### 1、为什么要拆分数据库

单体项目在构建之初，数据库的负载和数据量都不大，所以不需要对数据库做拆分。电商业务的数据量增长飞快，所以最开始的PHP+MySQL的架构已经不能满足实际要求了，于是淘宝想到的第一个办法就是把MySQL替换成Oracle。但是没过了多久，在08年前后，单节点的Oracle数据库也不好用了，于是淘宝终于告别了单节点数据库，开始拆分数据库。从一个节点，变成多个节点。

拆分数据库是有讲究的，比如说拆分方法有两种：垂直切分和水平切分。那你是先水平切分还是垂直切分呢？先水平切分，然后垂直切分。

### 2、垂直切分

垂直切分是根据业务来拆分数据库，同一类业务数据表拆分到一个独立的数据库，另一类的数据表拆分到其他数据库。

比如说一个新零售的电商数据库，我们可以吧跟商品相关的数据表拆分为一个数据库，然后在这些数据表的基础之上，构建出商品系统。比如用Java语言或者php语言，创建出一个商城系统。然后把跟进销存相关的数据表拆分到另一个数据库上，再用程序构建出仓库系统。



垂直切分解决了什么问题：

垂直切分可以降低单节点数据库的负载。原来所有数据表都放在一个数据库节点上，无疑所有的读写请求也都发到这个MySQL上面，所以数据库的负载太高。如果把一个节点的数据库拆分成多个MySQL数据库，这样就可以有效的降低每个MySQL数据库的负载。

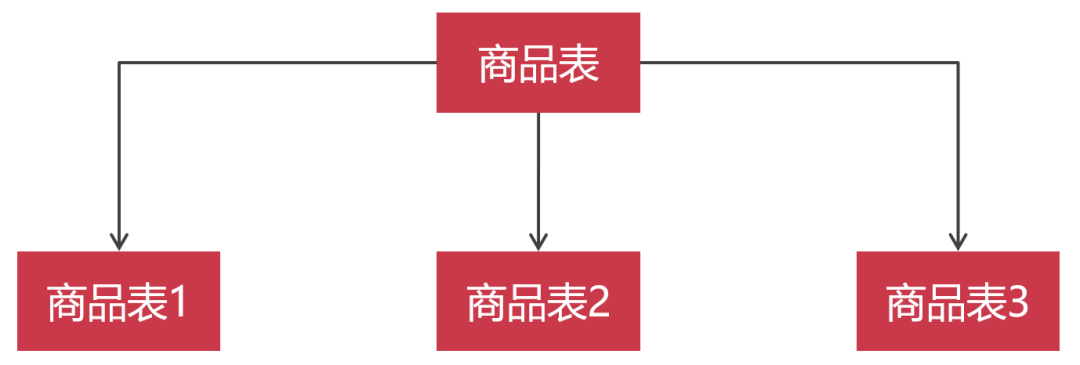
垂直拆分不能解决什么问题：

垂直切分不能解决的是缩表，比如说商品表无论划分给哪个数据库节点，商品表的记录还是那么多，不管你把数据库垂直拆分的有多细致，每个数据表里面的数据量是没有变化的。

MySQL单表记录超过2000万，读写性能会下降的很快，因此说垂直切分并不能起到缩表的效果。

### 3、水平切分

水平切分是按照某个字段的某种规则，把数据切分到多张数据表。一张数据表化整为零，拆分成多张数据表，这样就可以起到缩表的效果了。



很多人，都会水平切分存在误解，以为水平切分出来的数据表必须保存在不同的MySQL节点上。其实水平切分出来的数据表也可以保存在一个MySQL节点上面。不是水平切分一定需要多个MySQL节点。为什么这么说呢？

许多人不知道MySQL自带一种数据分区的技术，可以把一张表的数据，按照特殊规则，切分存储在不同的目录下。如果我们给Linux主机挂载了多块硬盘，我们完全可以利用MySQL分区技术，把一张表的数据切分存储在多个硬盘上。这样就由原来一块硬盘有限的IO能力，升级成了多个磁盘增强型的IO。

水平切分的用途：

水平切分可以把数据切分到多张数据表，可以起到缩表的作用。

但是也不是所有的数据表都要做水平切分。数据量较大的数据表才需要做数据切分，比如说电商系统中的，用户表、商品表、产品表、地址表、订单表等等。

水平切分的缺点：

不同数据表的切分规则并不一致，要根据实际业务来确定。所以我们在选择数据库中间件产品的时候，就要选择切分规则丰富的产品。常见的数据库中间件有：MyCat、Atlas、ProxySQL等等。数据库中间件的作用相当于SQL语句的路由器。你家路由器硬件配置不怎么高，但是不影响你享用百兆宽带。MyCat也是一个道理，它仅仅是起到SQL语句转发的作用，并不会实际执行SQL语句。我推荐使用MyCat最主要的原因是它自带了非常多的数据切分规则，我们可以按照主键求模切分数据，可以按照主键范围切分数据，还可以按照日期切分数据等等。

水平切分的另一个缺点就是扩容比较麻烦，日积月累，分片迟早有不够用的时候。这时候不是首先选择增加新的集群分片。因为一个MySQL分片，需要4~8个MySQL节点（最小规模），增加一个分片的投入成本是很高的。所以正确的做法是做冷热数据分离，定期对分片中的数据归档。把过期的业务数据，从分片中转移到归档库。目前来说数据压缩比最高的MySQL引擎是TokuDB，而且带着事物的写入速度是InnoDB引擎的6-14倍。用TokuDB作为归档数据库最适合不过。



### 4、为什么先做水平切分，后做垂直切分

随着数据量的增加，最先应该做的是①数据分片，利用多块硬盘来增大数据IO能力和存储空间，这么做的成本是最低的。几块硬盘的钱就能收获不错的IO性能。

进入到下一个阶段，数据量继续增大，这时候我们应该②把数据切分到多个MySQL节点上，用MyCat管理数据切分。当然还要做数据的读写分离等等，这里不展开讨论。在后台做水平切分的同时，业务系统也可以引入负载均衡、分布式架构等等。理论上，使用了冷热数据分离之后，水平切分这种方式可以继续维持很长一段时间，数据量再大也不怕，定期归档就好了。

数据库到了水平切分的阶段，数据量的增加已经不是更改架构设计的主要原因了。反而这个阶段业务系统承受不住了，如果再不对系统做模块拆分，业务系统也撑不下去了，所以按照模块和业务，把一个系统拆分成若干子系统。若干子系统之间，数据相对独立。比如淘宝不会跟支付支付宝分享全部数据，共享同一套数据表，这也影响各自业务的发展。所以就要弄垂直切分了，把数据表归类，拆分成若干个数据库系统。

讲到这里，你仔细想想。如果过早的对数据库做了垂直切分，势必要重新构建若干独立的业务系统，工作量太巨大。水平切分并不需要业务系统做大幅度的修改，因此说应该先从水平切分开始做。