



2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI

2023 · 上海

SkyWalking Summit





霍秉杰

青云科技可观测性与边缘计算团队负责人

“ KubeSphere 在可观测性领域的
探索与实践 ”

目录

CONTENTS

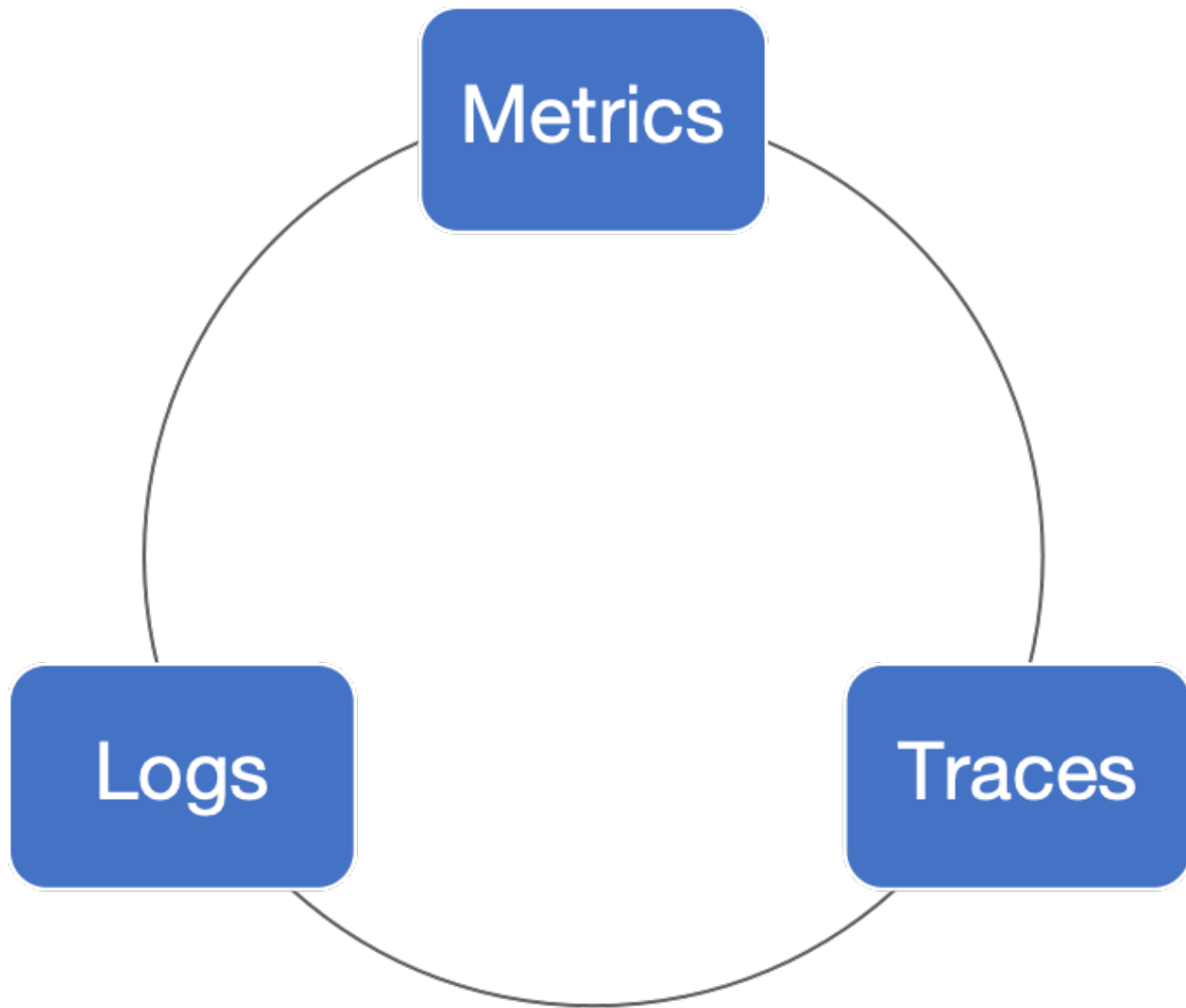
01. 可观测性的 3 大支柱？
02. KubeSphere 开源可观测性领域的实践
03. KubeSphere 商业可观测性产品的实践
04. Future Plan

01

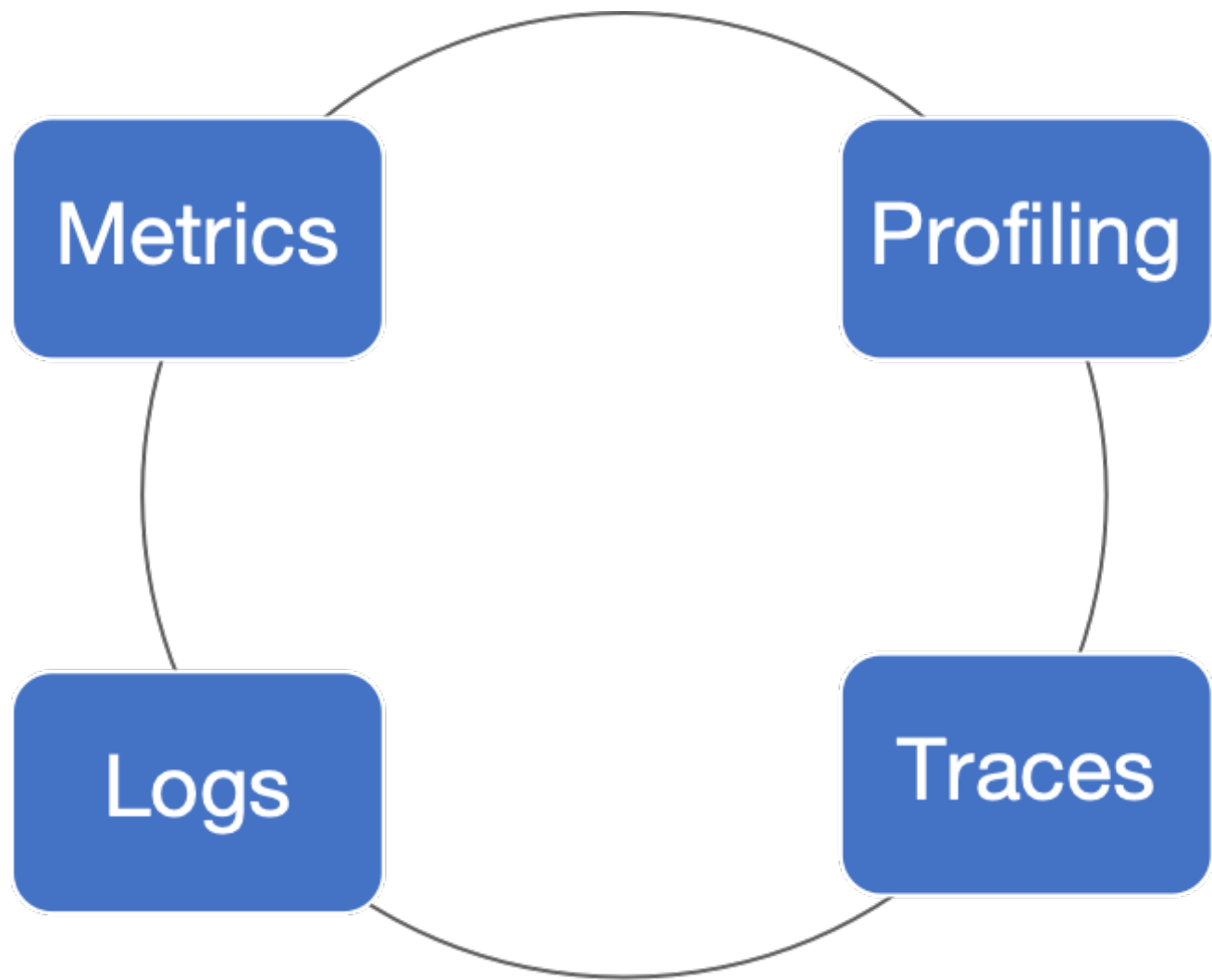
可观测性的 3 大支柱？

可观测性的 3 大支柱？

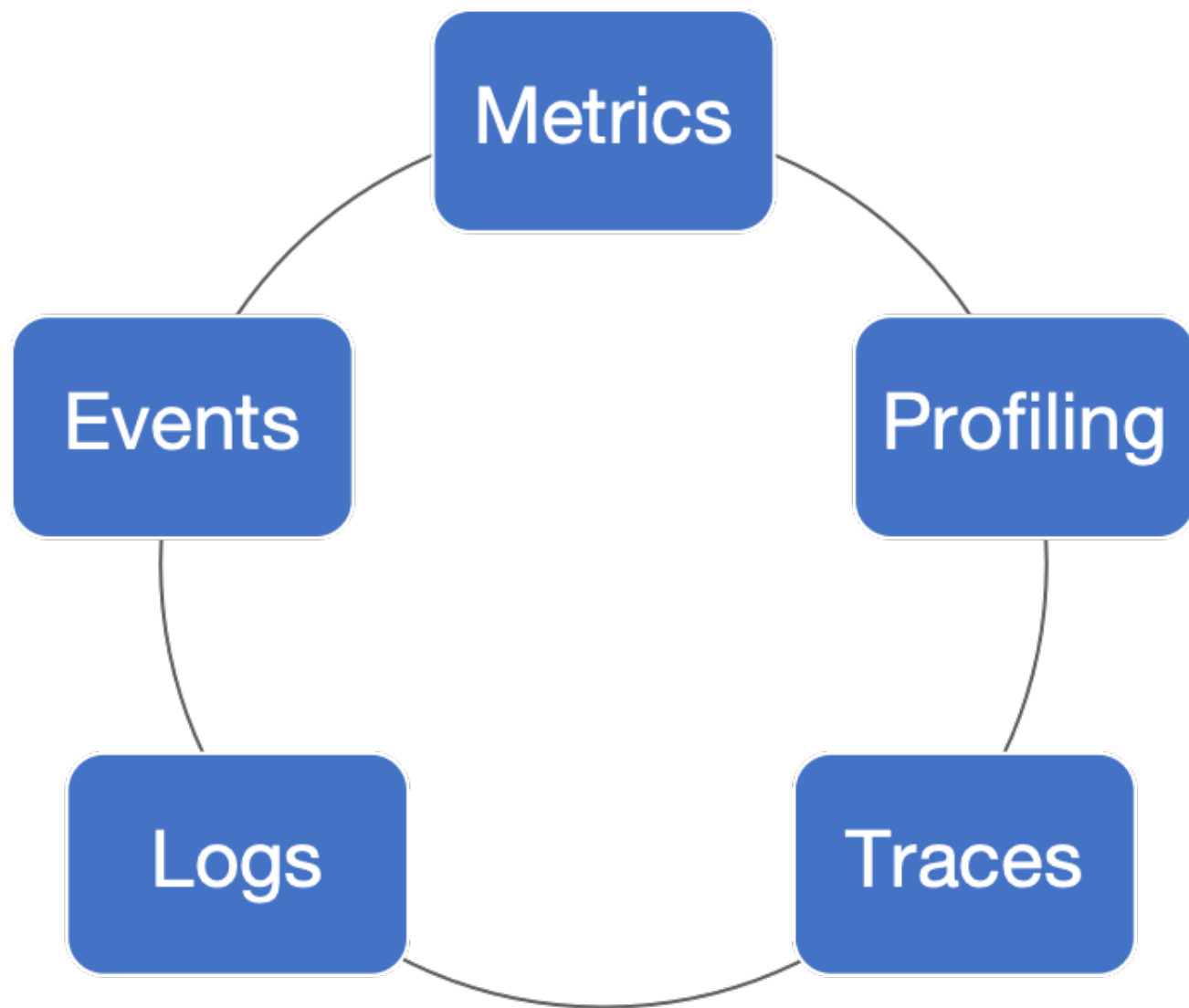
2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI



可观测性的 3 大支柱？



可观测性的 3 大支柱？

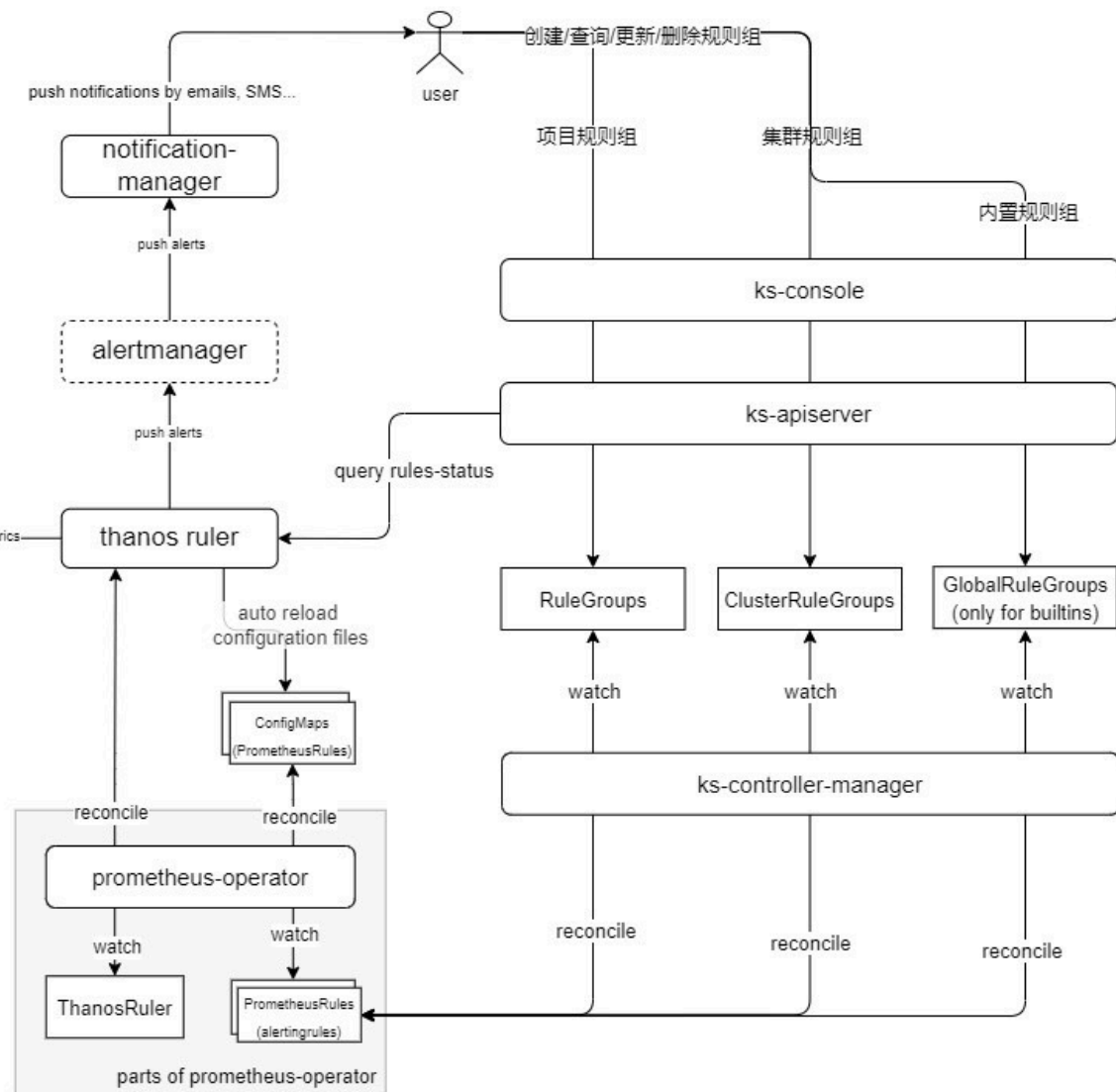
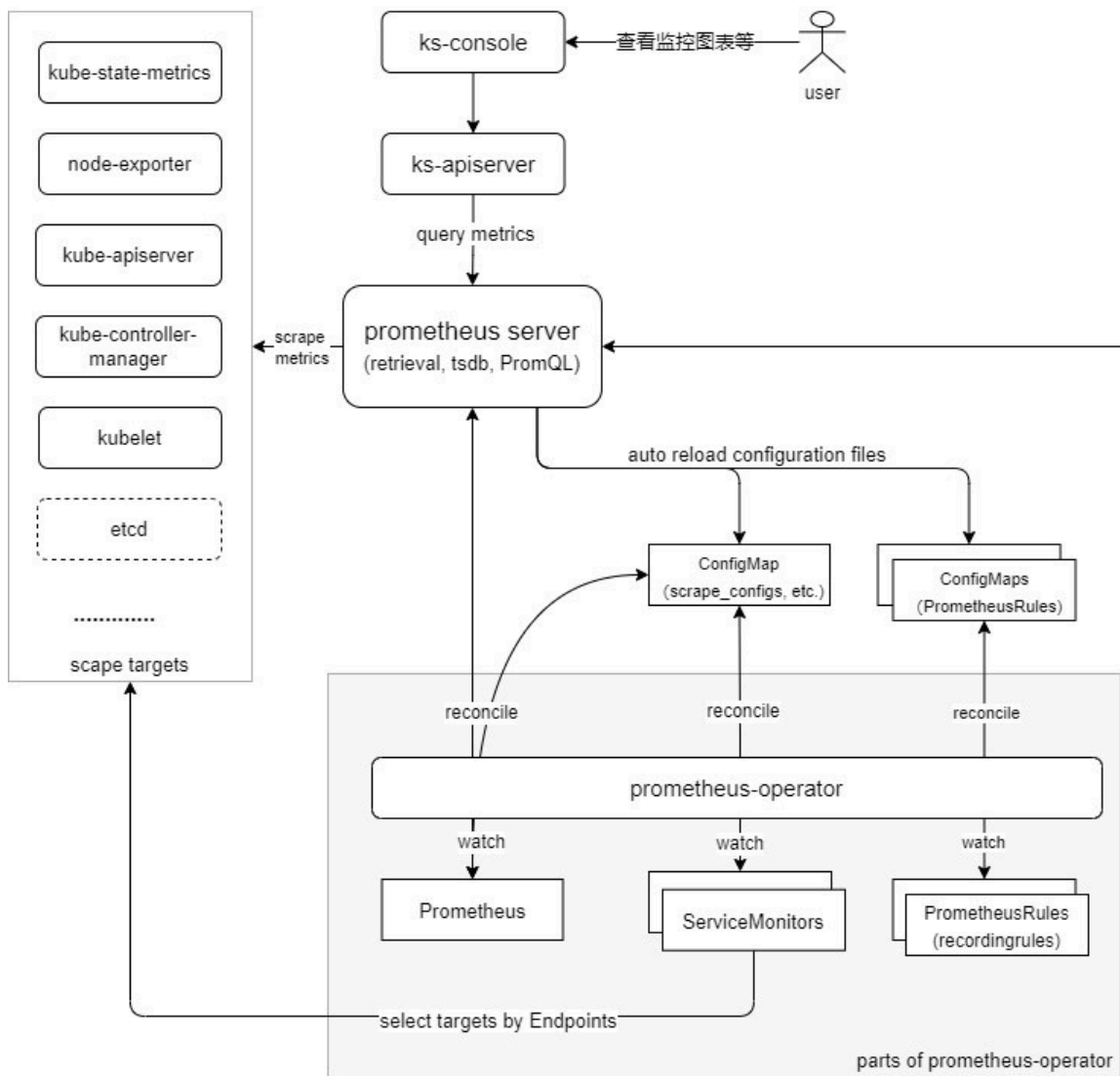


02

KubeSphere 开源可观测性领域的实践

KubeSphere Monitoring

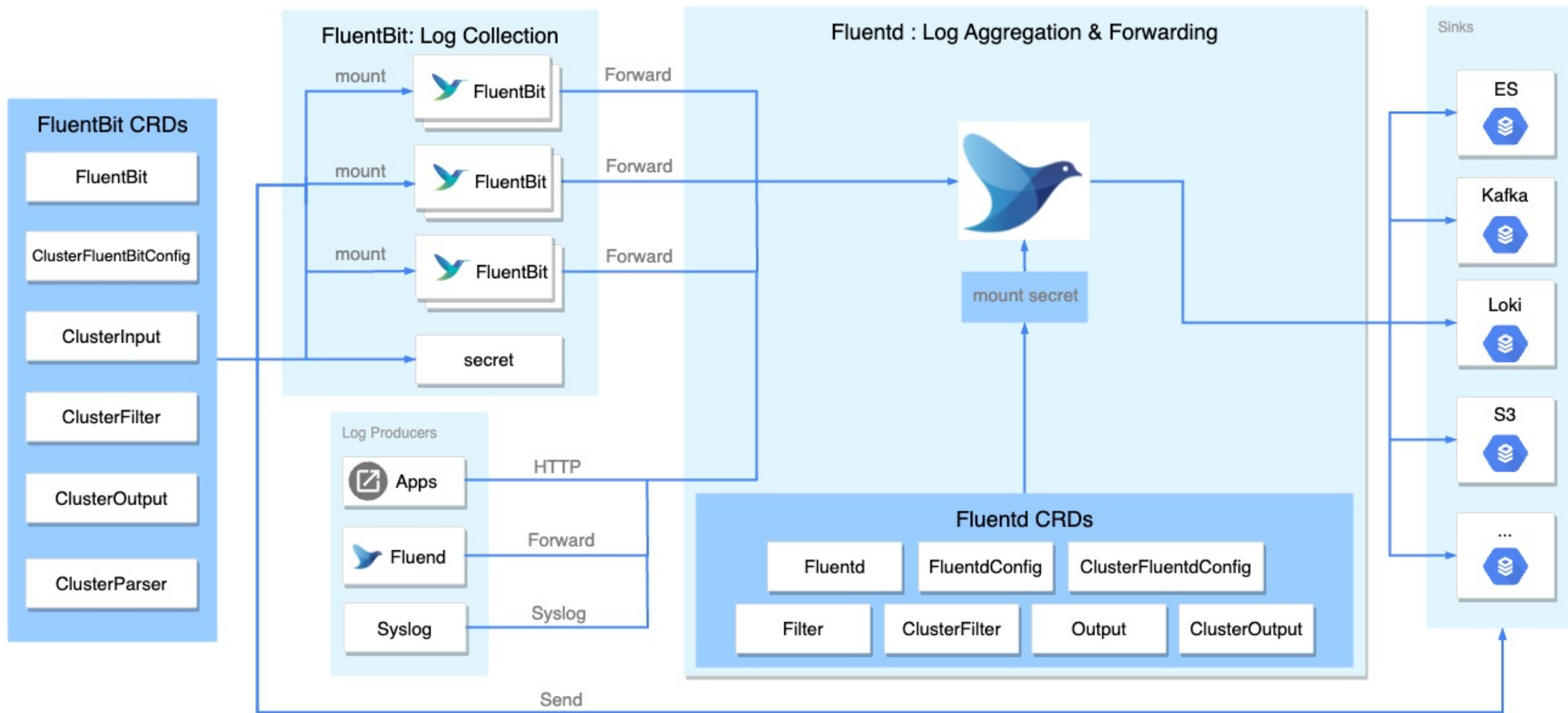
2023 | APACHE • SkyWalking
SUMMIT | CHINA • SHANGHAI



Logging: Fluent-Operator

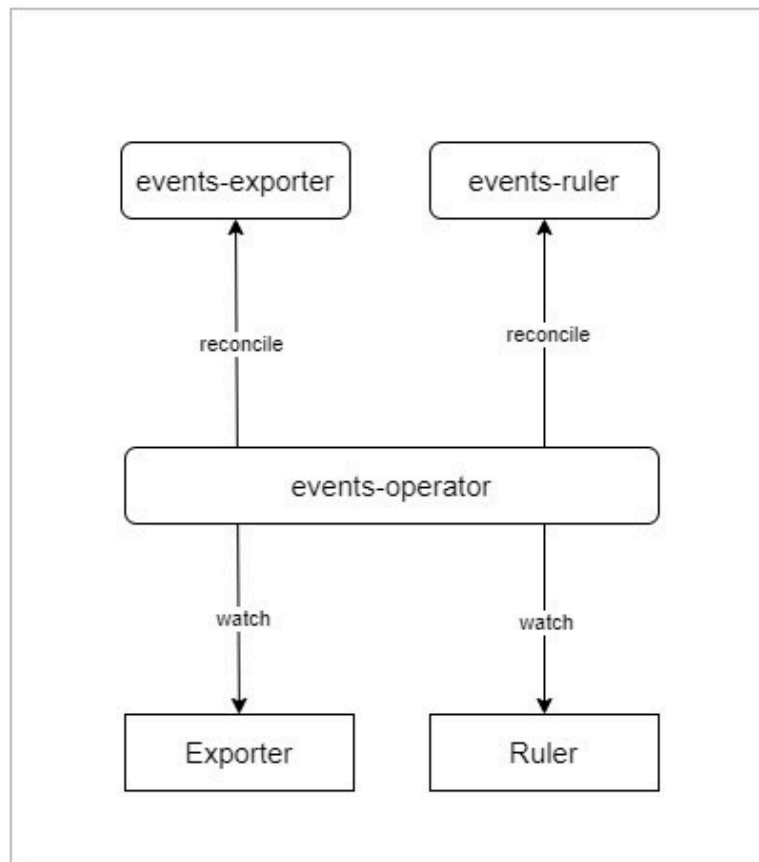
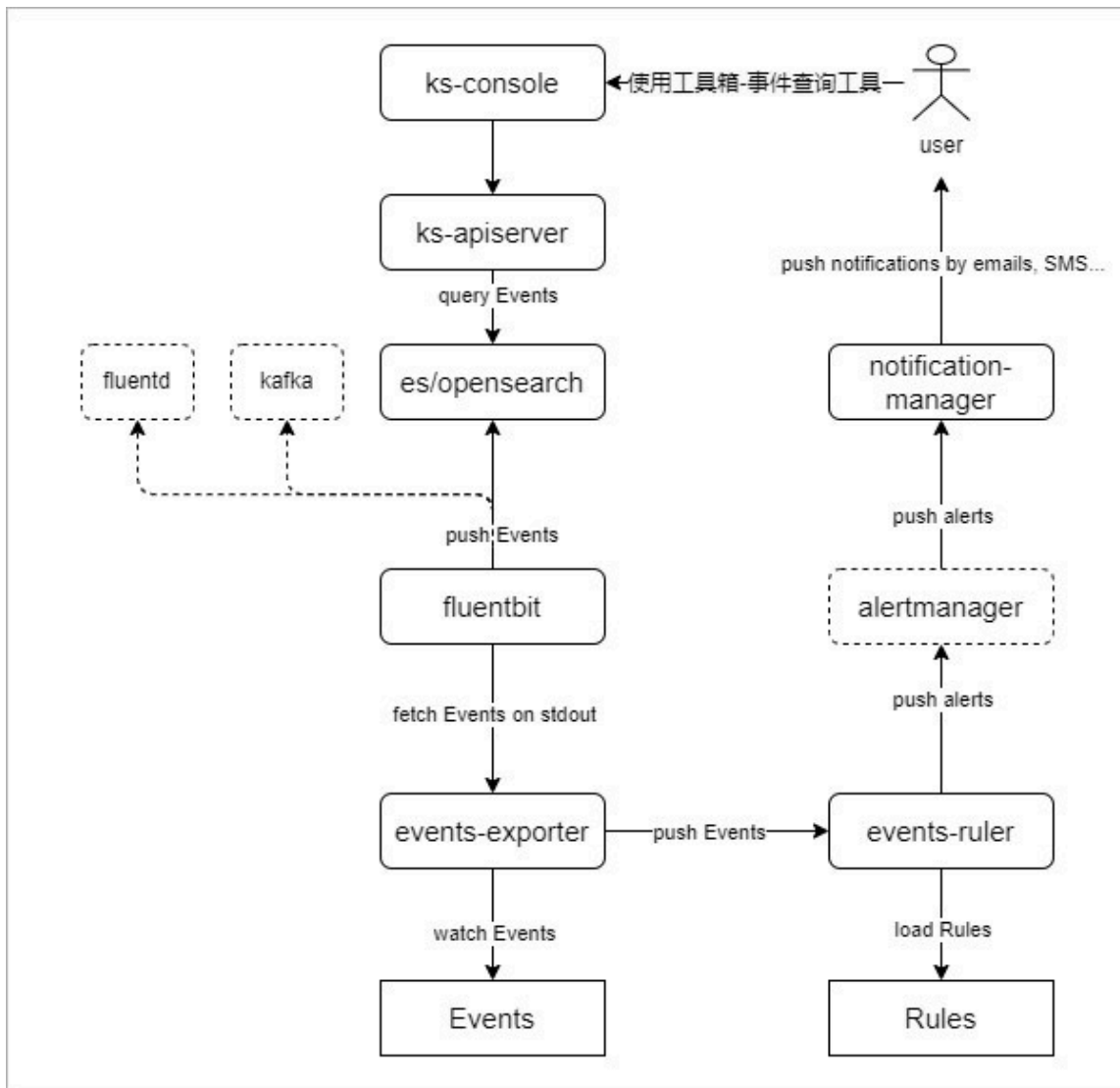
2023 | APACHE • SkyWalking
SUMMIT | CHINA • SHANGHAI

Fluent Operator : Build Cloud Native Log Processing Pipeline



<https://github.com/fluent/fluent-operator/>

Events : kube-events



Events : kube-events

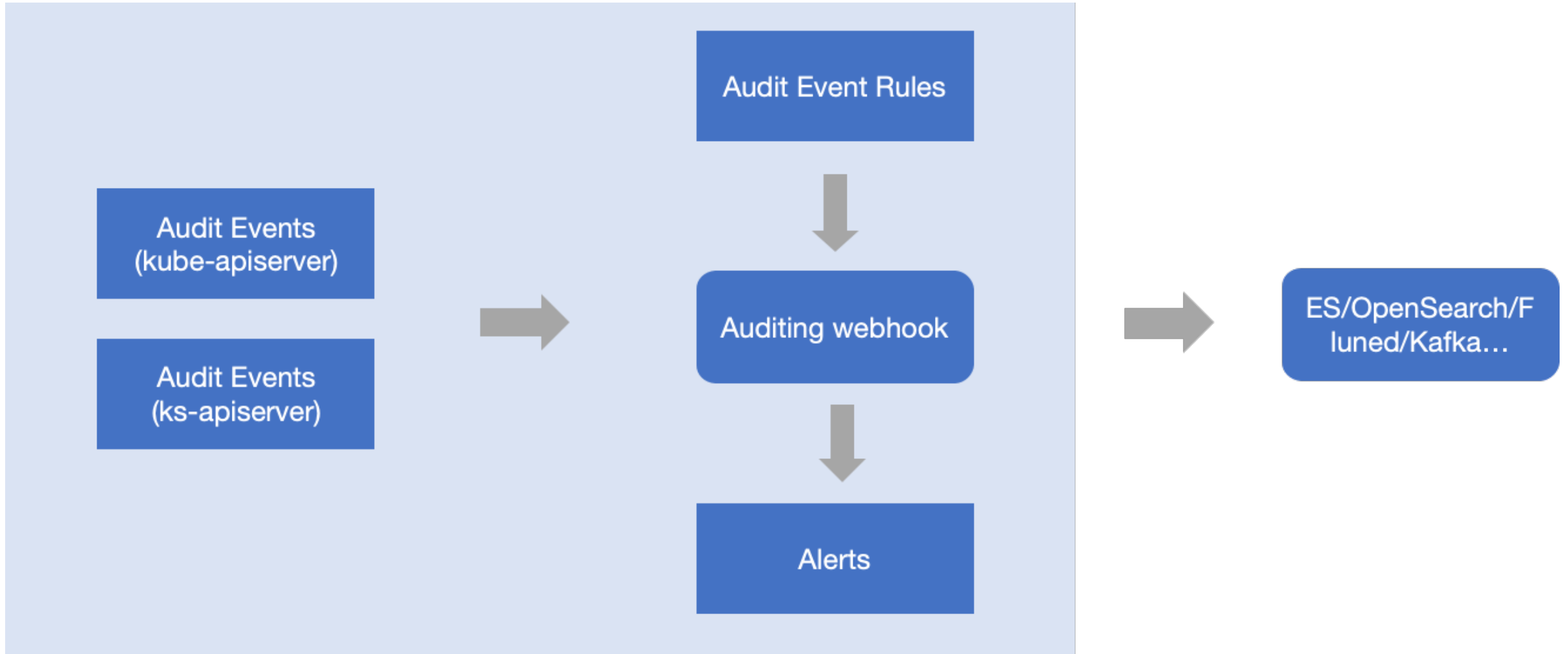
2023
SUMMIT

APACHE • SkyWalking
CHINA · SHANGHAI

```
apiVersion: events.kubesphere.io/v1 alpha1
kind: Rule
metadata:
  name: demo
  namespace: demo
  labels:
    kubesphere.io/rule-scope: cluster
spec:
  rules:
    - name: PodNetworkNotReady
      condition: |
        type="Warning" and involvedObject.kind="Pod" and
reason="NetworkNotReady"
      labels:
        severity: warning
      annotations:
        message: '%message'
        summary: Pod network is not ready
        summaryCn: Pod网络异常
      enable: true
      type: alert
```

Auditing: kube-auditing

2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI

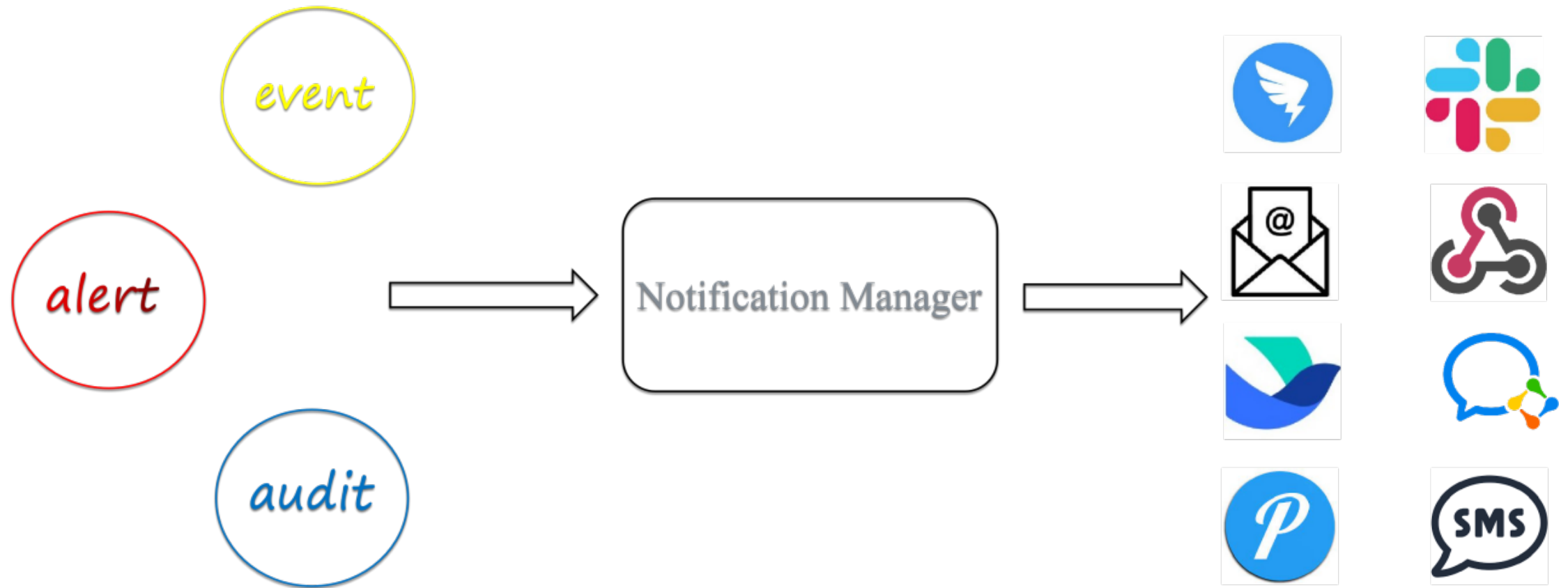


Auditing: kube-auditing

```
apiVersion: auditing.kubesphere.io/v1alpha1
kind: Rule
metadata:
  generation: 1
  labels:
    type: archiving
  name: archiving-rule
  namespace: kubesphere-logging-system
spec:
  rules:
    - desc: all action not need to be audit
      list:
        - get
        - list
        - watch
      name: ignore-action
      type: list
    - condition: Verb not in ${ignore-action}
      desc: All audit event except get, list, watch event
      enable: true
      name: archiving
      priority: DEBUG
      type: rule
```

Notification: Notification Manager

2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI



<https://github.com/kubesphere/notification-manager>

Notification: Notification Manager

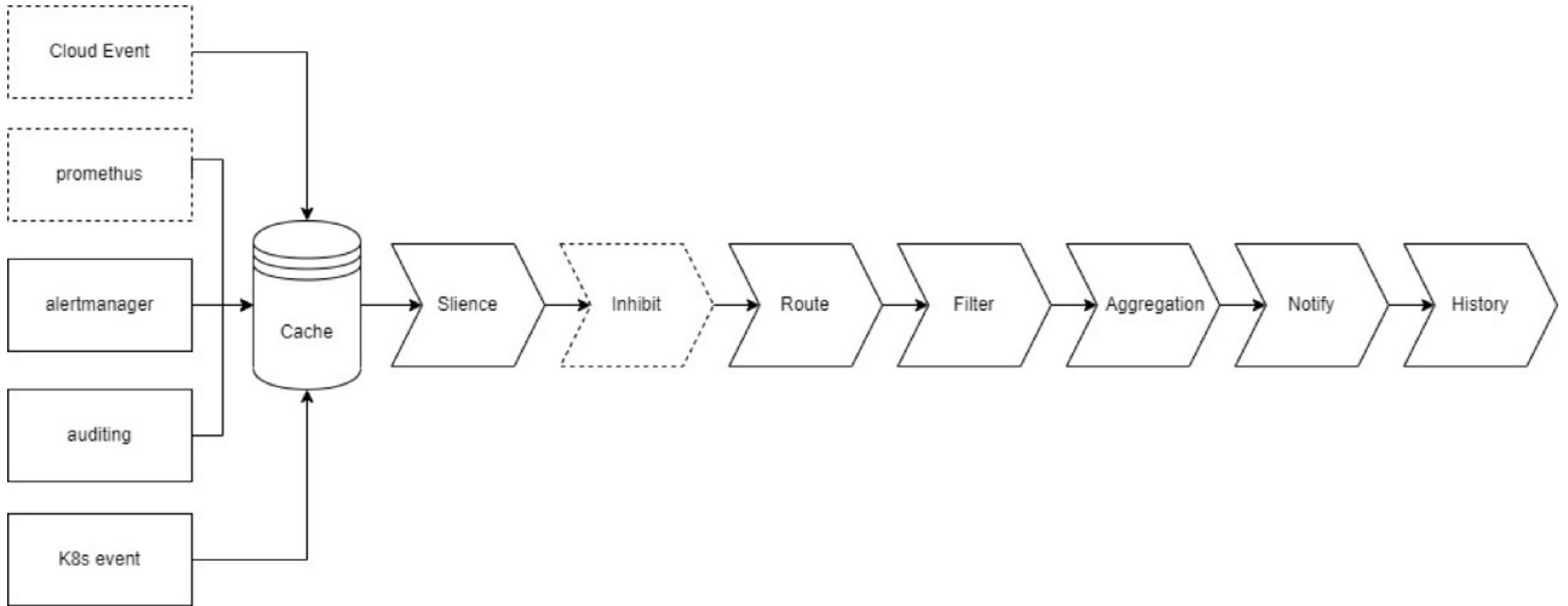
2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI

```
apiVersion: notification.kubesphere.io/v2beta2
kind: Config
metadata:
  name: default-config
  labels:
    type: default
spec:
  email:
    hello: "hello"
    authIdentify: nil
    authPassword:
      valueFrom:
        secretKeyRef:
          key: password
          name: default-config-secret
          namespace: kubesphere-monitoring-system
    authUsername: test
    from: test@kubesphere.io
    requireTLS: true
    smartHost:
      host: imap.kubesphere.io
      port: 25
    tls: {}
```

```
apiVersion: notification.kubesphere.io/v2beta2
kind: Receiver
metadata:
  name: global-receiver
  labels:
    type: tenant
    user: test
spec:
  email:
    alertSelector:
      matchExpressions:
        - key: namespace
          operator: DoesNotExist
    emailConfigSelector:
      matchLabels:
        type: tenant
        user: test
    enabled: true
    template: nm.default.html
    subjectTemplate: nm.default.subject
    tplType: html
    tplText:
      name: notification-manager-template
      namespace: kubesphere-monitoring-system
    to:
      - test@kubesphere.io
```


Notification: Notification Manager

2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI



<https://github.com/kubesphere/notification-manager>

03

KubeSphere 商业可观测性产品的实践

Prometheus Long-Term Storage 的崛起

2023
SUMMIT

APACHE • SkyWalking
CHINA · SHANGHAI

Remote Endpoints and Storage

The [remote write](#) and [remote read](#) features of Prometheus allow transparently sending and receiving samples. This is primarily intended for long term storage. It is recommended that you perform careful evaluation of any solution in this space to confirm it can handle your data volumes.

- [AppOptics](#): write
- [AWS Timestream](#): read and write
- [Azure Data Explorer](#): read and write
- [Azure Event Hubs](#): write
- [Chronix](#): write
- [Cortex](#): read and write
- [CrateDB](#): read and write
- [Elasticsearch](#): write
- [Gnocchi](#): write
- [Google BigQuery](#): read and write
- [Google Cloud Spanner](#): read and write
- [Grafana Mimir](#): read and write
- [Graphite](#): write
- [InfluxDB](#): read and write
- [Instana](#): write
- [IRONdb](#): read and write
- [Kafka](#): write
- [M3DB](#): read and write
- [New Relic](#): write
- [OpenTSDB](#): write
- [PostgreSQL/TimescaleDB](#): read and write
- [QuasarDB](#): read and write
- [SignalFx](#): write
- [Splunk](#): read and write
- [Sysdig Monitor](#): write
- [TiKV](#): read and write
- [Thanos](#): read and write
- [VictoriaMetrics](#): write
- [Wavefront](#): write

Prometheus 2017 年加入了 remote read/write

API, 成了 Prometheus 社区创新的起点

Prometheus Long-Term Storage 对比

2023
SUMMIT

APACHE • SkyWalking
CHINA · SHANGHAI

	M3	VictoriaMetrics	Thanos	Cortex	Mimir
Author	Uber	VictoriaMetrics	Improbable & Redhat	Grafana	Grafana
License	Apache-2.0	Apache-2.0	Apache 2.0	Apache 2.0	AGPL-3.0
中立性	Open Source	Open Core Commercial	CNCF incubation	CNCF incubation	Open Core Commercial
Stars	4.2k	6.8k	10.7k	4.8k	2.2k
Forks	405	648	1.6k	693	146
支持分布式模式	No	No	通过 Thanos Sidecar 支持	No	No
支持 Agent	?	vmagent	Prometheus Agent	Grafana/Prometheus Agent	Grafana/Prometheus Agent
社区及采用者	N/A	Very Popular	Very Popular	Popular	New Player
数据接入	M3 Coordinator	vminsert	Thanos Receive Router/Ingester	Distributor/Ingester	Distributor/Ingester
数据持久化	磁盘	磁盘	对象存储	对象存储	对象存储
数据查询	M3 Queries	vmselect	Thanos Querier/Query Frontend	Querier/Query Frontend	Querier/Query-Scheduler/Query-Frontend
规则计算与告警	N/A	vmalert	Thanos Ruler	Ruler	Ruler
存储管理与降采样	M3 Aggregator	降采样为收费功能	Thanos Compactor	Compactor	Compactor
数据接入可扩展性	Good	Very Good	Very Good	Very Good	Very Good
存储可扩展性	Good	Good	Very Good	Very Good	Very Good
数据查询可扩展性	Good	Very Good	Very Good	Very Good	Perfect
规则计算可扩展性	Poor	Good	Good	Good	Very Good
降采样可扩展性	Poor	降采样为收费功能	Good	Good	Very Good
多租户与权限控制	Poor	Very Good	Good	Good	Very Good
接入数据源多样性	Good	Very Good	Poor	Poor	Very Good
文档	Poor	Very Good	Good	Good	Very Good

Prometheus Long-Term Storage 对比

2023
SUMMIT

APACHE • SkyWalking
CHINA · SHANGHAI

- ❑ 数据持久化到硬盘的方案里面 VictoriaMetrics 是更好的选择
- ❑ 数据持久化到对象存储的方案里 Thanos 更受欢迎, Grafana Mimir 更有潜力
- ❑ Thanos/Cortex/Mimir 也可以不使用对象存储, 用本地盘存数据
- ❑ Grafana fork 了 Cortex 创建了 Mimir 并修改 License 为 AGPL-3.0, Cortex 不建议继续采用, 后续 Grafana 及社区的投入程度成疑
- ❑ Thanos/Cortex/Mimir 互相借鉴, 架构类似, 三个项目里有非常多的主力 Prometheus Maintainer 参与:
 - ✓ Cortex/Mimir 借鉴了 Thanos 的对象存储访问及持久化
 - ✓ Thanos 借鉴了 Cortex 的 QueryFrontend
 - ✓ Mimir 作为 Grafana Cloud 的开源版本, 基于 Thanos 和 Cortex 的架构做了更多的优化
- ❑ 总体来说不介意许可证的话可以采用 Mimir, 喜欢更宽松许可证的话已经是 CNCF 孵化项目的 Thanos 是更好的选择
- ❑ 没有对象存储推荐使用 VictoriaMetrics (有些重要功能没开源), 有对象存储尽量用 Thanos 或 Mimir

Thanos 并不完美

- ✓ Thanos 社区官方维护的部署方案只适合单租户或 POC 的场景: [kube-thanos](#)
- ✓ 多租户支持待完善，也缺乏租户数据生命周期管理机制
- ✓ 众多组件没有水平扩展机制：Ingester, Compactor, Store, Ruler
- ✓ Ruler 对基于原始 metrics 数据集中计算 recording rule 及 alert rule 的扩展性较弱
- ✓ 社区缺少合适的 Operator 以 CRDs 的方式管理各组件

KSE Whizard 可观测中心



04

Future Plan

支持 Vector 作为可观测数据收集 Agent

2023
SUMMIT

APACHE • SkyWalking
CHINA · SHANGHAI

Performance [↗](#)

The following performance tests demonstrate baseline performance between common protocols with the exception of the Regex Parsing test.

Test	Vector	Filebeat	FluentBit	FluentD	Logstash	SplunkUF	SplunkHF
TCP to Blackhole	86mib/s	n/a	64.4mib/s	27.7mib/s	40.6mib/s	n/a	n/a
File to TCP	76.7mib/s	7.8mib/s	35mib/s	26.1mib/s	3.1mib/s	40.1mib/s	39mib/s
Regex Parsing	13.2mib/s	n/a	20.5mib/s	2.6mib/s	4.6mib/s	n/a	7.8mib/s
TCP to HTTP	26.7mib/s	n/a	19.6mib/s	<1mib/s	2.7mib/s	n/a	n/a
TCP to TCP	69.9mib/s	5mib/s	67.1mib/s	3.9mib/s	10mib/s	70.4mib/s	7.6mib/s

To learn more about our performance tests, please see the [Vector test harness](#).

Correctness [↗](#)

The following correctness tests are not exhaustive, but they demonstrate fundamental differences in quality and attention to detail:

Test	Vector	Filebeat	FluentBit	FluentD	Logstash	Splunk UF	Splunk HF
Disk Buffer Persistence	✓	✓			△	✓	✓
File Rotate (create)	✓	✓	✓	✓	✓	✓	✓
File Rotate (copytruncate)	✓					✓	✓
File Truncation	✓	✓	✓	✓	✓	✓	✓
Process (SIGHUP)	✓				△	✓	✓
JSON (wrapped)	✓	✓	✓	✓	✓	✓	✓

To learn more about our correctness tests, please see the [Vector test harness](#).

Features [↗](#)

Vector is an end-to-end, unified, open data platform.

	Vector	Beats	Fluentbit	Fluentd	Logstash	Splunk UF	Splunk HF	Telegraf
End-to-end	✓							✓
Agent	✓	✓	✓			✓		✓
Aggregator	✓			✓	✓		✓	✓
Unified	✓							✓
Logs	✓	✓	✓	✓	✓	✓	✓	✓
Metrics	✓	△	△	△	△	△	△	✓
Traces	🚧							
Open	✓		✓	✓				✓
Open-source	✓	✓	✓	✓	✓			✓
Vendor-neutral	✓		✓	✓				✓
Reliability	✓							
Memory-safe	✓							✓
Delivery guarantees	✓					✓	✓	
Multi-core	✓	✓	✓	✓	✓	✓	✓	✓

△ = Not interoperable, metrics are represented as structured logs

Q&A

欢迎提问交流
(仅限2位提问)

请加我的企业微信



小kk@kubesphere.io

青云

加入 KubeSphere 社区



2023 | APACHE • SkyWalking
SUMMIT | CHINA · SHANGHAI

2023 • SkyWalking Summit

感谢您的观看

