KubeCon | CloudNativeCon

Europe 2023

KubeCon | CloudNativeCon

Europe 2023

Kubeadm Deep Dive

Paco Xu（DaoCloud）

Rohit Anand（NEC Corporation）

Intro us 😄

Paco Xu

DaoCloud Shanghai.

Mainly worked on kubeadm & sig-node

Twitter: xu_paco

Github: pacoxu

❤️ → ⚽ & PUBG

Rohit Anand

Technical Lead NEC

Mainly worked on SCL kubeadm

K8s Slack: @Rohit

LinkedIn:

❤️ → Squash & Badminton

Agenda

- What is Kubeadm?
- Project Highlights
  - Recent Updates
  - RoadMap
- Get Involved
  - How to contribute

What is Kubeadm?

Kubeadm is a node bootstrapper

Someone or something should provide the machines.

kubeadm creates a Kubernetes node on the machine

What Kubeadm Is Good for?

- A simple way for you to try out Kubernetes, possibly for the first time.
- A way for existing users to automate setting up a cluster and test their application.
- A building block in other ecosystem and/or installer tools with a larger scope.

Kubeadm A SIG Cluster Lifecycle project

SIG Cluster Lifecycle’s objective is to simplify creation, configuration, upgrade, downgrade and teardown of kubernetes clusters and their components.

--- The SIG ClusterLifecycle charter

Devstats (recent 2 years)

Etienne Champetier
Zhang Yu
RA489
Ma Xinjian
songxiao-wang87
Davanum Srinivas
Sascha Grunert
Vinayak Goyal
caodonghui
XinYang
Haleygo
wangyysde
Antonio Ojea
Sata Qiu
Octopusjust
Jordan Liggitt
cndoit18
Steve Kuznetsov
xin.li
wangguoyan
chymy
Lubomir Ivanov
tallaxes
Benjamin Elder
Marek Siarkowicz
Manish Kumar
Mark Rossetti
Paco Xu
pengpeng wu
Aaron Crickenberger
Patrick Ohly
Jian Zeng
Arnaud Meukam
Abirdcfly
Hao Diao 10284789
navist2020
Dave Chen
haoyun
Zhang Zhanpeng
Claudiu Belu
muyangren2
Stephen Augustus

~40 contributors

- cmd/kubeadm
- test/e2e_kubeadm
- kubeadm subproject

Kubeadm workflow

1. Initialize the cluster and the first control-plane node

```
$ sudo kubeadm init
```

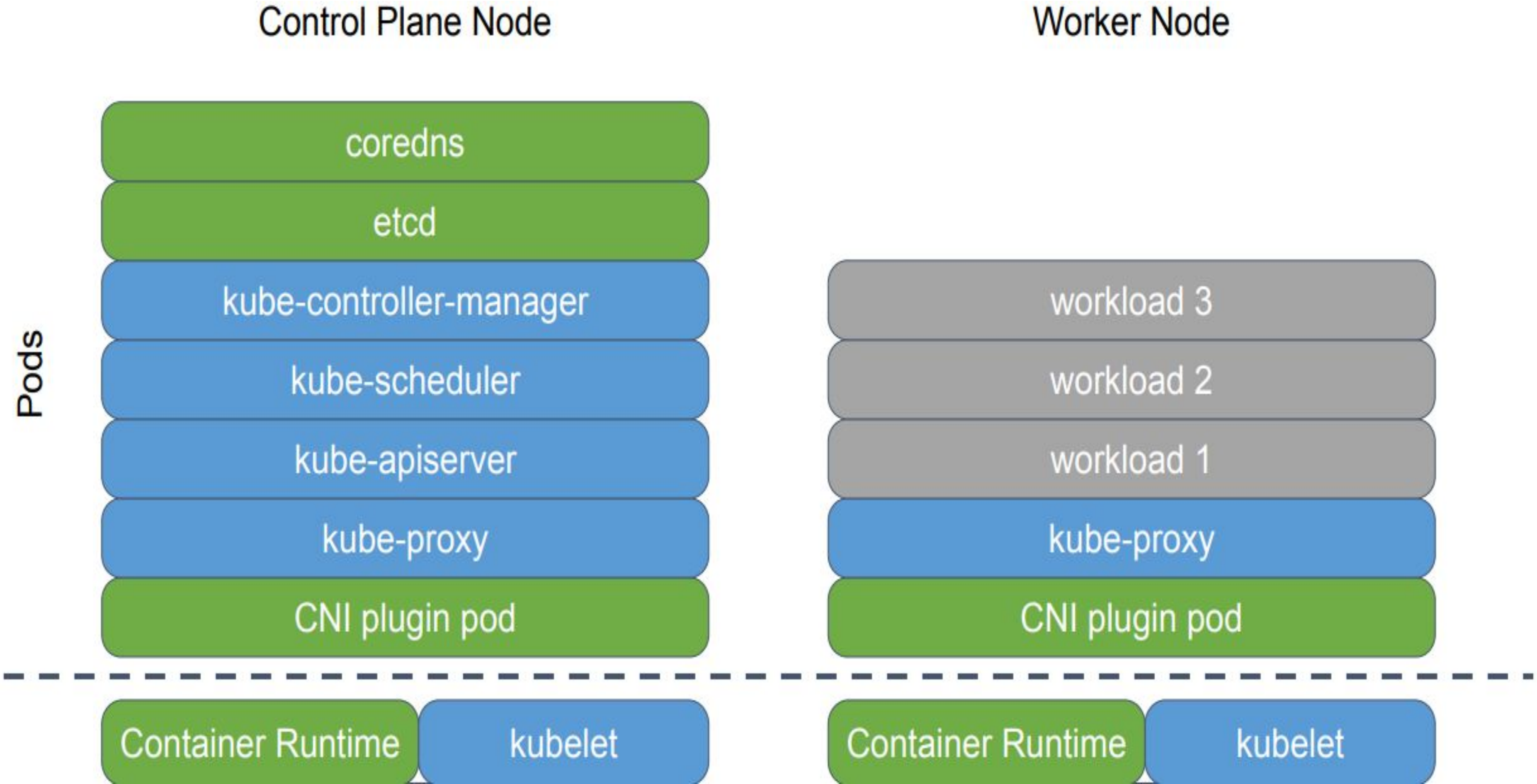
2. Install a POD network addon

```
$ kubectl apply …
```

3. Join worker nodes

```
$  sudo kubeadm join : <control-plane-host> :
<control-plane-port> --token
<token>--discovery-token-ca-cert-hash sha256:<hash>
```

What kubeadm deploy?



Kubeadm upgrade workflow

1. Check for available upgrades

```
$ sudo kubeadm upgrade plan
```

2. Upgrade the first control-plane node on the cluster

```
$ sudo kubeadm upgrade apply v1.24.0
```

3. Upgrade the rest of the nodes

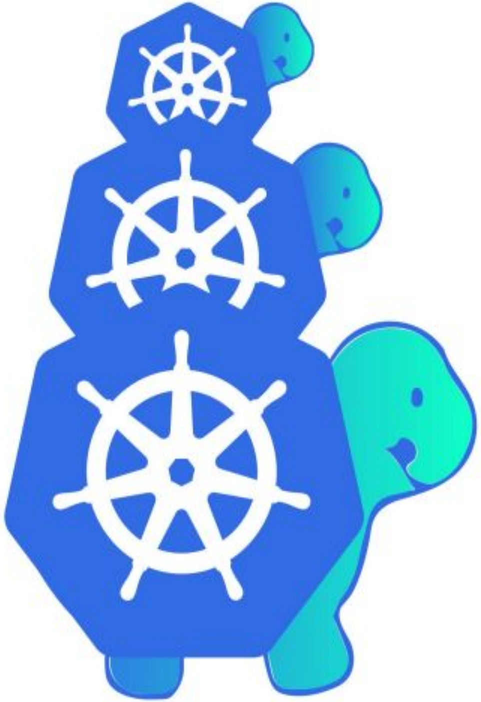
```
$  sudo kubeadm upgrade node
```

Kubeadm scope

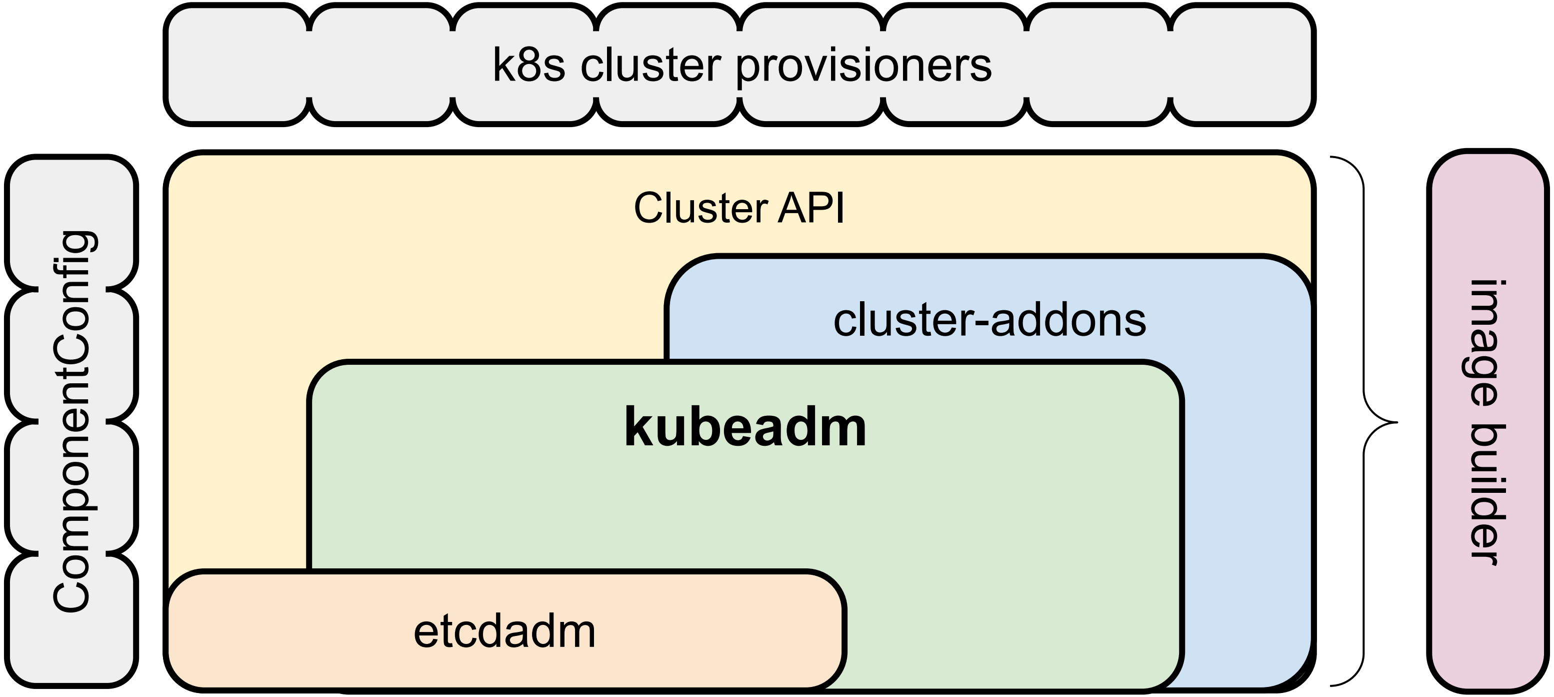
Be Simple

Be Extensible

Who uses kubeadm?



Part of the composable solution



Kubeadm Highlights

Recent KEPs & Features

- KEP-2915: Replace usage of the kubelet-config-x.y naming in kubeadm
  - UnversionedKubeletConfigMap alpha:v1.23; beta: v1.24; GA: v1.25; Removed: v1.26
- rename "node-role.kubernetes.io/master"
  - 1.20 deprecated; 1.24 handling master label; 1.25 taint only if master legacy label exists; 1.26 cleanup all.
- KEP-2568: Run control-plane as non-root in kubeadm.
  - RootlessControlPlane alpha: v1.22
  - Pending on user namespace support in kubelet(alpha v1.25).
- KEP-1739: kubeadm customization with patches
  - kubeadm: add support for patching a "kubeletconfiguration" target #110405 v1.25
- sig-cl/kubeadm: Use etcd's learner mode
  - v1.27 alpha

RootlessControlPlane

RootlessControlPlane is alpha since v1.22 using `securityContext.runAsUser`.

1. Run `kube-apiserver` as non-root in `kubeadm`.
2. Run `kube-controller-manager` as non-root in `kubeadm`.
3. Run `kube-scheduler` as non-root in `kubeadm`.
4. Run `etcd` as non-root in `kubeadm`.

Besides, to run Kubernetes with non root

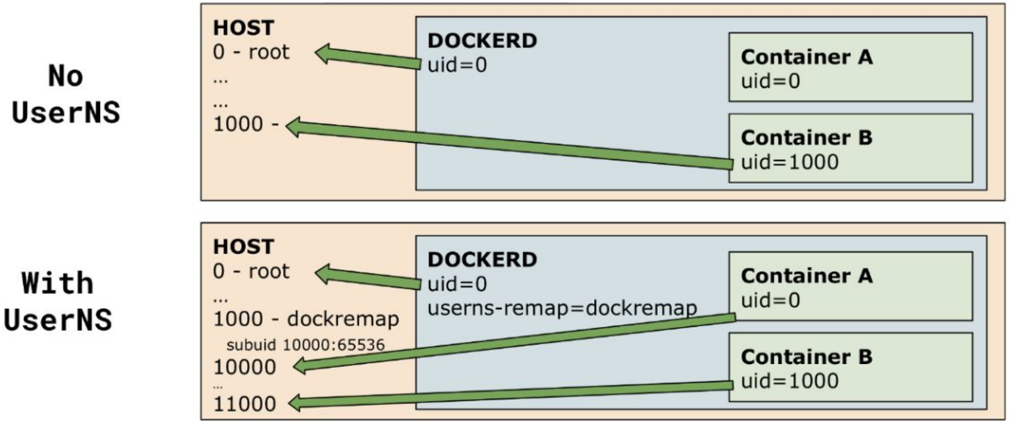
- kubelet can run in UserNS since Kubernetes 1.14.

Rootless Update in SIG Node

In sig-node, add user namespace support with FG UserNamespacesStatelessPodsSupport is alpha since v1.25.

- in 1.27, the feature supports userns in stateless pods with idmap mounts.

After the user namespace feature is promoted to beta or later, we can alter to use it in kubeadm then.



Join etcd in learner mode

Join control plane node

1. etcd join as a learner
2. promote to a voting member

The feature will be alpha in v1.27.

- [Support auto promoting learner member to voting member #15107]

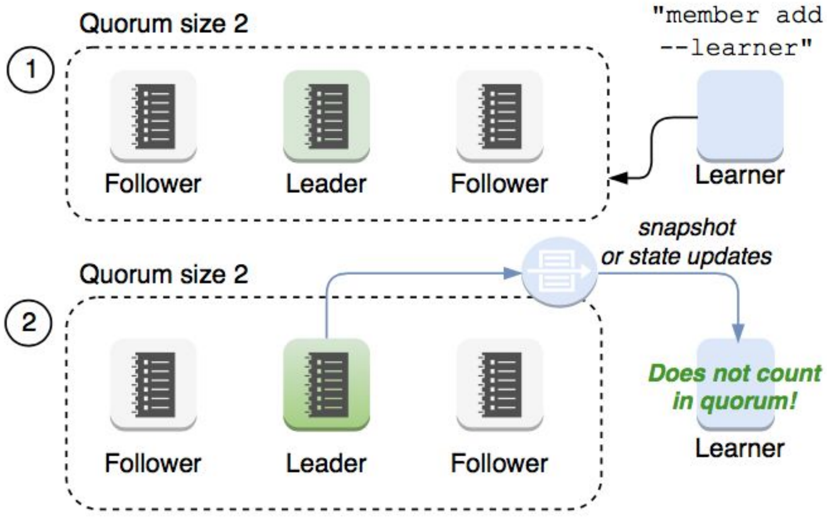


Figure 10. Add a learner node as a non-voting member. Wait until learner node catches up to leader's logs. Until then, learner node neither votes nor counts towards quorum.

--patch example

KEP-1739: kubeadm customization with patches

- v1.25: patching a "**kubeletconfiguration**" target #110405, so you can patch your specific node configuration in local dir.
- v1.22: `kube-apiserver`, `kube-controller-manager`, `kube-scheduler`, `etcd` are supported.

```
apiVersion: kubeadm.k8s.io/v1beta3
kind: InitConfiguration
patches:
  directory: /home/user/somedir
```

```
cat <<EOF >/tmp/kubeadm-patches/etcd+json.json
[{"op":"add","path":"/metadata/annotations/patched","value":"true"}]
EOF

cat <<EOF > /tmp/kubeadm-patches/kubeletconfiguration+strategic.yaml
shutdownGracePeriod: 1s
EOF
```

kubeadm Configuration (v1beta3)

The support matrix will look something like this now and in the future:

- v1.10 and earlier: v1alpha1
- v1.11: v1alpha1 read-only, writes only v1alpha2 config
- v1.12: v1alpha2 read-only, writes only v1alpha3 config. Errors if the user tries to use v1alpha1
- v1.13: v1alpha3 read-only, writes only v1beta1 config. Errors if the user tries to use v1alpha1 or v1alpha2
- v1.14: v1alpha3 convert only, writes only v1beta1 config. Errors if the user tries to use v1alpha1 or v1alpha2
- v1.15: v1beta1 read-only, writes only v1beta2 config. Errors if the user tries to use v1alpha1, v1alpha2 or v1alpha3
- v1.22: v1beta2 read-only, writes only v1beta3 config. Errors if the user tries to use v1beta1 and older

WIP : ResetConfiguration and UpgradeConfiguration .

Road Map

Road Map

- kubeadm configuration v1beta4
  - add UpgradeConfiguration /ResetConfiguration API types
- Kubeadm operator ⏸
  - Removed from kubeadm
  - The discussion is still open and needs more feedback

kubeadm Configuration (v1beta4 in discussion)

https://github.com/kubernetes/enhancements/issues/970

candidates for v1beta4 (in v1.28+)

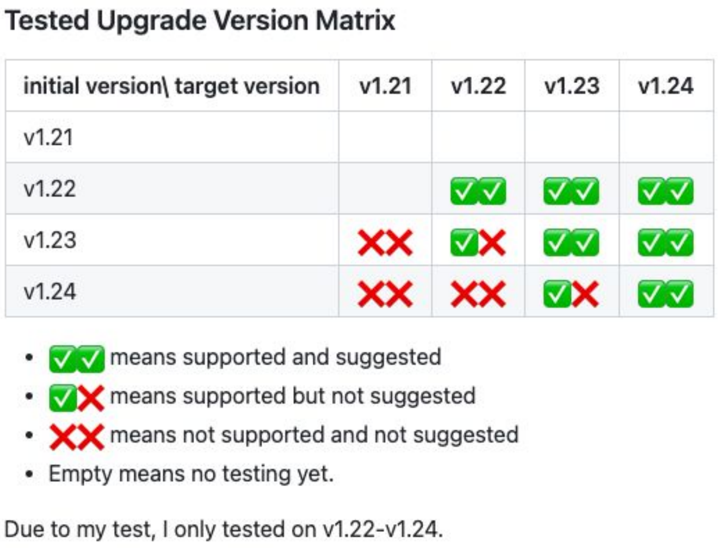
- kubeadm: Support skipping addons image pull kubernetes#114534 @ruquanzhao
- kubeadm: implementation of `UpgradeConfiguration`API types kubernetes#114062 @chendave
- kubeadm: implementation of `ResetConfiguration`API types kubernetes#113583 @chendave
- Support custom env in kubeadm `ControlPlaneComponent`kubeadm#2845
- add API support for controlling various timeouts during init / join kubeadm#2463
- handling of extraArgs which are allowed multiple times kubeadm#1601

Kubeadm operator

Currently the kubeadm-operator v0.1.0 can support upgrade cross versions like v1.22 to v1.24. It supports most kubeadm commands.

1. cluster upgrade (support dry run)
2. re-configuration
3. renew certs
4. upgrade addons: kube-proxy, coredns

See quick-start.



Tested Upgrade Version Matrix

initial version\ target version	v1.21	v1.22	v1.23	v1.24
v1.21				
v1.22		✅✅	✅✅	✅✅
v1.23	❌❌	✅❌	✅✅	✅✅
v1.24	❌❌	❌❌	✅❌	✅✅

- ✅✅ means supported and suggested
- ✅❌ means supported but not suggested
- ❌❌ means not supported and not suggested
- Empty means no testing yet.

Due to my test, I only tested on v1.22-v1.24.

kubeadm comparison

kubeadm	“Kubernetes **Node** operator”	simple and extensible, focus on kubelet/node
kubespray	“**OS** operators”	**ansible** + kubeadm bare metal and most clouds
kubeadm-operator*[1]	kubeadm-operator	focus on day-2 of kubeadm auto upgrade and configure
cluster-api	“Kubernetes Clusters operator”	uses Kubernetes-style APIs and patterns to automate cluster lifecycle management
kOps	“Kubernetes Operations”	focus on cloud infras (GAed in AWS and GCE) can generate **terraform**
kubean	“k8s-operator for kubespray”	no more ansible, just CRD.

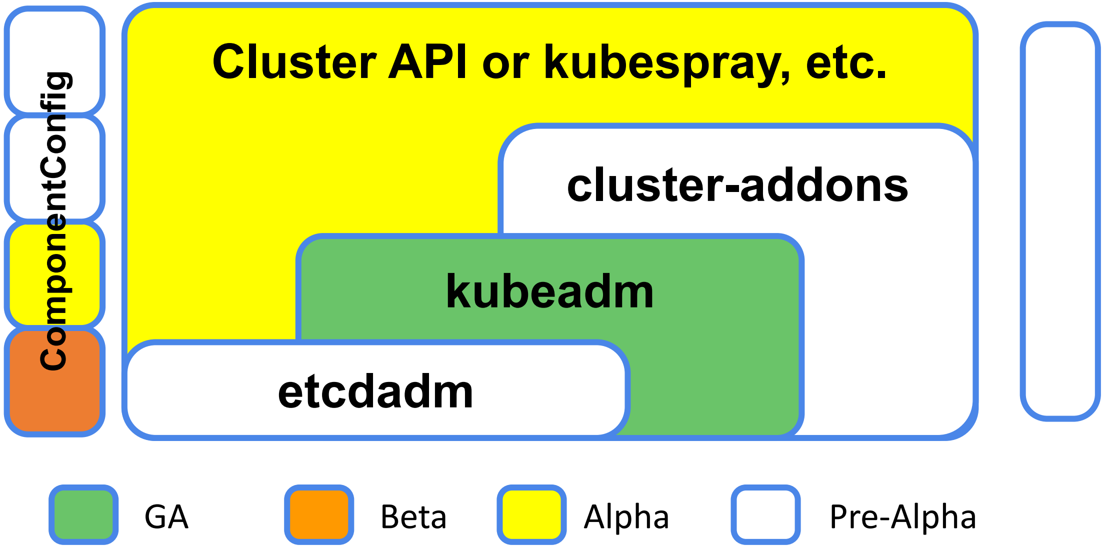
*[1] kubeadm-operator was originally incubated in the kubeadm project, but later removed due to lack of feedback.

Get Involved

Help Wanted

We need your help!

There is still a lot of work to do in order to build voltron!



Cluster API or kubespray, etc.

cluster-addons

kubeadm

etcdadm

ComponentConfig

🟩 GA

🟧 Beta

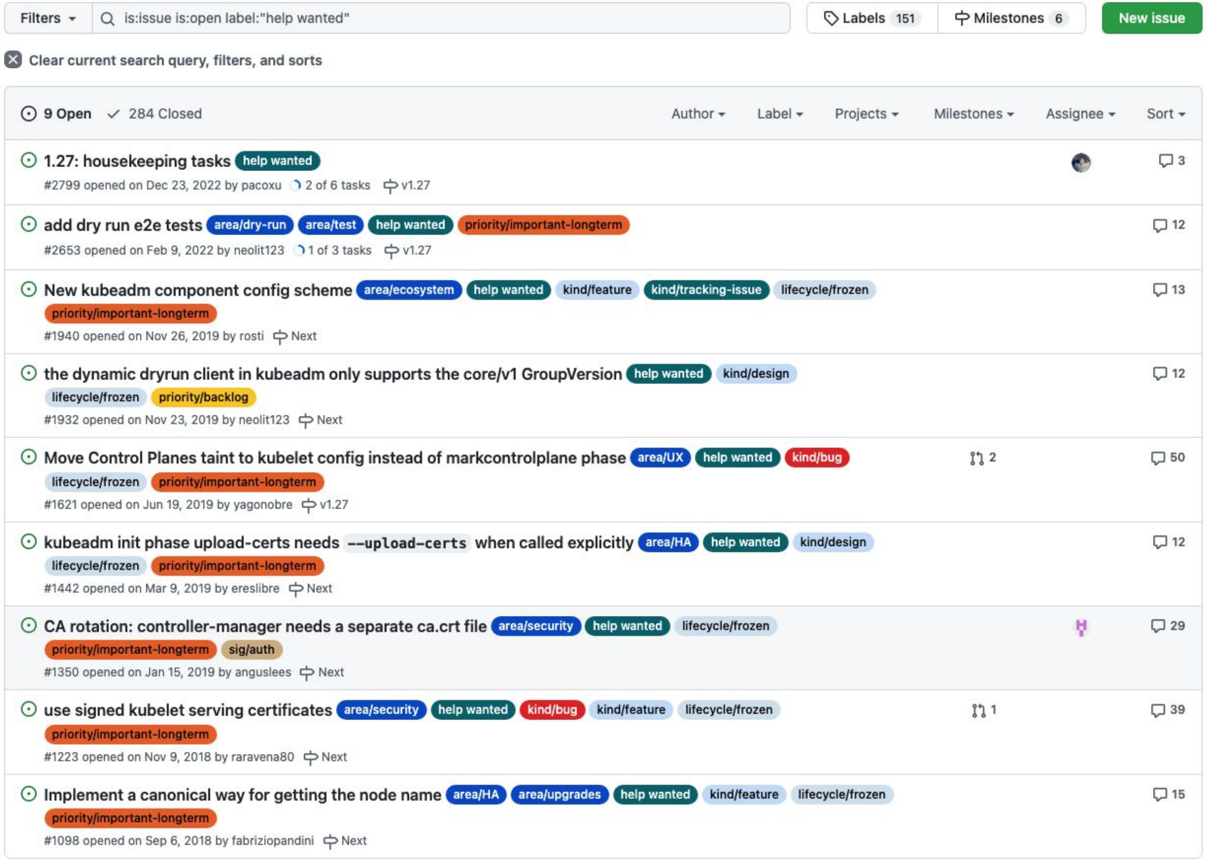
🟨 Alpha

⬜ Pre-Alpha

How you can contribute

- Kubeadm New Contributor Onboarding
- Navigate to our community page
- Look for “good first issue”, “help wanted” labeled issues in our repositories
- Help with docs and testing
- Attend our Zoom meetings, and ask questions
- Introduce yourself on Slack
- Attend/Watch new contributor sessions (SIG ContribEx)
- Chop wood, carry water, **be kind**
    - Everyone `**earns**` their place at the table (social capital)

Help Wanted



1.27: housekeeping tasks help wanted
#2799 opened on Dec 23, 2022 by pacoxu 2 of 6 tasks v1.27 — 3

add dry run e2e tests area/dry-run area/test help wanted priority/important-longterm
#2653 opened on Feb 9, 2022 by neolit123 1 of 3 tasks v1.27 — 12

New kubeadm component config scheme area/ecosystem help wanted kind/feature kind/tracking-issue lifecycle/frozen priority/important-longterm
#1940 opened on Nov 26, 2019 by rosti Next — 13

the dynamic dryrun client in kubeadm only supports the core/v1 GroupVersion help wanted kind/design lifecycle/frozen priority/backlog
#1932 opened on Nov 23, 2019 by neolit123 Next — 12

Move Control Planes taint to kubelet config instead of markcontrolplane phase area/UX help wanted kind/bug lifecycle/frozen priority/important-longterm
#1621 opened on Jun 19, 2019 by yagonobre v1.27 — 2 — 50

kubeadm init phase upload-certs needs `--upload-certs` when called explicitly area/HA help wanted kind/design lifecycle/frozen priority/important-longterm
#1442 opened on Mar 9, 2019 by ereslibre Next — 12

CA rotation: controller-manager needs a separate ca.crt file area/security help wanted lifecycle/frozen priority/important-longterm sig/auth
#1350 opened on Jan 15, 2019 by anguslees Next — 29

use signed kubelet serving certificates area/security help wanted kind/bug kind/feature lifecycle/frozen priority/important-longterm
#1223 opened on Nov 9, 2018 by raravena80 Next — 1 — 39

Implement a canonical way for getting the node name area/HA area/upgrades help wanted kind/feature lifecycle/frozen priority/important-longterm
#1098 opened on Sep 6, 2018 by fabriziopandini Next — 15

KubeCon | CloudNativeCon

Europe 2023

Please scan the QR Code above
to leave feedback on this session

FAQ

When does docker not supported?

The removal of dockershim was originally announced as a part of the Kubernetes v1.20 release. The Kubernetes v1.24 release actually removed the dockershim from Kubernetes. However, you can still use docker with cri-dockerd.

Why Container Runtime is using a different pause image?
For instance, containerd can configure the sandbox image separately. In v1.27, a warning message is added when detecting that the sandbox image of the container runtime is inconsistent with that used by kubeadm #115610

k8s.gcr.io Redirect to registry.k8s.io - What You Need to Know?
Traffic from the older k8s.gcr.io registry will be redirected to registry.k8s.io with the eventual goal of sunsetting k8s.gcr.io.