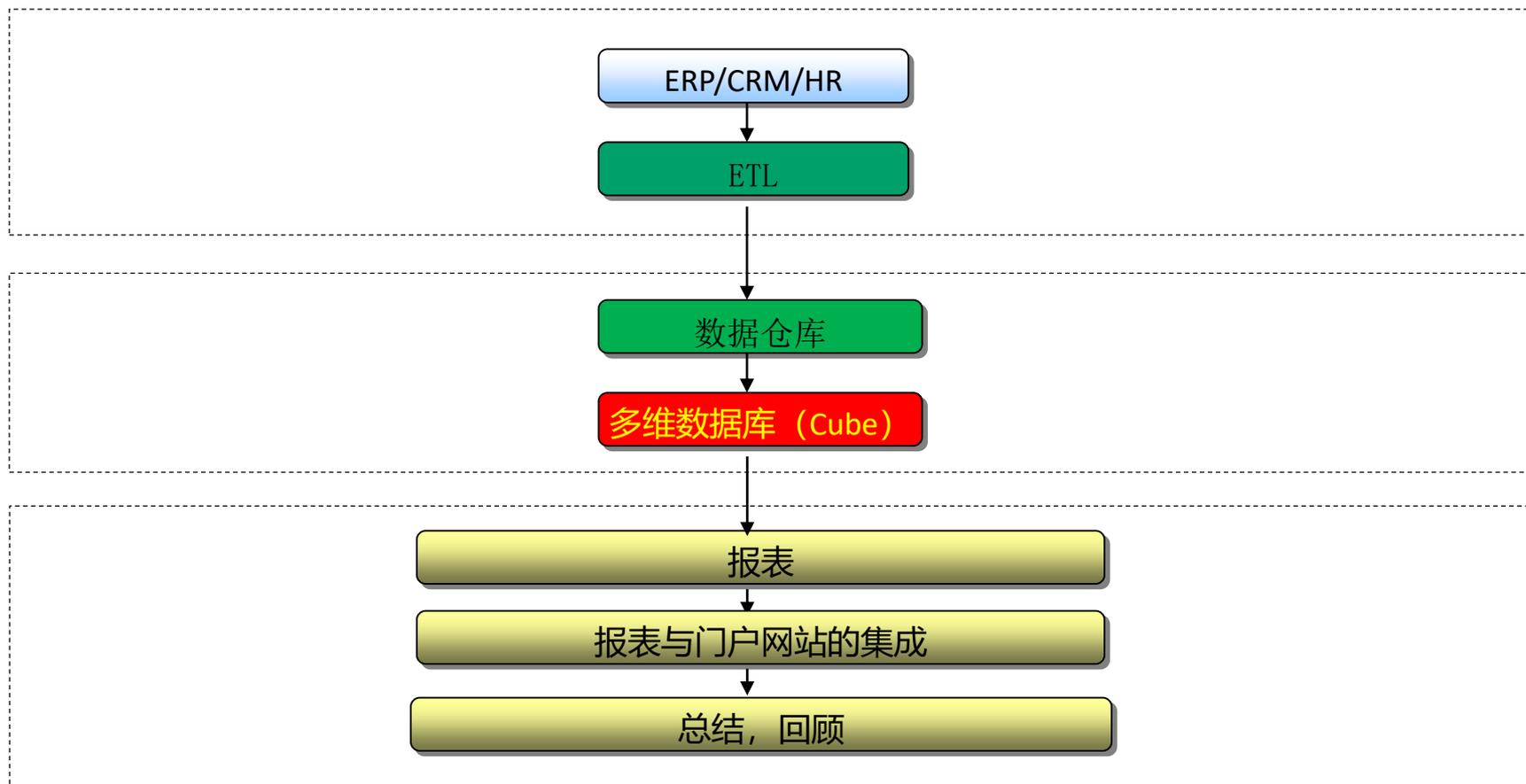




- 多维数据库在整个BI过程中所处的地位
- 课程回顾
- 本章任务
- 多维数据库概述及分析基础
- 多维数据库分析方法
- 多维数据库存储方式
- 使用 SSAS进行多维数据库开发



📁 ETL应用

📁 什么是ETL？

📁 ETL设计中需要注意哪几个步骤？

📁 ETL实现有哪几种方法？

📁 SSIS应用

📁 SSIS中有哪几种容器及各自作用？

📁 SSIS中有几种任务？

📁 SSIS中常用的控件有哪些？

- ❏ 掌握多维数据库设计与开发
 - ❏ 理解多维数据库概念
 - ❏ 使用SSAS进行多维数据库设计与开发

- ❏ 多维数据库（Multi Dimensional Database, MDD）可以简单地理解为：将数据存放在一个n维数组中，而不是像关系数据库那样以记录的形式存放。因此它存在大量稀疏矩阵，人们可以通过多维视图来观察数据。多维数据库增加了一个时间维，与关系数据库相比，它的优势在于可以提高数据处理速度，加快反应时间，提高查询效率。

- ❏ 多维数据分析是以数据库或数据仓库为基础的，其最终数据来源与OLTP一样均来自底层的数据库系统，但两者面对的用户不同，数据的特点与处理也不同。
- ❏ 多维数据分析与OLTP是两类不同的应用，OLTP面对的是操作人员和低层管理人员，多维数据分析面对的是决策人员和高层管理人员。
- ❏ OLTP是对基本数据的查询和增删改操作，它以数据库为基础，而多维数据分析更适合以数据仓库为基础的数据分析处理。

- ❏ 多维数据集由于其多维的特性通常被形象地称作立方体（Cube）
- ❏ 多维数据集是一个数据集集合，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

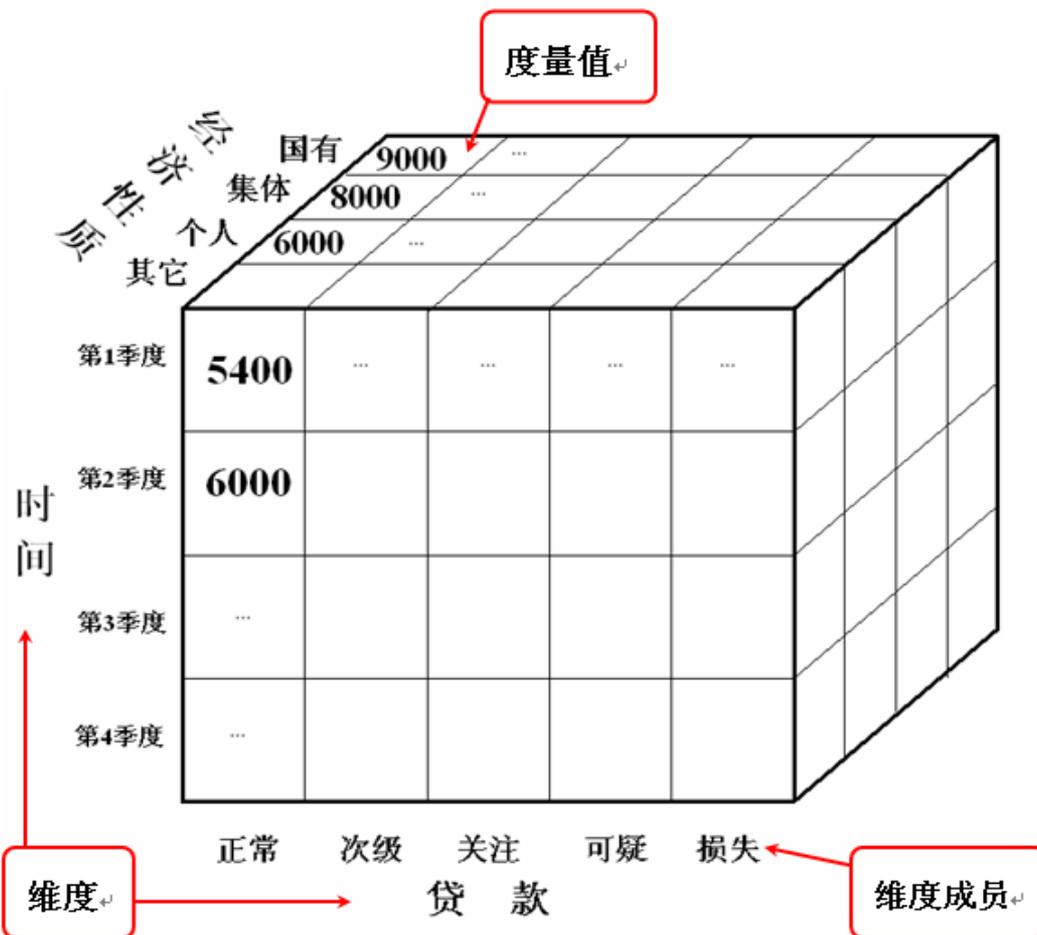
- 度量值是决策者所关心的具有实际意义的数值。
 - 例如，销售量、库存量、银行贷款金额等。
- 度量值所在的表称为事实数据表，事实数据表中存放的事实数据通常包含大量的数据行。
- 事实数据表的主要特点是包含数值数据（事实），而这些数值数据可以统计汇总以提供有关单位运作历史的信息。
- 度量值所分析的多维数据集的核心，它是最终用户浏览多维数据集时重点查看的数值数据。

- ❏ 维度（简称维）是人们观察数据的角度。
 - ❏ 例如，企业常常关心产品销售数据随时间的变化情况，这是从时间的角度来观察产品的销售，因此时间就是一个维（时间维）。
 - ❏ 例如，银行会给不同经济性质的企业贷款，比如国有、集体等，若通过企业性质的角度来分析贷款数据，那么经济性质也就成为了一个维度。
- ❏ 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。

- ❏ 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，可以从月、季度、年等不同级别来描述，那么月、季度、年等就是时间维的级别。

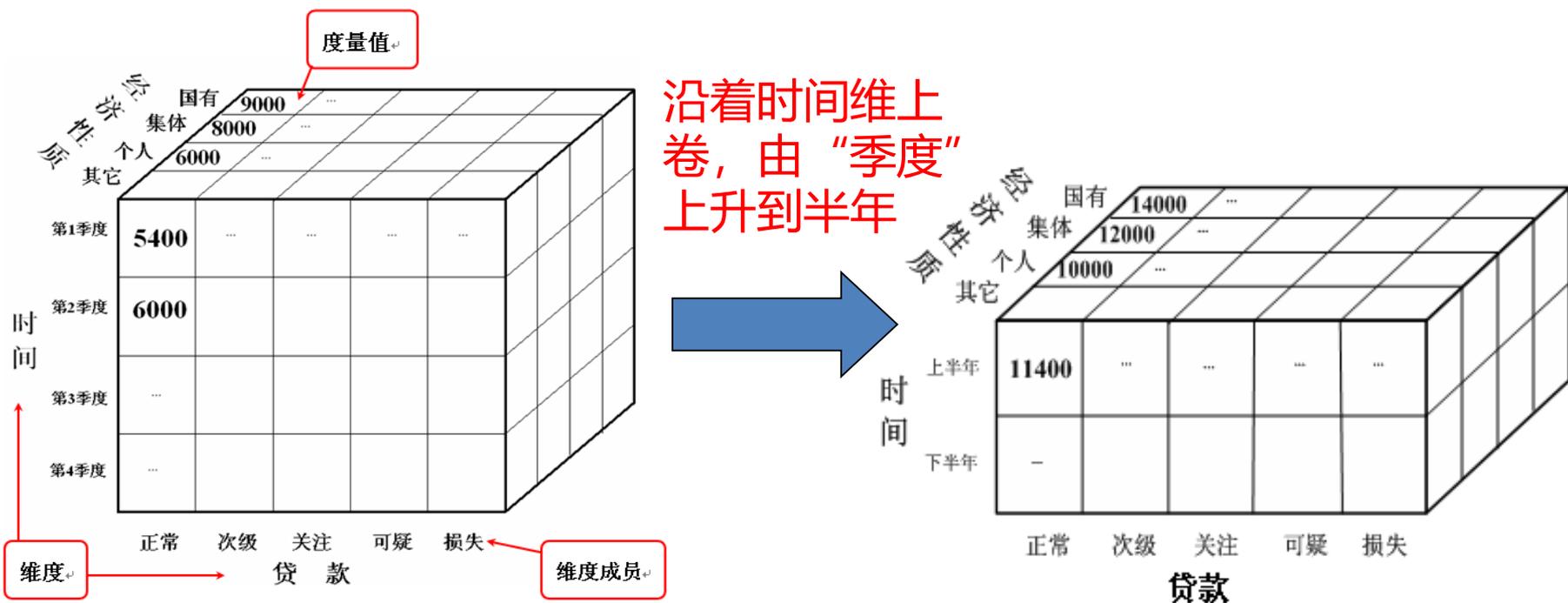
- ❏ 维的一个取值称为该维的一个维度成员（简称维成员）。
- ❏ 如果一个维是多级别的，那么该维的维度成员是在不同维级别的取值的组合。
 - ❏ 例如，考虑时间维具有日、月、年这3个级别，分别在日、月、年上各取一个值组合起来，就得到了时间维的一个维成员，即“某年某月某日”。

多维数据集示例

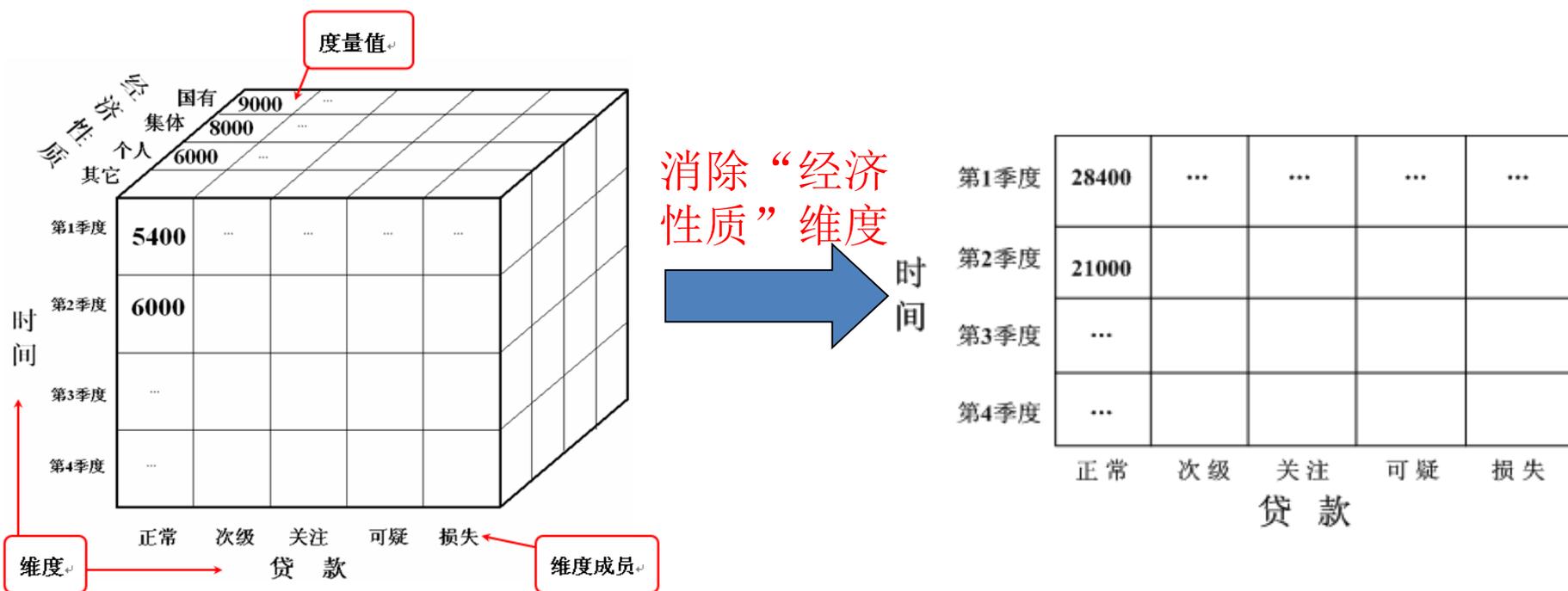


- ❏ 多维分析可以对以多维形式组织起来的数据进行上卷、下钻、切片、切块、旋转等各种分析操作，以便剖析数据，使分析者、决策者能从多个角度、多个侧面观察数据库中的数据，从而深入了解包含在数据中的信息和内涵。

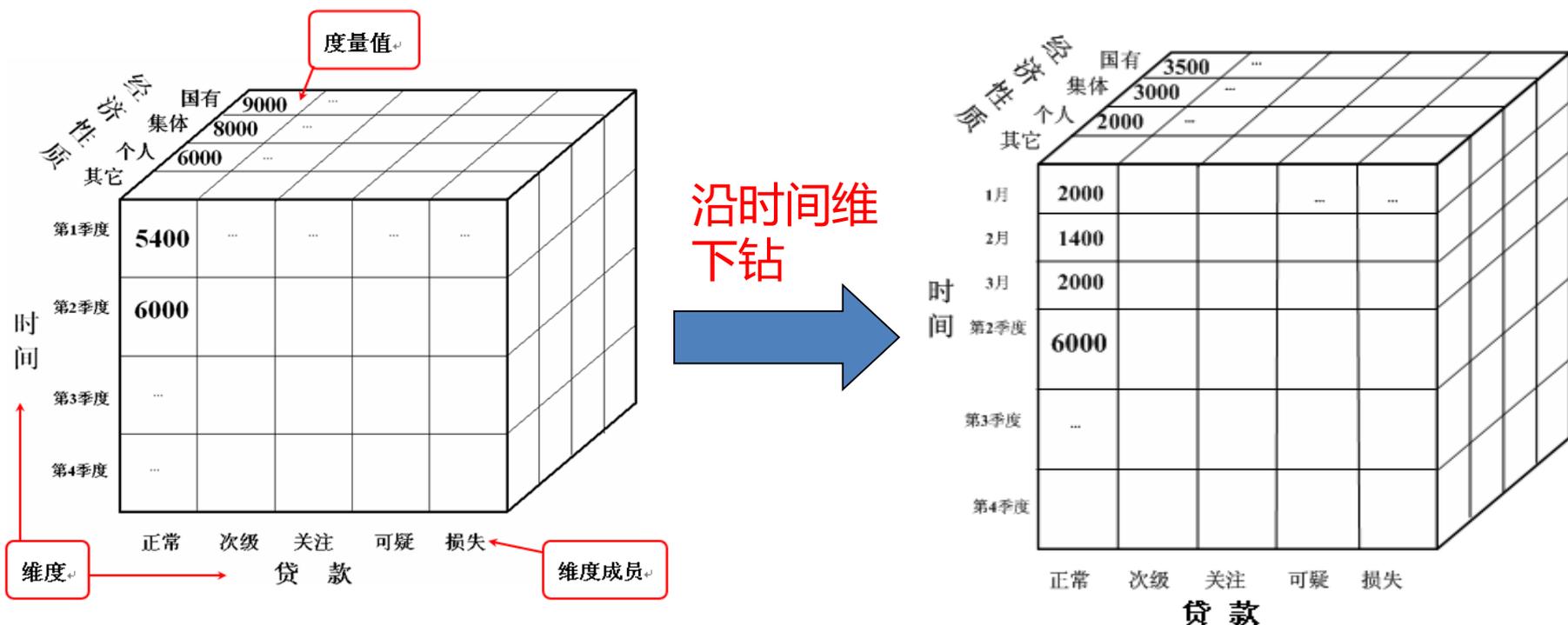
上卷是在数据立方体中执行聚集操作，通过在维级别中上升或通过消除某个或某些维来观察更概括的数据。



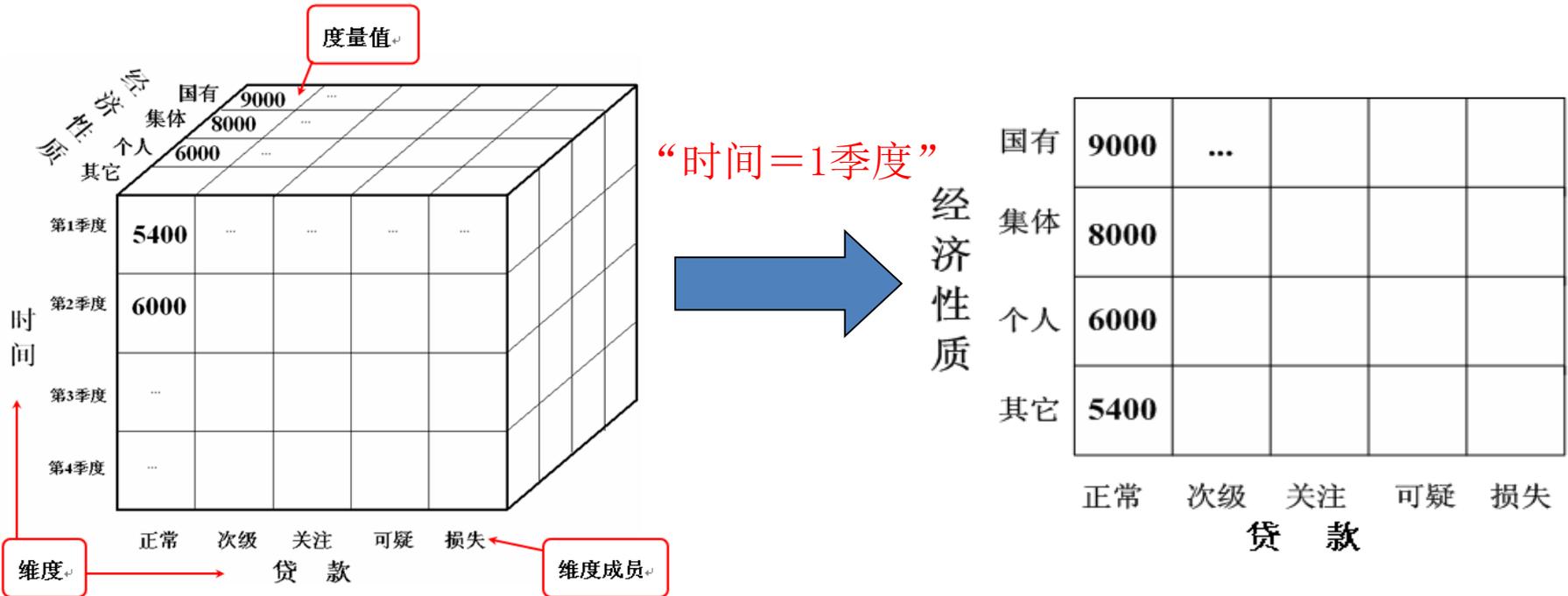
上卷的另外一种情况是通过消除一个或多个维来观察更加概况的数据。



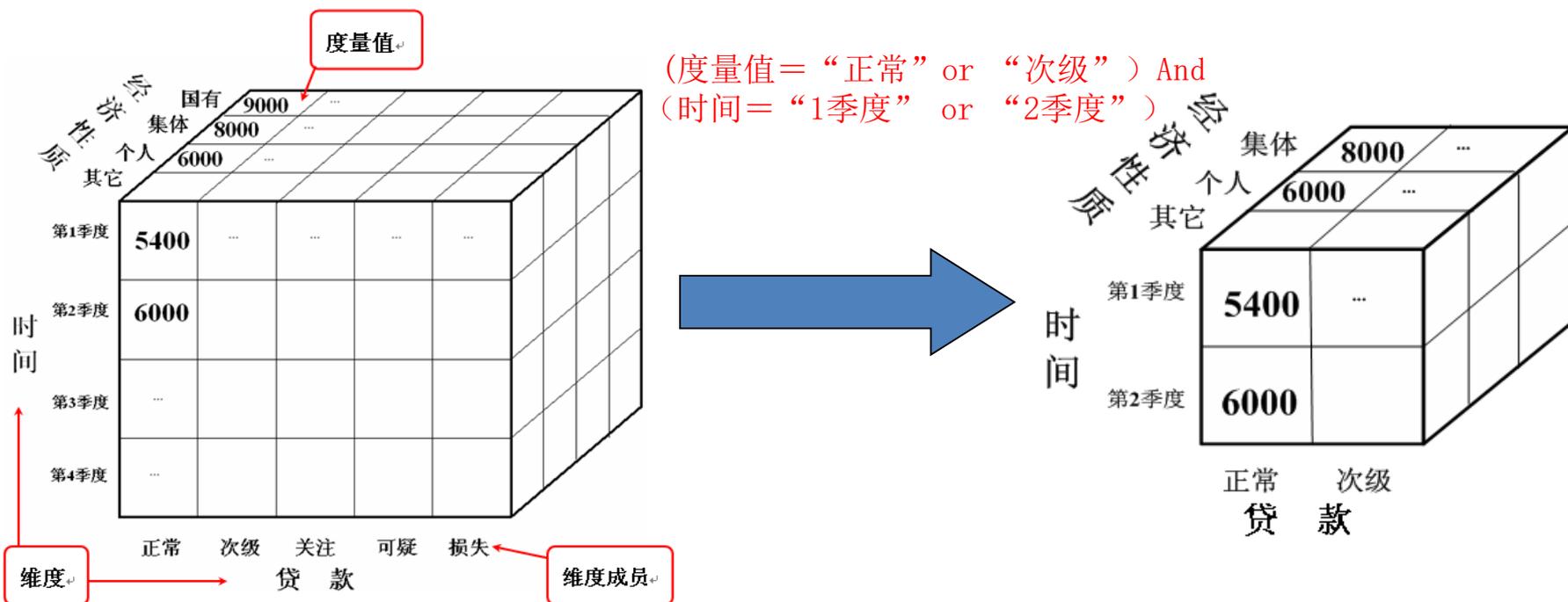
下钻是通过在维级别中下降或通过引入某个或某些维来更细致的观察数据。



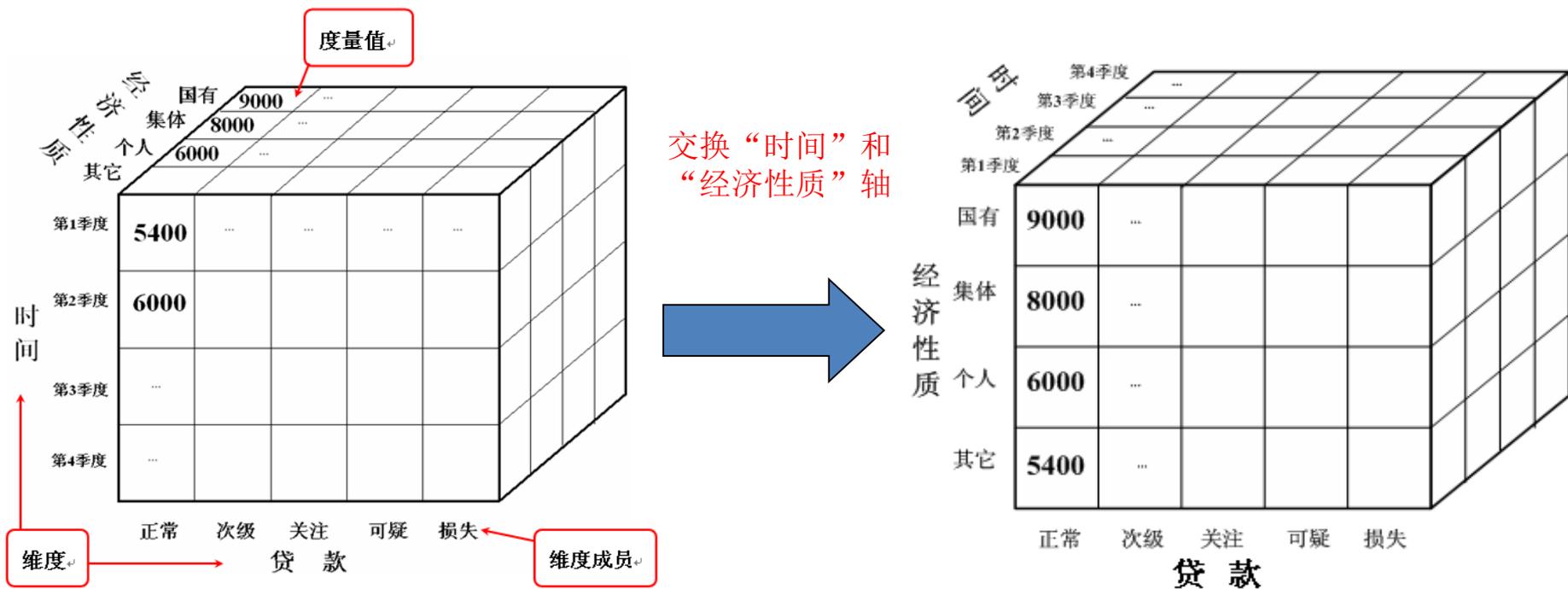
在给定的数据立方体的一个维上进行的选择操作。切片的结果是得到了一个二维的平面数据。



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。



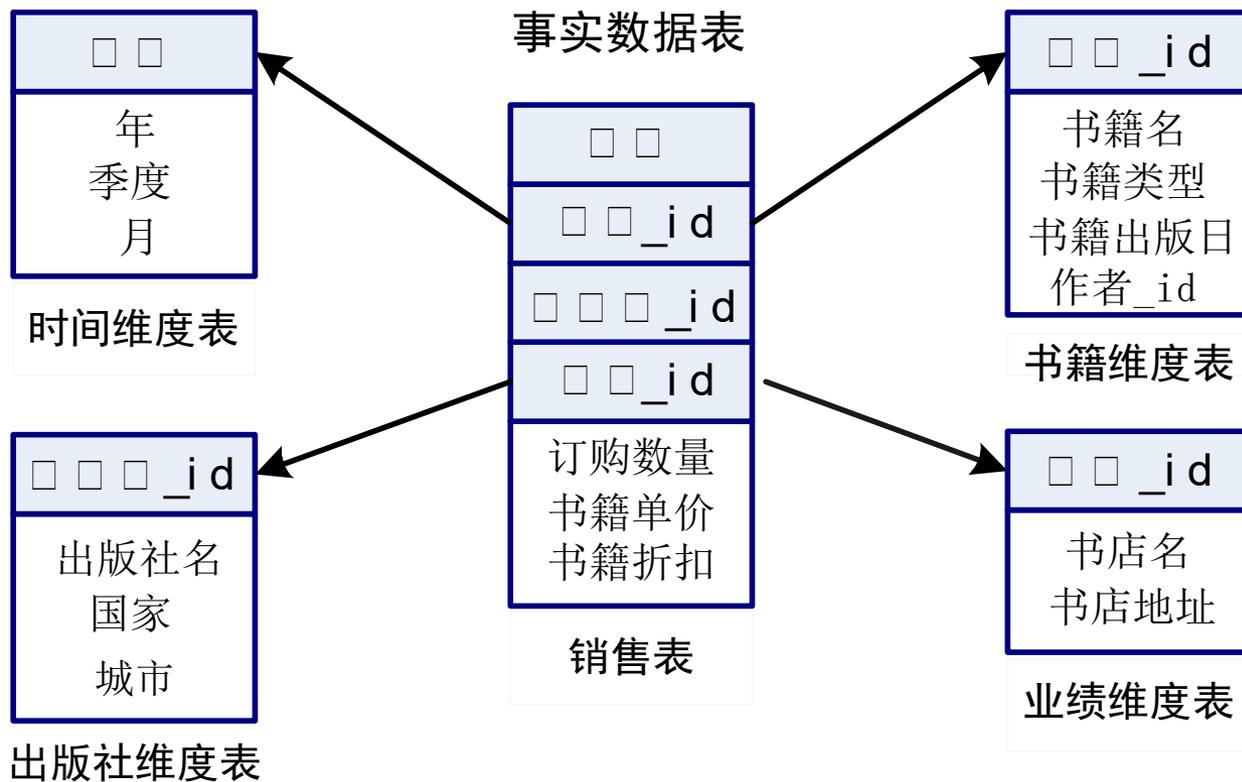
📁 转轴就是改变维的方向。



- ❏ 维度表和事实表相互独立，又互相关联并构成一个统一的架构。
- ❏ 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
- ❏ 在SQL Server 中，这些架构的中心都是一个事实数据表。

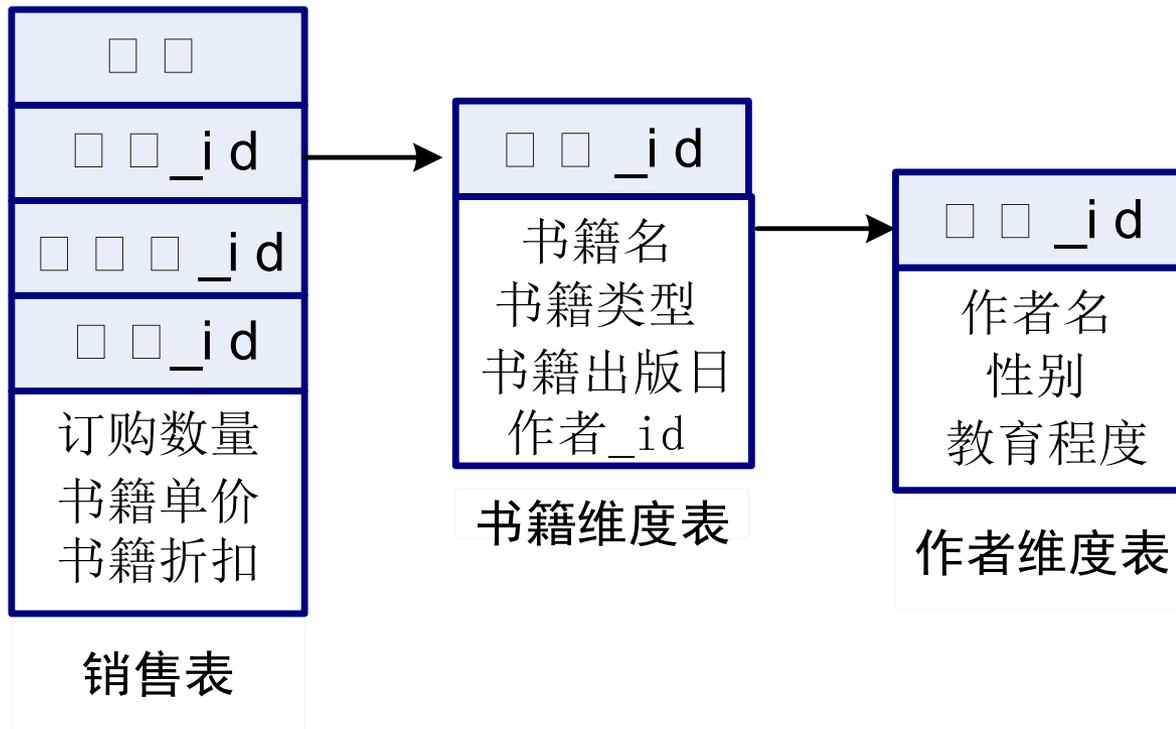
- ❏ 维度表只与事实表关联，维度表彼此之间没有任何联系。
- ❏ 每个维度表中的主码都只能是单列的，同时该主码被放置在事实数据表中，作为事实数据表与维表连接的外码。
- ❏ 星型架构是以事实表为核心，其他的维度表围绕这个核心表呈星形状分布。

星型架构示意图



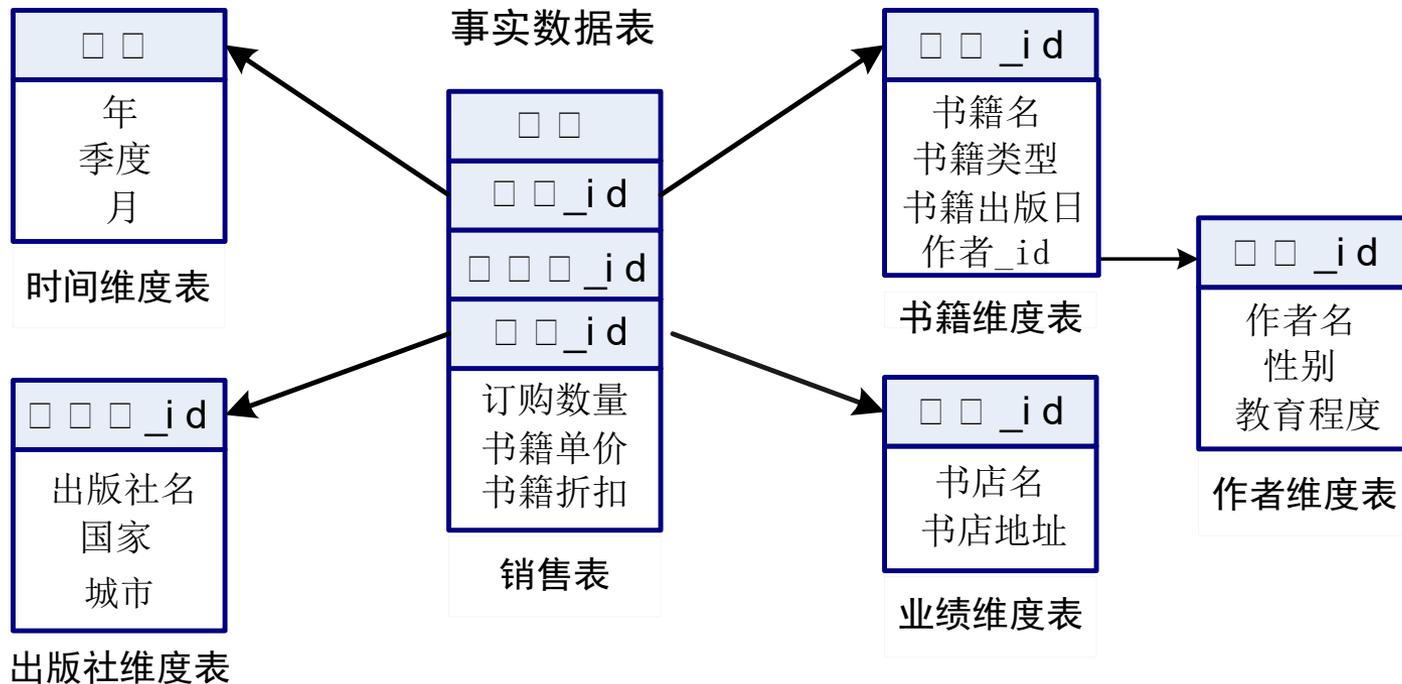
- ❏ 某个维度表不与事实表直接关联，而是与另一个维表关联。
- ❏ 可以进一步细化查看数据的粒度。
- ❏ 维度表和与其相关联的其他维度表也是靠外码关联的。
- ❏ 也以事实数据表为核心。

事实数据表



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。

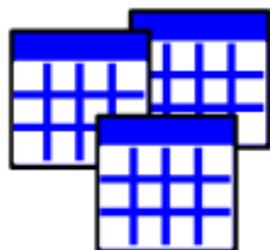


SQL Server 2008的Analysis三种多维数据存储方式：

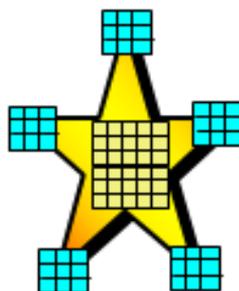
MOLAP（多维OLAP）

ROLAP（关系OLAP）

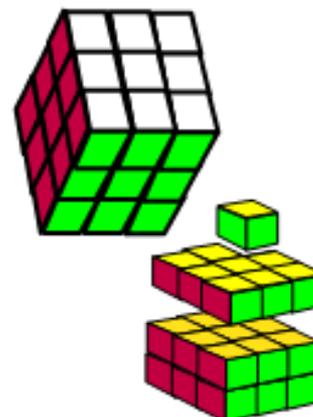
HOLAP（混合OLAP）



OLTP



ROLAP



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。

- ❏ ROLAP的数据与计算结果直接由原来的关系数据库取得，存储维度的数据以数据表形式存储在OLAP服务器上。
- ❏ 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。

- ❏ 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。
 - ❏ 这样它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。

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

- ❏ **多维数据集**是一个数据集合，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。
- ❏ **度量值**是决策者所关心的具有实际意义的数值。
- ❏ **维度**是人们观察数据的角度。
- ❏ **维的级别**是维度的不同的细节程度。
- ❏ **维度成员**是维的一个取值。
- ❏ **数据集合的常用操作**有上卷、下钻、切片、切块和转轴。
- ❏ **维度表和事实表的连接方式**主要有星型架构、雪花型架构以及星型雪花架构。
- ❏ **多维数据的存储模式**有ROLAP、MOLAP和HOLAP。