高等学校教材·计算机应用

多维数据分析原理与应用

姚家奕　等编著

清 华 大 学 出 版 社
北　京

内 容 简 介

数据仓库技术的研究和应用是当前数据管理领域的热点。随着信息化建设在企业、国家政府部门以及社会其他领域的不断普及，数据分析和决策支持系统的开发与建设逐渐纳入到信息化建设的过程中来。面向组织决策层的数据管理技术是决策者挖掘组织内部和组织外部经营管理信息价值的有利工具和方法。

本书从数据仓库理论、多维数据分析技术和多维数据分析应用 3 个方面，以某市地税局数据仓库的成功实施为背景案例，系统地介绍了基于 Microsoft SQL Server 2000 的 OLAP 多维数据引擎——Analysis Services 构建数据仓库多维数据集的全过程，旨在为读者提供从理论到应用的一整套数据仓库 OLAP 解决方案的清晰视图。

本书可以作为信息管理专业和计算机专业本科生或硕士生教材，也可以作为从事数据仓库建设和研究人员的参考书。

图书在版编目（CIP）数据

多维数据分析原理与应用/姚家奕等编著. —北京：清华大学出版社，2004.5
高等学校教材·计算机应用
ISBN 7-302-08377-0

Ⅰ. 多… Ⅱ. 姚… Ⅲ. 数据库系统–系统分析–高等学校–教材 Ⅳ. TP311.13

中国版本图书馆 CIP 数据核字（2004）第 026139 号

出 版 者：清华大学出版社　　地　　址：北京清华大学学研大厦
　　http://www.tup.com.cn　　邮　　编：100084
　　社 总 机：010-62770175　　客户服务：010-62776969
组稿编辑：丁　岭
文稿编辑：闫红梅
印 装 者：北京鑫海金奥胶印有限公司
发 行 者：新华书店总店北京发行所
开　　本：185×260　印张：12　字数：298 千字
版　　次：2004 年 5 月第 1 版　2004 年 5 月第 1 次印刷
书　　号：ISBN 7-302-08377-0/TP·6029
印　　数：1~5000
定　　价：20.00 元

前　言

目前，基于通用数据库技术的联机事务处理（OLTP）系统在大多数现代企业中的应用已经比较成熟，如企业管理信息系统、办公自动化系统、电子商务远程交易系统等。它们为企业信息化迈出了坚实的一步，由此带来的日常业务处理效率的大大提高给企业贴上了信息化的标签。然而，这仅仅是企业信息化的初期。

在 OLTP 系统逐渐走向成熟的进程中，业务高效处理的企业群体日益暴露出来的是决策滞后的新问题。抢先一步的企业已经清楚地认识到：因决策不当造成的人员、设备、资金等的不均衡配置渐渐冲淡了因业务处理效率的提高带来的效益；数百万美元的生意因为你犹豫了一分钟而被决策更快、更准的竞争对手拿走。这些信息技术引发的决策快速问题，常常令企业决策者措手不及。企业运作是一个循环移动的过程，任何一个复杂的企业运作模式都可简化为：业务处理——决策——业务处理。业务处理是执行前次的决策，决策是推进新的业务处理。实际企业中，这个循环链处处存在，小到推销员推销产品，大到整个企业的高利润目标。这个链条的任何一个环节出现问题，都会导致企业的不正常运转。OLTP 系统完善的现代竞争机制下，业务处理迅速而高效，决策频率也大大加快，决策者应接不暇是难免的。

另一个推动因素是决策深度问题。在企业得益于丰富的数据资源的同时，管理者更清楚地认识到，要竞争就决不能满足于简单的数据收集和整理。市场、销售、财务、产品、顾客、服务、供货、存货、销货等方面的数据的内在联系逐渐成为企业各层管理者致力挖掘的焦点。于是，如何从大量的历史数据中有效地提取有用的数据成为各企业所普遍面临的问题。这时，企业管理者已不满足于进行“本月销售额是多少，比上个月增长了多少？”等简单的数据处理，而是要清楚哪些客户会成为我们的利润源泉？哪些产品最有潜力？将会失去哪些市场？哪些地方花费最高？哪些市场运行得最好，为什么？我们进入新市场的胜算有几何，为什么？这些都是辅助决策问题，也是企业要追求的高利润问题。

信息技术引发的问题常常会推动信息技术的前进，正是由于这些困扰企业的实际问题的推动，一种新兴的数据仓库技术诞生了。OLTP 系统积累的大量历史数据为数据仓库提供了数据支持；成熟的通用数据库技术衍生出来的数据仓库专用数据库技术为数据仓库提供了技术背景；对决策质量、效率要求日益提高的现代企业为数据仓库提供了广阔的应用领域。在信息化高度发展的美国企业或企业联合中，大规模的集成数据库系统已经蕴涵着数据仓库思想，只是没有被提出来而已。当决策迟缓渐渐成为企业发展的“瓶颈”时，这种新技术被提上日程，许多专家、学者开始致力于如何应用数据仓库技术辅助企业决策的研究。

本由由北京交通大学经管学院姚家奕负责本书的统稿和定稿，吕希艳负责编写第 1 章和第 2 章，姜海负责编写第 3 章~第 5 章，姚家奕负责编写第 6 章和第 7 章，青岛大学于忠清负责编写第 8 章。

在本书的编写过程中，北京交通大学陈景艳教授对全书的结构和内容提出了许多合理的建议，同时得到多位同行和老师的大力支持，在此一并表示衷心的感谢！书中的“阅读材料”和“案例分析”摘自《计算机世界》和一些网站，在此，也对这些媒体表示感谢！书中内容若有不妥或错误之处，恳请各位读者不吝指正。

本书可以作为信息管理专业和计算机专业本科教材，也可以作为从事数据仓库建设和研究人员的参考书。

目　　录

第 1 章　数据仓库体系结构

数据仓库技术是随着计算机技术的飞速发展而产生的。传统的数据库技术是单一的数据资源，即数据库为中心，进行事务处理、批处理到决策分析等各种类型的数据处理工作。近年来，由于计算机和网络的广泛应用，计算开始向两个不同的方向拓展，一是广度计算，一是深度计算。广度计算的含义是把计算机的应用范围尽量扩大，同时实现广泛的数据交流，互联网就是广度计算的特征；另一方面就是人们对以往计算机的简单数据操作，提出了更高的要求，希望计算机能够更多地参与数据分析与决策的制定等领域。特别是数据库处理可以大致划分为两大类：操作型处理和分析型处理（或信息型处理）。这种分离划清了数据处理的分析型环境与操作型环境之间的界限，从而由原来的以单一数据库为中心的数据环境发展为一种以数据仓库为基础的体系化环境。

1.1　初识数据仓库

按“数据仓库之父”W.H.Inmon 的定义，数据仓库是一个面向主题的、集成的、非易失的且随时间变化的数据集合，用来支持管理人员的决策。以主题为导向的数据仓库是围绕着企业的基本实体设计的，如政府主管行业的城市数据仓库设计主题可考虑城市的人口总数、人均工资水平、失业人数 、税收情况及污染指数等，它们都是内在密切联系的，这样有利于实现数据的关系化 、规则化，且可提供动态 、多维的数据统计 、数据查询，建立关系数据模型，预测发展趋势等，若仅靠现有的在线事务处理系统是无法做到这一点的。数据集成需通过设计实现命名协议、关键字、关系、编码的一致；高度的集成性强调企业必须规划好大量的设计工作，才能真正成功地建立数据仓库。非易失性指数据仓库里的数据不进行实时更新，数据经过复杂的提取过程后定期进入数据仓库。一旦进入，就不能再由用户进行更新。随时间而变化，指数据仓库的设计要按不同时段组织数据，如按月、按季或按年。从历史数据中生成不同级别的汇总数据（或阶段数据），这一点很重要。

1.2　数据仓库解决的问题

数据仓库技术可解决事务处理系统处理不了的决策问题，具有动态集成和综合处理能力。具体能解决如下问题：

(1) 解决“业绩下降 10%”与“业绩上升 15%”的问题。

在线事务处理系统进行数据抽取时，传统上以放任自流的态度处理整个软、硬件体系，这会由于层层抽取的不一致性产生严重的“蜘蛛网”（spider web）问题，导致对同一个问

题不同部门的结论不同，且可能相距甚远，如部门甲认为公司业绩下降 10%，而部门乙可能得出公司业绩上升 15%的结论。数据仓库技术通过给数据加上时基，分离原始数据与导出数据，消除同类数据的算法的差异，提高了数据的可信性，有效避免了“蜘蛛网”问题。

(2) 解决企业环境中多数据源及数据不一致性问题。

数据仓库可以通过数据转移工具将位于不同的地理位置、不同平台、不同数据库中的数据按照一定的规则，高度集中在一个数据仓库中，达到充分利用各种数据源的目的。且在构建数据仓库的过程中，可充分考虑企业原环境数据的不一致性问题，将系统中不一致的数据，按数据的一致性原则转移到数据仓库中，从而保证数据仓库中数据的完全一致，这对作出正确的决策是至关重要的。例如，在电信市场，数据仓库可迅速而准确地向客户提供混合销售、留言、呼叫、等待等综合服务。

(3) 充分而高效地利用企业积累的大量历史数据

在传统在线事务处理系统中，历史数据大多被存储在光盘、磁带或其他大容量存储介质中，查询历史数据是费时、费力的事情，进行数据分析时就更不用说了，况且各年的数据可能存储在不同的介质上，导致数据处理效率较低。数据仓库中主要存储的就是历史数据和大量的汇总数据，因而基于历史数据的分析在数据仓库系统中则显得非常方便，且效率显著提高，因为在数据仓库中存储的就是大量预先处理的汇总数据。

(4) 进行辅助决策分析。

基于数据挖掘、数据抽取和决策支持上发展起来的数据仓库技术，使得决策支持系统进入实用化阶段。一个具有高效的数据分析功能的数据仓库可能会告诉用户（当然，用户事先并不知道它能做到这些）：如何防止丢失有利的客户？查明哪些客户会离开？为什么会离开？提高周末的票价将会带来什么好处？需要多大成本实现一项新服务？2001 年意外伤亡索赔份额为何增长一倍？等等。

1.3 一个成功的例子

本书将以海威决策支持系统（HDC3X）为例，分析数据仓库技术在商业企业信息化中的成功应用。该系统对企业积累的大量业务数据的处理游刃有余，使企业人员能快速、交互、方便有效地从这些大量杂乱无章的数据中获取有意义的信息，决策者能利用现有数据指导企业决策和发掘企业的竞争优势。现简介如下：

系统将现代企业需要的两种不同的数据环境清晰地区分为：日常操作型的数据环境和决策支持信息的数据环境。将企业整个数据仓库体系结构划分为 4 层：操作层、数据仓库层、部门层（或数据集市层）及个体层。

数据仓库体系结构的操作层即指日常操作型的数据环境，如企业的管理信息系统、办公自动化系统等。

数据仓库层建构在决策支持信息的数据环境中，包括以下 3 个主要功能。

(1) 数据抽取：负责从外部数据源抽取数据、区分数据、复制数据或重新定义数据格式等，以备装入数据仓库。

(2) 数据支持：负责数据仓库的内部数据服务，提供的服务包括数据存储的组织、数

据的维护、数据的分发、数据仓库的例行维护等，这些工作利用了数据库管理系统（DBMS）的功能。

(3) 工具平台：是面向不同类型的用户的数据仓库的前端。主要由查询生成工具、多维分析工具和数据挖掘工具等工具集组成，以实现决策支持系统的各种要求。

部门层可以认为是数据仓库的具体应用领域，用以存放大量的导出数据。该系统包括市场、销售、客户、商品、信用风险、消费者、库存及人力资源等多个方面导出数据，体现了系统面向主题的设计方案。

在市场分析方面，能使市场管理人员通过分析新产品的投放情况，不断优化产品定价和促销方式，并结合市场促销渠道和市场投资回报率等影响市场成败的因素，提出全面开拓市场，增加利润和提高影响力的多种解决方案供决策者选择；在销售分析方面，允许销售人员追踪企业的收入进度、比较自己的库存、其他库存和以前的库存，从而对来自市场的反馈，反应更加迅速和准确；在客户分析方面，提供强大的分析工具，能处理大量的产品、顾客的数据细节，对用户进行分类和分析，使理解顾客购买模型成为可能。在此基础上，商家可以提供更有针对性的商品，以最少的花费，满足老顾客和潜在顾客的要求；在商品分析方面，借助对销售、产品分类、价格和购买趋势等分析，制订有效的商品配比计划，并提供图表和统计报告来评价这个计划的成功与否，而且，基于数据仓库技术的海威决策支持系统还可以进行需求预测、商品潜力、地区购买力以及其他各式各样的商业因素分析，例如：地区周消费量、最佳卖点、季节性消费和异常消费等，进而提供一切可能的商机；在信用风险分析方面，该数据仓库解决方案能使企业的信用评测部门从客户交易、现金循环周期、风险得分、非常规购销模型等分析中得出日常经营的风险及即将面临的风险。通过这些信息，管理者能够提前做出准备，作出科学准确的决策。这样，当他们面临这些风险时，便可游刃有余；在消费者分析方面，通过对消费者市场，如：居民生活水平、消费能力、品牌忠诚度、企业形象等的分析，管理者可以更加准确地对未来消费市场的热点、冷点、转向、走势等行情进行预测，然后信心十足地做出决策；在库存分析方面，不仅能有效地控制资产和商品流通量，对比销售与成本的差价，分析销售百分比供其他分析使用，而且通过对库存历史数据的大量即时查询分析挖掘隐性的数据，辅助完成对仓储式销售的最有利的厂址选择、再分配的最佳方案和端到端传递的最优路线等的决策，使库存经营整体协调达到最优；在人力资源分析方面，数据仓库通过对超时和工作量的分析，糟糕的业绩和优秀人员的鉴别，人力满员和人力紧缺的预测等，辅助管理者进行人力资源的最佳配置。

数据仓库体系结构的个体层借助于一个或多个导出数据完成大多数的启发式分析，实际上它更倾向于执行支持，供最高层管理者使用。

基于数据仓库技术的商业企业决策支持系统解决方案，能有效地解决如何管理商业企业中浩如烟海的数据，以及如何从中提取有用的信息等问题，把企业网络中不同信息环境中的商业数据集中起来，生成不同级别的数据，汇总到集成的数据库中，并提供各种数据挖掘、抽取和分析技术，能有效地辅助企业管理者运筹帷幄，决胜千里，实现在商海竞争中不败的雄心壮志。该决策支持系统是数据仓库技术在商业企业中成功应用的一个典型例子。

1.4 数据仓库中心——操作型数据还是分析型数据

近年来，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保持旺盛活力。其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、准确和经过准确定义的，它们源于两种类型的数据：操作型数据和分析型数据。

操作型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、活期存款账户的当前余额或当前脱销产品的数量。这类数据可以表明某件事情的现状，而且它们在任何时候都是处于变化之中的。

另一方面，分析型数据是历史数据，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。某个时刻的销售额是最终数据，不可逆的。在这个时刻，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 月份的总销售额或者开发人员在过去 6 个月里的工资变化。操作型数据与分析型数据的主要区别如表 1-1 所示。

表 1-1 操作型数据与分析型数据的区别

操作型数据	分析型数据
表示业务处理的动态情况	表示业务处理的静态情况
在存取的瞬间是正确的	代表过去的数据
可更新，由录入人员或经过专门培训的输入事务而更新	不可更新，终端用户的访问权限常常是只读的
处理细节问题	受到更多关注的是结论性的数据，是综合的，或是提炼的
操作需求事先可知，系统可按预计的工作量进行优化	操作需求事先不知道，永远不知道下一步用户要做什么
有许多事务，每个事务影响数据的一小部分	有数目不多的一些查询，每个查询可访问大量数据
对性能要求高	对性能要求宽松
面向应用，支持日常操作	面向分析，支持管理需求
用户不必理解数据库，只是输入数据	用户需要理解数据库，以从数据中得出有意义的结论

分析数据通常采自操作数据，可用来对公司在某个时期的运营状况做全面分析。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 2000 Analysis Services 将使用户拥有建立此类数据的能力。用于分析的数据可以包括企业内部数据和企业外购数据。SQL Server 2000 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联机分析处理服务器 Analysis Services 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——可以建立 SQL Server 数据仓库和多维数据中心，从而按需求访问任何信息。

1.5　数据仓库体系结构

数据仓库不是软件产品也不是应用程序，它是系统体系结构。体系结构是按照优先原则对方法进行的安排。这些原则使得通过客户机、网络和数据库软件而执行业务过程成为可能。而这些业务过程又为处理基本业务的规则提供了启示。假如数据仓库市场有各种有用的和有创意的软件产品，而它们都宣称自己是解决方案，那么就有必要考虑一下了。毫无疑问，这些软件产品都有值得注意的优点。尽管许多产品都暗示自己是体系结构，因为没有一种产品是孤立的。但是，关键问题依然存在，即有一种产品本身就是体系结构。

作为体系结构，数据仓库包含了许多产品，如图 1-1 所示。每一种产品都有除数据仓库操作以外的功能。简单地说，数据仓库体系结构是蓝图，是一种安排或示意。这种体系结构提供基础设施，使得传达客户信息的企业应用程序得以实现。数据仓库系统体系结构提供一种模式，在这种模式中，应用程序之间互相紧密连接，而且与硬件、操作系统、数据库、网络及接口软件集成起来，并与业务过程交叉引用。

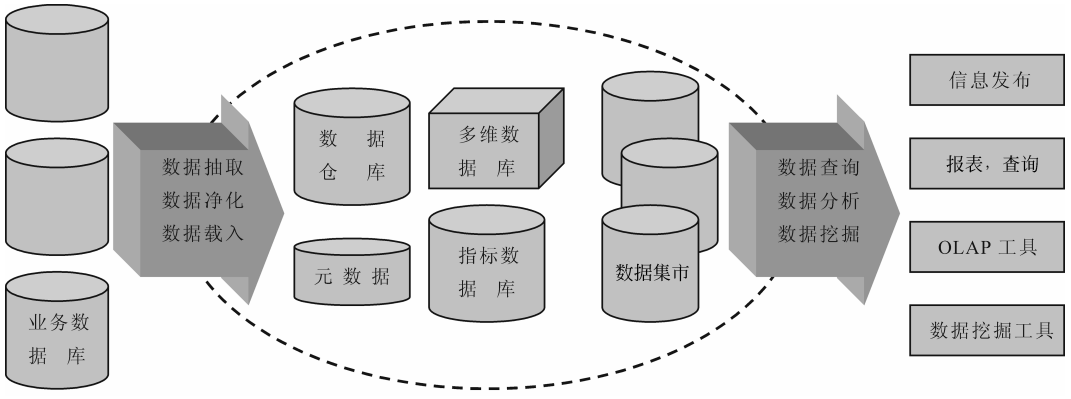

图 1-1　数据仓库的体系结构

1.5.1　数据仓库软件工具集

简单地说，数据仓库软件的工具集包括：一个数据库引擎，这个数据库引擎包括已经

存储起来的程序，其中包含数据和操作逻辑；一个分析应用程序服务器，用于处理复杂的业务过程，例如预测、客户关系管理或推销评估：最终用户界面，通常是客户工作站或桌面的图形用户界面(GUI)；客户机与服务器的连接，通常经由广域网(WAN)连接，并使用基于内部网的 Web 浏览器。为了方便数据仓库长期的生存发展，需要有能够提供导航层及所有重要元数据的软件。该软件必须可以提供用户定义的特别数据查询，并且可以随时进行查询。列出数据仓库的最小构成清单将对设计或购买产品有着积极的意义。但是，构成这个清单的主要目的并不是为了永远使用它，只是为了说明数据仓库的优势不在于这些软件，而在于这些软件表现业务维(并提供有关信息)的方式，例如客户、产品、渠道以及它们在时间和空间(及其他有关结构)的交互。这些维与业务之间联系的方式是承诺之源，是数据仓库的力量之源，这种力量使得数据仓库成为带动企业发展的知识源泉。

1.5.2 体系结构的稳定性

体系结构的本质特性是稳定性，需求的特性是流动。尽管在建设系统、实现以前和移植系统组件时，冻结需求是有益的，然而这种方法只是权益之计。同样，任何体系结构，不管它多么坚固和灵活，扩展的程度仍然是有限的。关键是，与数据仓库系统体系结构相比，最终用户的需求是动态的、波动的。由于这些需求持续地发展、变化与完善。要建立仅仅基于需求的系统，就要冒出错的风险。通常，在企业环境中，不可预见的事件对数据仓库的设计或建设不可避免地会造成外在的打击。如果体系结构很脆弱，它将被动态事件所压倒，使企业的目标丧失，甚至在实现这个目标前它已经过时。另一方面，如果体系结构提供一个灵活的平台，在高度连贯而松散结合的框架下包容许多可能性，那么，它将能够适应不可避免的奇怪事件的冲击，在发展的环境中持续提供服务。

数据仓库体系结构通过交叉引用信息系统与系统建设方法的中心特点来实现。通过这种方法，体系结构在概念和逻辑上的想象得以贯彻实现。例如，基本的结构包括数据、应用功能、连通性、表示(用户界面)、事件(时间序列数据或进度表)和业务驱动力。任何坚固的体系结构都有它自己的实现方法。这个前进过程类似于项目按时间来实现的过程。它从作用域开始，通过抽象的、逻辑的和物理的模型进行处理。从这一点来说，最终使用的技术——系统软/硬件的数据、功能、规则的体现——成为系统建设和配置的制约因素。尽管人机接口产生于客户机/服务器革命的冲击之前，但它们已经更新了这个时代。通信网络(包括 WAN)总是框架的一个必不可少的特性。这就是说，它始终友好地对待以网络为中心的 Internet 计算平台。事实上，人们已经出版了有关该体系结构的书。它是客户机/服务器体系结构的超集，是目前实现的系统最普遍的形式。

1.5.3 维数据结构

数据仓库侧重于维间数据结构的区分。这些维包括客户产品、时间、地点和事实，如卖出、送出或客户在特殊场合使用的产品质量。这种将不同维合成一个有意义的、独一无二的事实结构的连接，即著名的“星型模式”，后面将详细介绍。许多数据仓库功能都需要将事务系统中产生的原始数据转换成能支持决策的形式。其中包括将数据提取成有意义

的、能提供决策支持的聚合：另外还包括数据转换——去除在转换时产生的不一致的或不精确的数据。事务和决策支持系统间的数据转换要求追踪和同步从源到目标系统的语义。数据的收集需要追踪和同步数据摘要一直到数据细节。当运行数十个或上百个程序或系统时，需要具备检索、存储、编目和检索这些系统之间互操作的大量数据的能力。储存库的中心职能类似于图书馆学科中的协调功能。元数据是数据仓库最有挑战性的特性，它可以提高或限制整体数据仓库体系结构的可缩放性、灵活性和可维护性。通常在数据仓库中，人的因素就体现于在桌面上加工信息。这是 OLAP(联机分析处理)的领域。决策支持事件的时间水平与事务系统的截然不同。后者通常着重于 30 天的“ 公开存货清单 ”，而决策支持应用(例如预测)要求 3 到 5 年的数据。加速决策和做出更恰当的决策是配置数据仓库体系结构的重要目标。数据仓库致力于完善结构支持的业务目标，这不同于那些基本的日常商务运作。前者包括策略处理，例如品牌发展，基于客户知识的交叉销售，以及一系列供应链和价值链(逻辑和市场上)的主动性。

1.6　数据仓库体系结构的计算模式

数据仓库体系结构与客户机/服务器计算形式类似。后端进行数据存储，通常使用一个关系数据库；前端是桌面表示工具，用来分隔数据立方体，并聚合从数据存储库返回的数据：中间则是一系列的辅助工具导航、聚合、分析和元数据层，它们本身相当复杂，但这样设计的目的是用来隐藏复杂性，并向最终用户传送一个连贯而一致的界面。考虑在 3270 哑终端的时代，通过在大型机上执行的计算周期，可以实现表示层、应用层、数据库层。这里，企业服务器层的表示、应用和数据库都在后端。另外，表示层可以是中间层服务器的进程，例如，所谓的 X-Windows 系统或者以上 3 个功能都可以利用 OLAP 工具在超级工作站上实现。而且，不同应用功能的分割可以在桌面上的胖客户以及后端数据库上的触发器之间分布，其中胖客户包含了许多确认和应用逻辑以及进程(如存储过程)。

数据仓库应用包括以下一些基本问题，如营销、品牌发展、库存或资产管理、客户的满意程度以及价值链上下的风险承担等。它们使企业分析员、需求计划员、市场管理员、产品开发专家或者知识工人看到对公司非常重要的业务过程的基本定量特征。企业动力同企业本身的目标一样复杂多样。通常，为 OLAP，以便与传统的联机事务处理(OLTP)区别开来。

1.7　以数据为中心

数据仓库必须是以数据为中心的体系结构，也就是说，以数据为中心是与以计算为中心、以实时为中心或以运行控制为中心相对立的。但是，如果考虑最终用户对数据仓库的特别查询，就可以实现以用户为中心的连接，而以前只限于以信息为中心。这不是附加软件，也不是旧式系统的“ 解决方案 ”，因为许多企业系统给数据仓库提供数据，包括电子商务和企业资源规划系统。同样，数据仓库需要从一个或多个老系统以及其他企业系统中

提取、净化和转换数据。这些系统可能是本地的，也可能是远程的。因此，即使采用最先进的数据库管理系统工具和技术来进行数据密集型的处理，数据仓库的模型仍然需要很简单，这样才能发掘各种业务功能中的共同模式。

数据仓库系统向开放式的 Internet 标准和协议方向迁移。有时，这些标准和协议被描绘成“拯救信息技术产业的骑兵”(kmn，1997)，它们促进了客户与供应商之间、沿着原料和信息供应链跨越 WAN 进行的合作。尽管数据仓库内容往往不是那些在客户和供应商之间日常进行共享的敏感数据，但仍然存在一些例外情况，例如，在 Wal-Mart，供应商被邀请（或要求）去访问数据仓库来管理其存货控制和库存补给。

数据仓库信息通常处理时间序列数据——销售、交货或价格如何从一个时间段到另一时间段的变化。这些数据可能非常复杂。它是抽象的，可处理多个维，包括时间、地点(地理位置)、产品以及其他需要考虑的诸多因素。它非常庞大，数据数量非常多。另外，它还包括量的对比。为了管理这种复杂性，就要用到可视化的表示。对于某些人，可视化的表示甚至是必不可少的。通过清晰的图形表示，可以降低复杂性。通过简单的线性设计来表示复杂的时间序列数据，既直观又能使人感到愉悦。这些是桌面图形用户界面开发的重要原则。主要的界面设计原则是使界面与表示方法分开。换句话说，最好的界面是看不见的。当窗口中的小部件、按钮和框在前台出现时，其意义是无法得知的。界面应该像一个侍者一样，需要他时就出现，否则就应当悄无声息。

1.8 数据仓库工作流

数据仓库体系结构使得工作流可以有特殊变化。用三层客户机/服务器的术语来说，它最适于中间层，在那里，分析应用程序引擎实现与决策支持相关的业务过程，例如客户关系管理、预测和产品升级等。工作流作为协作方式，与其作为装配线是不同的。通过数据仓库系统集中在一起的一组人在不同的工作流下合作，而不是在运作事务系统的特性下协作。在信息装配中，重复性的过程自动化了，并通过输入-输出阶梯式的过程来传递。与信息装配线不同的是，数据仓库使工作流成为不同过程之间相互的承诺——一个过程对另一个过程提出请求并得到满足的合作关系。什么是承诺呢？举个例子，根据以前的经验和推销商品的知识，对产品需求的估计暗示了采购和交货产品的承诺。如果产品供应和来源之间相互没有关系，那规划操作就一无所用。数据和产品本质上并不流动。但是，由于采购和交货之间的相互连通性，供应链的交接情况就会得以改善，库存就会减少，风险承担者、雇主和客户的处理工作就可以及时完成。由于数据仓库系统使用的机制，承诺是相当重要的。承诺可作为共享的体系结构，使工作流成为承诺的协作方式。通常，承诺超出了单个企业的范围。人们通常并不认为承诺是体系结构的特性，但实际上它确实是。体系结构包含技术，但并不仅限于此。数据仓库包括市场本身——客户购买行为、交货和相关的服务行为的特性。所有这些都是通过系统体系结构而得以可见和获取的。

数据仓库体系结构支持那些时间范围与日常系统操作不同的过程。这个系统的目的是超出日常运行的范围。目的是减少失败的危险，增加预见未来的选择和商机，增强对业务业绩和运作的控制。认为未来与过去相似的观点是相当有用的假设。许多数据仓库使用 3