

你好，欢迎来到分布式存储模块的加餐环节，本课时我将和你一起梳理面试中分布式系统的数据库的高频考点，做到温故知新。

面试中如何考察分布式存储

广义的分布式存储根据不同的应用领域，划分为以下的类别：

- 分布式协同系统
- 分布式文件系统
- 分布式任务调度框架
- 分布式 NoSQL 存储
- 分布式关系数据库
- 各种消息队列 MQ
- 流式计算框架

当然，这只是一种划分方式，你也可以根据存储数据的特点，将分布式存储系统划分为块存储、对象存储等不同的分类。

可以看到，分布式存储技术的范围非常大，技术覆盖的广度和深度都很有料，比如分布式协同系统或者各种流计算框架，都可以单独作为一个专栏来进行展开讲解。

由于篇幅有限，我在分布式存储这个模块里，主要围绕分布式系统下的关系型数据库这一主题，选择了与大部分开发者直接相关的热点内容，包括数据库的读写分离、分库分表存储拆分后的唯一主键问题，以及典型的 NoSQL 数据库应用。另外，简单介绍了 ElasticSearch 技术、倒排索引的实现等。

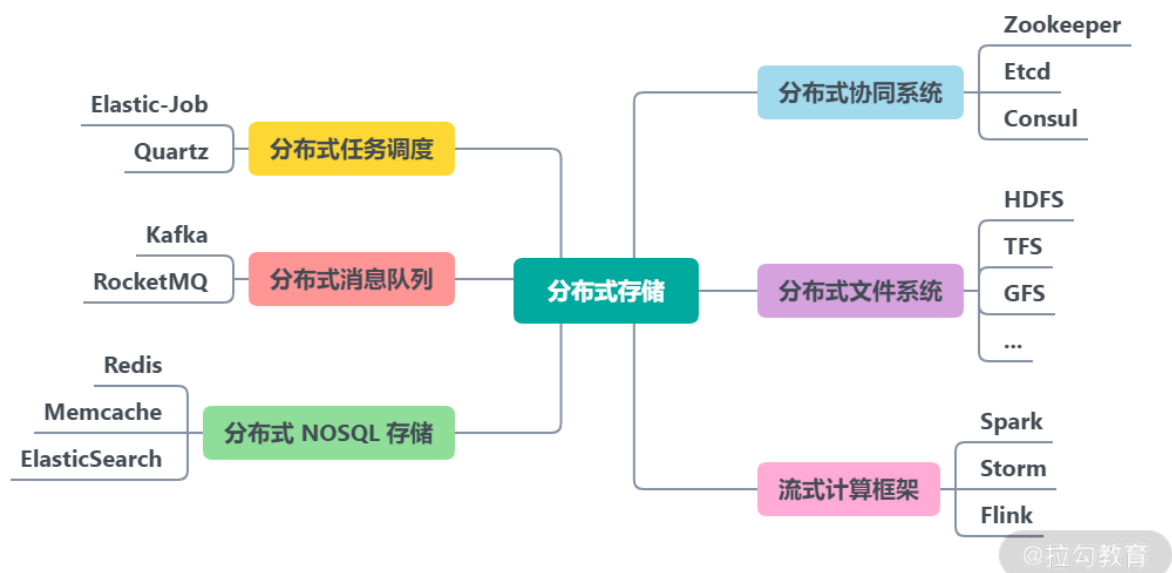
和之前一样，我在这里选择了一些热点技术问题，你可以考察一下自己的掌握程度。以分布式场景下的数据库拆分为例，面试官会对你进行下面的考察：

- 当高并发系统设计时，为什么要分库分表？
- 用过哪些分库分表中间件？
- 不同的分库分表中间件都有什么优点和缺点？
- 如何对数据库进行垂直拆分或水平拆分？
- 如果要设计一个可以动态扩容缩容的分库分表方案，应该如何做？
- 数据库分库分表以后，如何处理设计主键生成器？
- 不同的主键生成方式有什么区别？

上面的问题，都可以在“分布式存储”模块的内容中找到思路，你可以对照本模块学过的知识，整理自己的答案。

分布式存储有哪些高频考点

上面我提到过，分布式存储包含了非常丰富的技术栈，本模块的内容虽然在实际开发中有着高频应用，但只是分布式存储技术领域非常小的一部分。在下面这张思维导图中，除了分布式下的关系型数据库之外的内容，我还补充了一些经典分布式存储技术的部分，你可以对照这张思维导图，进行针对性的扩展。



以分布式文件系统为例，常见的分布式文件系统有 Google 的 GFS、Hadoop 实现的分布式文件系统 HDFS、Sun 公司推出的 Lustre、淘宝的 TFS、FastDFS 等，这几种存储组件都有各自的应用场景。

比如淘宝的 TFS 适合用于图片等小文件、大规模存储的应用场景，是淘宝专门为了支持电商场景下数以千万的商品图片而开发的；FastDFS 类似 GFS，是一款开源的分布式文件系统，适合各类规模较小的图片和视频网站。

比如流式计算框架，有著名的流式计算三剑客，Storm、Spark 和 Flink，这三个框架基本上覆盖了绝大多数的流式计算业务，适用于不同的大数据处理场景。

今天的内容就到这里了，也欢迎你留言分享自己的面试经验，和大家一起讨论。

精选评论