1. **说一下你们公司平台的架构**

十几台web服务器 两台redis 一台ES 一台Kafka

1. **es用在哪**

ElasticSearch的功能

分布式的搜索引擎和数据分析引擎

搜索：网站的站内搜索，IT系统的检索

数据分析：电商网站，统计销售排名前10的商家

全文检索，结构化检索，数据分析

全文检索：我想搜索商品名称包含某个关键字的商品

结构化检索：我想搜索商品分类为日化用品的商品都有哪些

数据分析：我们分析每一个商品分类下有多少个商品

对海量数据进行近实时的处理

分布式：ES自动可以将海量数据分散到多台服务器上去存储和检索

海联数据的处理：分布式以后，就可以采用大量的服务器去存储和检索数据，自然而然就可以实现海量数据的处理了

近实时：检索数据要花费1小时（这就不要近实时，离线批处理，batch-processing）；在秒级别对数据进行搜索和分析

ElasticSearch的应用场景

维基百科

The Guardian（国外新闻网站）

Stack Overflow（国外的程序异常讨论论坛）

GitHub（开源代码管理）

电商网站

日志数据分析

商品价格监控网站

BI系统

站内搜索

ElasticSearch的特点

可以作为一个大型分布式集群（数百台服务器）技术，处理PB级数据，服务大公司；也可以运行在单机上，服务小公司

Elasticsearch不是什么新技术，主要是将全文检索、数据分析以及分布式技术，合并在了一起

对用户而言，是开箱即用的，非常简单，作为中小型的应用，直接3分钟部署一下ES

Elasticsearch作为传统数据库的一个补充,比如全文检索，同义词处理，相关度排名，复杂数据分析，海量数据的近实时处理；

1. **es存储的数据量有多大**

一般回答几百个G

1. **kafka与redis区别**

redis 消息推送（基于分布式 pub/sub）多用于实时性较高的消息推送，并不保证可靠。  
其他的mq和kafka保证可靠但有一些延迟（非实时系统没有保证延迟）。redis-pub/sub断电就清空，而使用redis-list作为消息推送虽然有持久化，但是又太弱智，也并非完全可靠不会丢。

另外一点，redis 发布订阅除了表示不同的 topic 外，并不支持分组，比如kafka中发布一个东西，多个订阅者可以分组，同一个组里只有一个订阅者会收到该消息，这样可以用作负载均衡。

Redis，它首先是一个内存数据库，其提供的PUB/SUB功能把消息保存在内存中(基于channel)，因此如果你的消息的持久性需求并不高且后端应用的消费能力超强的话，使用Redis PUB/SUB是比较合适的使用场景。比如官网说提供的一个网络聊天室的例子：模拟IRC，因为channel就是IRC中的服务器。用户发起连接，发布消息到channel，接收其他用户的消息。这些对于持久性的要求并不高，使用Redis PUB/SUB来做足矣。

而Kafka是一个完整的系统，它提供了一个高吞吐量、分布式的提交日志(由于提供了Kafka Connect和Kafka Streams，目前Kafka官网已经将自己修正为一个分布式的流式处理平台，这里也可以看出Kafka的野心：-)。除了p2p的消息队列，它当然提供PUB/SUB方式的消息模型。而且，Kafka默认提供了消息的持久化，确保消息的不丢失性(至少是大部分情况下)。另外，由于消费元数据是保存在consumer端的，所以对于消费而言consumer被赋予极大的自由度。consumer可以顺序地消费消息，也可以重新消费之前处理过的消息。这些都是Redis PUB/SUB无法做到的。

**Redis PUB/SUB使用场景：**

1. 消息持久性需求不高  
2. 吞吐量要求不高  
3. 可以忍受数据丢失  
4. 数据量不大

**Kafka使用场景：**

上面以外的其他场景：）  
1. 高可靠性  
2. 高吞吐量  
3. 持久性高  
4. 多样化的消费处理模型

1. **kafka默认缓存时间 7天**

kafka的默认存放时间为7天

Kafka配置参数详解https://www.orchome.com/472

1. **kakka配置 分区**
2. **kafka有五个分区 没有指定分区 会写到哪一个分区**
3. **如果是读数据的话 会读哪一个分区 会读默认的0**
4. **kafka数据存在哪**
5. **redis和kafka队列的区别**
6. **redis有几种类型 string set zet list hash**
7. **Redis的数据存在哪 内存**
8. **set和zset区别**
9. **集合有什么特性 确定性 互异性 无序性**
10. **用户量，pv数据，存在哪**
11. **hadoop干嘛的 是一个大数据生态圈 用什么存储数据 HDFS 用来存海量数据的 hive hbase 进行高级应用**
12. **Hadoop**
13. **mysql数据量多大**
14. **mysql最左匹配原则的左是怎么定义的**
15. **索引的建立 如果不根据业务 DBA会根据什么来建立索引**
16. **索引怎么实现的**
17. **B+和b-有什么区别 谁查询的更快 为什么 树是怎么构建的**
18. **TCP/IP协议 三次握手 两个客户端同时建立连接 如何判断 ACK**
19. **分布式集群怎么部署的 部署了什么东西**
20. **PHP工厂模式 一般需要几个基类**
21. **什么是观察者模式**
22. **Laravel和YII和TP有什么区别 event job command**
23. **PHP7有什么新特性**
24. **linux命令 只知道文件内内容 如何查找文件位置**
25. **更改用户主 chown 更改所有 -R**
26. **PHP想把一维数组分割成二维数组**

珍岛

1. 风控是怎么实现的
2. Redis分布式锁
3. 手机消息推送怎么实现的
4. 对比几个mvc框架的优缺点TP,Laravel,YII
5. jquery取出标签里面的一段存文本怎么做
6. 跨域怎么产生的 jsonp原理是什么 两个一样的域名会不会参数跨域 什么样的情况下会产生跨域
7. 单点登录如何实现的
8. mysql连接用的什么连接的 为什么用PDO连接
9. Mysql长连接，短连接是什么以及应用场景
10. 除了PHP其他语言有接触过吗 GO
11. Redis有什么优势
12. 负载均衡原理是什么
13. 如何抵御短时高并发 比如活动期间 服务器
14. 用什么做缓存 redis redis一台扛不住怎么办
15. PHP优化做了哪些东西
16. 已有的一些模块是如何实现 速度的提升优化 比如
17. 服务器安全
18. XSS攻击
19. cookie如何保证安全性
20. Session回收机制
21. PHP垃圾回收机制
22. 关闭浏览器session还存在吗
23. nginx反向代理 代理别的站点时 那个站点通过什么找到反向代理服务器
24. 日志回收怎么处理的
25. Mysql优化
26. 新闻模块需要几张表
27. 大文件上传
28. PHP Curl
29. 做项目中的难点 推荐系统的实现 推荐算法
30. 对996有什么想法
31. 你们团队多少人 你觉得你属于什么水平
32. 有没有排查过线上比较难的bug
33. 技术难点 有挑战的地方
34. 对加班有什么想法 这个行业如果不加班的话 不太现实 最大的承受程度
35. 遇到自己解决不了的问题 怎么办
36. 提前完成了任务 一般会怎么做
37. 测试测出来你的bug多不多