In-place Upgrade Noway! Blue/Green Your Way to a New Kubernetes Version

Me?

Ricardo Aravena

SRE Manager - Rakuten

CNCF SIG-Runtime Chair

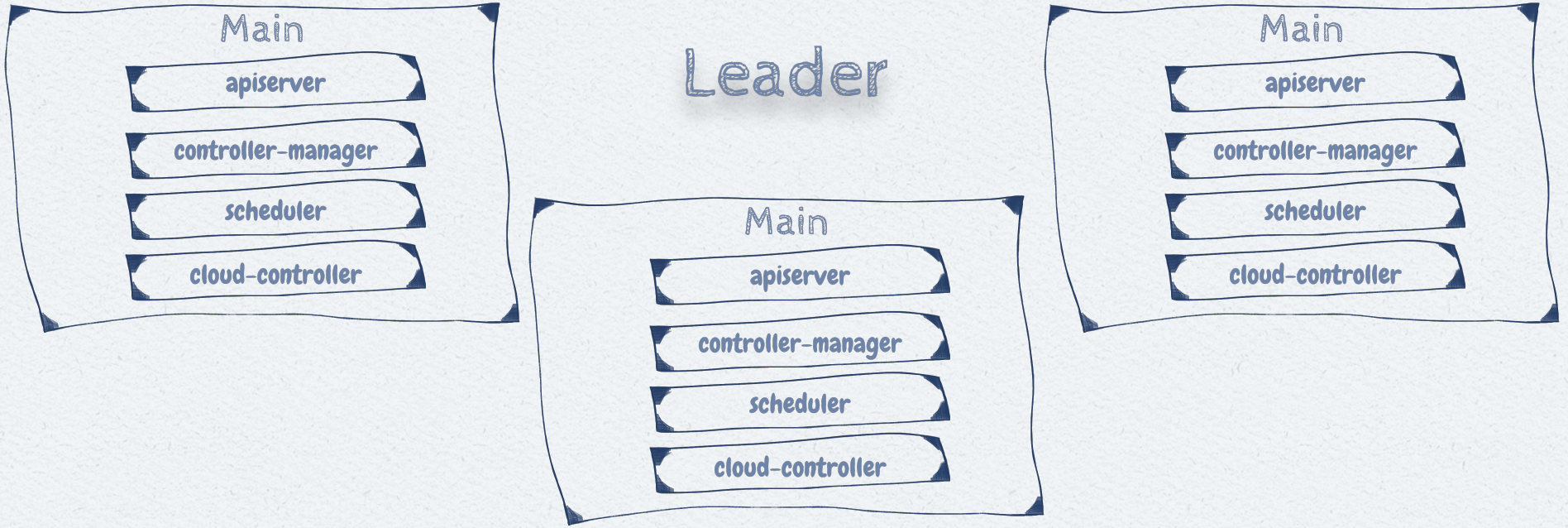
Kata Containers Contributor

@raravena80

Outline

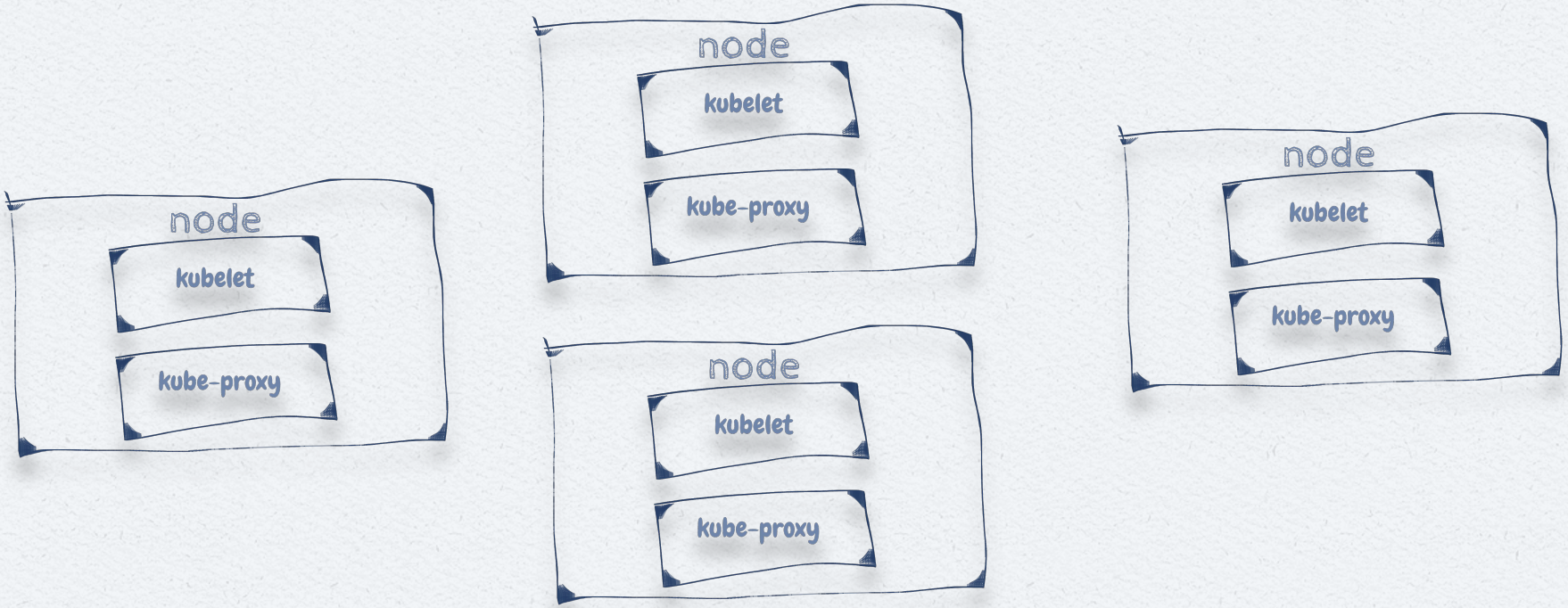
- Kubernetes Components
  - ✘ Control plane
  - ✘ Data plane
- Upgrade Problems
- Solutions & Tools
- Production
- Future
- Takeaways

K8s Control Plane

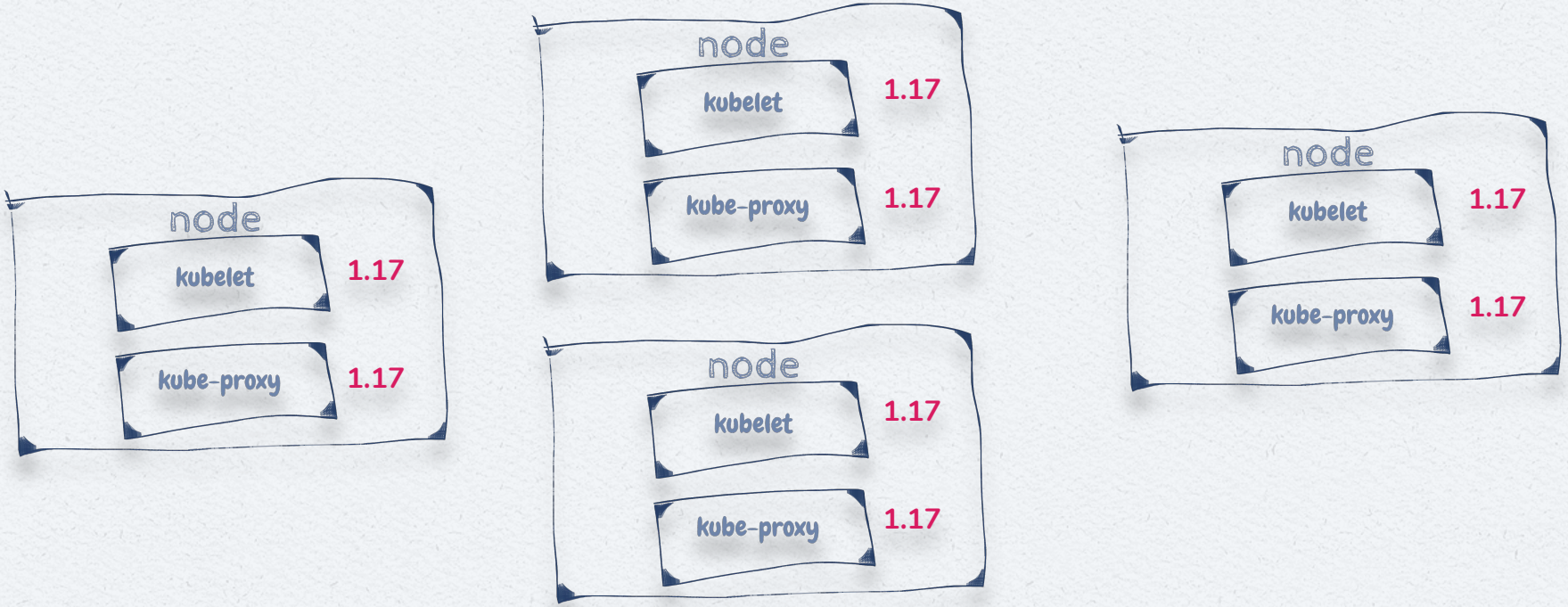


K8s Control Plane 1.17

K8s Data Plane



K8s Data Plane 1.17




K8s Upgrade Problems?

K8s Control Plane Upgrades

K8s API Versions

- ✘ Alpha newapigroup/v1alpha1
- ✘ Beta newapigroup/v1beta1
- ✘ Stable newapigroup/v1

K8s API Versions...

- ✘ newapigroup/v1alpha1 -> ... -> newapigroup/v1alphaN ->
- ✘ newapigroup/v1beta1 -> ... -> newapigroup/v1betaN ->
- ✘ newapigroup/v1 ->
- ✘ newapigroup/v2alpha1 -> ...

K8s 1.17 ... K8s 1.x Example

K8s 1.17 Release Notes

- ✘ All resources within the *rbac.authorization.k8s.io/v1alpha1* and *rbac.authorization.k8s.io/v1beta1* API groups are deprecated in favor of *rbac.authorization.k8s.io/v1*, and will no longer be served in *v1.20*. (#84758, @liggitt)
- ✘ The in-tree AWS EBS plugin *kubernetes.io/aws-ebs* is now deprecated and will be removed in *1.21*. Users that self-deploy Kubernetes on AWS should enable CSIMigration + CSIMigrationAWS features and install the AWS EBS CSI Driver to avoid disruption to existing Pod and PVC objects at that time. Users should start using the AWS EBS CSI CSI Driver directly for any new volumes. (#85237, @leakingtapan)

K8s 1.17 → K8s 1.20 RBAC Example...

K8s 1.17

```yaml
apiVersion: rbac.authorization.k8s.io/v1beta1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

K8s 1.20

```yaml
apiVersion: rbac.authorization.k8s.io/v1beta1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

K8s 1.17 → K8s 1.20 RBAC Example...

K8s 1.17

```yaml
apiVersion: rbac.authorization.k8s.io/v1beta1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

K8s 1.20

```yaml
apiVersion: rbac.authorization.k8s.io/v1beta1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

K8s 1.17 → K8s 1.20 RBAC Example...

K8s 1.17

```yaml
apiVersion: rbac.authorization.k8s.io/v1beta1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

K8s 1.20

```yaml
apiVersion: rbac.authorization.k8s.io/v1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

K8s 1.17 → K8s 1.20 RBAC Example...

K8s 1.17

```yaml
apiVersion: rbac.authorization.k8s.io/v1beta1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

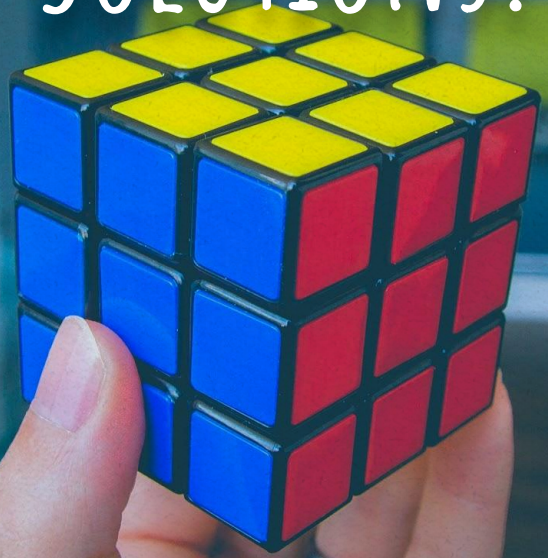
K8s 1.20

```yaml
apiVersion: rbac.authorization.k8s.io/v1
kind: Role
metadata:
  namespace: default
  name: pod-and-pod-logs-reader
rules:
- apiGroups: [""]
  resources: ["pods", "pods/log"]
  verbs: ["get", "list"]
```

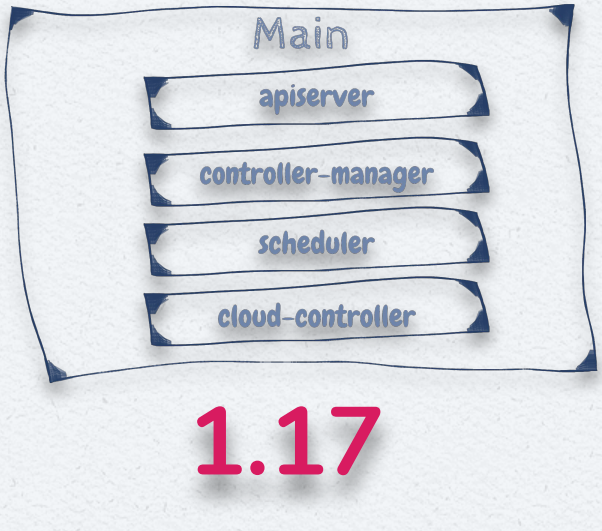
K8s Data Plane Upgrades

Supported version skews	kubelet 1.17, 1.16 → kube-apiserver 1.17
	kubectl 1.18, 1.17, 1.16 → kube-apiserver 1.17
	kube-controller-manager, kube-scheduler, cloud-controller-manager 1.16, 1.17 → kube-apiserver 1.17

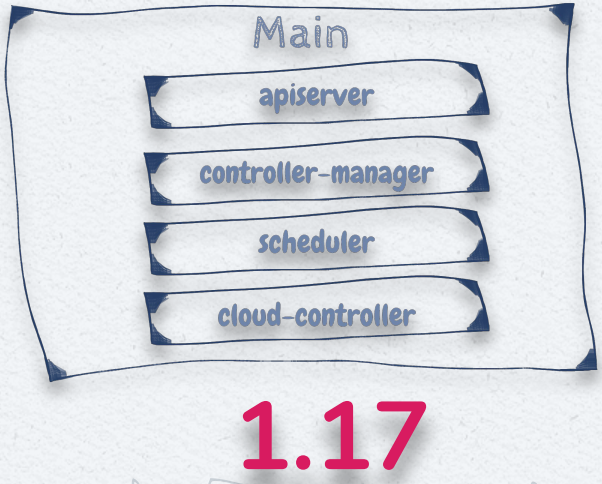
Solutions?

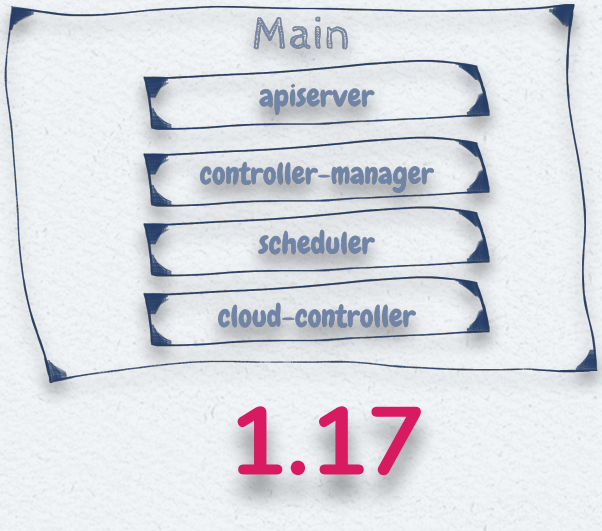


In-Place Upgrade

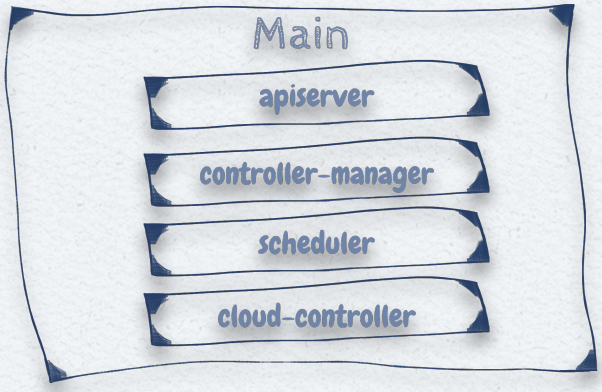


Leader



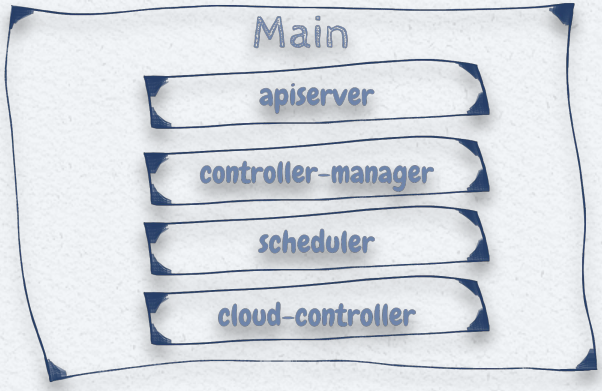


In-Place Upgrade

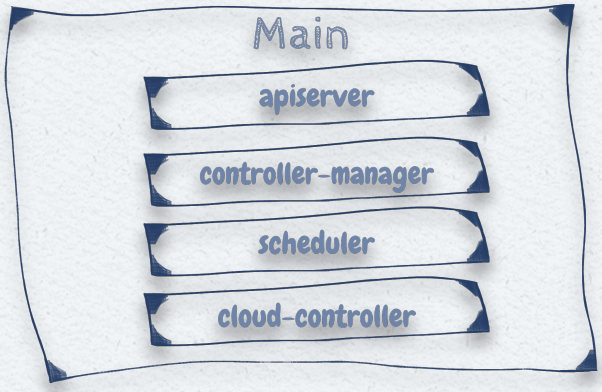


1.18

Leader

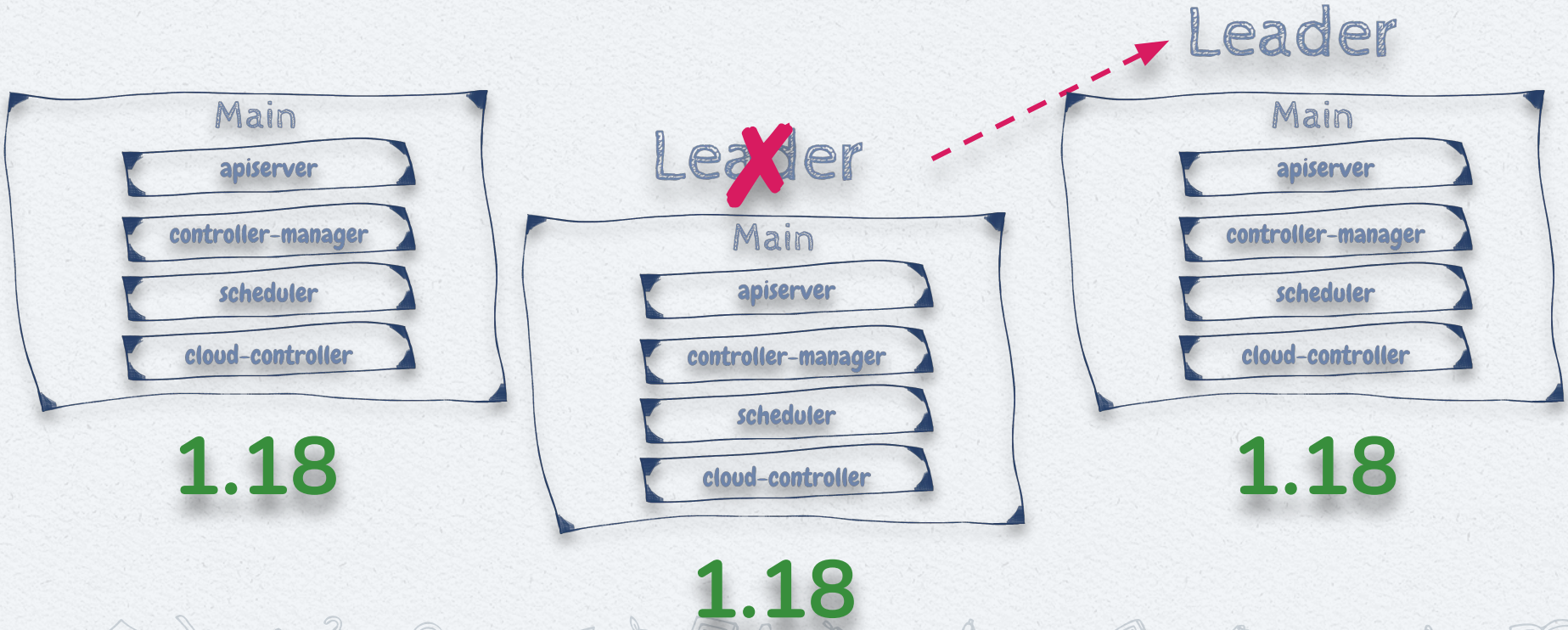


1.17



1.18

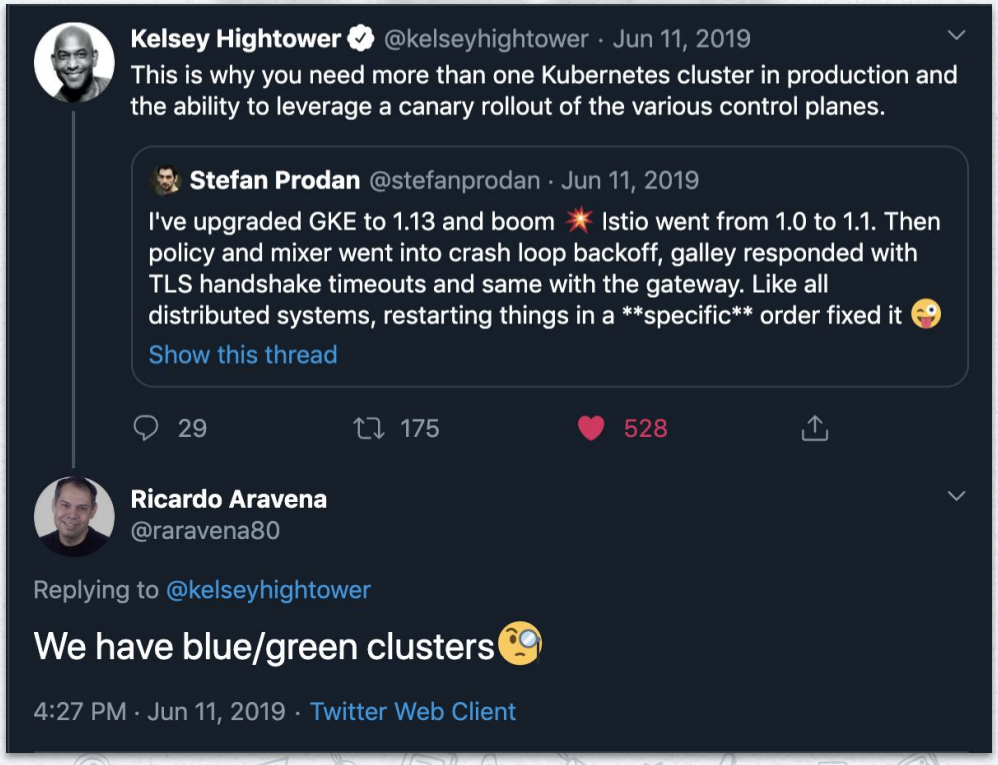
In-Place Upgrade



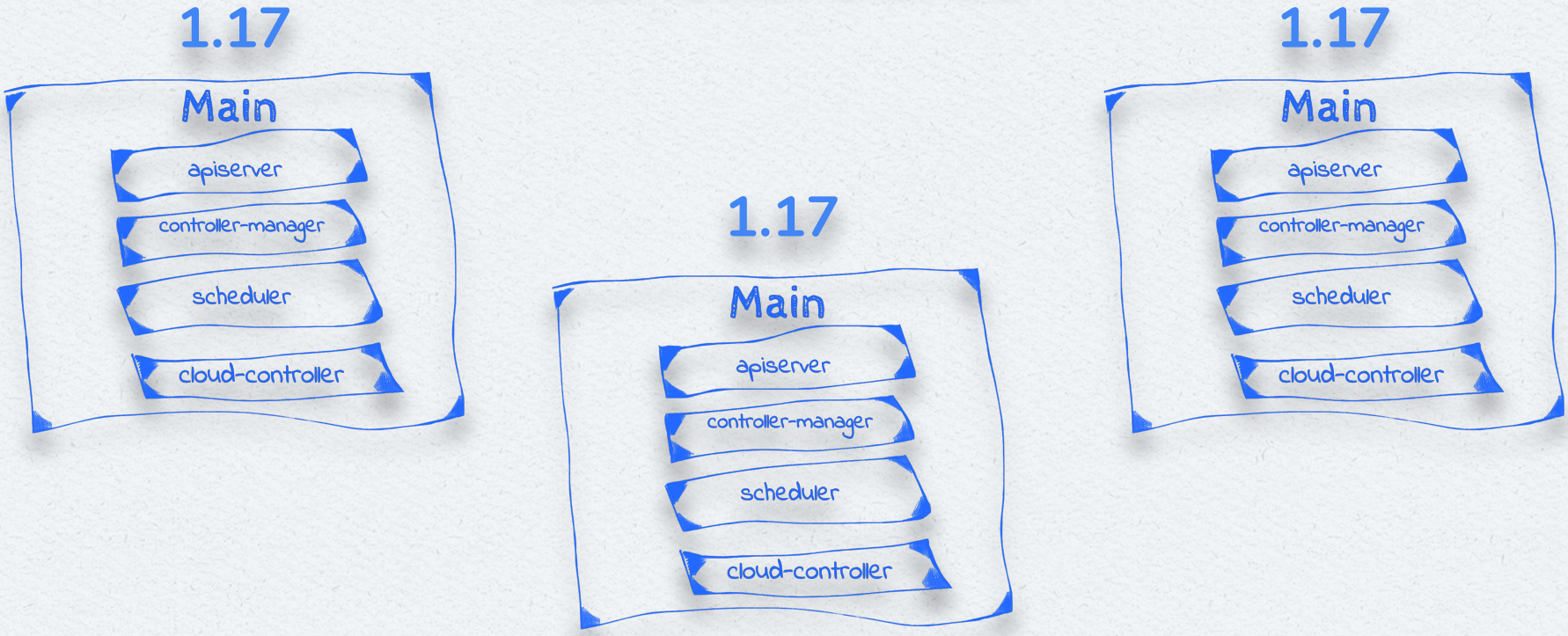
In-Place Upgrade

```
NAME                                          READY   STATUS                  RESTARTS   AGE
sitewhere-asset-management-d6d8f5fc4-6d2bn    0/1     Init:CrashLoopBackOff   4          2m35s
sitewhere-batch-operations-cb765d4f4-pws7q    0/1     Init:CrashLoopBackOff   4          2m35s
sitewhere-command-delivery-7c657bd7d6-wb22c   0/1     Init:CrashLoopBackOff   4          2m35s
sitewhere-device-management-78967857c6-s85fw  0/1     Init:CrashLoopBackOff   4          2m35s
sitewhere-device-registration-944fff8f6-gx8g6 0/1     Init:CrashLoopBackOff   4          2m35s
sitewhere-device-state-77f8b58dc8-2xs9j       0/1     Init:CrashLoopBackOff   4          2m35s
sitewhere-event-management-5f46d7676b-d9xg8   0/1     Init:CrashLoopBackOff   4          2m34s
sitewhere-event-search-6c55fd7f54-glkv7       0/1     Init:CrashLoopBackOff   4          2m34s
sitewhere-event-sources-96459bd4d-9x5sk       0/1     Init:CrashLoopBackOff   4          2m34s
sitewhere-inbound-processing-7756648fbb-6bkgm 0/1     Init:CrashLoopBackOff   4          2m34s
sitewhere-instance-management-5ffff59bbd-f9zmr 0/1    Init:CrashLoopBackOff   4          2m34s
sitewhere-jaeger-66c9b6769b-st24k             1/1     Running                 0          2m33s
sitewhere-kafka-0                             0/1     Pending                 0          2m34s
sitewhere-label-generation-5d9cc99c8b-mn8cd   0/1     Init:CrashLoopBackOff   4          2m33s
sitewhere-mosquitto-6db6959798-h2hnx          1/1     Running                 0          2m33s
sitewhere-outbound-connectors-f8cf79d76-ng9c6 0/1     Init:CrashLoopBackOff   4          2m33s
sitewhere-rule-processing-5465fb5bb4-rzrfz    0/1     Init:CrashLoopBackOff   4          2m33s
sitewhere-schedule-management-67fd6c5596-gjnsq 0/1    Init:CrashLoopBackOff   4          2m32s
sitewhere-streaming-media-6c4746d9b8-drf77    0/1     Init:CrashLoopBackOff   4          2m32s
sitewhere-tenant-management-7f9f89b6bd-8jnjv  0/1     Init:CrashLoopBackOff   4          2m32s
sitewhere-user-management-7b5f78c8bd-mmfvh    0/1     Init:CrashLoopBackOff   4          2m32s
```

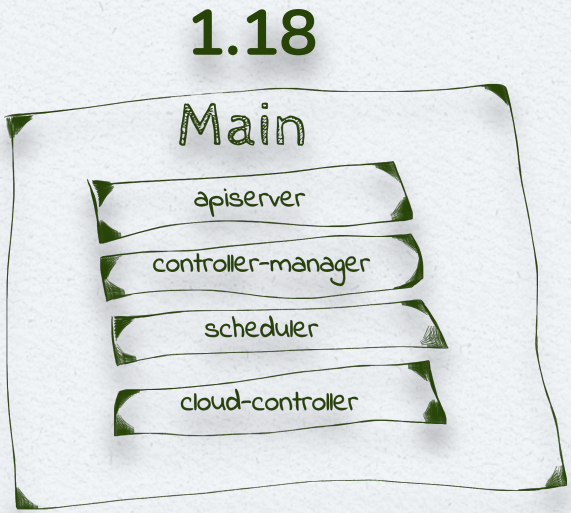
BLUE/GREEN CLUSTERS

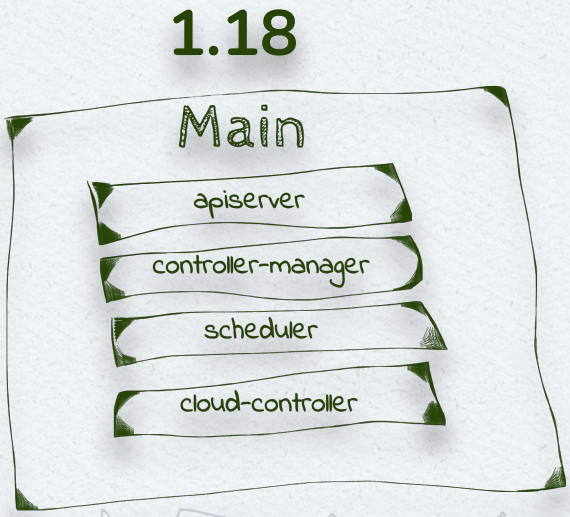


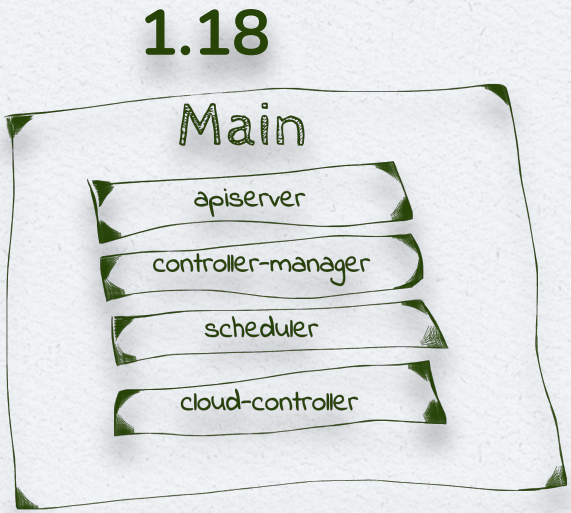
BLUE/GREEN CLUSTERS



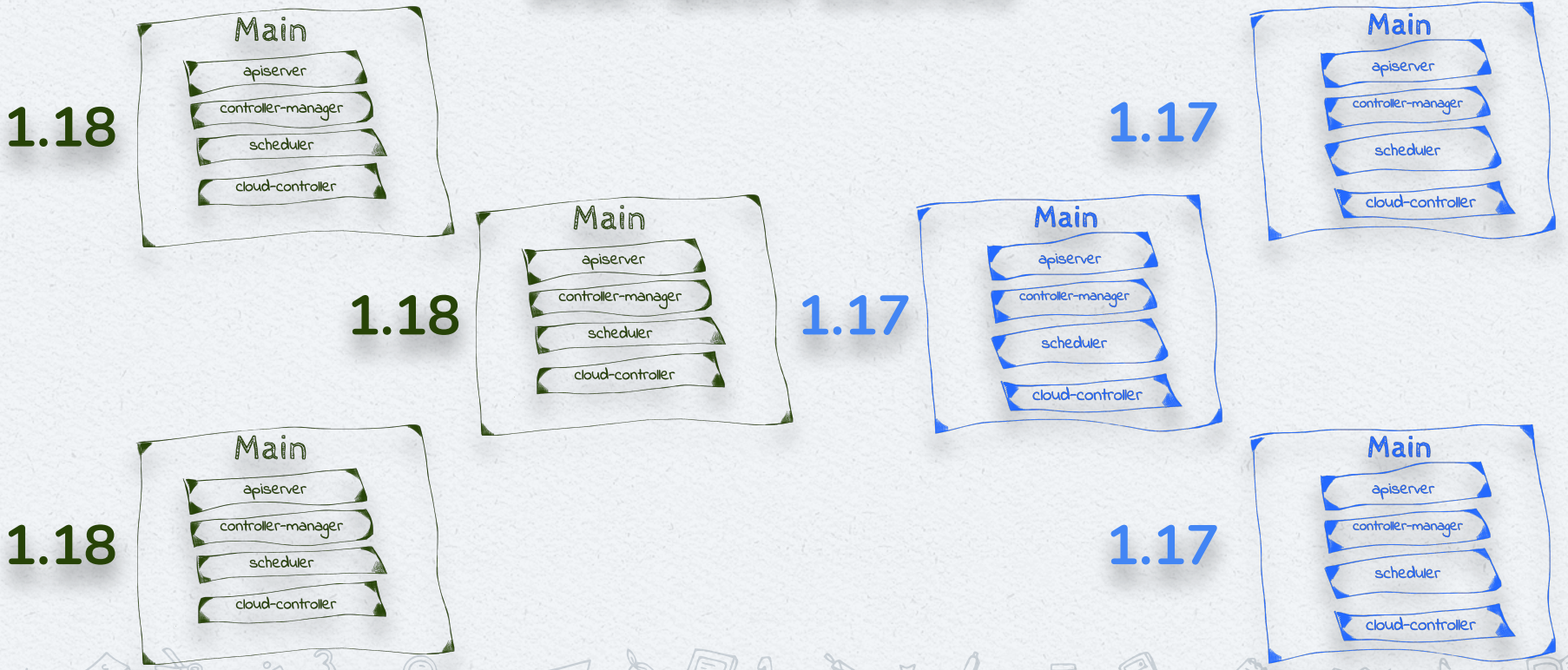
BLUE/GREEN CLUSTERS



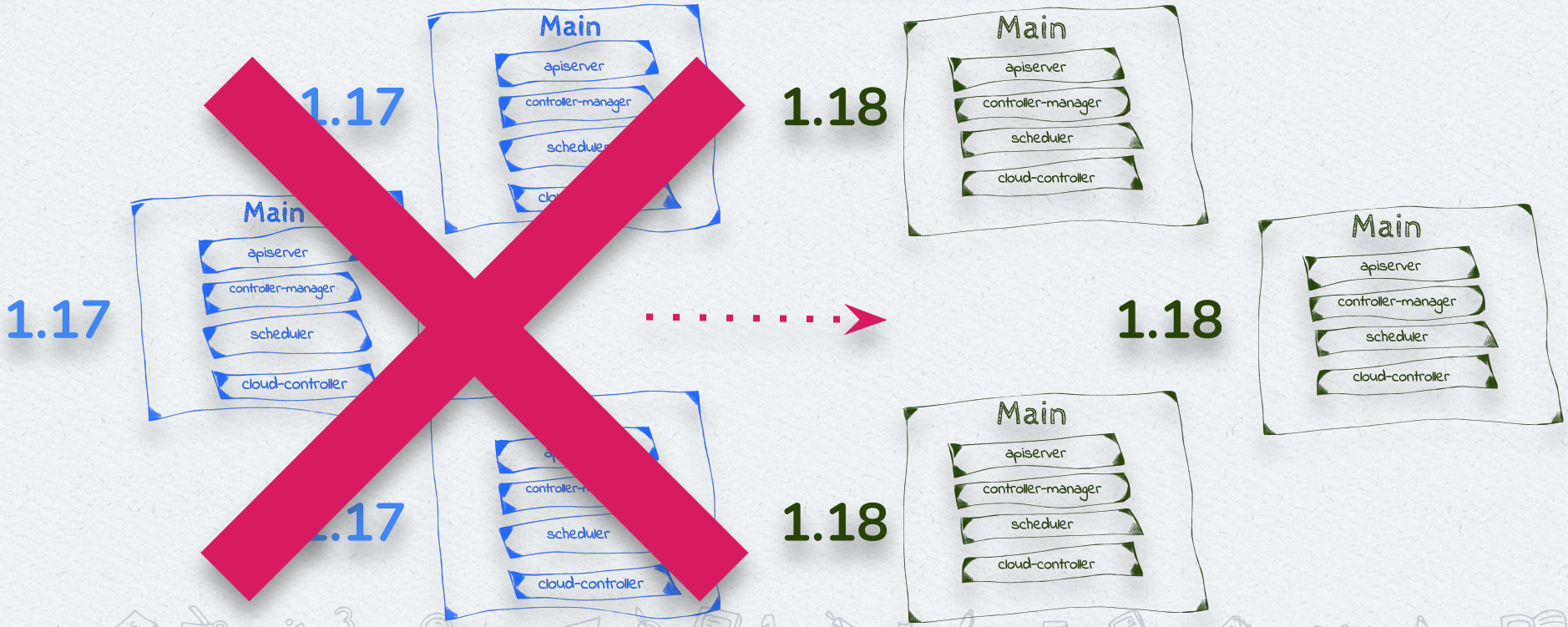




BLUE/GREEN CLUSTERS

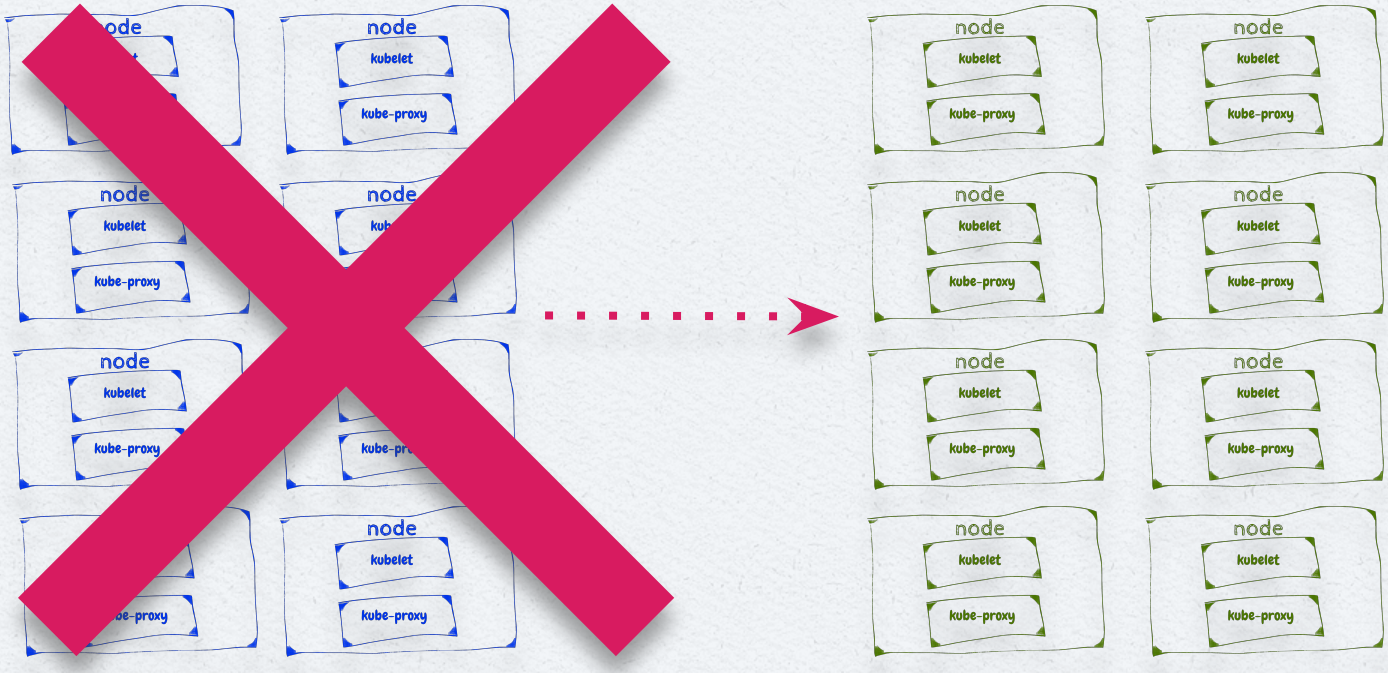


BLUE/GREEN CLUSTERS



BLUE/GREEN CLUSTERS - NODES

1.17

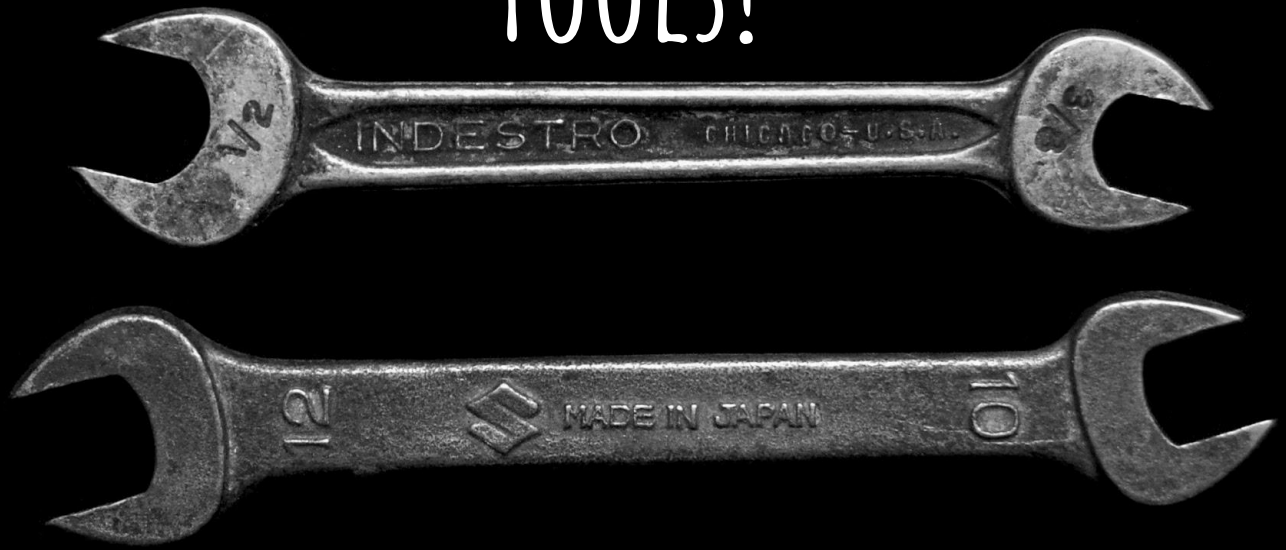


1.18

Other Blue/Green Approaches

- ✘ Federation V2
  - ✘ Alpha
  - ✘ Multi-region
- ✘ Cluster Mirroring
  - ✘ Linkerd
- ✘ Cluster API
  - ✘ v1alpha2
  - ✘ Management cluster

Tools!



Bash? Or Python or Go or Whatever

```
#!/bin/bash

all_update(){
    local local_dir=""
    local_dir="$( cd "$( dirname "${BASH_SOURCE[0]}" )" && pwd )"
    source "${local_dir}/../cli/update"
    source "${local_dir}/../networking/update"
    source "${local_dir}/../coredns/update"
    source "${local_dir}/../storage/update"
    source "${local_dir}/../access/update"
    source "${local_dir}/../tiller/update"
    source "${local_dir}/../flux/update"
    source "${local_dir}/../kube2iam/update"
    source "${local_dir}/../keiko/update"
    source "${local_dir}/../nginx-ingress/update"
    source "${local_dir}/../monitoring/update"
    source "${local_dir}/../sysdig/update"
    source "${local_dir}/../fluentd/update"
    source "${local_dir}/../kube-system-namespace/update"
    source "${local_dir}/../cluster-autoscaler/update"
    source "${local_dir}/../spot-termination-handler/update"
    source "${local_dir}/../fluent-bit/update"
}
all_update
```

Terraform/eksctl

```
module "my-cluster" {
  source          = "terraform-aws-modules/eks/aws"
  cluster_name    = "my-cluster"
  cluster_version = "1.16"
  subnets         = ["subnet-abcde012", "subnet-bcde012a"]
  vpc_id          = "vpc-1234556abcdef"

  worker_groups = [
    {
      instance_type = "m4.large"
      asg_max_size  = 5
    }
  ]
}
```

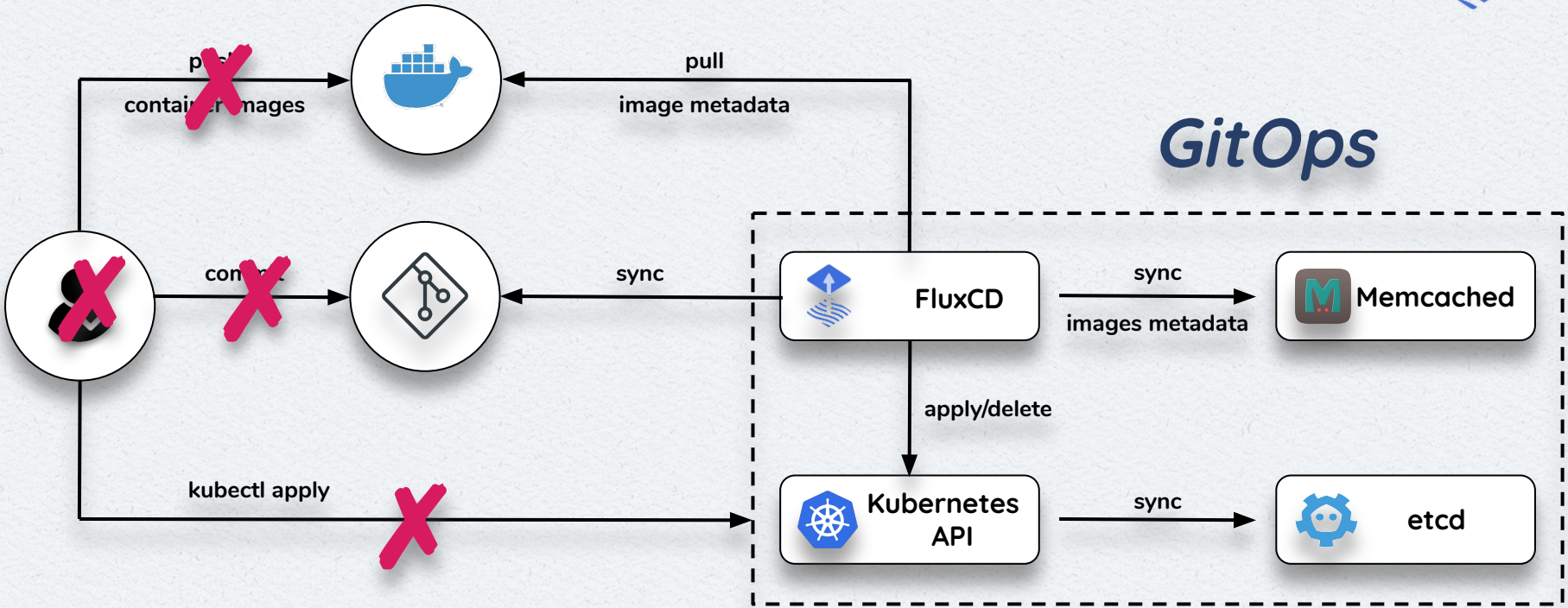
```
apiVersion: eksctl.io/v1alpha5
kind: ClusterConfig
metadata:
  name: basic-cluster
  region: eu-north-1
nodeGroups:
  - name: ng-1
    instanceType: m5.large
    desiredCapacity: 10
    volumeSize: 80
    ssh:
      allow: true
  - name: ng-2
    instanceType: m5.xlarge
    ...
```

Flux

Manage Add-ons

- ✘ Coredns
- ✘ Cluster-autoscaler
- ✘ Network overlay (Calico, Weave, CNI, Cillium, etc)
- ✘ Dashboard
- ✘ KubeVirt

Flux


pull
container images
pull
image metadata
GitOps
commit
sync
FluxCD
sync
images metadata
Memcached
apply/delete
kubectl apply
Kubernetes API
sync
etcd

WKSCTL

What is it?

- ✘ GitOps for clusters
- ✘ State in YAML file
- ✘ Supports
  - ✘ Firekube
  - ✘ Vagrant
  - ✘ Footloose
  - ✘ GCE
  - ✘ No AWS yet

Keiko

How I?

- ✘ Bootstrap and manage worker nodes for my cluster?
- ✘ Mitigate spurious pod/node failures as well as maintain SLAs and compliance?
- ✘ Manage critical cluster services required across all apps on clusters?
- ✘ Optimize cost of my cluster?
- ✘ Do forensic dumps?

Keiko...


Security
Kube-Forensics
Monitoring
Active-monitor
Reliability
Governor
Cost Eff
Minion Manager
Orchestration
Instance-manager
Upgrade-manager
Addon Manager

Other Tools...

- Kops
  - Golang
- Kubespray
  - Ansible
- Linkerd
  - Traffic mirroring

Pluto

```
○ → pluto list-versions
KIND                          NAME                                    DEPRECATED IN   REMOVED IN   REPLACEMENT                          COMPONENT
Deployment                    extensions/v1beta1                      v1.9.0          v1.16.0      apps/v1                              k8s
Deployment                    apps/v1beta2                            v1.9.0          v1.16.0      apps/v1                              k8s
Deployment                    apps/v1beta1                            v1.9.0          v1.16.0      apps/v1                              k8s
StatefulSet                   apps/v1beta1                            v1.9.0          v1.16.0      apps/v1                              k8s
StatefulSet                   apps/v1beta2                            v1.9.0          v1.16.0      apps/v1                              k8s
NetworkPolicy                 extensions/v1beta1                      v1.9.0          v1.16.0      networking.k8s.io/v1                 k8s
Ingress                       extensions/v1beta1                      v1.14.0         v1.22.0      networking.k8s.io/v1beta1            k8s
DaemonSet                     apps/v1beta2                            v1.9.0          v1.16.0      apps/v1                              k8s
DaemonSet                     extensions/v1beta1                      v1.9.0          v1.16.0      apps/v1                              k8s
PodSecurityPolicy             extensions/v1beta1                      v1.10.0         v1.16.0      policy/v1beta1                       k8s
ReplicaSet                    extensions/v1beta1                      n/a             v1.16.0      apps/v1                              k8s
ReplicaSet                    apps/v1beta1                            n/a             v1.16.0      apps/v1                              k8s
ReplicaSet                    apps/v1beta2                            n/a             v1.16.0      apps/v1                              k8s
PriorityClass                 scheduling.k8s.io/v1beta1               v1.14.0         v1.17.0      scheduling.k8s.io/v1                 k8s
PriorityClass                 scheduling.k8s.io/v1alpha1              v1.14.0         v1.17.0      scheduling.k8s.io/v1                 k8s
CustomResourceDefinition      apiextensions.k8s.io/v1beta1            v1.16.0         v1.19.0      apiextensions.k8s.io/v1              k8s
MutatingWebhookConfiguration  admissionregistration.k8s.io/v1beta1    v1.16.0         v1.19.0      admissionregistration.k8s.io/v1      k8s
ClusterRoleBinding            rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
ClusterRole                   rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
ClusterRoleBindingList        rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
ClusterRoleList               rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
Role                          rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
RoleBinding                   rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
RoleList                      rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
RoleBindingList               rbac.authorization.k8s.io/v1alpha1      v1.17.0         v1.20.0      rbac.authorization.k8s.io/v1         k8s
CSINode                       storage.k8s.io/v1beta1                  v1.17.0         n/a          n/a                                  k8s
```


https://github.com/FairwindsOps/pluto

Production...

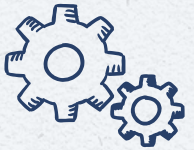
Full automation!

- ✘ K8s control plane
- ✘ K8s data plane
- ✘ Cluster add-ons
- ✘ Stateless applications
- ✘ Blue/Green traffic int/ext switchover

Production Stateful Apps?...

- ✘ Usually requires a maintenance window
- ✘ Backup data!
- ✘ Move data through snapshots, or re-use volumes
- ✘ Create replicas!
- ✘ Use master switchover
  - ✘ Multiples read replicas → one becomes new master in new cluster
- ✘ Multiple masters
  - ✘ Multiple read/write replicas → one by one in new cluster

Future Tools? - Gaps



- ✘ Automatic traffic switchover
- ✘ Stateful applications switchover
- ✘ Monitoring upgrades
  - ✘ Addon warnings
- ✘ Security checks in upgrades
- ✘ Service mesh multi-cluster
- ✘ Operator cluster upgrade awareness

Cloud providers Example: EKS

40 Researching

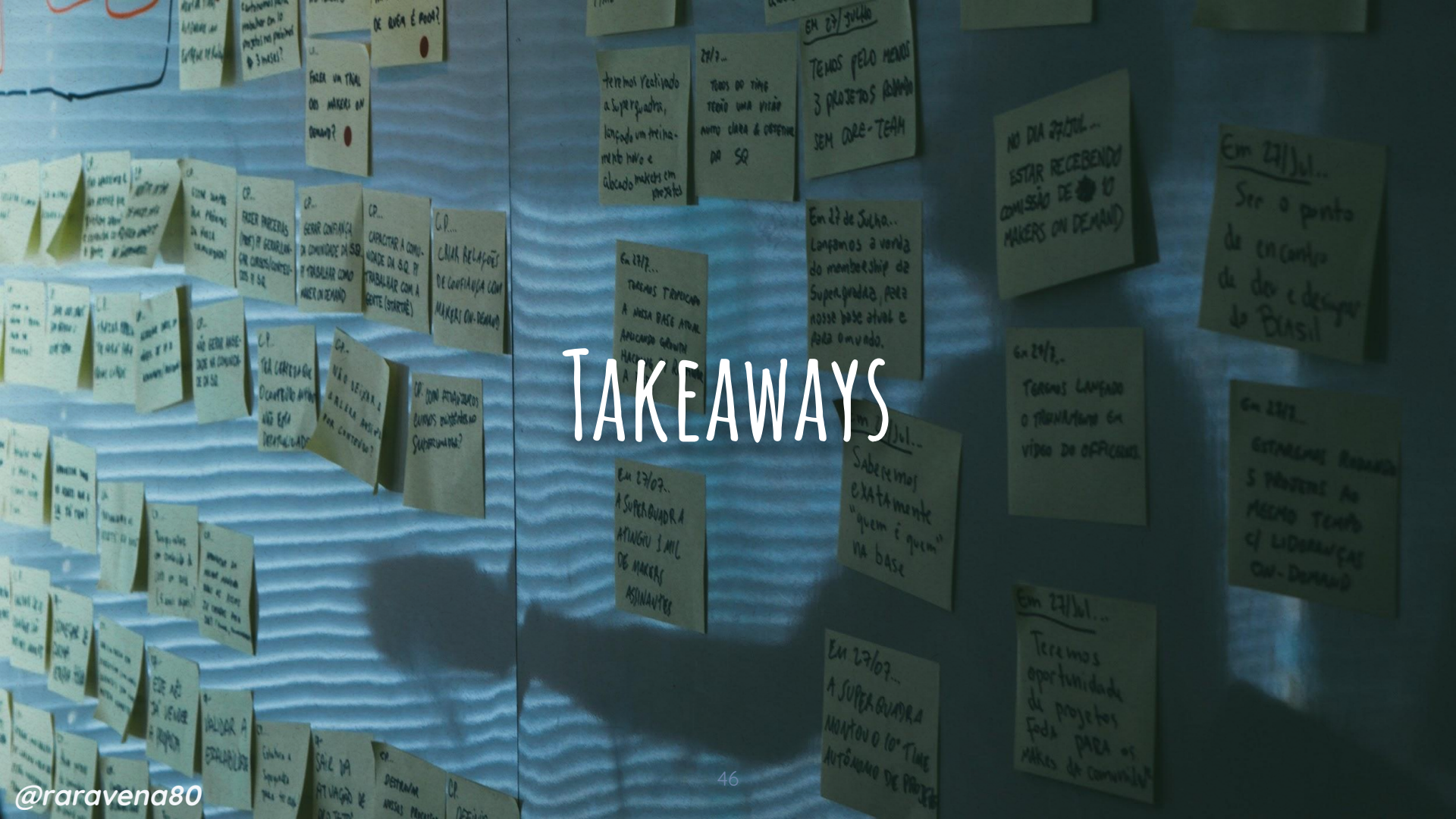
- [EKS] [request]: Notifications / More control over EKS Master Node Patch Rollouts
  #604 opened by johnpemberton
  EKS Proposed
- [EKS] [request]: EKS Managed Nodes should allow for custom security groups
  #609 opened by bhops
  EKS Proposed
- [EKS] [request]: On create: only return ACTIVE when endpoint actually usable

50 We're Working On It

- [EKS]: One-click Full Cluster Upgrade
  #600 opened by mohitanchlia
  EKS Proposed
- ECS Service Discovery not respecting TTL when updating service
  #343 opened by matthewduren
  ECS Proposed
- Targets not deregistered after task is stopped, serious routing issue with dynamic port mapping
  #347 opened by Vilsepi

10 Coming Soon

- [EKS] Managed Nodes Upgrade in Console
  #605 opened by tabern
  EKS
- New EKS Region : Beijing
  #219 opened by wholroyd
  EKS Proposed
- Interactively Upgrade EKS Worker Node
  #57 opened by mrichman
  EKS Proposed

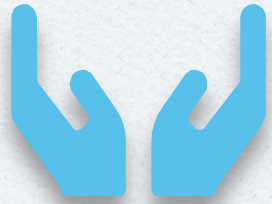


Takeaways

- ✘ Start in QA (never in production)
- ✘ Start small
- ✘ Leverage Keiko/Terraform - Flux/Terraform
- ✘ Fully create ‘Green’ both control and data plane
  - ✘ No mixed versions
- ✘ Leverage kubectl convert!
- ✘ Always test prior to ‘Green’ traffic switchover
- ✘ Backup, backup! if running stateful apps

References

- ✘ Keiko https://github.com/keikoproj
- ✘ Flux https://github.com/fluxcd/flux
- ✘ Pluto https://github.com/FairwindsOps/pluto
- ✘ Linkerd Multi-cluster https://linkerd.io/2/features/multicluster/
- ✘ KubeSpray https://github.com/kubernetes-sigs/kubespray
- ✘ Kops https://github.com/kubernetes/kops
- ✘ Cillium Multi-cluster https://cilium.io/blog/2019/03/12/clustermesh/



THANKS!

Any questions? I'd love to chat more

You can find me on Twitter

@raravena80