

自然键、代理键、持久键

参考答案：

键：键是唯一标识一个实体的一个或者多个数据属性。在物理数据库中，键可以由变的一个或者多个列组成，它们的值唯一标识关系表中的一行。

1、持久键

数据仓库为员工编号创建一个单一键，这个单一键保持永久性不会发生变化，有时也被叫做超自然持久键，比如身份证号。

2、自然键

现实中已经存在的属性组成的键，它在业务概念中是唯一的，并具有一定的业务含义，比如员工编号。

例如，如果雇员辞职，然后重新工作，则雇员号码（自然键）可能会发生变化。

3、自然键和持久键区别

比如说公司员工离职之后又重新入职，他的自然键也就是员工编号发生了变化，但是他的持久键（身份证号）是不变的。

4、代理键

字段本身不具有业务意义，只具有主键作用，比如自动增长的ID。代理键有许多其他的称呼：无意义键、整数键、非自然键、人工键、合成键等。在数仓中多数据来源统一编码ONEID为代理键，根据系统或业务需求，人为设计的编码。

5、自然键和代理键的选择

在阿里《阿里巴巴大数据实践》中，阿里数仓建设使用的是自然键，虽然在kimball维度建模中，使用代理键作为每个维表的主键，用于处理缓慢变化维。

但一般更建议使用自然键，主要有一下几个方面的理由：

- 生成代理键难度很大，大大增加ETL人物的开发和维护成本。
- 对于缓慢变化维的处理，也是建议使用快照表的方式，虽然相比拉链表，要多使用一些空间，但是维度表数据量不会太大。而且开发和维护成本低，操作也简单，对于投入到数仓开发的新手，理解性也好。
- 实际的开发中，也很少会遇到对来自多个操作型系统的数据整合，强行使用代理键，投入产出并不高；在遇到多操作系统数据整合的场景，可具体问题具体分析，可再考虑使用代理键的方式。