1. es 分布式架构原理
2. es写入数据的流程 查询原理什么

es有可能存在5秒的数据丢失

数据写入时-》随机找一台es(协调协作)-》hash 找到对应的primary shard写-》同步到其他副本shard

数据先写入系统buffer->每一秒刷入os cash ,每秒写入一秒的 translog到 os cash ->

写入数据Sagment file ,每过5秒translog 日志中-》日志每过30分钟flash一次

删除数据：标识写入磁盘.del文件,搜索时会剔除搜索结果

Merage:当segmentfile 文件多到一定程度，会触发文件的合并,在merage过程中和.del文件比较,物理删除数据

搜索：随机找到一个结点（协调节点）,然后找到每一个shard,查询所有的结果,然后再匹配 找到更精确的

查询：根据id找到对应的primary shard,得到数据

1. es如何提高查询性能

性能提高是没有绝对的解决方法的

读取数据时流程

es

shard

File system cash

Segment file

1. 保证filesystemcash 的大小
2. Es中存储的数据 尽量存关键的数据 +mysql
3. 后台 可以每过一段时间 预热一些高热点的数据,将数据刷到systemfilecash上
4. 水平拆分 热数据和一般数据 分开索引

Es数据模型 不要使用复杂的搜索语法,会拉低性能

Es分页 越深入,速度越慢，

app下拉 scroll api

4.es生产集群的架构是什么 每个索引的数据量有多大 每个索引有多大