


七个生产案例告诉你BATJ为何选择ElasticSearch！ 应用场景和优势！

原创 梵高的左耳 胖滚猪学编程 5月28日

来自专辑
死磕ElasticSearch

 点上方蓝字，将胖滚猪“设为星标，奥利给！

从今天开始，想和你一起死磕ElasticSearch，学习分布式搜索引擎，跟着胖滚猪就对了！

既然是ES的第一课，那么最重要的是让你爱上它！不想说那些单纯的优势、概念了，直接上大厂的生产案例，才是最能吸引你的！跟着大厂走，没问题的！

为啥选择ES？

一个技术服务组件，首先需要了解全面它的使用场景，才能更针对性的去研究及推广。因此第一要务是搞懂为什么要学习ElasticSearch，开头先po一张排行图，大哥的地位可不是瞎搞来的，没点实力能上位？凭这排名就是你要学习它的理由！

| Rank | | | DBMS | Database Model | Score | | |
|----------|----------|----------|------------------------|----------------------------|----------|----------|----------|
| Apr 2020 | Mar 2020 | Apr 2019 | | | Apr 2020 | Mar 2020 | Apr 2019 |
| 1. | 1. | 1. | Oracle + | Relational, Multi-model | 1345.42 | +4.78 | +65.48 |
| 2. | 2. | 2. | MySQL + | Relational, Multi-model | 1268.35 | +8.62 | +53.21 |
| 3. | 3. | 3. | Microsoft SQL Server + | Relational, Multi-model | 1083.43 | -14.43 | +23.47 |
| 4. | 4. | 4. | PostgreSQL + | Relational, Multi-model | 509.86 | -4.06 | +31.14 |
| 5. | 5. | 5. | MongoDB + | Document, Multi-model | 438.43 | +0.82 | +36.45 |
| 6. | 6. | 6. | IBM Db2 + | Relational, Multi-model | 165.63 | +3.07 | -10.42 |
| 7. | 7. | ↑8. | Elasticsearch + | Search engine, Multi-model | 148.91 | -0.26 | +2.91 |
| 8. | 8. | ↓7. | Redis + | Key-value, Multi-model | 144.81 | -2.77 | -1.57 |
| 9. | ↑10. | ↑10. | SQLite + | Relational | 121.92 | -3.22 | -22.73 |
| 10. | ↓9. | ↓9. | Microsoft Access | Relational | | | |

凭啥排这么前呢？不就是个搜索引擎吗。额，也许提到Elasticsearch，你第一反应就是"搜索引擎"。类似百度搜索、淘宝搜索那种。而我写这篇文章就是为了纠正你这个"错误"的观点。

Elasticsearch 确实是做搜索引擎出身的，但是到现在已经进化成了一个全能型的数据产品。因此你的思维决不能限制在搜索引擎上。

本文通过一线大厂的八个案例，全方位让你了解ElasticSearch的应用场景和优势，包括：

- 日志实时分析
- 搜索服务
- 数据分析
- 数据监控
- 查询服务
- 后端存储

ElasticSearch在腾讯的应用

ElasticSearch在腾讯的应用非常广泛，主要有三：日志实时分析场景、搜索服务、时序数据分析。

- 搜索服务：例如像腾讯文档基于 ES 做全文检索，电商客户拼多多、蘑菇街等大量的商品搜索都是基于 ES。
- 日志分析：这个是 ES 应用最广泛的领域，支持全栈的日志分析，包括各种应用日志、数据库日志、用户行为日志、网络数据、安全数据等等。ES 拥有一套完整的日志解决方案，可以秒级实现从采集到展示。
- 时序分析：典型的场景是监控数据分析，比如云监控，整个腾讯云的监控都是基于 ES 的。此外还包括物联网场景，也有大量的时序数据。时序数据的特点是写入吞吐量特别高，ES 支持的同时也提供了丰富的多维统计分析算子。

日志实时分析

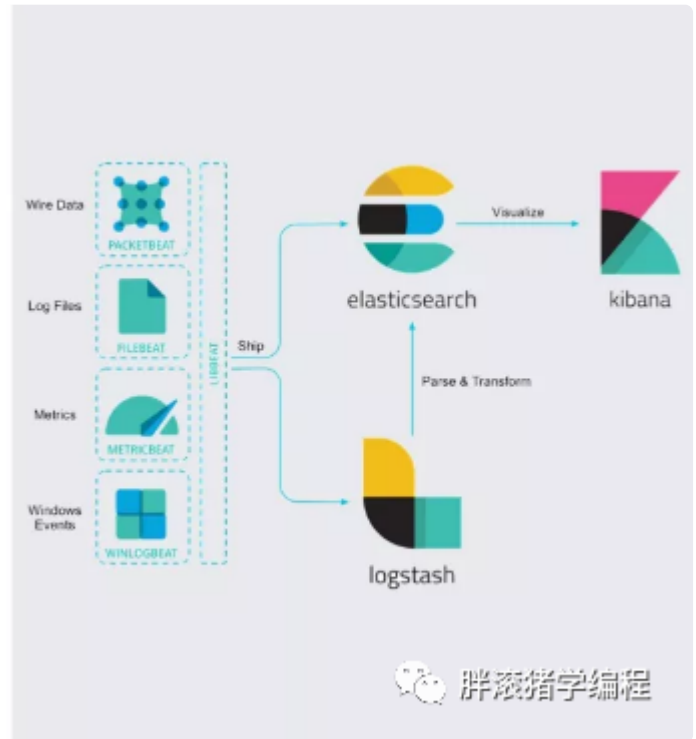
日志实时分析

典型场景

- 业务日志：用户行为
- 运营日志：慢查询、异常日志
- 审计日志
-

主要特点

- 完整的解决方案，易运维
- 实时性：从日志产生到可访问，十秒级
- 交互式分析：万亿级日志，搜索秒级响应
- 全文搜索：基于倒排索引，搜索分析灵活



4

典型日志如下：

- 运营日志，比如慢日志、异常日志，用来定位业务问题；
- 业务日志，比如用户的点击、访问日志，可以用来分析用户行为；
- 审计日志，可以用于安全分析。ES 很完美的解决了日志实时分析的需求，它具有如下特点：

Elastic 生态提供了完整的日志解决方案，任何一个开发、运维同学使用成熟组件，通过简单部署，即可搭建起一个完整的日志实时分析服务。

- 在 Elastic 生态中，日志从产生到可访问一般在 10s 级。相比于传统大数据解决方案的几十分钟、小时级，时效性非常高。ES 拥有一套完整的日志解决方案（ELK），可以秒级实现从采集到展示。
- 由于支持倒排索引、列存储等数据结构，ES 提供非常灵活的搜索分析能力。
- 支持交互式分析，即使在万亿级日志的情况下，ES 搜索响应时间也是秒级。

日志是互联网行业最基础、最广泛的数据形式，ES 非常完美的解决了日志实时分析场景，这也是近几年 ES 快速发展的一个重要原因

搜索服务

搜索服务

典型场景

- 商品搜索
- APP搜索
- 站内搜索
-

主要特点

- 高性能：10w级QPS，20ms级平响
- 强相关：搜索结果高度匹配用户意图
- 高可用：可用性达9999，跨机房容灾



胖猴猪学编程

搜索服务，典型场景包含：商品搜索，类似京东、淘宝、拼多多中的商品搜索；APP 搜索，支持应用商店里的应用搜索；站内搜索，支持论坛、在线文档等搜索功能。我们支持了大量搜索服务，它们主要有以下特点：

- 高性能：单个服务最大达到 10w+ QPS，平响 20ms~，P95 延时小于 100ms。
- 强相关：搜索体验主要取决于搜索结果是否高度匹配用户意图，需要通过正确率、召回率等指标进行评估。
- 高可用：搜索场景通常要求高可用性，支持单机房故障容灾。任何一个电商服务，如淘宝、京东、拼多多，只要故障一个小时就可以上头条。

时序数据分析

时序数据分析

典型场景

- Metrics
- APM
- 物联网数据
-

主要特点

- 高并发写入：1000w/s，600~节点
- 高查询性能：十毫秒级
- 多维分析：灵活、多维度的统计分析



6

时序数据分析，典型的时序数据包含：Metrics，即传统的服务器监控；整个腾讯云的监控都是基于 ES 的。APM，应用性能监控；物联网数据，智能硬件、工业物联网等产生的传感器数据。时序数据的特点是写入吞吐量特别高，ES 支持的同时也提供了丰富的多维统计分析算子。这类场景具有以下特点：

- 高并发写入：线上单集群最大规模达到 600+节点、1000w/s 的写入吞吐。
- 高查询性能：要求单条曲线 或者单个时间线的查询延时在 10ms~。
- 多维分析：要求灵活、多维度的统计分析能力，比如我们在查看监控的时候，可以按照地域、业务模块等灵活的进行统计分析。

上面通过腾讯的案例我们了解了三大应用场景，

- 日志实时分析场景
- 搜索服务
- 时序数据分析

另外从这三大应用场景我们也可以归纳出ES的几大优势：

- 1、具有高可用性、高扩展性；
- 2、查询速度快，性能佳；
- 3、搜索功能强大，高度匹配用户意图。

因此，可以看出，ES在日志实时分析和搜索方面的应用优势简直是无敌的！起码目前，在这两方面，还没有强劲的对手！

ElasticSearch在京东的应用

通过京东的案例，聊一聊ES在查询、检索、数据分析方面的应用场景

由于较高的性能和较低的使用门槛，京东内部有很多的场景都在使用 Elasticsearch。覆盖了京东多条业务线，同时也覆盖了很多应用场景：

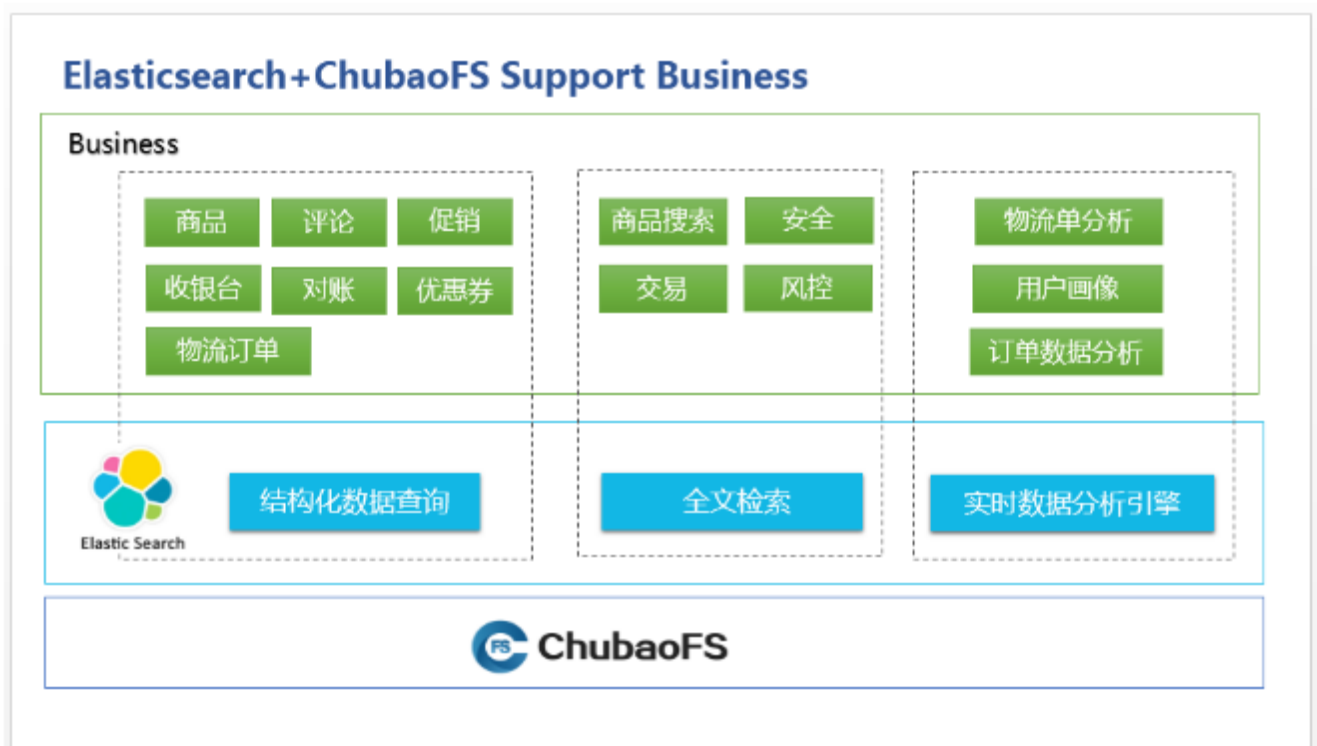


Image1: Elasticsearch +ChubaoFS 支持京东商城应用场景

补充关系型数据库的结构化数据查询

主要应用的业务是商品、促销、优惠券、订单、收银台、物流、对账、评论等大数据量查询。此场景的核心诉求是高性能、稳定性和高可用性，部分场景会有检索要求，通常用于加速关系型数据库，业务系统通过 binlog 同步或业务双写完成数据同步。

全文检索功能

主要的应用场景是应用、安全、风控、交易等操作日志，以及京东部分品类商品搜索。此类日志化场景对写要求很高，查询性能及高可用等要求相对较低，大的业务写会达到数千万 / 秒，存储以 PB 为单位来计算。这些场景对磁盘、内存有比较高的要求，因此，京东也做了相应优化，用于减少内存消耗，提升磁盘整体使用率，使用更廉价的磁盘来降低成本等等。

实时数据分析引擎，形成统计报表

主要应用的业务是物流单的各种分析、订单数据分析、用户画像等。因为业务数据分析纬度较多，flink、storm 等流式分析对于某些报表场景不太适用，批处理实时性又成为问题，所以近实时分析的 Elasticsearch 就成为了这些业务的选择。

从京东的案例中，我们似乎看到了，可以利用ES在某些场景下代替关系型数据库哦！不仅如此，ES在实时数据分析领域，居然也有一席之地！

ElasticSearch在去哪儿的应用

通过去哪儿的案例，聊一聊ES在查询方面的应用场景，可以简单的理解为"代替"mysql。注意代替加了引号，闭着眼睛想都不可能完全代替。比如事务性。

15年去哪儿网酒店日均订单量达到30w+，随着多平台订单的聚合日均订单能达到100w左右。

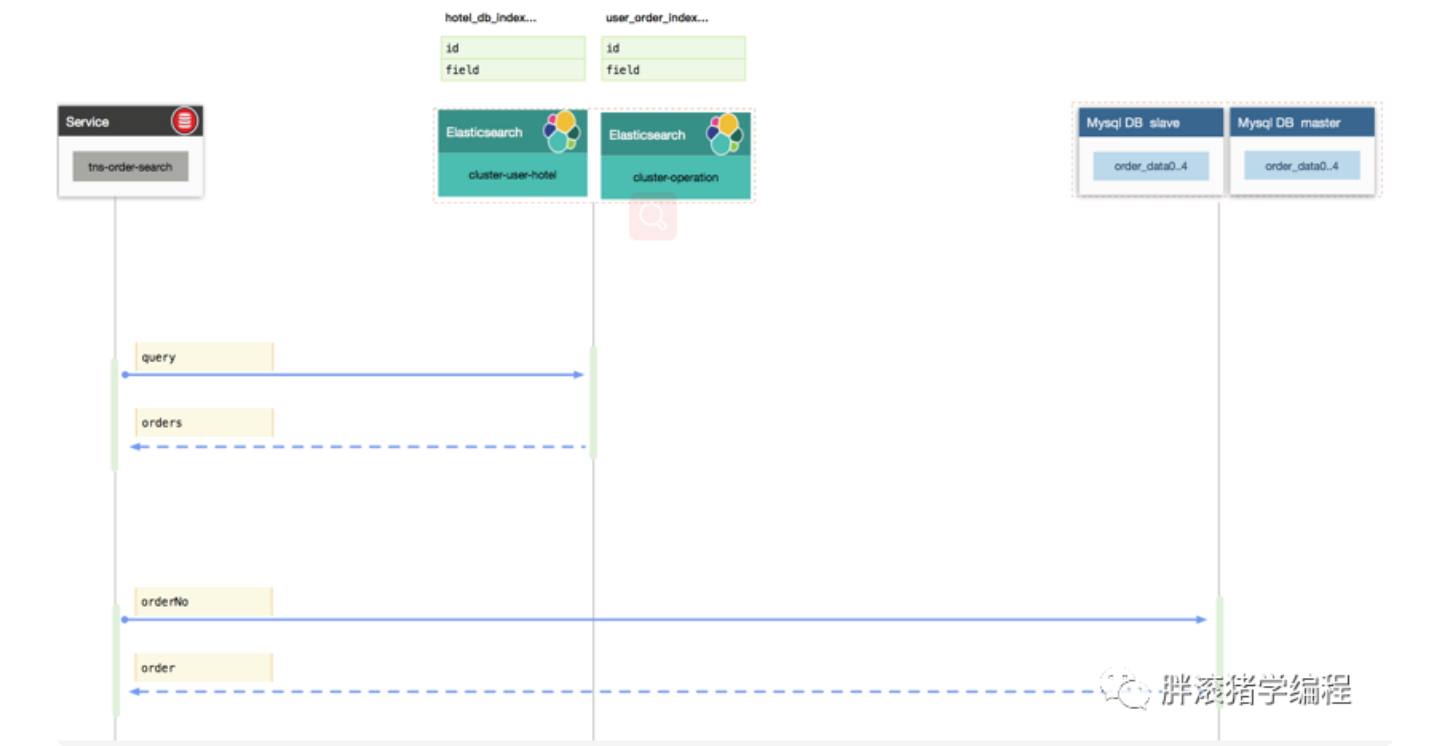
原来采用的热表分库方式，即将最近6个月的订单的放置在一张表中，将历史订单放在在 history表中。history表存储全量的数据，当用户查询的下单时间跨度超过6个月即查询历史订单表，此分表方式热表的数据量为4000w左右，当时能解决的问题。但是显然不能满足携程艺龙订单接入的需求。

如果继续按照热表方式，数据量将超过1亿条。全量数据表保存2年的可能就超过4亿的数据量。所以寻找有效途径解决此问题迫在眉睫。由于对这预计4亿的数据量还需按照预定日期、入住日期、离店日期、订单号、联系人姓名、电话、酒店名称、订单状态.....等多个条件查询。所以简单按照某一个维度进行分表操作没有意义。

显然只通过DB来支撑大量的查询是不可取的，同时对于一些复杂的查询，Mysql支持得不够友好，所以Elasticsearch分布式搜索储存集群的引入，就是为了解决订单数据的存储与搜索的问题。

对订单模型进行抽象和分类，将常用搜索字段和基础属性字段剥离。DB做分库分表，存储订单详情；Elasticsearch存储搜索字段。

订单复杂查询直接走Elasticsearch，基于OrderNo的简单查询走DB，如下图所示。



从去哪儿的案例中，我们似乎看到了，关系型数据库撑不起的复杂查询，ES可以胜任。

总结

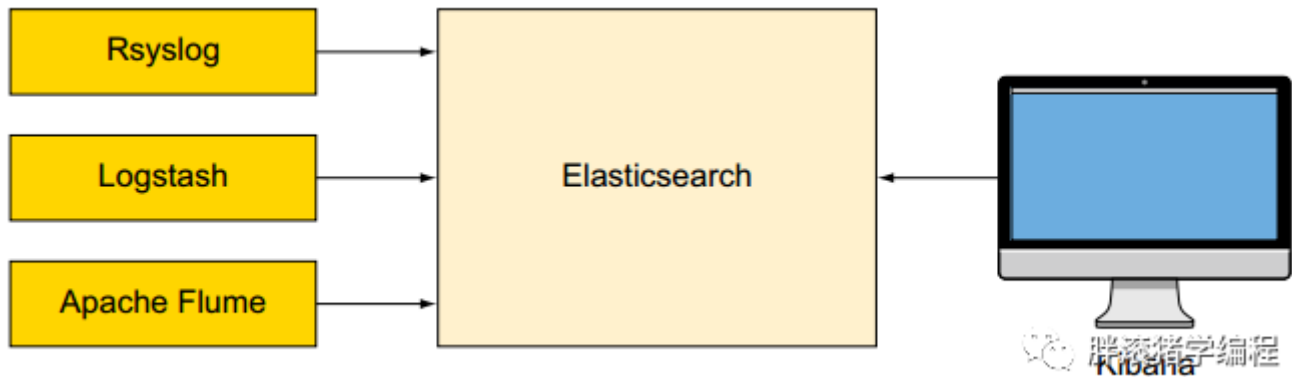
什么时候应该用ElasticSearch？

- 1、典型搜索场景：闭着眼用它！
- 2、典型日志分析场景：闭着眼用它！
- 3、关系型数据库查询有瓶颈：考虑下用它！为啥是考虑？ES的优点在于查询，然而实践证明，在被作为数据库来使用，即写完马上查询会有延迟。
- 4、数据分析场景：考虑下用它！为啥是考虑？简单通用的场景需求可以大规模使用，但在特定业务场景领域，还是要选择更加专业的数据产品，如复杂聚合，ClickHouse相比Elasticsearch 做亿级别数据深度聚合需求会更加合适。

ElasticSearch有什么优势呢？

- 1、很简便的横向扩容，分布式的架构，可以轻松地对资源进行横向纵向扩缩容，可以满足不同数据量级及查询场景对硬件资源的需求。能由数百台到万台机器搭建满足PB级的快速搜索，也能搭建单机版服务小公司。
- 2、查询速度快：ES底层采用Lucene作为搜索引擎，并在此之上做了多重优化，保证了用户对数据查询数据的需求。可"代替"传统关系型数据库，也可用于复杂数据分析，海量数据的近实时处理等。
- 3、相关性高：ES内部提供了完善的评分机制，会根据分词出现的频次等信息对文档进行相关性排序，保证相关性越高的文档排序越靠前。另外还提供了包括模糊查询，前缀查询，通配符查询等在内的多种查询手段，帮助用户快速高效地进行检索。
- 4、功能点多但使用比较简便，开箱即用，性能优化比较简单
- 5、生态圈丰富，社区活跃，适配多种工具。如下图，处理日志和输出到Elasticsearch，您可以使用日志记录工具，如Logstash（www.elastic.co/products/logstash），搜索和可视化界

面分析这些日志，你可以使用Kibana（www.elastic.co/产品/kibana），即传说中的ELK技术栈。另外当前主流的大数据框架也几乎都支持ES，比如Flink和ES就是个完美搭档。



本文参考：

[腾讯万亿级 Elasticsearch 技术解密](#)

[Elasticsearch在京东的使用场景](#)

END

点击查看往期内容回顾

- 【漫画】ES必知必会的倒排索引和分词
- Kafka消息分区机制的原理及分区策略
- 面试官：说出八种消息队列的应用场景

原创不易，养成习惯，点个在看！

