

Elasticsearch 20:30开始

好消息!!!

大数据线上班随到随学! 火热报名中!!!

-- 终身免费重学

-- 老师一对一辅导! 电脑远程协助解决问题!

大数据线下班将于3月28日开班! 火热报名中!!!

-- 老师面授课程! 传统式教室教学已开班多期! 学习完美就业!

大数据周末班将于5月7日再次开班! 火热报名中!!!

贾老师: 1786418286

何老师: 1926106490

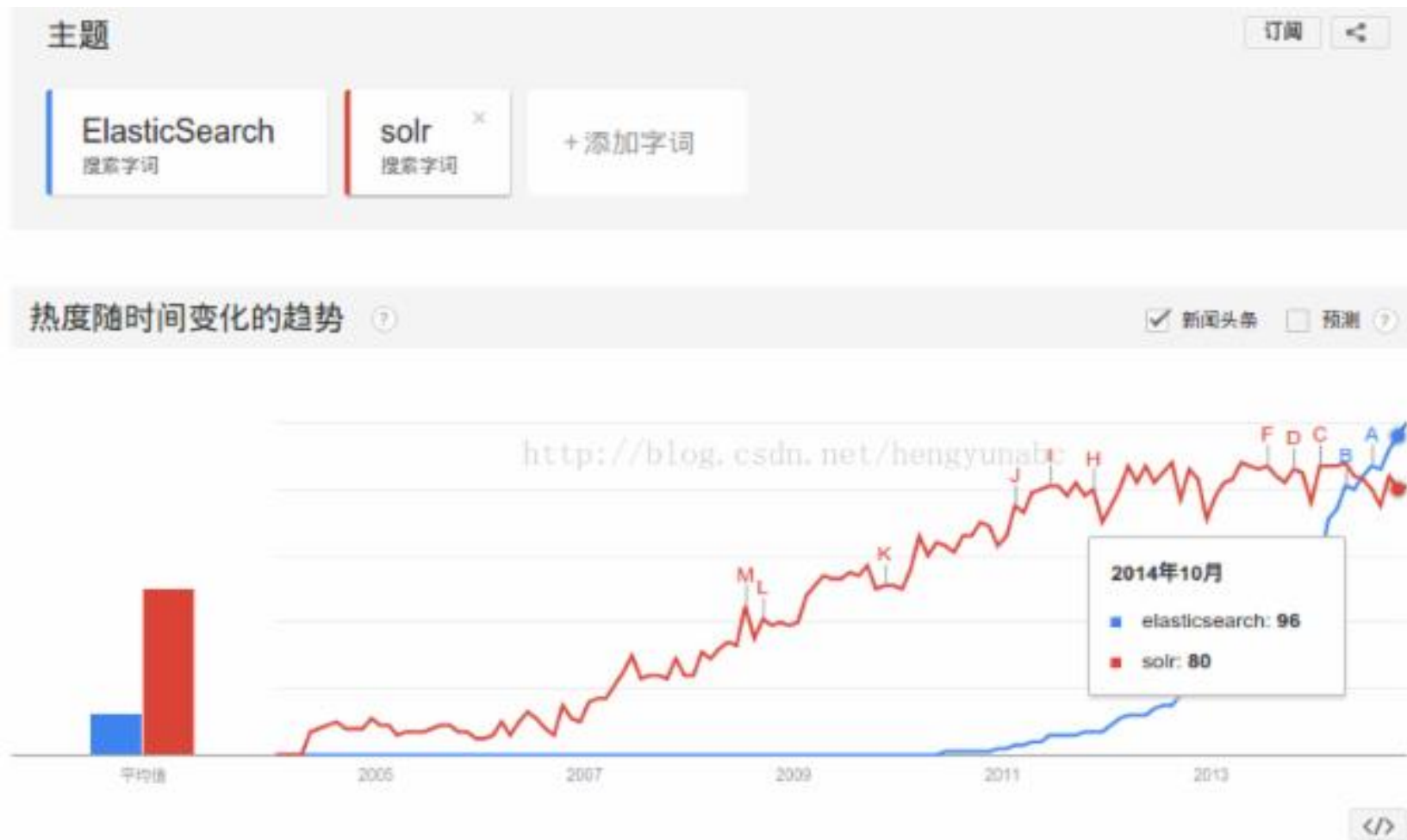
詹老师: 2805048645

讨论技术可以加入以下QQ群: 172599077, 156927834

讲师: 君临天下

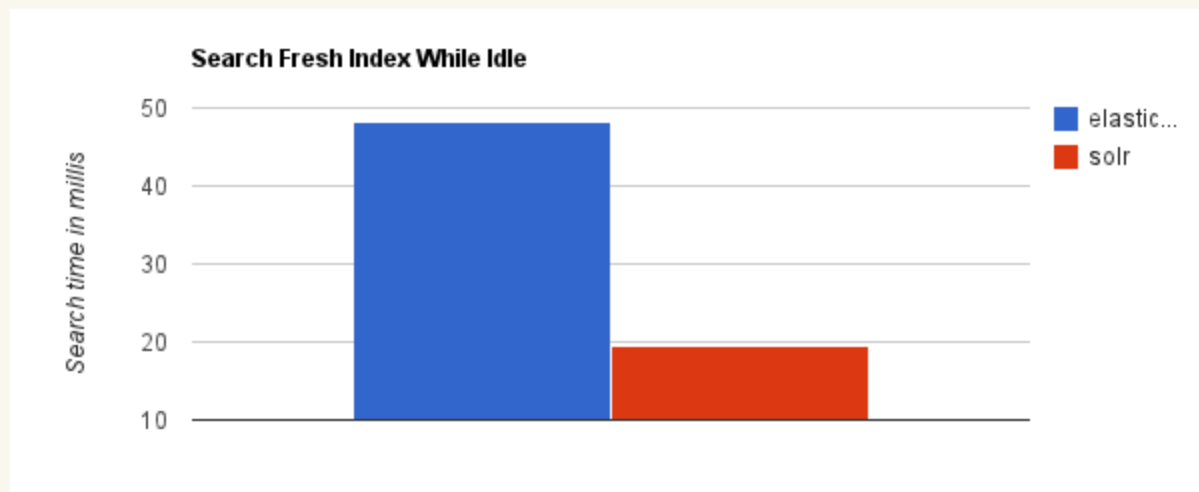
- **什么是Elasticsearch , 和lucene以及solr的关系 ?**
- **官网**<https://www.elastic.co/products>
- **特点 :**
 - 动态
 - 分布式

- Es和solr百度指数对比



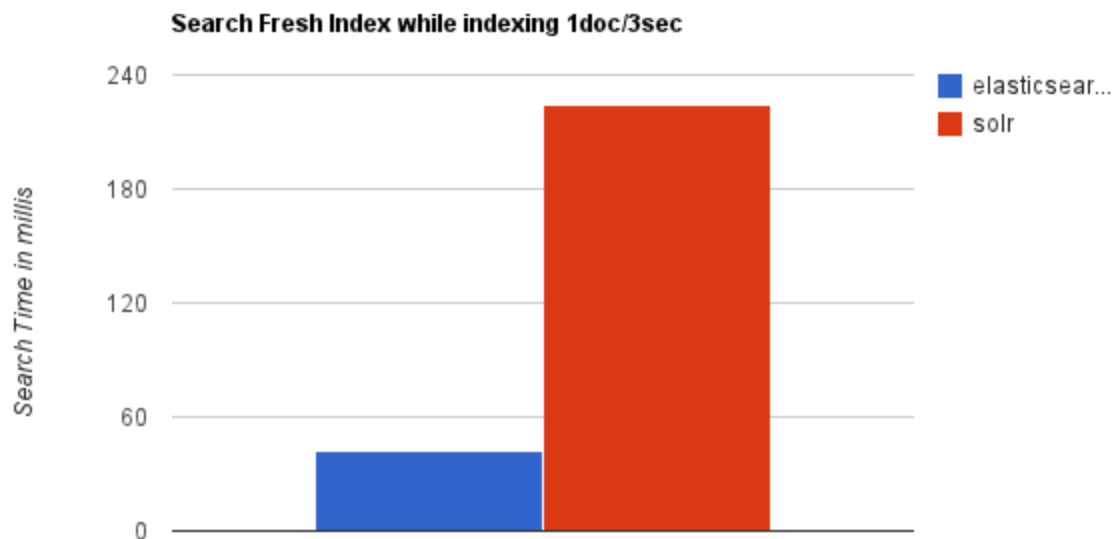
- **Elasticsearch和solr性能对比**

当单纯的对已有数据进行搜索时，Solr更快。



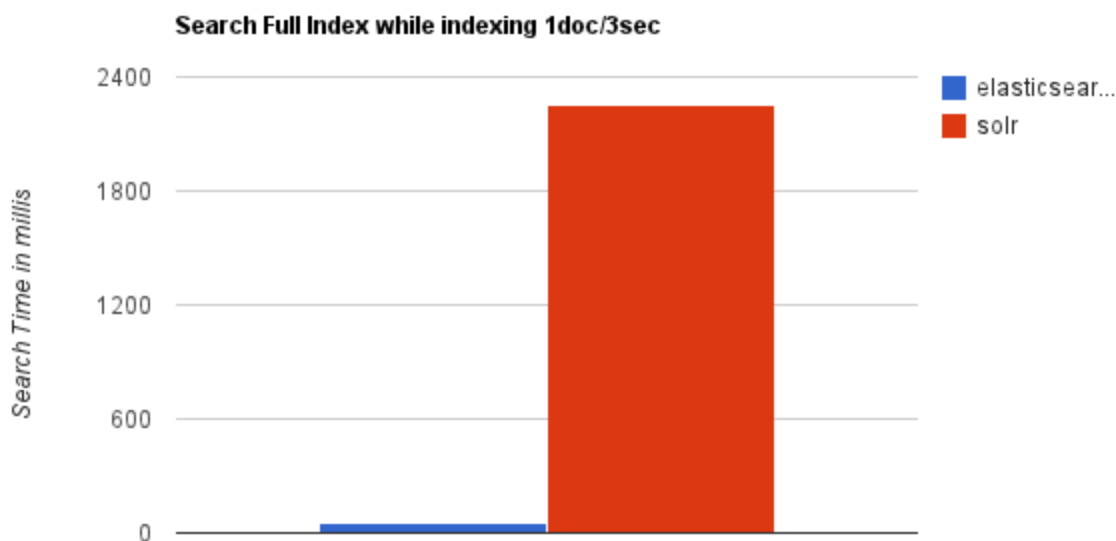
- **Elasticsearch和solr性能对比**

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性能对比**

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。



- Solr和elasticsearch的性能对比

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。



- **Lucene的核心理论——倒排索引**

- **根据属性的值来查找记录。这种索引表中的每一项都包括一个属性值和具有该属性值的各记录的地址。由于不是由记录来确定属性值，而是由属性值来确定记录的位置，因而称为倒排索引(inverted index)**
- **单词——文档矩阵**

单词-文档矩阵					
	文档1	文档2	文档3	文档4	文档5
词汇1	✓			✓	
词汇2		✓	✓		
词汇3				✓	
词汇4	✓				✓
词汇5		✓			
词汇6			✓		

- **Lucene索引中的概念**

- document

用户提供的源是一条条记录，它们可以是文本文件、字符串或者数据库表的一条记录等等。一条记录经过索引之后，就是以Document的形式存储在索引文件中的。用户进行搜索，也是以Document列表的形式返回。

。

- field

一个Document可以包含多个信息域，例如一篇文章可以包含“标题”、“正文”、“最后修改时间”等信息域，这些信息域就是通过Field在Document中存储的。

- segment

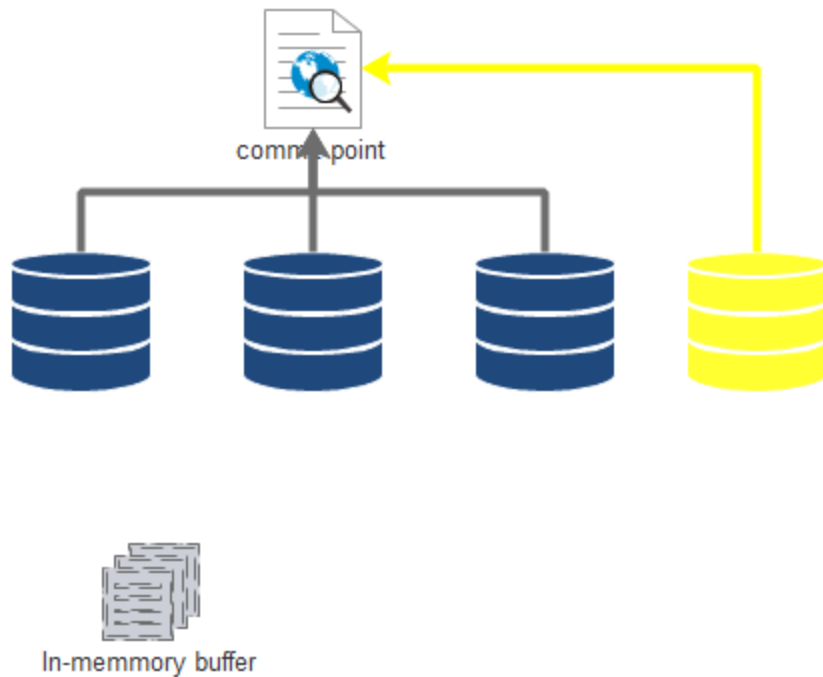
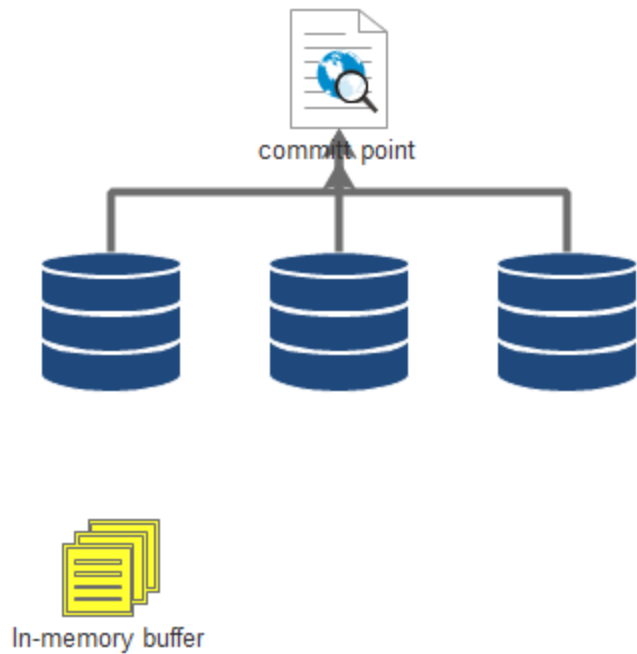
添加索引时并不是每个document都马上添加到同一个索引文件，它们首先被写入到不同的小文件，然后再合并成一个大索引文件，这里每个小文件都是一个segment。

- **Elasticsearch数据流向**

- 动态更新lucene索引，规则：**新收到的数据写入到新的索引文件里面**

- **步骤：**

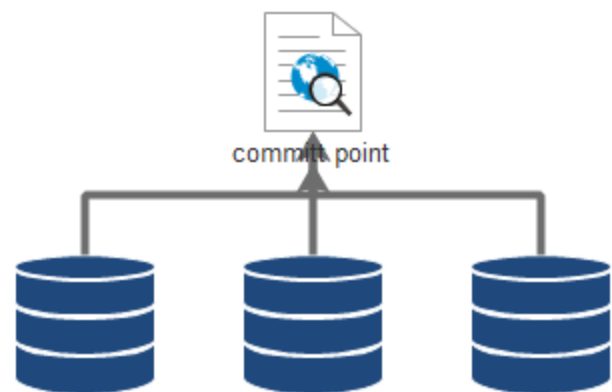
- 每次生成的倒排索引叫一个段（segment）然后另外使用一个commit文件记录索引内所有的segment，生成segment的数据来源是内存buffer




默认每隔一秒刷一次到文件系统缓存，文件系统缓存再到磁盘，可以调用/_refresh手动刷新

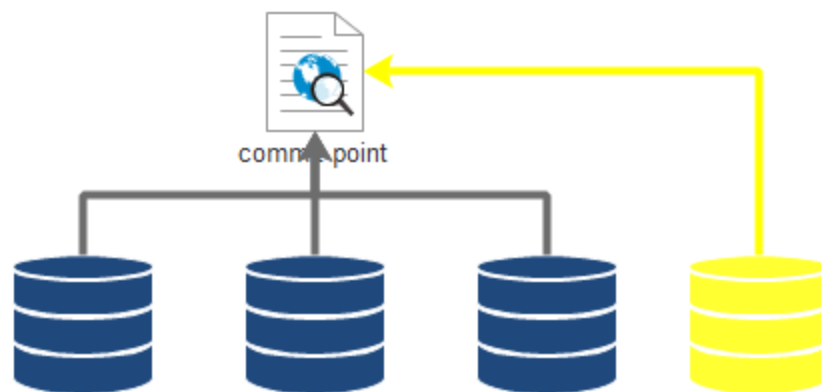
- **思考：既然每隔一秒只是写到文件系统缓存，那么最后一步写到实际磁盘是什么来控制的？如果中间出现主机错误、硬件故障等异常，数据会不会丢失呢？**

- Translog提供磁盘同步控制
 - 数据写入内存buffer同时记录了一个translog日志





In-memory buffer


Translog




In-memory buffer


Translog

- **Translog提供磁盘同步控制**

- 等到commit文件更新的时候，translog才清空，这一步叫做flush，默认每半小时刷新一次，也可以手动调用
- 也可以通过配置index.translog.flush_threshold_ops参数，控制每多少条刷新一次

- **Segment merge**
 - 独立线程做merge工作



- **Elasticsearch 环境安装**

- **Elasticsearch javaAPI**
 - 见附件

- **明天课程预告：**
- **Elasticsearch的核心概念**
- **Elasticsearch的操作**