Follow the Data from the Darkest Corners of K8s

Frederic Branczyk (@fredbrancz) & Piotr Szcześniak (@piosz)

Production survey

- Do you run production workloads?

Production survey

✓ Do you run production workloads?

Production survey

- ✓ Do you run production workloads?
- ❑ Do you have monitoring of the workloads?

Production survey

- ✓ Do you run production workloads?
- ✓ Do you have monitoring of the workloads?

Production survey

- ✓ Do you run production workloads?
- ✓ Do you have monitoring of the workloads?

Production survey

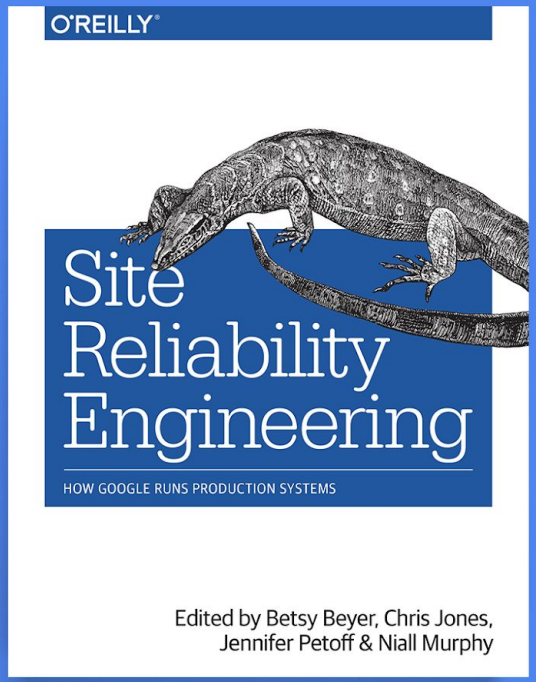
✓ Do you run production workloads?

✗ Do you have monitoring of the workloads?

Production survey

✓ Do you run production workloads?

✗ Do you have monitoring of the workloads?



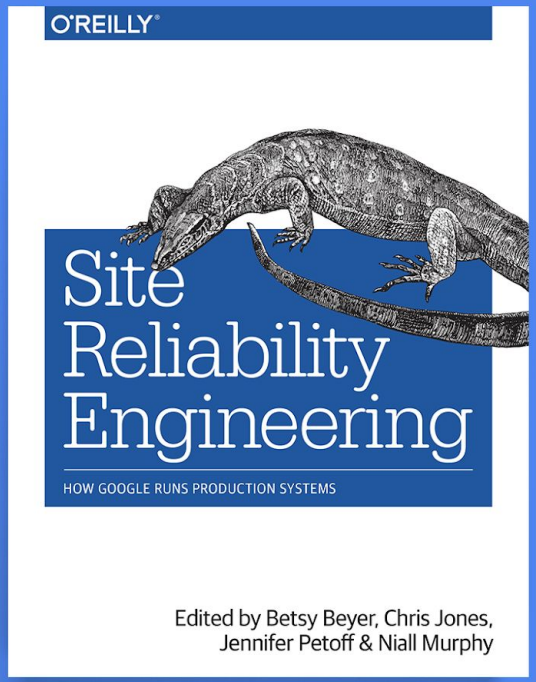
Site Reliability Engineering

Edited by Betsy Beyer, Chris Jones, Jennifer Petoff and Niall Richard Murphy

Members of the SRE team explain how their engagement with the entire software lifecycle has enabled Google to build, deploy, monitor, and maintain some of the largest software systems in the world.

READ ONLINE FOR FREE

BUY FROM GOOGLE BOOKS



Site Reliability Engineering

Edited by Betsy Beyer, Chris Jones, Jennifer Petoff and Niall Richard Murphy

Members of the SRE team explain how their engagement with the entire software lifecycle has enabled Google to build, deploy, monitor, and maintain some of the largest software systems in the world.

READ ONLINE FOR FREE

BUY FROM GOOGLE BOOKS

“Kubernetes is not about monitoring”

“Kubernetes is not about monitoring”
but ...

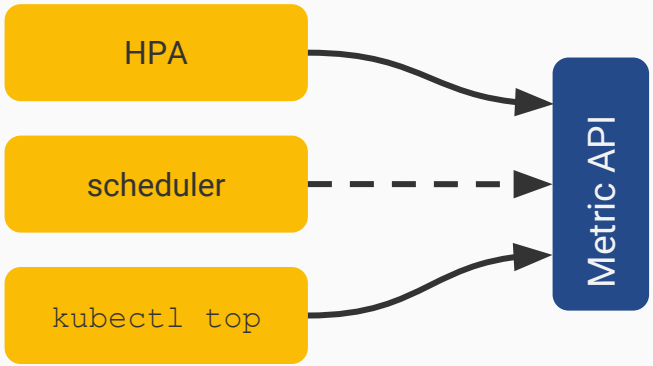
Components need metrics

HPA

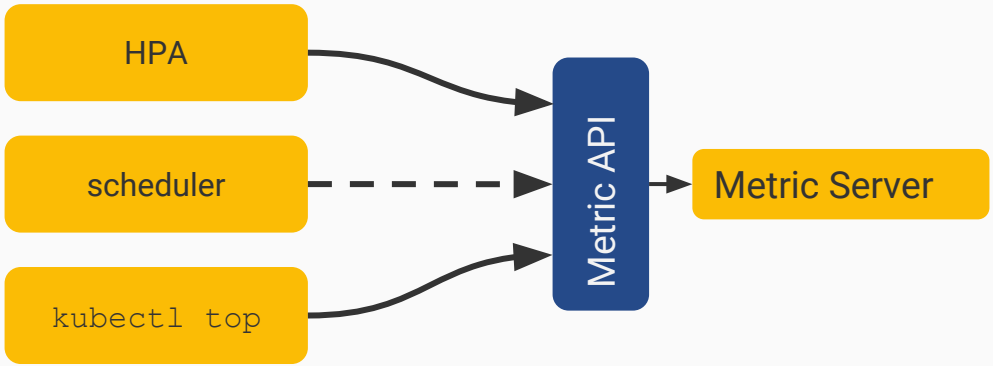
scheduler

`kubectl top`

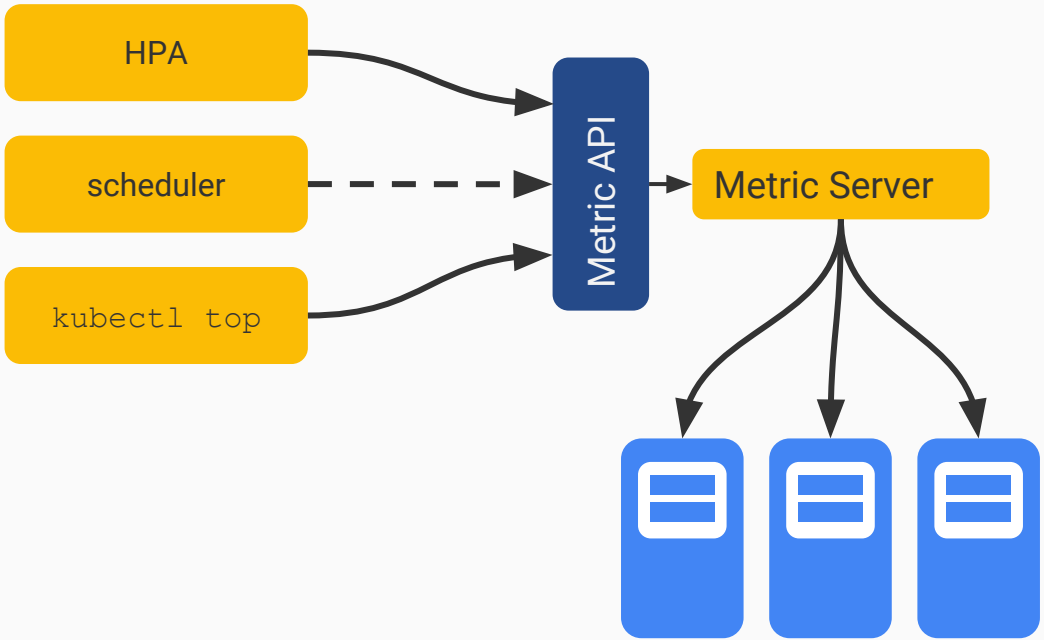
Metrics API



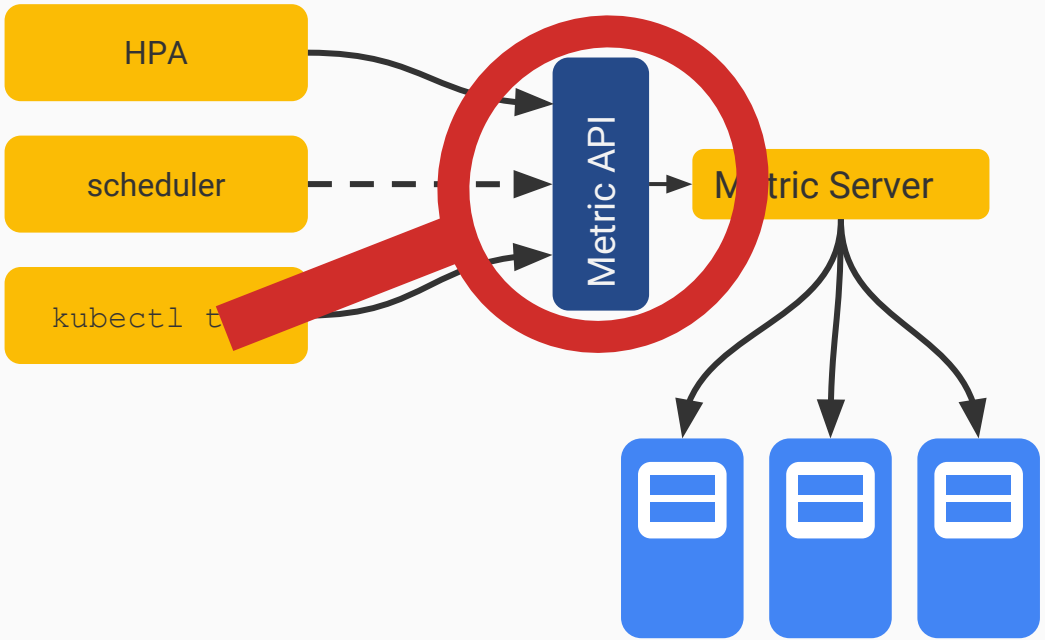
Metrics Server



Metrics Server scrapes nodes



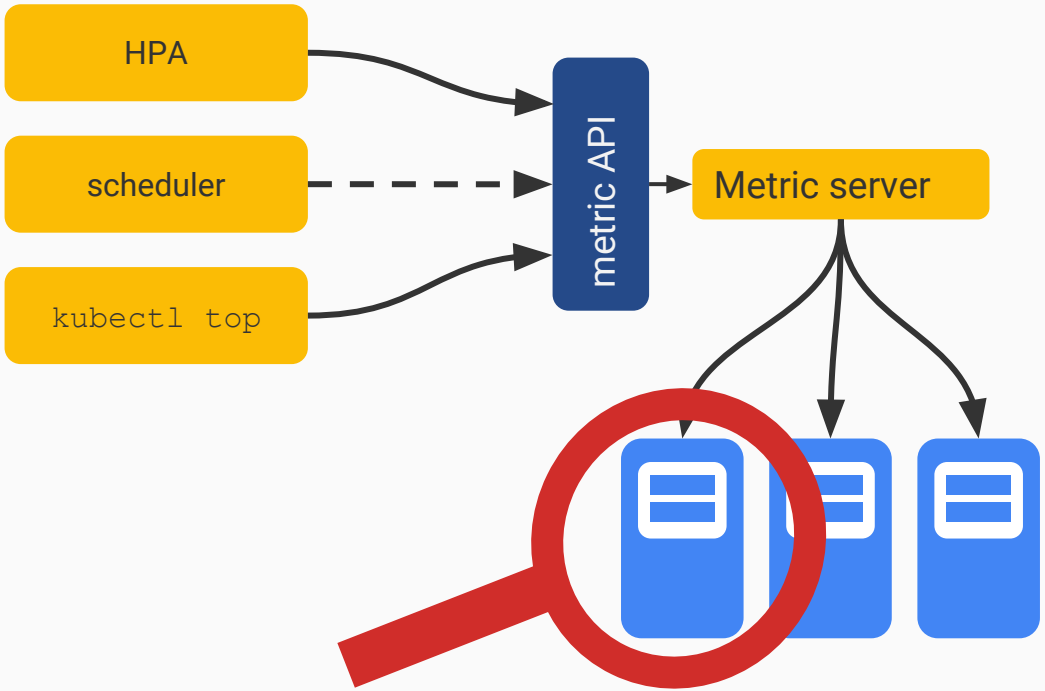
Metrics API closer look



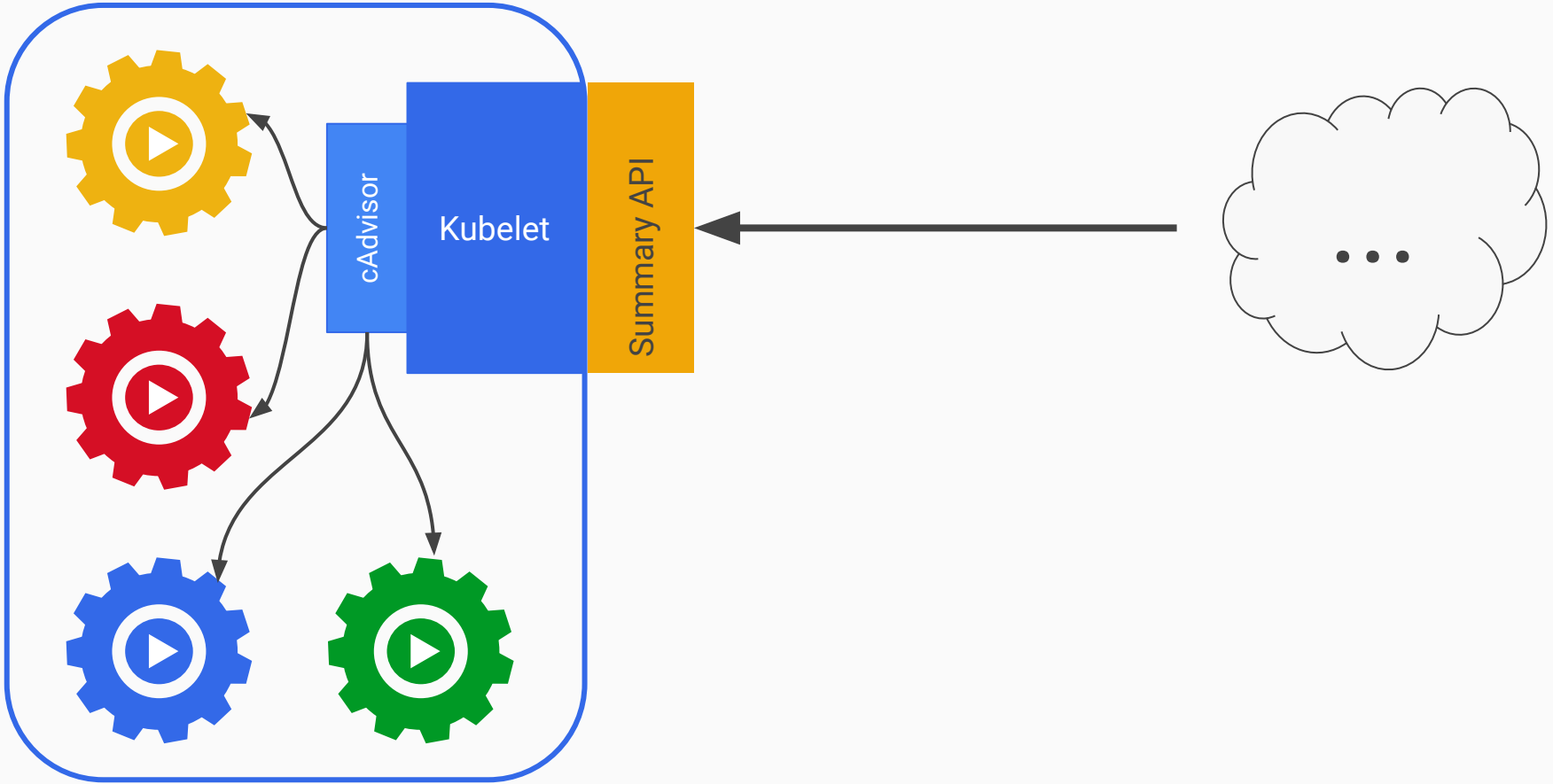
Metrics API closer look

```
type NodeMetrics struct {
	Timestamp metav1.Time
	Window    metav1.Duration

	Usage v1.ResourceList
}

type PodMetrics struct {
	Timestamp metav1.Time
	Window    metav1.Duration

	Containers []ContainerMetrics
}

type ContainerMetrics struct {
	Name string
	Usage v1.ResourceList
}
```

Node closer look



Node closer look

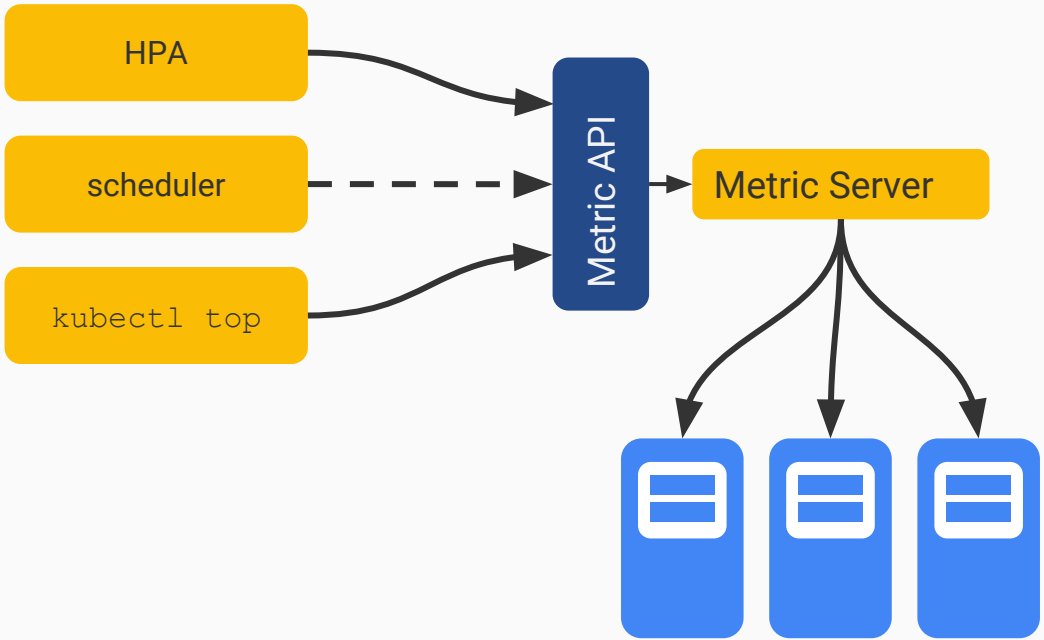


“Kubernetes is not about monitoring”

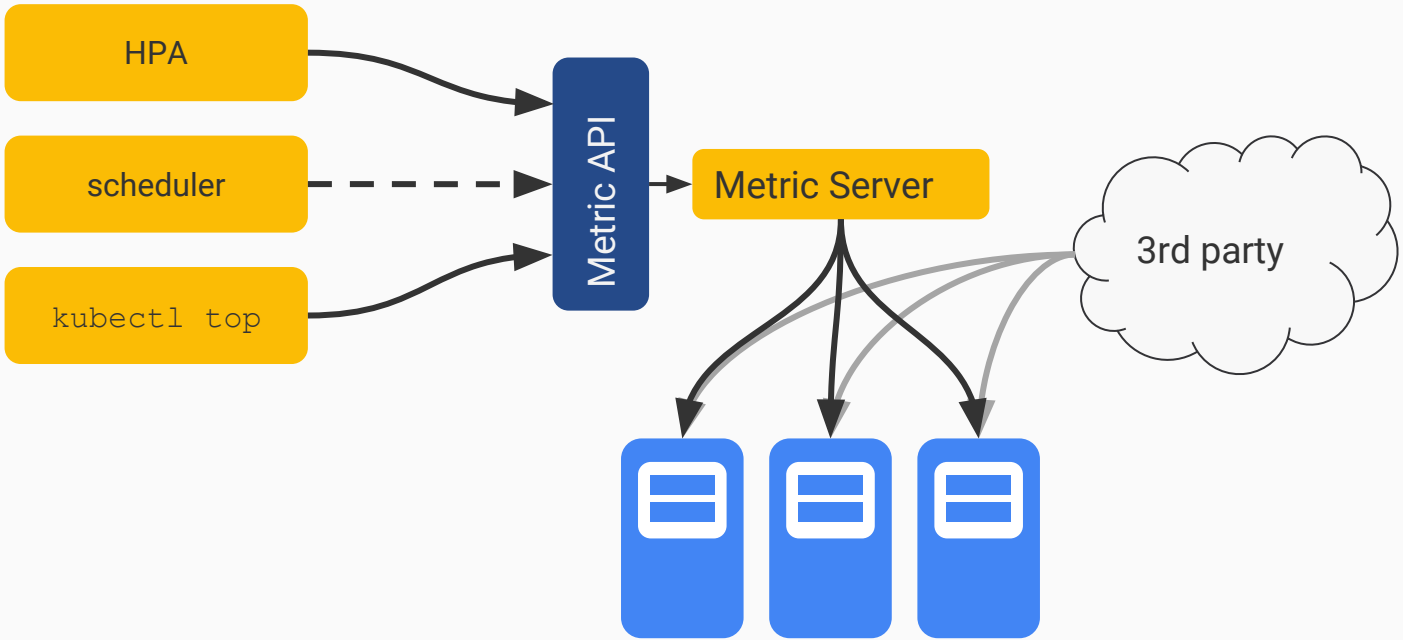
“Kubernetes is not about monitoring”

so ...

Core metrics pipeline



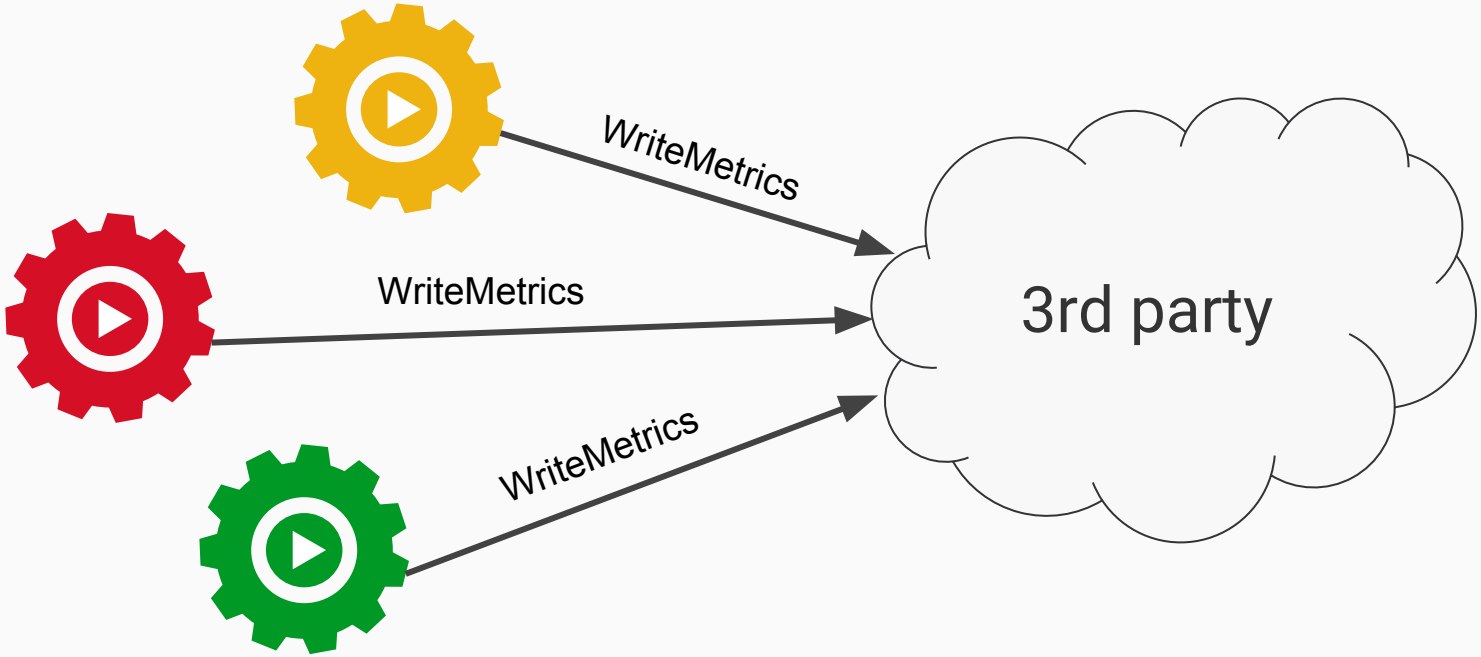
3rd party monitoring solution



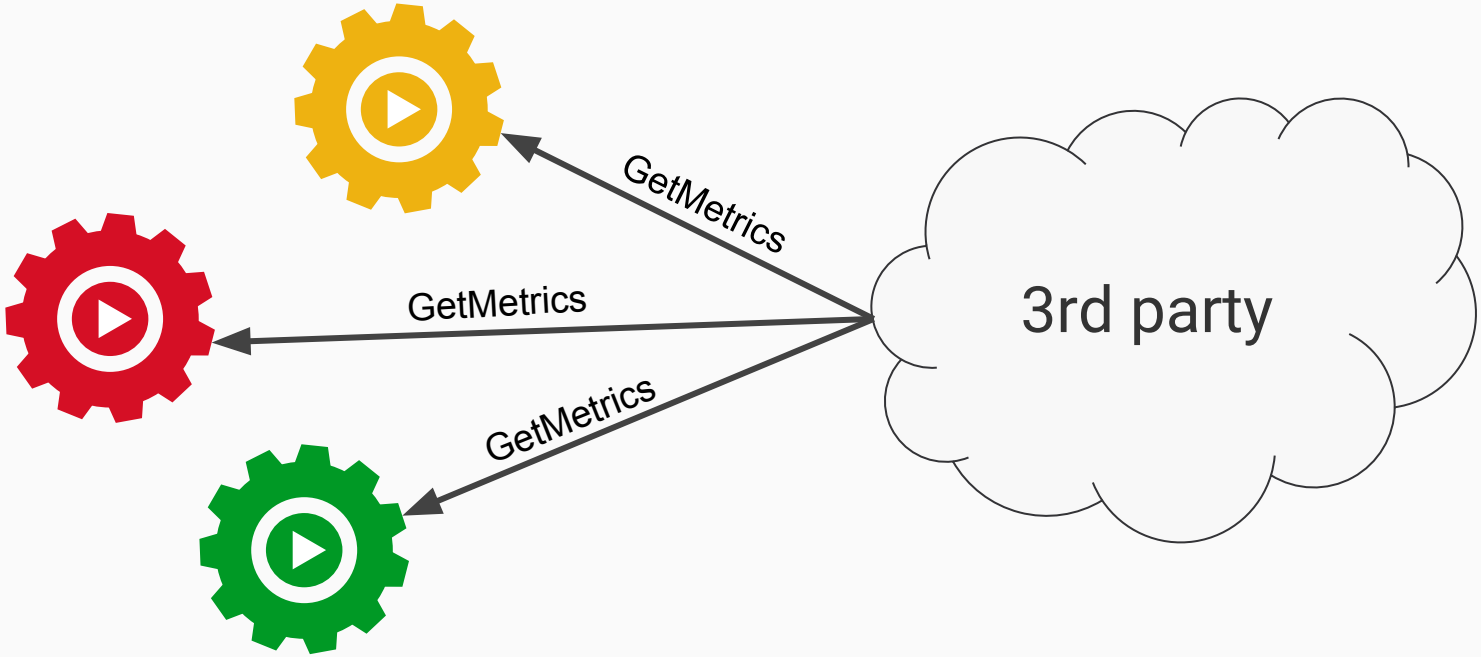
Push vs. Pull

Push vs. Pull

Push



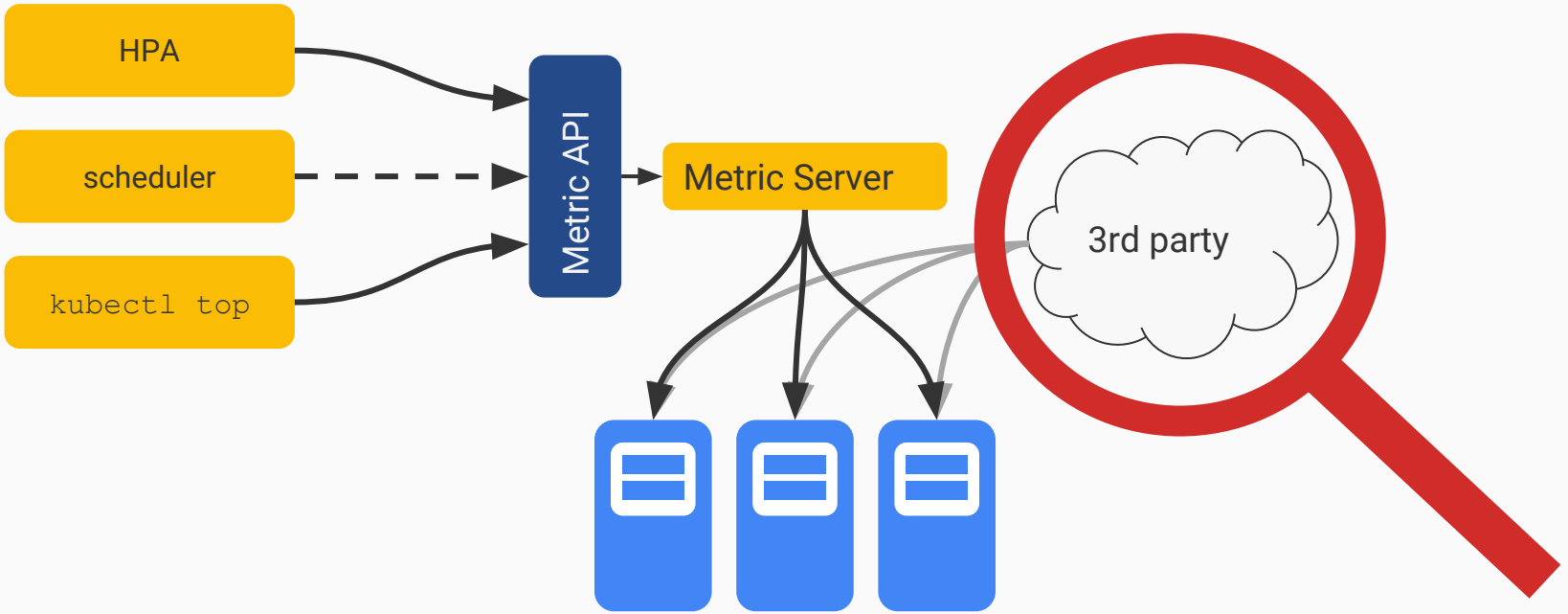
Pull



The right approach is...

The right approach is...

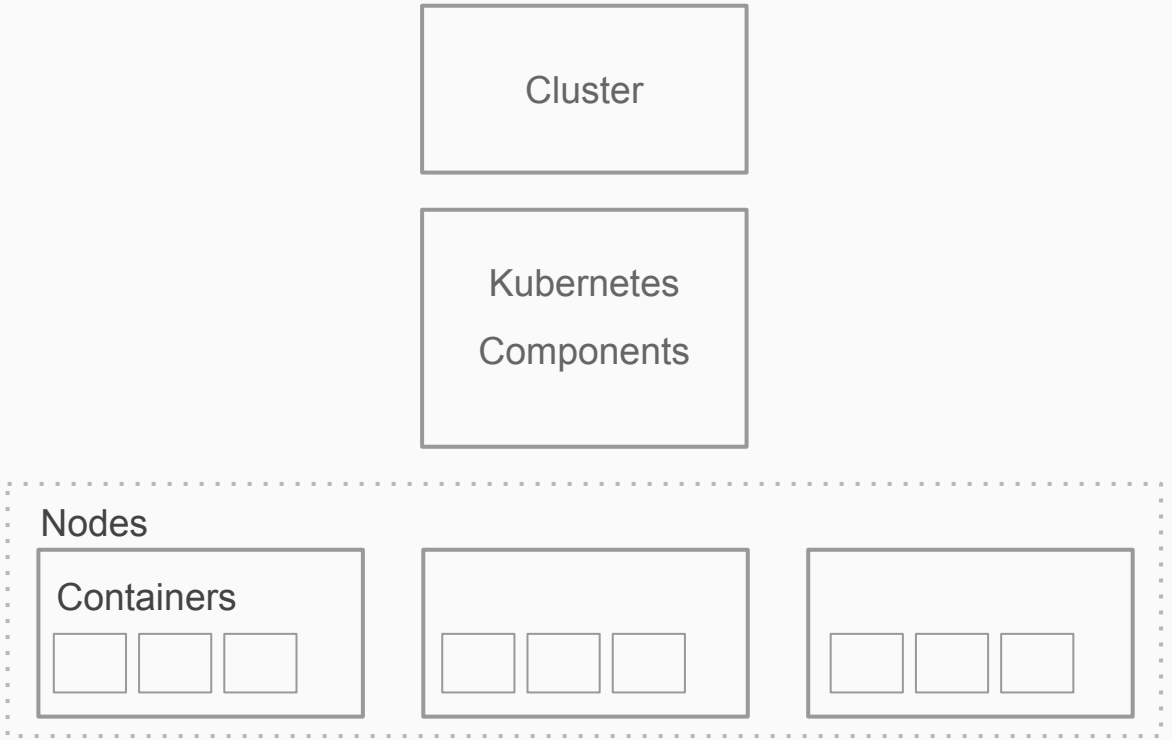
3rd party monitoring solution - closer look



Prometheus

Bottom up

- Nodes
- Pods/Containers
- Kubernetes
- More



Node level

Nodes

- Prometheus node-exporter

```
predict_linear(node_filesystem_free[6h], 3600 * 24) < 0
```

Pods/Containers

- cAdvisor (soon deprecated)
- See Solly's proposal for Kubernetes node metrics!
  - https://docs.google.com/document/d/1_CdNWIjPBqVDMvu82aJICQsSCbh2BR-y9a8uXjQm4TI/edit
- Join sig-instrumentation!

Dashboards!


CPU Usage
1.00
0.75
0.50
0.25
0
06:40
06:45
06:50
06:55
07:00
07:05
07:10
07:15
07:20
07:25
07:30
07:35
kube-apiserver-6wvr7
kube-apiserver-dhbqb
kube-apiserver-f6xd7
kube-controller-manager-5587f5cdb4-c9zgn
kube-controller-manager-5587f5cdb4-v79k9
kube-dns-7cd69df56-mfl26
kube-etcd-endpoints-f8d44b879-5ncs6
kube-flannel-54tmj
kube-flannel-77xkh
kube-flannel-fbqr4
kube-flannel-rq9gp
kube-proxy-k7ndp
kube-proxy-kg42n
kube-proxy-rgzkh
kube-proxy-s4cnt
kube-scheduler-6587d4d955-2fv8p
kube-scheduler-6587d4d955-qk788
pod-checkpointer-9lgn8
pod-checkpointer-9lgn8-president-1.brancz.com
pod-checkpointer-l267x
pod-checkpointer-l267x-president-0.brancz.com
pod-checkpointer-qr452
pod-checkpointer-qr452-president-2.brancz.com
CPU Quota
(1 hidden panels)
Memory Usage
Memory Usage
4.0 Bil
3.0 Bil
2.0 Bil
1.0 Bil
0
06:40
06:45
06:50
06:55
07:00
07:05
07:10
07:15
07:20
07:25
07:30
07:35
kube-apiserver-6wvr7
kube-apiserver-dhbqb
kube-apiserver-f6xd7
kube-controller-manager-5587f5cdb4-c9zgn
kube-controller-manager-5587f5cdb4-v79k9
kube-dns-7cd69df56-mfl26
kube-etcd-endpoints-f8d44b879-5ncs6
kube-flannel-54tmj
kube-flannel-77xkh
kube-flannel-fbqr4
kube-flannel-rq9gp
kube-proxy-k7ndp
kube-proxy-kg42n
kube-proxy-rgzkh
kube-proxy-s4cnt
kube-scheduler-6587d4d955-2fv8p
kube-scheduler-6587d4d955-qk788
pod-checkpointer-9lgn8
pod-checkpointer-9lgn8-president-1.brancz.com
pod-checkpointer-l267x
pod-checkpointer-l267x-president-0.brancz.com
pod-checkpointer-qr452
pod-checkpointer-qr452-president-2.brancz.com

Kubernetes components

Kubelet

- Container runtime operations
- Kubelet stats (running containers, request latencies, etc.)

```
rate(kubelet_docker_operations_errors[5m])
```

```
kubelet_pod_start_latency_microseconds
```

API

- Central to all performance and errors in Kubernetes

```
sum(rate(apiserver_request_count{code=~"[45][0-9][0-9]"}[1h]))
                    by(resource, verb) /
  sum(rate(apiserver_request_count[1h])) by(resource, verb)
```

Scheduler

- Scheduling latency

```
histogram_quantile(0.99,
sum(scheduler_e2e_scheduling_latency_microseconds_bucket)
by (le, cluster)) / 1e+06
```

Controller manager

- A consumer of the API

```
rate(rest_client_requests_total{code!~"2.."}[5m])
```

Cluster level

kube-state-metrics

- Metrics from Kubernetes API objects

```
kube_deployment_spec_replicas{job="kube-state-metrics"}
                              !=
kube_deployment_status_replicas_available{job="kube-state-metrics"}
```

etcd

etcd

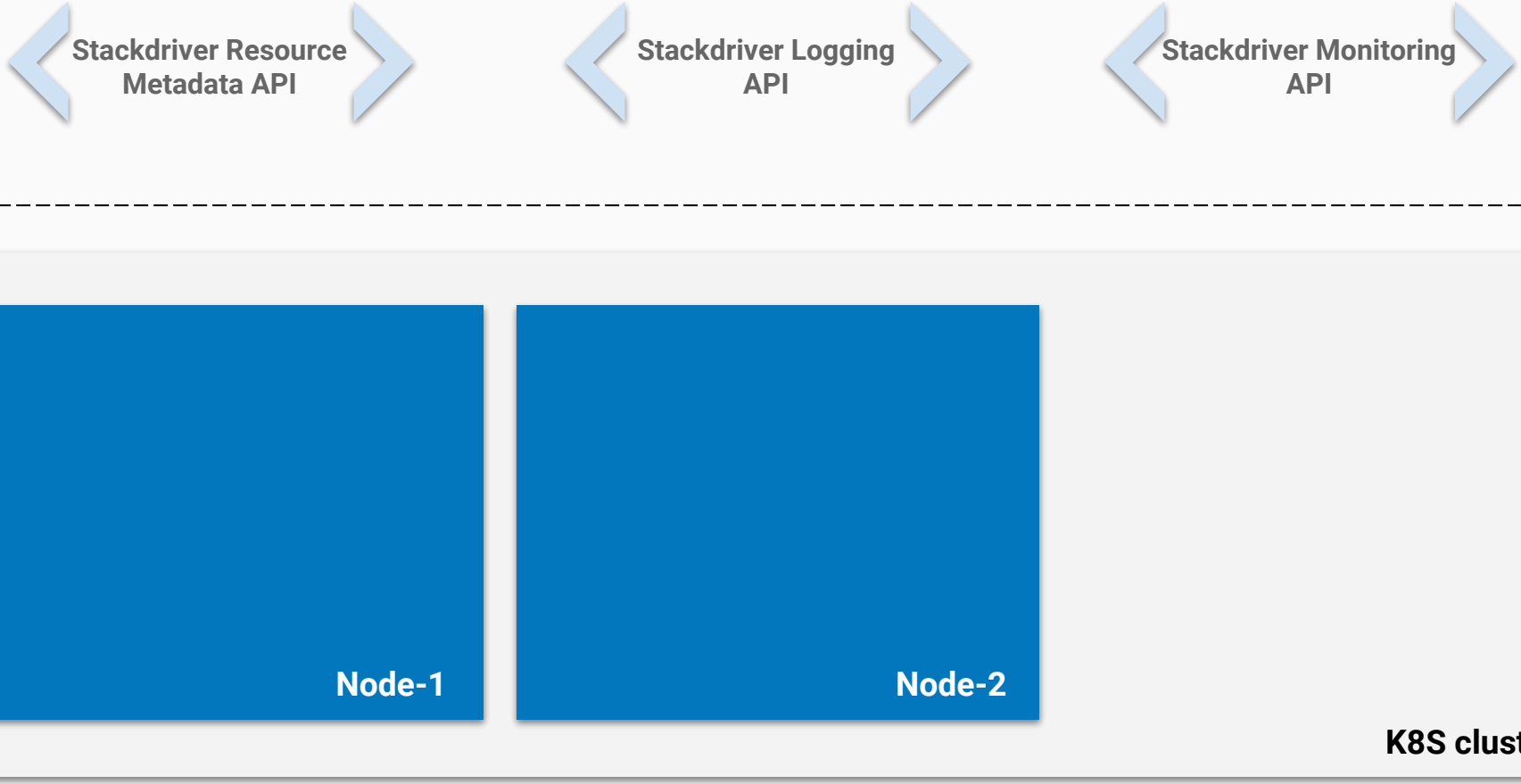
- Heart of Kubernetes

```
histogram_quantile(0.99,
rate(etcd_disk_wal_fsync_duration_seconds_bucket[5m]))
```

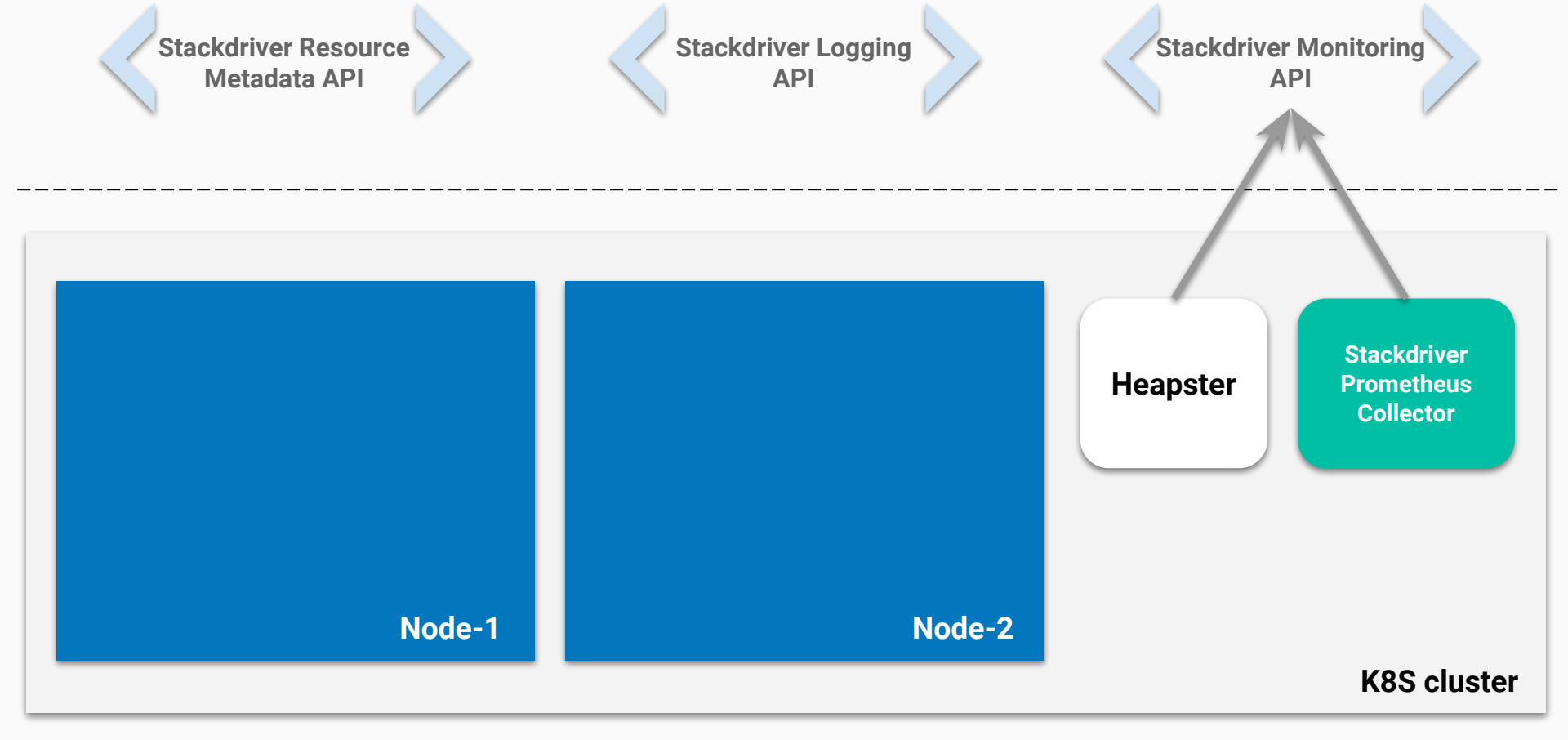
Let’s look at some real data!

Stackdriver

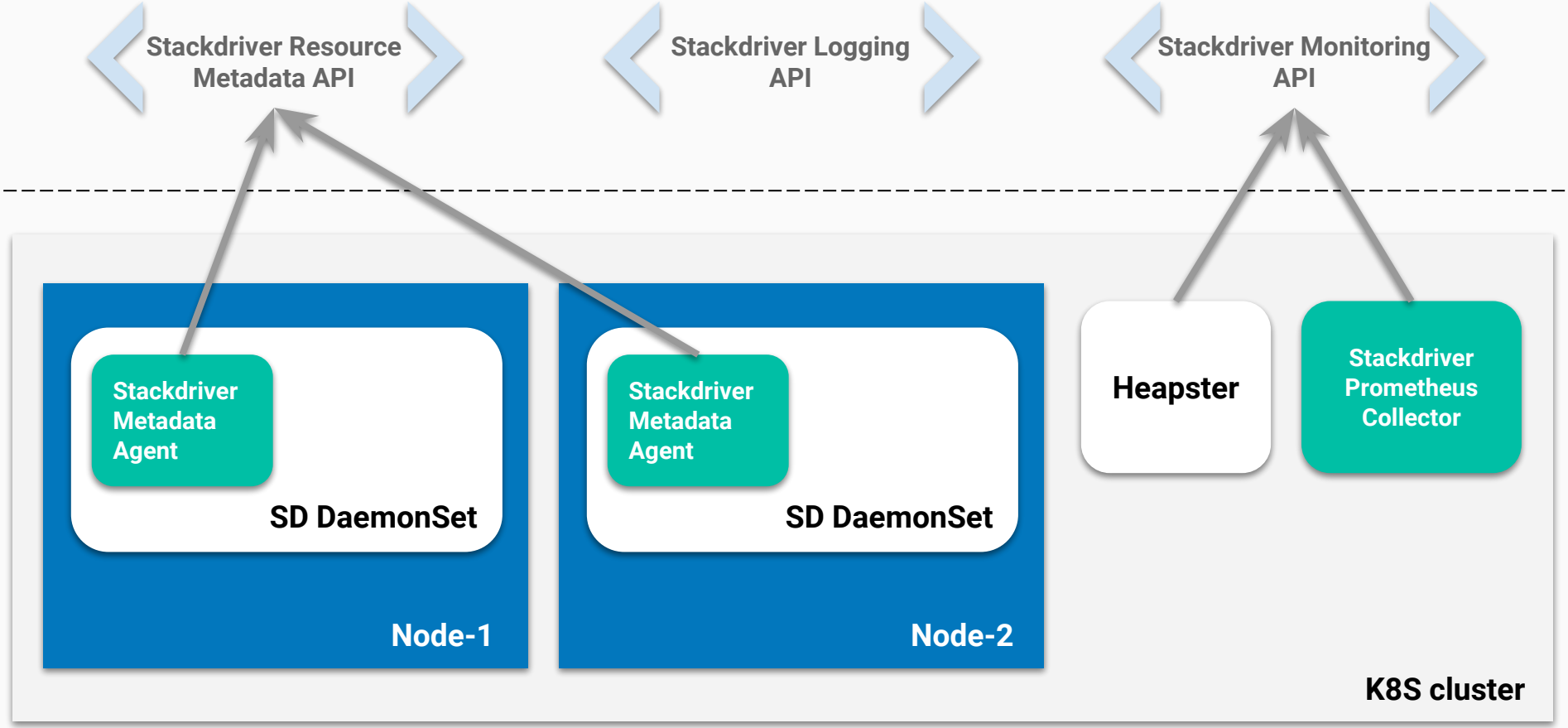
Stackdriver-Kubernetes integration overview



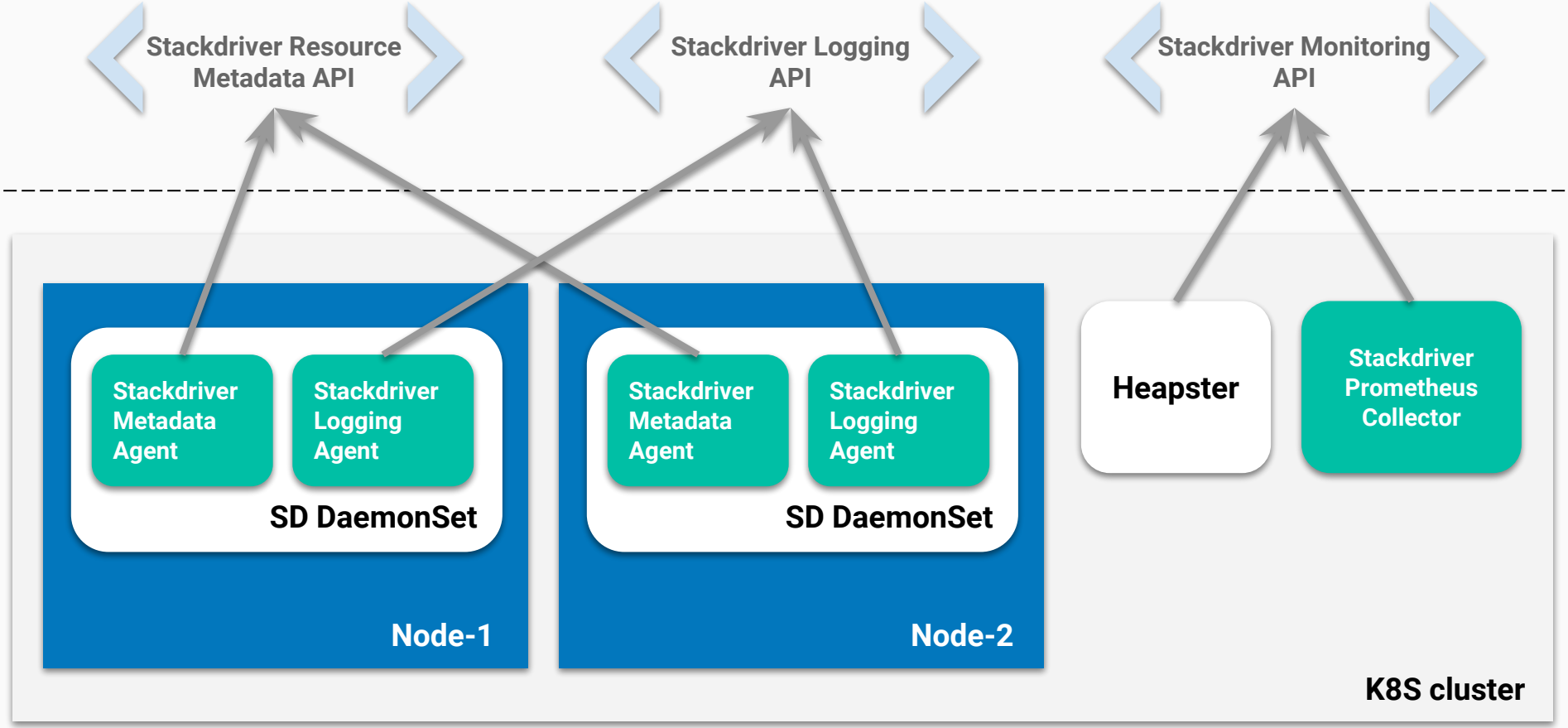
Stackdriver-Kubernetes integration overview



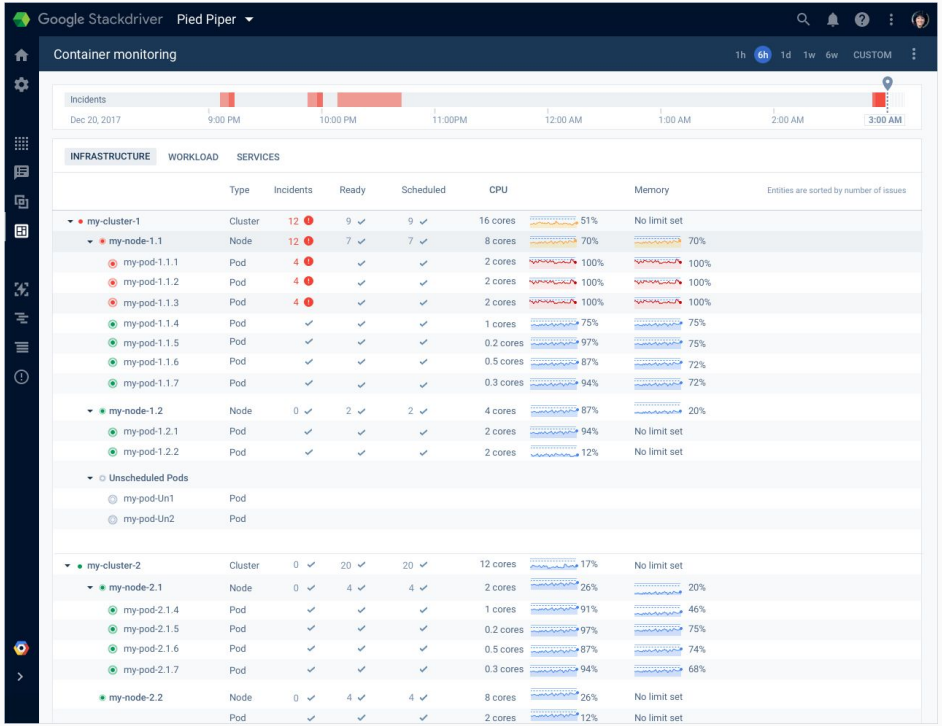
Stackdriver-Kubernetes integration overview

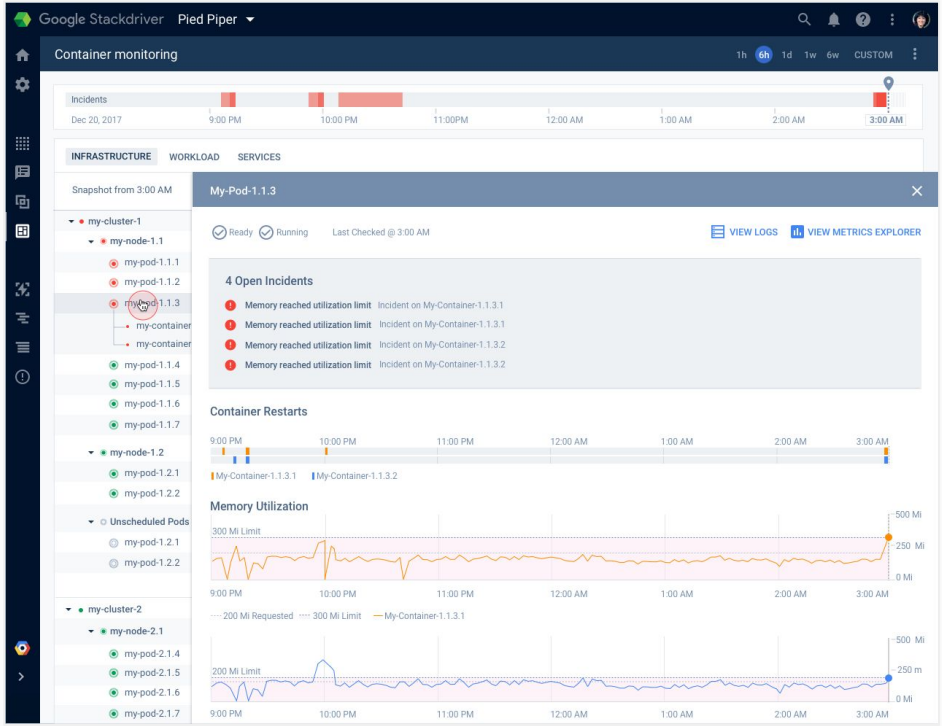


Stackdriver-Kubernetes integration overview



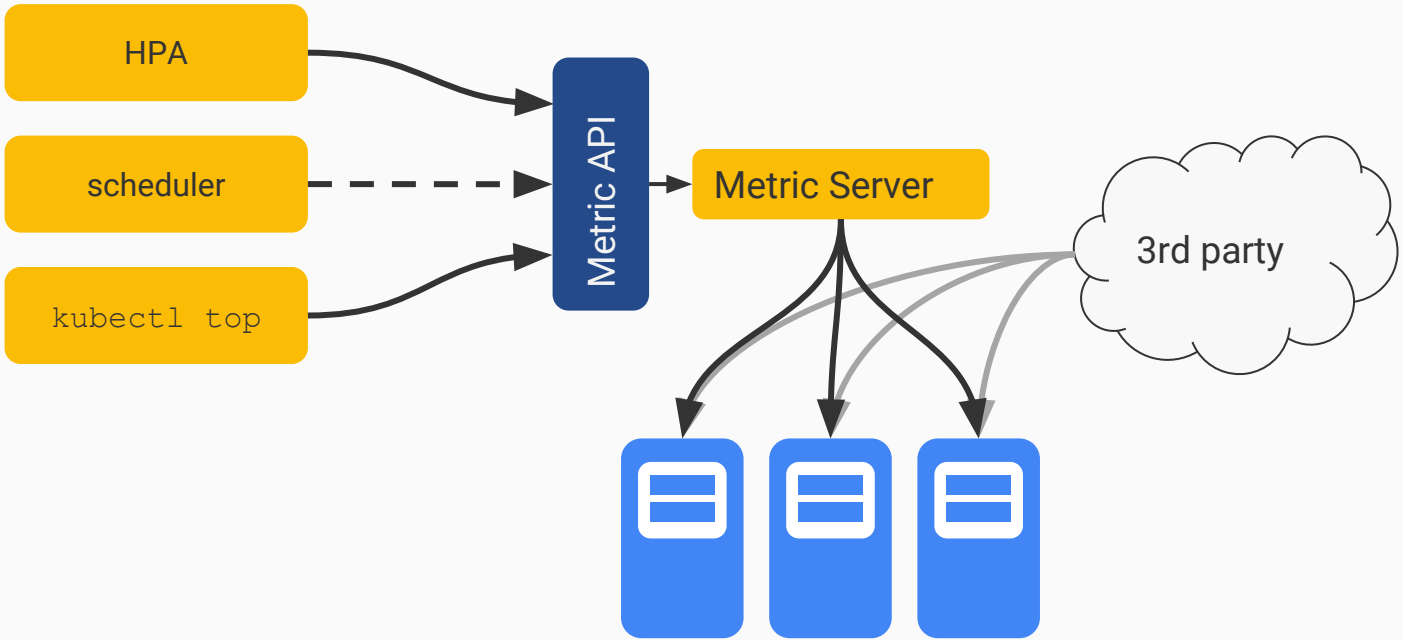
New Stackdriver dashboards



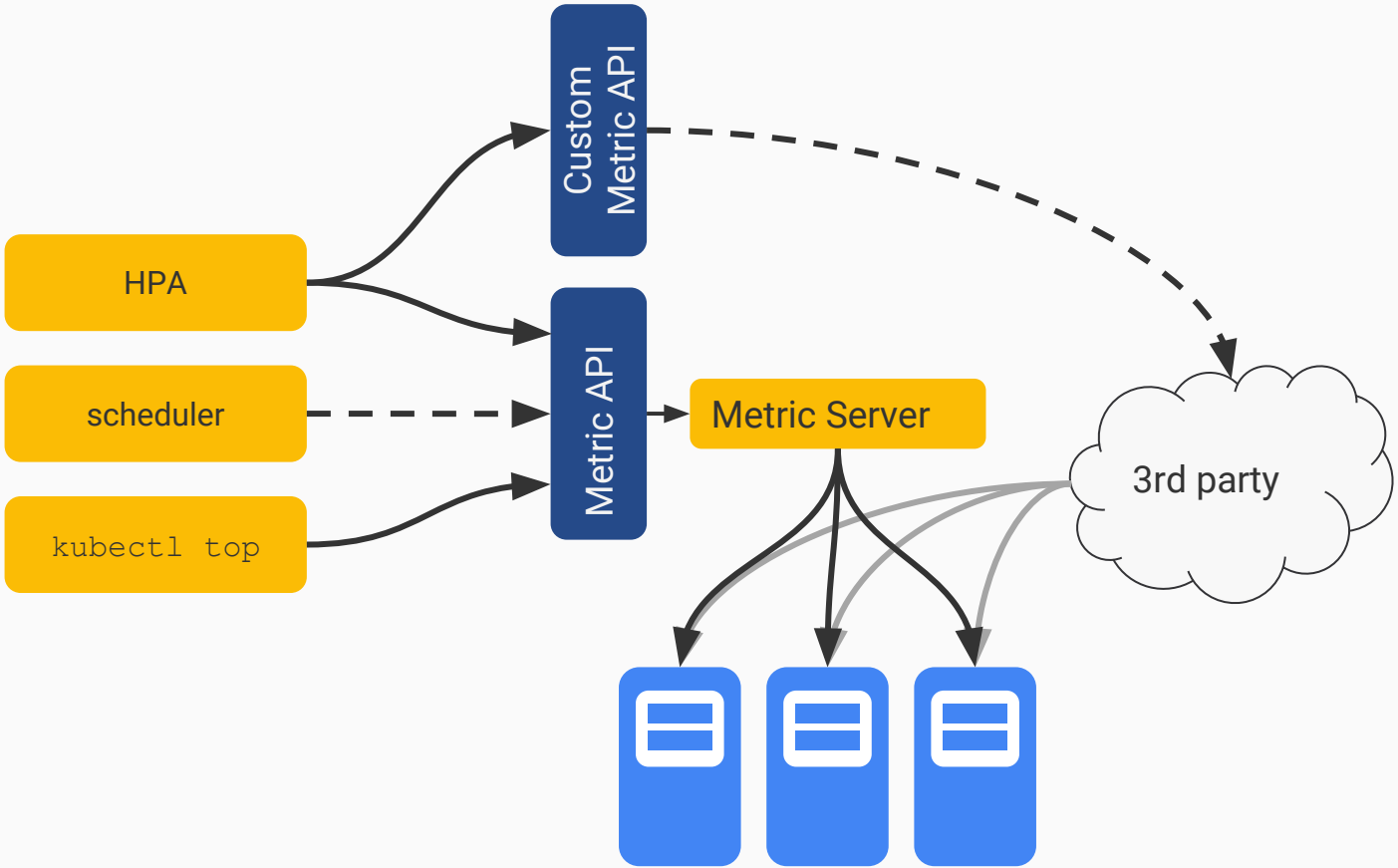


“Kubernetes is not about monitoring”

“Kubernetes is not about monitoring”
however ...


HPA
scheduler
kubectl top
Metric API
Metric Server
3rd party

Custom Metrics API



Thank you!

Questions?