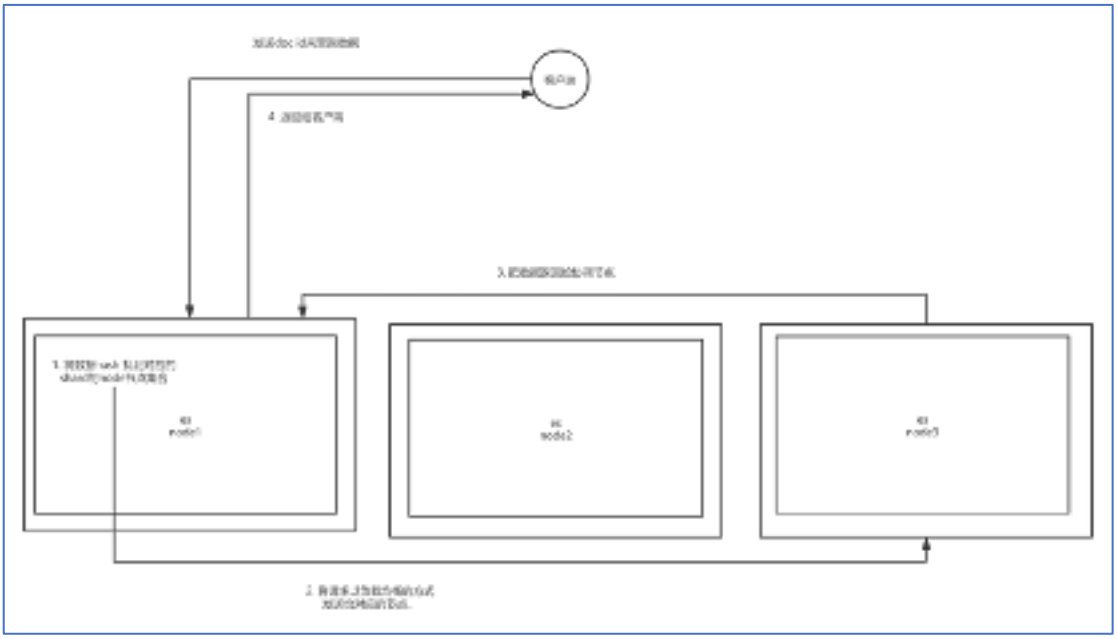
Elasticsearch 面试题

1、Elasticsearch 读取数据

使用 RestFul API 向对应的 node 发送查询请求，根据 did 来判断在哪个 shard 上，返回的是 primary 和 replica 的 node 节点集合。

这样会负载均衡地把查询发送到对应节点，之后对应节点接收到请求，将 document 数据返回协调节点，协调节点把 document 返回给客户端。



2、您能解释一下 X-Pack for Elasticsearch 的功能和重要性吗？

X-Pack 是与 Elasticsearch 一起安装的扩展程序。

X-Pack 的各种功能包括安全性（基于角色的访问，特权/权限，角色和用户安全性），监视，报告，警报等。

3、Elasticsearch 中的节点（比如共 20 个），其中的 10 个选了一个 master，另外 10 个选了另一个 master，怎么办？

- 当集群 master 候选数量不小于 3 个时，可以通过设置最少投票通过数量（discovery.zen.minimum_master_nodes）超过所有候选节点一半以上来解决脑裂问题；
- 当候选数量为两个时，只能修改为唯一的一个 master 候选，其他作为 data 节点，避免脑裂问题。

4、解释一下 Elasticsearch 集群中的 索引的概念 ？

Elasticsearch 集群可以包含多个索引，与关系数据库相比，它们相当于数据库表。

5、你可以列出 Elasticsearch 各种类型的分析器吗？

Elasticsearch Analyzer 的类型为内置分析器和自定义分析器。

Standard Analyzer

标准分析器是默认分词器，如果未指定，则使用该分词器。

它基于 Unicode 文本分割算法，适用于大多数语言。

Whitespace Analyzer

基于空格字符切词。

Stop Analyzer

在 simple Analyzer 的基础上，移除停用词。

Keyword Analyzer

不切词，将输入的整个串一起返回。

自定义分词器的模板

自定义分词器的在 Mapping 的 Setting 部分设置：

```
PUT my_custom_index
{
  "settings":{
    "analysis":{
    "char_filter":{},
    "tokenizer":{},
    "filter":{},
    "analyzer":{}
    }
  }
}
```

其中：

“char_filter”:{},——对应字符过滤部分；

“tokenizer”:{},——对应文本切分为分词部分；

“filter”:{},——对应分词后再过滤部分；

“analyzer”:{}——对应分词器组成部分，其中会包含：1. 2. 3。

6、解释一下 Elasticsearch Node？

节点是 Elasticsearch 的实例。实际业务中，我们会说：ES 集群包含 3 个节点、7 个节点。

这里节点实际就是：一个独立的 Elasticsearch 进程，一般将一个节点部署到一台独立的服务器或者虚拟机、容器中。

不同节点根据角色不同，可以划分为：

主节点

帮助配置和管理在整个集群中添加和删除节点。

数据节点

存储数据并执行诸如 CRUD（创建/读取/更新/删除）操作，对数据进行搜索和聚合的操作。

1、 客户端节点（或者说：协调节点） 将集群请求转发到主节点，将与数据相关的请求转发到数据节点。

2、 摄取节点

用于在索引之前对文档进行预处理。

7、在安装 Elasticsearch 时，请说明不同的软件包及其重要性？

这个貌似没什么好说的，去官方文档下载对应操作系统安装包即可。

部分功能是收费的，如机器学习、高级别 kerberos 认证安全等选型要知悉。

8、Elasticsearch 在部署时，对 Linux 的设置有哪些优化方法？

1、 关闭缓存 swap;

2、 堆内存设置为：Min（节点内存/2, 32GB）;

3、 设置最大文件句柄数；

4、 线程池+队列大小根据业务需要做调整；

5、 磁盘存储 raid 方式——存储有条件使用 RAID10，增加单节点性能以及避免单节点存储故障。

9、请解释有关 Elasticsearch 的 NRT？

从文档索引（写入）到可搜索到之间的延迟默认一秒钟，因此 Elasticsearch 是近实时（NRT）搜索平台。

也就是说：文档写入，最快一秒钟被索引到，不能再快了。

写入调优的时候，我们通常会动态调整：refresh_interval = 30s 或者更达值，以使得写入数据更晚一点时间被搜索到。

10、elasticsearch 的 document 设计

在使用 es 时 避免使用复杂的查询语句（Join 、聚合），就是在建立索引时，就根据查询语句建立好对应的元数据。