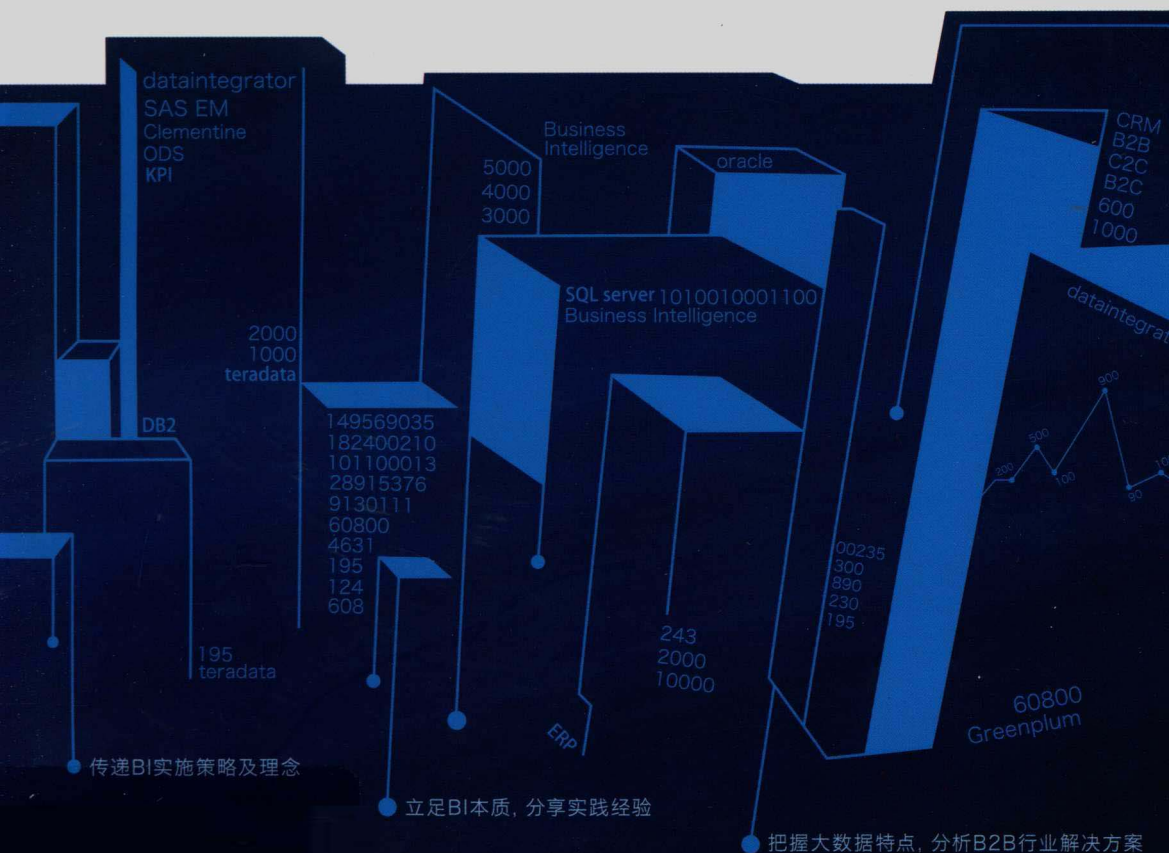


商业智能 在电子商务中的 实践与应用

黄建鹏/徐晓冬/魏宝军 著



公司

东南大学出版社

商业智能在电子商务中的 实践与应用

黄建鹏 徐晓冬 魏宝军 著

焦点科技股份有限公司

东南大学出版社

·南京·

内容提要

本书全面详细地叙述了商业智能在电子商务行业的应用以及发展情况,重点探讨了适用于电子商务行业的 BI 整体解决方案,结合云技术打造了分布式数据仓库架构。同时对数据仓库、OLAP 分析、数据挖掘等技术进行了深入探讨,并结合应用帮助读者了解技术的本质。最后,笔者总结了 BI 对于电子商务的应用价值和电子商务行业成功实施 BI 的经验,对于读者实施或学习 BI 都有实际的指导意义。

本书内容深入浅出,是企业领导决策层、商业智能开发人士和初学者的宝贵参考资料。

图书在版编目(CIP)数据

商业智能在电子商务中的实践与应用 / 黄建鹏, 徐晓冬, 魏宝军著. —南京: 东南大学出版社, 2012. 12

ISBN 978-7-5641-3797-7

I. ①商… II. ①黄…②徐…③魏… III. ①电子商务—研究 IV. ①F713.36

中国版本图书馆 CIP 数据核字(2012)第 277151 号

商业职能在电子商务中的实践与应用

出版发行	东南大学出版社
出 版 人	江建中
网 址	http://www.seupress.com
社 址	南京市四牌楼 2 号(邮编:210096)
经 销	全国各地新华书店
印 刷	南京玉河印刷厂
开 本	700mm×1000mm 1/16
印 张	7.25
字 数	135 千字
版 次	2012 年 12 月第 1 版
印 次	2012 年 12 月第 1 次印刷
书 号	ISBN 978-7-5641-3797-7
定 价	30.00 元

本社图书若有印装质量问题,请直接与营销部联系。电话(传真):025-83791830。



前 言

当前国内最具竞争与活力的行业毫无疑问是电子商务,大家可以随手列出淘宝、京东、拍拍、中国制造网、卓越、苏宁易购、库巴等一大批知名网站。在网站建设初期,功能开发主要用来满足大家发布、查阅及交易,但随着十多年的发展,一些领先企业包括中国制造网当前的发展重心已经进入到信息优化阶段,为买卖双方提供精确化信息支持,提升买卖双方网站应用效果,其核心思想是借助商业智能技术对海量数据进行分析与挖掘,提升对网站、网站用户的了解,提取用户感兴趣的信息,让卖方知道什么商品更好卖,不断优化自身产品和服务,让买方也能找到更合适的商品,最终提升了网站对用户的黏度。

来自焦点科技分享

为此焦点科技股份有限公司(简称“焦点科技”、“焦点”)自 2010 年开始全方位实施 BI,经过两年的工作,已经建立起领先的企业级商业智能系统,同时积累了相当可观的技术、项目及业务经验,根据公司在技术方面提出的开放、务实、卓越的口号,特别组织了本次专著的撰写。本书作者均来自焦点科技研发中心 BI 研发部,成员都具备丰富的 BI 项目实施经验。焦点科技作为国内外贸 B2B 电子商务领先企业,在 BI 应用方面已经走到了行业前面,6 个月时间建立起成熟稳定的数据仓库,一年时间 BI 应用覆盖了 80% 的业务部门,并推广到高层领导的决策分析,真正实现企业级自上而下 BI 应用,并且拥有众多的数据挖掘成功案例,有力地证明了在当前条件下 BI 实施不仅没有风险,还能做到高效。当前正值公司大范围推广及应用 BI 阶段,为了提升自我实力及在行业内分享成功经验,这里特别以专著方式起草本书,同时也是所有作者在百忙之中的一次难得总结。

内容特点

本书既较为详尽地介绍电子商务的商业智能应用,又从商业智能技术及应用的基本目的来谈及一整套的 BI 实施方案、设计、技术及理念,这本专著主要具有下面的鲜明特点:

1. 成体系:成体系介绍 BI,涵盖数据仓库、ETL、OLAP 分析、数据挖掘,涉及完整的企业级 BI 内容。

2. 认识深:对 BI 有深刻的认识,不拘泥于软件和产品,也不一味地介绍新技术,更注重 BI 的本质和对最佳实践的追求,重视新技术的合理应用及设计。
3. 实战经验丰富:既有通用的方案和技术标准,又有相应的技术实现。
4. 强调策略:注重 BI 实施策略及理念,传递的统一规划、逐步实施 BI 的经验具有较高的现实性。
5. 注重 B2B 行业的解决方案,大数据处理及应用是其一大特点。

本书内容简介

各章的简要内容如下:

第一章对商业智能的应用背景、发展历史、应用现状以及数据仓库、ETL、报表和数据挖掘等商业智能相关技术进行了介绍。

第二章分析了商业智能在电子商务行业的应用前景。该章讨论了商业智能在电子商务平台优化、产品创新、客户分析以及网络安全方面的应用。

第三章介绍 BI 在电子商务企业的整体解决方案。详细介绍了 BI 的系统架构的构建;结合云技术讨论了分布式数据仓库的架构;介绍了实践中 ETL 的开发以及测试、多维分析技术;深入探讨了数据挖掘技术的价值和应用;对 BI 在商业智能的应用领域进行了详细分析;最后,还讨论了 BI 的软硬件实施方案、BI 实施策略和风险控制等实施中的具体问题。

第四章、第五章分别介绍了 BI 对内、对外的支持。

第四章介绍了 BI 对内的决策分析支持,包括高层仪表盘、企业级绩效、主题性 OLAP 分析以及分析性工作流程,各节详细介绍了分析的维度和指标。

第五章介绍了 BI 对外对于网站用户的支持,包括网站推广效果优化、市场分析、行业及竞争对手分析和增值服务效果及推荐,各节详细介绍了分析的维度和指标。

第六章、第七章对目前焦点科技实施 BI 以来的一些经验进行总结,介绍了 BI 成功实施的几个关键点,BI 在焦点科技的应用展望。

最后,第八章介绍了商业智能的最新发展方向。

本书适合对象

本书内容包含方案、设计、技术、应用、案例,力求从实际应用及本质上对 BI 进行归纳及理解,可为专业人士提供参考,实战内容比如数据仓库设计、ETL 开发、OLAP 分析、数据挖掘都有实用简单开发标准,从而受众的门槛大大降低。总体来看,主要对下面一些人员价值较大。

1. BI 从业人员:书中有完整的 BI 解决方案,又针对电子商务行业提供了大数



据的特别解决方案,这些方案实用性非常高;有配备完善的技术实现,针对 ETL、OLAP 分析、数据挖掘还提供了实用型技术标准,该技术标准对项目的复制、人员的培养、代码的规范都有一定意义;另外书中非常注重关键技术的深入分析,对于 BI 从业人员来讲,如果具备一定的基础,通过该内容学习可以很快明白自己工作过程中的一些疑虑,快速掌握这些关键技术,避免低水平摸索,甚至走弯路;书中对实施方法论有非常多的现实考量,应该说是我们 BI 从业人员技术之外的有益补充,这方面有的时候直接关系到项目的成败。

2. 电子商务相关人士:对这部分人员来讲主要是 BI 的普及,通过本书大家可以全面了解 BI 在电子商务的应用和价值,并且当前 BI 实施能力已经不再有风险,一切变得容易。

3. IT 管理人员:IT 领导一定会或多或少了解 BI,那么本书对 IT 领导最大的价值就是其实战性和中立性。内容包含大量 BI 核心技术和实施方法论,图示丰富,如果有一定 IT 基础,浏览本书后基本上可以对 BI 了解了七八成;另外书中没有太多的具体软件,更注重 BI 技术的本质问题和解决思路,这有助于 IT 领导们选择适合自己公司的 BI 产品。这里不得不说的是虽然现在 BI 各大厂商看起来都很强大并且差异不大,但事实上项目成功率、实施成本相差很大。

4. 数据挖掘研究者:数据挖掘比较讲究算法,实际工作中数据分析应用门槛较低,需求也较大,如果对常规的 OLAP 分析有较多理解,将有助于挖掘算法寻找应用空间,同时数据仓库设计及数据加载技术是数据挖掘准备数据的必备过程。总的来说了解基本的 BI 实践对于数据挖掘应用具有较大促进作用,现在有很多挖掘人士对存储过程、ETL 技术非常熟练,因为他们认为这是数据挖掘必备的技能。

5. 大学毕业生:BI 的入门门槛很低,基本上 1~2 天就能开始做事,是非常适合刚毕业的学生去从事的一门工作。但客观地说,BI 要做得好必须懂架构、调优及分析,甚至数据挖掘,对数学、统计等常规冷门理科生来讲,不失为一个好的选择,BI 行业入门快,工作长远后能用得上自己的专业知识。本书有 BI 详尽的讲述,涉及面广,对于毕业生来讲得先找一个点来开始工作更为现实,相信里面的实战性还是可以帮助大家快速入门并且做些事情,从长远考虑,从业 BI 需懂得关于数据仓库、ETL、报表分析、数据挖掘等知识。

致谢

本文的撰写得到了焦点科技院士工作站、研究院的大力支持,在此表示感谢。同时 BI 研发部的其他同事也参与了该书的编写,包括:陈静、张登峰、肖洁岑、夏平嵩、丁艳艳、胡春香、王婷、房鹏展、余华等。



目 录

第一章 商业智能概述	1
1.1 什么是商业智能	1
1.2 商业智能在国内应用背景	2
1.3 商业智能发展历史和国内应用现状	4
1.4 商业智能相关技术介绍	5
1.5 小结	7
第二章 BI 在电子商务的应用前景	9
2.1 电子商务发展对商业智能的必然需求	9
2.2 商业智能在电子商务的应用情况及价值	9
2.3 小结	11
第三章 商业智能在电子商务的解决方案	12
3.1 系统架构	12
3.2 数据仓库及模型	13
3.2.1 数据仓库的定义	14
3.2.2 数据仓库的架构	14
3.2.3 数据仓库模型	17
3.3 ETL	21
3.3.1 ETL 目标和基本要求	22
3.3.2 ETL 的开发	22
3.3.3 ETL 的测试	25
3.4 多维分析	27
3.4.1 多维数据分析技术	27
3.4.2 多维数据分析的实现	28
3.4.3 多维分析在商业智能中的应用	30
3.5 数据挖掘	31
3.5.1 价值体现	31



3.5.2	主要应用	32
3.5.3	应用实现要求	33
3.5.4	常见问题	34
3.6	BI 在电子商务领域的应用	35
3.6.1	电子商务运营决策分析	35
3.6.2	对外-支撑网站用户分析	36
3.7	软硬件方案	38
3.8	实施策略及风险控制	40
3.8.1	自上而下策略	41
3.8.2	自下而上策略	43
3.8.3	整体规划、逐步实施	44
3.9	BI 项目风险控制	46
3.9.1	组织风险	46
3.9.2	技术风险控制	47
3.9.3	业务风险	48
3.9.4	实施风险	49
3.10	小结	50

第四章	BI 对内决策支持平台	52
4.1	高层仪表盘	52
4.1.1	定义	52
4.1.2	目标和价值	52
4.1.3	设计关键点	53
4.1.4	高层仪表盘在焦点科技的应用	54
4.2	企业级绩效	56
4.2.1	定义	56
4.2.2	目标和价值	57
4.2.3	设计关键点	57
4.2.4	企业级绩效管理在焦点科技的应用	58
4.3	主题性 OLAP 分析	60
4.3.1	目的和作用	60
4.3.2	销售分析	60
4.3.3	运营分析	62
4.3.4	订单分析	64



4.3.5 客服分析	65
4.3.6 买家服务分析	67
4.4 分析型工作流程	68
4.4.1 分析型工作流程是什么?	68
4.4.2 分析型流程的应用价值	68
4.4.3 分析型流程在电子商务的应用	69
4.5 小结	70
第五章 BI 对外支持网站用户	71
5.1 网站推广效果及优化	71
5.1.1 用户网站推广目标	71
5.1.2 推广效果影响因素	72
5.1.3 BI 推广效果分析和优化支持	72
5.2 市场分析	74
5.2.1 平台行为分析	75
5.2.2 宏观市场分析及预测	76
5.3 行业及竞争对手分析	77
5.3.1 行业经济特性分析	77
5.3.2 行业竞争环境分析	77
5.3.3 竞争对手分析	78
5.4 增值服务效果及推荐	79
5.4.1 增值服务	79
5.4.2 电子商务推荐系统	80
5.5 小结	82
第六章 B2B 电子商务 BI 成功实施的关键点	84
6.1 是否有强大的 BI 技术架构?	84
6.2 是否有大数据的存储设计能力?	85
6.3 是否能构建适合公司自身的软硬件平台?	88
6.4 是否懂得合理利用数据挖掘技术?	89
6.5 数据质量怎么解决?	91
6.6 B2B 行业的数据增值服务?	93
6.7 怎样获得持续的投入和支持?	94
6.8 小结	96

第七章 焦点科技 BI 应用价值及展望	98
7.1 应用价值	98
7.2 BI 应用展望	100
7.3 小结	101
 第八章 商业智能最新发展方向	 102
8.1 信息集成度更高	102
8.2 操作更简单、体验更人性化	102
8.3 速度更快、性能更优:BI 充分融合云技术	103
8.4 可视化技术及交互式分析成技术发展亮点	103
8.5 小结	104
 参考文献	 105



第一章 商业智能概述

1.1 什么是商业智能

商业智能,又称商务智能,英文为 Business Intelligence,简写为 BI。目前,学术界还没有统一的商业智能的定义。商业智能的概念于 1996 年最早由加特纳集团(Gartner Group)提出,加特纳集团将商业智能定义为: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。商业智能技术提供企业迅速分析数据的技术和方法,包括收集、管理和分析数据,将这些数据转化为有用的信息,然后分发到企业各处。

在信息化高度发达的今天,商业智能被越来越多的企业所重视。但在 IT 业界,对于商业智能有着各种各样的叫法和理解,有人说商业智能就是数据仓库,也有人说商业智能就是做报表,如果从现有的工具和项目划分来理解,确实很容易陷入纷争。这里我们先来解释几个相关名词:

数据仓库,英文为 Data Warehousing,简写为 DW,是企业级分析型需求的公共数据源。

ETL 是数据从业务系统抽取(Extraction)、转换(Transformation)和装载(Load)到数据仓库的过程^[1]。

OLAP 是联机在线分析(On-line Analytical Processing)^[2]的简称,就是基于数据库、大数据的联机快速分析,分析一般指多维分析,可以根据维度汇总下钻等操作,主要特点还是要求快,数据量大,其他功能和 Excel 的数据透视表差不多。

数据挖掘:简单理解就是通过人工智能技术从海量数据里面提取有用知识。

直观的来说,商业智能指综合利用数据仓库和 ETL 技术、OLAP 分析以及数据挖掘技术对数据进行有效的整合和存储,并对数据进行分析,提取出其中蕴藏的知识,从而帮助客户进行决策分析。

各大 IT 厂商如 IBM、ORACLE、SAP、微软等著名企业都纷纷提供整体的 BI 解决方案和产品;具备一定规模的公司都已经或者准备做 BI 方面的工作。目前,BI 已经为很多 IT 公司带来了实实在在的价值,腾讯公司在其官方访谈中也强调“数据挖掘”是其秘密武器;淘宝网借助 BI 技术实现的数据魔方、量子统计等产品



具有强大的创收能力;焦点科技也通过 BI 技术实现了企业内决策管理的智能化,为企业的快速响应、精准决策提供了信息平台,另外在信息自动审核、网站效果优化等方面借助数据挖掘技术也已经取得了良好的效果,既可以对内提升企业效率,又可以对外帮助客户提升推广效果。

1.2 商业智能在国内应用背景

本节主要基于国内传统行业以及新兴的互联网行业的具体情况来谈商业智能的出现和应用背景,避免使用国外的经典案例和空泛的理论。以下经验都是团队成员在从事 BI 的工作中曾经遇到和处理过的实际问题。

1) 迫切的分析型需求

中国经济经历 90 年代以来快速发展的 20 年,涌现出了一大批有财力、有活力的公司,特别是改制后的垄断企业、沿海知名民营企业整体建设都逐步向国际领先企业靠拢,在 2000 年前后都陆续积极建设核心业务系统,电讯行业的计费系统、生产制造的 MES、零售分销行业的 ERP 系统等都基本在这一时期快速得以实现,这些系统让操作人员在工作流程化规范化方面取得了长足进步,资金流得到严格控制和监管。

随着这些系统的完善特别是业务系统数据的积累,公司管理决策者以及业务分析人员越来越发现基于业务系统基础数据的各种分析对科学地制定企业管理决策非常有帮助,在工作过程中有越来越依赖系统数据的趋势,这一时期其分析工作普遍是如下流程:

- 先从业务系统导出数据到 Excel,然后通过灵活的 Excel 进行加工,最后生成报表。
- 生成的报表除了支持客户自身的工作之外,还需要传递给相关部门和领导,辅助管理决策。

在这个分析过程中普遍存在下面一些突出问题。

- 业务系统压力大:业务系统数据量越来越大,从业务系统导出数据越来越慢,并且频繁的导出数据给业务系统带来巨大压力,甚至已经影响到业务系统正常使用。
- 手工报表耗时长:导入数据,再加上 Excel 数据加工时间,通常出个周报、月报最少需要 1~2 天时长,耗时太长。
- 数据不统一:每个部门都有分析人员,大家都在各自出报表和做分析,数据和分析结果难于共享,并且汇总的数据也常常容易产生矛盾,数据不统一。
- 决策难于深化:大量时间在做数据处理,没时间分析,分析发现问题之后难



于做相关分析,需要持续投入大量时间,最终造成员工很忙、领导很急的状况。

- 在电子商务行业,其核心数据还是一些日志,比如访问日志、搜索日志等,这些数据蕴藏着丰富的用户行为,但其特点就是数据量大,利用 Excel 来处理数据、分析数据已经完全不可能,所以这时候分析往往只能退而求其次,比如通过用户量这些小数据来近似把握运营情况,从而企业要做到精细化运营、精确化服务基本无从谈起。

这些问题在业务系统运营之后都存在,并且随着时间的推移,特别对于快速发展的公司来说,随着数据量的激增,问题更加明显。

2) 企业精细化管理需要 BI 支撑

精细化管理的核心思想还是快、精、准,这几个基本要求都需要 BI 的强力支撑。

- 快,要求自上而下的实时把控,第一时间发现问题需要 BI 支持。特别是高层领导,对整个公司的重大问题都需要在发现问题的第一时间就能够把各种流程控制系统里面的业务数据提取出来进行进一步的分析,发现导致问题发生的异常点,以便及时处理问题。

- 精,能对问题追根溯源,需要 BI 的向下钻取、向上汇总、交叉分析、关联分析等基本技术支撑,否则发现问题也不知道原因在哪里。

- 准,要求问题落实到人,问题出在谁身上,谁该受到惩罚,谁该进行工作改进,这些问题需要 BI 对组织架构进行多层级分析。

3) 数据里面的知识可以帮助企业进行优化升级

数据—知识—操作—数据这样一个信息闭环其实就是实践—总结—再实践的一个螺旋式上升过程,在这个过程中如果缺少数据到知识的一步那就是简单重复的操作,而加入知识总结这样一个分析之后的实践,则是有提升的实践,其必然结果是促进我们工作不断得到修正和优化。

4) 知识还可以产品化

在电子商务行业,BI 用处还不太一样,传统行业的 BI 主要集中在企业内报表和分析,而电子商务行业不论是 B2B 还是 B2C,BI 的主要用处还是支持网站用户的需求。知识可以提升网站的友好性,让网站设计者了解用户的习惯和行为,设计符合用户操作的流程和功能;根据买家的行为提供个性化的商品推荐可以帮助用户提升工作效率和采购效果,提升他们对网站的认可度;另外对于卖家来说在网站进行推广之后,他们迫切希望知道有没有效果,改进机会在哪里。

这些需求通过 BI 从数据里面提取知识,这些知识将提升网站对买卖双方的支持能力,在这个过程中也可以对部分知识进行产品化,从而为电子商务企业在搜索排名等功能性收费之外创造第二条盈利模式——知识收费,这点淘宝做得比较充



分,数据魔方是其典型代表。

1.3 商业智能发展历史和国内应用现状

1946 年第一台电子计算机问世,这意味着信息时代即将来临。

1946 年人类历史上第一台电子计算机问世,随之而来的是全球信息化的到来,计算机在世界的各行各业扮演着越来越重要的角色,信息从纸质的书本、文档中转移到计算机中,并通过网络进行共享、互动。

20 世纪 80 年代,各行各业实施了众多的信息系统,提高了企业的信息传播速度,提升了工作的效率,随着信息系统的逐步完善、成熟,企业不仅仅满足于计算机带来的流程的便捷和信息的积累,他们想利用信息系统中的信息来辅助高层决策,这就是商业智能系统产生的原因。

20 世纪 90 年代,数据库技术的高速发展,使得多维分析工具逐步进入商用,打破了商业智能技术上的瓶颈,随后,其发展取得了前所未有的加速度,并在 21 世纪的第一个十年蓬勃发展。

过去 10 年电子商务在国内一样大获成功,全面进入规模企业,也必将在未来有更深入的应用,因为 BI 的厂商和核心技术基本在国外,这里我们主要从应用角度来总结了一下 BI 的发展情况。

2000 年至 2004 年,国内 BI 应用还是初期阶段,主要是一些国有大型企业在进行尝试,其中较为明显的成果是建立起数据仓库,OLAP 分析也获得一定应用,特别是出报表方面已经凸显 BI 的价值;在这个时期通过一些先行者的 BI 实施带来了 BI 整体的解决方案和实施技术,为后期 BI 快速推广积累了技术和人才基础,其中一些先行的软件外包公司形成了强大的 BI 实施能力;这个时候国内规模企业也逐步开始接触 BI,慢慢知道了什么是数据仓库、ETL、OLAP、数据挖掘等;BI 方案虽然在这个阶段比较成熟,但产品还是参差不齐,总体来讲数据仓库及 ETL 基本可以满足需要,而 OLAP 工具实施效果仍然是 BI 项目最大风险,数据挖掘基本还处于概念阶段。

2004 年至 2008 年国内 BI 进入全面发展及应用阶段。基本上有一定规模的企业都开始实施 BI,跨国公司的中国本部也开始实施本土 BI;OLAP 分析在这个阶段达到比较高的水平,OLAP 产品都有不同程度的升级,功能上基本能满足用户的需求,虽然性能还是一个瓶颈,但 BI 项目经过合理设计及实施普遍对企业分析效率和分析能力有较大提升,一些领先企业还实现了企业自上而下的全面 BI 分析应用,高层仪表盘,中层 OLAP 分析,基层工作的监控、预警及考核;另外数据挖掘在这个时期得到切切实实应用,在银行、通信、大型网站等海量数据行业的企业为了



自身安全或者核心竞争力的塑造都开始挖掘上有投入,并且取得了不少成果,这些经典案例也为接下来数据挖掘在国内推广及应用打下基础,让数据挖掘从概念性阶段进入实实在在应用阶段。这个时期在国内 BI 应用中还必须提及一些世界级的软件公司,国外有 IBM、HP、Bearing Point 及印度几大软件外包公司等,国内有东南融通、文思创新、华为等都在 BI 方面提供强大的 BI 解决方案和实施能力,为国内的 BI 普及及应用提供了技术上的保证。

2008 年之后 BI 应用开始往系统化发展,不仅注重知识的提取,更注重知识的应用及系统化封装。这个时期 BI 应用产生了新的分支,即分析型流程定制(Analytic Processing),其目标就是在工作流程中加入分析支持,提升业务操作的能力和效率,实现分析结果的流程化应用。2008 年后数据挖掘开始得到深入应用,2000 年之后各大院校及培训机构培养的数据挖掘人才在这个阶段充分得到机会,特别是在电子商务等大数据行业,数据挖掘的应用已经非常普遍。在产品上也呈现出几大趋势,BI 原始厂商纷纷被收购,BI 产品和技术处于大融合阶段,几大厂商都能够提供整体 BI 解决方案,特别是 IBM。技术上最大机会还就是内存 OLAP 技术和云计算,前者让 OLAP 体验更佳,后者让 ETL 及查询更快。

2010 年之后的 10 年内笔者认为主要还是云计算带来技术架构升级和基于成熟 BI 应用的企业级管理提升。云计算让企业海量数据分析不再是挑战,也将给我们数据仓库、ETL 设计带来技术上的升级,现在流行的 Hadoop 技术就是一个例子。技术创新带来的管理变革是必然,不管是企业基于绩效管理的内控,还是基于分析挖掘的业务执行能力都将得到系统化的提升,BI 技术系统化的必然结果就是实现执行—数据—知识—优化执行的信息闭环,在管理上不仅信息获取更及时,甚至实时,更能提升我们决策分析的能力,让知识来指导和帮助我们提升工作效率。

1.4 商业智能相关技术介绍

虽然 BI 产品过去 10 年让人眼花缭乱,但从基本技术来讲还是比较稳定,基本可以分为数据仓库、ETL、OLAP 和数据挖掘。

1) 数据仓库

数据仓库(Data Warehouse,简称 DW),就是面向主题的、集成的、相对稳定的、随时间不断变化(不同时间)的数据集合,用以支持经营管理中的决策制定过程、数据仓库中的数据面向主题,与传统数据库面向应用相对应。主题是一个在较高层次上将数据归类的标准,每一个主题对应一个宏观的分析领域,数据仓库的集成特性是指在数据进入数据仓库之前,必须经过数据加工和集成,这是建立数据仓库的关键步骤,首先要统一原始数据中的矛盾之处,还要将原始数据结构做一个从

面向应用向面向主题的转变;数据仓库的稳定性是指数据仓库反映的是历史数据,而不是日常事务处理产生的数据,数据经加工和集成进入数据仓库后是极少或根本不修改的;数据仓库是不同时间的数据集合,它要求数据仓库中的数据保存时限能满足进行决策分析的需要,而且数据仓库中的数据都要标明该数据的历史时期。

数据仓库最根本的特点是物理地存放数据,而且这些数据并不是最新的、专有的,而是来源于其他数据库的。数据仓库的建立并不是要取代数据库,它要建立一个较全面和完善的信息应用的基础上,用于支持高层决策分析,而事务处理数据库在企业的信息环境中承担的是日常操作性的任务。数据仓库是数据库技术的一种新的应用,而且到目前为止,数据仓库还是用关系数据库管理系统来管理其中的数据。

2) ETL

ETL(Extract-Transform-Load 的缩写,即数据抽取、转换、装载的过程)作为 BI/DW(Business Intelligence)的核心和灵魂,能够按照统一的规则集成并提高数据的价值,是负责完成数据从数据源向目标数据仓库转化的过程,是实施数据仓库的重要步骤。如果说数据仓库的模型设计是一座大厦的设计蓝图,数据是砖瓦的话,那么 ETL 就是建设大厦的过程。在整个项目中最难部分是用户需求分析和模型设计,而 ETL 规则设计和实施则是工作量最大的,占整个项目的 60%~80%,这是国内外从众多实践中得到的普遍共识。

ETL 是数据抽取(Extract)、清洗(Cleaning)、转换(Transform)、装载(Load)的过程,是构建数据仓库的重要一环,用户从数据源抽取出所需的数据,经过数据清洗,最终按照预先定义好的数据仓库模型,将数据加载到数据仓库中去。

信息是现代企业的重要资源,是企业运用科学管理、决策分析的基础。目前,大多数企业花费大量的资金和时间来构建联机事务处理(OLTP)的业务系统和办公自动化系统,用来记录事务处理的各种相关数据。据统计,数据量每两到三年时间就会成倍增长,这些数据蕴含着巨大的商业价值,而企业所关注的通常只占总数据量的 2%~4%。因此,企业仍然没有最大化地利用已存在的数据资源,以至于浪费了更多的时间和资金,也失去制定关键商业决策的最佳契机。于是,企业如何通过各种技术手段把数据转换为信息、知识,已经成了提高其核心竞争力的主要瓶颈。而 ETL 就是主要的一个技术手段。

3) 多维分析

多维分析(OLAP-Online Analysis Processing,联机分析处理)是一种获取、分析数据的技术,是指根据业务需要,从数据库中提取相关业务数据,并且对所提取的数据进行不同角度(指标),不同深度(维度)的分析,最终形成多维业务报表,并根据报表中数据的客观规律,发现数据中的趋势和异常,给决策者提供科学的决策



依据。大家常见的 Excel 数据透视表就是比较简单的 OLAP 分析工具。

4) 数据挖掘

数据挖掘,在人工智能领域,习惯上又称为数据库中的知识发现(Knowledge Discovery in Databases, KDD),也有人把数据挖掘视为数据库中知识发现过程的一个基本步骤。知识发现过程由以下三个阶段组成:①数据准备;②数据挖掘;③结果表达和解释。数据挖掘可以与用户或知识库交互^[3]。

数据挖掘是通过分析每个数据,从大量数据中寻找其规律的技术,主要有数据准备、规律寻找和规律表示三个步骤。数据准备是从相关的数据源中选取所需的数据并整合成用于数据挖掘的数据集;规律寻找是用某种方法将数据集所含的规律找出来;规律表示是尽可能以用户可理解的方式(如可视化)将找出的规律表示出来。

数据挖掘的任务有关联分析、聚类分析、分类分析、异常分析、特异群组分析和演变分析等。并非所有的信息发现任务都被视为数据挖掘。例如,使用数据库管理系统查找个别的记录,或通过因特网的搜索引擎查找特定的 Web 页面,则是信息检索(information retrieval)领域的任务。虽然这些任务是重要的,可能涉及使用复杂的算法和数据结构,但是它们主要依赖传统的计算机科学技术和数据的明显特征来创建索引结构,从而有效地组织和检索信息^[4]。尽管如此,数据挖掘技术也已用来增强信息检索系统的能力。

1.5 小结

商业智能,又称商务智能,英文为 Business Intelligence,简称为 BI。商业智能指综合利用数据仓库、ETL 技术、OLAP 分析以及数据挖掘技术对数据进行有效的整合和存储,并对数据进行分析,提取出其中蕴藏的知识,从而帮助客户进行决策分析。

业务系统的完善使操作人员在工作流程化规范化方面取得了长足进步,然而随着时间的推移,业务系统积累了大量的数据,这些数据包含了对管理决策和优化工作流程的信息。然而,使用传统的手工报表方式,却难以从业务系统的数据中完整和准确地获取有用的信息,这样既不能帮助企业深化决策,也不能支持企业的精细化管理,而这一切将在引入商业智能以后得以有效解决。商业智能对数据进行知识化提取是企业优化升级的必然需求,并且对于电子商务行业来说,BI 对于买家和卖家的支持可以提升网站的客户满意度,甚至于利用 BI 技术开发出数据分析产品,帮助企业开创新的盈利模式。

2000 年至 2004 年,是国内商业智能应用的初期,主要是建立起数据仓库和