

私塾在线 《高级软件架构师实战培训 阶段二》 ——跟着cc学架构系列精品教程

1010101010101010101010101

大数据量基础

n 何谓大数据量

通俗的说，就是数据很多，数据量很大，记录数一般在千万级或者亿级甚至更多；存储在一个或多个服务器上；存储体量一般在TB级甚至PB级以上。

n 跟“大数据”的区别

“大数据”通常指的是对大数据量进行分析和挖掘，发掘出数据中蕴含的有意义的东西，比如：规律、趋势、喜好等等，并能据此做出一定的推理和预测，通常会涉及数据仓库、数据挖掘、人工智能等方面的东西。

n 从服务端应用的视角看大数据量

服务端应用在处理业务逻辑时，会多次操作数据，如果数据量太大，那么每次对数据进行操作会消耗大量的资源，性能也比较低下，从而导致整个应用性能下降。

从这个角度看，大数据量问题的本质就是：要操作的数据的基数太大

n 大数据量带来的问题

一个就是慢；

另一个就似乎：多次操作的叠加可能导致数据库崩溃

n 大数据量问题并不是互联网应用独有

大数据量处理的基本思路-1

n 一：分字诀

- 1: 用和不用分开，常用和不常用分开
- 2: 对数据库存放的数据：分区、分库、分表
- 3: 对文件存放的数据：拆文件
- 4: 考虑分批处理

原则就是：尽量使每次操作的数据的基数减少

n 二：合理使用缓存

n 三：数据库优化

- 1: 合理设计数据库结构
- 2: 合理构建索引
- 3: 数据库集群

大数据量处理的基本思路-2

n 四：优化算法

- 1: 优化操作数据的算法
- 2: 优化Sql
- 3: 考虑使用临时表、中间表

n 五：合理使用NoSql

- 1: 比如：Mongodb、Redis、HBase等等

n 六：使用分布式大数据处理方案

- 1: 比如：Hadoop、Spark、Storm等等