

SQL Server 2000 OLAP 服务设计与应用

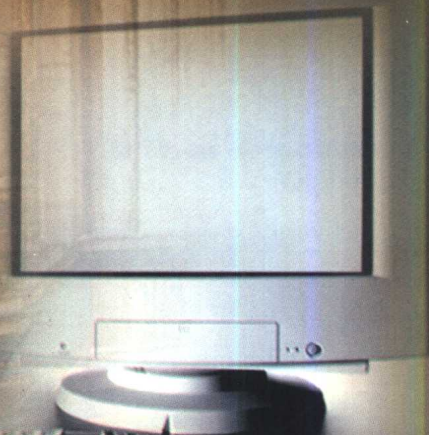
飞思科技产品研发中心 编著



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>



开发专家之数据库

SQL Server 2000 OLAP 服务设计与应用

飞思科技产品研发中心 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书属于《开发专家之数据库》系列。SQL Server 2000 是 Microsoft 公司最新的数据库软件，它不仅支持传统的关系数据库，还支持知识仓库及功能强大的数据分析服务。

本书针对 DTS、OLAP 模型与设计、SQL 分析服务、数据透视表和 Office 2000、开发决策支持程序、OLAP 商业应用程序、知识库及数据仓库管理等内容，以实例的方式加以介绍，使读者能够充分掌握 SQL Server 2000 分析服务的设计与实际应用，进而构建企业级决策支持系统。

本书适合于学习数据库技术的中级读者，对于高级读者来说，也是一本优秀的参考书。

本书所用源程序请到 <http://www.fecit.com.cn> 网站的“下载专区”下载。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，翻版必究。

图书在版编目 (CIP) 数据

SQL Server 2000 OLAP 服务设计与应用/飞思科技产品研发中心编著. —北京: 电子工业出版社, 2002.1
(开发专家之数据库)

ISBN 7-5053-7080-4

I. S... II. 飞... III. 关系数据库—数据库管理系统, SQL Server 2000 IV. TP311.138

中国版本图书馆 CIP 数据核字 (2001) 第 075703 号

丛 书 名: 开发专家之数据库

书 名: SQL Server 2000 OLAP 服务设计与应用

编 著: 飞思科技产品研发中心

责任编辑: 卢国俊

排版制作: 电子工业出版社计算机排版室监制

印 刷 者: 中国科学院印刷厂

出版发行: 电子工业出版社 <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销: 各地新华书店

开 本: 787×1092 1/16 印张: 27 字数: 691.2 千字

版 次: 2002 年 1 月第 1 版 2002 年 1 月第 1 次印刷

书 号: ISBN 7-5053-7080-4

TP · 4060

印 数: 5000 册 定价: 38.00 元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁盘或光盘有问题者,请向购买书店调换。若书店售缺,请与本社发行部联系调换。电话 68279077

出版说明

现代社会的发展，珠玑纵横。曾几何时，境由心起，想登高远望，却见雾影渺渺，难得分明。于是惊叹，面对科技的百年一诺，虽意气峥嵘，竟落得难逐新流，无言徘徊。

迷茫、困惑、烦闷、愁怀，一时间接踵而至，该何从何去？

理想在右，信念在左，电子工业出版社计算机图书研发部适时播种，又一棵桃李——《开发专家》破土发芽。根虽浅，却足以令您抑郁释怀；枝虽嫩，却可使您壮然前行。

《开发专家之数据库》那妩媚的花蕾虽然刚刚昭然入目，却已将这一径新世纪的长途点缀得春意融融。而 Visual C++编程、Visual Basic 编程、网络编程和 Delphi 编程如即将舒展的花枝，更将使天下穿枝拂叶的行人感到柳暗花明，峰回路转。

我们在随时耕耘。当花枝上鲜花乱颤，芬芳醉人的时候，也就是您分列珠玑，追世逐流的时候。

我们希望随时收获。当天下桃李纷至沓来，向我们诉说成功时，即使以后踏着荆棘工作，我们也不会觉得痛楚。

我们更希望《开发专家》能够为您的理想插上腾飞的翅膀，帮您开发出人生的另一重境界。

品牌标识：



电子工业出版社计算机图书研发部
于北京

关于飞思

世纪之交的北京，一群满怀共同理想的年轻人聚集在飞思教育产品研发中心的旗帜下，他们将新的希望和活力注入了中国 IT 教育产品开发领域。从那时起，飞思人一直在为把自己打造成为中国 IT 教育产品开发的精英团队而不懈努力。

21 世纪的今天，飞思人在多元化教育产品的开发和出版等方面已经迈出了坚实的第一步，开拓出属于自己的一片天空，初步赢得了涓涓细流。

如今，本着教育为科技服务的宗旨，飞思教育产品研发中心拓展为飞思科技产品研发中心，并以崭新的面貌等待您的支持与关注。

飞思人理念

我们经常感谢生活的慷慨，让我们这些原本并不同源的人得以同本，为了同一个梦想走到一起。

因为身处科技教育前沿，我们深感任重道远；因为伴随知识更新节奏，我们一刻不敢停歇。虽然我们年轻，但我们拥有

“严谨、高效、协作”的团队精神

全方位、立体化的服务意识


实力雄厚的作者群和开发队伍

当然，最重要的是我们拥有：

恒久不变的理想和永不枯竭的激情和灵感

正因如此，我们敢于宣称：

飞思教育 = 丰富的内容 + 完美的形式

这也是你和我共同精心培育的品牌  的承诺。

“问渠哪得清如许，为有源头活水来”。路再远，终需用脚去量；风景再美，均需自然抚育。

年轻的飞思人愿作清风细雨、阳光晨露，滋润你发芽，成长；更愿作坚实的铺路石，为你铺就成功之路。

前 言

关于数据库

世界信息化的飓风，掀起中国信息技术迅速发展的浪潮，一时间群贤毕至、冠盖云集，使得数据库开发及编程人才也供不应求，越来越多的人不断地寻求机会，以期能够在这个信息的大潮中找到属于自己的那片浪花。

基于此，飞思科技产品研发中心策划并组织编写了主要涉及 SQL Server、MySQL 等数据库应用的编程丛书，以不断满足广大用户对编程知识的需求。

SQL Server 2000 是为创建可伸缩电子商务、在线商务和数据仓储解决方案而设计的真正意义上的关系型数据库管理与分析系统。SQL Server 2000 中包含许多新特性，这些特性使其成为针对电子商务、数据仓库和在线商务解决方案的卓越的数据库平台。

SQL Server 2000 能提供超大型系统所需的数据库服务。大型服务器可能有成千上万的用户同时连接到 SQL Server 2000 的情况，SQL Server 2000 为这些环境提供了全面的保护，具有防止问题发生的安全措施，例如，可以防止多个用户试图同时更新相同的数据。SQL Server 2000 还在多个用户之间有效地分配可用资源，比如内存、网络带宽和磁盘 I/O 等。

SQL Server 2000 不仅能作为一个功能强大的数据库服务器有效地工作，而且数据库引擎也用于需要在客户端本地存储独立数据库的应用程序中。SQL Server 2000 可以动态地将自身配置成能有效地使用客户端桌面或膝上型电脑中的可用资源，而不需要为每个客户端专设一个数据库管理员。应用程序供应商还可以将 SQL Server 2000 作为应用程序的数据存储组件嵌入到应用程序中。

总之，对于那些希望确保数据库解决方案具备伸缩性、可靠性及灵活性的客户而言，SQL Server 2000 提供了最强的支持。

我们推出的这套《开发专家之数据库》丛书，有如下几个特点：

- 知识全面 整套丛书涵盖了 SQL Server 2000 的全部内容。
- 实例丰富 对重点知识的讲解都精心设计了典型的例子。通过研读这些例子，读者就可以轻松地攻克所有的难关。
- 内容新颖 整套丛书对 SQL Server 2000 的新特性进行了重点讲解，同时介绍了微软的 .Net 平台：包括最新的数据库访问接口 ADO.Net、XML 编程组件 XML.Net 等。

岁月荏苒，科技也与您的青春一样稍纵即逝。适时把握科技发展的命脉，您也就把握住了您花样般的青春年华。陈旧的编程概念，如今已换上新颜，吐出新蕾，想马上采撷吗？路径是：电子工业出版社飞思科技产品研发中心\“开发专家之数据库”系列。

关于本书

对于一名专业的信息技术人员来说，如何最大限度地利用公司各个部门多年来在数据

库系统上的投资，把公司花了大气力搜集来的数据变成对公司未来发展有巨大价值的信息，这是每一个从事信息技术工作的专业人员都应该考虑的问题。

在这种背景下，OLAP（联机分析处理）的概念应运而生，并且蓬勃发展起来。微软也不失时机地在 SQL Server 中推出了 DTS（Data Transformation Service，数据转换服务），结合 SQL Server 2000 提供的 Analysis Service 和数据透视表服务，SQL Server 优秀的数据库技术为企业提供了一个完备的解决方案。

本书详细地介绍了 DTS 和 Analysis Service，以及如何使用 Microsoft Office 作为 OLAP 系统的前端用户界面，并且介绍了多维 SQL 语法扩展（MDX）及 Microsoft 知识库等技术。

如果您准备从事数据仓库或者分析服务方面的研究，那么，这本书完全可以为您提供关于这方面知识的完整参考。本书的目的是通过实例，详细地、循序渐进地介绍 Microsoft SQL Server 2000 分析服务的设计与应用，使信息技术专业人员能够在短时间内掌握微软提供的解决方案，解决那些也许困扰他们很长时间的难题。

本书最大的特色就是实例，我们在讲解中尽可能使用实例，而不是深奥的技术术语。希望通过这种方式，让读者在学习完本书后，能够立刻开始应用本书中学习到的技术。在本书的第 8 章，我们为读者布置了一个大作业，希望读者用前面几章所学到的知识独立完成一个基于 Web 的数据仓库系统，当然，我们还是给出了必要的提示。

关于本书所涉及的源程序，如需要，请到 <http://www.fecit.com.cn> 网站的“下载专区”下载。

本书由飞思科技产品研发中心组织编写，主要部分由李辉、任永政主笔，同时参加编写工作的有郝艳芬、胡松龄、张青、王惠娟、王子龙等。由于本书涉及的内容丰富，加之篇幅、时间所限，书中不足之处，敬请读者批评指正。我们的联系方式：

电话：(010) 68131648 (010) 68134545

E-mail: fecit@fecit.com.cn fecit@sina.com


<http://www.fecit.com.cn> <http://www.fecit.net>


丛书约定


对本丛书统一运用的符号解释如下：


【 】 表示命令、快捷键。


【 】 → 【 】 表示打开某一菜单下的菜单或命令。

 **说明** 表示需要解释说明的部分。

 **步骤** 表示某一个例子的操作步骤。

 **技巧** 表示操作过程中的技巧部分。

 **注意** 表示需要引起注意的地方。

 **提示** 表示某一步骤的需要提示的部分。

飞思科技产品研发中心

目 录

第 1 章 导论.....	1
1.1 构建企业级决策支持系统的主要步骤.....	2
1.2 数据仓库概述.....	2
1.3 数据仓库的形式.....	5
1.4 数据仓库组件.....	6
1.5 考察实际所需的时间和费用.....	7
1.6 微软的数据仓库框架.....	8
1.7 数据仓库的分析过程.....	9
1.8 数据仓库与 OLAP.....	14
第 2 章 数据转换服务.....	15
2.1 DTS 包概述.....	16
2.2 DTS 工具.....	20
2.3 DTS 包的制作与执行.....	20
2.4 数据泵 (Data Pump).....	45
2.5 高级 DTS 技术.....	48
第 3 章 OLAP 结构与与设计.....	51
3.1 OLTP 设计.....	52
3.2 OLAP 设计.....	55
3.3 设计 OLAP 维模型.....	62
3.4 多维模型的设计步骤.....	66
3.5 OLAP 模型实践 (图书销售系统).....	71
第 4 章 SQL Server 分析服务与多维服务.....	111
4.1 Microsoft SQL Server 2000 分析服务.....	112
4.2 分析服务的安装.....	120
4.3 设计与创建多维数据集.....	124
4.4 存储多维数据集.....	126
4.5 设计与创建多维数据集实践.....	138
第 5 章 OLAP 设计与实践.....	185
5.1 商业 OLAP 构建分析 (使用 Northwind 数据库).....	186
5.2 建立 NW_Mart 数据库.....	189
5.3 设计 NW_Mart 的星型模型.....	189
5.4 建立 OLAP NW_Mart 的 DTS 包.....	194
5.5 设定数据库与数据来源.....	236
5.6 建立销售分析多维数据集.....	239
5.7 建立运费分析多维数据集.....	278
5.8 设定销售分析及运费分析的计算成员.....	287

5.9	建立销售运费分析的虚拟多维数据集	291
5.10	使用 VB6 制作多维数据集处理程序	295
第 6 章	数据透视表和 Office 2000	303
6.1	概论	304
6.2	Microsoft SQL Server Analysis Service 结构	304
6.3	数据透视表服务介绍	305
6.4	客户端用户工具	309
6.5	用 Excel 2000 分析多维数据	312
第 7 章	开发决策支持系统应用程序的概念	333
7.1	OLE DB 与 OLAP 应用程序结构	334
7.2	多维模式名词介绍	335
7.3	多维 SQL 语法 (MDX)	336
7.4	MDX 相关函数	341
7.5	ADO MD 对象介绍	342
7.6	MDX DDL 语法与本地多维数据集	351
7.7	MDX 和 ADO MD 语法练习——使用 Excel 2000 VBA	362
7.8	MDX 和 ADO MD 语法练习——使用 Active Server Page	370
第 8 章	商业应用程序设计	373
8.1	概论	374
8.2	用 Active Server Page 制作出版社订单处理系统 (OLTP)	374
8.3	用 Active Server Page 制作出版社 OLAP 系统	377
第 9 章	Microsoft 知识库	391
9.1	什么是 Microsoft Repository	392
9.2	Repository 的对象	395
9.3	开放信息模型 (OIM)	396
9.4	数据仓库的信息模型	399
第 10 章	分析服务器的管理	407
10.1	OLAP 多维数据集用户账户管理	408
10.2	多维数据集的使用分析	413
10.3	多维数据集的优化	417

第 1 章 导 论

内容提要:

- 构建企业级决策支持系统的主要步骤
- 数据仓库概述
- 数据仓库的形式
- 数据仓库组件
- 考察实际所需的时间和费用
- 微软的数据仓库架构与工具
- 数据仓库的分析过程
- 数据仓库与 OLAP

1.1 构建企业级决策支持系统的主要步骤

到目前为止，企业数据的分析已经从早期的书面报告、电子表格、决策信息分析逐步发展到联机分析处理（Online Analytical Processing，简称 OLAP）。自从 Arbor Software 为了同联机事务处理（Online Transaction Processing，简称 OLTP）相区别而创造了“联机分析处理”（前者更强调实时的数据更新）这个名词以来，世界上主要的数据库管理系统（DBMS）制造商已经花费了巨大的人力物力来创建以数据仓库（Data Warehousing）以及 OLAP 为基础的决策支持系统（Decision Supporting System，简称 DSS）。1994 年 10 月 Sybase 推出了 ExpressWay；1995 年 7 月 Oracle 推出了 Express；1995 年 11 月 Informix 推出了 Metacube；Microsoft 在 1996 年 10 月同 Panorama 正式开始合作；到了 1997 年 1 月 Arbor 成为 IBM 合作伙伴的时候，数据仓库已经成为各大数据库管理系统制造商的必备功能。

Microsoft 强调在不同制造商的平台之间的集成与互操作性，以便在异构环境中实现分析、管理与存储数据的功能。因此，基于 OLE DB 的 OLAP 也就应运而生了，现在已经成为绝大多数制造商所支持的行业标准。

Microsoft SQL Server 7.0 是完全支持这些标准的高性能、低成本数据库管理系统产品。

在微软的决策支持系统框架中，构建一个企业级的决策支持系统被分为三个步骤：

步骤

（1）创建企业级数据仓库。

从各种联机事务处理（OLTP）系统的异构数据库中，通过数据转换服务（Data Transformation Service，简称 DTS）的数据验证、数据合并、数据分离与数据清理等过程来汇集有助于决策支持分析的原始数据。

（2）联机分析处理。

通过预先规划的多维分析来利用数据仓库，将数据仓库的“数据”转换为决策支持分析所需要的“信息”，并存储在 SQL Server 联机分析处理服务的多维数据集（Cube）中。

（3）利用前端应用工具进行分析。

利用诸如 Visual Basic 或 Visual InterDev 等工具，使用 ADOMD 数据对象接口（Data Object Interface，简称 DOI）以及 MDX（Multi-Dimension extension，简称多维扩展）语法，通过 OLE DB for OLAP Server 来读取多维数据集的数据，或者使用诸如 Excel 2000、Access 2000（现已升级为 2002 版）这样的成套工具来进行分析。

我们将在这本书中按照上述步骤为大家一一介绍。

1.2 数据仓库概述

数据仓库（Data Warehouse）的提出是以关系数据库、并行处理和分布式技术的飞速

发展为基础的，是解决在信息技术发展中存在的拥有大量数据，而有用信息相对贫乏的综合解决方案。

从目前形势看，数据仓库技术已紧跟 Internet，成为信息社会中获得企业竞争优势的又一关键。全球许多大公司已经将 Internet 网络和数据仓库技术列入其企业计划，而且有很多企业为使自己在竞争中处于优势也已经率先采用。

1.2.1 数据仓库的概念

“什么是数据仓库？”这恐怕是每一个刚刚开始接触数据仓库的技术人员都会提出的一个问题。数据仓库和通常所说的数据库有什么不同？或者只是前者存储的数据更多？在向你的客户或者上司介绍数据仓库解决方案的时候，你必须能够回答这个问题。

为了回答这个问题，首先让我们看一下 20 世纪 90 年代初期，业界公认的数据仓库之父 W.H.Inmon 在“*Building the Data Warehouse*”一书中给出的数据仓库定义：

“数据仓库是面向主题的、集成的、随时间变化的、稳定的数据集合，用以支持管理决策的制订过程。”（原文：A Data Warehouse is a subject-oriented, integrated, time variant, nonvolatile collection of data in support of management's decision-making process.）

这可以说是对数据仓库的一个概念性的定义，它全面地描述了一个理想的数据仓库具有的特征，因此可以作为数据仓库的一个很好的判断依据。

这个定义很好地解释了什么是数据仓库，但在某种程度上来讲，又显得太理论化。相比之下，Ralph Kimball 在“*The Data Warehouse Toolkit*”一书中给出的定义更加实用：

“数据仓库是针对查询和分析而架构的事务数据的拷贝。”（原文：Data Warehouse is a copy of transaction data specifically structured for query and analysis.）

对这个定义需要补充说明的两点是：

- 虽然数据仓库中 95%~99% 的数据是事务数据，但是还是有部分数据不是事务数据。
- 除了查询和分析之外，最重要的还有报表，因为数据仓库系统最主要的输出就是经过格式化的数据透视表。

综上所述，我们可以说数据仓库是将企业中异构的数据加以合并，将历史数据抽取出来辅助决策分析。

一般说来，数据仓库提供了决策支持分析应用程序所需要的历史资料。这里提到的应用程序包括：报表（Reporting）、联机分析处理、主管信息系统（Executive Information System，简称 EIS），还包括数据挖掘（Data Mining）。数据仓库是一个集中的、集成的信息存储。所谓集成，是指整理、合并以及再设计。集成的难易程度依赖于有多少异构系统需要导入到数据仓库中，以及相似的信息在处理过程中是否会产生很大的差异。

数据仓库和一般事务系统的数据库是不同的，在一般事务系统的数据库中，更关心的是满足联机事务系统（OLTP）的要求。其设计要求是为了让数据的即时添加、删除与更新的速度最优。而数据仓库则需要取还其中的集成的信息并输出报表，而且数据仓库是一个只读的系统。OLTP 系统包含了企业每天都要使用或需要操作的数据，而数据仓库包含的数据则是要用来进行分析的。所以，OLTP 系统是即时的、经常变动的，而且在使用过



程中，某些事务数据也可能是不完整的或者甚至是未知的。而数据仓库中的数据是历史数据，不仅不会常常变动，而且还将某些事务错误的部分做了调整。OLTP 和数据仓库的目标是不同的，而且其数据模型也不相同。在 OLTP 系统的设计目标中，消除冗余的数据非常重要，因为数据的冗余会严重地影响数据更新的速度，因此 OLTP 系统都是高度正规的关系型数据库结构。而对于数据仓库来说，冗余的数据却是必需的，因为这样才能简化用户的访问，提高性能，也减少表格 (Table) 需要进行的连接 (join) 操作次数。因此，常常采用反正规的多维模型 (Multi-Dimensional Model) 来设计。

1.2.2 数据仓库的特征

数据仓库具有如下的特征：

1. 数据仓库是面向主题的

操作性的数据，例如订单处理和制造数据库，是围绕着业务活动或者不同的功能领域组织的。典型情况下，它们是针对某个单一的、固定的应用程序优化的。应用程序的功能区别导致公司需要在不同的位置存储同样的数据。信息格式的多样化和交流的需要常常是一对矛盾。例如，在一个交货数据库中，客户列表具有详细的地址信息，一般是按照客户的“邮政编码”来索引的。而同样的客户列表在发票系统中则可能具有不同的开票地址，而且常常是按照客户的“账户号”来进行索引的。两种情况下客户的名称是一样的，但是具体内容是有区别的，并且存储在不同的位置上。要想从这两个数据库中抽取的数据之间寻找任何相互关系都可能是一个巨大的挑战。相反，数据仓库是围绕主题组织的。面向主题使数据在格式上保持一致，并且对于最终用户来说也更好理解。主题可以是“产品”、“客户”、“订单”，而不是“采购”、“薪水册”。

2. 数据仓库是集成的

数据仓库中的数据集成是通过规定，在格式、命名等等上面保持一致来实现的。由于历史的原因，操作性数据通常在数据表述上存在矛盾。例如，一组操作性数据库也许会用“男”和“女”、“1”或“2”、“x”或“y”来表示男性和女性。一般情况下，这种不一致性远比这个例子复杂，而且常常很难察觉。在数据仓库中，通过定义，数据则总是保持一致的风格。

3. 数据仓库是随时间变化的

数据仓库是随时间变化的，因为它们同时维护着历史的和当前的数据。相反，操作性的数据库只包含最近的数据，一般不超过 1 年（常常更短）。数据仓库按照每日、每周或者每月的频率从操作性数据库中导入数据，而且一般要保存 3~5 年的时间。这是两种不同环境的一个显著区别。历史信息对决策制订者来说非常重要，因为他们需要了解趋势和数据之间的关系。例如，某个饮料生产商的产品经理需要了解促销手段和销售量之间的关系。这种信息一般是不可能从只保存当前数据的操作性数据库中获得的。

4. 数据仓库是稳定的

稳定性是数据仓库的又一个主要特征，它表示信息在被导入数据仓库之后，就很少进

行修改、插入或者删除之类的操作。导入的数据是从原来存储在操作性数据库中的数据转换而来的。数据仓库把从操作性数据库中转换来的或者总结出来的数据按照时间顺序添加到前一个时期的数据之后。除了这个过程，数据仓库中的信息一般保持固定不变。稳定性使得我们可以对数据仓库进行最大限度的优化。

5. 数据仓库是支持管理决策的制定过程

这是数据仓库的最终目的，数据仓库不是用来支持业务上的日常操作的，而是用来支持战略性的或者战术性的管理决策的。这也是数据仓库有别于操作性数据库的主要特征。

1.2.3 数据仓库给企业带来的利益

现在，你应该能够基本了解究竟什么是数据仓库了。随着你对本书的深入阅读，你会对这个概念有越来越深的体会。在本节的最后，我们简单地讨论一下相关的另一个问题：数据仓库究竟能给企业带来什么样的好处？为什么像 GE、P&G 这样的公司要在全世界范围内推广数据仓库？

答案是：因为数据仓库可以让企业的高级主管更好地了解企业现在拥有的数据，从过去的和现在的数据中看出未来的趋势以及企业运作的某个环节对这种趋势的影响，因此能更好地进行决策分析，进而更好地为客户提供服务，从而提升企业自身的竞争优势。

在了解了什么是数据仓库以及数据仓库究竟能为企业带来什么样的利益之后，很自然地我们会开始关心另一个问题：数据仓库都有哪些类型？哪个是适合我的企业的呢？我们将在下一节讨论这个问题。



1.3 数据仓库的形式

企业所需要的数据仓库的形式是根据企业及其决策支持分析的需要而定的。

最简单的数据仓库形式之一是操作性数据存储（Operational Data Store，简称 ODS）。ODS 将 OLTP 数据库进行复制，并将不适合的字段或者数据进行调整和修正。ODS 的最初用途是产生一般报表并且提供每日事务的详细的、概括性的分析（因为 ODS 是从 OLTP 系统复制来的，因此有些专家认为它不是一种数据仓库形式，但我们认为，因为 ODS 的范围很广而且具备很多数据仓库的功能，因此还是需要在此提及）。

根据企业本身对报表的需求，ODS 可能要每月、每周或更短的时间就要进行更新，有时候甚至是即时更新。其最主要的优点在于提高 OLTP 系统的性能，因为报表和查询的操作是从 OLTP 系统中脱机加载到 ODS 的。

如果你的企业只需要产生日常上的报表，外加一些分析和市场调查的话，ODS 应该就可以满足你的需求了。甚至，如果你的企业每天只需要处理少许事务，那么就是 ODS 都显得多余了，这种情况下，生产系统产生的报表就足够了。

另一种形式的数据库就是数据集市（Data Mart）。一般来说，数据集市的信息来自单一部门或是日常工作的处理事项。这些数据通常用来做特殊范围的销售分析或生产线分析。数据集市通常只包括摘要信息，如果需要的话，这些摘要信息也可以连接到 ODS 通

过向下展开 (Drilling down) 的方式来查看事务的细节。

在数据集中可以同时运行多个 OLTP 应用程序, 跨部门的 (cross-departmental) 分析、主管信息系统和数据挖掘应用程序都需要将整个企业的信息集成在一起才能发挥最大的功效, 因此企业的数据仓库也需要汇集很大范围的数据并加以分析。这是一件非常复杂的工作, 因此通常会由企业的 IT 小组来负责管理。同时, 由于数据仓库中的数据是从不同的系统集成到某一公用的主题领域, 因此也使得数据仓库的维护和管理更加复杂。不同的系统往往会产生数据不相容或不一致的情况, 因此这些数据在集成到数据仓库之前必须经过多次转换。根据保守的估计, 创建一个数据仓库的过程中, 大约 80% 的时间都花在数据萃取、整理和数据导入上了。

如果你的企业能从数据挖掘中获取利益, 那么, 创建数据仓库就是值得努力的。数据建模 (Modeling Data)、估计或预测数据结果等等, 都需要结合使用数据挖掘工具及各种统计技术, 尤其对于存在巨量数据的情况。Microsoft 在 SQL Server 的下一个版本 SQL Server 2000 中提供了数据挖掘功能。

了解了数据仓库的形式之后, 顺便提一下数据仓库的创建方法。

数据仓库的创建方法包括集中式架构——将整个企业的数据放在单一且巨大的存储中; 分布式架构——将数据分布在多个数据集中, 因此也产生了不同的自顶向下 (Top-down)、自底向上 (Bottom-up) 以及数据商场 (Data Mall) 等方法。在本书的第 3 章将进行进一步的介绍。

1.4 数据仓库组件

可以想像, 数据仓库这么复杂的东西一定包含了许多组件。事实上也的确如此, 数据仓库的每个组件可能会由一个或多个软硬件来实现。到目前为止, 没有一家制造商能独立提供所有的数据仓库设备。

数据仓库把数据从操作系统中抽取出来放入清理 (scrubbed) 保留区中, 然后将数据有关时间的字段合并, 再装载到数据库中供用户访问。

转换后的数据存储的位置与原来存储数据的位置不一样, 这些变化被记录在目录 (catalog) 中, 而元数据 (metadata) 工具负责管理这些目录。所谓元数据是指用于在数据仓库中定义或描述数据的数据。我们已经很熟悉的诸如表格、字段名称的定义叫做前端 (frontend) 元数据, 还有一种是映射到原始数据库的被称为后端 (backend) 元数据。

设计和管理的工具对于数据仓库而言也是非常重要的, 无效的数据在数据仓库中是不会被统计的。数据仓库的维护任务非常之重, 当公司的需求增加时, 数据仓库的运作也会随之扩展, 而且在 OLTP 数据库改变之后, 数据仓库也会有相应的变化, 同时, 这些变化也必须记录在元数据中。我们可以说, OLTP 系统只是当前状况的一个快照 (snapshot), 而数据仓库则提供了整个时间段的数据。这要求无论 OLTP 系统的数据如何改变, 只要数据仓库的数据在某个时间有所变化, 都应该把这些变化记录下来, 否则, 在查询的时候就会产生不正常的结果。

1.5 考察实际所需的时间和费用

根据 1996 年的一份研究资料表明, 构建数据仓库的平均开销是 220 万美元, 一般需要 2.3 年的时间才能完成。在构建数据仓库的企业中, 90% 以上投资回报率在 40% 以上, 而 50% 的企业投资回报率更高达 160%, 3 年后的投资回报率将累积到 400%。由此可以看出, 在数据仓库上的投资将给整个企业带来极大的回报; 同时也看出, 构建数据仓库不是一朝一夕的事情, 对企业来说, 投资也不是一个小数目, 因此在构建数据仓库之前, 必须先对数据仓库的投资花费以及预期的结果进行分析。

1. 对存储空间的要求

数据仓库需要巨大的硬盘空间。在规划存储空间的时候, 不要简单地沿用过去设计 OLTP 数据库的经验, 请记住数据仓库存储历史数据, 大部分的企业一般会保存最少一年且最有价值的历史数据。而如果要进行任何趋势分析的话, 则需要保存数年的数据。同时, 报表和分析所需要的索引也不可能只有一个, 因此, 必须考虑需要的硬盘空间。目前, 企业的数据仓库所需要的空间用 Terabyte (1 Terabyte = 2^{40} byte) 为单位计算, 某些企业甚至要用 Petabyte (1 Petabyte = 1 024 Terabyte) 进行计算。幸好现在存储技术的发展使得硬盘技术和价格都算合适。

2. 数据合并

在把不同的 OLTP 系统的数据合并的过程中, 不相容或不一致的情况肯定会出现。数据挖掘对数据的一致性要求很高, 因此必须调整 OLTP 系统上庞大的数据, 并整理出数据仓库所需要的数据, 这样才能保证成本的合理应用。

3. 数据的安全性

数据仓库的安全考虑与 OLTP 的安全考虑不一样。由于用户从数据仓库将获得更为高级的信息, 因此访问数据仓库的用户将会比访问 OLTP 系统的用户多。因此安全性考虑也更复杂。“需要者知情”这种保密工作原则对于数据仓库来说是非常重要的, 只有这样才能相对减少访问数据仓库的用户数量, 从而提高效率。企业必须从“有权利就知道”(right to know) 转变为“有需要才知道”(need to know), 否则, 创建数据仓库也就失去了其意义。

4. 用户界面

数据仓库必须具有友好的终端用户界面。因为跟 OLTP 数据库不一样, 数据仓库的使用对象往往是不大会使用计算机的管理及决策阶层。因此, 在用户界面的考虑上也需要更深入一些。

5. 规划方案

一个数据仓库的构建需要花费 2.3 年的时间, 而用于规划的时间也不会少于这个数量。数据仓库是一项庞大的工程, 在具体实现之前, 必须投入足够的精力考虑构建数据仓库所要达到的目标, 有什么潜在的成本与利益, 什么资源是需要的, 有多少工作是你需要外包



给其他服务提供商的等等。完善的规划是非常必要的，如果只是一味地花钱，数据仓库是很难成功的。

企业每天都在产生数据，这些数据对于企业来说意味着无形的资产。越来越多的企业认识到这一点，因此也越来越依赖于数据仓库。但是，必须认识到，先进的技术不会自动转变为生产力，如何运用和管理这些技术对于成功是必不可少的。

1.6 微软的数据仓库框架

微软的数据仓库如图 1-1 所示，在 SQL Server 上创建数据仓库数据库（或者数据集市）。通过数据转换服务利用 OLE DB 这个应用程序编程接口（API）来有效地访问异构数据。系统的不同组件之间通过微软中心库（Repository）共享元数据。前端工具可以使用 Internet Explorer、Access、Excel、自然英语查询（Microsoft English Query）等等来通过 OLE DB 存取数据仓库的数据。

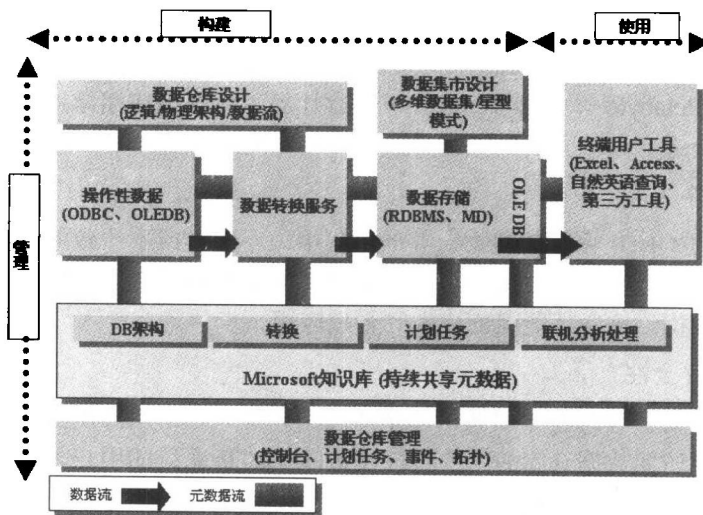


图 1-1 微软的数据仓库框架

Microsoft SQL Server 2000 提供了功能强大的数据仓库与决策支持工具。包括：

1. 数据转换服务

让用户可以通过 OLE DB 将来自多个异构数据库来源的数据，汇集成一个或分离成多个目标。数据转换服务不仅可以在数据转换的同时通过 ActiveX 脚本使数据具有一致性，还可以根据基本数据来汇总数据。

2. 联机分析处理服务

通过联机分析处理服务工具能够分析维，并建议应该首先存储哪些值，而其他值可以在执行时间动态生成。并且联机分析处理服务还提供了向导程序帮助你提高查询速度或节省硬盘存储空间。