

# 2019

A low-angle, upward-looking shot of a modern glass skyscraper with a grid-like facade, set against a clear blue sky. The building's reflection is visible in the lower part of the frame.

## 企业级数据仓库实战

A light gray world map is centered in the background, composed of a grid of small dots. The map is partially obscured by the text and the building image above it.



# 常见的维度表



## 缓慢变化维(SCD)

维度的属性并不是静态的，会随着时间的流逝发生缓慢的变化。这种随着时间流逝的维度，我们一般称之为缓慢变化维，并且把处理维度表的历史变化信息的问题称为处理缓慢变化维的问题。

常见的处理方式有五种：

Type1：直接覆盖。常见的场景如个人的余额

Type2：拉链表方式。通过添加记录来讲每一次变化都记录到SCD中，每条记录都有三个字段，如 start\_time, end\_time 表名有效期，设定一个active\_flag的布尔型字段，用来标识该字段是否为最新的状态

Type3：添加属性列，来记录最近变化而非全部变化

Type4：新建一个表来保存历史记录。即为Type1和Type2的混合

Type5：即为Type1/Type2/Type3的混合

在实际的工作中，如果使用的是大数据(比如Hive)，有时候也会使用分区表每日一个快照的方式来保留这些维表的历史状态，一般以天为单位来进行保留。由于这些快照是以天为单位，一天之内发生的变化就不发记录下来了。



## 退化维

退化维是指只有编号，但是不需要单独的维表的维度，比如订单编号、支付编号等，属于一种特殊的维度



## 微型维度

微型维度是指从大型维表中提取出来的常用的维度，独立成一个维表的维度，这类型的维度和具体的业务紧密结合。比如，我们统计用户常用的应用下载列表TOP10，只需要维护一张日常使用的应用维表即可，而全量的应用维表可能是数以千万计的





## 杂项维度

杂项为是指一些非常规的维表，其数据量很小，不用独立成维表，这些维表我们可以统一存放在一张表中。一般这类型维表的的存放至少包括三个字段信息：维度类型type，维度key，维度value。用户可以根据不同的维度type获取自己需要的维度



## 日期维度

连接到实际事实表的日期维度，使得能够对事实表，按照熟悉的日期、月份、财务周期、节假日等进行导航。常见的日期维度的值有：

1. 日期
2. 周几
3. 第几周
4. 第几季度
5. 第几个月
6. 第几年
7. 节假日
8. ...

日期维度往往需要手工提前维护，有些字段例如周数，是存在不同的算法，所以如果有统一的日期维度，就可以很好的进行统一

THANK YOU FOR YOUR GUIDANCE.

谢谢