
elastic
中文社区
vivo
腾讯云·云社区
ELASTIC 深圳
MEETUP
2019年4月20日
广东深圳市南山区高新科技园中区一路腾讯大厦2F多功能厅
http://elasticsearch.cn

Elasticsearch

在腾讯的大规模实践

johngqjiang（姜国强）

腾讯 - 云架构平台部 – 高级工程师

腾讯Elasticsearch、时序数据库等系统的技术负责人

目录

日志实时分析

典型场景

- 访问日志
- 慢日志、错误日志
- 审计日志
- ……

主要特点

- 时效性：分钟级可见
- 高并发写入：100w/s，PB级~
- 秒级交互式分析
- 完整的解决方案，易运维

非适用场景

- 离线计算场景
- 海量日志冷备
- ……

时序数据分析

典型场景

- 监控统计
- 物联网传感器数据
- ……

服务器监控　　智能硬件

应用监控　　工业流水线

主要特点

- 高并发写入：1000w/s+
- 海量存储：与保存时间成正比
- 多维度、交互式、灵活可扩展的统计分析

竞品：InfluxDB

- 分布式架构商业化
- 读写性能接近
- 更高的压缩方案

搜索服务

典型场景

- 商品搜索
- 文档搜索
- 站内搜索
- ……

主要特点

- 高并发查询：50w/s
- 平稳低延时：平响15ms，P95 75ms
- 高可用：可用性4个9，跨机房容灾

目录

挑战性的环境

混合云

公有云
私有云
内部云

丰富场景

日志实时分析
时序数据处理
搜索服务

高压力

400+节点
1000w/s写入
50w QPS

可用性

成本

性能

可用性挑战

系统健壮性

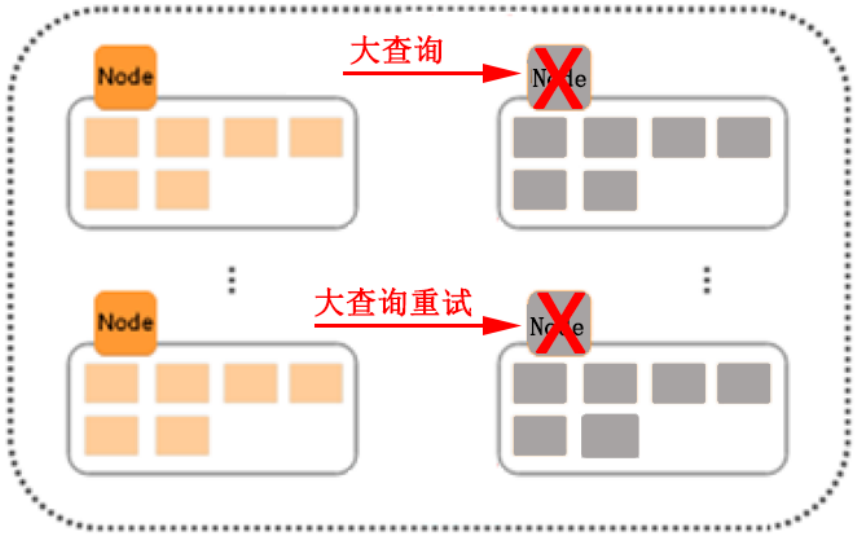
- 容忍异常：大查询、压力过载、网络分区等
- 集群均衡：多节点、多盘等

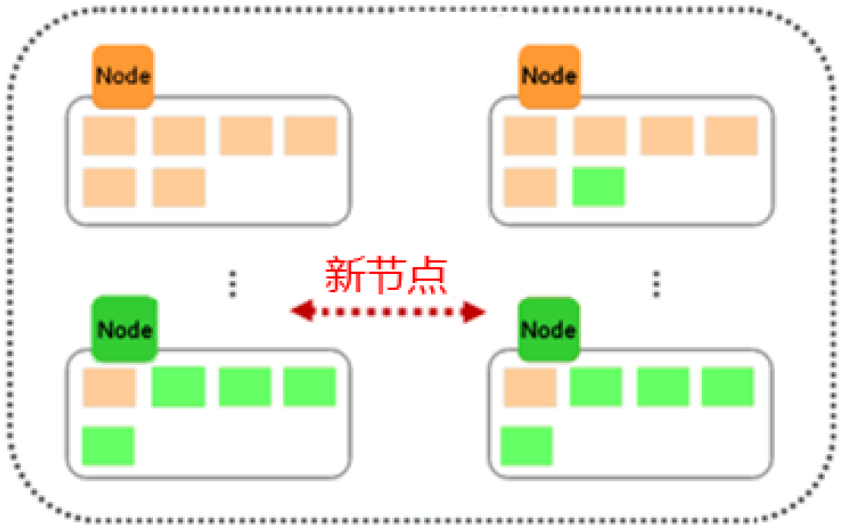
容灾方案

- 集群长时间不可用
- 数据损坏

Bug

- Master堵塞
- Recovery分布式死锁
- 异常链接复用
- ……





成本挑战

矛盾

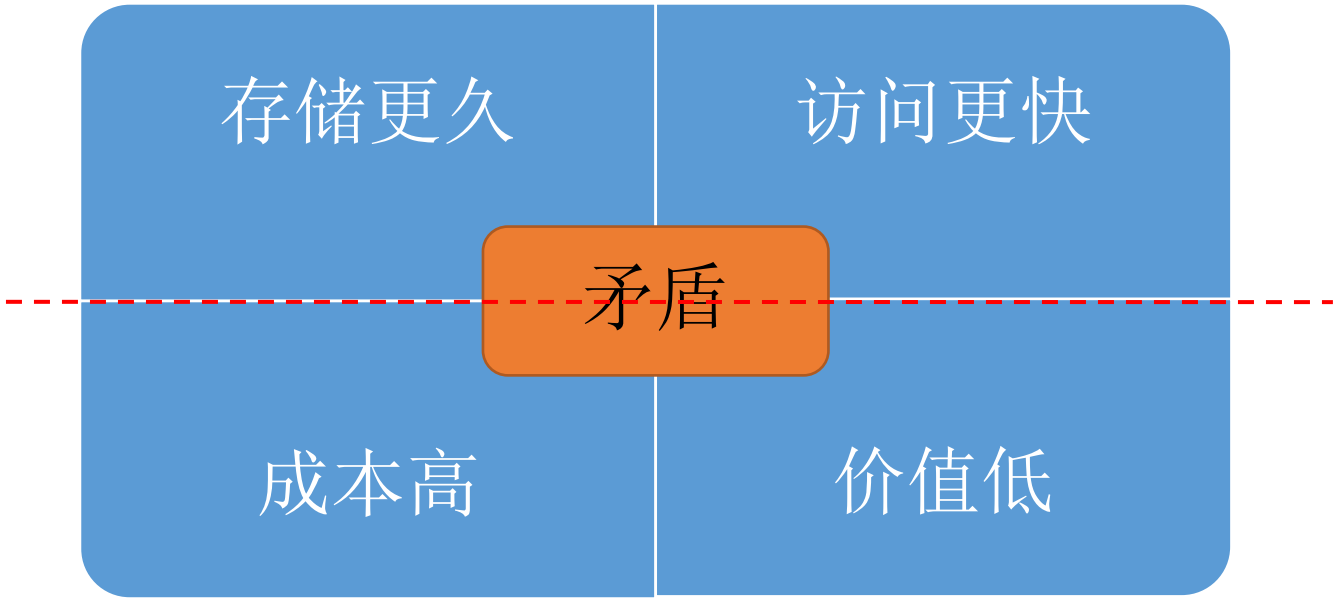
- 用户需求 VS 成本
- 日志、监控等时间序数据量大

硬盘成本

- 时间序数据：海量存储
- 存储放大：3倍

内存成本

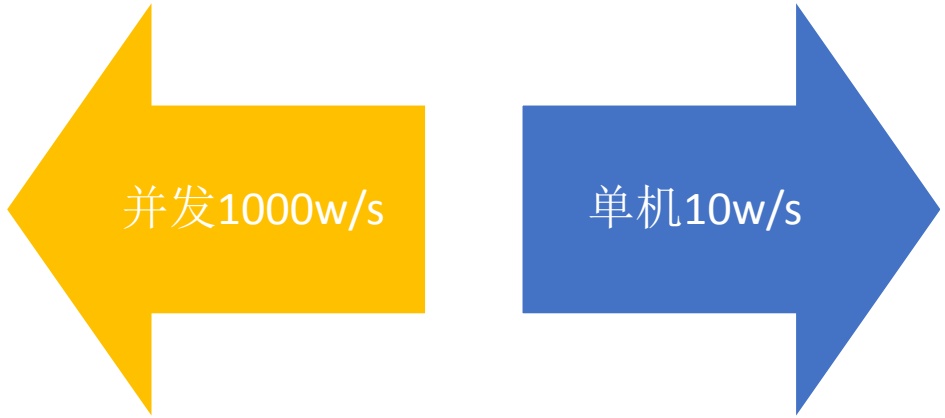
- JVM堆内存使用率高
- 扩容、关闭历史索引
  - 磁盘利用率低、维护成本高



性能挑战

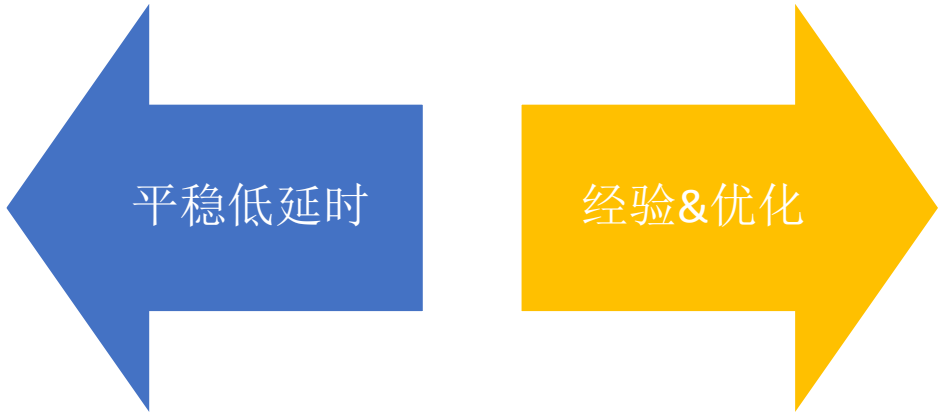
写入

- 典型场景：时序数据、日志
- 高并发写入 VS 单机低吞吐
  - 自带主键(_id)性能衰减1倍~



查询

- 典型场景：搜索服务
- 搜索高要求 VS 延时&毛刺
  - 期望：平响10ms~，P95 100ms~，50w QPS
  - 丰富的维护经验
  - 深度调优



目录

高可用方案

系统健壮性

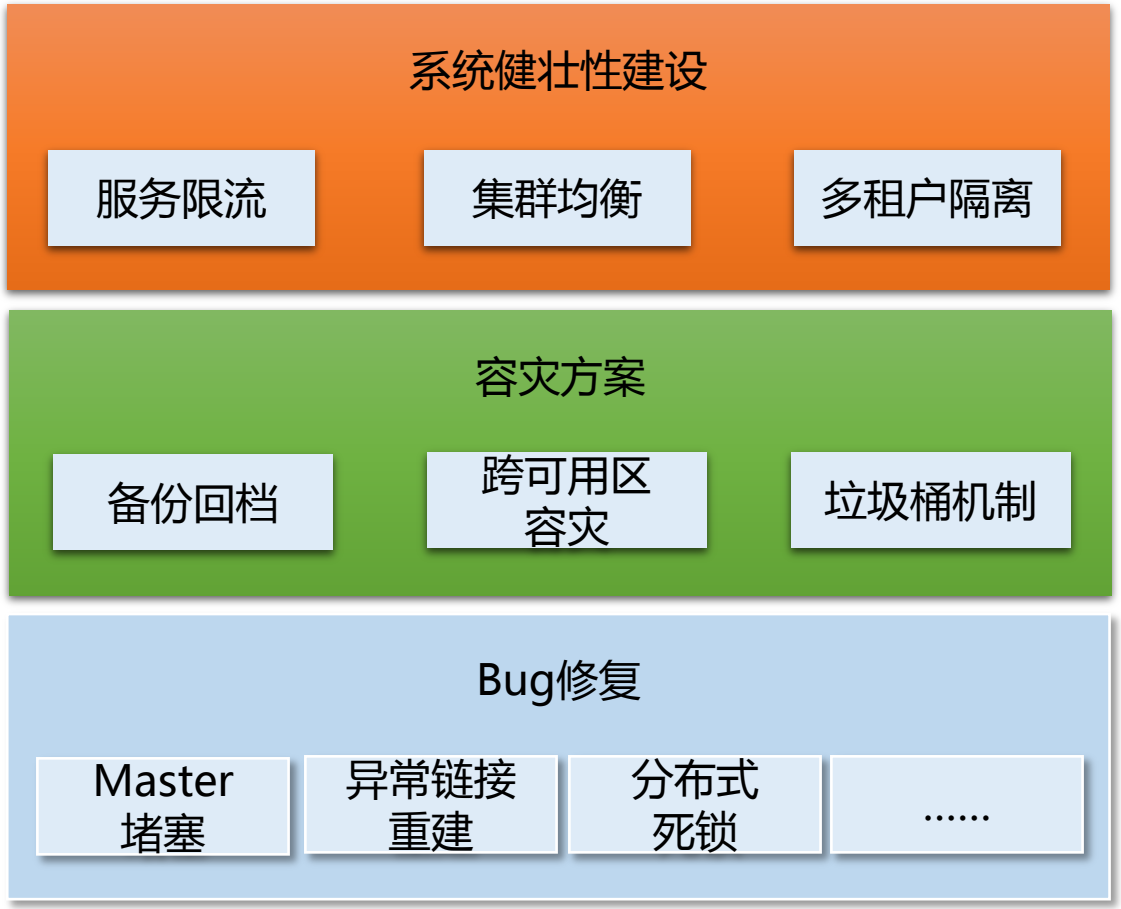
- 容忍大查询、压力过载、网络分区等
- 优化集群多节点、多盘的分片/空间/压力均衡
- 多租户资源隔离：Cgroups

容灾方案

- 保障灾难场景下的数据可恢复
- 容忍单机房故障
- 实例销毁保护

Bug修复

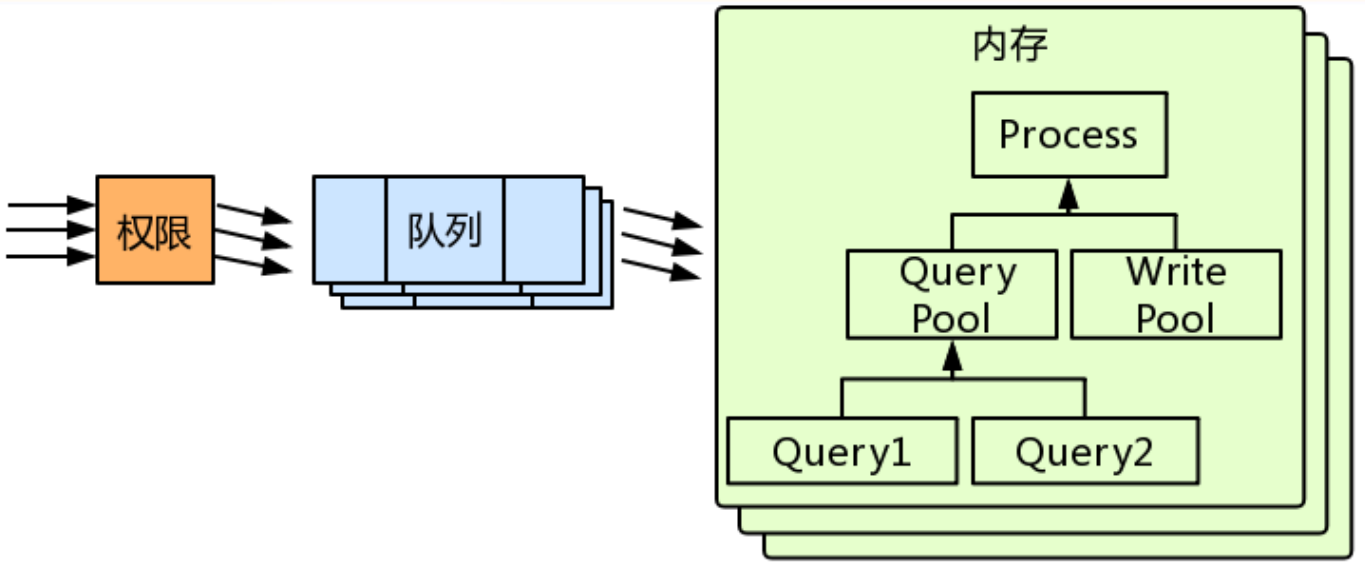
- 已知问题修复



高可用方案 – 服务限流

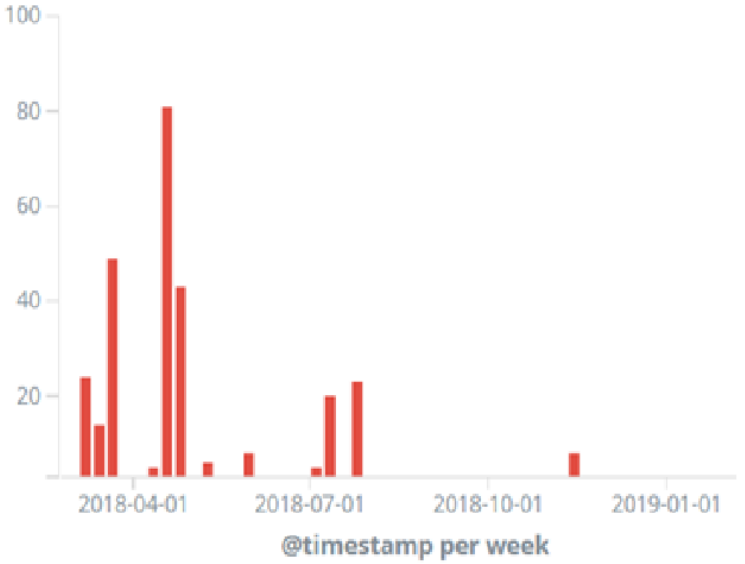
问题

- 压力过载、大查询等导致的集群崩溃
- 网络分区结束后集群不可恢复



方案：多维度、适应性强

- 权限系统：限制攻击、误操作等
- 并发限流：降低资源抢占，优化队列模型、任务去重
- 内存限流：进程/内存池/查询分层，兼顾执行/协调节点
- 多租户隔离：Cgroups



服务不可用告警降低

成本优化解决方案

存储优化：时间序数据存储成本高

- 使用混合存储进行性能、成本的平衡
- 聚合数据降低成本的同时，提高查询性能
- 更多冷数据处理方式：生命周期管理/备份归档

内存优化：JVM内存不足

- 提升内存利用效率：淘汰冷内存
- 降低堆内存使用：降低GC，提高节点规格

存储成本 - Rollup

问题

- 存储更久数据 VS 降低存储成本&提高访问性能

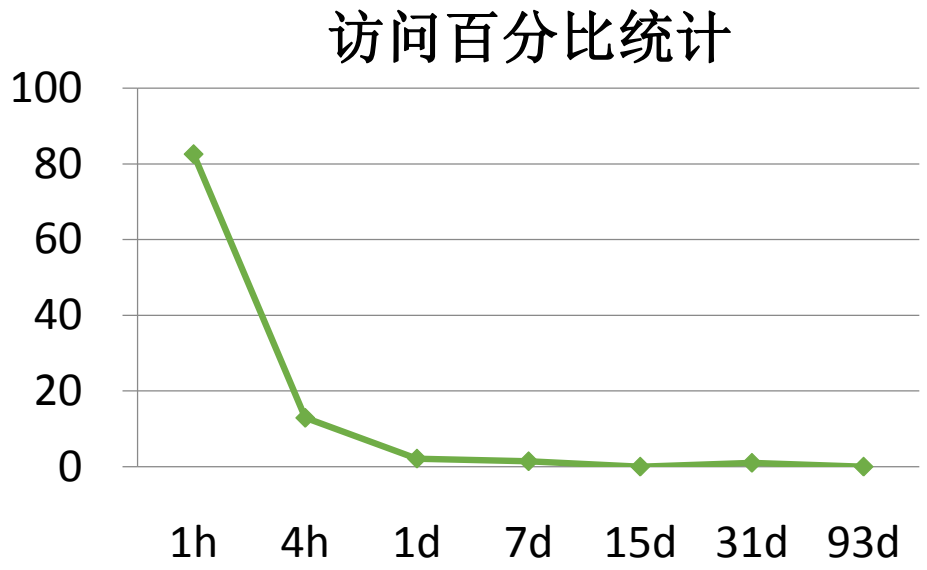
场景分析

- 冷热特性明显
- SSD&HDD性能、价格互补
- 历史数据查询统计结果

方案

- 数据预排序
- 流式Rollup
- 开销：计算<10%写入、内存100MB

粒度	1分钟	5分钟	1小时	1天
数据量(GB)	916	255	30	3



	region	host	time	cpu_usage
1	region	host	time	cpu_usage
2	gz	10.11.12.13	2017-07-01 10:00:00	20
3	bj	14.15.16.17	2017-07-01 10:00:00	20
4	gz	10.11.12.13	2017-07-01 10:00:10	30
5	bj	14.15.16.17	2017-07-01 10:00:10	40
6	gz	10.11.12.13	2017-07-01 10:00:20	25

	region	host	time	cpu_usage
1	region	host	time	cpu_usage
2	gz	10.11.12.13	2017-07-01 10:00:00	25
3	bj	14.15.16.17	2017-07-01 10:00:00	30

内存成本

问题

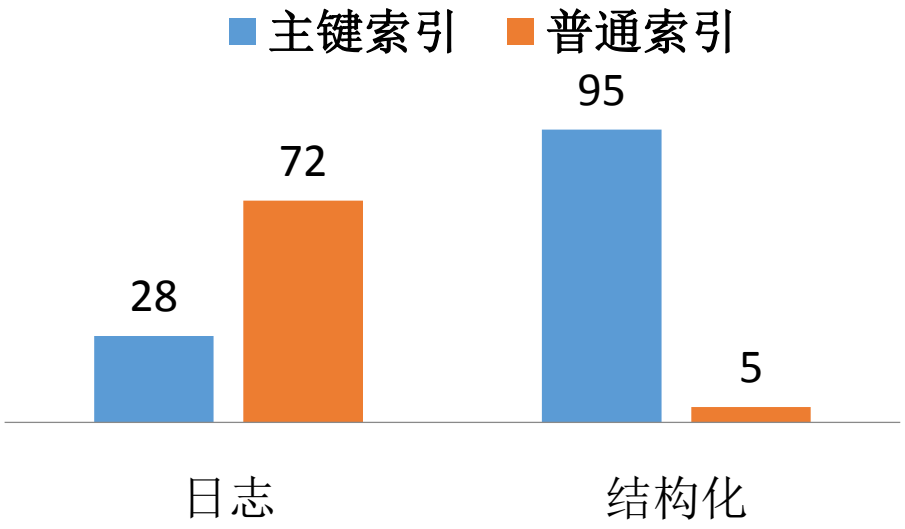
- 降低内存使用 VS 保持查询性能

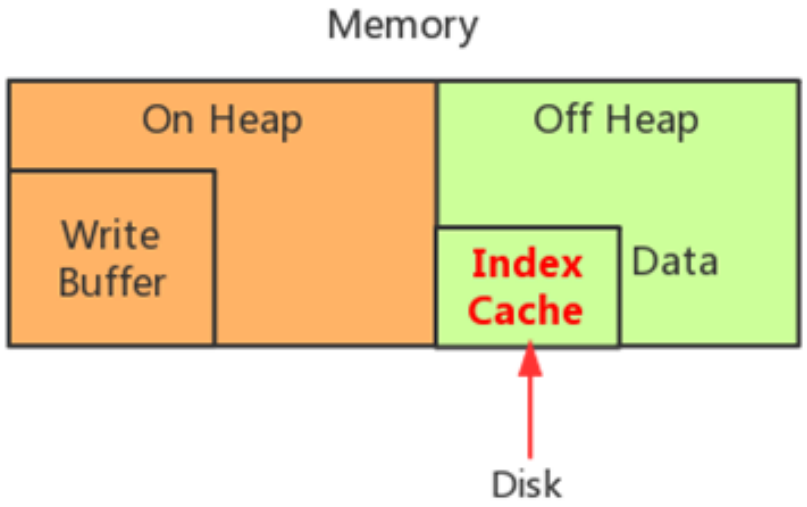
场景分析

- 主要开销：索引，64G内存支持3T存储
- 内存利用效率差：历史数据、主键索引

方案

- LRU Cache：淘汰历史数据、主键索引
- Off Heap：降低GC，增强扩展能力





性能优化方案

写入：时序场景1000w+

- 优化主键去重性能，性能提升~1倍
- 减少分支跳转、指令Miss
- 减少冗余计算

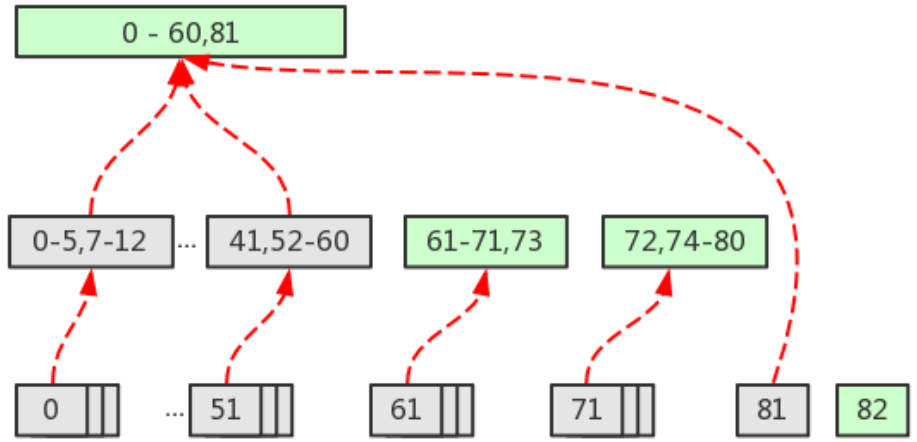
查询：搜索服务平稳低延迟

- 优化Merge策略、冷数据自动Merge
- 基于CBO模型的查询计划优化
- 基于索引范围的请求剪枝

Merge策略

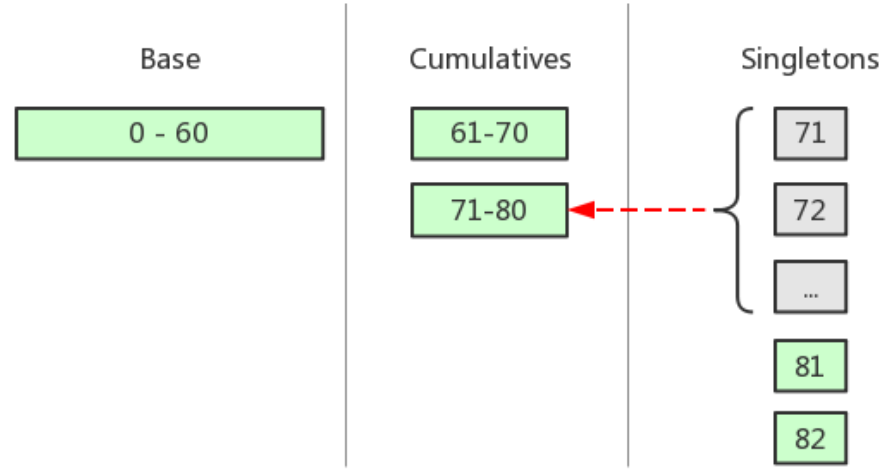
时间序Merge

- 原生策略：大小相似性 + 最大上限
- 问题：时间乱序，不利于裁剪、加重随机io
- 分层Merge：增加时间邻近特性，滚动Merge



冷索引自动Merge

- 不活跃索引：Segment数量30~
- 冷索引、写入较少索引（搜索服务）



效果

- 搜索、时序等场景：性能提升1倍~

目录

技术及产品建设

产品场景

Elasticsearch

日志分析

时序数据服务

全文检索

生态融合

ELK生态

Hadoop生态

ES 内核

可用性	成本	性能	功能特性
服务限流	冷热分离	查询优化	权限系统
均衡策略	Rollup	写入优化	时序接口
多租户隔离	内存优化	Merge策略	中文分词

管控平台

管控操作	运维支撑	任务执行引擎	运维工具
资源管理	监控告警	备份回档	数据迁移
实例管理	审计日志	数据导出	压测工具
作业管理	健康巡检	生命周期管理	

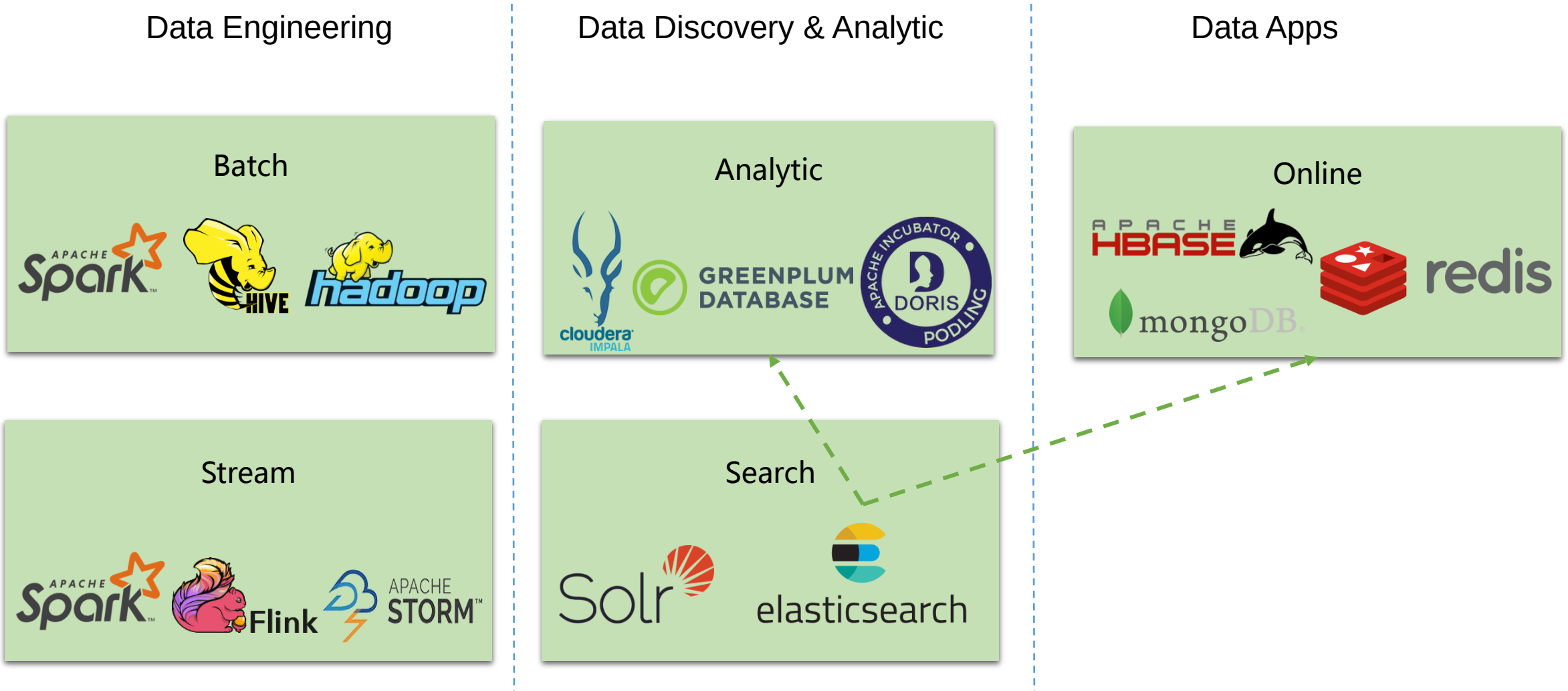
腾讯云

Elasticsearch服务

CTSDB服务

长期探索思考

大数据图谱



开源社区协同

腾讯云 和 Elastic 公司更紧密的合作……

更丰富的产品能力

引入X-Pack

提供更多的商业特性

技术贡献

更深入的参与、回馈 ES 社区，助力Elastic生态建设

开发者资源

举办主题展会、线下沙龙、线上活动、提供开发者试用/购买优惠等

Thank You

腾讯云
ES 介绍

腾讯云
云+社区 ES 专栏