

维度重塑

数据重塑

数据重塑是指将数据从一种结构化形式转换为另一种结构化形式的过程，核心在于改变数据的组织方式（如行、列、索引、层级），以适应不同的分析或处理需求，同时保持数据的完整性。

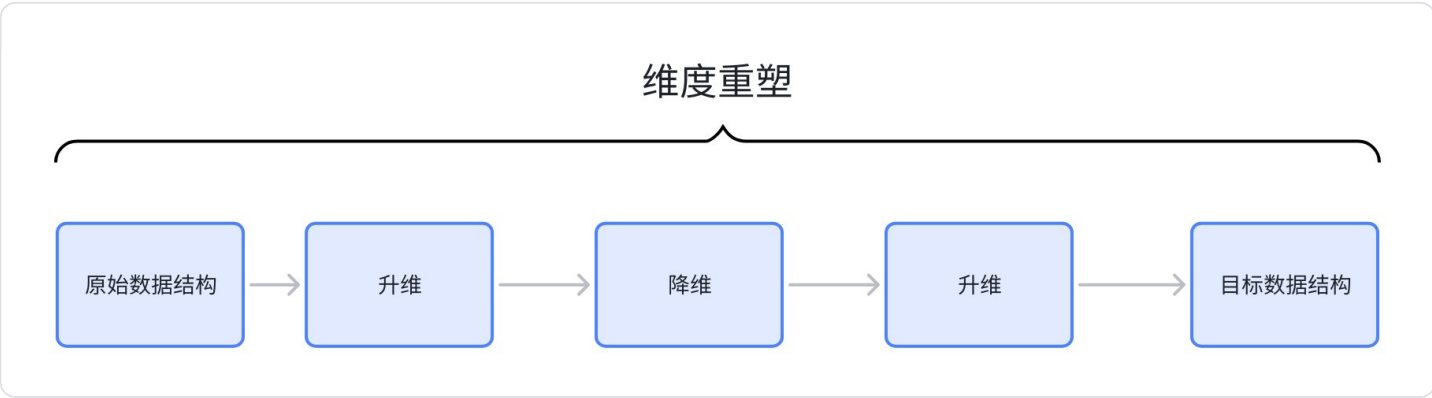
这是一个宽泛的概念，我们聚焦于 BI 领域，维度建模领域，则应该讨论“维度重塑”

维度重塑

在这里提出一种全新的维度重塑方法，其目标是简化数据可视化领域的数据处理复杂度。

Python 和 R 语言有工具已经支持了维度重塑

- 1. Python Pandas 提供了 pivot 与 melt 进行维度重塑
- 2. R tidyverse 提供了 pivot_longer 与 pivot_wider 进行维度重塑



升维与降维

1个维度: 店名
2个指标: 西瓜、香蕉

店名	西瓜	香蕉
水果店	15	18
大果店	20	25
小果店	12	15

升维

降维

2个维度: 店名、水果类型
1个指标: 售价

店名	水果类型	售价
水果店	西瓜	15
水果店	香蕉	18
大果店	西瓜	20
大果店	香蕉	25
小果店	西瓜	12
小果店	香蕉	15

关键认知

升维、降维在精神上符合范畴论的思想（对象与态射，以及同构），但在实现上并不严格遵循范畴论。

特殊强调：

- 1. 升维时，会“凭空”创建不存在的“指标名称”与“指标值”信息
- 2. 降维时，会“移除”数据中存在的“指标名称”与“指标值”信息

升维可以完整的转换数据，但维度列名称会出现空值，因此支持填补额外的信息。

降维会丢失信息内容，因此需要额外的保存转换信息，才能达到真正意义上的同构转换，否则信息一定会丢失。

分组升维与降维

1个维度: 店名
4个指标: 西瓜、哈密瓜、水晶葡萄、茉莉葡萄

店名	西瓜	哈密瓜	水晶葡萄	茉莉葡萄
水果店	15	18	18	25
大果店	13	15	13	20

分组升维

分组降维

分组1:
2个维度: 店名、瓜类型
1个指标: 瓜价格

店名	瓜类型	瓜价格
水果店	西瓜	15
水果店	哈密瓜	18
大果店	西瓜	13
大果店	哈密瓜	15

分组2:
2个维度: 店名、葡萄类型
1个指标: 葡萄价格

店名	葡萄类型	葡萄价格
水果店	水晶葡萄	18
水果店	茉莉葡萄	25
大果店	水晶葡萄	13
大果店	茉莉葡萄	20

关键认知

同普通的升维和降维，有类似的 信息增加 或 信息丢失 场景。此外由于分组的引入，会产生更多的 空数据

作用意义：

- 1. 指标分组：以轻松通过分组升维，快速处理明细数据
- 2. 多组查询：通过多条 SQL 可以轻松获取到多份明细数据，它们可以按照分组降维的方式，合并为一份数据。

规律推导

升维

$C \text{ 维度} \times N \text{ 指标} \implies C + 1 \text{ 维度} \times 1 \text{ 指标}$

维度数量	指标数量	所有指标 升维	维度数量	指标数量
0	1		1	1
1	2		2	1
1	N		2	1
C	N		C+1	1

2个维度
4个指标

店名	城市	西瓜	香蕉	哈密瓜	葡萄
水果店	北京	15	18	15	22
小果店	北京	12	15	13	20

升维

3个维度
1个指标

店名	城市	水果类型	售价
水果店	北京	西瓜	15
水果店	北京	香蕉	18
水果店	北京	哈密瓜	15
水果店	北京	葡萄	22
小果店	北京	西瓜	12
小果店	北京	香蕉	15
小果店	北京	哈密瓜	13
小果店	北京	葡萄	20

关键点

- 1. 多指标升维， 指标数量变为一， 1 个指标升维后， 指标也还是 1。
- 2. 多维度升维， 维度多一个， 0 个维度也会加 1
- 3. 0个维度1个指标, 可以反复升维, 得到任意个维度和1个指标(从而一个指标, 也能画个柱状图)

降维

$C \text{ 维度} \times N \text{ 指标} \implies C - 1 \text{ 维度} \times N \text{ 指标} * Y(\text{维值数})$

维度数量	指标数量	一个维度 降维	维度数量	指标数量
1	0		0	Y
1	2		0	2*Y
C	N		C-1	N*Y

2个维度
8个指标

店名	城市	西瓜-售价	香蕉-售价	哈密瓜-售价	葡萄-售价	西瓜-成本	香蕉-成本	哈密瓜-成本	葡萄-成本
水果店	北京	15	18	15	22	10	15	13	11
小果店	北京	12	15	13	20	8	9	11	14
大果店	上海	10				10			

← 降维

3个维度
2个指标

店名	城市	水果类型	售价	成本
水果店	北京	西瓜	15	10
水果店	北京	香蕉	18	15
水果店	北京	哈密瓜	15	13
水果店	北京	葡萄	22	11
小果店	北京	西瓜	12	8
小果店	北京	香蕉	15	9
小果店	北京	哈密瓜	13	11
小果店	北京	葡萄	20	14
大果店	上海	西瓜	10	10

关键点

- 1. 多指标降维， 维值与指标会笛卡尔积， 成为新的指标
- 2. 多维度降维， 多个维度值会笛卡尔积， 成为新的维度

实战案例

当前结构:

3个维度: 年+类别+子类别

3个指标: 销售额+利润+折扣

目标结构:

2个维度(X + Group)

1个指标(Y)



第一步升维:

4个维度: 年、类别(3)、子类别(17)、指标名称(3)

1个指标: 指标值

第二步降维:

1个维度: 年

51个指标: 51个指标

第三步升维:

2个维度: 年 + 指标名称维度

1个指标: 指标值

特殊情况

维度数量	指标数量	能否绘图?	信息
0	0	✗	
1	0	✓	表格、词云
0	1	✓	表格、饼图, 所有支持1个指标任意维度的图表. 因为0个维度1个指标, 可以反复升维得到任意个维度1个指标.