white duck

Kubernetes 1.23 – What’s new?

Cloud Native Rosenheim Meetup, Februar 2022

CLOUD NATIVE ROSENHEIM MEET UP

Microsoft Partner

Microsoft

Gold Cloud Platform
Gold DevOps
Silver Application Development
Silver Security
Silver Application Integration

GitHub

Who we are?

Philip Welz (Senior Kubernetes & DevOps Engineer, GitLab Hero, CKA, CKAD & CKS)

Email:	philip.welz@whiteduck.de
Twitter:	@philip_welz
LinkedIn:	https://www.linkedin.com/in/philip-welz

Nico Meisenzahl (Senior Cloud & DevOps Consultant, MVP, GitLab Hero)

Email:	nico.meisenzahl@whiteduck.de
Twitter:	@nmeisenzahl
LinkedIn:	https://www.linkedin.com/in/nicomeisenzahl

twitter.com/CloudNative_Ro

twitter.com/whiteduck_gmbh

twitter.com/AzureMeetup

Housekeeping

- this meetup will be streamed on YouTube!
- want to participate?
  - join our meetup to get access to the Zoom meeting
    - https://www.meetup.com/CloudNative-Rosenheim-Meetup
  - we do also monitor the comments on YouTube

Agenda

- Kubernetes 1.23 release overview
- new features – our picks
- deprecations
- further topics

white duck

Kubernetes 1.23

- “The Next Frontier”
  - the release logo continues with the theme's Star Trek reference
  - the ship represents the collective teamwork of the release team
  - every star is a Kubernetes logo
- has been released on December 7, 2021
  - third release in 2021
  - second longer-cycle release after the change from four to three yearly releases

kubernetes 1.23
The Next Frontier

Updates & changes

- Kubernetes 1.23 introduces 47 enhancements
  - 11 have graduated to stable
  - 17 are moving to beta
  - 17 are entering alpha
  - 1 features has been deprecated
- Kubernetes now complies with level 1 of the SLSA
  - “Supply-chain Levels for Software Artifacts”
  - the build process must be fully scripted/automated and provide evidence
  - https://slsa.dev/spec/v0.1/levels

NEW FEATURES – OUR PICK

Kubernetes feature states

- alpha
  - aren’t enabled by default
  - opt-in via feature gate flag on Kubernetes components
  - managed Kubernetes: vendor may decide what feature gates are enabled
- beta
  - are enabled by default
  - opt-out via feature gate flag available
- stable (general availability)
  - commitment that they are staying in place throughout the current major version

#1 IPv4/IPv6 Dual-stack Networking

- graduates to stable
- with this you can natively run your cluster in dual-stack mode (IPv4/IPv6)
  - Container Network Interface (CNI) network plugin needs to support this
    - e.g., Kubenet, Calico (https://projectcalico.docs.tigera.io/networking/ipv6)
  - nodes must have routable IPv4/IPv6 network interfaces
  - Services continue to default to single-stack
- more details on the how and why
  - https://www.infoq.com/news/2021/12/dual-stack-kubernetes
  - https://kubernetes.io/docs/concepts/services-networking/dual-stack

#2 Ephemeral Containers

- graduates to beta
- allows you to add ephemeral containers to your pods using “kubectl debug”
- great way to debug running Pods
  - debug with all your favourite tools and dependencies
  - allows to share process namespaces
- for more details and a demo, see the recording of our Kubernetes 1.22 meetup
  - https://www.youtube.com/watch?v=YmGiIRj9tdM

#3 Structured logging

- graduates to beta
- alpha was introduced with 1.19
- structured logs natively support (key, value) pairs and object references
- logs can also be outputted in JSON format
- more Details
  - https://kubernetes.io/blog/2020/09/04/kubernetes-1-19-introducing-structured-logs/

#4 Generic Ephemeral Volume feature

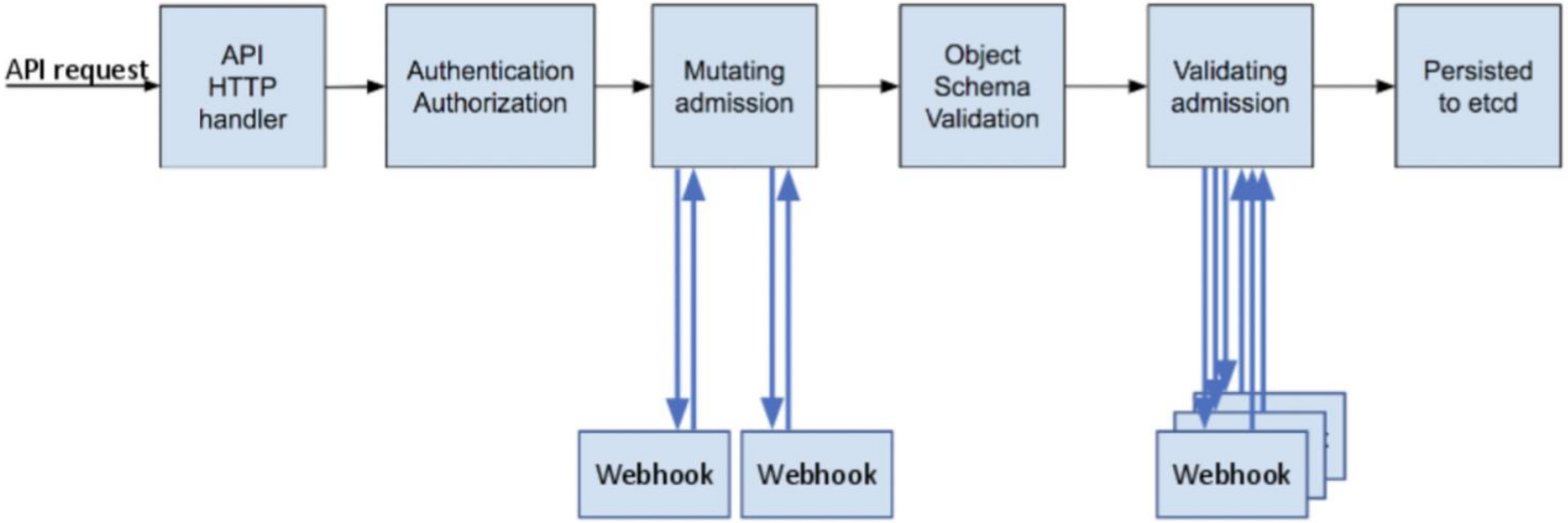
- graduates to stable
- defines inline ephemeral volumes that will work with any storage driver (CSI) that supports dynamic provisioning
  - storage can be local or network-attached
  - volumes can have a fixed size that Pods are not able to exceed
  - Volumes may have some initial data, depending on the driver and parameters
  - snapshotting, cloning, resizing supported (assuming that the driver supports them)

#5 PodSecurity admission controller

- graduates to beta
- offers a built-in Pod Security admission controller as a successor to PodSecurityPolicies
  - three different policies to broadly cover the security spectrum: Privileged, Baseline & Restricted
  - the enforcement can be done at three levels: enforce, audit & warn

#5 PodSecurity admission controller

- for more details, see the recording of our Kubernetes 1.22 meetup
  - https://www.youtube.com/watch?v=YmGiIRj9tdM



Admission Controller Phases

#6 Skip Volume Ownership

- graduates to stable
- allows to configure volume permission and ownership change policy
  - allows to speeds up the pod start up time on very large volumes
- can be defined via *pod.Spec.SecurityContext.FSGroupChangePolicy*

#7 TTL after finish

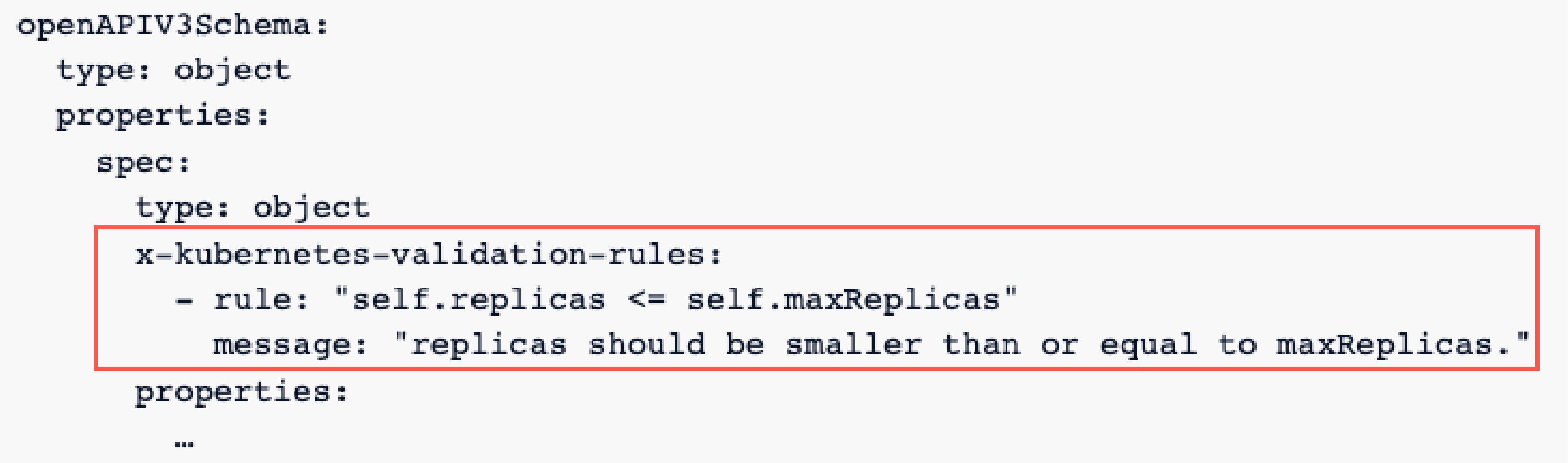
- graduates to stable
- TTL controller clean up finished or complete jobs automatically
  - gets specified in .spec.ttlSecondsAfterFinished
    - is this field set to 0, the job will be immediately deleted
- helps to remove load from the API server as sometimes those lingering pods may cause cluster performance degradation

#8 Expression language validation for CRD

- new feature (alpha)
  - needs to be enabled
    via “CustomResourceValidationExpressions” feature gate
- custom resources will be validated by validation rules using the Common Expression Language (CEL)
  - allows to build self-contained CRDs (everything is defined in the CRD)
  - can simplify deployments by not needing webhooks

#8 Expression language validation for CRD

- more details
  - https://opensource.google/projects/cel



```
openAPIV3Schema:
  type: object
  properties:
    spec:
      type: object
      x-kubernetes-validation-rules:
        - rule: "self.replicas <= self.maxReplicas"
          message: "replicas should be smaller than or equal to maxReplicas."
      properties:
        …
```

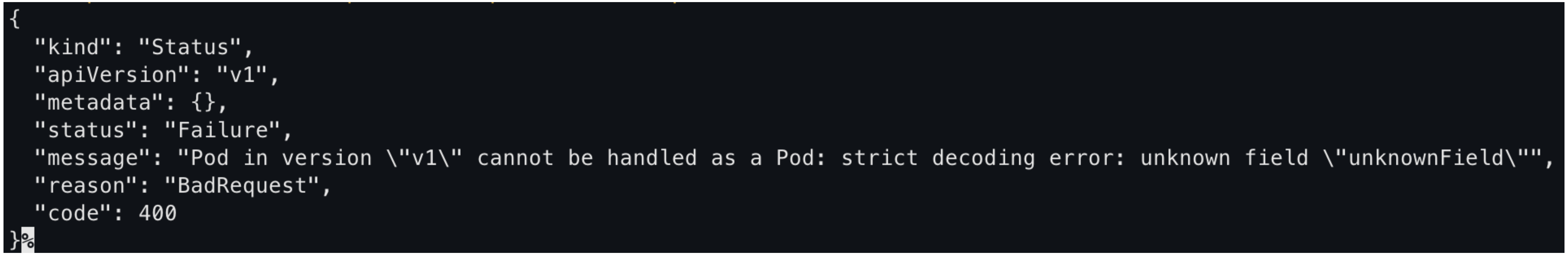
#9 Support for Windows privileged containers

- graduates to beta
- allows to run privileged container (like with Linux)
  - not required for “normal” workload
  - but opens flexibility for third-party and integrations (security, monitoring, …)

#10 Server Side Field Validation

- new feature (alpha)
  - needs to be enabled via “ServerSideFieldValidation” feature gate
- users will receive warnings from the server when they send Kubernetes objects that contain unknown or duplicate fields
  - server-side version of “kubectl--validate=true”
  - no client-side implementation required
- the query parameter “fieldValidation” can be specify to ignore, warn or deny the request

#10 Demo: Server Side Field Validation



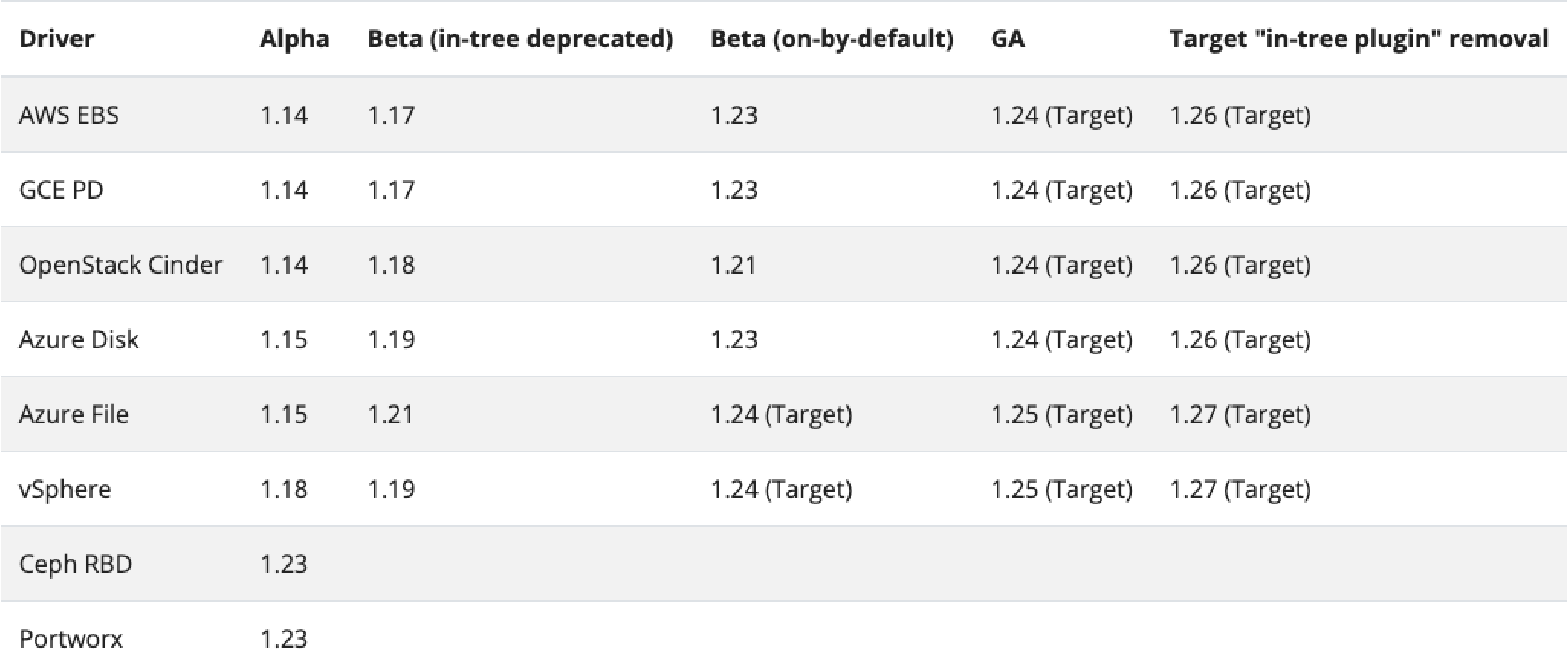
```
{
  "kind": "Status",
  "apiVersion": "v1",
  "metadata": {},
  "status": "Failure",
  "message": "Pod in version \"v1\" cannot be handled as a Pod: strict decoding error: unknown field \"unknownField\"",
  "reason": "BadRequest",
  "code": 400
}%
```

#11 CronJobs

- finally graduates to stable
  - was introduced in Kubernetes 1.4
  - in beta since Kubernetes 1.8
  - v2 implementation is default since Kubernetes 1.21
- just in case: a CronJob creates Jobs on a repeating schedule

#12 CSI Migration updates

- enables the replacement of existing in-tree storage plugins with Container Storage Interface (CSI)

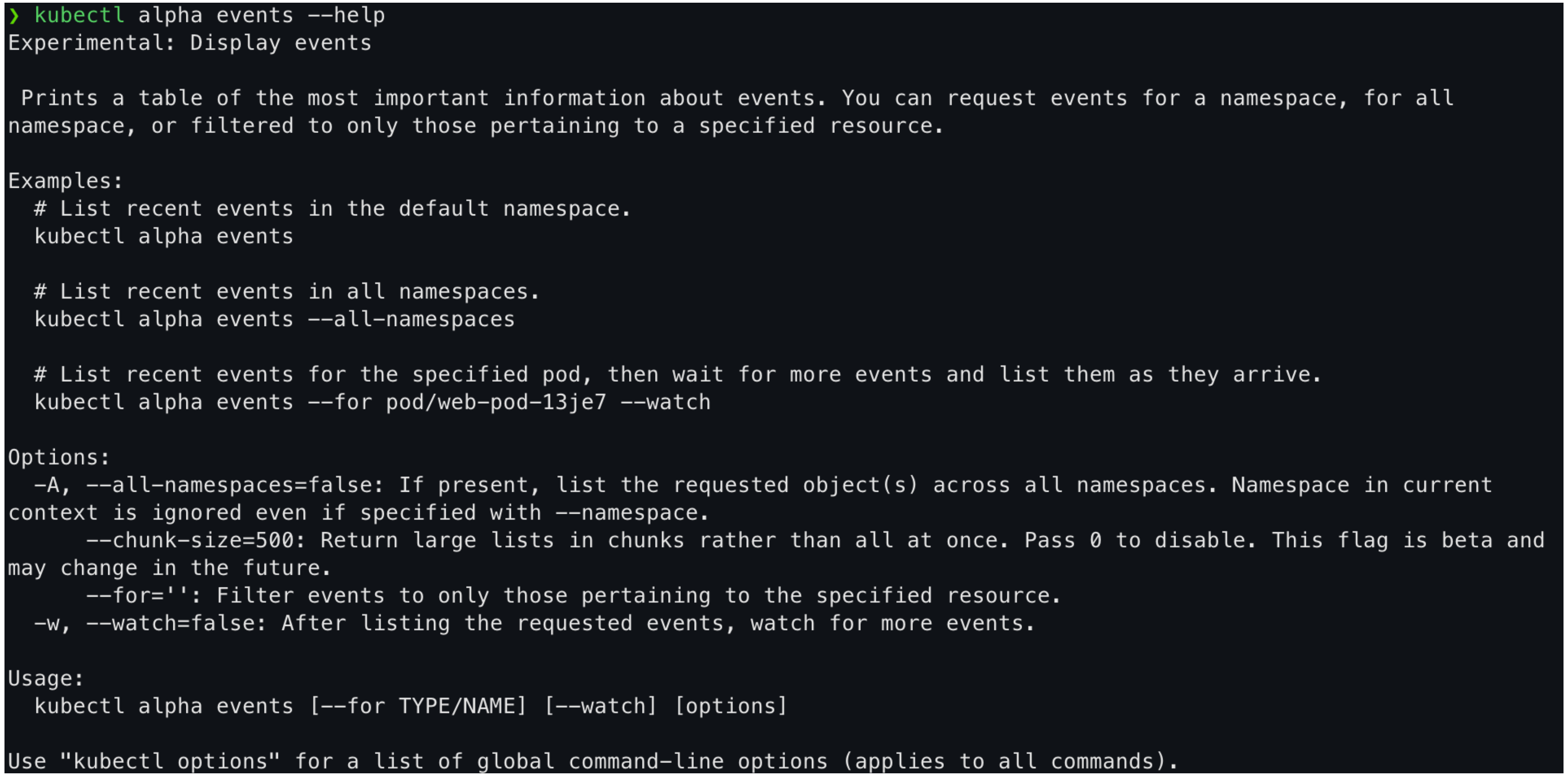


Driver	Alpha	Beta (in-tree deprecated)	Beta (on-by-default)	GA	Target "in-tree plugin" removal
AWS EBS	1.14	1.17	1.23	1.24 (Target)	1.26 (Target)
GCE PD	1.14	1.17	1.23	1.24 (Target)	1.26 (Target)
OpenStack Cinder	1.14	1.18	1.21	1.24 (Target)	1.26 (Target)
Azure Disk	1.15	1.19	1.23	1.24 (Target)	1.26 (Target)
Azure File	1.15	1.21	1.24 (Target)	1.25 (Target)	1.27 (Target)
vSphere	1.18	1.19	1.24 (Target)	1.25 (Target)	1.27 (Target)
Ceph RBD	1.23				
Portworx	1.23				

#13 “kubectl events”

- still in alpha
  - “kubectl alpha events”
- reimplementation of “kubectl get events”
  - addresses long standing issues (filtering, watching, sorting,…)
  - the events sub-command will help improve user experience
- more details
  - https://github.com/kubernetes/enhancements/tree/master/keps/sig-cli/1440-kubectl-events

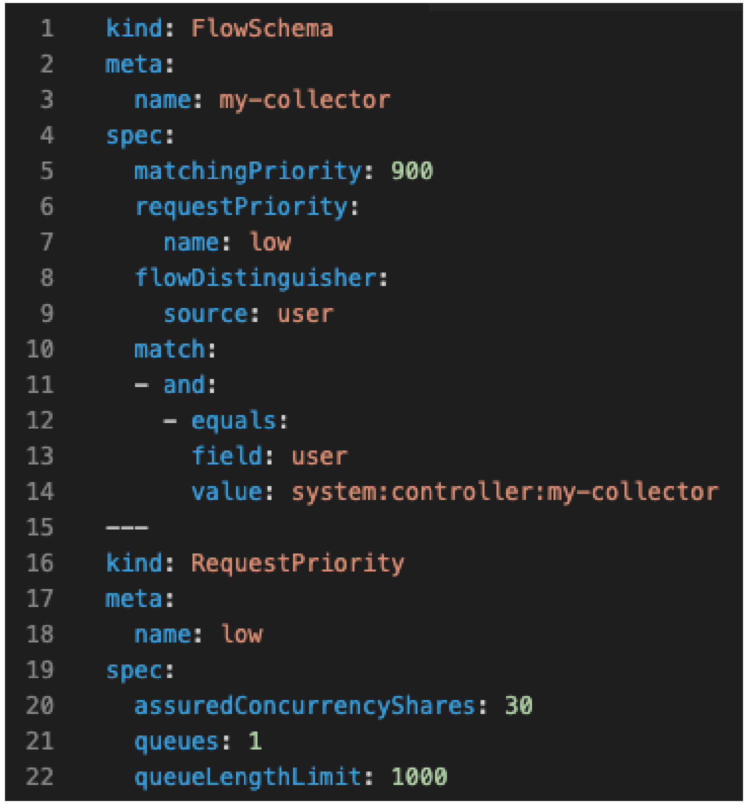
#13 Demo: “kubectl events”



```
❯ kubectl alpha events --help
Experimental: Display events

 Prints a table of the most important information about events. You can request events for a namespace, for all
namespace, or filtered to only those pertaining to a specified resource.

Examples:
  # List recent events in the default namespace.
  kubectl alpha events

  # List recent events in all namespaces.
  kubectl alpha events --all-namespaces

  # List recent events for the specified pod, then wait for more events and list them as they arrive.
  kubectl alpha events --for pod/web-pod-13je7 --watch

Options:
  -A, --all-namespaces=false: If present, list the requested object(s) across all namespaces. Namespace in current
context is ignored even if specified with --namespace.
      --chunk-size=500: Return large lists in chunks rather than all at once. Pass 0 to disable. This flag is beta and
may change in the future.
      --for='': Filter events to only those pertaining to the specified resource.
  -w, --watch=false: After listing the requested events, watch for more events.

Usage:
  kubectl alpha events [--for TYPE/NAME] [--watch] [options]

Use "kubectl options" for a list of global command-line options (applies to all commands).
```

#14 Priority and Fairness for API Server Requests

- graduates to beta
- allows granular option to prioritise API requests during high load
- major API changes (since alpha)



```yaml
kind: FlowSchema
meta:
  name: my-collector
spec:
  matchingPriority: 900
  requestPriority:
    name: low
  flowDistinguisher:
    source: user
  match:
  - and:
    - equals:
      field: user
      value: system:controller:my-collector
---
kind: RequestPriority
meta:
  name: low
spec:
  assuredConcurrencyShares: 30
  queues: 1
  queueLengthLimit: 1000
```

#15 Auto remove PVCs created by StatefulSet

- new feature (alpha)
  - needs to be enabled via “StatefulSetAutoDeletePVC” feature gate
- controls the lifetime of PVCs generated from the StatefulSet spec template
  - allows to automatically delete a PVC is a StatefulSet is deleted or scaled down
  - the retention policy is defined within the StatefulSet spec
- more details
  - https://kubernetes.io/blog/2021/12/16/kubernetes-1-23-statefulset-pvc-auto-deletion

#15 Demo: Auto remove PVCs created by StatefulSet

#16 OpenAPI v3 support

- new feature (alpha)
  - needs to be enabled via “OpenApiv3” feature gate
- Kubernetes API server support for OpenAPI v3
- OpenAPI v3 allows more complex definitions that are especially helpful for CRDs
  - before CRDs could be defined in v3, they would have been exported with v2, resulting in missing information
- exposed on /openapi/v3/apis/{group}/{version}
  - new spec with improved performance and discovery

#17 Recovering from PVC resize failures

- new feature (alpha)
  - needs to be enabled via "RecoverVolumeExpansionFailure" feature gate
- allows to reverse a PVC resizing if failed
  - helps to easily recover from common errors related to an unsupported PVC size
- reduction is only possible
  - if the new value is higher than capacity used
  - the previous expansion failed

#18 gRPC probe support

- new feature (alpha)
  - needs to be enabled via “GRPCContainerProbe” feature gate
- implementation of gRPC Health Checking Protocol needed
  - https://github.com/grpc/grpc/blob/master/doc/health-checking.md
  - no need to implement the grpc_health_probe wrapper CLI
- use gRPC probes as you know them from HTTP/TCP
  - readiness, liveness and startup probes
  - no support for named ports
  - no TLS support

#18 Demo: gRPC probe support

- Demo app
  - https://github.com/nmeisenzahl/grpc-health

DEPRECATIONS

#1 Deprecation of FlexVolume

- Maintainers of FlexVolume drivers should implement a CSI driver verion
- Users should migrate from FlexVolume to CSI
- Container Storage Interface (CSI) is the successor of FlexVolume
  - more flexible out-of-tree implementation
  - recommended way to write volume drivers in Kubernetes
- More details
  - https://github.com/kubernetes/community/blob/master/sig-storage/volume-plugin-faq.md#kubernetes-volume-plugin-faq-for-storage-vendors

#2 HorizontalPodAutoscaler v2

- HorizontalPodAutoscaler v2 graduates to GA (stable)
  - autoscaling/v2
- deprecations
  - autoscaling/v2beta2
- notable changes
  - “Resources” renamed to “PodResource”
  - “Disabled” renamed to “ScalingDisabled”
  - “Min/Max” renamed to “Min/MaxChange”

#3 kubectl --dry-run

- “kubectl --dry-run” without a value is deprecated now
  - client, server or none are allowed values
  - one of them must be defined
- “kubectl --dry-run=none”
  - disables dry-run
- “kubectl --dry-run=client”
  - only print the object that would be sent, without sending it
- “kubectl --dry-run=server”
  - submit server-side request without persisting the resource
  - default fields, validation, admission chain (validating, mutating)

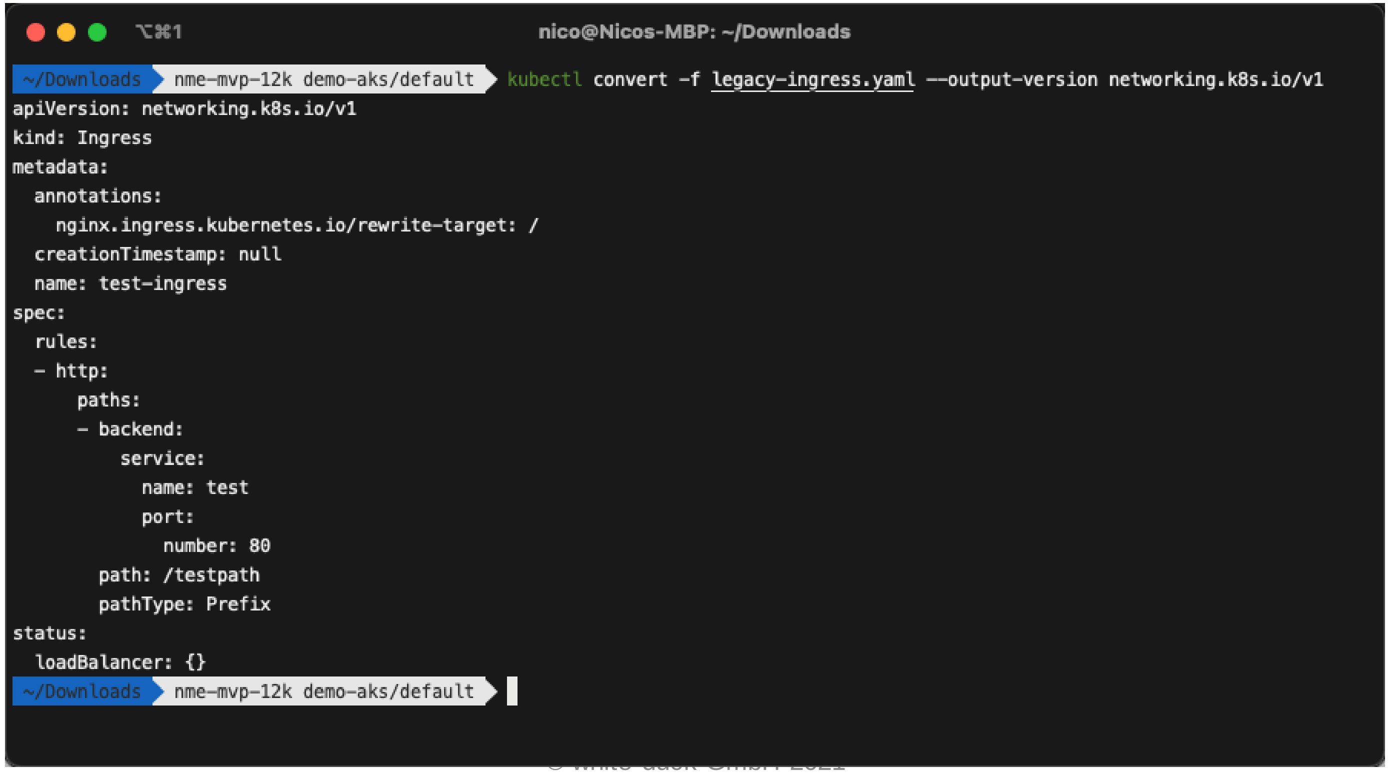
FURTHER TOPICS

CRD handling issue

- Kubernetes drops unknown fields from list items in CRD
  - if “x-kubernetes-preserve-unknown-fields: true” is set on the list and not on the individual items
- impacts Kubernetes 1.23.0 – 1.23.2
  - fixed with 1.23.3 (released in January 25)
- make sure to review your third-party deployments prior upgrade
- more details
  - https://github.com/kubernetes/kubernetes/issues/107690

kubectl convert

- a plugin for kubectl
- helps you to update your manifests to a newer API version



```
~/Downloads  nme-mvp-12k demo-aks/default  kubectl convert -f legacy-ingress.yaml --output-version networking.k8s.io/v1
apiVersion: networking.k8s.io/v1
kind: Ingress
metadata:
  annotations:
    nginx.ingress.kubernetes.io/rewrite-target: /
  creationTimestamp: null
  name: test-ingress
spec:
  rules:
  - http:
      paths:
      - backend:
          service:
            name: test
            port:
              number: 80
        path: /testpath
        pathType: Prefix
status:
  loadBalancer: {}
~/Downloads  nme-mvp-12k demo-aks/default 
```

Further links

- https://github.com/whiteducksoftware/cloud-native-rosenheim-meetup
- https://kubernetes.io/blog/2021/12/07/kubernetes-1-23-release-announcement
  - https://kubernetes.io/blog/2021/12/08/dual-stack-networking-ga
  - https://kubernetes.io/blog/2021/12/09/pod-security-admission-beta
  - https://kubernetes.io/blog/2021/12/10/storage-in-tree-to-csi-migration-status-update
  - https://kubernetes.io/blog/2021/12/15/kubernetes-1-23-prevent-persistentvolume-leaks-when-deleting-out-of-order
  - https://kubernetes.io/blog/2021/12/16/kubernetes-1-23-statefulset-pvc-auto-deletion
- https://sysdig.com/blog/kubernetes-1-23-whats-new

Questions?

Philip Welz (Senior Kubernetes & DevOps Engineer, GitLab Hero, CKA, CKAD & CKS)

Email:	philip.welz@whiteduck.de
Twitter:	@philip_welz
LinkedIn:	https://www.linkedin.com/in/philip-welz

Nico Meisenzahl (Senior Cloud & DevOps Consultant, MVP, GitLab Hero)

Email:	nico.meisenzahl@whiteduck.de
Twitter:	@nmeisenzahl
LinkedIn:	https://www.linkedin.com/in/nicomeisenzahl

twitter.com/CloudNative_Ro

twitter.com/whiteduck_gmbh

twitter.com/AzureMeetup