HZ BOOKS

原创精品系列

商务智能与数据挖掘
Microsoft SQL Server应用

谢邦昌　主编
华通人Data Mining团队　参编

机械工业出版社
China Machine Press

TP311.138/530
2008

原创精品系列

商务智能与数据挖掘
Microsoft SQL Server应用

谢邦昌　主编
华通人Data Mining团队　参编

机械工业出版社
China Machine Press

本书主要讨论数据挖掘技术的基本原理与应用，可以解决企业运营中遇到的各种问题，并介绍了SQL Server 2005 处理这些问题的方法。内容主要包括数据仓库、数据挖掘中的主要方法、SQL Server 2005 中的商业智能与数据挖掘功能、决策树模型、聚类分析、神经网络模型和时间序列模型等，并配有相关的范例分析与实例练习。

本书内容翔实，示例丰富，结构合理，可作为各类开发人员及企业管理人员的参考用书。

版权所有，侵权必究。

本书法律顾问　北京市展达律师事务所

图书在版编目（CIP）数据

商务智能与数据挖掘Microsoft SQL Server应用/谢邦昌主编. —北京：机械工业出版社，2008.2

ISBN 978-7-111-23241-4

Ⅰ.商… Ⅱ.谢… Ⅲ.关系数据库－数据库管理系统，SQL Server Ⅳ.TP311.138

中国版本图书馆CIP数据核字（2008）第007173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：王春华

北京京北制版厂印刷 · 新华书店北京发行所发行

2008年3月第1版第1次印刷

186mm × 240mm · 22印张

标准书号：ISBN 978-7-111-23241-4

定价：49.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线（010）68326294

推 荐 序

过去20年，企业已累积了大量的商业数据，并运用数据仓库来分析过去的信息，然而，对于过去的了解，并不表示就拥有丰富的商业知识，数据挖掘提供的预测功能，可协助企业洞悉商机，也是现今提升企业竞争力的重要技术。

要发挥数据挖掘的最大功效，有3项要素不可或缺：了解算法并加以运用、具备Domain Know-How分析对的问题、熟悉工具的使用并与现行系统整合。这本书不但深入浅出地介绍了这些关键要素，还配有范例，帮助读者融会贯通。微软也期望借由Microsoft SQL Server易于建立的、高弹性的平台，让那些过去从未想过使用数据挖掘的企业，通过完善的架构、与IT信息系统深度的整合、搭配完整的分析工具以及Office易于使用的前端接口来建立全新的商业智能系统，从而提升企业利润并有效降低成本。企业可以针对各种常见的商业问题自定义数据挖掘的解决方案。

谢邦昌教授是数据挖掘领域专家中的专家，听过他演讲的客户都称赞其演讲内容、实际案例分析和技术应用等非常契合企业的需求。而谢邦昌教授对新技术的引进及学习的积极态度，更让后生晚辈的我们敬佩及推崇。本书由数据仓库、数据挖掘与商业智能概念开始谈起，进而深入地探讨了如何利用Microsoft SQL Server建立数据仓库、数据挖掘与商业智能之平台。期望借由谢邦昌教授对九大算法的精辟介绍及实际案例讲解，并搭配Microsoft SQL Server平民化的价格与易用整合的平台及Office前端，读者能有效达到数据挖掘的三大关键要素，快速进入数据挖掘的神秘殿堂。诚挚推荐这本不可错过的好书。

微软大中国区　市场战略部
策略规划经理
黄淑翠

序

从事市场调研行业这么多年，一直都在面临一个挑战：市场调研公司能够为客户带来什么价值？或者说客户花的钱是否能够得到更高的回报？

要解决好这个问题，我想取决于两点：第一，提供的数据是准确的；第二，能够从这些数据中得到有价值的信息，从而为客户提供高附加值的服务。

我们一直在为这两个方面的完善而努力。华通人公司拥有丰富的宏观行业、企业以及多年来积累下来的调查数据，我们都坚信这些数据就是一个宝藏，并且这个宝藏中积累的财富会越来越多。但如何将这些宝藏挖掘出来，从中得到更高价值的信息，我们曾经很困惑，感觉无从下手，因为这些数据貌似彼此间孤立，并且摆在一起更显得纷繁复杂。

直到我们接触到数据挖掘（Data Mining），这个困扰已久的问题才得以解决，这都受益于我们与谢邦昌教授进行的深度合作。谢教授作为辅仁大学统计信息学系教授以及中华资料采矿协会理事长，长期以来从事数据挖掘的理论教学和推广应用工作，并从最近几年开始，着力将数据挖掘的理论与方法向学术届和商业应用领域推广。正是在和谢教授的合作中，我们发现数据挖掘正是我们需要的工具，我们可以将杂乱无章的数据建立起数据仓库（Data Warehouse），并在此基础上进行深度挖掘。我们的员工也如获至宝，开始潜心学习数据挖掘技术，组成了单独的数据挖掘团队，不但对我们自己的数据资源进行挖掘，也将这一技术融入到各个项目中去，越来越多的数据挖掘技术的应用，极大提高了项目的质量与水平，实现了为客户提供高附加值服务的期望。

我们对数据挖掘如饥似渴，也认为数据挖掘技术对市场调研行业价值的升华起到了真正意义上的帮助，但在我们面前还有一个问题，那就是现在还没有一本优秀的、基于商业应用的数据挖掘教材。谢教授的这本著作正好解决了这个难题，这本书深入浅出地讲授了如何实现数据挖掘的商业应用，书中的例子更是让我们直观地了解了数据挖掘的应用价值。

我们非常荣幸能够为本书的出版出一份力，我们公司的数据挖掘团队负责这本书的编著和修改，以适应读者的阅读习惯和需求。为此他们付出了很多努力，在这里对他们所做的工作表示感谢，华通人数据挖掘团队包括——组长：赵巍；组员：池碧云、付强、洪晶、武文捷、郑韶霞（顺序按拼音首字母排列）。

这本书的出版是数据挖掘技术普及与推广应用的一个美好开端。

北京华通人商用信息有限公司名誉董事长

高余先

试读结束：需要全本请在线购买：www.ertongbook.com

目　录

第1章 绪 论

1.1 商业智能

1. 何谓商业智能

根据IBM的统计，在电子商务时代，数据量正以每年1.3倍的速度迅速扩增。然而在大量数据中，真正被利用来分析与运用的部分却不到一成。因此如何将这些庞大的数据快速转换成决策者所需的信息，以作为提升企业运营所需的企业智慧，已经成为经营管理的一大挑战，而商业智能的应用也因此而逐渐受到企业界的重视。通过商业智能的运用，企业可将原始的顾客数据做更深入的分析，进而建立有效的预测模式及顾客市场区分，使CRM（Customer Relationship Management）的运用更具成效，也有助于未来KM（Knowledge Management）的落实。商业智能的观念是指利用组织化及系统化的流程来取得、分析、发布对其商业活动有重大影响的信息；利用商业智能的协助来预测客户或竞争者的行动，及市场活动或趋势的变化情形。

所谓商业智能（Business Intelligence, BI）指的是企业利用信息科技以企业内部及外部既有的数据库数据为基础，根据所需解决的问题进行数据的汇总，整合成数据仓库后，利用适当的工具进行数据处理，利用联机分析（OLAP）及数据挖掘（Data Mining）等技术分析数据，将所发现的潜在特性或是建立的预测模型传递给决策者，以提供协助其进行决策，并达到企业目标。

一般认为，商业智能是一种利用信息科技，将现今分散于企业内部、外部结构化数据加以汇总，并依据某些特定需求进行分析与运算，再以最佳的方法，将结果呈现给决策者、管理者或是知识工作者的一种分析机制。换句话说，企业将可通过商业智能使得企业中的决策者能够获得适当的信息，以协助其做出最正确的决策。

2. 商业智能的作用及意义

商业智能之所以重要，不外乎是由于企业之间的激烈竞争，企业经营者为求生存不得不竭尽所思协助企业持续生存下去，因此企业主们必须随时随地根据所掌握的信息作出实时决定。但是事后回过头来审视这些实时决策，会发现其中包含了有效解决问题以及无法解决问题的决策。除了决策者本身个性会影响决策外，影响决策有效性最重要的因素就是做决策时所掌握信息的充足性及正确性，而商业智能的含义就是指通过企业所拥有的数据和数据仓库的汇总，结合联机分析及数据挖掘分析技术挖掘出潜藏在数据库中的有用信息，并将其提供给决策者或部门主管作为平时运营的决策依据，而当企业面临危机时或必须立即做出重大决策时，也能依据

数据仓库所提供的正确数据及时做出正确的决策，协助企业顺利解决问题，化危机为转机，更可见商业智能的重要性。王茁在《商业智能》一书中提到“商业智能所争取的就是充分利用企业在日常经营过程中搜集的大量数据，并将它们转化为信息和知识来避免企业中的猜测行为和无知状态。”

对一般企业来说，商业智能主要可以应用在以下几个方面。

（1）了解运营状况

商业智能可以用来帮助企业了解本身运营状况及其推动力量，协助用户清楚了解产品未来趋势、运营上出现哪些异常情况和哪些行为正对业务产生影响。

（2）衡量绩效

商业智能可以用来确立对员工的期望，追踪并管理其绩效。

（3）改善关系

商业智能可通过顾客关系管理（CRM）的整合运用，有效为顾客、员工、供货商、股东和大众提供关于企业及其业务状况的有用信息，从而提高企业的知名度，强化整体信息的一致性。利用商业智能，企业可以在问题变成危机之前，很快地侦测出来问题所在并提出相关建议方案加以解决。商业智能也有助于加强顾客忠诚度，一个参与其中并充分掌握信息的顾客更加有可能会购买你的产品或服务。

（4）创造获利机会

掌握各种商务信息的企业可以出售这些信息从中获取利润。但是，企业需要发现信息的买主并找到合适的传递方式。

近年来，企业发展的节奏越来越快，商业复杂性越来越高。虽然许多因特网的企业都消失了，但是因特网的速度不仅没有减慢而且更加突显出其意义。不论企业规模的大小，都需要面对瞬息万变的市场趋势，并根据既有信息做出决策，而这些决策的依据是正确无误的信息。由此可见，信息在企业经营管理中的重要性仅次于人才。

3. 商业智能架构

有许多人会将商业智能误认为企业中技术性层次的电子化解决方案，然而商业智能却是整合了“管理”、“决策”及“信息科技”三项要素的有效分析机制，因此企业必须以策略层次的观点来考察商业智能才能了解其重要性。就应用方面来看，因为现今信息科技与因特网的兴起，商业智能的应用范畴日益增加，不论是企业界中众人熟知的顾客关系管理、供应链管理、企业资源规划或者是知识管理，都是商业智能实务上的运用。为了要使得企业中的决策人员实时地取得正确及其所需要的数据，商业智能的操作性层面工具可以说是商业智能中最重要的核心，这些工具包含了数据仓库（Data Warehouse）、联机分析（OLAP）、数据挖掘（Data Mining）等。

在实务上，以商业智能在顾客关系管理上的应用为例，企业常通过数据仓库技术汇总来自

不同数据库的信息，利用数据挖掘技术来进行各项分析，并由此针对顾客过去购买记录、个人基本数据等，分析顾客的产品贡献度、建立市场细分，以便于销售方案的制定，或是针对不同特性顾客进行交叉销售（Cross selling）与向上销售（Up selling）以提升产值。

4. 商业智能流程

商业智能于企业中的实施流程如图1-1所示。由图中我们可以了解，企业导入商业智能应用方案过程前，必须清楚地了解企业本身对于导入商业智能的需求是什么，也就是必须明确企业导入商业智能解决方案的原因、整合的组织层级、各部门支持的程度和企业主本身对此的重视程度等。若企业主不重视，各部门不提供协助，或是项目主持者的层级不够，即使商业智能解决方案再完整，也无法解决企业的问题，无法达成企业需求，而终将面临失败。

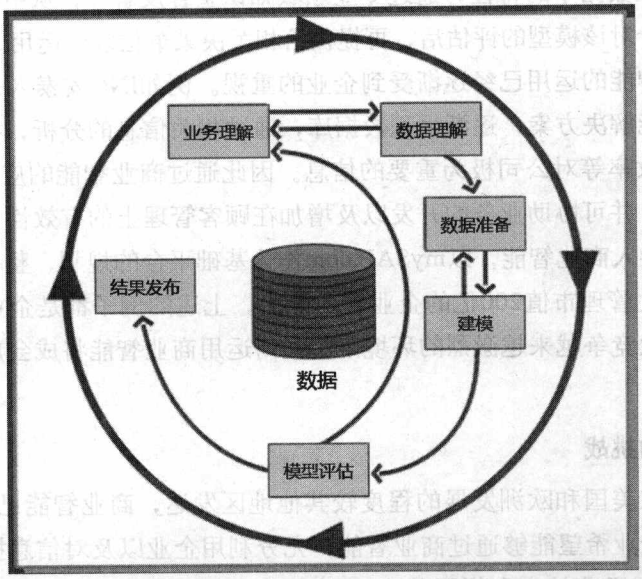


图 1-1

资料来源：www.crisp-dm.org

当企业已经了解导入商业智能的目的及其需求时，下一步则是要了解企业本身所拥有的数据。商业智能的解决方案不外乎是针对企业既有的数据通过加值分析探索出潜藏的特性，因此对企业本身数据的掌握就更显重要。

以往的企业中，各部门常会根据本身的需求，对顾客通过促销活动等方式搜集顾客的信息，但往往是仅根据部门本身需求而制定，而缺乏整体性考虑，无法将信息与企业整体的顾客数据库进行整合，造成许多重要的顾客信息不完整，所分析出的信息也是容易造成偏差，无法真正为企业解决问题。商业智能的核心工作在于，根据企业数据库整合成可以作为分析用的数据仓库，再进一步通过分析技术来探索数据。而对于数据仓库的建立，《Building the Data

试读结束，需要全本请在线购买：www.ertongbook.com

Warehouse》[⊖]的作者William Inmon认为数据仓库必须具备有“面向主题（Subject-oriented）”、“整合性（Integrated）”、“时间转化（Time-variant）”及“不易变化（Non-volatile）”四个特性。因此事实上数据仓库是有别于传统的数据库系统，这点是必须特别注意的。

企业在建构商业智能基础建设的过程中，实时数据查询分析功能扮演非常重要的角色。简单来说，联机分析就是能让用户依据本身决策需求来浏览数据、动态且实时的产生其所需的报表，以提高分析效率的技术。事实上，它除了能提供在线实时数据分析模块外，更重要的是能展示多维度（Multi-Dimensional）的数据。

然而商业智能的另一项重要技术是数据挖掘的分析技术，主要是在大量数据库中寻找有意义或有价值的信息的过程。通过机器学习技术或是统计分析方法论，根据整合的数据加以分析探索，发掘出隐含在数据中的特性，通过专业领域知识整合及解释，从中找出合理且有用的信息，经过相关部门针对该模型的评估后，再提供给相关决策单位加以运用。

近年来，商业智能的运用已经逐渐受到企业的重视。例如ING安泰人寿自1998年起，逐步导入IBM的商业智能解决方案，逐渐累积数据库，通过相关信息的分析，找出顾客区分、消费行为、业务成本与效率等对公司极为重要的信息。因此通过商业智能的应用，ING安泰人寿能够更深入了解客户，并可协助业务的开发以及增加在顾客管理上的有效性。另外，全球企业可口可乐公司也通过导入商业智能，以mySAP.com作为基础平台的规划，整合财务信息，提升财务规划能力，以强化管理市值200亿的企业管理能力。上述的例子都是企业运用商业智能的最佳典范，因此在企业竞争越来越激烈的环境下，如何运用商业智能将成会成为企业强化竞争力的重要关键之一。

5. 商业智能中的挑战

商业智能活动在美国和欧洲发展的程度较其他地区发达，商业智能已经变成企业e化的主要项目之一。欧美企业希望能够通过商业智能，充分利用企业以及对信息技术的投资、改善决策、提高利润、提高运营效率和增强信息透明度。然而针对欧美企业应用商业智能的目的，Gartner在2002年进行的商业智能调查中发现，美国企业与欧洲企业对于商业智能工具的使用略有不同，美国企业主要是利用商业智能作联机分析，而欧洲企业则是通过商业智能进行高级分析。

较可惜的是，纵观欧美企业对商业智能的应用层面，商业智能的运用并未被广泛提升到策略性层面。这导致有些企业即使导入商业智能，也不一定可以成功运用商业智能。有些企业这方面的项目由于资金不足、人员不足或者由于采取了未能与策略性的运营目标整合一致的操作而终遭失败。由于缺乏有效、适当的规划，很多项目变得僵化、孤立，无法适应不断变化的市场环境，最后不得不面对废止的命运。

美国著名商业智能专家Shaku Atre于2003年所提出的商业智能白皮书中明白指出，企业的

⊖ 本书中文版《数据仓库》（原书第4版）（书号7-111-19194-3, 定价39.00）已由机械工业出版社出版。——编者注

商业智能项目之所以失败的十大原因，包含：

未能认识到商业智能项目是跨部门的商务整合计划，未能理解商业智能不同于孤立的解决方案：

- 缺乏积极参与的支持者或支持者在企业中没被充分授权。
- 缺乏来自业务部门的代表或参与者不够主动积极。
- 缺乏有技术、有执行能力的人员或者未充分利用人力。
- 缺乏有回馈机制的软件开发方法。
- 缺乏分工、缺乏方法论。
- 缺乏商务分析或活动标准。
- 缺乏对“劣质数据影响一切”的认知和对策。
- 缺乏对元数据(Metadata)的必要性和使用方式。
- 过分依赖分散的方法和工具。

从上述原因中，不难看出其主要原因不外乎是企业没有把商业智能看成是影响企业兴衰和存亡的大事。如果企业把商业智能和数据仓库看成是策略性问题而不是一般性或不重要的问题，就会提升实行商业智能项目成功的可能性。

微软公司的Microsoft SQL Server 2005是一个完整的商业智能平台，为用户提供了可用于构建典型和创新的分析应用程序所需的各种特性、工具和功能。其中引入了大量成熟可靠的数据挖掘技术，帮助企业求解企业运营中遇到的种种问题。本书将讨论数据挖掘可以解决的各种问题，并介绍SQL Server 2005处理这些问题的方法。

1.2 数据挖掘

在信息科技发展日新月异的今天，在软件技术不断改良和硬件购置成本大幅降低的条件下，数据处理与存储管理不再成为问题。这也间接带动了企业对运营相关数据库的建立与应用。

知识经济时代的来临，企业间的竞争模式，从传统的“红海策略”，即采取压低成本与价格的杀价流血竞争，到近来倡导以创新为核心竞争力的“蓝海策略”。不论哪一种策略模式，都是不断地从研发、制造、营销、客服或资源配置等运营的相关问题上，寻求问题的发生原因，并尝试找出解决方案。而在整个运营阶段中，陆续累积的庞大数据，往往就是答案的隐身之处！因此，如何善用数据，从运营历史的记录里，挖掘出深藏其中的宝贵经验(金矿)，就是数据挖掘(Data Mining)的目的。

企业在尝试分析其数据时都面临若干问题。一般而言，并不缺乏数据。事实上，很多企业感觉到自己被数据淹没了，没有办法完全利用所有的数据，将其变成有用的信息，尤其是当数据从不同的操作系统涌入时，如何得到一致性的信息，这是一直困扰的问题。为了处理这方面的问题，专业人员开发了数据仓库（Data Warehousing）技术，以便企业从来源于各种不同操作系统间的数据，加以处理并将其变成有用的信息。

一个适当运作的数据仓库是具有强大功能的解决方案。公司可以对信息进行分析，并将其加以利用，以进行明智的决策。通过使用数据仓库，可以为您提供诸如以下问题的答案：

- 哪些产品最受15～20岁的女性欢迎？
- 特定消费者的订单前置时间和按时交付的百分比与所有消费者的平均值相比如何？
- 病房花在每个患者身上的成本和时间是多少？
- 在签约阶段停滞时间超过十天的项目所占的百分比为多少？
- 如果某个特定的实验室在某类特定的药品上投入了较多的资金，临床试验结果是否显示病人健康状况好于其他实验室？

除了这些通常可使用分析应用程序得出答案的问题之外，数据仓库还支持各种数据交换格式。分析应用程序设计供分析人员使用，分析人员会对数据进行分类，研究有助管理与决策的分析结果；报表应用程序会产出书面报表或在线报表，这些报表功能要求略低的用户使用，提供静态内容，或提供有限的深入挖掘功能；对于业务决策者而言，计分卡是非常强大的功能，可以提供公司关键性能指示（Key Performance Indicator, KPI）的概况，使决策者知道其身处何处。

尽管数据仓库功能强大而实用，但其自身有一个局限，它实质上反映的是过去的历史。由于数据仓库经常在特定周期或时间点进行数据加载和处理，因此它只是表示一个时间点上的快照（Snapshot）。即使建构了实时（real-time）或近似实时（near real-time）的数据仓库，其数据仍然只表示当前和历史的数据，无法达到“预测”的需要。因此，为发现数据的因果关系，数据仓库需要利用其他科学方法，进行定量分析。

与传统的统计分析方法不同的地方是，数据挖掘不是让人提出假设，然后据此去找相关数据，而是让数据仓库确定数据关联性，并允许采用与以往不同的模式对数据进行分析。通过数据挖掘，可以得出诸如以下这些问题的答案：

- 客户将购买什么产品？哪些产品将一起销售？
- 如何预测哪些消费者可能会流失？
- 市场状况如何，将会如何发展？
- 企业如何对其网站使用模式进行最佳的分析？
- 组织如何确定营销活动是否成功？
- 什么是分析非结构化数据（如文本文件、视频文件等）的最好技术？

第2章 数据仓库

2.1 数据仓库定义

数据仓库（Data Warehousing）是运用新信息科技所提供的大量数据存储、分析能力，将以往无法深入整理分析的客户数据建立成为一个强大的顾客关系管理系统，以协助企业制定精准的运营决策。“数据仓库”对于企业的贡献在于“效果”，能适时地提供高级主管最需要的决策支持信息，做到“在适当的时间将正确的信息传递给适当或需要的人。”简单地说，就是运用信息科技将宝贵的运营数据，建立成为协助主管做出各种管理决策的一个整合性“支持系统”，利用这个“支持系统”，企业可以灵活地分析所有细致深入的客户数据，以建立强大的“顾客关系管理”优势。

2.2 数据仓库特点

数据仓库与传统的数据库是有所不同的。数据库是未经整理后的一大堆数据集；而数据仓库是从数据库中萃取出来，经过整理、规划、建构而成的一个有系统的数据库的子集合。数据仓库具有下列几个特点：

1. 面向主题（Subject Orient）

数据仓库的信息系统，数据建立的着重点在于以重要的主题组件为核心，作为建构的方向。数据需求者只要把要研究的相关主题数据，从数据库中撷取、整合之后就可以做研究分析。

2. 整合性（Integrated）

各应用系统的数据须经过整合，以便利执行相关分析操作。

3. 长期性（Time Variance）

为了执行趋势的分析，数据仓库系统常须保留1~10年的历史数据。这与数据库为日常性的数据有所不同。

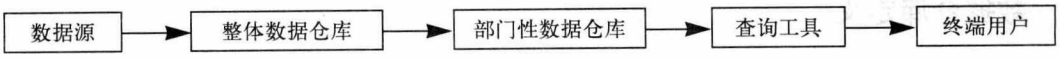
4. 稳定性（Non-Volatile）

数据库（Database）的数据可以随时被修改，但是数据仓库的数据，并非日常性的数据而是历史性的数据，通常作为长期性分析用途，只有内部相关人员会定期性的修改数据结构，但频率不会太高。数据仓库并不允许用户去做更新的动作，所以其数据是较少有变动。

由于数据仓库内的数据具备上述特点，所以数据仓库必须通过一连串的程序（配合良好的软硬件设备）方可建立完成，而非一个即买即可使用的产品。

2.3　数据仓库架构

建立数据仓库（Data Warehousing）是一种能正确地组合与管理不同数据来源的技术，其目的在于回答您业务经营上的问题以便让您做出正确决定。数据仓库的整体架构如下：



数据仓库的基本架构及整体概念，我将它区分为以下几个基本组件来加以说明：

专业顾问通过与企业进行需求访谈，建立数据仓库的Model，然后将企业内各种数据整合于数据库中，并建立前端分析数据的工具以及管理工具，这样的过程即为建立数据仓库的基本过程。

- 设计（Design）。即数据仓库的数据Model的设计，这部分是最重要的。若Model设计的不够周全或不理想，不管之后的报表设计如何精美，也有可能跑出错误的信息，这也是需要选择有经验的专业顾问建立数据仓库的一个重要原因。
- 整合（Integrate）。即数据的整合转换过程，包含数据解释（Data Extraction）、数据转换（Data Transformation）、数据清理（Data Cleaning）、数据加载（Data Load）。将各种来源的数据整合、转换并加载数据仓库中，程序编写较为繁杂，自动化处理困难，经常要人工参与操作，这一部分约占DW项目的60%~70%的人力及时间。
- 管理（Management）。即数据仓库的中心，是一个巨大容量及提供突发定向（ad-hoc）查询的数据库。
- 可视化（Visualize）。即前端呈现给用户看的形式，例如数据挖掘（Data Mining）及OLAP工具，用以呈现分析过的数据形式。
- 调度（Administration）。为管理的工具，例如网络监控流量、安全管理等。图2-1是一完整数据仓库的逻辑架构。

由IT用户（IT Users）将平日操作数据存入至操作/数据源数据库（Operational/Source Data），通过多种数据转换工具将数据以各种转换方式汇总至整体数据仓库。再由整体数据仓库，使用数据复制、分布工具，依其需求将数据复制并散布至各部门的数据仓库。提供业务用户，以各种不同的信息存取方式及工具，完成各类业务信息需求。其中信息的存取工具必须能够提供至部门及整体数据仓库的存取功能，否则数据仓库将因其本身的架构及组成工具限制了用户对信息取得及整体仓储的价值。如图2-2所示。

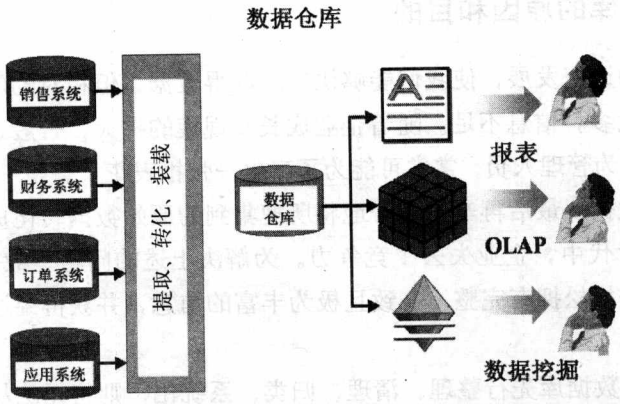


图2-1 数据仓库逻辑架构

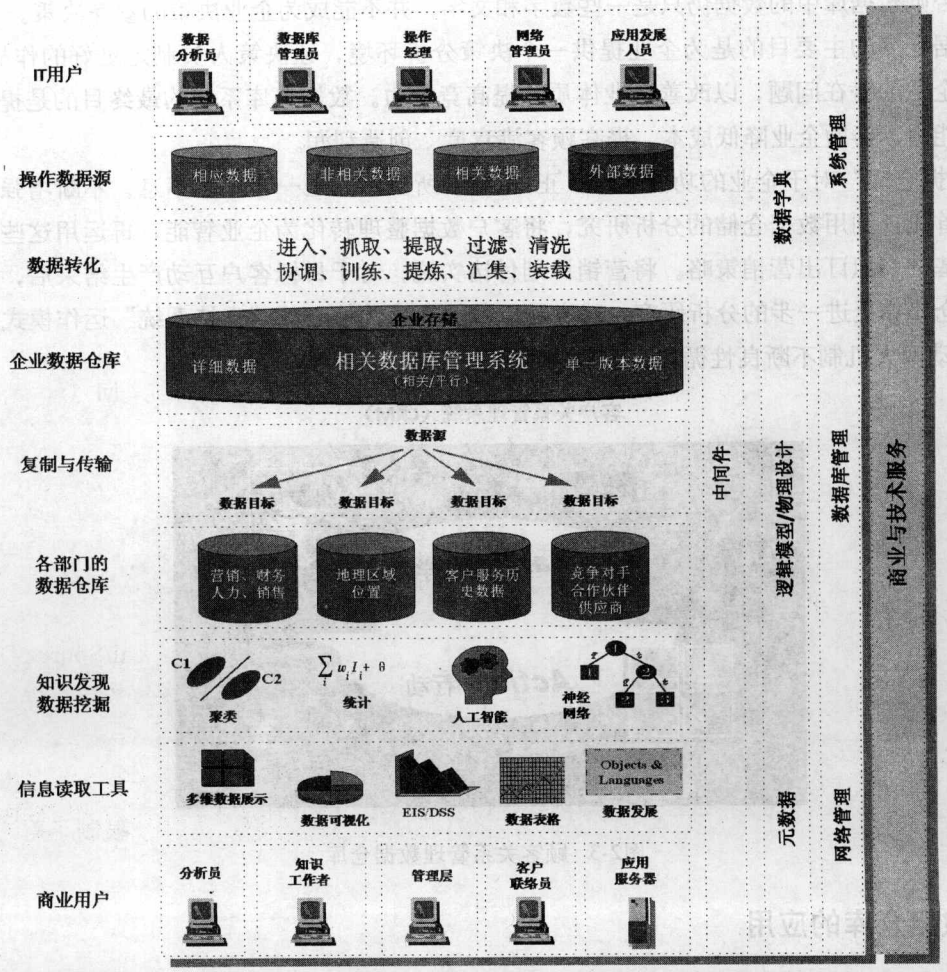


图2-2 数据仓库流程