



IBM软件学院

商业智能

设计、部署与实现

张云涛 龚 玲
飞思科技产品研发中心

编著
监制



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



IBM软件学院

商业智能

设计、部署与实现

张云涛 龚 玲
飞思科技产品研发中心

编著
监制

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING



内 容 简 介

商业智能利用已有的数据资源进行各种有效的业务决策。它不仅是数据管理、信息处理最活跃的研究领域之一,而且在各行各业中都得到了广泛的应用。本书共分 12 章,全面介绍了商业智能系统的设计、部署、数据处理及系统管理等相关知识,包括数据仓库的架构和模型设计、数据仓库的构建、数据仓库项目的开发和实施、数据仓库的应用、在线分析技术、数据挖掘技术、智能挖掘器、知识管理、数据仓库的运行和管理,以及 OLAP 服务器的运行和管理等技术内容。

此外,本书对商业智能系统所涉及的主流产品技术进行了深入浅出的介绍。通过本书的学习,读者对商业智能的整体结构、技术和产品等将会有深入的了解和认识。

本书可作为商业智能领域的技术人员的参考书,也可作为相关专业的高年级本科生或研究生课程的辅助教材。

未经许可,不得以任何方式复制或抄袭本书的部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

商业智能设计、部署与实现 / 张云涛, 龚玲编著. —北京: 电子工业出版社, 2004.11

(IBM 软件学院)

ISBN 7-121-00473-9

I. 商... II. ①张...②龚... III. 数据库系统 IV. TP311.13

中国版本图书馆 CIP 数据核字 (2004) 第 107324 号

责任编辑: 王树伟 孙伟娟

印 刷: 北京智力达印刷有限公司

出版发行: 电子工业出版社

北京海淀区万寿路 173 信箱 邮编: 100036

经 销: 各地新华书店

开 本: 787×980 1/16 印张: 15.25 字数: 341.6 千字

印 次: 2004 年 11 月第 1 次印刷

印 数: 5000 册 定价: 25.00 元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。
联系电话: 010-68279077。质量投诉请发邮件至 zlt@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

出版说明

随着中国正式加入 WTO，如何提高中国企业的全球竞争力成了政府、企业和媒体关注的焦点。几乎所有的企业都意识到，企业信息化是必由之路，其中，企业电子商务当是信息化建设中的一项重要内容。建立网络环境下的企业管理的新型模式，是电子商务系统实施的“本质”。建立企业管理新型模式的任务，就是要将电子商务系统成为企业信息系统的有机组成部分，这样才能尽享电子商务对时空突破的优势。目前全球电子商务高速发展，预计到 2003 年，其收入将达 32 000 亿美元，将占世界经济总量的 5%。中国企业，无论是大型企业还是中小型企业，越来越意识到，要在激烈的市场竞争中保持优势，企业信息化、管理科学化已成为必然，应用网络实现企业现代化管理，使用电子商务实现企业重组，已经成为诸多企业的中心议题。我们有理由相信，中国企业一定会应用这种方式，提高自身的竞争力，这是企业发展的必然。

长期以来，IBM 在企业应用领域所付出巨大的努力，使得它具有极强的生命力。今天，我们看到大量的政府部门、金融、电信和企业用户使用 IBM DB2 作为数据存储平台，使用 IBM Lotus Domino 作为办公自动化的支撑环境，使用 WebSphere 作为企业电子商务的基础平台等。

适逢政府办公自动化、企业信息化、电子商务在中国深入人心，同时看到全球 IT 产业的迅猛发展态势，作为全球大型企业应用产品提供商，IBM 又将新版 DB2、WebSphere 和 Lotus/Domino 产品推向市场。与 IBM 前期产品一样，它们将受到更加广泛的注意，成为计算机领域一颗耀眼的明星。

今天，我们与 IBM 联合出版“IBM 软件学院”系列丛书，将 IBM 企业应用软件的最新技术，连同多年的研究成果奉献给广大读者。

本套丛书将成为广大数据库开发人员，数据库管理人员，网上应用系统开发人员，电子商务应用开发人员，大专院校计算机、信息、管理的相关专业的广大师生和数据库爱好者的的重要参考读物。

在此，我们临出版之残酷竞争而不惧，旌旗猎猎而异军突起，这与广大读者的支持是分不开的。为使我们的脚步更坚实、使我们的队伍永远保持活力和创造力，我们期待着您能为我们的前进贡献出您的意见和建议。同时，我们也在等待着您的加入。

我们的联系方式：

电 话：(010) 68134545 68131648

电子邮件：support@fecit.com.cn

飞思在线：<http://www.fecit.com.cn> <http://www.fecit.net>

答疑网址：<http://www.fecit.com.cn/question.htm>

通用网址：计算机图书、飞思、飞思教育、飞思科技、FECIT

电子工业出版社计算机研发部

关于飞思

新世纪之初的北京，一群满怀共同理想的年轻人聚集在飞思教育产品研发中心的旗帜下，他们将新的希望和活力注入了中国 IT 教育产品开发领域。飞思人在为把自己打造成为中国 IT 教育产品研发的精英团队而更加不懈努力。

21 世纪的今天，飞思人在多元化教育产品的开发和出版等方面已经迈出了坚实的第一步，开拓出属于自己的一片天空，初步赢来了涓涓细流。

如今，本着教育为科技服务的宗旨，飞思科技产品研发中心以崭新的面貌等待您的支持与关注。

飞思人理念

我们经常感谢生活的慷慨，让我们这些原本并不同源的人得以同本，为了同一个梦想走到一起。

因为身处科技教育前沿，我们深感任重道远；因为伴随知识更新节奏，我们一刻不敢停歇。虽然我们年轻，但我们拥有：

“严谨、高效、协作”的团队精神

全方位、立体化的服务意识

实力雄厚的作者群和开发队伍

当然，最重要的是我们拥有：

恒久不变的理想和永不枯竭的激情和灵感

正因如此，我们敢于宣称：

飞思科技 = 丰富的内容 + 完美的形式

这也是我们共同精心培育的品牌  的承诺。

“问渠哪得清如许，为有源头活水来”。路再远，终需用脚去量；风景再美，终需自然抚育。

年轻的飞思人愿为清风细雨、阳光晨露，滋润您发芽、成长；更甘当坚实的铺路石，为您铺就成功之路。

前 言

在当前全球化竞争日益激烈的经济环境下，一个企业的生死存亡，关键在于它是否能够为用户提供不同的需求做出快速的反应及正确的决策并提供优秀的服务和产品。商业智能的实质是从数据中有效地提取信息，从信息中及时地发现知识，为人类的思维决策和企业战略发展服务。

目前，在很多领域，已成功地引入了商业智能技术，尤其是在如银行、电信、保险、交通、零售等商业领域。商业智能所能解决的典型商业问题包括：直接营销、面向细分市场客户群体划分、用户背景分析、交叉销售等市场分析行为，以及客户流失性分析、客户信用评估、欺诈发现等。

商业智能系统将信息转换为知识。商业智能是在正确的时间将准确的信息交给合适的用户，从而支持决策过程。因此商业智能系统需要大量有用的、精确的数据及合适的分析工具。商业智能系统通常包括数据仓库、仓库管理和分析工具三部分。目前开发的商业智能系统通常是以数据仓库技术为基础，以联机分析处理和数据采掘工具为手段进行实施的一整套解决方案。

本书全面讨论了商业智能领域的相关技术及商业智能解决方案，并详细讨论了商业智能系统的设计、部署与实现的各个方面。全书共分 12 章，全面介绍了商业智能系统的设计、部署、数据处理及系统管理等相关知识，包括数据仓库的架构和模型设计、数据仓库的构建、数据仓库项目的开发和实施、数据仓库的应用、在线分析技术、数据挖掘技术、智能挖掘器、知识管理、数据仓库的运行和管理，以及 OLAP 服务器的运行和管理等技术内容。

本书系统地分析了与商业智能有关的新技术、新进展和新产品，并对商业智能的主流产品和技术，尤其是对 IBM 商业智能解决方案和相关技术进行了论述。

本书可作为商业智能领域的技术人员的参考书，不论是技术负责人、架构师、系统管理员还是开发人员，本书都具有一定的参考价值，另外，也可作为相关专业的高年级本科生或研究生课程的辅助教材。

本书作者多年来从事相关领域的研究和工程实践工作，并在最近两年给研究生讲授了数据挖掘和商业智能课程，本书的出版正是基于这些工作。在作者的研究工作及本书的撰写过程中，许多专家提出了中肯的建议及给予了热诚的帮助，在此作者表示衷心的感谢。

数据挖掘是一个迅速发展的研究领域，并不断有新内容、新方法、新技术等涌现，书中的不当之处在所难免，敬请专家和读者朋友指正。您的任何建议和批评都是我们极为宝贵的财富。作者将努力和广大的读者朋友一起分享您的真知灼见，并在新版本中改进不足之处。

我们的联系地址是：

电 话：(010) 68134545 68131648

电子邮件：support@fecit.com.cn

飞思在线：<http://www.fecit.com.cn> <http://www.fecit.net>

答疑网址：<http://www.fecit.com.cn/question.htm>

源代码下载：<http://www.fecit.com.cn/download.htm>

通用网址：计算机图书、飞思、飞思教育、飞思科技、FECIT

飞思科技产品研发中心

目 录

第 1 章 商业智能导论.....	1
1.1 数据、信息和知识.....	2
1.1.1 知识价值链: 数据.....	3
1.1.2 知识价值链: 信息.....	3
1.1.3 知识价值链: 知识.....	3
1.2 数据决策面临的挑战.....	4
1.2.1 业务和数据的关系.....	4
1.2.2 不同的用户对数据的需求不同.....	5
1.2.3 操作型数据和信息型数据的比较.....	5
1.3 IBM 商业智能.....	6
1.3.1 商业智能系统需求.....	6
1.3.2 DW+DSS=商业智能系统.....	7
1.3.3 IBM 商业智能产品.....	8
第 2 章 数据仓库导论.....	11
2.1 概述.....	12
2.1.1 从数据库到数据仓库... ..	12
2.1.2 数据仓库的功能需求... ..	14
2.2 数据集市.....	15
2.2.1 什么叫数据集市.....	15
2.2.2 数据集市和数据仓库的比较.....	16
2.3 操作型数据存储.....	17
2.4 数据仓库的结构.....	18
2.4.1 数据仓库的总体参考框架.....	18
2.4.2 两层数据仓库.....	19

2.4.3 三层数据仓库.....	19
第 3 章 数据仓库的模型.....	23
3.1 数据仓库建模的基本原则.....	24
3.2 实体——联系模型.....	25
3.3 逻辑模型.....	26
3.3.1 关系模型.....	26
3.3.2 多维模型.....	27
3.3.3 星型模式.....	27
3.3.4 雪花模式和事实星座... ..	28
3.3.5 时态建模.....	29
3.3.6 规范化/反规范化.....	29
3.4 物理模型.....	31
3.4.1 簇集设计.....	31
3.4.2 索引设计.....	32
3.4.3 分区设计.....	33
3.4.4 RAID 配置.....	33
3.5 元数据模型.....	35
3.5.1 元数据的类型.....	35
3.5.2 构造元数据.....	36
3.5.3 元数据在数据仓库中的作用.....	37
3.5.4 元数据的存储、管理与维护.....	38
3.5.5 元数据的使用.....	38
3.5.6 元数据的分布性.....	39
3.6 数据粒度模型.....	39
3.6.1 数据粒度的划分.....	39
3.6.2 确定粒度的级别.....	39
3.7 数据存储.....	40

第4章 数据仓库的构建.....	41
4.1 数据准备.....	42
4.1.1 数据的复制和抽取.....	43
4.1.2 数据转换.....	44
4.1.3 空缺值的处理.....	47
4.1.4 数据标准的不一致.....	48
4.1.5 解决数据质量的关键.....	48
4.2 IBM 数据装载方案.....	50
4.2.1 EXPORT.....	50
4.2.2 IMPORT.....	52
4.2.3 LOAD.....	53
4.3 IBM 复制体系结构.....	56
第5章 数据仓库的开发过程.....	59
5.1 项目组的组成.....	60
5.1.1 项目业务小组.....	60
5.1.2 项目开发小组.....	60
5.2 项目规划.....	61
5.2.1 项目规划简介.....	62
5.2.2 选择数据仓库实现策略.....	62
5.2.3 确定数据仓库的开发目标和实现范围.....	62
5.3 数据仓库项目管理.....	63
5.3.1 项目进度管理.....	64
5.3.2 项目质量管理.....	64
5.3.3 项目风险管理.....	65
5.4 数据仓库的需求分析.....	66
5.4.1 数据仓库的用户.....	66
5.4.2 需求类型.....	66
5.4.3 需求的确定.....	67
5.4.4 需求的分析.....	67
5.5 数据仓库实施.....	68
5.5.1 原型法.....	68

5.5.2 选择数据仓库的实现技术.....	69
5.5.3 设计数据仓库模型.....	71
5.5.4 创建数据准备区.....	71
5.6 数据仓库化.....	72
5.7 规划、设计和建立数据仓库应用.....	74
第6章 数据仓库的应用.....	75
6.1 可视化.....	76
6.2 数据仓库的测试.....	77
6.3 数据仓库的维护.....	80
6.3.1 数据的追加和刷新.....	80
6.3.2 参照完整性维护.....	80
6.4 数据仓库的应用与数据挖掘中的法律问题.....	81
第7章 OLAP 技术.....	83
7.1 概述.....	84
7.1.1 什么是 OLAP.....	84
7.1.2 OLAP 和 OLTP 的比较.....	87
7.2 什么是多维.....	87
7.2.1 维的层次关系.....	87
7.2.2 维的类关系.....	87
7.2.3 多维数据库.....	88
7.2.4 多维视图.....	88
7.2.5 立方体和超立方.....	89
7.3 多维分析.....	89
7.3.1 下钻和上卷.....	89
7.3.2 切片和切块.....	91
7.3.3 旋转.....	91
7.4 OLAP 的实现技术.....	91
7.4.1 MOLAP 和 ROLAP 的比较.....	91

7.4.2	OLAP 的组合结构.....	92
7.5	DB2 OLAP 方案.....	93
7.5.1	DB2 OLAP 的 体系结构.....	93
7.5.2	DB2 OLAP 的 存储结构.....	94
第 8 章	数据挖掘技术.....	97
8.1	概述.....	98
8.2	联机分析处理和数据挖掘的 比较.....	99
8.2.1	验证型分析.....	100
8.2.2	发现型分析.....	100
8.3	数据挖掘技术.....	100
8.3.1	关联分析.....	100
8.3.2	聚类.....	103
8.3.3	分类和预测.....	105
8.3.4	发现序列模式.....	108
8.3.5	统计函数.....	109
第 9 章	IBM 智能挖掘器.....	113
9.1	IBM 智能挖掘器的安装.....	115
9.1.1	AIX 环境下的安装.....	115
9.1.2	Windows 环境下的 安装.....	116
9.2	系统结构.....	122
9.3	图形用户界面.....	123
9.3.1	主窗口.....	123
9.3.2	对象创建向导.....	124
9.4	挖掘处理过程.....	125
9.4.1	标识业务问题.....	125
9.4.2	收集数据.....	125
9.4.3	数据准备.....	133
9.4.4	挖掘数据.....	136
9.4.5	分析结果.....	137

9.5	聚类.....	138
9.6	分类.....	144
9.7	预测值.....	145
第 10 章	知识管理.....	147
10.1	概述.....	148
10.1.1	知识的定义.....	148
10.1.2	知识管理定义.....	148
10.1.3	为什么知识 管理很热.....	149
10.2	知识管理的内容.....	150
10.2.1	目标和需求.....	150
10.2.2	显性知识.....	151
10.2.3	隐性知识.....	151
10.2.4	知识创新.....	151
10.3	知识管理的难点.....	152
10.3.1	为什么共享 很困难.....	152
10.3.2	知识管理的关键.....	153
10.4	知识的生命周期.....	153
10.5	知识获取的方式.....	154
10.6	知识管理技术.....	154
10.6.1	业务环境.....	155
10.6.2	合作技术.....	155
10.6.3	内容管理.....	156
10.6.4	知识管理的 效果评估.....	156
第 11 章	数据仓库的运行和管理.....	157
11.1	DB2 数据仓库中心.....	158
11.1.1	定义数据仓库的 用户和组.....	160
11.1.2	定义数据仓库的 主题.....	162

11.1.3	定义数据仓库的 数据源	163
11.1.4	定义数据仓库目标 ..	168
11.1.5	定义数据的抽取和 转换	171
11.1.6	创建星型模式	174
11.2	DB2 数据仓库管理器	176
11.3	信息目录管理器	177
11.3.1	发布元数据	177
11.3.2	使用元数据	179
11.4	数据维护	181
11.4.1	数据存储的统计 分析	181
11.4.2	数据库重组	182
11.5	备份和恢复	183
11.5.1	数据仓库的备份	183
11.5.2	数据仓库的恢复	184
第 12 章	OLAP 服务器的 运行和管理	193
12.1	Essbase/DB2 OLAP 服务器组件	194
12.1.1	OLAP 引擎	194
12.1.2	DB2 OLAP 分析 服务器	194
12.1.3	应用程序管理器	195
12.1.4	OLAP 集成服务器 ..	196
12.1.5	OLAP 偏差发现 挖掘器	197
12.1.6	Essbase-Ready 工具	197
12.2	DB2 OLAP 服务器的 安装	198

12.2.1	在 Windows NT/2000 上安装	198
12.2.2	在 UNIX 上安装	206
12.3	在 DB2 OLAP 服务器上 配置 Java	208
12.3.1	在 Windows 上配置 Java	208
12.3.2	在 UNIX 上配置 Java	209
12.4	在 DB2 OLAP 服务器上配置 SQL Interface 和数据源	213
12.4.1	在 Windows 上配置 SQL Interface 和数据源 ..	213
12.4.2	在 UNIX 上配置 SQL Interface 和数据源 ..	215
12.5	创建 OLAP 元数据目录	219
12.5.1	自动创建 OLAP 元 数据目录	219
12.5.2	手动创建 OLAP 元 数据目录	220
12.6	手动创建数据库轮廓	221
12.6.1	创建维	222
12.6.2	创建维成员	222
12.7	动态创建数据库轮廓	224
12.7.1	映射机制	224
12.7.2	创建数据加载规则 ..	225
12.7.3	关联轮廓与数据 加载规则	226
12.8	DB2 OLAP 服务器应用 程序工具	228
参考文献		229

1

商业智能导论

- 1.1 数据、信息和知识
- 1.2 数据决策面临的挑战
- 1.3 IBM 商业智能

1.1 数据、信息和知识

商业智能 (Business Intelligence) 简称 BI, 但 BI 是一个容易被误解的缩写。在人工智能等领域的学术著作中, BI 通常指生物智能 (Biological Intelligence)。随着商业智能技术的普及, 尤其是 IBM 在该领域的技术推广, 目前 BI 一般指代商业智能, 这和 IBM 将电子商务的概念由 E-Commerce 拓展为 E-Business 有异曲同工之处。

商业智能尚未有统一的定义, 但商业智能解决方案的领导者 IBM 提出的定义已被大多数人所接受。在 IBM 官方网站上对商业智能的定义如下:

“商业智能是指利用已有的数据资源做出更好的商业决策。它包括数据访问、数据和业务分析, 以及发现新的商业机会”。

商业智能的实质是从数据中有效地提取信息, 从信息中及时地发现知识, 为人类的思维决策和战略发展服务。数据、信息和知识的关系如图 1-1 所示。

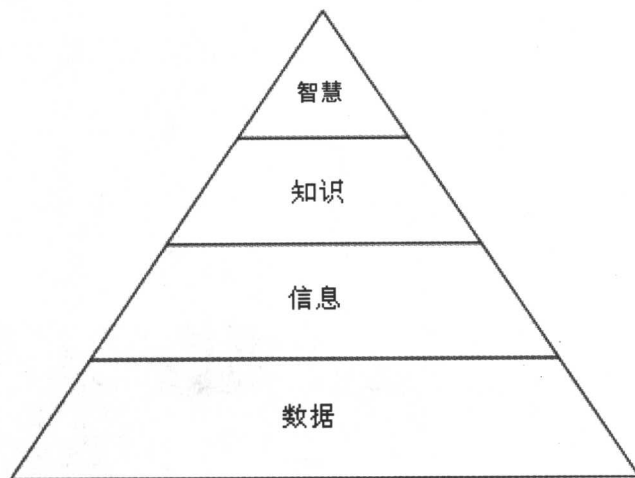


图 1-1 数据、信息和知识间的关系

在数据、信息和知识间的层次结构中, 最低层的是具有潜在意义的数据。对数据进行加工则可得到具有明确意义的信息。知识是人们对可重复信息之间的联系的认识, 它只是信息中已被人们认识了极小的一部分。智慧则是所有知识的顶峰, 智慧指导人们如何获取知识和应用知识。

1.1.1 知识价值链：数据

数据是孤立的、互不关联的客观事实、文字、数据和符号，没有上下文和解释。数据是知识的表现形式，它是客观事物属性、数量、位置及其相互关系等的抽象表示。人们将数据看做形成知识的源泉。

由于数据库等信息技术的广泛应用，在过去的 30 年中，积累了大量的数据，并且数据总量呈现指数级增长之势，然而存储在信息系统中的数据本身——二进制比特并没有直接的价值，数据的价值是通过其所携带的信息表现出来的。

在数据量迅猛增长的同时，如果没有合适的数据处理和数据分析工具，则海量数据不仅不是一笔财富，而且还会带来梦魇，即所谓的“处于数据的海洋中，却缺少知识”。

1.1.2 知识价值链：信息

信息是指人们对数据进行系统地采集、组织、整理和分析的结果，目的是使数据结构化、有序化。如果将宇宙看成是一个封闭系统，则由三部分组成：物质、能量和信息。信息是物质与能量运动的形式，是以物质和能量为载体的客观存在。单独使用数据的可能性很小，一般都是将数据转换成信息的形式以便于使用。有的专家认为信息是数据和上下文的结合，可用于进行决策。信息是数据所标示的含义，信息也是知识的表现形式。因此可以将信息视做对数据的解释。人们认识世界是通过物质和能量运动的形式所反映出来的各种信息实现的。一般信息都可以用一组词及其值来描述。

从理论上说，信息伴随着不确定性。信息的定量表征联系着不确定性的度量。信息将能减少不确定性，换句话说，不确定性减少得越多，所具有的信息也就越多。因此，信息可以帮助人们进行决策。

1.1.3 知识价值链：知识

关于知识，人们有不同的理解和定义，比较有代表性的如：Feign 认为知识是经过削弱、塑造、解释、选择和转换了的信息；Bernstein 定义知识是由特定领域的描述、关系和过程组成的；Hayes-Roth 则认为知识=事实+信念+启发式。有些专家对知识的定义是经验、价值、有条理的信息、专家的见解和本能的直觉的结合体，它可以提供评价和吸收新的经验和信息的环境和架构。一般而言，知识是人们对自然现象的认识和从中总结出来的规律、经验。

虽然关于知识没有统一的定义，但一般将事实、规则、模式、规律和约束等看做知识。事实是指人类对客观事物属性的值或状态的描述，可以用一个值为真的命题陈述或一种状

态的描述来表达。规则可以分为前提条件和结论两部分，用于表示因果关系的知识。如果规则中含有可以实例化为不同具体值的变量，则这种规则称为规律。模式是指符合事物生存运行的内在规律，具有正确的发展导向和行为要求的统一式样、运行机制、管理体制、解决方案的综合，一般指可以作为范本、模本、变本的式样。

广义上，知识指类别特征的概括性描述。根据数据的微观特性发现其表征的、带有普遍性的、较高层次概念的、中观和宏观的知识，反映同类事物的共同性质，是对数据的概括、精炼和抽象。知识是以多种方式把一个或多个信息关联在一起的信息结构。

1.2 数据决策面临的挑战

1.2.1 业务和数据的关系

任何好的决策都需要数据的支持。一个决策的正确程度取决于所使用的数据的正确程度。从大量的数据中经过深层分析，获得有利于商业运作、提高竞争力的信息，从而为决策提供真正有价值的信息。例如对于银行来说，可以根据现有客户信息数据库和客户的消费行为来判定新客户可能的信用程度，减小信用卡恶意透支的可能性。对于保险行业可以对以往的数据进行挖掘从而得到知识，减小潜在的风险。

虽然数据本身并不直接用于决策，然而当数据转化为信息时就可用于决策分析，直接服务于业务，提高企业竞争力。数据、信息和决策间的关系如图 1-2 所示。

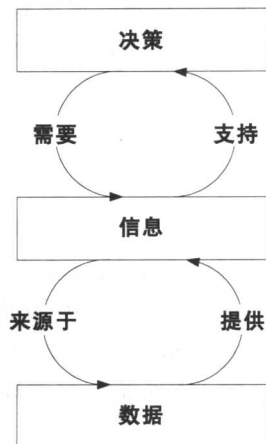


图 1-2 数据、信息与决策间的关系

1.2.2 不同的用户对数据的需求不同

早期的信息系统只能被熟悉数据且有相当的计算机经验的信息提供者使用。企业中的大多数人,诸如管理人员与业务人员却很少能够使用这些系统,他们必须依靠信息提供者来解答他们的问题,并提供给他们所需要的信息。然而,这些用户对数据和信息的需求各不相同。例如,业务操作人员通常需要详细的业务操作数据,而管理人员则往往需要统计和分析数据。这些不同的需求很难完全通过一些固定的报表或查询直接实现。

目前,企业需要能够迅速地建立起信息系统,以便业务用户能够方便、快捷地访问业务信息,这些信息必须能够反映瞬息万变的业务环境。

如今,在IT系统的投资往往占企业支出相当大的比重,因此必须要从IT系统已有的信息中获取最大的收益。企业内的信息系统面向更广泛的用户,诸如管理人员、业务人员等信息消费者。业务中的不同的角色,对信息和信息处理的需求也将有所不同。例如:业务人员需要了解和处理与他们负责的业务紧密相关的数据;需要预定义的报表格式的用户则需要数据精确的报表;业务分析人员则通常需要从不同的角度对数据进行分析;而企业决策人员则往往需要信息来帮助他们进行决策。

用户各种新的需求很难通过老的信息系统满足,只有融合了多种传统的信息技术和新兴的信息技术的商业智能系统才能满足企业内这些用户的新的需求。

1.2.3 操作型数据和信息型数据的比较

通常数据库系统也称做操作型数据库,所谓的操作型数据库是面向细节数据的数据库,收集的大多数数据用于处理企业正在进行的业务,可以满足企业内的复杂的处理流程,这种类型的数据可称为操作型数据。在操作型数据中,数据高度范式化,从而避免了数据冗余。用于处理操作型数据的系统称做OLTP(在线事务处理)。

数据仓库为了分析目的收集、组织数据。用于分析的这类数据称为信息型数据或分析型数据。处理信息型数据的系统称为OLAP(在线分析处理)。

操作型数据与信息型数据的差别主要在以下几个方面:

- 目的
- 内容
- 粒度
- 来源
- 处理
- 更新

➤ 用户

操作型数据和信息型数据的区别首先是其目的不同，操作型数据用于常规的事务处理，而信息型数据则主要用于决策等分析型应用。

其次是数据的内容不同，操作型数据是有关业务目前状态的数据，而信息数据则主要是历史数据。

再次是粒度不同，操作型数据一般是低层的细节数据。而信息型数据则通常是高层的汇总信息。相应地，操作型数据通常是原始数据，而信息型数据则是经过加工导出的数据。

操作型数据通常是业务中的原始数据，而信息型数据则往往是导出或汇总数据。

因为操作型数据通常用于事务处理。在企业中，事务处理的特征是事务数量很多，但每个事务的处理量较小（即涉及的数据量较少）；而 OLAP 则相反，通常是分析请求数量不多，但每个分析请求通常涉及大量的数据，有时甚至需要遍历整个企业的数据。

在操作型数据库上每小时都有大量的事务发生。这种数据库经常是“最新”的，它表示的是当前业务情况的一个快照，经常是一个时间点。而信息型数据总是和时间相关，并且更改后的数据并不会覆盖原先的数据，而是以基于时间的数据的不同版本形式保存下来。

此外，操作型数据主要面向企业中进行操作处理的人员，而信息型数据主要面向企业管理人员和分析人员。分析型数据库通常在一段时间是稳定的，它表示过去特定时间点的情况，可看做为历史数据。

1.3 IBM 商业智能

在当前全球化竞争日益激烈的经济环境下，一个企业的生死存亡，关键在于它是否能够为各种不同的用户的需求做出快速的反应及正确的决策并提供优秀的服务和产品。IBM 商业智能系统提供了一系列技术和产品，并为用户提供了进行业务分析和做出战略决策所需要的信息，从而提高企业的竞争力。

1.3.1 商业智能系统需求

商业智能是一门面向应用的技术。商业智能处理过程首先必须明确需要解决的商业问题。商业智能按企业既定业务目标，对大量的企业数据进行探索和分析，揭示隐藏的、未知的或验证已知的规律性，并进一步将其模型化。分析这些数据也不是单纯为了研究的需要，更主要是为商业决策提供真正有价值的信息，提高企业竞争能力，进而获得利润。传统的决策制定中心是主观制定决策，主观决策存在一些不足。因为数据可以提供比较客观