

High Available + Scalable Prometheus with Thanos in Alibaba

Guo'an Qin & Tao Li, Alibaba Group

About US



秦国安

花名 炎烈，高级技术专家

十二年工作经历，先后在百度、奇虎360、豌豆荚从事运维与开发工作

2016年加入阿里巴巴，主导完成数据库容器化及资源调度、存储计算分离

2018年加入容器平台团队，先后负责 kubelet & 数据生态及基础技术 SRE 相关工作

李涛

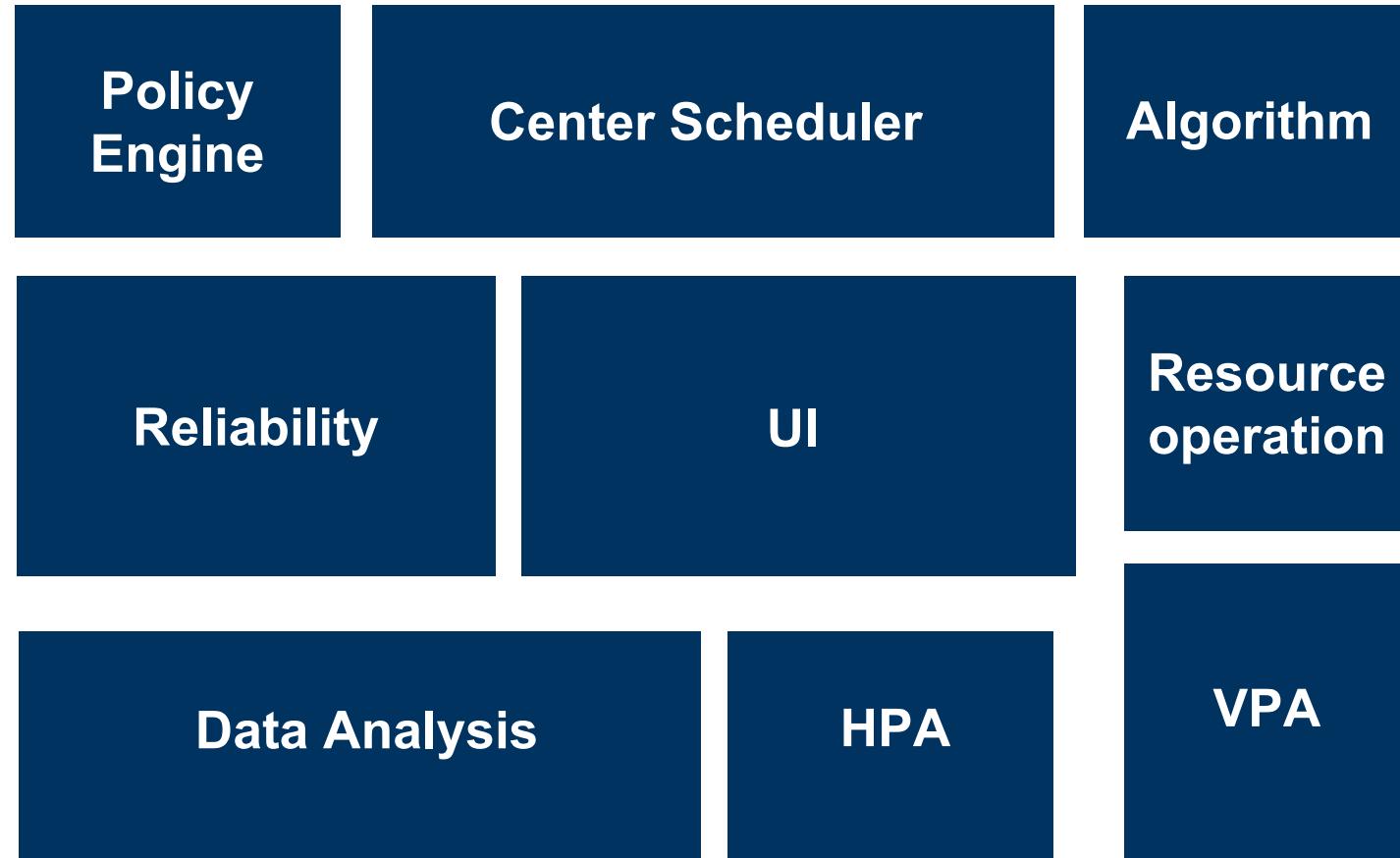
花名 吕风，高级开发工程师

2018年加入容器平台团队，负责数据生态相关工作



Background

Use Case



Our Plan

Monitoring dashboard

Systems (data points trend/aggregation)

- Autoscaling
- Alerting
- Scheduler

Operational dashboard

Human (interactive arbitrary query focused)

- System operator
- Algorithm engineer
- Data analyst

Instrumentation

Insight

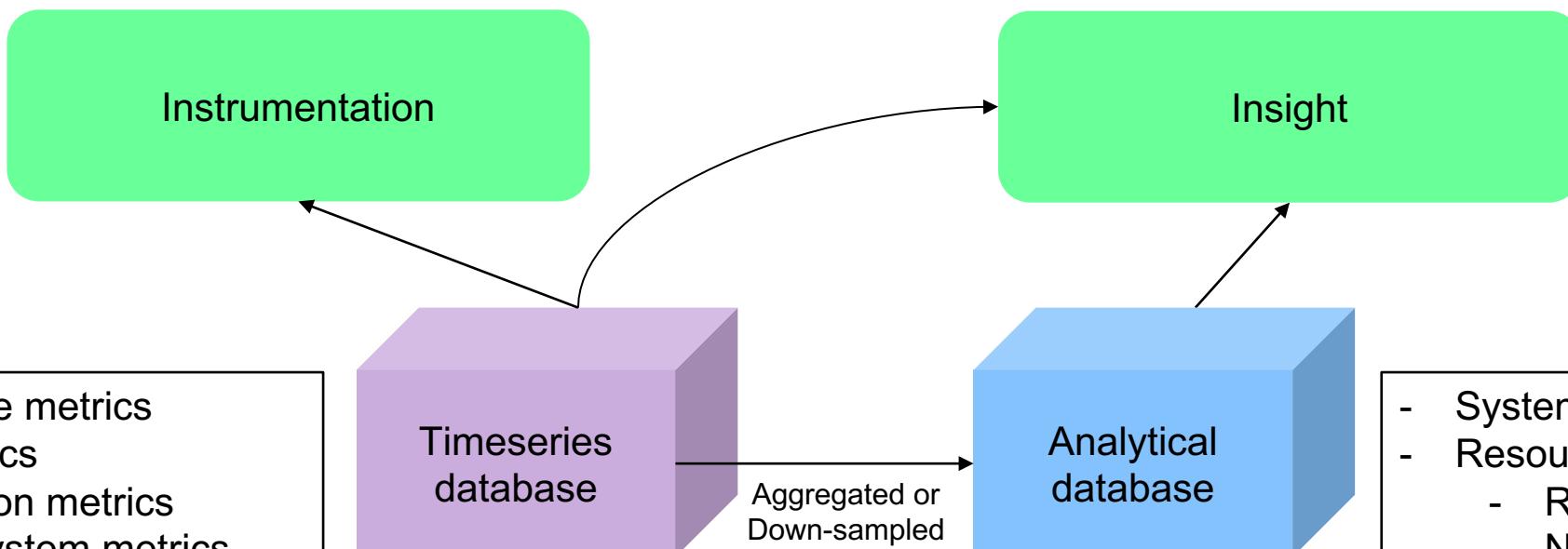
Timeseries
database

Analytical
database

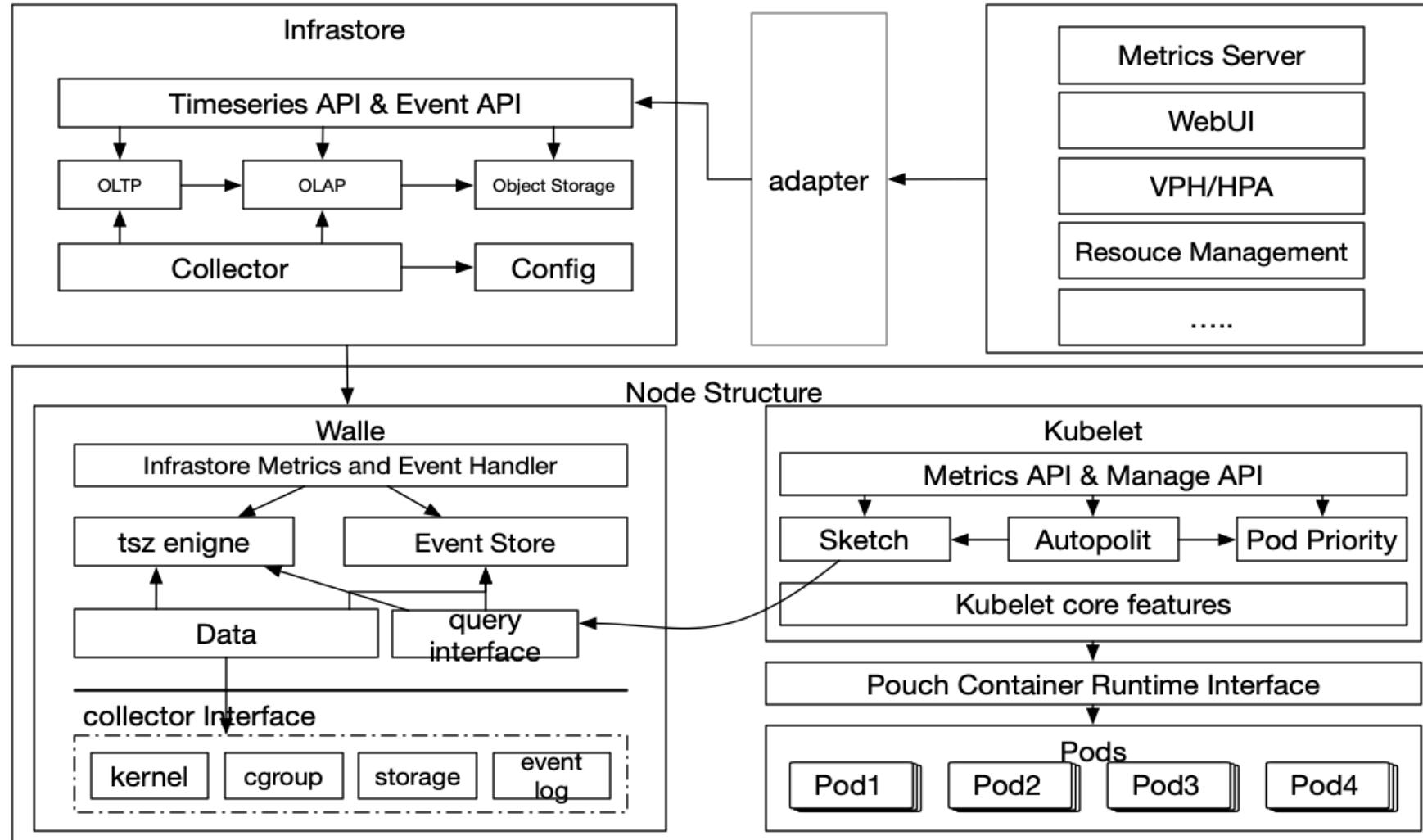
- Hardware metrics
- OS metrics
- Application metrics
- Sigma system metrics
- ...

Aggregated or
Down-sampled

- System information
- Resources information
 - Running pods
 - Number of apps
 - ...
- Events
- ...



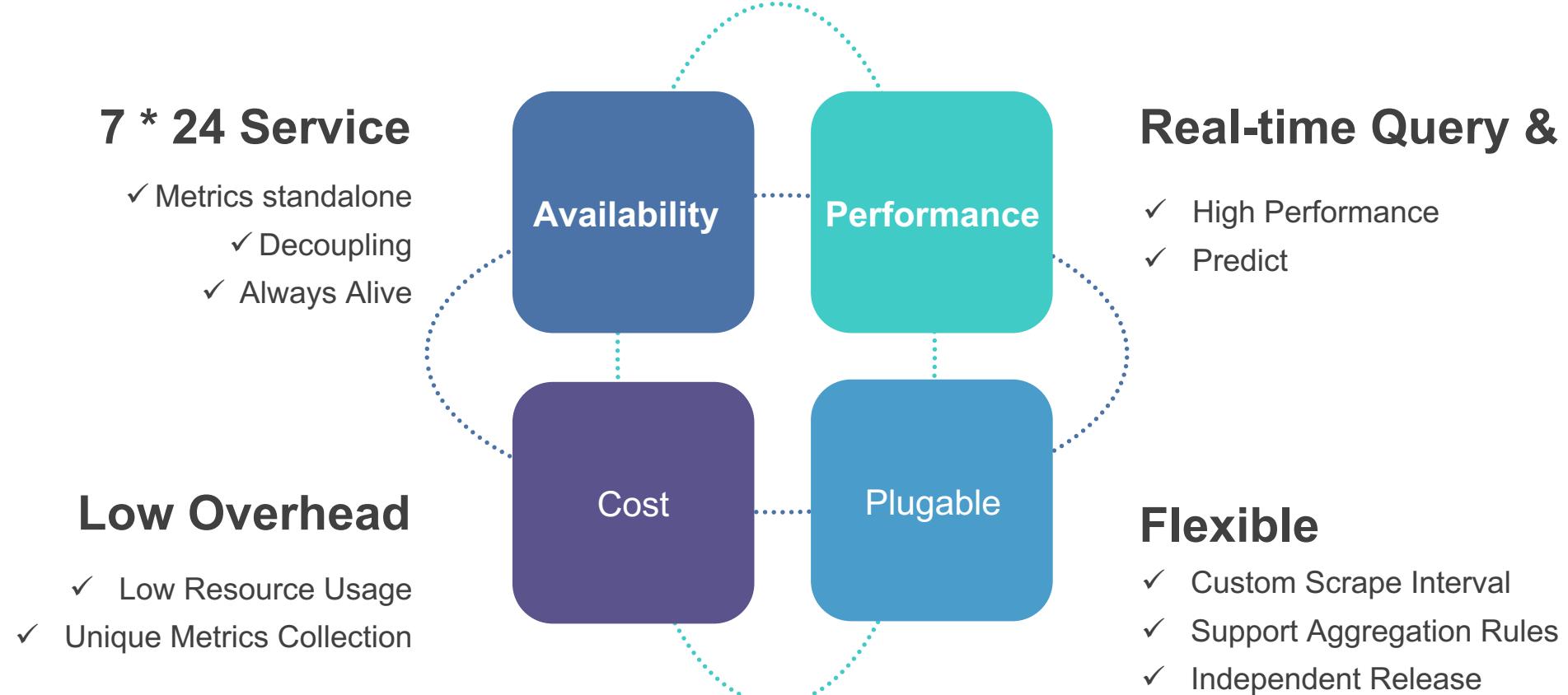
Architecture



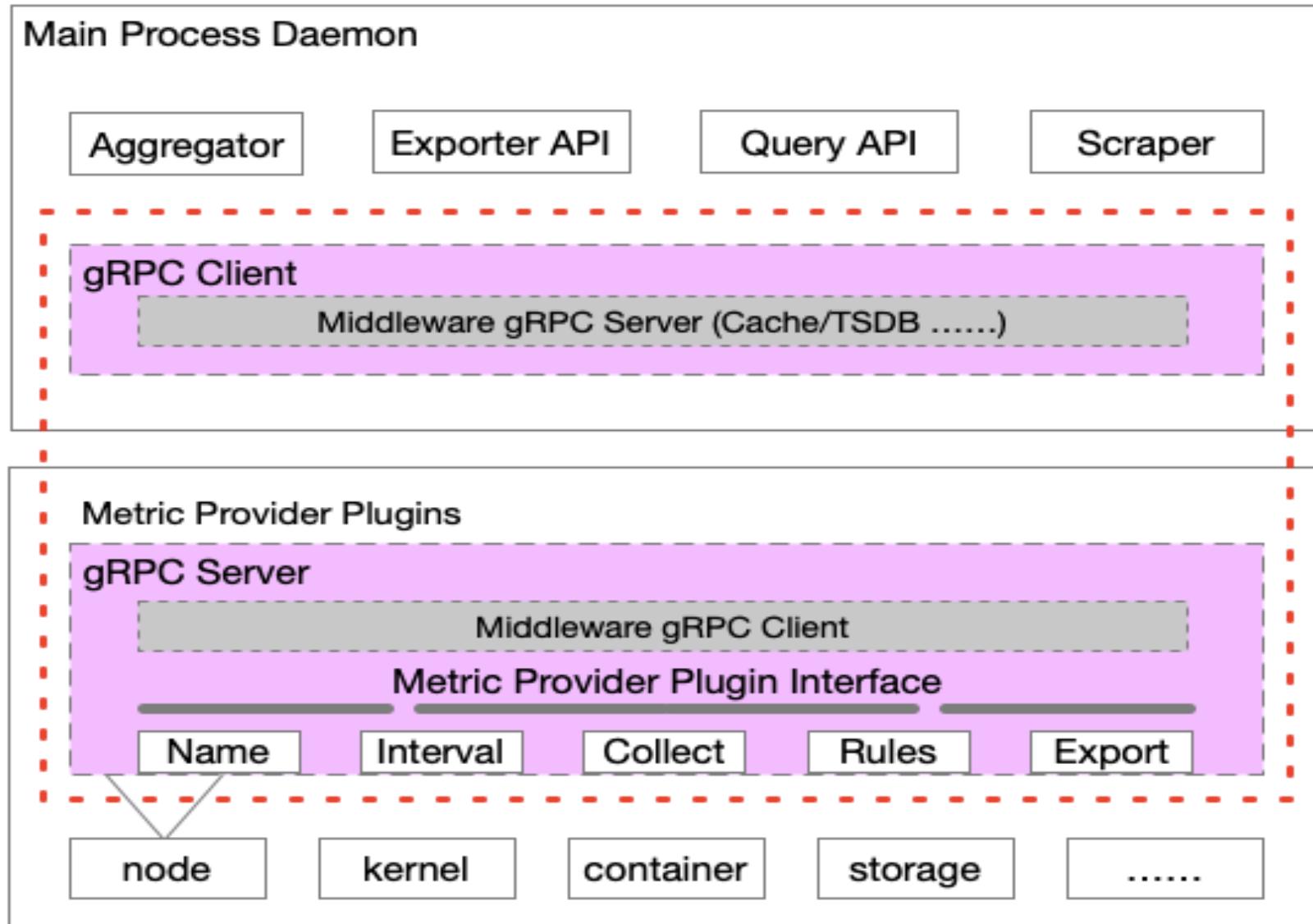


Node Agent

Design Principles



Agent Architecture



Implementation

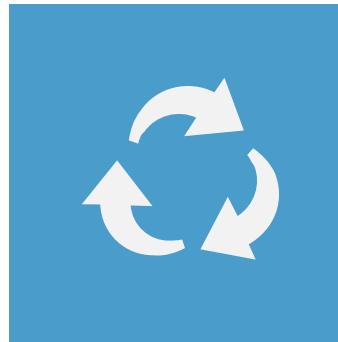
TSDB Engine

- ✓ High Performance & Low Overhead
 - ✓ Built-in Predict Function
 - ✓ TSZ Compression



HTTP API

- ✓ Support Local Query
- ✓ Prometheus Exporter



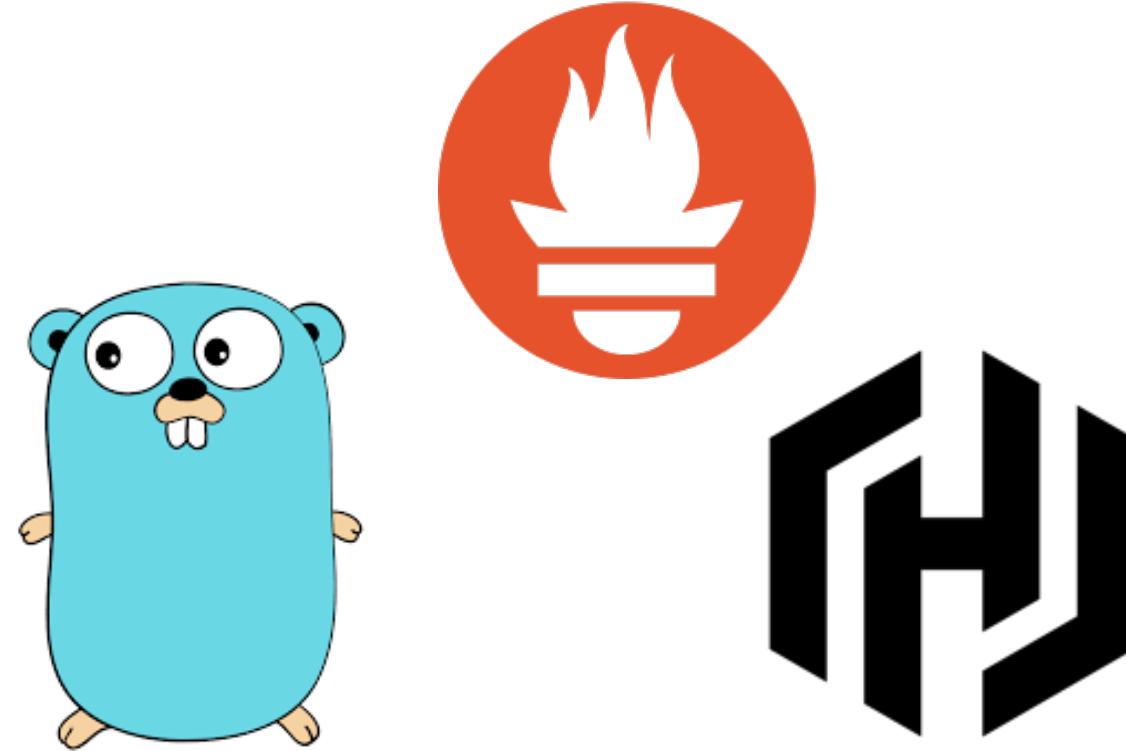
Plugins

- ✓ gRPC Plugins
- ✓ Standardized Interface
- ✓ Host + Plugins Pattern

Event Storage

- ✓ OS
- ✓ Kernel
- ✓ Hardware
- ✓ Troubleshooting

Standing on the shoulders of giants



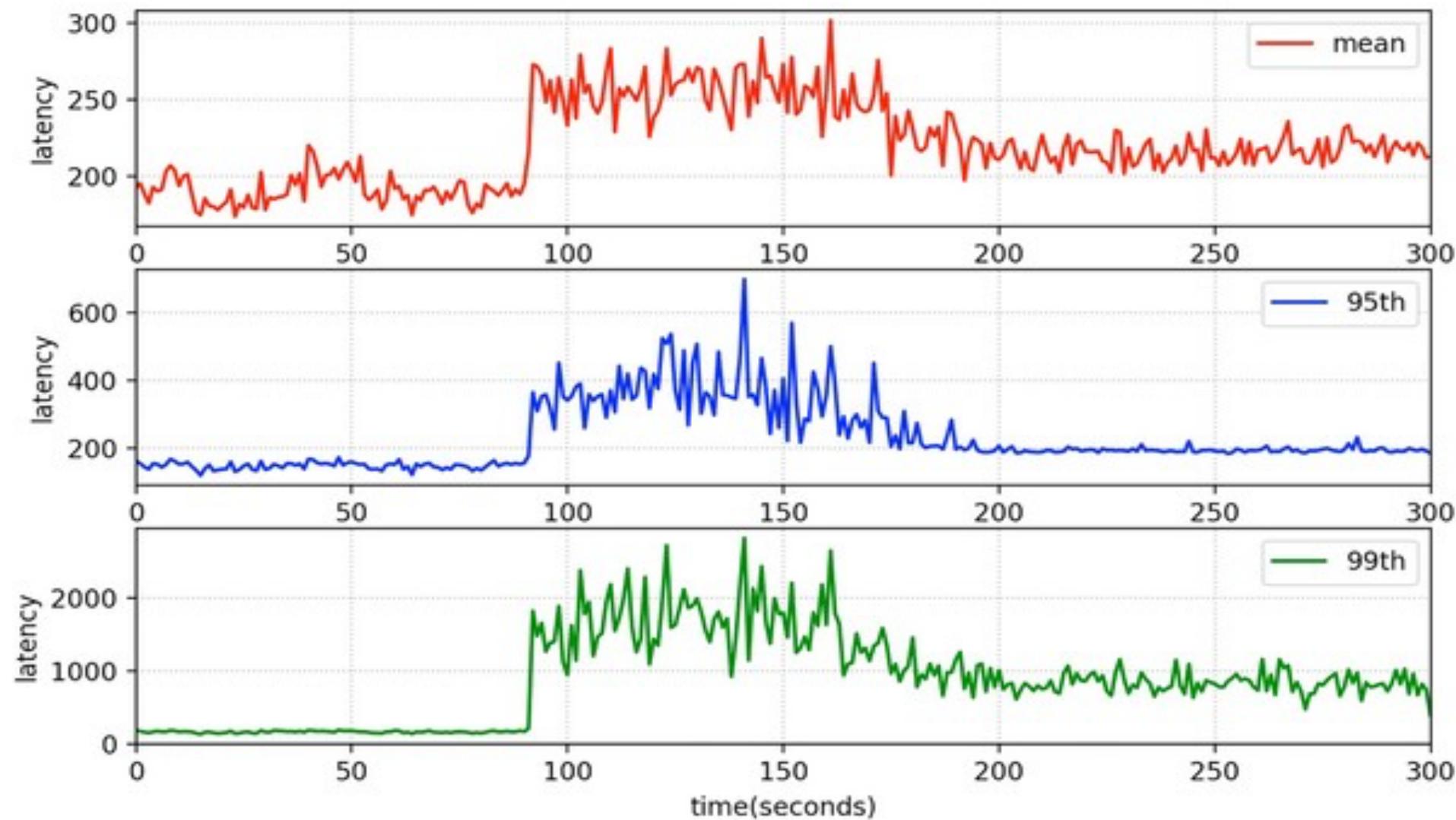
<https://golang.org/>

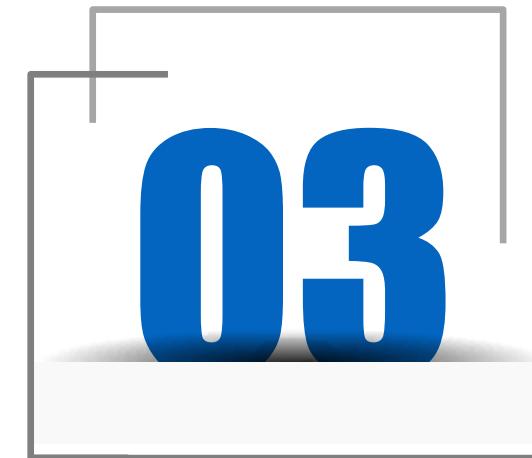
<https://github.com/hashicorp/go-plugin>

<https://github.com/prometheus/tsdb>

<https://github.com/prometheus/prometheus/tree/master/promql>

Data Achievement

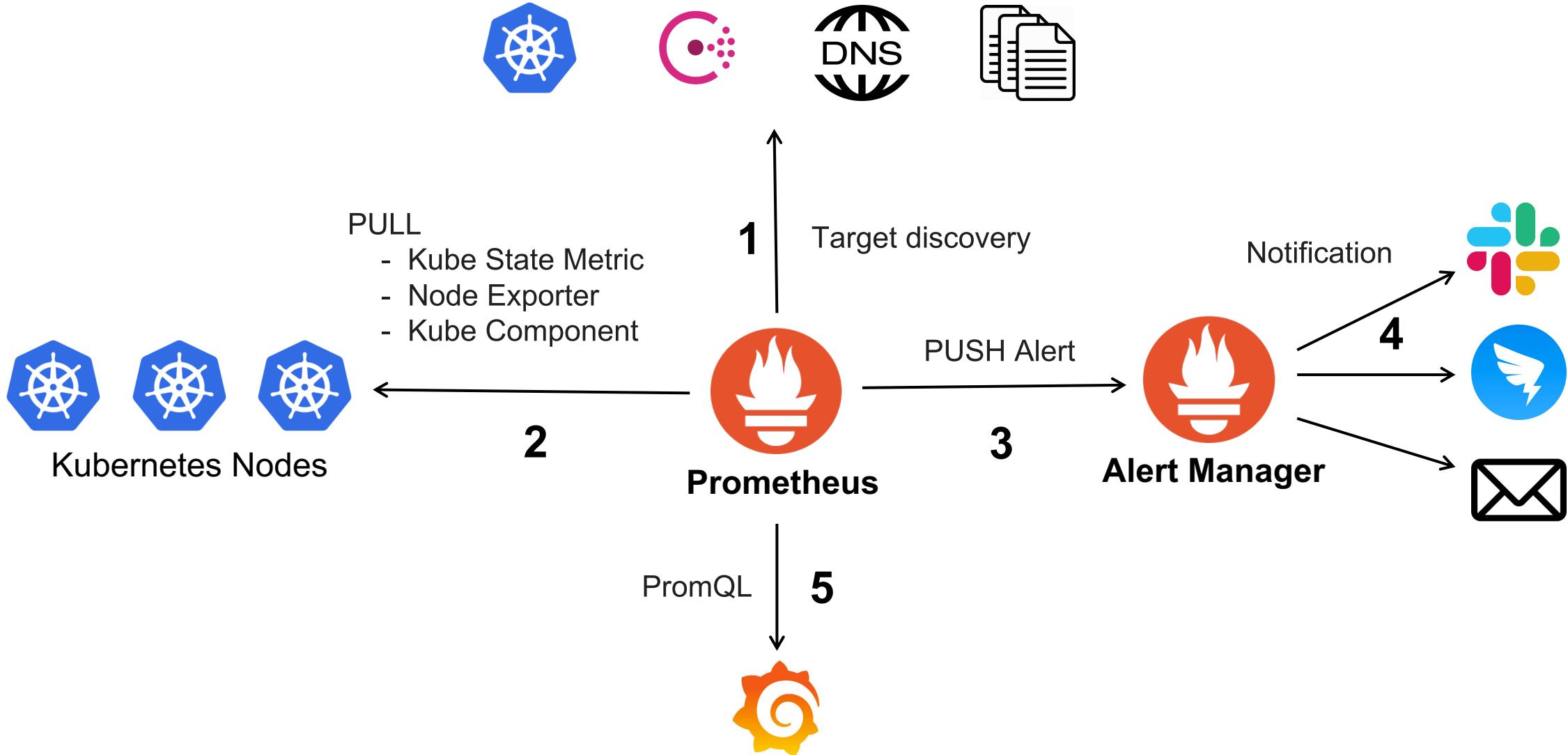




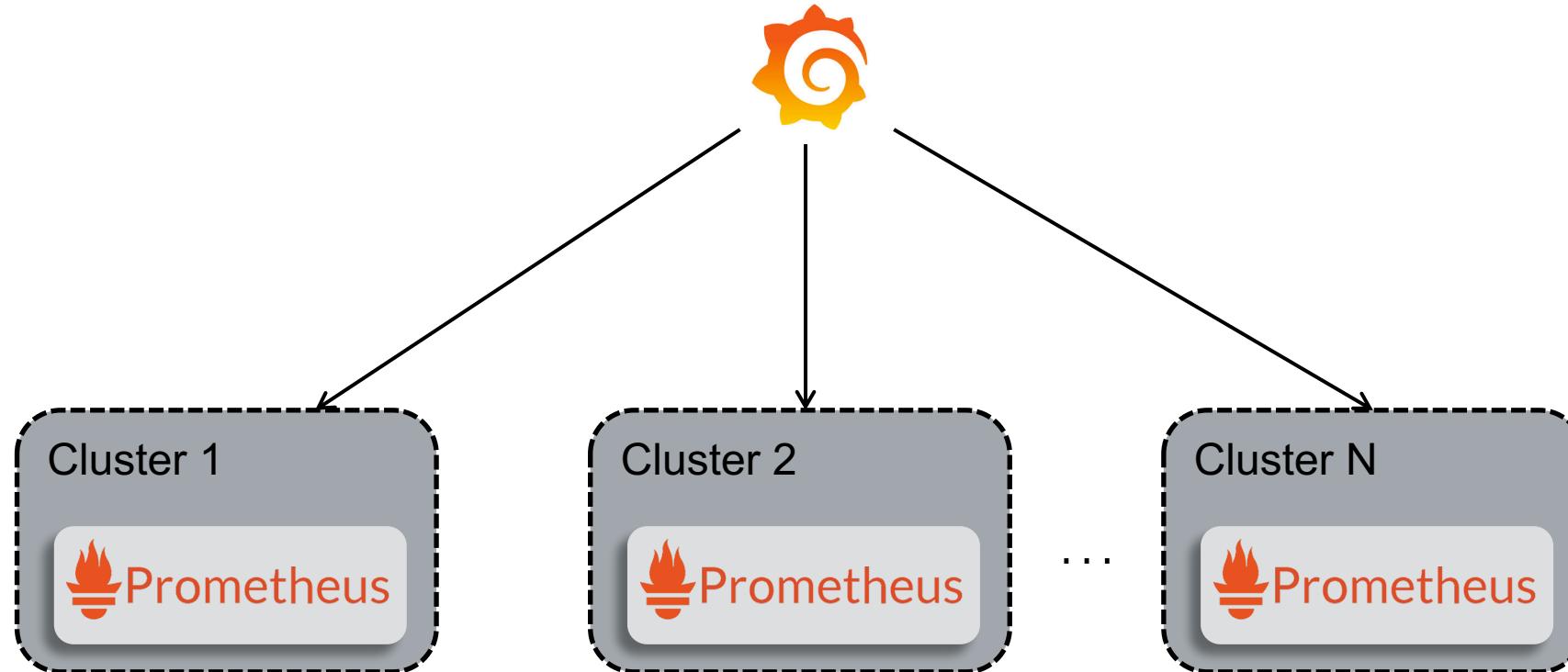
03

Prometheus & Thanos in Practice

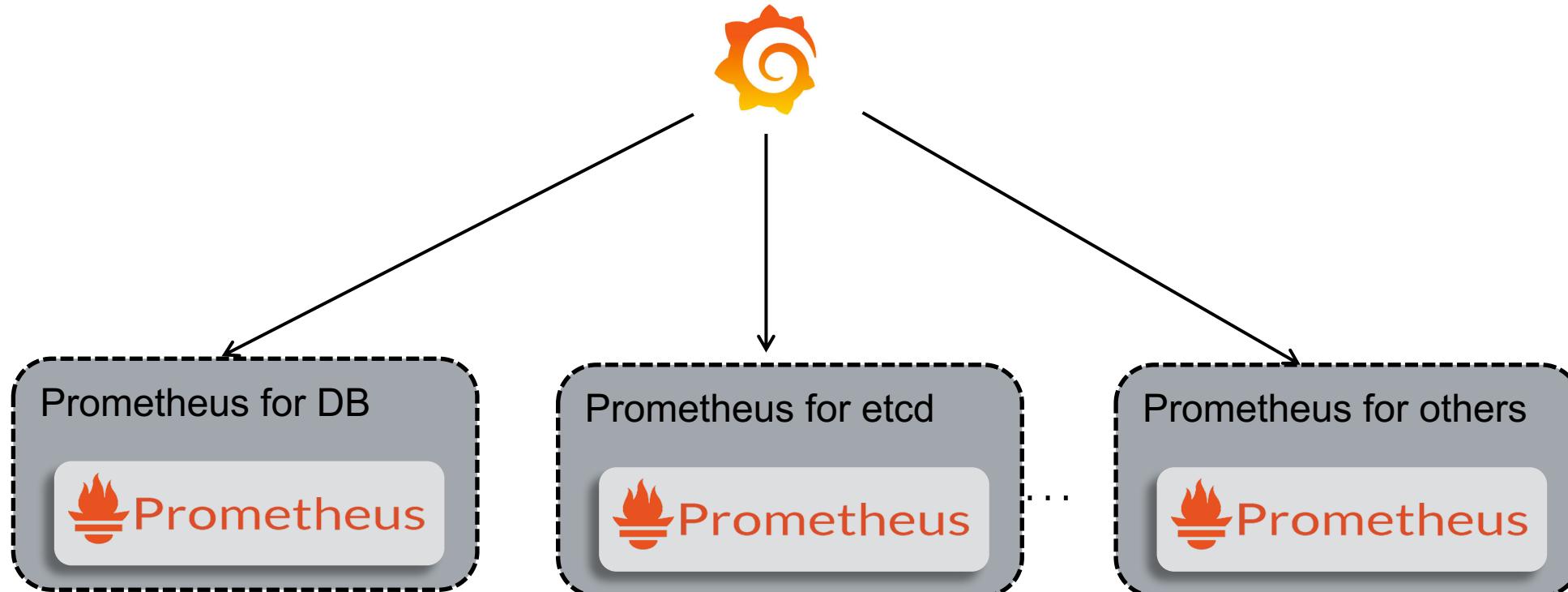
Prometheus – Architecture overview



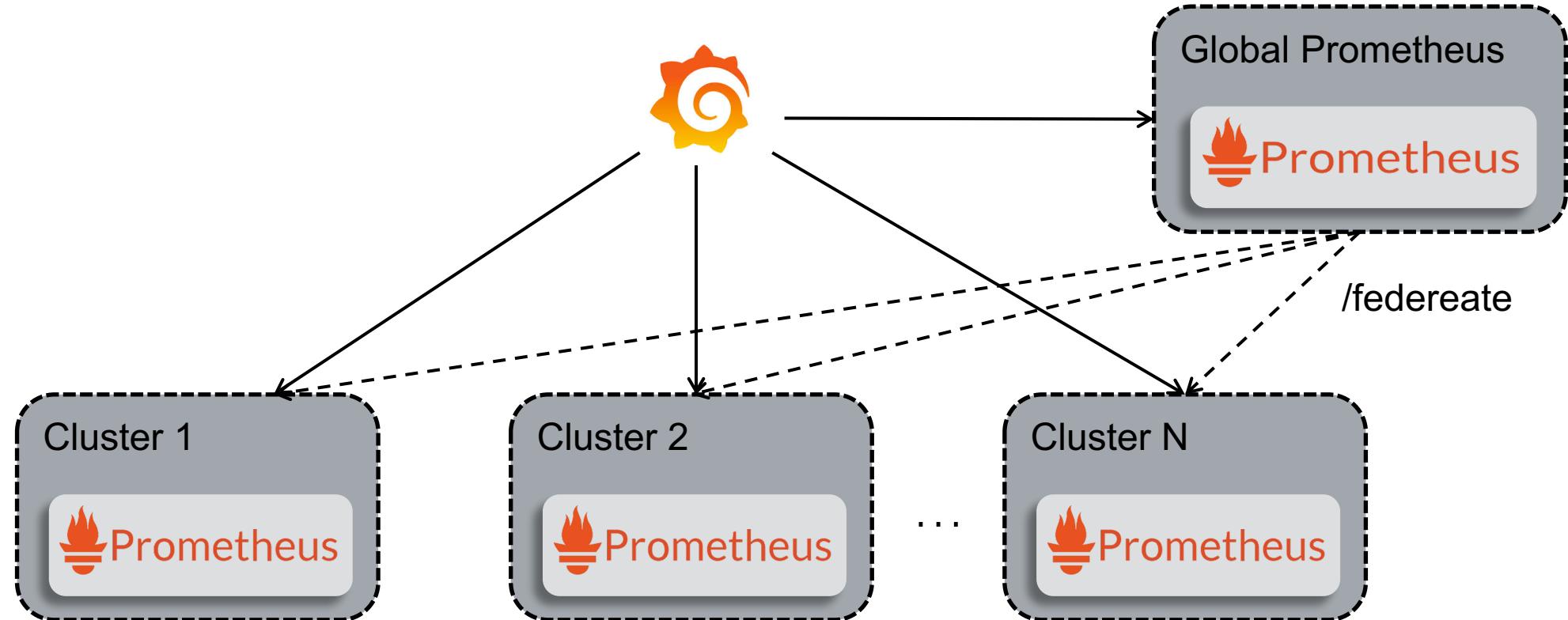
Prometheus at Scale



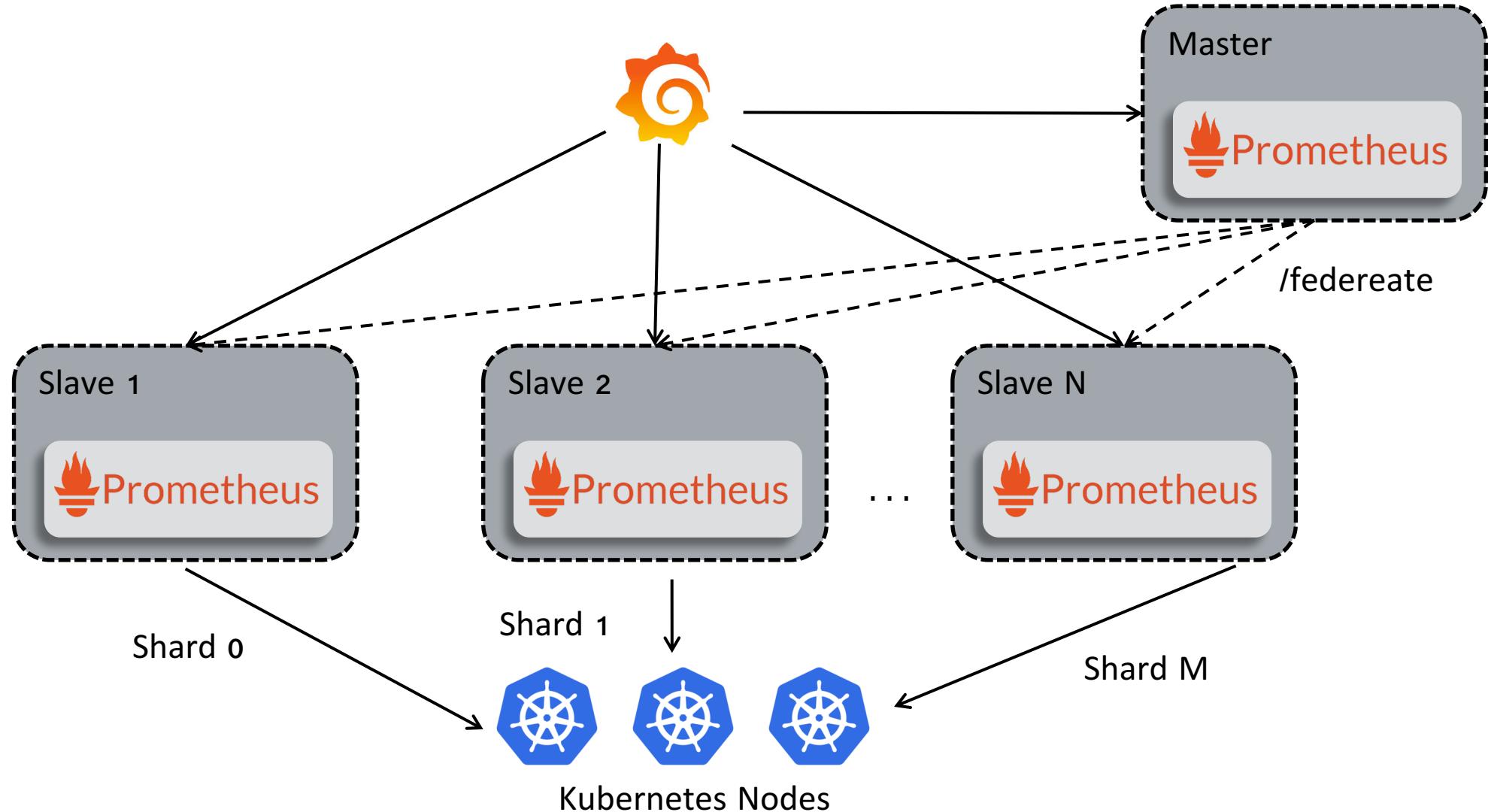
Prometheus Splitting By Use



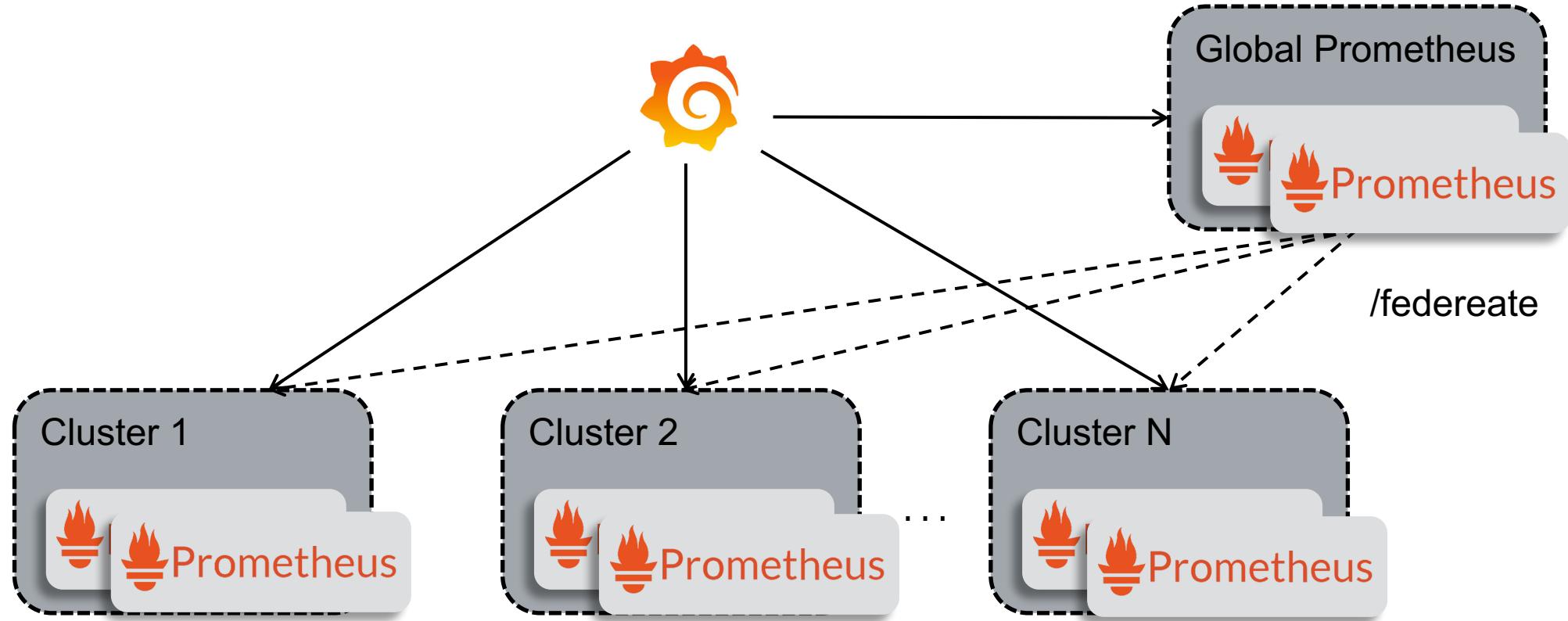
Prometheus Hierarchy



Prometheus Horizontal Sharding



Prometheus High Availability



Hello, Thanos



Highly available Prometheus setup with long term storage capabilities.



Global Query View



Unlimited Retention



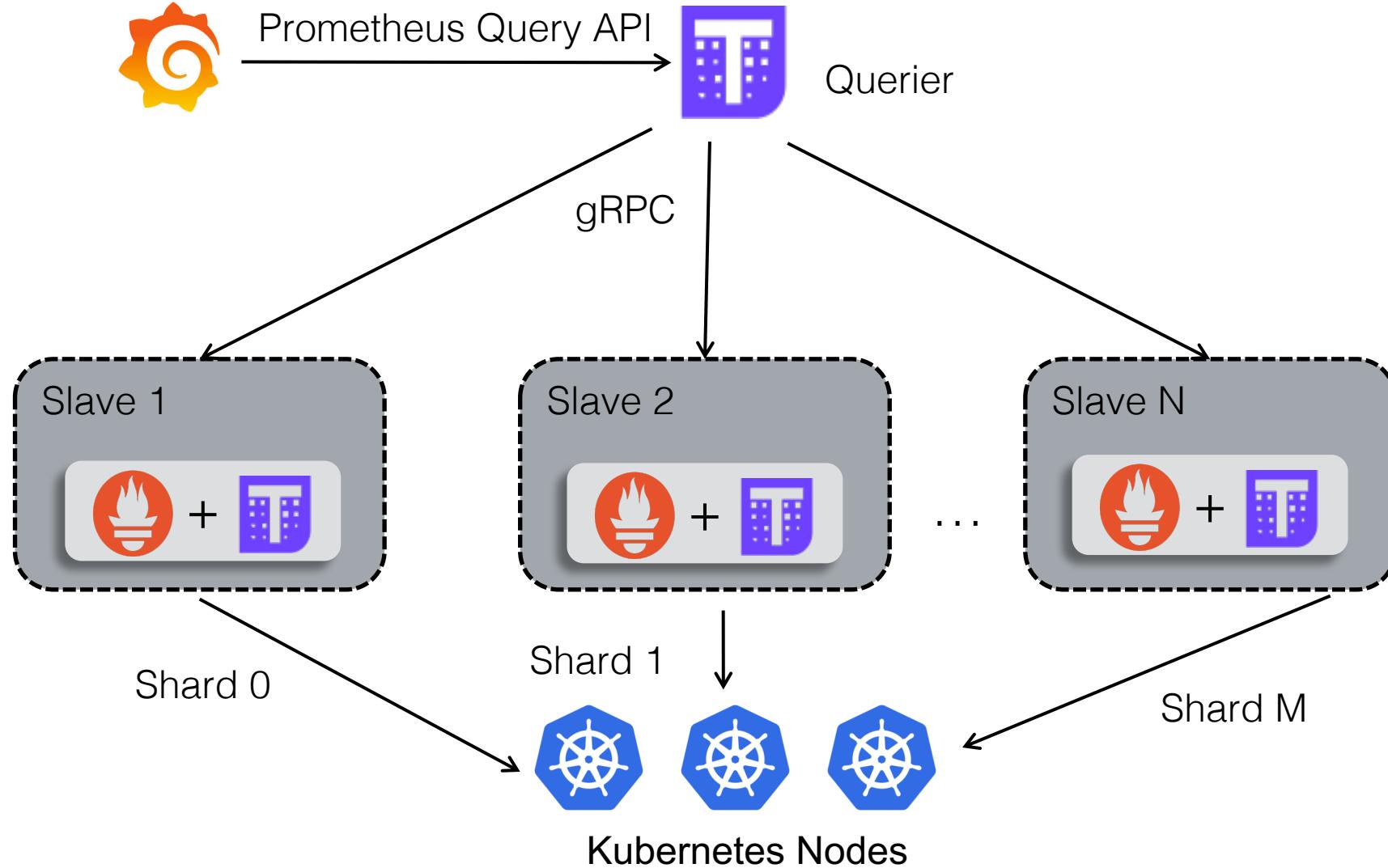
Prometheus Compatible



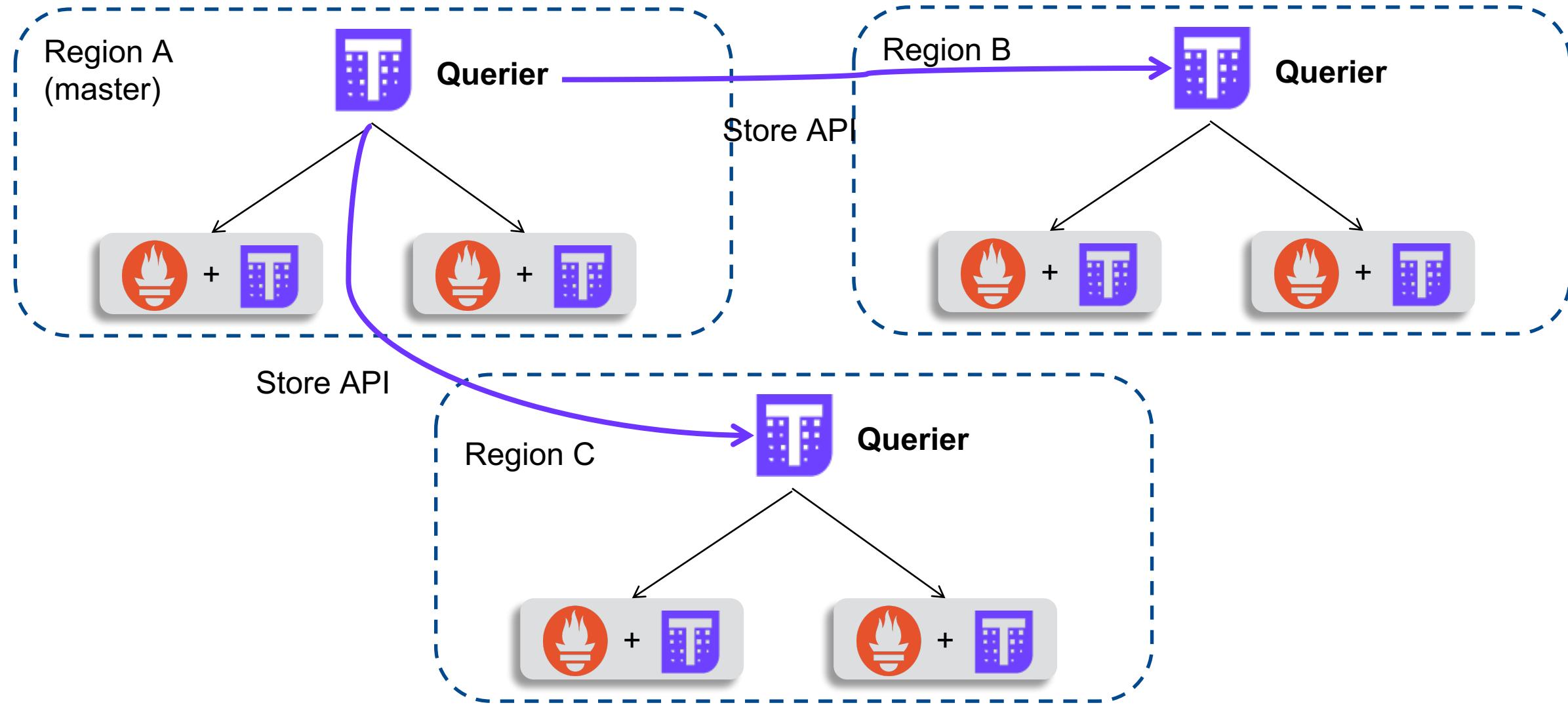
Downsampling & Compaction

<https://github.com/improbable-eng/thanos>

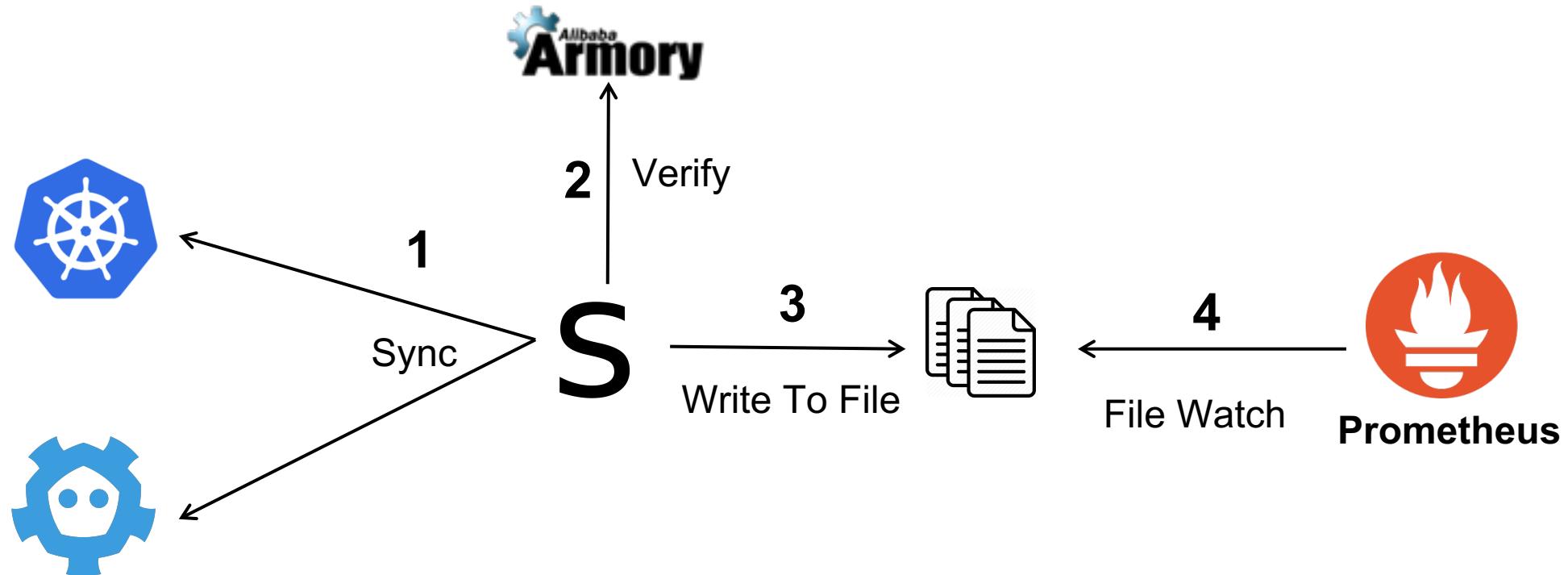
Global Querier



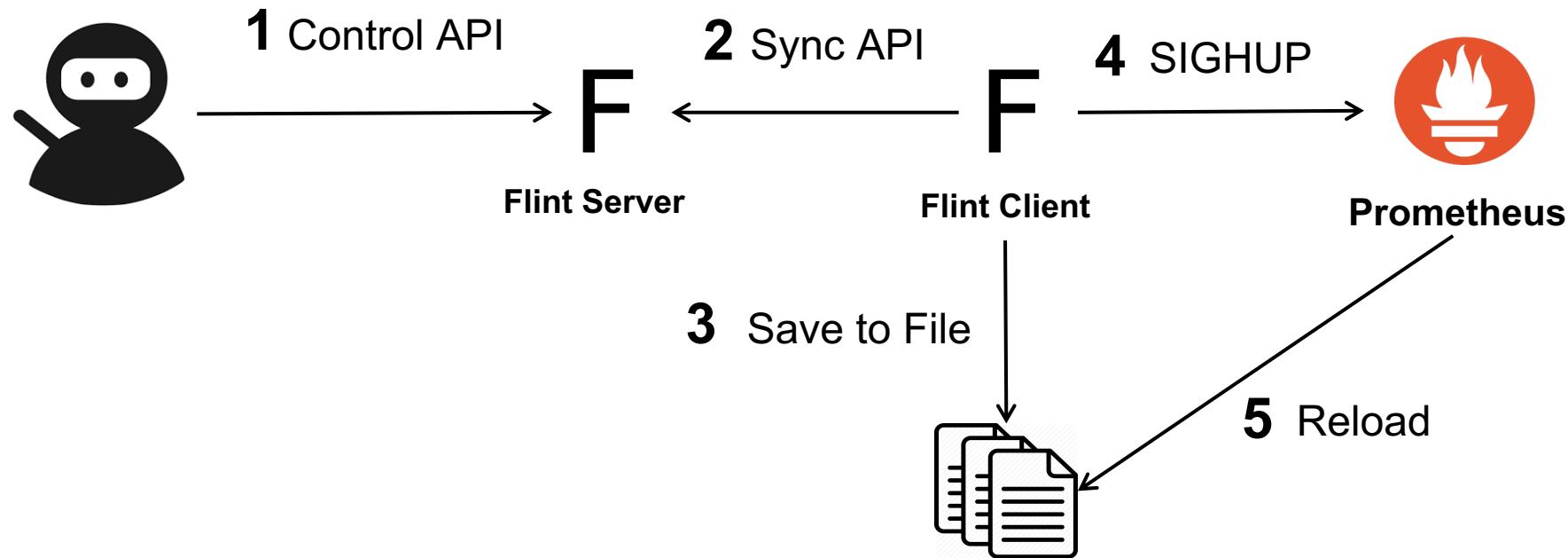
Federated Querier



Custom Target Discovery



Configuration Management



04

Thinking Future

Enhancement

- Authentication
- Query cache
- Query safeguards
- Multi Tenancy
- Integrated with Apache Flink

Challenge

- Data migration & Recovery
- Scaling along with Kubernetes

Thanks !



奥运会全球指定云服务商