

问题日志

Kerwin:20:33:04 ES 几节课?

答：两节，后边的项目中实战还会再讲两节

jackhu:20:55:26

Elk, es 实战： 拼音搜索，联想词语，api，索引库设计

李文峰-长沙:20:35:20

ES 前面几个月公网环境爆出了安全漏洞

答：爆出漏洞肯定会修复的，不会影响他的地位

大鹏 ~:20:58:57

newsq1 阿里的 oceanDb 江湖上处在什么位置啊？

答：目前流行度不高

354 systems in ranking, March 2020

Rank			DBMS	Database Model	Score		
Mar 2020	Feb 2020	Mar 2019			Mar 2020	Feb 2020	Mar 2019
1.	1.	1.	Oracle	Relational, Multi-model	1340.64	+4.11	+61.50
2.	2.	2.	MySQL	Relational, Multi-model	1259.73	-7.92	+61.48
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1097.86	+4.11	+50.01
4.	4.	4.	PostgreSQL	Relational, Multi-model	513.92	+6.98	+44.11
5.	5.	5.	MongoDB	Document, Multi-model	437.61	+4.28	+36.27
6.	6.	6.	IBM Db2	Relational, Multi-model	162.56	-2.99	-14.64
7.	7.	9.	Elasticsearch	Search engine, Multi-model	149.17	-2.98	+6.38
8.	8.	8.	Redis	Key-value, Multi-model	147.58	-3.84	+1.46
9.	9.	7.	Microsoft Access	Relational	125.14	-2.92	-21.07
10.	10.	10.	SQLite	Relational	121.95	-1.41	-2.92
11.	11.	11.	Cassandra	Wide column	120.95	+0.60	-1.84
12.	12.	13.	Splunk	Search engine	88.52	-0.26	+5.42
13.	13.	12.	MariaDB	Relational, Multi-model	88.35	+1.01	+4.04
14.	14.	15.	Hive	Relational	85.38	+1.85	+12.38
15.	15.	14.	Teradata	Relational, Multi-model	77.84	+1.03	+2.63
16.	16.	21.	Amazon DynamoDB	Multi-model	62.51	+0.38	+8.02
17.	17.	16.	Solr	Search engine	55.09	-1.07	-4.92
18.	18.	20.	SAP HANA	Relational, Multi-model	54.27	-0.70	-1.24
19.	19.	18.	FileMaker	Relational	54.16	-0.72	-3.97
20.	21.	19.	SAP Adaptive Server	Relational	52.77	+0.04	-3.27

张勇:21:10:00

定位静夜思，怎么找到的呢？

答：在索引里从前往后找

Jaye:21:10:30

索引存储是有顺序的。

马瑞谦:21:10:32

倒排的索引是提前建立的

答：是的

李东雪:21:10:35

这个 es 用了分词？

答：要使用中文分词器

peter:21:16:18

数据库如果有建索引，检索是采用什么扫描？也是顺序扫描吗？

答：倒排索引

jackhu:21:30:48

手机 app

每一个 app 都是一个搜索

每一个互联网软件都有搜索

jackhu:21:38:57

solr 适合做静态 少量数据检索

elasticSearch 适合对实时产生数据进行搜索

Kerwin:21:41:09

多少数据属于海量呢

jackhu:21:44:54 TB

铭斌:21:48:59

索引库 是用什么数据库

答：不是其他的数据库，es 存储索引的就叫索引库

Kerwin:21:54:09

es 不是更新索引性能很差吗，是根据 alias 命令来更新吗？

答：可以用 kafka 之类的做缓存，然后批量更新

李振华:22:02:32

月 是一个词么？

答：可以自定义扩展词汇

Kerwin:22:04:39

爬虫的数据灌入 es，不是属于更新吗

答：属于

刘盼:22:05:11

网页一直变 那要一直更新吗

答：是的，更新规则自己来定，es 只是存储

jackhu:22:09:32

中文环境下，数据量非常大，索引库是否会无限制膨胀？

答：索引数量有限，但是文档可能无限

小黄:22:12:06

索引怎么和文档关联？

答：索引存储文档 ID 即可

peter:22:13:10

1 条记录是一个文档，那 2 条记录是不是就是 2 个文档了？

答：是的

马瑞谦:22:20:05

ik 分词器怎么做到分成静夜思和李白的？

答：分词器有词库

郭子林:22:22:02

李白和静夜思是不是会对应一个文档？

答：可能会，但是会合并

姜朋-上海:22:22:37

还有一个李白怎么匹配，两个词检索出来的需要合并么

答：需要

李振华:22:22:49

那这个搜索结果 是 静夜思 + 李白 倒排的链表合集么？

答：是的

马瑞谦:22:22:52

类似词典？是可以自定义更新的吗

答：可以

大鹏 ~:22:23:20

搜索结果顺序是什么呢？

答：匹配度，得分情况

小黄:22:23:44

如果输入的是“静夜思李”呢

答：分成静夜思，词库有李就有李

姜朋-上海:22:23:50

是并集还是交集

答：并

李东雪:22:25:04

加了索引也是全表扫描进行匹配？

答：扫描的是索引

姜朋-上海:22:26:17

得分的算法会讲么

答：项目会讲一点

默轩-梁明昱-杭州:22:29:30

图片搜索 用的是什么技术？

答：图片搜索的原理是根据特征值搜索的

Java.Gibson:22:35:52

倒排链表是什么时候创建的？

答：更新索引的时候

大鹏 ~:22:41:25

es 为什么依赖 lucene 不是其他 或者 更底层呢？

答：lucene 已经是底层了，是最好的选择

默轩-梁明显-杭州:22:45:44

ES 是不是只支持文字搜索？

答：可以用文字搜图片

默轩-梁明显-杭州:22:46:25

百度可用图片搜索图片

答：图片是用特征值搜索的，如果传入图片，提取特征值，也可以以图搜图

大鹏 ~:22:52:45

es 是搜索领域唯一方案吗？

答：不是，常用的还有 solr