**1 全文检索的原理: 倒排索引原理;**

1-1: 数据分为

结构化数据:固定格式/有限长度的数据: 数据库/元数据 处理方式: 数据库通过SQL语句/ 元数据的话,利用windows搜索对文件名;

非结构化数据;:不定长度/无固定格式; 邮件/word文档; 处理方式: 顺序扫描法/ 从非结构化数据中提取重新组织(索引)

综上所述: **先建立索引,再对索引进行搜索的过程称为:全文检索(Full-text-Search)**

1-2: 全文检索过程分为: 索引创建: 将显示世界中所有的结构化/非结构化数据提取信息,创建索引的过程;

搜索索引: 得到用户的查询请求,搜索创建的索引,返回结果的过程;

倒排索引理解: 从字符串到文件的映射 是文件到字符串映射的反向过程,于是保存这种信息的索引称为反向索引;

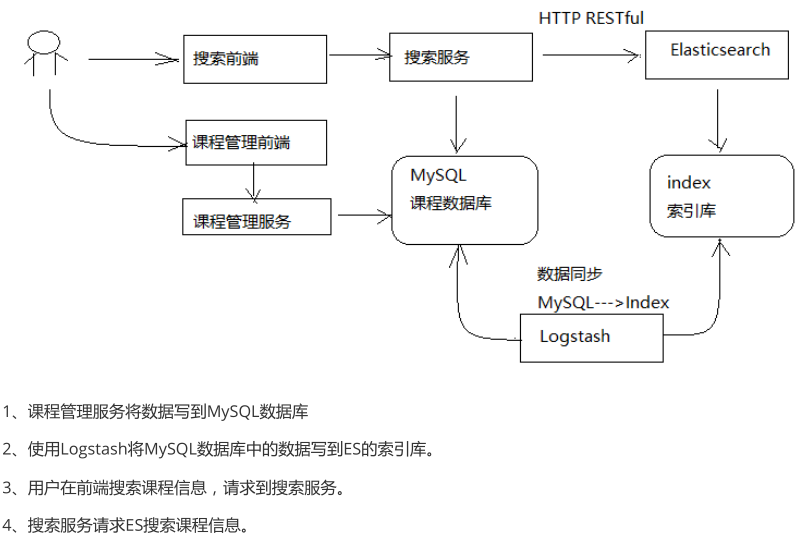
优势: 一次索引, 多次使用;

2 实际开发流程:

2-1: 根据页面需求分析;



2-2: 搜索流程:



2-3: 技术方案: **先把数据索引到索引库中;**

1 当课程向Mysql 添加后,同时将课程信息添加到索引库--**-采用Logstash实现,从Mysql中将数据采集到ES 索引库;**

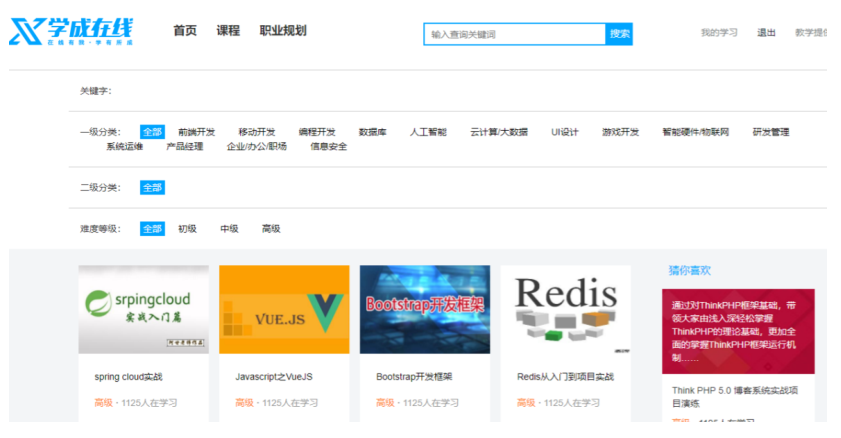
2 当课程向MySQL 更新课程的同时,需要更新索引库的信息----**采用Logstash实现;**

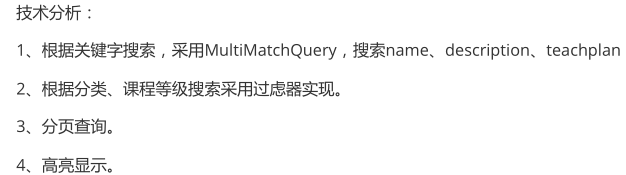
3 当课程在Mysql删除后同时将该课程从索引库删除----------**手动程序实现,在删除课程后将索引库中课程信息删除;**

2-4:

使用Logstash,解压 安装logstash-input-jdbc插件,创建模板文件/my.conf;

**2-5: 具体的分析:**





3 创建搜索服务工程; **从索引库中查询需要的数据; 主要使用DSL;**

4

ES 工作原理:

4-2 ES 核心概念:

Cluster:集群; Es工作在多个服务器上; Node:节点, 形成集群的每个服务器;

Shard: 分片; 数据分为较小的分片,每个分片放到不同的服务器上. 在查询的时候,每个相关的分片,组合到一起;

Replia: 副本; 为了提高高可用性,可以使用分片副本/ 当主分片丢失时,集群会将副本提升为新的主分片;

ES特点和优势: 索引分拆成多个分片, 每个分片可能有多个副本.

4-3: ES与Solr 比较:

Solr 利用Zookeeper进行分布式管理; Elasticsearch 自身带有分布式协调管理功能;

Solr支持更多格式数据;Elasticsearch仅仅支持Json文件格式;

Solr在传统的搜索应用中表现良好;Elasticsearch 在处理实时搜索应用是效率高;

随着数据量的增加,Solr的搜索效率会更低;Elasticsearch则没有明显的变化;

5 ES面试题集锦:

<https://juejin.im/entry/5c46d7c2e51d4551df6f2338> BAT等一线大厂 Elasticsearch面试题解读

<https://blog.csdn.net/lby0307/article/details/79677803> Elasticsearch面试题;

<https://blog.csdn.net/suchahaerkang/article/details/87531647> Elasticsearch;