

Monitoring *Cloud Native* applications with Prometheus

Aaron Kirkbride @ Weaveworks





Time Series Database

```
time_series_1 => [(t0, 0), (t1, 100), (t2, 150), (t3, 170), (t4, 300), ...]
```

```
time_series_2 => [(t0, 0), (t1, 0), (t2, 0), (t3, 4), (t4, 2), (t5, 2), ...]
```



- *Incident* at Weaveworks
- Integrations with **Kubernetes**



Monitoring

Alerts when there is a **user impact**



Weave Cloud APP Instance: Weave Cloud (prod)

Release all latest to cortex:deployment/alertmanager, cortex:deployment/distributor, and cortex:deployment/configs.

11:38 AM **prod-alert** APP

AuthFEErrorRate (FIRING)

default/authfe: high error rate -

Impact: Some or all of Weave Cloud is inaccessible to many users

Details: The authfe service has an error rate (response code ≥ 500) of 78.1%.

[Playbook](#) [Dashboard](#)

CortexDistributorErrorRate (FIRING)

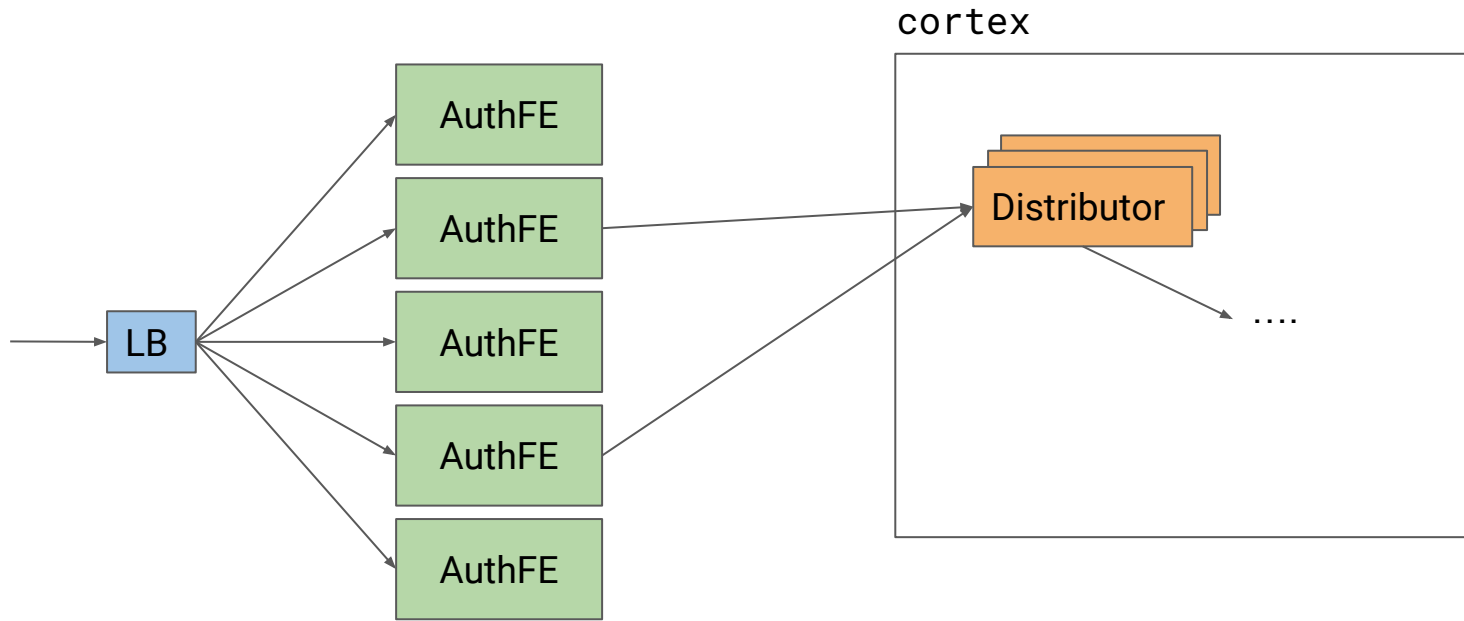
cortex/distributor high error rate -

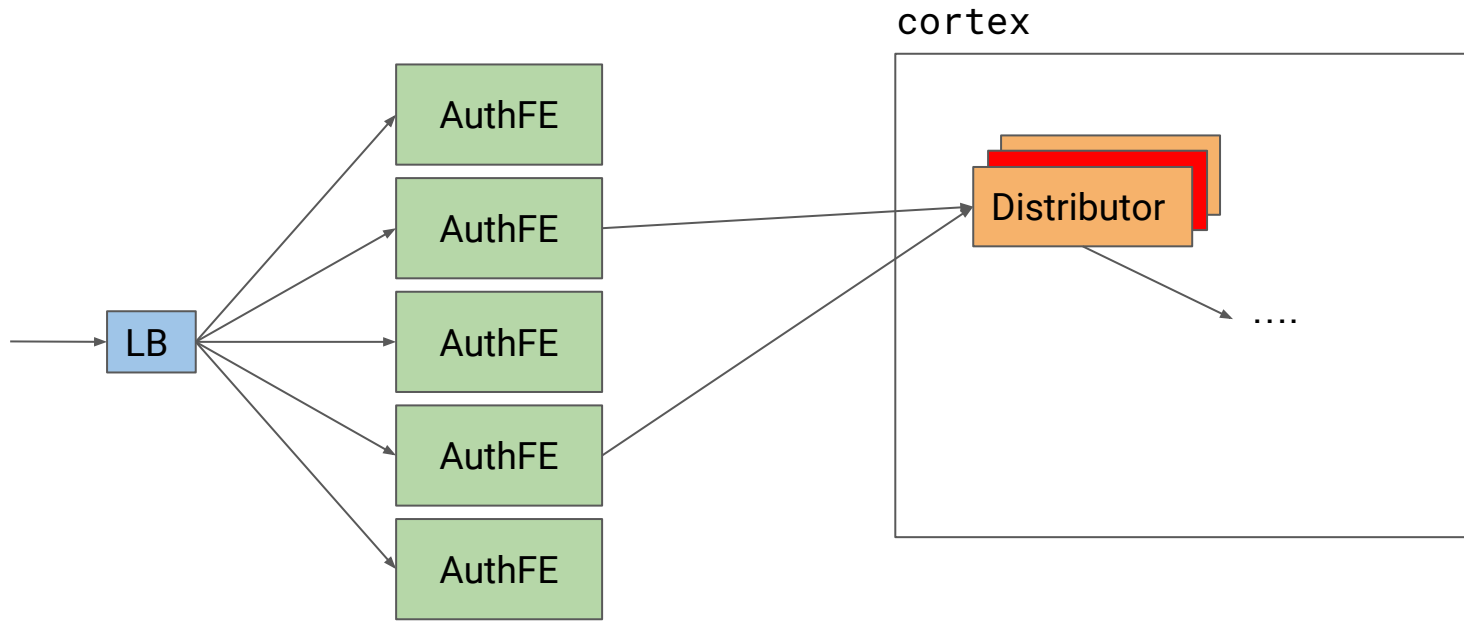
Impact: Dropping end-user data

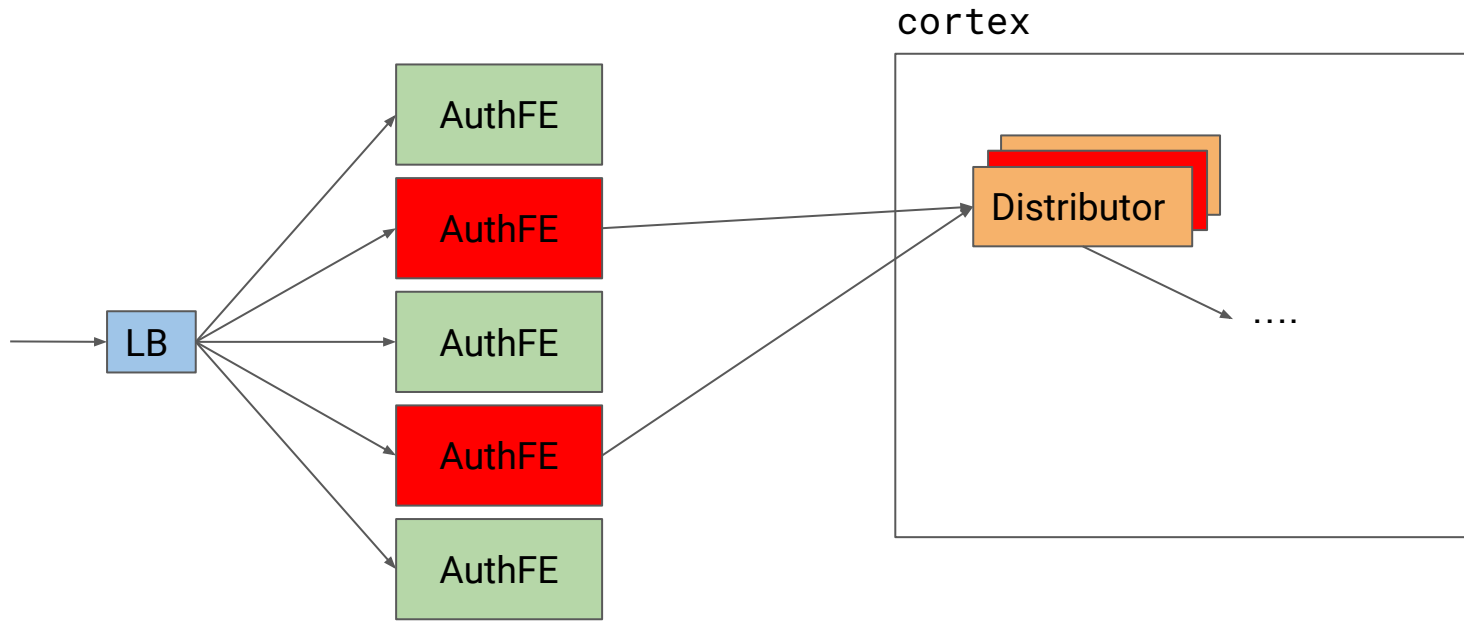
Details: Writes to Cortex are failing with error rate of 100%.

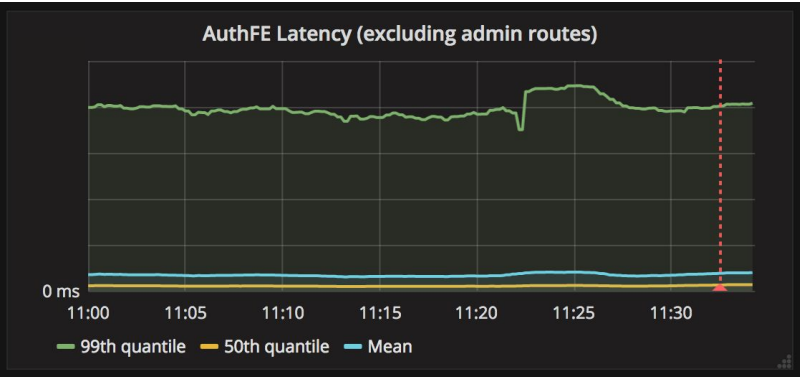
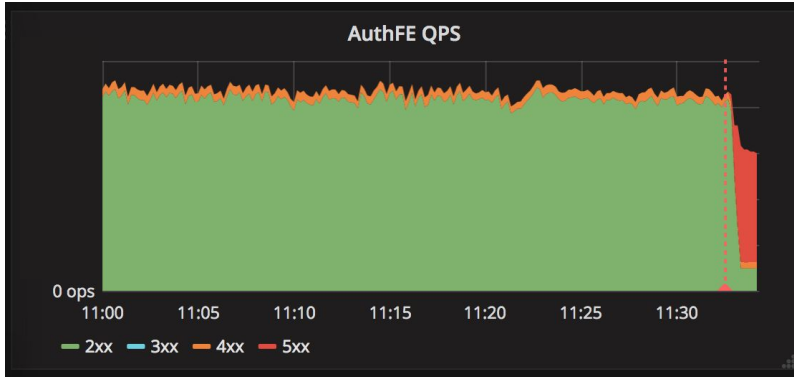
[Playbook](#) [Dashboard](#)



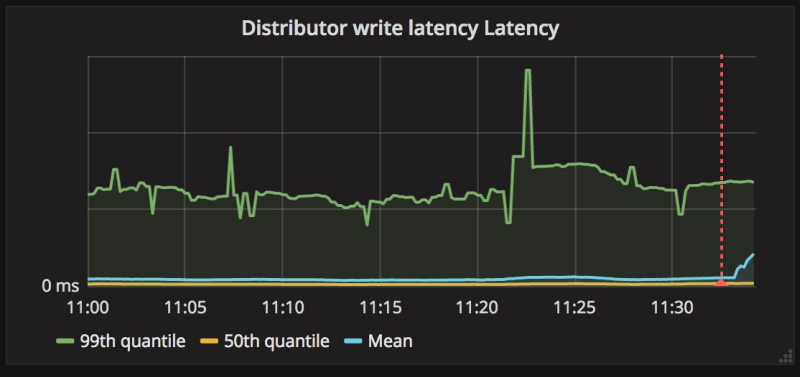
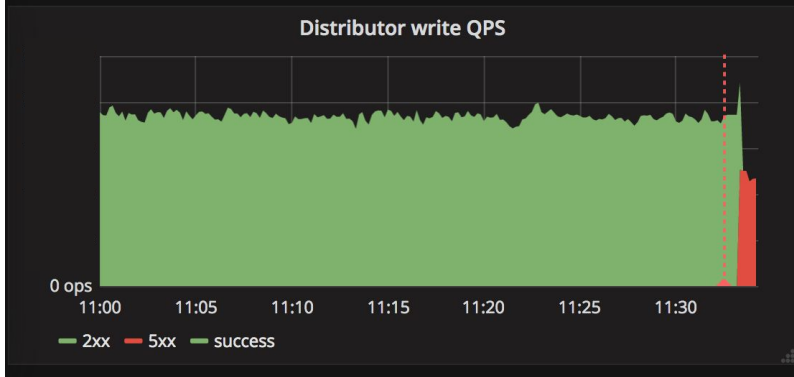


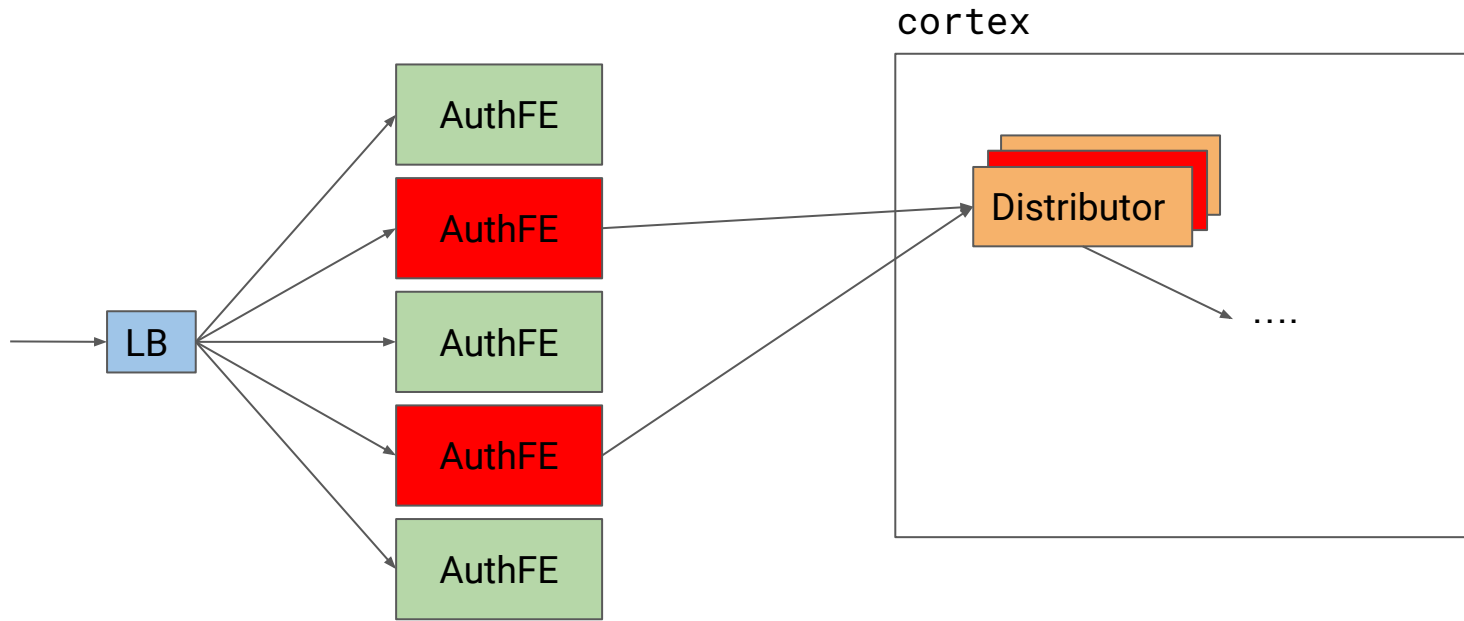






▼ Distributor





stats.default.authfe.get.200 => $(t_0, 2000)$, $(t_1, 4021)$, ...
stats.default.authfe.get.500 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 10)$, ...
stats.default.authfe.get.502 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 0)$, ...

stats.cortex.distributor.get.200 => $(t_0, 100)$, $(t_1, 120)$, ...
stats.cortex.distributor.get.500 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 10)$, ...
stats.cortex.distributor.get.502 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 0)$, ...



NS	Pod name	Ready	Status	Age
cortex	distributor-6476689b4d-54bt7	2/2	Running	2h
cortex	distributor-6476689b4d-6m49h	2/2	Running	2h
cortex	distributor-6476689b4d-9kkfw	2/2	Running	2h
cortex	distributor-6476689b4d-r4k94	2/2	Running	2h
cortex	distributor-6476689b4d-96w6g	2/2	Running	2h
cortex	distributor-6476689b4d-rckzb	2/2	Running	2h
cortex	distributor-6476689b4d-z4zsr	2/2	Running	2h
cortex	distributor-6476689b4d-88nxc	2/2	Running	5m
cortex	distributor-6476689b4d-9c54c	2/2	Running	3m
...



stats.default.authfe.get.200 => $(t_0, 2000)$, $(t_1, 4021)$, ...
stats.default.authfe.get.500 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 10)$, ...
stats.default.authfe.get.502 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 0)$, ...

stats.cortex.distributor.get.200 => $(t_0, 100)$, $(t_1, 120)$, ...
stats.cortex.distributor.get.500 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 10)$, ...
stats.cortex.distributor.get.502 => $(t_0, 0)$, $(t_1, 0)$, $(t_2, 0)$, ...



```
service_request_duration_seconds_count{
  method="GET",
  route="/push",
  status_code="500",
  job="cortex/distributor",
  instance="distributor-6476689b4d-9c54c",
  node="ip-172-20-2-91.ec2.internal",
}
```

} Labels - (key, value) pairs

=> (t0, 1028), (t1, 2060), (t2, 3094), ...



```
service_request_duration_seconds_count{status_code=~"5.."} 
```




```
rate(service_request_duration_seconds_count{status_code=~"5.."}[1m])
```



```
sum(rate(service_request_duration_seconds_count{status_code=~"5.."}[1m])) by (job)
```

job

default/authfe	(t0, 0), (t1, 0), (t2, 20), (t2, 18), (t2, 20), ...
cortex/distributor	(t0, 0), (t1, 0), (t2, 50), (t3, 54), (t2, 51), ...



```
sum(rate(service_request_duration_seconds_count{status_code=~"5.."}[1m])) by (job)
```

```
=> (t0, 0.05), (t1, 0.06), (t2, 0.05), (t3,0.07), (t4,0.07), ...
```



Evaluate every *n* seconds into a new time series



```
job:service_request_errors:rate1m
```



11:38 AM

prod-alert APP

AuthFEErrorRate (FIRING)

default/authfe: high error rate -

Impact: Some or all of Weave Cloud is inaccessible to many users

Details: The authfe service has an error rate (response code ≥ 500) of 78.1%.

[Playbook](#) [Dashboard](#)

CortexDistributorErrorRate (FIRING)

cortex/distributor high error rate -

Impact: Dropping end-user data

Details: Writes to Cortex are failing with error rate of 100%.

[Playbook](#) [Dashboard](#)



Derived timeseries for fast querying

```
- alert: AuthFEErrorRate
  expr: job:service_request_errors_percent:rate1m{job="default/authfe"} > 0
  for: 1m
  labels:
    severity: critical
  annotations:
    summary: 'default/authfe: high error rate'
    description: The authfe service has an error rate (response code >= 500) of
      {{$value | printf "%.1f"}}%.
    impact: Some or all of Weave Cloud is inaccessible to many users
    playbookURL: ...
    dashboardURL: ...
```

Condition

prod-alert APP

AuthFEErrorRate (RESOLVED)

default/authfe: high error rate -

Impact: Some or all of Weave Cloud is inaccessible to many users

Details: The authfe service has an error rate (response code ≥ 500) of 0.781 errors per second.

[Playbook](#) [Dashboard](#)

CortexDistributorErrorRate (RESOLVED)

cortex/distributor high error rate -

Impact: Dropping end-user data

Details: Writes to Cortex are failing with error rate of 1.000 per second for more than 5m

[Playbook](#) [Dashboard](#)



Instrumenting



common/middleware/instrument.go:

```
// RequestDuration is our standard histogram vector.
var RequestDuration = prometheus.NewHistogramVec(
    prometheus.HistogramOpts{
        Namespace: cfg.MetricsNamespace,
        Name:       "request_duration_seconds",
        Help:      "Time (in seconds) spent serving HTTP requests.",
    },
    []string{"method", "route", "status_code", "ws"} } Labels
)

func (i Instrument) Wrap(next http.Handler) http.Handler {
    return http.HandlerFunc(func(w http.ResponseWriter, r *http.Request) {
        ...
        RequestDuration.WithLabelValues(r.Method, route, status, isWS).Observe(took.Seconds())
    })
}
```


Key Metrics

- **R**ate - number of requests per second
- **E**rrors - number of those requests which are failing
- **D**uration - the amount of time those requests take

<https://www.weave.works/blog/the-red-method-key-metrics-for-microservices-architecture/>

<https://www.youtube.com/watch?v=TJLpYXbnfQ4>

<https://landing.google.com/sre/book/chapters/monitoring-distributed-systems.html>

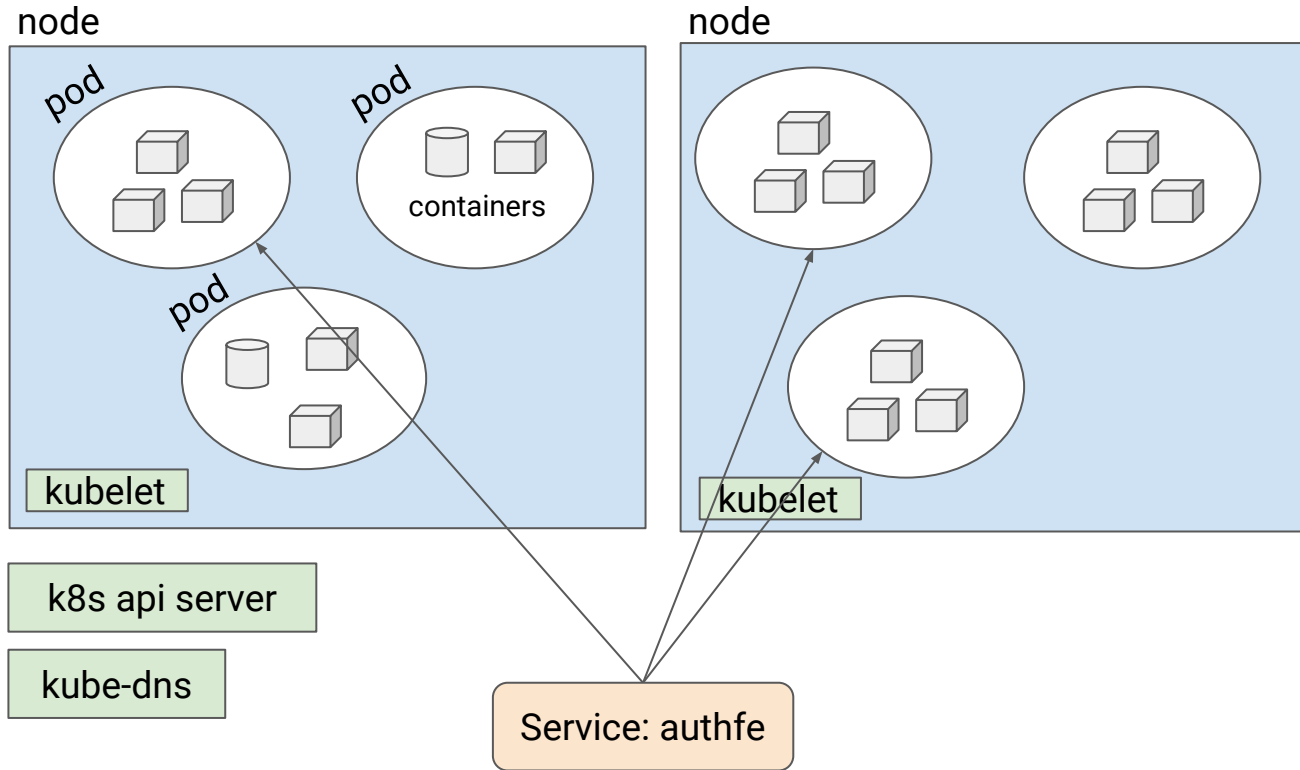


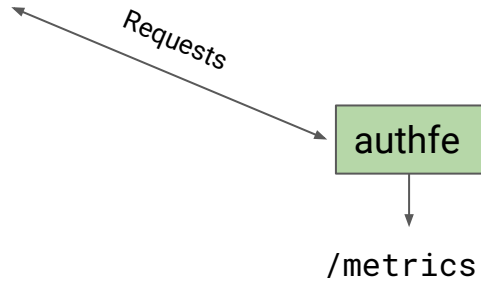
Kubernetes



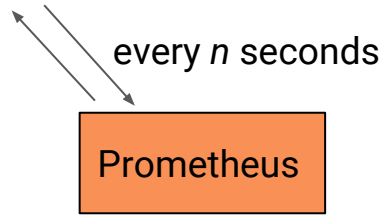


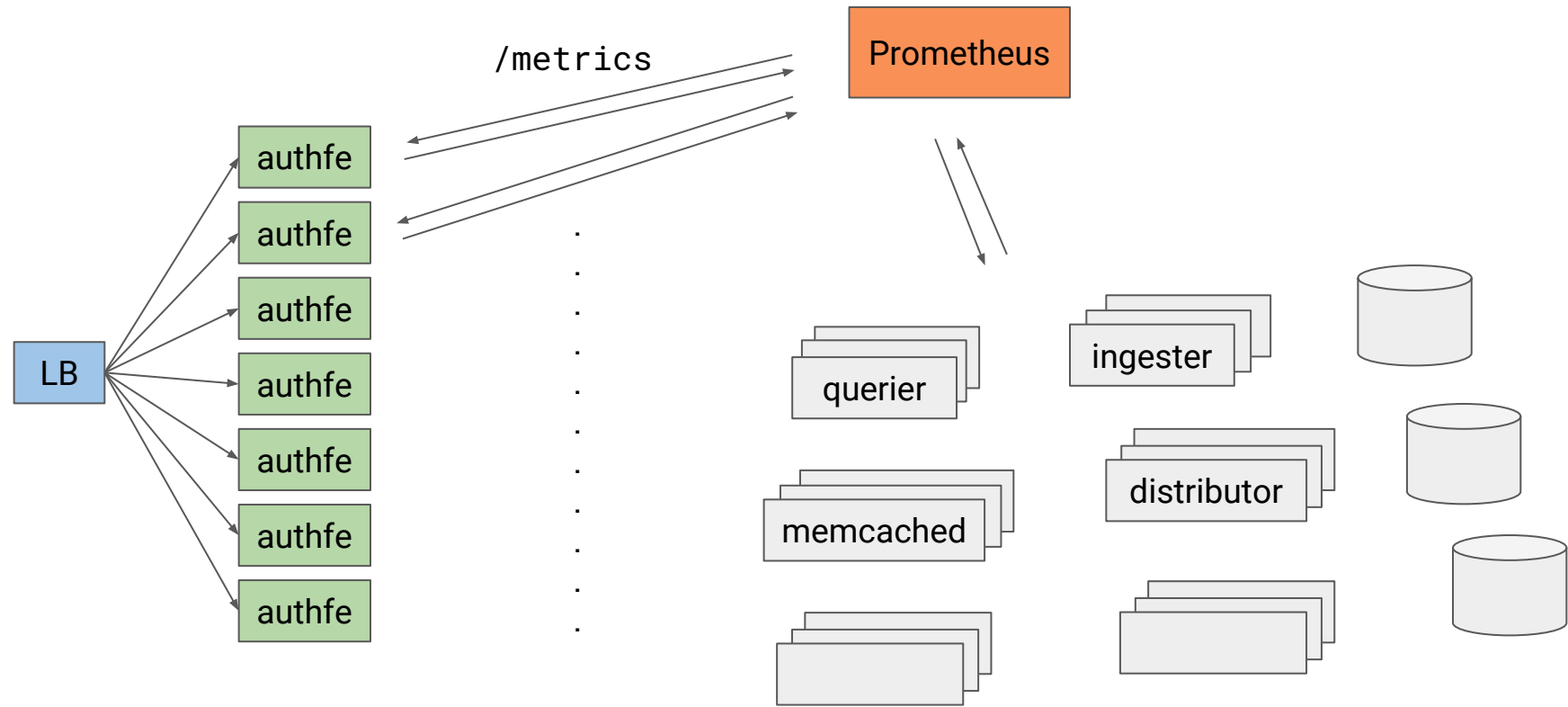
Kubernetes

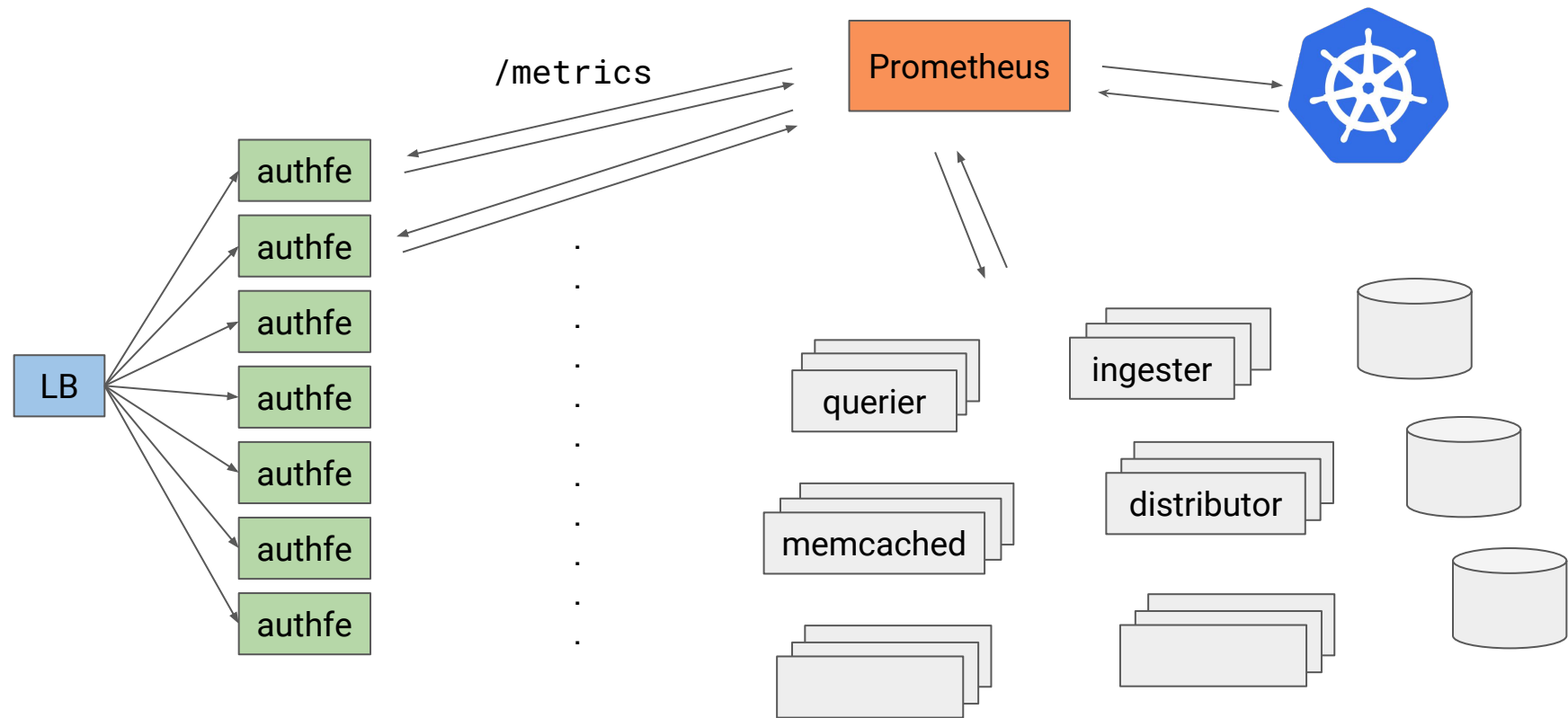




```
# HELP service_request_duration_seconds_count Time (in seconds) spent serving HTTP requests.  
# TYPE service_request_duration_seconds_count histogram  
service_request_duration_seconds_count{route="/prom/push", method="POST", ...} 120001  
service_request_duration_seconds_count{route="/users", method="GET", ...} 32001  
...
```







```
scrape_configs:
```

```
- job_name: kubernetes-pods
```

```
  kubernetes_sd_configs:
```

```
    - role: pod
```



default/authfe (4/4 up)

Endpoint	State	Labels	Last Scrape	Error
http://10.244.254.183:24231/metrics	UP	<code>_weave_service="authfe"</code> <code>instance="authfe-86888865c6-47pd5"</code> <code>kubernetes_namespace="default"</code> <code>kubernetes_pod_name="authfe-86888865c6-47pd5"</code> <code>node="ip-172-20-1-135.ec2.internal"</code>	9.084s ago	
http://10.244.254.183:8080/metrics	UP	<code>_weave_service="authfe"</code> <code>instance="authfe-86888865c6-47pd5"</code> <code>kubernetes_namespace="default"</code> <code>kubernetes_pod_name="authfe-86888865c6-47pd5"</code> <code>node="ip-172-20-1-135.ec2.internal"</code>	3.199s ago	
http://10.244.204.141:24231/metrics	UP	<code>_weave_service="authfe"</code> <code>instance="authfe-86888865c6-bl1nnn"</code> <code>kubernetes_namespace="default"</code> <code>kubernetes_pod_name="authfe-86888865c6-bl1nnn"</code> <code>node="ip-172-20-3-30.ec2.internal"</code>	4.792s ago	
http://10.244.204.141:8080/metrics	UP	<code>_weave_service="authfe"</code> <code>instance="authfe-86888865c6-bl1nnn"</code> <code>kubernetes_namespace="default"</code> <code>kubernetes_pod_name="authfe-86888865c6-bl1nnn"</code> <code>node="ip-172-20-3-30.ec2.internal"</code>	7.811s ago	

cortex/distributor (4/4 up)

Endpoint	State	Labels	Last Scrape	Error
http://10.244.229.112:80/metrics	UP	<code>_weave_service="distributor"</code> <code>instance="distributor-5f76f588f-7fvsw"</code> <code>kubernetes_namespace="cortex"</code> <code>kubernetes_pod_name="distributor-5f76f588f-7fvsw"</code> <code>node="ip-172-20-2-209.ec2.internal"</code>	1.666s ago	
http://10.244.229.112:24231/metrics	UP	<code>_weave_service="distributor"</code> <code>instance="distributor-5f76f588f-7fvsw"</code> <code>kubernetes_namespace="cortex"</code> <code>kubernetes_pod_name="distributor-5f76f588f-7fvsw"</code> <code>node="ip-172-20-2-209.ec2.internal"</code>	8.087s ago	
http://10.244.254.177:80/metrics	UP	<code>_weave_service="distributor"</code> <code>instance="distributor-5f76f588f-mkwr1"</code> <code>kubernetes_namespace="cortex"</code> <code>kubernetes_pod_name="distributor-5f76f588f-mkwr1"</code> <code>node="ip-172-20-1-135.ec2.internal"</code>	8.522s ago	
http://10.244.254.177:24231/metrics	UP	<code>_weave_service="distributor"</code> <code>instance="distributor-5f76f588f-mkwr1"</code> <code>kubernetes_namespace="cortex"</code> <code>kubernetes_pod_name="distributor-5f76f588f-mkwr1"</code> <code>node="ip-172-20-1-135.ec2.internal"</code>	10.08s ago	



cortex/distributor



/metrics

```
# HELP service_request_duration_seconds_count Time (in seconds) spent serving HTTP
requests.
# TYPE service_request_duration_seconds_count histogram
service_request_duration_seconds_count{route="/prom/push", method="POST", ...} 120001
service_request_duration_seconds_count{route="/users", method="GET", ...} 32001
...
```



Prometheus



```
service_request_duration_seconds_count{
  route="/prom/push",
  method="POST",
  job="cortex/distributor",
  instance="distributor-6476689b4d-9c54c",
  node="ip-172-20-2-91.ec2.internal"
} 120001
```



```
scrape_configs:  
- job_name: 'kubernetes-pods'  
  kubernetes_sd_configs:  
  - role: pod  
    relabel_configs:  
  - source_labels: [__meta_kubernetes_pod_name]  
    action: replace  
    target_label: instance
```



```
service_request_duration_seconds_count{  
  route="/prom/push",  
  method="POST",  
  instance="distributor-6476689b4d-9c54c",  
} 120001
```



```
- source_labels: [__meta_kubernetes_namespace, __meta_kubernetes_pod_label_name]
  action: replace
  separator: /
  target_label: job
  replacement: $1
```



```
service_request_duration_seconds_count{
  route="/prom/push",
  method="POST",
  job="cortex/distributor",
} 120001
```



K8s Deployment config:

```
    annotations:  
      prometheus.io.scrape: false
```

- source_labels: [__meta_kubernetes_pod_annotation_prometheus_io_scrape]
 action: drop
 regex: false



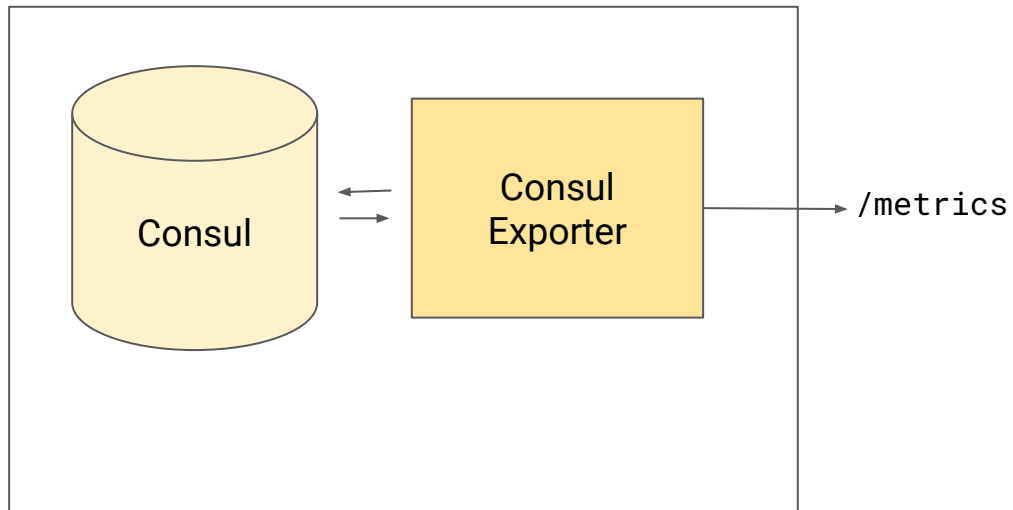
K8s Deployment config:

```
annotations:  
  prometheus.io.port: 5678
```

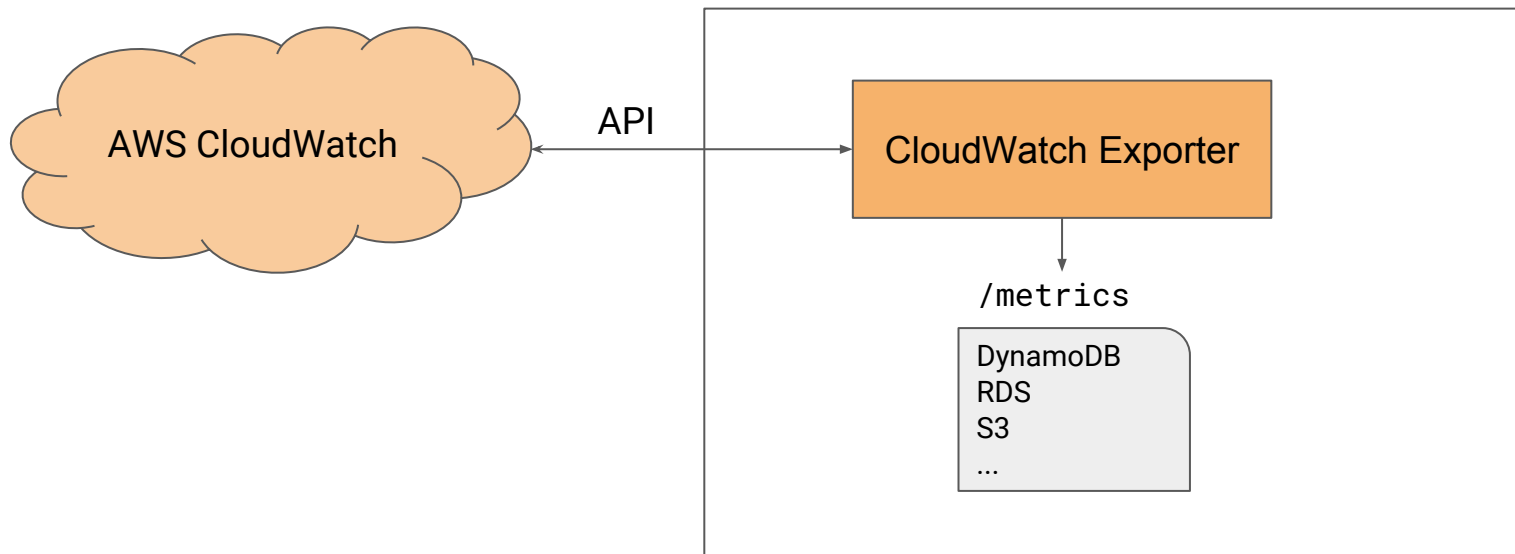
```
- source_labels: [__address__, __meta_kubernetes_pod_annotation_prometheus_io_port]  
  action: replace  
  target_label: __address__  
  regex: "(.+?)(\\:\\\\d+)?;(\\d+)"  
  replacement: $1:$3
```



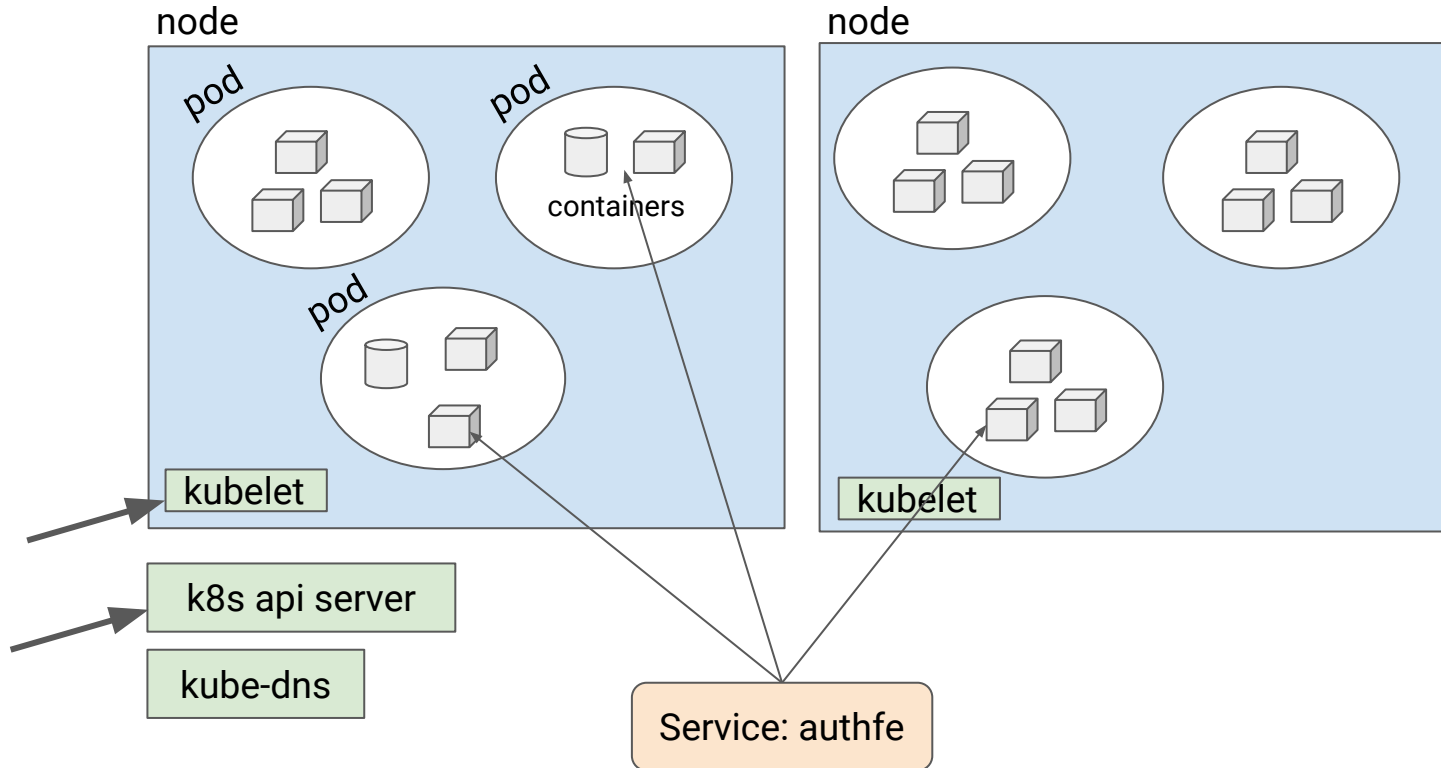
Exporters



<https://prometheus.io/docs/instrumenting/exporters/>

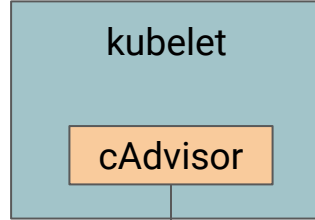


Kubernetes



kube-state-metrics





/metrics/cadvisor



```
- job_name: cadvisor
  kubernetes_sd_configs:
  - role: node
    metrics_path: /metrics/cadvisor
    scheme: https
    tls_config:
      ca_file: /var/run/secrets/kubernetes.io/serviceaccount/ca.crt
      insecure_skip_verify: true
```



`container_cpu_usage_seconds_total`

`container_memory_usage_bytes`




```
namespace_label_name:container_cpu_usage_seconds_total:sum_rate =>
```

```
sum by (namespace, label_name) (  
    label_replace(  
        sum(rate(container_cpu_usage_seconds_total{image!=""}[5m])) by (pod_name, namespace),  
        "pod",  
        "$1",  
        "pod_name",  
        "(.*)"  
    )  
    * on (pod) group_right(pod_name) kube_pod_labels{job="weave/kube-state-metrics"}  
)
```

```
namespace_label_name:container_memory_usage_bytes:sum =>
```

```
sum by (namespace, label_name) (  
    label_replace(  
        sum(container_memory_usage_bytes{image!=""}) by (pod_name, namespace),  
        "pod",  
        "$1",  
        "pod_name",  
        "(.*)"  
    )  
    * on (pod) group_right(pod_name) kube_pod_labels{job="weave/kube-state-metrics"}  
)
```

Takeaways

- System based metrics and **alerting**
- Flexible data model
- Perfect match with **Kubernetes**



Thanks!

<https://prometheus.io>

Thursday Workshop (9:00am - 12pm):
Mastering Microservices Monitoring with Prometheus
By: Brice Fernandes & Ilya Dmitrichenko



We're hiring in London, Berlin and SF!

<https://weave.works/hiring>