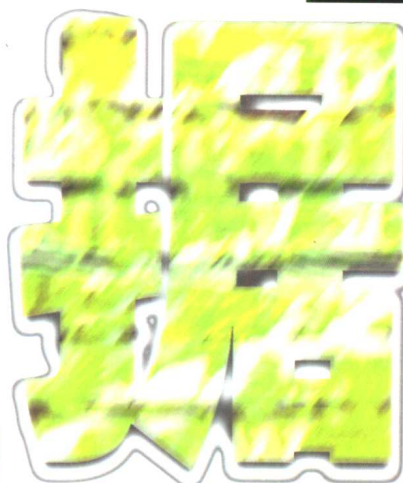




创世纪工商管理译库

主 编：廖 理

仓库

结 构 和 实 现

Architecture and Implementation

马克·汉弗莱斯 (Mark Humphries)

[美] 迈克尔·W. 霍金斯 (Michael W. Hawkins) 著

米歇尔·C. 戴 (Michelle C. Dy)

陈光群 译

哈里斯·科恩企业计算技术协会
IT专业人士解决方案

社 会 科 学 文 献 出 版 社
Social Sciences Documentation Publishing House

创世世纪工商管理译库

数据仓储

结构和实现

Data Warehousing

Architecture and Implementation

马克·汉弗莱斯 (Mark Humphries)

[美] 迈克尔·W. 霍金斯 (Michael W. Hawkins) 著

米歇尔·C. 戴 (Michelle C. Dy)

陈光群 译

数据仓储:结构和实现

·创世世纪工商管理译库·

著 者 / [美] 马克·汉弗莱斯 迈克尔·W. 霍金斯 迈克尔·C. 戴
译 者 / 陈光群

出 版 人 / 谢寿光
出 版 者 / 社会科学文献出版社
地 址 / 北京市东城区先晓胡同 10 号
邮政编码 / 100005
网 址 / <http://www.ssdph.com.cn>
责任部门 / 财经与管理图书事业部
(010)65286768
责任编辑 / 屠敏珠
文稿编辑 / 冯韵文
责任校对 / 闫晓琦
责任印制 / 同 非

总 经 销 / 社会科学文献出版社发行部
(010)65139961 65139963
经 销 / 各地书店
读者服务 / 客户服务中心
(010)65285539
法律顾问 / 北京建元律师事务所
排 版 / 北京中文天地文化艺术有限公司
印 刷 / 北京智力达印刷有限公司

开 本 / 787×1092 毫米 1/16 开
印 张 / 18.75
字 数 / 343 千字
版 次 / 2003 年 12 月第 1 版
印 次 / 2003 年 12 月第 1 次印刷

书 号 / ISBN 7-80190-055-3/F·024
著作权合同 / 图字 01-2002-6260
登 记 号
定 价 / 35.00 元

本书如有破损、缺页、装订错误,
请与本社客户服务中心联系更换



版权所有 翻印必究

译者序

在竞争日益激烈和瞬息万变的市场经济条件下，企业的各级管理人员迫切需要对不同层次的大量信息迅速做出抉择。这就要求企业的各级管理人员能够从企业信息技术系统的大量复杂的业务数据中获取各自权限内的决策信息，及时把握市场变化的脉搏，做出正确有效的判断和抉择。企业决策是否及时和正确，是企业在激烈竞争的商海中能得到生存和发展的必要条件。因而，从企业各级决策者的角度来看，对企业信息技术系统要求的重点，已经从传统的业务数据操作过程扩展到对业务数据的联机分析处理，并从中得到面向各种管理主题的统计信息和决策支持信息。

随着数据库技术和互联网技术的发展而日益发展和流行的数据仓储技术正是解决上述问题的一种技术方案。它是基于大规模数据库的决策支持系统环境的核心。数据仓库是一系列集成的、面向主题的数据仓库，并设计用来提供决策者所需要的信息。

本书详尽地讲述了数据仓库的基本概念和企业尽可能地保留已有的信息技术系统投资下，如何经济有效地建立企业的数据仓库的方法。主要包括：企业的技术信息体系结构；数据仓库的概念；企业开发和建立数据仓库的关键人员的选择及其责任；建立数据仓库的策略；数据仓库的规划；数据仓库的实现；数据仓库需要的硬件和操作系统；数据仓库的软件；数据仓库的模式设计；数据仓库的元数据和数据仓库的应用。本书还讲述了数据仓库建立后的维护和演进以及数据仓库技术的发展倾向。本书在附件中还提供了极其有用的联机分析处理用的 R/OLAPXL 程序和用户手册以及仓库设计师用户手册。因而本书对企业的决策者、数据仓库的开发和管理者、信息技术工程师和计算机技术专业的教师、研究生和本科生都有重要的参考价值。

本书在翻译过程中，由清华大学计算机专业硕士陈刚先生对技术名词术语进行了统一和校正。

由于译者水平有限，以及一些术语的翻译目前尚缺乏规范，错误之处望广大读者批评指正。

译者

2002 年 12 月

前 言

本书适用于那些已经对数据仓储技术有一定了解或将要评价、学习、实现数据仓储技术的信息技术专业人员。

数据仓储技术正方兴未艾，在过去的几年中，它的规模和影响在不断地增加。有关数据仓储技术的产品、销售商、组织，当然还有图书都层出不穷。那些成功实施了数据仓库的企业都认为数据仓库是具有战略意义的。他们甚至会想以前没有数据仓库的时候企业究竟是如何生存下来的。

早在 1995 年，一份加特纳集团（Gartner Group）对《财富》500 强的信息技术主管的调查表明，90% 的企业到 1998 年会实施数据仓库。事实上，到 1998 年美国排名前 100 位的银行都积极使用基于数据仓库的收益率应用程序。大约 30% 的公司都积极跟踪这项技术，成立长期或临时的部门来规划、创建、维护、升级和支持数据仓库。

如果您是一位信息技术专业人员，需要为您所在的组织规划、管理、设计、实施、支持或维护数据仓库的话，那么，本书对您十分适合。

第一部分介绍了企业体系结构和数据仓库的概念，也是本书的理论基础。

本书的第二部分注重于数据仓储启动的 3 个关键人物：项目主办者、首席信息官和项目经理。这部分用来解决这 3 类人所关心的问题。

第三部分讲述了规划和实施数据仓库的过程，并提供了相关的指导。这对于首次进行数据仓库开发的人员和那些已经有经验的人员都是非常有用的。

本书的第四部分注重于数据仓储的技术方面。它让那些令人头晕目眩的各种技术部件变得井井有条，使您可以从中选择来建立自己的数据仓库。

本书的第五部分为数据仓储的未来打开了一扇窗。

本书附带一张软盘，其中包含两个软件产品。请查看软盘中的 readme. txt 文件来获得相关的最新信息。

附带的软件产品是：

- **R/OLAPXL**。R/OLAPXL 是强有力的查询和报表工具。它允许用户把放在 ODBC 兼容的数据库里面的数据集市和数据仓库中的数据读取到微软 Excel 电子表格

中。数据读取到了微软 Excel 中，您就可以随意使用 Excel 的标准功能来分析、报表和图示这些数据。

- **数据仓库设计师。**数据仓库设计师能够生成数据定义语言的语句，这些语句可以创建数据仓库和数据集市的维表。用户通过图形用户界面的前端来说明需要的数据结构。软件生成的语句可以创建主键、外键、索引、约束和表结构。它能够识别主要的维模型概念，如事实 and 维表，核心和定制模式，以及基础和合计模式。

在使用光盘上的软件之前，您必须先阅读并同意附上的许可协议。软件的使用手册包含在本书的附录之中。关于这些软件产品的最新信息请访问内联网商用系统公司的网站。网址是 <http://www.intranetsys.com>。

目 录

译者序	1
前 言	1

第一部分 引 言

第一章 企业的信息技术 (IT) 体系结构	3
过去：企业体系结构的演进	3
现在：信息技术专业人员的责任	3
业务角度	5
技术角度	7
组织结构迁移方案	11
迁移策略：我们怎样前进？	20
小 结	22
第二章 数据仓库的概念	24
计算重点的逐步变化	24
数据仓库的定义	26
动态的专项报表	28
数据仓库的目的	29
关于数据集市的话	30
关于操作性数据库的话	31
数据仓库的成本 - 效益分析/投资回报	34
小 结	35

第二部分 人 员

第三章 项目主办者	39
数据仓库将怎样影响我的决策过程?	39
数据仓库怎样改进我的财务过程、市场营销、操作?	40
数据仓库项目什么时候才被认为是合理的?	41
包含哪些支出?	43
风险是什么?	45
减少风险的方法	50
我的组织是否为数据仓库做好了准备?	51
我怎样衡量成果?	52
小 结	53
第四章 首席信息官	54
我怎样支持数据仓库?	54
我的数据仓库将怎样演进?	55
谁应当参与数据仓库项目?	56
小组的结构是怎样的?	60
我的人员需要什么新的技能?	60
怎样把数据仓储纳入我们的信息技术结构?	63
我需要找多少家供应商?	64
我应当怎样寻找数据仓库的供应商?	65
数据仓储怎样影响我的现有系统?	69
数据仓库及其对其他企业初始阶段的影响是什么?	69
在什么情况下数据仓库是不适合的?	70
我怎样管理或控制数据仓库的启动?	72
小 结	74
第五章 项目经理	75
我怎样展开数据仓库的启动?	75
硬件平台有多重要?	78

包含什么技术?	79
我在数据仓储中仍然用关系数据库吗?	80
一个数据仓储项目要持续多长时间?	85
数据仓库和其他信息技术项目有什么不同?	86
什么是数据仓储项目成功的关键因素?	87
小 结	88

第三部分 过 程

第六章 仓储战略	91
战略的组成部分	91
决定组织的环境	92
进行初步的需求调查	92
进行初步的源系统审查	94
识别外部数据源 (如果使用)	95
定义仓库的迭代阶段 (分阶段实施)	96
确定初步的数据仓库体系结构	96
评估开发和生产的环境及工具	97
小 结	98
第七章 仓库管理和支持过程	99
确定问题的跟踪和解决过程	99
制定容量规划	100
确定仓库清除规则	101
确定安全措施	101
确定备份和恢复策略	102
建立仓库使用统计的集合	103
小 结	103
第八章 数据仓库规划	104
组成小组和给小组定向	104

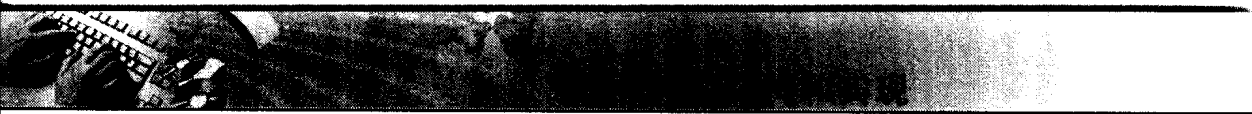
进行决策需求分析	105
进行决策源系统审计	106
设计逻辑的和物理的仓库模型	108
生成从源到目标字段的映射	109
选择开发和生产的环境和工具	111
为本次迭代建立原型	111
制定本次迭代的实施计划	112
仓库规划的提示和忠告	114
小 结	117

第九章 数据仓库的实施	119
获得和建立开发环境	119
获得操作表的拷贝	119
最终决定物理上的仓库方案设计	120
建立或装配提取和转换子系统	120
建立和装配数据质量子系统	122
建立仓库的装载子系统	126
制定数据仓库的模式	129
设定仓库的元数据	129
设定数据接入和检索工具	130
完成生产仓库的装载	132
进行用户培训	132
进行用户测试和验收	134
小 结	134

第四部分 技 术

第十章 硬件和操作系统	139
并行的硬件技术	139
硬件选择标准	139
小 结	141

第十一章 仓储的软件	142
中间件和连接工具	142
提取工具	143
转换工具	145
数据质量工具	146
数据装载器	147
数据库管理系统	147
元数据资料档案库	149
数据存取和检索工具	149
数据建模工具	153
仓库管理工具	153
源系统	154
小 结	154
第十二章 仓库模式设计	155
联机交易处理系统使用标准化的数据结构	155
决策系统的维数造型	157
两种表格：细节和维数	157
一种模式是事实细节表格加上有关的维数表格	159
细节是全部标准化的，维数是非标准化的	159
维数层次和分层次操作	160
时间维数	162
细节表格的粒度	162
细节表格键码连接着维数键码	162
集合或汇总	163
维数特征	165
多种星形模式	165
核心和定制表格	166
小 结	167
第十三章 仓库的元数据	168
元数据是一种提取的形式	168
为什么元数据是重要的	169



元数据的类型	170
版本化	172
元数据作为仓储任务自动化的基础	173
小 结	173

第十四章 仓储的应用	175
早期的采用者	175
仓储应用的类型	175
仓储技术的专门应用程序	177
小 结	179

第五部分 现在向何处去

第十五章 仓库的维护和演进	183
定期的仓库装载	183
仓库统计数据的收集	183
仓库用户概况	184
安全和接入层面	185
数据质量	185
数据增长	186
更新仓库子系统	186
数据库的优化和调整	187
数据仓库的员工安排	188
用户和仓库员工的培训	188
随后的仓库迭代	189
充电反馈模式	189
灾难恢复	190
小 结	190

第十六章 仓储的趋势	191
数据仓储行业的持续发展	191

更多的行业采用仓储技术	191
数据挖掘技术日趋成熟	192
出现和使用元数据交换标准	192
可使用万维网的解决方案可用性增加	192
数据集市项目普遍使用 Windows NT	193
应用软件包中仓储模块的可用性	193
在仓库供应商中有更多的合并和收购	193
小 结	194

第六部分 附 录

附录 1 R/OLAPXL® 用户手册	197
欢迎使用 R/OLAPXL!	197
安装	200
教程	202
用户指南	205
使用 R/OLAPXL 的列	207
设定 R/OLAPXL 的选项	216
R/OLAPXL 的工具栏	225
宏编程	227
R/OLAPXL 的信息	234
附录 2 仓库设计师用户手册	241
欢迎使用仓库设计师!	241
业务概念	244
仓库设计师的工具栏	245
应用程序	246
维	247
模式	249
定制维	251



定制模式	251
合计维	251
合计模式	252
附录 3 在线数据仓储资源	253
附录 4 工具和供应商目录	255
附录 5 软件许可证协议	260
附录 6 专业用语中英文对照表	262
附录 7 外国人名中英文对照表	276
附录 8 外国公司名中英文对照表	277
许可证协议和有限保证	279

第一部分 引言

企业体系结构 (Enterprise Architecture) 这个术语是指部件之间的相互联系, 这些部件集成在一起, 用以满足企业的信息需求。本部分会引入企业信息技术体系结构的概念, 并提供一个可以包容企业计算所需的各种技术的框架。

数据仓储 (Data Warehousing) 技术只不过是信息技术体系结构中诸多部件的一种。这一部分是要阐明数据仓储是怎样纳入整个信息技术体系结构中的。希望信息技术专业人员在将数据仓储技术与企业中其他信息技术部件进行集成的时候, 能够处于比较有利的位置。

第一章 企业的信息技术（IT）体系结构

本章先简单地介绍一下变化中的业务要求如何影响企业体系结构的演进。引入了 InfoMotion（运作中的信息）企业体系结构，它为信息技术专业人员提供了一种框架，当前的各种技术都可以纳入到这种框架中去。

过去：企业体系结构的演进

在一定时期内，一个企业的信息技术体系结构依赖于 3 个主要因素：

- 企业的业务需求；
- 当时可用的技术；
- 企业对早期技术累积的投资。

企业的业务需求在不断地变化，并且这种变化是指数级的。多年来，业务需求已经从日常交易的书面记录转变为业务过程的自动化。例外报告已经从跟踪日常交易并校正其错误，转向可自行调整的业务过程的开发。

技术同样也从呈指数级增长的计算和通信能力中得到了发展。然而，相对于计算机硬件上的进步，在软件开发和结构定义领域存在着严重的滞后。因此企业体系结构就无法按照业务要求自然演进，要么是迁就早先的技术投资，要么是严重地限制了自身进一步演进的能力。

后来才认识到，典型的企业体系结构的演进反映了信息技术专业人员方方面面的努力，他们不断利用最新的技术来改进对业务操作的支持。不幸的是，这种努力往往会导致企业陷入各部件不能兼容的困境。

现在：信息技术专业人员的责任

今天，信息技术专业人员通常肩负双重责任，使用信息技术满足业务需求和把新