



国外经典教材·计算机科学与技术

PEARSON
Prentice
Hall

Modern Data Warehousing, Mining, and Visualization Core Concepts

数据仓库、挖掘和 可视化——核心概念

(美) George M. Marakas 著

敖富江 译



清华大学出版社

国外经典教材·计算机科学与技术

数据仓库、挖掘和可视化 —— 核心概念

(美) George M. Marakas 著

敖富江 译

清华大学出版社

北 京

Simplified Chinese edition copyright © 2004 by PEARSON EDUCATION ASIA LIMITED and TSINGHUA UNIVERSITY PRESS.

Original English language title from Proprietor's edition of the Work.

Original English language title: Modern Data Warehousing, Mining, and Visualization: Core Concepts by George M. Marakas, Copyright © 2003

EISBN: 0-13-101459-5

All Rights Reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Region of Hong Kong and Macao).

本书中文简体翻译版由培生教育出版集团授权给清华大学出版社在中国境内(不包括中国香港、澳门特别行政区)出版发行

北京市版权局著作权合同登记号 图字: 01-2003-1793

版权所有, 翻印必究。举报电话: 010-62782989 13901104297 13801310933

本书封面贴有 Pearson Education(培生教育出版集团)激光防伪标签, 无标签者不得销售。

图书在版编目(CIP)数据

数据仓库、挖掘和可视化——核心概念/(美)麦瑞卡斯(Marakas,G.M.)著; 敖富江译.

—北京: 清华大学出版社, 2004.10

书名原文: Modern Data Warehousing, Mining, and Visualization: Core Concepts

(国外经典教材·计算机科学与技术)

ISBN 7-302-09261-3

I. 数… II. ①麦… ②敖… III. ①数据库系统—教材②数据采集—教材 IV. ①TP311.13 ②TP274

中国版本图书馆 CIP 数据核字(2004)第 084533 号

出版者: 清华大学出版社 地址: 北京清华大学学研大厦

<http://www.tup.com.cn> 邮编: 100084

社总机: 010-62770175 客户服务: 010-62776969

组稿编辑: 曹 康

文稿编辑: 李 阳

封面设计: 久久度文化

版式设计: 康 博

印装者: 清华大学印刷厂

发行者: 新华书店总店北京发行所

开 本: 185×260 印张: 19 字数: 394 千字

版 次: 2004 年 10 月第 1 版 2004 年 10 月第 1 次印刷

书 号: ISBN 7-302-09261-3/TP·6501

印 数: 1~5000

定 价: 36.00 元

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题, 请与清华大学出版社出版部联系调换。联系电话: (010)62770175-3103 或(010)62795704

出版说明

近年来，我国的高等教育特别是计算机学科教育，进行了一系列大的调整 and 改革，急需一批门类齐全、具有国际先进水平的计算机经典教材，以适应当前我国计算机科学的教學需要。通过使用国外先进的经典教材，可以了解并吸收国际先进的教学思想和教学方法，使我国的计算机科学教育能够跟上国际计算机教育发展的步伐，从而培育出更多具有国际水准的计算机专业人才，增强我国计算机产业的核心竞争力。为此，我们从国外知名的出版集团 Pearson 引进这套“国外经典教材·计算机科学与技术”教材。

作为全球最大的图书出版机构，Pearson 在高等教育领域有着不凡的表现，其下属的 Prentice Hall 和 Addison Wesley 出版社是全球计算机高等教育的龙头出版机构。清华大学出版社与 Pearson 出版集团长期保持着紧密友好的合作关系，这次引进的“国外经典教材·计算机科学与技术”教材大部分出自 Prentice Hall 和 Addison Wesley 两家出版社。为了组织该套教材的出版，我们在国内聘请了一批知名的专家和教授，成立了一个专门的教材编审委员会。

教材编审委员会的运作从教材的选题阶段即开始启动，各位委员根据国内外高等院校计算机科学及相关专业的现有课程体系，并结合各个专业的培养方向，从 Pearson 出版的计算机系列教材中精心挑选针对性强的题材，以保证该套教材的优秀性和领先性，避免出现“低质重复引进”或“高质消化不良”的现象。

为了保证出版质量，我们为这套教材配备了一批经验丰富的编辑、排版、校对人员，制定了更加严格的出版流程。本套教材的译者，全部来自于对应专业的高校教师或拥有相关经验的 IT 专家。每本教材的责编在翻译伊始，就定期不间断地与该书的译者进行交流与反馈。为了尽可能地保留与发扬教材原著的精华，在经过翻译、排版和传统的三审三校之后，我们还请编审委员或相关的专家教授对文稿进行审读，以最大程度地弥补和修正在前面一系列加工过程中对教材造成的误差和瑕疵。

由于时间紧迫和受全体制作人员自身能力所限，该套教材在出版过程中很可能还存在一些遗憾，欢迎广大师生来电来信批评指正。同时，也欢迎读者朋友积极向我们推荐各类优秀的国外计算机教材，共同为我国高等院校计算机教育事业贡献力量。

清华大学出版社

国外经典教材·计算机科学与技术

编审委员会

主任委员:

孙家广 清华大学教授

副主任委员:

周立柱 清华大学教授

委员（按姓氏笔画排序）:

王成山	天津大学教授
王 珊	中国人民大学教授
冯少荣	厦门大学教授
冯全源	西南交通大学教授
刘乐善	华中科技大学教授
刘腾红	中南财经政法大学教授
吉根林	南京师范大学教授
孙吉贵	吉林大学教授
阮秋琦	北京交通大学教授
何 晨	上海交通大学教授
吴百锋	复旦大学教授
李 彤	云南大学教授
杨宗源	华东师范大学教授
沈钧毅	西安交通大学教授
邵志清	华东理工大学教授
陈 纯	浙江大学教授
陈 钟	北京大学教授
陈道蓄	南京大学教授
周伯生	北京航空航天大学教授
孟祥旭	山东大学教授
姚淑珍	北京航空航天大学教授
徐佩霞	中国科学技术大学教授
徐晓飞	哈尔滨工业大学教授
秦小麟	南京航空航天大学教授
钱培德	苏州大学教授
曹元大	北京理工大学教授
龚声蓉	苏州大学教授
谢希仁	中国人民解放军理工大学教授

前 言

概念和目的

本书提供了讲授数据仓库(data warehousing)、数据挖掘(data mining)和数据可视化(data visualization)基本概念的基础。内容既着重于这些现代新兴技术的技术问题和管理问题,也强调与这些问题有关的隐含意义。

商业学院的主要目的是,从用户角度开发透彻理解这些新兴技术的教材,而可用于在商业学院中向学生讲授这个主题的课本总是从设计者的角度编写的。笔者发现,这在一定程度上是矛盾的。因此,本书特别着力于帮助学生透彻地理解数据仓库及其相关技术的价值。其内容明显地面向实际,强调在所有主题域的设计和开发之上的应用和实现。未来的管理者必须理解数据仓库的设计,但他们也必须掌握这些技术有效、关键的应用所必需的技巧,以预先估计问题描述及其相关解决方案的质量。本书采用多学科用户/管理者的方式介绍了支持 21 世纪业务处理所必需的数据仓库技术。

总而言之,使未来的管理者熟悉和掌握数据仓库部署、数据挖掘和可视化技术,是非常重要的。正是对这种重要性的信念激发笔者编写了本书。对现在和未来的管理者而言,应用以及对使用的理解比设计更为重要,现在如此,将来依然如此。

读者对象

本书适合于商业学院的所有学生,可以作为本科高年级或研究生的信息系统课程、执行或决策选修课程的教材。四年制的大学和社区学院通常提供这些课程。在学习本课程之前,学生先最好完成 MIS 程序导论、系统分析和设计以及数据库概念等课程。此外,学生学习商业课程越深入,那么与本书介绍的数据挖掘和可视化的相关程度就越高。书中的很多章节除了面向学生以外,还可用于日常管理的课程中,为数据仓库从业者提供参考。

教学要素

本书适当地利用了顶级商业学院课程通常使用的传统教学要素。书中使用图形和示例来介绍概念。每一章都包含一个介绍性的小案例,以突出该章介绍的概念。每章的

最后都包含关键概念小结、复习题和问题、参考文献和补充读物。下面简单介绍这些要素。

本章学习目标

每一章都用性能术语和操作术语提出该章的学习目标。换句话说，目标指出作为学习成果，学生应该能够理解的内容和能够完成的操作。

本章小案例

小案例都来自实际应用，它们为学生提供该章材料的参考。此外，每一个小案例都涉及到案例应用场合的关键角色，以便使单个学生或学生小组使用各种研究工具进行进一步的调查以扩展案例。

图形和表格

显然，精心设计的图表有助于学生对内容的理解。一旦有可能，每一章包含的图表不但在正文中引用，而且它们还作为对文字讨论的可视化引用出现。

描述性的插图

为了进一步解释与决策过程相关的概念，本书采用了描述性插图的技巧。使用描述性插图是为了让学生不仅弄清楚讨论的特定技巧是如何应用的，而且还能明白应用它的场合和环境。

数据挖掘和数据可视化练习

本书的一个特别有价值之处是与 Megaputer 公司的 PolyAnalyst 和 TextAnalyst 数据挖掘和数据可视化软件的捆绑。该软件包代表了数据挖掘和数据可视化系统的技术前沿。其价值为 10 000 美元，它的销售与本书无关。

关键概念

在每一章的小节后面是关键概念，它们以在该章中出现的先后顺序列出。帮助学生复习该章的内容，以准备班级讨论或测验。

复习题

每章都包含 10~20 道复习题，以帮助学生理解和记忆该章的内容。每一个问题的详细、准确答案都可以很容易地在该章中找到。

进一步讨论

每章最后的一些问题扩展了该章所提供的材料，让学生进行更深入的思考和讨论。每一个讨论题都可以用来组织班级公开讨论，其中很多问题可以扩充为个人或小组小

项目。

各章概述

第 1 章——数据挖掘、数据仓库和数据可视化概述

第 1 章概述了数据挖掘、数据仓库和数据可视化技术,为下一步详细研究打下基础。此外,还介绍了每一种技术发展的简要历史,从而使读者能更清楚地认识到这些强大的业务工具的发展。

第 2 章——数据仓库

第 2 章详细介绍现代数据仓库,解释数据仓库的目的及其相关特性。此外,还介绍了各种数据仓库配置的定义和示例,以及在数据仓库的成功设计和实现中元数据的作用。

第 3 章——数据挖掘和数据可视化

该章扩展了第 1、2 章介绍的概念,深入探讨了数据挖掘领域,引入了联机分析处理(Online Analytical Processing, OLAP)及其变种的概念。此外,该章还讨论了用于挖掘数据的技巧、它们目前的限制及其在数据可视化环境中的应用。

该章的第 2 部分主要分析数据可视化和数据挖掘。本书是这些领域的第一本教材,还包含了一个实际的数据挖掘和数据可视化商业软件包,并在内容中集成了这个软件包。

第 4 章——机器学习

人工智能和数据挖掘的最新成果——机器学习是本章的内容。该章详细介绍了决策树、模糊逻辑和语言多义性等概念,作为详细讨论人工神经网络和遗传算法的基础。在典型的数据挖掘环境中通常采用这些分析方法。该章简单介绍了这个领域的常见应用程序。此外,还提供了最常见的人工神经网络学习算法的数学推导。

第 5 章——执行信息系统

该章主要关注在现代数据环境中最常见的实现接口——EIS(Executive Information System, 执行信息系统),详细介绍了执行域以及各种使能技术在 EIS 的开发和应用程序中的应用。内容包括 EIS 技术的定义、EIS 发展的简明历史,执行级决策和决策生成器的独特属性,以及把 EIS 引入组织环境所引发的问题。

第 6 章——设计并构造数据仓库

第 6 章深入探讨了在组织数据仓库的开发中常见的进程、过程、工具和技巧。该章将使读者更好地理解与这种新颖而强大的数据存储方法相关的一些独特挑战。

第 7 章——数据挖掘、仓库和可视化的未来

该章是本书的最后一章，揭示了数据仓库、数据挖掘和数据可视化的发展趋势和未来应用，也讨论了这些技术在其他领域的进展。

相关资源

配套的 Web 站点(WWW.PRENHALL.COM/MARAKAS)为教师提供的支持材料包括非常丰富、有用的支持信息。每章都配有一个 PowerPoint 文件，用于课堂教学。本书中的复习题也可在线提供，学生可以方便地把答案直接通过电子邮件发送给教师。

对 Web 站点教师部分的访问需要一个有效的用户 ID 和口令。教师需要通过登录到 Web 站点，完成初始的教师注册过程，把自己注册为教师。完成注册过程后，注册请求将转发给销售商进行确认。如有问题，请填写本书最后所附的“教辅资料申请及教师信息反馈表”中的相关表项，我们将就本书给您提供所需的服务。一旦取得访问 Web 站点教师部分的许可，就可以下载本书的教师手册和测试项目文件。

目 录

第 1 章 数据挖掘、数据仓库和数据可视化概述	1
1.1 现代数据仓库	4
1.2 数据仓库的作用和结构	5
1.3 数据仓库可以完成的工作	7
1.3.1 快捷的信息交付	7
1.3.2 企业内/外的数据集成	8
1.3.3 从历史趋势展望未来	8
1.3.4 用新方式浏览数据的工具	8
1.3.5 免受 IS 部门资源限制	8
1.3.6 对日常工作有何意义	9
1.3.7 没有什么 是免费的	10
1.3.8 是否存在数据仓库无法完成的工作	10
1.4 数据仓库的成本	11
1.4.1 不定成本	12
1.4.2 成本合理性	12
1.4.3 没有数据仓库时如何访问信息	12
1.4.4 存在数据仓库时如何访问信息	13
1.4.5 成本的差距	14
1.4.6 多报表	15
1.4.7 普通用户和高级用户	15
1.4.8 普通用户、高级用户与成本合理性	15
1.4.9 数据集市和数据仓库	16
1.5 数据挖掘的基础	18
1.6 数据挖掘的起源	18
1.7 数据勘查和数据挖掘的方法	22
1.7.1 数据相关	23
1.7.2 数据仓库和数据挖掘/数据勘查	27
1.7.3 最大的挑战——数据容量	28
1.8 数据可视化的基础	29
1.8.1 数据可视化背后的历史	30
1.8.2 走在知识前沿的数据可视化	34
1.9 小结	34

第2章 数据仓库	39
2.1 存储、仓库和集市	41
2.1.1 数据存储	42
2.1.2 数据集市	43
2.1.3 元数据	43
2.1.4 DW 环境	43
2.1.5 DW 的特征	44
2.1.6 数据冗余问题	48
2.2 数据仓库体系结构	49
2.2.1 操作型数据库层和外部数据库层	50
2.2.2 信息访问层	50
2.2.3 数据访问层	50
2.2.4 元数据层	50
2.2.5 过程管理层	51
2.2.6 应用通信层	51
2.2.7 物理 DW 层	51
2.2.8 数据分段层	51
2.2.9 数据仓库技术	52
2.3 描述数据的数据——元数据	53
2.3.1 “抽象”的概念	53
2.3.2 数据的键	54
2.3.3 运转中的元数据	54
2.3.4 一致性——避免事实出现多个版本	55
2.3.5 同意事实的一个版本	55
2.4 元数据的提取	56
2.5 实现数据仓库	59
2.5.1 过失1：“如果创建了仓库，它们的用途就自然显现出来了”	59
2.5.2 过失2：遗漏了体系结构框架	59
2.5.3 过失3：理解用文档对假设进行说明的重要性	60
2.5.4 过失4：没有使用正确的工具	60
2.5.5 过失5：滥用生命周期	61
2.5.6 过失6：忽略了数据冲突问题	61
2.5.7 过失7：没有从失败中吸取经验教训	61
2.6 数据仓库技术	61
2.7 小结	63
第3章 数据挖掘和数据可视化	66
3.1 数据挖掘的概念	69

3.1.1	验证型分析与挖掘型分析	71
3.1.2	DM 的逐渐普及	71
3.1.3	使用 DM 进行更精确的个人预测	72
3.2	联机分析处理	72
3.2.1	MOLAP	73
3.2.2	ROLAP	74
3.3	用于数据挖掘的技术	76
3.3.1	分类	76
3.3.2	关联	77
3.3.3	序列	77
3.3.4	聚类	78
3.3.5	DM 技术	78
3.3.6	KDD 过程	79
3.3.7	新型 DM 应用	81
3.4	市场购物篮分析: DM 算法之王	82
3.4.1	市场购物篮分析的优点	83
3.4.2	市场购物篮分析的关联规则	84
3.4.3	利用虚项进行市场购物篮分析	88
3.4.4	使用这些结果	91
3.5	DM 目前的限制和面临的挑战	92
3.5.1	标识出遗漏信息	92
3.5.2	数据噪声和遗漏值	93
3.5.3	大型数据库和高维度	93
3.6	数据可视化	93
3.6.1	历史	95
3.6.2	人类视觉感知和数据可视化	95
3.6.3	地理信息系统	98
3.6.4	数据可视化技术的应用	99
3.7	Software 技术	100
3.7.1	Red Brick	100
3.7.2	Oracle	102
3.7.3	Informix	103
3.7.4	Sybase	103
3.7.5	Silicon Graphics	104
3.7.6	IBM	104
3.8	小结	105

第 4 章 机器学习	121
4.1 模糊逻辑和语义模糊性	123
4.1.1 语义模糊性	123
4.1.2 模糊逻辑的基础	124
4.1.3 模糊与概率	126
4.1.4 模糊逻辑的优缺点	126
4.2 人工神经网络	128
4.2.1 神经计算的基础	129
4.2.2 人工神经网络的训练	132
4.2.3 人工神经网络学习范式	133
4.2.4 神经计算的优点和局限性	135
4.3 遗传算法与遗传进化网络	137
4.3.1 最优理论	137
4.3.2 介绍遗传算法	138
4.3.3 遗传算法的优点和局限性	143
4.4 机器学习的应用	144
4.4.1 机器学习的应用领域	144
4.4.2 机器学习系统的未来	147
4.5 小结	148
第 5 章 执行信息系统	155
5.1 EIS 的概念	157
5.1.1 典型的 EIS 会话	158
5.1.2 EIS 无法完成的工作	159
5.2 EIS 历史	159
5.3 高级执行官与众不同的原因	160
5.3.1 执行官的信息需求	160
5.3.2 执行信息的类型和源	163
5.4 EIS 组件	165
5.4.1 硬件组件	166
5.4.2 软件组件	167
5.4.3 目前的 EIS 技术	168
5.5 使 EIS 开始工作	169
5.5.1 EIS 开发框架	169
5.5.2 一些需要避免的 EIS 局限和缺陷	172
5.5.3 失败是无法接受的	173
5.6 执行官决策和 EIS 的未来	174
5.6.1 转化条件	174

5.6.2 未来的 EIS	175
5.7 小结	177
第 6 章 设计并构造数据仓库	182
6.1 DW 设计的企业模型方法	184
6.1.1 定义 EM	184
6.1.2 自上而下的 DW 设计	185
6.1.3 自下而上的 DW 设计	185
6.1.4 企业集成的概念	185
6.2 DW 项目规划	187
6.2.1 DW 项目定义和就绪度评估	187
6.2.2 克服 DW 就绪度的不足	190
6.2.3 定义新建 DW 项目的作用域	192
6.2.4 为 DW 项目定义商业认证	193
6.2.5 使 DW 项目成为商业案例所面临的问题	193
6.2.6 经济可行性分析	195
6.2.7 经济可行性指标	199
6.2.8 开发 DW 项目规划	203
6.3 指定分析和设计工具	205
6.3.1 DW 逻辑设计的维度建模方法	205
6.3.2 设计 DW 事实表	208
6.4 DW 体系结构规范和开发	209
6.4.1 通用源	210
6.4.2 通用维	210
6.4.3 通用业务规则	210
6.4.4 通用语义	210
6.4.5 通用度量单位	211
6.4.6 DW 体系结构关键组件区域	212
6.4.7 开发 DW 技术体系结构	214
6.5 DW 项目的成功因素	215
6.6 小结	216
第 7 章 数据挖掘、数据仓库和数据可视化的未来	219
7.1 数据仓库的未来	223
7.1.1 规章约束	223
7.1.2 非结构化数据的存储	223
7.1.3 World Wide Web	224
7.2 可选存储器和数据仓库	226
7.2.1 可选存储器	226

7.2.2 使可选存储器工作	228
7.3 数据仓库的趋势	229
7.4 数据挖掘的未来	232
7.4.1 在个性化和隐私之间权衡	232
7.4.2 “信息是自然资源”的概念	235
7.4.3 大型数据集的爆炸	236
7.5 使用数据挖掘保护隐私	237
7.5.1 目前的 IDS 系统如何检测入侵	237
7.5.2 目前的 IDS 系统具有的弊端	237
7.5.3 我们在寻找什么	238
7.5.4 DM 如何提供帮助	239
7.5.5 入侵检测 DM 所面临的困难	239
7.6 影响 DM 未来的趋势	240
7.6.1 数据趋势	240
7.6.2 硬件趋势	240
7.6.3 网络趋势	240
7.6.4 业务趋势	241
7.6.5 对 DM 的未来所预期的可能情况	241
7.7 数据可视化的未来	242
7.7.1 武器性能和安全仿真	242
7.7.2 医学损伤治疗	243
7.7.3 未来将会展现的技术能力	244
7.7.4 未来的医学外伤场景展望	245
7.8 未来可视化应用的各个组件	246
7.9 小结	247

第1章 数据挖掘、数据仓库 和数据可视化概述

学习目标

- 解释开发数据仓库的目的和动机。
- 理解操作性数据存储与企业数据存储之间的差异。
- 理解数据仓库在企业技术基础结构中的作用。
- 理解数据仓库与业务数据集市之间的关系。
- 熟悉数据仓库能为企业完成的各项工作和不能完成的工作。
- 理解数据挖掘的起源。
- 熟悉与典型数据挖掘项目相关的事件顺序。
- 理解相关(correlation)这一概念。
- 熟悉与现代数据仓库相关的一些数据可视化技术的历史和应用。

数据挖掘小案例

构建 InsureCo 保险公司多维数据仓库

保险行业是数据仓库(Data Warehouse, DW)市场中一个重要的并且逐渐发展壮大的部分。保险企业面临着几种必须以多种方式分析的复杂事务。以前,考虑存储数以百万的或者数以十亿的在线访问事务是不切实际的。然而现在,随着强大的处理器的诞生,以及复杂的数据库查询软件的出现,大型的复杂数据库已经可以用于数据仓库中。同时,保险企业也承受着降低成本的巨大压力。这一行业中的成本几乎完全来自于索赔或者险损赔偿。

InsureCo 是一家大型保险企业,年平均营业收入超过 20 亿美元,并为大约 200 万客户提供汽车、房产以及个人财产保险。InsureCo 面临的困难是构建一个能够分析各类保险业务索赔的企业数据仓库,而其中一些业务可能要追溯到 15 年以前。

为了达到这个目标,需要做的第一步是通过对索赔分析、索赔处理、字段操作、欺诈和安全管理、财务以及营销等的各个未来的终端用户进行探询来收集信息。可以从每一组客户中得到他们的一些描述,例如:他们每天完成哪些日常工作,他们如何衡量自己完成的工作是否成功,以及他们认为怎样才可以更好地进行自己的业务。

探询中出现的三个主要的议题会极大地影响最终的数据仓库设计。第一，要透彻地理解自己的索赔单，客户需要看见每一项可能的事务。这就排除了呈现综合数据的需要。很多终端用户分析师需要对巨大的事务处理池进行切片和切块操作。

第二，客户需要定期地每月查看一次业务。因此，每月需要对索赔进行一次分组，并在月末将它与同年的其他月份或者与前几年的同期进行比较。但是，只为获取每月的保险金和每月的索赔费用而不断重复复杂的事务处理，这是不切实际的，也将与保存每个事务的需要发生冲突。

第三，这个设计需要满足 InsureCo 各类业务的多样性。记录汽车事故索赔的事实跟记录房产火灾损失或者住室被窃索赔的事实是完全不同的。

很多种行业里都出现了这些数据冲突，而对数据仓库的设计者来说这也是一个常见的问题。细节的事务浏览和每月的快照浏览间的冲突几乎总是需要设计者在数据仓库中准备两种表格。

InsureCo 的解决方案是通过小心地对基级索赔分析事务进行维化(dimensinalizing)，首先获取业务的事务浏览和每月快照浏览。每个索赔处理事务都能够符合图 1-1 中的星形联接模式。这个结构是典型的事务级别的数据仓库模式。

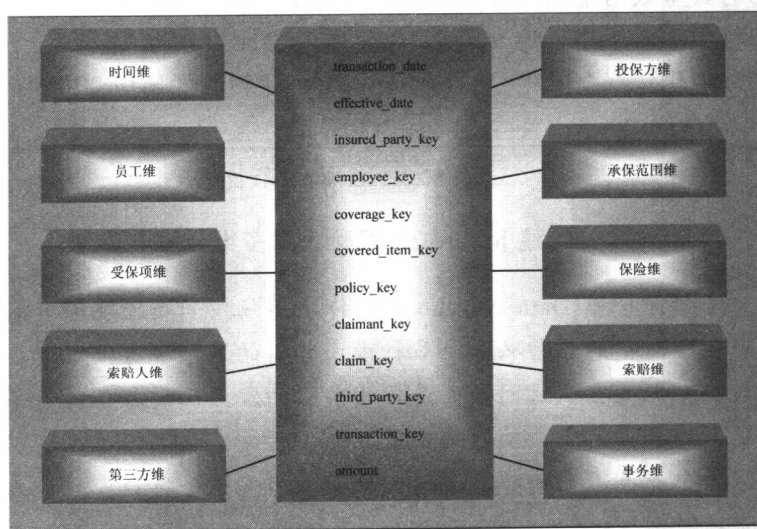


图 1-1 InsureCo 的星型联接模式

事务级别的星型联接模式为 InsureCo 分析索赔提供了一种功能非常强大的方式。索赔人的数目，索赔的时机，赔付的时机，以及第三方(如目击人和律师)的介入，都可以从数据浏览中容易地推导出来。奇怪的是，推导出到当前为止的索赔量度(如每月快照)却不太容易，因为这需要检查从第一天开始的每一个细节的事务。解决办法是向 InsureCo 的数据仓库添加数据的每月快照版。每月快照在添加更多事实的同时删除了一些维度。每月快照表非常灵活，因为可以随意地将有兴趣的汇总作为事件添加进去。