KubeCon | CloudNativeCon

North America 2018

Deep Dive: kubespray

Who are we?

Chad Swenson

- Lead Software Engineer at AT&T
  - Running kubespray clusters in production for 2.5 years
- Kubespray maintainer
- Active in SIG Cluster Lifecycle

Antoine Legrand

- Software Engineering Manager at Red Hat
- Kubespray maintainer and founder
- Active in SIG Apps and SIG Cluster Lifecycle

Goals of this session

Highlight the core values of kubespray and how it influenced design decisions

2018 achievements and 2019 Roadmap

The community behind Kubespray

The beginning

September 2015, two engineers started an internal campaign to use Kubernetes 1.0 in production....

The truth

... but miserably failed to convince anyone because

Managing Kubernetes in production is hard !

Mission of kubespray

Easily install and manage
Kubernetes clusters

Goals of kubespray

Production Readiness

On-premise (baremetal) first

Composable

Production Readiness

High-Availability

Security

Maintenance

Observability

High-Availability

Kubernetes Nodes

etcd

Kubernetes Control Plane

Cluster DNS and other cluster services

High-Availability - Nodes

- Node components must stay functional
  - Resource QoS
  - Priority Class
  - Reserved system and kube resources
- Multiple nodes needed to provide HA for applications
- Different use cases need a wide variety of options in their cluster topology, # and types of nodes
- Node Problem Detector - early work, more coming in 2019

High-Availability - etcd

- Etcd is a Raft based cluster
  - HA by default with 3+ nodes
  - Only use odd numbers for cluster size
- Smart-clients
  - Accepts list of IPs and manage the Load balancing client-side

High-Availability - Control Plane

- Inside Cluster
  - Node-level proxy
    - keepalived
    - nginx
  - Smart client or internal service?
    - Will follow community progress on this
- Outside cluster
  - External LB
  - External DNS
  - Many others, depends upon your infrastructure

High-Availability - Cluster Services

- Cluster services, like DNS, require multiple replicas
  - Replica count can be autoscaled varying on cluster size
  - kubernetes-incubator/cluster-proportional-autoscaler
- We may add vertical pod autoscaling as it matures

Security

- RBAC on by default
- Configurable Authentication
  - WebHook
  - Token
  - Basic
- Pod Security Policy
- Individual certificates for each node

Maintenance - Cluster Lifecycle

- Support full lifecycle of cluster operations
  - New cluster
  - Upgrade cluster
  - Scale a cluster
  - Remove nodes or an entire cluster
- Backup and restore
  - etcd snapshots taken during upgrade
  - Expect improvements here in 2019

Maintenance - Playbooks

- cluster.yml
  - Install or reconfigure a cluster
- upgrade-cluster.yml
  - Graceful rolling upgrade to a new version of kubespray
- scale.yml
  - Add a node to an existing cluster
- remove-node.yml
  - Remove a particular node from a cluster
- reset.yml
  - Uninstall kubespray from an entire cluster

Observability

- Metrics server
- Coming in 2019: Prometheus configurations by default
  - Alerts
  - Scrape endpoints
  - Will likely use Prometheus Operator
- Various dashboards

Goals of kubespray

Production Readiness

On-premise (baremetal) first

Composable

On-premise

- Bring your own Machines
- Good fit for Ansible
  - ssh
  - idempotent playbooks account for wide variety across host OS images
- Means we must install and tune OS packages and config that cloud provider images might already handle
- Bare metal compatibility usually guarantees other infrastructure types as a bonus

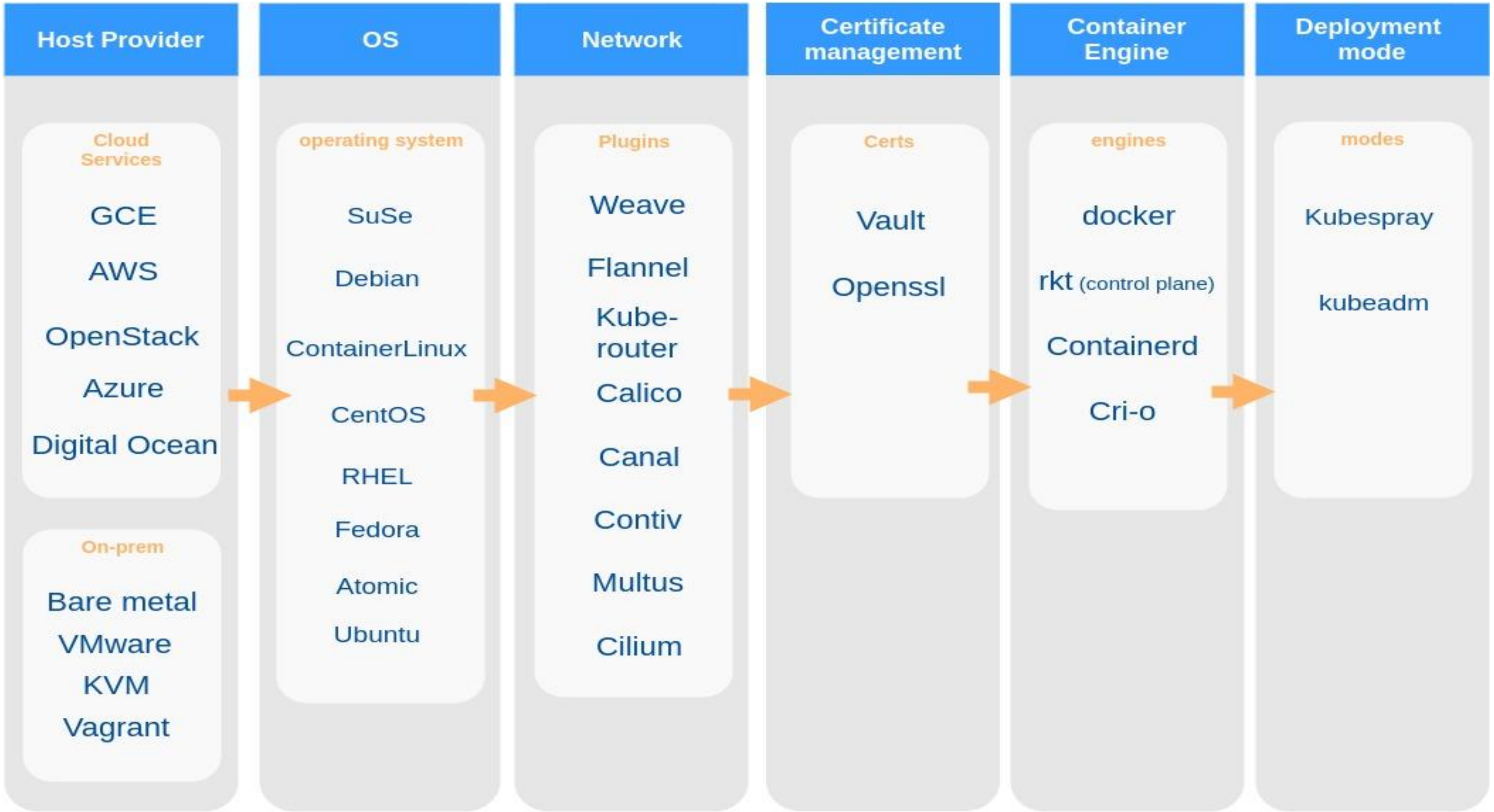
Goals of kubespray

Production Readiness

On-premise (baremetal) first

Composable

User Options



Host Provider

Cloud Services

GCE
AWS
OpenStack
Azure
Digital Ocean

On-prem

Bare metal
VMware
KVM
Vagrant

OS

operating system

SuSe
Debian
ContainerLinux
CentOS
RHEL
Fedora
Atomic
Ubuntu

Network

Plugins

Weave
Flannel
Kube-router
Calico
Canal
Contiv
Multus
Cilium

Certificate management

Certs

Vault
Openssl

Container Engine

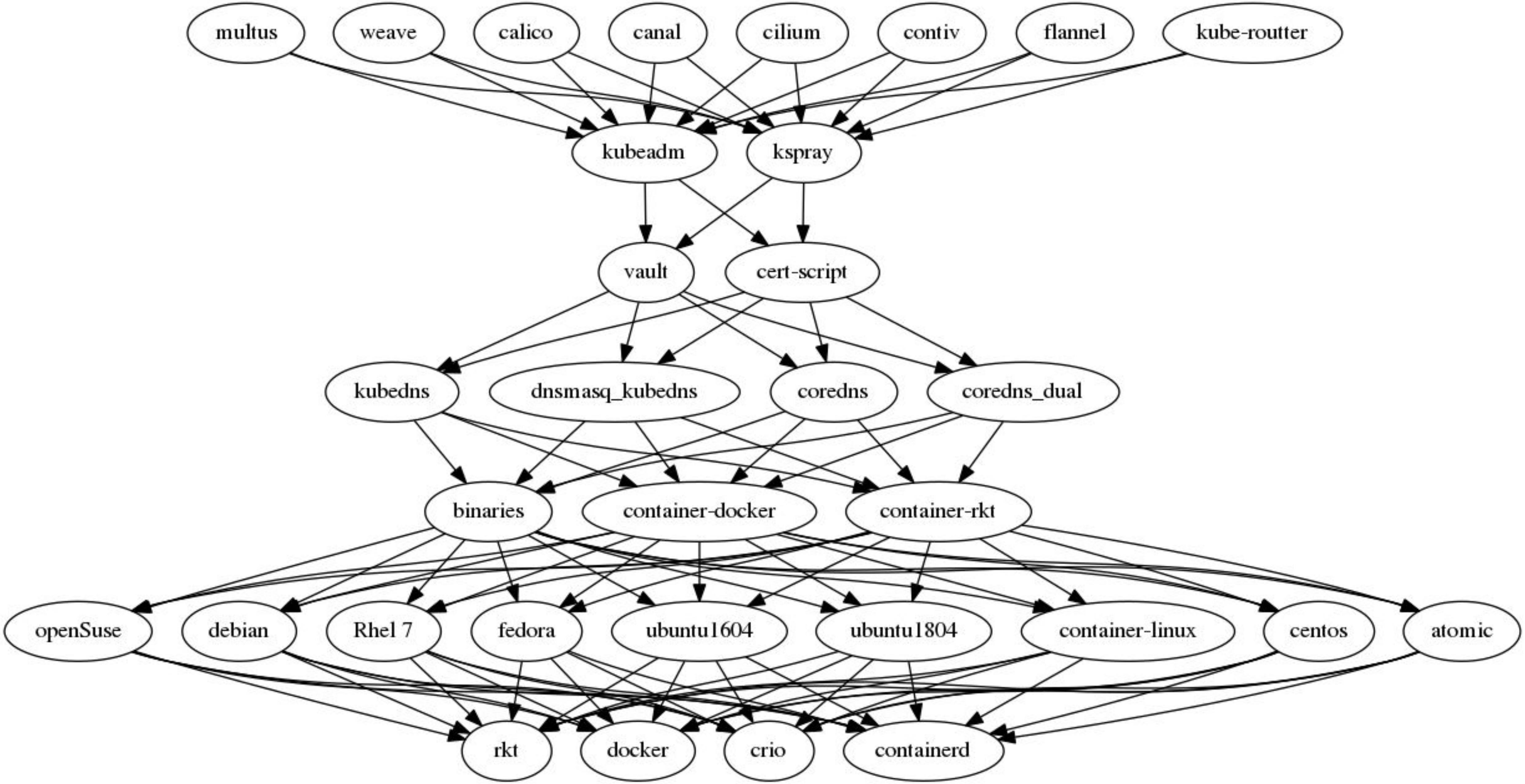
engines

docker
rkt (control plane)
Containerd
Cri-o

Deployment mode

modes

Kubespray
kubeadm


multus
weave
calico
canal
cilium
contiv
flannel
kube-routter
kubeadm
kspray
vault
cert-script
kubedns
dnsmasq_kubedns
coredns
coredns_dual
binaries
container-docker
container-rkt
openSuse
debian
Rhel 7
fedora
ubuntu1604
ubuntu1804
container-linux
centos
atomic
rkt
docker
crio
containerd

It's Chaos

14000+ legit paths to test, maintain and support !

Composability or Production Ready

Composability conflicts with the production-readiness Goal

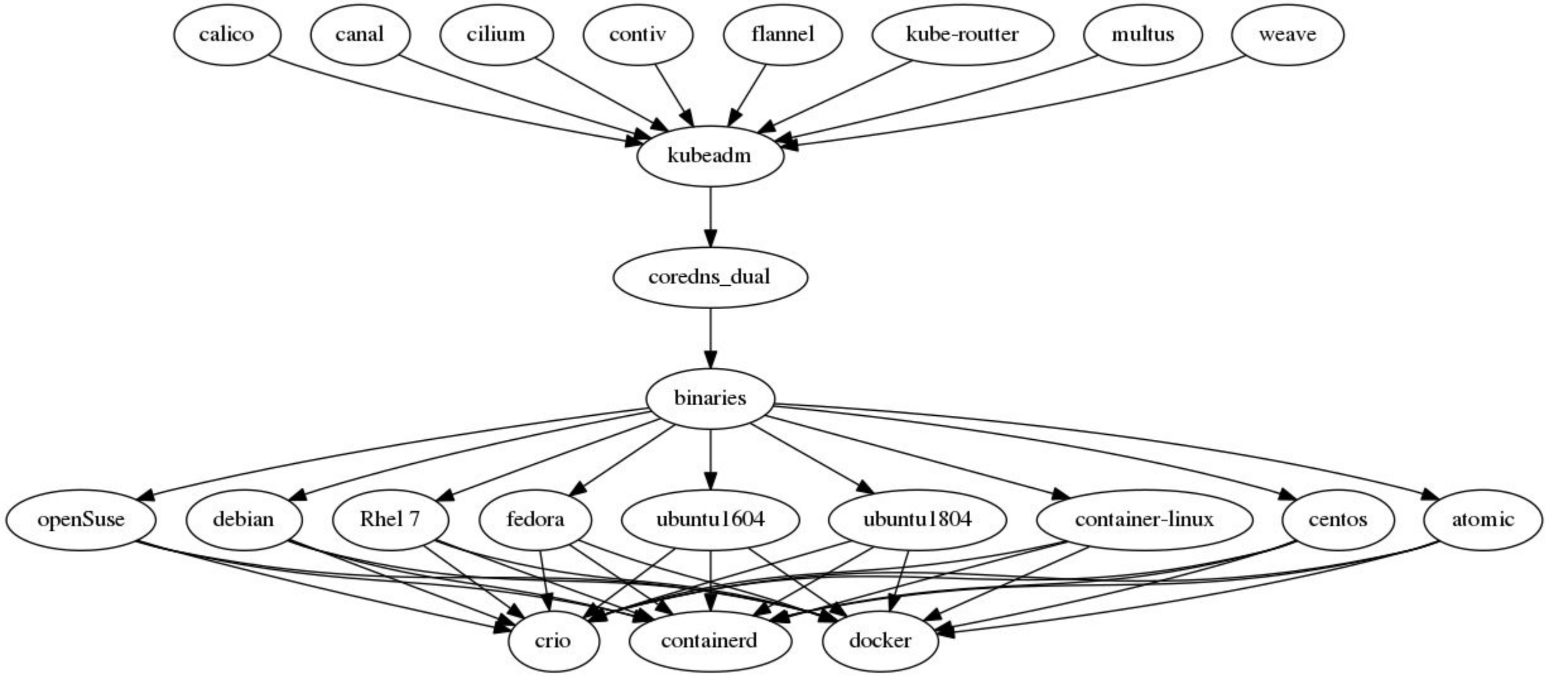
Keep Option Chaos Under Control

https://github.com/kubernetes-sigs/kubespray/issues/3508

- DNS: coredns, kubedns, kubedns_dnsmasq, coredns_dual,
- Deployment: Kubespray, Kubeadm
- Install-modes:
  - rkt+systemd
  - docker+systemd
  - binary+systemd
  - staticpod+kubelet
- cert-management: vault, script(kubespray)

Keep Option Chaos Under Control

14000 down to 200 paths



Kubespray Project In 2018

- kubeadm is the default deployment mode in v2.8+
  - Non-kubeadm is now deprecated and will be removed in v2.9
  - Upgrades from non-kubeadm are supported, please report issues!
- kubespray repo moved to the kubernetes-sigs GitHub project!
  - https://github.com/kubernetes/org/issues/208
- New Network Plugin Support
  - Cilium, Kube-router, Multus
- Many new k8s features and community components now integrated
  - PSP, PriorityClass, Dynamic Kubelet, and much more!
- Removed non-core components (add-ons)
- cri-o

Roadmap 2019

- Improve observability options out of the box
- Adopt (and build) new tools, best practices and features in alignment with the rest of SIG Cluster Lifecycle
  - CI
  - kubeadm
  - ComponentConfig
  - Bundles
  - etcdadm
- Decentralized orchestration
  - Auto-scaling
  - Automatic upgrades
  - Fast provisioning at scale
- Whatever you (the community) decide!

Who uses Kubespray?

- Startups, small, medium, large, and enterprise businesses, Governments
  - Commonly chosen option for anyone deploying consistently configured clusters on-prem or across multiple infrastructure providers
- Open-source projects
- Kubernetes distribution vendors

Kubespray Community In 2018

- 4000+ unique users directly participated in the repo in 2018
  - 857 contributed code, issues or comments.
  - 309 unique contributors created pull requests
  - CI built 14000 cluster VMs testing PRs
  - ~10,000 unique visitors per week to kubespray on GitHub
- 4,807 Stars
- Kubespray is frequently second the most active Kubernetes repository by commits (out of ~70) after kubernetes/kubernetes https://k8s.devstats.cncf.io/
- Growing fast
  - More than half of lifetime activity within the past six months
- Very active Slack channels #kubespray and #kubespray-dev on slack.k8s.io
- No company behind kubespray, 100% community driven!!

We ❤ new contributors!

Where are my contributors at?

Who here has contributed to kubespray?

KubeCon | CloudNativeCon

North America 2018