



Declarative Multi Cluster Monitoring

Matthias Loibl - @metalmatze - Loodse
Frederic Branczyk - @fredbrancz - Red Hat

Prometheus Intro



Quick Prometheus intro

HTTP Get /

Application
Example Metric:
Request Count

0

HTTP Get /metrics
every 15s

Prometheus

Application's
Request Count

Time	Value
T0	0
T1	2

...



What's a target?

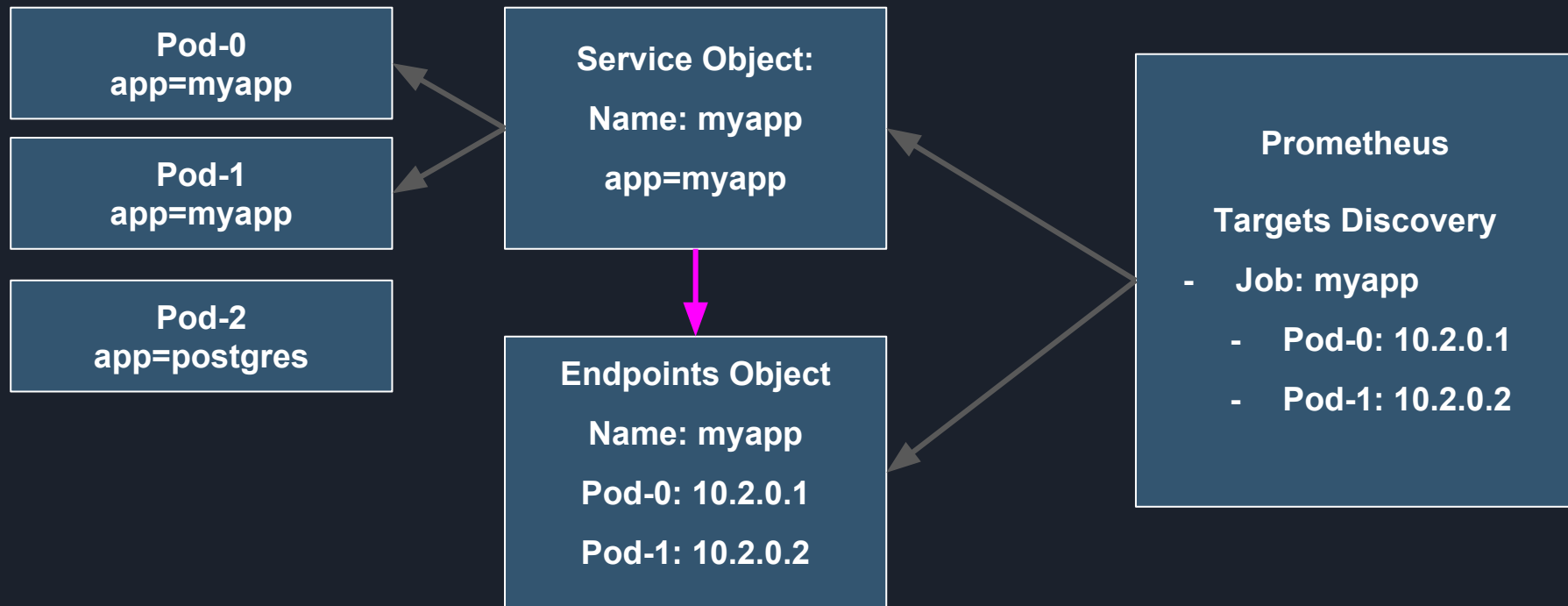
- HTTP Server with /metrics endpoint
- Discovered by a SD mechanism
 - Static target list
 - DNS discovery
 - Kubernetes discovery



Kubernetes Discovery

- Targets to discover
 - Pods
 - Nodes
 - Endpoints/Services
- Automatically reconfigure
 - Add, update, remove

Discovery via Kubernetes Services



Target overview

Pods

Services/Endpoints

frontend-0

frontend-1

api-0

api-1

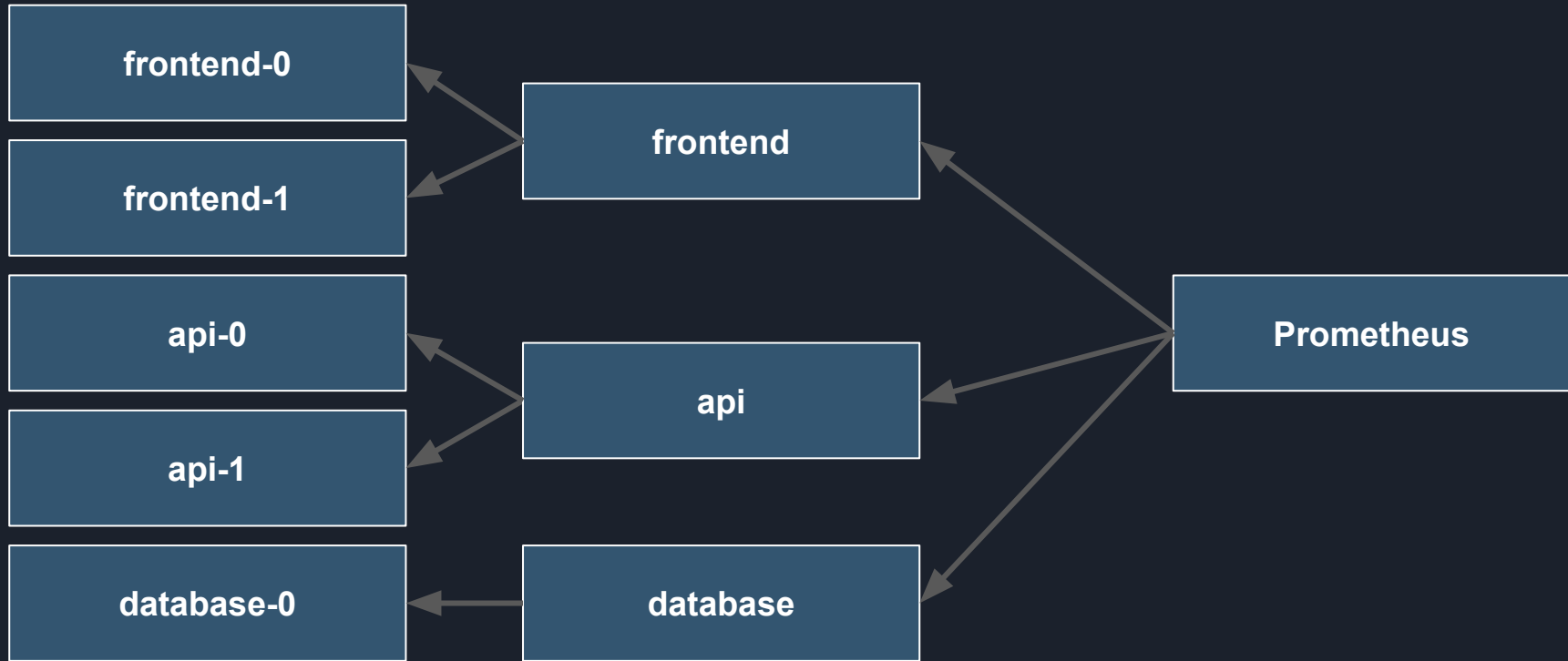
database-0

frontend

api

database

Prometheus



A Kubernetes config (1/n)

```
global:
  scrape_interval: 30s
  scrape_timeout: 10s
  evaluation_interval: 30s
alerting:
  alertmanagers:
  - kubernetes_sd_configs:
    - api_server: null
      role: endpoints
      namespaces:
        names:
        - tectonic-system
  scheme: http
  path_prefix: /
  timeout: 10s
  relabel_configs:
```


A Kubernetes config (2/n)

```
- source_labels: [__meta_kubernetes_service_name]
  separator: ;
  regex: alertmanager-main
  replacement: $1
  action: keep
- source_labels: [__meta_kubernetes_endpoint_port_name]
  separator: ;
  regex: web
  replacement: $1
  action: keep
rule_files:
- /etc/prometheus/rules/rules-0/*.rules
scrape_configs:
- job_name: tectonic-system/alertmanager/0
  scrape_interval: 30s
  scrape_timeout: 10s
```

A Kubernetes config (3/n)

```
metrics_path: /metrics
scheme: http
kubernetes_sd_configs:
- api_server: null
  role: endpoints
  namespaces:
    names:
    - tectonic-system
relabel_configs:
- source_labels: [__meta_kubernetes_service_label_alertmanager]
  separator: ;
  regex: main
  replacement: $1
  action: keep
- source_labels: [__meta_kubernetes_service_label_k8s_app]
  separator: ;
```

A Kubernetes config (4/n)

```
- source_labels: [__meta_kubernetes_service_name]
  separator: ;
  regex: alertmanager-main
  replacement: $1
  action: keep
- source_labels: [__meta_kubernetes_endpoint_port_name]
  separator: ;
  regex: web
  replacement: $1
  action: keep
rule_files:
- /etc/prometheus/rules/rules-0/*.rules
scrape_configs:
- job_name: tectonic-system/alertmanager/0
  scrape_interval: 30s
  scrape_timeout: 10s
```

A Kubernetes config (5/n)

```
regex: alertmanager
replacement: $1
action: keep
- source_labels: [__meta_kubernetes_endpoint_port_name]
  separator: ;
  regex: web
  replacement: $1
  action: keep
- source_labels: [__meta_kubernetes_namespace]
  separator: ;
  regex: (.*)
  target_label: namespace
  replacement: $1
  action: replace
- source_labels: [__meta_kubernetes_pod_name]
  separator: ;
```



**That was 66 out of
613 Lines of config**



**There has got to be a
better way!**

Enter

Prometheus Operator



ServiceMonitor

Pods

Services / Endpoints

frontend-0

frontend-1

api-0

api-1

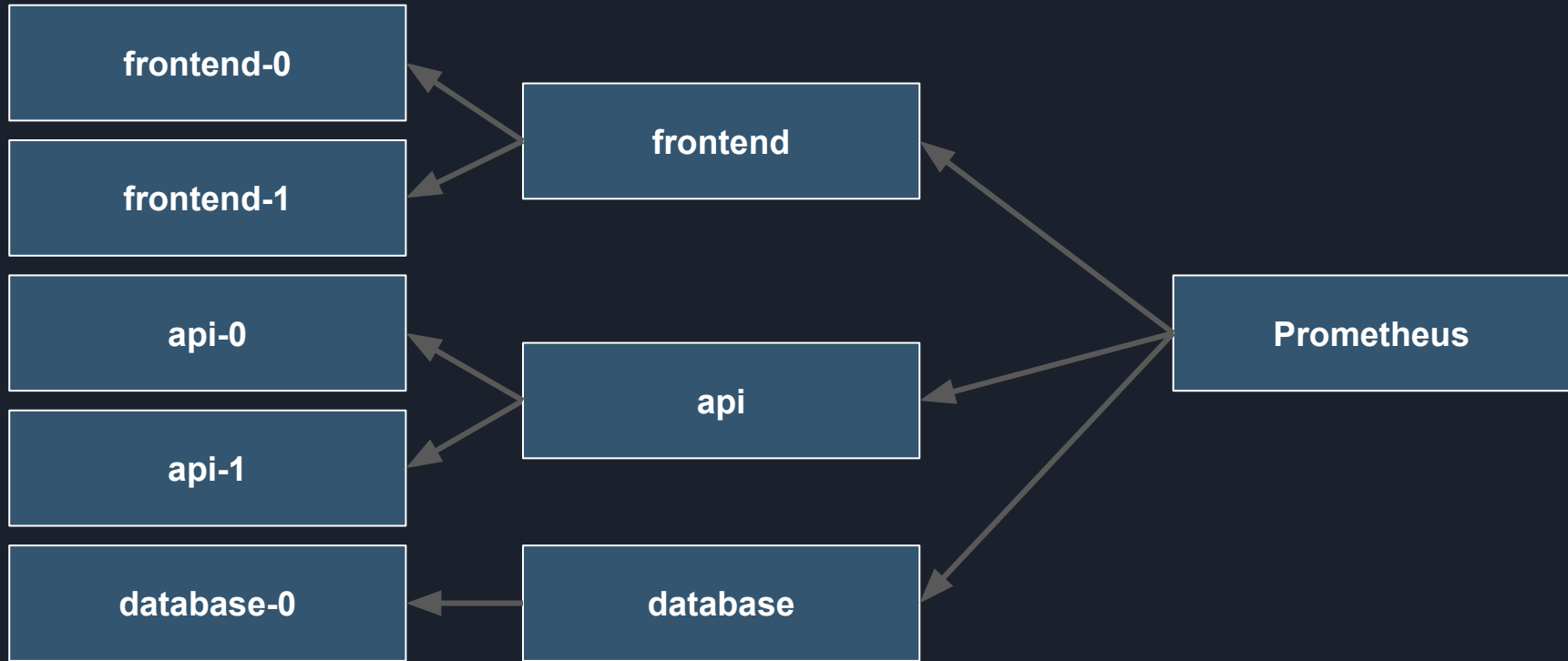
database-0

frontend

api

database

Prometheus



ServiceMonitor

Pods

Services/Endpoints

ServiceMonitor

frontend-0

frontend-1

api-0

api-1

database-0

frontend

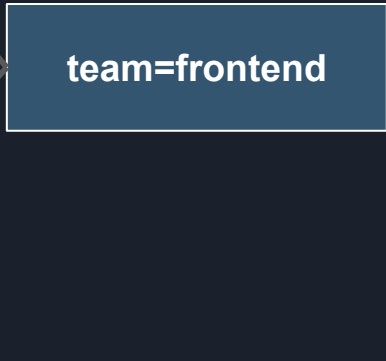
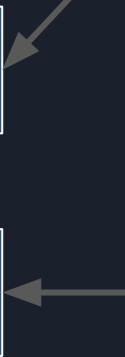
api

database

team=frontend

team=infra

Prometheus





Kubernetes native configuration

- **Prometheus is environment agnostic**
 - **We know we're running in Kubernetes**
- **Complicated configuration paradigms**
- **Abstractions!**



Prometheus Operator

- github.com/coreos/prometheus-operator
- **Kubernetes native objects**
 - **Operating Prometheus / complex stateful apps in code**
- **Graceful upgrades, migrations, operational knowledge**
- **1.4.x, 1.5.x, 1.6.x, 1.7.x, 1.8.x, 2.1.x, 2.2.x**
- **Kubernetes native configuration**



Declarative everything

- Declarative Kubernetes APIs
- Declarative Target configuration with logical grouping
- Declarative Alerting configuration
- Declarative ... everything

Declarative Grafana Dashboards





Declarative Grafana Dashboards

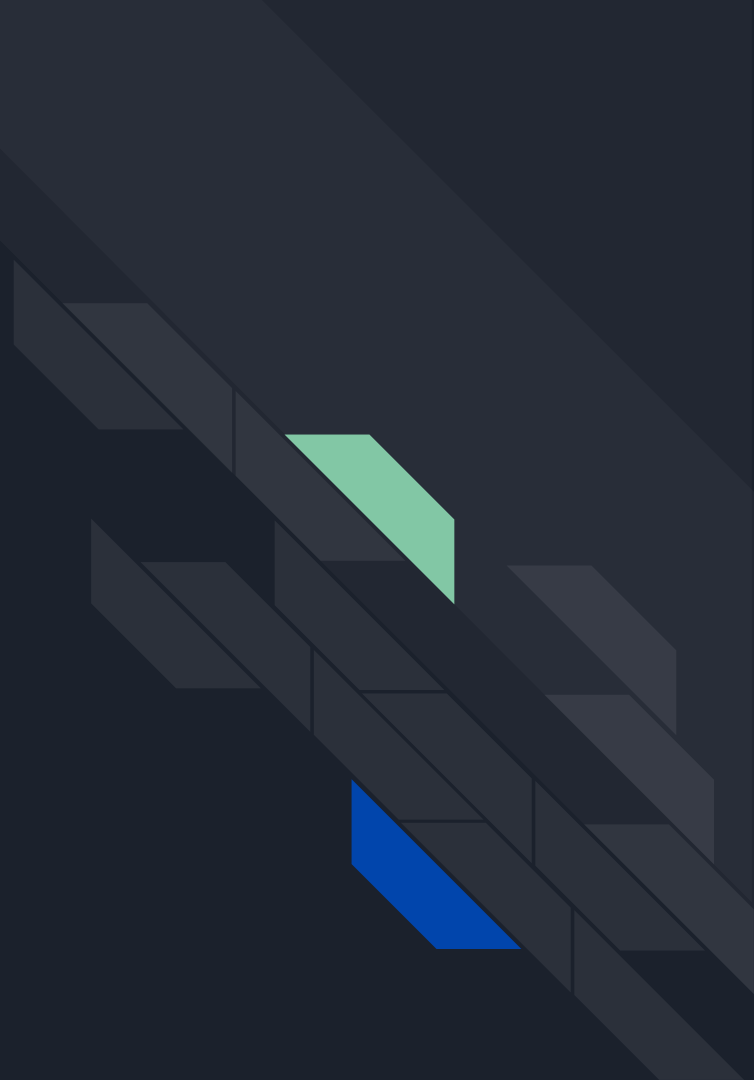
Grafana v5 supports provisioning with files

Use Jsonnet - Tom Wilkie showed that in an earlier talk

github.com/brancz/kubernetes-grafana

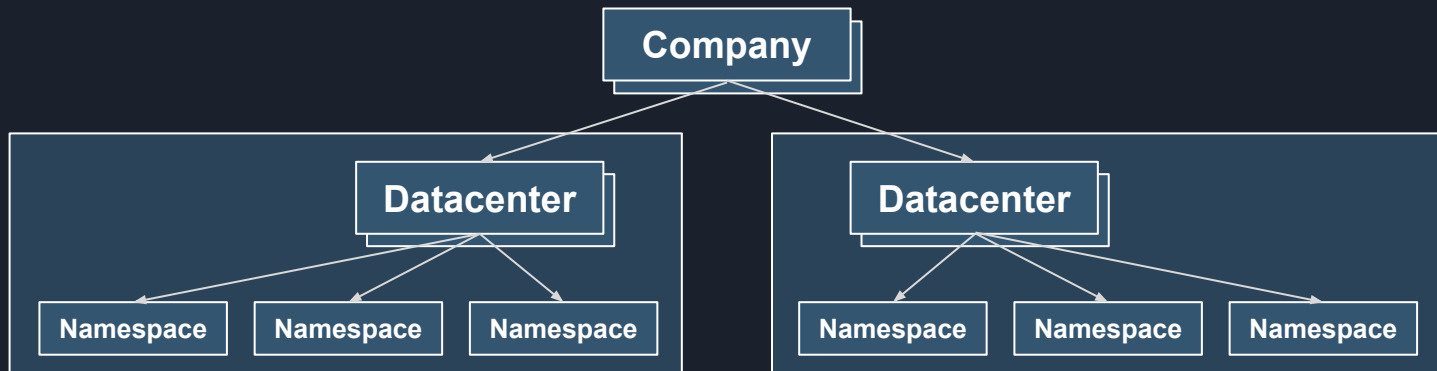
TODO Graph Panel inside Dashboard as jsonnet example

Multi-Cluster Monitoring



Federation

- Federation allows a Prometheus server to scrape selected time series from another Prometheus server.
- High retention at the root, with only very specific metrics





Prometheus per Namespace

- Deployed automatically by our API
- Scrapes all Kubernetes master components in its namespace
- Send alerts to a datacenter Alertmanager



Prometheus per Cluster

- Deploy manually at cluster creation
- Federates all Namespace Prometheus with one single ServiceMonitor

ServiceMonitor

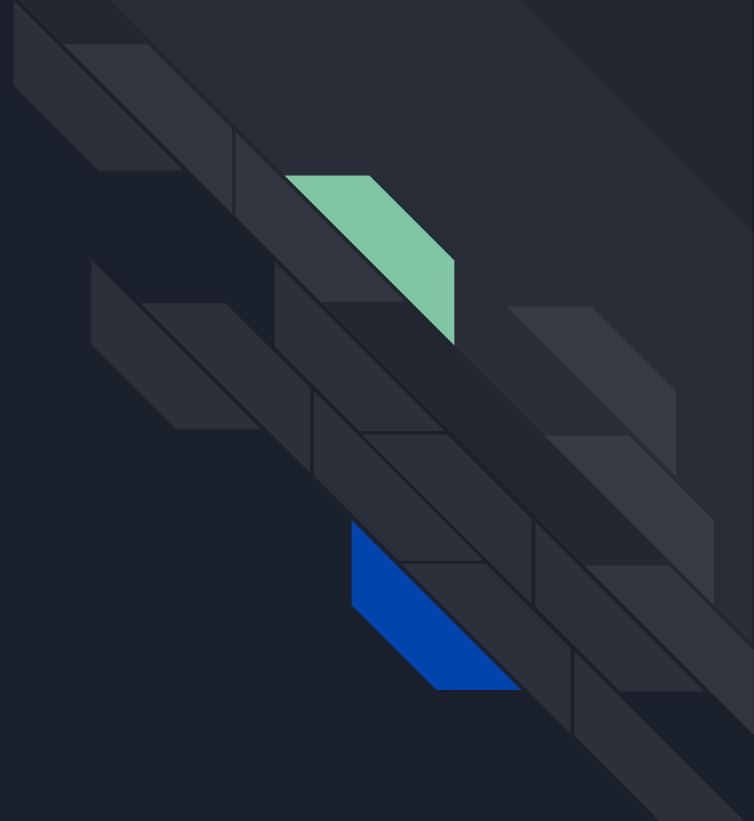
```
apiVersion: monitoring.coreos.com/v1
kind: ServiceMonitor
metadata:
  name: clusters
  labels:
    team: kubermatic
spec:
  selector:
    matchLabels:
    cluster: user
  namespaceSelector:
    any: true
  endpoints:
    - port: web
      interval: 30s
      path: /federate
      honorLabels: true
  params:
    'match[]':
      - '{__name__=~"machine_controller.*"}'
```



Prometheus per Company

- Scrapes all Prometheus in different cluster / data centers
- High retention
- Very few but specific metrics
- Mostly for Dashboarding
- Useful for global SLAs/ SLOs/ SLIs

Short Multi-Cluster Demo





Drawbacks with current Federation

- Need to use Sticky Sessions...
- Scraping multiple replicas of Prometheus per Cluster?

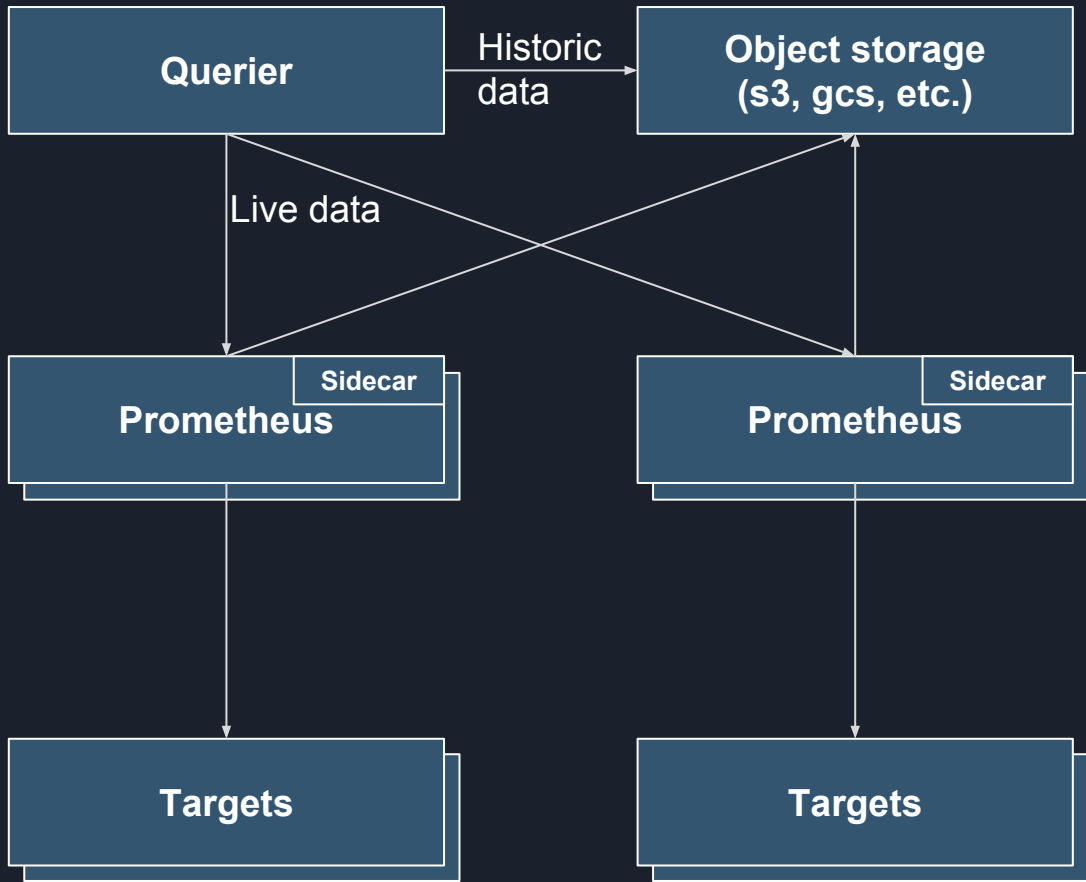
The Future





Thanos

- Long term storage
- Global view of data at real time
- Downsampling
- Builds on Prometheus 2.0 storage engine
 - Mapped into memory



Thanos support in Prometheus Operator


coreos / [prometheus-operator](#) [View Repository](#) [Unwatch](#) 78 [Unstar](#) 1,060 [Fork](#) 455

[Code](#) [Issues](#) 93 [Pull requests](#) 18 [Projects](#) 0 [Wiki](#) [Insights](#) [Settings](#)

Enable Thanos sidecar #1219 [Edit](#)

[Merged](#) mxinden merged 4 commits into `master` from `thanos4` 3 days ago

[Conversation](#) 2 [Commits](#) 4 [Files changed](#) 14 +336 -87



mxinden commented 3 days ago Member




prometheus: enforce external labels

Add external labels `prometheus` and `prometheus_replica` via Pod env variables and Prometheus config expansion.

contrib: add Thanos example manifests

prometheus: Drop 'prometheus_replica' label on alerts

Reviewers

-  brancz ✓
-  ant31 ●
-  fabxc ●

Assignees

No one—assign yourself

Labels



Summary

- Everything declarative (Rules)
- Global view
- Long term storage

Thank you

The background features a series of dark grey, 3D-style rectangular blocks arranged in a descending staircase pattern from the top right towards the bottom left. Two blocks are highlighted with color: a light green block and a blue block, both positioned on the right side of the staircase.

Matthias Loibl - @metalmatze - Loodse

Frederic Branczyk - @fredbrancz - Red Hat