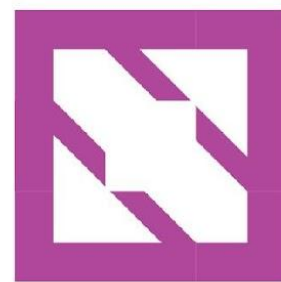




**KubeCon**



**CloudNativeCon**

**Europe 2025**





KubeCon



**CloudNativeCon**

## Europe 2025

# etcd v3.6.0-rc.3 and etcd-operator v0.1.0

Thursday April 3, 2025 16:00 - 16:30 BST

- Arka Saha (@ArkaSaha30)
- Benjamin Wang (@ahrtr)
- Ciprian Hacman (@hakman)
- Ivan Valdes Castillo (@ivanvc)
- Siyuan Zhang (@siyuanfoundation)



# Speakers



Arka Saha  
(@ArkaSaha30)



Benjamin Wang  
(@ahrtr)



Ciprian Hacman  
(@hakman)



Ivan Valdes Castillo  
(@ivanvc)



Siyuan Zhang  
(@siyuanfoundation)



# New etcd maintainer/approver

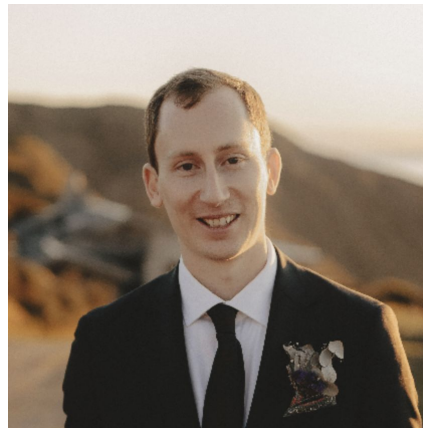


Fu Wei (@fuweid)

## etcd release leads



Ivan Valdes Castillo  
(@ivanvc)



James Blair (@jmhbzn)

# etcd v3.6.0-rc.3 and etcd-operator v0.1.0



KubeCon



CloudNativeCon

Europe 2025

## Content

- etcd v3.6.0 features
  - migration to v3store
  - downgrade
  - feature gate
  - livez/readyz checks
  - v3discovery
  - upgrade issue
- etcd v3.6.0 performance
  - performance tool/measure
  - performance improvement (memory)
- release
  - new release team
  - release process etc.
- WG etcd-operator
- etcd-operator v0.1.0 & future work
  - reconciliation workflow
  - data persistence
  - future work
- Q/A

## etcd v3.6.0 features: migration to v3store



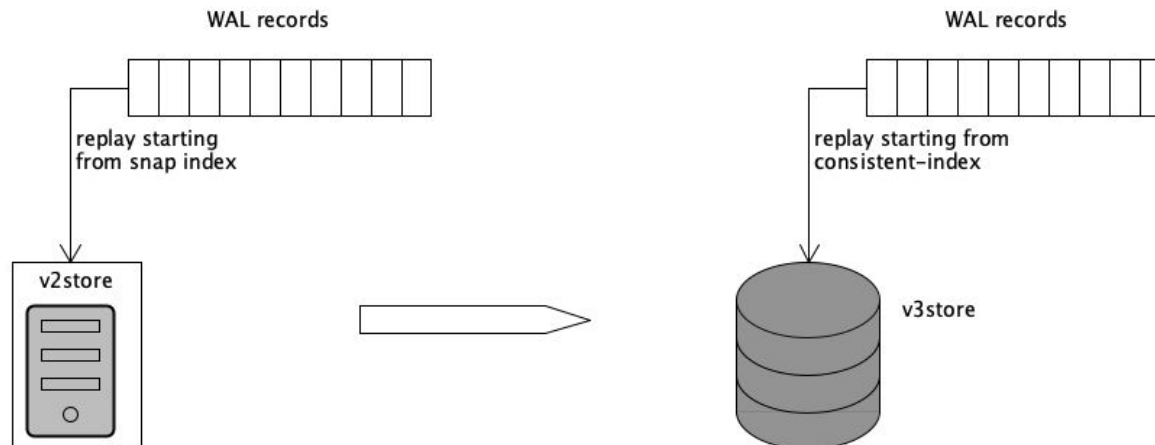
KubeCon



CloudNativeCon

Europe 2025

- v2store:
  - example: 000000000000000002-000000000000186a1.snap
  - deprecated since release-3.4, but it can be enabled via flag “--enable-v2” up to release-3.5.
  - source of truth for membership data
- v3store: bbolt db file
  - `${data-dir}/member/snap/db`
  - source of truth for the membership data
  - The flag “--enable-v2” has been removed in release-3.6, so no way to enable v2store anymore.
- future work in release-3.7
  - bootstrap etcd from `consistent_index`



# etcd v3.6.0 features: downgrade

- etcd v3.6.0 is the first version to fully support downgrade
- At a high level, the process involves
  - migrating the data schema to the target version (e.g. 3.5)
  - rolling downgrade each member
- From user perspective
  - Step 1: validate the downgrade
    - etcdctl command
      - \$ etcdctl downgrade validate 3.5  
Downgrade validate success, cluster version 3.6
  - Step 2: enable the downgrade
    - etcdctl command
      - \$ etcdctl downgrade enable 3.5  
Downgrade enable success, cluster version 3.6
    - wait for the storage version being changed to the target version

```
F5WPF69Q7H:etcd wachao$ ./bin/etcdctl endpoint status -w table --cluster
```

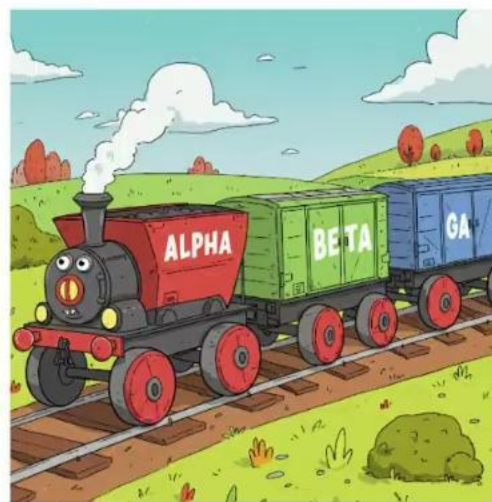
ENDPOINT	ID	VERSION	STORAGE VERSION	DB SIZE	IN USE	PERCENTAGE NOT IN USE	QUOTA	IS LEADER	IS LEARNER	RAFT TERM	RAFT INDEX	RAFT APPLIED INDEX	ERRORS	DOWNGRADE TARGET VERSION	DOWNGRADE ENABLED
http://127.0.0.1:2379	8211fid0f64f3269	3.6.0-alpha.0	3.5.0	98 kB	98 kB	0%	0 B	false	false	2	10	10		3.5.0	true
http://127.0.0.1:22379	91bc3c398fb3c146	3.6.0-alpha.0	3.5.0	98 kB	98 kB	0%	0 B	false	false	2	10	10		3.5.0	true
http://127.0.0.1:32379	fd422379fda50e48	3.6.0-alpha.0	3.5.0	98 kB	98 kB	0%	0 B	true	false	2	10	10		3.5.0	true

- Step 3: Replace image or binary for the members one by one

## etcd v3.6.0 features: feature gate



- Kubernetes style feature gates([KEP-4578](#))
  - Alpha, Beta, GA feature stages
  - new cmd line flag: `--feature-gates=CompactHashCheck=true`
  - Migrated existing `--experimental` feature flags into feature gates
- Kubernetes Enhancement Proposals (KEPs) process for new feature development
- Benefits:
  - No more flag breaking change for graduating a feature from `--experimental` stage
  - Less error prone and less risk of new feature breaking etcd
  - More trackable feature stage and graduation criteria
  - Faster development cycle





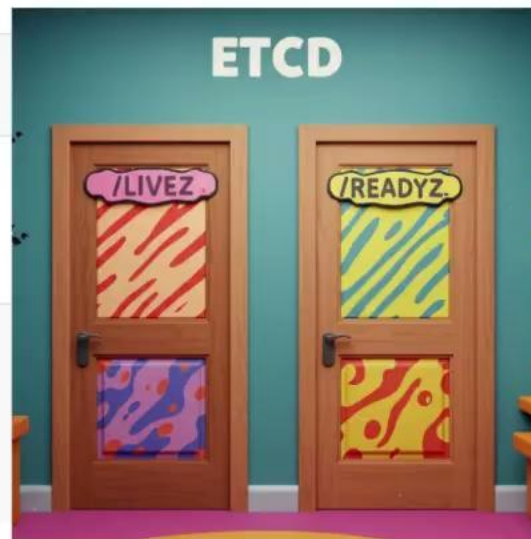
## etcd v3.6.0 features: livez/readyz checks

- new /livez endpoint reflects whether the process is alive or if it needs a restart.
- new /readyz endpoint reflects whether the process is ready to serve traffic.

```
curl -k http://localhost:2379/readyz?verbose
```

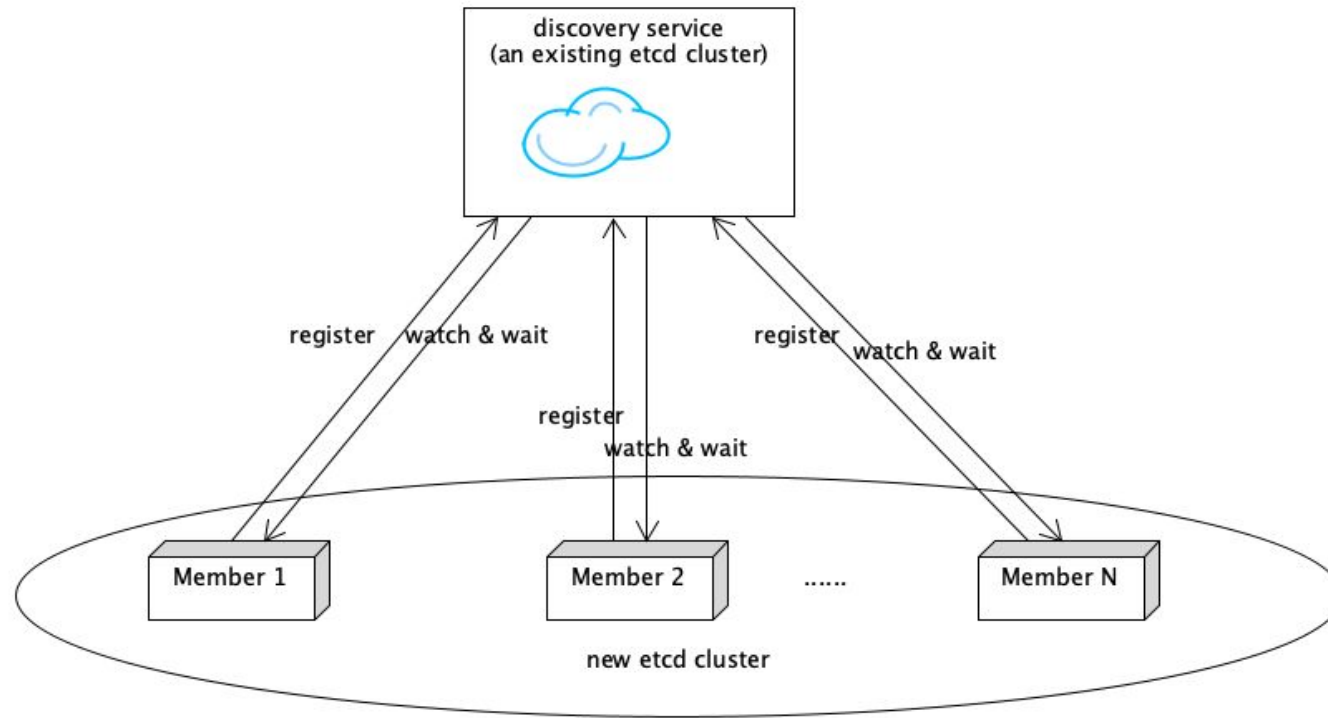
and you would see the response similar to

```
[+]data_corruption ok  
[+]serializable_read ok  
[+]linearizable_read ok  
ok
```



## etcd v3.6.0 features: v3discovery

- Discovery service protocol V3:
  - bootstrap new etcd clusters with the help of the discovery service
  - only used in the first cluster bootstrap
  - clientv3



- v2discovery:
  - clientv2
  - deprecated
  - <https://discovery.etcd.io>: not maintained anymore

etcd v3.6.0 features: upgrade issue

<https://github.com/etcd-io/etcd/issues/19557>

- Symptom
  - Upgrading from 3.5 to 3.6 may fail due to “membership: too many learner in cluster”
- Root cause
  - learner promotion operation isn’t persisted into v3store
- Affected versions
  - All patch versions in v3.5.1 - v3.5.19
- Fixed version
  - v3.5.20
  - v3.6.0-rc.3

Users are required to upgrade to 3.5.20 (or a high patch) first, before upgrading to v3.6.0 !!!

# etcd v3.6.0 performance: tool/measure



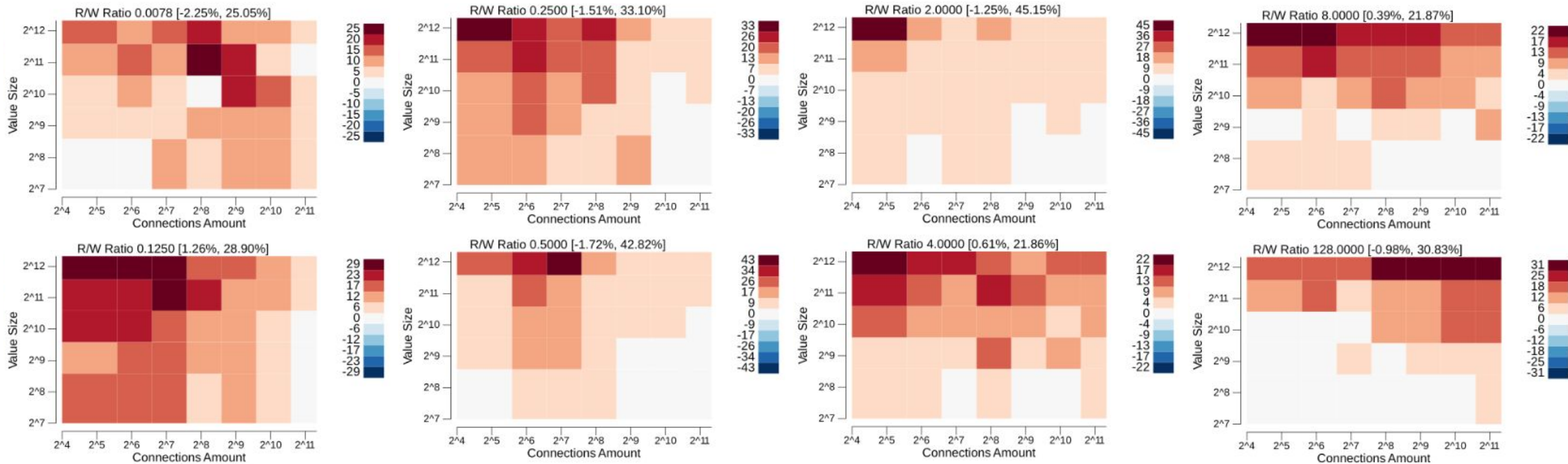
KubeCon



CloudNativeCon

Europe 2025

- Read Write Heatmaps
  - Rewrote plotting code from Python into Go
- We're expanding the plots from heatmaps to power line charts
- Exploration into collecting resource usage (CPU/RAM) to plot the results
- Read performance v3.5.20 vs. v3.6.0-rc.2:





# etcd v3.6.0 performance: QPS improvement



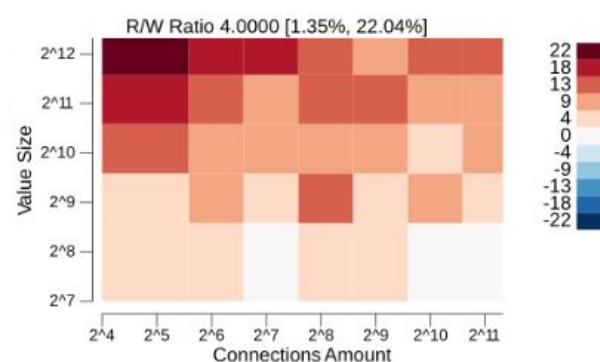
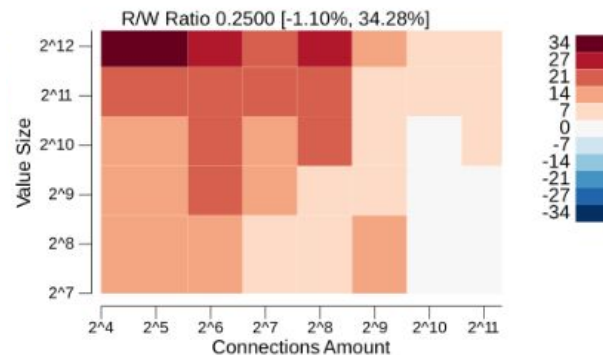
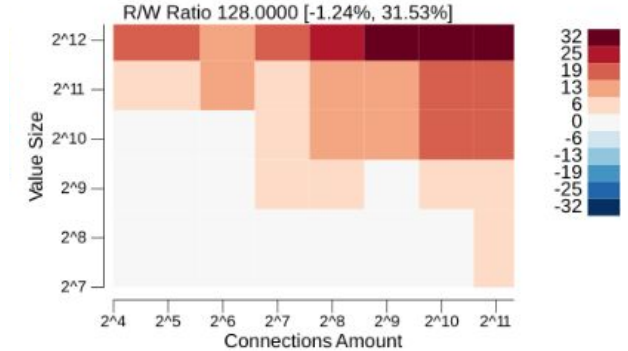
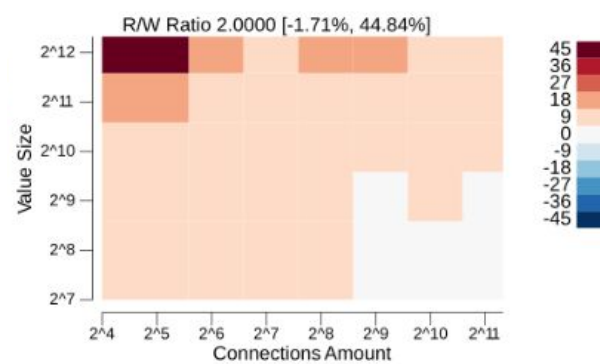
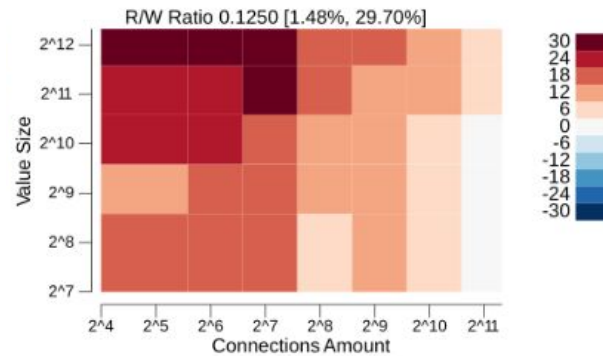
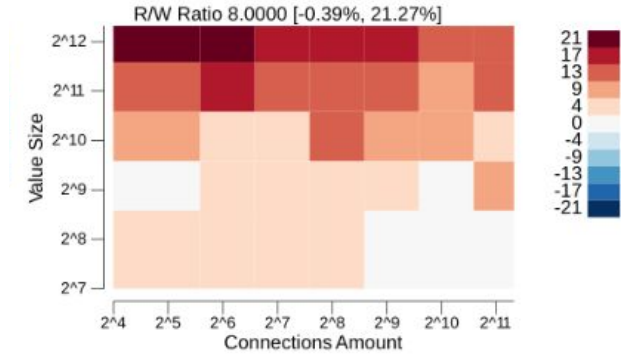
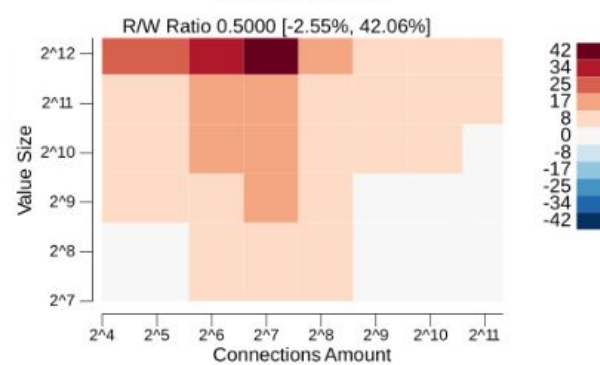
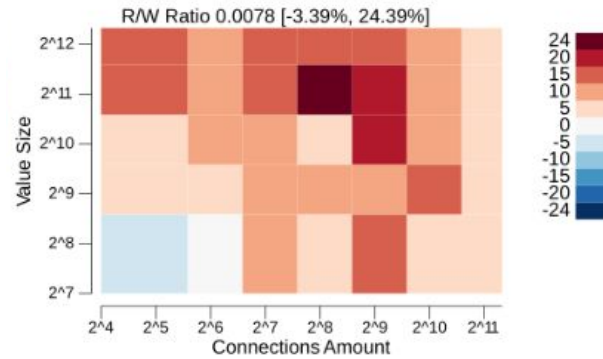
KubeCon



CloudNativeCon

Europe 2025

- Write performance v3.5.20 vs. v3.6.0-rc.2:



# etcd v3.6.0 performance: QPS improvement



KubeCon

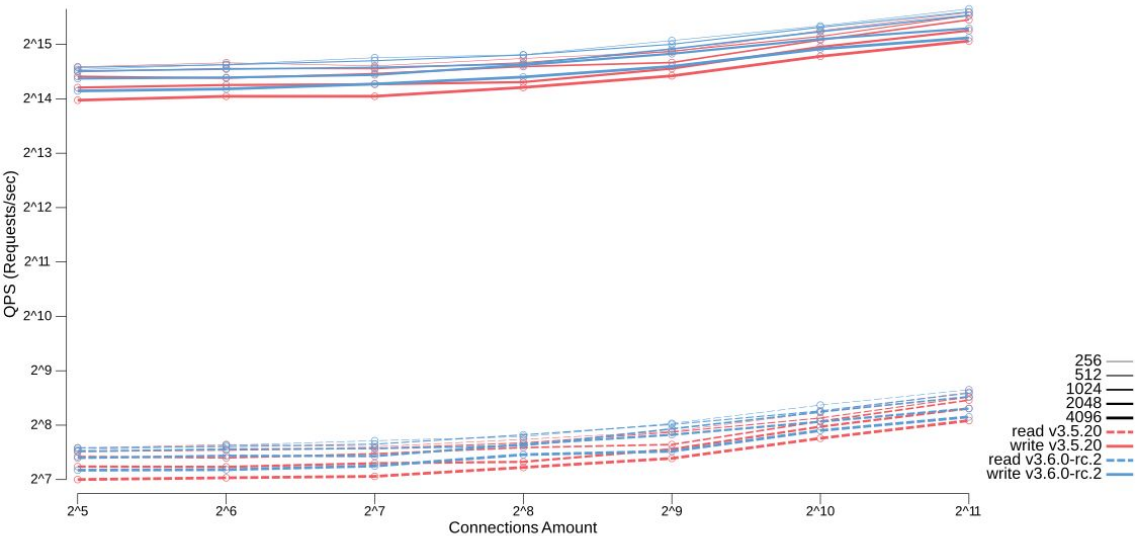


CloudNativeCon

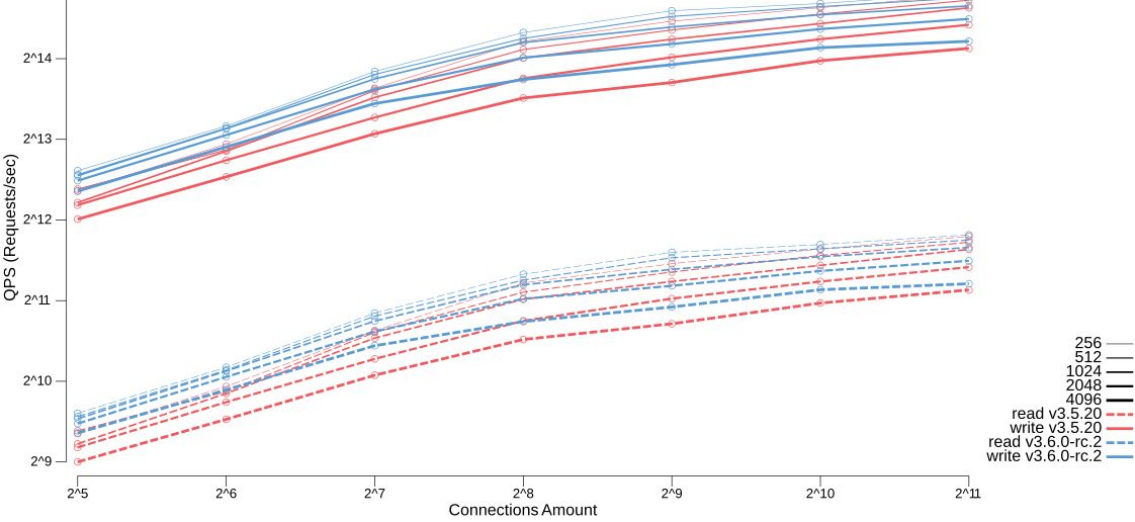
Europe 2025

- Power line charts

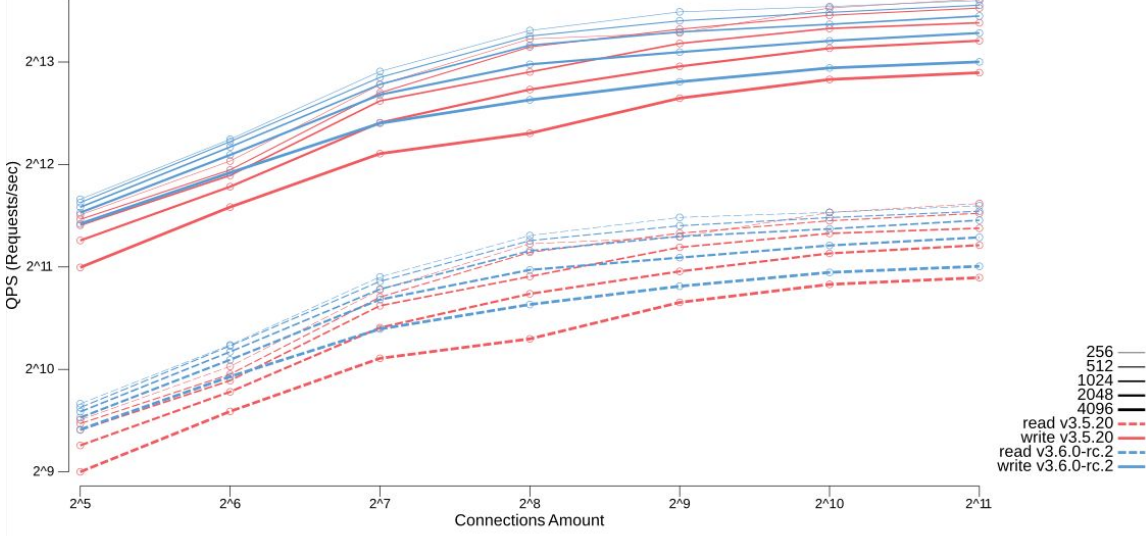
R/W Ratio 0.0078



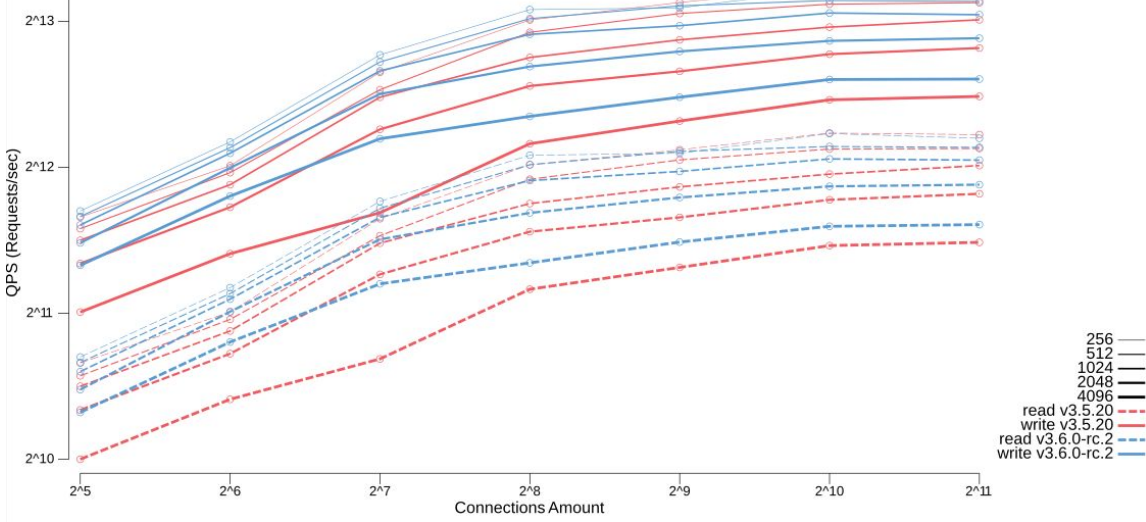
R/W Ratio 0.1250



R/W Ratio 0.2500

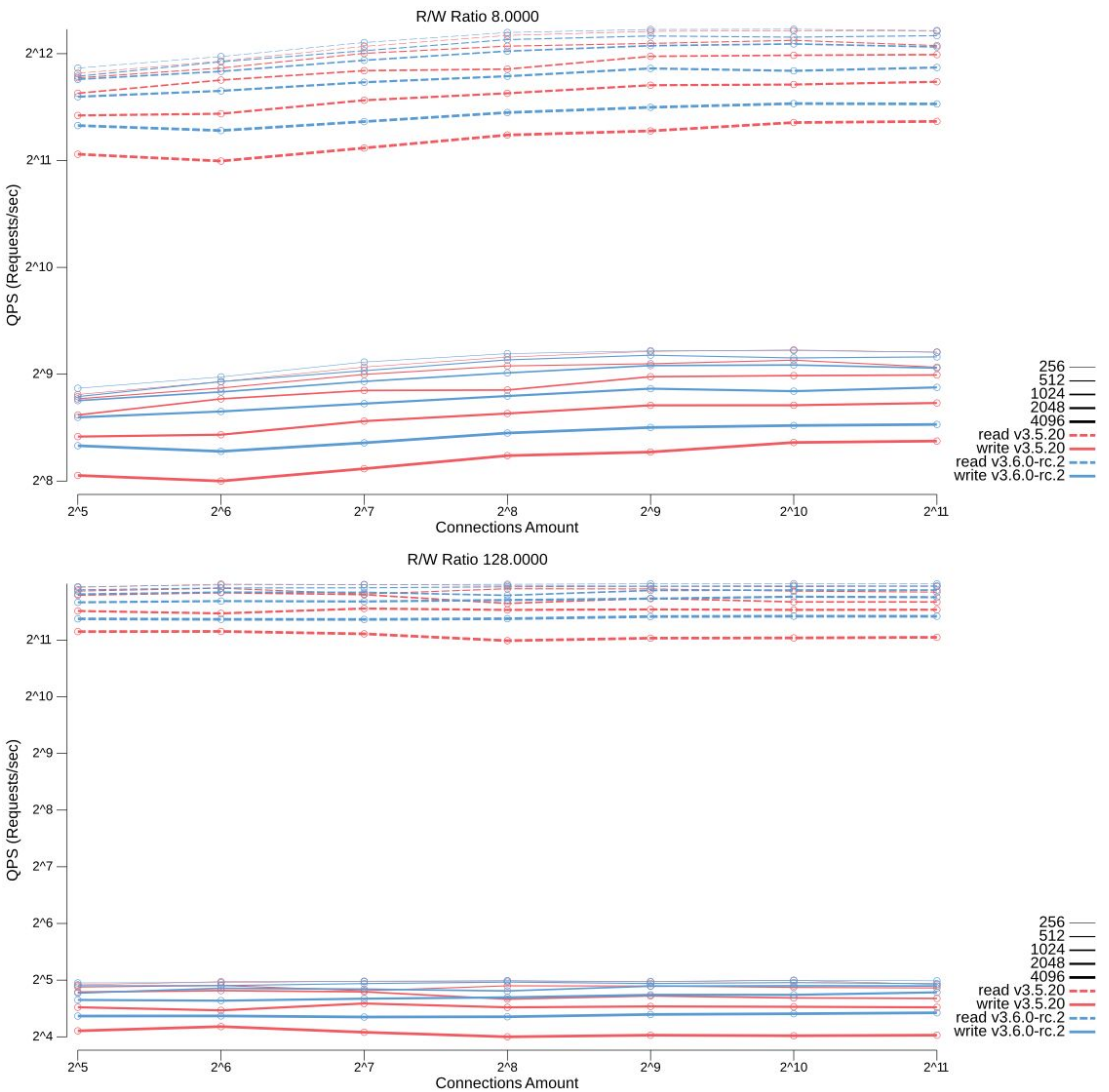
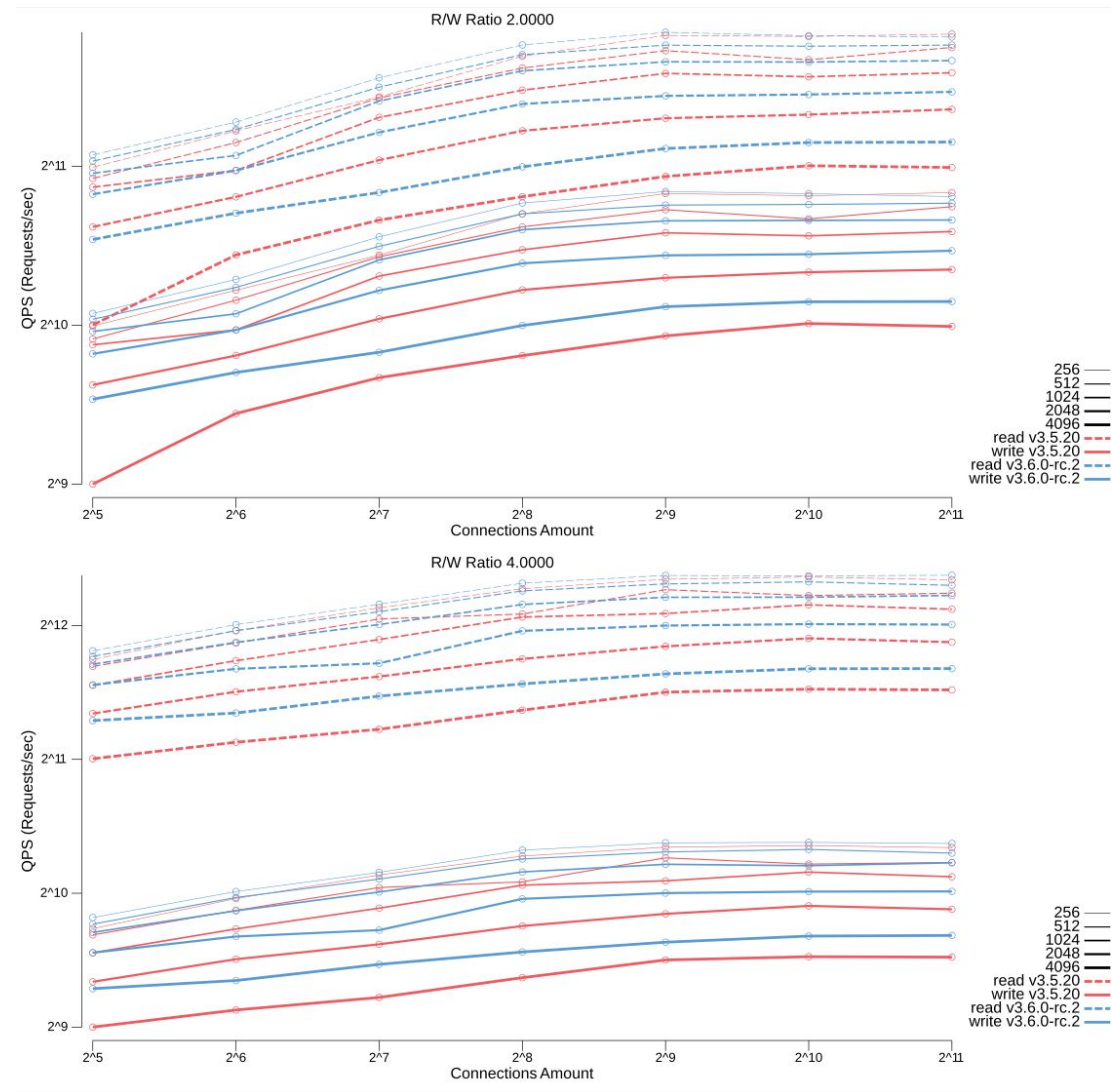


R/W Ratio 0.5000



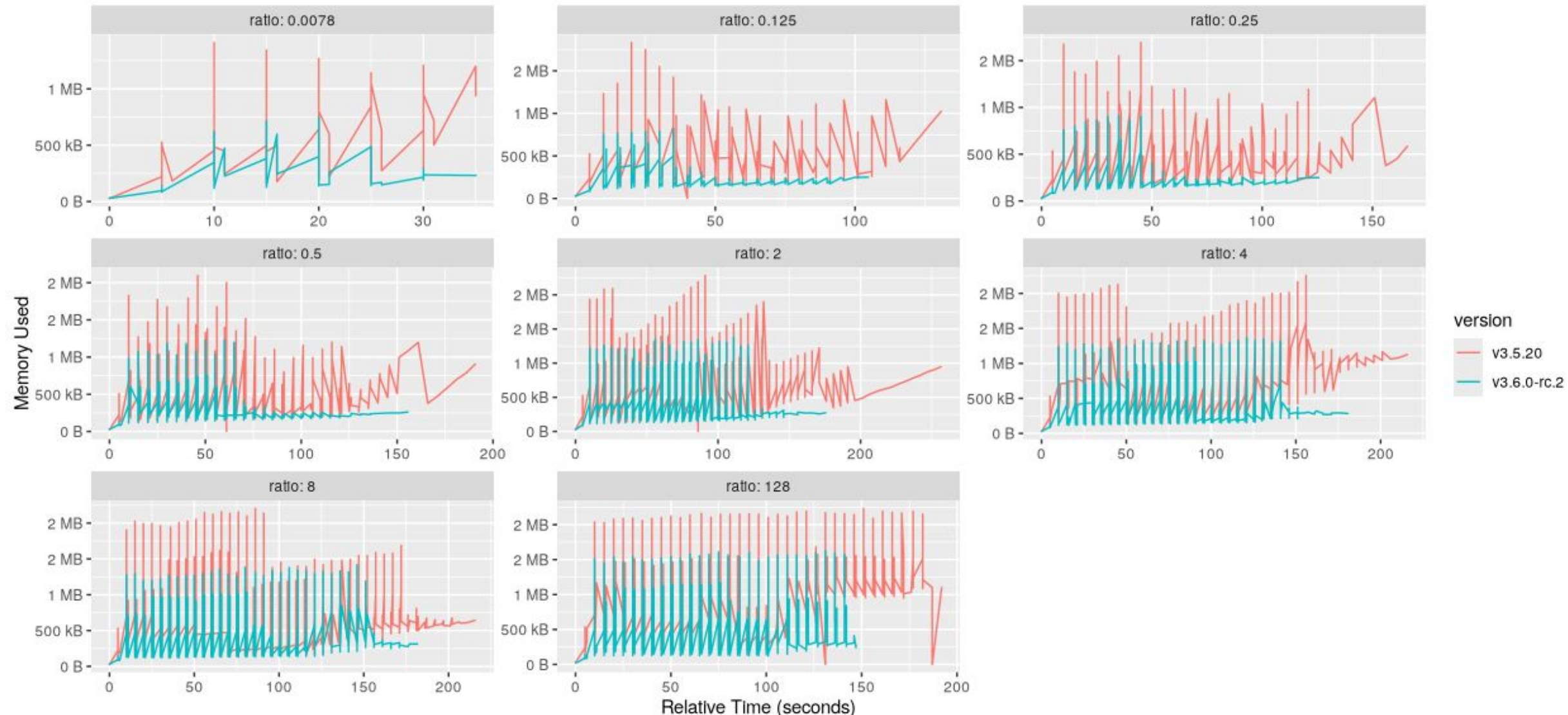
# etcd v3.6.0 performance: QPS improvement

- Power line charts



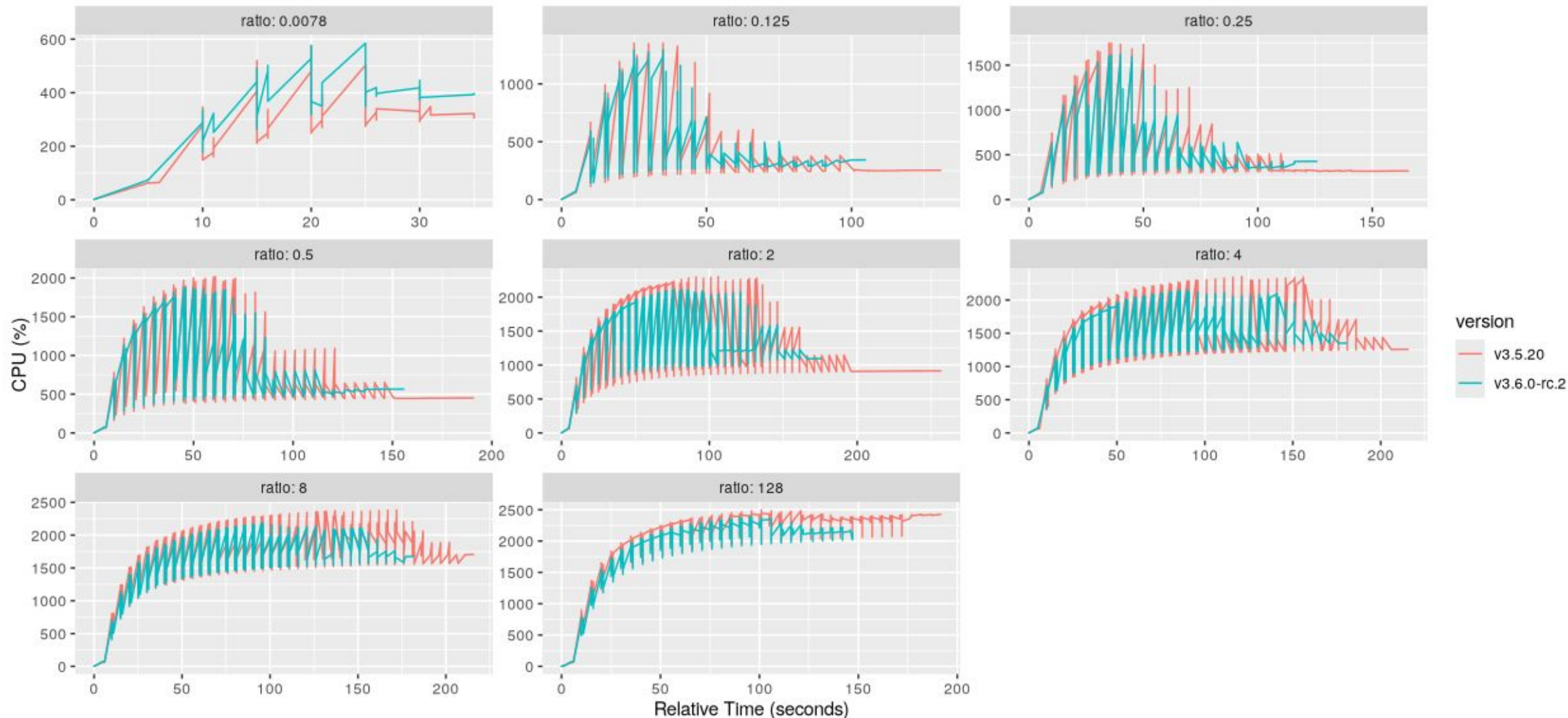


# etcd v3.6.0 performance: RAM improvement



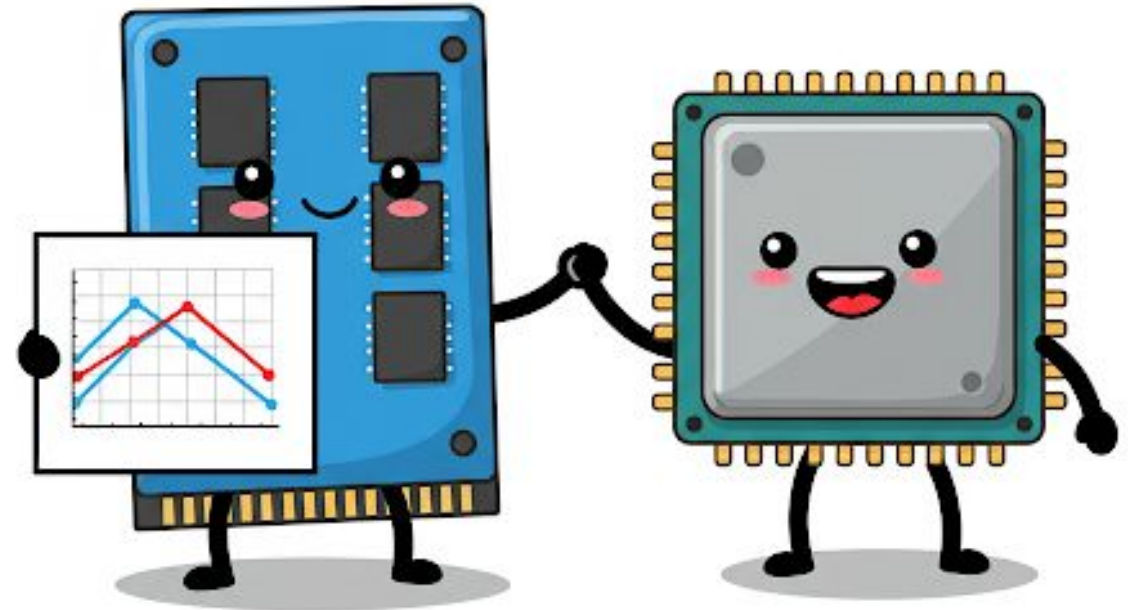


# etcd v3.6.0 performance: CPU improvement



## etcd v3.6.0 performance: RAM/CPU improvement

- v3.5 default `--snapshot-count` value: 100,000
  - v3.6: 10,000
- Reduced the number of Raft snapshots in memory making the compact history more frequent [#18825](#) ([@serathius](#))
- We're still working on memory usage improvements



# Release Process



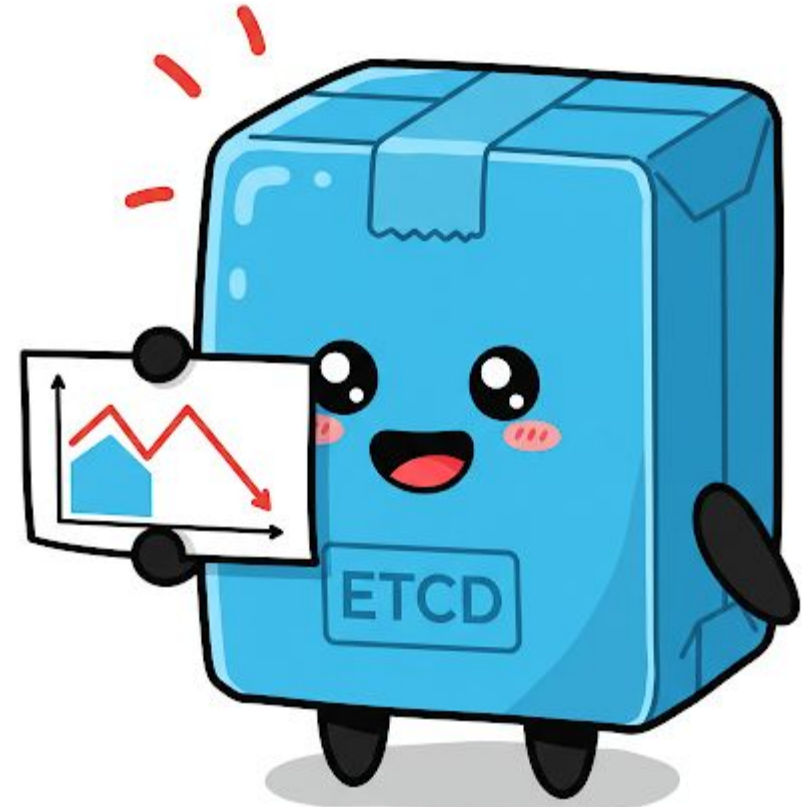
KubeCon



CloudNativeCon

Europe 2025

- We're working on streamlining the process
  - Decreased the release time from more than a day to minutes
  - Made the process "more boring" :)
- Working on bringing more automation to the process
- Working on periodic jobs to test images for vulnerabilities



# WG etcd-operator

- In Scope:
  - Collect requirements & use cases for an etcd operator (running inside Kubernetes)
  - Prioritize the tasks and create a roadmap
  - Bootstrap the **etcd-operator** project, owned by SIG etcd
  - Review existing etcd operators to see if any could be used as starting point
  - Discuss and design the core reconciliation workflow, and provide a PoC
- Out of Scope:
  - Manage etcd clusters running within non-Kubernetes environments
  - Manage etcd clusters which are used as the storage backend of a host kube-apiserver
- Bi-weekly [Zoom meeting](#) on [Tuesdays at 11:00 PT](#)



# etcd-operator v0.1.0



KubeCon



CloudNativeCon

Europe 2025

- **v0.1.0**

- Bootstrap the repo using **Kubebuilder**
- Add automated deps update
- Add basic tests
- Add build, tag and release scripts
- Create a new etcd cluster
- Customize an etcd cluster
- Use CSI drivers for storage

Grateful to all contributors, with special mentions to [@ivanvc](#), [@abdurrehman107](#), [@ArkaSaha30](#) and [@gdasson](#)

```
apiVersion: operator.etcd.io/v1alpha1
kind: EtcdCluster
metadata:
  labels:
    app.kubernetes.io/name: etcd-operator
    app.kubernetes.io/managed-by: kustomize
  name: etcdcluster-sample
spec:
  size: 3
  version: "v3.5.20"
  storageSpec:
    accessModes: "ReadWriteOnce"
    storageClassName: "standard"
    volumeSizeRequest: "8Gi"
```



KubeCon



CloudNativeCon

Europe 2025

## etcd-operator: future work

- **v0.2.0**

- Enable TLS communication through Certificate Management, starting with support for providers:
  - Auto
  - Cert-Manager

spec:

size: 3

version: "v3.5.19"

tls:

provider: "cert-manager"

providerCfg:

certManagerCfg:

commonName: "etcd.etcd-operator-system"

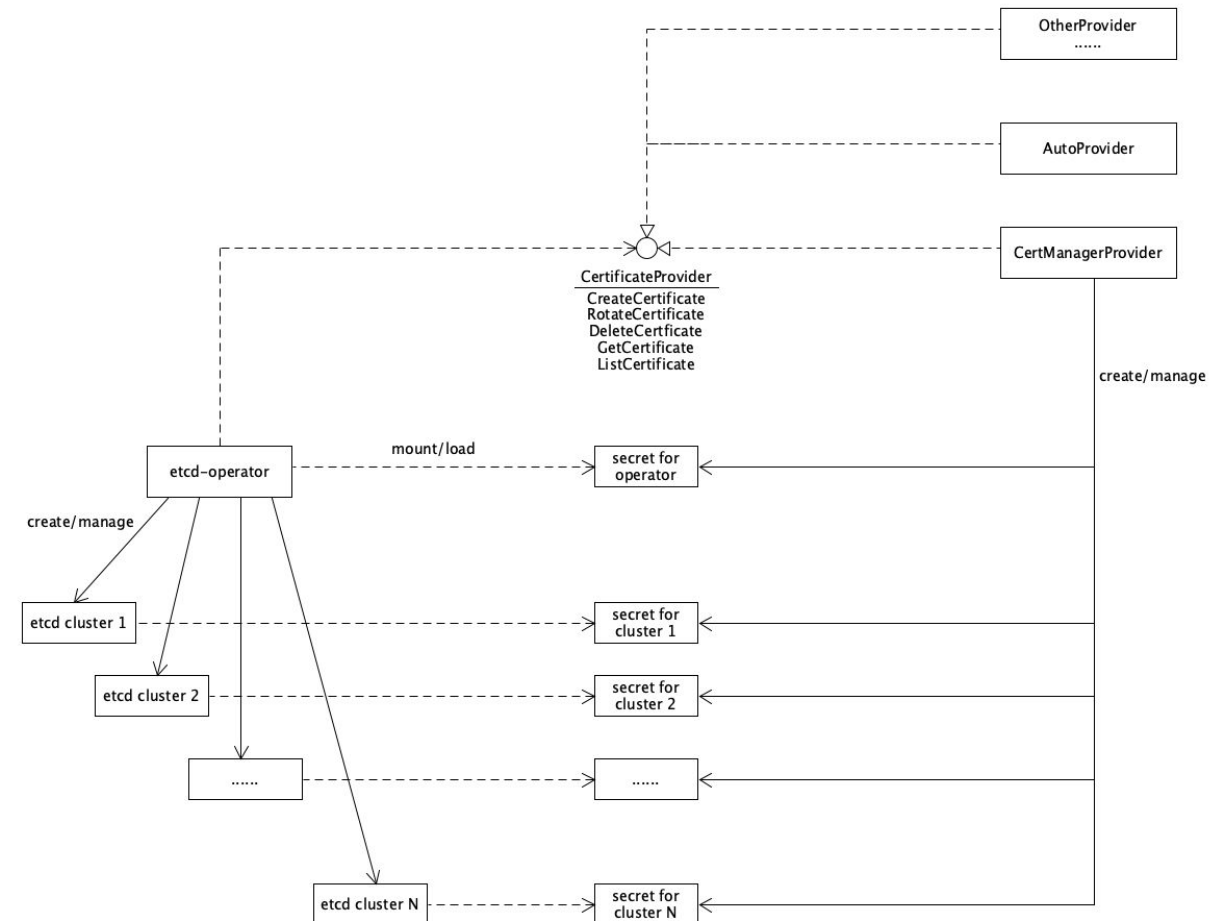
organizations:

– "etcd-operator"

duration: 90d

issuerName: "selfsigned"

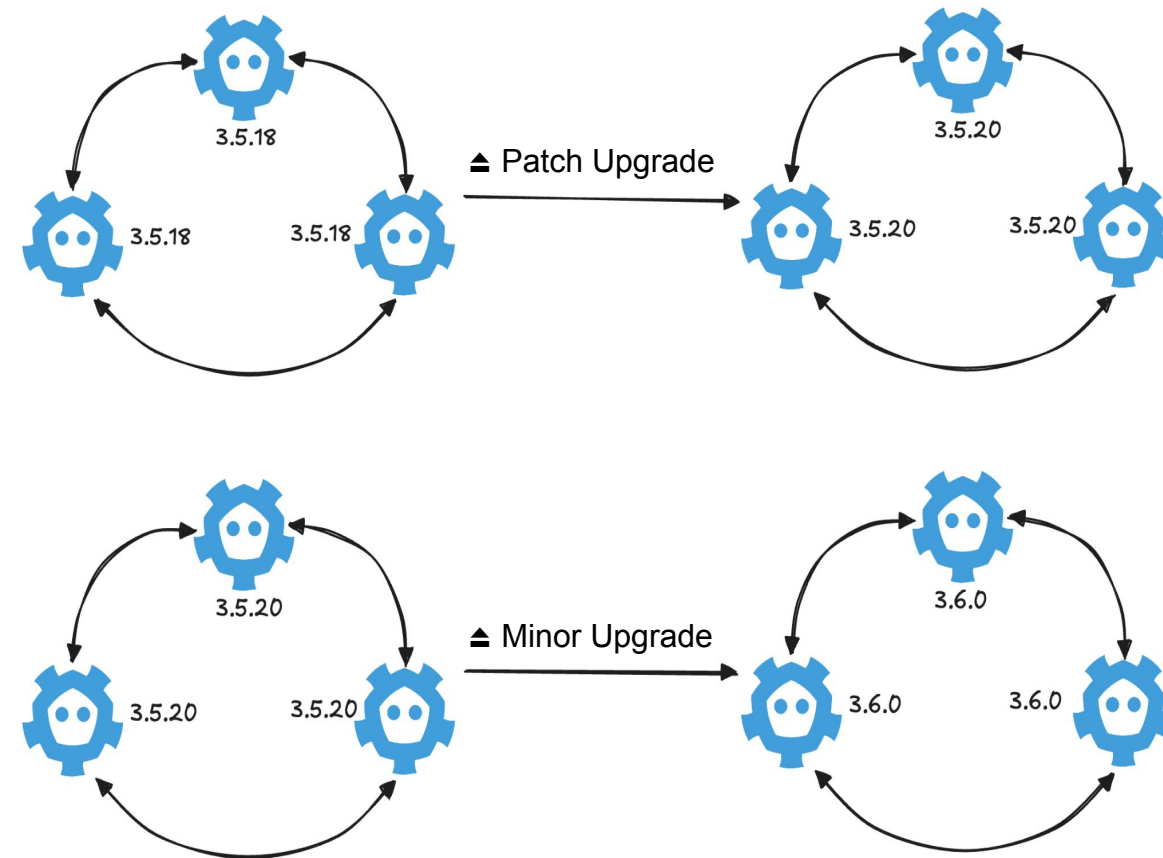
issuerKind: "ClusterIssuer"



## etcd-operator: future work

- **v0.2.0**

- Enable TLS communication through Certificate Management, starting with support for providers:
  - Auto
  - Cert-Manager
- Upgrades
  - Patch upgrades  
(M.m.p  $\rightarrow$  M.m.p+n, n=1,2,3...)
  - Minor upgrades  
(M.m  $\rightarrow$  M.m+1)



KubeCon



CloudNativeCon

Europe 2025

## etcd-operator: future work

- **v0.2.0**

- Enable TLS communication through Certificate Management, starting with support for providers:
  - Auto
  - Cert-Manager
- Upgrades
  - Patch upgrades  
(M.m.p  $\rightarrow$  M.m.p+n, n=1,2,3...)
  - Minor upgrades  
(M.m  $\rightarrow$  M.m+1)
- Customizations
  - Enable configuration options via flags/ENV variables to customize etcd
- Introduce Helm charts

- **v0.3.0 and beyond**

- Recovery of cluster members
- Backup cluster creation
  - On-demand
  - Periodic
- New cluster from the available backup



Q & A

Thank you!