback a Deployment; for example, when the Deployment is not stable, such as crash looping



Deployment ensures that only a certain number of Pods are down while they are being updated.

By default, it ensures that at least 25% of the desired number of Pods are up (25% max unavailable).

Deployments keep the history of revision which had been made.

Module 4: Deployment Configuration

While performing a rolling update, there are two important configurations to know.

Configuration Parameter	Description
maxSurge	Maximum Number of PODS that can be scheduled above original number of pods.
maxUnavailable	Maximum number of pods that can be unavailable during the update

maxUnavailable=0 and maxSurge=20% << Full Capacity is maintained.

maxUnavailable=10% and maxSurge=0 << Update with no extra capacity. In-place updates.

If you want fast rollout, make use of maxSurge.

If there might be a resource quota in place and partial unavailability is acceptable, maxUnavailable can be used.

Module 5: Important Pointer - Deployments

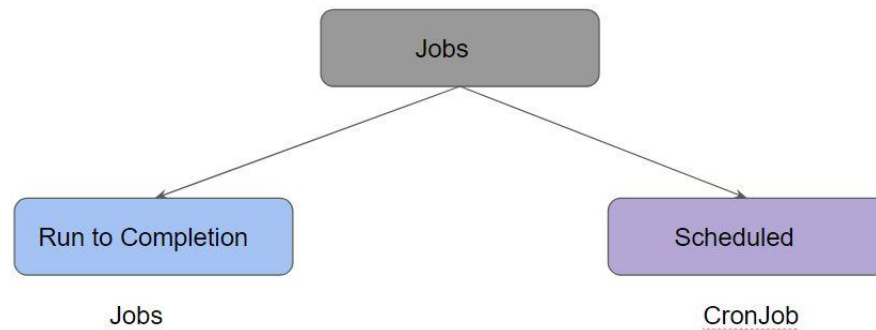
1. You should know how to set a new image to deployment as part of rolling update.
2. You should know the importance of --record instruction.
3. You should know how to rollback a deployment.
4. You should be able to scale the deployment

Following are some of the important commands to remember:

Command	Command
kubectl set image deployment nginx-deployment nginx=nginx:1.91 --record	Set Image
kubectl scale deployment nginx-deployment --replicas 10	Scale Deployment
kubectl rollout undo deployment nginx-deployment	Rollout Undo

Module 6: Jobs

Job is responsible for creating one or more pods to run the instruction which has been specified.



There are two types of Jobs available:

- Jobs (Run to completion)
- CronJobs

Once the task is completed, the pods are not terminated.

It is up to the user to delete old jobs after taking note of the status.

When you delete the job, all the pods associated with the job will be deleted.

In a multiple-worker node cluster, if one node fails where the job is running, that job will start the pod in the available node.

Module 7: CronJobs

CronJob allows us to run jobs based on a time schedule.

Cron jobs are useful for creating periodic and recurring tasks, like running backups or sending emails.

You can specify cron schedule similar to the cron in Linux.

Join Our Discord Community

We invite you to join our Discord community, where you can interact with our support team for any course-based technical queries and connect with other students who are doing the same course.

Joining URL:

<http://kplabs.in/chat>

