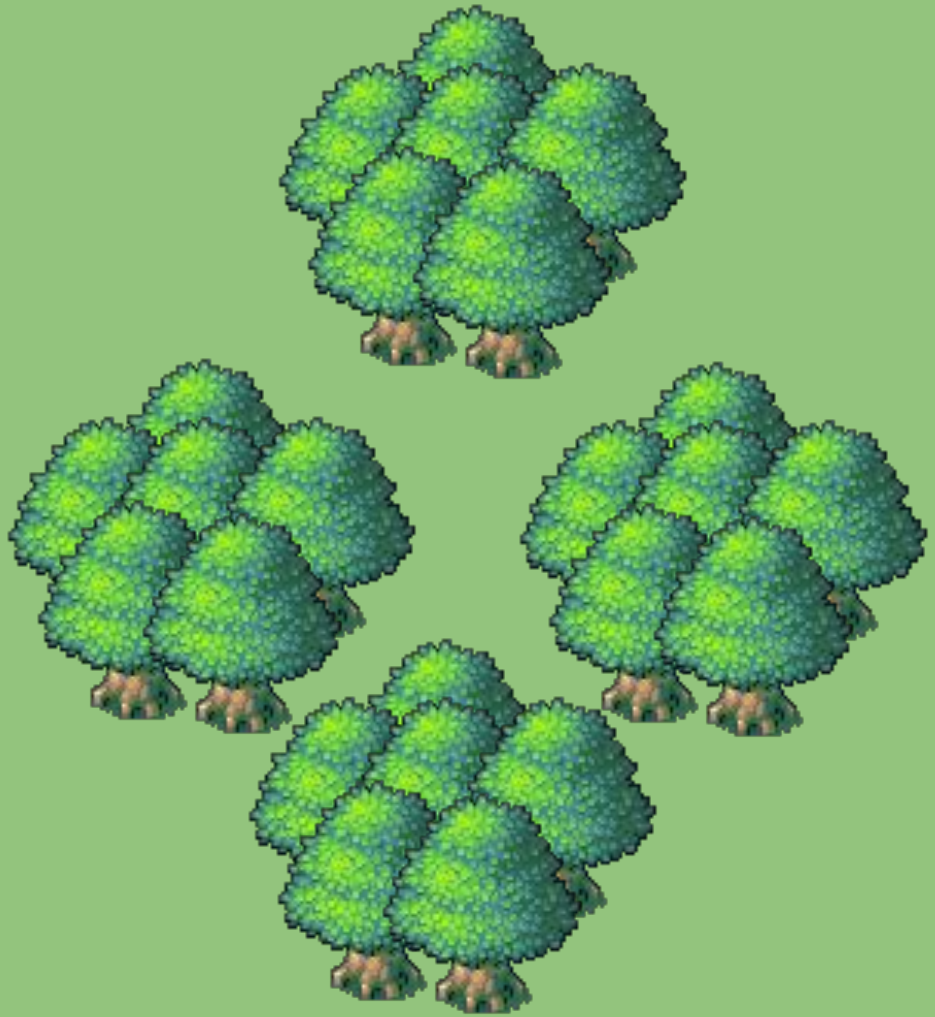
Building a Global Cross-Cloud Monitoring Platform

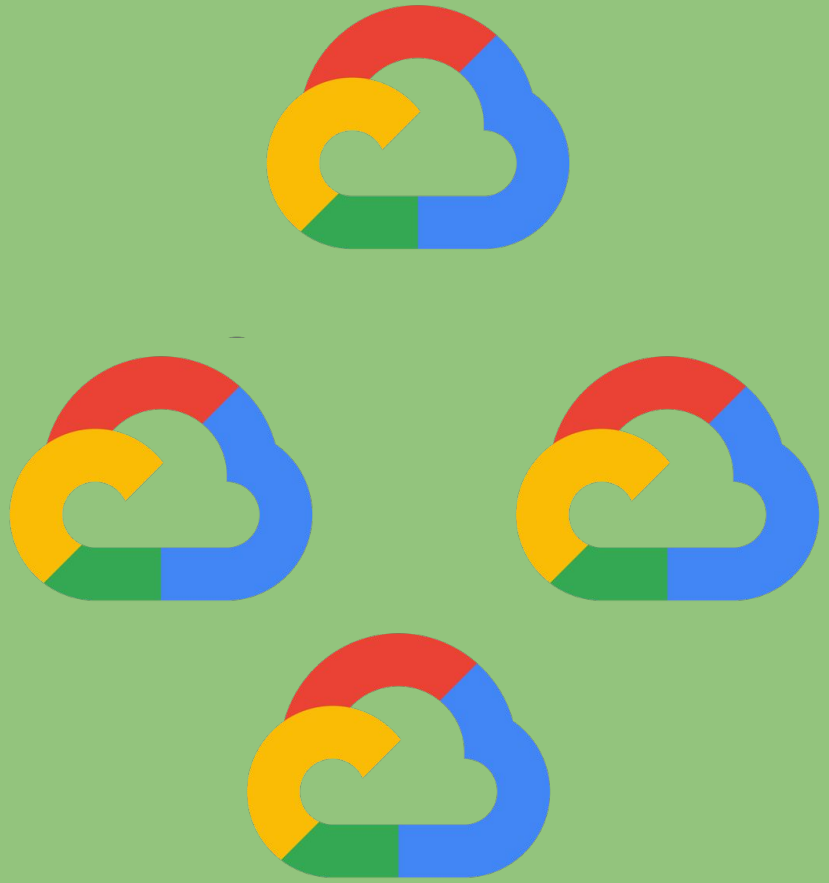
Dominic Green, Software Engineer
dom@improbable.io, @domgreen
25th June 2019, Shanghai China

Yifan Zhao, Software Engineer
yifan@improbable.io

If a tree falls in a forest and no one is around to hear it, does it make a sound?

如果一棵树落在森林里，周围没有人听到它，它会发出声音吗？





aws

Alibaba Cloud

Dominic Green

- Software Engineer @ Improbable
- Observability Team
- OSS Contributor
  - Thanos
  - go-grpc-middleware
  - go-httpwares
- Meetup Organiser
  - Prometheus London
  - London Gophers

Yifan Zhao

- Software Engineer @ Improbable
- Founder of Improbable China

Agenda

Single Cluster

Multi-Cluster

Multi-Cloud

IMPROBABLE

Our Mission: Make Impossible Games Possible

"Improbable's platform, SpatialOS, is designed to let anyone build massive simulations, running in the cloud: imagine Minecraft with thousands of players in the same space.... Its ultimate goal: to create totally immersive, persistent virtual worlds."

- WIRED, May 2017

Founded:
2012

Employees:
310

Games in Development:
+19

Define: Monitoring

> Collecting, processing, aggregating, and displaying real-time quantitative data about a system, such as query counts and types, error counts and types, processing times, and server lifetimes.

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

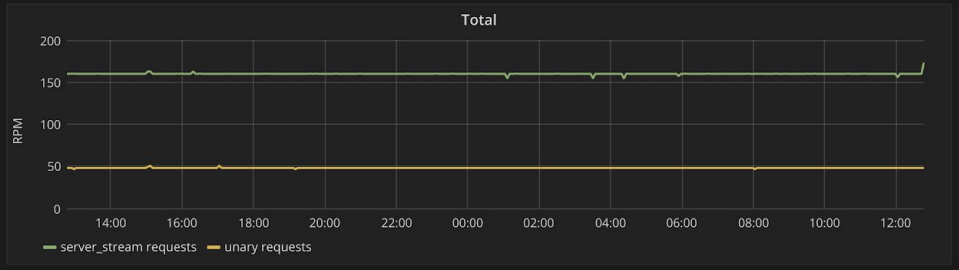
Generic gRPC Server

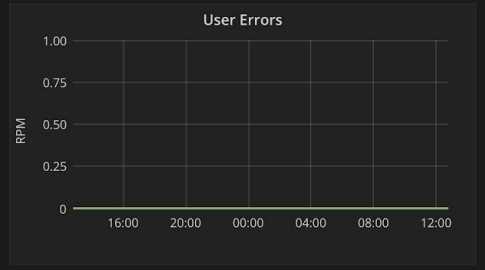
Zoom Out | Last 24 hours UTC

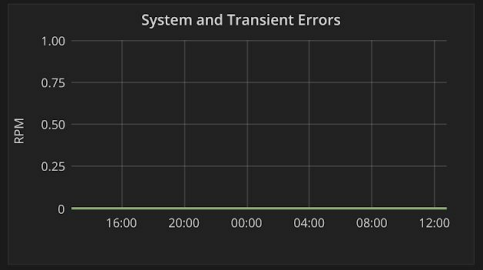
datasource: sanctuary | cluster: h-eu1 | job: thanos-mon-sidecar | env: production | service: All

thanos-mon-sidecar Dashboards | Kibana Logs

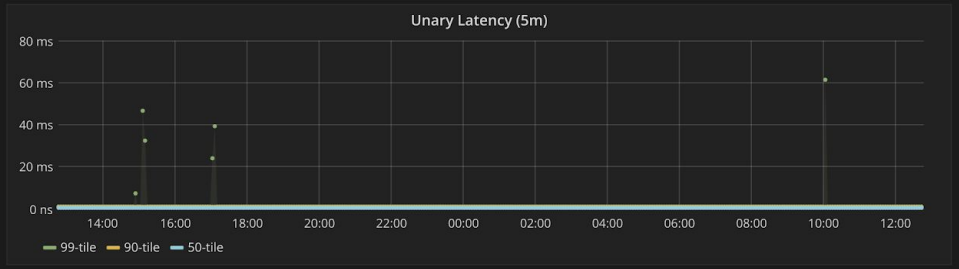
Requests (All)

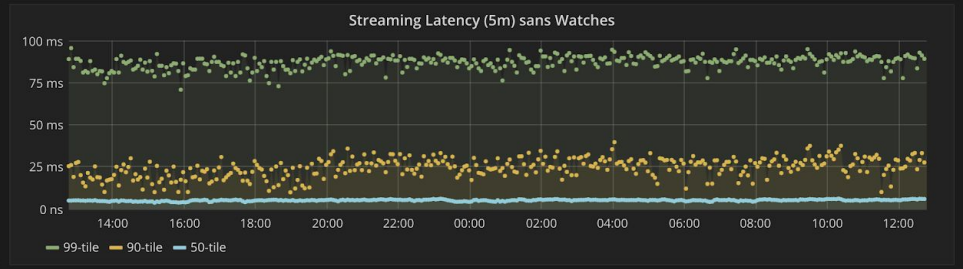






Latencies (All)






Prometheus

Prometheus

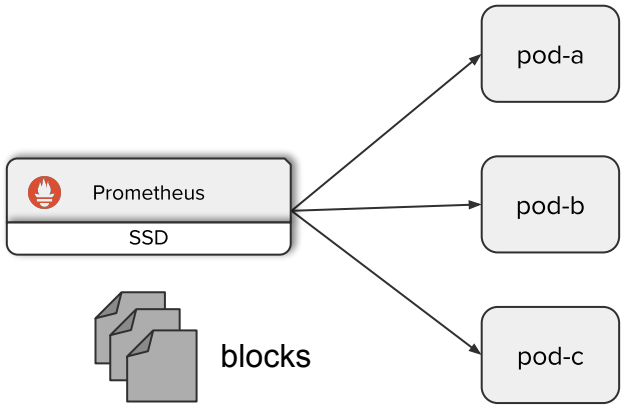
/metrics

```
# TYPE counter
app_request_total 1337

# TYPE guage
app_request_in_flight_total 3

# TYPE histogram
app_request_duration_ms_bucket {le="0.005"} 500
app_request_duration_ms_bucket {le="0.01"} 213
```

kubernetes

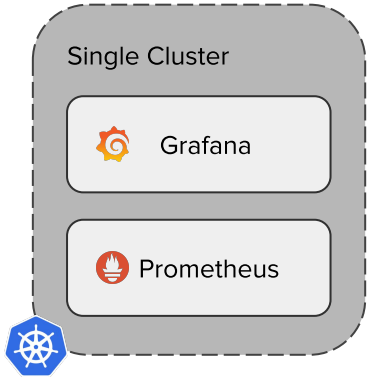
Prometheus

```
scrape_configs:
- job_name: 'kubernetes-pods'
  kubernetes_sd_configs:
  - role: pod
  relabel_configs:
  ...
```




```
apiVersion: v1
kind: Pod
metadata:
    annotations:
      prometheus.io/path: /metrics
      prometheus.io/port: "8080"
      prometheus.io/scheme: http
      prometheus.io/scrape: "true"
```

1 Single Cluster



1 Single Cluster

Kubernetes

- Basis for workload management
- Kubernetes Service Discovery for discovering workloads
- Mature tooling and automation

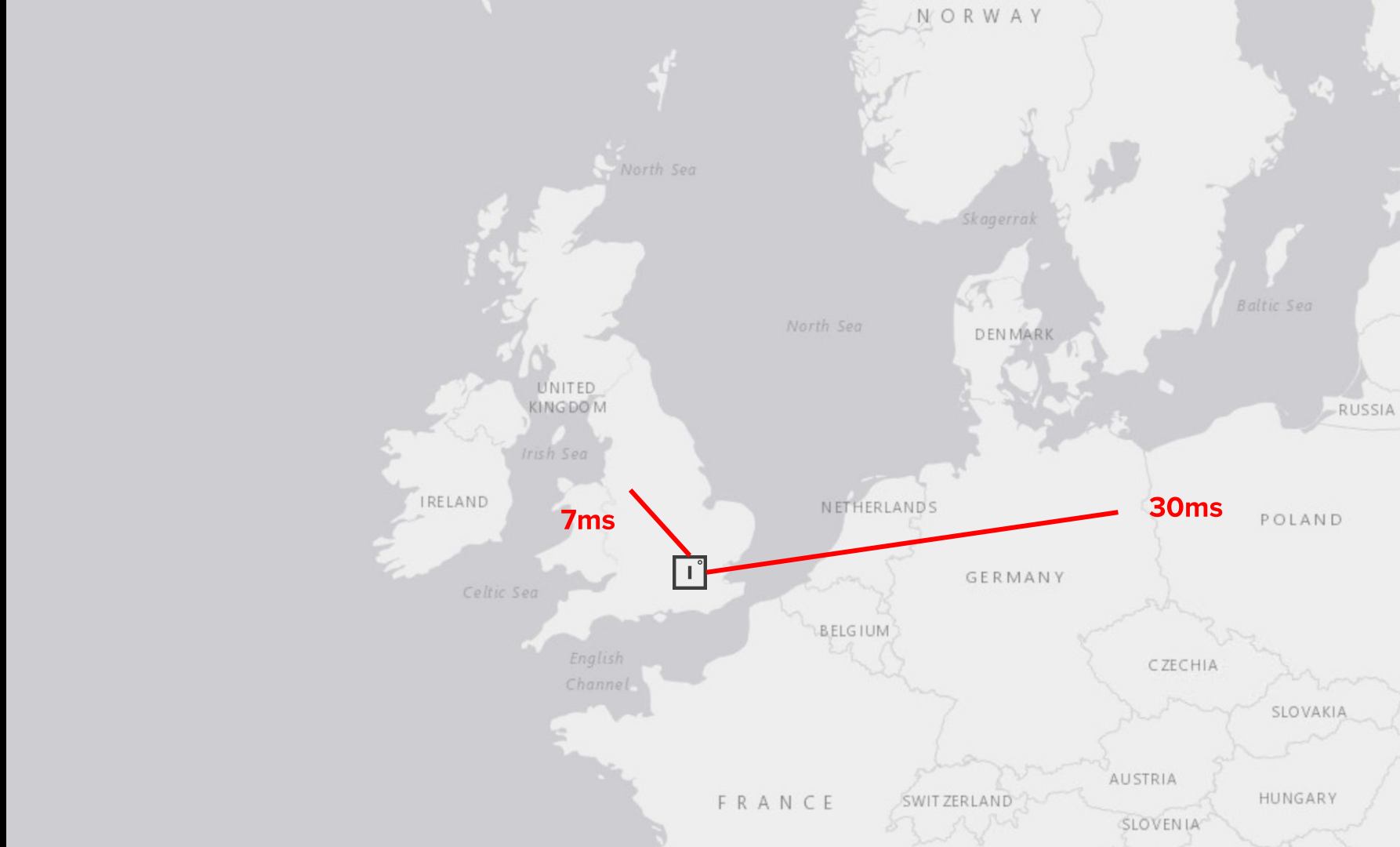
Prometheus

- Collection of data from workloads
- Data queried directly from Prometheus Scrapers
- Fast becoming industry standard for metric collection

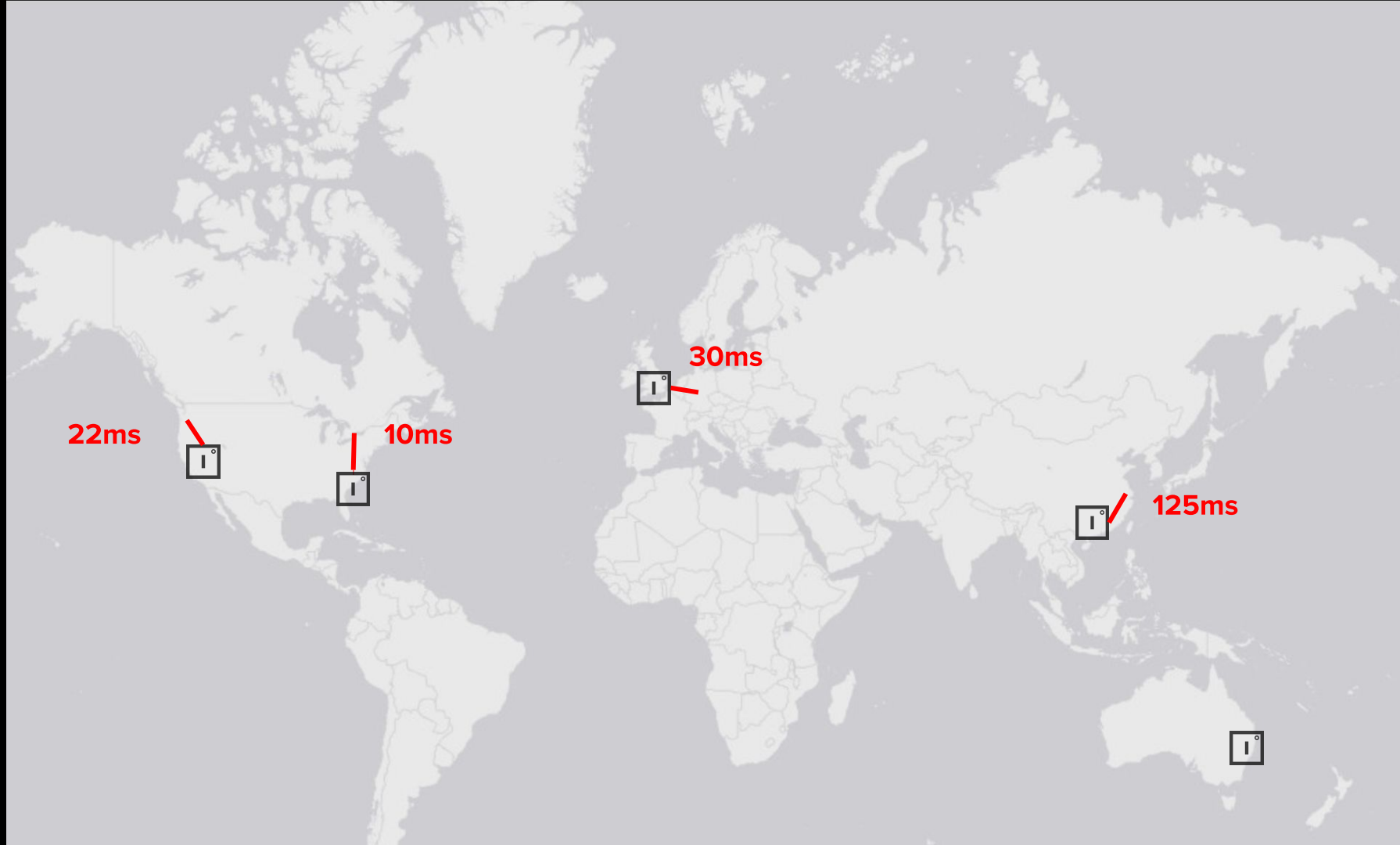
Single Cluster

- Pros
  - Simple
  - Easy to Monitor

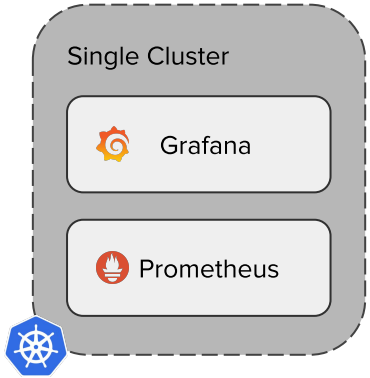
- Cons
  - Redundancy
  - Latency


NORWAY
North Sea
Skagerrak
Baltic Sea
North Sea
DENMARK
UNITED KINGDOM
RUSSIA
Irish Sea
IRELAND
7ms
NETHERLANDS
30ms
POLAND
GERMANY
Celtic Sea
BELGIUM
English Channel
CZECHIA
SLOVAKIA
AUSTRIA
FRANCE
SWITZERLAND
HUNGARY
SLOVENIA

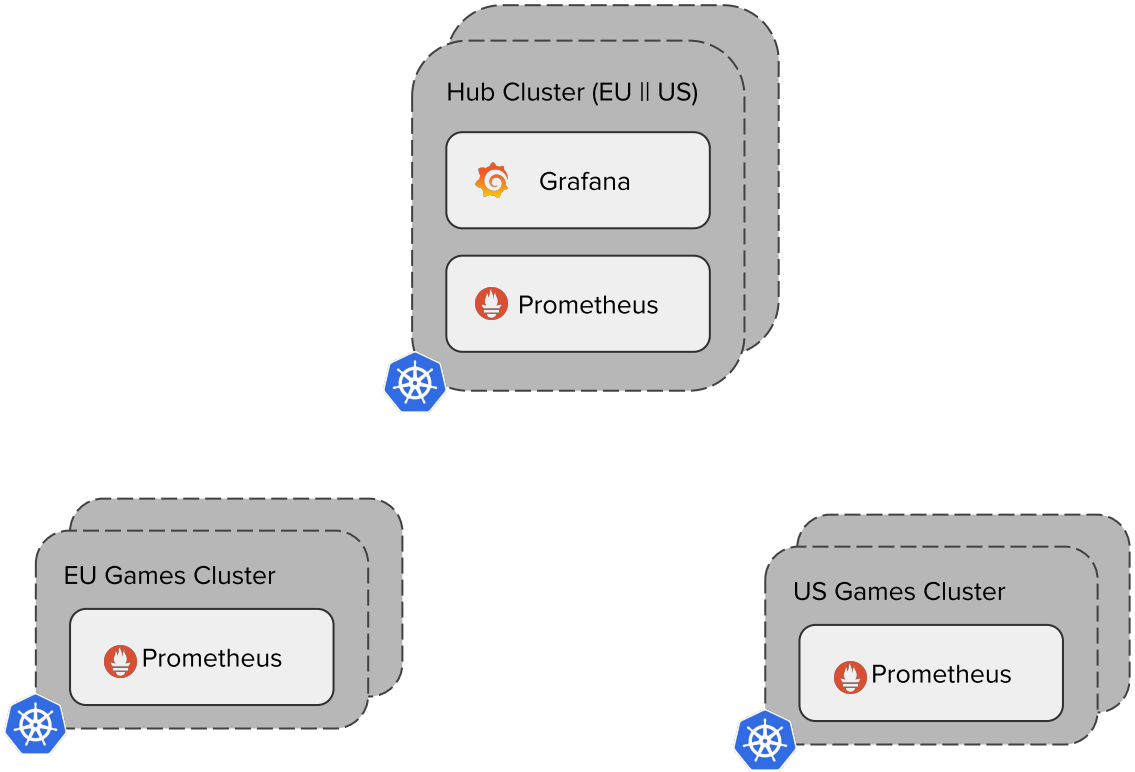

130ms
70ms
300ms
280ms


22ms
10ms
30ms
125ms

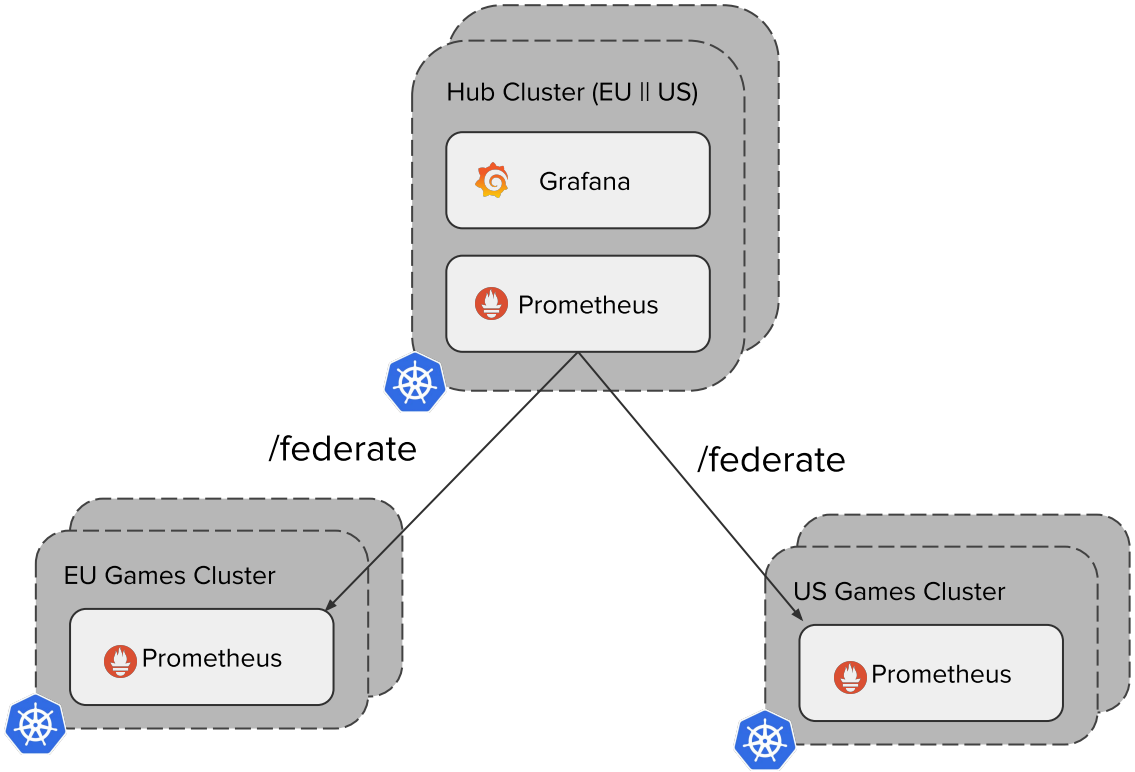
Multi-Cluster

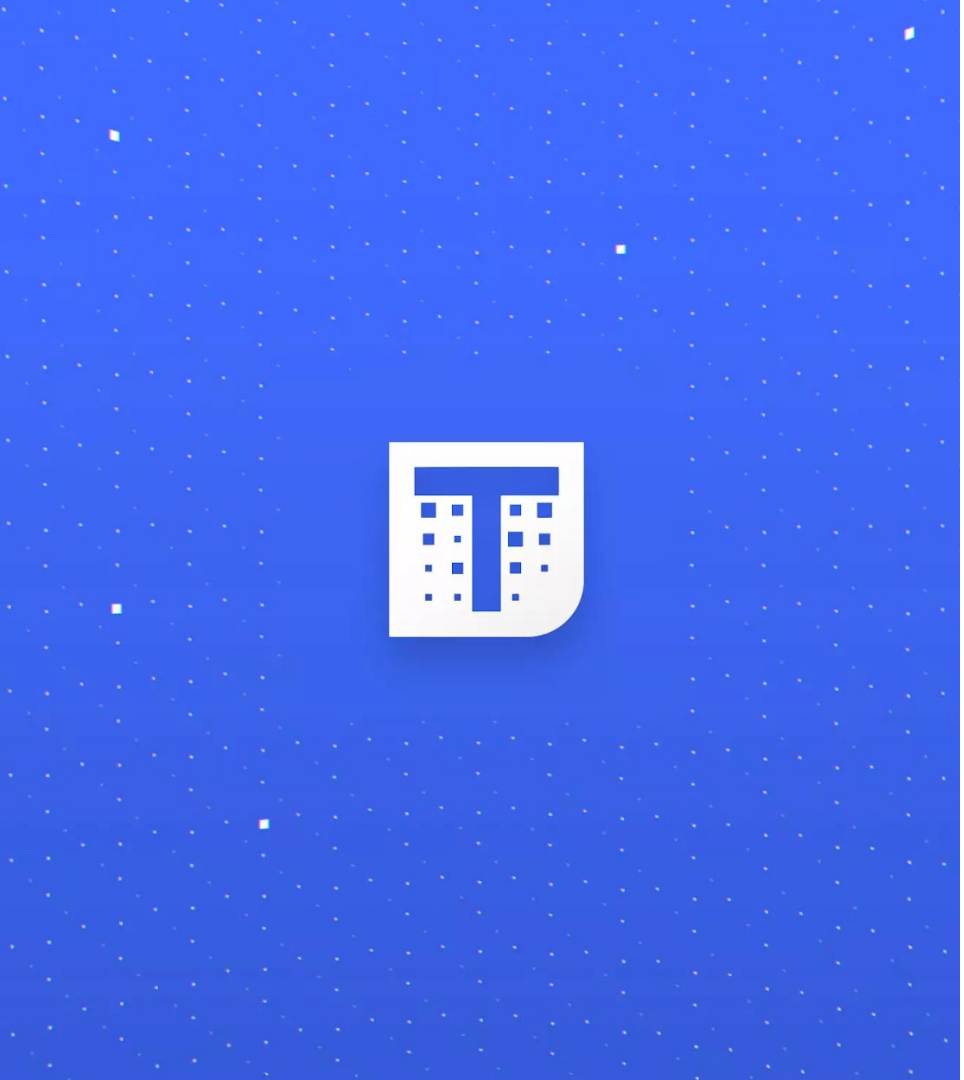


Multi-Cluster - Prometheus



Multi-Cluster - Federation



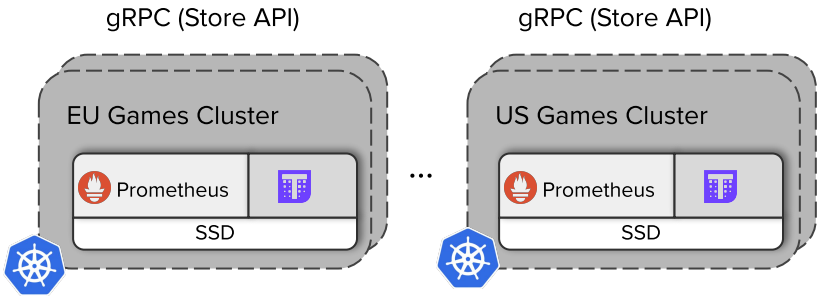


Thanos Project

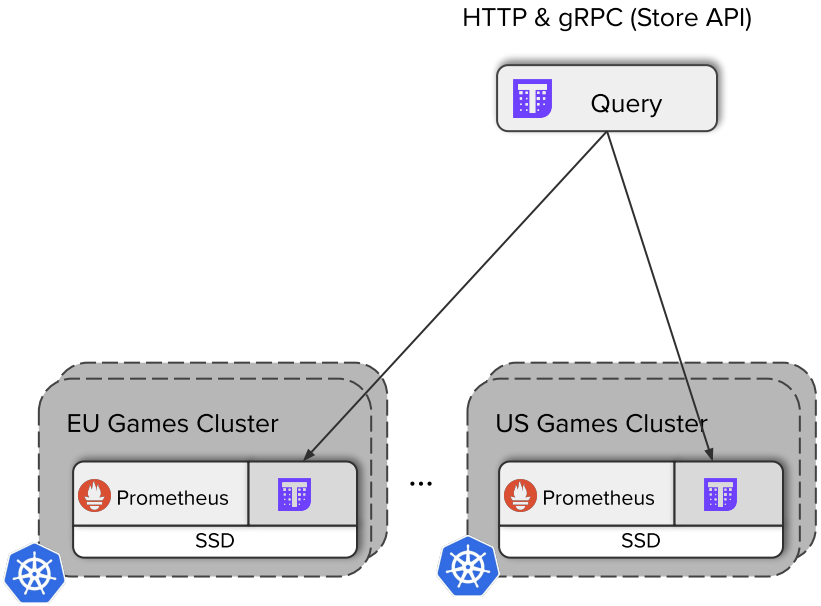
- Global query view of Metrics.
- Unlimited retention of Metrics.
- High availability of components, including Prometheus.

- Downsampling of Metrics.

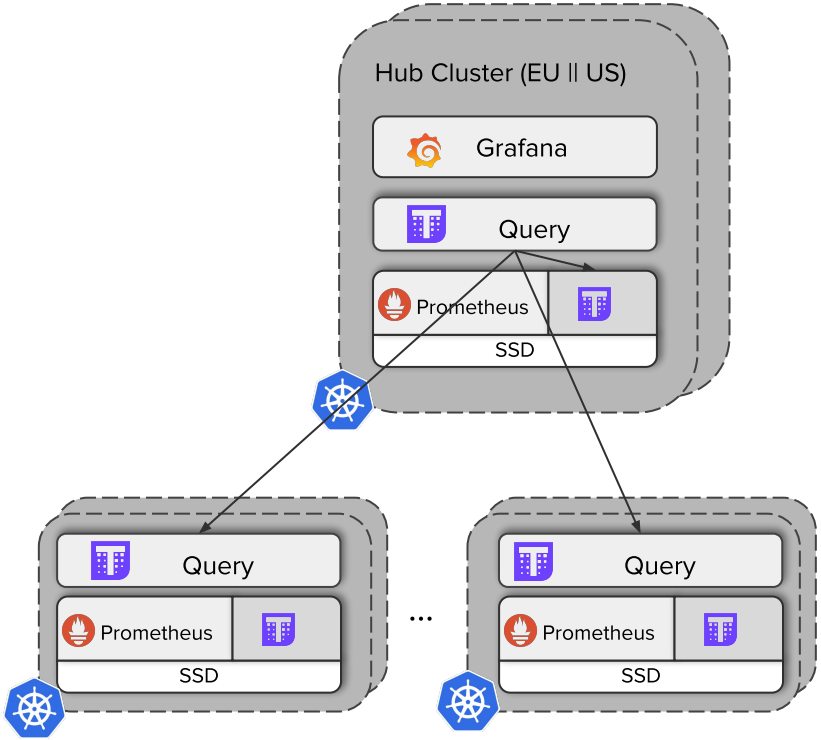
Multi-Cluster - Thanos Sidecar



Multi-Cluster - Thanos Query

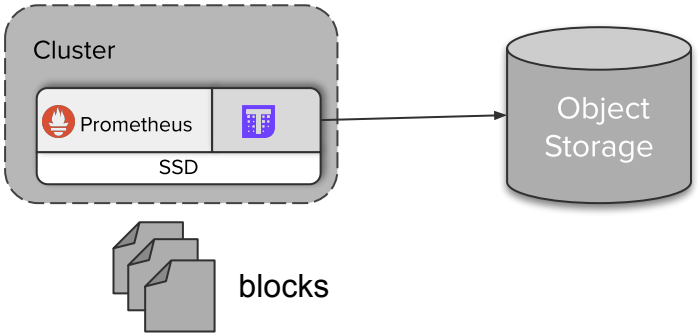


Multi-Cluster - Thanos Query



Multi-Cluster - Thanos Query

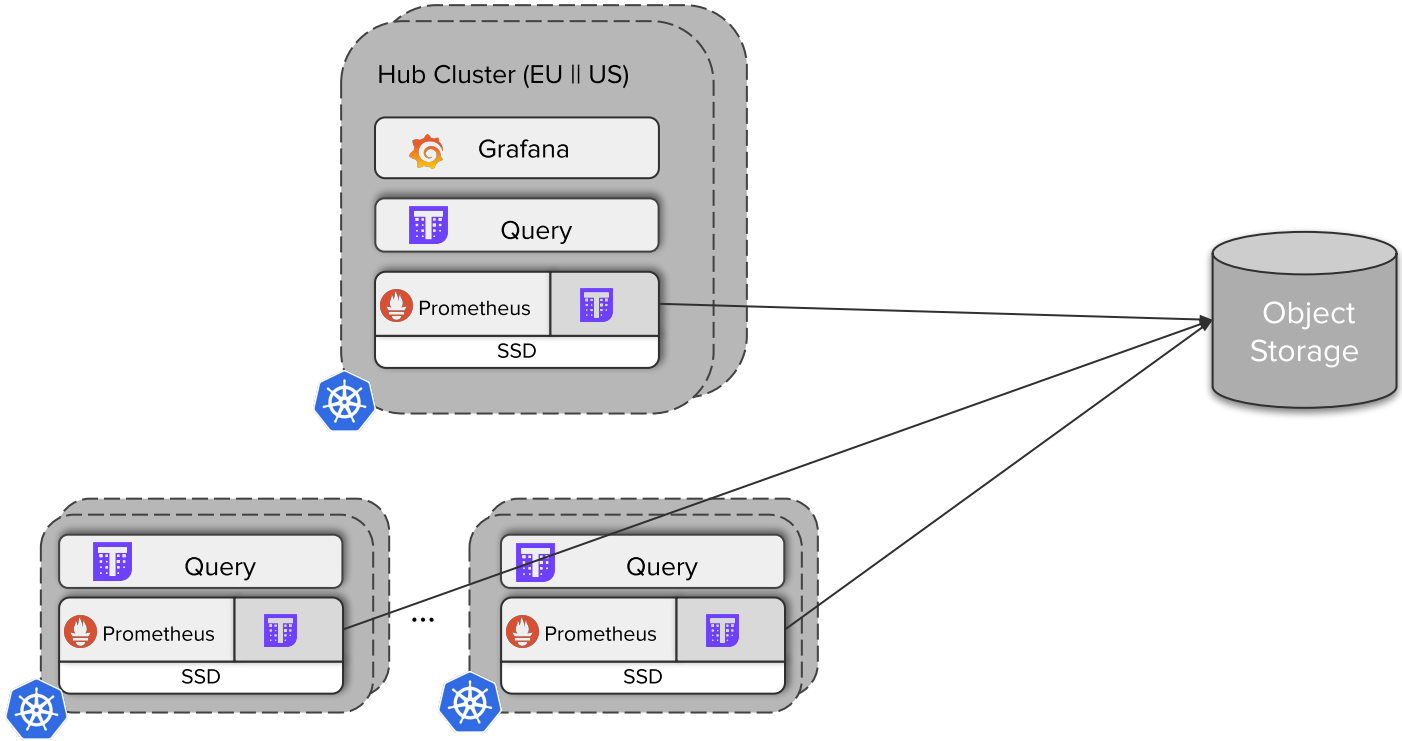
Multi-Cluster - Thanos Sidecar



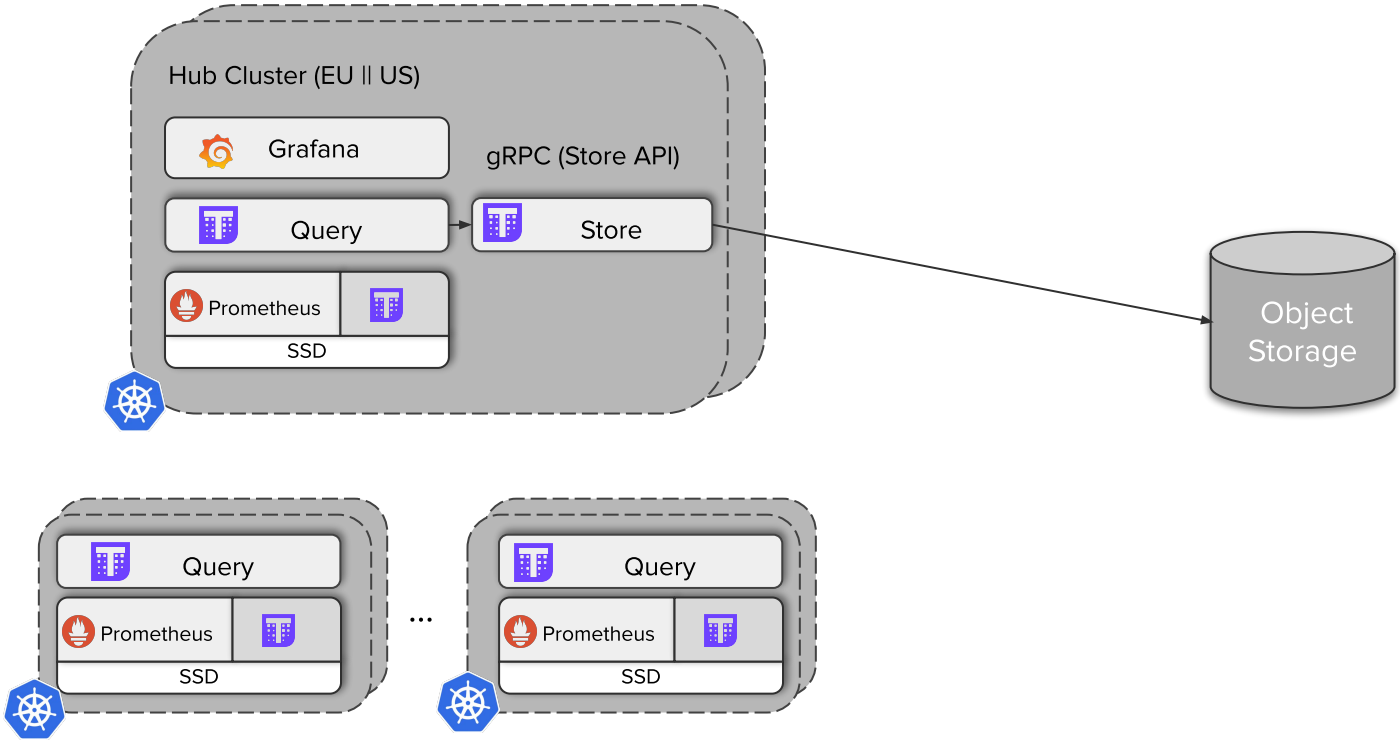
Currently Supported:

- Google Cloud Storage
- S3
- Azure Blob Storage
- Tencent
- Aliyun OSS (soon)

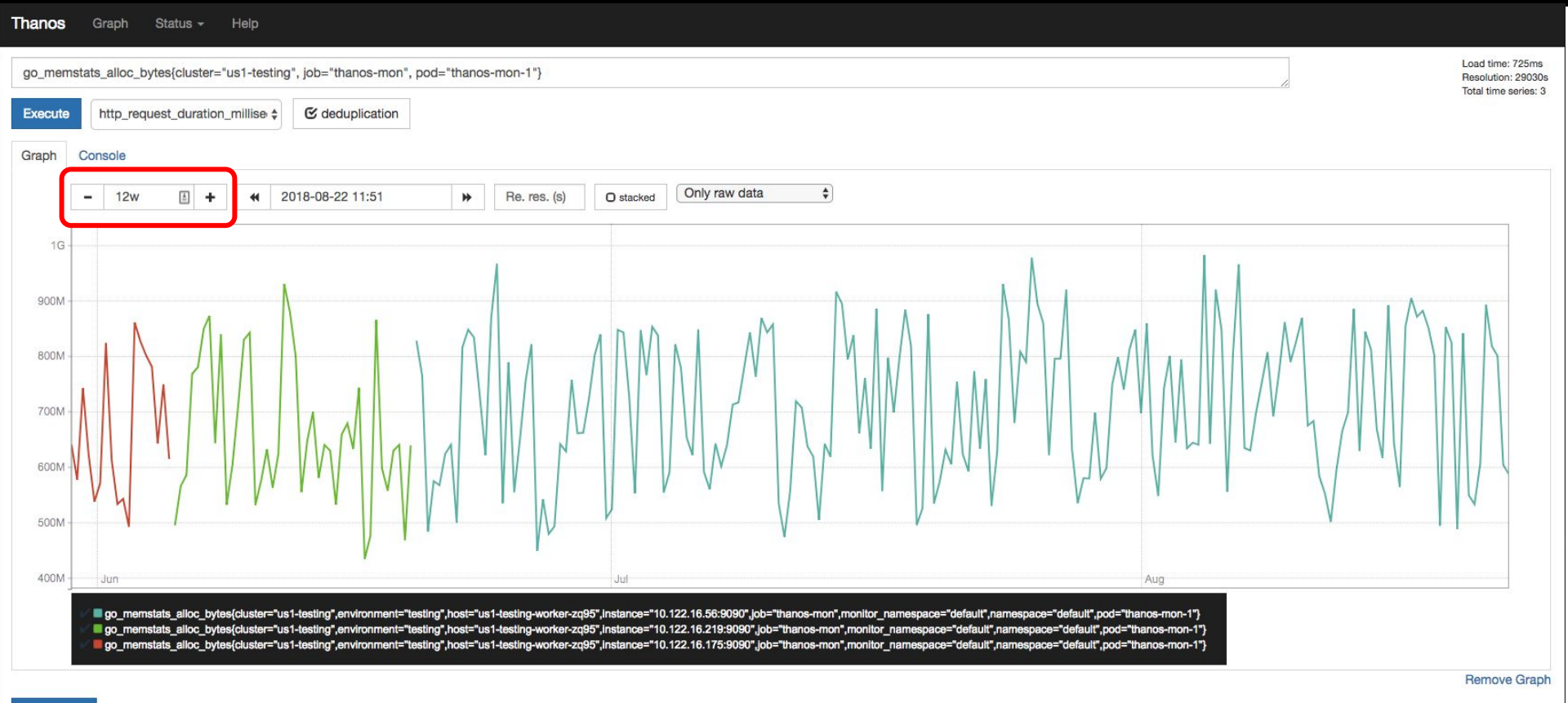
Multi-Cluster - Storage



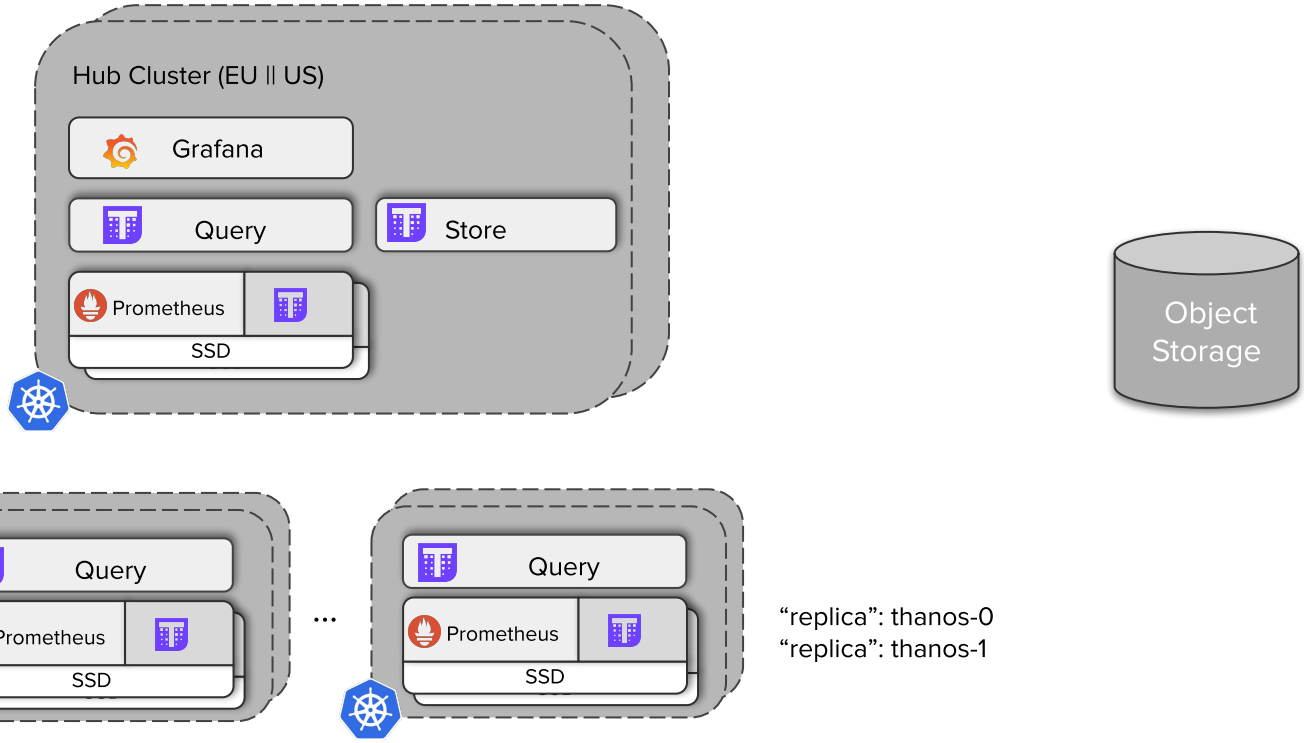
Multi-Cluster - Storage



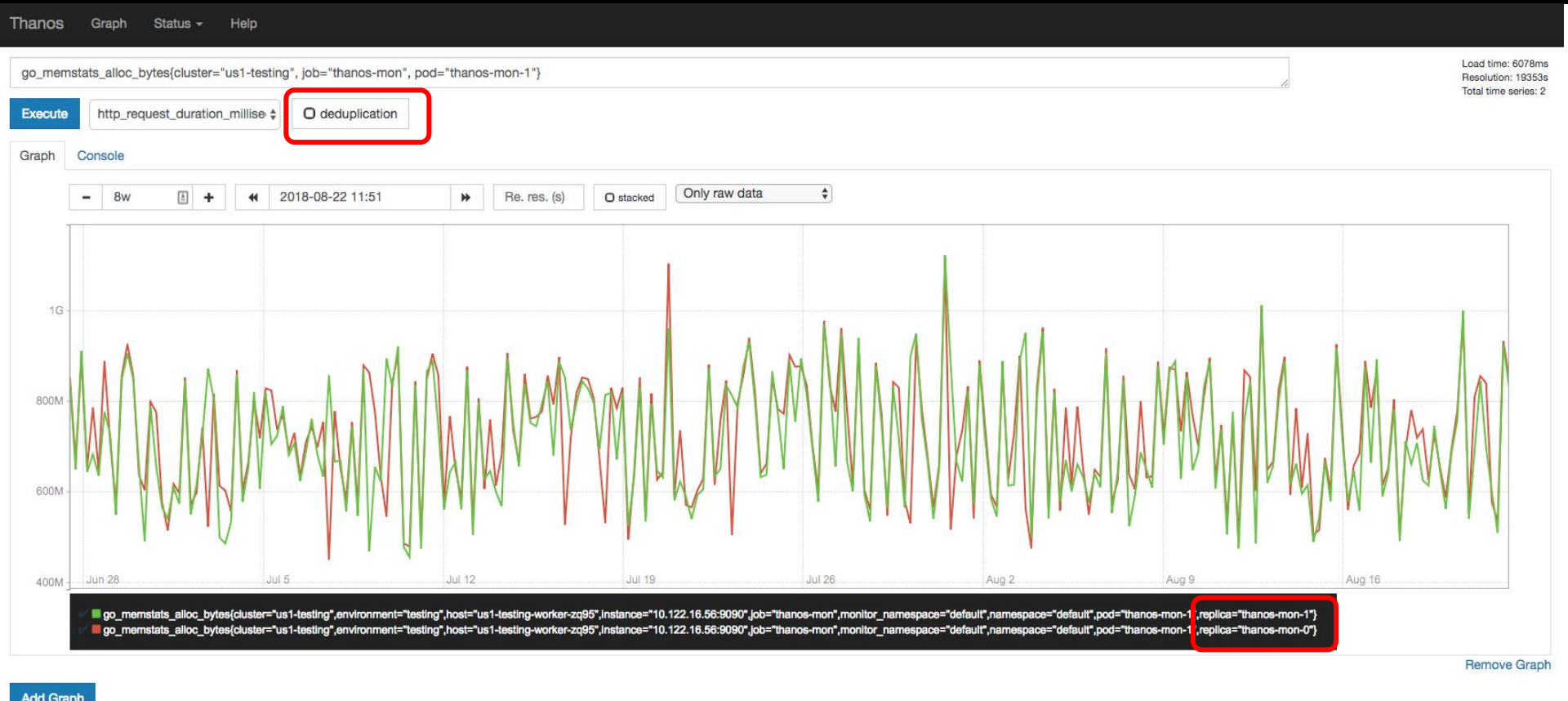
Multi-Cluster - Storage



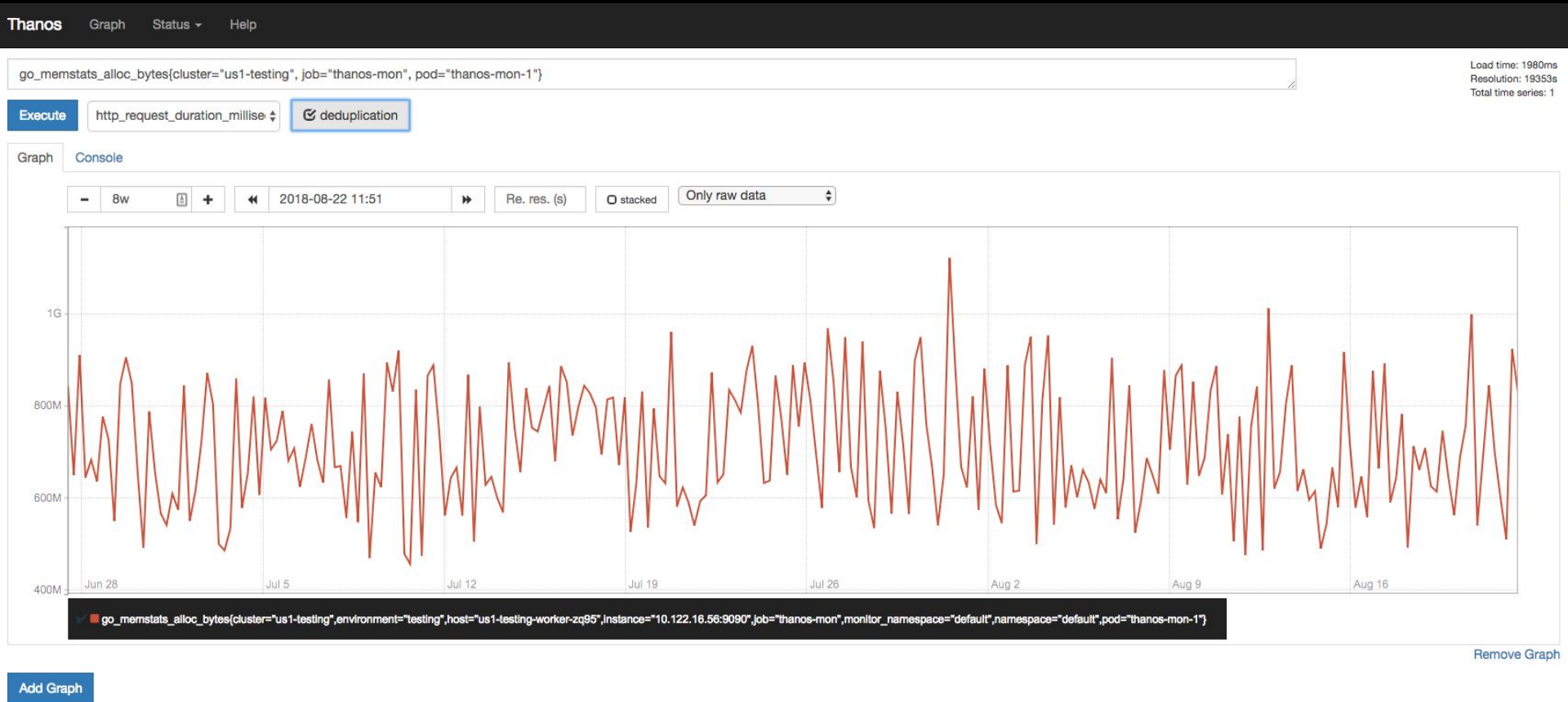
Multi-Cluster - High Availability



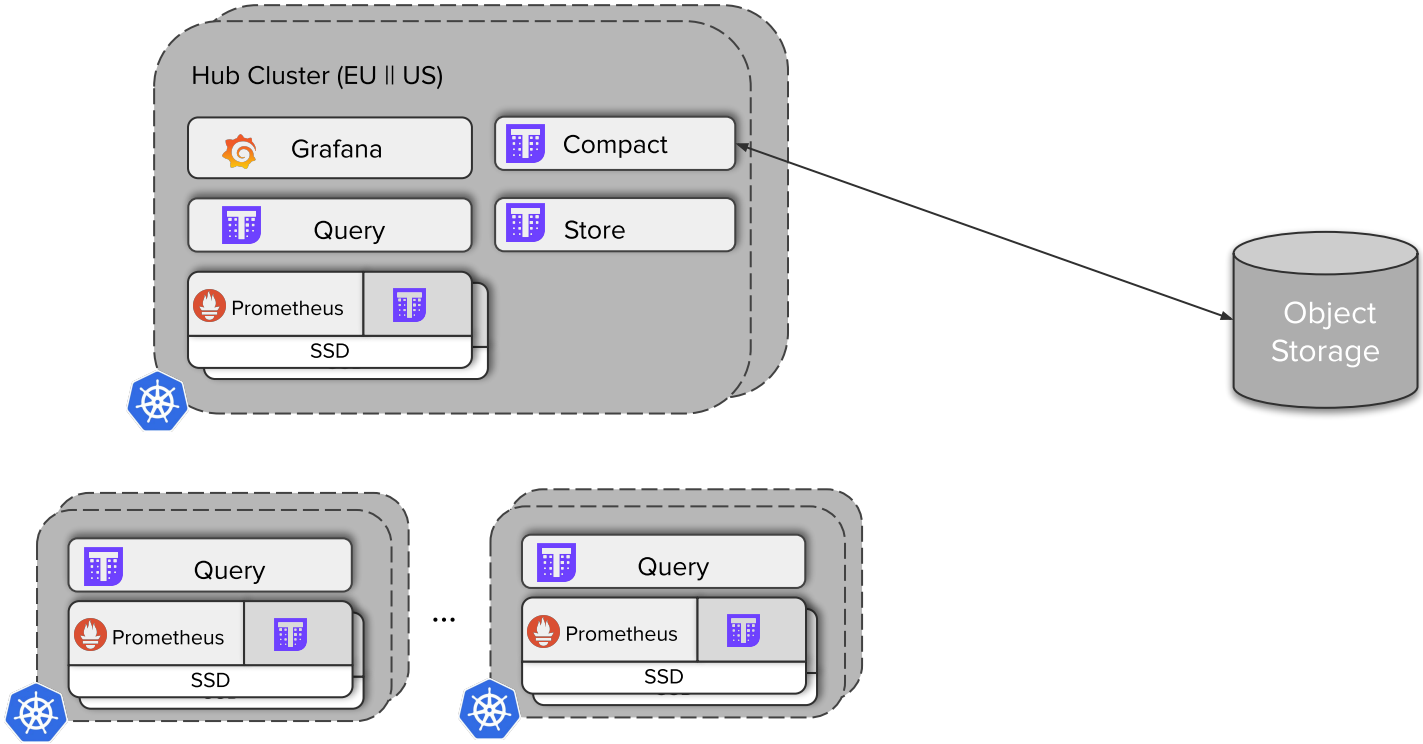
Multi-Cluster - High Availability



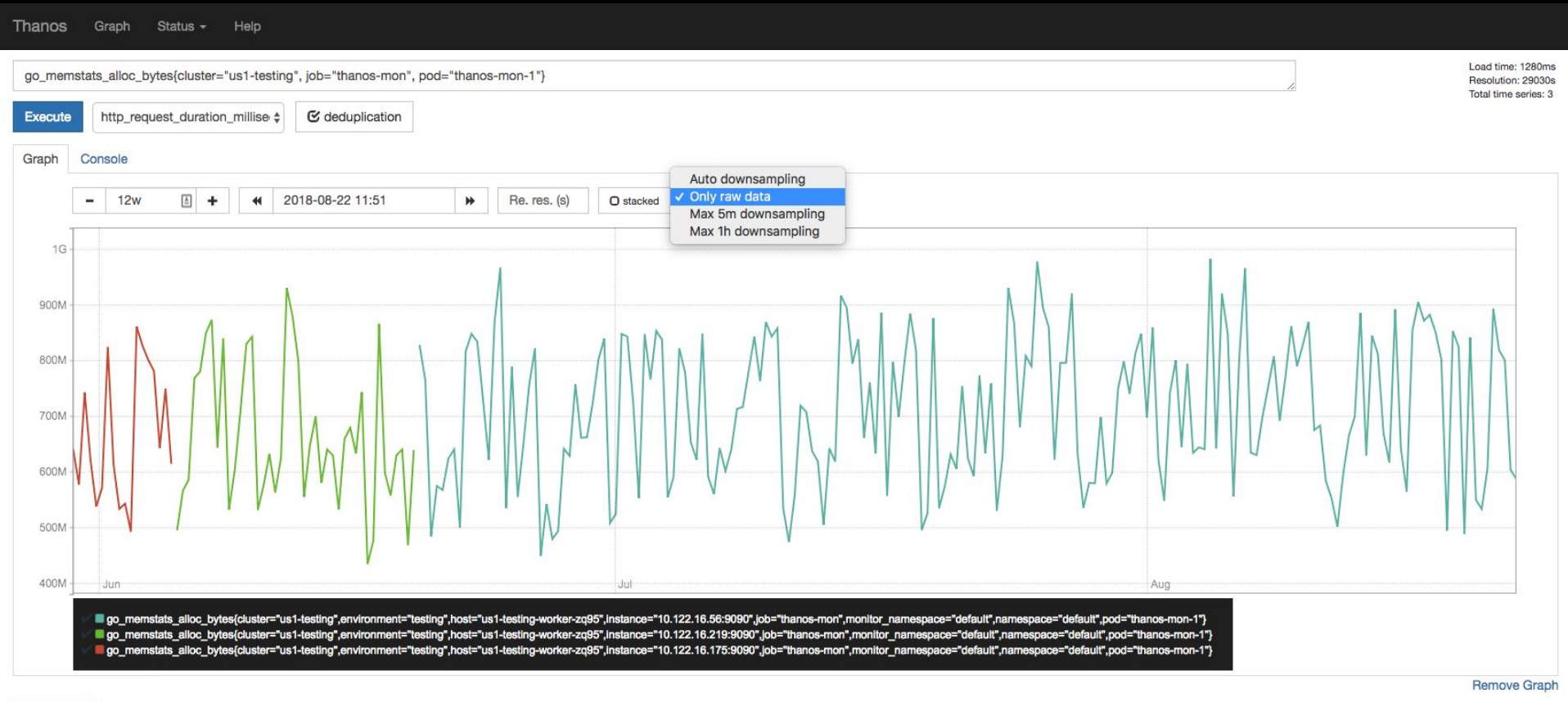
Multi-Cluster - High Availability



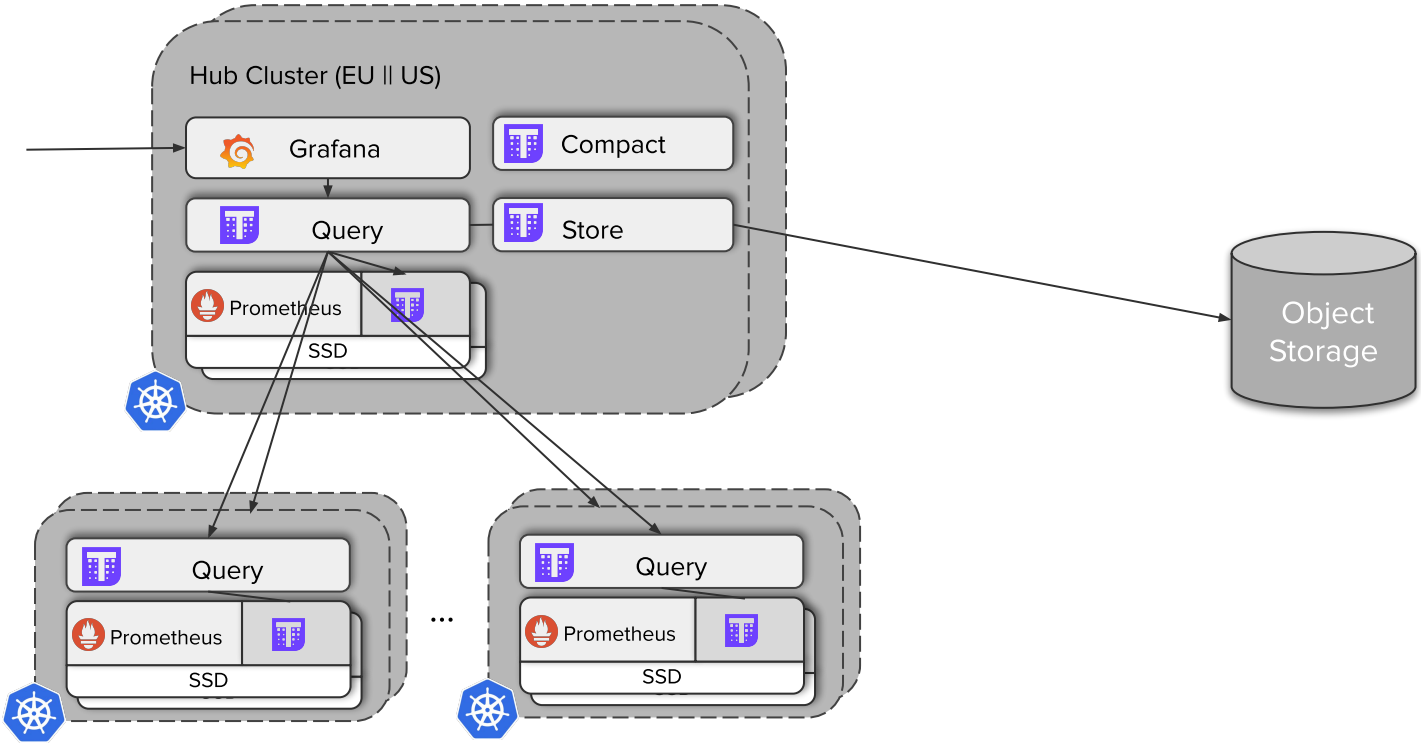
Multi-Cluster - Compaction



Multi-Cluster - Compaction



Multi-Cluster



Multi-Cluster

Kubernetes

- Consistent approach in all clusters
- Kubernetes Service Discovery for discovering workloads
- Mature tooling and automation

Prometheus

- Collection of data from workloads
- Federation can be problematic

Thanos

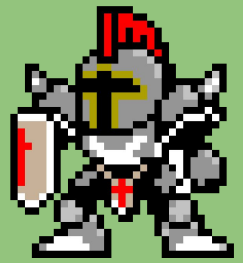
- Global View, Retention, HA, Downsampling
- Pulls Metrics from Object Storage or Thanos Sidecar
- Builds on existing Prometheus infrastructure

Multi-Cluster

- Pros
  - Reduced Latency
  - High Availability
    - Cluster Level
    - Workload level
  - Global Query
  - Long Term Metrics

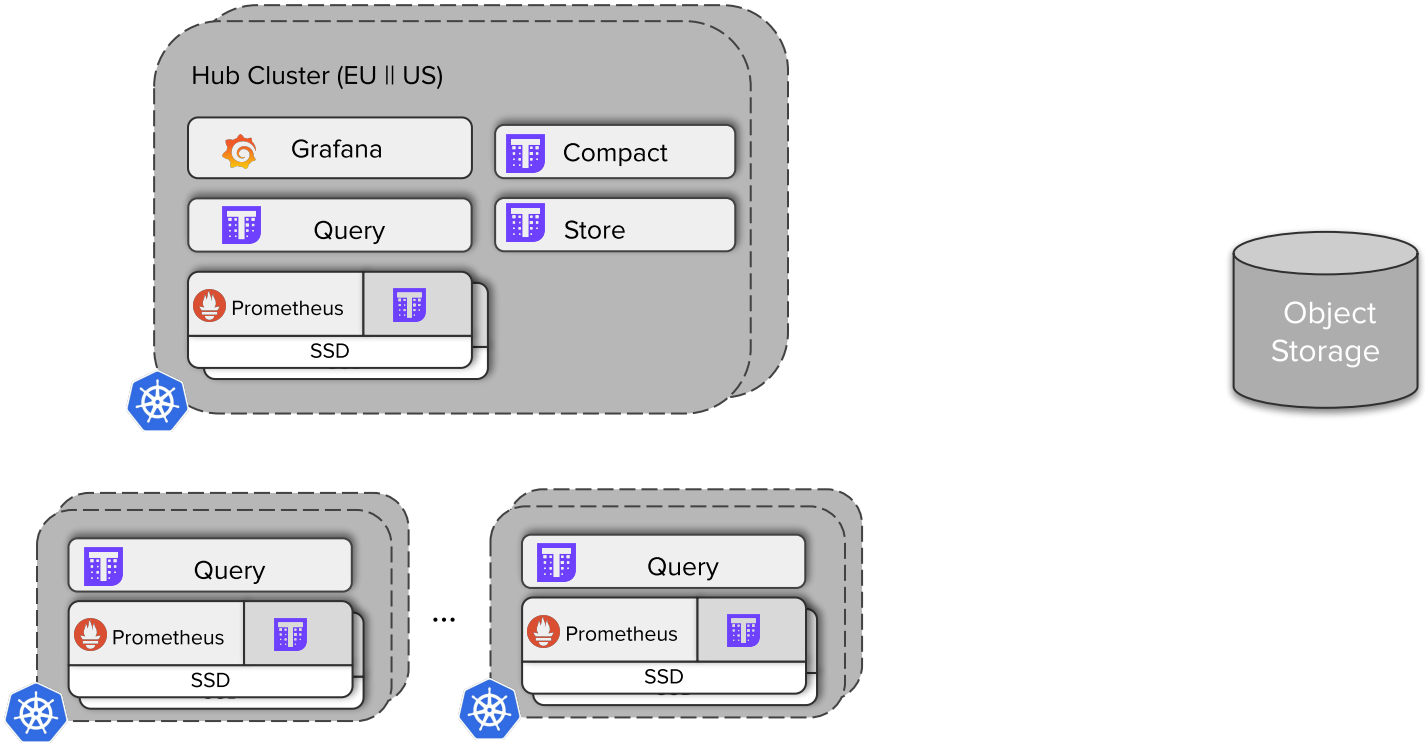
- Cons
  - Observability is harder
  - Increased Complexity
  - Automation?
  - Tooling?

No plan survives first contact with the enemy.

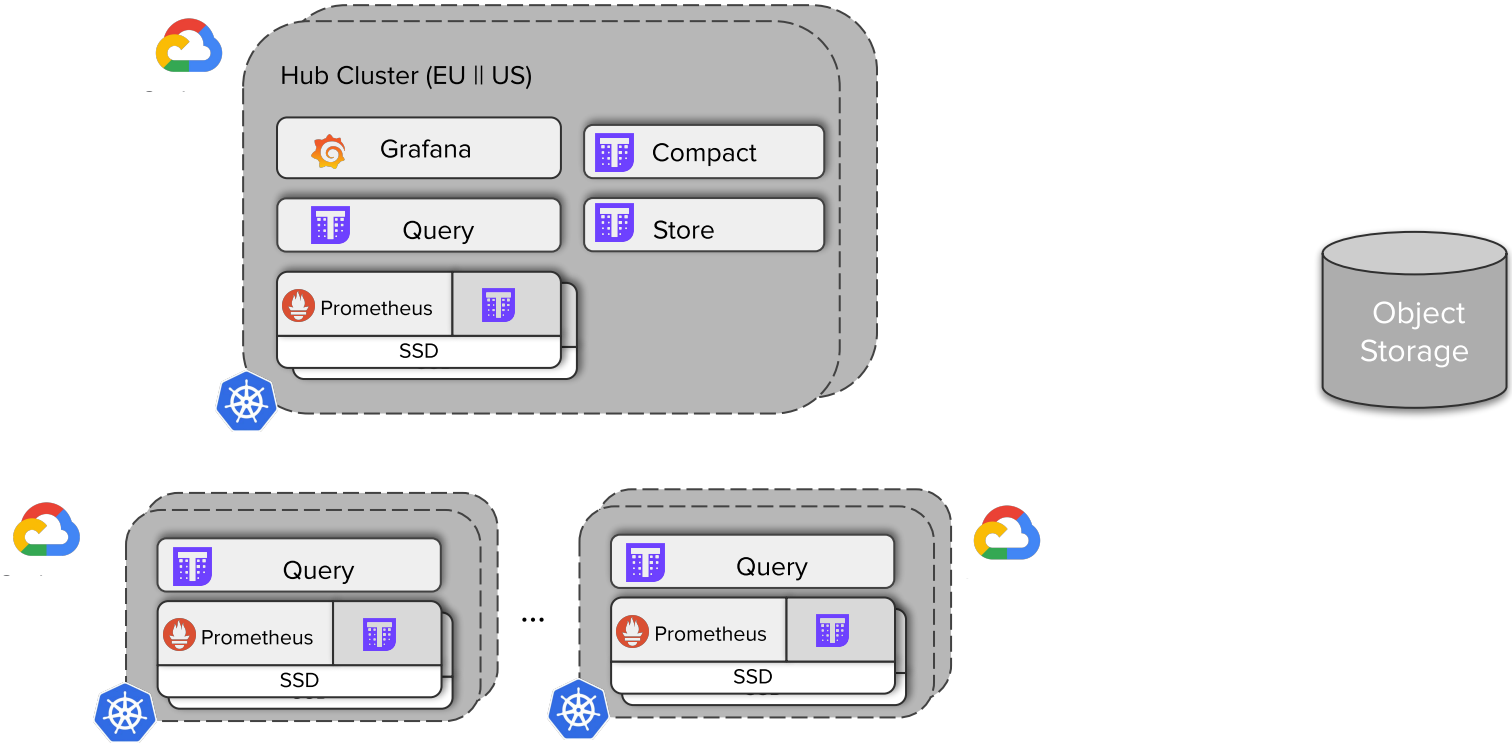


没有计划在与敌人的第一次接触中幸存。

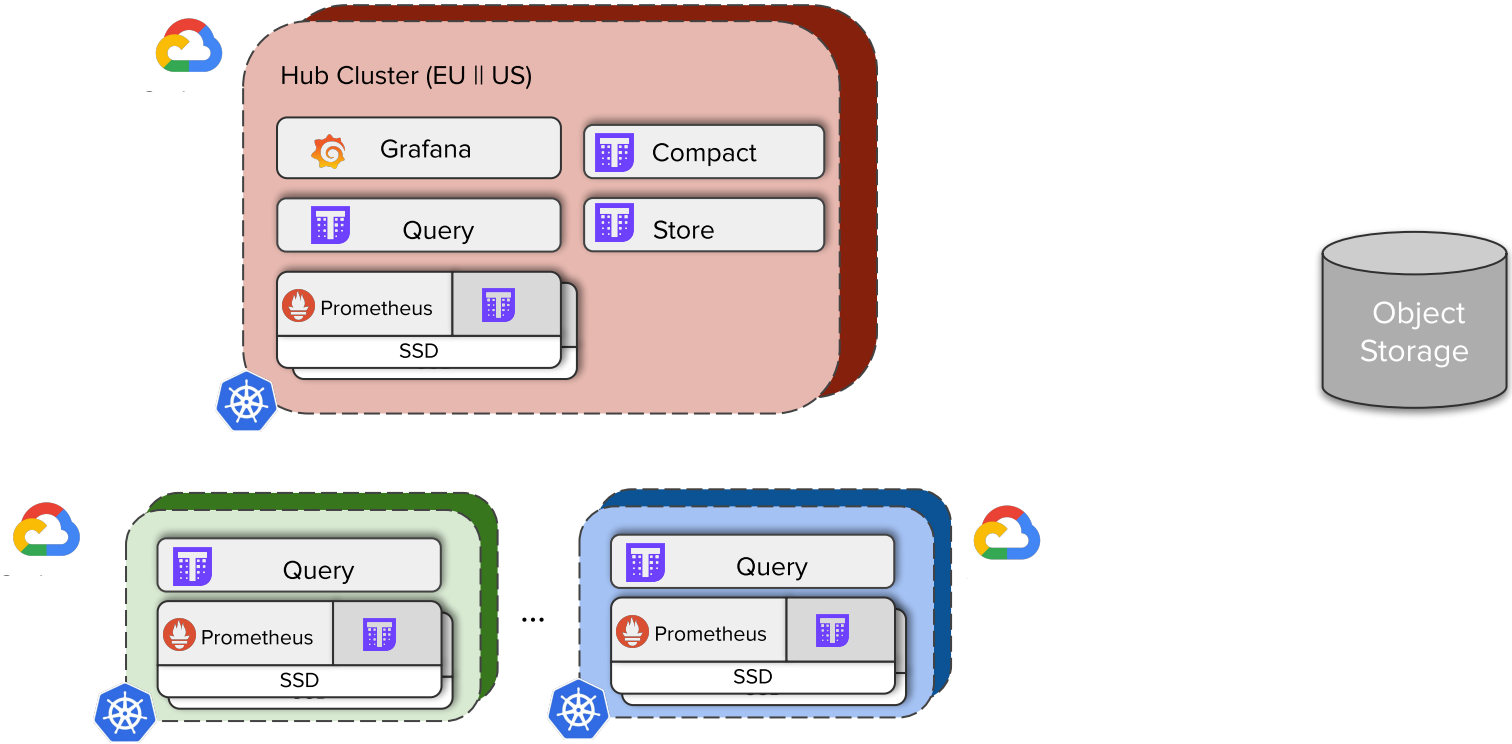
Multi-Cluster - Networking



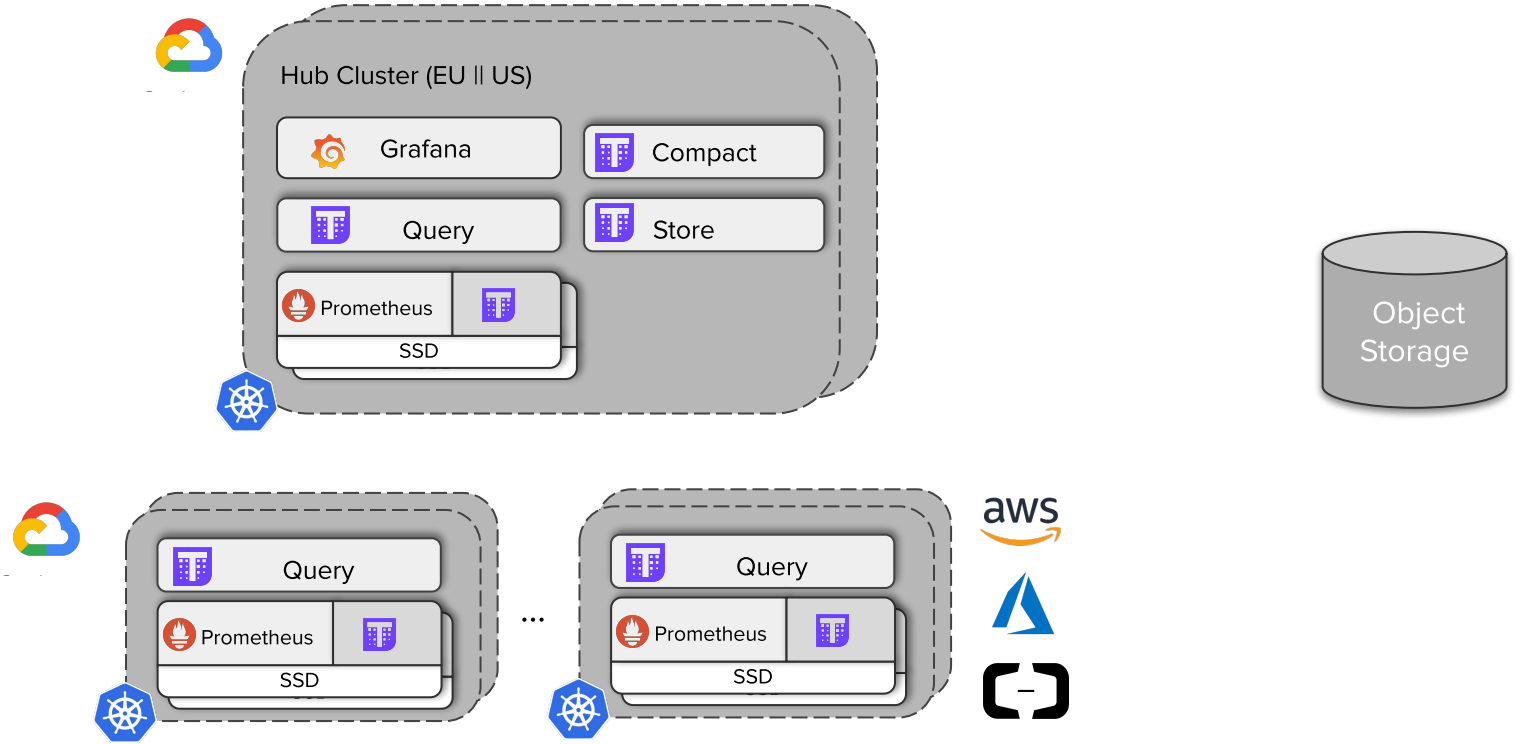
Multi-Cluster - Networking



Multi-Cluster - Networking

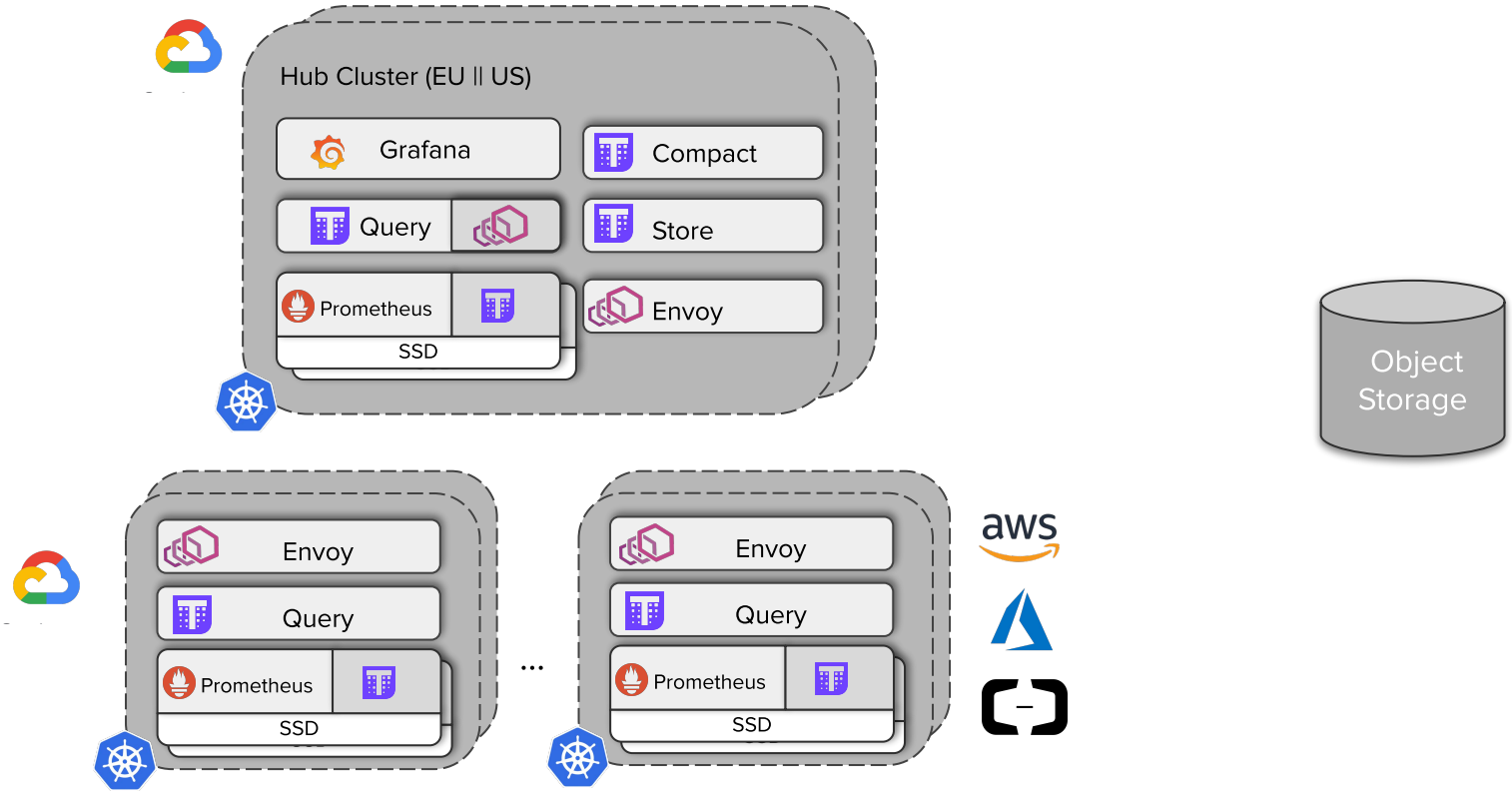


Multi-Cloud - Networking

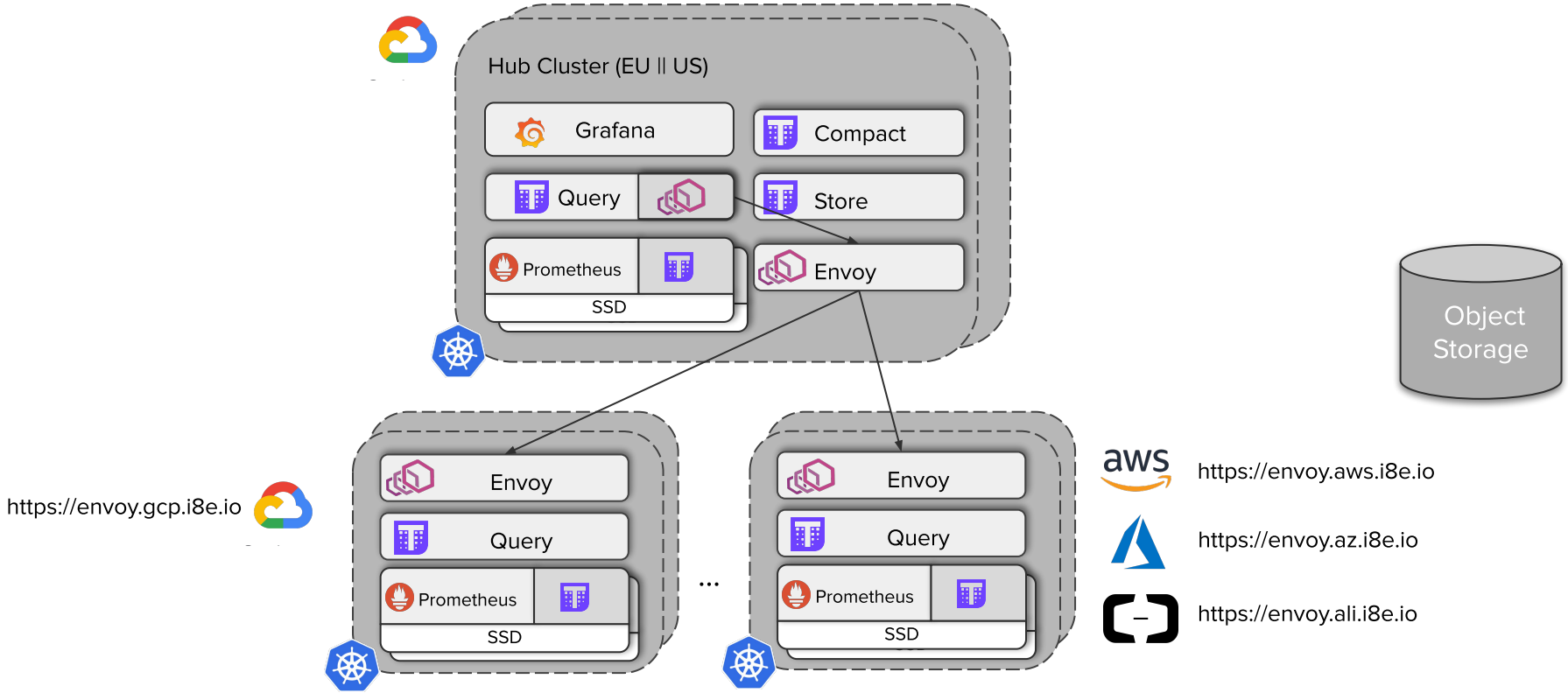



envoy

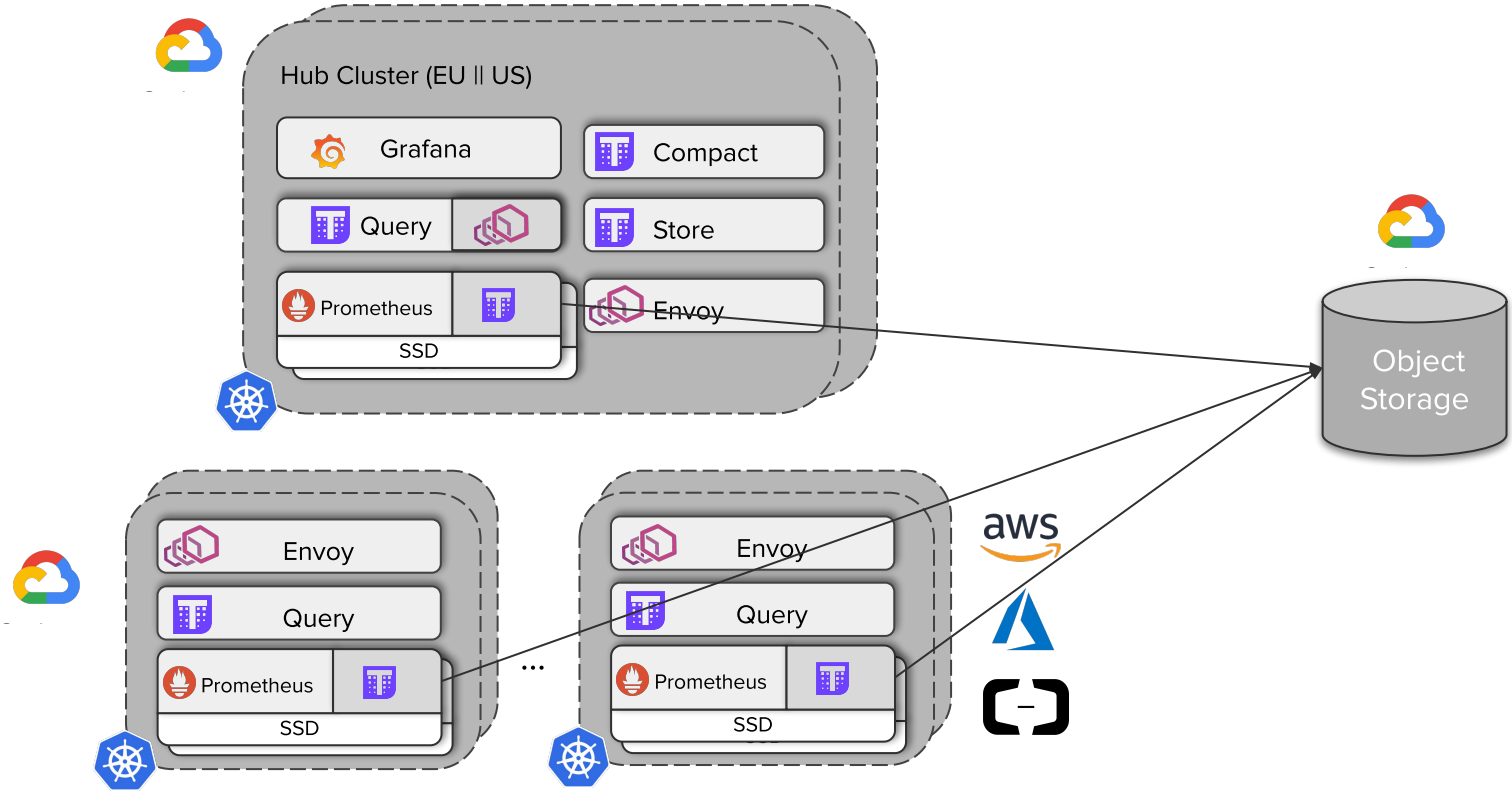
Multi-Cloud - Envoy



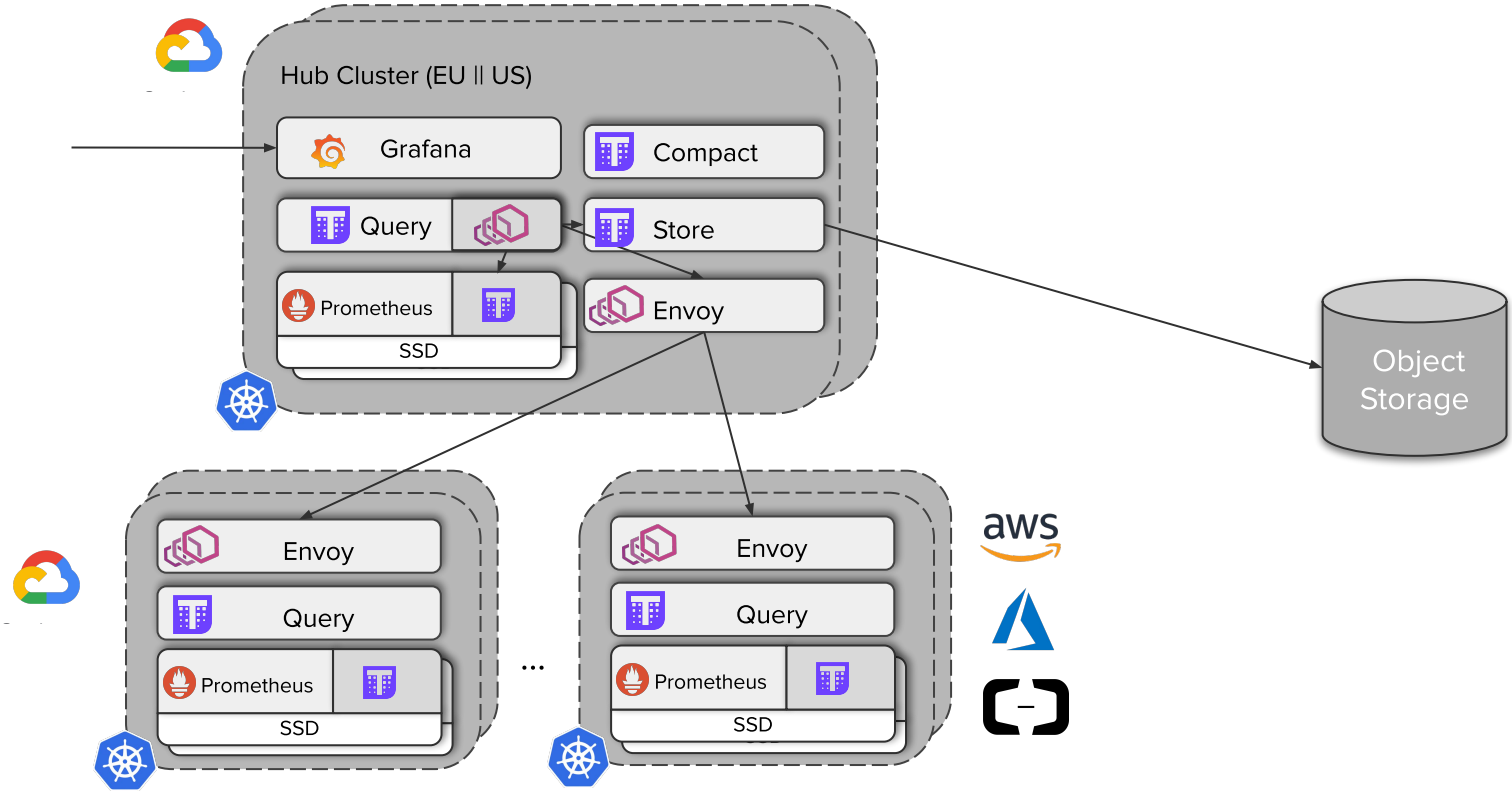
Multi-Cloud - Envoy



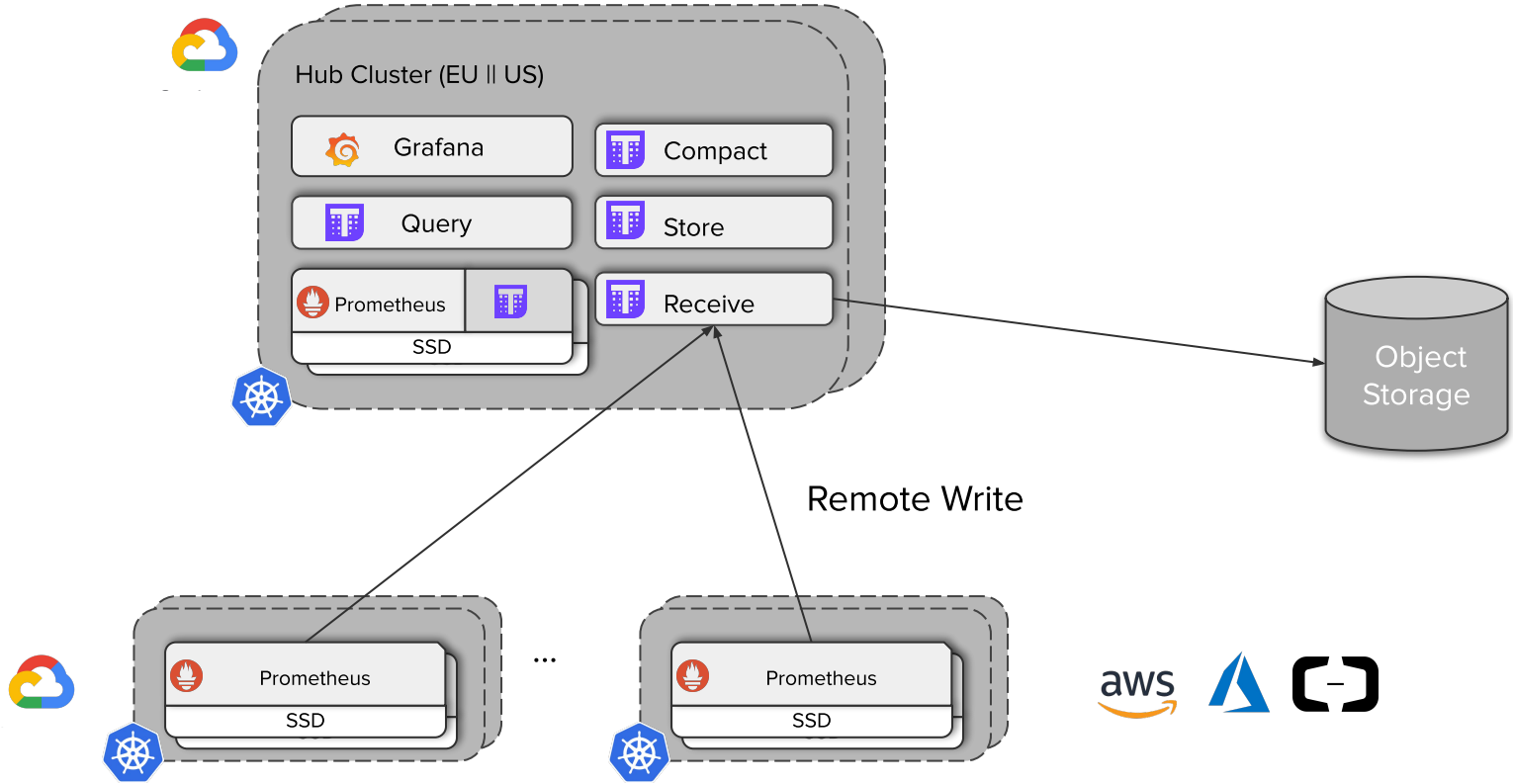
Multi-Cloud - Persistence



Multi-Cloud - Querying



Multi-Cloud - Thanos Receive



Multi-Cloud

Kubernetes

- Consistent approach in all clouds & clusters
- Kubernetes Service Discovery for discovering workloads

Prometheus

- Collection of metrics from workloads
- TSDB Storage format

Thanos

- Global View, Retention, HA, Downsampling
- Flexible StoreAPI allows for different usage scenarios

Envoy

- Edge Proxy for same approach cross-cluster and cross-cloud communication

Multi-Cloud

- Pros
  - Reduced Latency
  - High Availability
    - Cluster Level
    - Workload level
  - Global Query
  - Long Term Metrics

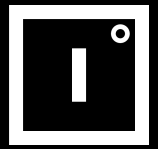
- Cons
  - Observability is harder
  - Increased Complexity
  - Automation??
  - Tooling??

Summary

Single Cluster

Multi-Cluster

Multi-Cloud



improbable.io