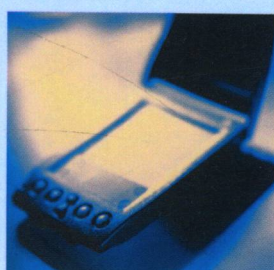


商业智能 理论与应用实践



苏力萍 主编



中国科学技术出版社
CHINA SCIENCE AND TECHNOLOGY PRESS

商业智能理论与应用实践

苏力萍 主编

中国科学技术出版社

· 北 京 ·

图书在版编目 (CIP) 数据

商业智能理论与应用实践 / 苏力萍主编. —北京: 中国科学技术出版社, 2012. 6

ISBN 978 - 7 - 5046 - 6112 - 8

I. ①商… II. ①苏… III. ①电子商务 IV. ①F713.36

中国版本图书馆 CIP 数据核字 (2012) 第 131603 号

责任编辑 吕秀齐

封面设计 李 丽

责任校对 韩 玲

责任印制 张建农

出 版 中国科学技术出版社

发 行 科学普及出版社发行部

地 址 北京市海淀区中关村南大街 16 号

邮 编 100081

发行电话 010 - 62173865

传 真 010 - 62179148

投稿电话 010 - 62176522

网 址 <http://www.cspbooks.com.cn>

开 本 787mm × 1092mm 1/16

字 数 300 千字

印 张 20

版 次 2012 年 8 月第 1 版

印 次 2012 年 8 月第 1 次印刷

印 刷 北京长宁印刷有限公司

书 号 ISBN 978 - 5046 - 6112 - 8/F · 748

定 价 38.00 元

(凡购买本社图书, 如有缺页、倒页、脱页者, 本社发行部负责调换)

本社图书贴有防伪标志, 未贴为盗版

编辑委员会

主 编 苏力萍

副 主 编 雷雪梅 董 平 陶丽红

参编人员 苏力萍 雷雪梅 董 平 陶丽红

白 莉 杨 勇 李水珍 夏 晖

胥 杰 阮进喜 王 蕾 薛 婷

张鹏亮 刘鹏春

前 言

商业智能（Business Intelligence，BI）的概念最早是在 1996 年由 Gartner Group 提出来的。它将商业智能定义为一类由数据仓库（或数据集市）、查询报表、数据分析、数据挖掘、数据备份和恢复等部分组成，以帮助企业决策为目的的技术及其应用。

商业智能描述了一系列的概念和方法，通过应用基于事实的支持系统来辅助商业决策的制定。换句话说，商业智能就是能够帮助企业对业务经营智能地做出正确决定的一系列工具及其使用策略。商业智能的核心是同任何其他管理软件一样的，关键都是一种管理思想。从技术层面上讲，商业智能也不是什么新技术，它是将数据仓库（DW）、联机分析处理（OLAP）、数据挖掘（DM）等技术与资源管理系统（ERP）结合起来，应用于商业活动实际过程当中，最终达到技术服务于决策的目的。

其实，商业智能一直存在于企业的日常工作当中。比如对数据的简单整理，对报表的分析，通过这些分析做出未来若干时间内的工作规划等，这些都是商业智能的表现。随着企业信息化的发展，在应用 ERP 过程中，大量的数据积累，大量的信息涌现，造成了企业对 ERP 数据信息的困惑，从而引发了企业对于专业商业智能软件产品的需求。商业智能不再仅仅是一种概念、一种技术，它更多地成为一种业务层面的需求，为企业应用服务。

本书是作者长期从事数据仓库、数据挖掘与数据展现的研究工作，在数据处理各个阶段进行了探索和尝试，并在此基础上构建了不同的商业智能应用系统。书中实践部分的案例全部来源于本人的实际研究工作，具有很好的实践指导意义。通过对商业智能系统的建设实践，不断总结经验，总结出能够对商业智能项目前期规划和项目实施进行指导的综合知识是本书最显著的特点。

本书共分三大部分，分别从商业智能基础知识、核心技术和应用实践三个方面对商业智能的概念、相关理论知识、构建实际应用系统的过程进行了

阐述。

第一部分是商业智能基础知识介绍：

第1章商业智能概述，介绍了商业智能的概念、研究内容，并阐述了商业智能与知识管理、数据质量的关系。第2章对商业智能的应用现状进行了描述，分别从商业智能与企业商业决策、商业智能应用类型、应用领域及其扩展，以及商业智能的发展趋势等方面进行阐述。

第二部分是商业智能核心技术介绍：

第3章数据仓库、数据挖掘，对数据仓库的基本知识，数据提取、转换和加载过程、数据展现、多维分析与展现、数据挖掘与预测等数据仓库相关理论知识加以介绍。第4章元数据，介绍了元数据的基本概念、元数据收集过程、元模型以及元数据管理体系。第5章地理信息系统，对近年来快速发展的GIS（地理信息系统）作为一种新型的数据展现形式的相关知识加以介绍，并对其发展趋势进行了探讨。第6章分布式系统整合，对分布式、异构数据源数据进行整合的架构加以介绍，以及WEB服务的相关知识的描述。第7章系统安全，从商业智能系统的安全弱点出发，介绍了构建BI系统安全防护时应考虑的诸多因素及方法。第8章系统技术架构，讲述了构建BI系统的过程，以及各步骤应遵循的架构。

第三部分是商业智能应用实践介绍：

第9章保险项目实例建设，以构建一个保险公司BI项目的应用为实例，阐述了构建BI应用实例的设计和实现过程。第10章BI项目中的GIS应用，描述了以GIS方式展示BI系统的实际应用的构建过程。第11章分布式系统的数据整合，以多个实际应用为基础，阐述了对于分布式、异构数据源为数据来源的BI系统的构建架构和过程。

本书的出版得到了中国科学技术出版社的大力支持，在此表示衷心的感谢。

在信息量巨大，新技术发展变化迅速的今天，新的理念和技术不断涌现，时刻影响着BI系统的应用模式和发展进程。尽管本人已付出艰苦的努力，但由于水平有限，疏漏之处在所难免，恳请读者批评指正。另外，本人也在商业智能移动化方面做了些尝试，但由于案例还不够完善，无法在本书中进行展开讨论。愿借本书抛砖引玉，与各位同行交流、学习。来信请寄 lsu@ustb.edu.cn。

作者

2012年5月10日

目 录

第一部分 商业智能基础

第 1 章 商业智能概论	2
1.1 商业智能的概念	2
1.2 商业智能的研究内容	4
1.3 商业智能与知识管理	6
1.4 商业智能与数据质量	12
第 2 章 商业智能的应用	15
2.1 企业的困惑	15
2.2 商业决策与商业智能	16
2.3 智能型企业	21
2.4 商业智能应用类型	23
2.5 商业智能领域扩展	26
2.6 商业智能发展	37

第二部分 商业智能核心技术

第 3 章 数据仓库与数据挖掘	46
3.1 数据仓库的概念及特征	46
3.2 数据仓库的体系结构	47
3.3 数据提取、转换和加载	48

3.4	多维分析与数据展现	54
3.5	数据挖掘与数据预测	56
第4章	元数据	61
4.1	元数据概念	61
4.2	元数据收集过程	66
4.3	CWM 元模型	71
4.4	元数据管理体系	73
第5章	地理信息技术	75
5.1	WebGIS 的概念与特点	75
5.2	WebGIS 实现方法与比较	76
5.3	WebGIS 实现模型	78
5.4	WebGIS 发展趋势	79
第6章	分布式系统整合	82
6.1	分布式数据库	82
6.2	异构数据源	85
6.3	面向服务的体系架构 (SOA)	90
6.4	Web 服务 (Web-service)	96
第7章	系统安全	103
7.1	BI 系统的安全弱点	103
7.2	BI 系统安全防护	105
7.3	安全计划	111
第8章	系统技术架构	112
8.1	Kimball 生命周期过程	112
8.2	技术架构简介	116
8.3	数据架构	119
8.4	应用程序架构	126
8.5	底层基础设施	139

8.6	元数据	142
8.7	安全性	145
8.8	架构开发过程	145
8.9	创建底层设施图	147

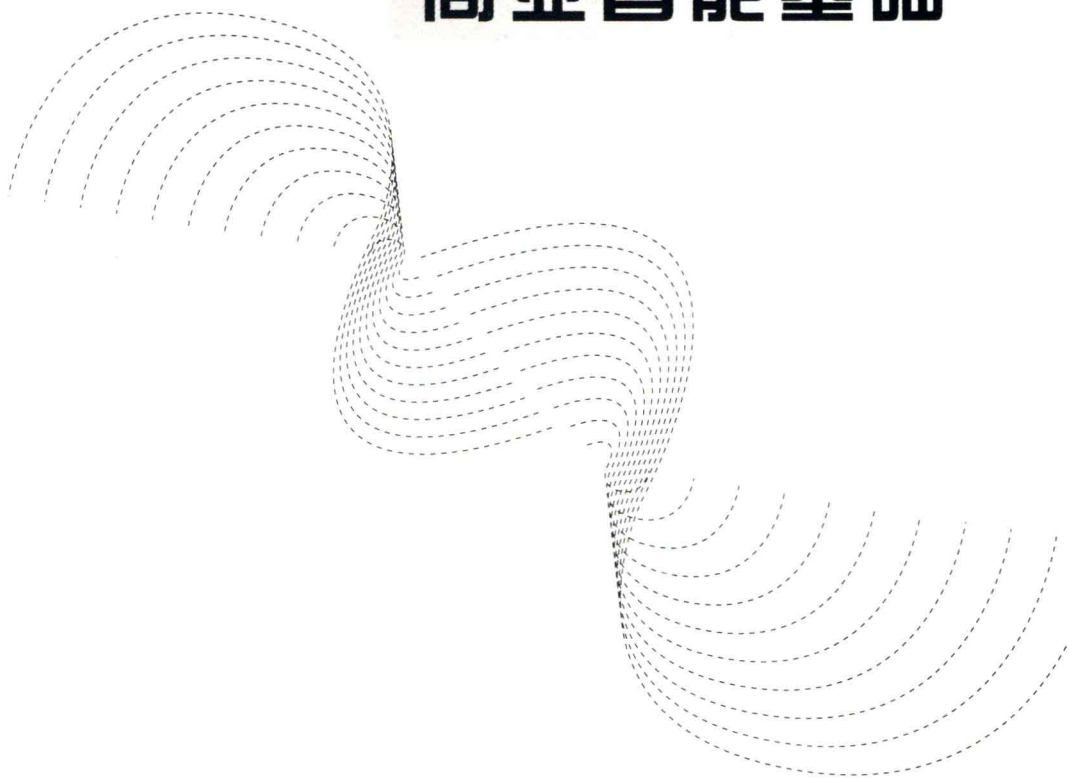
第三部分 商业智能应用实践

第9章	保险 BI 项目建设示例	152
9.1	项目启动	153
9.2	需求调查与分析	155
9.3	BI 架构设计	161
9.4	元数据管理	163
9.5	产品选择	165
9.6	维度模型设计和开发	166
9.7	数据库设计	176
9.8	应用开发	186
9.9	BI 应用程序开发	188
第10章	BI 项目中的 GIS 应用	192
10.1	系统需求	192
10.2	系统框架设计	193
10.3	功能模块	194
10.4	数据准备与 ETL	201
10.5	数据仓库设计和实现	206
10.6	应用系统的设计与实现	209
10.7	基于 GIS 的 BI 数据展现示例	219
第11章	分布式系统的数据整合	224
11.1	项目简介	224

11.2	主要问题分析与解决思路	225
11.3	解决方案	228
11.4	分布式异构数据源的集成	236
11.5	目标数据仓库的设计与实现	238
11.6	学生子系统的数据集成	241
11.7	资金监控子系统	258
11.8	工程管理子系统	263
11.9	决策支持平台的实现	282
主要参考文献及网址		300
缩略语解释		307

第一部分

商业智能基础



第 1 章 商业智能概论

1.1 商业智能的概念

1.1.1 商业智能的定义

商业智能，又称商务智能，英文为 Business Intelligence，简称为 BI。商业智能不是一个新名词。商业智能的概念最早由 Gartner Group 于 1996 年提出，它将商业智能定义为一类由数据仓库（或数据集市）、查询报表、数据分析、数据挖掘、数据备份和恢复等部分组成，以帮助企业决策为目的的技术及其应用。商业智能描述了一系列的概念和方法，通过应用基于事实的支持系统来辅助商业决策的制定。早在 20 世纪 80 年代，“商业智能”的标准是能比较容易地获得想要的数据和信息。20 世纪 90 年代是商业智能真正起步的阶段。到目前为止，关于 BI 还没有统一的定义，不同的人只是从不同的方面表达了对 BI 的理解。在 90 年代初，Gartner Group 的 Howard Dresner 把 EUQR（终端查询和报表）、DSS、OLAP 称为商业智能。企业使用这些工具使企业获得的优势也被称为商业智能。后来，出现了数据仓库、数据集市技术，以及与之相关的 ETL（抽取，转换，加载）、数据清洗、数据挖掘、商业建模等，人们也将这些技术统归为商业智能的领域。目前，存在将商业智能与数据仓库和基于数据仓库的分析方法等同起来的认识趋势。

换句话说来说，商业智能就是能够帮助企业对业务经营智能地做出正确决定的工具。商业智能的核心是同任何其他管理软件一样的，关键都是一种管理思想。从技术层面上讲，商业智能也不是什么新技术，它只是数据仓库、OLAP 和数据挖掘等技术的综合运用。

对商业智能定义，也可以从以下四个层面来解释：

（1）信息系统层面。信息系统层面是商业智能系统（BI system）的物理基础，表现为具有强大决策分析功能的单独的软件工具和面向特定应用领域的信息系统平台，如 SCM、CRM、ERP。与事务型的 MIS 不同，商业智能系统能提供分析、趋势预测等决策分析功能。

(2) 数据分析层面。数据分析层面是系列算法、工具或模型。数据分析层面首先获取与所关心主题有关的高质量的数据或信息,然后自动或人工参与使用具有分析功能的算法、工具或模型,帮助人们分析信息、得出结论、形成假设、验证假设。

(3) 知识发现层面。与数据分析层面一样,知识发现层面也包含系列算法、工具或模型。知识发现层面将数据转变成信息,而后通过发现,将信息转变成知识;或者直接将信息转变成知识。

(4) 战略层面。战略层面将信息或知识应用在提高决策能力和运营能力上。商业智能的战略层面是利用多个数据源的信息以及应用经验和假设来提高企业决策能力的一组概念、方法和过程的集合。它通过对数据的获取、管理和分析,为贯穿企业组织的各种人员提供信息,以提高企业战略决策和战术决策能力。

1.1.2 商业智能是解决方案

目前,很多厂商活跃在商业智能领域,提供了很多BI产品。商业智能帮助企业和管理层迅速地发现企业中的问题,提示管理人员加以解决。但事实上,商业智能软件系统不能代替管理人员进行决策,不能自动处理企业运行过程中遇到的问题。商业智能为企业带来的是一种科学的管理思想,给整个企业带来的是决策的快速性和准确性,发现问题的及时性,以及发现那些对手未发现的潜在的知识和规律,而这些信息是企业产生经济效益的基础。比如:通过对销售数据的分析可发现各类客户的特征和喜欢购买商品之间的联系,这样就可进行更有针对性的精确的促销活动或向客户提供更具有个性的服务等,这都会为企业带来直接的经济效益。

目前,学术界对商业智能的定义并不统一。商业智能通常被理解为将企业中现有的数据转化为知识,帮助企业做出明智的业务经营决策的工具。这里所谈的数据包括来自企业业务系统的订单、库存、交易账目、客户和供应商等来自企业所处行业和竞争对手的数据以及来自企业所处的其他外部环境中的各种数据。而商业智能能够辅助的业务经营决策,既可以是操作层的,也可以是战术层和战略层的。为了将数据转化为知识,需要利用数据仓库、联机分析处理(OLAP)工具和数据挖掘等技术。正如前面所说,商业智能不是什么新技术,它只是数据仓库、OLAP和数据挖掘等技术的综合运用。

可以认为,商业智能是对商业信息的搜集、管理和分析过程,目的是使企业的各级决策者获得知识或洞察力(insight),促使他们做出对企业更有利的决策。商业智能一般由数据仓库、联机分析处理、数据挖掘、数据备份和

恢复等部分组成。商业智能的实现涉及到软件、硬件、咨询服务及应用，其基本体系结构包括数据仓库、联机分析处理和数据挖掘三个部分。

因此，把商业智能看成是一种解决方案应该比较恰当。商业智能的关键是从许多来自不同的企业运作系统的数据中提取出有用的数据并进行清理，以保证数据的正确性，然后经过抽取（Extraction）、转换（Transformation）和装载（Load），即 ETL 过程，合并到一个企业级的数据仓库里，从而得到企业数据的一个全局视图，在此基础上利用合适的查询和分析工具、数据挖掘工具、OLAP 工具等对其进行分析和处理（这时信息变为辅助决策的知识），最后将知识呈现给管理者，为管理者的决策过程提供支持。

1.1.3 商业智能系统主要功能

商业智能功能非常强大，对数据的处理有独特之处。主要功能包括：

（1）数据管理功能。从多个数据源抽取、转换和加载数据以及清理、集成数据的能力，大量数据高效存储与维护的能力。

（2）数据分析功能。具备多种数据分析功能，同时具有终端信息查询及报表生成功能和数据可视化功能。

（3）知识发现功能。从大型数据库数据中提取人们感兴趣的的知识的能力。这些知识以是隐含的、事先未知的、潜在有用的信息，提取的知识表示为概念、规则、规律、模式等形式。

总之，商业智能的目标是将企业所掌握的信息转换成竞争优势，提高企业决策能力、决策效率、决策准确性。为完成这一目标，商业智能必须具有从实现数据分析到知识发现的算法、模型和过程，决策的主题具有广泛的普遍性。

1.2 商业智能的研究内容

商业智能为更好地制订战略和决策提供良好的环境，为特定的应用系统（如客户关系管理 CRM、供应链管理 SCM、企业资源计划 ERP）提供数据环境和决策分析支持。当面向特定应用的特定战略和决策问题时，商业智能从数据准备做起，首先建立或虚拟一个集成的数据环境，然后在集成的数据环境中，利用科学的决策分析工具，通过数据分析、知识发现等过程，为制订战略和决策提供支持。

数据仓库提供数据存储环境，而且是面向特定主题的决策支持环境。来自各种数据源中的数据经过清理、抽取、转换和加载，按照某一主题存储。

数据集市是面向特定主题的小型数据仓库，解决了企业级数据仓库存储大量数据而带来的建设周期长、造价高、可扩展性差等缺陷。

商业智能是利用当今计算机前沿技术作支撑、运用现代管理技术进行指导的应用系统，它的研究热点集中在三个方面：支撑技术的研究、体系结构的研究、应用系统的研究。

1.2.1 支撑技术的研究

商业智能作为一个在 20 世纪 90 年代末期出现的跨学科新兴领域，必须借鉴两方面的先进成果，一是计算机技术的前沿技术，一是企业管理方面的新理论、新观点。企业管理方面的新理论、新观点为战略制订和决策提供先进的管理模式，帮助企业更好地运营；先进的计算机技术是提高系统性能的有力手段。

商业智能的技术主要包括以下几项：一是计算机技术，包括：数据仓库、数据集市技术；数据挖掘技术；OLTP、OLAP 等分析技术；数据可视化技术；计算机网络与 WEB 技术。二是企业管理，包括：统计、预测等运筹学方法；客户管理、供应链管理、企业资源计划等管理理论和方法；企业建模方法。

支撑技术的研究主要围绕两部分展开：决策支持工具研究和企业建模方法研究。企业建模是为解决如何建立特定企业模式的辅助工具。IDEF 等研究方法是较程式化的企业建模方法，比较新的建模方法包括基于 UML 的企业建模等方法。

数据挖掘算法的研究是目前计算机界研究的热点之一，它逐渐成为一个跨越人工智能、数据统计等多学科的研究领域。决策分析工具的研究还包括各种分析方法的研究。

1.2.2 体系结构的研究

商业智能系统主要包括数据预处理、建立数据仓库、数据分析及数据展现四个主要阶段。数据预处理是整合企业原始数据的第一步，它包括数据的抽取、转换和装载三个过程。建立数据仓库则是处理海量数据的基础。数据分析是体现系统智能的关键，一般采用联机分析处理和数据挖掘两大技术。联机分析处理不仅进行数据汇总/聚集，同时还提供切片、切块、下钻、上卷和旋转等数据分析功能，用户可以方便地对海量数据进行多维分析。数据挖掘的目标则是挖掘数据背后隐藏的知识，通过关联分析、聚类和分类等方法建立分析模型，预测企业未来发展趋势和将要面临的问题。在海量数据和分析手段增多的情况下，数据展现则主要保障系统分析结果的可视化。一般认

为数据仓库、OLAP 和数据挖掘技术是商业智能的三大组成部分。

面向特定应用会有相应改进的体系结构,使商业具有良好的性能,例如:建立数据存储和数据模型能很好地支持数据分析和知识发现的需要;选择何种决策分析工具,包括选择实现何种任务、选择实现这种任务的何种工具;将分析和发现的信息和知识通过何种接口达到需要的用户等等。

1.2.3 应用系统的研究

对应用系统的研究重点在于对各个应用领域所面临的决策问题的分析。根据对各类问题的解决方式和解决方案来决定商业智能系统应该提供的功能以及具体实现方法。目前,商业智能被广泛应用于与企业运营过程相关的各个领域,并且在很多领域已经形成其特有体系。目前具有代表性的应用领域包括:企业资源计划(ERP)、客户关系管理(CRM)、企业性能管理(BPM)、人力资源管理(HRM)、供应链管理(SCM)、电子商务(E-business)。

1.3 商业智能与知识管理

1.3.1 数据、信息、知识与智慧

1. 数据与信息

数据是记录客观事物的、可鉴别的符号,也是人们测量所得到的最原始结果,比如高度、时间、观点等等。数据具有客观性。

信息是关于客观事实的可通信的知识,是指有新内容、新知识的消息,是经过加工以后、对客观世界产生影响的数据,是对数据全面的综合所得的结果。比如表格和统计结果等等。信息具有事实性、时效性、不相关性、等级性。

信息与数据既有联系,又有区别,主要表现在:

(1) 信息是加工后的数据。信息是一种经过选摘、分析、综合的数据,它使用户可以更清楚地了解正在发生什么事。所以,数据是原材料,信息是产品,信息是数据的含义。

(2) 数据和信息是相对的。表现在一些数据对某些人来说是信息,而对另外一些人而言则可能只是数据。例如,在运输管理中,运输单对司机来说是信息,这是因为司机可以从该运输单上知道什么时候要为什么客户运输什么物品。而对负责经营的管理者来说,运输单只是数据,因为从单张运输单中,他无法知道本月经营情况,他并不能掌握现有可用的司机、运输工具等。

(3) 信息是观念上的。因为信息是加工了的数据，所以采用什么模型（或公式）、多长的信息间隔时间来加工数据以获得信息，是受人对客观事物变化规律的认识制约，由人确定的。因此，信息是揭示数据内在的含义，是观念上的。

(4) 表现形式。数据是对事件审慎、客观的记录，是一种结构化的记录，它有四种基本形式：数字、词汇、声音和图像。离散、互不相关的客观事实，以及孤立的文字、数据和符号，均表明数据的特征是缺乏关联性和目的性的。信息是已经排列成有意义的重要数据。比如，数字是数据，一张随机数字表则是信息。声音是数据，而经过排列并转换成一定韵律和声调的声音则是音乐信息。

2. 知识与信息

知识来自人们的思想，信息是知识的“子集”或“基石”，知识是信息应用和生产性使用。人们吸收信息，并对它进行处理、排序、分类、储存，然后用来构建知识。知识是经过加工提炼，将很多信息材料的内在联系进行分析而得出的系统结论。知识，是人们对信息加以评估和重新组织后可以有目的地用于个人或组织，这些“信息”就是知识。

对于研发人员而言，杂志名称及其网址可能就是信息，这些杂志或者网站的内容可以提供知识，当然这些知识可能是有用的也可能没有任何价值。对于设计者而言，信息可能包括确定一本在线手册及获取它的方法，在产品中如何以及为什么使用某种材料就是知识。对于管理者而言，竞争对手发布的财务报告就是信息，通过该报告分析得出的竞争对手的销售成本、综合开销及行政费用在过去的几年中持续降低就是知识。信息回答了“是什么”的问题，而知识回答了“怎么做”和“为什么”这样做的的问题。

3. 智慧与知识

与此相关，人们常说的智慧，从知识的角度来理解，是一种激活了的的知识。主要表现为收集、加工、传播、应用信息与知识的能力，对事件发展的前瞻性看法等。智慧得益于人的内在价值和信仰。

综合来看，资料是无相关性的事实；信息为经过分析与解释的资料，而知识则为一种信息与经验的结合体，是经验、价值、文字化信息的综合体。当知识具备有效应用的机制后，进一步可转化为一种智慧（智力资本），成为组织核心竞争力的来源。

1.3.2 知识管理和知识管理系统

较早提出知识管理概念的哈佛大学教授彼得·F·德克鲁认为，知识管理