

# [VIP] ElasticSearch基本原理

做技术人的指路明灯，职场生涯的精神导师！

➤ Tom老师QQ号：441221062



01

# Elastic Stack生态

# 大数据分析利器





# 快!快!快!



## 第一快：学习门槛低，开发周期短，上线快

Hadoop

30 Day

Elastic Stack

7 Day

## 出活快！靠谱！



第二快：性能好，查询快，实时展示结果

**对T + 1 说不！**

现在，立刻马上就要看到数据



## 第三快：迅速扩容，快速支撑增长迅猛的数据量

GB



TB/PB



# Elastic Stack之ELK

## ElasticSearch

### Java语言编写

实时的分布式搜索和分析引擎，它可以用于全文搜索，结构化搜索以及分析。

## Logstash

### JRuby语言编写

是一个具有实时渠道能力的**数据收集**引擎，包含输入、过滤、输出模块，一般在过滤模块中做日志格式化的解析工作。

## Kibana

### JavaScript编写

为ElasticSearch提供分析和**可视化**的Web平台。  
它可以ElasticSearch的索引中查找，交互数据，并生成各种维度的表图。





# 02

## ElasticSearch的兴起

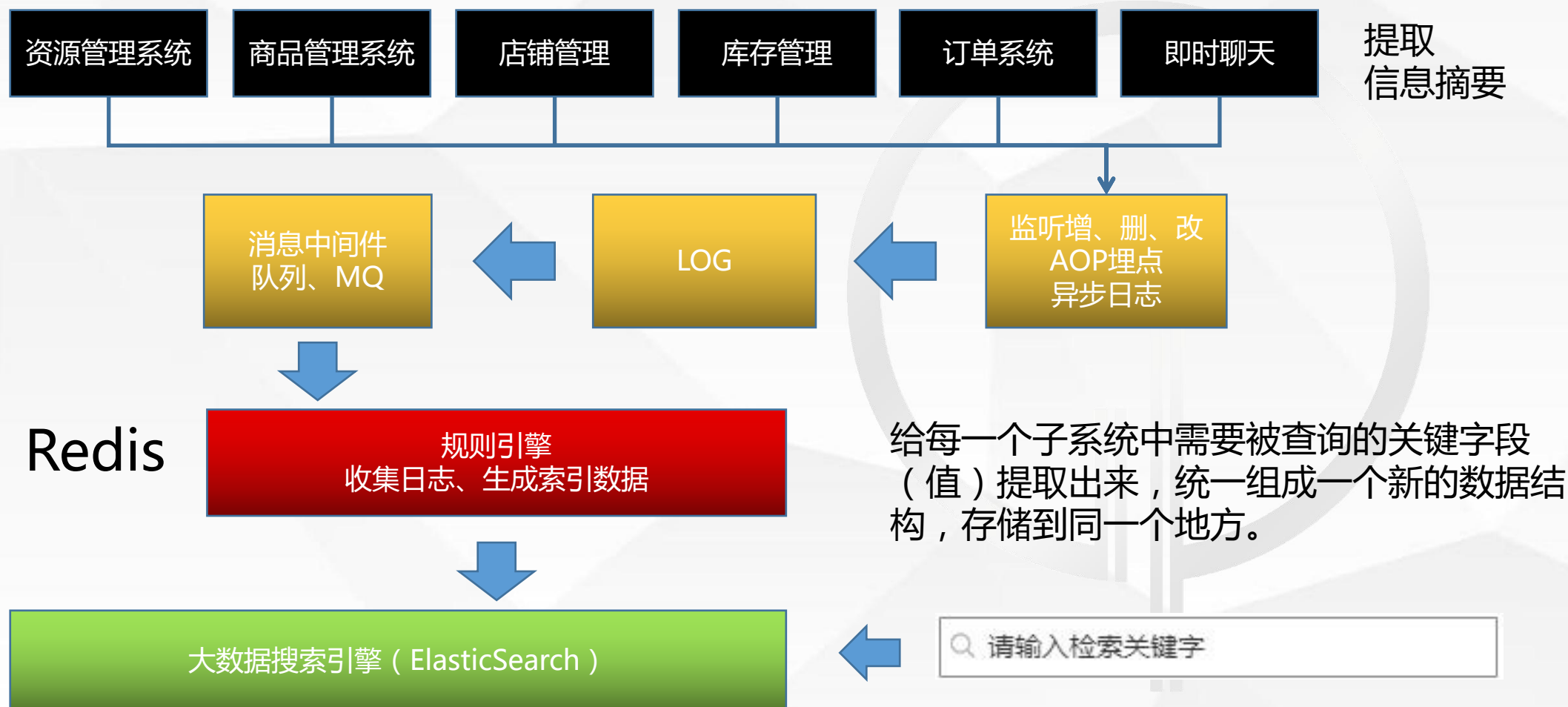
Search is something that any application should have !

-- Shay Banon ( 沙伊巴农 )  
Creator of Elasticsearch





# 电商平台的分布式聚合搜索引擎架构



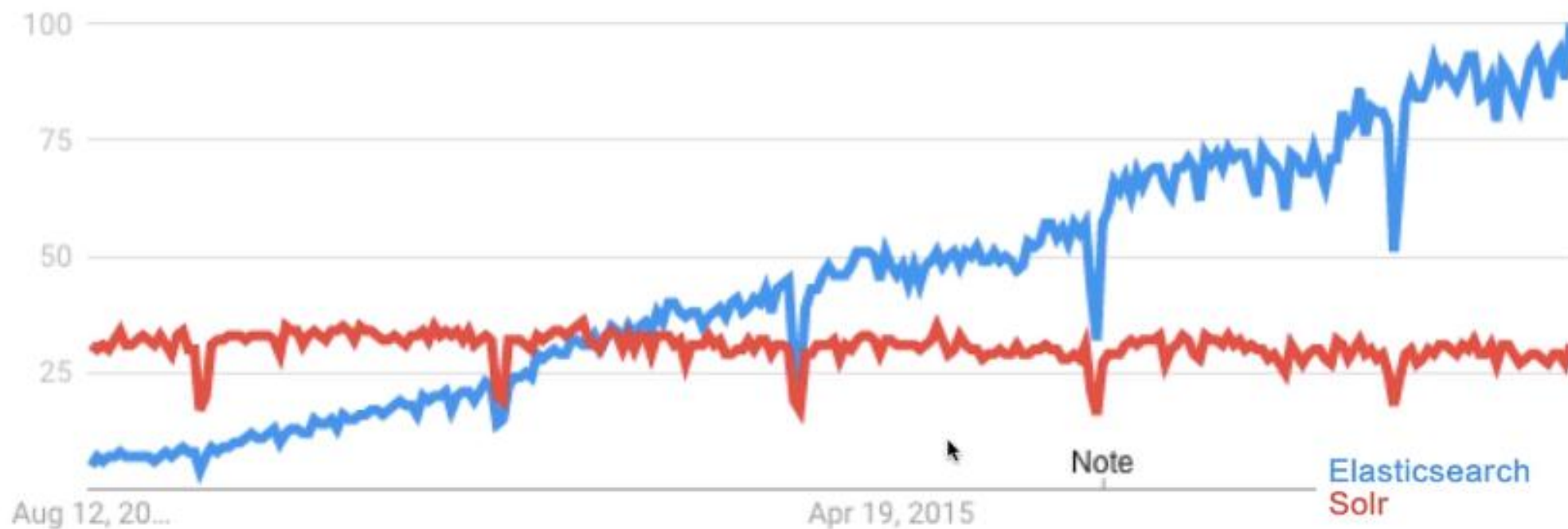
给每一个子系统中需要被查询的关键字段（值）提取出来，统一组成一个新的数据结构，存储到同一个地方。

分布式搜索引擎，基于Lucene进行开发，同类产品还有Solr

# 02

## 为什么要选择ElasticSearch

# 用数据说话





# 用数据说话



# 用数据说话



阿里巴巴



Tencent 腾讯



滴滴

滴滴一下 美好出行



ctrip

携程

平安银行

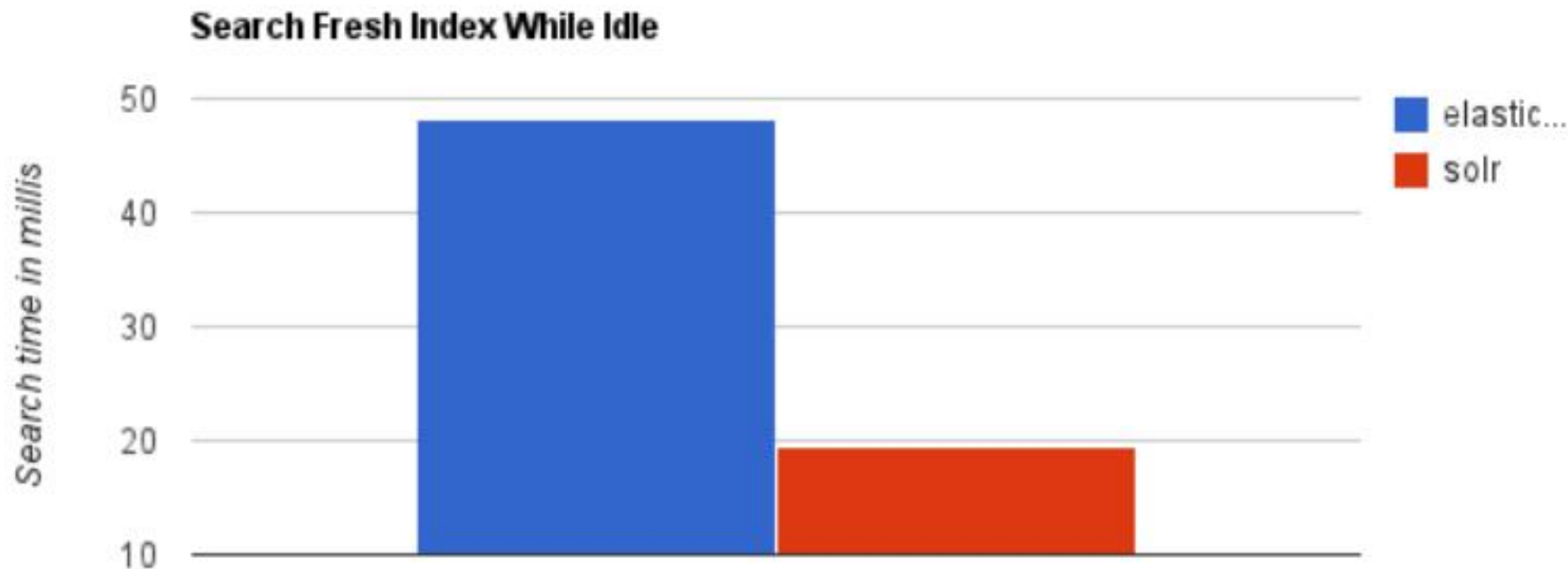
PINGAN BANK

ofo



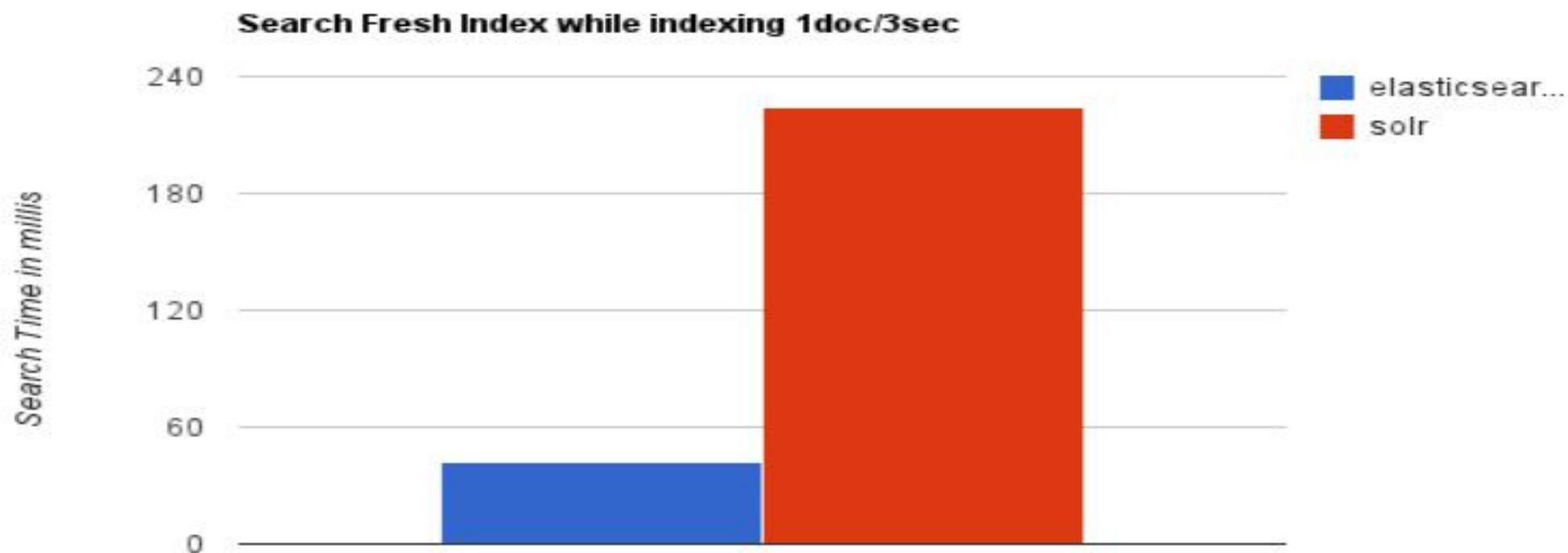
# ElasticSearch和Solr运行数据对比

当单纯的对已有数据进行搜索时，Solr更快。



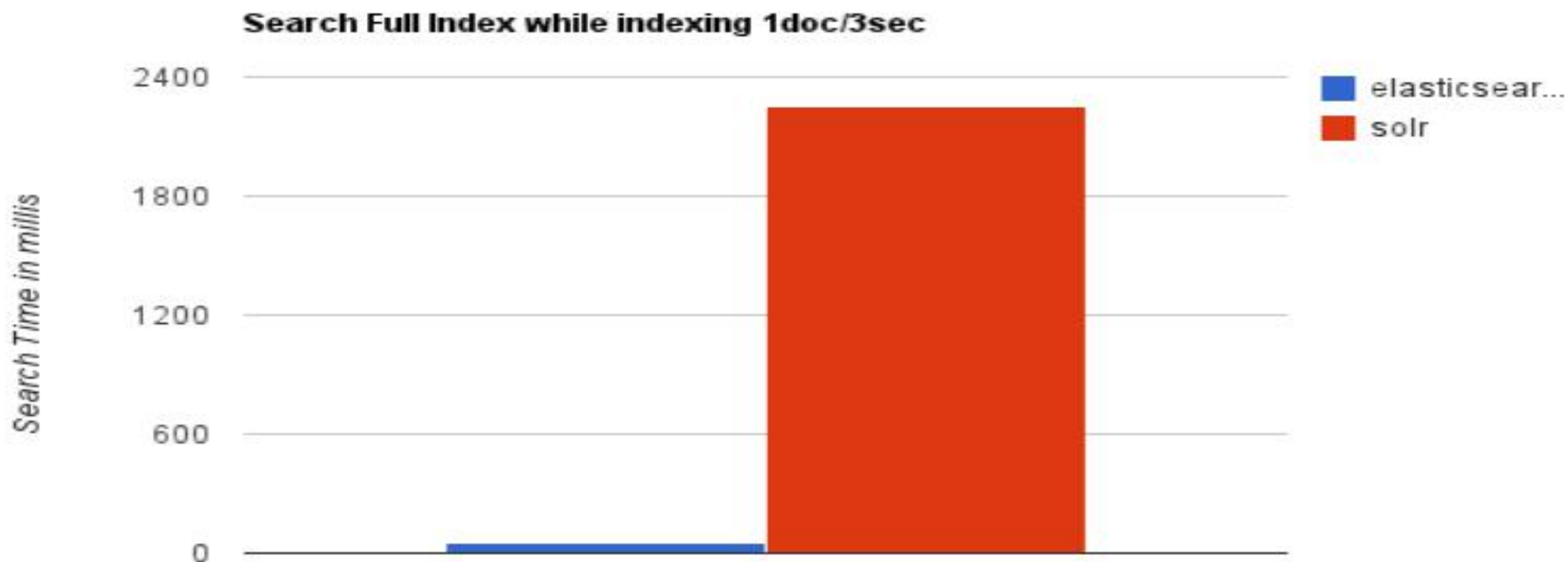
# ElasticSearch和Solr运行数据对比

当实时建立索引时, Solr会产生io阻塞, 查询性能较差, Elasticsearch具有明显的优势



# ElasticSearch和Solr运行数据对比

随着数据量的增加，Solr的搜索效率会变得更低，而Elasticsearch却没有明显的变化。





# ElasticSearch和Solr运行数据对比

综上所述，Solr的架构不适合**实时**搜索的应用。

## 实际生产环境测试

下图为将搜索引擎从Solr转到Elasticsearch以后的平均查询速度有了50倍的提升。



03

## ElasticSearch学习方法论

## 类比学习法

利用已有的知识和新的知识多维度对比  
产生差异化的结果，达到加深印象的目的



- 1、Lucene 是一个 JAVA 搜索类库，它本身并不是一个完整的解决方案，需要额外的开发工作。
- 2、Document文档存储、文本搜索。
- 3、Index索引，聚合检索。
- 4、Analyzer分词器，如IKAnalyzer、word分词、Ansj、Stanford等
- 5、大数据搜索引擎解决方案原理
- 6、NoSQL的兴起（Redis、MongoDB、Memecache）





## 处理文本的最高效做法就是：正则匹配





# 关于NoSQL数据库

No SQL  
内存数据库

Redis  
MongoDB  
Memecache  
HBase  
...

Relational DB  
关系型数据库

MySQL  
Oracle  
SQL Server  
Access  
...



# 关系型数据库和ES操作姿势对比

关系型数据库

建库(DB)  
建表(Table)  
键约束(Constraint)

ElasticSearch

建库 ( Index )  
建表 ( Index Type )  
主键(ID)



# 关系型数据库和ES操作姿势对比

## JDBC操作

- 1、加载驱动类(JDBC驱动)
- 2、建立连接 ( Connection )
- 3、创建语句集 ( Statement )
- 4、执行语句集 execute()
- 5、获取结果集 ( ResultSet )
- 6、关闭结果、语句、连接

## ES Client操作

----

- 1、建立连接 ( TransportClient )
- 2、条件构造 ( SearchRequestBuilder )
- 3、执行语句 execute()
- 4、获取结果 ( SearchResponse )
- 5、关闭以上操作





# 谢谢观看

咕泡学院，只为更好的你！

➤ Tom老师QQ号：441221062

