[宽表和窄表的建设该如何选择？](http://www.cnblogs.com/super-d2/p/3384938.html)

参考：<https://www.cnblogs.com/super-d2/p/3384938.html>

**宽表和窄表的建设该如何选择？**

这个问题相信纠结了很多从是数据库开发、数据仓库开发和后台开发人员；单单考虑这个问题，难给出一个绝对的答案；本人从事数据仓库开发工作到现在已经有一年半时间了，对于这个问题，我也曾经纠结过，但是是否有绝对的答案呢？事实上任何东西都没有绝对的说法。

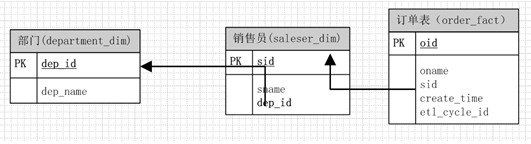
**考虑这样的一个问题，一个公司有这样的一个需求：**

设计销售领域的订单事实表，该事实表应该包含哪些维度和度量？事实表和维表该分别如何去设计？

好了，我们把关键信息拿出来，首先我们要有维度包括：销售员、销售员所属部门、下订单的时间；度量：销售量；

那么，订单事实表，其实就是一个商品销售的清单；

依照这个思路，我们建立的第一个模型可能是以下这样的：



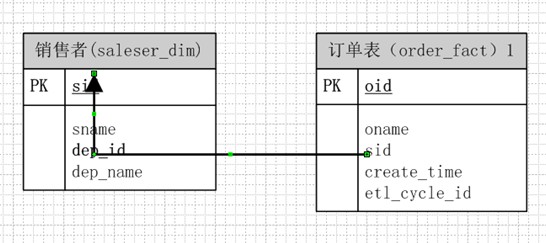
单单看上去，貌似是符合我们的问题的需要，而且符合数据库的范式设计：没有冗余字段；但是情况真的就是这样吗？

答案是否定的，确实对于一般的OLTP系统而言这样的表设计确实减少了冗余，增删改查等操作也很方便，但是往往对于我们的统计系统、OLAP、数据挖掘而言，情况却并非如此，举个例子：我们要统计每个部门各自的销售量为多少？那么对于上表,sql是这样的：

select a.\*,b.sid into #dep\_saleser from department a,saleser\_dim b on a.dep\_id = b.dep\_id;

select count(1),a.dep\_name from #dep\_saleser a,order\_fact b on a.sid=b.sid group by a.dep\_name;

对于这么一个简单的需求已经要写两个sql去实现了，其实数据库表模型的的设计是灵活的，我们完全可以根据我们的业务去设计我们的数据表；考虑到部门和销售员可以是同属于销售者这个维度，只是他们是有上下级别关系的那么依照这个思路，我们的模型可以建立为下面这样：



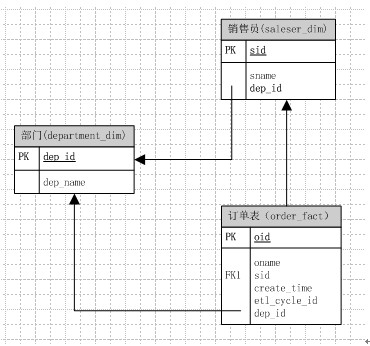
那么统计每个部门各自的销售量，可以用如下sql去实现：

select count(1),a.dep\_name from saleser\_dim a,order\_fact b

on a.sid=b.sid group by a.dep\_name;

确实对于这个模型而言，有些情况下会出现冗余（填写用户，没有填写部门；填写部门没填写用户）；但是对于提取数统计的逻辑又相对来说要简单了好多；

考虑到要实现取数简单，我们还可以想出另外一种方法：



　　看上去好像不错哦~~，取数据也就一句sql就搞定了，但是却是最最槽糕的情况，有可能一个销售员，前几天登记的部门是a，但是其实他的所属于的部门为b，那么对于上面这个模型，我们得改动销售员和订单表；而对于上面的其他两个模型都仅仅需要改动一张表就行了，造成查询数据不一致往往也就是这种数据模型所造成的。

    所谓的宽表就是字段比较多的表，包含的维度层次比较多，造成冗余也比较多，毁范式设计，但是利于取数统计，而窄表往往对于OLTP（联机事务处理过程(OLTP)）比较合适，符合范式设计原则；