



Designed for

Microsoft®
Windows NT®
Windows 98



附 赠
CD-ROM

Data Warehousing with Microsoft SQL Server 7.0 Technical Reference

微软公司
核心技术书库

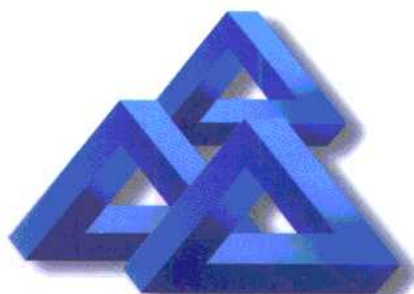
(美) Jake Strum 著

刘汉字 郝春明 裴莹 来强 译

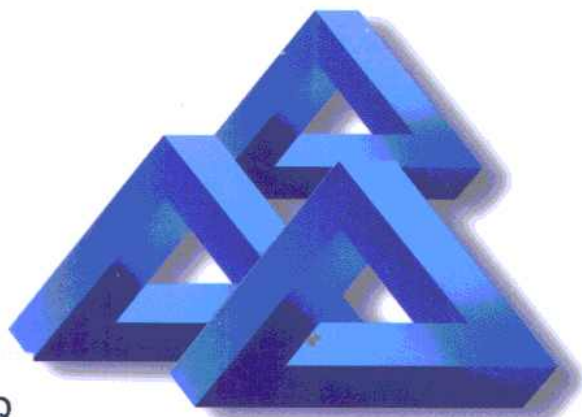


Microsoft

SQL



Server 7



数据仓库技术指南

1385Q



机械工业出版社
China Machine Press

Microsoft Press

TP311.1385Q 209
S81b

微软公司核心技术书库

Microsoft SQL Server 7.0

数据仓库技术指南

(美) Jake Sturm 著
刘汉宇 郝春明
裴莹 来强 译

本书附盘可从本馆主页 <http://lib.szu.edu.cn/>
上由“馆藏检索”该书详细信息后下载,
也可到视听部复制



A0939716



机械工业出版社
China Machine Press

本书首先介绍了数据仓库的基础知识,包括操作数据和分析数据的比较、数据仓库和数据市场、成功数据仓库的特征、数据仓库的规划和设计;然后介绍了对复杂关系型数据进行快速、完善分析的在线分析处理(OLAP)服务,可以对开放型数据库连接(ODBC)和OLE DB 所支持的各种资源进行转换的数据转换服务(DTS),先进的查询处理器以及用来执行复杂查询的多维表达式(MDX)查询语言,SQL Server立方体和PivotTable 服务,管理和协调资源的高效公用程序等等。

本书是有效解决企业数据仓库建设问题的最佳参考资料。书中不仅讲述了SQL Server 7.0如何帮助你创建数据仓库,而且介绍了其强大功能的执行情况。对从事数据库管理的人员来说,是一本极好的参考书。

Jake Sturm: Data Warehousing with Microsoft SQL Server 7.0 Technical Reference.

Copyright © 2000 by Microsoft Corporation.

Original English language edition copyright © 2000 by Jake Sturm.

Published by arrangement with the original publisher, Microsoft Press, a division of Microsoft Corporation, Redmond, Washington, U. S. A. All rights reserved.

本书中文简体字版由美国微软出版社授权机械工业出版社出版。未经出版者书面许可,不得以任何方式复制或抄袭本书内容。

版权所有,侵权必究。

本书版权登记号:图字:01-2000-0802

图书在版编目(CIP)数据

Microsoft SQL Server 7数据仓库技术指南/(美)斯蒂姆(Sturm, J.)著;刘汉字等译.

- 北京:机械工业出版社,2000.9

(微软公司核心技术书库)

书名原文:Data Warehousing with Microsoft SQL Server 7.0 Technical Reference

ISBN 7-111-08201-X

I. M… II. ①斯… ②刘… III. 关系数据库-数据库管理系统, SQL Server 7 IV. TP311.138

中国版本图书馆CIP数据核字(2000)第40988号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:赵红燕

北京忠信诚胶印厂印刷·新华书店北京发行所发行

2000年9月第1版第1次印刷

787mm×1092mm 1/16·21.5 印张

印数:0 001-6 000册

定价:49.00元(附光盘)

凡购本书,如有倒页、脱页、缺页,由本社发行部调换

前言

新千年已经来临，公司企业要想取得成功，必须解决如何恰当管理公司信息这一极为关键的问题。技术、市场以及公司开展业务的整体方式可能在几个月乃至几个星期内发生根本性变化。一个在昨天只能靠连锁店出售产品的公司到了今天完全可能通过因特网进行营销活动。对于一个公司来说，要想作出明智决策、永远不落后于飞速变化的潮流，首先就必须拥有它自己能够有效应用的准确信息。

一般而言，公司企业都拥有大量准确信息，但往往缺乏知识管理技术，然而正是这些知识管理技术可以使它们极为方便地访问会对其规划和决策产生影响的重要数据。所谓知识管理，要做的是使用可让公司员工访问以各种文档格式排列的信息的系统。这些信息可能包括以往的经验教训、当前或以前的开发项目文件、数据库中的数据、网络结构文件、最佳方案等内容。任何用于公司决策的信息都应该成为知识管理系统的组成部分。在浩繁的信息散布于文本文件、数据库、邮件服务器或其他数据存储器中的情况下，如何找到自己需要的准确信息即便不是根本无法完成的任务，也是相当艰巨的。在竞争激烈、瞬息万变的环境下，决策时如果没有足够的信息作依据，往往有可能断送一个公司的前程。

有的时候，问题不是信息太少，而是太多。凡是有过一次上网就能查询出数百种结果的经历的人都知道，信息过滥其实与缺乏信息一样不能给人以帮助。投入使用的数据库都必须具有适当的过滤功能，这样才能只把使用者需要的信息提供给他们。综合考虑了使用者各种使用要求的Microsoft SQL Server 7.0数据仓库就具有这种过滤功能，可以把适当的信息在适当的时间提供适当的对象——而这正是公司取得成功的关键所在。

微软公司为使用者建立知识管理系统提供了4个强有力的资源：对象链接与嵌入数据库（OLE DB）、数据转换服务（Data Transformation Services, DTS）、Microsoft SQL Server 7.0 联机分析处理服务（OLAP Services）和数据透视表服务（PivotTable Services）。OLE DB可以与任何数据库建立通信联系。有了OLE DB，便可以访问文本文件、Microsoft Exchange Servers、SQL Server 7.0数据库、Microsoft Access数据库和Oracle 8数据库中的数据。也就是说，可以访问OLE DB供应商提供的任何数据库中的数据。在本书编写付印期间，OLE DB供应商已将几乎所有重要数据库都推向了市场。拥有访问数据的能力是建立知识管理系统的第一步。

由于数据可能以多种不同的格式散布于公司的各个部门，经理们往往需要有一种工具，用以随意传送数据并在必要时将其转换为一种标准格式供全公司人员使用。作为SQL Server 组成部分之一的DTS可以使使用者拥有传送、转换、清除和验证数据的能力。

在能够访问和传送数据之后，你还需要各种高级工具来查询和分析数据，然后将数据以准确且实用的格式提取出来。OLAP Services可以执行查询和分析数据仓库中数据的任务，而PivotTable Services则可以把数据显示给使用者。把OLE DB、DTS、OLAP Services和PivotTable Services结合为一体，可以形成一个功能强大的系统，使你能够把适宜的信息在适宜的时间传送到适宜的地点。这样的信息管理可以帮助公司经理们在较短时间内作出更富远见卓识的决策。本书将说明如何利用这4个强有力的资源建立起覆盖公司所有部门的系统，通

过Microsoft SQL Server 7.0数据仓库进行知识管理。

本书适用对象

本书是为所有希望深入了解SQL Server 7.0数据仓库的人士编写的。对于SQL Server 7.0数据仓库的管理员，本书提供了从创建和设计数据仓库的基本概念到优化数据库等高级主题的全部内容。对于开发人员，本书有一章专门介绍了ADO MD，并在一个附录中大量列举了Visual Basic代码样本。对管理员和开发人员，本书中有三章专门介绍MDX内容。

本书主要内容

本书由13章和5个附录组成。开头的几章论述了建立数据仓库的基本概念，向不太了解数据仓库的读者提供了他们建立SQL Server 7.0数据仓库时所必须具备的基础知识。对于较有经验的读者，这几章可以作为复习资料。本书说明了与数据仓库关系密切的SQL Server 7.0的新特点，其中包括DTS、PivotTable Services、OLAP Services以及它们的使用方法。读者可以通读全书，也可以略过论述MDX的3章；对于这3章的内容，读者完全可以根据自己的兴趣和相关知识水平来决定是否需要阅读。

第1章回顾了公司知识管理的发展史。有关SQL Server数据仓库建库系统的所有重要组件的定义均在本章列出，其中包括OLAP Services、OLE DB、DTS、向导和Microsoft Office 2000。本章还定义了分析数据和操作数据的差别，并对分析数据进行了详细描述。

一个数据仓库的建立绝对离不开可以满足公司业务要求和使用者需要的高超设计。你可以利用第2章介绍的微软解决方案框架（Microsoft Solutions Framework, MSF）确定自己的系统目标和设计方案。MSF还可以帮助你把设计思想转变为配置于公司各部门的高效率工作系统。本章说明了如何应用MSF来创建数据仓库，同时还讨论了所有相关人员应如何组合成一个整体高效率工作的团队的问题。

第3章提出了在线分析处理（OLAP）理论（这一理论是建立分析数据仓库系统的基础），同时还介绍了标准和非标准表、维、信息对象、星型模式、立方体等OLAP基本概念。

第4章有关建立OLAP Services数据仓库的内容是第3章所述OLAP理论的实际应用。本章说明了应该如何使用OLAP Services中的向导建立数据仓库。

第5章介绍了MDX的基本语法；MDX是一种查询语言，可用来对数据仓库中的数据进行分析。本章还列举了MDX的使用实例。

第6章扩展了MDX语言的语法，包括计算成员。计算成员使你可以通过数据仓库中已有的成员创建新的数据仓库成员，从而得以创建任何定制成员，例如，它具有根据总销售额和货物的总成本计算利润的能力。

第7章讨论的MDX表达式是该语言的一个补充部分，可以使你运用数据仓库中的数据进行各种操作。通过这些表达式，可以进行强有力的数据查询，为哪怕是最复杂的分析问题找到答案。

第8章涉及的内容是ADO Multidimensional（ADO MD）。这是一种接口，可把Visual Basic等开发工具通过OLE DB与数据仓库连接起来。ADO MD可以使你非常方便地把功能强大的Visual Basic、Microsoft Office、ASP（Active Server Pages）等工具用于各种用途。本章讨论了ADO MD的对象层次、它可以检索的信息、它的工作原理以及你可以利用ADO MD层次建立什么类型系统等内容。

数据仓库是较大型企业系统的组成部分。这样的系统还需包含从数据仓库中提取数据并将其展示给使用者等的大量应用。第9章讨论了数据仓库管理员经常会遇到的几种应用程序设计问题，同时还说明了Visual Basic项目设计的各种设计原则。

第10章讨论的是Data Transformation Services (DTS)。这是一种可以使你非常方便地传送、清除和转换数据的工具。DTS 可以通过OLE DB把信息从一个数据库传送到另一个数据库。你可以根据本章的提示，在随本书所赠的CD光盘上对含有1998年12月份数据的Microsoft Access数据库进行操作。你可以先把数据传送到一个升级数据库中 (FoodMart.mdb)，然后用新信息对数据仓库进行更新。

第11章讨论的数据透视表服务 (PivotTable Services) 可以把OLAP服务器和所有需要与之通信的外部应用程序连接起来。ADO MD、Microsoft Excel、Visual Basic等应用程序都需要使用PivotTable Services。在这一章，你可以学会如何使用PivotTable Services在应用程序与客户端之间建立立方体。PivotTable Services具有许多新的特性，如客户端的高速数据缓存和元数据等，从而大大提高了分析查询的效率。

Microsoft SQL Server 7.0带有各种可供你用来优化数据仓库的工具。第12章对这些工具做了介绍，同时还列出了一整套优化SQL Server数据仓库的最佳方案。

最后一章对通过SQL Server数据仓库建立起整个系统的过程进行了总结，讨论了数字神经系统和数字仪表板 (dashboard) 这两个构成了知识管理系统基础的概念。

附录A列举了MDX的实例。附录B介绍了应该如何把来自Microsoft知识库的各种信息以一种叫做开放信息模型 (Open Information Model, OIM) 的标准格式储存起来。在附录C中，你将学会如何创建三个组件把ADO MD包装起来。附录D介绍了Microsoft SQL Server 7.0 的英语查询 (English Query)，通过这种工具，使用者可以用英语而不必用正规SQL语句进行数据库查询。附录E介绍了决策支持对象 (Decision Support Objects DSO)，这是一系列对象、接口、类集和枚举，供使用者对SQL Server数据仓库进行有计划的管理。你可以在DSO内建立自己的OLAP管理程序，以确保各单元的安全。

随书附赠光盘的使用

随书附赠光盘中含有白皮书、SQL Server 7.0的Service Pack 1、帮助文件、代码样本和MDX查询文件。光盘上的材料是为帮助你理解本书内容而设计的，同时还提供了你在把Microsoft SQL Server 7.0用做数据仓库时所需要的各种资源。在使用光盘资料时，需要满足以下系统要求：

- 486或以上处理器。
- Microsoft Windows 95、Windows 98、Windows NT 4.0或以上版本。
- 32MB内存。
- CD-ROM驱动器。
- 鼠标或其他指示器 (推荐)。
- 要处理Visual Basic文件，需安装Visual Basic 6。
- 要处理doc文件，需安装Microsoft Word。
- 要处理Foodmart.mdb数据库文件和MDX文件，需在Windows NT 4.0或以上版本上安装SQL Server 7.0。

第1章 企业信息共享

10年来,科学技术飞速发展。许多几年前还不知为何物的技术如今已经风靡全球,其中包括利用电子数据交换(EDI)和因特网商务(电子商务)技术进行的即时交易。当前的市场形势促使公司企业必须保持旺盛活力。今天的公司要想在激烈的竞争中立于不败之地,就必须做到:

- 认清公司可能面临的所有风险。
- 优化业务程序。
- 在竞争白热化的市场上尽力实现利润最大化。
- 认清会对公司造成影响的市场变化和市场机会并及时作出反应。
- 在飞速变化的市场上控制风险和把握机会。
- 在不必对公司结构作重大调整的前提下适应市场和技术的变化。

其实,达到这些目的的关键就在于适时掌握准确信息,利用这些信息作出正确决策以及最终及时贯彻这些决策。

对于公司企业来说,获得和利用信息的方式就是建立覆盖公司所有部门的企业综合信息系统。拥有了这种企业综合信息系统的公司更有可能在现代经济中保持竞争优势和盈利能力。而没能成功建立起这种系统的公司则可能始终在现代经济中苦苦挣扎。

企业综合信息系统需要使用多种技术、不同类型的计算和操作系统以及各类数据仓库。要想在一个综合系统中把所有各个部分结合为一个整体,首先需要对各种应用程序和基础结构作周密规划,以使公司提出的各项任务都能高效率完成。这种系统的核心是存储于各个数据源中的所有信息的可访问性。这些信息必须是一致、准确和经过精细定义的。

Microsoft SQL Server 7.0可以储存、管理和分析小到几兆字节容量、大到数兆兆(a few terabytes)字节容量的各种数据库。把Microsoft Visual Studio和Microsoft BackOffice的强大功能结合起来之后,就可以建立起一个企业综合信息系统,它不仅会覆盖企业的所有部门,同时还可以在必要的时间和必要的地点提供你所需要的任何信息。此外,SQL Server 7.0、Visual Studio和BackOffice还可以提供把信息转变为适时工作方案的资源。

本书将说明创建可使相关人员访问企业内任何信息的SQL Server分析数据仓库的方法。通过微软开发工具建立的应用可以访问系统中的信息并以使用者要求的任何格式把信息提取出来。建立这样的数据仓库是公司保持可持续发展的关键所在。本书向你提供的,是对于公司开发人员来说极其重要的信息资源。

1.1 方案核心:操作数据和分析数据

公司在作出明智决策和圆满贯彻决策的过程中需要使用两种类型数据:操作数据和分析数据。这两种数据都可以存储在SQL Server 7.0中。

1.1.1 操作数据

操作数据是处于不断变换和更新之中的,属于动态数据。订单输入数据库中的当前订单

就是这类数据的一个例子。操作数据所代表的是某一时间点的当前信息。操作数据可以表明待发订单的状况、活期存款帐户的当前余额或当前脱销产品的数量。这类数据可以告诉我们某件事情的现状，而且它们在任何时候都是处于变化之中的。

1.1.2 分析数据

另一方面，分析数据是历史数据，通常都不会随着时间的推移而发生变化，因此属于静态数据。只有在原始信息错误的情况下，分析数据才会有变动。某个时间点的销售额是最终数据，不可逆的。在这个时间点，信息变成了静态，因此可以从动态数据源迁移到静态数据源。分析数据可用于查看跨时间段的信息。例如，我们可以查看1月份的总销售额、7月4日在华盛顿市出售国旗的数量或者开发人员在过去6个月里的工资变化。

分析数据通常来自操作数据，可用来对公司在某个时期的运营状况作全面分析。这种数据必须准确、可访问，并且能以实用的格式显示出来。SQL Server 7.0将使你拥有建立此类数据的能力。

分析数据并不一定要完全来自公司内部。几年以后，社会上就会有大批外部数据中心出现，向各公司企业提供相关服务（或许是交纳一定费用后，便可通过因特网访问这些数据中心）。到那时，来自外部的额外信息可以补充到公司的数据库之中。

SQL Server可以访问各种数据源的数据，把它们转变为一种统一的格式，校验数据的完整性，最终把数据存储到一个在线分析处理（OLAP）服务器中供使用者方便访问。通过对象的链接和嵌入数据库（OLE DB）接口——这是一种API接口，可访问任何数据仓库中的任何信息——你可以建立SQL Server数据仓库和数据中心，从而按公司的要求访问任何信息。

1.1.3 分析数据的历史：从电子表格到数据仓库

许多年来，对数据进行分析一直是公司企业工作进程中的一个基本环节。如今，公司为了在现代市场的激烈竞争中立于不败之地，对获取信息的速度要求大大超过了以往。在过去20年里发生的巨变之一便是数据分析的速度有了极大的提高。

20世纪60年代以前，所有分析工作都是在纸张文件上进行的。像合计月总销售额这样的简单计算工作既费时又费工，而比较两个地区间销售额之类的事情当然就更复杂了。许多公司意识到，要想进行这样的比较，首先必须在公司内部建立一种统一的系统。没有这样的统一系统，就不可能确定一种产品在加利福尼亚一家商店的销售量是否高于纽约的另一家商店。

许多公司采取了标准部件和产品号码、统一编写报告等措施。有了这样的系统之后，相关人员就能比较容易地把各种信息收集到综合报告里，给出详细和归纳性的信息。利用人力处理这些信息难度很大，同时又很费时间，但是在一种统一的系统帮助下，这样的工作至少有了完成的可能性。在公司里开发一种表述信息的统一系统到现在仍旧是分析工作的一个基本组成部分，同时也是创建分析数据库之前必须完成的工作之一。

到了60年代，一些大公司开始利用非常昂贵的计算机进行分析工作。当时，这些系统运行速度缓慢，一次简单的查询就要耗费很长时间。储存和检索数据的效能之低，以至于配备了计算机的公司还是把大部分信息储存在基于纸张的系统中。

在最近的20年里，高效率计算机系统纷纷问世，其中包括电子表格、执行信息系统和OLAP。

1. 电子表格

电子表格是展示分析信息的一种绝妙工具，但是它的最大问题在于输入信息非常麻烦。

在许多公司，信息都是靠人工输入的。这样的系统其实并不比前期基于纸张的系统方便多少。电子表格具有计算功能，但是它需要有人把所有信息收集并输入到电子表格中。执行信息系统便是对电子表格的改进。

2. 执行信息系统

随着计算机在公司里应用得越来越普遍以及存储数据越来越容易并且更加便宜，大量信息都被存储到了数据库里。公司出于对自身业务情况进行评价的目的，需要对这些数据进行分析。要想完成这样的分析工作，必须用一整套应用程序把数据库中的信息经过归纳融入一份报告之中。在经过漫长的信息存储过程之后，对所存储的信息进行极端复杂的查询可以为这样的报告提供所需的信息。这类查询工作由于需要使用大量时间和资源，因此必须通宵达旦地进行。称作执行信息系统的专用软件就是为进行这样的分析而专门开发的。

开发人员通常要用若干天乃至数周时间来建立复杂的查询系统并编写将信息输入电子表格的代码。为了提供适用于公司不同部门的报告，往往需要开发出几十种应用程序。而为了支持这些应用程序，公司的各个部门也需创建规模较小的数据库以容纳与自身业务相关的信息。这些小数据库可以在同一核心信息的基础上创建，但是它们往往拥有完全不同的数据格式、字段和表格名，通常含有专门的字段。由于这些小数据库之间存在着这样的差别，因此它们之间不可能实现信息共享，即使它们是在同一大规模核心数据的基础之上建立的。这便形成了所谓的“数据岛”。

来自不同部门并建立在不同数据库基础上的报告往往会对相同内容的计算得出不同的数值。由于每个部门都就自己的关键性业务活动提出相关数字，而且各部门的数字不尽吻合，公司开起会来自然就妙趣横生了。一般来说，区分哪个部门的数字准确、哪个部门的数字不准确是不可能的事。这些差异通常会导致更加专门化的报告和数据库，甚至会出现数字自相矛盾的现象。改变这种状况的第一步是开发OLAP工具。

3. OLAP工具：创建数据市场和数据仓库

OLAP工具利用一种叫做多维结构的专用数据结构对历史数据进行复杂分析。这些结构存储在一个叫做数据仓库的专用数据库里。数据仓库在以恰当的方式建立起来之后，可以为整个公司提供前后一致的企业历史数据。还可以使用数据市场——数据中心与数据仓库基本类似，只是它仅收入一个机构（或部门）的数据，或者只适用于一项业务内容。第3章将用专门篇幅讨论这些结构以及用它们建立数据仓库的过程。

OLAP工具免除了为每一次分析专门开发应用程序的需要。它还允许使用者提出“5月份在纽约和芝加哥向30岁以上消费者出售了多少小器具”之类的分析要求。微软的OLAP工具（SQL Server 7.0 OLAP Services）是与另一种工具PivotTable Services一同问世的，后者可以把信息直接放入Excel电子表格。此外还有许多第三方工具，它们可以用来查看电子表格格式的数据。由此可见，OLAP Services工具可以使使用者快速高效地进行任何分析查询工作。PivotTable Services可以把得自OLAP Services的信息转换为使用者熟悉的格式。

OLAP Services可令你只需开发一种应用程序便能进行任何分析查询，而不必像以前那样需要为每一种查询专门开发专用的应用程序。OLAP Services还可以提供比SQL Server存储程序效率更高的数据查询方式。

不过, OLAP Services并不能解决不同部门各自建有“数据岛”的问题。只有在组建一个负责公司总体数据仓库创建工作的小组, 并且公司所有部门都通力合作, 为创建一个统一的企业系统而共同努力的情况下, 数据岛的现象才有可能避免。通过Data Transformation Services (DTS), 你可以把所有数据岛以一种统一的格式组合到一个中央数据仓库中, 建立起一个通用企业系统, 把数据岛连接成一个数据大陆。这样, 尽管OLAP Services不能自动解决这一问题, 负责创建数据仓库的工作小组与各部门相关人员之间通过密切合作可以找到解决问题的办法。

你必须与用户合作, 在你的公司与用户之间建立起连接各自数据岛的桥梁。OLAP Services仅仅是一种工具。如果使用者不能或不愿意建立统一的企业系统, OLAP Services也无法耍魔术般变出一个来。

建立数据仓库首先需要有一张总体规划图。你拥有一家公司, 又拥有一个包含数据仓库在内的企业系统。要想确保你的数据仓库计划取得成功, 必须要解决公司环境和技术两个方面的问题。下面, 我们看看一种规模较大的系统, 这种系统叫做决策支持系统, 我们的数据仓库就是其中的基本成分。

1.2 创建SQL Server 7.0决策支持系统

我们在讨论SQL Server 7.0和建立决策支持系统时, 内容绝不仅仅涉及分析数据库。分析数据库只是一个更大系统的组成部分。这一系统将含有存储和管理多维数据立方体的工具 (OLAP Server和OLAP Manager), 显示数据的工具 (PivotTable Services、Excel、Visual Basic等) 和在需要时转换数据的工具 (DTS)。它是微软总体方案中多个功能强大的支持点中的一个; Microsoft BackOffice、微软开发工具、SQL Server第三方应用程序和SQL Server的结合使用, 将使你拥有建立决策支持系统每个元素所需的全部工具。我们的系统将如图1-1所示。

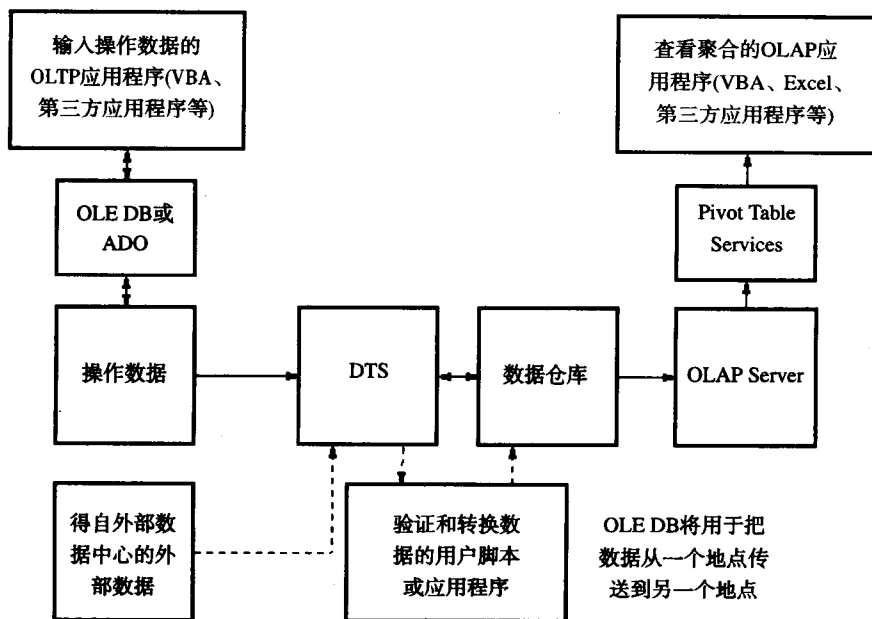


图1-1 构成强大决策支持系统的数据库和工具

1.2.1 决策支持系统的目标

决策支持系统可以实现公司的以下目标：

- **信息的深度** 决策支持系统必须允许使用者查看到从最高层次到最低层次的所有信息。决策支持系统的某个使用者可能需要先从东海岸的消费者数据查看起，然后到纽约州的消费者数据，最终到纽约市的消费者数据。而当该使用者到达纽约市这一层次后，他又可能需要首先查看消费者中把男性工资收入分为高、中、低三个等级的数据，然后把把这些数据与女性消费者中相同工资收入等级的数据进行比较。此后，该使用者还可能需要将相同的分类数据与芝加哥的相关数据进行比较。这就是决策支持系统所必须向使用者提供的东西：任何层次的所需信息。
- **信息比较** 前面提到的数据比较，如纽约市男性与女性消费者之间的工资收入比较，是决策支持系统的基本特性之一：具有对各类信息进行比较的功能。最常见比较之一是对业务活动的比较，如某一时期的销售额与另一时期销售额的比较。某个使用者可能需要把本年5月份的销售额与去年5月份的销售额进行比较。或者，某个使用者可能需要把一次大型广告宣传活动之前与之后的销售额进行比较。决策支持系统必须使使用者能够轻而易举地进行这些比较。
- **可用的信息** 如果信息不准确或者不适于需要解决的业务问题，系统即便是具有检索大量信息的功能也毫无用处。因此，建立的系统必须能够提供足够有效的信息，以帮助使用者作出可以给公司带来收益的业务决策。
- **及时的信息** 如果使用者在几个小时之内就需要某种信息，然而系统用了几天乃至几个星期才把数据传送到数据仓库，那么这样的系统不会带来任何益处。相关的信息即便再准确，来得过晚对公司也毫无意义。在有竞争者刚刚侵入你的市场、仅夺走你的市场份额百分之几之时，就找出这个对手并分析出他这样做的原因，可以帮助公司及时制定出应付这一新竞争情况的对策。如果已有四分之一乃至更多的消费者被别人挖走之后你才有所察觉，这时无无论你采取什么措施恐怕都于事无补了。
- **快速分析** 使用者可能要用多种方法分析数据，以达到对某一业务活动进行评价的目的。每次查询都必须能够快速进行，这样才能使使用者在合理的时间内作出几十次查询。
- **可访问的信息** 信息应该用商业语言而非分析系统语言显示。接口应该允许使用者用他们熟悉的词汇和语言进行数据分析。

以这种方式建立起来的系统可以用来对公司的销售额、人员现状以及消费者欺诈行为的可能性进行周密分析。它还可以识别高风险消费者和因特网购买趋势等情况。总而言之，满足上述要求的决策支持系统可以为公司几乎所有业务活动提供有用的详细信息。

1.2.2 规划决策支持系统

建立这样一个系统需要周密的规划和对现有不同选择方案的深刻了解。得到微软其他产品支持的SQL Server 7.0可以满足上述所有要求。

必须解决的难度最大的问题之一，是如何把散布于各种潜在来源的数据集合为一个整体。其中有些数据所使用的可能并非实用格式，在用于数据仓库之前可能需要进行强化处理。有些数据将采自为公司各部门的特殊用途而建立的数据岛。这些都是非常困难的问题，必须在

建立数据仓库之前得到解决。通过对企业情况的仔细评价、周密规划、风险分析和追踪调查,能够识别出这些问题并制定出计划解决它们。

下面,让我们来看一看能为你建立决策支持系统提供帮助的SQL Server的特点。

1.3 SQL Server 7.0决策支持系统的工具

SQL Server 7.0分析系统能够有效利用系统的多个不同的组成部分。微软为创建决策支持系统提供了以下工具:OLAP Server、PivotTable Services、OLE DB、SQL Server Data Transformation Services、SQL Server多维数据库、大量SQL Server向导和设计软件、Office 2000和Microsoft Repository。本书将对这些工具逐一加以介绍。下文将对其中的几种工具作一扼要介绍。

1.3.1 在线分析处理服务

如上文所述,在线分析处理服务(OLAP Services)可以使我们快速高效地进行复杂的分析查询。OLAP Services将找到效率最高的方式进行查询。在OLAP Services的帮助下,你可以在存储于物理硬盘驱动器上的数据量与通过所谓聚合进行分析的速度之间进行选择。

所谓聚合,是指对数据的合计。这是我们在分析数据(如总销售额)时常常要做的工作。OLAP Services具有最常用的聚合功能,可以把聚合后的数据存储到数据库中,而这正好可以节约分析数据的时间。当然,你在数据库中保存的聚合数据越多,数据库就会越庞大。因此,我们应该在数据的规模与进行分析查询时所需耗费的时间之间作一权衡,找到一种折衷的办法。通过OLAP Services,可以非常方便地在数据库规模与分析速度之间找到最佳设置。

1.3.2 数据透视表服务

数据透视表服务(PivotTable Services)允许应用程序从OLAP Server上请求调用一套数据,在Excel等应用程序把数据显示给使用者后再将数据归还。PivotTable Services还可以把信息高速缓存在用户的计算机上。这就是说,使用者可以连接、请求调用并在线查看信息。此外,PivotTable Services还可以通过Visual Basic应用程序进行数据访问,如此一来,便可以建立起自己的用户应用程序进行分析查询了。

1.3.3 对象链接和嵌入数据库

对象链接和嵌入数据库(OLE DB)API是通用访问的基础。OLE DB是一种接口,数据库应用程序可以通过它访问OLE DB供应商提供的任何类型数据。有了OLE DB,可以对数据仓库的多维结构进行检索和处理。OLE DB还允许我们查看和处理其他数据库的数据,如ORACLE数据、电子邮件文件中的数据、文本文件等。ActiveX Data Objects(ADO)是Visual Basic编程人员据此访问OLE DB接口的工具;也就是说,ADO是OLE DB API的封装。在我们的编码实例中,我们将频繁使用ADO。

1.3.4 数据转换服务

数据转换服务(DTS)允许我们在任何数据源之间进行数据的虚拟输入、输出和转换。为了达到这一目的,DTS通过OLE DB把不同的数据源连接起来。DTS支持多种数据格式,其中包

括美国国家信息交换码(ASCII)。可以参照DTS向导进行DTS转换,也可以通过VBScript或JavaScript等脚本语言进行转换编码。亦可编写出Visual Basic应用程序进行DTS转换。

1.3.5 SQL Server 7.0多维数据库

SQL Server 7.0允许你使用它的关系数据库存储分析数据,或者把信息存储在SQL Server多维数据库中一种叫做立方体的特殊结构中。如本书第3章所述,多维数据库是专为提高分析查询效率而设计的。多维数据库与关系数据库之间的差别将在第4章中讨论。

1.3.6 向导和设计软件

SQL Server 7.0中含有多种设计工具。这些工具可以使你非常方便地备份数据库、创建聚合、查看数据仓库的结构以及建立和设计立方体。通过易用的图形用户接口,你几乎可以做所需要做的事情。

1.3.7 Office 2000

通过Excel等办公产品,你可以查看电子表格中的立方体数据。你还可以编写Visual Basic for Applications(VBA)代码和建立Excel模块,用以检索和处理数据。Office 2000允许你为数据仓库建立用户化的前端。

1.3.8 微软知识库

微软知识库(Microsoft Repository)可向你提供以一种标准格式存储数据仓库结构的场所。统一建模语言(Unified Modeling Language, UML)工具、其他数据库或任何可以解读本知识库的附加系统均可解读此种标准格式。总之,本知识库具有信息共享的功能,其中包括你的数据仓库结构。

1.4 小结

与微软的其他产品结合使用的SQL Server 7.0可向你提供功能强大的一整套工具,使你能够创建一个完整的决策支持系统。此决策支持系统能够以易于访问的格式在使用者需要的时候向他们提供准确的信息。只要规划和设计得当,公司决策者对创建决策支持系统有清醒的认识,SQL Server数据仓库就能对如何提高公司运营效率、优化业务活动的决策提供巨大帮助。

第2章 规划和设计企业分析系统

周密的规划和严格的管理，是设计和创建包含Microsoft SQL Server 7.0数据仓库的分析系统所必不可少的因素。创建分析系统应遵照以下步骤进行：

- 设计数据仓库的数据结构。
- 设计用于查看数据仓库中数据的用户应用程序。
- 设计数据转换方式，以将历史数据库或在线交易处理（OLTP）数据库中的数据传送到数据仓库。
- 推广应用数据仓库。
- 配置应用程序和传送数据。

通常，数据仓库也是规模更大的企业系统项目的组成部分之一，该系统项目的内容包括创建或升级OLTP系统以及用于查看分析信息的前端应用程序。如果数据仓库的规划、设计和创建缺乏经过精确定义的框架，项目可能会归于失败。所谓框架，是指可以向你提供据此创建一个完整系统的结构的某个东西。建立数据仓库的适宜框架应该具有三大基本特点：

- 可以使你为项目作出技术决策的观察点。
- 可以使你确定项目是否能准时完成的计量参照点。
- 以灵活方式反复运用现有知识的能力。

Microsoft Solutions Framework（MSF）就是具有这三大特点的一种框架。通过MSF创建的系统还应具有以下特点：

- 系统可以被分解成若干小系统。这些小系统既可以逐一运行，也可以同时运行。因此，系统可以覆盖很长一个时期，但所获得的依然是短期结果。
- 系统是按不同的部分分别建成的，既可以一部分一部分地创建，也可以各部分同时创建。
- 系统是分多个阶段建成的，每个阶段都建成系统的一个部分。
- 系统项目靠各工作小组协同合作完成。

MSF把系统项目分成了设想、规划、开发和稳定四个阶段。这四个阶段如图2-1所示。

注意，上述过程是呈螺旋上升式的，并非简单的循环往复，因为任何企业系统通常都不可能在一个阶段内创建完成，相反，需要经过多个阶段，而每个阶段都会给系统增添新的功能。

例如，这样的系统可能涉及新建一个电子商务网站。第一阶段包括创建一个具有基本指令功能的万维网网站。第二阶段包括建立一个数据仓库，以存储有关系统使用者的信息，如如何浏览网页等。在第二阶段开展工作的同时，建立用于分析和查看数据仓库中信息的应用程序的第三阶段的工作也开始进行。第四阶段根据对数据仓库数据的分析给系统添加功能。电子商务网站是通过不断反复的步骤建成的，每个步骤都能给整个系统增加新的功能。

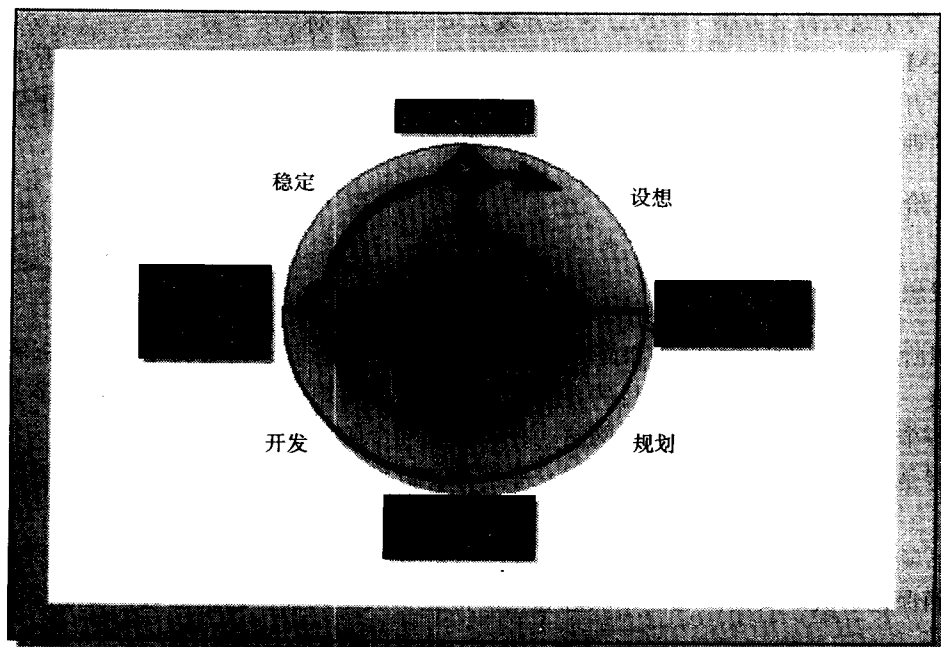


图2-1 设想、规划、开发和稳定

网站建成后，各阶段的工作也未必就完全结束。技术、市场以及公司本身都是在不断变化的。这些变化可能要求给网站添加新的功能，如此一来就要开始新的循环。从这个角度看，各阶段的工作永远不会有终结；而系统为了适应当前的业务需要必须不断发展和变化。由此可见，你将处于一个永远不会结束的规划、创建、管理、再规划……的循环往复之中。

2.1 里程碑

上述四个阶段中每个阶段工作的完成都是整个系统项目中的一个经过精确定义的计量点。这些经过精确定义的计量点叫作里程碑。里程碑通常都规定有交付使用指标，可用来确定系统项目的工作是否在按计划进行。一个里程碑所涉及的工作既可能已经完成，也可能并未完成。例如，一个里程碑可能涉及把所有数据输入数据仓库。你可以为此规定一个日期，譬如5月5日。在5月5日这一天，数据可能已经全部存储到数据仓库里了，也可能还没有。如果数据均已输入数据仓库，你就按计划完成了工作；如果没有，则你就没有按时完成计划。各阶段工作都能达到里程碑可以确保你圆满完成整体工作计划。

在某一阶段的工作达到里程碑后，你便可以进入下一阶段的工作。开始下一阶段工作的唯一标准是前面这个阶段的工作必须达到了相应里程碑的要求。项目实施的四个阶段中的每个阶段都有四个里程碑：设想/范围得到批准、项目计划得到批准、按原定范围建成/首次试用、获准投入使用。所有里程碑都应建立在二进制的基础之上，即或真或假。

在MSF中，所有工作小组成员共同承担系统项目的责任。每个工作小组成员都对系统项目的成功直接负责。MSF可以帮助开发人员对系统的各个组成部分进行设计，以实现最终用户和管理人员的目标。MSF要求开发人员具有创新能力，充分发挥自己的才智，在系统项目规定的范围内尽可能设计出质量上佳的系统组成部分。根据最终用户和用户管理人员的需要

设计系统各个组成部分或整个系统，这是开发人员的第一目标。

使用MSF的一大优势是它既可以用于开发软件，也可以用于配置基础结构。由于分析系统涉及了开发用户软件、数据转换软件和创建数据仓库设施等项内容，因此，MSF是实施系统项目的理想选择。

2.2 各阶段工作综述

2.2.1 设想阶段

设想阶段的工作是定义系统和提出设想/范围陈述。设想陈述将定义系统的目标和提出系统的发展。范围陈述定义当前项目的界限。例如，某个设想可能是建立一个需要多个阶段建成系统各个部分的完整的电子商务网站。在经过头几个阶段建好基本网站后，下一个项目的范围可能是为该电子商务网站建立数据仓库部分。以后的项目可能是完成设想陈述的其他部分。

如果项目建成了系统各个组成部分，则添加功能的工作可以留待以后进行。设想/范围陈述可以确保把解决必要的问题。当设想和范围都被精确定义并得到批准时，设想阶段便告结束。

2.2.2 规划阶段

规划阶段由编写一套定义使用者要求的文件、确定需用来满足这些要求的系统各个组成部分和服务以及对这些系统组成部分的周密设计进行论证三个部分组成。如果项目内容是建立数据仓库，这些系统组成部分应该包括用于转换和传送数据的Visual Basic程序、用于查看数据的Excel程序以及用于创建数据仓库表格结构的工具等。上述文件完成之时就是本阶段结束之日。规划阶段可以确保你对整个系统作出恰当定义。

2.2.3 开发阶段

开发阶段由开发和测试系统各个组成部分以及就系统各组成部分对整个系统的适用性进行评价三个部分组成。如果项目内容是创建数据仓库，本阶段的工作应该包括建立数据仓库和开发用户应用程序。当所有新开发工作完成时，本阶段便告结束。开发阶段可以使你有机会创建出质量最佳的系统。开发阶段结束后，给系统添加任何功能都意味着一个新的阶段的开始。将来非常有可能出现需要给系统增加功能的情况，因为设想陈述的内容往往会超出本阶段的范围。因此，一个项目可能会多次循环往复。

2.2.4 稳定阶段

稳定阶段侧重于测试完全建成的系统，寻找和确定所存在的任何问题，并且配置系统的各个组成部分。这一阶段的工作还包括向数据仓库传送数据。当系统测试完毕交付使用时，本阶段便宣告结束。

2.2.5 目标和交付使用指标

每个阶段的工作都有各自的目标。其中每个目标都应该有：

- 一个目标。
- 一系列为达到目标而实施的行动。
- 一系列达到目标时的完工交付使用指标。

交付使用指标是指完工后的系统项目的组成部分；它们由UML图表、数据库模式、代码模块等组成，需经工作小组全体成员检验合格后方能批准系统投入使用。

2.2.6 三类相关人员

大多数系统项目基本上都有三类人员涉足其中：管理人员、最终用户和系统开发人员。每类人员对系统项目提出的目标各不相同。客户管理人员的目标侧重于系统项目将如何提高公司的运作效率和盈利能力。管理人员要求有证据证明系统项目将能对投资作出回报。最终用户的兴趣则在于交付使用指标，如他们可以用来完成自身工作的软件以及他们将怎样利用软件使自己工作起来更方便等。最终用户还对系统项目如何帮助他们进行日常工作感兴趣。开发人员的兴趣在于如何创建起整个系统。在数据仓库项目中，最终用户可能同时也是管理人员。

在设想和规划阶段，所有三类人——管理人员、最终用户和开发人员都会对项目目标的定义作出贡献，但是每类人都各自从不同的角度给出定义。

系统项目的总体目标和目的以及最终的交付使用指标都在设想阶段写入相关文件，而实现总体目标所必须采取的各种行动则在规划阶段写入相关文件。

1. 管理人员

管理人员的目标侧重于系统项目对公司经营活动的影响，尤其是系统能如何最大限度地提高经营活动的效率。他们的目标还侧重于系统项目的总体目标、目的及其交付使用指标。这些目标需要在设想阶段经过精确定义后写入相关文件之中，因此，管理人员提出的目标应在项目建设的设想阶段之初得到透彻分析研究。最终用户在设想阶段的后半部分提出自己的目标，对管理人员的目标起到补充作用。最终用户侧重于最终建成的系统将会怎样提高他们的工作效率。系统开发人员的目标则侧重于系统的总体结构、系统运行平台等。在设想阶段，管理人员就系统项目的设想和范围提出详细意见，而最终用户和系统开发人员则给这些设想和范围补充新的内容。

2. 最终用户

最终用户主要在系统的外围部分工作，因此他们的目标一般都涉及系统的这一部分——通常是使用者接口所能提供的服务。在设想/服务以及系统的外围部分被精确定义之后，系统开发人员便开始工作，对系统的内部部分进行详细定义。

3. 系统开发人员

与最终用户一样，系统开发人员在设想阶段提出他们的总体目标，如为Excel数据仓库开发用户应用程序等。只有在开发人员提出相关目标之后，有关系统各组成部分及其为实现最终用户和管理人员目标而所必须提供的服务的文件才能够真正完成。

在管理人员、最终用户和系统开发人员的目标都被定义之后，系统开发人员对实现这些目标的必要步骤作详细阐述，然后再对能够帮助实现这些目标的系统各组成部分和服务作出定义。

在开发阶段，系统开发人员针对在规划阶段后期得到定义的系统各组成部分进行开发工