阿里云Elasticsearch实时计算平台实践

阿里巴巴搜索事业部 夏宏举

1. 批次/实时搜索数据场景痛点
2. 实时计算平台架构
3. Elasticsearch内核深入优化解密
4. 强大的数据处理能力
5. 展望未来

阿里云Elasticsearch

1000+

集群数量

6000+

Node数

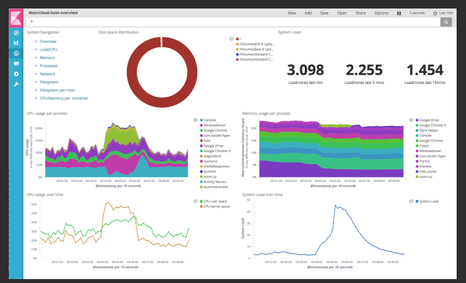
1PB+

数据量

典型场景

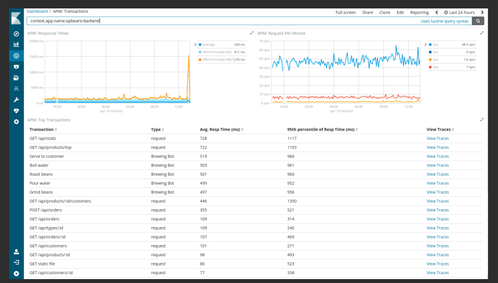
日志分析

应用与网页搜索



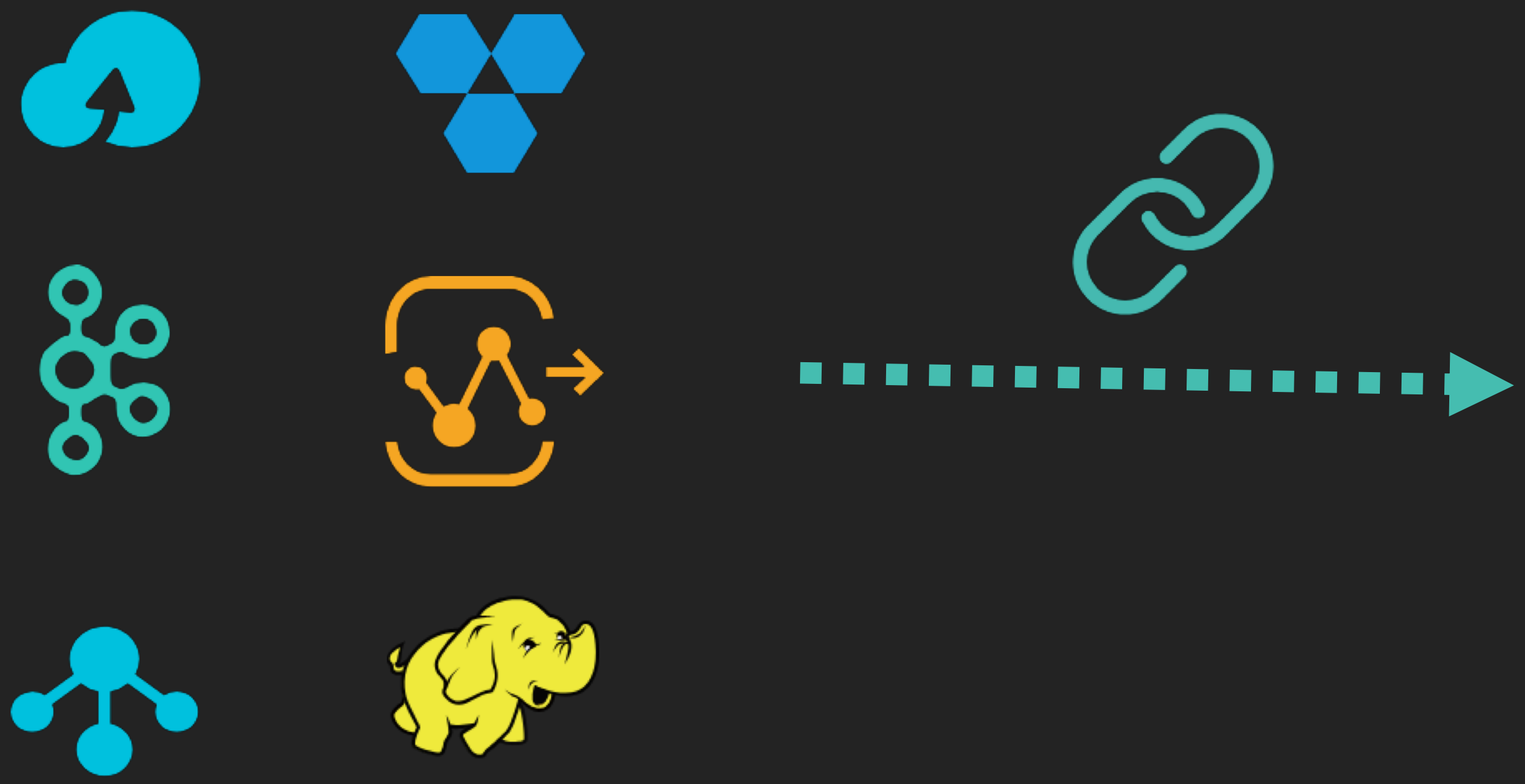
Metrics

安全与业务分析

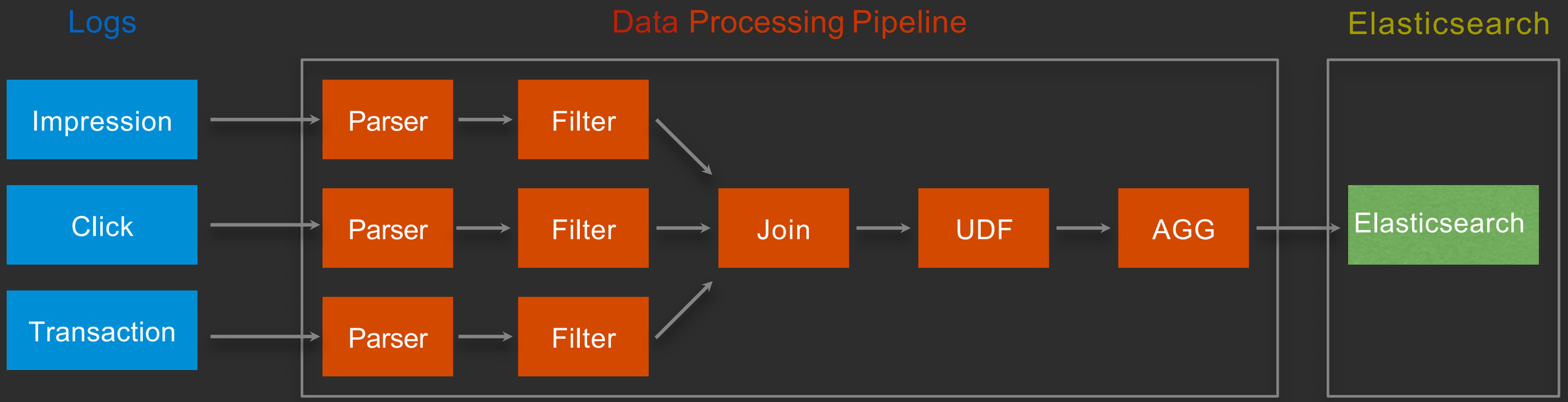


APM

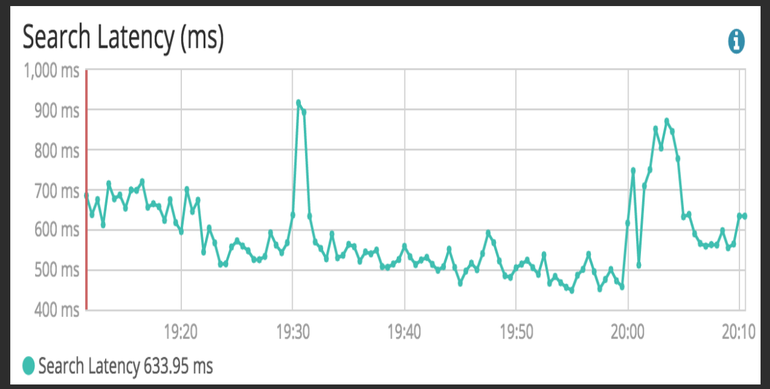
海量数据源如何对接



复杂数据处理逻辑

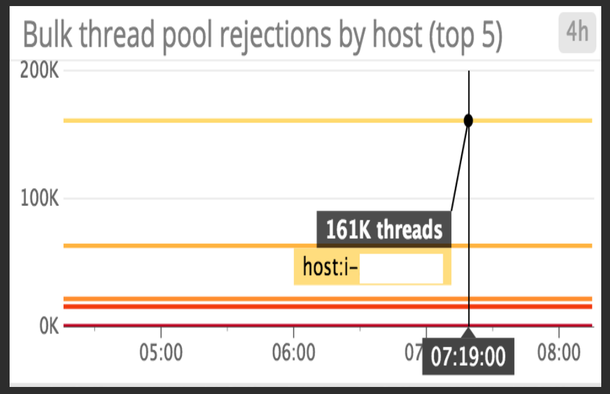

Logs
Data Processing Pipeline
Elasticsearch
Impression
Click
Transaction
Parser
Parser
Parser
Filter
Filter
Filter
Join
UDF
AGG
Elasticsearch

数据导入痛点



批次数据导入影响查询

数据导入争抢在线资源



数据导入速度慢

海量数据越导越慢

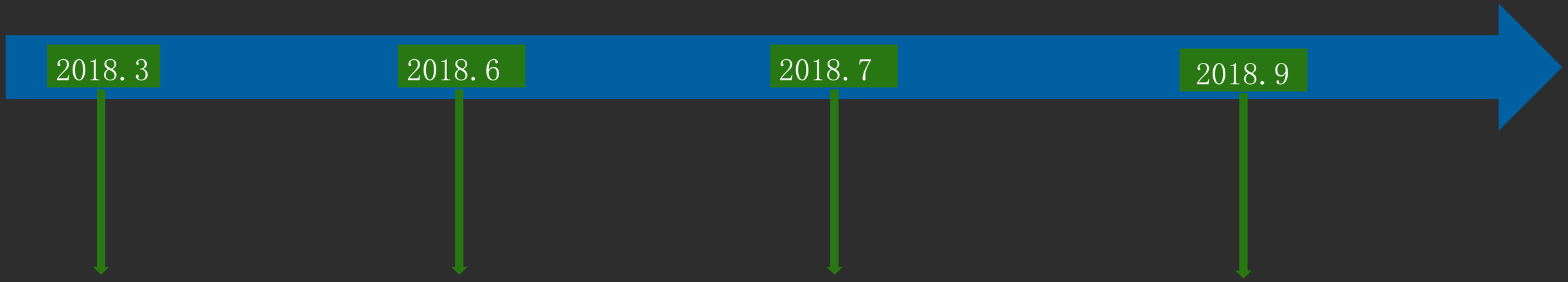
批次/实时数据无损切换

批次数据做完需要追增量保持数据完整无损

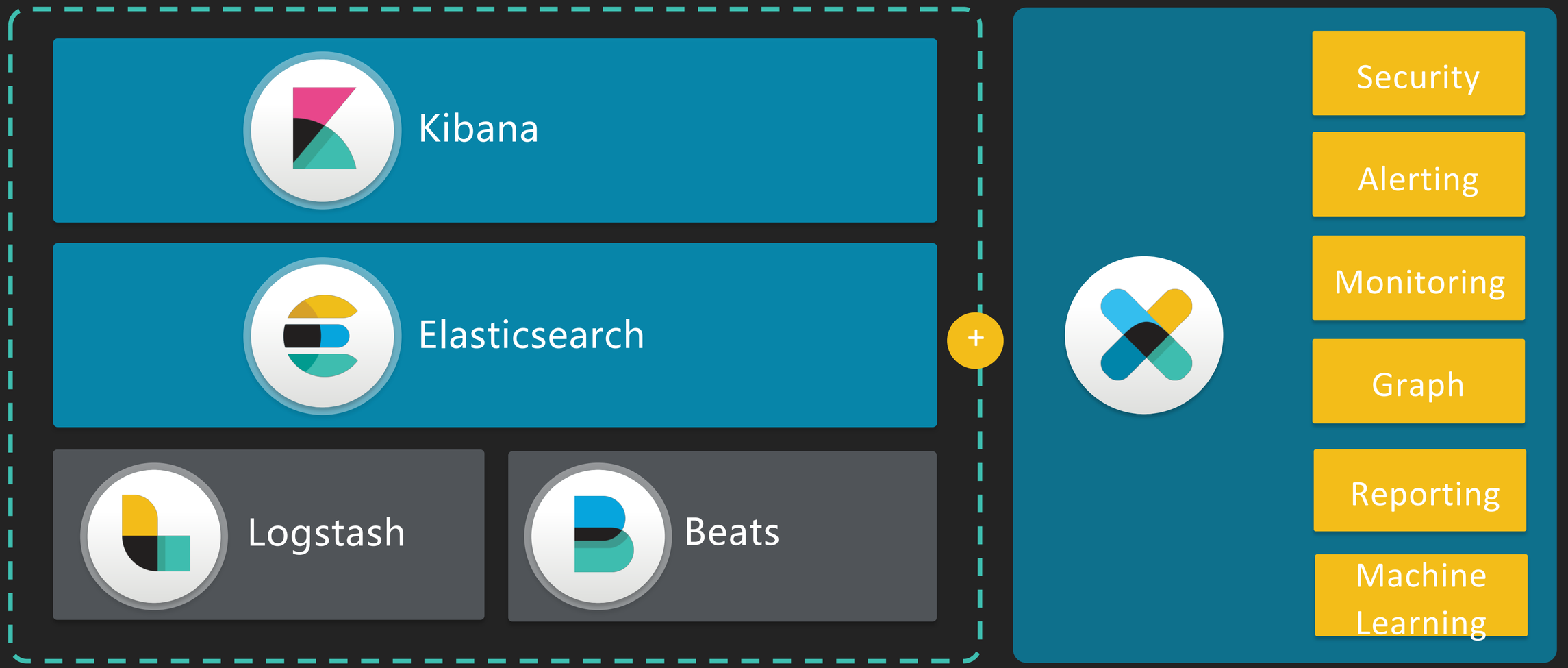
实时计算平台应运而生

致力于解决搜索场景下海量数据批次/实时计算问题, 基于实时计算引擎Blink提供高可用、高性能的搜索离线和实时数据处理能力

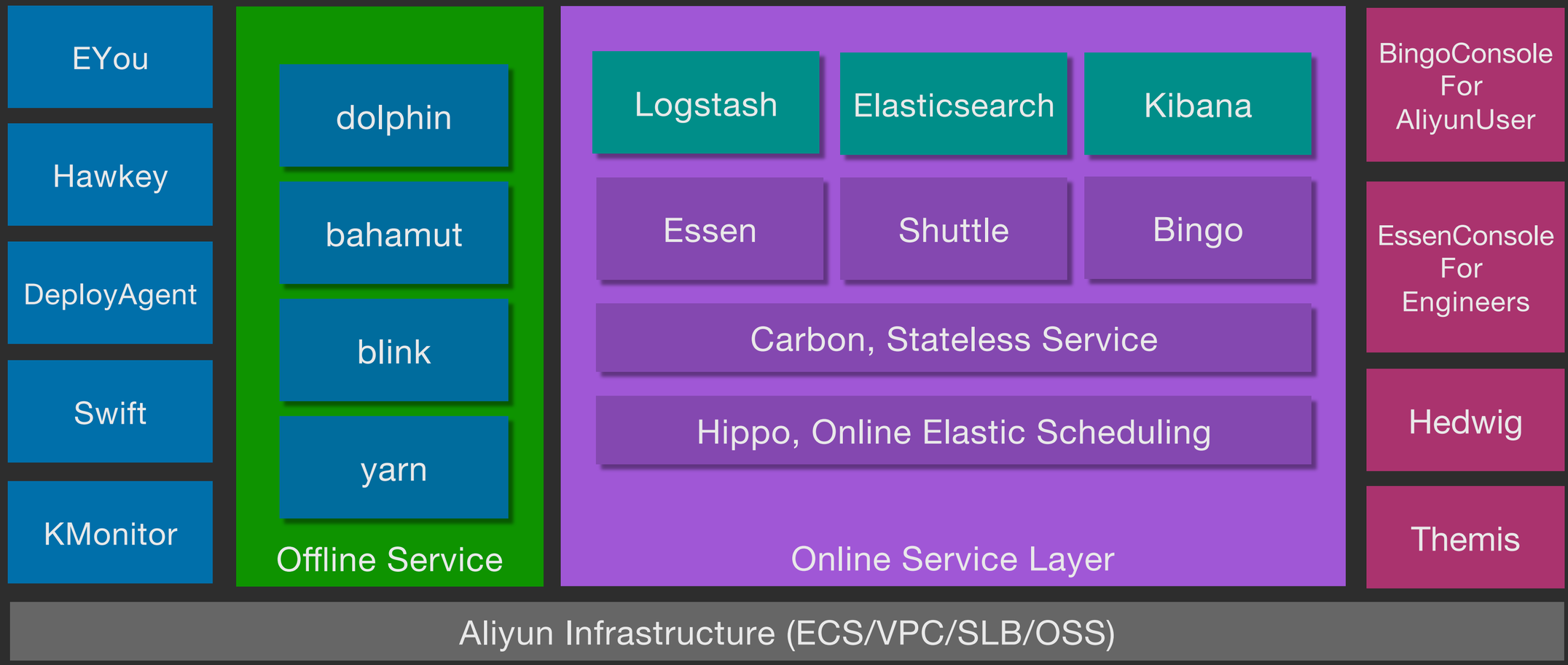
实时计算平台演进



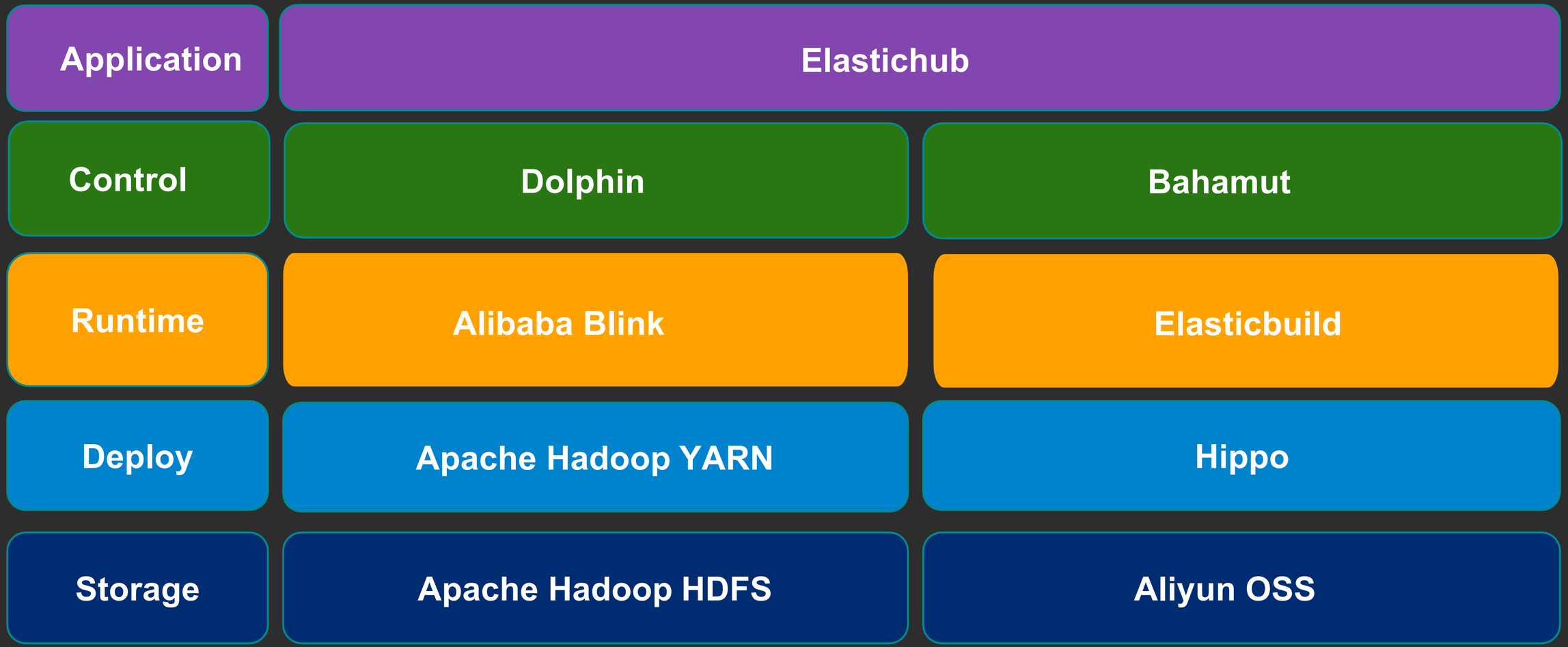
- 调研用户痛点
  - 数据源对接困难
  - 全量数据导入慢
  - 全量/增量切换缺乏管控
- Elasticbuild启动
  - 基于Blink的离线索引构建
  - 针对ES内核进行一系列优化
  - 写入性能提升至在线的4倍
- 与离线组件平台深度合作
  - Dolphin和Bahamut作为离线管控
  - 打通公有云网络环境，明确云上部署方案
- Elastichub发布
  - 常用数据源对接
  - 批次构建索引加速
  - 实时数据计算


Kibana
Elasticsearch
Logstash
Beats
+
Security
Alerting
Monitoring
Graph
Reporting
Machine Learning

阿里云Elasticsearch产品架构



阿里云Elasticsearch实时计算平台



Elasticsearch内核深入优化思路

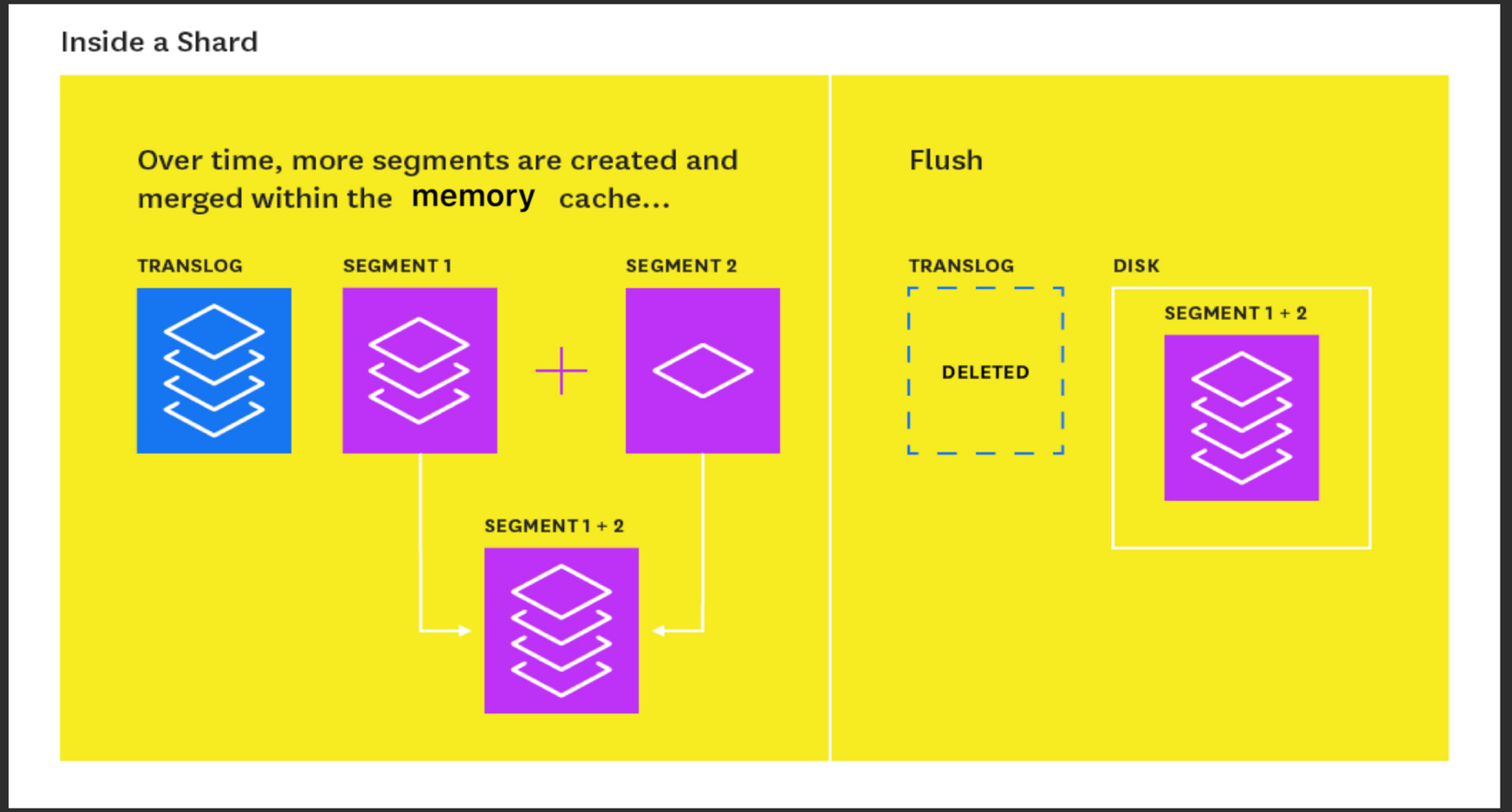
- 硬件方向
  - Elasticsearch支持读写HDFS, 分布式存储, 高可用
- 软件方向
  - 系统架构
    - 规模弹性 ：当Shard数固定后, 缺乏了弹性扩缩容的能力
    - 优化 ：将Shard数据进一步拆分, 达到并发构建, 随时扩缩集群规模的能力
  - Elasticsearch内核级别
    - Translog ：Failover需要通过Translog回追数据, 增加了大量IO消耗
    - 优化 ：利用Blink checkpoints机制替代Translog
    - 索引合并 ：索引合并的IO消耗巨大
    - 优化 ：通过内存Merge可以减少Merge索引的IO读写

HDFS Directory

- 扩展Lucene的BaseDirectory
  - Elasticsearch可以直接读写HDFS
  - 分布式弹性扩展能力
  - 良好的数据读写性能
  - 使用高效云盘成本大大降低
- Block Cache优化
  - 缓存热点HDFS Block块
  - 等价文件系统的PageCache

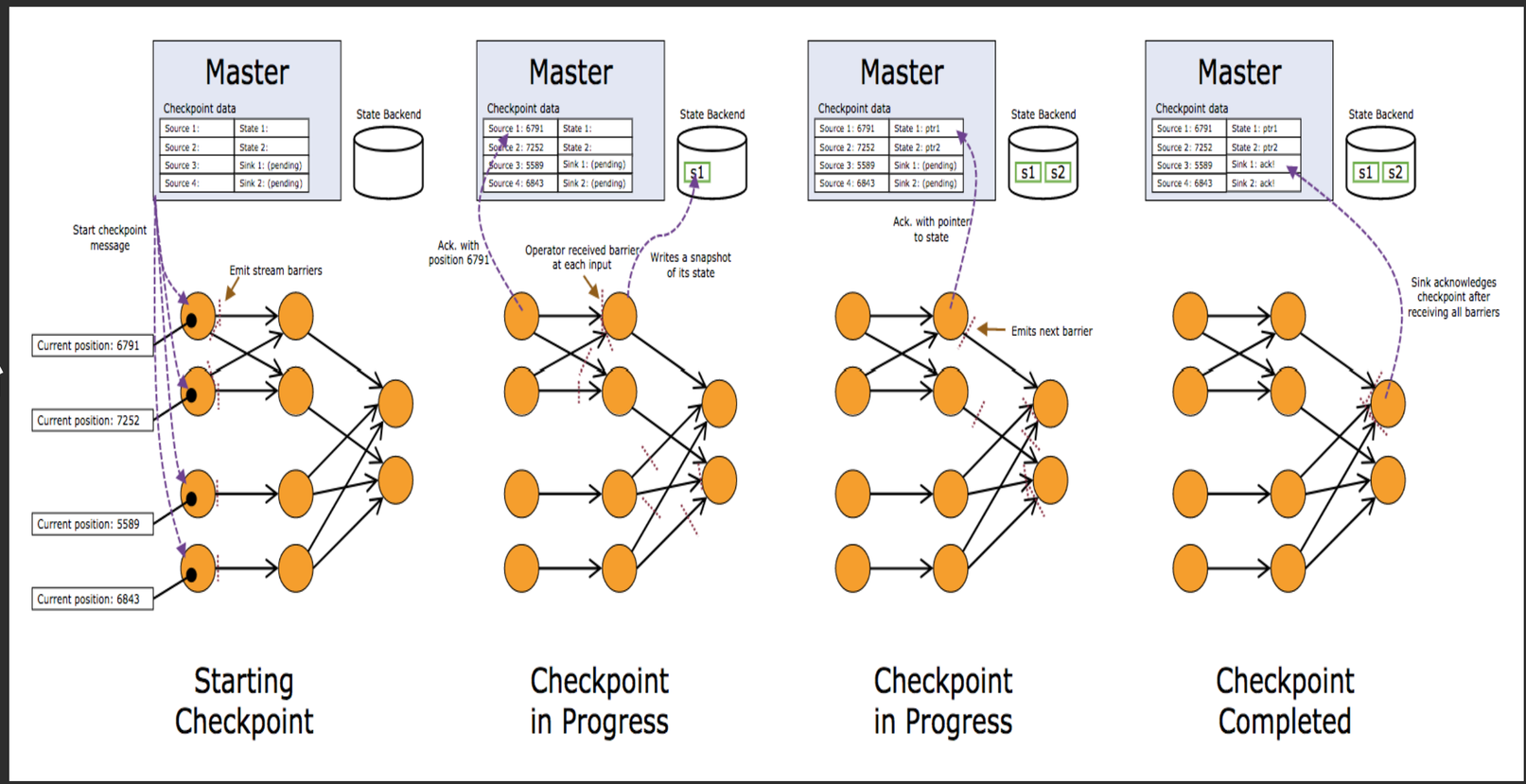
NRTCache实现内存索引合并

- RAMDirectory机制实现
- 不落地直接在内存中合并
- Cache内存大小可配置



优化Failover机制

- 基于Blink Checkpoints
  - 秒级Failover恢复
  - Exactly Once保证数据准确
  - 与HDFS Directory配合使用
  - Recovery从上次Checkpoint开始
- 移除Translog显著降低IO消耗



Blink Checkpoints & Recovery

Elasticsearch内核优化解密

并发索引构建

Scaling能力，速度随资源可扩展

Local模式无网络开销

集群节点间无数据传输

自定义Snapshot

Shard级别Snapshot，自动形成完整Snapshot

一系列优化后相比在线构建，性能提升4倍

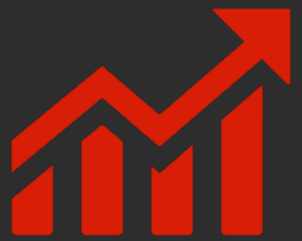
相同资源情况下

强大的数据处理能力

SQL & Logstash Filter Plugins

Why SQL ?

Declarative



Optimized

Understandable

Stable

Unify

SQL功能


DML & DDL

UDF/UDTF/UDAF

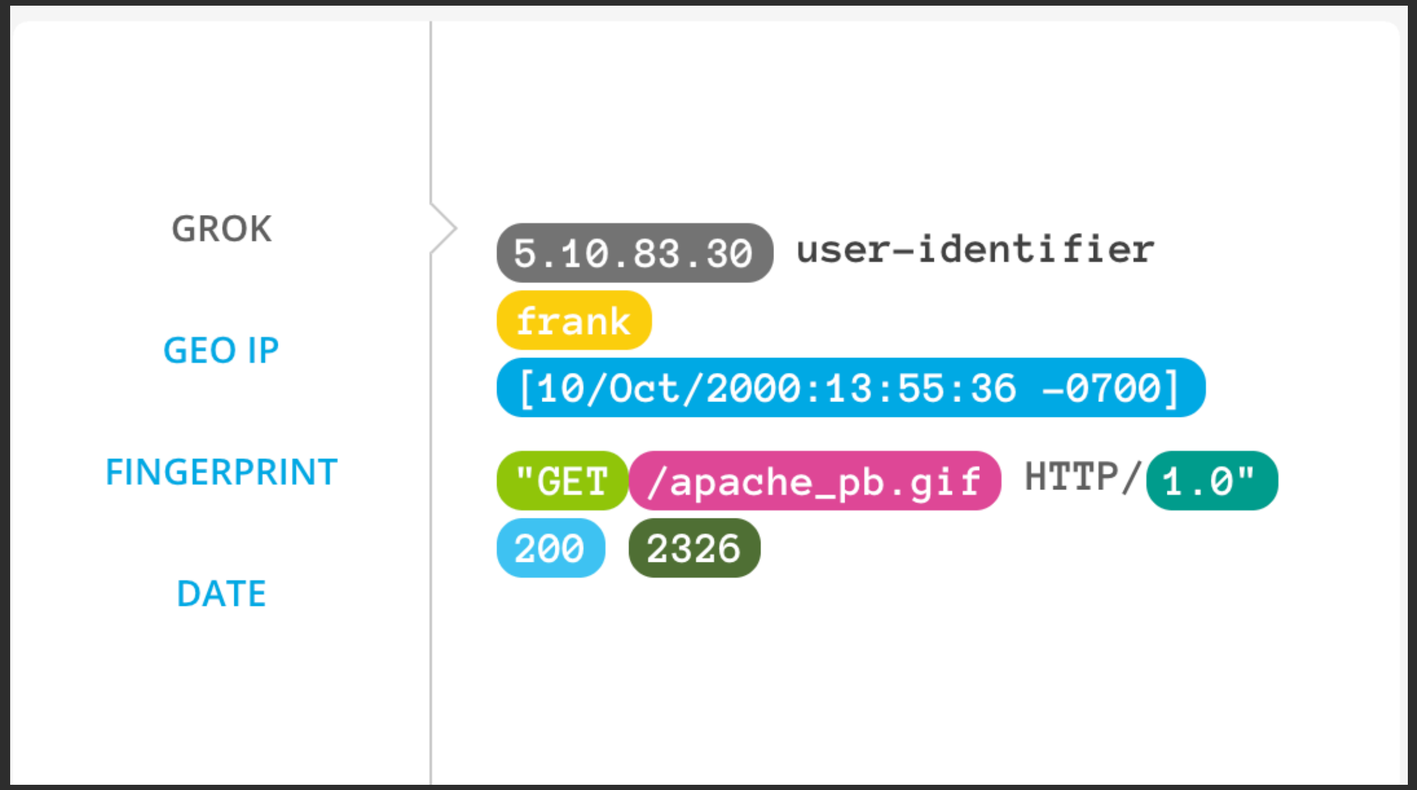

JOIN


Window AGG

ETL & Query Optimization

Logstash Filter Plugins

兼容Logstash Filter Plugins的数据处理协议



展望未来

毫秒级实时可见

从原生的秒级可见提升至毫秒级别

日志场景批次增量

加速数据写入吞吐

OSS作为底层存储

显著降低成本

2018 杭州·云栖大会
THE COMPUTING CONFERENCE

Alibaba Group
阿里巴巴集团

驱动数字中国

EMPOWER DIGITAL CHINA

专业、垂直、纯粹的 Elastic 开源技术交流社区

https://elasticsearch.cn/