SQL Server 2000 数据仓库应用与开发

罗运模 等 编著

人民邮电出版社

需要全本请访问 www.ertongbook.com

图书在版编目(CIP)**数据**

SQL Server 2000 **数据仓库应用与开发/罗运模等编著.—北京：人民邮电出版社，**2001.7
ISBN 7-115-09437-3

I. S… II. **罗**… 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** 039258 **号**

内 容 提 要

本书全面介绍了 Microsoft SQL Server 2000 **数据仓库开发和应用的最新技术，重点介绍了查询和分析数据仓库中有关商业决策数据的具体方法。主要内容包括** SQL Server 2000 Analysis Service **的安装和** Analysis Manager **的启动、数据仓库数据分析实例、连接数据仓库数据源的方法、创建多维数据集、数据仓库结构、维度的编辑与处理、多维数据集的编辑与处理、数据挖掘技术、客户端数据分析、分析数据库的安全管理、数据仓库的设计与组织、数据仓库的管理与维护等。**

本书适合于企业管理人员、数据仓库管理员、开发人员、应用程序设计人员及大专院校相关专业的师生阅读，亦可作为相关培训班的教材。

SQL Server 2000 数据仓库应用与开发

◆ **编　　著　罗运模 等**
责任编辑　张瑞喜　赵桂珍

◆ **人民邮电出版社出版发行　　北京市崇文区夕照寺街** 14 **号**
邮编 100061　**电子函件** 315@pptph.com.cn
网址 http://www.pptph.com.cn
北京汉魂图文设计有限公司制作
北京顺义向阳胶印厂印刷
新华书店总店北京发行所经销

◆ **开本：**787×1092 1/16
印张：21
字数：507 **千字**　　2001 **年** 7 **月第** 1 **版**
印数：1 – 0 000 **册**　　2001 **年** 7 **月北京第** 1 **次印刷**

ISBN 7-115-09437-3/TP·2321

定价：34.00 **元**

前　言

本书的主要目的是将数据仓库应用技术从数据仓库专家那里解放出来，成为大众可以掌握或使用的数据分析和数据挖掘工具。

本书全面、通俗地介绍了 Microsoft SQL Server 2000 数据仓库开发和应用的过程和方法。书中特别注重向读者介绍如何查阅和分析数据仓库中关于商业决策的数据，是一本介绍最新的数据仓库开发和应用技术的基础性读物。

读者或许对数据库非常熟悉而对数据仓库却不甚了解，这是因为数据库的开发和应用已经有几十年的历史，而数据仓库则只有几年的历史。事实上，Microsoft 公司在 1998 年发行的 SQL Server 7.0 中才引入 OLAP Services，从而使 SQL Server 7.0 具有了基本的数据仓库功能。

数据库系统所解决的任务是实现企业业务的在线实时处理（OLTP），比如银行的储蓄存款和取款事务。数据仓库系统所解决的任务是实现企业业务活动数据的实时在线分析（OLAP），为企业决策者提供一种处理企业业务数据、洞悉企业商务详情、探究商业规律与趋势、优化商务决策、并预测未来行动的科学分析工具。

为使企业管理者的决策活动更加迅速、准确、有效，企业需要具备对不同信息来源迅速进行分析和访问的能力。为迎接上述挑战，企业必须对不同的信息来源进行集成、合并与汇总摘要。这样一来，企业便需要创建一个数据仓库——一个大型的、集中的数据存储库。在数据仓库中，信息被收集、组织，并可供随时调用和分析，以便决策者能够根据实际数据进行正确的商务决策，并在商战中处于不败之地。

通过使用数据仓库功能，企业能够为整个组织机构和商业伙伴、顾客及供应商提供先进的交互式分析功能。

Microsoft SQL Server 2000 提供了一套全新的综合分析服务工具。分析服务工具（以前称为 OLAP 服务）可以为商业活动提供了集成的 OLAP 和数据挖掘功能。OLAP 可以通过多维存储技术对大型、复杂数据集进行快速、高级的分析工作。数据挖掘功能能够揭示出隐藏在大量数据中的倾向及趋势，它允许组织或机构最大限度的从数据中获取价值。通过对现有数据进行有效分析，这一功能可以对未来的趋势进行预测。Microsoft SQL Server 2000 针对包括集成数据挖掘、OLAP 服务、安全性服务及通过 Internet 对多维数据集进行访问和链接等在内的分析服务提供了新的数据仓库功能。

Microsoft SQL Server 2000 支持万亿字节级别的数据库，能够对规模最大且最为灵活的数

据仓库实施管理；能够通过企业部门服务器、便携式电脑及移动设备实现对大规模、集中化数据源的访问；能够优化并执行诸如星型查询连接等典型复杂化查询；内部查询并行机制通过将复杂单一的查询分解为若干组件，进而能够将有关工作负载分配给多个处理器的方式来提高查询性能；经过索引的视图能够以包含集合的方式来改进性能表现；端到端平台集成工具能够使企业为实现快速营销、提高工作效率并降低总体拥有成本而创建、管理和使用数据仓库。

数据仓库的传统任务是收集并组织历史业务数据，以便对其进行分析并帮助管理业务决策。直到最近，还只有数据库专家才能访问数据仓库，这些专家可以创建检索、汇总和格式化信息所需的复杂查询，供分析人员和高级决策者使用。随着数据仓库变得越来越普通且业务在决策管理过程中涉及的级别较低，只具备浅显的数据库专门技术的人直接对数据仓库进行最终用户访问的需要日益增加。也就是说，随着数据仓库使用技术门坎的降低，会有越来越多的人需要和掌握数据仓库使用技术。特别是 Microsoft SQL Server 2000 版的正式发行，将更会有越来越多的人需要并较容易地掌握数据仓库使用技术。这些人就是本书的主要读者基础，这也就是本书写作的主要原因之一。

通俗地讲，数据库是获取数据的，而数据仓库则是分析数据的。

全书共分 12 章。第 1 章介绍 SQL Server 2000 Analysis Service 的安装和 Analysis Manager 的启动方法。第 2 章介绍一个已经存在的数据仓库应用实例，通过浏览其中的分析数据，预先感受一下数据仓库的意义和价值。第 3 章介绍连接数据仓库数据源的方法。第 4 章介绍创建一个简单的多维数据集的方法，也是创建和使用多维数据集的一般过程。第 5 章介绍有关数据仓库方面的概念和术语。第 6 章介绍维度的编辑与处理方法。第 7 章介绍多维数据集的编辑与处理方法。第 8 章介绍数据仓库中的数据挖掘技术。第 9 章介绍在客户端通过 Microsoft Excel 查询远程数据仓库中的多维数据集中数据的方法。第 10 章介绍分析数据库的安全管理。第 11 章介绍数据仓库的设计与组织。第 12 章介绍数据仓库的管理与维护方法。

本书所介绍的操作步骤和操作过程都是作者亲自实践过的，所有举例，都是经过实际运行的，其结果都是直接运行所得到的。建议读者在阅读本书时，跟随书中内容的介绍进行具体的操作实践，这样会收到事半功倍的效果。

本书读者对象为企业管理人员（尤其是企业的经理、总经理等中高层管理人员）、数据仓库管理员、数据仓库开发人员、数据仓库应用程序设计人员和大专院校相关专业的师生。

参加本书编写的人员有：罗运模、崔小兵、谢志敏、罗华宾、尚文化、唐宾、刘志方、李华刚、文中、董云、章元、蔡京华、白涛、李军、艾华、李华海、陶元庆、王国华、陶宗明、吴江北、黄海涛、李文娟。

感谢深圳远望城多媒体电脑有限公司及同事们的支持和帮助。

由于本书涉及面广，加之作者经验所限，书中难免有不足或错误之处，敬请读者批评指正。

作者

编 后 语

通过本书（《SQL Server 2000 数据仓库应用与开发》）的学习，读者已经是一个优秀的数据仓库应用能手了，不但能够自己在数据仓库中创建多维数据集，而且还能进行数据分析和数据挖掘操作。但如果读者希望自己能够开发一个完整的数据仓库项目，则还需要阅读有关 SQL Server 2000 数据库方面的书籍（比如，本书的姐妹篇《完全掌握 SQL Server 2000》），这样不但会成为一个优秀的数据仓库应用能手，而且还会是一个优秀的数据仓库开发能手。读者如果再仔细阅读《Microsoft OLAP 解决方案》一书，则又可以成为一个数据仓库开发、数据分析和数据挖掘的高手。

需要全本请访问 www.ertongbook.com

目　录

第 1 章　安装和启动 …… 1
1.1　SQL Server 2000 的新特性及数据仓库功能 …… 1
1.1.1　SQL Server 2000 的新特性 …… 1
1.1.2　SQL Server 2000 不同版本比较 …… 2
1.1.3　SQL Server 2000 的数据仓库功能简介 …… 3
1.2　SQL Server 2000 的系统需求 …… 7
1.3　安装 SQL Server 2000 …… 8
1.4　安装 SQL Server 2000 Analysis Service …… 9
1.5　打开 Analysis Manager 窗口 …… 12
1.6　关闭 Analysis Manager 窗口 …… 13
第 2 章　数据分析实例 …… 14
2.1　进入多维数据集浏览界面与多维数据集浏览界面介绍 …… 14
2.1.1　进入多维数据集浏览界面 …… 14
2.1.2　多维数据集浏览界面介绍 …… 16
2.2　浏览多维数据集 …… 18
2.3　替换和添加维度 …… 23
第 3 章　连接数据源 …… 27
3.1　创建数据库 …… 27
3.2　导入数据 …… 29
3.3　FoodMart 的数据结构 …… 32
3.4　创建数据分析数据库 …… 32
3.5　连接数据源 …… 34
第 4 章　创建多维数据集 …… 37
4.1　创建维度 …… 37
4.1.1　创建顾客维度 …… 37
4.1.2　创建产品维度 …… 43
4.1.3　创建时间维度 …… 46
4.2　创建多维数据集 …… 49
4.3　处理多维数据集 …… 53

4.3.1 关闭多维数据集编辑窗口时进行处理……53
4.3.2 通过“Analysis Manager”窗口进行处理……57
4.4 浏览数据……58
第 5 章 数据仓库结构……60
5.1 数据仓库数据库……60
5.2 事实数据表……61
5.3 维度表……62
5.4 维度……63
5.5 维度创建方式……65
5.6 维度类型……66
5.6.1 按使用方式分类……66
5.6.2 按创建方式分类……66
5.7 维度存储模型……67
5.8 度量值……67
5.9 多维数据集……68
5.10 多维数据集的架构……68
5.11 多维数据集分区……69
5.12 多维数据集存储模式……70
5.13 多维数据集数据模型……72
5.14 数据仓库总体结构……73
5.15 数据仓库与客户查询关系……74
第 6 章 维度的编辑与处理……75
6.1 编辑维度……75
6.1.1 启动维度编辑器……75
6.1.2 修改维度属性……76
6.1.3 保存修改……85
6.1.4 修改成员键列和成员名称列……86
6.1.5 添加维度级别……88
6.1.6 创建成员属性……91
6.1.7 添加表……93
6.2 处理维度……98
6.3 删除维度……99
6.4 浏览维度数据……100
6.5 查看维度元数据……101
6.6 创建父子维度……103
6.6.1 父子维度举例……103
6.6.2 创建父子维度……104
6.6.3 父子维度的层次结构……107
6.7 创建虚拟维度……108

6.7.1 基于成员属性创建虚拟维度……109
6.7.2 基于列创建虚拟维度……111
6.7.3 关于虚拟维度……116
6.8 创建专用维度……117
6.9 添加父级别数据……119
6.9.1 创建商店维度……119
6.9.2 创建多维数据集……120
6.9.3 添加和查看父级别数据……123
6.10 设置维度高级特性……125
6.11 定义高级计算……130
6.11.1 创建新维度……131
6.11.2 创建多维数据集……133
6.11.3 设置维度的高级属性……133
6.11.4 查阅费用预算数据……135
6.11.5 定义高级计算……137
第 7 章 多维数据集的编辑与处理……143
7.1 编辑多维数据集……143
7.1.1 启动多维数据集编辑器……143
7.1.2 修改多维数据集属性……145
7.2 处理多维数据集……152
7.3 删除多维数据集……152
7.4 浏览多维数据集数据……153
7.4.1 通过多维数据集编辑器浏览数据……153
7.4.2 通过 Analysis Manager 窗口浏览数据……153
7.5 查看多维数据集元数据……154
7.6 创建计算成员……155
7.6.1 创建计算成员……155
7.6.2 查看计算成员数据……158
7.7 创建计算单元……159
7.7.1 在销售多维数据集中添加商店维度……160
7.7.2 创建计算单元……161
7.7.3 查看计算单元数据……168
7.8 钻取数据……169
7.8.1 设置钻取选项……169
7.8.2 查看钻取数据……171
7.9 创建对策……172
7.9.1 创建对策……172
7.9.2 触发对策……175
7.10 创建分区……176

7.11 创建和处理虚拟多维数据集……182
7.12 链接多维数据集……186
7.12.1 创建新用户……186
7.12.2 授予操作权限……189
7.12.3 链接多维数据集……192
7.12.4 查看链接多维数据集数据……194
第 8 章 数据挖掘……198
8.1 关于数据挖掘模型……198
8.1.1 数据挖掘特性……198
8.1.2 数据挖掘模型……199
8.2 决策树数据挖掘技术……199
8.3 浏览挖掘维度和挖掘多维数据集数据……204
8.3.1 浏览挖掘维度数据……204
8.3.2 浏览挖掘多维数据集数据……205
8.4 编辑挖掘多维数据集……207
8.5 编辑挖掘模型……209
8.6 浏览决策树……210
8.7 更新挖掘模型……218
8.8 聚集挖掘技术……219
8.8.1 创建聚集挖掘模型……220
8.8.2 浏览聚集挖掘模型树……222
8.8.3 设置聚集挖掘模型的分组数……225
8.9 关系挖掘技术……226
8.10 浏览数据挖掘相关性网络……232
第 9 章 客户端数据分析……235
9.1 连接 Internet……235
9.2 连接多维数据集……235
9.2.1 启动 Microsoft Excel……235
9.2.2 连接多维数据集……236
9.3 透视多维数据集数据……241
9.3.1 定制透视表……241
9.3.2 保存透视表……249
第 10 章 分析数据库的安全管理……251
10.1 关于角色……251
10.1.1 分析数据库角色……251
10.1.2 多维数据集角色……251
10.1.3 挖掘模型角色……252
10.1.4 最终用户权限……252
10.2 多维数据集角色的创建和管理……252

10.2.1 创建多维数据集角色 ……252
10.2.2 查看和编辑角色 ……256
10.3 数据库角色的创建和管理 ……261
10.4 挖掘模型角色的创建和管理 ……266
第 11 章 数据仓库的设计与组织 ……270
11.1 OLTP 与 OLAP ……270
11.1.1 联机事务处理 ……270
11.1.2 联机分析处理 ……271
11.1.3 联机分析处理的必要性 ……273
11.2 数据仓库与数据集市 ……273
11.2.1 数据仓库与数据库 ……274
11.2.2 数据集市 ……276
11.3 设计数据仓库 ……277
11.3.1 数据仓库设计 ……277
11.3.2 数据集市设计 ……277
11.4 创建数据准备区 ……278
11.4.1 关于数据准备区 ……278
11.4.2 创建数据准备区 ……278
11.4.3 创建数据仓库数据库 ……279
11.5 析取数据 ……280
11.6 清理和转换数据 ……281
11.7 将数据装入数据仓库 ……282
11.8 准备显示数据 ……283
11.9 索引及视图索引 ……283
11.10 将数据分发到数据集市 ……285
11.11 设计 Analysis Services 应用程序 ……285
11.11.1 T-SQL 语言编程 ……285
11.11.2 VBA 语言编程 ……288
11.11.3 Visual Basic 语言编程 ……289
11.12 数据仓库及相关环境 ……289
11.13 数据仓库的一般开发过程 ……290
第 12 章 数据仓库的管理与维护 ……294
12.1 数据仓库管理概述 ……294
12.2 数据仓库管理 ……295
12.2.1 数据仓库管理与联机事务处理系统数据库管理的比较 ……295
12.2.2 备份数据仓库 ……296
12.2.3 删除数据仓库 ……299
12.2.4 还原数据仓库 ……300
12.3 多维数据集的管理 ……302

12.3.1 打开 DTS 包窗口……302
12.3.2 创建处理任务……303
12.3.3 安排处理任务……306
12.3.4 编辑工作流关系……308
12.3.5 保存包……310
12.3.6 调度包……313
12.4 优化数据仓库性能……315
12.4.1 基于使用优化性能……315
12.4.2 基于使用分析向导……320
编 后 语……323

需要全本请访问 www.ertongbook.com

第 1 章　安装和启动

SQL Server 2000 是美国微软公司在 SQL Server 7.0 的基础上经多年潜心研究而开发成功的最新、功能最强的数据库管理系统。用户不但可以应用 SQL Server 2000 开发数据库应用项目，而且还可以开发数据仓库应用项目(需要再安装 SQL Server 2000 Analysis Service)。这样，就扩大了数据库软件系统的应用范围，使用户获得更大效益。

本章介绍 SQL 2000 Analysis Service 的安装和 Analysis Manager 的启动方法，主要包括 SQL Server 2000 的新特性及数据仓库功能、SQL Server 2000 的系统需求、安装 SQL Server 2000、安装 SQL 2000 Analysis Service、打开 Analysis Manager 窗口和关闭 Analysis Manager 窗口等内容。

1.1　SQL Server 2000 的新特性及数据仓库功能

SQL Server 2000 不但在很多方面增强了 SQL Server 7.0 的原有功能，而且新增了许多新的功能，特别是在数据仓库方面增添了许多强大的数据分析和挖掘功能，它已经成为功能强大的数据仓库系统。

注意：本节中有些概念可能不易理解，但通过本书的学习后，读者就可以完全理解这些概念了。所以，读者在阅读本节时，不要急于完全理解其中的所有概念。

1.1.1　SQL Server 2000 的新特性

建立在 SQL Server 7.0 可扩展基础上的 SQL Server 2000 代表着下一代 Microsoft .NET Enterprise Servers(企业服务器)数据库的发展趋势。SQL Server 2000 是为创建可伸缩电子商务、在线商务和数据仓库的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。

1. 电子商务

电子商务解决方案对数据库应用程序提出了更高要求。由于来自世界各地的用户以全天不间断的方式对电子商务网站进行访问，因此必须确保由此产生的繁重事务性负载得到高效处理。而编程方面的灵活性对确保与现有系统间的协同工作能力和电子商务所需的可靠性而言可谓至关重要。除了提供电子商务所需的可伸缩性与可扩展性之外，SQL Server 2000 还提供了丰富的数据库编程能力(基于 Web 标准)，以确保系统的协同工作和灵活性。与此相关，SQL Server 2000 还具有许多新特性，例如：丰富的扩展标记语言(XML)、Worldwide Web Consortium(万维网社团，W3C)标准支持、利用 Transact SQL(T-SQL)实现的 XML 数据操作能力、灵活而强大的 Web 分析功能以及使用 Hypertext Transfer Protocol(超文本传输协议，HTTP)进行的安全 Web 数据访问等。

2. 在线商务

SQL Server 已成为颇受众多客户青睐的数据产品，这主要应归功于该产品在对称多处理(SMP)硬件方面所具有的可伸缩性和较低的数据库维护需求。通过使用 SQL Server 2000，将使用户获得增强的可靠性、可伸缩性和简化的管理与优化功能，而这些恰恰是成功快速部署在线商务应用程序所必需的。SQL Server 2000 不仅支持集中化数据库管理功能，而且还最大程度地实现了管理与优化工作的自动化，从而减轻了有关管理人员的负担。

3. 数据仓库

Microsoft SQL Server 2000 提供了一套全新的综合分析服务系统。分析服务(以前称为 OLAP 服务)为商业活动提供了集成的 OLAP 服务和数据挖掘功能。OLAP 可以通过多维存储技术对大型、复杂数据集进行快速、高级的分析工作。数据挖掘功能能够揭示出隐藏在大量数据中的倾向及趋势，它允许组织或机构最大限度的从数据中获取价值。随着 SQL Server 7.0 联机分析处理(OLAP)服务的引入，企业需要对来源各异的数据信息进行集成、合并与汇总摘要，而数据仓库则通过使用大型、集中的数据存储来提供上述功能。在这种数据存储中，信息可以被收集、组织，并可供决策者随时调用。通过 Microsoft SQL Server 2000 的数据仓库功能的应用，决策者便可洞悉详情、探究规律与趋势、优化商务决策并预测未来的行动。Microsoft SQL Server 2000 针对包括集成数据挖掘、OLAP 服务、安全性服务及通过 Internet 对多维数据集进行访问和链接等在内的分析服务提供了新的数据仓库功能。

1.1.2 SQL Server 2000 不同版本比较

SQL Server 2000 的数据仓库功能的强弱与其不同的版本有关，不同版本的 SQL Server 2000，其数据仓库功能有所差别，表 1-1 是不同版本的 SQL Server 2000 数据仓库功能的比较(其中，打“√”表示该版本具有所列特性)。

表 1-1

特性	SQL SERVER 2000 企业版	SQL SERVER 2000 标准版	SQL SERVER 2000 个人版[3]
可伸缩性	√	√	
综合性分析服务	√		
数据转换服务	√	√	√
数据挖掘	√	√	√
联机分析处理(OLAP)	√	√[1]	√[1]
OLAP 灵活性	√	√	√
闭合循环分析	√	√[2]	√[2]
索引视图	√		
元数据服务	√	√	√
Office 2000 集成	√	√	√
数据仓库联盟	√	√	√
可支持的 Web 分析	√		
英语查询	√	√	√

说明 1. 表格中的 1 表示缺少一些企业版中的特性。

2. 表格中的 2 表示 需要 Commerce Server 2000 的支持。

3. 表格中的 3 表示分析服务(包括 OLAP、数据挖掘及其他数据仓库特性)包含在个人版中，但这些服务不能安装在 Windows 98 和 Windows Me 上。

1.1.3 SQL Server 2000 的数据仓库功能简介

SQL Server 2000 具有许多新的数据仓库功能，本小节将做简单介绍。

1. 可伸缩性

可伸缩性是 Microsoft SQL Server 7.0 所具有的功能，并且很受欢迎。一般来说，数据仓库需要存储和处理大量的数据，并需要不断地从其他数据库或数据存储系统转换数据。因此，数据库服务器的可伸缩性对数据仓库来说具有十分重要的意义。在 Microsoft Windows 2000 Datacenter Server 上，Microsoft SQL Server 2000 企业版可以伸缩至 64 GB 内存和 32 个 CPU。这一可伸缩性可以与向外扩充(scale-out)技术(如分布式分区视图)组合使用，以处理最大的数据集和事务负荷。

SMP 硬件包括从流行于小型企业中的双处理器系统到财富 500 强公司数据中心内的 16 路和 32 路系统。通过更好地利用这些越来越常见的并行多处理(SMP)硬件，SQL Server 2000 可以完成更多的操作。例如，多 TG 数据仓库可以从并行索引创建中获益，其作用就是极大地减少创建事实数据表索引所需的时间。

2. 综合性分析服务

Microsoft SQL Server 2000 中的分析服务就是原来的联机分析处理(OLAP)服务，其作用就是提供一个完整的、包含 OLAP 特性和数据挖掘功能的端到端分析平台。用户可以混合使用和比较来自 Microsoft 和第三方的算法和工具，以定制分析应用程序。分析服务是专为数据库管理员和应用程序开发人员准备的，而不是为统计专家们准备的。任何具有 SQL(结构化查询语言)和 Microsoft Visual Basic 开发系统基础的个人都可以理解、使用并以编程方式扩展分析服务的功能。分析服务包含在 SQL Server 2000 企业版、SQL Server 2000 标准版、SQL Server 2000 开发人员版中。

3. 数据转换服务(DTS)

DTS 特性简化了使用 OLE DB、ODBC(开放式数据库连接)或文本文件异类数据的提取、转换以及加载到任意支持 OLE DB 的数据库或多维存储中的操作。DTS 还允许用户定期导入或变换数据，以实现数据转换的自动化。

与 Microsoft SQL Server 7.0 版本相比，SQL Server 2000 的 DTS 在可用性和可编程性方面得到了改善。通过一些允许程序在数据转换过程中以多点方式进行交互的接口，开发人员可以访问并操纵多相数据库提取(pump)的操作。多相数据提取功能实现了数据库转换和错误/故障处理的灵活性。DTS 还允许用户利用新的参数化源选择语句和改良的错误日志，将“重新启动”内建在 DTS 程序包中。

DTS 程序包现在可以保存为基于 Visual Basic 的代码。这就简化了通过编程接口开发 DTS 程序包的工作。此外，在与 Visual SourceSafe 版本控制系统组合使用时，DTS 还提供了程序包版本控制和备份的替代方法。因为程序包现在可以在执行时调用其他程序包，所以可以在重用程序包方面具有更大的灵活性。新的任务和定制转换特性扩展了数据操作的功能、性能、

灵活性，以及 DTS 与来自其他厂商产品的集成性。

数据转换服务是数据仓库所需要的一个重要功能，因为数据仓库中的许多数据都需要从其他数据库系统或数据存储系统转换而得到。

4. 数据挖掘

数据挖掘技术帮助用户分析关系数据库和多维 OLAP 数据集中的数据，以获取用于预测用途的模式和趋势。SQL Server 2000 中的数据挖掘功能与关系数据源和 OLAP 数据源紧密集成在一起。

SQL Server 2000 包含由 Microsoft 研究院开发的两种数据挖掘算法：Microsoft 决策树挖掘算法和 Microsoft 聚集挖掘算法。此外，SQL Server 2000 中的数据挖掘支持由第三方开发的数据挖掘算法。

(1) Microsoft 决策树算法：该算法基于分类。算法建立一个决策树，用于按照事实数据表中的一些列来预测其他列的值。该算法可以用于判断最倾向于单击特定标题(banner)或从某电子商务网站购买特定商品的个人。

(2) Microsoft 聚集算法：该算法将记录组合到可以表示类似的、可预测的特征的聚集中。通常这些特征可能是隐含或非直观的。例如，聚集算法可以用于将潜在的汽车买主分组，并创建对应于每个汽车购买群体的营销活动。

5. 联机分析处理(OLAP)

SQL Server 2000 分析服务提供了最完整的、最具集成性的、可支持 Web 的分析服务。OLAP 组件包括一个中间层服务器，它可以让用户对大量数据进行分析而只花费很少的数据检索执行时间。数据源可以包含各种系统提供的数据源，如 SQL Server、Oracle、DB2 及其他关系数据库和平面文件。SQL Server 2000 中的数据挖掘特性可以让公司发现其发展模式和趋势，并对企业未来的趋势进行预测。

6. 分析 OLAP 灵活性

Microsoft SQL Server 2000 除了支持原 SQL Server 7.0 已有的“星型架构”维度和“雪花架构”维度外，还特别支持一些新的维度类型。表 1-2 是随 SQL Server 2000 推出的新的维度类型及描述。

表 1-2

维度类型	描述
父子维度	该维度类型支持基于源表列中成员间父子链接的层次结构。结构化图表、零件组装及账单材料系统都是这一层次结构的例子。这种维度有时也被称为“不均衡”维度
不规则维度	该维度类型至少有一个成员的逻辑父代不直接位于该成员的上层。在国家与城市之间没有州或省一级建制的国家就是不规则维度的一个例子
可变维度	与其他维度类型相比，该维度类型允许更大范围的更改，而无需在更改之后完全处理一个多维数据集。这就增加了客户端应用程序可以利用的多维数据集
写启用维度	该维度类型可以通过分析管理器或任意支持维度回写的客户端应用程序来更新。角色被用于控制来自客户端应用程序的维度写访问
虚拟维度	该维度类型拥有由其他维度成员决定的成员

7. 闭合循环分析

OLAP 操作是 Microsoft SQL Server 2000 分析服务中的一个新特性，它允许最终用户按照分析结果自动驱动商业流程。操作(Action)就是由最终用户引发的对特定多维数据集或多维数据集的一部分的动作。该操作可以启动一个以特定数据项为参数的应用程序，或者检索关于特定数据项的信息。通过实现操作(可以利用操作向导方便的定义)，开发人员可以将来自复杂数据翻译工具的客户端应用程序转换成企业级操作系统中反馈循环的组成部分。通过使用操作，最终用户可以超越传统的分析并启动查找问题和缺陷的解决方案。

通过一个简单的如何使用操作(Action)的例子，研究鞋店中的一个最终用户如何浏览存货清单并发现白色鞋子的缺货状况。该最终用户选择了白色鞋子的订购操作。这一操作通过定单输入系统引发了订购更多白色鞋子的定单。

8. 索引视图

复杂报表的场景经常会在数据仓库应用程序中遇到，它在查询过程中会对数据库服务器产生大量请求。当这些查询访问视图时，因为数据库系统要将建立视图结果集所需的逻辑合并到从基本表数据建立完整查询结果集所需的逻辑中，所以性能将会下降。这一操作的开销可能会比较大，尤其当视图涉及到复杂的大量行处理，如大量数据聚合或多表联结时。因为结果集并不永久存放在数据库(标准视图)中，所以以后对该视图的访问可能会导致每次执行查询都建立结果集的代价。

SQL Server 2000 允许为视图创建独特的聚集索引，从而使访问此类视图的查询的性能得到极大的改善。在创建了这样一个索引后，视图将被执行，结果集将被存放在数据库中，存放的方式与带有聚集索引的表存放方式相同。这样，就可以在数据库中实现高效率的查询。对于那些在 FROM 子句中不直接指定视图名的查询，SQL Server 查询优化器将使用视图索引。现有查询将受益于从索引视图检索数据，而无需重新编写程序原码。对于某些特定类型的视图，甚至可以获得指数级的性能改善。

从易用性的角度来看，SQL Server 2000 中索引视图的实现方式非常独特。具有索引视图功能的同类竞争产品需要查询语句使用特殊的代码来索引视图，因此现有查询必须用提示进行更新，以强制查询处理器识别出索引视图。而 SQL Server 2000 中索引视图的实现方式，不需要重新编码查询语句——SQL Server 查询处理器在适当的时候自动识别和利用索引视图，无需数据库管理员的任何干预。

9. 元数据服务

Microsoft SQL Server 2000 元数据服务是一组帮助用户管理元数据的服务。元数据描述了数据的结构和意义，同时也可以描述应用程序和流程的结构和意义。需要牢记的是，元数据是一个抽象，而非一个上下文，它可以在开发环境中得到应用。

SQL Server 2000 元数据服务提供了一种存储和管理关于信息系统和应用程序的元数据的方法。该技术担当数据和组件定义、开发和部署模型、可重用软件组件及数据仓库描述的中枢。

元数据服务由知识库引擎、工具、API(应用程序编程接口)、标准信息模型、浏览器、SDK(软件开发工具包)等组成。

10. Office 2000 集成

Microsoft Office 2000 提供对 OLAP 和分析服务的支持，这一支持允许用户使用常见的工

具对大量数据进行复杂的分析，并且得到非常好的执行性能。因为 Office 2000 与基于 Microsoft SQL Server 的数据仓库紧密集成在一起，所以更多用户群可以访问企业信息并进行分析操作。Office 的 Web 组件还支持 OLAP，可以在浏览器内提供浏览和制图功能。

Microsoft Excel 2000 提供了一个丰富的分析前端，它可以访问企业 SQL Server 数据，以作出更好的决策。此外，用户可以在 OLAP 多维数据集上创建 OLAP 和 Microsoft 数据透视表动态视图，以达到提高工作效率的目的。Excel 2000 使任何用户都可以方便地将电子数据表、图表及数据透视表保存为 Web 组件或交互式 Web 页，其他人可以访问 SQL Server 数据并查看和分析这些数据。

11. 数据仓库联盟

Microsoft 数据仓库联盟为软件开发商提供了完善的数据仓库软件框架。该框架起到技术和市场营销路标的作用。数据仓库联盟为客户提供了更多的最佳解决方案的分析工具的选择。

12. 支持 Web 分析

当公司收集到越来越多的有关客户的信息时，他们寻求分析和发掘这些数据是否有价值的新方法。SQL Server 2000 中的分析服务推出了链接的多维数据集和对多维数据集的 HTTP(超文本传输协议)访问。这两个技术实现了 Web 分析，因此用户可以利用由合作伙伴拥有的多维数据集或由研究公司通过 Internet 销售的多维数据集。

链接的多维数据集是在其他分析服务器上定义和存储的数据集。这些服务器包括位于企业防火墙之外的服务器。对最终用户而言，链接的多维数据集类似于常规数据集。链接的数据集允许数据供应商在一台分析服务器上创建、存储和维护一个多维数据集，同时该多维数据集还可以作为多台分析服务器上的一个或多个链接的多维数据集使用，使用 HTTP 和 HTTPS 的数据源也可被支持。链接的多维数据集使用其源多维数据集的聚合，因此没有自己的数据存储需求。该技术允许组织更新一个多维数据集并保留该多维数据集的所有权，同时允许多个客户同时利用这一多维数据集。利用 SQL Server 2000，敏感信息可以存放在安全服务器的数据源和多维数据集中，但仍然可以以链接的多维数据集的形式提供给其他服务器。此外，用户可以利用虚拟多维数据集特性将来自市场研究公司的外部链接的多维数据集与包含销售数据的内部多维数据集组合在一起，通过直接评估市场份额、相关趋势、市场普及率及市场潜力获得新的分析深度。

SQL Server 2000 还允许经由 HTTP 的多维数据集访问。这就允许企业通过 HTTP 和一个防火墙安全地共享数据集或访问远程多维数据集，而无需在 Web 服务器开启一个端口。

13. 英语查询

利用英语查询，用户可以创建能接受自然语言查询(以普通英语表达的问题)的应用程序，并用这种查询取代原先复杂的 SQL 查询。这一理想的机制允许来自新用户的无格式查询，因而可以让开发人员将精力集中在应用程序开发的其他部分，而不是将精力花费在建立大量预定义报表并期待它们满足用户的需要上。现在，在流行的 Web 上，用户无格式查询的需要更为迫切。用户希望 Web 站点能提供综合性的、易于使用的搜索功能。

英语查询特性让数据库开发人员和管理员为各种最终用户提供存放于关系数据库中数据的方便访问，同时还支持高级用户及那些请求复杂的数据子集的用户。英语查询将用户的问题转换成可将所需结果从数据库中提取出来的 SQL 查询。

在 SQL Server 2000 中，英语查询应用程序的创建和部署已经得到极大的简化，具体体现

需要全本请访问 www.ertongbook.com