吉林大学哲学社会科学学术文库

商务智能
概念、方法及在管理中的应用

Business Intelligence
Concept, Method and Application in Management

刘伟江／著

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

吉林大学哲学社会科学学术文库

商务智能
概念、方法及在管理中的应用

Business Intelligence
Concept, Method and Application in Management

刘伟江 / 著

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

图书在版编目(CIP)数据

商务智能：概念、方法及在管理中的应用／刘伟江著 —北京：社会科学文献出版社，2012 1

（吉林大学哲学社会科学学术文库）

ISBN 978－7－5097－2947－2

Ⅰ ①商· Ⅱ ①刘· Ⅲ ①电子商务－研究 Ⅳ ①F713.36

中国版本图书馆 CIP 数据核字（2011）第 253825 号

·吉林大学哲学社会科学学术文库·

商务智能

——概念、方法及在管理中的应用

著　　者／刘伟江

出 版 人／谢寿光

出 版 者／社会科学文献出版社

地　　址／北京市西城区北三环中路甲 29 号院 3 号楼华龙大厦

邮政编码／100029

责任部门／财经与管理图书事业部（010）59367226　　责任编辑／陶　璇

电子信箱／caijingbu@ssap.cn　　责任校对／孙光迈

项目统筹／恽　薇　陶　璇　　责任印制／岳　阳

总 经 销／社会科学文献出版社发行部（010）59367081　59367089

读者服务／读者服务中心（010）59367028

印　　装／三河市文通印刷包装有限公司

开　　本／787mm×1092mm　1/16　　印　张／13

版　　次／2012 年 1 月第 1 版　　字　数／163 千字

印　　次／2012 年 1 月第 1 次印刷

书　　号／ISBN 978－7－5097－2947－2

定　　价／39.00 元

本书如有破损、缺页、装订错误，请与本社读者服务中心联系更换

前　言

随着信息技术的不断发展，人们拥有大量的数据，如何运用这些数据并从中挖掘出有意义的信息和知识，进而将之作为人们决策的依据是很多人关注的重点。基于此，本书力图通过介绍与商务智能相关的内容来达到这一目标。

本书注重理论与实践相结合，以商务智能的三个组成部分——数据仓库、联机分析和数据挖掘为主线，分别介绍了各部分所涉及的相关概念、方法和技术。本书的特点是：

第一，面向管理中的问题，注重实际应用。数据仓库、联机分析和数据挖掘都是应用性很强的技术。本书在对这些方法和技术进行介绍的过程中，力图结合管理中的实际问题，以期为企业的决策提供有价值的参考。

第二，可操作性强。本书在介绍相关技术时，根据网上公开的 Foodmart 2000. mdb 数据集中的数据，在 SQL Server 2005 操作环境下，作了丰富的图例和操作讲解，读者可以模拟本书的操作步骤，并把这些技术方法应用到自己需要处理的问题中。

本书主要分为四部分内容。

第一部分为概述，是第 1 章的内容，这章对商务智能的概念、发展等作了相应的介绍。

第二部分为数据仓库和联机分析，包含第 2、3、4 章的内容，

在这部分中主要对数据仓库的建立、数据清理、数据查询等做了说明。

第三部分为数据挖掘，包含第5、6、7章的内容，这部分主要对数据挖掘方法进行了详细的介绍，涉及决策树、神经网络、关联规则、聚类等方法。

第四部分为商务智能在管理中的应用，主要是第8章的内容。在这部分中，根据商业企业在实际中可能遇到的商业问题，通过选择一定的数据挖掘方法来进行处理，并对得到的数据结果进行了分析说明。

由于作者的水平和经验有限，有些方面的研究还有待深入和提高，难免有错讹之处，真诚欢迎广大读者批评指正。

刘伟江

2011年9月于长春

目录
CONTENTS

目录
CONTENTS

第1章 概述

1.1 商务智能简介

1.1.1 商务智能概念

“啤酒与尿布”的故事最初是美国学者 Agrawal 从沃尔玛商店中大量的顾客消费数据中，利用计算机通过商品关联关系的计算方法——Aprior 算法[1]分析得到的一种现象：在某些特定的情况下，啤酒和尿布这两种看起来完全不相干的商品会出现在一个购物篮中，这一现象的发现有助于商家安排商品的合理摆放和选择相应的促销活动。沃尔玛随后对啤酒和尿布进行了捆绑销售，不出意料，两种商品的销售量双双增加。这个有趣的故事带给了人们深深的思考：如何利用信息技术收集数据并采取有效的方法处理这些大量的、有噪声的、不完全的数据，进而从中找出未知的、有用的信息和知识来帮助人们作决策？因此，商务智能（Business Intelligence，简称 BI）作为一种能帮助人们将大量数据转化成有价值的信息和知识的有效过程越来越引起人们的广泛重视，很多公司和个人都根据自己的理解给出了商务智能的定义。

IBM 公司对商务智能的定义是：商务智能是指利用已有的数据资源作出更好的商业决策，它包括数据访问、数据和业务分析及发现新的商业机会。该定义认为商务智能可以从根本上帮助企业把其运营数据转化为高价值的可以获取的知识（或信息），并且在恰当的时候通过恰当的方式把恰当的信息传给恰当的人。Gartner Group 的 Howard Dresner 在 1989 年将商务智能定义为一类由数据仓库（或数据集市）、报表查询、联机分析、数据挖掘等部分组成，以帮助企业决策的技术及应用。该定义的基本层次结构如图 1－1 所示。

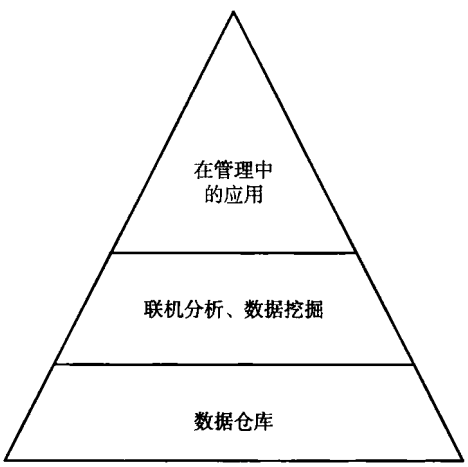

图 1－1　商务智能的基本层次结构

从图 1－1 可以看出，数据仓库中的数据是商务智能的基础，一般来说，数据分为三类：第一类是结构化数据，如数字、符号等，这类数据可以存储在关系型数据库里，用二维逻辑表来表现；第二类是半结构化数据，如文本、网页等；第三类是非结构化数据，如图像、视频、音频等。联机分析、数据挖掘等是商务智能的核心，其中联机分析（Online Analysis Process，简称 OLAP）技术是用来帮助分析人员、管理人员从多种角度把原始数据中转化

出来的、能够真正为用户所理解的，并真实反映数据维特性的信息进行快速、一致、交互的访问，从而让用户可以从各个角度更深入地了解数据的一类软件技术；数据挖掘（Data Mining，简称DM）是与数据仓库密切相关的一种信息技术，近几年来，很多专家和学者从不同的角度给出了数据挖掘的定义，如从技术角度，数据挖掘（Data Mining）就是从大量的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的，但又是潜在有用的信息和知识的过程。该定义包括好几层含义：数据源中的数据是真实的、大量的、含噪声的；发现的内容是用户感兴趣的知识，这些知识是可接受的、可理解的和可运用的；在知识发现过程中，并不要求发现放之四海皆准的知识，这些知识仅支持特定问题的处理就可以。从商业角度来讲，数据挖掘是一种新的商业信息处理技术，其主要特点是对商业数据库中的大量操作型数据进行抽取、转换、分析和其他模型化处理，从中提取辅助商业决策的关键性信息。总之，数据挖掘就是要从大量数据中提取出隐藏在其中的有用信息。

正因为如此，近年来商务智能在管理中有着极其广泛的应用，商务智能的应用问题主要从以下几个方面进行考虑：一是要解决什么样的问题；二是围绕要解决的问题进行相关数据的收集；三是选取合适的数据挖掘方法对数据进行分析；四是对分析的结果有个合理的说明并提出切实可行的建议。

1.1.2 商务智能的发展

商务智能的发展主要经历了以下几个阶段[2]：

1. 萌芽期

1947年卡内基梅隆大学的赫伯特·西蒙（Herbert Simon）教授在《行政组织的决策过程》一书中提出：如果能利用存贮

在计算机里的信息来辅助决策，人类理性的范围将会大大扩大。他的这个观点奠定了决策支持系统的基础，而他对决策支持系统的研究被学术界公认为现代商务智能概念最早的源头和起点。

2. 发展期

随着决策支持系统的发展，如何把多个不同数据源的数据有机地整合起来是决策支持系统面临的难题。1988 年，为解决企业中的数据集成问题，IBM 公司的研究员 Barry Devlin 和 Paul Murphy 创造性地提出了一个新的术语：数据仓库（Data Warehouse）。但 IBM 公司只是把数据仓库这个词当做一个新的概念来宣传，而没有进一步提出实际的架构和设计。1992 年，比尔·恩门（Bill Inmon）在《如何构建数据仓库》一书中首次给出了关于数据仓库的清晰定义和操作性极强的建议，从而为数据仓库的广泛应用奠定了基础，也为商务智能的发展提供了支撑。随后的联机分析技术的发展使得数据仓库有了更广泛的用武之地，它使人们可以从多个视角、多个维度观察数据，从而清楚地了解事情的发生过程。

3. 成熟期

随着数据仓库、联机分析技术的发展和成熟，商务智能的框架基本形成，但真正给商务智能带来活力的是数据挖掘技术的出现。数据挖掘技术以其能通过对大量数据的分析来揭示数据之间隐藏的、有意义的关系、模式和趋势，为决策者提供新的信息和知识的能力，使得商务智能真正有了“智能”的内涵。

进入 21 世纪以来，随着技术的发展，信息可视化使得商务智能的应用又有了更进一步的发展。所谓信息可视化（Information Visualization）是指以图形、图像、动画等更为生动、易于理解的方式来展现和诠释数据之间的复杂关系和发展趋势，以