KubeCon | CloudNativeCon

Europe 2018

SIG API Machinery Deep Dive

Stefan Schimanski – sttts@redhat.com – @the_sttts

Agenda


KubeCon | CloudNativeCon Europe 2018

- **Outlook** to Kubernetes 1.11+
- **Deep Dive** into CustomResourceDefinitions
- **Questions**

Outlook – Custom Resources

- Kubernetes 1.11+
  - α: **Multiple versions without conversion** – design proposal
  - α: **Pruning** – in validation spec unspecified fields are removed – blocker for GA
  - α: **Defaulting** – defaults from OpenAPI validation schema are applied
  - α: **Graceful Deletion** – maybe, to be discussed – #63162
  - α: **Server Side Printing Columns** – "kubectl get" customization – #60991
  - β: **Subresources** – α since 1.10 – #62786

  - `OpenAPI additionalProperties` allowed now (mutually exclusive with `properties`)

- Kubernetes 1.12+
  - **Multiple versions** with **declarative field renames**
  - **Strict create mode?** Discuss: #5889 – my favorite CRD UX issue
    Related: CRD OpenAPI validation spec not served by kube-apiserver

The Future: Versioning

- Most asked for feature for long time
- It is coming, but slowly

"NoConversion": maybe in 1.11

```yaml
apiVersion: apiextensions.k8s.io/v1beta1
kind: CustomResourceDefinition
metadata:
  name: contibutorsummit.kubecon.io
spec:
  group: kubecon.io
  version: v1
  versions:
    - name: v1
      storage: true
    - name: v1alpha1
```

"Declarative Conversions": maybe in 1.12+

```yaml
apiVersion: apiextensions.k8s.io/v1beta1
kind: CustomResourceDefinition
metadata:
  name: contibutorsummit.kubecon.io
spec:
  group: kubecon.io
  version: v1
  conversions:
    declarative:
      renames:
        from: v1alpha1
        to: v1
        old: .spec.foo
        new: bar
```

Outlook – Prepare for Pruning

- **Deep change of semantics** of Custom Resources
- From JSON blob store to schema based storage

```
OpenAPIv3Schema: {
      properties: {
            foo: {}
      }
}
```

- Example CR: { **"foo": 1**, **"bar": 2** } → { **"foo": 1** }

Opt-in in CRD v1beta1
Mandatory in GA

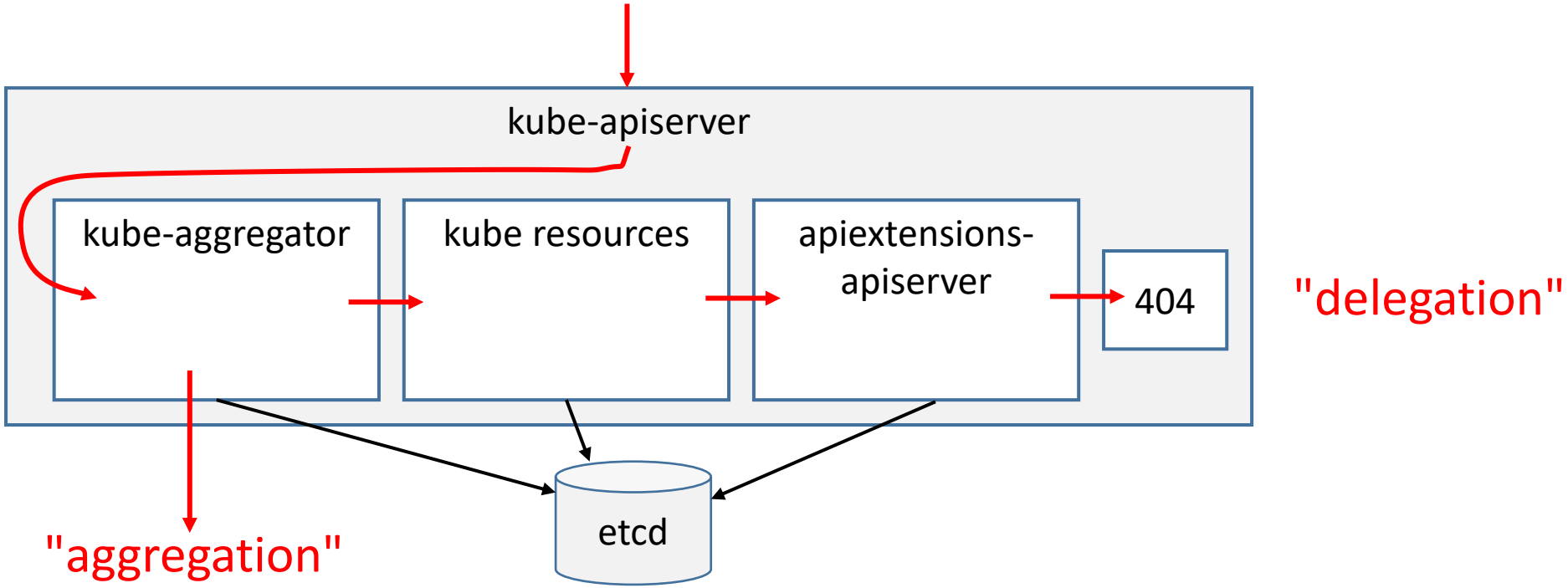
Deep Dive

apiextensions-apiserver

CustomResourceDefinitions are served by

https://github.com/kubernetes/apiextensions-apiserver

usually embedded into kube-apiserver via **delegation**.



api-machinery-session.kubecon.io.yaml

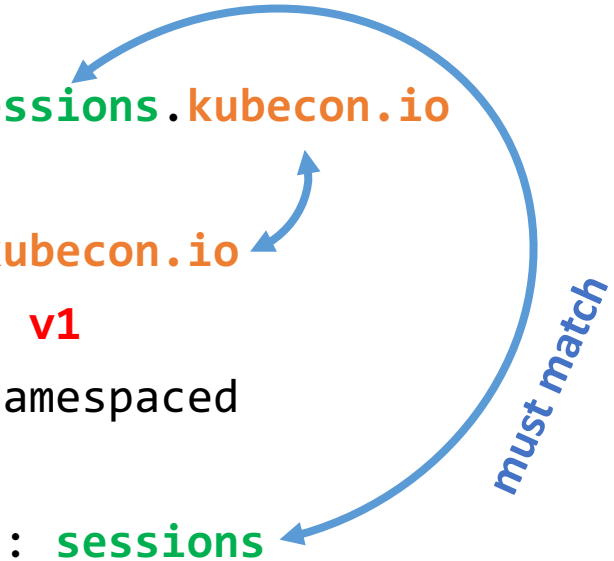
KubeCon | CloudNativeCon Europe 2018

```yaml
apiVersion: kubecon.io/v1
kind: Session
metadata:
  name: api-machinery
  namespace: eu2018
spec:
  type: deepdive
  title: "SIG API Machinery Deep Dive"
  capacity: 42
status:
  attendees: 23
  conditions:
    - lastTransitionTime: 2018-05-04T12:47:54Z
      status: "True"
      type: Started
```

sessions.kubecon.io.yaml


```yaml
apiVersion: apiextensions.k8s.io/v1beta1
kind: CustomResourceDefinition
metadata:
  name: sessions.kubecon.io
spec:
  group: kubecon.io
  version: v1
  scope: Namespaced
  names:
    plural: sessions
    singular: session
    kind: Session

    # shortNames:
    # - talks
```



the **resource:**

- usually lower-case singular
- in http path

the **kind:**

- usually capital singular
- like the Go type

Create & wait for Established


```
$ kubectl create -f sessions.kubecon.io.yaml
```

... and then watch `status.conditions["Established"]`.

Conditions: → **NamesAccepted** → **Established**

= no name conflicts = CRD is served*

* There is a race – to be fixed in #63068.
  Better wait 5 seconds in ≤1.10.

kubectl get sessions –v=7

- I0429 21:17:53.042783 66743 round_trippers.go:383] GET https://localhost:6443/apis
- I0429 21:17:53.135811 66743 round_trippers.go:383] GET https://localhost:6443/apis/kubecon.io/v1
- I0429 21:17:53.138353 66743 round_trippers.go:383] GET https://localhost:6443/apis/kubecon.io/v1/namespaces/default/sessions

LIST

discovery

No resources found.

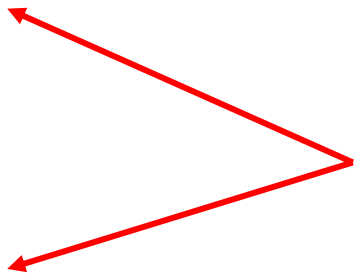
sessions → **kind** Session
resource sessions
} in **API group** kubecon.io/v1

note: we also support "shortNames"

We call this "REST mapping"

api-machinery-session.kubecon.io.yaml


```yaml
apiVersion: kubecon.io/v1
kind: Session
metadata:
  name: api-machinery
  namespace: eu2018
spec:
  type: deepdive
  title: "SIG API Machinery Deep Dive"
  capacity: 42
status:
  attendees: 23
  conditions:
    - lastTransitionTime: 2018-05-04T12:47:54Z
      status: "True"
      type: Started
```



Recommended to follow the **spec+status pattern**.
Important for /status subresource.

etcd Storage


KubeCon | CloudNativeCon Europe 2018

```
$ export ETCDCTL_API=3
$ etcdctl get / --prefix --keys-only | grep kubecon
/registry/apiextensions.k8s.io/customresourcedefinitions/sessions.kubecon.io
/registry/apiregistration.k8s.io/apiservices/v1.kubecon.io
/registry/kubecon.io/sessions/eu2018/api-machinery
```

```
$ etcdctl get /registry/kubecon.io/sessions/eu2018/api-machinery
{"apiVersion":"kubecon.io/v1","kind":"Session","metadata":{"clusterName":"","creat
ionTimestamp":"2018-04-29T20:30:27Z","generation":1,"name":"api-
machinery","namespace":"eu2018","resourceVersion":"","selfLink":"","uid":"273a1ae3
-4bec-11e8-8d91-4c3275978b79"},"spec":{"capacity":10,"title":"SIG API Machinery
Deep
Dive","type":"deepdive"},"status":{"attendees":10,"conditions":[{"lastTransitionTi
me":"2018-05-04T12:47:54Z","status":"True","type":"Started"}]}}
```

unverified JSON blob

unstructured.Unstructured

Internally, **CustomResources** are

```
import "k8s.io/apimachinery/pkg/apis/meta/v1/unstructured"

unstructured.Unstructured{
  Object: map[string]interface{}
}
```

json.Unmarshal

i.e. maps+slices+values.

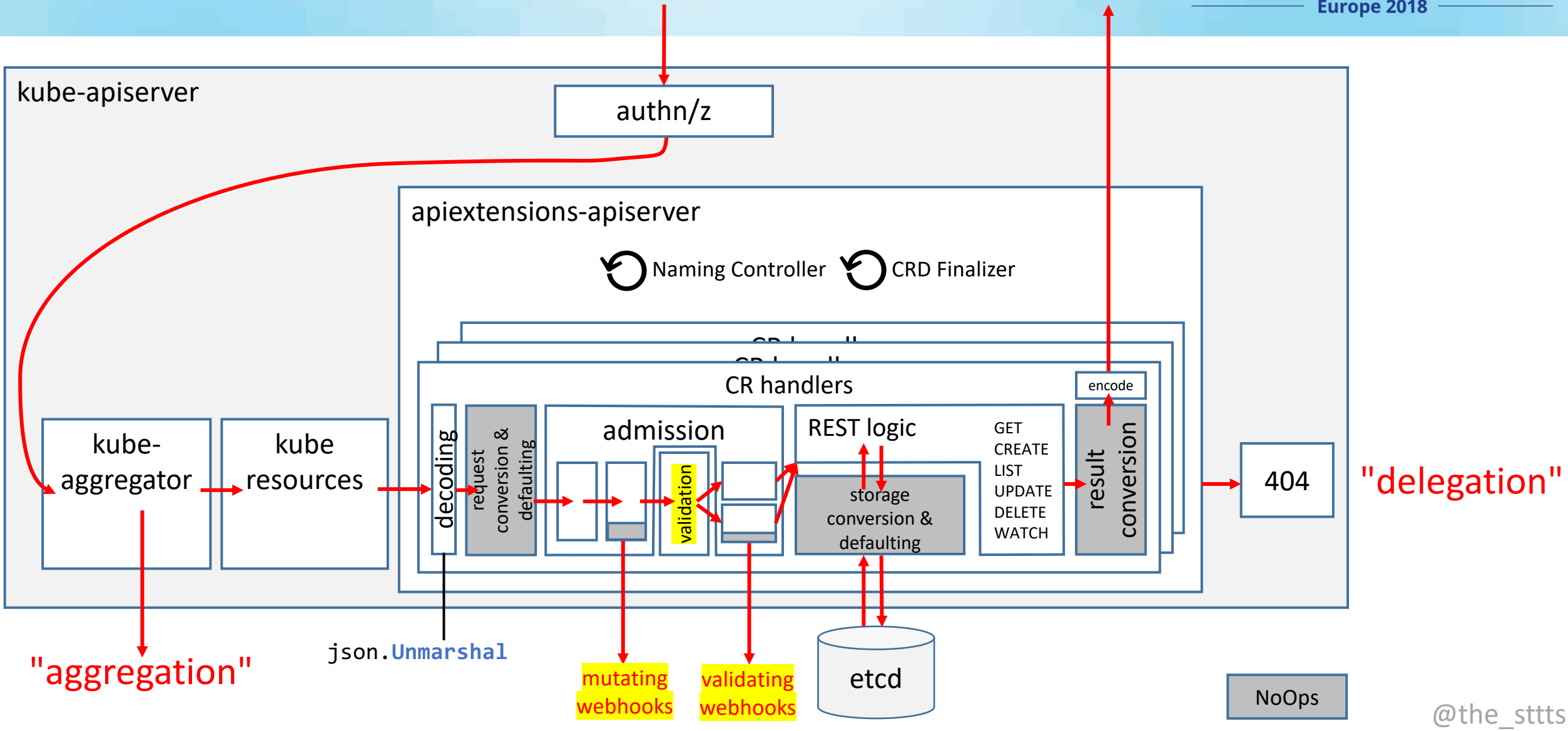
Dynamic Client

- client-go counterpart: k8s.io/client-go/dynamic
- in 1.11+ with sane interface #62913:

```
dynamic.NewForConfig(cfg).Resource(gvr).Namespace(ns).Get(name, opts)
```

- generated, typed clients are generally preferred

Zoom into apiextensions-apiserver



Validation

- The standard: **OpenAPI v3 schema**

  https://github.com/OAI/OpenAPI-Specification/blob/master/versions/3.0.0.md#schemaObject

- based on JSON Schema:

  https://tools.ietf.org/html/draft-wright-json-schema-validation-00

Custom Resource

```yaml
spec:
  type: deepdive
  title: "SIG API Machinery De…
  capacity: 42
status:
  attendees: 23
  conditions:
    - lastTransitionTime: 2018…
      status: "True"
      type: Started
```

Helpful tools:

kubernetes/kube-openapi#37

tamalsaha/kube-openapi-generator

Some other tool from prometheus-operator?

Rancher has another one, speak to @lemonjet

```yaml
properties:
  spec:
    properties:
      type:
        anyOf: [{"pattern": "^deepdive$"}, …]
      title: {"type": "string"}
      capacity: {"type": "format": "integer", "minimum": 0, "default": 0}
    required: ["type", "title", "capacity"]
  status:
    properties:
      attendees: {"type": "number", "format": "integer", "minimum": 0}
      conditions:
        type: "array"
        items:
          properties:
            lastTransitionTime: {"type": "dateTime"}
            status:
              anyOf: [{"pattern": "^True$"}, …]
            type:
              anyOf: [{"pattern": "^Started$"}, …]
          required: ["lastTransitionTime", "status", "type"]
```

a quantor (anyOf, oneOf, allOf exist)

note: enum is forbidden (why?)

maybe in 1.11+

regular expression

OpenAPI v3 Schema

etcd Storage – Pruning


```
$ export ETCDCTL_API=3
$ etcdctl get / --prefix --keys-only | grep kubecon
/registry/apiextensions.k8s.io/customresourcedefinitions/sessions.kubecon.io
/registry/apiregistration.k8s.io/apiservices/v1.kubecon.io
/registry/kubecon.io/sessions/eu2018/api-machinery

$ etcdctl get /registry/kubecon.io/sessions/eu2018/api-machinery
{"apiVersion":"kubecon.io/v1","kind":"Session","metadata":{"clusterName":"","
creationTimestamp":"2018-04-29T20:30:27Z","generation":1,"name":"api-
machinery","namespace":"eu2018","resourceVersion":"","selfLink":"","uid":"273
a1ae3-4bec-11e8-8d91-4c3275978b79"},"spec":{"capacity":10,"title":"SIG API
Machinery Deep
Dive","type":"deepdive"},"status":{"attendees":10,"conditions":[{"lastTransit
ionTime":"2018-05-04T12:47:54Z","status":"True","type":"Started",
"someUnknownField":"someValue", "someFutureField":"dangerous value"}]}}
```

~~un~~verified JSON blob
with possibly unspecified fields

we need pruning!
Kube 1.11+

Deeper Dive – go-openapi/validate


```
validator := validate.NewSchemaValidator(schema, …)
result := validator.Validate(obj)
```

= OpenAPIv3Schema

= JSON object

```
specSchema := result.FieldSchemata()[ validator.NewFieldKey(obj, "spec") ]
```

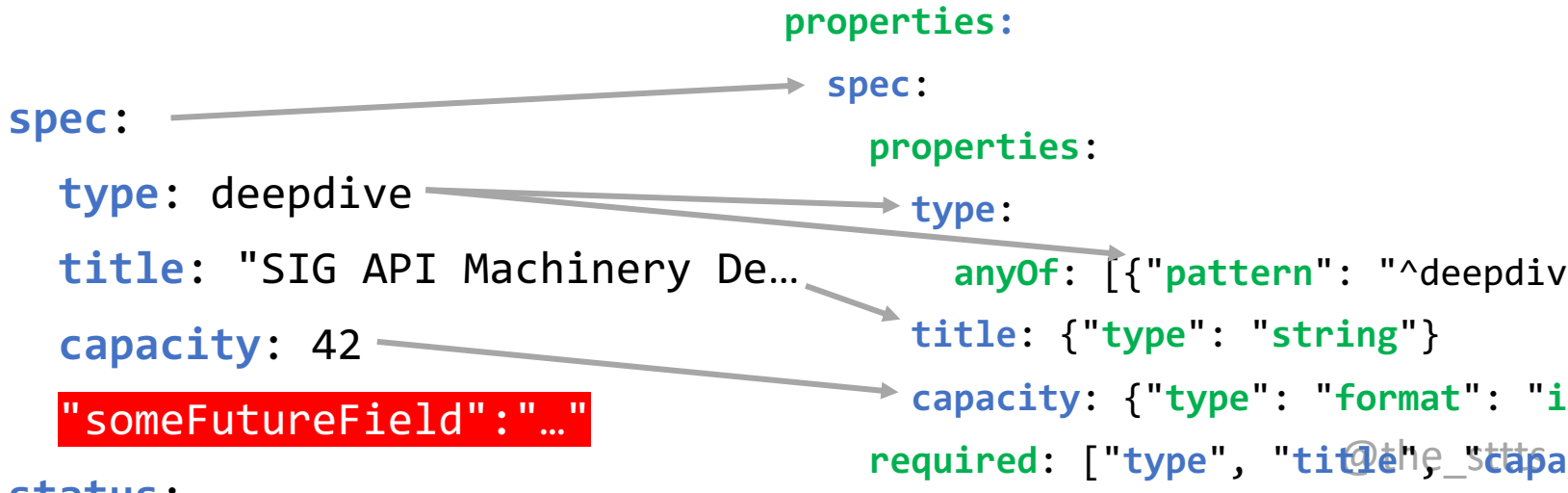
OpenAPI validation **result** gives us a **mapping**: JSON nodes → OpenAPI schemata:

Deeper Dive – go-openapi/validate

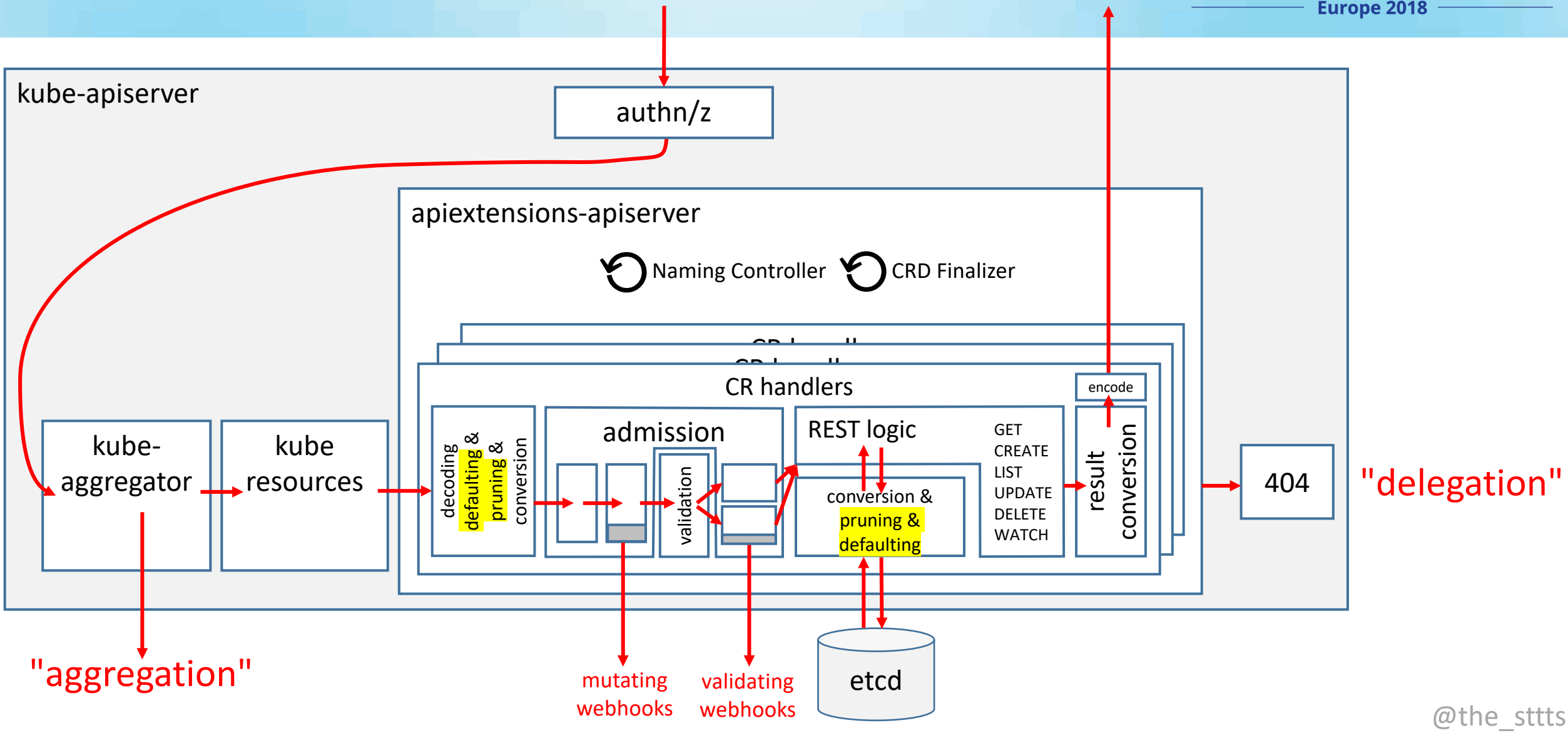

```
func ApplyDefaults(r *validate.Result) {
  fieldSchemata := r.FieldSchemata()
  for key, schemata := range fieldSchemata {
    LookForDefaultingScheme:
    for _, s := range schemata {
      if s.Default != nil {
        if _, found := key.Object()[key.Field()]; !found {
          key.Object()[key.Field()] = s.Default
          break LookForDefaultingScheme
        }
      }
    }
  }
}
```

← defaulting algorithm on half a slide

sketch of pruning→



Zoom into apiextensions-apiserver



Scaling the session


```
$ kubectl scale --replicas=10 -n eu2018 sessions/api-machinery --v=7
```

I0429 22:33:03.083150 74535 round_trippers.go:383] GET https://localhost:6443/apis/kubecon.io/v1/namespaces/eu2018/sessions/api-machinery/scale

I0429 22:33:03.083725 74535 round_trippers.go:408] Response Status: 404 Not Found in 0 milliseconds

We call this "subresource /scale".

```yaml
spec:
  type: deepdive
  title: "SIG API Machinery De…
  capacity: 42
status:
  attendees: 23
  conditions:
    - lastTransitionTime: 2018…
      status: "True"
      type: Started
```

alpha in 1.10
hopefully beta in 1.11

```yaml
apiVersion: apiextensions.k8s.io/v1beta1
kind: CustomResourceDefinition
metadata:
  name: sessions.kubecon.io
spec:
 …
 subresources:
   scale:
     specReplicasPath: .spec.capacity
     statusReplicasPath: .status.attendees

   # status: {}
```

JSON paths

Scaling the session


```
$ kubectl scale --replicas=10 -n eu2018 sessions/api-machinery --v=7
```

- I0429 22:43:14.757286 80725 round_trippers.go:405] GET https://localhost:6443/apis/kubecon.io/v1/namespaces/eu2018/sessions/api-machinery/scale 200 OK in 0 milliseconds
- I0429 22:43:14.757318 80725 request.go:897] Response Body:

```
{
        "kind": "Scale",
        "apiVersion": "autoscaling/v1",
        "metadata": {...},
        "spec": {"replicas":42},
        "status":{"replicas":23}
}
```

- PUT https://localhost:6443/apis/kubecon.io/v1/namespaces/eu2018/sessions/api-machinery/scale 200 OK in 2 milliseconds

session.kubecon.io/api-machinery scaled

(polymorphic) scale client


```
import (
        "k8s.io/client-go/discovery/cached"
        "k8s.io/client-go/scale"
)

cachedDiscovery    := discocache.NewMemCacheClient(hpaClientGoClient.Discovery())
restMapper         := discovery.NewDeferredDiscoveryRESTMapper(cachedDiscovery)
scaleKindResolver  := scale.NewDiscoveryScaleKindResolver(hpaClientGoClient.Discovery())
scaleClient, err   := scale.NewForConfig(cfg, restMapper, dynamic.LegacyAPIPathResolverFunc, scaleKindResolver)
```

```yaml
spec:
  type: deepdive
  title: "SIG API Machinery De…
  capacity: 42
status:
  attendees: 23
  conditions:
    - lastTransitionTime: 2018…
      status: "True"
      type: Started
```

alpha in 1.10
hopefully beta in 1.11

```yaml
apiVersion: apiextensions.k8s.io/v1beta1
kind: CustomResourceDefinition
metadata:
  name: sessions.kubecon.io
spec:
 …
 subresources:
   scale:
     specReplicasPath: .spec.capacity
     statusReplicasPath: .status.attendees

   status: {}
```

JSON paths

spec/status split

main endpoint only changes .spec
/status changes .status

Recap

Outlook – Prepare for Pruning

- **Deep change of semantics** of Custom Resources
- From JSON blob store to schema based storage

```
OpenAPIv3Schema: {
      properties: {
            foo: {}
      }
}
```

- Example CR: { **"foo": 1**, **"bar": 2** } → { **"foo": 1** }

Opt-in in CRD v1beta1
Mandatory in GA

Outlook – Custom Resources

- Kubernetes 1.11+
  - α: **Multiple versions without conversion** – design proposal
  - α: **Pruning** – in validation spec unspecified fields are removed – blocker for GA
  - α: **Defaulting** – defaults from OpenAPI validation schema are applied
  - α: **Graceful Deletion** – maybe, to be discussed – #63162
  - α: **Server Side Printing Columns** – "kubectl get" customization – #60991
  - β: **Subresources** – α since 1.10 – #62786

  - `OpenAPI additionalProperties` allowed now (mutually exclusive with `properties`)

- Kubernetes 1.12+
  - **Multiple versions** with **declarative field renames**
  - **Strict create mode?** Discuss: #5889 – my favorite CRD UX issue
    Related: CRD OpenAPI validation spec not served by kube-apiserver