Elasticsearch 多版本多集群运维管理实践

魏彬

2019.12.07,普翔科技

公司简介

Intro

上海 I 北京 I 深圳 I 香港

Elastic 战略级合作伙伴

普翔科技

htttp://www.puxiangtech.com
http://elastictech.cn

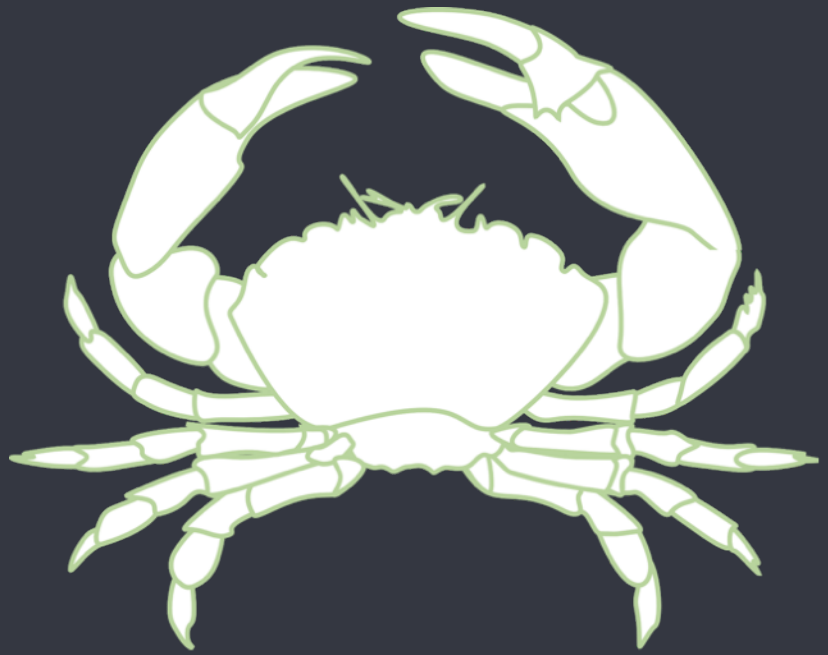
国内最早的一批 Elastic 认证工程师

elastic

议程

Agenda

1. Elasticsearch 在一家公司的发展路径
2. Elasticsearch 支持团队的痛点和解决方案
3. Elasticsearch 多集群监控和运维的方案探索
4. 未来规划

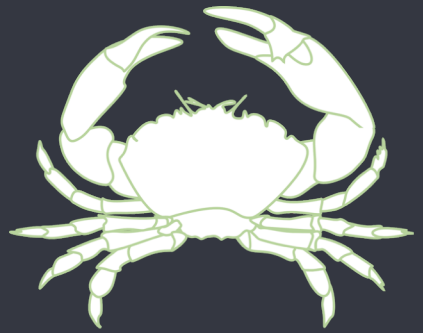


第一个吃螃蟹的人

Logging/运维部门

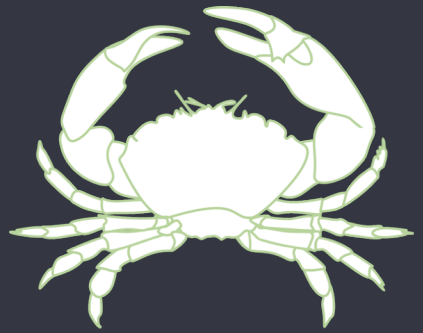
第一个吃螃蟹的人

Search/业务部门



第一个吃螃蟹的人

SIEM/安全部门



第一个吃螃蟹的人

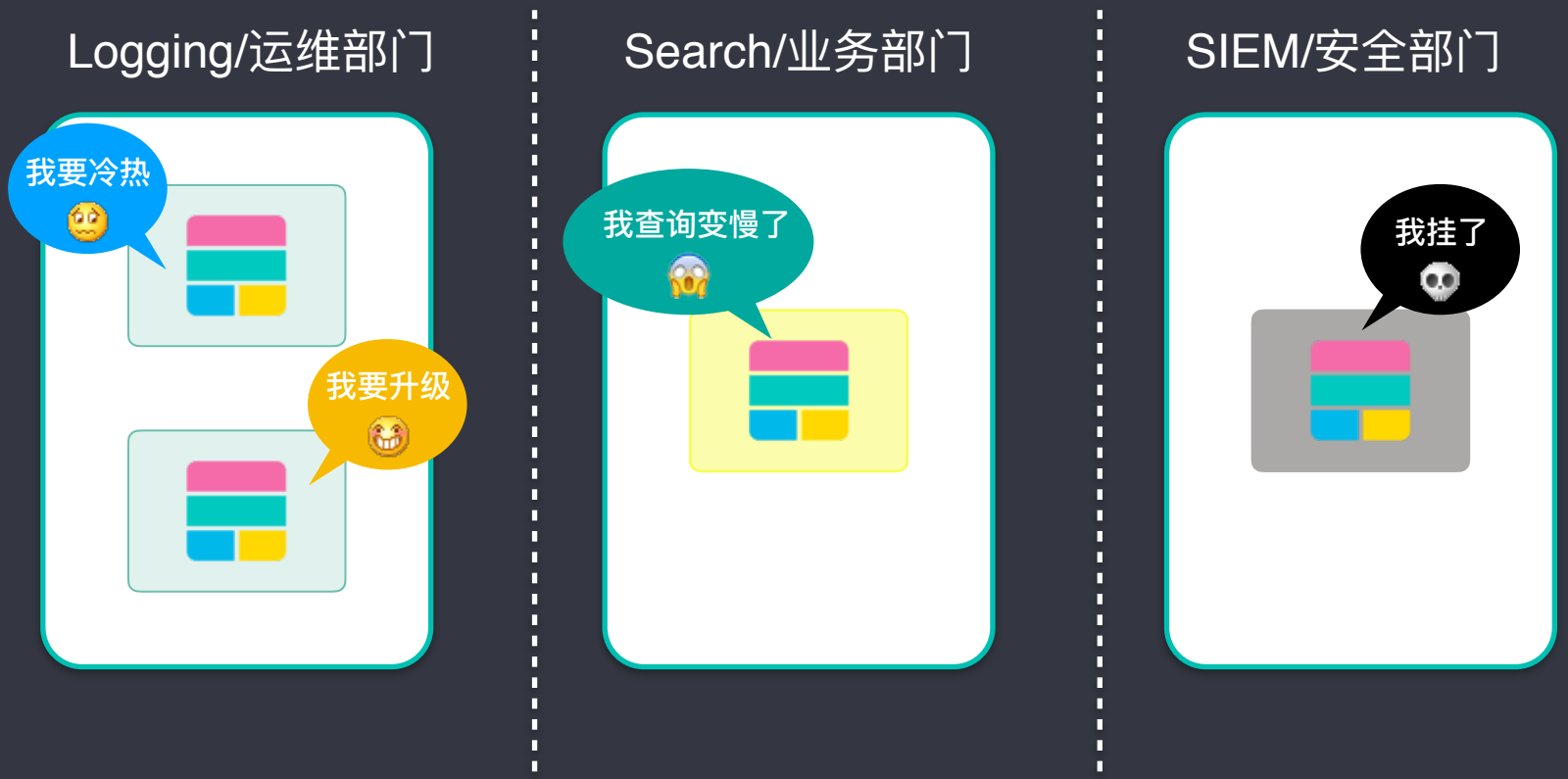
第一阶段

野蛮生长

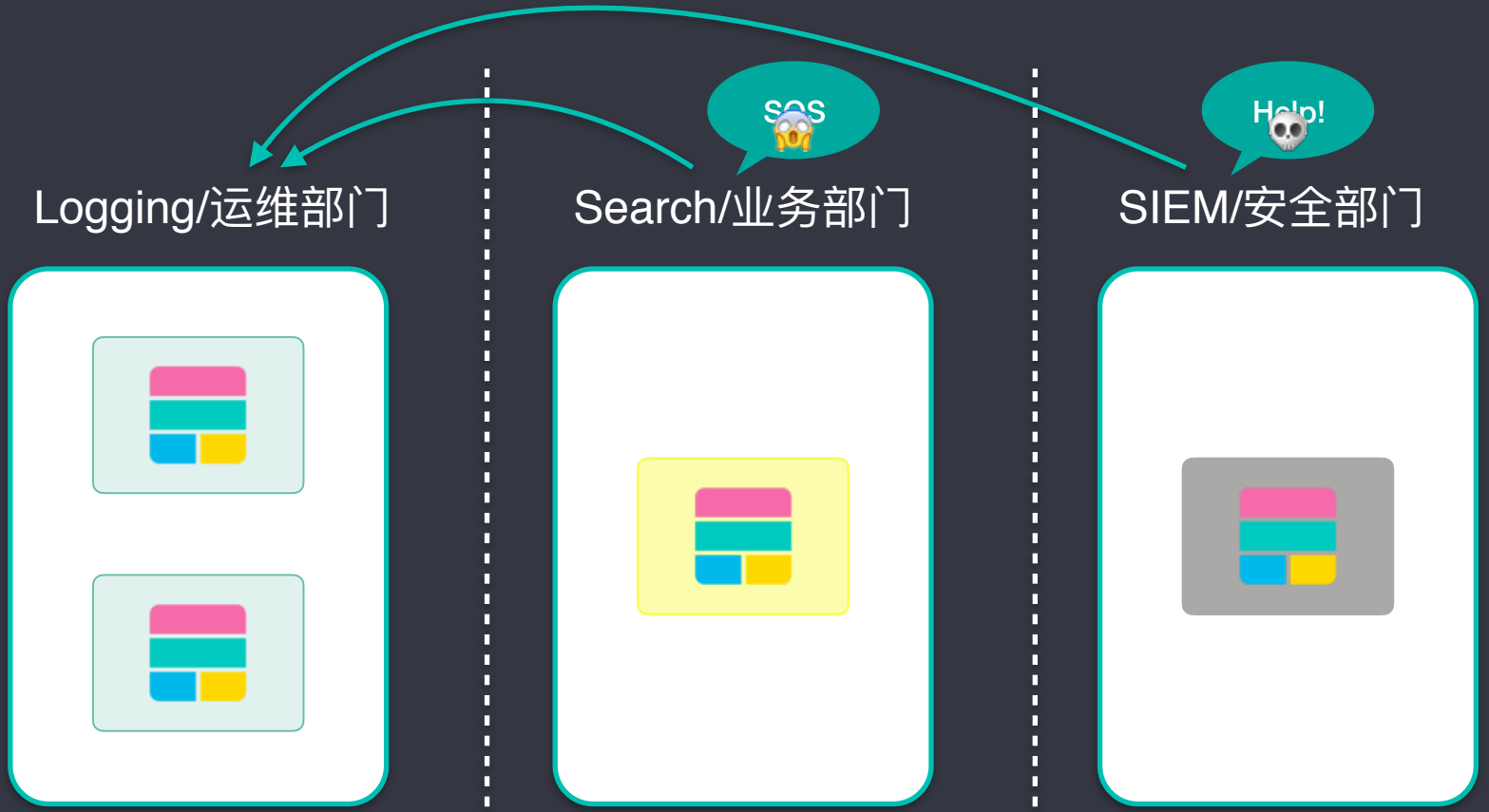
Logging/运维部门

Search/业务部门

SIEM/安全部门



各部门各项目组架构和运维能力不同



Elasticsearch 技术支持

破壁

第二阶段

专职支持

ES架构师

ES运维


MSA

业务使用方式多样

运维管理方式多样

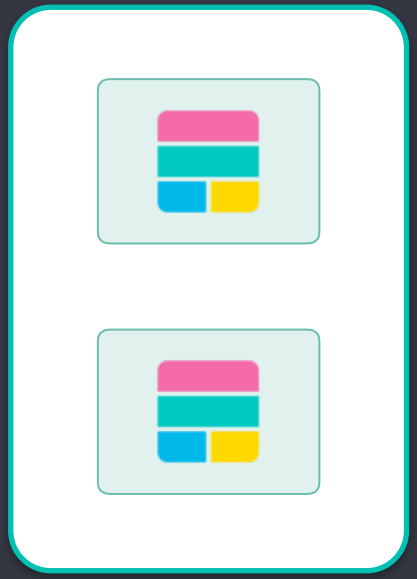
业务已上线，变更困难

集群过多，难以掌握全貌

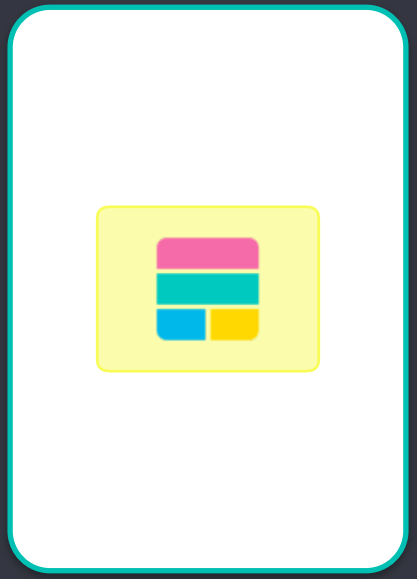
跨部门沟通困难

。 。 。 。 。 。

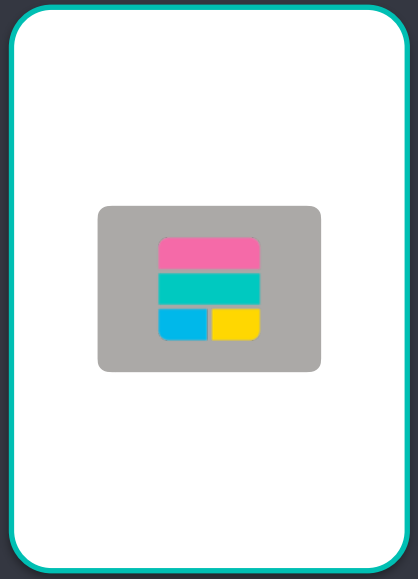
Logging/运维部门



Search/业务部门



SIEM/安全部门



是时候集中管控了!


Logging/运维部门
Search/业务部门
SIEM/安全部门
ES 服务平台

第三阶段

服务中心化

PASS/SASS

内部/私有云

¥¥¥¥¥¥
¥¥¥¥¥¥
第三阶段
服务中心化

¥¥¥
¥¥¥
第二阶段
专职支持

¥¥¥
第一阶段
野蛮生长

议程

Agenda

1 Elasticsearch 在一家公司的发展路径

2 Elasticsearch 支持团队的痛点和解决方案

3 Elasticsearch 多集群监控和运维的方案探索

4 未来规划

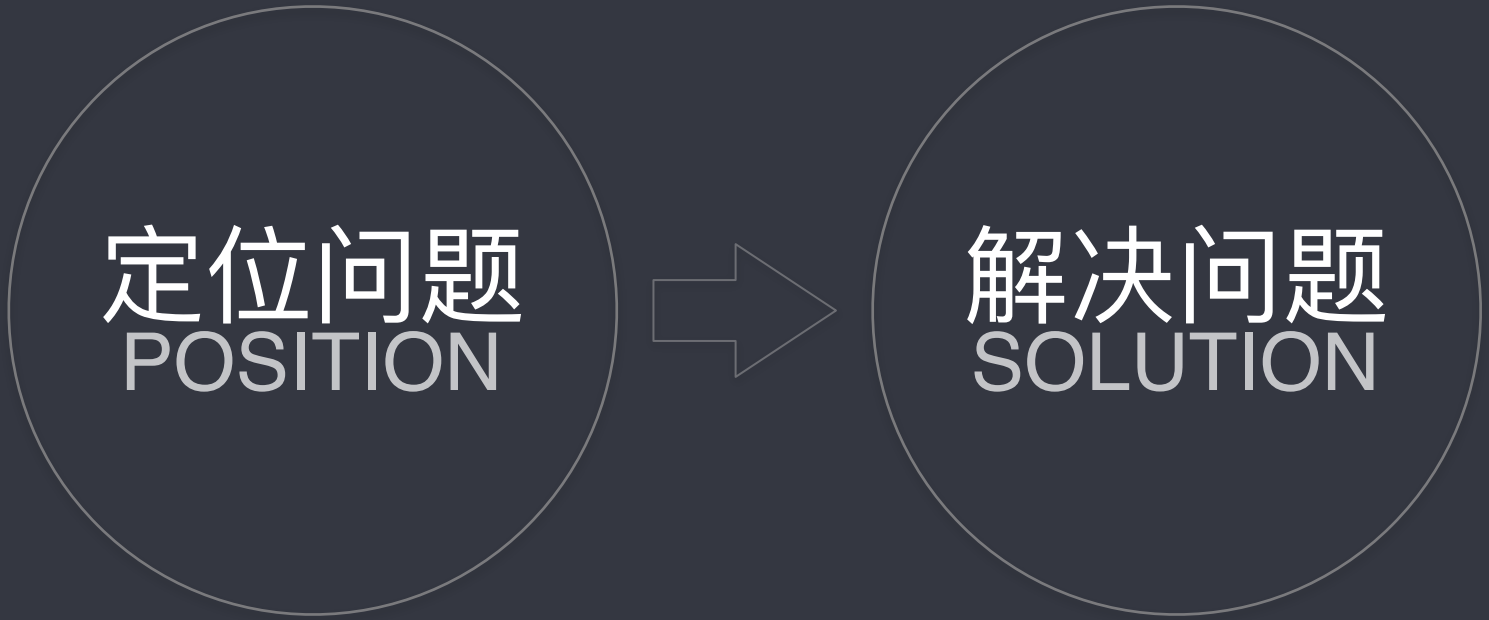
Elasticsearch 支持团队的痛点

Painpoint

- 各集群使用版本多样，从 2 到 7 ，频繁踩老坑，升级困难
- 各集群配置不统一，硬件、软件、数据目录、日志目录等，排查问题困难
- 问题表现都是读写问题，内在原因多种多样，检查项众多
- 基础监控未覆盖所有集群，排查问题缺乏关键时间的关键指标
- ……

Elasticsearch 支持团队的痛点

Painpoint



elastic

Elasticsearch 支持团队的痛点

Painpoint

- 获取详细的集群信息
  - Cluster
  - Node
  - Index
  - Shard
  - Log
- 获取方式
  - 连接集群 - API 请求

Elasticsearch 支持团队的痛点 - 获取详细的集群信息

Painpoint

https://github.com/elastic/support-diagnostics

```
sudo ./diagnostics.sh --host 10.0.0.20 --port 9201
```

diagnostics-20191014-160605.tar.gz

diagnostics-20191014-160605.tar.gz

alias.json	cat_thread_pool.txt	huge_pages.txt	netstat.txt	shard_stores.json
allocation.txt	cluster_health.json	indices_stats.json	network-cache-settings.properties	shards.json
allocation_explain.json	cluster_pending_tasks.json	iostat.txt	nodeattrs.txt	ss.txt
allocation_explain_disk.json	cluster_settings.json	jps.txt	nodes.json	sysctl.txt
cat_aliases.txt	cluster_settings_defaults.json	jstack.txt	nodes_hot_threads.txt	system-digest.json
cat_fielddata.txt	cluster_state.json	licenses.json	nodes_stats.json	system-digest.txt
cat_health.txt	cluster_stats.json	limits.txt	pipelines.json	tasks.json
cat_indices.txt	count.json	logs	plugins.json	templates.json
cat_master.txt	cpu.txt	manifest.json	proc-limit.txt	top.txt
cat_nodes.txt	cpu_governor.txt	mapping.json	process-list.txt	top_threads.txt
cat_pending_tasks.txt	diagnostics.log	master.json	recovery.json	uname.txt
cat_recovery.txt	dmesg.txt	ml_anomaly_detectors.json	sar.txt	version.json
cat_segments.txt	fielddata.json	ml_datafeeds.json	segments.json	watcher_stats.json
cat_shards.txt	fielddata_stats.json	ml_stats.json	settings.json	xpack.json

诊断：到底哪里出了问题?

数据建模不合理?

查询语句需要调优?

索引分片太少了还是太多了?

Shard 分布不平均?

主节点是否有任务堆积?

节点 Heap 设置不正确?

节点角色分配是否合理?

基础的写入优化是否做了?

是否有频繁地长时间 GC?

是否有大量慢查询?

Elasticsearch 支持团队的痛点

Painpoint

- 人工诊断
  - 依赖诊断人员的经验
  - 难以形成有效的知识传递
  - 效率低，且有遗漏的情况
- 机器诊断
  - 将经验沉淀为机器可识别的规则列表
  - 效率高，可复用，对使用人员要求低
  - 可自动化，定期诊断多集群

Elasticsearch 支持团队的诊断列表 Cluster

Painpoint

1. 集群健康状态检查
2. 集群节点角色分配的合理性检查
3. 集群最小可选主节点数配置的合理性检查
4. 集群索引删除保护配置检查
5. 集群写请求处理压力检查
6. 集群读请求处理压力检查
7. 集群分片迁移检查
8. 集群备份检查
9. 集群高可用架构设计检查
10. 集群主节点 Pending Task 检查

Elasticsearch 支持团队的诊断列表 Node

Painpoint

1. 文件句柄数调优检查
2. swap 内存关闭
3. 虚拟内存大小调优
4. 线程数调优
5. JVM Heap 大小设定
6. 各节点 JVM Heap 大小设定一致性
7. 各节点硬件配置检查
8. 各节点内存磁盘比检查
9. 各数据节点分片分布数检查
10. 单数据节点分片数检查
11. 节点读写压力检查
12. 节点内存压力检查

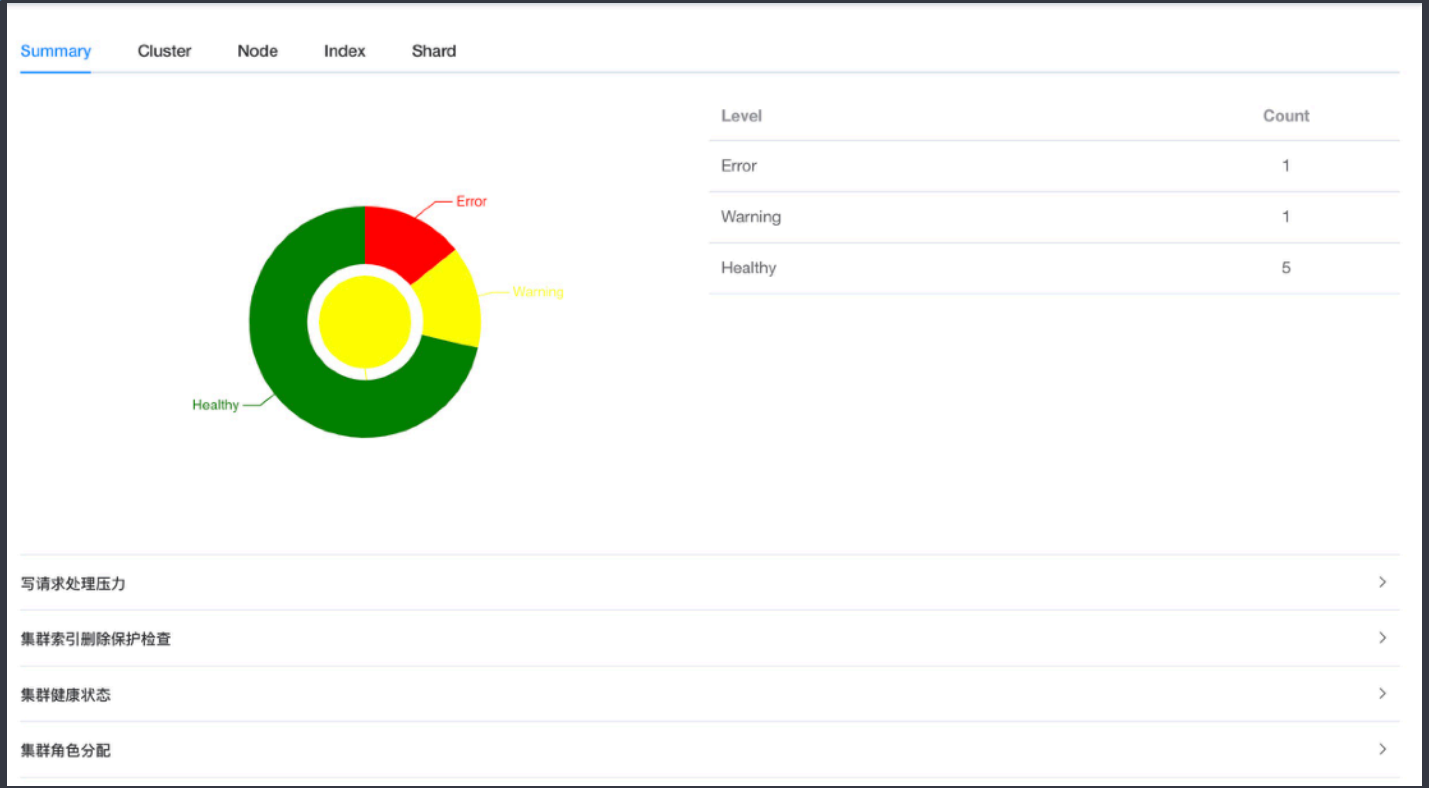
Elasticsearch 支持团队的诊断列表 Index

Painpoint

1. 索引分片数设定合理性检查
2. 字段类型和相关设置是否最优
3. 副本设置合理性检查
4. Shard 分配是否平均到所有节点
5. 别名使用检查
6. 字段数目检查
7. 模板使用检查
8. 字段自动添加功能是否关闭

Elasticsearch 支持团队的诊断结果页面

Painpoint



elastic

Elasticsearch 支持案例分享1

Painpoint

- 问题描述:
  - 集群在增加节点后，在业务高峰期集群大量响应延迟，读写流量没有显著增长
- 问题分析:
  - 2 版本的集群，还没有很好的指标监控，通过日志查看，部分节点有严重的 gc
- 根因:
  - 由于部署人员没有采用该部门先前的节点部署方式，导致 JVM HEAP 的环境变量没有生效，继而新增节点都是默认 Heap 大小，大数据量迁移到该节点后由于 Heap 过小，引发频繁 GC
- 解决方案:
  - 按照原有方式重启有问题节点

Elasticsearch 支持案例分享1

Painpoint

- 规则：
  - Node #6 各节点 JVM Heap 大小设定一致性

Elasticsearch 支持案例分享2

Painpoint

- 问题描述:
  - 集群写速率没有达到峰值，但有 reject 发生
- 问题分析:
  - 通过分析发现所有的 reject 集中在一个节点上
- 根因:
  - 某大索引分片分布不平均，导致某个节点分配了过多的分片，成为了写入热点，降低了整个集群的处理能力
- 解决方案:
  - 引入 total_shards_per_node 参数规避该问题

Elasticsearch 支持案例分享2

Painpoint

- 规则:
    - Index #4 Shard 分配是否平均到所有节点

Elasticsearch 支持案例分享3

Painpoint

- 问题描述:
  - 集群写速率骤降，新数据无法正常写入
- 问题分析:
  - 通过分析发现大量 pending task 在堆积，大部分为 update mapping 的 task
- 根因:
  - 业务直接将原始数据做 json decode 后存入 es，由于业务变更，某个 object 的 key 会以用户 id 作为 key name，导致生成大量的新字段，堵塞了 master 处理队列
- 解决方案
  - 将索引 dynamic 参数设置为 false
  - 与业务沟通修改该 key 的命名方式

Elasticsearch 支持案例分享3

Painpoint

- 规则:
    - Cluster #10 集群主节点 Pending Task 检查

议程

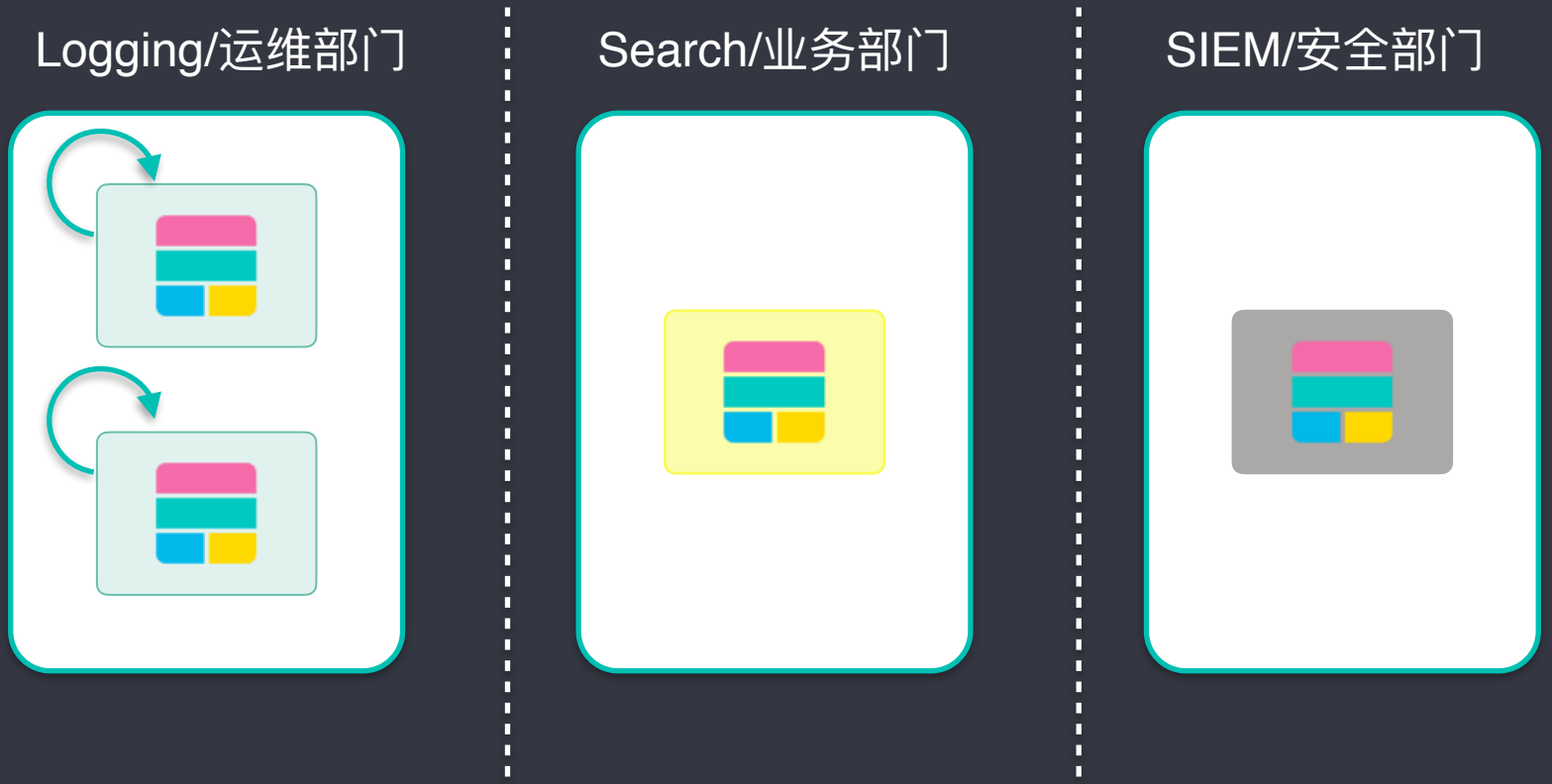
Agenda

1 Elasticsearch 在一家公司的发展路径

2 Elasticsearch 支持团队的痛点和解决方案

3 Elasticsearch 多集群监控和运维的方案探索

4 未来规划

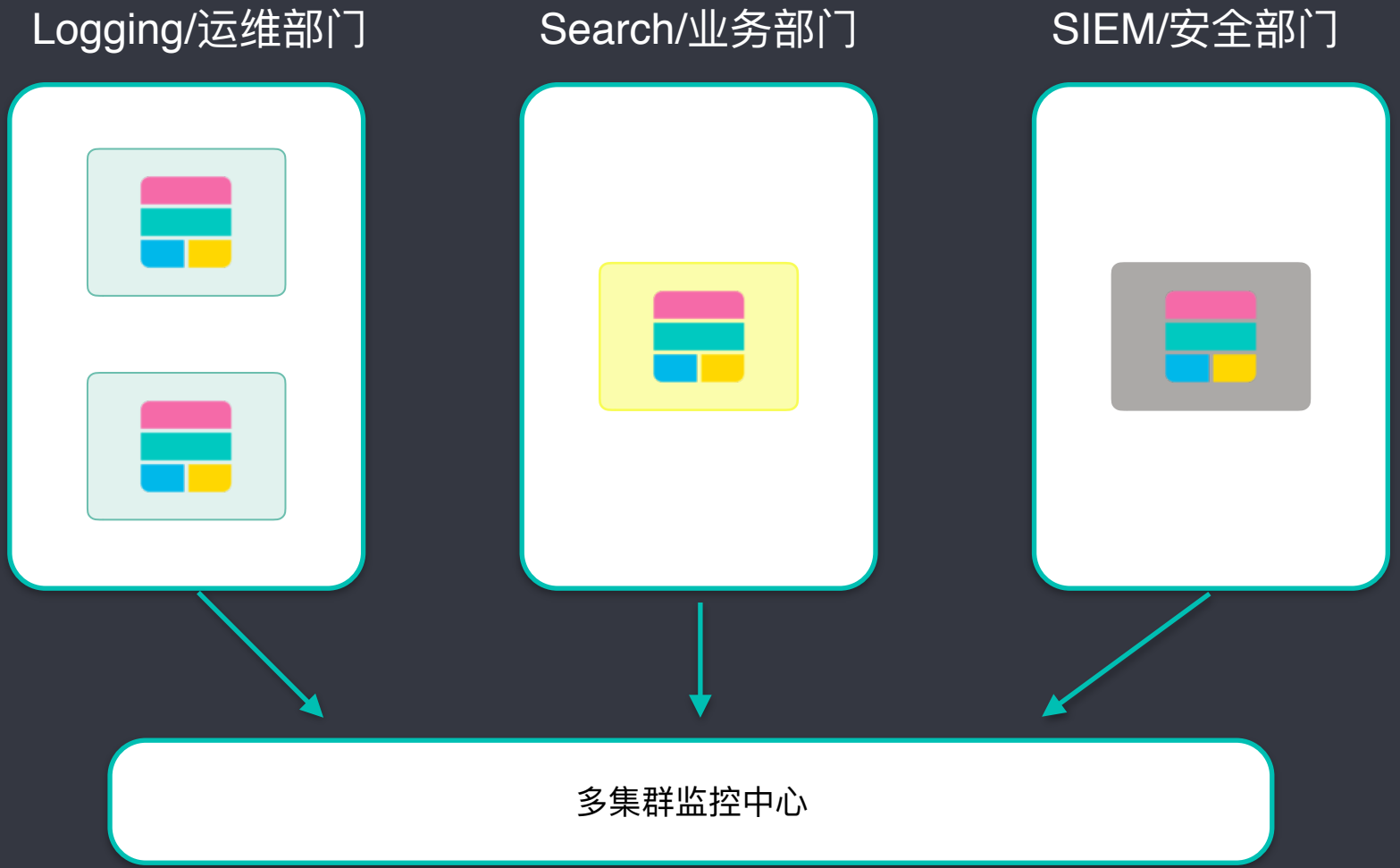


多数集群是自监控甚至未监控!

Elasticsearch 多集群统一监控的好处

Monitor A Lot of Clusters

- 减轻业务人员的运维成本和压力
- 实时掌控所有集群的运行状态，及时发现异常
- 逐步掌握各部门各项目的集群情况，为下一步集中管控做准备


Logging/运维部门
Search/业务部门
SIEM/安全部门
多集群监控中心

Elasticsearch 集群监控的维度

Monitor ES

- 指标 Metrics
- 日志 Logs

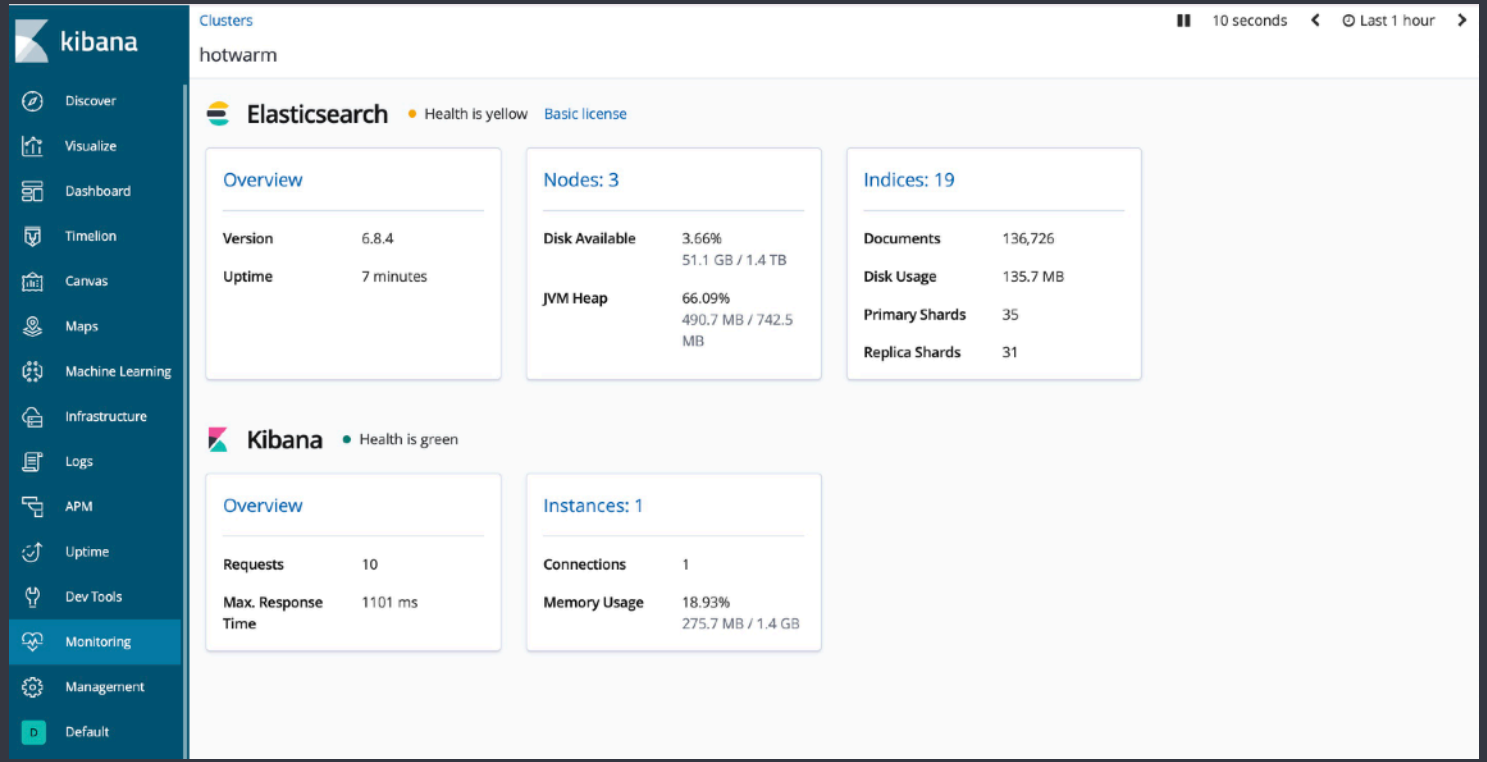
Elasticsearch 集群监控的常见方案

Monitor ES

- X-Pack Monitoring
- Prometheus + Grafana
- Metricbeat Elasticsearch Module
- Zabbix

Elasticsearch 集群监控的常见方案 X-Pack

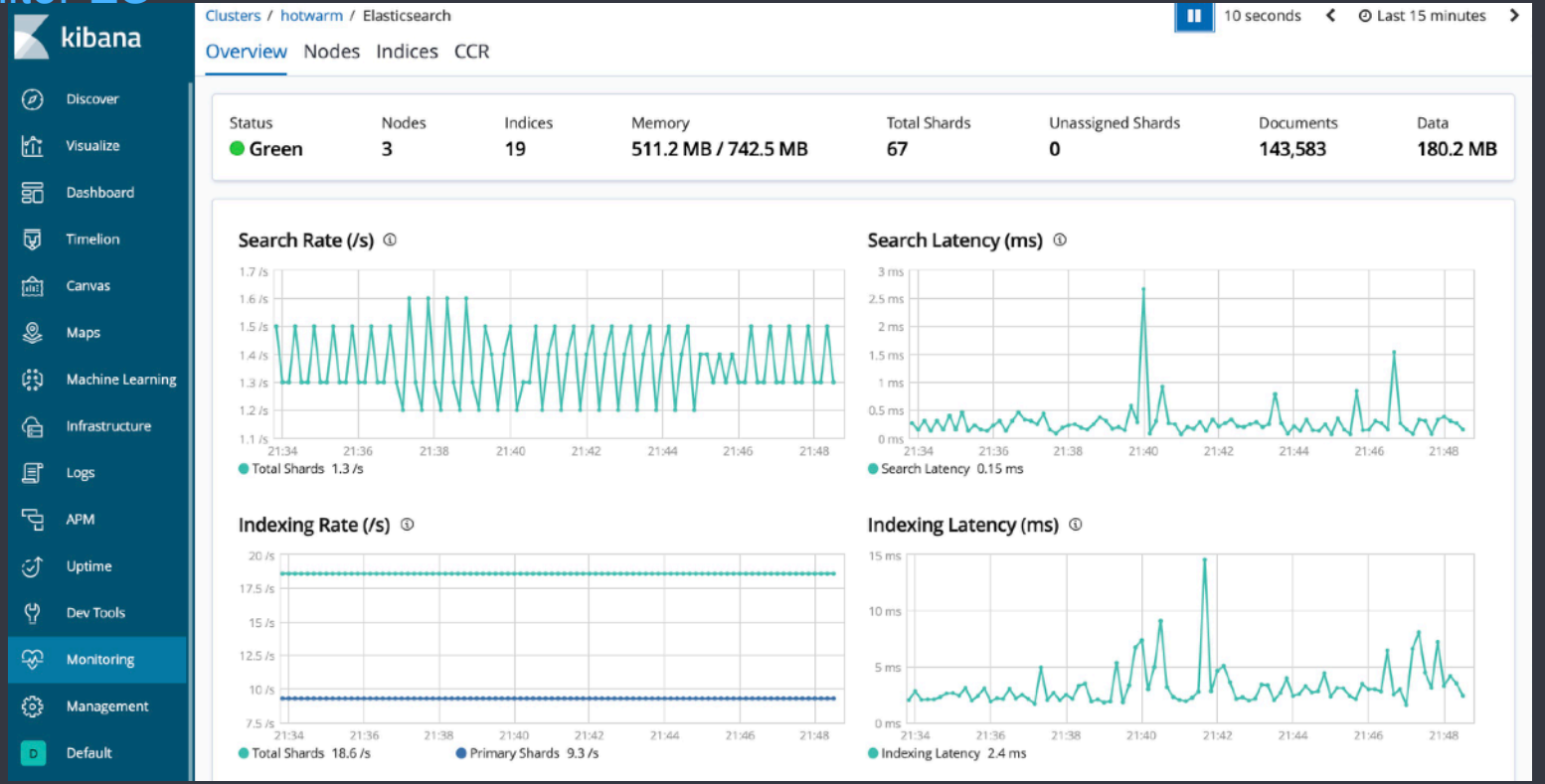
Monitor ES



elastic

Elasticsearch 集群监控的常见方案 X-Pack

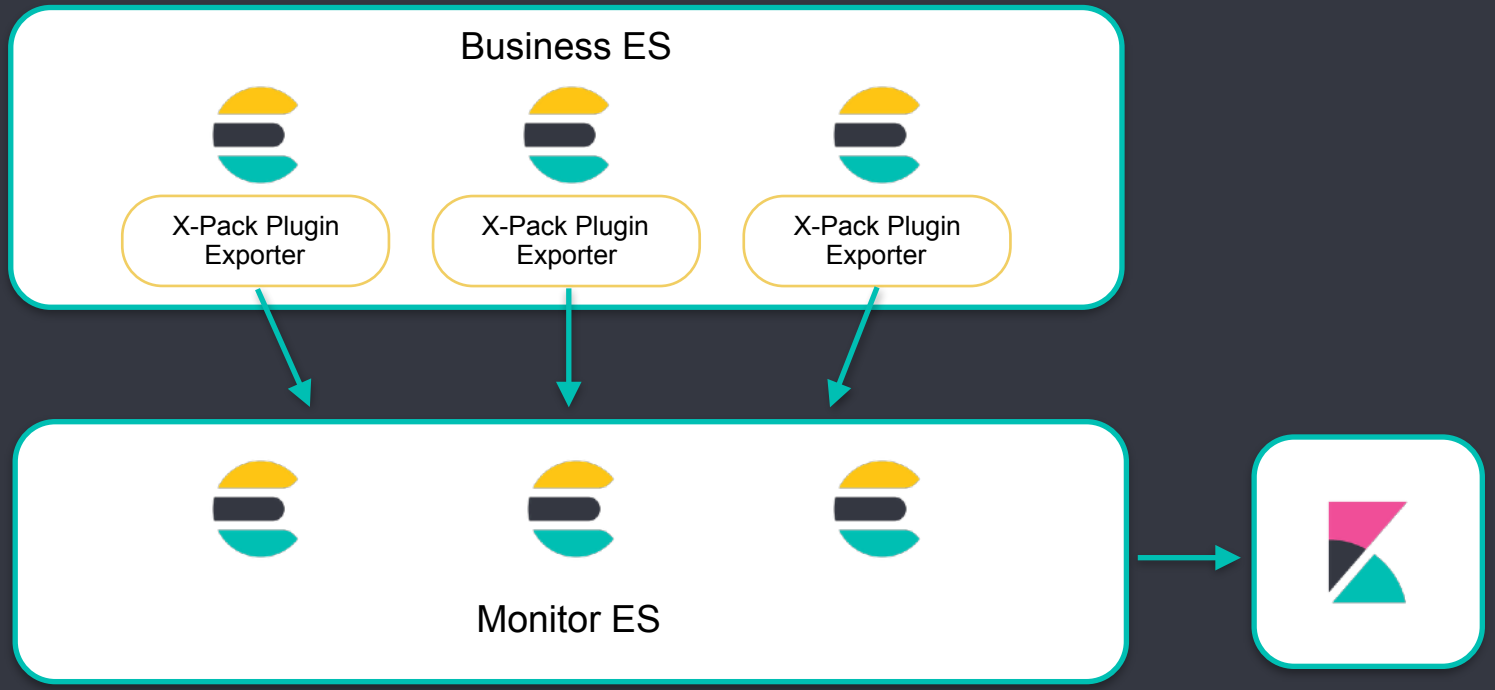
Monitor ES



elastic

Elasticsearch 集群监控的常见方案 X-Pack

Monitor ES



elastic

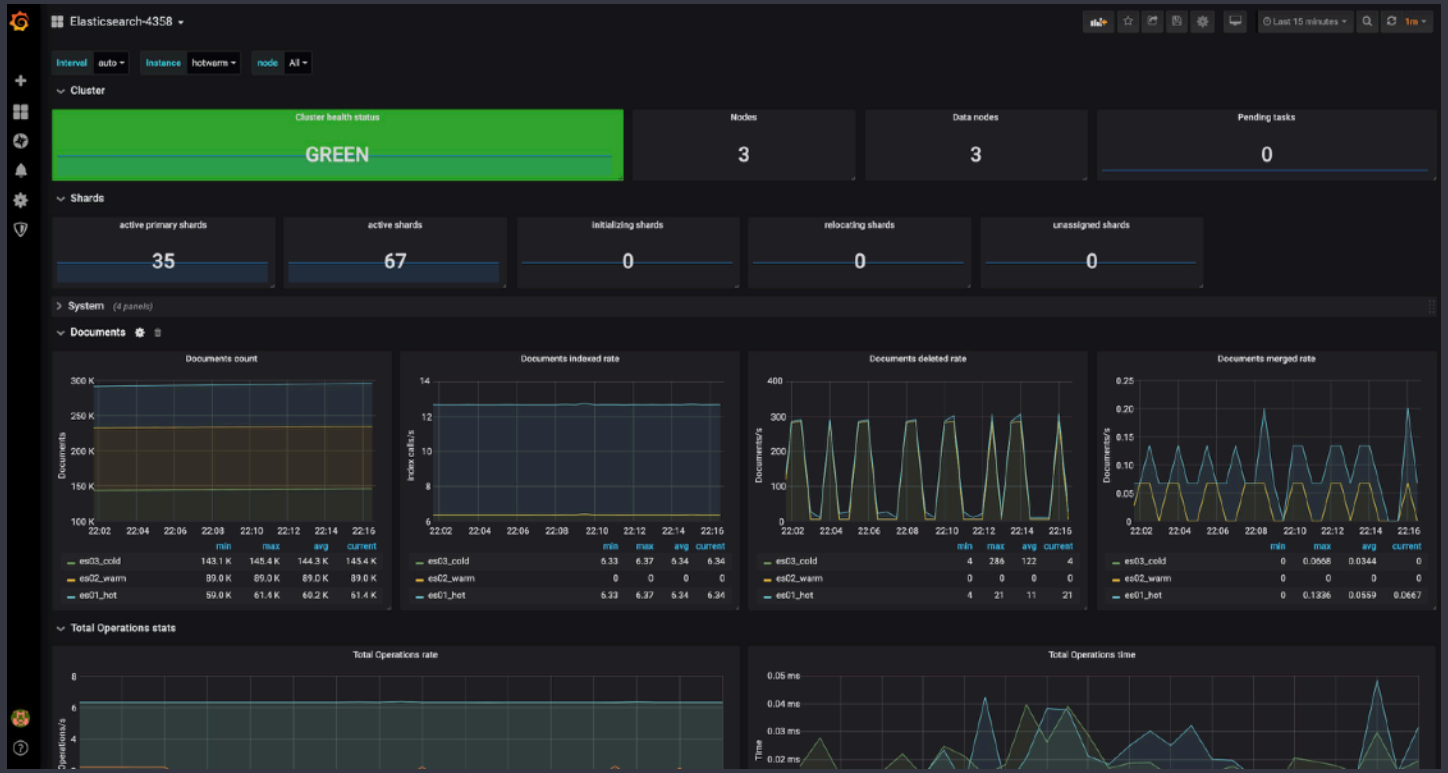
Elasticsearch 集群监控的常见方案 X-Pack

Monitor ES

- 简单易用，指标丰富
- 运行机制
    - 通过插件形式定时执行 exporter 采集指标
    - PUSH 到当前集群或专用监控集群
- 缺点：
    - 单集群免费，多集群收费
    - 6.3 后内置，启用即可，低版本需要额外安装插件
    - 可视化图表固定，可通过 Kibana 自行拓展

Elasticsearch 集群监控的常见方案 - Prometheus + Grafana

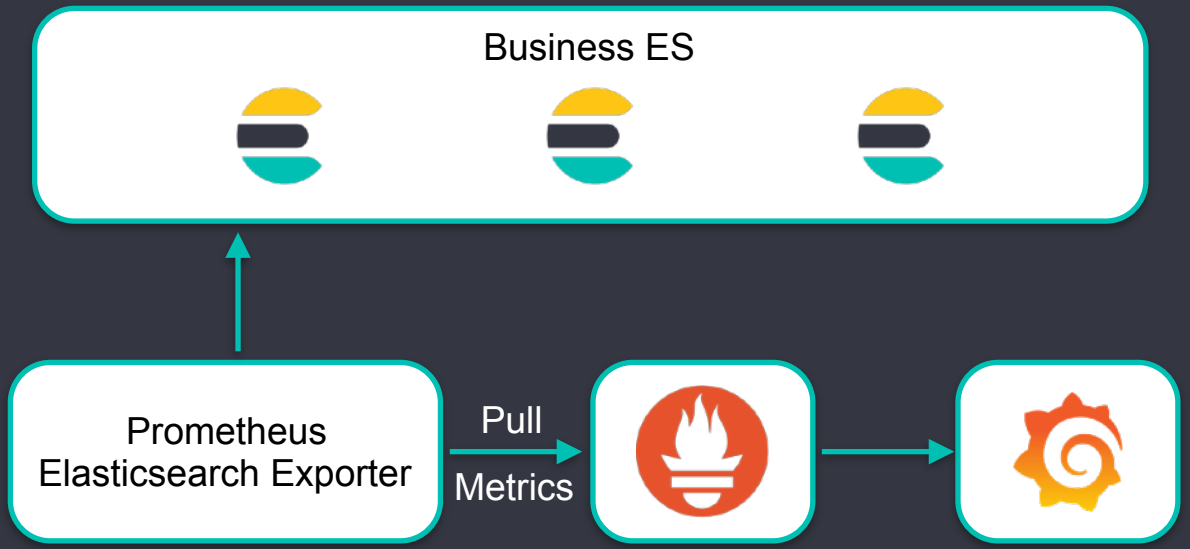
Monitor ES



elastic

Elasticsearch 集群监控的常见方案 - Prometheus + Grafana

Monitor ES



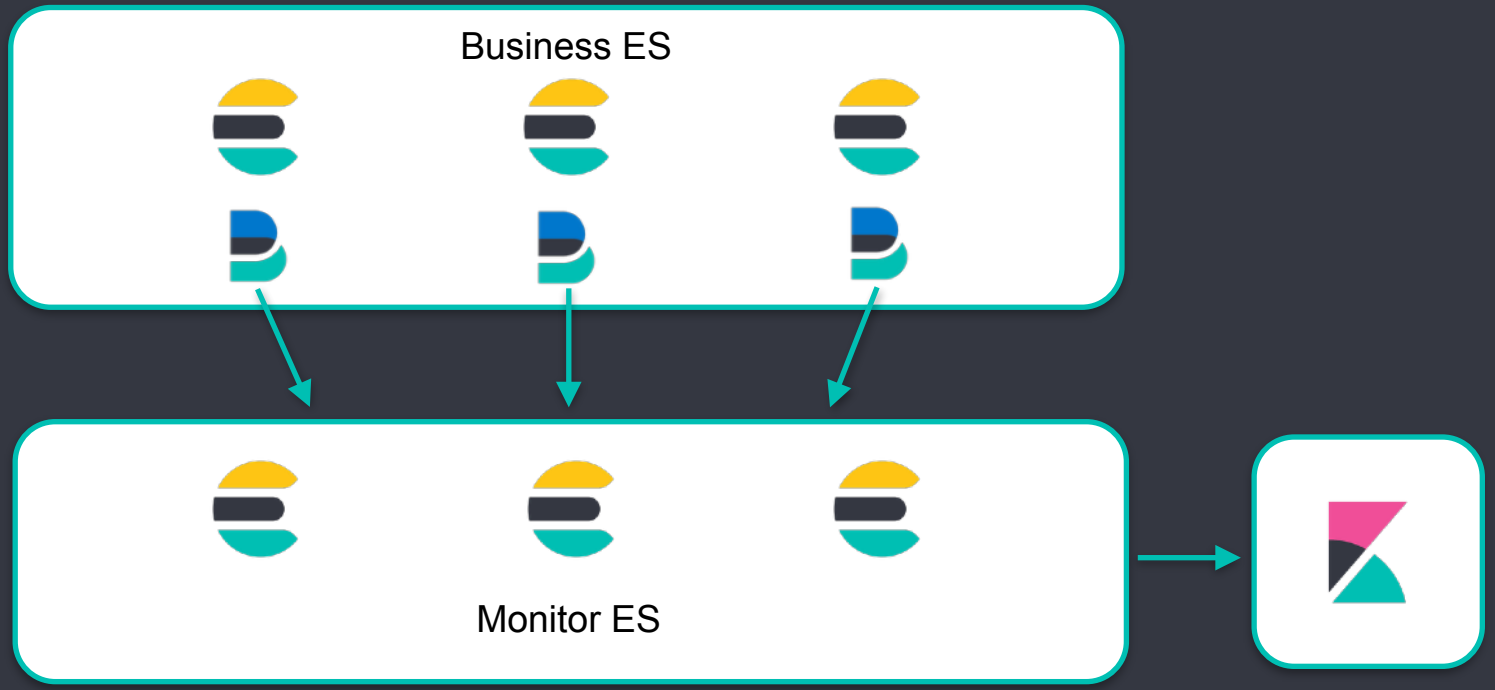
Elasticsearch 集群监控的常见方案 - Prometheus + Grafana

Monitor ES

- 无侵入性，部署灵活，支持多版本集群，指标丰富
- 可灵活由单集群扩展到多集群监控
- 运行机制
    - 基于 elasticsearch exporter 抓取指标
        - https://github.com/justwatchcom/elasticsearch_exporter
    - 基于 Prometheus 存储和查询数据
    - 基于 Grafana 实现报表展示
        - https://grafana.com/dashboards/2322
        - https://grafana.com/dashboards/4358
- 缺点：
    - 引入新的技术栈，需要额外维护，有一定学习成本
    - 只针对指标，无法存储日志

Elasticsearch 集群监控的常见方案 - Metricbeat

Monitor ES



Elasticsearch 集群监控的常见方案 - Metricbeat

Monitor ES

- 无侵入性，部署灵活，指标丰富
- 运行机制
    - 基于 metricbeat elasticsearch module 抓取指标
    - 基于 Elasticsearch 存储和查询数据
    - 可以兼容 Monitoring App
- 缺点：
    - 只支持 6 以上的版本
    - Monitoring App 多集群属于收费功能

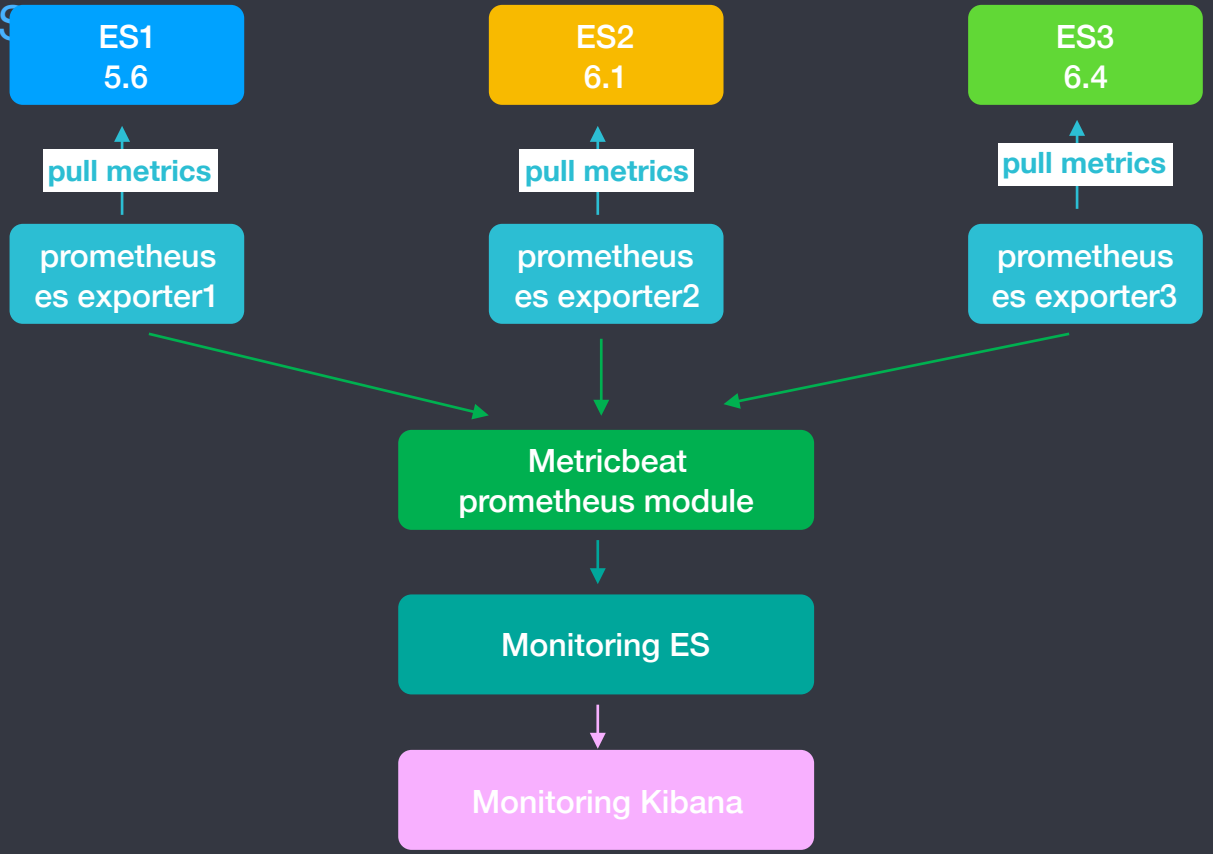
Elasticsearch 集群监控的理想方案

Monitor ES

- 无侵入性，部署灵活
- 支持多版本，指标丰富
- 支持与日志进行关联分析
- 尽量不引入新的技术栈
- 报表可自主定制化制作

Elasticsearch 集群监控的理想方案

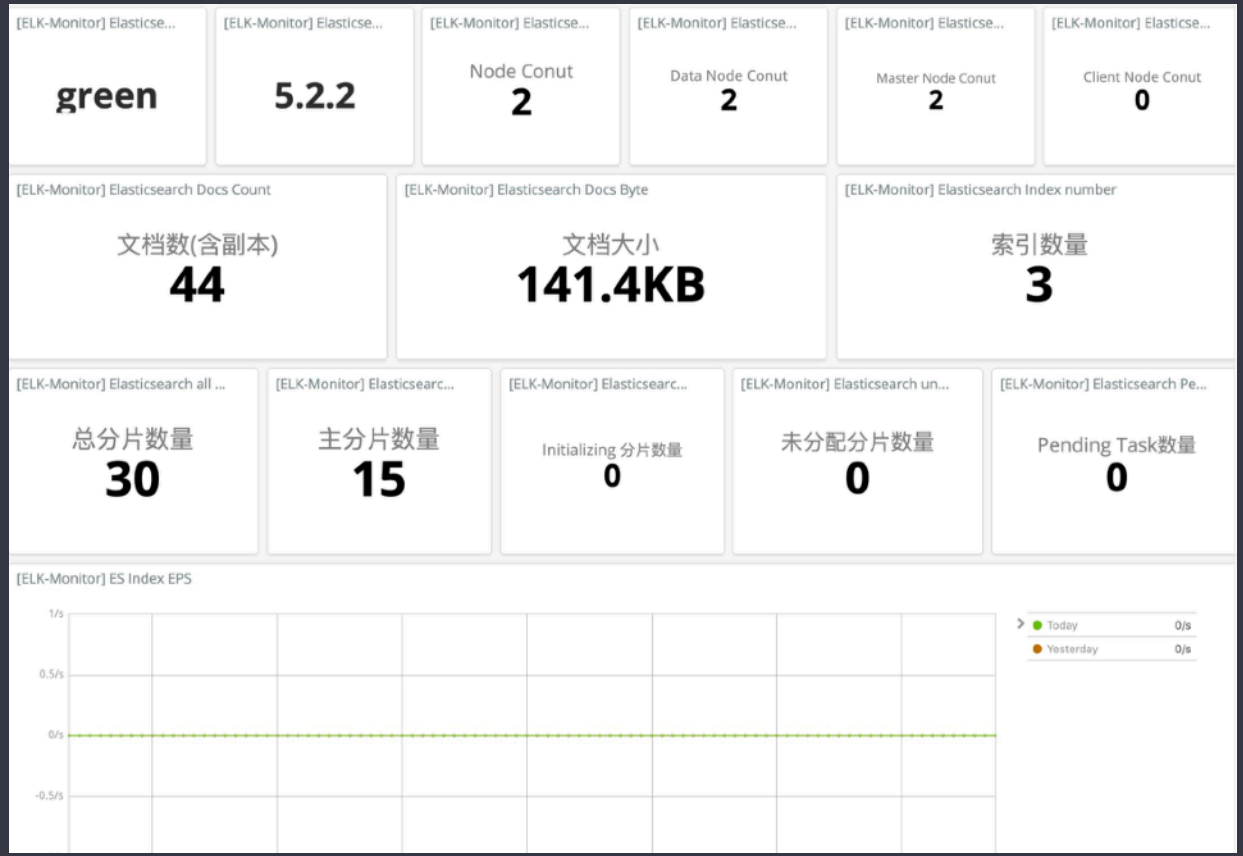
Monitor ES



elastic

Elasticsearch 集群监控的理想方案

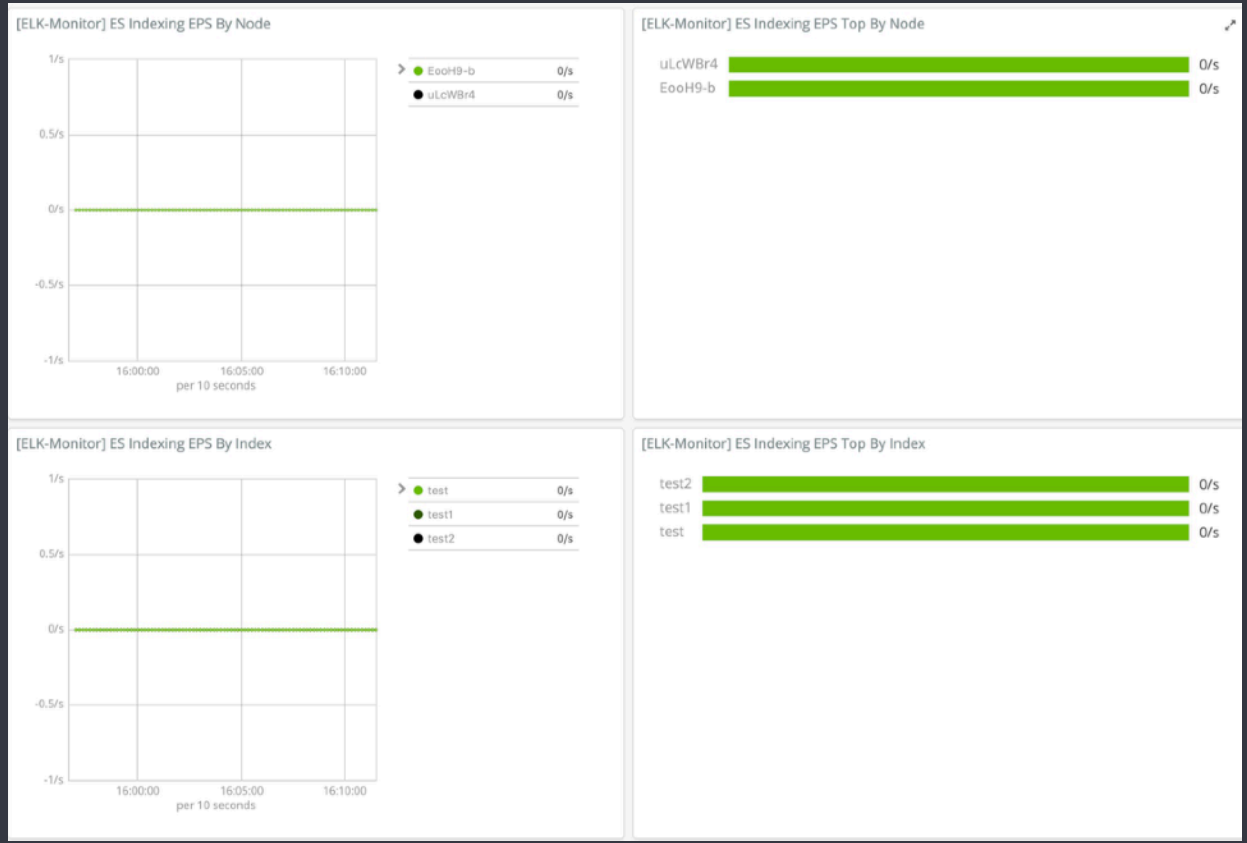
Monitor ES



elastic

Elasticsearch 集群监控的理想方案

Monitor ES



elastic

议程

Agenda

1 Elasticsearch 在一家公司的发展路径

2 Elasticsearch 多集群监控和运维的方案探索

3 Elasticsearch 多集群监控和运维的方案探索

4 未来规划

elastic

未来规划

Plan

- 多集群数据生命周期管理
- 多集群代理网关
- 集群容器化管理