

第2章 多维数据分析基础与方法

- ❖ **1** 多维数据分析基础
- ❖ **2** 多维数据分析方法
- ❖ **3** 维度表与事实表的连接
- ❖ **4** 多维数据的存储方式
- ❖ **5** 常用的服务器端分析工具
- ❖ **6** 常用的客户端分析工具
- ❖ **7** 各种工具的准备和安装
- ❖ 小结

1 多维数据分析基础

❖ 多维数据分析（OLAP）

- 概念：
 - 以海量数据为基础的复杂数据分析技术，侧重决策支持。
- 与OLTP的区别
 - 面向对象不同
 - 操作不同

1 多维数据分析基础

❖ 多维数据集（Cube）

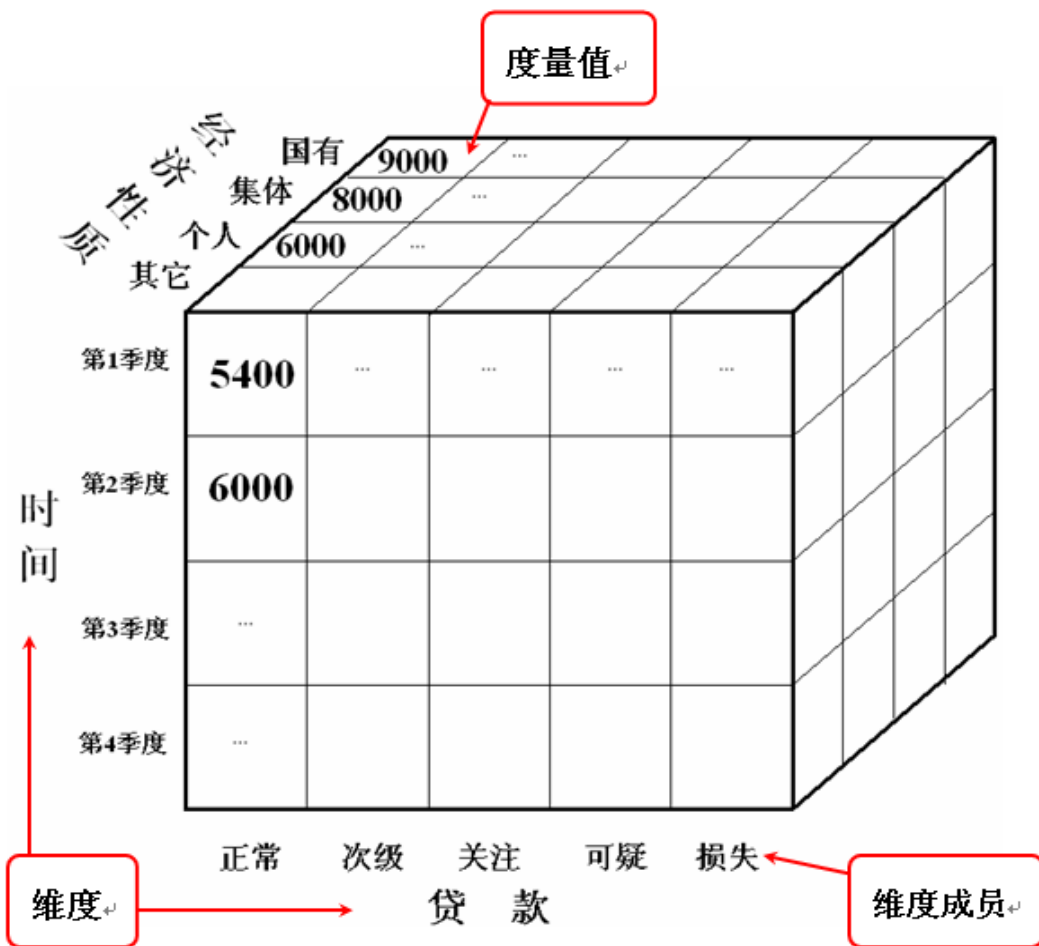
■ 概念

- 是一个数据集，通常从数据仓库的子集构造，并组织 and 汇总成一个由一组维度和度量值定义的多维结构。

■ 特性

- 多维，也称作立方体（Cube）
- 提供一种便于使用的查询数据的机制。

1 多维数据分析基础



1 多维数据分析基础

❖ 度量值 (Measure)

- 度量值是一组值，是客户发生事件或动作的事实记录。
- 如：
 - 客户打电话，记录次数和费用等；
 - 超市客户销售，记录销售数量、金额等等。
- 度量值所在的表称为事实数据表，常规多维数据集的结构中只能有一个事实数据表。

1 多维数据分析基础

❖ 维度（Dimension）

- 维度（也简称为维）是人们观察数据的角度。
- 例如，时间维，地区维等。
- 包含维度信息的表是维度表，维度表包含描述事实数据表中的事实记录的特性。
- 描述维度的字段称为维度属性

1 多维数据分析基础

❖ 维度成员（Dimension Member）

- 维的一个取值称为该维的一个维度成员（简称维成员）。
- 例如，考虑时间维具有日、月、年这3个级别，分别在日、月、年上各取一个值组合起来，就得到了时间维的一个维成员，即“某年某月某日”。

1 多维数据分析基础

❖ 维度的层次结构

■ 定义

- 层次结构是维度中成员的集合以及这些成员之间的相对位置。
- 如，地址维：洲、国家、城市、区、街道、门牌号

■ 类型

• 均衡层次结构

- 层次结构的所有分支都降至同一级别，每个成员的逻辑父代就是其上级成员。

• 不均衡层次结构

- 层次结构的所有分支都降至不同级别，每个成员的逻辑父代就是其上级成员。如CEO/部门经理（/员工）、执行秘书

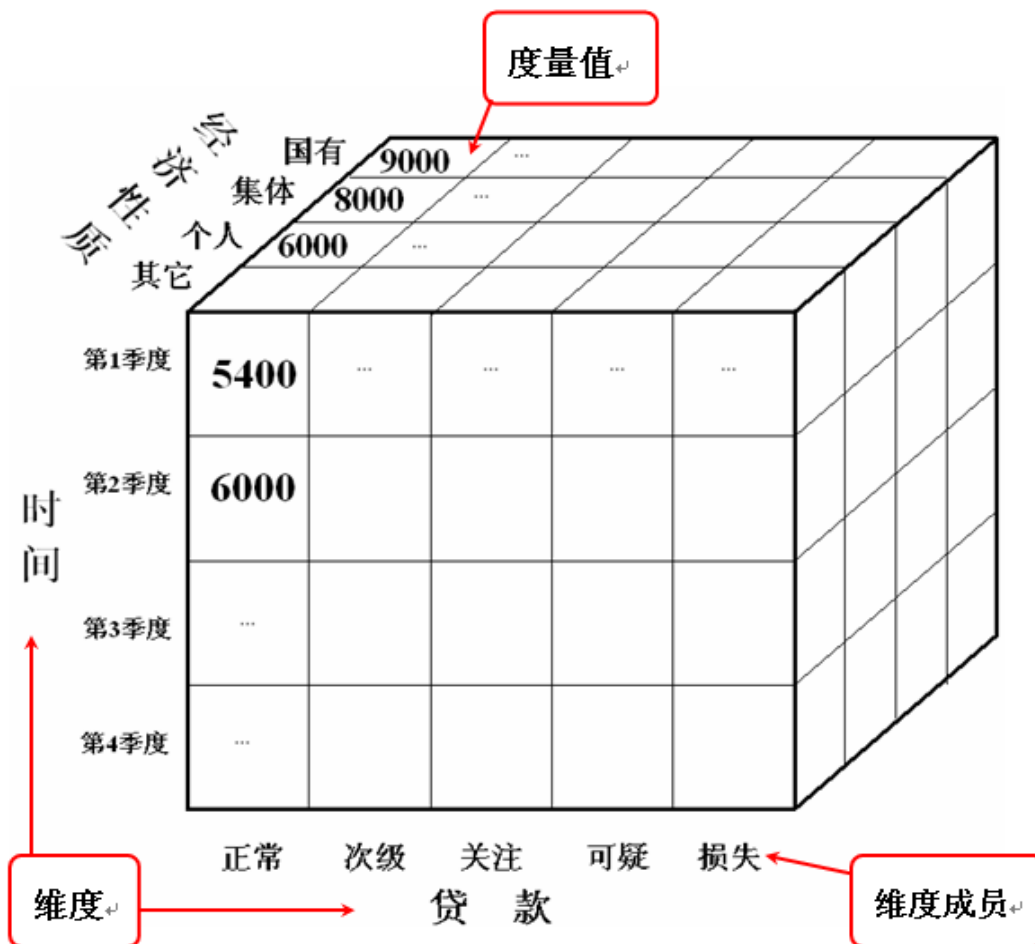
• 不规则层次结构

1 多维数据分析基础

❖ 维的级别（Dimension Level）

- 人们观察数据的某个特定角度（即某个维）还可以存在不同的细节程度，我们称这些维度的不同的细节程度为维的级别。
- 一个维往往具有多个级别。
- 例如描述时间维时（“月、季度、年”为层次结构），则月、季度、年为3个级别，可从不同级别来描述。

1 多维数据分析基础



2 多维数据分析方法

❖ 多维分析方法

■ 方法

- 上卷
- 下钻
- 切片
- 切块
- 旋转

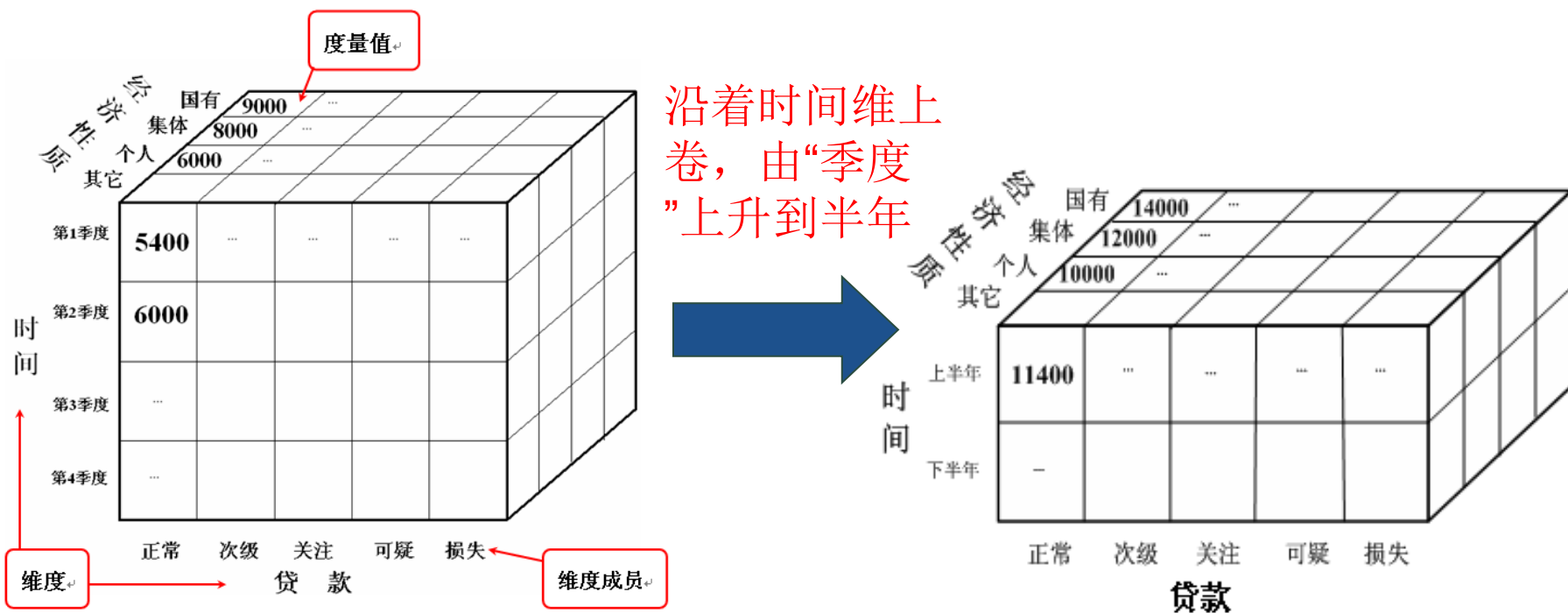
■ 目的

- 从多个角度、多个侧面观察数据库中的数据。

2 多维数据分析方法

❖ 上卷 (Roll-Up)

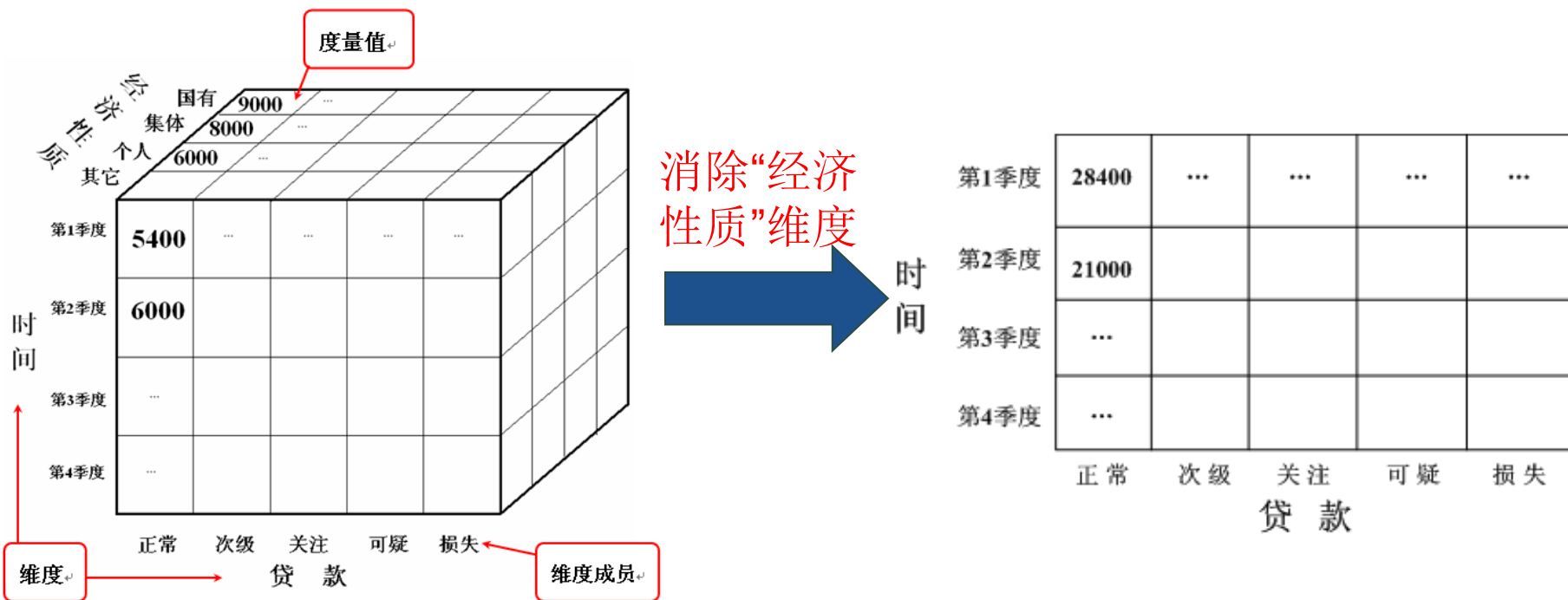
- 在数据立方体中执行聚集操作，通过在维级别中上升来观察更概括的数据。



2 多维数据分析方法

❖ 上卷

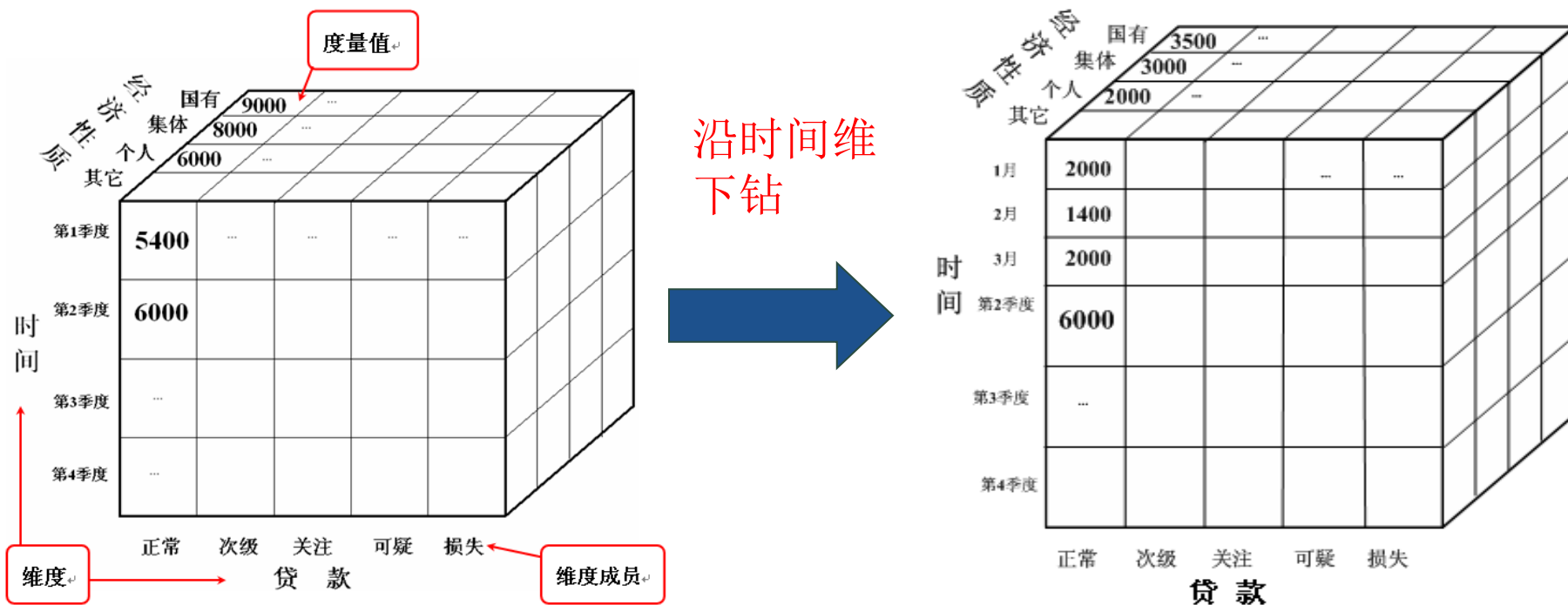
- 通过消除一个或多个维来观察更加概况的数据。



2 多维数据分析方法

❖ 下钻 (drill-down)

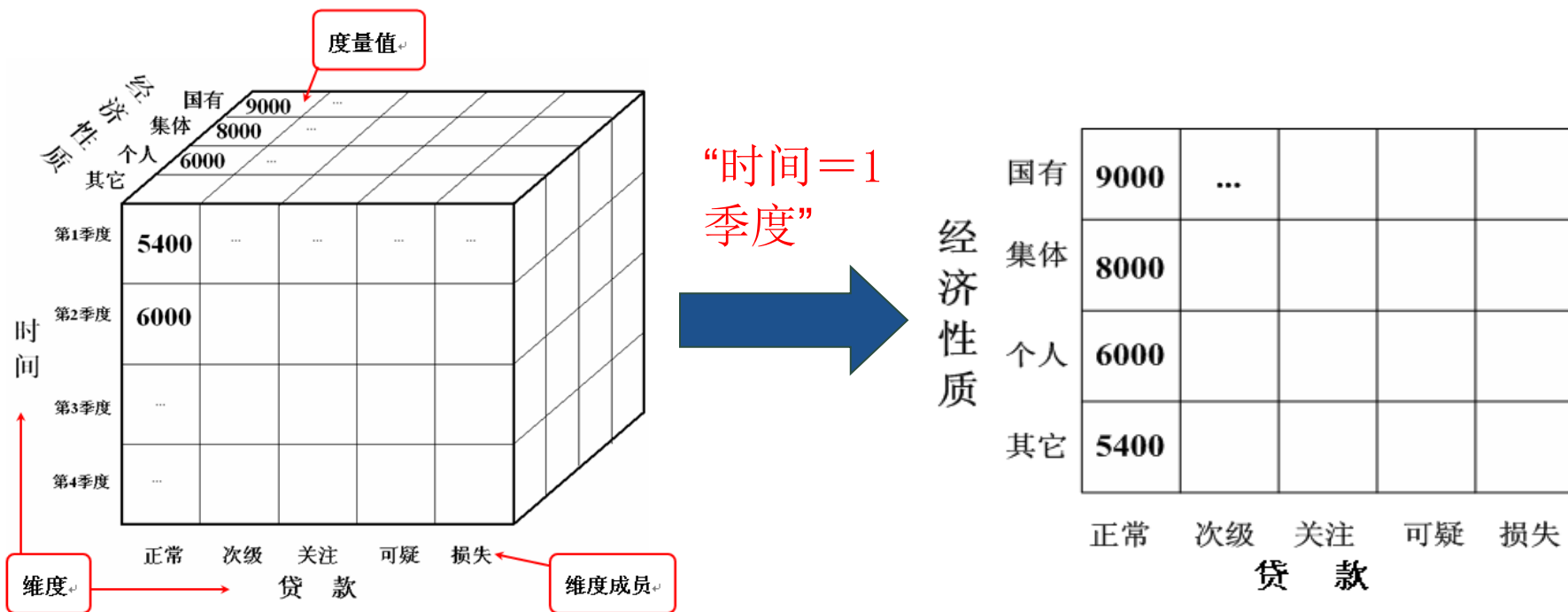
- 通过 在维级别中下降或通过引入某个或某些维来更细致的观察数据。



2 多维数据分析方法

❖ 切片 (slice)

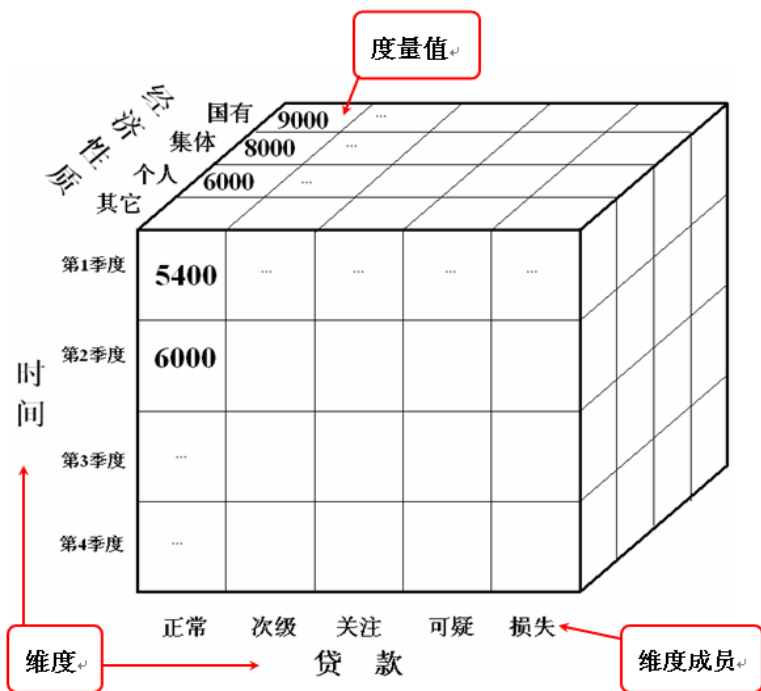
- 在给定的数据立方体的一个维上进行的选择操作。切片的结果是得到了一个二维的平面数据。



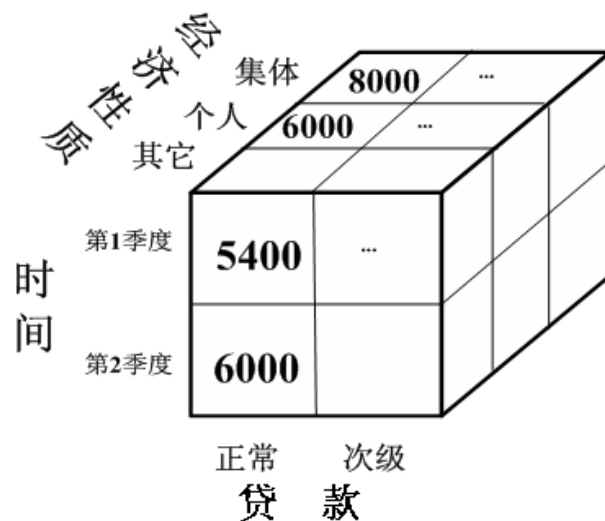
2 多维数据分析方法

❖ 切块 (dice)

- 在给定的数据立方体的两个或多个维上进行的选择操作。切块的结果是得到了一个子立方体。



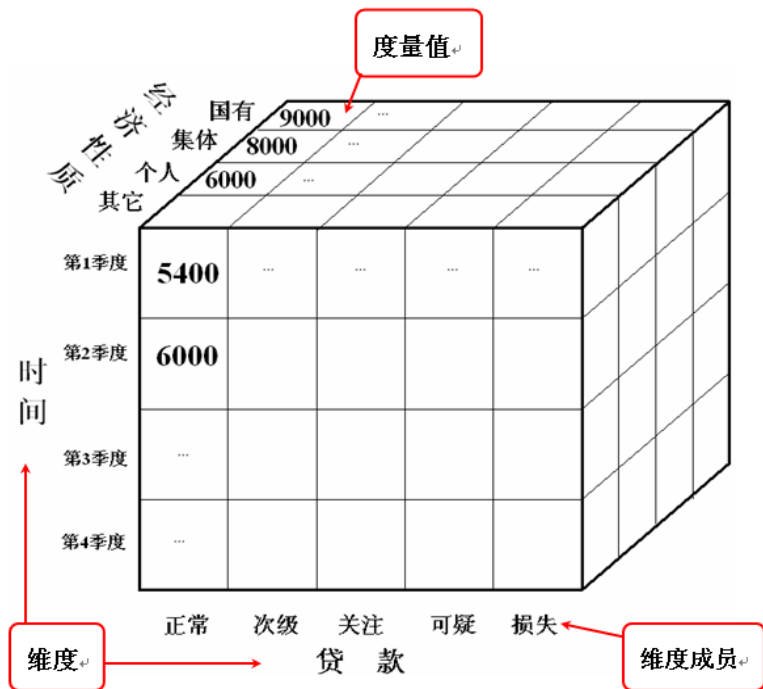
(度量值 = “正常” or “次级”) And (时间 = “1季度” or “2季度”)



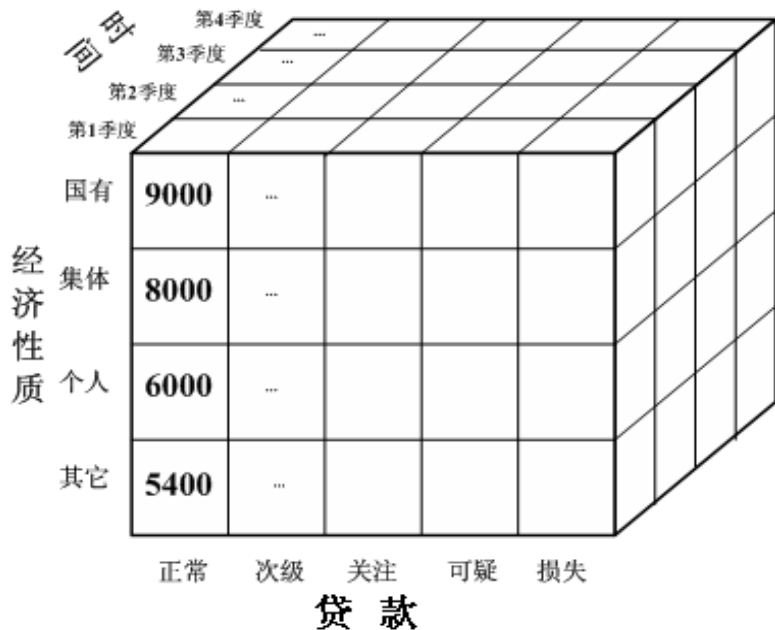
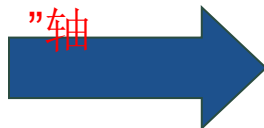
2 多维数据分析方法

❖ 转轴 (pivot or rotate)

- 转轴就是改变维的方向。



交换“时间”和“经济性质”轴



3 维度表与事实表的连接

❖ 维度表和事实表的联系

■ 维度表

- 包含某维度信息的表；
- 维度由主键和维属性构成。维属性是维表里的列。

■ 事实表

- 包含度量值（事实）的表
- 特征
 - 记录数量多
 - 事实表除了度量外，其他字段都是同维表或者中间表的关键字。
 - 如果事实相关的维度很多，则事实表的字段数也会很多。

3 维度表与事实表的连接

❖ 维度表和事实表的联系

▪ 联系

- 相互独立，又互相关联并构成一个统一的架构。
- 事实表处在架构的中央。

❖ 构建多维数据集时常用的架构：

- 星型架构
- 雪花型架构
- 星型雪花架构

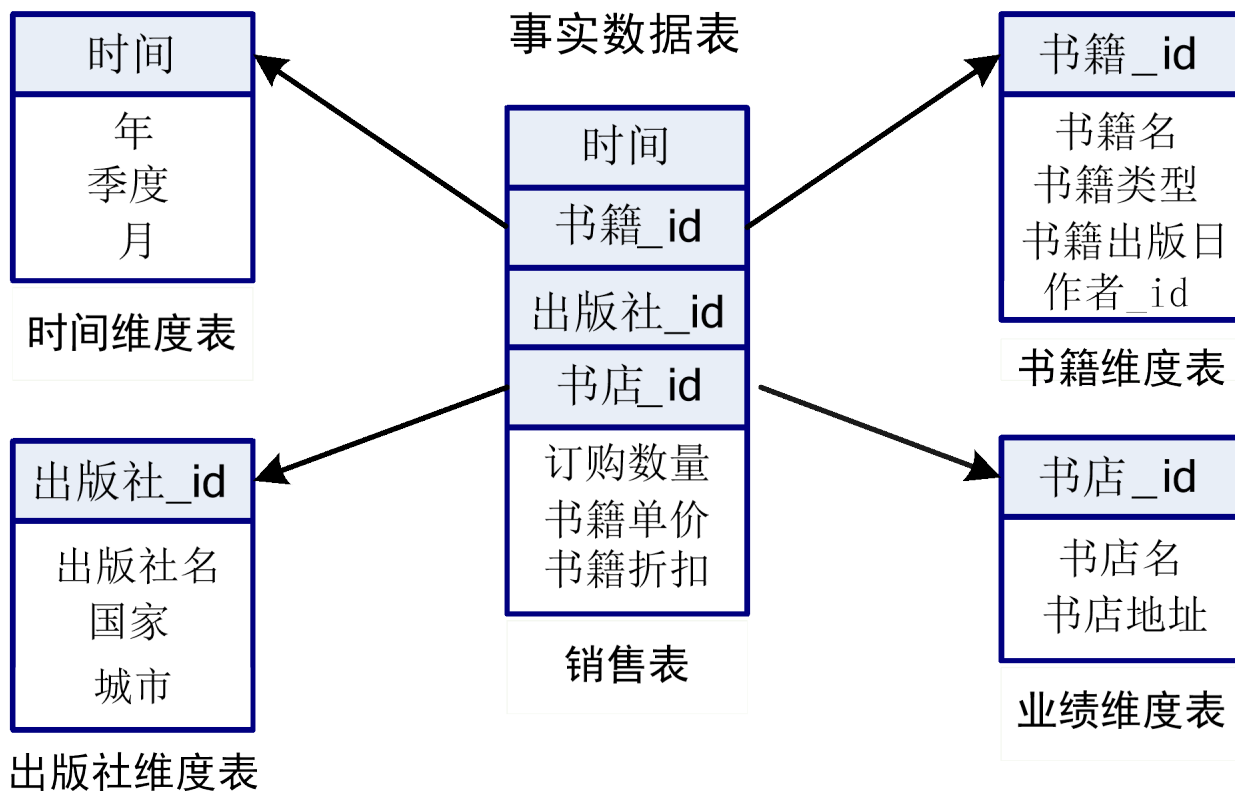
3 维度表与事实表的连接

❖ 星型架构

- 星型架构是以事实表为核心，其他的维度表围绕这个核心表呈星型状分布。
- 每个维度表中的主码都只能是单列的，同时该主码被放置在事实数据表中，作为事实数据表与维表连接的外码。
- 维度表只与事实表关联，维度表彼此之间没有任何联系。

3 维度表与事实表的连接

❖ 星型架构示意图



3 维度表与事实表的连接

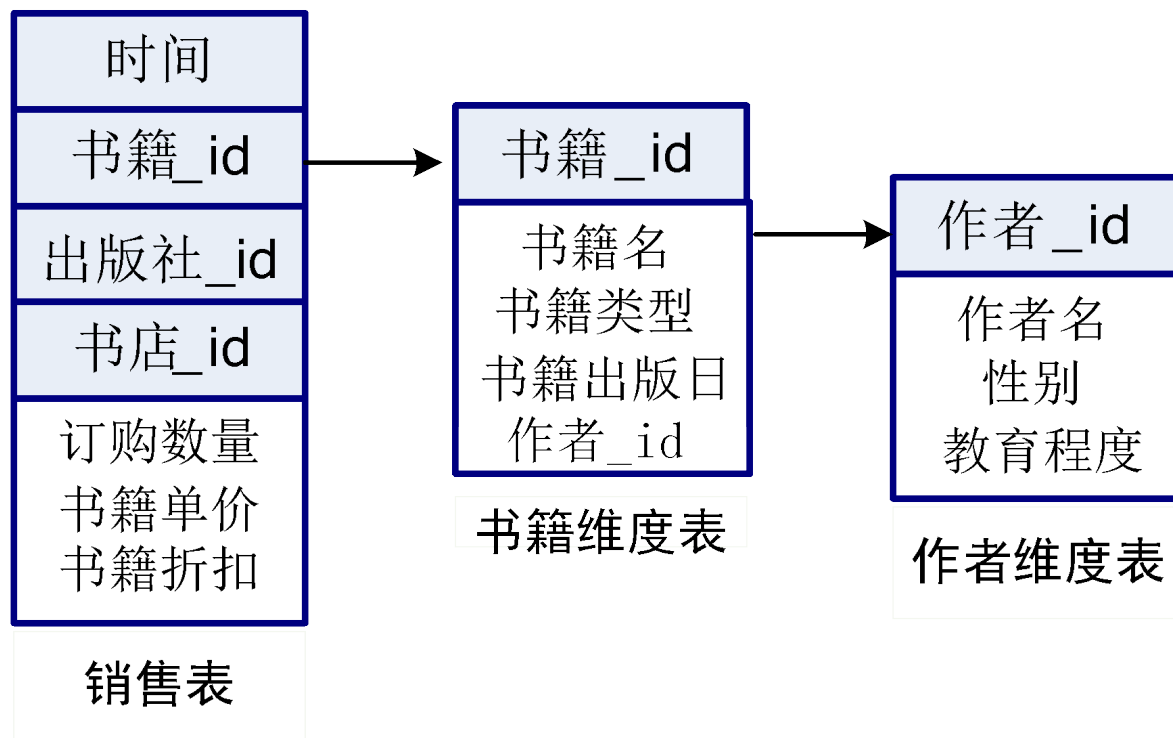
❖ 雪花型架构 (Snow Schema)

- 以事实数据表为核心；
- 某个维度表不与事实表直接关联，而是与另一个维表关联
- 维度表和与其相关联的维度表通过外码关联；
- 可以进一步细化查看数据的粒度。

3 维度表与事实表的连接

❖ 雪花型架构示意图

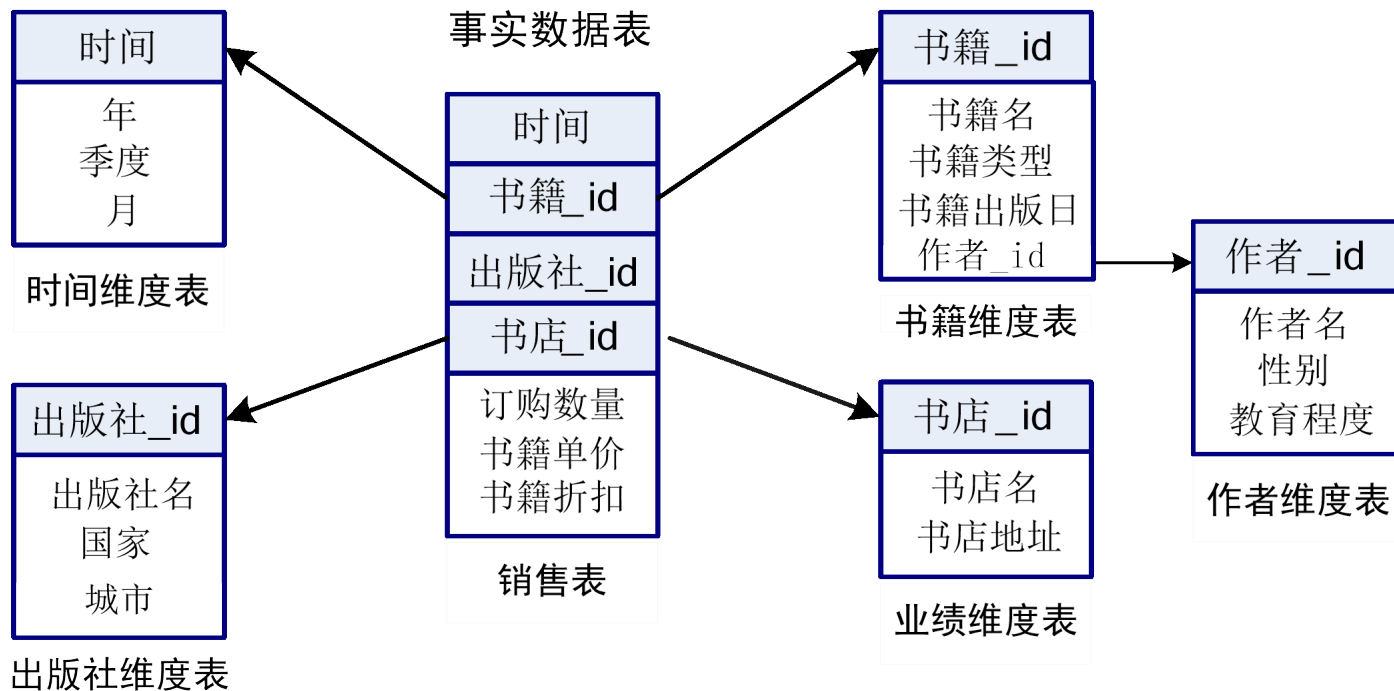
事实数据表



3 维度表与事实表的连接

❖ 星型雪花架构(Star-Snow Schema)

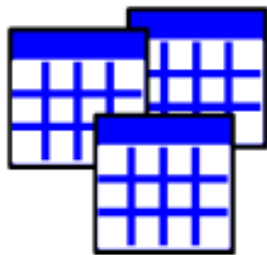
- 将星型架构和雪花式架构合并在一起使用，而成为星型雪花架构。



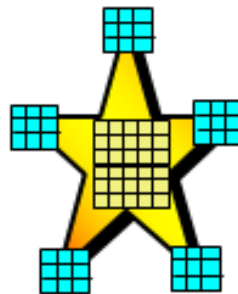
4 多维数据的存储方式

❖ SQL Server 的 Analysis 三种多维数据存储方式:

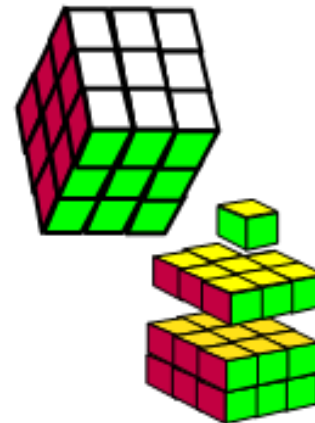
- MOLAP (多维OLAP, Multidimensional OLAP)
- ROLAP (关系OLAP, Relational OLAP)
- HCLAP (混合OLAP, Hybrid OLAP)



OLTP



ROLAP



MOLAP

4 多维数据的存储方式

❖ ROLAP

- ROLAP的数据与计算结果直接由原来的关系数据库取得。
- ROLAP将支撑多维数据的原始数据、多维数据集数据、汇总数据和维度数据都存储在现有的关系数据库中，并用独立的关系表来存放聚集数据。
- 不存储源数据副本，占用的磁盘空间最少，但存取速度也比较低。

4 多维数据的存储方式

❖ MOLAP

- **MOLAP**使用多维数组存储数据，它是一种高性能的多维数据存储格式。
- 多维数据在存储中将形成“立方体”的结构。**MOLAP**存储模式将数据与计算结果都存储在立方体结构中，并存储在分析服务器上。
- 该结构在处理维度时创建。
- 存取速度最快，查询性能最好，但占用磁盘空间较多。

4 多维数据的存储方式

❖ **HOLAP**

- **ROLAP**与**MOLAP**存储方式的结合。
- 原始数据和**ROLAP**一样存储在原来的关系数据库中，而聚合数据则以多维的形式存储在分析服务器。
- 优点：
 - 既能与关系数据库建立连接，同时又利用了多维数据库的性能优势。
- 缺点
 - 是在**ROLAP**和**MOLAP**系统之间的切换会影响它的效率。

4 多维数据的存储方式

❖ 三种存储方式的比较

内容	MOLAP	ROLAP	HOLAP
源数据的副本	有	无	无
占用分析服务器存储空间	大	小	小
使用多维数据集	小	较大	大
数据查询	快	慢	慢
聚合数据的查询	快	慢	快
使用查询频度	经常	不经常	经常

5常用的服务器端分析工具

❖ **SQL Server 2005**

- **Management Studio(数据管理)**
- **Analysis Services (分析服务)**

❖ **IBM公司的DB2 OLAP Server (OLAP服务器)。**

5.1 SQL Server Analysis Services

❖ 组件

- **SQL Server Integration Services(SSIS)**
- **SQL Server Analysis Services(SSAS)**
- **SQL Server Reporting Services(SSRS)**

❖ 平台

- **Business Intelligence Development Studio**

5.1 SQL Server Analysis Services

❖ 组件的结构



以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/765120212302011214>