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？
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
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
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.

Thanos Begins
Prometheus

/metrics

# TYPE counter
app_request_total 1337

# TYPE gauge
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
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"
Single Cluster

- Grafana
- Prometheus
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
Multi-Cluster

Single Cluster

- Grafana
- Prometheus
Multi-Cluster - Federation

- Hub Cluster (EU || US)
  - Grafana
  - Prometheus

- EU Games Cluster
  - Prometheus

- US Games Cluster
  - Prometheus

/federate
Thanos Begins
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 Sidecar

Currently Supported:
- Google Cloud Storage
- S3
- Azure Blob Storage
- Tencent
- Aliyun OSS (soon)
Multi-Cluster - Storage
Multi-Cluster - High Availability

Hub Cluster (EU & US)
- Grafana
- Query
- Store
- Prometheus
- SSD

Object Storage

"replica": thanos-0
"replica": thanos-1

Query
- Prometheus
- SSD

...
Multi-Cluster - High Availability

```
$ go_memstats_alloc_bytes {cluster="us1-testing", job="thanos-mon", pod="thanos-mon-1"}
```

```
http_request_duration_milliseconds deduplication
```

Graph View

```
$ go_memstats_alloc_bytes {cluster="us1-testing", environment="testing", host="url-testing-worker-xq38f", instance="10.122.16.56:9090", job="thanos-mon", namespace="default", namespace="default", pod="thanos-mon-1"}
```
Multi-Cluster - Compaction
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

Hub Cluster (EU II US)

- Grafana
- Compact
- Query
- Store
- Prometheus
- SSD

Object Storage
Multi-Cloud - Envoy

Hub Cluster (EU II US)

- Grafana
- Compact
- Query
- Store
- Prometheus
- Envoy

SSD

Object Storage

https://envoy.gcp.i8e.io

https://envoy.aws.i8e.io

https://envoy.az.i8e.io

https://envoy.ali.i8e.io
Multi-Cloud - Persistence
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