

1.为什么要使用 ES?

虽然 lucene 是性能最先进的、功能完善的搜索引擎，但是它只能有一个库，而且使用它需要集成到 java 项目才能使用，配置还特别复杂,而 ES 语法简单、是基于 Resultful api 风格的,它在 lucene 所有功能基础上还能集成多台服务器进行搜索，也就是集群;

2.ES 和 lucene 的区别是什么？

lucene 是现存功能最强大、最先进搜索库，直接基于 lucene 开发，api 非常复杂大量的 java 代码、需要深入了解原理

对比处:

只有一个索引库、
传统的遍历搜索方式
采用 B+树索引;

Es 是基于 lucene 的，隐藏了 lucene 复杂部分的一个分布式全文检索框架

对比处:

一个 es 的集群包含多个索引库、
分布式搜索

Es 是采用倒排索引

es 没有事物概念,删除不能恢复

es 开源免费

正排索引: id ---> value

倒排索引: value ---> id

3.你还了解哪些全文检索工具？

Lucene, Solr, HadoopContrib, Katta

4.Bulk 一次最大处理多少数据量？

bulk 会把将要处理的数据载入内存中，所以数据量是有限制的

最佳的数据量不是一个确定的数值，它取决于你的硬件，你的文档大小以及复杂性，你的索

引以及搜索的负载。

一般建议是 1000-5000 个文档,如果你的文档很大,可以适当减少队列,大小建议是 5-15MB,默认不能超过 100M,可以在 es 的配置文件中修改这个值 `http.max_content_length: 100mb`

5.ES 在高并发的情况下如何保证数据线程安全问题?

在读数据与写数据之间如果有其他线程进行写操作,就会出问题,es 使用版本控制才避免这种问题

在修改数据的时候指定版本号,操作一次版本号加 1

6.ES 管理的工具有哪些?

1) BigDesk Plugin (作者 Lukáš Vlček): 监控 es 状态的插件,推荐! 主要提供的是节点的实时状态监控,包括 jvm 的情况,linux 的情况,elasticsearch 的情况

2) Elasticsearch Head Plugin (作者 Ben Birch): 很方便对 es 进行各种操作的客户端。?

3) Kibana(ES 官方推荐): Es 基于 RestAPI 调试、数据报表、数据搜索

7.ES 自动映射的规则?

Es 映射是用于确定字段类型,将新增的每个字段数据类型映射后确定的字段类型,常见的映射类型有基本数据类型、复杂数据类型映射有数组对象(但对象映射扁平化后相关性会丢失,不能确定哪个是哪个属性)、

全局映射的方式默认映射继承 `_default_` 的配置、动态模板映射: `dynamic_templates`,使用动态模板映射会覆盖默认模板,可设置关键字分词,默认模板是 `string` 类型的。