



Elastic 中文社区

深圳 Meetup

2019/11/16

Saturday 13:00

合作伙伴



合作社区



Tencent 腾讯

Elasticsearch 在腾讯的优化实践



morningchen 陈曦

高级开发工程师

TEG . 云架构平台部

腾讯云Elasticsearch Service、CTSDB后台开发者。在日志分析、大数据搜索、时序数据分析等领域有丰富的实践经验。

Tencent 腾讯

1

ES 的应用场景

2

遇到的挑战

3

开源贡献 & 优化实践

1 ES 的应用场景

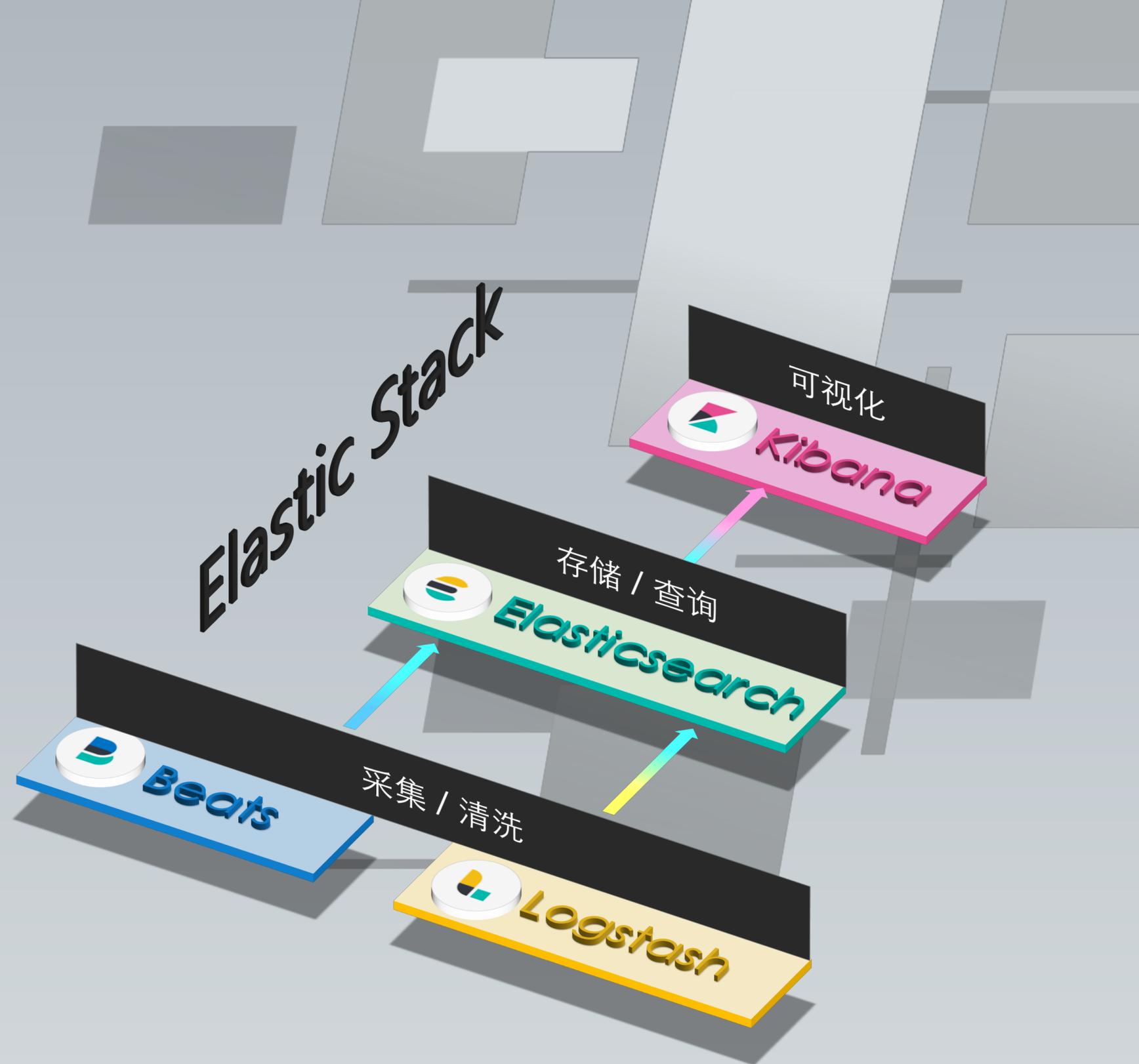
日志实时分析 – 首选开源方案

典型场景

- 业务日志：用户行为
- 状态日志：慢查询、异常探测
- 系统日志
-

主要特性

- Elastic Stack: 完整解决方案 (纯文本 -> 结构化数据)
- 实时性：从日志产生到可访问，秒级
- 全文搜索：基于倒排索引，支持灵活的搜索分析
- 交互式分析：万亿级日志，搜索秒级响应



1 ES 的应用场景

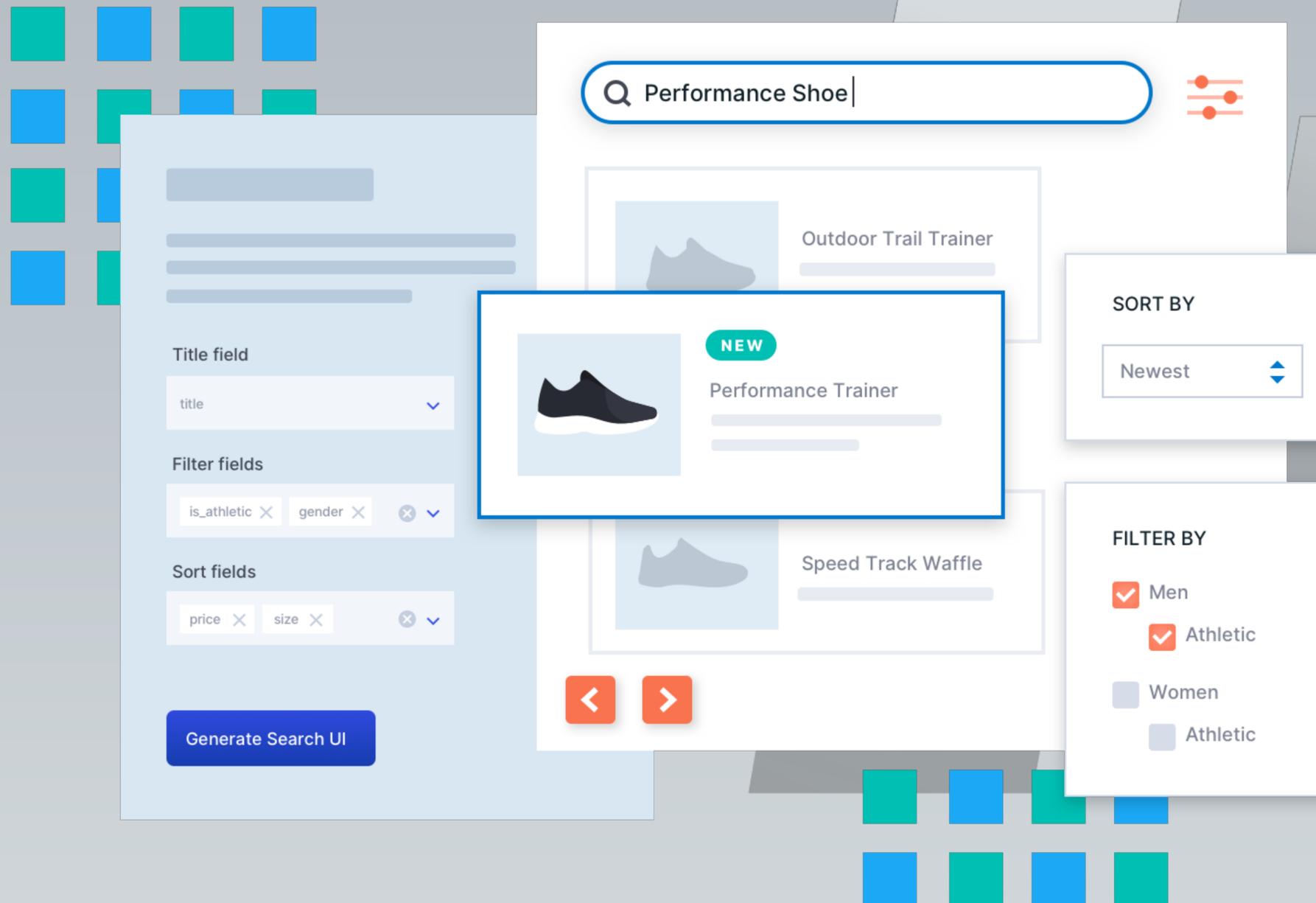
搜索服务

典型场景

 仪表板搜索	 电子商务
 手机应用搜索	 地理位置搜索

主要特性

- 高性能：高并发、低延迟的搜索体验
- 强相关：自定义打分、排序机制
- 高可用：机房、机架感知，异地容灾



时序数据分析

典型场景

- Metrics
- APM
- 物联网数据
-

主要特性

- 高并发写入：支持每秒千万级的时序写入
- 多维分析：多维度、灵活可扩展的统计分析
- 高性能：毫秒级查询
- 可视化：Iframe 可嵌入页面



Tencent 腾讯

1

ES 的应用场景

2

遇到的挑战

3

开源贡献 & 优化实践

丰富的业务需求

公有云

场景丰富

大量中小客户
不同的应用场景
不同的熟悉程度

内部云

超大规模

500+节点
千万级写入
百万级查询

私有云

标准化、自动化

不同行业/领域
完全隔离的环境
标准化交付
自动化运维

2 遇到的挑战

搜索场景下的挑战

代表：电商、站内搜索等在线服务



2 遇到的挑战

日志、时序场景下的挑战

数据量与时间成正比



1

ES 的应用场景

2

遇到的挑战

3

开源贡献 & 优化实践

开源贡献

	8 PR
	5 Merged
	2 Contributors

Contribution



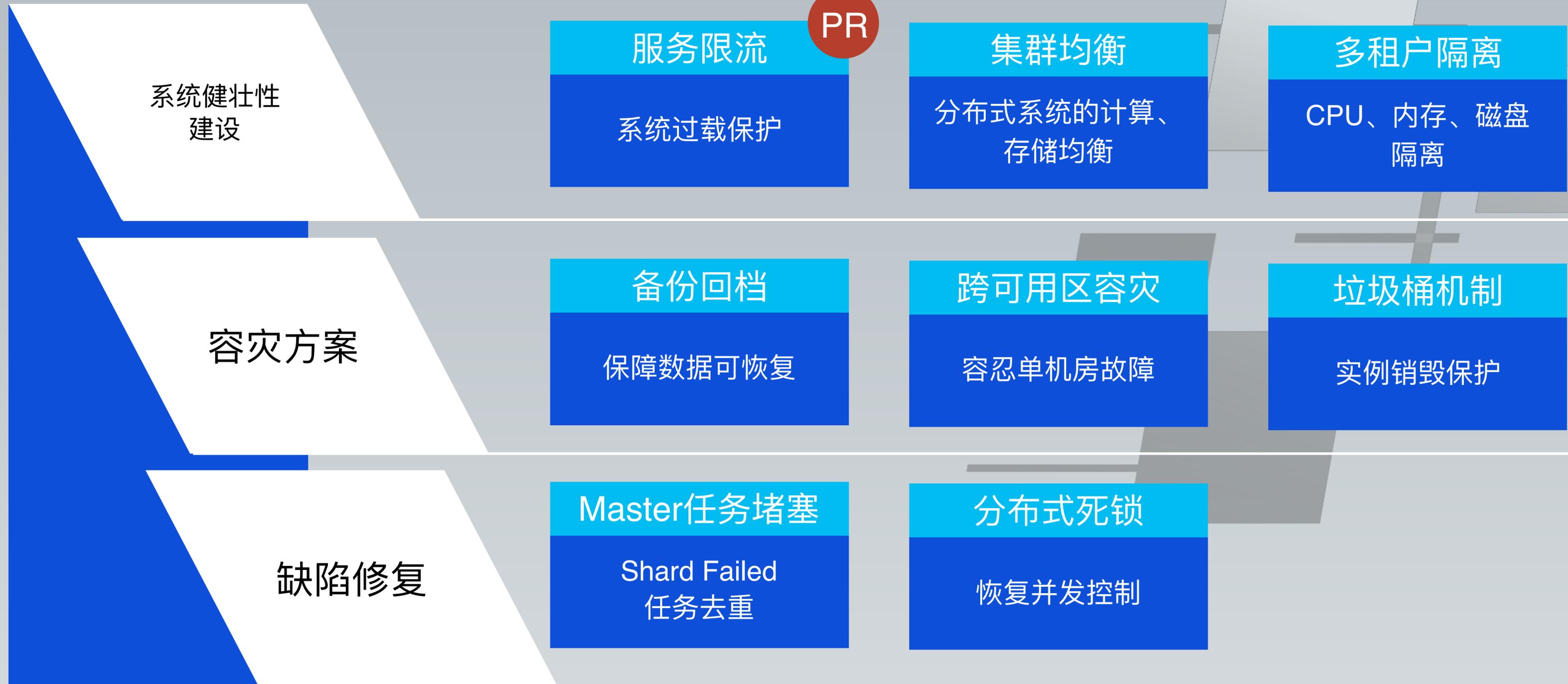
- 写入
- 查询
- 集群管理



可用性优化 – 面临挑战

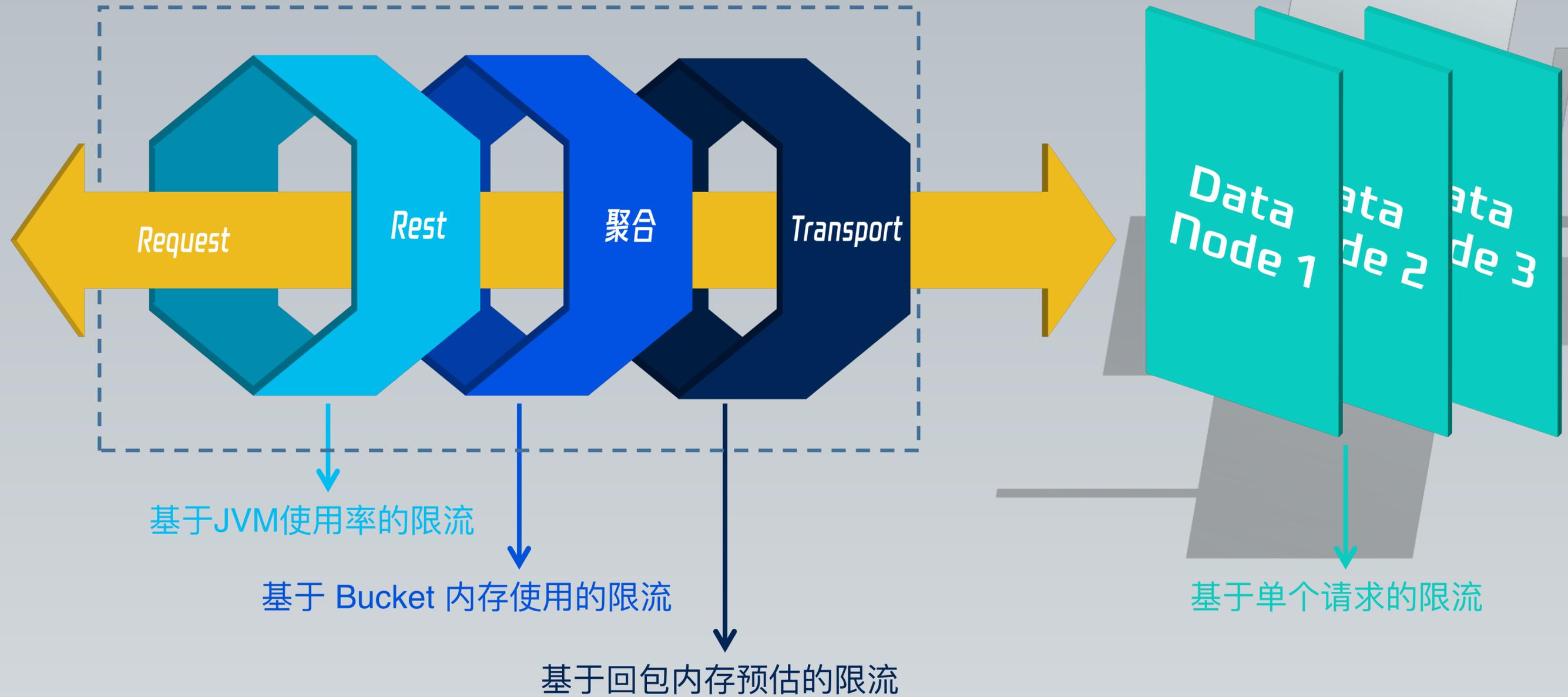


可用性优化 – 解决方案



可用性优化 – 服务限流

Coordinate Node



可用性优化 – Shard分配算法优化

避免新增节点压力过大

• 原生逻辑

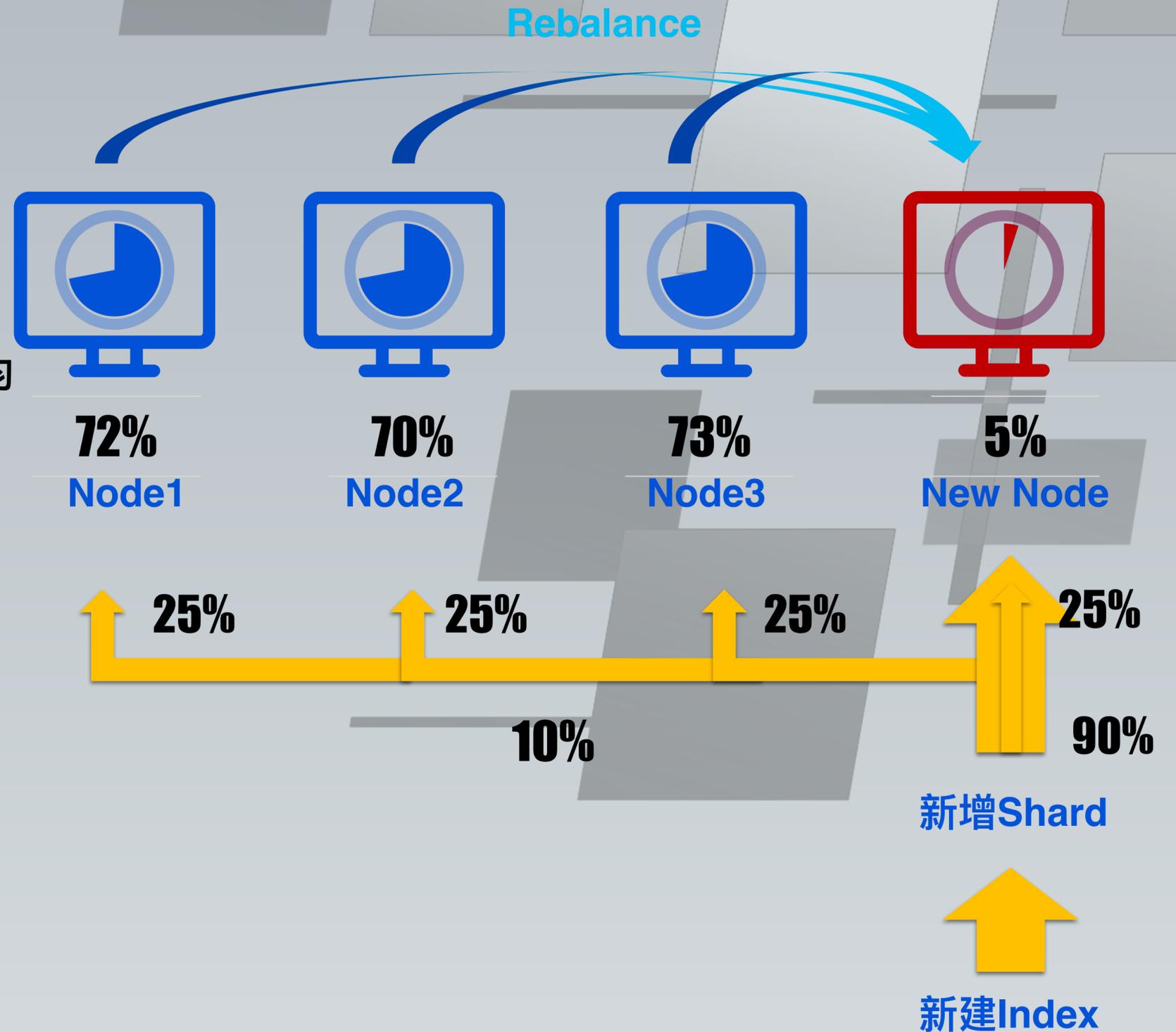
新增shard大概率被分配到新节点

调整参数，不考虑shard分布，导致shard分配不均

```
100 node-debug  
200 node-new-1  
200 node-new-3
```

• 我们的逻辑

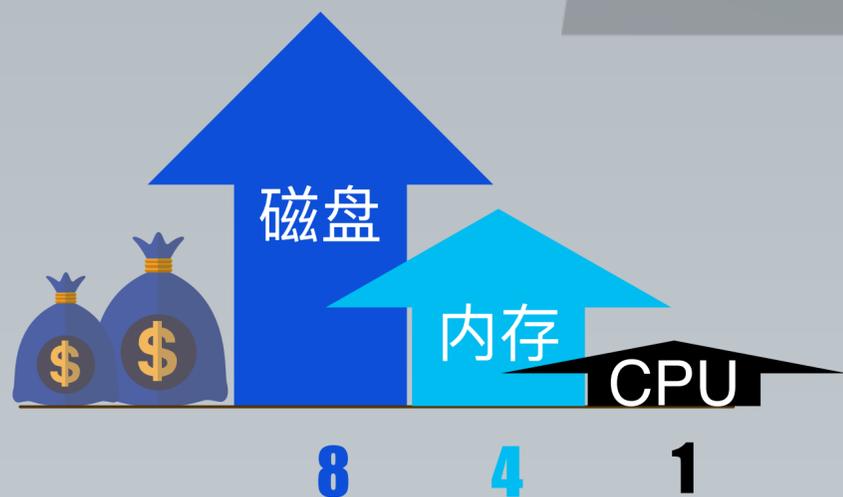
在保持原有逻辑的基础上，尽量随机分配shard



成本优化 – 面临挑战

成本分析

- 代表场景：日志、时序数据
- 成本比例关系

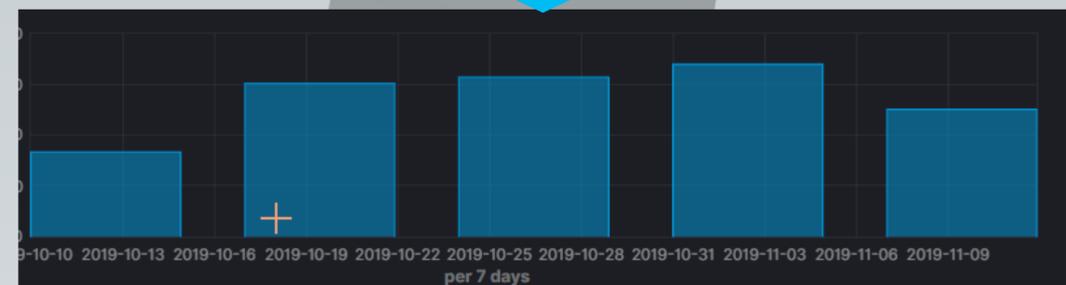
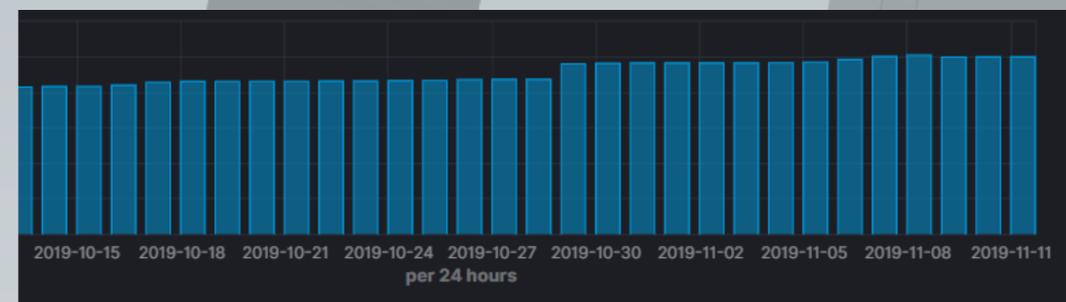
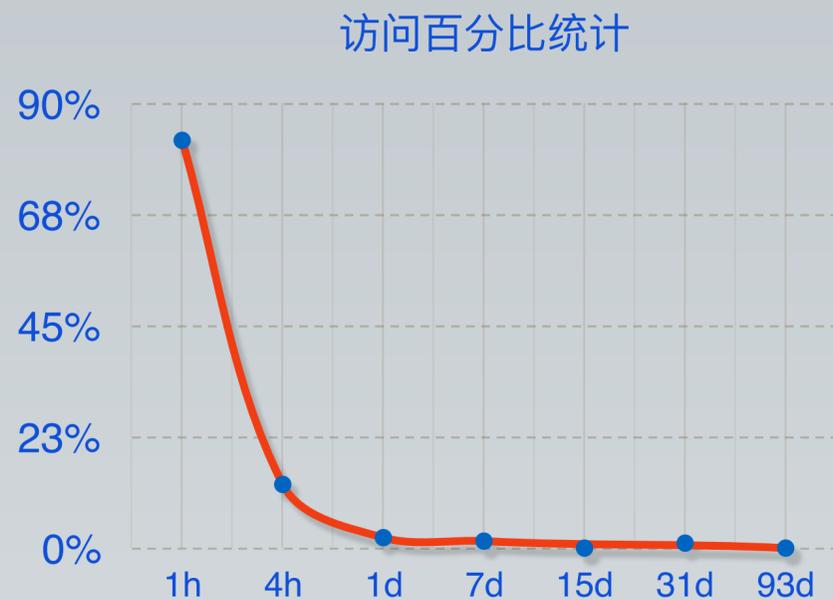


主要瓶颈

- 磁盘
- 内存

数据分析

- 冷热特性明显：访问近多远少
- 历史数据查询聚合结果
- 长时间范围查询性能差



成本优化 – 解决方案

硬盘成本

冷热分离

混合存储平衡
成本、性能

自研 Rollup

预计算
降低成本、提升性能

冷数据处理

备份归档、
生命周期管理

内存成本

Off Heap

减少堆内存，降低GC、
提高节点规格

LRU Cache

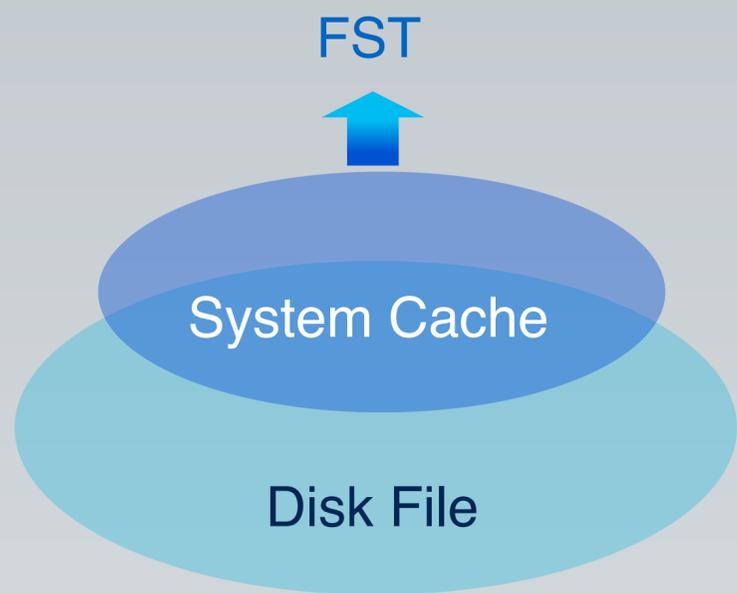
提升内存利用效率

成本优化 – 内存优化方案：

FST Off Heap 释放堆内空间，提升内存使用效率

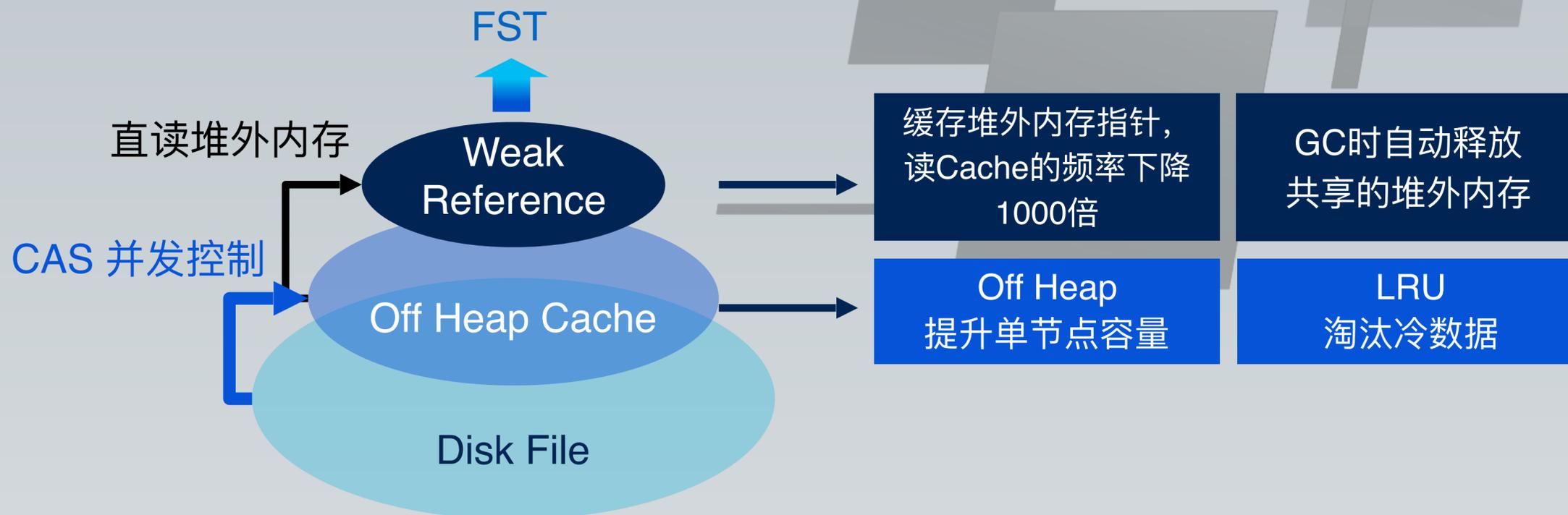
官方实现：

- 依赖系统缓存，按需加载
- 查询性能N倍损耗



我们的优势：

- Off Heap Cache 保证热数据 FST 的缓存空间
- 直读堆外内存，性能损耗控制在2%以内
- Cache 容量可动态调整



成本优化 – Rollup

方案：预计算：降低历史数据量，换取成本和性能（数量级）

我们的优势：

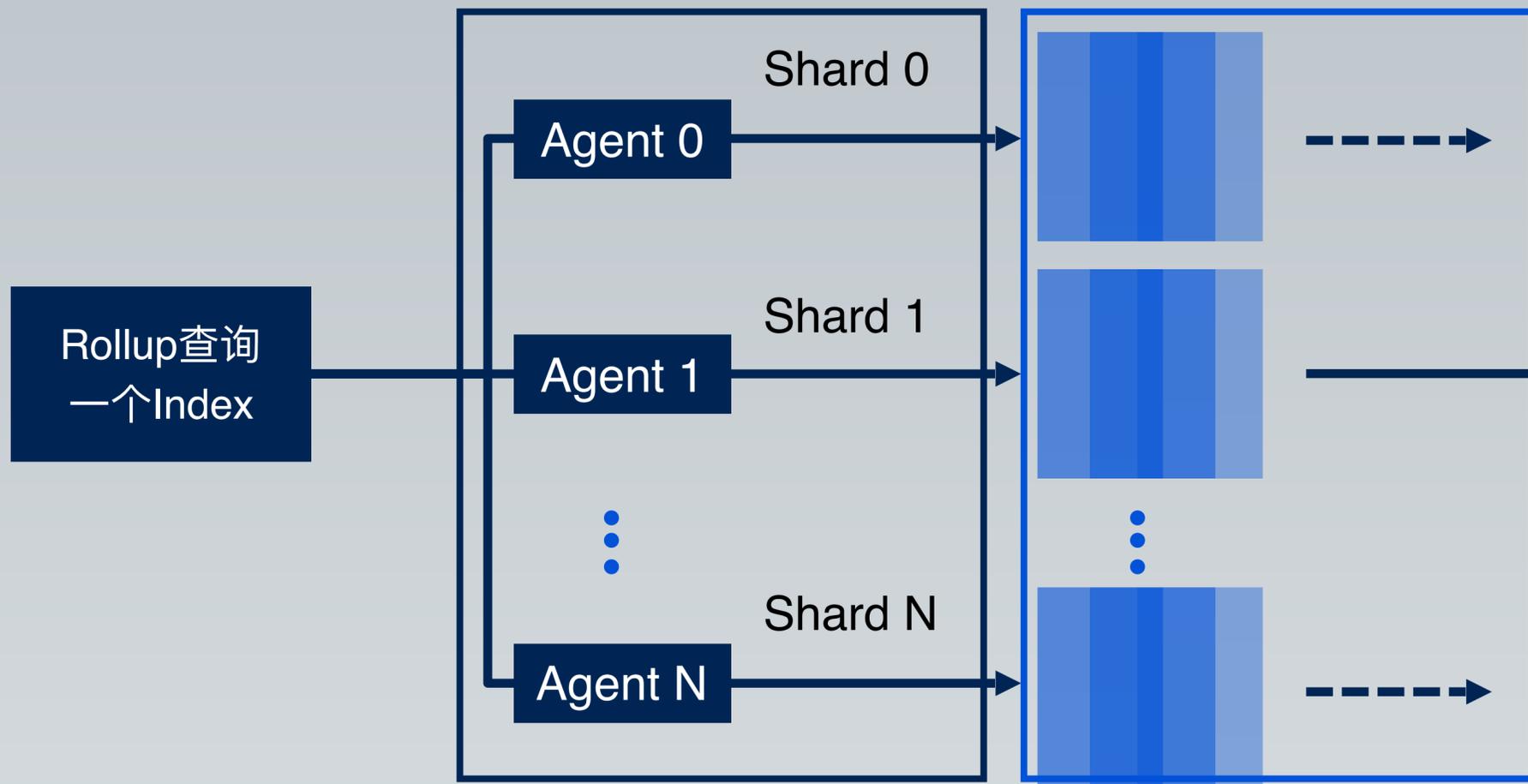
计算、内存开销低

Index

Shard

多路并发提升查询效率

流式查询降低内存消耗

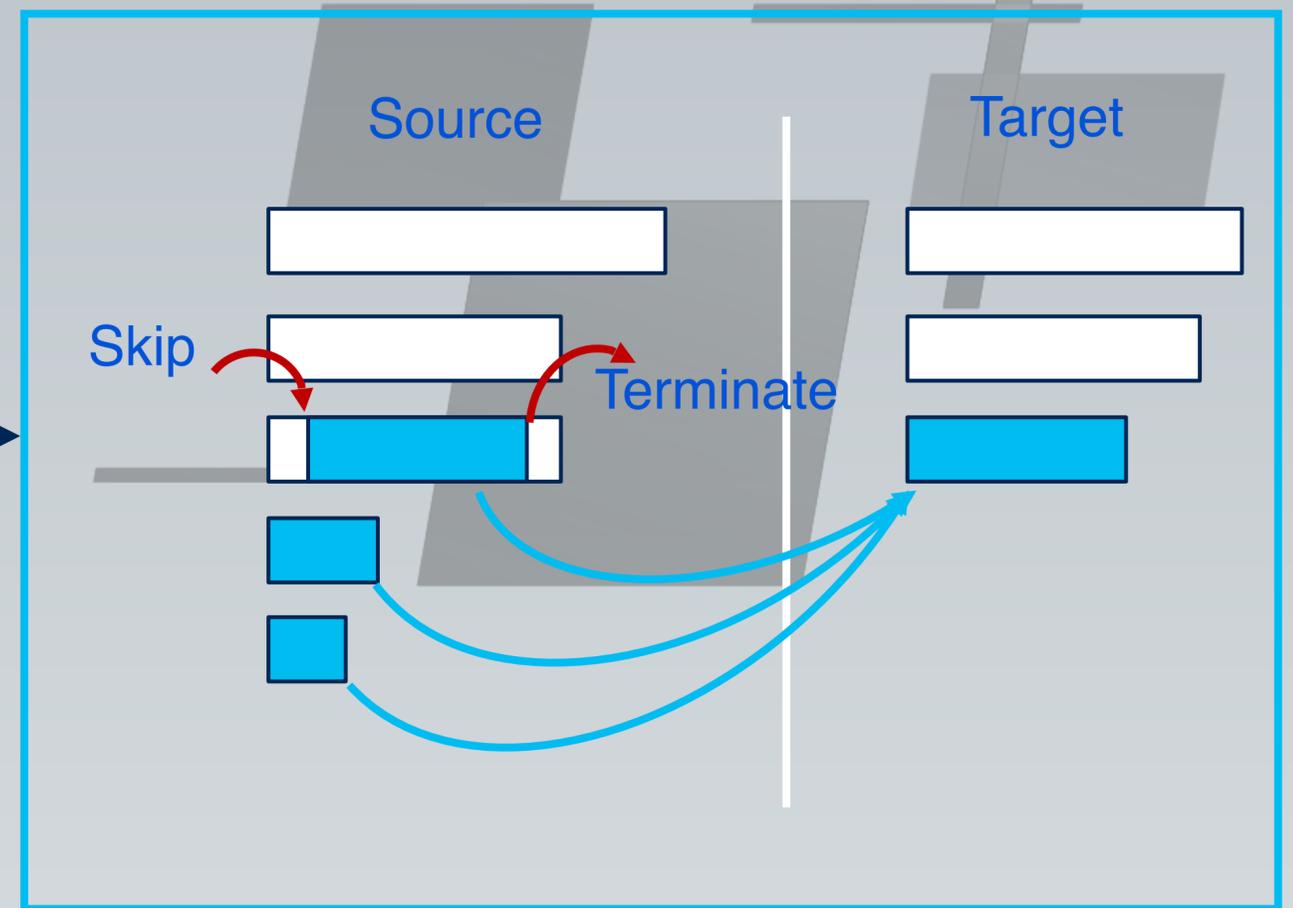


原始数据					
TIME	VALUE	TIME	MAX	MIN	AVG
2019-08-01 00:00:00	10.0	2019-08-01 00:00:00	15.0	2.0	8.0
2019-08-01 00:01:00	12.0				
2019-08-01 00:02:00	11.0				
2019-08-01 00:03:00	9.0				
2019-08-01 00:04:00	10.0				
....					
2019-08-01 23:56:00	15.0				
2019-08-01 23:57:00	2.0				
2019-08-01 23:58:00	3.0				
2019-08-01 23:59:00	4.0				
.....					

↓
1440

Lucene

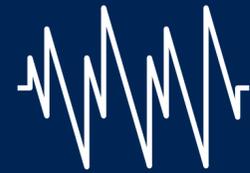
数据排序 + Skip/Terminate 降低查询耗时



性能优化 – 面临挑战

写入

- 代表场景：
日志、时序数据
- 高并发写入：
1000w级写入吞吐
自带主键(去重需求)性能衰减1倍



查询

- 代表场景：
搜索服务
- 高并发、低延时查询：
平响20ms, P95 100ms,
100w级 QPS
避免查询毛刺

性能优化 – 解决方案

数据恢复优化

提升节点重启速度 **PR**

尽可能使用本地文件
恢复数据

写入优化

主键去重

带主键写入
性能提升 50%

向量化

减少分支跳转、
指令 Miss

存储裁剪

减少冗余计算

查询优化

Merge 策略
优化Merge策略、
冷数据自动Merge

执行计划 **PR**
基于CBO模型的
执行计划优化

查询剪枝 **PR**
基于索引范围的
请求剪枝

查询毛刺优化 **PR**
优化查询
缓存逻辑

性能优化 – Merge 策略

原生策略:

大小相似性 + 最大上限 [5GB]



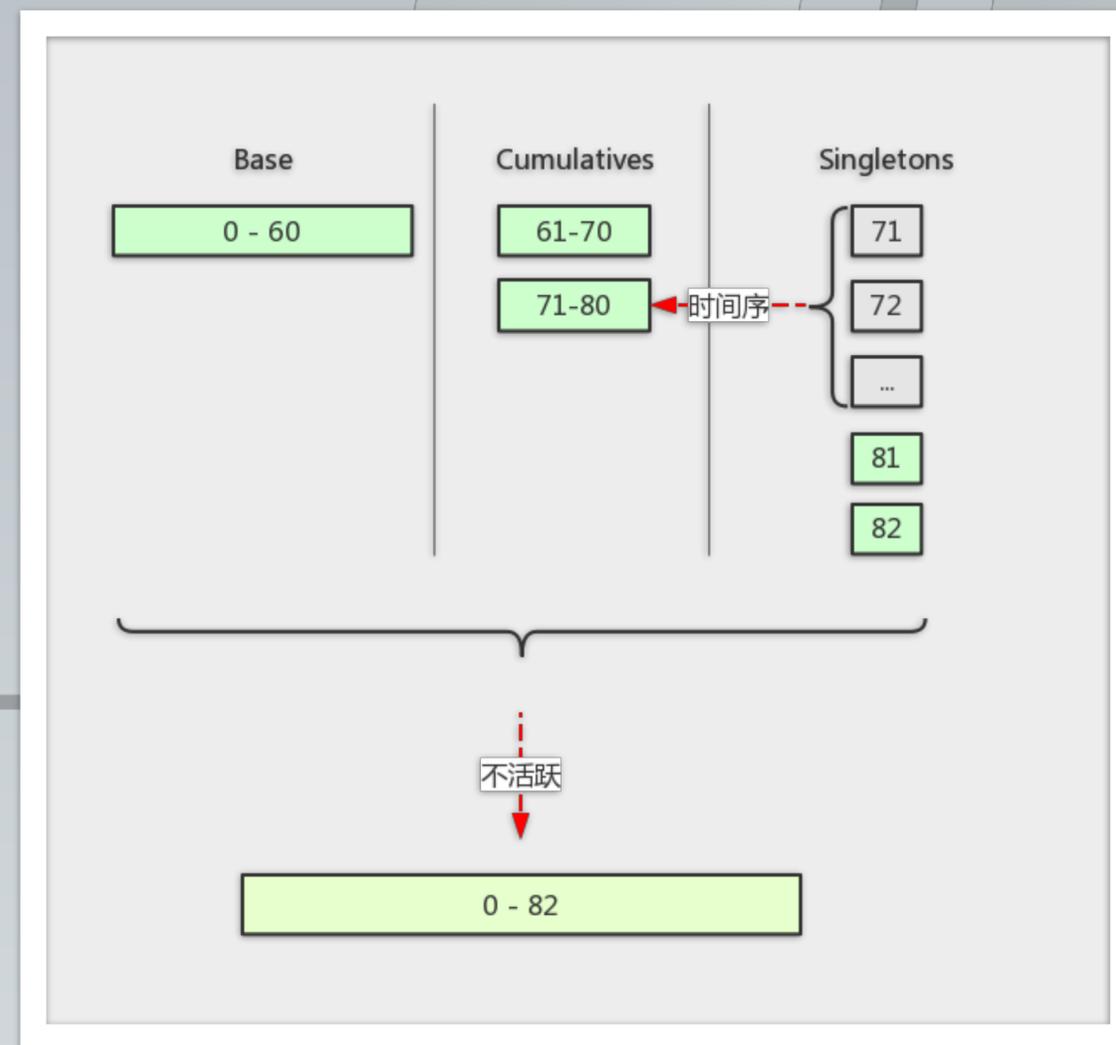
时间序Merge:

增加时间邻近特性, 方便查询裁剪

冷数据Merge:

提高查询性能, 降低资源开销

效果: 搜索场景性能提升1倍



技术及产品持续建设



Thanks



腾讯云
ES 产品



腾讯云
云+社区 ES 专栏

Elastic 中文社区：<http://elasticsearch.cn>

Elastic Meetup 是由 Elastic 官方中文社区定期举办的线下交流活动，主要围绕 Elastic 的开源产品

(Elasticsearch、Logstash、Kibana 和 Beats) 及周边技术，探讨在搜索、数据实时分析、日志分析、安全等领域的实践与应用。

欢迎加入 Elastic 中文社区，参与分享交流或赞助我们的社区活动

深圳联络人：杨振涛

微信：nodexy

邮箱：nodexy@qq.com



微信扫码关注

本次活动回顾 以及活动现场的照片等 会在“vivo 互联网技术”公众号发布 敬请关注