

关注：Oracle数据库、Oracle数据库集群、Weblogic中间件、Weblogic中间件集群、软件自动化发布、软件持续集成、软件配置管理、Linux操作系统

昵称：心随梦飞[周路敏]
园龄：5年10个月
粉丝：26
关注：0
[+加关注](#)

< 2017年8月 >						
日	一	二	三	四	五	六
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

搜索

找找看

谷歌搜索

常用链接

我的随笔
我的评论
我的参与
最新评论
我的标签
[更多链接](#)

文章分类(56)

Delphi(1)
jenkins(21)
Maven(21)
Oracle(9)
SQL Server(1)
SVN(1)
测试(1)
缓存
平台搭建
前端开发(1)

文章档案(59)

2015年7月 (1)
2014年9月 (1)
2014年8月 (7)
2014年7月 (37)
2012年9月 (1)
2012年8月 (9)
2012年3月 (1)
2011年11月 (1)
2011年10月 (1)

C语言相关链接

万一博客---->C语言基础

delphi相关链接

万一博客---->JSON
万一博客---->Windows编程
万一博客---->Windows消息
万一博客---->多线程
万一博客---->钩子相关函数
万一博客---->接口相关
万一博客---->类相关
万一博客---->容器类
万一博客---->使用泛型



搜索引擎选择： [Elasticsearch](#)与[Solr](#)

搜索引擎选择： [Elasticsearch](#)与[Solr](#)

搜索引擎选型调研文档

Elasticsearch简介*

Elasticsearch是一个实时的分布式搜索和分析引擎。它可以帮助你前所未有的速度去处理大规模数据。

它可以用于全文搜索，结构化搜索以及分析，当然你也可以将这三者进行组合。

Elasticsearch是一个建立在全文搜索引擎 Apache Lucene™ 基础上的搜索引擎，可以说Lucene是当今最先进，最高效的全功能开源搜索引擎框架。

但是Lucene只是一个框架，要充分利用它的功能，需要使用JAVA，并且在程序中集成Lucene。需要很多的学习了解，才能明白它是如何运行的，Lucene确实非常复杂。

Elasticsearch使用Lucene作为内部引擎，但是在使用它做全文搜索时，只需要使用统一开发好的API即可，而不需要了解其背后复杂的Lucene的运行原理。

当然Elasticsearch并不仅仅是Lucene这么简单，它不但包括了全文搜索功能，还可以进行以下工作：

- 分布式实时文件存储，并将每一个字段都编入索引，使其可以被搜索。
- 实时分析的分布式搜索引擎。
- 可以扩展到上百台服务器，处理PB级别的结构化或非结构化数据。

这么多的功能被集成到一台服务器上，你可以轻松地通过客户端或者任何你喜欢的程序语言与ES的RESTful API进行交流。

Elasticsearch的上手是非常简单的。它附带了很多非常合理的默认值，这让初学者很好地避免一上手就要面对复杂的理论，

它安装好了就可以使用了，用很小的学习成本就可以变得很有生产力。

随着越学越深入，还可以利用Elasticsearch更多高级的功能，整个引擎可以很灵活地进行配置。可以根据自身需求来定制属于自己的Elasticsearch。

使用案例：

- 维基百科使用Elasticsearch来进行全文搜做并高亮显示关键词，以及提供search

万一博客---->数组相关
万一博客---->正则表达式
万一博客---->资源文件

JavaWeb相关链接

PHP相关链接

PHP100视频教程

工具软件相关链接

快递

快递100
快递查询

数据库相关链接

海南.胡勇博客---Oracle
海南.胡勇博客

英语相关链接

VOA慢速英语视频
VOA美国之音
VOA美国之音慢速英语网
VOA英语网
新版新概念英语1-4册

积分与排名

积分 - 22863
排名 - 13780

最新评论

1. Re:搜索引擎选择： Elasticsearch与Solr
夜行侠老师视频教学：es问题汇总索的精准性控制3、解决搜索中英文数字混合4、搜索中in查询5、安装问题6、es源码内部是如何做到插件化扩展7、统计时展示不全8、搜索条件过多，内容过长报错9、索引某一字.....

--smart152819

2. Re:搜索引擎选择： Elasticsearch与Solr
博主，能提供下关于solr与es的数据分析对比报告么

--易连山

3. Re:搜索引擎选择： Elasticsearch与Solr
《浅入深出ElasticSearch构建高性能搜索架构》课程学习地址：课程出自学途无忧网；讲师：夜行侠老师一、课程用到的软件ElasticSearch5.0.0Spring Tool Suite 3.....

--哈雷兄弟

4. Re:搜索引擎选择： Elasticsearch与Solr
mm

--禅道

5. Re:搜索引擎选择： Elasticsearch与Solr
介绍的好！

--坦荡

6. Re:搜索引擎选择： Elasticsearch与Solr
浅入深出ElasticSearch构建高性能搜索架构
课程学习地址：

--无名小卒的苦

7. Re:搜索引擎选择： Elasticsearch与Solr
@napoay英文原文在这里...

--杨Coder

8. Re:搜索引擎选择： Elasticsearch与Solr
博主，博客中solr和es性能对比的数据出处能告知一下吗？

--napoay

9. Re:搜索引擎选择： Elasticsearch与Solr
能详细说一下 "随着数据量的增加，Solr的

h-as-you-type、did-you-mean等搜索建议功能。

- 英国卫报使用Elasticsearch来处理访客日志，以便能将公众对不同文章的反应实时地反馈给各位编辑。
- StackOverflow将全文搜索与地理位置和相关信息进行结合，以提供more-like-this相关问题的展现。

- GitHub使用Elasticsearch来检索超过1300亿行代码。

- 每天，Goldman Sachs使用它来处理5TB数据的索引，还有很多投行使用它来分析股票市场的变动。

但是Elasticsearch并不只是面向大型企业的，它还帮助了很多类似DataDog以及Klout的创业公司进行了功能的扩展。

Elasticsearch的优缺点**：

优点

- Elasticsearch是分布式的。不需要其他组件，分发是实时的，被叫做” Push replication”。
- Elasticsearch 完全支持 Apache Lucene 的接近实时的搜索。
- 处理多租户（multitenancy）不需要特殊配置，而Solr则需要更多的高级设置。
- Elasticsearch 采用 Gateway 的概念，使得完备份更加简单。
- 各节点组成对等的网络结构，某些节点出现故障时会自动分配其他节点代替其进行工作。

缺点

- 只有一名开发者（当前Elasticsearch GitHub组织已经不只如此，已经有了相当活跃的维护者）
- 还不够自动（不适合当前新的Index Warmup API）

Solr简介*

Solr（读作“solar”）是Apache Lucene项目的开源企业搜索平台。其主要功能包括全文检索、命中标示、分面搜索、动态聚类、数据库集成，以及富文本（如Word、PDF）的处理。Solr是高度可扩展的，并提供了分布式搜索和索引复制。Solr是最流行的企业级搜索引擎，Solr4 还增加了NoSQL支持。

Solr是用Java编写、运行在Servlet容器（如 Apache Tomcat 或Jetty）的一个独立的全文搜索服务器。 Solr采用了 Lucene Java 搜索库为核心的全文索引和搜索，并具有类似REST的HTTP/XML和JSON的API。Solr强大的外部配置功能使得无需进行Java编码，便可对 其进行调整以适应多种类型的应用程序。Solr有一个插件架构，以支持更多的高级定制。

因为2010年 Apache Lucene 和 Apache Solr 项目合并，两个项目是由同一个Apache软件基金会开发团队制作实现的。提到技术或产品时，Lucene/Solr或

搜索效率会变得更低，而Elasticsearch却没有明显的变化。" 这里的数据量是什么数量级么？多谢。

--光阴四溅

10. Re:搜索引擎选择： Elasticsearch与 Solr

@王安琪没配错，那是时间。...

--skywalker0011

11. Re:搜索引擎选择： Elasticsearch与 Solr

我写的Ubuntu下ES集群部署博客SEO太差了，百度必应谷歌都搜不到，来推广下方便初学者搜到(*^__^*)Ubuntu下Elasticsearch 2.1集群部署过程与遇到的问题及解决方法（开机自.....

--张冬

12. Re:第十四章：详解Jenkins节点配置 Jenkins分布式配置和java或者用c#有关系吗？我开发用的.net平台，不大懂的配置。。

--Hedy徐

13. Re:搜索引擎选择： Elasticsearch与 Solr

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。这段话的配图错了。

--王安琪

Solr/Lucene是一样的。

Solr的优缺点

优点

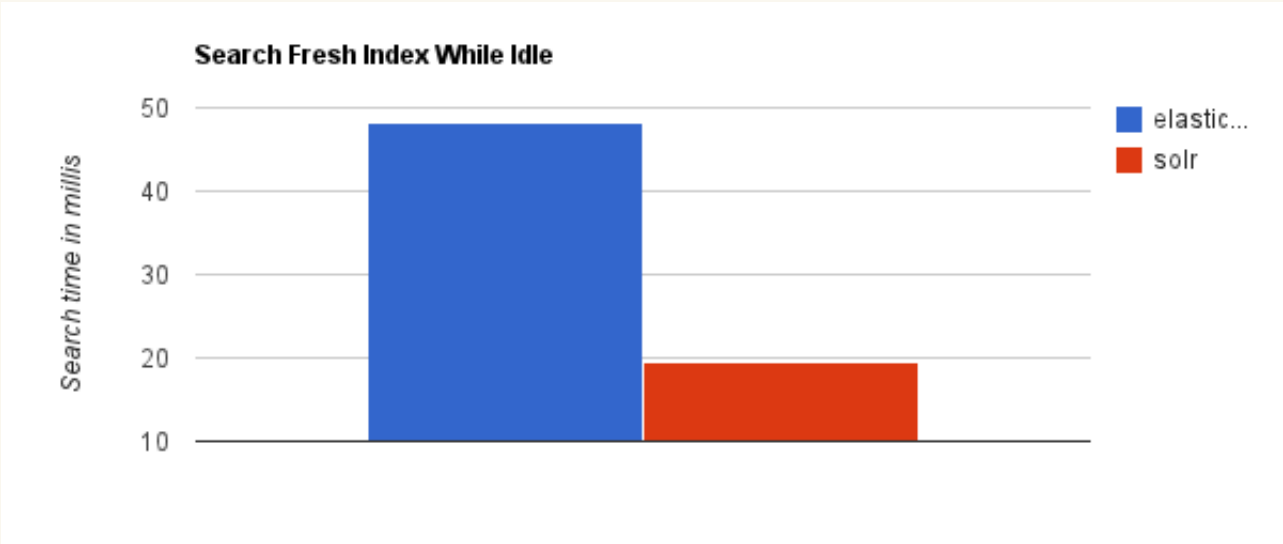
1. Solr有一个更大、更成熟的用户、开发和贡献者社区。
2. 支持添加多种格式的索引，如：HTML、PDF、微软 Office 系列软件格式以及JSON、XML、CSV 等纯文本格式。
3. Solr比较成熟、稳定。
4. 不考虑建索引的同时进行搜索，速度更快。

缺点

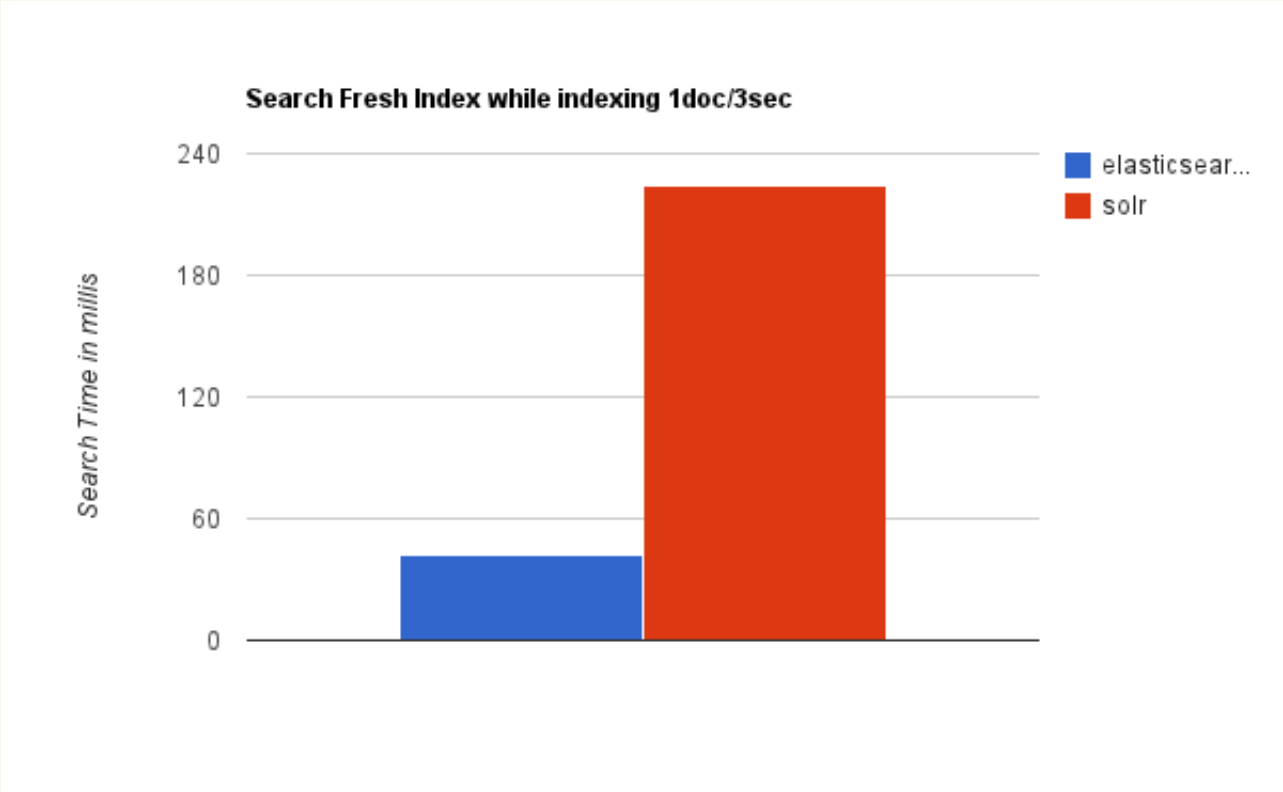
1. 建立索引时，搜索效率下降，实时索引搜索效率不高。

Elasticsearch与Solr的比较*

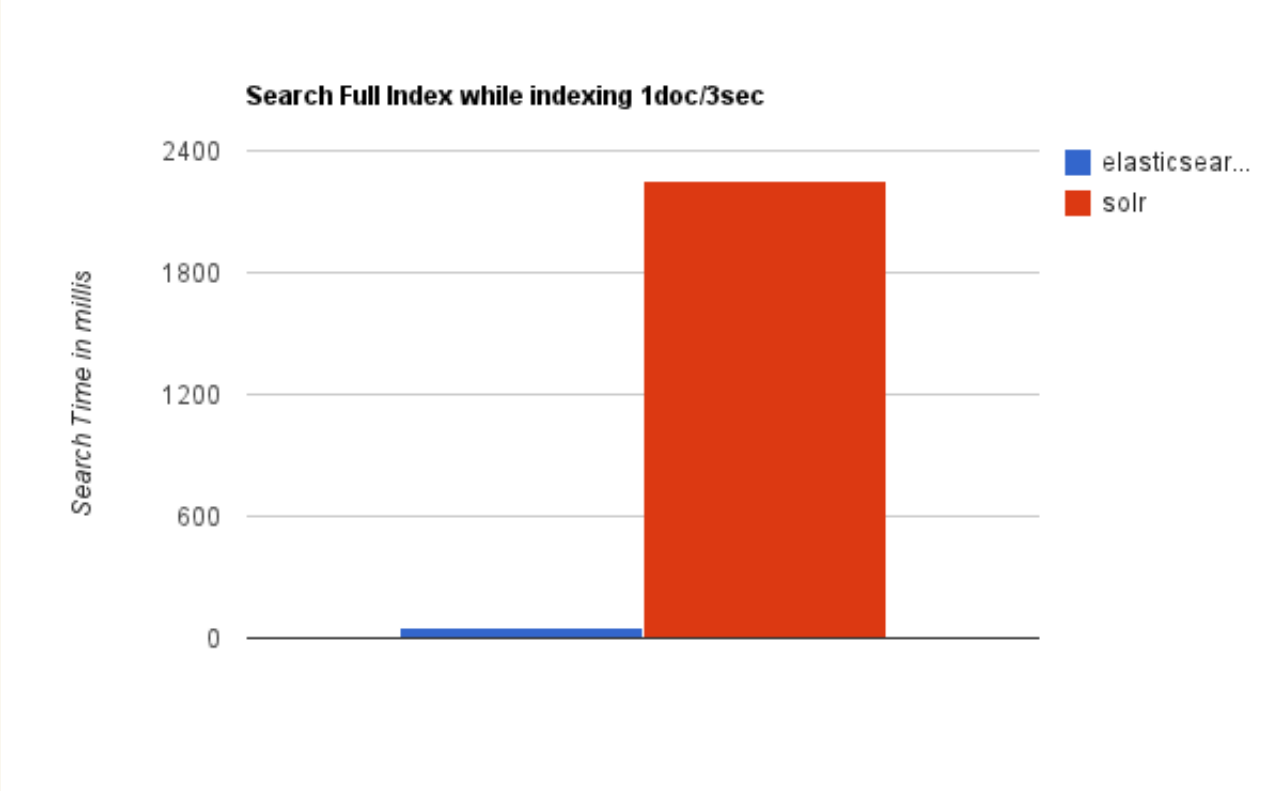
当单纯的对已有数据进行搜索时，Solr更快。



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。



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。



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。



Elasticsearch 与 Solr 的比较总结

- 二者安装都很简单；
- Solr 利用 Zookeeper 进行分布式管理，而 Elasticsearch 自身带有分布式协调管理功能；
- Solr 支持更多格式的数据，而 Elasticsearch 仅支持json文件格式；
- Solr 官方提供的功能更多，而 Elasticsearch 本身更侧重于核心功能，高级功能多有第三方插件提供；
- Solr 在传统的搜索应用中表现好于 Elasticsearch，但在处理实时搜索应用时效率明显低于 Elasticsearch。

Solr 是传统搜索应用的有力解决方案，但 Elasticsearch 更适用于新兴的实时搜索应用。

其他基于Lucene的开源搜索引擎解决方案*

1. 直接使用 [Lucene](#)

说明：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。

优点：成熟的解决方案，有很多的成功案例。apache 顶级项目，正在持续快速的

进步。庞大而活跃的开发社区，大量的开发人员。它只是一个类库，有足够的定制和优化空间：经过简单定制，就可以满足绝大部分常见的需求；经过优化，可以支持 10亿+ 量级的搜索。

缺点：需要额外的开发工作。所有的扩展，分布式，可靠性等都需要自己实现；非实时，从建索引到可以搜索中间有一个时间延迟，而当前的“近实时” (Lucene Near Real Time search)搜索方案的可扩展性有待进一步完善

- [Katta](#)

说明：基于 Lucene 的，支持分布式，可扩展，具有容错功能，准实时的搜索方案。

优点：开箱即用，可以与 Hadoop 配合实现分布式。具备扩展和容错机制。

缺点：只是搜索方案，建索引部分还是需要自己实现。在搜索功能上，只实现了最基本的需求。成功案例较少，项目的成熟度稍微差一些。因为需要支持分布式，对于一些复杂的查询需求，定制的难度会比较大。

- [Hadoop contrib/index](#)

说明：Map/Reduce 模式的，分布式建索引方案，可以跟 Katta 配合使用。

优点：分布式建索引，具备可扩展性。

缺点：只是建索引方案，不包括搜索实现。工作在批处理模式，对实时搜索的支持不佳。

- [LinkedIn 的开源方案](#)

说明：基于 Lucene 的一系列解决方案，包括 准实时搜索 zoie ， facet 搜索实现 bobo ，机器学习算法 decomposer ，摘要存储库 krati ，数据库模式包装 sensei 等等

优点：经过验证的解决方案，支持分布式，可扩展，丰富的功能实现

缺点：与 linkedin 公司的联系太紧密，可定制性比较差

- [Lucandra](#)

说明：基于 Lucene，索引存在 cassandra 数据库中

优点：参考 cassandra 的优点

缺点：参考 cassandra 的缺点。另外，这只是一个 demo，没有经过大量验证

- [HBasene](#)

说明：基于 Lucene，索引存在 HBase 数据库中

优点：参考 HBase 的优点

缺点：参考 HBase 的缺点。另外，在实现中，lucene terms 是存成行，但每个 term 对应的 posting lists 是以列的方式存储的。随着单个 term 的 posting

lists 的增大，查询时的速度受到的影响会非常大

转载：http://blog.csdn.net/jameshadoop/article/details/44905643

好文要顶

关注我

收藏该文

心随梦飞[周路敏]

关注 - 0

粉丝 - 26

+加关注

11

推荐

0

反对

posted @ 2015-07-08 09:22 心随梦飞[周路敏] 阅读(131703) 评论(12) 编辑 收藏

- 发表评论
- #1楼

2015-09-10 08:56 | 王安琪

回复 引用

“

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。这段话的配图错了。

支持(1) 反对(8)
- #2楼

2016-01-12 19:55 | 张冬

回复 引用

“

我写的Ubuntu下ES集群部署博客SEO太差了，百度必应谷歌都搜不到，来推广下方便初学者搜到(*^__^*)
Ubuntu下Elasticsearch 2.1集群部署过程与遇到的问题及解决方法（开机自启动、root用户启动） - 张冬 - 博客园
<http://www.cnblogs.com/zhangdong92/p/5056344.html>

支持(0) 反对(0)
- #3楼

2016-05-30 11:25 | skywalker0011

回复 引用

“

@ 王安琪没配错，那是时间。

支持(0) 反对(0)
- #4楼

2016-06-23 03:38 | 光阴四溅

回复 引用

“

能详细说一下 "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。" 这里的数据量是什么数量级么？多谢。

支持(0) 反对(0)
- #5楼

2016-07-15 14:45 | napoay

回复 引用

“

博主，博客中solr和es性能对比的数据出处能告知一下吗？

支持(1) 反对(0)
- #6楼

2016-09-10 19:03 | 杨Coder

回复 引用

“

@ napoay
英文原文在这里
<http://blog.socialcast.com/realtime-search-solr-vs-elasticsearch/>

支持(0) 反对(0)
- #7楼

2016-11-30 15:04 | 无名小卒的苦

回复 引用

“

浅入深出ElasticSearch构建高性能搜索架构
课程学习地址：<http://www.xuetuwuyou.com/course/161>

支持(0) 反对(0)
- #8楼

2017-02-05 21:56 | 坦荡

回复 引用

“

介绍的好！


支持(0) 反对(0)
- #9楼

2017-02-20 10:33 | 禅道

回复 引用

“

mm

#10楼 2017-03-01 11:34 | 哈雷兄弟 

回复 引用



《浅入深出ElasticSearch构建高性能搜索架构》
课程学习地址：<http://www.xuetuwuyou.com/course/161>
课程出自学途无忧网：<http://www.xuetuwuyou.com>
讲师：夜行侠老师

一、课程用到的软件

ElasticSearch5.0.0
Spring Tool Suite 3.8.2.RELEASE
Gradle2.13
Maven3.0.5
Spring4
Netty4
Hadoop2.7.1
Kibana5.0
JDK1.8.0_111

二、课程目标

- 1、快速学习ElasticSearch
- 2、了解ElasticSearch内部原理
- 3、构建高性能ES搜索架构

三、适合人群

- 1、想学习搜索引擎
- 2、想构建类似淘宝主页的搜索接口
- 3、想深入了解ES底层原理，玩转ES

四、课程目录

- 课时1：重要声明
- 课时2：es与solr对比介绍
- 课时3：centos的安装虚拟机网络配置
- 课时4：elasticsearch5.0集群安装
- 课时5：elasticsearch head插件讲解
- 课时6：elasticsearch kibana 安装与使用
- 课时7：ik分词器插件的编译安装
- 课时8：ik分词器插件从mysql中定时获取扩展词停顿词
- 课时9：ik分词器插件从mysql中定时获取扩展词停顿词
- 课时10：ik分词器扩展词和停顿词动态加载演示
- 课时11：ik分词器获取扩展词后的使用
- 课时12：实战es的java客户端
- 课时13：es的java客户端5.0与spring整合
- 课时14：es的rest客户端使用，以及与spring整合
- 课时15：Netty服务器构建
- 课时16：netty http服务器搭建并与客户端通信
- 课时17：netty http rpc构建
- 课时18：Rest客户端单个文档插入到索引中
- 课时19：Rest客户端批量插入多个文档并优化架构
- 课时20：免费 关键词索引
- 课时21：过滤器索引，Rest响应结果格式化
- 课时22：内嵌过滤器,整理返回结果
- 课时23：组合查询（Bool）组合内嵌查询上
- 课时24：组合查询（Bool）组合内嵌查询下
- 课时25：分组查询(agggregation)
- 课时26：agggregation 内嵌分组
- 课时27：常用分词器讲解（正则过滤器）
- 课时28：修改ik分词器插件加上同义词过滤器
- 课时29：ik分词器插件同义词应用
- 课时30：elasticsearch实现拼音搜索
- 课时31：es实现拼音搜索+ik分词器+同义词
- 课时32：in查询，高亮，和排序
- 课时33：es插件讲解
- 课时34：suggest推荐索引操作
- 课时35：suggest推荐索引架构实现
- 课时36：elasticsearch实现架构分析
- 课时37：guice 讲解，实例分析
- 课时38：elasticsearch6.0源码在eclipse上构建
- 课时39：elasticsearch6.0源码在eclipse调试与分析
- 课时40：X-Pack在es与Kibana上的安装
- 课时41：es的权限控制
- 课时42：集群健康状况，各个节点状态，索引状态查看
- 课时43：线上服务器硬件配置要求，jvm的配置，es重要配置讲解
- 课时44：es的备份和恢复，以及如何与hadoop，spark等整合

#11楼 2017-05-09 10:08 | 易连山

回复 引用

“
博主，能提供下关于solr与 es的数据分析对比报告么

支持(0) 反对(0)

#12楼 2017-06-25 20:55 | smart152819

回复 引用

“
夜行侠老师视频教学：es问题汇总
<http://www.itjoin.org/>索的精准性控制
3、解决搜索中英文数字混合
4、搜索中in查询
5、安装问题
6、es源码内部是如何做到插件化扩展
7、统计时展示不全
8、搜索条件过多，内容过长报错
9、索引某一字段内容太大
10、自动补全设计
11、es的分布式如何实现RPC底层
12、es的分片机制
13、head插件的使用
14、es分片和副本级的是原理
15、如何做到只更新部分数据
16、数组在index中如何制定
17、es搜索结果窗口太大
18、group by里如何排序
19、es底层新增和删除索引的全过程
20、es底层检索的全过程

支持(0) 反对(0)

刷新评论 刷新页面 返回顶部

发表评论

昵称: 克己、

评论内容:

提交评论 退出 订阅评论

[Ctrl+Enter快捷键提交]

- 最新IT新闻：
- 手机预装软件卸载难：系统升级后一些APP又回来了
 - 特斯拉强化客服措施 允许客户直接向公司高管投诉
 - 广达获第三代苹果手表订单 进行测试即将大规模量产
 - 亚马逊推开发包：任意外部硬件都能植入Alexa语音功能
 - 信而富第二季度营收1520万美元 同比增长9%
- » 更多新闻...

- 最新知识库文章：
- 做到这一点，你也可以成为优秀的程序员
 - 写给立志做码农的大学生
 - 架构腐化之谜
 - 学会思考，而不只是编程
 - 编写Shell脚本的最佳实践
- » 更多知识库文章...

