

中国高等院校信息系统学科课程体系（CIS2011）规划教材

丛书主编：陈国青

商务智能方法与应用

刘红岩 编著

- 根据《中国信息系统学科课程体系2011》(CIS2011)的精神组织编写
- 教育部高等学校管理科学与工程类学科专业教学指导委员会、国际信息系统协会中国分会(CNAIS)推荐教材
- 与ACM/AIS最新推出的信息系统学科课程体系《IS2009/IS2010》同步

清华大学出版社

013039204

F713.36
823

中国高等院校信息系统学科课程体系 (CIS2011) 规划教材

商务智能方法与应用

刘红岩 编著



清华大学出版社
北京



北航

C1647829

F713.36
823

305000010

内 容 简 介

商务智能是从大量数据中发现隐含的知识,辅助管理人员做出科学决策的方法、系统和应用。本书主要介绍商务智能的基本概念、主要功能、系统架构,以及数据分析和数据管理的主要方法和技术。全书内容分为5个部分,分别介绍了数据仓库、在线分析处理以及数据挖掘的建模、分析和评价方法,涵盖多维数据模型的建模、多维分析方法以及各种知识发现方法(包括关联分析、分类、聚类、数值预测、序列模式挖掘、社会网络分析、数据流数据挖掘、多关系数据挖掘以及协同过滤和意见挖掘等);通过案例介绍了商务智能系统的应用;还介绍了常用软件系统及其使用方法,并对商务智能对社会的影响和未来发展进行了分析和展望。

本书内容具体、新颖、丰富、易于理解,反映了商务智能的最新发展趋势,适合作为信息管理、计算机应用、电子商务以及管理专业本科生和硕士生的教材,也可以作为数据分析人员的参考资料。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

商务智能方法与应用/刘红岩编著.--北京:清华大学出版社,2013

中国高等院校信息系统学科课程体系(CIS2011)规划教材

ISBN 978-7-302-31009-9

I. ①商… II. ①刘… III. ①电子商务—高等学校—教材 IV. ①F713.36

中国版本图书馆CIP数据核字(2012)第304166号

责任编辑:索梅 王冰飞

封面设计:常雪影

责任校对:李建庄

责任印制:宋林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:13

字 数:323千字

版 次:2013年5月第1版

印 次:2013年5月第1次印刷

印 数:1~3000

定 价:25.00元

产品编号:044644-01

中国高等院校信息系统学科课程体系(CIS2011)规划教材

编写委员会

主 任 陈国青 清华大学

副 主 任 毛基业 中国人民大学

委 员 (按姓名拼音排序)

陈 红	中国人民大学	赖茂生	北京大学
李 纲	武汉大学	梁昌勇	合肥工业大学
刘红岩	清华大学	刘咏梅	中南大学
卢 涛	大连理工大学	鲁耀斌	华中科技大学
牛东来	首都经贸大学	孙建军	南京大学
王 君	北京航空航天大学	张 新	山东经济学院
左美云	中国人民大学		



序

PREFACE

信息技术的飞速进步以及管理实践的发展演化,使得技术与管理的融合达到了一个前所未有的深度。一方面,移动性(如泛在互联、移动商务)、虚拟性(如虚拟体验、赛博空间)、个性化(如精准营销、推荐服务)、社会性(如社交媒体、社会商务)、极端数据(如富媒体、大数据)等特征以及电子商务、物联网、云计算等应用,不仅丰富了信息化的内涵和外延,也催生了许多新型管理模式和商务机会;另一方面,许多传统的业务决策在今天已经变成了或正在变成信息分析决策,许多惯常的组织管理活动在今天已经变成了或正在变成信息管理活动。

信息系统作为一个承载组织业务模式的人—机计算平台,其主流学科领域呈现出技术与管理并重的特点。在技术视角层面,信息系统的研究和应用关注信息是如何提取和处理、系统是如何开发和构建;在管理视角层面,信息系统的研究和应用关注系统是如何被采纳和使用、组织是如何被影响和改变。如同产品开发和产品使用一样,信息系统的技术与管理视角反映出“造”和“用”的某种分野,但信息系统却又是“造”和“用”的统一体,反映出技术与管理的密切内在联系。

信息系统的新拓展涉及深度和广度两个维度。在深度上,将沿着事务处理、分析处理和商务智能的轨迹提升,以逐步回答管理决策者在经营运作中提出的“发生了什么”、“为什么会发生”、“将发生什么”的问题。在广度上,拓宽企业内部业务信息化的领域并进行必要的集成;同时,向企业外延展信息化的触角,以支撑与客户和供应商的业务活动。这就需要重视企业外数据的分析与处理,如用户生成数据(评论、口碑等)、社交网络和媒体反馈、商誉和企业舆情分析等。

在这样的背景下,研究探讨技术与管理融合的理论与实践、推动信息化建设和人才培养是我国信息系统相关专业领域的教育工作者面临的任务。“中国高等院校信息系统学科课程体系(CIS2011)”是在教育部管理科学与工程类专业教学指导委员会、国际信息系统协会中国分会(CNAIS)的委托下成立课题组编纂完成的,旨在更好地反映国际信息系统领域的发展趋势和专业课程体系变化,体现我国信息化人才的需求模式和专业课程体系要素,有助于促进我国相关专业领域的课程建设。

本教材系列是“中国高等院校信息系统学科课程体系(CIS2011)”的规划教材系列,涵盖下列主要教材书目,以体现 CIS2011 课程体系的整体框架、知识模块和时代特点:管理信息系统(MIS)、信息资源管理(IRM)、计算机网络与应用(CN)、数据结构(DS)、数据库系统原理(DB)、信息系统分析与设计(SA)、信息组织(IO)、商务智能方法与应用(BI)、信息系统项目管理(PM)、信息检索(IR)、电子商务(EC)、企业信息系统及应用(EIS)。

本教材系列主要作为信息系统及其相关专业本科生专业核心课和主干课程教材,也可以作为相关领域研究生的参考教材。同时可供企业信息化实践人员参考。



(陈国青)

清华大学经济管理学院 EMC 讲席教授
教育部管理科学与工程类学科专业教学指导委员会副主任
国际信息系统协会中国分会(CNAIS)主席

2013 年 3 月于清华园

前言

FOREWORD

企业运营过程中不断积累的数据越来越多,形成了传统数据管理系统无法有效管理和使用的大数据。同时,企业运营环境越来越复杂,竞争越来越激烈,商业机会稍纵即逝,企业管理者需要全面掌握企业各方面的运营状况,及时做出反应,进行快速科学的决策。商务智能系统具有数据集成、数据管理以及数据分析的功能,可以从大量数据中发现隐含的模式,辅助管理者发现业务运营的规律和异常,降低运营成本和减少风险,以增加收入,同时提高客户满意度,已经被越来越多的企业采纳,成为提高企业竞争力的重要方法,在许多重要的商业应用中发挥着重要的作用。

全球管理资讯公司麦肯锡在 2011 年发布了一个题目为《大数据:创新、竞争和生产力的下一个前沿领域》的研究报告。报告中指出,大数据可以为世界经济创造重要价值,提高企业和公共部门的生产率和竞争力,并为消费者创造大量的经济剩余。报告中总结了利用大数据创造价值的 5 种方法和途径。同时指出,为了充分挖掘大数据的潜力需要解决的几个问题之一就是技术,包括数据集成、分析、可视化等方面的技术。利用这些技术才能从大数据中发现相关信息和知识。这些技术正是本书将重点介绍的内容,属于商务智能的技术范畴。报告中预测,到 2018 年美国将有 140 000~190 000 的数据分析方面的人才缺口。

2012 年 3 月美国白宫科学与技术政策办公室(OSTP)发布了《大数据研究和开发计划》(Big Data Research and Development Initiative)。该计划由美国国家科学基金会、国防部、能源部等 6 个联邦部门和机构投资两亿美元,旨在推进用于收集、存储、管理、分析以及共享大量数据的先进核心技术的发展,增强国家安全性,扩大使用 and 开发大数据技术的人才队伍。

本书作者自 20 世纪 90 年代末就开始在清华大学为研究生开设商务智能课程,十几年来一直从事商务智能领域的教学和科研工作。本书内容是作者近十几年教学和科研内容的总结,旨在为管理专业、信息系统专业、计算机专业以及其他相关专业介绍商务智能的基本概念、功能和分析方法。全书分为 5 个部分,共 14 章内容。第 I 部分介绍了商务智能概念和商务智能过程,主要讲解了商务智能的基本概念、系统构成、发展历史以及系统开发过程。第 II 部分详细介绍了商务智能的各种数据分析方法,包括关联分析、分类、聚类、数值预测等方法。对于每种方法,通过介绍几个经典易懂的分析方法,使读者能够了解每类方法的主要概念和实现的基本原理,不求技术上的深度和广度。第 III 部分是基础技术,介绍了实现数据的分析需要的基础支撑技术,包括数据预处理方法、数据仓库、在线分析处理和可视化技术。第 IV 部分是商务智能应用与系统,介绍了商务智能的应用领域,并以推荐系统和意见挖掘为例,介绍了典型的应用方法。另外,还介绍了常用的商务智能系统,包括开源软件 and 商品化的商务智能系统。为了便于理解前面介绍的数据分析方法,此部分还介绍了开源数据挖掘软件 Weka 的使用方法。第 V 部分是深度应用与发展,进一步介绍了针对复杂数据的分析

方法,包括序列数据分析、社会网络分析、数据流分析、多关系数据分析方法等。最后,对商务智能的进一步发展及其对社会的影响进行了分析和展望。

本书在编写过程中得到了教育部高等学校管理科学与工程类学科教学指导委员会副主任、国家信息化专家咨询委员会成员、国际信息系统协会中国分会主席清华大学的陈国青教授的帮助和指导,得到了国家自然科学基金委(71272029)的资助,在此表示感谢。鉴于作者水平有限,书中难免有不当及不足之处,望广大读者批评指正。

刘红岩

2013年春于清华园

目录

CONTENTS

第 I 部分 商务智能概念及过程

第 1 章 导言

1.1 商务智能的基本概念

1.1.1 数据

1.1.2 信息和知识

1.2 商务智能的系统构成

1.3 商务智能的发展历史

练习题 1

第 2 章 商务智能过程

2.1 商务智能系统的开发方法

2.1.1 商务智能系统的开发过程

2.1.2 商务智能系统成功的关键因素

2.2 数据库与数据仓库

2.3 在线分析处理与在线事务处理

2.4 商务智能与决策支持系统

练习题 2

第 II 部分 商务智能方法

第 3 章 关联分析

3.1 频繁模式与关联规则

3.2 频繁项集的典型挖掘方法

3.2.1 逐层发现算法 Apriori

3.2.2 无候选集发现算法 FP-growth

3.3 关联规则的生成方法

3.4 关联规则的其他类型

3.4.1 多层次关联规则

3.4.2 负模式

3.4.3 结构化数据中的关联分析

3.5 关联规则的兴趣度的其他度量

练习题 3

第 4 章 分类

4.1 分类的概念

4.2 决策树分类方法

4.2.1 决策树的构建过程

4.2.2 属性的类型及分裂条件

4.2.3 决策树的剪枝

4.3 朴素贝叶斯分类

4.4 k 近邻分类

4.5 分类性能的度量方法

4.5.1 测试数据集的构造

4.5.2 分类性能的度量指标

4.5.3 不同分类模型的比较

练习题 4

第 5 章 数值预测

5.1 数值预测的概念

5.2 回归方法

5.2.1 一元线性回归

5.2.2 多元线性回归

5.2.3 非线性回归

5.3 回归树与模型树

5.3.1 模型树的构建

5.3.2 模型树的剪枝

5.3.3 算法

5.4 k 近邻数值预测

5.5 预测误差的度量

练习题 5

第 6 章 聚类分析

6.1 概述

6.1.1 聚类的概念

6.1.2 聚类方法分类

6.2 相似度衡量方法

6.2.1 数据类型

6.2.2 基于内容的相似度衡量

6.2.3 基于链接的相似度衡量

6.3 k 均值聚类法

6.4 层次聚类方法

6.5 DBSCAN 方法

6.6 聚类效果衡量方法

练习题 6

第Ⅲ部分 基础技术

第 7 章 数据预处理

7.1 数据预处理的原因和任务

7.2 数据规范化

7.3 数据离散化

7.3.1 分箱离散化

7.3.2 基于熵的离散化

7.3.3 离散化方法 ChiMerge

7.4 数据清洗

7.5 特征提取与特征选择

7.5.1 特征选择

7.5.2 特征提取

练习题 7

第 8 章 数据仓库

8.1 数据仓库的基本概念

8.2 数据仓库的体系结构

8.3 多维数据模型

8.3.1 多维数据模型的概念

8.3.2 多维数据模型的构建方法

8.4 数据仓库项目的开发

8.4.1 数据仓库的开发模式

8.4.2 数据仓库开发过程

练习题 8

第 9 章 在线分析处理

9.1 在线分析处理简介

9.2 多维数据模型中的层次设计

9.3 立方体的定义和计算

9.4 OLAP 的多维数据分析

练习题 9

第 10 章 商务智能可视化

10.1 商务智能可视化的类型

10.2 数据可视化

10.3 过程和结果可视化

10.4 积分卡和仪表盘

练习题 10

第Ⅳ部分 应用与系统

第 11 章 商务智能应用

11.1 商务智能应用领域

11.1.1 关系营销

11.1.2 生产管理

11.2 推荐系统

11.2.1 基于用户的协同过滤

11.2.2 基于产品的协同过滤

11.2.3 基于内容的推荐方法

11.3 意见挖掘

11.3.1 特征和意见的抽取

11.3.2 意见极性判断

练习题 11

第 12 章 商务智能系统

12.1 开源数据挖掘软件

12.1.1 概述

12.1.2 Weka

12.2 商品化的商务智能系统

练习题 12

第 V 部分 深度应用与发展

第 13 章 复杂数据的商务智能分析方法

13.1 序列模式挖掘

13.1.1 序列模式的定义

13.1.2 序列模式挖掘算法

13.2 社会网络分析

13.2.1 中心度分析

13.2.2 链接分析

13.3 数据流数据挖掘

13.4 多关系数据挖掘

练习题 13

第 14 章 商务智能的社会影响与发展

14.1 商务智能中的隐私保护

14.2 移动商务智能

14.3 云商务智能

练习题 14

参考文献

第 I 部分

商务智能概念及过程



第1章 导言

当今的企业处在信息爆炸、经济全球化的背景下,面对瞬息万变、竞争激烈的业务环境,面对日益挑剔的“强大”客户,企业面临着重重的压力和巨大挑战,如何顶住压力、迎接挑战关系到企业的生死存亡。在这种情况下,企业的管理人员必须能够对企业的运营状况洞察秋毫,迅速地、敏捷地、准确地应对各种变化,做出科学的决策。明智正确的决策不是仅凭直觉就能做出的。传统的在线事务处理系统可以帮助企业高效率地开展日常业务,在提供决策支持信息方面却无能为力。然而企业内部的业务运营系统记录的大量业务数据,企业外部竞争者的相关信息、客户的信息、浩如烟海的互联网信息等,以及不断发展的信息技术给企业带来了机会。如何从这些海量数据中挖掘出有价值的知识并辅助管理人员进行决策,是每个企业亟须解决的问题。商务智能正是在这样的背景下诞生了。本章将介绍商务智能的基本概念、系统构成和发展历史。

1.1 商务智能的基本概念

商务智能(Business Intelligence)这个术语最早由 Gartner Group 的分析师 Howard Dresner 提出。在 1996 年,他将商务智能定义为由数据仓库(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的、以辅助企业决策为目的的一类技术及其应用。自此之后,不同的权威机构、不同的学者从各个方面赋予了商务智能不同的内涵。对工业界来说,商务智能是一类技术或工具,利用它们可以对大量的数据进行收集、管理、分析和挖掘,以改善业务决策水平,增强企业的竞争力。对学术界来说,商务智能是一套理论、方法和应用,通过它们可以快速地发现海量数据中隐含的各种知识,有效地解决企业面临的管理和决策问题,支持企业的战略实施。商务智能可以被看做是从海量数据中发现知识的过程,也可以被看做是企业拥有的一种能力,一种通过对数据的管理、分析和挖掘从而快速地做出正确决策的能力。

1.1.1 数据

数据是对事物描述的符号。在计算机科学中,数据是数字、文字、图像、声音等可以输入计算机,然后被识别的符号。数据是未经加工和修饰的原料。企业运营离不开数据。企业运营的各个环节每天都在积累数据,如供应商、客户的数据,销售、生产以及库存数据等。社会化媒体、智能化手机等使得全世界不计其数的个体也在不断地产生数据。这些数据往往超出了传统的数据库系统能够存储、管理和分析的范畴,也因此促进了商务智能技术的发展。

一个企业的各个业务部门通常都有相应的信息系统支持,这些系统一般通过关系数据

库存储和管理数据。这些数据是通过统一的二维表格的形式存储的,这类数据称为结构化数据。例如,表 1.1 存放的是有关交易细节的数据。这样的数据结构在关系数据库中称为表(Table),表中的每列代表一个属性。表 1.1 所示的表含有 5 个属性:交易号、商品号、单价、折扣和数量。表中的每一行描述了一笔交易的相关信息。

表 1.1 交易细节表

交易号	商品号	单价	折扣	数量
00587205	133792	2.99	0	1
00587205	146775	2	0	1
00587205	000700	1	0	4.5
00587206	147525	1	0	5.9
00587206	113838	1	0	7.5

还有一类数据无法利用统一的二维表格的形式存储。例如,文档、HTML 文件以及各种多媒体数据等,这类数据称为非结构化数据(Unstructured Data)。非结构化数据中一种常见的数据是文本数据。例如,客户对所购买商品的评论信息就属于文本数据,如图 1.1 所示。

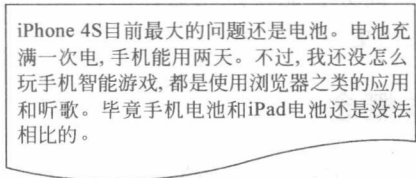


图 1.1 有关苹果手机 iPhone 4S 的一段评论文本

2011 年 6 月,麦肯锡全球研究院(MGI)发布了一份报告——《大数据:创新、竞争和生产力的下一个前沿领域》。该报告指出,大数据(Big Data)是指大小超出了典型数据库软件获取、存储、管理和分析的能力的数据集。不存在一个固定的符合大数据的数据集大小的标准。不同行业采用的数据管理软件工具不尽相同,其通常拥有的数据库大小也不相同。而且随着时间的推移,数据集的大小在不断增长,数据管理技术也在不断进步。就目前来说,大数据的数据集大小从几个 TB(TeraByte)到数个 PB(PetaByte,1PB=1024TB)都有可能。

据麦肯锡估计,全球企业 2010 年硬盘上存储了超过 7EB(ExaByte 1EB=1024PB)的新数据,而消费者在个人计算机及笔记本电脑等设备上存储了超过 6EB 的新数据。1EB=2³⁰ GB,1EB 的数据相当于美国国会图书馆中存储数据的四千多倍。很多研究机构都预测,全球产生、存储和消耗的数据正在呈指数增长。

大数据中蕴藏着商机,可以变为一个企业的核心竞争力。报告中指出,大数据可以发挥重要的经济作用,不但有利于私人商业活动,也有利于国民经济和公民。数据可以为世界经济创造重要价值,提高企业和公共部门的生产率和竞争力,并为消费者创造大量的经济剩余。此报告深入分析了 5 个领域,以说明大数据是如何创建价值的。这 5 个领域分别包括:美国的医疗卫生行业、欧洲的公共管理领域、美国零售业、全球制造业和个人地理位置信息。其要点总结于图 1.2 中。从图中可以看到,如果能够有效地利用大数据来提高效率和质量,

预计美国医疗行业每年通过数据获得的潜在价值可超过 3000 亿美元,能够使得美国医疗卫生支出降低超过 8%;通过利用大数据提高政府行政管理方面的运作效率,可以为欧洲国家创造超过 2500 亿欧元的价值;充分利用大数据的零售商有可能将其经营利润提高 60% 以上;利用大数据可以使制造业每年在产品开发、组装方面节省成本 50%,降低运营资本 7%;对于个人地理位置信息的有效利用可以为服务提供商创造 1000 亿美元的收入。



图 1.2 大数据在 5 个领域可以创造的价值

(图片来源,麦肯锡报告《大数据:创新、竞争和生产力的下一个前沿领域》)

该报告中总结了利用大数据创造价值的 5 种方法和途径,如创造透明度、发现需求、改进性能、客户分群、用算法替代或支持人工决策,开发新的业务模式、产品、服务等。下面是该报告对上述几个方面内容的总结和精简。

1) 创造透明度

使得一个企业的利益相关者能够及时、方便地访问大数据可以创造巨大的价值。例如,在制造业,将研发部门、工程、制造等部门的数据进行集成可以显著节省时间,从而可利用更多的时间改善质量和进行营销。

2) 通过实验发现需求、揭示性能变化、改进性能

企业内部存储了大量详细的事务以及性能数据。利用 IT 技术可以记录业务过程中的参数,进而进行控制实验。利用数据分析业务性能的变化性,从而使得管理者更好地进行性能管理。

3) 客户分群

利用大数据可以对客户进行细分,创建非常细的客户群体,从而根据客户群体的特点提供满足需求的产品和服务。这种方法在营销和风险管理领域应用广泛,同时可以应用到其他领域,如公共部门。现在,有些公司(如生活消费品和服务业提供公司)甚至利用大数据处理技术来实现客户的实时分群,以进行有效地促销或推送广告。

4) 用算法替代或支持人工决策

通过分析大数据可以发现潜在的、有价值的知识,可以显著改进决策质量、降低风险。这种分析有许多应用,例如,可以帮助税收部门发现需要进一步检查的缴税者,可以帮助零售商根据实时的库存量和销量自动调整库存。通过对整个大数据进行分析,而不是对小部分抽样数据进行分析,可以辅助管理人员做出更好的决策。

5) 开发新的业务模式、产品和服务

利用大数据可以使公司开发出新的产品和服务,提升已有产品和服务,开创崭新的业务模式。制造业企业根据用户使用产品的体验可以改进下一代新产品的设计,开发新的售后服务模式。新涌现的大量位置信息开创了崭新的、基于位置的服务模式,例如导航服务、推荐服务等。

为了充分挖掘大数据的潜力需要解决的几个问题之一是技术,包括数据集成、分析、可视化等方面的技术。利用这些技术才能从大数据中发现信息和知识。这些技术也正是本书在下面的章节中将要重点介绍的内容。

1.1.2 信息和知识

数据可以看做是未经加工的原料,通过一定的技术和方法,对数据进行集成和分析,挖掘其潜在的规律和内涵,得到的结果是信息(Information)。信息是具有商务意义的数据。例如,通过对零售信息的集成和分析发现,某超市的客户群根据其消费行为可以分为若干群体,每个群体具有一些明显的特征。例如,其中一个群体是单身女性,喜欢经常购买化妆品,消费金额高。客户群的这些特征对于销售管理人员来说是非常有意义的,因此,可以称之为信息,有别于原始的销售交易记录和客户的基本特征方面的数据。

当信息用于商务决策,并基于决策开展相应的商务活动时,信息就上升为知识(Knowledge)。信息转化为知识的过程不仅需要信息,而且需要结合决策者的经验和能力,用以解决实际问题。例如,某连锁超市的经理发现,近期化妆品的销售业绩下降了,为了解决该问题,决定采取促销措施。根据对数据的分析得到的客户分群的信息,销售经理锁定了促销的目标客户群,最终开展了为这部分客户邮寄优惠券的促销活动。这与随机选取一部分用户进行促销比较起来,显然要有效得多。由此可以看到,通过对数据的分析得到的信息,原本看起来是缺乏生命力的,但是一经利用,就会给企业带来利润。

1.2 商务智能的系统构成

一个商务智能系统通常包含6个主要组成部分:数据源、数据仓库、在线分析处理、数据探查、数据挖掘以及业务绩效管理。

1. 数据源

商务智能的根本是数据。数据的来源主要包括企业内部的操作型系统,即支持各业务部分日常运营的信息系统,以及企业外部的信息,如人口统计信息、竞争对手信息等。综合多方面的数据进行分析,可以有效提高信息的有效性。

2. 数据仓库

各种数据源的数据经过抽取、转换之后需要被放到一个供分析使用的环境,以便对数据进行管理,这就是数据仓库(Data Warehouse)。通常针对单个部门的数据仓库称为数据集