10亿订单的分库分表

1.订单数据的划分

订单数据分为热数据和冷数据

热数据:3个月内的订单数据,查询实时性较高

冷数据:之前的订单,查询频率不高

个别用户需要查看一年前的数据,可以走离线数据查看.

热数据:使用mysql存储,

冷数据(一年内):存储在ES中,搜索引擎可以较快查询.(EC属于副本存储,存储三个数,通过一定的规则进行存储,通过矩阵的形式存储,可以通过相互计算获取原有数据)

冷数据(一年后):存放在Hive中

2.分库分表

2.1 按业务拆分

将用户,商品,订单分为三个库(涉及到三个库的具体联系???????????)

2.2 分表策略

分表策略需要一个不可重复的id shard key(片键),

单个库分配了100张订单表通过取模计算出来.订单在那张子表中

order\_id %100,确定表的位置

注意问题在查询的时候,只能通过片键(shard key)查询,如果需要通过某个实体查询,需要将更改实体的id设为shard key.

2.3 分库策略

分表策略能够解决数据量的效率问题.,无法给数据库的并发操作,带来效率上的提高.,分表实际上还是在一个数据库进行操作,很容易烧到数据库io性能的影响.

分库和分标相似,通过驱魔的方式进行路由

2.4分库分表策略

扩展系统并发能力和查询性能.

不能简单的使用片键进行取模操作,需要一个中间变量打散到不同的子包

中间变量:shard\_key%( 库数量\* 单个库的表数量)

库序号:取整(中间变量/单个库的表数量)

表序号:中间变量%单个库的表数量



0代表索引,

计算感觉这个计算公式不对