



博研文化

SQL 2000 决策分析 OLAP 建置与应用

林宏谕◎编著

专业经理人与
高级幕僚
的必备工具

内附本书范例光盘



中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

SQL 2000 决策分析

OLAP 建置与应用

林宏谕 编著

中国铁道出版社

2001年·北京

(京)新登字 063 号

北京市版权局著作权合同登记号: 01-2001-2073 号

版 权 声 明

本书中文繁体字版由台湾博硕文化股份有限公司出版, 2001。本书中文简体字版经台湾博硕文化股份有限公司授权由中国铁道出版社出版, 2001。任何单位或个人未经出版者书面允许不得以任何手段复制或抄袭本书内容。

本书贴有博硕文化激光防伪标签, 无标签者不得销售。版权所有, 侵权必究。

图书在版编目(CIP)数据

SQL 2000 决策分析: OLAP 建置与应用/林宏谕编著. —北京: 中国铁道出版社, 2001. 7

ISBN 7-113-04258-9

I. S… II. 林… 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 039462 号

书 名: SQL 2000 决策分析——OLAP 建置与应用

作 者: 林宏谕

出版发行: 中国铁道出版社(100054, 北京市宣武区右安门西街 8 号)

策划编辑: 苏 茜

特邀编辑: 王占清

封面设计: 张国东 孙天昭

印 刷: 北京市兴顺印刷厂

开 本: 787×1092 1/18 印张: 19 字数: 457 千

版 本: 2001 年 8 月第 1 版 2001 年 8 月第 1 次印刷

印 数: 1~5000 册

书 号: ISBN 7-113-04258-9/TP·582

定 价: 33.00 元

版权所有 盗印必究

凡购买铁道版的图书, 如有缺页、倒页、脱页者, 请与本社计算机图书批销部调换。

510103

出版说明



近年来,知识数据库(Knowledge Database-KDD)的概念与技术盛行,衍生出许多的概念与相对的软件工具,每一种工具的处理功能与获得成本,都有相当大的差别。本书就是通过 Microsoft SQL Server 2000 中的 Analysis Service, 配合如 Excel、Web 等前端分析工具,告诉你如何利用现有的数据库服务器与软件工具,来建置企业内部的数据库分析机制。

作者从数据仓库的概念着手,并从 OLAP 分析数据库的建置开始,由浅入深地带领用户建立分析多维数据集与数据挖掘模型,说明如何通过前端的分析工具如 Excel、Web 来获取分析的结果。让你不但可以用最少的时间,处理决策分析所需的资料,并从中得到各种不同的分析结果,作为经营决策的最佳参考。

本书由台湾博硕文化股份有限公司提供版权,经中国铁道出版社计算机图书项目中心审选;李自运、梁秀玲、关超、史广顺、张瀚文、陈辑超、马超、杨小平、杨小军、段小明、杨军、陈贤淑及孟丽花等同志参与了本书的整稿及编排工作。

中国铁道出版社

2001 年 8 月

目 录

第 1 章 SQL 2000 与分析管理器	1
1-1 SQL 2000 中的“分析管理器”简介	2
1-2 “分析管理器”的功能架构	5
1-2-1 易于使用	5
1-2-2 灵活性的数据模型	6
1-2-3 可调整性	7
1-2-4 集成性	8
1-3 OLAP、数据挖掘与分析模式多维数据集	10
1-3-1 分析模式多维数据集	10
1-3-2 OLAP 与 Data Mining 简介	14
1-3-3 OLAP 数据实体存储方式	15
第 2 章 建立 OLAP 分析数据库	19
2-1 设置新的服务器与对应的数据库	20
2-2 设置数据源	22
2-3 简谈 ODBC	24
第 3 章 个案简介与基本维度建立	27
3-1 个案范例简介	28
3-2 设置范例数据库文件为数据源	28
3-3 建立共享的维度	29
3-3-1 维度的简介	29
3-3-2 共享维度的建立	31
3-3-3 建立时间维度	32
3-3-4 建立标准维度	36
第 4 章 多维数据集的建立与应用	47
4-1 建立基本分析多维数据集模式	48
4-2 多维数据集分析模式的应用	55
4-2-1 直接使用“分析管理器”进行数据浏览	55

4-2-2	使用 Excel 作为前端分析工具	60
4-2-3	在 Web 上建立 OLAP 数据透视	64
第 5 章	多维数据集的编辑与管理	71
5-1	维度的编辑	72
5-1-1	维度的新建与删除	72
5-1-2	维度的编辑	73
5-1-3	更改维度数据源	74
5-1-4	新建或删除维度中的层次	75
5-2	多维数据集编辑	76
5-2-1	多维数据集中维度的编辑	78
5-2-2	多维数据集中度量值的编辑	79
第 6 章	父子维度的建立与应用	83
6-1	建立父子维度	84
6-2	父子维度的深入探讨	88
6-2-1	“层次性维度”与“父子维度”	88
6-2-2	结点是否包含数据的模式建立	89
第 7 章	计算成员的建立与应用	99
7-1	度量值成员的导出与应用	101
7-1-1	建立度量值计算成员	101
7-1-2	编辑度量值计算成员	104
7-2	维度成员的导出与应用	106
7-3	合并计算成员	109
第 8 章	成员属性与虚拟维度	111
8-1	成员属性的建立与应用	112
8-1-1	成员属性的意义	112
8-1-2	建立成员属性	113
8-2	虚拟维度的建立与应用	117
8-2-1	使用成员属性建立虚拟维度	118
8-2-2	使用“列”建立虚拟维度	121
第 9 章	多维数据集的分区与合并	127
9-1	多维数据集分区	128
9-1-1	多维数据集分区的概念	128

9-1-2 建立多维数据集分区.....	129
9-1-3 编辑分区与筛选条件设置.....	134
9-2 分区的合并	136
第 10 章 虚拟多维数据集与链接多维数据集的建立与应用	141
10-1 虚拟多维数据集的建立	142
10-1-1 虚拟多维数据集的意义.....	142
10-1-2 建立虚拟多维数据集.....	142
10-2 链接多维数据集的建立	146
10-2-1 链接多维数据集的概念.....	146
第 11 章 分析模式的管理	151
11-1 管理与使用权限的设置	152
11-1-1 系统管理员的安全性控制.....	152
11-1-2 数据库角色定义与管理.....	153
11-1-3 多维数据集角色的管理.....	161
11-2 数据库管理议题	167
11-2-1 数据源的重整与分析数据库处理.....	167
11-2-2 数据库的存档.....	171
11-2-3 分析对象的复制.....	175
11-3 多维数据集使用分析与优化处理.....	176
11-3-1 使用分析.....	176
11-3-2 基于使用的优化.....	180
11-4 整体工作环境设置	183
11-4-1 常规性设置.....	184
11-4-2 运算环境.....	187
11-4-3 处理设置.....	188
11-4-4 日志记录设置.....	188
第 12 章 高级多维数据集的处理	191
12-1 设置钻取选项	192
12-1-1 钻取的基本概念.....	192
12-1-2 启用“钻取”功能.....	193
12-1-3 设置分区的钻取选项.....	195
12-2 “命名集”的建立与应用	196
12-2-1 命名集的概念.....	196

12-2-2	建立命名集.....	197
12-2-3	命名集在 MDX 中的应用.....	198
12-3	“对策”的建立与应用.....	200
12-3-1	“对策”的概念.....	200
12-3-2	建立与应用“对策”.....	200
12-4	计算单元的应用.....	205
12-4-1	建立计算单元.....	205
12-4-2	编辑计算单元.....	208
第 13 章	汇总公式的建立与应用.....	211
13-1	自定义成员公式的建立与应用.....	214
13-1-1	建立或选择列以保存自定义成员公式.....	214
13-1-2	使用 MDX 公式建立自定义成员.....	216
13-1-3	使用一元运算符公式建立自定义成员.....	222
13-2	自定义汇总公式的建立与应用.....	228
13-2-1	汇总的意义.....	228
13-2-2	建立自定义汇总公式.....	230
第 14 章	MDX 的概述与常用函数.....	235
14-1	MDX 概述.....	236
14-2	MDX 的处理目标表示方式.....	236
14-2-1	成员的表达方式与函数.....	237
14-2-2	Tuple 与集合.....	240
14-2-3	SQL 与 MDX.....	241
14-2-4	坐标轴与子多维度数据集.....	243
14-2-5	注释的表达式.....	244
14-2-6	空数据集的处理.....	244
14-3	使用 MDX 查询.....	246
14-3-1	SELECT 语句.....	246
14-3-2	使用函数与表达式.....	249
14-3-3	数值函数.....	251
14-3-4	集合函数.....	254
14-3-5	判断式函数.....	260
第 15 章	数据挖掘.....	263
15-1	使用挖掘模型向导建立“决策树”挖掘模型.....	264

15-2	数据挖掘算法	271
15-2-1	决策树算法.....	271
15-2-2	聚集算法.....	272
15-2-3	建立挖掘模型的基本概念.....	272
15-3	建立“聚集”挖掘模型与编辑挖掘模型	273
15-3-1	建立“聚集”挖掘模型.....	273
15-3-2	挖掘模型的编辑.....	276
15-3-3	相关性网络的应用.....	280
15-4	建立 OLAP 挖掘模型	281
15-4-1	建立挖掘模型用的多维数据集.....	282
15-4-2	建立决策树 OLAP 挖掘模型.....	283
15-4-3	建立聚集 OLAP 挖掘模型.....	290
15-5	挖掘模型角色管理	291

4-1 SQL 2000 中的“分析管理器”简介



SQL 2000 与分析 管理器

注意

注意

1-1 SQL 2000 中的“分析管理器”简介

本书主要介绍的对象——“分析管理器”(Analysis Services, 本书以下将用中文“分析管理器”来说明), 是在 SQL 2000 架构中, 用来提供“联机分析处理”(OLAP)与“数据挖掘”(Data Mining)的中介层服务器。“分析管理器”为前一版本 SQL 7.0 中“OLAP Service”的升级功能, 有关整体的功能简介, 请参考 1-2 节“分析管理器”的功能架构。

备注

note

“分析管理器”服务器必须安装在“Microsoft Windows® 2000 Server”上或搭配有“Service Pack 5”的“Microsoft Windows NT® Server 4.0 的操作系统”。严格来说, 本书所谓的“分析管理器”, 包含“分析管理器”服务器(Analysis Service Server)及其所提供的服务、管理功能, 包括管理 Analysis Server、建立数据库与指定数据源、建立与处理多维数据集、建立与处理数据挖掘 (Data Mining) 模型、指定数据存储类型选项及基于使用的优化查询效率、安全性管理、浏览数据源、共享维度、安全性角色以及其他对象等, 这些也就是本书要介绍的主要内容。

“分析管理器”会预先计算 SQL 2000 数据仓库(Data Warehouse)的“汇总”数据, 并以事先定义的方式(也就是“维度”), 组织并存储到多维数据集中, 以便快速地响应决策分析者复杂而多维的分析查询。

“分析管理器”也可以让用户从多维数据集或关系型数据库中建立数据“挖掘模型”, 以进行数据的探索与预测。此外, “分析管理器”还搭配提供了 PivotTable® 的功能(数据透视表), 它与“OLE DB 提供者”(OLEDB Provider)兼容, 以便让分析的数据可以提供如 Excel 的数据透视表, 或可以在其他应用程序中用来进行数据分析, 甚至是建立脱机多维数据集分析模式。

当我们完整安装了 SQL 2000 中的“分析管理器”之后, 便可以使用它了。同时我们也可以于 Windows 2000 或是 Windows NT 的“开始/程序”中的“Microsoft SQL Server/Analysis Service”命令下选择“分析管理器”来打开“分析管理器”。打开后的“分析管理器”画面, 如图 1-1 所示。

备注

note

在不侵犯知识产权的情况下, 用户可以在没有安装 SQL 2000 任何版本的环境中, 单独安装“分析管理器”并且执行“分析管理器”的大部分功能。但是, 有些功能将会因为用户没有安装 SQL 2000 或安装的版本不同而受限制, 例如根据 SQL 2000 的说明, 12-4-1 节里的“建立计算单元”功能, 便必须在 SQL 2000 企业版中才可以顺利执行。而每一个 SQL 2000 版本可以支持“分析管理器”的功能, 根据“SQL 联机丛书”的说明简化成表 1-1。

表 1-1 每一版本对于“分析管理器”中功能的支持

SQL 2000 不同版本对于“分析管理器”中功能的支持	企业版	标准版	个人版
Analysis Services	支持	支持	支持
用户自定义的 OLAP 分区	支持	N/A	N/A
多维数据集数据分区向导	支持	N/A	N/A
链接的 OLAP 多维数据集	支持	N/A	N/A
ROLAP 维度支持	支持	N/A	N/A
HTTP Internet 支持	支持	N/A	N/A
自定义汇总	支持	支持	支持
计算成员	支持	N/A	N/A
写回	支持	N/A	N/A
超大型维度支持	支持	N/A	N/A
对策	支持	支持	支持
分布式分区多维数据集	支持	N/A	N/A
数据挖掘 (Data Mining)	支持	支持	支持

在 SQL2000 版本中的“分析管理器”，比 SQL 7.0 的“OLAP Service”新建了许多功能。除了更灵活性地增加了许多分析的层次，以便让企业或是一般需要进行决策分析的决策人员可以从完整的企业数据仓库中，以符合决策的观点，分析过去的历史数据外，更包含了所谓的“Data Mining”（在“分析管理器”中称之为“数据挖掘”，而将于第 15 章中说明，而有关 OLAP 与 Data Mining 的概念，请参考 1-3 节），让决策者可以根据过去的历史数据，从不同的维度中，根据所观察的度量值去预测企业发展未来的可能趋势。

重点

important

OLAP (On-Line Analytical Processing)——联机分析式处理程序，我们可以将它解读成一种“数据分析的程序”与“数据实体存储方式”，将来为数据库中的数据和使用特殊的数据结构，存储各种不同的“维度”（决策者分析数据的角度）与查询对象（度量值）的结果值。

当我们打开“分析管理器”之后，系统会参照目前机器上的名称作为默认的“分析管理器”的主要服务器名称，例如目前所在的机器上，名称为“VIVIAN”，则默认的分析管理器服务器，名称定为“VIVIAN”。但如果在整个联机的环境中，还包含有其他的可连接数据库（也就是说其他的机器上也安装了“分析管理器”），则在本机器的“分析管理器”中，也可以注册并连接到其他机器的“分析管理器”服务器上。

倘若在整个企业的联机环境上，包含前一版本的数据库服务器(SQL 7.0)，则“分析管理器”一样可以注册并连接，只是进行分析时，便只能用到前一版本 OLAP Service 的功能。同时，系统也会在“分析管理器”主环境中加以区别的，如图 1-1 所示。

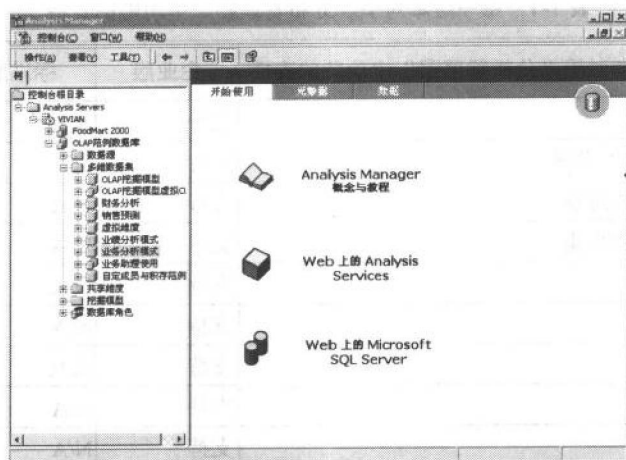


图 1-1 打开“分析管理器”之后所看到的画面

“分析管理器”可以帮我们管理不同版本的数据库分析(如 SQL 7 与 SQL 2000)服务器，每一个数据库服务器下，又可以管理多个机器上的“分析管理器”；每一个计算机上，又可以包含多个“数据库”，每一个数据库可以设置几个不同的数据源和多个多维数据集分析对象、共享维度、挖掘模型与管理角色。

技巧

Trick

在“分析管理器”中，每一个不同的组件都有不同的图标代表。例如，图 1-1 中的  图标就代表“数据库”，要取得这部分的信息可以由“SQL Server 联机丛书”中的“用户界面帮助参考”里取得相关信息，如图 1-2 所示。

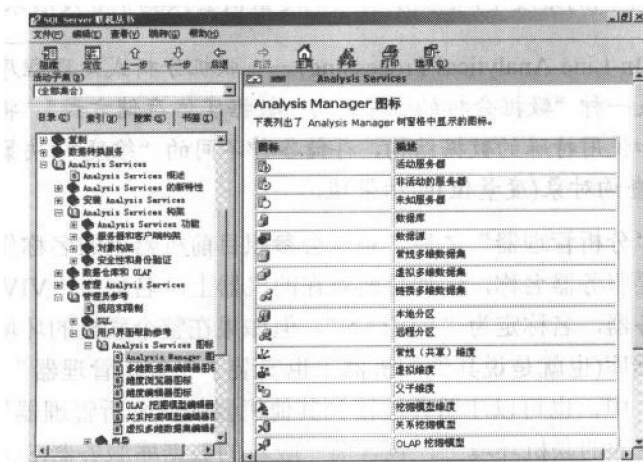


图 1-2 运用 SQL Server 取得整理信息

因此，我们知道，应用“分析管理器”的第一步骤便是建立连接的机器服务器，当然本机上的服务器为必然的服务器，如目前笔者所使用的机器名称 VIVIAN，则安装好计算机后，便会建立至少一个机器服务器。在该服务器中，也会包含有一个默认的范例数据库 FoodMart 2000，其中也包含已经建立的数据源与分析多维数据集及相关的共享维度。

1-2 “分析管理器”的功能架构

根据 SQL 2000 里的“SQL Server 联机丛书”所提供的功能架构，“分析管理器”的功能大致上可以从以下几个角度说明。

1-2-1 易于使用

如现在的一般应用软件一样，“分析管理器”也提供完整的“在线教程”，可以让用户在短短时间内，即可对“分析管理器”的用法有一个初步的概念。我们打开“分析管理器”就可以看见有一个“Analysis Manager 概念与教程”，激活后，系统会以一般网页的方式逐步介绍、引导用户建立基本多维数据集及其他高级工作 (如建立分区、虚拟多维数据集、安全性角色、可写入维度、对策及数据挖掘等) 的步骤，如图 1-3 所示。

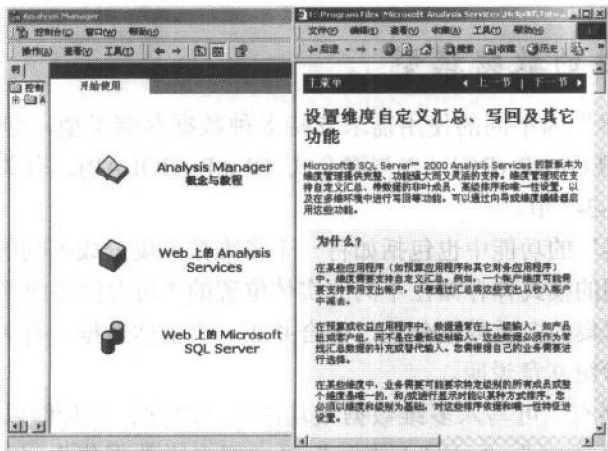


图 1-3 利用“分析管理器的概念与教程”进行学习

“分析管理器”主要支持的两大功能分别为：“联机分析处理”(OLAP) 与“数据挖掘”(Data Mining)。为了用户更易于使用此两种技术，在“分析管理器”中，多以“向导”、“编辑器”以及完整的说明信息来协助相关用户。其中“分析管理器”控制台提供完善的用户接口，以便模式建立者建立分析模式。如图 1-4 所示。在本书的各章

节中，我们也多是用“向导”和“编辑器”来说明问题。

此外，在 4-2-1 节中，我们也会学习到，用户不需要离开“分析管理器”便可以查看多维数据集和维度分析的结果。甚至还可以查看数据挖掘模型的数据。

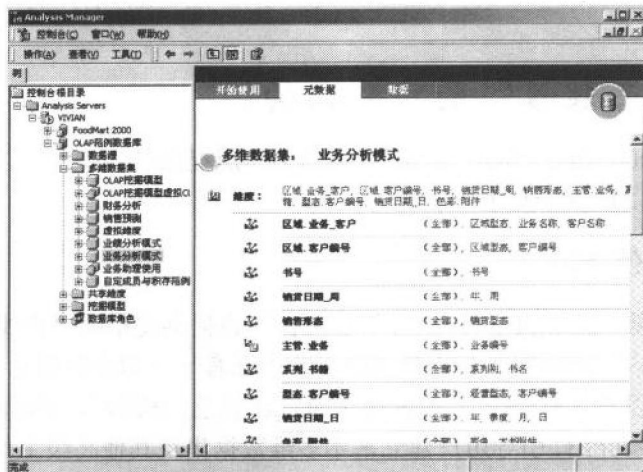


图 1-4 “分析管理器”控制台提供清晰的层次对象架构与完整的功能，以“向导”与“编辑器”辅助分析模式的建立者与用户

1-2-2 灵活性的数据模型

“分析管理器”因不同的使用需求支持多种数据存储类型，包括多维度 OLAP (MOLAP)、关系型 OLAP (ROLAP)与混合式 OLAP (HOLAP)。有关 OLAP 模式数据存储类型请参考 1-3-3 节。

“分析管理器”的功能中也包括如将一个多维数据集分成不同的分区，而每一个分区可以使用不同的模式体存储在不同的实体位置的“可分区多维数据集存储”上，也可以将多维数据集的多重数据分区，再合并成一个实体数据。有关多维数据集的分区与应用，请参考第 9 章说明。

另外，还包括有“可写入多维数据集功能”、“对称、不对称以及参差不齐的层次式架构”、“父子维度”、“可写入维度”、“虚拟多维数据集”，以及应用多维表达式(MDX) (如图 1-5)、数学公式与用户自定义函数以建立的“导出量值成员”和“计算成员”、“计算单元”、“成员属性”与“虚拟维度”等。这些议题都将在本书中一一介绍。



图 1-5 使用 MDX 语法建立导出度量值，维度成员以作为自定义的分析对象

1-2-3 可调整性

“分析管理器”也提供了可调整性、具有延展性的架构和符合各种数据分析与应用的需求。包括：

存储设计功能

在 OLAP 分析模式中，系统会将决策分析者最后所需的汇总数据，根据所期望分析的角度予以计算并存储，但是究竟要先完成多少计算，以便可以在有限的存储空间中，得到最佳的查询处理效率，便是一个重点。“存储设计向导”可以在系统效率与配置给存储汇总的磁盘空间之间取得最佳的平衡点。

基于使用的优化

建立多维数据集分析模式时，会以各种“可能”的分析角度来建立维度，以“可能”的汇总计算方式作为量值或计算成员的设置。但是并不是所有组合被查询或应用的频率都一样。“分析管理器”利用“基于使用的优化向导”（如图 1-6 所示）调整多维数据集的效率，以便为最常执行的查询提供快速的响应。如此，用户便可以很快地建立一个汇总数最少的系统(所占空间最少)，以后再按照系统的实际使用状况逐步调整。这一部分请参考 11-3 节说明。

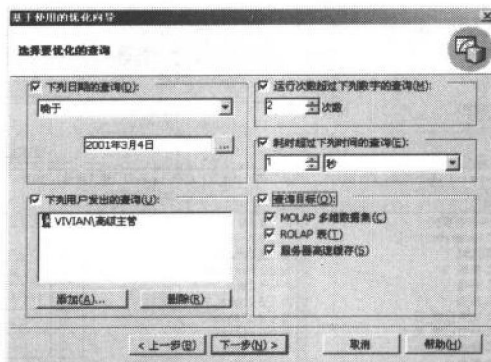


图 1-6 系统提供基于使用的优化功能，
可以不断调整执行性能达到优化

此外，还包括有“分区”、“数据压缩与存储最优化”、“分布式查询”、“链接多维数据集”和“增量更新”等，甚至还包括对于 LAN、WAN、Internet 以及动态通讯的支持与连接的 HTTP 验证，主要是让用户可以创建一个更具有灵活性与效率的分析环境。

1-2-4 集成性

“分析管理器”除了本身提供强大的模式建立与分析功能外，还可以配合其他组件和程序使用，以强化企业数据集成的功能。

集成的安全性

在“分析管理器”中对于数据的安全性，不论是多维数据集或是数据“挖掘模型”，都以 Microsoft Windows NT 或是 Windows 2000 的安全性考虑为基础，也就是操作系统层次的用户或组的认证，以便使用 SQL Server 2000 中的数据、多维数据集等，如图 1-7 所示。有关角色类别的定义与安全性的设置，请参考 11-1 节说明。

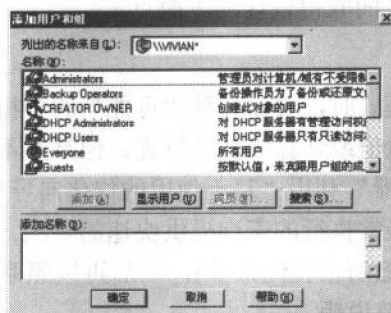


图 1-7 设置类别角色时，将会以 Windows NT 的
认证为基础，进行数据的保护