第 32 卷　第 23 期　　计　算　机　工　程　　2006 年 12 月
Vol.32　No.23　　Computer Engineering　　December 2006
·软件技术与数据库·　文章编号：1000—3428(2006)23—0076—02　文献标识码：A　中图分类号：TP311.12

# 基于层次主键模型的多维数据概念模型

严金贵[1,2]，罗　军[2]，周娜娜[2]
(1. 武警福州指挥学校训练部，福州 350000；2. 重庆大学计算机学院，重庆 400044)

**摘　要：**以关系数据库为基础，介绍了基于 E-R 模型的层次主键模型及其优点，提出了 E-R 模型到层次主键模型的转换步骤，对层次主键模型的维层次关系进行了探讨，分析了层次主键模型到多维模型的转换方法。
**关键词：**层次主键模型(HPKM)；数据仓库；概念模型；多维模型；E-R 模型

# Conceptual Model for Multidimensional Data Based on HPKM

YAN Jingui[1,2], LUO Jun[2], ZHOU Nana[2]
(1. Training Division, Fuzhou Command School of Armed Police, Fuzhou 350000; 2. College of Computer, Chongqing University, Chongqing 400044)

**【Abstract】** Based on RDB, this paper introduces the hierarchic primary key model (HPKM) based on E-R model and its strongpoints, presents the approach of the transformation from E-R model to HPKM, discusses the dimensional hiberarchy of HPKM. It analyzes the method of the transformation from HPKM to multidimensional model.
**【Key words】** Hierarchic primary key model(HPKM); Data warehouse; Conceptual model; Multidimensional model; E-R model

建立数据模型是构造数据仓库的重要步骤之一，多维数据模型是数据仓库设计中广泛采用的概念模型。数据仓库的建立应该以数据库系统为依托，不应该脱离原有的数据库系统来实现[1]。因为原有的数据库系统在设计过程中已经花费了大量的人力与时间进行系统调研、系统分析与设计，从而构造出E-R模型。数据仓库概念模型的设计如果以原有的E-R模型为基础开始进行，势必可以缩短数据仓库系统开发周期，节省开发费用。

E-R模型对所建模的现实世界具有很强的语义表述能力，但E-R模型并不能对数据仓库概念模型进行充分的描述。为了能够自然地表达出数据仓库模式具有的多维语义，近几年来，许多学者从E-R模型出发，相继提出了一些模型和方法，如ME/R[2]、构造属性树的方法[3]、StarER方法[4]等。但这些方法和模型都不是很完备，还存在诸多问题：有的要求E-R图要比较规范，有一定的限定条件；有的构造过程过于烦琐复杂，达不到形式化和规范化的要求；有的构造方法还是半自动化。

## 1 层次主键模型(HPKM)

### 1.1 HPKM 概念

E-R 数据模型能够有效和自然地模拟现实世界，抽象现实世界，在数据库设计中占有重要地位；但是由 E-R 数据模型得到的逻辑模型中各实体之间的关系像“蜘蛛网”一样繁杂，即使是相同的需求得到的 E-R 模型也是相差甚远，最重要的是 E-R 模型不能满足多维建模的需要，因此有必要对 E-R 模型进行扩展，进一步形式化 E-R 模型。分析关系数据库发现存在如下规律：

(1)关系数据库通常由若干张二维表构成，每个表都有一个主键，主键唯一标识某个实体。这是表的共同性所在，也是层次主键模型的设计基础。

(2)任何主键均由表中属性组成，可以是单个属性，也可以是属性集。属性集中属性的数量不会超过表中属性的数量。

(3)抽象意义上任何构成主键的属性都有来源表，要么来自本表，要么来自其它表。

有了以上的规律，就可以按照主键中属性的个数对表进行分类，把表分为主键只包含一个属性的表(简称一主键表)、主键包含 2 个属性的表(二主键表)、……、主键包含 N 个属性的表(N 主键表)，将所有主键具有相同属性个数的表放在同一层次，从上往下依次排列一主键表、二主键表、……、N 主键表，这样表之间便有了层次性。这样的数据模型称之为 HPKM。此外，为了讨论的需要，在第 0 层把所有主键的属性列出来，每个属性独立组成一个表(虚表)，虚表中除主键外没有其它属性，所有一主键表的主键属性均来自这些虚表。这样既保证了 HPKM 的完备性，又保证了任何构成主键的属性都有来源表这条规律的正确性。

HPKM 的示意图如图 1 所示。其中 ⟶ 表示构成主键的属性的来源。

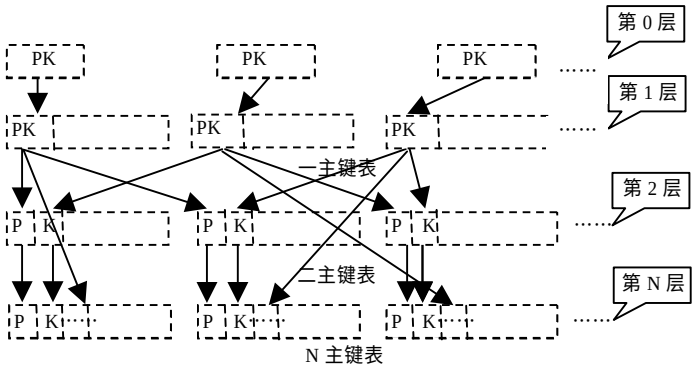


**图 1　层次主键模型**

**作者简介：**严金贵(1974－)，男，硕士、讲师，主研方向：信息管理与决策支持技术；罗　军，副教授；周娜娜，硕士
**收稿日期：**2006-07-24　**E-mail：**yanjingui327@163.com

### 1.2 E-R 模型到 HPKM 的转换

基于上面对 HPKM 的描述，可以很容易得到从 E-R 模型到 HPKM 的转换步骤：(1)列出所有表的主键属性，一个属性独立组成一个虚表，构成 HPKM 的第 0 层；(2)依次列出所有的一主键表、二主键表、……，分别构成 HPKM 的第 1 层、第 2 层、……；(3)标识所有表主键的属性来源(第 0 层表除外)，主键属性来自本表的其来源表可以从第 0 层的虚表中获得。

### 1.3 HPKM 的优点

与前面提到的其它模型相比，HPKM 具有如下优点：(1)任意 E-R 模型都可以形式化成 HPKM，没有任何限定条件；(2)构造过程直观、简单，不需要依靠直觉和经验，不存在不规范和形式化方面的不足；(3)可以自动地进行转换处理，基本可以实现自动化。

由此可见，HPKM 在 E-R 模型和多维模型之间架起了桥梁，使企业原始 E-R 模型到数据仓库多维模型得以顺利过渡，如图 2 所示。


**图 2 E-R 模型、HPKM 与多维数据模型之间的关系**

## 2 HPKM 的维层次关系

在 HPKM 中没有直接的事实表和维表概念，但是一旦用户确定了自己所关注的主题或分析的对象后，就可以在 HPKM 中找到相对应的表，这个表就相当于多维结构中的事实表；而“事实表”中主键属性的来源表就相当于多维结构中的维表。在 HPKM 中，“维”的层次关系主要是通过外键引用关系来体现的。在关系数据库中表与表之间的引用关系主要有 2 种：(1)引用被引用表的主键属性作为引用表主键属性的一部分；(2)引用被引用表的主键属性作为引用表的非主键属性。上述介绍的 HPKM 实际上只是针对第 1 种情况而建立的模型，而与第 2 种情况相对应的同样可以建立层次模型，不妨称为层次外键模型，即外键也可以层次化。其模型结构的构造过程与 HPKM 相类似。例如，描述地理的维可以从地区、省、国家等不同层次来描述，其子层次外键模型见图 3。

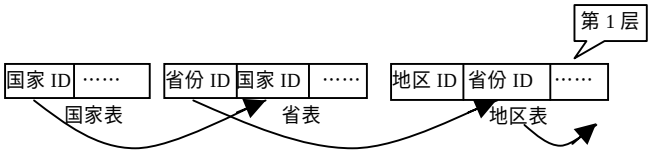


**图 3 地理维子层次外键模型**

“国家 ID”、“省份 ID”、“地区 ID”分别为国家表、省表和地区表的主键属性，地区表中非主键属性“省份 ID”引用省表的主键属性，省表中非主键属性“国家 ID”引用国家表的主键属性。在层次外键模型中，为了区别起见，用表示构成外键的属性的来源。

## 3 HPKM 到多维模型的转换

### 3.1 转换步骤

HPKM 从本质上讲是一种多维模型，只是表现形式有所不同，因此由 HPKM 转换为数据仓库广泛采用的多维模型比较自然和简单。大体上分为以下几个步骤：

(1)识别事实表。事实表是多维数据结构的核心，存储了维表关键字和决策者所关心的真正数据，是多维查询的焦点，可以根据用户需要对数据进行各种操作和统计。HPKM 中用户所关心的主题和分析的对象即是多维结构的事实 3 表。比如商品销售表、货物定购表、股票交易表等，它们是决策者理解和分析的对象，记录了商业过程的操作细节信息，并且数据量随着时间的推移而剧增。具体哪些表是事实表要对用户的需求进行系统分析后才能确定。

(2)识别维表和维层次。维表一般用来存放维的层次、成员类别等维描述信息，HPKM 中主键属性的来源表即是多维结构的维表；维层次主要是通过前面提到的外键引用来识别的，在层次外键模型中可以很容易地得到维的层次关系。

(3)添加时间维。HPKM 是从 E-R 模型转换过来的，而在 E-R 模型中的关系一般没有时间属性，即使存在时间属性，也只是记录操作的时间，并没有根据时间进行深层次的分析。比如，在一个销售系统中，库存关系存储的是当前最新数据，每进行一次产品销售就伴随着一次更新。因此，从 HPKM 转换到多维模型过程中，通过添加时间维来保留历史信息，进而对历史数据进行统计分析。此时事实表中的主键要进行相应的修改，增加一个时间维主键的属性，而时间维通过一个主键链接到事实表中与此对应的一个外键上。

(4)消除必要的层次关系。在 E-R 模型中，由于规范化的要求，对一些实体进行了分割。但在数据仓库中，主要是通过牺牲空间的方法来换取查询上的效率，因此适当的冗余是允许的，也是必要的。这样，在一个层次关系中，可以把高层实体转入到低层实体中，通过减少数据表的数目，降低查询的复杂性，提高查询速度。这个操作在多维数据建模的关键技术中，它属于反规范化的方法。

(5)构造多维数据模型。经过上述步骤处理后，可以容易地把一个 HPKM 转换为若干个用户所需要的多维模型。值得注意的是，HPKM 是对整个商业过程进行建模，而多维模型只围绕特定的商业过程或主题展开，每个多维模型只对一个商业过程建模。因此，一个 HPKM 可以分解成多个多维模型。

### 3.2 示例

图 4 是某电信公司利润子 HPKM，它希望从不同时间段、不同地区、不同品牌(固定电话、小灵通、宽带)、不同业务(通话、短信)角度来考察公司的利润情况。该模型涉及 5 个表：时间表、地理表、品牌表、业务表和利润表。前 4 个表属于一主键表，模型中只列出它们的主键属性；利润表主键有 4 个属性，分别是“时间 ID”、“地理 ID”、“品牌 ID”、“业务 ID”。

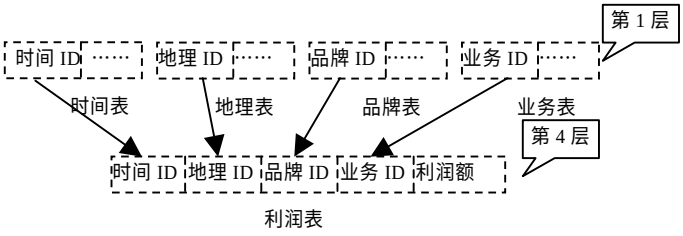


**图 4 电信公司利润子 HPKM**

按照上述转换步骤的描述，可以对电信公司利润子 HPKM 进行转换。由于星型模型是多维数据模型广为采用的一种模式，在这里只给出星型模型的转换结果。图 5 即为转换后的模型，它包含有 4 个维表(时间维、地理维、品牌维、业务维)，维表和事实表通过主外键关系建立连接。通过这种设计，决策者可以灵活地利用各个维之间的组合观察事实数据的变化和趋势，辅助分析决策。

（下转第 81 页）