How to use Elastic Stack?

白凡@Sunlands

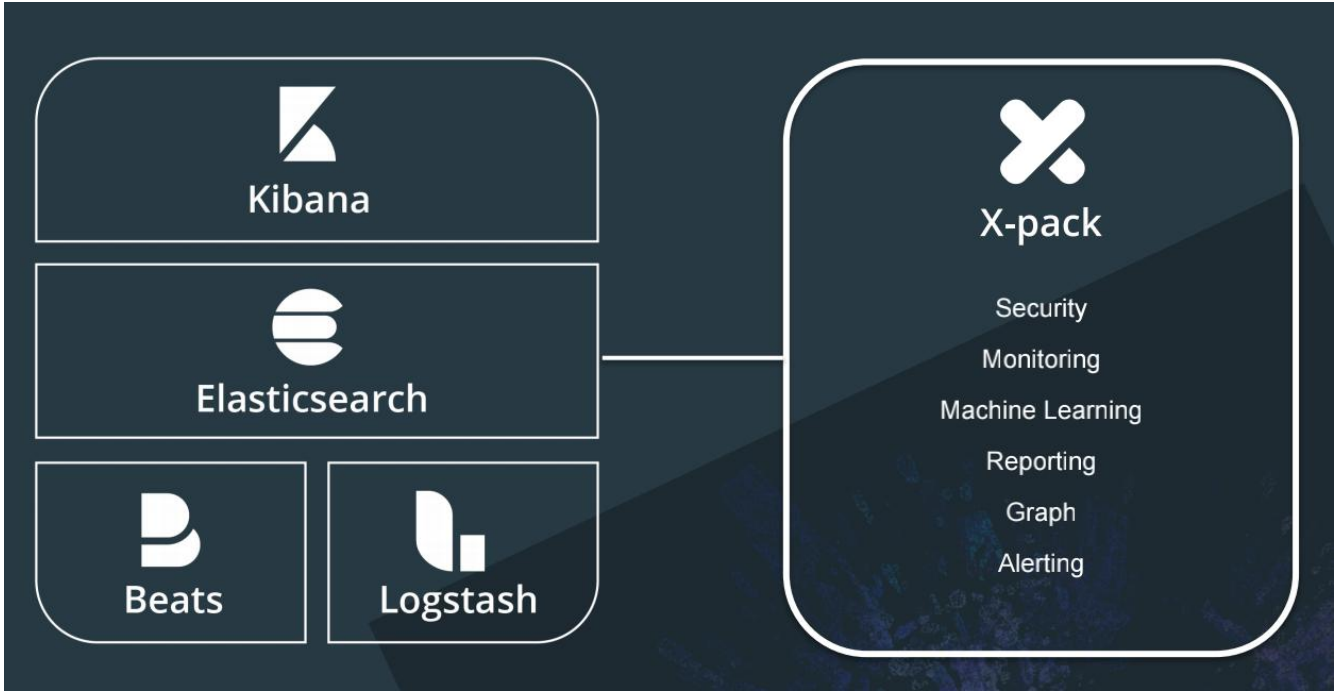
Nov 4, 2017

主要内容

- What is Elastic Stack?
- How to use ES?
- OneIndex
- Sunlands ES Roadmap

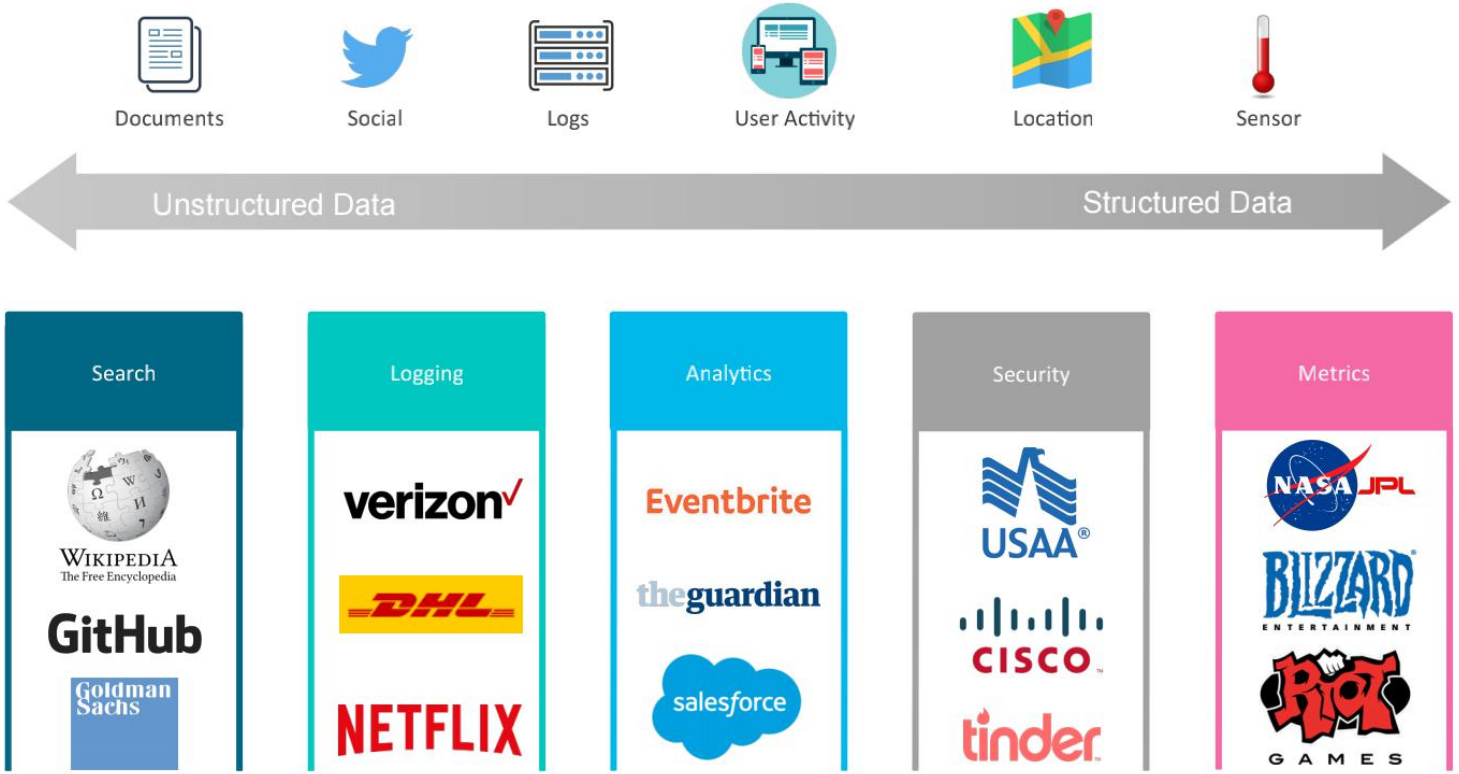
What is Elastic Stack?

Elastic Stack是整个Elastic框架集合，其中主要成员包含：
Elasticsearch,Beats,Logstash,Kibana以及X-pack（收费产品，不开源）



elastic

What is Elastic Stack?



elastic

Elasticsearch

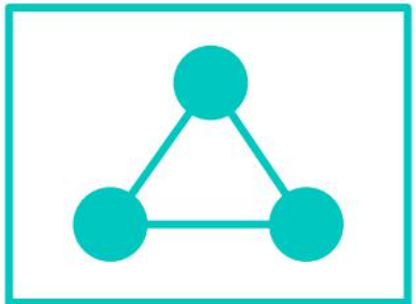
Elasticsearch全文检索功能基于Apache Lucence，其相关度计算算法依据：词条频度/倒排文档频度(TF/IDF)

词条频度(Term Frequency)词条在当前文档中越频繁的话，那么权重就越高。在一个字段中出现了 5 次的词条应该比只出现了 1 次的文档更加相关。

字段长度归约(Field-length Norm)字段越短，那么其权重就越高。如果一个词条出现在较短的字段，如 title 字段中，那么该字段的内容相比更长的 body 字段而言，更有可能是关于该词条的。

Elasticsearch

Platform around a distributed data store



Distributed

Scaleable

Highly Available

API

JSON

Language clients

Search

Aggregations

Geospatial

elastic

How to use ES?

ELK

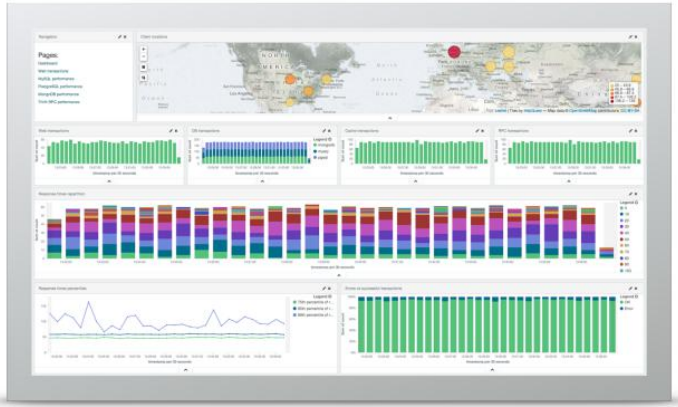
- 目前Elastic Stack最广泛的应用场景之一，用于实时收集监控日志数据。
- Beats/Logstash：对服务器性能日志，业务日志等进行采集。
- Elasticseach：接收采集来的日志数据，并以文件形式写入磁盘存储。
- Kibana：数据可视化平台，对ES数据进行查询展示，具有丰富的Dashboard模板。

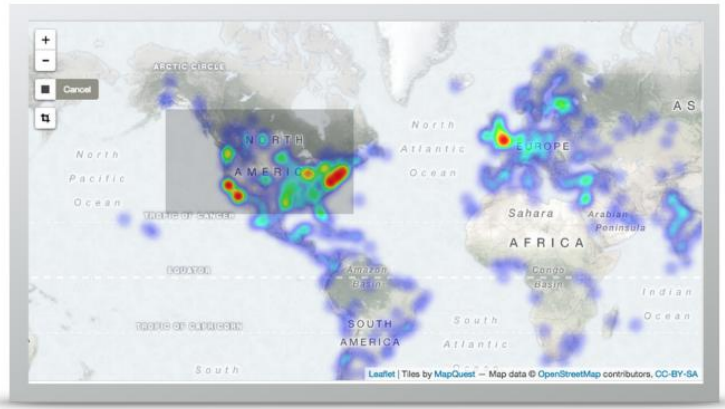
How to use ES?

ELK

Visualize and Explore

kibana





Easily create bar charts, line and scatter plots, histograms, pie charts, and maps

elastic

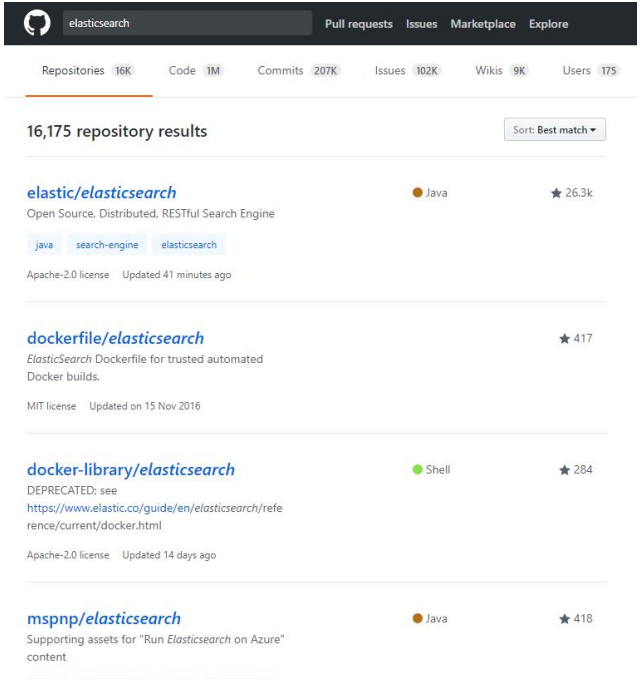
How to use ES?

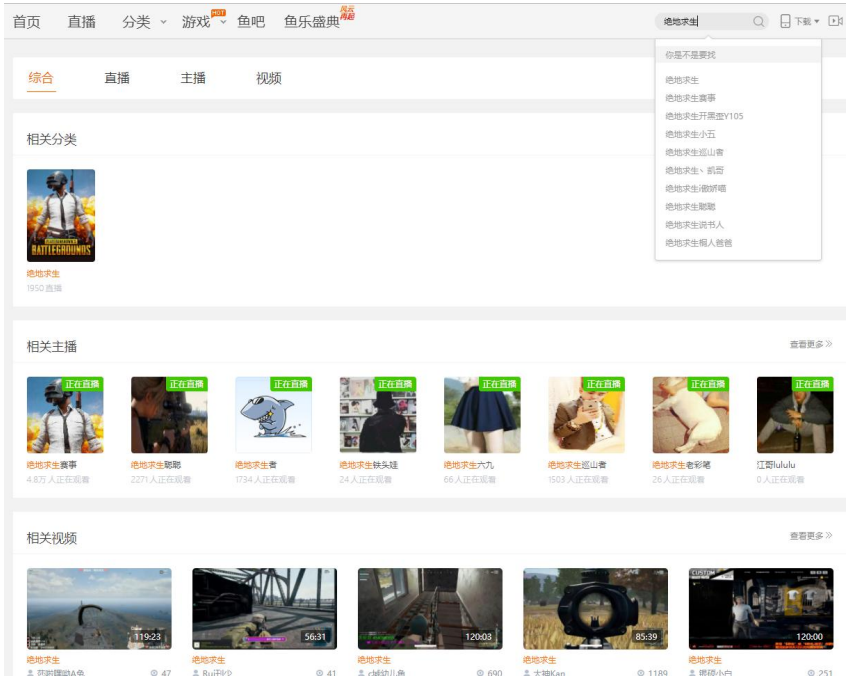
Search

- 基于Elasitcsearch的全文检索系统，也是ES目前重要的应用场景。
- 丰富的查询API，支持RESTful web接口，Java Client API
- 安装搭建方便，配置相对简单
- 分布式存储结构，集群可轻松水平拓展
- 功能全面的分词系统，可满足大部分检索需求
- 倒排索引存储，保证搜索性能要求
- head,cebro等插件齐全
- 社区完善，有问必答

How to use ES?

Search





elastic

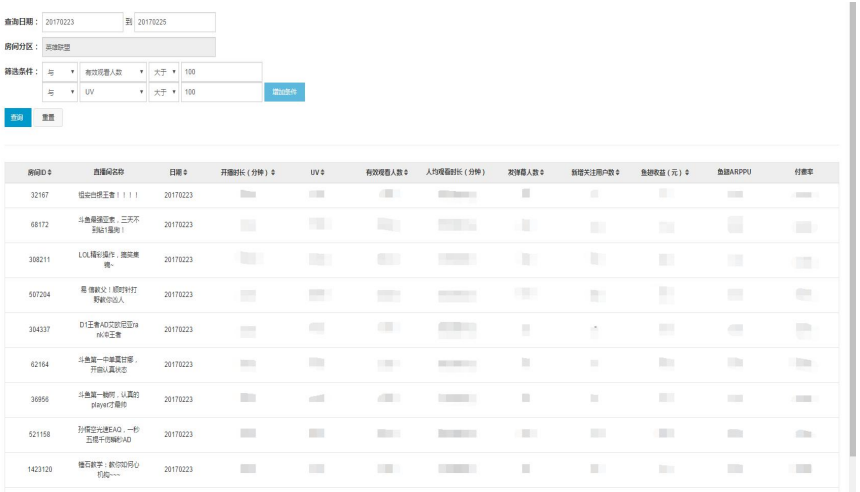
How to use ES?

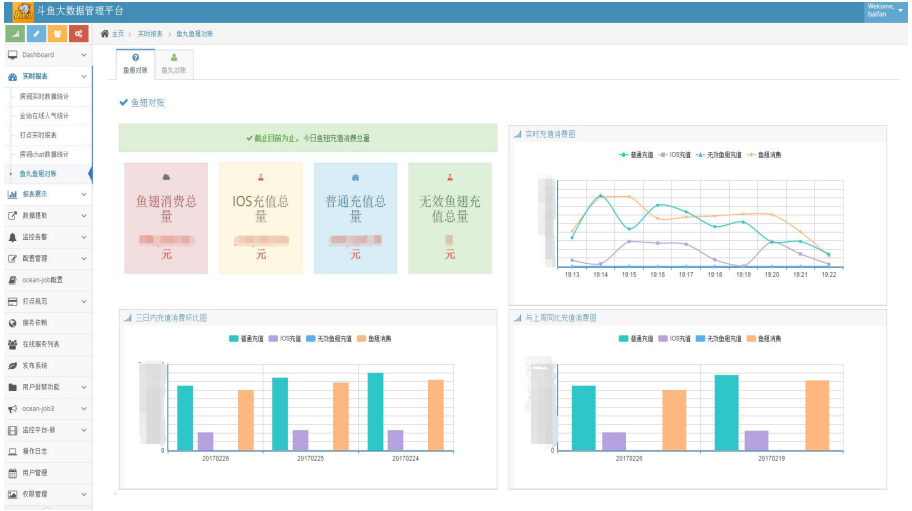
数据可视化

- Elasticsearch具有良好的可读性，足以应付绝大多数的高并发请求，并能保持较好的接口性能
- 查询API丰富，可通过简单逻辑运算，实现多维度高复杂的数据查询
- Kibana具有丰富的Dashboard模板，简单配置，即可实现。也可通过ECharts等插件，自定义Dashboard
- 轻松实现数据的环比，同比，实时计算，趋势图等方案

How to use ES?

数据可视化





elastic

How to use ES?

风控系统

- 风险控制是所有公司系统不可忽视的环节

- 平台用户或咨询师作弊行为监控
- 平台系统对可能的存在风险问题预警
- 平台系统风险出现后损失挽回及进一步扩散控制
- 平台系统风险保留，记录留存

How to use ES?

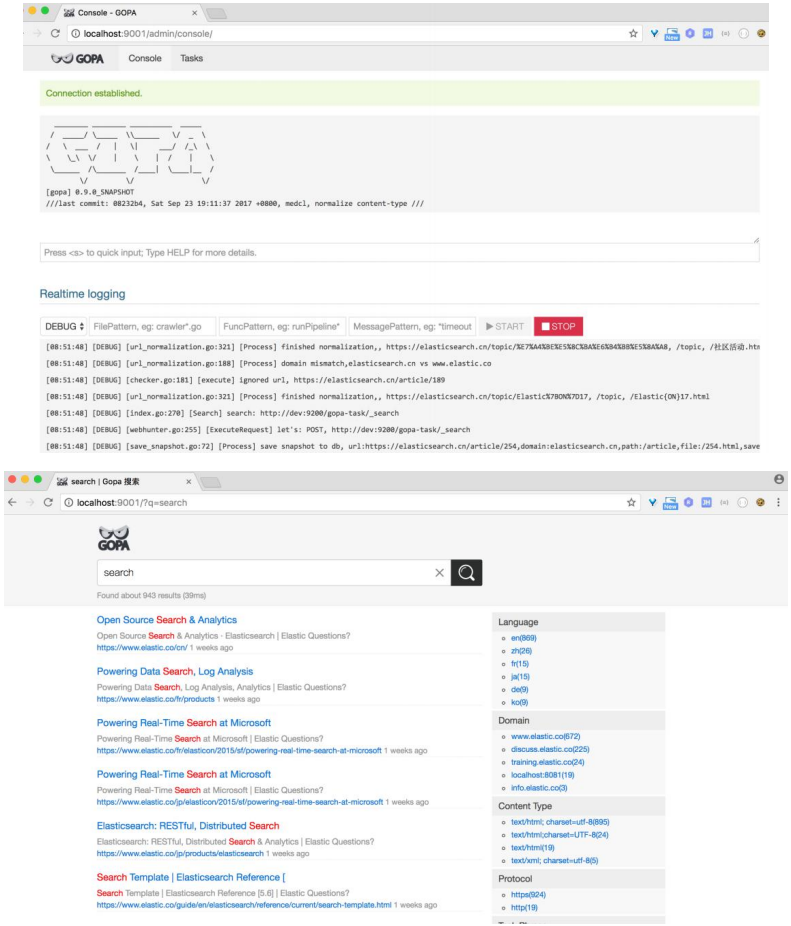
风控系统

elastic

How to use ES?

舆情系统

- 舆情系统数据可采用ES做中间存储查询
- GOPA，相比于Nutch,Scrapy更加轻量级，搭建使用简单，无缝连接ES
- 亦可使用Nutch等爬虫工具，写入ES



elastic

How to use ES?

自定义词库

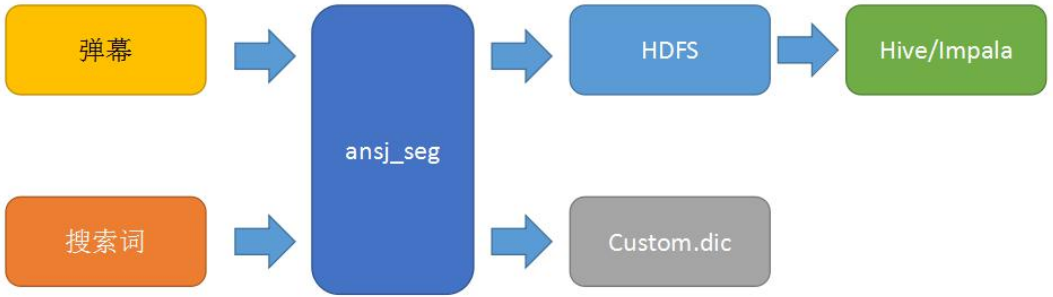
- 生成符合自己生产环境的自定义词库
- 自定义词库可用于全站关键词分析，用户关键词分析以及全站热门词汇分析等
- 生成的自定义词库用于ES分词，提高检索系统可用性
- 用户画像

elastic

How to use ES?

自定义词库

- 采用Ansj中文分词器
- HDFS，离线统计分析
- Coustom.dic，供索引分词



How to use ES?

推荐系统

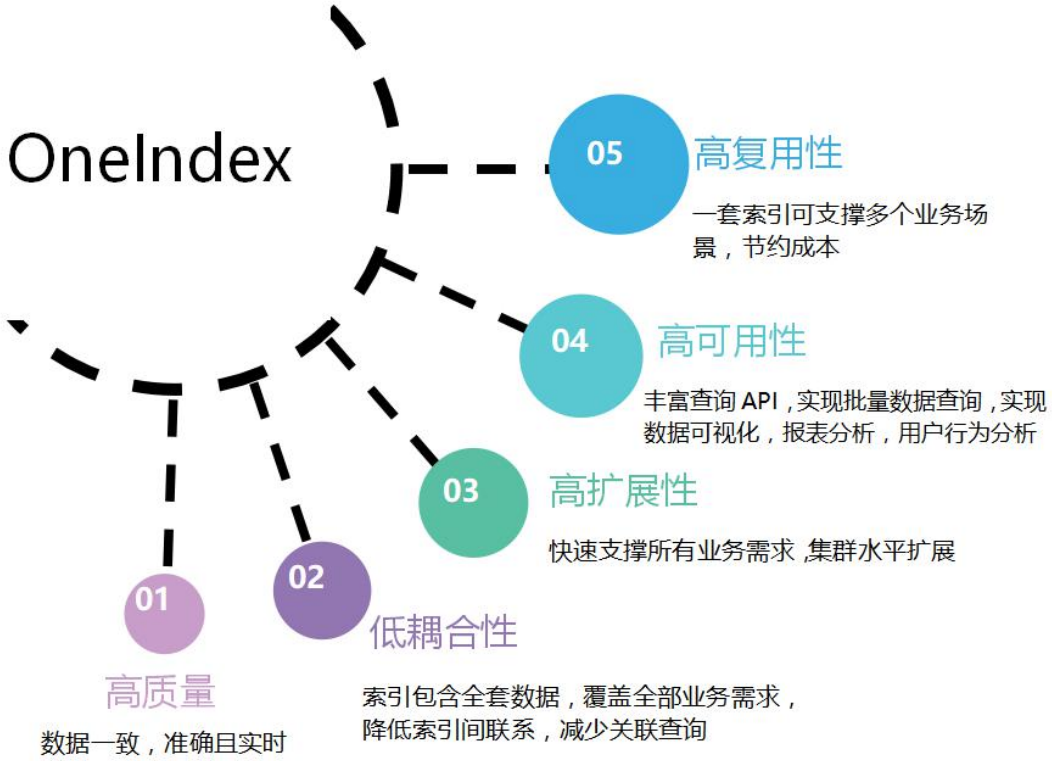
- 制定有效推荐规则
- 用户画像数据写入ES
- 通过ES boost权重配置，可以轻松实现推荐业务
- 无需重写大量逻辑代码
- 高并发环境下，能保证系统的可用性，可读性

elastic

OneIndex?!

OneIndex?!

- 概念来源于阿里“OneData”
- 大数据之痛：
- 高人力成本、数据错误、浪费资源、杂乱无章、效率低下、批量多维度查询问题
- 基于Elasitcsearch的“OneIndex”

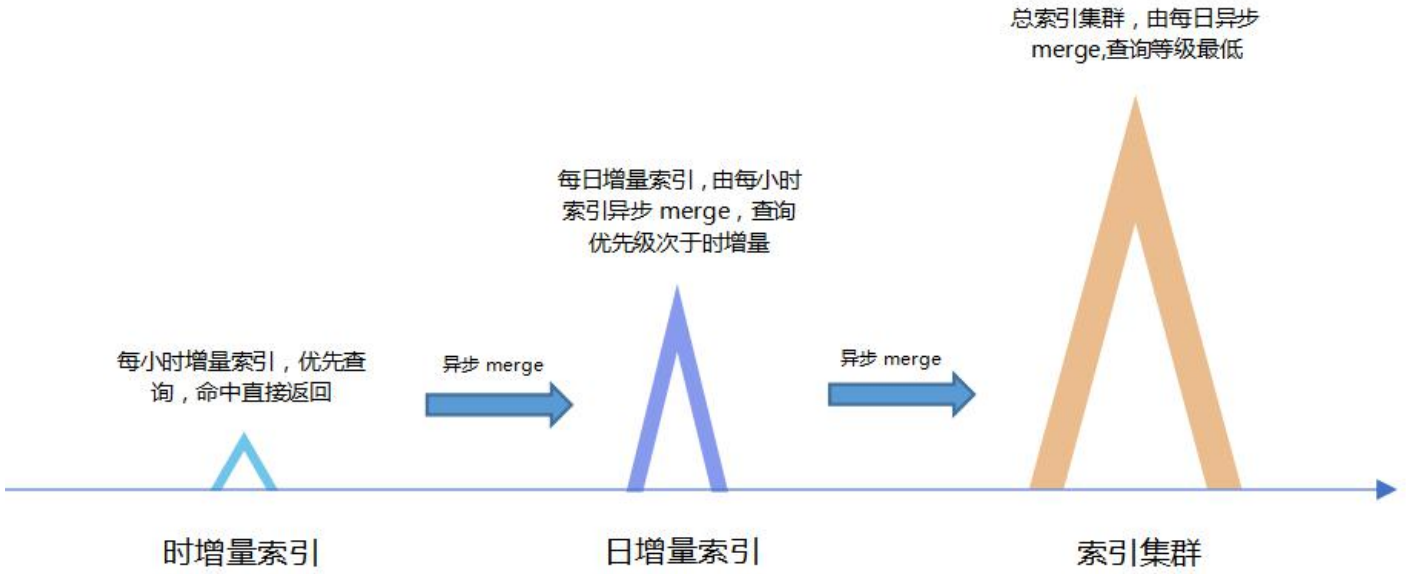


elastic


日志统计监控
系统全文检索
数据可视化
风险控制
个性推荐
...
监控系统
Zabbix
Kibana
API Check
支持微信，
邮件，短信，
电话告警
日志收集系统
（ELK）
服务器运维日志
平台业务日志
接口性能日志
统一数据服务中间件
用户画像索引
订单流水索引
OA 平台索引
风控索引
资产管理索引
舆情索引
业务日志索引
知识图谱索引
...
...
统一规范
数据模型
任务调度平台
数据清洗落地
(HBase)
离线计算(Spark,Hive,Impala)
实时计算(Spark,Strom,Canal)

OneIndex?!

- 海量数据检索方案



elastic

Sunlands ES Roadmap


初具规模
初步搭建 ES 集群
创建用户订单索引
搭建数据可视化平台
全文检索
扩展索引，加强业务覆盖
提供全站全文检索平台
基本数据落地系统
ELK 日志收集
ELK 日志收集系统引入
完善数据清洗落地
基本离线实时计算，监控
OneIndex 初现
完善离线实时计算，监控系统
基本风控，个性推荐
任务调度平台
OneIndex
统一规范，索引全覆盖
日志监控，全文检索，数据可视化，风控，推荐，舆情全方位展开
Step 1
Step 2
Step 3
Step 4
Step 5

Thanks!