

黄建鹏/徐晓冬/魏宝军著



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

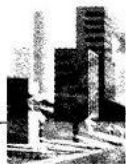
I. ①商… II. ①黄…②徐…③魏… III. ①电子商
务—研究 IV. ①F713.36

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。

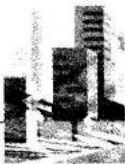
第六章、第七章对目前焦点科技实施BI以来的一些经验进行总结，介绍了BI成功实施的几个关键点，BI在焦点科技的应用展望。

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。

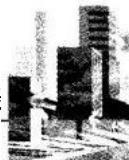
3.IT 管理人员：IT 领导一定会或多或少了解 BI, 那么本书对 IT 领导最大的价值就是其实战性和中立性。内容包含大量 BI 核心技术和实施方法论，图示丰富，如果有一定IT 基础，浏览本书后基本上可以对 BI 了解了七八成；另外书中没有太多的具体软件，更注重 BI技术的本质问题和解决思路，这有助于IT 领导们选择适合自己公司的BI产品。这里不得不说的是虽然现在BI 各大厂商看起来都很强大并且差异不大，但事实上项目成功率、实施成本相差很大。

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、肖洁琴、夏平嵩、丁艳艳、胡春香、王婷、房鹏展、余华等。



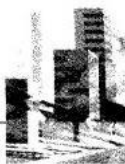
目 录

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

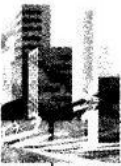


3.5.2	主要应用	32
3.5.3	应用实现要求	33
3.5.4	常见问题	34
3.6	BI在电子商务领域的应用	35
3.6.1	电子商务运营决策分析	35
3.6.2	对外-支撑网站用户分析	36
3.7	软硬件方案	38
3.8	实施策略及风险控制	40
3.8.1	自上而下策略	41
3.8.2	自下而上策略	43
3.8.3	整体规划、逐步实施	44
3.9	BI项目风险控制	46
3.9.1	组织风险	46
3.9.2	技术风险控制	47
3.9.3	业务风险	48
3.9.4	实施风险	49
2 3.10	小结	50

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, 简称为 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, 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, 简称为 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]。

OLAP 是联机在线分析(On-line Analytical Processing)]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) 急切的分析型需求

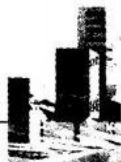
中国经济经历90年代以来快速发展的20年，涌现出了一大批有财力、有活力的公司，特别是改制后的垄断企业、沿海知名民营企业整体建设都逐步向国际领先企业靠拢，在2000年前后都陆续积极建设核心业务系统，电讯行业的计费系统、生产制造的MES、零售分销行业的ERP 系统等都基本在这一时期快速得于实现，这些系统让操作人员在工作流程化规范化方面取得了长足进步，资金流得到严格控制和监管。

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，从业务系统查数据导出数据越来越慢，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。

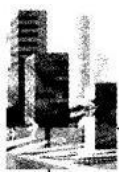
3) 数据里面的知识可以帮助企业进行优化升级

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。

5

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



OLAP 分析的初级应用，数据挖掘应用基本还处于概念阶段。2004年至2008年，国内BI进入全面发展和应用的时期，OLAP 分析达到了比较高的水平，数据挖掘在银行、通信、大型网站等海量数据行业取得了许多切实的应用成果。2008年之后BI应用开始往系统化发展，并且产生了新的分支即分析型流程定制(Analytical Processing)，此时数据挖掘也开始得到了广泛和深入地使用。笔者认为2012年之后的10年内，云技术将带来 BI 技术架构升级和基于成熟BI应用的企业级管理提升。

商业智能涉及的相关技术有数据仓库、ETL 技术、多维分析和数据挖掘：

数据仓库就是面向主题的、集成的、相对稳定的、随时间不断变化(不同时间)的数据集合。

ETL 是数据抽取(Extract)、清洗(Cleaning)、转换(Transform)、装载(Load)的过程。

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。

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。



第二章 BI 在电子商务的应用前景

2.1 电子商务发展对商业智能的必然需求

在这个信息科技和互联网逐渐深入生活的时代，电子商务的发展在消费拉动中已经成为了举足轻重的角色。经过早期粗放式的电子商务发展，企业该如何充分利用电子商务的第二个发展阶段，在消费经济转型的过程中争取更优势的地位呢？

中国制造网的买家群体和卖家群体每年都呈几何规模的数字增长，很多新的问题不断地涌现，急需电子商务从业者来解决，以B2B行业来看，主要有如下的一些问题：

- (1)B2B 盈利模式单一，大部分依靠收取会员费和广告费。
- (2)服务信息反馈渠道单一，仅仅依赖服务人员和客户的沟通。
- (3)平台同质化现象严重，需要新的产品服务客户。
- (4)买卖双方沟通渠道缺乏，效率低下。

目前，各大电子商务企业主要依靠提高用户体验，升级搜索引擎，应用商业智能来解决这些问题。商业智能作为一个很重要的解决方案，目前在电子商务企业发挥着越来越大的价值。

阿里巴巴、淘宝和中国制造网纷纷推出各自的数据分析产品，目前的“数据魔方”、“量子统计”和“数据罗盘”都有了各自稳定的用户，部分还实现了在线销售，为B2B平台只收取平台服务费这一单一盈利点提供了一个新的盈利模式。下节将着重介绍商业智能在电子商务行业中的应用范围及产生的价值。

2.2 商业智能在电子商务的应用情况及价值

电子商务行业依托于电子信息和网络技术形成了一种新的商业模式，电子商务本身就是信息化时代的产物，它传递信息、推广信息，同时也利用信息。商业智能综合利用数据仓库、OLAP 技术和数据挖掘技术实现从数据转化为信息，帮助企业利用信息为企业运营和管理提供分析和决策的依据。

1) 商业智能在电子商务平台优化中的应用

电子商务企业依托于网站平台为客户提供各种服务，客户在平台上可以搜索、浏览以及进行询盘和交易。合理的页面布局和业务流程能够让用户快速找到对自己有用的信息，并且能够适当增加用户在网站上停留的时间。商业智能通过网站的流量分析与用户行为分析，监控网站各功能的效果，为合理优化网站功能模块，提供依据。比如通过分析用户的访问日志，可以分析出用户普遍的网站浏览习惯和偏好，从而完善和优化网站平台的结构，获得更好的用户体验。

2) 商业智能在产品优化及创新中的应用

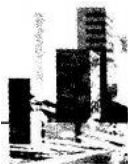
对于竞争激烈的电子商务企业来说，产品就是企业的生命。如何能够设计出一个用户满意的产品，已经不能完全凭借需求人员的直觉和行业经验，更多的还是需要数据挖掘和分析的支持，例如产品的结构设计，产品如何定价，以及该产品适用人群分析等。并且每一产品都有其生命周期，一个过去很好的产品现在可能已经过时了，所以对于老产品的效果分析与挖掘是必不可少的，怎样正确的对当前产品的效果做出客观评价，同时对产品将来的发展趋势作出准确的预测都需要商业智能提供技术和数据的支持。

3) 商业智能在客户分析中的应用

电子商务企业进行客户分析，一方面可以让自己了解客户，另一方面可以帮助客户了解客户的客户。挖掘客户的兴趣爱好，可以为客户进行个性化的服务和产品推送，提升用户的满意度。利用聚类分析等数据挖掘技术可以根据客户的属性进行客户细分，从而定制精确化营销的方案。电子商务平台还可以利用自身掌握的信息服务客户，比如分析供应商的买家分布、买家喜好等信息，为供应商提供推广和营销的数据支持，不仅可以提升供应商对于网站的黏性和满意度，同时可以为企业创造新的赢利点。

4) 商业智能在交易欺诈及网络安全中的应用

网络欺诈、钓鱼等异常行为的发生不仅会使用户承受损失，也会降低电子商务平台本身的价值和客户满意度。网络交易安全是电子商务行业良好发展的重要保障。运用数据挖掘技术，可以帮助企业发现网络钓鱼、网络诈骗等异常事件，通过关联分析、分类等挖掘方法识别和预测异常用户。比如网络反欺诈，如果依靠人工核查，不仅需要消耗大量的人力与物力，而且大多是在发生欺诈行为之后才能发现。利用数据挖掘技术，可以对历史欺诈会员的行为特征进行提取和总结，从中提炼出欺诈分类规则，建立数据挖掘模型形成欺诈预警系统，不仅能帮助审核人员快速锁定目标，并且能对异常用户进行预测，能够在欺诈行为开始之前进行预警，做到事前发现，从而有效防止欺诈的产生。



2.3 小结

电子商务企业想要在消费经济转型的过程中争取更优势的地位，必须解决如下问题：①B2B盈利模式单一，大部分依靠收取会员费和广告费；②服务信息反馈渠道单一，仅仅依赖服务人员和客户的沟通；③平台同质化现象严重，缺少新的产品服务客户；④买卖双方沟通渠道缺乏，效率低下。商业智能可以帮助企业提供以上问题的解决方案，首先通过网站的流量分析与用户行为分析，监控网站各功能的效果，为合理优化网站功能模块提供依据；其次，利用BI技术，可以分析出产品的结构设计、定价以及适用人群，并且评价产品效果，帮助产品创新；第三，进行客户分析，一方面可以让企业了解客户，另一方面可以帮助客户了解客户的客户，提升沟通的效率，并且BI对客户的支持可以发展成企业新的盈利模式；最后，运用数据挖掘技术，可以帮助企业发现网络钓鱼、网络诈骗等异常事件，保障客户的利益和平台的良好发展。

第三章 商业智能在电子商务的解决方案

电子商务行业和传统的制造业、零售业无论从商业模式、组织架构等各个方面都有着很大的区别。早期，很多BI咨询公司或者实施公司把原先在传统行业的解决方案直接搬到电子商务行业，企图复制在传统行业的成功，最后发现，不管是项目实施的成功率还是项目后期的效果都非常的不理想，从而让整个电子商务行业对于此行业是否适合利用BI产生了很大的质疑。本章将结合焦点科技实施 BI 以来的经验，系统讨论适用于电子商务行业的 BI 整体解决方案。

3.1 系统架构

传统行业的BI解决方案中核心的数据仓库架构设计如图3.1所示：

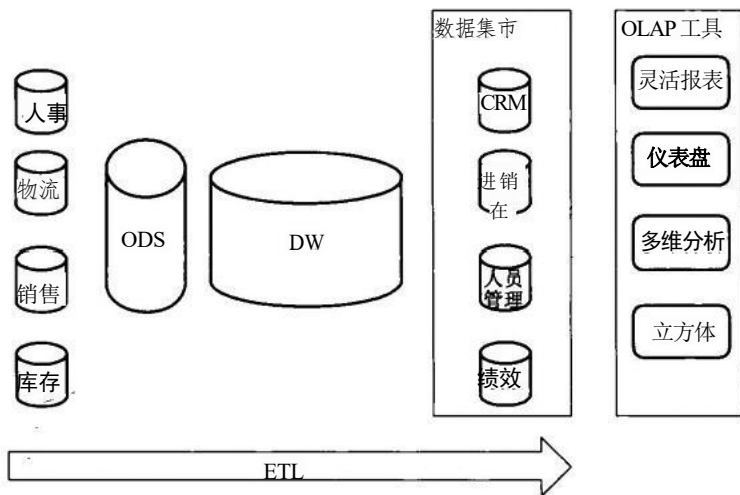
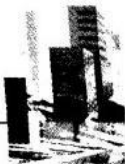


图3.1 传统行业数据仓库架构示意图

这个数据仓库架构移植到电子商务行业就会产生严重的“水土不服”现象，通过总结，发现是下面几种情况限制了整个架构发挥正常的作用。

(1) 电子商务行业数据源多种多样，如各种关系型数据库、社交产品产生的数据、网络日志数据等。单一的 ETL 工具很难很好的处理所有数据源。



(2) 电子商务数据量巨大。“大数据”这个词的流行就是因为目前各类网站产生了巨大的网络日志，并且各个公司在处理这么大规模的数据时都遇到了很大的问题。传统的关系型数据库目前来看已经无法满足处理如此大规模数据的要求。

(3) 电子商务对于数据分析的实时性要求极高，有很大部分的数据只在短时间内才能体现数据的价值，处理数据速度的快慢就显得尤为重要，传统的离线 ETL 处理方式已经无法满足这个需要。

(4) 用户量巨大，电子商务的 BI 应用很大一部分是面向网站的用户群体的，这个群体数目巨大。传统的 BI 展现工具设计的时候是为了服务企业的中高层用户，移植到电子商务行业，无论从用户体验还是交互效率都存在巨大差距。

以上四个问题就是传统的 BI 解决方案在电子商务行业屡屡失败的根本原因，根据焦点科技两年以来的 BI 实施经验来看，这些问题都是可以解决的。数据仓库架构的设计本质上都是和传统行业的架构是一样的，只是其中的一些解决方法和技术有所不同，具体架构设计如图 3.2 所示：

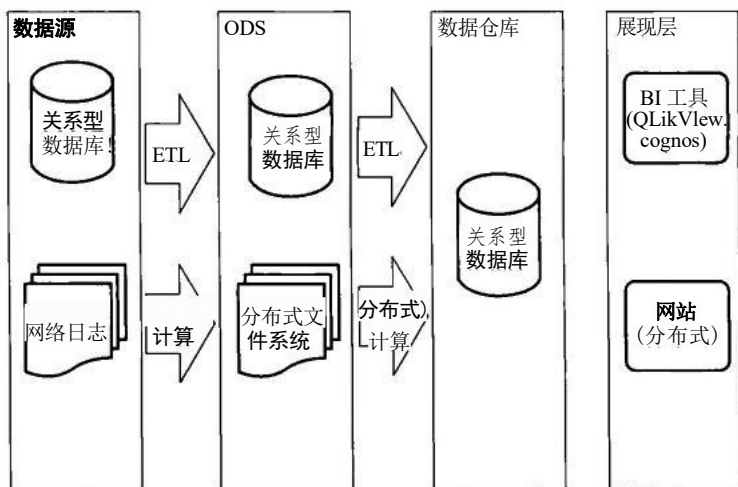


图 3.2 电子商务行业数据仓库架构示意图

电子商务企业实施 BI 不能简单的照搬传统行业的 BI 实施经验，要在理解数据仓库架构本质的基础上根据行业的特点进行不同技术的采用和架构的升级。本章将详细论述在电子商务行业实施 BI 的成功经验。

3.2 数据仓库及模型

与传统商业模式不同，对于电子商务行业来说，WEB 用户与网站信息提供者

不能进行直接的信息交流和意见反馈。而经营者非常希望了解到用户喜欢什么样的信息,不同用户对信息的不同需求等这些作为网站发展决策的重要依据。因此建立数据仓库成为企业管理决策支持系统的基础。数据仓库通过有效地组织和存储数据,把其内部隐藏的信息转化为商业价值,可以帮助企业从海量数据中找到真正有用的决策信息,为提升企业效益提供服务,解决决策者们迫切关心的问题。

3.2.1 数据仓库的定义

数据仓库并不是为了存储数据,而是为更好地利用企业内所有可能收集到的数据进行决策支持。数据仓库的概念是 W.H.Inmon 在其《数据仓库》一书中提出的,他指出“数据仓库是面向主题的、集成的、具有时间特征的、稳定的数据集合,用以支持经营管理中的决策制定过程”[5]。简单地讲,数据仓库就是企业内部一个专门的、大型统一的数据存储系统,支持快速、灵活、有效的分析型数据查询。

数据仓库的基本功能包括数据抽取、数据筛选和清洗、数据加载、建立数据集市、完成数据仓库的查询、决策分析和数据挖掘等操作。

3.2.2 数据仓库的架构

1) 数据仓库的设计思路

对于传统 BI 来说,数据仓库主要使用关系型数据库来支持离线分析和复杂查询。而随着电子商务行业的发展,网站需要立即对用户的网站行为进行在线分析,比如显示某个到访顾客的所有历史来访记录,同时实时跟踪显示某个访客在一个店铺正在访问的页面等信息,而关系型数据库的查询性能难以满足这样的需求。内存数据库(如 HBase 等)的数据处理速度比传统数据库的数据处理速度要快很多,一般都在10倍以上。因此在线分析,实时 ETL 和多用户在线等对于数据查询和处理效率要求较高的需求可以采用内存数据库来完成。

同时,电子商务行业会面临越来越多的比如网站日志、用户行为这样的大数据的处理,还要进行复杂的数据挖掘,还有半结构化数据的处理,因此在数据仓库系统中引入分布式数据库成为必然的趋势。比如,以 Hadoop[6]为代表的海量数据处理开源工具无疑是吸引人的。

综上所述,对于数据仓库的设计思路如图3.3所示,我们可以结

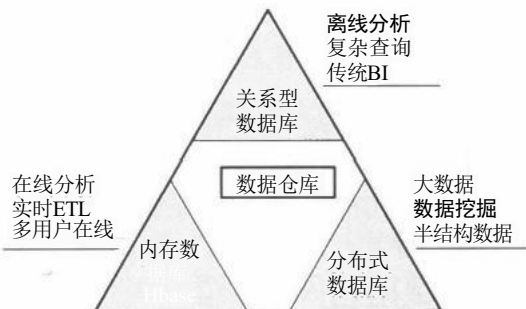
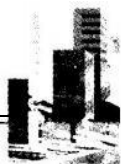


图3.3 数据仓库的设计思路



合使用关系型数据库、内存数据库和分布式数据库以满足电子商务行业对于商业智能在数据处理、数据存储等方面的要求。

2) 企业级数据仓库架构

数据仓库的目的是构建面向分析的集成化数据环境，为企业提供决策支持(Decision Support)。数据仓库的数据来源于外部，并且开放给外部应用。数据仓库按照数据流向可以分为三层结构，包括数据层、信息层和分析层(如图3.4所示)。

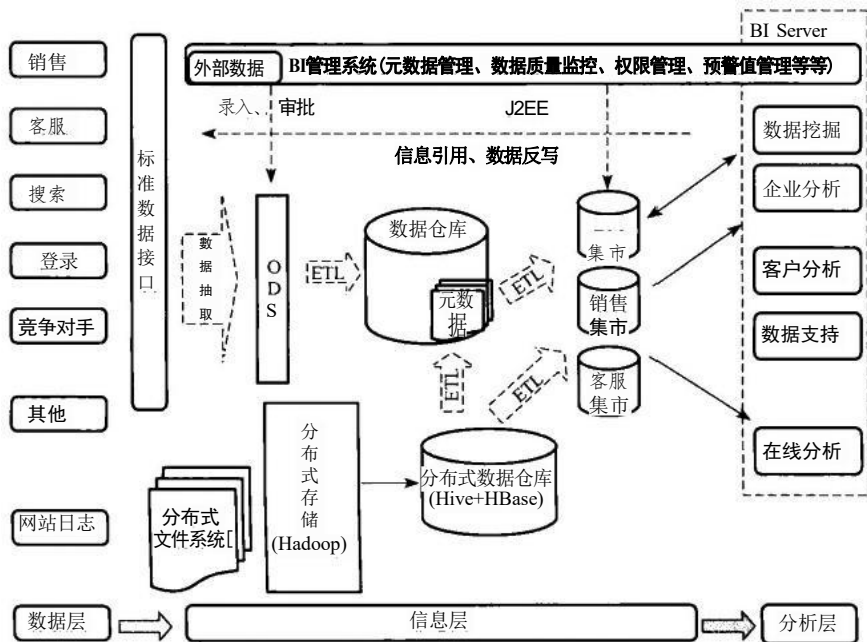


图3.4 企业级数据仓库架构

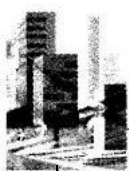
(1) 数据层

数据层主要是数据源和标准数据接口。

数据源：对于电子商务行业来说，数据源包括网站日志、用户网站行为(如登录、搜索等)、竞争对手数据以及其他记录网站运营(如客服、销售等)的数据等。

标准数据接口：标准数据接口主要是为了标准化数据，将源数据按照BI的数据类型和存储结构进行标准化，从而利于BI应用。BI用于决策分析的数据大多都存在于企业的其他业务系统中，为此其他系统都需要提供数据接口给BI系统，支持其源数据。通常出于对源业务系统的安全和对BI的稳定的考虑，我们对数据接口有如下基本要求：

- 通过数据接口，BI抽取数据所施加给业务系统的压力要减到最小，要充分评估数据接口对业务系统的安全性。



- 数据接口可以是数据库的形式，也可以是文件的形式，常用的文件形式有 XML、TXT 等格式，不管哪种格式，都要求数据接口有实际的数据落地，并且和业务系统不同一个库，甚至不同一台服务器。

- 数据接口要求稳定，不随业务系统变化而随意变化模型，数据接口的标准和一切数据模型由 BI 系统实施人员决定；数据接口数据最好采用推送的方式到达，完全受控于业务系统代码，而不是 BI 系统人员开发代码到业务系统任意进行抓取。

- 数据接口有一定的更新周期要求，通常是一天更新一次。

- 数据接口可以采用和源业务系统完全一样的模型，也可以是中间表，一般建议企业级 BI 系统实施的时候数据接口保持为业务系统表结构。

(2) 信息层

信息层包括 ODS、数据仓库、数据集市、ETL、元数据。

ODS: 操作数据存储(Operation Data Storage)主要用途是将多个数据源的数据集成到一个临时缓冲区中供数据仓库使用。ODS 是数据源和目标数据仓库之间的缓冲库，能有效减轻数据源的压力和 ETL 的压力，其表结构和数据接口表结构尽量保持一致。比如：我们从源数据向数据仓库中加载事实表数据时，这时候我们需要进行聚合操作，如果没有 ODS 层，那么所有聚合操作的压力是在源系统形成的，这就会给客户源系统带来很大的压力，这是在项目实施过程中经常遇到的一个问题。

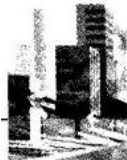
通常 ODS 可以包括下面三层。

- 映射层：把所有要用到的数据源表的所有字段映射到本地数据库，实现数据的落地，对于外部数据也可以通过系统管理到这一层，实现所有数据的集中。

- 数据预处理层：在这一层可以进行数据整合、数据筛选和增加关联表等预处理，所有的预处理都是为了简化和提升 ETL 过程。

- 数据清洗：对于有质量问题的数据进行数据清洗，清洗规则以业务部门定义为准。

数据仓库：如图 3.3 所示，我们建议数据仓库采用关系型数据库、内存数据库和分布式数据库的综合结构。关系型数据库用以支持传统 BI 数据需求。云技术的出现打开了大数据处理的新篇章，采用分布式数据库来支持 BI 大数据的存储与计算，可以显著的提升数据搜索的性能。在 Hadoop 分布式系统架构中，提供了一个数据仓库工具 Hive，Hive 本身建立在 Hadoop 的体系结构上，能很好地处理不变的大规模数据(例如网络日志)上的批量任务；但是，Hive 不提供在线事务处理和实时查询，这一切可以由 Hadoop 的另一个成员 HBase 来完成，HBase 为查询而生的，它通过组织起节点内所有机器的内存，提供一个超大的内存 Hash 表，它需要组织自己的数据结构，包括磁盘和内存中的，表在 HBase 中是物理表，而不是逻辑表，搜索引擎使用它来存储索引，以满足查询的实时性需求。



数据仓库是根据实际的业务主题逐步进行设计的，一般开始阶段都会选择核心业务进行设计。

数据集市(Data Mart,简称 DM):数据集市是面向特定的主题进行设计,根据主题业务的实际需要可以对仓库数据进行筛选、过滤、汇总、表合并等操作。数据集市的设计需要选择合适的集市粒度,但没必要保留最明细的数据,根据实际分析需要我们选择最为合适的粒度,只有这样才能最大限度提高性能和整体设计的合理性。

ETL:ETL 占据整个 BI实施最少50%的工作量。现在有大量的工具可以用来开发ETL 过程,也可以直接基于存储过程进行开发,各有优缺点。前者易于维护和管理,但是要对工具非常熟悉才能确保有比较好的性能;后者必须对所有开发人员进行培训,培训开发中的设计标准,这样才能避免因为过多存储过程代码而最终无法管理的局面,如果管理恰当可以带来较高的性能。

元数据:元数据(Meta Date),其实应该叫做解释性数据,即数据的数据。主要记录数据仓库中模型的定义、各层级间的映射关系、监控数据仓库的数据状态及ETL 的任务运行状态。一般会通过元数据资料库(Metadata Repository)来统一地存储和管理元数据,其主要目的是使数据仓库的设计、部署、操作和管理能达成协同和一致。

(3) 分析层

分析层主要是数据的应用,如数据挖掘,报表分析,数据支持等。

17

3.2.3 数据仓库模型

数据模型是对现实事物的反映和抽象,它可以帮助我们更加清晰地了解客观世界。在分析完业务需求之后,我们就需要建立数据仓库模型以反映业务逻辑。正确和完备的数据模型,是BI 进行一切数据分析的基础,是决定数据仓库项目成功与否的重要因素。

采用第三范式(3NF)建模对事务处理来说很有好处,但是对于数据仓库来说,要求模型能够提供较好的查询性能,因此数据仓库建模通常使用维度建模[。接下来,将介绍数据仓库维度建模的方法。

1) 数据仓库建模的原则

电子商务行业的信息系统具有业务复杂、数据量大、系统庞大等特点。电子商务行业的数据仓库模型设计应该遵循以下原则:

- 满足不同用户的需求

电子商务行业的BI用户包括“买家”、“卖家”以及平台供应商自己,平台供应商内部还可以分为不同的业务部门(比如客服、销售等),或者分为高层领导、中层领导和普通员工。不同层次、不同级别、不同岗位、不同立场的用户关注的信息不



同。数据仓库必须能够支持不同用户的信息需求。

- 兼顾查询效率与数据粒度的需求

过细的数据粒度会占用大量的存储空间，并且会降低查询速度。但是高度汇总的数据却不能保证数据分析的灵活性。数据仓库模型的设计需要同时兼顾查询效率与数据粒度。电子商务行业每天将会产生大量的日志数据，比如用户访问日志，如果都将这些数据保留在数据仓库中，将会消耗大量的存储空间。但是，对于数据分析来说，当前一个时间段内(比如最近一年内)的明细日志数据是有分析价值的，但是在这之前的日志数据可能我们只关心用户在每一天的点击量，那么可以将这些历史数据按天汇总后再存储，这样将会降低存储并且方便查询。

- 支持用户需求变化

用户的需求会随着业务或者市场的变化而变化。电子商务行业正处于高速发展阶段，同时也带来了激烈的竞争，用户的需求变化也会越来越频繁。数据仓库模型必须能够支持用户不断变化的需求。

- 避免业务运营系统性能影响

电子商务平台的用户可能来自不同的国家，在任何时候都会有用户访问平台，业务运营系统可能在24小时内都处于忙碌状态。数据仓库每天都需要抽取业务系统的数据，在设计数据仓库模型时应该考虑尽量减少和缩短对于业务系统的访问时间和资源占用，从而降低对业务系统性能的影响。

- 提供可扩展性

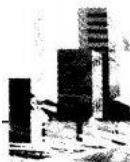
数据模型的可扩展性决定了数据仓库对新的需求的适应能力，建模既要考虑眼前的信息需求，也要考虑未来的需求。比如对于电子商务行业的竞争对手数据的分析，现在只有A企业是我们的竞争对手，一个月后新加入的B企业可能也成了我们的竞争对手，我们可以把A、B的数据都放在同一个表中，但这样显然不方便进行数据分析(比如数据对比)。如果再增加竞争对手C、D、E、...，这个表将会变得越来越庞大，更加不利于进行数据分析。在这种情况下，针对不同的竞争对手建立不同的模型，会更加方便数据存储和数据分析。

2) 星形模型和雪花形模型

目前常用的数据仓库数据模型为多维数据模型，这种模型主要以星形模式和雪花形模式在关系数据库系统中存在。

(1) 星形模型

星形模型是一种多维的数据关系，它由一个事实表(Fact Table)和一组维表(Dimension Table)组成，如图3.5所示。每个维表都有一个维作为主键，所有这些维则组合成事实表的主键，换言之，事实表主键的每个元素都是维表的外键。事实表的非主属性称为事实(Fact)，它们一般都是数值或其他可以进行计算的数据；而



维大都是文字、时间等类型的数据。

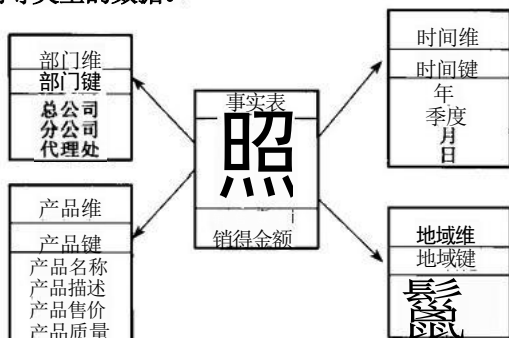


图3.5 星型模型

使用星形模式主要有两方面的原因：

- 提高查询的效率。采用星形模式设计的数据仓库的优点是由于数据的组织已经过预处理，主要数据都在庞大的事实表中，所以只要扫描事实表就可以进行查询，而不必把多个庞大的表连接起来，查询访问效率较高。同时由于维表一般都很小，甚至可以放在高速缓存中，所以与事实表作连接时其速度较快。

- 便于用户理解。对于非计算机专业的用户而言，星形模式比较直观，通过分析星形模式，很容易组合出各种查询。

19

(2) 雪花形模型

在实际应用中，随着事实表和维表的增加和变化，星形模型会产生多种衍生模型，包括星系模型、星座模型、二级维表和雪花模型。雪花模型是对星形模型维表的进一步层次化，将某些维表扩展成事实表，这样既可以应付不同级别用户的查询，又可以将源数据通过层次间的联系向上综合，最大限度地减少数据存储量，因而提高了查询功能。如图3.6所示，将地域维表分解为国家、省份、城市等维表。

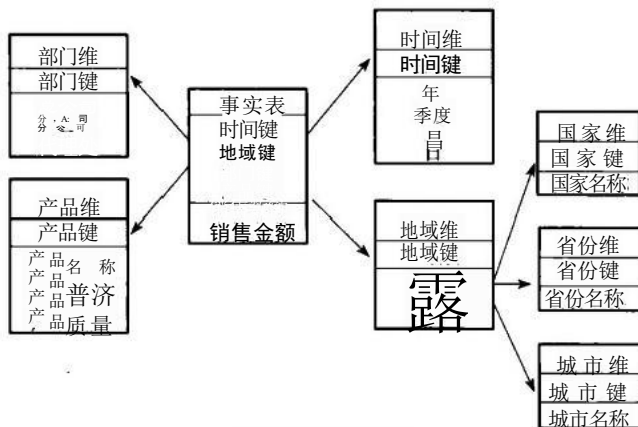


图3.6 雪花模型

雪花模型的维度表是基于范式理论的，因此是介于第三范式和星形模式之间的一种设计模式，通常是部分数据组织采用第三范式的规范结构，部分数据组织采用星形模式的事实表和维表结构。在某些情况下，雪花模式的形成是由于星形模式在组织数据时，为减少维表层次和处理多对多关系而对数据表进行规范化处理后形成的。

雪花模型的优点是：在一定程度上减少了存储空间；规范化的结构更容易更新和维护。同样雪花模式也存在不少缺点：雪花模式比较复杂，用户不容易理解；浏览内容相对困难；额外的连接将使查询性能下降。

在数据仓库中，通常不推荐“雪花化”。因为在数据仓库中，查询性能相对OLAP系统来说更加被重视，而雪花模式会降低数据仓库系统的性能。

3) 数据仓库的模型设计

数据仓库的设计以支持企业进行有效的BI实施为目的，企业BI对内、外部的用户提供统一的分析平台和有价值的信息与知识。以中国制造网为例，BI系统提供从市场分析、产品开发与推广、销售、客服和客户等全程分析与决策支持。我们建议以星形结构模型建立数据仓库。设计一个数据仓库主要要考虑如下方面：

(1) 主题设计

数据仓库是根据实际的业务主题逐步进行设计的，数据仓库的开发也有主题性，一般开始阶段都会选择核心业务进行设计。根据上文提到的业务点，进行数据仓库的主题划分，如图3.7所示：



图3.7 数据仓库主题划分

- 客户主题

客户主题下主要存储客户的详细信息，记录客户的基本属性以及注册信息等数据。客户信息数据体现最新的客户状态，当客户资料发生变化时，旧的客户信息将被新的客户信息更新掉。客户信息数据中保留每一个生命周期的客户的最新数据。

- 市场主题

市场主题下主要存储竞争对手数据，如竞争对手客户信息、竞争对手的产品目



录设置情况等。

- 产品主题

产品主题下存储客户的产品信息，包括产品的基本属性，产品信息数据体现最新的产品状态，当产品信息发生变化时，旧的产品信息将被新的产品信息更新掉。产品信息永久保存。

- 推广主题

推广主题下存储业务推广产生的数据信息，比如展会信息等。

- 销售主题

销售主题下记录产品(这里的产品指网站提供的服务)的销售情况，销售人员的业绩等信息，主要存储了销售人员详细信息、销售合同信息、销售合同产品信息、销售订单信息、服务信息表、销售回款信息等数据。

- 客服主题

客服主题下主要存储了客服客户服务信息、客服工作记录等数据。

(2) 事实设计

事实表是历史数据的事实，不具有在业务上重复出现的性质，比如销售数据，事实表更多的是体现主题特征。事实表是星形结构的核心，记录主题的主干内容。比如销售订单数据，主要内容有日期、订单ID、客户、订单状态、订单金额等。订单ID为主键，订单金额为事实表的主干信息，其他字段为维度。

(3) 维度设计

维度表是存放分析角度数据的表，维度数据具有高重现性，比如产品数据维度表经常为不同主题所共用；根据以往经验每个企业维度都不会太多，每个主题虽然都可能有自己特征的维度，但是公共维度占大部分。常用的公共维度有时间维度(年、月、日、周等)、产品数据维度(产品ID、公司ID、产品名称、产品状态等)、客户信息维度(公司ID、公司名称、公司所在地区、公司联系人、公司主营行业)等。

(4) 代理关键字

维表关键字只能用代理关键字。使用代理关键字可以避免不同数据源数据的统一性问题，因为不同数据源的数据可能缺乏统一的关键字编码；代理关键字可以处理缓慢变化维，因为代理关键字可以描述历史变化信息，区分同一维度值ID在属性变化前后的区别。例如员工维表，某员工变换部门，从销售一部换到销售二部，那么在表中就会记录下该人员的两条信息，该员工属于销售一部的记录作为历史记录，该员工属于销售二部的记录作为当前记录。

3.3 ETL

数据由各种系统数据库、文件、网页等存储方式进入数据仓库需要经过抽取

(Extraction)、转换(Transform) 和装载(Load) 这样一个过程, 简称 ETL, 这个过程非常具有挑战性, 但是只要能认识 ETL 的本质和核心问题, 还是能把 ETL 简单完成。

焦点科技的 ETL 完全基于数据库存储过程进行开发, 在这个过程中非常重视完整的 ETL 设计思想、方案、架构及技术标准, 下面我们从各个角度来介绍焦点科技 BI 人员对 ETL 的设计及开发的认识, 从中大家或许可以领略一点技术的魅力。

3.3.1 ETL 目标和基本要求

简单点理解 ETL 的目标就是保证数据仓库、数据集市数据的及时性和准确性, 从而保证我们各种分析需求的数据要求。基本要求包括:

(1) 数据完全准确。数据完全准确是 ETL 的第一要求, 数据不准的系统没人敢用, 不准的数据甚至会带来巨大危害, 毕竟高层领导的分析和战略决策都需要 BI 数据来支撑。

(2) 更新及时。及时更新数据会给 BI 带来很好的用户体验, 但不太现实, 因为 ETL 运算量过于庞大, 所有 ETL 都实施或者频繁更新难于实现, 但是对于实时性要求比较高的分析, 提高更新频率是有必要的; 对不同需求有不同的更新策略就非常有必要了。

(3) 系统稳定。区别一般系统稳定性这里有两层含义, 除了服务不能停之外还要求 ETL 在更新数据的时候不会经常中断, 如果中断也能快速恢复。

(4) 对其他系统影响小。虽然每天要抽取的数据很多, 但是任何系统都不能容忍对自身影响太大的请求, 快速从业务系统获取数据是 ETL 第一步要解决的问题。

当然如果能做到如下两条, 那更为完美, 焦点科技这方面已经达到。

(5) 代码简单并且能够标准化。

(6) 维护简单。

3.3.2 ETL 的开发

1) 企业级 ETL 开发基本思路

• ETL 包的分解: 企业级 ETL 开发首先要做的事情是 BI 项目和企业业务主题规划, 然后按不同主题来划分 ETL 包, 最后得到分主题的设计 ETL 包, 包之间数据可以单向依赖, 不能互相依赖, 这样我们在后面就可以按照依赖关系进行调度。

• 分包分项目开发 ETL: 根据项目进度我们安排包的开发, 先上项目先开发, 后上后开发, 如果项目为同一主题, 我们可以共用一个包, 这样在数据统一性方

面会更好。

- ETL 更新顺序：根据表之间的数据传递关系，建立起表之间的依赖关系，根据表之间的依赖关系建立起包之间的依赖关系，如表 A 有数据来源于表 B，那么我们认为表 A 依赖表 B，对包同理。原则上表和表、包和包之间不互相依赖。

- ETL 中断纠正机制：ETL 中断是商业智能系统常见的问题，建立起简单有效的 ETL 中断处理机制是必不可少的工作之一。

- ETL 统一调度：确定更新时间、更新频率、更新对象。

2)ETL 的挑战及应对思路

- ETL 技术架构：技术架构只有合不合适，没有说哪个最好，前面我们在系统架构方面基本了解到现在我们针对电子商务行业采用了 ODS、DW、DM 这样三层架构，不管什么工具都要符合这样三层架构思想进行设计，这样我们在架构层面就有明确的目标定位，相当于统一了思想，从而为后面的整体设计，分步、分工实施提供了可能。系统架构也是我们后面制定技术标准的大框架。

- ETL 更新机制：ETL 开发除了数据质量之外，最具挑战的就是数据更新机制的设计，ODS、DW、DM 的每个表都面临数据怎么更新的问题。只有最合适，没有说哪个方法最优，不管哪个层级的表数据更新策略可以简单分为：

- I. 全量更新，每次把表数据清空，重新生成，对 ETL 来讲小表我们可以每次全量更新，而对于每天增量数据较多的表则不能全量更新，只能利用增量更新办法。

23

- II. 增量更新，其中增量更新分完全增量更新和相对增量更新，完全增量更新是指加载到目标表的数据和目标表原数据既不出现重合也不出现缝隙，这是最理想的一种更新方式，但是因为增量判断机制复杂，增量方式众多(新增、更新、删除)，每种增量方式的增量算法完全不一样，开发难度巨大，特别是数据源如果不稳定，那将出现致命性错误，甚至导致数据仓库必须重新初始化这样大的事件；相比较相对增量更新可以允许数据有一定重合度，通过关键字等判断自动删除重合部分再简单加载新数据，不用面临一个指标更新要几个算法应付各种更新机制的问题，这里只需要一种更新算法就可以，设计简单，性能也不会有太大牺牲，如果大数据引入分区技术，每次通过移除分区来删除重复数据那性能几乎没有损失。

- 异常中断纠正机制：很多企业在开发 BI 之初，面临 ETL 中断之后是重新跑数据还是哪里断了哪里重启这样一个问题。一般来讲重新跑风险小，但是时间上要允许，而断点处开始执行则面临人工干预的问题，即使再有经验的人操作这个风险也始终存在。事实上 ETL 开发到一定程度可以用点时间来解决问题，现在焦点科技通过合理设计在效果上实现了融合两者优点的方法：

I. 建立表之间依赖关系。

II. 任何报错只影响对该 ETL 对象表有依赖关系的表，受影响的表不执行，节省时间，因为跑了也是白跑，不受影响的表继续往下执行。

III. 跑完之后系统自我检查是否有报错，有报错再次重跑，重跑过程中系统自动调度已发生错误的 ETL 进行重跑，其他成功的表不再重跑，依赖于该 ETL 的表其 ETL 都需要重跑。

IN. 如果再次重跑还是失败那么通过邮件预警到相关人员，检查错误排除错误直到解决后再次重跑。

• 大数据的 ETL 技术：解决大数据，是电子商务行业必须面临的问题，现在主流的办法还是分布式计算，其中 Hadoop 最为常用，当然 EMC 的 Greenplum 更为成熟，其费用也不菲，有条件的建议大家应用后者。

• 数据质量：我们可以认为 BI 产生的数据和业务部门需要的数据不一致问题都是数据质量问题，所以在开发 ETL 过程中处处是风险：

I. 首先是需求问题，BI 需求最大难点是难于统一各个独立业务部门的指标定义，除非自上而下强行统一，但一般来讲 BI 实施之初不可能有这种力度，比较可行的办法还是在数据仓库放基础数据，面向不同业务部门设计独立的集市来满足其需求，当 BI 在各个业务部门应用到一定程度，他们会自动统一这些定义和需求，否则他们将无法沟通，当然 BI 团队可以加快这个进程。

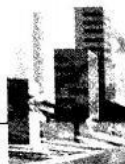
II. 其次就是架构问题，在数据仓库层面要放基础数据，基础数据不要有冗余，一个字段在多个表出现将会导致数据难于统一，这是最大忌讳，为了提升仓库到集市的数据统一性，我们可以在仓库层生成结构性的中间表，让大部分基础运算都统一到这层，这样为仓库进集市的数据统一性又降低了难度。集市设计最好是分主题，在同一主题下，数据定义要求完全一致，表尽量共用设计，这样不管是哪个项目在同一体系内数据都是一致的，另外就是不同定义的数据哪怕在不同主题间也要区别命名。

III. 最直接的数据质量问题还是 ETL 开发和测试，ETL 测试不仅要看数据总量，必要时还得交叉维度和层级进行数据测试，简单测试指标包括汇总、计数等。

IN. 系统维护要能自动预警数据质量问题，这样可以相关人员第一时间发现问题、解决问题。

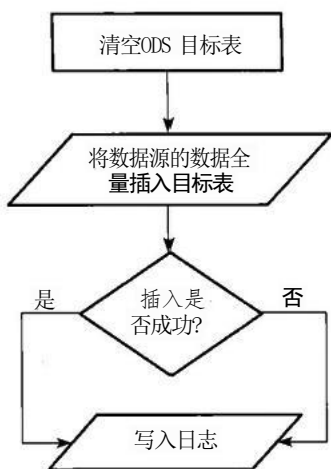
3) 电子商务行业可采用的 ETL 开发

一般的设计数据仓库都可分为 ODS、DW、DM 三层，下面我们来看一下这三层我们是怎么做的。



(1) ODS 层全量加载、增量加载 ETL 开发流程 (图3.8)

ODS 全量加载:



ODS 增量加载:

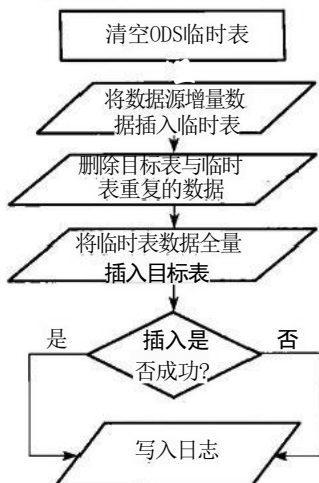


图3.8 ODS_ETL 开发流程图

- 全量加载将清空原有数据，重新载入最新的源数据。
- 增量加载每次只更新历史数据中有更新的记录，并且载入历史数据中没有的新数据。

25

(2) DW 层 ETL 开发

- 事实表先更新，再更新维表。
- DW 层事实表全量加载 ETL 开发：与 ODS 层全量加载 ETL 开发方法类似，在此不再重回复。
- DW层事实表增量加载 ETL 开发：与 ODS 层增量加载ETL 开发方法基本一致，但是需要把相关维表字段关联修改为维表的代理关键字。
- DW层维表增量加载 ETL 开发如图3.9所示。

(3) DM层

- DM层全量加载 ETL 开发：与ODS层全量加载 ETL 开发方法类似，在此不再重复。
- DM 层增量加载 ETL 开发如图3.10所示。

3.3.3 ETL 的测试

ETL 数据测试采用黑盒测试[8], 包括 ETL 常规检查和业务逻辑测试。

ETL 常规检查包括①ETL 脚本是否有运行错误；②ETL 编码是否符合 ETL 技术规范。

ETL 业务逻辑测试是 ETL 测试的重点，主要包括：①业务逻辑测试：指标设

DW 维表增量加载：

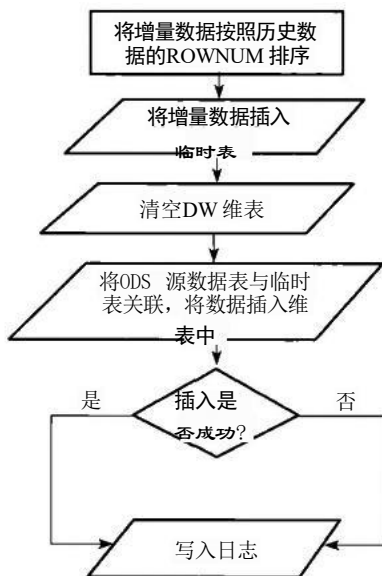


图3.9 DW_维表增量加载 ETL 开发流程图

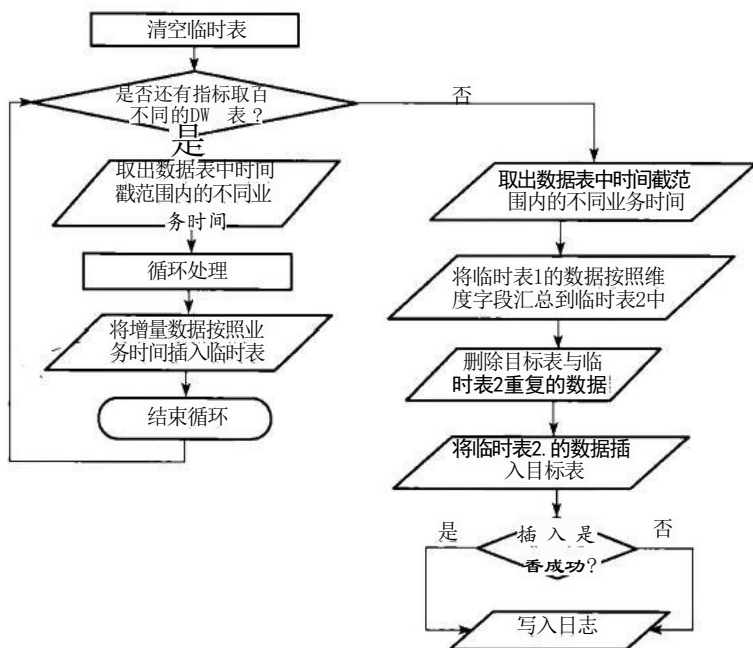
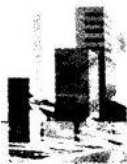


图3.10 DM 层增量加载 ETL 开发流程图



计是否符合业务逻辑；②数据量测试：目标表数据量是否与数据源数据量一致；③唯一性检查：检查 ETL 重复调度是否会出现重复数据；④性能测试：测试 ETL 的效率是否在可接受的范围内。

ETL 的测试流程包括：理解业务逻辑，开发测试用例，编写测试文档，通知开发人员修改 BUG,BUG 验证等。

3.4 多维分析

OLAP 的目标是满足决策支持或者满足在多维环境下特定的查询和报表需求，它的技术核心是“维”这个概念。多维数据分析工具的集合构成了OLAP。存储在各种数据源中的数据通过ETL 存储到数据仓库中，对数据仓库中的数据进行多维建模，便于利用数据挖掘技术和工具发现隐含的信息，还可以利用前端展现工具将数据分析结果展现出来，以供用户查阅。

3.4.1 多维数据分析技术

如何理解多维数据中的维呢？维是我们分析问题和观察事物的角度，从不同的角度对事物进行分析观察可以得到不同的结果，也能够让我们能够更加全面和真实的了解事物的本质。比如公司销售业务数据中，记录了公司产品(会员)详细情况，则我们可从以下几个方面来对销售数据进行分析：

27

(1)从产品的角度，可以按产品的类别、品牌来查看产品的销售情况。

(2)从客户的角度，可以按客户的类别、地区等来查看产品的购买情况。

(3)从销售代表的角度，可以按销售代表的部门、级别等来查看产品销售业绩。

(4)从时间的角度，可以按年度、季度、月份等来观察产品销售的变动情况。

其中产品、客户、销售代表、时间分别是四个不同的维度，每个维度都从不同方面体现了销售数据的特征，而每个维度又可按粒度的不同划分成多个层次，称为维度成员。

多维分析中另一个重要的概念是数据指标，简称指标，指标代表了数据中的可度量的属性，在上面的销售数据中有两个重要的指标是销售数量和销售金额。

了解了多维数据的维的概念之后，便可对数据进行多维分析操作，常见的多维分析操作主要有：钻取(上钻和下钻)、切片、切块、旋转。焦点科技目前使用的 QlikView 软件即可实现上述多维分析操作。

钻取是改变维的层次，变换分析的粒度。钻取包括上钻和下钻，上钻是在某一维上将低层次的细节数据概括到高层次的汇总数据的过程，减少了分析的维数；下钻则是相反，它从汇总数据深入到细节数据进行观察和增加新维。比如：将各销售

人员的销售业绩汇总为各个销售部门的销售业绩，整个销售部的销售情况可以细分为各销售部的销售情况，进一步细分到各个销售人员的销售业绩。在多维分析中，如果在某一维度上限定了一个值，则称为对原有分析的一个切片；如果对多个维度进行限定，每个维度限定为一组取值范围，则称为对原有分析的一个切块。如果变换维度的顺序和方向，或交换两个维度的位置，则称为旋转。

3.4.2 多维数据分析的实现

要进行多维数据分析，首先需要提取维度和指标信息并且确立维度与指标之间的关系，然后根据统计规则统计出指标的结果，最后将统计结果展现出来。这三个步骤分别对应于数据仓库建模、ETL 开发、前端展现的BI 项目开发过程。下面以焦点科技销售BI 系统下的销售业绩统计分析为例(为了便于描述，对数据进行了简化)，详细阐述多维数据分析的实现。

1) 提取维度和指标信息

数据仓库中销售业绩数据主要包括部门人员信息和销售数据。部门人员信息为：销售人员、所在部门。销售数据为：日期、销售人员、销售额。销售提成是销售额的3%。进行销售业绩统计分析的维度有：时间维、人员维，数据指标有销售额、销售提成。

2) 实现主题扩展和关联

由于各个主题间维度在定义时都是相互独立的，其间并无隐含的关联信息，但是为了数据一致性和关联性，我们将会建立各个分析主题维度的关系信息。比如销售的提成主题和部门的业绩主题就是上下层的关系，我们只要在销售提成主题上加入部门维度，既可以实现提成主题分析也可以向上钻成部门业绩的多维分析主题，如图3.11所示：

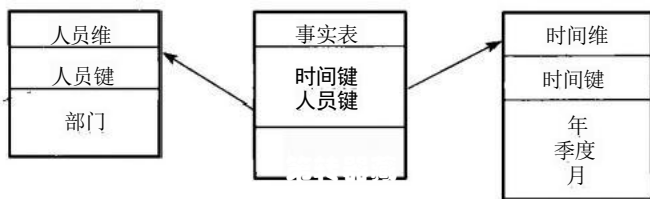


图3.11 销售业绩数据模型

3) 执行数据整合

建立好数据模型之后，利用 ETL 将数据仓库中的数据按照客户需求进行相应的统计汇总得到多维分析数据。最终形成QlikView 报表，实现钻取、分区、分块、旋转等操作。销售业绩统计的结果可以生成如下的 QlikView 报表。

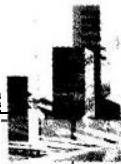


图3. 12展现了各部门各销售代表每月的销售业绩情况，可以在图3. 12 的基础上进行上钻得到各部门各销售代表的年度销售情况(如图3. 14), 以供各部门主管或相关人员查看和分析。在图3. 13基础上再进行上钻，即可得到各销售部的年度销售情况。相反的，如果在图3. 13基础上，进行下钻，可以得到如图3. 12所示的明细数据。在图3. 12中想要查看各部门各季度的销售情况，则可以将时间维度与人员维度进行旋转(交换顺序), 结果如图3. 15所示。

销售情况明细表						
部门	销售代表		单度	月份	销售超	销售提成
销售一部	李明	一季度	=	一月	1000	30
				二月	1200	36
				三月	1400	42
		二季度	=	四月	1100	33
				五月	1500	45
				六月	1700	51
		三季度	=	七月	1100	33
				八月	1100	33
				九月	1000	30
		四季度	=	十月	1600	48
				十一月	1300	39
				十二月	1500	45
	王丽	一季度	日	一月	1200	36
				二月	1700	51
				三月	1100	33
		二季度		四月	1100	33
				五月	1000	30
				六月	1100	33
		三季度		七月	1000	30
				八月	1600	48
				九月	1300	39
		四季度		十月	1500	45
				十一月	1400	42
				十二月	1500	45
销售二部	刘阳	季度		一月	1000	30
				二月	1200	36
				三月	1000	30
		二季度		四月	1100	33
				五月	1500	45
				六月	1500	45
		三季度		七月	1100	33
				八月	1100	33
				九月	1100	33
		四季度		十月	1500	45
				十一月	1500	45
				十二月	1000	30
	张三	季度	西	一月	1400	42
				月	1500	45
				三月	1400	42
		二季度		月	1600	48
				五月	1100	33
				六月	900	27
		三季度		七月	1100	33
				八月	1100	33
				九月	1000	30
		四季度		十月	1100	33
				月	1500	45
				十二月	1500	45
合计					60800	1824

图3. 12 销售业绩分析报表_明细表

QlikView 报表中可以便捷地使用图形分析来展现多维数据，如图3. 13所示，利用柱形图展现出各季度各部门的销售情况，便于对比分析。并且在柱形图上仍然可以实现钻取、分区、分块等多维分析操作。比如，将季度下钻到月份，如图3. 16所示。

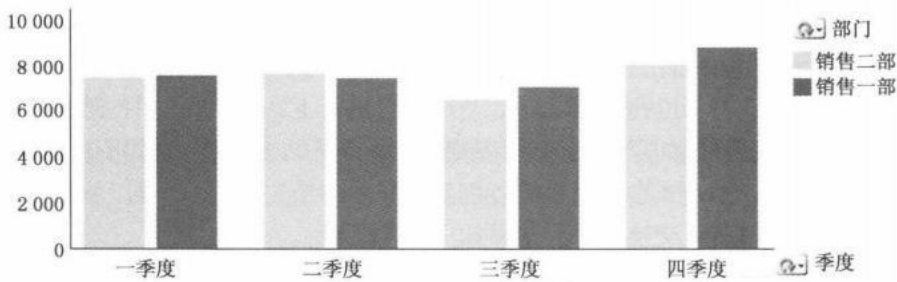


图3.13 销售业绩分析报表_柱形图

销售情况明细表					
部门	销售代表	季度	月份	销售额	销售提成
销售一部	李明			15500	465
	王商			15500	465
销售二部	刘阳			14600	438
	张三			15200	456
合计				60800	1824

图3.14 销售业绩分析报表_上钻

销售情况明细表					
部门	年度	月份	销 要 代 表	销售额	销售提成
销售一部	2	一季度	进	7600	228
		二季度		7500	225
		三季度	生	7100	213
		四季度		8800	264
销售二部		一季度		7500	225
		二季度	生	7700	231
		三季度	+	6500	195
		四季度	注	8100	243
合计				60800	1824

图3.15 销售业绩分析报表_旋转

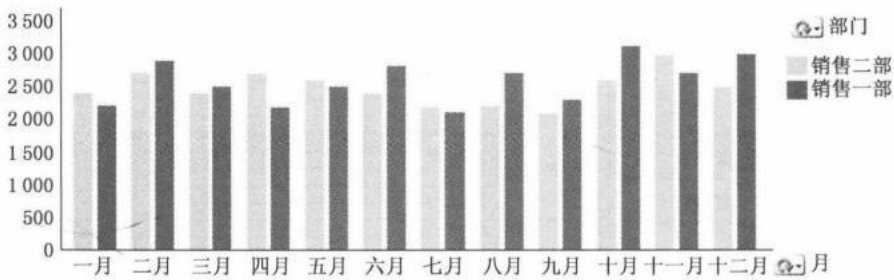


图3.16 销售业绩分析报表_柱形图下钻

3.4.3 多维分析在商业智能中的应用

1)覆盖各层用户需求

BI 在电子商务行业中的用户有买家、卖家、网站内部中高层管理人员以及一线操作员。多维数据分析可以从不同的客户角度进行数据分析，满足各个层次、各

个级别用户的统计分析需求。比如，销售业绩分析中，底层销售代表需要知道一定的时间周期内的自己的销售额和自己的提成情况；而部门主管们更加关注的是本部门的月、季度或者年的计划及绩效。

2) 支持数据挖掘

多维分析的结果可以用于支持数据挖掘，我们可以基于现有分析结果，挖掘出更有价值的隐含在数据中的信息。比如，基于销售数据的多维分析，进行简单的数据挖掘就可以发现出每个季节或月份比较畅销的产品类型、不同的用户对于产品的喜好等，从而可以根据这些信息制订合适的产品销售计划。

3.5 数据挖掘

数据挖掘是从海量数据中挖掘潜在的信息与知识，挖掘结果的应用往往可以带来产品的创新以及业务的创新。B2B、B2C 等电子商务平台在每天的运营过程中，信息系统都会产生大量数据信息，数据挖掘技术能够把历史积累的大量数据进行分析以及模型化的挖掘和处理，从中发现隐藏的规律或模式，提取辅助商业决策的关键性数据，为决策提供支持。MIC 挖掘价值体现见图3. 17所示。



图3. 17 MIC 挖掘价值体现

3.5.1 价值体现

1) 产品深度支持

新产品设计：对于竞争激烈的互联网企业，好的产品就是企业的生命。随着信息化的发展，如何能够设计出一个用户满意的产品，已经不再完全凭借个人的直觉和行业经验，更多的还是需要数据挖掘和分析的支持，例如产品的结构设计、产品如何定价，以及该产品适用人群分析等。

旧产品优化：每一产品都有其生命周期，一个过去很好的产品现在可能已经过时了，所以对于老产品的效果分析与挖掘是必不可少的，怎样正确的对当前产品的



效果作出客观评价，同时对产品将来的发展趋势作出准确的预测都需要数据挖掘技术的支持。

2) 客户深度认识

服务客户：俗话说知己知彼，百战不殆。电子商务对于用户来讲提供的也是一种服务，如何能够让用户对我们的服务认可，首先就需要对客户有所了解，这样我们才能够有针对性地为客户服务。例如客户细分、流失客户挖掘等都是通过数据挖掘的技术使我们对客户能有一个更深入地了解。

客户风险控制：对客户深度认识除了能够为客户更好的服务以外，还可以控制由客户带来的风险，尤其是互联网这样一个特殊的行业，欺诈等现象时有发生，影响非常严重，可是少量的异常客户隐藏在海量的客户当中我们很难区分，但有了数据挖掘技术的支持，我们可以对客户有更深入地认识，就可以把异常客户区分出来，并对异常客户进行监控，从而对客户可能带来的风险进行控制。

3) 精确化营销与服务

精确化营销与服务是基于对客户深度认识的基础上，针对于不同的客户采用不同的营销策略和服务策略，提高营销的成功率和服务效果。要做到精确化单靠人工去完成成本非常大，或者说根本无法实现，但数据挖掘技术可以精确地发现用户的需求并且挖掘出最合适的产品与营销服务策略，销售与客服人员可以通过数据挖掘结果的数据支持，自动的实现精确化的营销与服务。例如现在大型网站都非常流行的个性化推荐服务以及客户维系挽留系统，都是数据挖掘技术在精确化营销与服务的应用。

4) 精确化推广与市场培养

精确化推广：网络推广是电子商务网站得到迅猛发展的重要手段之一，网络推广最大的难题就是推广成本与推广效果的平衡问题，如何能够在有限的推广成本前提下取得最好的推广效果，数据挖掘技术可以通过公司内部以往推广的一些历史数据挖掘出最有效的推广方式，提供辅助精确化推广的决策支持。

精确化市场培养：市场是公司赖以生存的外部环境，数据挖掘技术能够对目前市场情况作出准确评估，并对市场将来的可能的变化趋势作出预测，使决策层根据市场分析结果作出正确的市场决策，抓住市场机会，规避市场风险。

3.5.2 主要应用

上文中我们在讨论数据挖掘价值体现的时候实际上已经提及了一些数据挖掘的应用，例如客户细分、客户流失预警等。就是通过这些方方面面的数据挖掘应用，数据挖掘的价值才能在电子商务行业中体现出来。经过总结，数据挖掘在电子商务行业主要有如图3.18所示的应用。

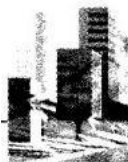


图3.18 数据挖掘的应用

3.5.3 应用实现要求

1) 完善的开发规划

应用之间的关系要求：数据挖掘的多项应用并不是独立的，绝大部分应用需要联合起来才能发挥价值，例如客户维系挽留系统一般就需要以客户流失预警系统为前提，客户细分是大部分其他应用的基础，这种依赖关系存在但是又不绝对，所以导致不同应用之间的关系非常复杂。必须制订完善的开发计划。

应用的价值要求：对于一般的应用开发计划肯定是价值大、难度低的先开发，但是数据挖掘的应用有时候价值的大小很难直接评估，例如客户细分从短期内看价值不大，但是它是很多挖掘应用的基础，长期来看价值非常大。因此应用的价值和长远的开发计划关系非常密切，只有详细的开发计划出来了，才能准确地评判什么样的应用应该先做，什么样的应用应该后做。

数据要求：数据挖掘一般对历史数据都有要求，虽然现在的电子商务都有自己的数据仓库，储存了大量的历史数据，但是具体到某一个实际的应用，我们还是经常会发现数据不足的问题，原因就是历史数据往往是业务系统实现现有业务而产生的，不一定会考虑到后续的数据分析与挖掘，缺少一些关键的分析指标，因此经常为了能够收集到这些数据而将一些很重要的数据挖掘应用延后，如果我们有完整的数据挖掘应用开发规划，就可以根据计划提前准备必要的数据库。

2) 与时俱进的挖掘技术

虽然多数的数据挖掘应用只用到了最基础的挖掘算法，基础的挖掘算法非常



有用，但是挖掘技术毕竟是一门难度非常高的技术，一旦需要用到比较高端的挖掘技术时才来学习是来不及的，因此在平时的工作当中，保持挖掘团队技术的先进性是非常必要的，这些技术是数据挖掘应用能够稳定实现的有力保证。

3.5.4 常见问题

1) 执行效率

复杂算法的执行效率：电子商务行业是业内公认的海量数据的行业之一，在这里有很多效果非常好的算法因为效率的问题而搁浅，特别是有些实际的数据挖掘应用不能用抽样的方式解决效率问题。因此在电子商务行业有时候并不是结果最准确的挖掘算法才是最好的。一定要找到成本与效果的平衡点，这样的挖掘算法才是最好。

业务系统实时要求：很多挖掘算法不是把挖掘结果分析出来就算结束，这些结果还要提供给业务系统，供业务系统实时处理事务时使用。这类需求要求实时性要求非常高，一种比较实用的方法是在BI系统中离线训练挖掘模型，等模型规则评估通过以后，将规则封装到业务系统中进行实时处理。另一种方法就是尽量提升挖掘算法的效率，BI实时处理数据并实时同步到业务系统当中。

2) 数据质量

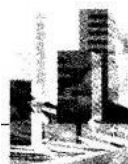
数据缺失：电子商务行业大部分用户是虚拟的，这就导致了用户产生的数据经常是部分信息丢失，例如用户如果不登录就在网站上进行浏览操作，那么这些操作的主体信息就会缺失，间接地与主体相关的信息也会缺失。

数据失真：电子商务行业并不是完全需要实名制，因此很多信息用户自愿填写，填写只要符合规则就可以，对填写的真实性一般是不做验证的，这就导致了大量数据失真。

做电子商务行业的数据挖掘的时候，由于数据的缺失与失真导致数据的预处理一般要比其他行业工作量大很多，挖掘的难度也大很多。在这一点上在做电子商务行业数据挖掘项目的时候需要充分评估项目的难度与实现周期，提前做好准备。

3) 云计算

近年来云计算越来越流行，特别是发展迅速，数据量暴增的电子商务行业，云计算的应用取得了长足的进步；相对于传统的数据挖掘技术，能不能更有效的和云计算相结合，开发出基于云计算技术的数据挖掘应用，是电子商务行业未来数据挖掘新的发展方向。焦点科技目前做的云计算平台下的个性化推荐系统就是基于云计算平台的数据挖掘应用。



3.6 BI 在电子商务领域的应用

3.6.1 电子商务运营决策分析

随着整个电子商务行业格局的不断调整，B2C、B2B、C2C网站层出不穷，我国的电子商务运营企业，正面临激烈的竞争和前所未有的挑战。如何提高对市场的快速反应能力，改善企业管理水平，提高经营效率，达到建设行业领先商务平台的战略目标，成为我国电子商务行业当前面临的一个重要课题。

商业智能作为一种企业辅助决策的解决方案，在传统制造业取得了显著的效果。电子商务行业和传统的制造业有着很大的不同，电子商务行业不仅要辅助公司的中高层做决策，还要和网站的运营部门结合起来，做网站的分析。如图3.19所示，电子商务决策分析系统为分析层、决策层提供决策分析平台，为操作层以及各业务系统提供执行优化的智能帮助。

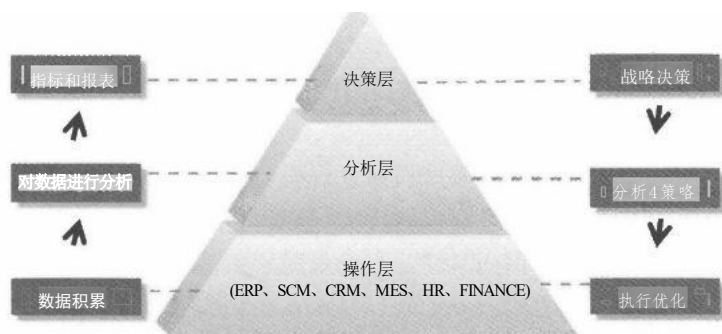


图3.19 电子商务决策分析系统

1) 企业级 KPI 决策系统

满足企业信息整合，由粗放型经营和管理向集约型转变。对于企业决策者也就是公司的高层人员而言，必须对企业运作中的各种信息和数据进行监控、分析、管理，并据此进行判断、决策，进而采取行动。随着各种数据处理和分析应用程序的涌现和升级，企业发现需要处理和分析的信息和数据越来越多，越来越复杂，在处理和历史信息中所投入的时间成本更大。而使用仪表盘能可视化地呈现信息，且有助于判断、监控并支持决策，从而有效地提升信息系统的实时信息处理能力，这一特性使得“仪表盘”成为缓解和解决上述问题的重要途径，并日益赢得使用者的青睐。

2) 部门级绩效考核系统

目前很多企业在绩效考核中面临的问题是这些企业大多建设实施了 ERP 系

统,实现了业务处理的自动化,但在管理决策上很多企业并未真正实现管理信息自动化。这样企业绩效管理的处理过程是,ERP 系统进行业务处理产生了大量业务数据,管理部门在此基础上把数据从 ERP 中倒出,利用数据库处理或 EXCEL 手工进行数据整理和分析,按照企业的绩效管理的考核方法,如 KPI 的各种指标等进行计算和分析,最后得出绩效考核的“数量”结果。

3) 业务监控及分析

对于电子商务网站,主营业务就是网站,了解网站运营情况,对运营进行系统有效的分析至关重要。运用 BI 对运营数据进行分析,从而提升网站整体运营效果,监控并分析网站运营状况,找出存在的问题,提出改进优化意见,实施改进措施,进一步提升网站整体性能,获得更高的知名度,留住老用户,吸引新用户。

4) 分析型工作流程

目前市场成形的大型分析型系统主要有 ERP 系统、SCM 系统、CRM 系统等,这些系统都给公司运营管理带来了巨大的价值。因此要提高公司在电子商务领域的核心竞争力,分析型流程定制是公司运营发展的必然趋势。分析型工作流程纳入到运营系统后,会把 BI 和业务紧密结合,使公司运营系统化、体系化,同时也会提高运营系统的响应效率与质量,从而提升操作人员的工作效率与工作质量,规范化运营体系保障业务运营的正常进行。

3.6.2 对外-支撑网站用户分析

BI 在支撑网站上有其自身的优势,无论是完整的成熟仓库架构还是先进的数据挖掘技术,都是其他系统无法取代的。网站作为服务客户的一个平台,网站运营无非是以下几个方面:用户对象、服务工具、服务方式,而 BI 在电子商务服务中的这几个方面都有应用(图3.20),下面是其具体应用框架。

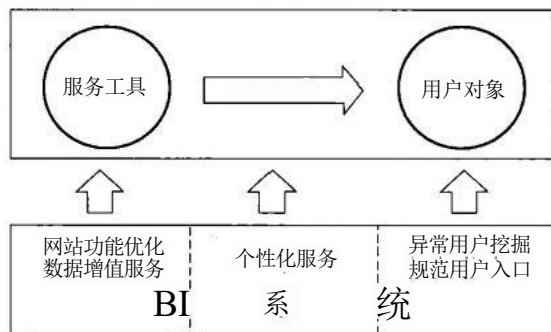
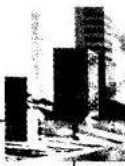


图3.20 BI 网站支撑系统框架



1) 个性化服务

电子商务竞争日趋激烈,网站的服务质量对于网站的发展至关重要。网站的用户千差万别,采用统一的服务不可能满足所有用户的需求,如何对每一个用户采取有针对性的个性化服务就是BI的应用领域之一。

个性化服务主要可以区分为:问候、用户定制、实时消息和特色推荐。但是无论哪一种个性化服务都必须面对服务时机的选择、服务对象的识别、用户偏好的总结以及服务与对象匹配几个环节,只是有时候某些情况下某些环节相对简单比较容易实现。

个性化服务做得好,可以大大缩减用户浏览网站所花费的时间,提高用户的效率,增加客户对站点的好感,提高用户黏性,对增加网站用户规模、提升网站知名度都有好处。

2) 网站异常用户挖掘

电子商务网站欺诈现象一直是用户非常关注的一个问题,很多电子商务企业甚至把欺诈问题看成是关乎网站生死的大问题来抓,但是在互联网这种特殊的环境模式下欺诈问题一直很难解决。

电子商务的欺诈现象,主要形式有诈骗、钓鱼以及一些虚假注册、重复注册等。这些异常用户隐藏在海量的用户群体当中,利用传统的统计方法和人工识别方法很难区分,BI可以利用数据挖掘技术,根据用户的历史行为数据的挖掘分析,以及用户基本信息的文本挖掘等手段,将异常用户与其他的正常用户进行区分,并将分析结果提供给相应的负责人进行处理,大大提高异常用户的甄别能力。

网站欺诈用户的减少,为其他用户的正常交易提供一个安全的环境,使用户在网站上交易更放心,无疑会大大提高用户对网站的认同感,对于网站的发展有着非常积极的作用。

3) 网站效果分析及优化

任何一个网站都不可能很完美,网站都是随着功能的不断改进和完善而不断进步的,那么功能好不好,应该如何完善就成了网站时时刻刻需要面对的问题;同时网站作为电子商务企业获利的工具,网站现有资源是否合理利用也是企业面临的另外一个问题。

BI通过数据挖掘与分析,剖析用户使用网站功能的每一个细节,通过流量分析,与用户访问路径的挖掘,合理评估网站功能的好坏,并对网站功能可能存在的问题进行定位,为产品人员合理的优化和改进相应的功能提供参考依据;BI通过网站资源的流量监控,对网站资源的价值进行评估,并发现新的有潜在价值的优秀资源,为运营部门对资源产品的合理定价提供参考依据。

网站功能的改进和优化可以保证网站能够与时俱进,确保网站能够有持

续的竞争力，使网站能够不断地做大做强；网站资源的合理利用则是网站持续盈利能力的保证，用户认可并且具有强大盈利能力的网站才是真正意义上好的网站。

4) 数据分析增值与服务

电子商务网站多年来沉淀了大量的数据资源，这些数据资源就好比有价值的矿石，除了利用这些资源优化网站间接的为企业创造价值以外，能不能运用这些丰富的数据资源，提炼出对用户有价值的信息，提供给需要的客户，从而直接为企业创造收益成为了很多企业的目标。

电子商务网站可分析的有价值信息很多，根据电子商务行业的特点这些信息可以分为下面几类。

- 可以帮助用户更加了解市场的市场分析：如用户所处行业的供求变化趋势，哪些产品热门等。
- 可以帮助用户更加了解他的客户的客户分析：例如用户来源于哪里，他们是如何找到我的，哪些客户需要重点关注等。
- 可以帮助用户提高他在网站上推广效果的效果评估及优化：例如用户购买推广产品的推广效果，哪些推广方式可以帮助用户提高推广效果等。

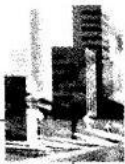
这些分析既可以帮助用户发现市场机会规避市场风险，又能帮助用户提高推广效果，核心就是在为用户创造价值。用户能够通过我们的数据分析增值与服务创造价值，那么他们自然愿意花费这些价值中的一部分来购买这些服务，因此数据分析增值与服务所能为公司创造的价值多少完全取决于增值服务为客户带来的价值的多少，为客户带来的价值越高，那么增值服务本身的价值就越大。

从上面的分析我们不难发现，网站作为电子商务企业服务于用户的一个工具，核心还是服务于用户，因此 BI 对网站的支撑实际上就是对网站用户的支撑，用户的需求在变化，那么 BI 对网站的支撑也要变化，以用户的价值为导向的 BI 的价值才能够得以体现。

3.7 软硬件方案

商业智能在电子商务行业中实施面临着较多的挑战和问题，如何选用合适的软硬件产品来支持企业的 BI 实施是一个重要的问题。

首先是处理海量数据的问题，随着业务的发展，电子商务企业每天都会产生巨大的数据量，比如网站用户信息、网站访问日志等记录，这些数据少则数十万数百万，多则上亿；如何选择合理的软硬件方案来高效准确地处理这些数据是商业智能成功实施的一个关键。商业智能在电子商务行业中会面对不同层次的用户，从



底层的操作人员到公司高层领导，还包括网站用户。不同层次和类型的用户会产生大量的差异化的需求，用户对实时性、数据粒度、用户终端等各方面的需求各不相同，因此BI实施过程中需要选择合适的软件和硬件，才能保证数据顺利的从业务系统、数据仓库、数据集市一直到最后的报表展现都能满足不同用户的需求。

电子商务行业的 BI 涉及的关键技术有数据仓库及 ETL 技术、OLAP 技术和数据挖掘技术，市场上已经存在着许多厂商的BI产品以及一些开源的项目来支持这些技术(表3.1)。BI产品种类繁多，每家厂商几乎都会宣称自己的产品是如何的好，在构建企业自身的 BI系统时应该根据企业自身的需求考虑软硬件产品的选择。

表3.1 主流 BI 工具

数据库	ETL工具	OLAP 工具	数据挖掘工具
DB2	DataStage	BO	SAS EM
Oracle	PowerCenter	Cognos	Clementine
SQL server	SSIS	QlikView	Weka
Teradata	DataIntegrator		
Greenplum			

BI产品主要涉及数据库、ETL 工具、OLAP 工具及数据挖掘工具，有的厂商有很全的产品线，从数据库、ETL 到前端展现及数据挖掘都有涉及，能够提供全面的解决方案，比如Cognos 和 Oracle;用户可以购买厂商的整体解决方案实现 BI。很多BI产品都是很优秀的，但是对于企业来说要考虑各方面问题，好的产品不一定会适合自身。

数据库方面 DB2、Oracle都是很好的产品，Teradata 和 EMC 公司的 Greenplum 在大数据处理方面有着很强的能力。根据 Gartner 在2012年2月发布的数据仓库魔力象限[]显示 Teradata在数据仓库方面处于领导者地位， Teradata 在实际的应用中也显示出其强大的数据处理能力，另外 Oracle、IBM也处于领导者象限，EMC 公司的 Greenplum 凭借其云计算方面的能力也进入领导者象限(图 3.21)。ETL 工具方面比较主流的产品有 DataStage 和 PowerCenter, 这两者占据了国内主要市场份额，DataStage 在被 IBM 收购之后显得后劲十足。分布式部署云计算是解决大数据最好的方法，Greenplum 在云计算方面表现出色，但是这是需要付费的，当前主流的 Hadoop 也很流行，可以让企业以较低的硬件投入实现分布式，当然这对开发人员的要求更高。

主流的 OLAP 工具有 BO、Cognos 以及QlikView 等，要评价工具好不好除了看数据准不准这类开发问题，更重要的就是要在考虑功能强大、开发简单之外，速度、美观、简单必不可少；很多厂商在性能上有自己解决办法，但大部分对开发者设

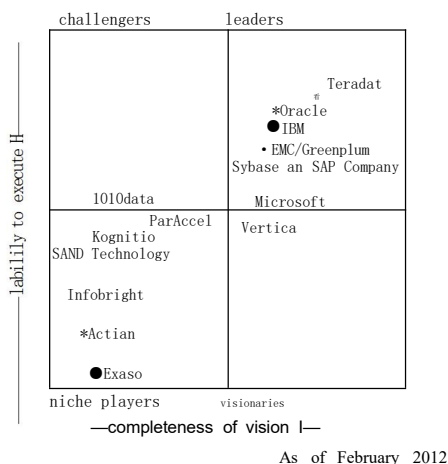


图3.21 Gartner 数据仓库魔力象限

计能力要求较高，比如Cognos 产品就要求编程实现，这样就需要开发者对产品非常熟悉才可行。Gartner 前端报表工具最具魔力象限连续几年纳入了 QlikView 这个产品，这个产品代表了基于内存的报表分析技术，美观简单，用户体验极佳。

数据挖掘产品领域，有 SAS、SPSS 等两大厂家，而像 IBM、Teradata 也都有自己的挖掘工具。相对而言，SAS EM功能更全面，性能更加优秀，算法也更全面，具有图形化界面，可视化操作。SPSS Clementine 开发更简便，功能上集成度更高，其中 SAS 需要编程，对开发者的要求更高一些。做好数据挖掘仅仅有工具是不够的，更需要数据挖掘人员对数据挖掘的理解和认识。

3.8 实施策略及风险控制

客观地说，懂得安装数据库、建表、写存储过程，再用展现工具开发报表基本就算 BI入门了，但要实施企业级的BI项目，只满足于这几点是远远不够的。如果实施策略不当，很可能造成投入巨资硬性推动后产生不了价值，或者东西很好并且有价值，因为方法不得当最后还是得不到客户认可，甚至被否定。所以这里我们来讲一下BI的实施策略，有代表性的恰恰是数据仓库支持OLAP 理念的两位创始人Bill Inmon 和 Ralph Kimball, Bill Inmon倡导自上而下实施策略(图3.22), 而 Ralph Kimball 则提出不同看法，认为数据仓库应该实施自下而上的实施策略。但经过那么多年的BI实施，对于一位有经验的BI实施人员我们认为可以采用第三种策略，那就是整体规划、分步实施，这种方式可以融合两种策略的优点，又可以避免其中的短处。

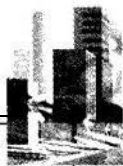


图3.22 传统的两大BI实施策略：自上而下、自下而上

3.8.1 自上而下策略

自上而下实施策略(图3.23)最开始由 Bill Inmon 提出，其核心思想还是突出前期的设计进而达到开发一步到位，实施的时候先统一规划业务，整体设计数据仓库，在数据比较稳定后再面向各个部门及各个层级人员整体实施企业级 BI应用。

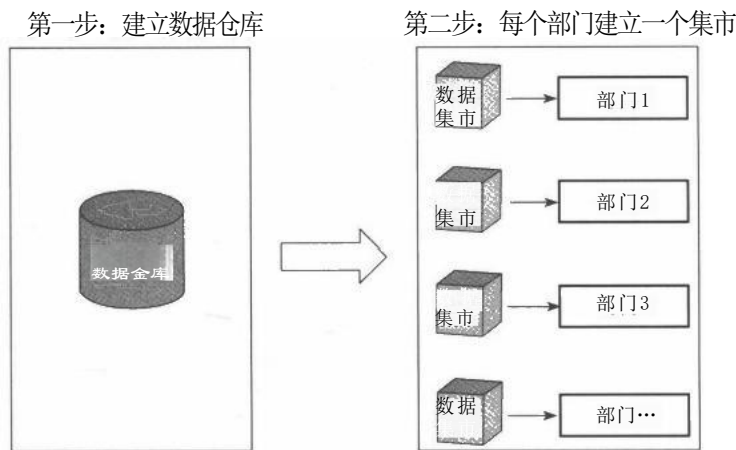


图3.23 自上而下BI实施策略

这种策略优点是：

- (1) 数据仓库模型统一设计，结构简单易于管理及扩展，冗余数据少。
- (2) 虽然有数据集市，但是数据集市是基于数据仓库的基础上进行设计，数据来源于数据集市，可以说是数据仓库派生出来面向特定主题的一个子集。
- (3) 数据仓库和数据集市在设计上属于两个不同层，只要数据仓库质量比较高，增加主题应用会变得相对简单。

其缺点是：

(1) 初期就大规模实施数据仓库，投资大。

(2) 需要较长时间才能使数据仓库开发到比较完善的程度，从而导致项目周期长，见效慢。

(3) 业务需求过于追求全面，导致项目难度大大提升，普遍情况是在 BI 实施之前大部分用户是很难想象 BI 有什么功能，能对他们有什么帮助，更难提出合理意见，这种时候唯一的办法就是引入高端咨询人员进行需求规划及设计。可想而知对于塑造差异化竞争力的今天来说，企业采用自上而下的实施策略存在巨大的风险。

从上面我们知道，自上而下的实施策略要发挥其优点避免风险必须具备如下条件：

(1) 有足够的财力。

(2) 直接打造企业级数据仓库，需要高层领导支持，解决跨部门协调资源的困难。

(3) 业务形态较为常见，有强大的成熟的业务咨询团队和实施团队，有能力化解整体设计及开发过程中的任何业务及技术问题。

(4) 业务系统较为统一，数据仓库数据质量有足够保证。

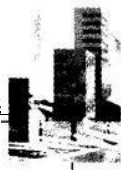
另外 Bill Inmon 提出自上而下的实施策略同时要求数据集市能够上下钻取，这种功能刚好数据立方体(Cube) 工具能够实现，最著名的有 Hyperion 公司的 Essbase, 还有就是 Cognos 的 Powerplay, 这些数据立方体工具笔者估计一开始应该是为了实现 Planning 应用而开发的，我们知道 Hyperion 和 Cognos 的 Planning 软件都是很知名的，Planning 软件在时间和组织架构上非常需要往上汇总、往下分摊，后面这种上下钻取的功能因为刚好满足 BI 的基本需求，过去十年被 BI 厂商广泛采纳和推广，并取得了不错成绩；但是从解决方案和工具来看，现在技术和工具只能说广义上的 Bill Inmon 的上下钻取，现在上下钻取在技术和架构上有如下三大流派：

(1) Cube 数据集市支撑的物理 OLAP。

(2) 关系型数据库加前端缓存支撑的逻辑型 OLAP。

(3) 关系型数据库加内存支撑的逻辑型 OLAP。

从上面的划分来看，当前数据模型设计和前端展现工具之间、软件和硬件之间互相支撑，这也是一种趋势，特别是现在云计算的成熟也必将促使 BI 在未来 3~5 年里其技术及工具方面发生巨大变化；从这个层面笔者再次提醒我们开发者要真正理解 BI 的本质和问题，了解主流的工具和技术，只有正确认识前者，合理利用后者才能做好企业级 BI 实施，特别是具有大数据的电子商务行业更是如此。



3.8.2 自下而上策略

虽然都是数据仓库理念的创始人,但不同于 Bill Inmon,Ralph Kimball 提倡自下而上的数据仓库实施策略(图3.24),Ralph Kimball 认为“不同数据集市合起来就是我们要的数据仓库”,这种策略的优缺点也同样明显。

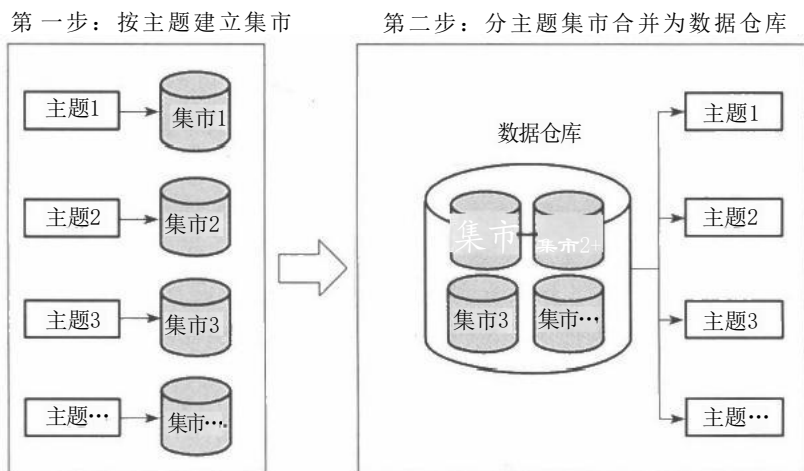


图3.24 自下而上 BI 实施策略

43

这种策略优点是:

- (1) 直接设计数据集市进行OLAP 分析,让BI快速产生价值,项目见效快。
- (2) 不断叠加的数据仓库建设策略能保证数据仓库设计更接近实际需要,避免仓库特别是模型的不合理设计,从而降低风险。
- (3) 另外在自下而上的策略里面Kimball提到维度建模的概念,从而产生了当前最为流行的数据仓库按照维度和事实的设计,特别是Kimball 对维度建模的若干建议对数据仓库设计具有非常高的指导价值。

其缺点是:

虽然前期集市设计速度快、风险小,但数据集市设计更多是在部门级或者比较单一的业务领域,这种设计难于推广到其他业务部门和领域,所以前期只注重数据集市设计是难于形成企业级数据仓库,具体来说有以下三个问题难于统一:

- (1) 数据粒度:数据仓库一般尽量保持必要的明细,而数据集市则是看当前业务需要。
- (2) 数据范围:数据集市只选择当前需要的数据,哪怕是同一个表在不用应用的时候也会在不同程度上进行过滤。
- (3) 数据定义:这个问题最大的,业务部门的指标选择和定义往往有差别,

所以不同业务部门的需求在集市设计上必然会有区别。

这里不得不说的是 Bill Inmon 和 Ralph Kimball 提出的策略都没有绝对意义上的自上而下或者自下而上,但在实际工作中,BI 项目组、IT 领导往往都喜欢绝对的自上而下,能够获取到公司最大的资源支持,结果也容易导致投入大,但是见效小而慢,甚至失败;而业务部门在久久看不到效果的时候又总是喜欢绝对的自下而上,每个部门自成一派,最终企业级 BI 迟迟建立不起来。为此我们还是要从本意上理解两种策略,特别是吸取其长处。

从上面我们知道自上而下、自下而上两种策略都有可取之处,也各有致命弊端,显然实际开发及应用过程中,谁都不会简单使用其中一种,都会两者兼备,这里最为提倡的是整体规划,逐步实施,下面进行详细介绍。

3.8.3 整体规划、逐步实施

简单概述不难发现 Bill Inmon 重视数据仓库的整体设计,从而保证系统的完整性,而 Ralph Kimball 重视快速推进,快速出效果,重视项目的实用性。应该说两者意图都值得肯定,所以我们提倡两者兼备的整体规划、逐步实施,如图3.25所示。

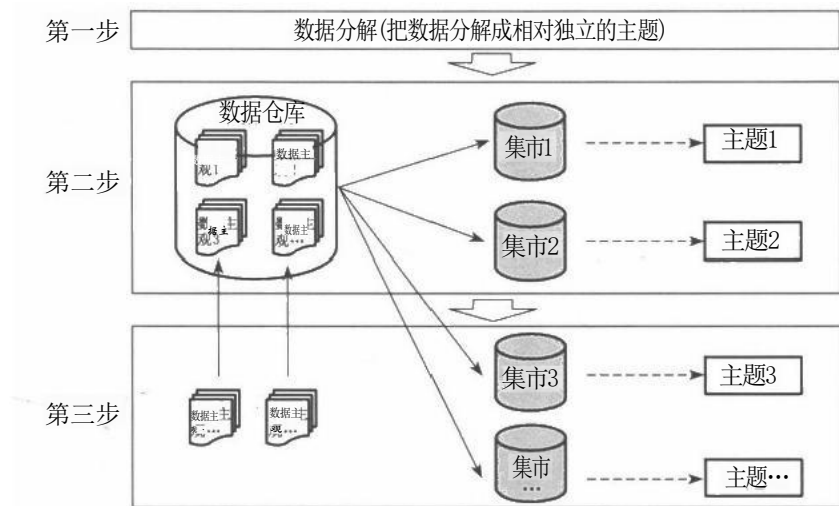
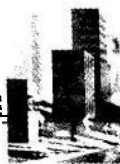


图3.25 整体规划、分步实施

这里我们建议把 BI 实施分成三步走。

第一步：把具有相互独立性的业务数据分成不同的主题，特别是要明确核心业务主题及维度信息表。

第二步：选定1~2个分析主题，对核心业务进行数据仓库设计并且实现数据入库；维表设计和开发尽可能一步到位，满足主题分析的集市数据要求；一个分析



主题一个集市。

第三步：根据用户使用情况及规划可以不断增加分析主题，扩展主题集市，如果数据仓库数据不够，那么需要不断补入数据，一般尽可能同一主题数据一起补入。

分步实施在推进整体规划上需要遵循下面的原则。

(1)基本原则：分步实施虽然要满足整体的规划，体现其整体性，但更要重视每一期项目的实效性，项目实施过程中可以根据实际情况对规划及计划进行合理调整。

(2)制订实施计划：基于整体规划和主题划分，对业务主题进行重要性分级及差距分析，从而形成分步实施计划。

(3)基本任务：针对不同的业务需求建立独立的集市，以满足不同的数据分析需求及性能需求；对数据仓库基本数据进行逐步完善，重视模型的完善和数据质量管理。

(4)初期目标：BI 初期开发偏向于系统的基础性建设及快速出效果。在初期一般都会把数据仓库建设到70%左右，完成公共维度和基本事实数据的入库；BI 实施效果得到体现，可以让企业对于 BI价值有直观认识，便于后面工作开展。

(5)中远期目标：中远期重视数据质量管理 and BI 整体应用体系。数据质量问题对 BI应用是致命的，长期来讲需要建立起数据质量管理体系；还要成体系的推进 BI应用，让BI 支持的分析点互相支撑，形成统一的企业级决策分析平台。

优点：

(1)整体规划、分步实施很好地融合了自上而下及自下而上BI 建设思路的各自优点。

(2)注重数据仓库的整体设计，70%左右的前期设计可以建立起数据仓库的基本框架。

(3)分步实施的业务应用可以让 BI不断得到回报，从而让企业更有信心。

(4)同步推进的业务应用可以不断检验数据仓库模型的正确性，从而避免在仓库模型建设上出现大的反复。

缺点及应对措施：

这种策略的不足就是怎么来保证项目的整体规划能力和分步实施效果。

(1)业务规划能力：BI 应用规划不得不面临不同主题间的横向区别，同一主题业务对于不同级别用户的需求也是有区别的，这就要求 BI 规划人员既要懂业务又要懂BI应用，且要有较高的 BI行业咨询能力。

(2)技术方案能力：因为数据仓库的完整性，BI 分步实施是对技术方案的一个挑战，如果没有领先的技术方案是难以保证项目的独立性和设计的整体性，所以团

队必须要有领先的 BI 技术方案。

(3) 项目实施能力：分步实施重视不断出效果，如果前期效果连续失败，那整体规划还是一句空谈，因此保证项目实施效果，特别是初期效果尤其关键，所以要求项目组有较高的项目实施能力；另外可以选择对报表分析需求较大的核心业务部门开始，并获得业务部门的支持，这样成功可能性要大。

3.9 BI 项目风险控制

BI 实施风险包含技术、业务、实施、推广等方面，但 BI 最大风险还是来自于很多人认为 BI 可有可无，有则应该更好，无也没关系的这样一个定位。事实上 BI 的应用是不能将就的，必须实实在在有价值，确实好用才行，所以其难度和风险非同一般。

3.9.1 组织风险

不管什么事情最终都是人在落实，所以组织的合理性是控制风险的第一步，要快速实现企业级 BI 应用需要非常合理的划分组织功能，让每一块的人做自己专业的事，并且每一块的要求门槛要尽量低。如图 3.26 所示的企业级 BI 组织具有较大优势：

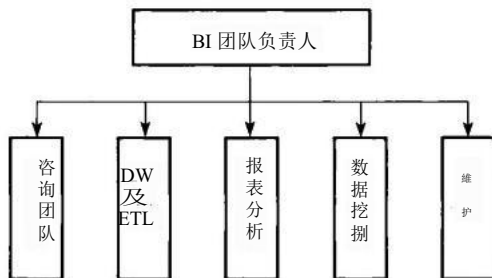
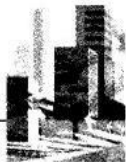


图3.26 BI 团队组织结构

在团队负责人下面设立咨询团队、DW 及 ETL 团队、报表分析团队、数据挖掘团队及 BI 维护团队，咨询团队的设立是我们的一大特点，并且为项目快速出效果起了决定作用。

(1) 咨询团队：是该组织结构里面最为重要的团队，要求团队成员精通 BI 技术及业务，并且有很强的项目规划及推广能力，软实力上乘；能准确对公司 BI 业务进行规划，对规划点能有序的安排实施，随时把控项目存在风险，能根据用户实际情况快速合理地做出项目调整，做到快速推进项目，并取得实际应用成果；这个团队要求有基本的咨询素质、很强的做事欲望、丰富的 BI 实施经验、上乘的协调沟通能力、熟悉咨询方法论且有较高的文档录入及表达能力；这个团队是所有项目的龙



头，事实证明咨询团队建立得好则BI项目的成功率就高，反之失败可能性就大。

(2)DW 及 ETL 团队：团队任务是负责数据仓库、ETL 技术架构及标准定制，负责BI项目的 ETL 开发以及负责 ETL 及 DW 技术研究。在电子商务行业里，DW 面临的大数据处理问题是该团队区别一般行业的重头所在，在 DW 相对较为稳定之后，建议把 DW团队独立出来，专门从事大数据存储及运算的工作，而项目中的ETL 和数据集市设计则由 ETL 团队负责设计及开发。

(3)报表分析：这里报表分析广义上包含大数据查询与导出、报表、多维分析及仪表盘。传统的对内报表分析在电子商务行业也同样需要，没有什么太多不同，都需要企业级仪表盘、OLAP 分析、分析型流程定制等。不同于对内的报表分析，电子商务的 BI应用重心在于对网站用户分析支持，所以要求报表分析的功能更加强大和灵活，网站的报表分析偏向简单及合理设计，用户体验是该工作的重大挑战，从事这部分工作的人员必须有很强的网站用户体验背景或能力。

(4)数据挖掘：专门从事项目的数据挖掘技术开发，承担企业级数据挖掘模型研究和人才培养。因为数据挖掘的特殊性，该团队非常有必要和大学、科研院所进行合作，促进自身技术不断提升，随时掌握最前沿的挖掘技术和应用。

(5)BI 维护：不同于一般业务系统，BI 项目每天有大量数据需要处理，很容易出错，如果团队成员没有较好的BI基础，那么在遇到问题时候只能束手无策，所以需要配备有BI背景的人员进行维护才比较可行，当然这部分维护人员也可以放到公司维护团队去。BI 项目没有数据是万万不行的，所以如何为BI项目快速提供样板数据是维护团队必须解决的问题。

BI 项目团队组织架构如图3. 27所示，采用咨询人员负责效果，项目经理负责实施的组合方式。这样可以保证企业内实施 BI, 在BI团队内部既要负责开发，还要对结果负责，让最懂BI的人来总控有利于保证项目效果和进度。因为咨询推广人员不用参与到每个项目开发，所以往往一个咨询推广人员可以同时带几个项目，这样可以最大限度发挥咨询推广人员的高端及关键作用，有利于项目的并行开发。

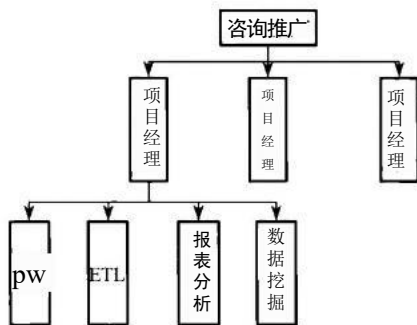


图3. 27 BI项目团队组织架构

3.9.2 技术风险控制

到目前为止，BI在国内已经快速发展了近十年，软件、实施、人才都已经较为成熟，所以只要方法得当技术上不应该存在太大风险，笔者这里建议我们企业要把



下面几件事情做好。

(1)深度理解 Gartner 关于BI产品的评价,慎重选择合适产品,多看看专业点评,应该说BI产品发展变化快, Gartner 的排名就变化快,要特别注意一些领先技术在BI中的应用,比如基于内存的分析、云计算、数据挖掘等。

(2)注意用户体验,很多BI项目失败就在于其用户体验差。

(3)注意实施难度,好的产品开发几天能上手,有些产品开发复杂,甚至安装都很困难。

(4)找个BI实施高手,保证项目的技术架构,这一点非常重要。最忌讳的是业务系统的开发人员转行做BI,因为BI不管是在应用还是在设计上都与业务系统有很大的区别,甚至有矛盾之处。

如果有资源,可以考虑内外结合地实施BI,提高BI项目前期的成功率。目前,市场上已经有很多领先的公司具备较高的BI实施水平,比如东南融通、文思创新、华为、亚信等。

3.9.3 业务风险

业务风险是BI实施的巨大风险。我们最开始在公式化的实施BI,完全遵循用户的要求,可能花了半年做了三个项目,到最后却没有用户使用,后来进行总结并且调整项目经理之后很快重新得到用户认可,并且应用得越来越好,这里不得不说BI业务风险巨大。如何降低业务风险,笔者认为应该注重下面几点。

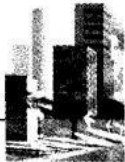
(1)普及BI知识:在实施BI初期,业务人员一般很难想象BI是什么,能给他们带来什么价值,更提不出具体的需求,为此一定要提供一个直观的东西让用户知道BI是什么,比如做一个Demo,在Demo里面把BI的功能充分展现一下,拖拉旋转、上钻下钻、图表制作等。

(2)用户决定需求:虽然用户提不出完整的BI需求,但BI人员要善于理解用户的意图,合理设计其需求,并且最终由用户来决定。

(3)有用就做,不要等:如果用户提不出完整需求,可以把对用户有用的先做,不求全,只要做出来之后用户会真正用起来就行,随着用户对BI的使用越来越多会对BI越来越熟悉,这样离提出全面完善的需求也就不远了。

(4)BI目标是帮助用户:BI功能很强大,但是BI需求来自用户,BI应用需要用户,BI提升完善更是需要用户,BI不可能取代用户的工作,只可能帮助用户更好地工作,让用户可以把更多的时间用来做更有价值的东西。

(5)业务规则需要统一:来自不同部门不同层面的需求,如果是同一个问题,最好用相同的指标定义,并且具有相同的指标名称,不同定义的指标名称一定要区别开来,避免部门间、上下间的信息不一致。



(6) 选择合适的项目经理：因为BI 业务需求来源比较灵活，实施力度比较有弹性，所以 BI 项目经理需要具备较好的沟通理解能力。

3.9.4 实施风险

企业级BI实施我们建议采用整体规划、分步实施的策略，这种策略要成功实施需要很多配套的管理办法。通过这一年多的BI 实施我们形成了完整的以效果为导向的 BI实施方法，基本包括下面几点。

1)QA 关键点评审及控制

- 立项：主要看项目需求价值和实施的可行性，其中项目需求价值一定要非常清楚，没有明确价值点和用户对象的项目不立项，最好是评审人员有丰富的分析基础，并且了解公司业务。项目可行性一般较容易控制，不是太难。

- 需求及设计评审：和一般项目不同，BI 项目DEMO 和实际开发很接近，难以快速模拟，基本上需求、设计、开发一起出来，所以需求和设计放在一起评审较为可行，这样在立项时候的项目实施范围及需求价值要求就有较高预判性，这恰好是我们咨询推广人员的价值所在。

- 试运行：BI 项目最难的地方就在于数据能否准确，一般来讲项目都需要一个试运行阶段，在试运行阶段业务部门安排几个人员进行试用，将发现的问题反馈给项目组进行修改，一般来讲致命性问题(比如数据正确性问题)都要解决了才能上线，其他不是很严重的问题(比如格式调整)可以放到下一期项目进行完善，从而保证项目总体进度较为顺利，不会因为局部小问题导致项目周期大幅延长；另外BI功能对很多用户来讲也是陌生的，用户还是要有初步的使用之后才能提出更多的改进建议，所以没有致命性问题，一般建议还是先推给用户进行使用，在使用中进行完善。

- 正式上线：在试用一段时间之后我们需要给用户发调查问卷，如果用户认为项目对他们有价值并且没有太大的问题那么就可以上线，这是以效果为导向的项目开发流程中重要的一环。项目是否上线绝对不是项目周期说了算，而是真正由用户的需求来决定；项目上线之后我们要持续关注用户的使用效果，监控项目的访问情况，同时也要注意用户意见的收集；另外使用效果不好的项目，如果没有足够的理由将不给予后期开发，如果是BI 人员问题那么需要对BI 人员进行调整，这样能促进用户和开发团队更加重视项目的使用效果。

2)技术方案及标准

整体规划、分步实施最为考验的是项目的技术方案，要求方案能够有前瞻性，这里需要一系列的解决方案；另外快速实施及出效果是分步实施的关键，这样对技术标准也有很高的要求。因此对于BI 项目最起码必须具备如下方案 and 标准：

- BI 系统架构方案;
- BI 系统数据接口技术标准;
- 数据模型设计技术标准;
- ETL 技术标准;
- 系统调度技术标准;
- 前端展现技术标准;
- 数据挖掘技术标准。

3) 项目效果评估及监控

BI项目的效果评估非常有必要,可以取访问人次、访问时长、访问人数、用户覆盖率等来作为评估的关键指标,对这些指标值及趋势进行监控,将有助于评估出项目的价值,通过趋势的预警还便于我们及时发现问题,改进项目。

3.10 小结

传统的数据仓库架构直接移植到电子商务行业会出现“水土不服”的现象,原因在于电子商务行业的数据源多而杂,需要处理和分析的数据量巨大,对数据分析的实时性要求较高,并且需要支持的用户量巨大。要解决这些问题,需要对数据仓库的整体架构做出调整和优化。传统 BI 的数据仓库主要使用关系型数据库来支持离线分析和复杂查询,难以满足大数据处理和实时查询等需求,因此可以结合使用关系型数据库、内存数据库和分布式数据库以满足电子商务行业对于商业智能在数据处理、数据存储等方面的要求。

数据仓库按照数据流向可以分为三层结构,包括数据层、信息层和分析层。

数据层主要是数据源和标准数据接口。信息层包括 ODS、数据仓库、数据集市、ETL、元数据。分析层主要是数据的应用,如数据挖掘、报表分析、数据支持等。

在数据仓库架构上,笔者建议在传统数据仓库的基础上,同时采用分布式数据仓库来支持BI 大数据的存储与计算,可以显著地提升数据处理和数据搜索的性能。在 Hadoop 分布式系统架构中,数据仓库工具 Hive 和基于内存的数据库 HBase 可以分别解决大数据批量任务和实时性查询需求。

正确和完备的数据模型,是BI 进行一切数据分析的基础,是决定数据仓库项目成功与否的重要因素。目前常用的数据仓库数据模型为多维数据模型,这种模型主要以星形模式和雪花形模式在关系数据库系统中存在,两种模式各有优缺点,但在数据仓库中,通常不推荐“雪花化”,查询性能相对OLAP 系统来说更加被重视,而雪花模式会降低数据仓库系统的性能。

ETL 负责将分布的、异构数据源中的数据如关系数据、平面数据文件等抽取

到临时中间层后进行清洗、转换、集成，最后加载到数据仓库或数据集市，成为联机分析处理、数据挖掘的基础。

存储在各种数据源中的数据通过 ETL 存储到数据仓库中，对数据仓库中的数据进行多维建模，再利用多维分析技术满足决策支持或者满足在多维环境下特定的查询和报表需求。多维分析技术的核心是“维”的概念，维指的是分析问题和观察事物的角度。

在数据仓库或多维分析的基础上，对数据进行更深层次的分析 and 知识提取，需要应用到数据挖掘技术。数据挖掘的价值体现于发现深度隐藏的规律或模式，数据挖掘在电子商务行业有广阔的运用空间，如精细化营销、市场评估与预测、客户分析(客户分群、流失预警、忠诚度分析等)、个性化推荐、网站异常预警与监控等。

BI 在电子商务的应用包括对内的运营决策支持和对外的网站用户支持。电子商务对内决策分析系统为分析层、决策层提供决策分析平台，为操作层以及各业务系统提供执行优化的智能帮助，如企业级 KPI 决策系统、部门级绩效考核系统、业务监控和分析以及分析型工作流程。BI 对于网站用户的支持主要体现在个性化服务、异常用户挖掘、网站效果分析与优化和提供数据分析增值与服务。

选择合适的软硬件产品是支持企业实施 BI 的关键问题。BI 产品种类繁多，每家厂商几乎都会宣称自己的产品是如何的好，在构建企业自身的 BI 系统时应该根据企业自身的需求考虑软件硬件产品的选择。数据库方面 DB2、Oracle 都是很好的产品，Teradata 和 EMC 公司的 Greenplum 在大数据处理方面有着很强的能力。根据 Gartner 在2012年2月发布的数据仓库魔力象限显示，Teradata 在数据仓库方面处于领导者地位。主流的 OLAP 工具有 BO、Cognos 以及 Qlikview 等，Gartner 前端报表工具最具魅力象限连续几年纳入了 Qlikview 这个产品，这个产品代表了基于内存的报表分析技术，美观简单、用户体验极佳。数据挖掘产品领域，有 SAS、SPSS 等两大厂家，而像 IBM、Teradata 也都有自己的挖掘工具。做好数据挖掘仅仅有工具是不够的，更需要数据挖掘人员对数据挖掘的理解和认识。

BI 的实施策略，有代表性的是数据仓库支持 OLAP 理念的两位创始人 Bill Inmon 和 Ralph Kimball, Bill Inmon 倡导自上而下实施策略，而 Ralph Kimball 则提出不同看法，认为数据仓库应该实施自下而上的实施策略。笔者认为可以采用第三种策略，那就是整体规划、分步实施，这种方式可以融合两种策略的优点，又可以避免其中的短处。

BI 实施风险包含技术、业务、实施、推广等方面，但 BI 最大风险还是来自于很多人认为 BI 可有可无，有则应该更好，无也没关系的这样一个定位。笔者总结了 BI 实施风险主要有组织风险、技术风险、业务风险、实施风险。

第四章 BI 对内决策支持平台

BI 通过整合和分析企业运营和业务系统产生的数据能够对企业内部的高层领导、中层管理人员以及底层操作员提供决策支持和分析支持。高层仪表盘以及企业级绩效分析可以帮助公司高层对公司的经营状况有一个清晰的认识，并作出相应的决策和计划。各部门的主题性 OLAP 分析则为中层管理人员提供部门绩效以及关键 KPI 的监控，便于中层管理人员了解任务完成情况和质量。如果在一线操作人员的工作流程中加入 BI 的分析功能，则能帮助其提高工作效率和工作质量。下文将从以上几个方面进行展开。

4.1 高层仪表盘

商业智能的应用目的是为了给用户提供了数据分析支持。从公司整体的角度来讲，同样为用户提供数据支持，对公司高层领导的支持无疑会比对普通员工的支持产生的价值更大。因此我们首先开始介绍商业智能是如何对公司高层领导进行服务的。

4.1.1 定义

仪表盘 (Dashboard) 的本义是通过将各种可视化的指示器、控件有效地组织成一个便于驾驶员随时获取交通工具运行状态，并根据具体情况及时做出判断，进行决策以及采取相应措施的事物。仪表盘在商业智能系统中是一个比较有特色的内容。通过像汽车油表类似的图形，系统可以向用户形象地展现数据分析的结果，避免用户看到过多的数字。

仪表盘所具备的可视化呈现信息，且有助于判断、监控并支持决策的特性，让它逐渐成为商业智能中不可缺少的一个组成部分。仪表盘能使数据可视化，从而使数据变得更有意义，更容易理解，方便用户通过数据表现为相关工作制定相应策略。

4.1.2 目标和价值

对于企业决策者也就是公司的高层人员而言，必须对企业运作中的各种信息



和数据进行监控、分析、管理，并据此进行判断、决策，进而采取措施。随着各种数据处理和分析应用程序的涌现和升级，企业发现需要处理和分析的信息以及数据越来越多，越来越复杂，在处理和与分析信息中所投入的时间成本也越来越大。而使用仪表盘能可视化地呈现信息，且有助于判断、监控并支持决策，从而有效地提升信息系统的实时信息处理能力，这一特性使得“仪表盘”成为缓解和解决上述问题的重要途径，并日益赢得使用者的青睐。

高层仪表盘的目标是以简洁、直观的界面，展现企业各环节的经营数据，并以丰富的展现形式为企业决策者提供分析和管理上的帮助，方便决策者洞察企业的运营状况。

4.1.3 设计关键点

高层仪表盘的用户为公司的高层，BI 人员可能没有太多的机会直接向高层领导了解需求，因此就要求 BI 人员自己要站在高层领导的角度来设计需求。笔者对高层仪表盘的设计，总结了下面几个关键点。

1) 监控关键绩效指标

仪表盘中呈现的必须是对于客户来说最重要和最关心的指标，这是吸引客户和留住客户的关键。特别对于高层仪表盘来说，高层用户更加关注企业的宏观指标，不会花时间来关心一些无关大局的指标。如果在实现仪表盘时没有将指标需求整理好，将一大堆不重要的指标掺杂在重要的指标当中，一方面会影响用户的体验，另一方面可能会影响用户对系统专业性方面的信心，最终导致用户对系统认可程度下降，导致项目的失败。

2) 准确性

信息仪表盘上呈现的所有信息必须完全准确。这一点对于每一个商业智能项目来说都是必需的，支撑信息仪表盘的数据必须经过测试和校验。一个错误的数字有时候甚至比没有这个数据产生的负面影响更加大。很有可能一百个正确的指标不会增加用户的满意度，而一两个错误的指标就会使得某个重要的用户对系统失去信心。

3) 方便直观

方便直观的展现方式才能让客户轻松上手。对系统要展现的各种数据要用最合适的表现形式进行展现，相对于数据本身，各种指标的趋势图、累计趋势图、饼图都是直观的表现形式；太过明细的数据在界面上应尽量避免；界面必须友好和便于操作，尽量保证用户不需要通过任何的操作就能看到想看的数据，尽量减少用户的操作。

同时，要求系统能够对实现定义的阈值作出响应，为用户实时的、可视化的呈现异常预警，同时启动相应的预警机制(如声音报警、电子邮件告知、高亮度闪烁

等),及时触发用户对于关键事件的警觉。

4) 分析与追踪

虽然数据应该以简单显示为主,但仪表盘还要具备让用户进行假设分析、假定推测等导向性分析功能,让用户可以通过可视化方式选择各种钻取方式,进行比较、分析和推理。同时要具备允许用户定制并追踪其关注的各项指标体系的能力。这样,仪表盘可以提升用户对于各种相互关联的业务变量的理解力和洞察力。在设计中对一些关键的维度要实现钻取功能,比如与销售相关的指标中不但要有时间指标的趋势图,对地区、部门等维度都需要提供查看功能。

5) 快速响应

用户在信息仪表盘上进行检索、筛选操作及生成报表的时候不应该感到明显延时。这里主要是用户体验方面的问题,为实现这一目的,在后台数据处理时要更多的将数据进行汇总,减小相关报表的数据量,并最好采用内存分析型的商业智能报表工具,将用户界面端响应提高到更快。

6) 基于 Web

这点也是用户体验相关的功能,尽量让用户能够通过Web 浏览器来登录仪表盘,而不需要安装客户端。一方面方便报表功能的升级,另一方面减少用户操作和减少培训工作的时间。

7) 安全

最后商业智能系统管理员需要方便地管理仪表盘软件、进行权限授权和异常跟踪。信息仪表盘软件也要提供数据加密功能,以确保敏感数据在Web 上传输的安全。

4.1.4 高层仪表盘在焦点科技的应用

在焦点科技的商业智能应用中,我们将这个仪表盘的概念进行实现,将企业高层领导关心的指标进行整理和显示。下面我们将焦点科技在公司高层领导仪表盘部分的应用进行介绍。

1) 功能划分

首先根据B2B行业的特点以及公司的实际情况,将公司分解为销售、运营、平台、人力资源、客户、市场分析等六个模块(图4.1)。这里的模块划分是根据公司的实际情况进行调整的。对于互联网B2B行业来说,客户和交易行为都是在网络平台上实现的,这样平台的使用情况展现就是一个关键的模块,因此我们将“平台”这个模块独立出来,与销售运营等模块并列展现,而像传统行业的供应商管理这样的模块就可以说是不重要的,可以不放入到高层领导仪表盘功能中。

2) 确定指标

因为我们前期整理了多个模块,每个模块都有各自的侧重点,都有各自的指

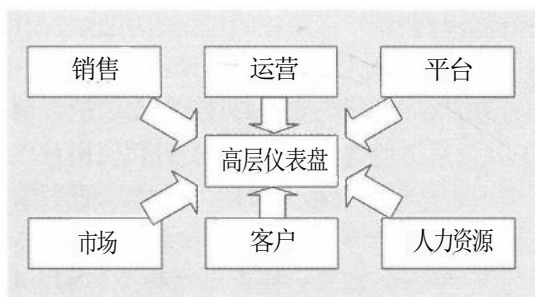


图4.1 模块划分

标。通过前期的一系列研究，所有的模块共整理出来一百多项指标。而从上百个指标集中如何选取对于领导来说最关心的和最关键的指标进行归纳和总结是至关重要的。这里介绍一些在项目实施过程中的经验。

首先要尽量选择结果型指标，而不是过程型指标。这样能够使得报表重点突出，而不是仅仅罗列数据。比如销售部门中“销售金额”就是一个很重要的指标，而“销售人员出勤率”对于公司高层领导来说就相对不重要，可以不必展现出来。

其次要尽量选择直接指标，而不是间接指标。这样能够方便领导快速得到结果、定位问题。比如使用“续约率”就不比“未续约率”更能直观表现出客户效果。

同时指标定义要明确，容易理解，不要含糊或者有歧义。比如平台部分有“活动用户数”（指统计时间段内有过登录行为的用户）和“活跃用户数”（指统计时间段内登录次数达到一定阈值的用户）这两个指标，都是从用户登录网站角度对平台进行衡量的指标。从字面上理解就比较容易混淆，那么在界面上这两个指标就不要全显示出来。

3) 报表设计

焦点的高层领导仪表盘在表现形式上还是一系列的报表，报表分为近期指标日监控、年度指标监控、综合指标列表这三个主题呈现给用户。

- 近期指标日监控功能：虽然对于高层来说月度指标才更加有用，但最重要的四个到五个指标领导们还是需要每日监控的，这里我们设计时展现的是近两周的指标趋势对比，而选取的指标都是网站平台的关键指标。通过这些指标的监控，能够及时发现异常、掌握网站动态、制定临时策略。采用的界面表现方式为两周日趋势对比图。

- 年度指标监控功能：这个首页模块将销售、运营、平台、客户、人力资源、市场分析等各个模块进行了整合，每个模块选择两个到三个指标进行详细的展示，展现的形式包括月度趋势图、月度累计趋势图。一般来讲界面上展示的都是本年度指标趋势与去年指标趋势的对比，方便用户观察数据趋势。而且提供通过一个按

钮实现查看去年指标趋势与前年指标趋势的对比的功能。采用的表现方式为当前指标同比对比图、两年月趋势对比图、两年月累计趋势对比图和饼图等形式。

这里需要根据实际的指标进行确定图形类型,比如销售部分的销售金额就需要使用月度趋势图来表现每月收入趋势情况,还需要使用月度累计趋势图来查看销售目标完成情况,而人力资源部分的人力结构情况就需要用饼图形式来表现了。

- 综合指标列表功能:最后还挑选出三十个到四十个的关键指标显示在表格中,这个表格按照月维度来显示,主要将本月的指标数值与上月指标值、去年同期指标值进行对比,并将波动异常的指标用醒目的颜色明确的表现出来。

通过这样一系列的图表,我们实现了焦点科技的高层领导仪表盘。这个系统能够满足领导们的日常数据需要,获取信息方便及时。对公司领导在更好地制定公司决策方面带来了很大的价值。

4.2 企业级绩效

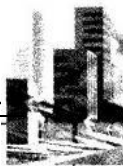
绩效管理是个大题目,是个系统工程,包括绩效目标、绩效辅导、绩效考核、绩效控制、绩效评估、绩效沟通等多个方面。商业智能在绩效管理上面也有很大的发挥空间,主要应用在绩效目标制定和绩效考核中。在企业进行绩效管理过程中 KPI 指标的制定是关键点。

4.2.1 定义

关键绩效指标法(Key Performance Indicator,简称 KPI)[10],它是把对绩效的评估简化为对几个关键指标的考核,将关键指标当作评估标准,把员工的绩效与关键指标作出比较地评估方法,在一定程度上可以说是目标管理法与帕累托定律的有效结合。关键指标必须符合 SMART 原则:具体性(Specific)、衡量性(Measurable)、可达性(Attainable)、现实性(Realistic)、时限性(Time-based)。

建立 KPI 指标的要点在于流程性、计划性和系统性。首先明确企业的战略目标,并在企业会议上利用“头脑风暴法”和“鱼骨分析法”找出企业的业务重点,也就是企业价值评估的重点。然后,再用“头脑风暴法”找出这些关键业务领域的关键业绩指标(KPI),即企业级KPI。

一般来说,指标指的是从哪些方面衡量或评价工作,解决“评价什么”的问题;而标准指的是在各个指标上分别应该达到什么样的水平,解决“被评价者怎样做,做多少”的问题。最后还需要对指标进行审核。审核主要是为了确保这些关键绩效指标能够全面、客观地反映被评价对象的绩效,而且易于操作。



4.2.2 目标和价值

目前很多企业在绩效考核中面临着问题，这些企业大多建设实施了很多系统，实现了业务处理的自动化，但在管理决策上很多企业并未真正实现管理信息自动化。这样企业绩效管理的处理过程是，系统进行业务处理产生了大量业务数据，管理部门在此基础上把数据从业务系统中倒出，利用数据库处理数据或利用 EXCEL 手工进行数据整理和分析，按照企业的绩效管理的考核方法，如 KPI 的各种指标等进行计算和分析，最后得出绩效考核的“数量”结果。

这个过程的后果是：一方面 ERP 产生了海量的业务数据，另一方面企业却不能自动从这个海量数据中获得管理决策、绩效管理有用及时的信息，必须借助于数据二次“开发”，即数据库编程或手工处理，缺少数据挖掘的技术支持。这样也就产生了海量数据和管理层信息孤岛的现象，管理层决策还要大量依靠主观的经验判断。而这些问题能够通过 BI 绩效管理的解决方案来解决。

4.2.3 设计关键点

商业智能系统在解决企业 KPI 管理的问题大概可以分为以下几个方面：

1) 发生了什么？

提供事先预制好的报告、报表或者仪表盘，利用集中管理的 KPI，解决企业运营绩效问题，监控企业的发展，实现复杂的报告用简单的方式表象出来。

2) 为何发生？

业务部门可以从固定的报表、报告和一些关键的 KPI 中得到很多相关的信息或者计算机自动预警，但是当发现问题时，需要了解为何发生了问题。这时，就需要即席查询和 OLAP 分析。业务分析员经常需要自己根据问题的需要完成自己的分析和报告。在很多情况下，业务分析员和决策制定者需要 BI 工具，通过访问集成好的数据仓库，获得需要的信息。

3) 现在发生了什么？

这个层次的 BI 是实时的信息分析。BI 帮助决策层建立当前情况下的业务战略和决策。即将会发生什么？客户发现仅仅了解现在还远远不够，将来会发生什么，预算、市场趋势分析、风险的预测和评估都属于该范畴。

4) 还会发生什么？

在现有的 BI 系统中，还需要统计分析的功能，来帮助分析客户的行为、预测客户的趋势、预言企业业务的发展、辨认欺诈行为、制定预算等。

需要这种模型的客户群是非常大的，但是在企业中应用的用户不会太多。这种模型需要复杂的算法、统计模型和大量的数据，所以需要具备大数据量的处理能

力，并行算法和网格计算是必要的。

总体来说商业智能系统能够满足公司在运营过程中事先预测，事中监控，事后分析各阶段对数据统计的要求，从而对企业的绩效管理带来很大程度上的推动效果。

4.2.4 企业级绩效管理在焦点科技的应用

在焦点科技的商业智能应用中，企业级的绩效管理也发挥了很大的作用。实施步骤分为确立 KPI 指标、建立指标维度、报表展现等几个部分。

1) 确立 KPI 指标

首先根据业务的实际情况和各部门制定的部门长期目标，以及前文中提到的 SMART 原则，确立了一系列的部门指标，也就是制定了部门 KPI。根据二八法则，这些 KPI 指标需要能够满足80%的业务需求。同时制定 KPI 指标不能纸上谈兵，还是要根据实际情况和业务部门多多沟通，达成一致。

同时在建立 KPI 指标时有几点需要注意：

(1)KPI 是关键绩效指标，不是一般所指的绩效指标。

(2) 指标的确立要根据实际数据情况，不能凭空想象。因为我们提出的是指标，而不是目标。

(3) 对于绩效结果需要客观理性地看待。KPI 是绩效指标，不是能力或态度指标。

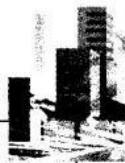
2) 建立指标维度

确定了KPI 指标之后，还需要对这些指标进行层层分解。在系统的表现就是能够通过多个维度来描述这个指标。这样在发现某个指标有异常的时候，能够通过各个维度的检查，发现产生问题的直接原因。

比如在销售金额这个指标的考察中发现，某个月的销售金额对比前面几个月和去年同期的月度都有了一定的波动，那么考核者和被考核者都需要及时发现问题所在，及时制定出相关的策略。而商业智能系统会将与销售金额相关的维度展现出来，比如查看哪个部门或分公司的销售业绩出现了下滑，哪个省市或者哪个行业的销售金额下降比较明显等。同时还会提供相关业务的走势，与销售金额在逻辑上有一定关联的指标，比如客户活跃度、客户续约率、市场宏观走势等指标，也可以一起提供给用户。这样可以使得用户能够快速定位问题，找到合适的解决方法。从这个角度来说，商业智能系统不但可以让用户得到绩效考核的最终结果，同时还能够在绩效考核的过程中给用户提供更多的支持，帮助用户更好地完成制定的 KPI 指标。

3) 报表展现

在确定了KPI 指标和指标相关的维度后，如何进行指标的展现也是一个很关



键的地方。好的设计只有让用户实实在在地感受到才能提供更好的价值。

在焦点的企业级KPI 报表系统设计中，我们对功能进行了划分。报表界面显示由近期 KPI 指标日监控、年度 KPI 指标监控、KPI 指标维度报表等部分组成。

(1) 近期 KPI 指标日监控功能

对于比较关键的四个到五个指标是需要进行每日监控的，这里每日监控的指标不一定是设计的最终的KPI 指标，可以是 KPI 实现过程中涉及的指标，也可以是对考核结果产生影响的指标。

设计时展现的是近两周的指标趋势对比。通过对这些指标的监控，能够及时发现异常、掌握数据的最新动态、制定临时策略。

采用的界面表现方式为指标的两周日趋势对比图。

(2) 年度 KPI 指标监控功能

考核往往有一个时间的过程。如是以一年作为一个考核周期，还是以一个月作为考核周期，或者是以一个星期作为一个考核周期等。不同的企业对于这个周期的定义是不同的。不过一般来说采用一年作为考核周期是最常见的。因此我们设计了年度 KPI 指标监控功能，展现的形式包括月度趋势图、月度累计趋势图。一般来讲界面上展示的都是本年度 KPI 指标趋势与去年 KPI 指标的对比趋势图。掌握了这些信息后，用户对个人或者部门的年终绩效指标会有一个直观的把握。根据实际的数据情况作出相应的响应。

59

(3) KPI 指标维度报表功能

这个功能主要是指指标维度在界面上的展示。如前面功能介绍中所述，当用户发现一个指标有问题时，就会到与维度相关的系列报表中查找问题所在。

报表实现上通常会在 KPI 指标报表上加入钻取功能。分析各种维度上的变化趋势，在维度比较多时，我们在设计上还加入了多个报表联动的功能。如在地区销售饼图中选择了北京地区，那么月度回款金额指标的趋势图就会相应的变为北京地区的月度回款金额指标趋势图。这样会更加方便用户对问题的查找。

当然，这样的功能对于一些商业智能软件来说实现起来比较麻烦，有时会带来大量的代码编写，很难维护。但如果没有这样的功能，用户在信息的获取方面就不是那么透明，甚至会带来对系统的不信任。焦点商业智能系统采用的是将Qlik-View 作为报表开发工具，不但能够很方便地完成上述的钻取旋转等功能，而且支持用户在线灵活调整报表，获得自己想看的数据。很大程度地减小了开发人员的工作量，同时带来了更好的用户体验。从这个角度来说一个好的报表工具的选取，对项目的成功确实具有相当好的推动意义。

通过使用上面所述的一系列 KPI 绩效考核的报表功能，用户能够更加明确地了解自己在实际工作中的进度，能够主动的根据数据寻求变化调整工作重心，而不

是被动的跟随领导的指示，对工作的促进作用还是比较明显的。

4.3 主题性 OLAP 分析

4.3.1 目的和作用

企业决策者对各种信息和数据进行监控、分析、管理、判断、决策。根据上文叙述，BI 为企业决策者提供了高层仪表盘与企业级的绩效，对于各部门来说同样需要对部门信息与数据进行监控、分析、管理、判断与决策，BI 为部门提供了主题性的 OLAP 分析。

BI 部门级 OLAP 分析，更专注于一个特定部门，同样提供决策层、分析层、操作层的支持。主要对部门工作情况进行监控、监督，通过报表的分析，有效地发现问题，改进问题，对部门工作实施改进提供数据依据，更好地安排工作计划，提升员工工作效率，有效地激励员工工作积极性。

根据公司组织架构，以下分别详细介绍 BI 提供的部门级 OLAP 分析，主要包括销售分析、运营分析、订单分析、客服分析、买家服务分析。

4.3.2 销售分析

销售部门是企业效益的最直接的实现者。如何提升企业的销售额，是每家企业时刻都在思考的问题，那么 BI 又能为企业销售的提升做些什么呢？

按照各种角色划分，销售的职能可以简单归纳为以下几点：

(1) 销售部门管理者(部门经理): 制订年度销售计划，并进行任务分解；管理、监督销售计划执行情况；研究把握销售人员需求，调动销售人员积极性。

(2) 销售人员: 执行销售工作安排，催收回款。

部门经理制订销售计划和分解任务需要参考历史的计划完成情况，并且会根据当前的执行情况调整计划。部门经理同时需要关注销售人员为了达到目标所进行的努力，并结合计划完成情况综合度量个人的绩效；部门经理还需要激励销售人员的工作热情，比如给予业绩较好的销售人员一定的奖励，鼓励销售人员之间的良性竞争。对于销售人员来说，如果能够及时了解自己的任务完成情况，就能合理安排自己的工作计划。销售人员的主要工作是向客户销售产品，为公司创造盈利。如果销售人员能够了解客户，对客户进行有针对性的销售活动，就能提高销售的成功率。

结合销售业务进行综合分析，销售的 BI 分析可以从以下几个方面进行：

1) 销售计划监控以及业绩统计

销售的计划及业绩统计目的是为了能够更好地监控销售计划完成度，有合理的依



据修改计划，评估销售业绩。

通过报表销售主管可以清晰地掌握整体的销售情况，进行环比、同比分析，了解销售额是否增长还是下滑。动态、多维地展现销售变化情况。销售人员则可了解自己的销售金额、排名、完成度等个人的销售情况。

按照年、季度、月这三个时间维度展现销售计划与完成度，提供部门以及个人的计划完成情况。根据公司销售提成制度和销售奖励规则，自动统计出销售人员的提成以及应该获得奖励的销售人员。部门业绩可以从合同情况、工作量、部门回款等指标进行展现。

2) 销售工作过程监控

销售过程的监控主要目的是针对销售工作量与工作品质的统计。通过报表销售主管可监控销售人员工作量的各项指标，完成绩效评估、修改完善绩效制度，并对绩效制度的效果进行评估，报表更可动态、多维地展现统计分析结果，为主管提供数据支持。

工作量统计是按月进行分部门统计销售人员的电话时长、唯一电话量、呼出呼入比、分配添加客户数、手动添加客户数、手动开放客户数、系统开放客户数、手动置撞单、添加客户总数、系统客户联系客户数、老客户总数、老客户联系客户数、老客户流失数。

工作电话明细统计是统计每天每个销售人员的通话记录，主要显示第一通电话和最后一通电话时间。

工作品质统计主要是按月进行分部门统计销售人员的添加客户留存率、被系统开放客户数、系统客户联系率、老客户回访率、老客户断约率。

3) 销售业务支持-精确化营销

精确化营销的目的是更有针对性地销售我们的产品，恰当地、贴切地对市场进行细分。通过电子媒介、电话访问、邮寄、互联网等方式，建立客户资料库，通过挖掘分析，确定可能购买的客户，为销售部提供更可靠的推广方案。

精确化营销分析是针对会员特征向销售人员推荐营销策略，帮助其快速了解会员特征，提升销售效果；渠道拓展分析，能帮助销售人员在快速选择好的销售渠道进行营销，节省销售时间；销售跟踪分析，能帮助销售人员理顺各个会员的销售历史信息，让销售人员进行跟踪时把握跟踪动向及矫正跟踪策略。会员效果分析能帮助客服做好客户维系工作，把握客户目前的体验情况，有针对性地制定维系策略，提升会员的满意度水平。

精确化营销对用户进行分类，向更具针对性的用户推荐针对性的产品。可针对用户的访问、询盘、已购买服务、用户行业、用户基本信息等多个指标进行分析，挖掘出具有销售潜力的用户对象，提升销售效果。

4.3.3 运营分析

运营分析主要针对信息审核部门(简称信息部),提供其他部门OLAP 分析。信息部主要对注册网站的会员提交的信息进行审核,包括提交的公司、产品、图片、商情等信息审核工作。

信息部的工作是繁杂的,每个员工一天中需要完成多项任务,每项任务都需要记录完成时长。主管则要统计汇总每个员工完成工作的情况,监督工作完成质量,主要监控审核的工作量与审核质量。

对于信息部工作内容的整体 BI 分析框架为以下几点(图4.2):

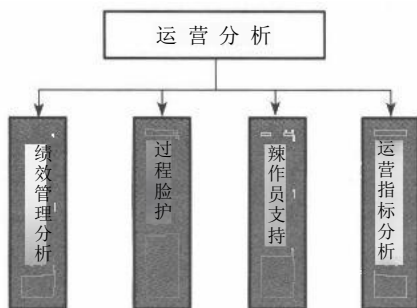


图4.2 BI运营分析

1) 部门绩效管理分析

绩效管理分析主要为资讯发布量、审核量提供统计分析。

通过报表分析信息部主管监督信息部员工工作量的各项指标,完成绩效评估、修改完善绩效制度,并对绩效制度的效果进行评估,报表更可动态、多维地展现统计分析结果,为主管提供数据支持。

资讯统计分析主要是针对信息部员工在 MIC 商业资讯中发布的文章数量进行统计、监控并实施员工绩效。传统的资讯统计信息部有专人在编辑系统中手工汇总统计需要的发布量。任何条件的变更都需要重新查询汇总。

通过资讯报表分析主要了解每个员工发布资讯的数量详情、每天信息部资讯发布的总量并与之前的相应统计时间对比,控制发布数量、跟踪变化趋势。可有效地提高资讯发布统计的便捷性、快速性,并提供 OLAP 分析,支持部门绩效管理。

审核量统计分析主要是针对信息部员工审核公司、产品、商情等信息数量进行统计、监控并实施员工绩效。传统的资讯统计信息部有专人在编辑系统中手工汇总统计需要审核量,或向系统运维部提出数据需求,汇总后供信息部统计使用。任何条件的变更都需要重新查询汇总。



审核报表最重要的是关注一段时间的审核总量，并细分到公司审核总量、产品审核总量、商情审核总量等。审核后会被员工判定为通过、返回、暂停、删除这四种状态，统计通过量、返回量、暂停量、删除量，并细分到公司、产品、商情，对通过率、返回率等做出计算。统计的指标主要就是以上这些，只是根据查询的维度不同，给出不同的统计数量。维度有信息部各部门、行业组、行业队列、员工个人、时间。BI审核统计分析可有效地提高审核量统计的便捷性、快速性，并提供 OLAP 分析，支持部门绩效管理。

2) 过程监控

过程监控为信息部员工各项工作内容与工作时长进行统计汇总，从而有效地提升信息部员工汇总个人工作所花费的工作时间，更简洁、准确地提供报表数据。

统计员工除审核以外每天的工作内容、各项工作内容工作时长。传统工作时长统计由员工自己每天手工记录，并发给自己的主管，主管进行手工汇总，对员工工作内容、完成量、完成时长进行月、年的汇总与记录。

工作时长报表统计可详细提供各员工各项工作内容的时长，员工只需登录报表查看自己的工作情况即可。主管则可直接在报表中查看每个员工的工作情况，自由进行各维度时间、内容的汇总分析，可有效地提高工作时长统计的便捷性、快速性，提供部门工作的过程监控分析。

3) 操作员支持

操作员支持主要由BI数据挖掘项目承担，根据用户的行为、基本信息进行分类，分析异常的用户特征，挖掘异常用户。

信息部数据挖掘项目主要目的是减少员工对异常用户行为的检测困难度与检测时间，提升操作员工作效率的过程。检测出用户的异常行为能够更好的净化 MIC 平台的运营环境，保证会员的利益，提升 MIC 信誉。但是长期以来信息部检测异常用户主要靠员工手工比对，效率低下、检测不全面，工作量非常巨大，人力成本过高，对 MIC 平台的稳定运营大为不利。

焦点科技信息部挖掘项目有：反欺诈、反“钓鱼”、自动化编辑。主要为信息部员工有效地查找出异常用户提供高效的解决方案。员工可登录 BI 系统，查看挖掘出的异常用户，再进行手工确认，大大减少其工作量。

4) 运营指标分析

运营分析统计主要针对 MIC 网站运营情况进行统计。传统的统计主要是由系统运维部门汇总后定期提供，数据不够灵活，分析维度不够多样。BI 报表可有效提高统计便捷性与快捷性，多角度展现、自由汇总数据。历史数据无需在本机保存，数据仓库可高效安全地保存大量数据，用户只需在BI系统中直接进行分析即可。主要包括产品数量统计、产品属性使用量统计、邮箱有效性统计、会员类型统

计、公司活跃度统计等。为网站运营的监控、分析、管理、判断、决策提供依据。

4.3.4 订单分析

随着电子商务的不断发展成熟,平台起家的服务提供商往往会随着自身实力的不断发展长大,结合原有业务,逐步涉足产业链中其他业务领域,追求产业链式的健康发展。与此同时,产品与服务也必将更加丰富多样。但产品链不断优化升级的同时,也给电子商务企业的订单部门带来了巨大的工作压力。

如焦点科技旗下就有:中国制造网、新一站保险、百卓采购网、领动、爱聘才、百分百物流网、文笔天天网、全球汽配网、焦点商学院、商聚园等产品,其中,就中国制造网而言,就可以衍生出如 Toprank、AS 认证、AR 报告等其他多种形式的子产品和服务。因此,对于一个规模化的电子商务企业而言,其订单生成量以及订单处理量都是相当大的。

1) 目标和价值

面对日益多元化的产品订单以及日益复杂的订单分析,此时,如果订单部依旧依靠传统的手工方式统计和分析各类订单,那么不仅效率低下、工作量巨大、人力成本过高,而且无法达到实时掌握订单动态、及时发现异常信息以及多维度的深入分析订单的效果。

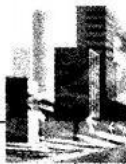
(1)对于操作员:不利于订单处理人员实时了解自身的工作情况,需对历史订单处理量进行二次转换才可转换为自己的工作量,繁琐的同时也无法直观地了解自身的工作效果以及在部门中所处位置等信息,不利于员工根据客观实情调整自己的工作。

(2)对于决策者、管理者:管理者也需每月手工汇总统计下属的工作量数据,报表时间大大地影响了其分析、管理、判断和决策的时间,且每月固定报表时间导致的信息延后性无法帮助管理者实时地掌握部门的整体工作情况,及时地调整和控制。

上述问题可以在引入商业智能技术后将得到解决。通过 BI 技术中的数据仓库、数据挖掘以报表展现技术可以实现自动化统计和分析订单信息、订单审核量、续约率等数据,并且可以根据需求从时间、地域、行业等维度全方位展现。

2)BI 订单部门人员绩效统计

利用BI技术,订单人员不需要手工汇总工作数据,只需直接登录BI系统就可以实时地了解自己的工作情况、部门的整体情况以及自身所处的位置,有助于其实现自我管理和自我控制。而部门管理者也可通过随时登录 BI 系统,实现了部门整体工作情况的实时把握,大大降低其报表时间,将更多的精力放在绩效考核、工作指导以及工作效果提升上。具体可以从对工作量、工作效果(一般为审核合格率



等)两个指标的统计分析以及监控上实现。

(1)部门级完成订单量:以产品、部门以及时间为维度,统计和汇总完成的订单量。该指标有助于管理者了解各个子部门的工作动态。同时设定不同产品线的完成规格,有助于管理者掌握不同产品线工作量差异,针对具体的优劣情况,进行相应的调整,展开对相应部门的管理和指导。

(2)人均完成订单量:以部门、产品以及时间为维度,统计人均订单完成量。该指标有助于管理者制定和调整工作量标准,同时有助于其了解部门每个工作人员工作负荷情况、工作稳定性以及成长空间,有助于展开相应奖惩或激励机制对员工进行管理;对个人而言,该指标有助于员工进行自我定位,并据此进行自我学习和自我提升。

(3)人均订单审核合格率:以部门、产品线以及时间维度,统计人均订单审核合格率。该指标有助于管理者了解部门工作质量,从工作效果角度分析部门和员工的工作表现,进行相应的控制和调整,并据此指导员工学习和提升。对个人而言,通过该指标结合工作量,有助于其更加全面的定位和提升。

(4)个人完成订单量/个人订单审核合格率:以个人为纬度,以订单量和审核合格率为指标,统计个人的工作完成情况。通过登录 BI 系统,部门管理者可以实时访问每个订单人员历史至今的所有工作明细记录,而订单人员则可以获取个人历史至今的所有工作明细记录。这些个人工作量指标,有助于管理者了解每个人的实际情况、成长情况。而对于个人而言,结合人均数据,可以更科学而准确地了解自己的工作实情,对于再返岗的工作将产生积极地指导作用。

65

3)BI 订单数据分析

(1)订单量分析:对用户购买网站服务的订单进行数量上的统计和分析,以产品类别、时间、地域以及行业等为维度,新订单数、续约订单数为指标。通过订单量的统计可以直接反应销售和客服人员的工作情况。而结合不同纬度进行分析,则可以了解公司产品在不同市场上的表现,利于公司了解对应市场的渗透率,有针对性地调整市场策略。

(2)续约率分析:对用户购买网站服务的续约情况进行分析,以时间、行业、地区为维度,当期续约率、半年续约率、延期续约率为指标。续约率的高低直接反映了客户对网站产品的满意度,同时也从侧面反映出销售人员和客服人员的工作情况。续约率分析可以让相关人员了解网站服务各个行业和各个地区受用户认可的程度,从而为网站有针对性地调整和实施网站服务优化和调整提供依据。

4.3.5 客服分析

BI 的运用越来越广泛,BI 对内和对外的支持功能也日益强大,目前公司的对

内BI实施已经步入正轨。BI对客服的作用可以从大方向上概括为两大块,即内部控制管理、外部工作支持。

BI对客服的内部工作支持,主要在对客服工作的过程及结果各项指标监控,这要涉及客服平时工作的经常性内容,如电话量、邮件量、通话品质等,结果指标如续约率、客户满意度评分等。

BI对客服的外部工作支持,主要在于对客户的多角度分析,这些深入分析能够有效地帮助客服了解自己的客户,做到知己知彼,高效服务。具体BI对于客户可以从下面的层面进行:客服考核系统实现了对部门及人员的监控、考核、辅助指导;客户推广效果分析实现了对客户的深度分析。

1) 客服考核系统

客服考核系统旨在辅助客服部门各层主管把握部门总体情况及各人员表现,融合了BI的自上而下,由战略策略—分析决策—执行的逐层下发的思想。

针对部门层面的表现监控,首先进行部门之间的工作过程指标、结果指标的对比,体现出各部门差异及优势;其次展现全部部门工作过程指标、结果指标的表现优异的员工,让领导层了解人员的情况,便于开展激励措施。

针对各分部门的表现监控,首先进行部门总体的工作过程指标的展现,让各自部门领导把握自己部门的总体工作趋势,进行管理和指导;其次展现部门全体员工结果指标,让各主管能够清晰地把握部门各人员的表现情况,进行激励考核工作。

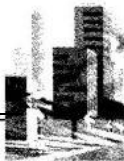
2) 客户推广效果分析系统

客户推广效果分析系统旨在分析客户的网站信息及表现情况,为客服人员提供有效的辅助服务工具,使其能够及时掌握客户的信息,做到有的放矢,精确化服务。BI从客户网站的各种行为信息着手,对客户进行深入的分析,辅助客服深度了解自己服务的客户,从更专业的角度更好地服务于客户,提高客户的满意度,进而提高客户忠诚度和价值。客户推广分析系统主要分为客户网站推广表现分析、客户产品被关注分析、客户账户完善程序分析。

客户网站推广表现分析:此部分对客服在服务期内在网站上获取的访问、询盘进行多维分析(时间趋势、区域分布),客户通过分析结果可以了解到自己目前在网站的推广效果的优劣,以及自己的产品在地区上受关注的差异;客服可以针对客户的表现提出一些建议,更好地服务于客户,帮助客户提升推广效果。

客户产品被关注分析:此部分针对客户产品进行细化分析,以此辅助客户了解自己产品的访问、询盘情况,帮助客户进行产品设置优化。

客户账户完善程序分析:此部分将客户的在网站的全部信息综合展现出来,主



要从账户、产品、商情、展示厅模块，提示客户哪部分需要完善，实现服务品质的提升。

4.3.6 买家服务分析

电子商务网站的盈利来源主要是收取会员(供应商)的会员费，一般情况下，企业设立的客户服务部也主要是针对供应商的服务。然而，越来越多的电商企业已经注意到为买家服务的重要性，丰富的买家资源是平台吸引供应商的关键竞争力之一，随着电子商务的进一步发展，对买家进行收费服务，将买家纳入收费会员的行列将成为企业新的利润点。

买家服务的目的在于帮助买家快速找到合适的供应商(买卖匹配)，降低采购风险，节约采购成本，提高采购体验。买家服务部门的工作职责包括买卖匹配、买家客户的日常维护、处理买家咨询和投诉以及了解买方市场，为公司扩展买家业务提供支持等。

BI 对于买家服务部的OLAP 分析可以从以下几个方面进行：

1) 买家客服工作绩效统计以及过程监控

买家客服的工作绩效统计主要是帮助主管代替人工统计方式，自动化统计每个买家服务人员的日常工作量，比如每月维护买家数、月匹配买家数、月处理买家邮件数等。过程监控是为每一个买家客服提供每日的工作量与工作任务的完成情况，帮助买家客服实时了解自身的工作情况，及时调整工作计划，完成工作任务。

2) 业务数据多维分析，提升工作效率与品质

买家客服日常工作中涉及的业务数据有买家信息、买家咨询与投诉、买家采购需求等。买家信息包括买家所在的地区、行业，以及买家的平台登录情况等，对买家进行地区与行业的分布统计，可以帮助了解买家市场的分布情况，帮助企业制定针对性的宣传与营销策略；买家的平台登录情况可以用来衡量买家的活跃度，买家活跃度的分析可以帮助企业实时监控买家动态，了解平台的运营情况。

如果企业建立了完善的买家咨询与投诉机制，就可以方便 BI 取出这种数据，比如投诉和咨询的内容进行了有效分类，那么对各种类型的投诉或咨询进行统计，就可以知道买家在哪些方面遇到困难，对哪些方面比较感兴趣，从而帮助买家客服调整工作策略，提升工作品质。

买家的采购需求是买家意愿的直接反映，买家采购需求表达了买家对于行业、产品、供应商资质(公司规模、业务类型等)、产地等市场因子的偏好，通过对以上维度进行分析，可以对买家市场有更加深入和细致的了解。

4.4 分析型工作流程

4.4.1 分析型工作流程是什么？

分析型工作流程简单点理解就是在工作流程中加入BI的分析功能，实现分BI数据支持和BI应用集成两种方式，前者由数据仓库、数据挖掘等提供分析结果和数据给业务系统，后者则直接由BI系统提供强大的报表和OLAP分析应用给业务系统进行集成，现在主流的BI工具基本都提供JAVA/.NET等集成接口，集成工作主要包括用户集成、权限集成、参数传递。

4.4.2 分析型流程的应用价值

通过ETL让BI系统可以源源不断获得业务系统的信息数据，进而加工成分析结果和挖掘知识，初期这些知识主要通过报表展现来服务于分析人员，这种方式比较适合中高层领导，其分析往往具有较高独立性，不需要其他人员和其他部门的参与即可完成，但对于基层业务人员来说，在工作过程中也非常需要一些智能决策来提升其执行力度。

- 个人绩效：可以让操作人员实时掌握自己工作业绩。
- 智能预警：可以让操作人员纠正不当操作。
- 工作流程点的分析支持可以让业务操作有据可循，执行更精准。
- 数据挖掘知识在业务流程中的智能应用可以节省大量的人工操作，提升工作效率，比如焦点科技利用数据挖掘技术可以自动化审核90%的用户，从而大大提升了用户信息审核速度。

所以我们非常有必要把BI系统的知识和分析结果应用到业务系统，也就是分析型工作流程的实现，技术实现包括数据反写和应用集成。如图4.3所示，一旦实现了分析型工作流程那么业务系统和BI系统将形成一个闭环，在企业信息化战略方面将有如下两大好处：

- 分析功能集成到业务执行过程中，将极大提升业务系统的分析能力。
- 对于执行一数据一分析一决策一执行这样完整的决策闭环，都有BI分析的支持，为整个流程的优化提供了基础。



图4.3 构建业务系统和BI系统的信息闭环



4.4.3 分析型流程在电子商务的应用

电子商务主要借助网络平台进行运营，运营环节中产生了大量的数据，而信息整理也是日常运营工作最重要的组成部分，在业务运营中也涉及大量的信息整理工作，要提高运营的效率与质量，把 BI 融入业务系统形成分析型工作流就显得尤为重要。在 B2B、B2C、C2C 领域中：企业战略决策、平台运营、销售管理、客服维系、信息管理、财务管理和人事管理体系中都需要基于信息做出准确及时的决策，这些都为分析型流程定制开拓了广阔的市场，目前分析型流程在电子商务领域应用广泛，这里介绍几种主要的应用类型。

1) 战略决策分析型流程

应用 BI 技术，可以把电子商务市场信息利用数据抓取技术进行收集整理，针对战略和决策事件主题把市场信息整合成分析型系统，实时提供给战略决策部门感兴趣的信息量，使得战略规划、市场动态分析、决策分析、实施执行一体化。切实提高战略决策部门的市场敏锐性，及时做出正确的决策。

2) 销售分析型流程

为提升公司销售业绩，可以利用 BI 技术对潜在销售用户进行挖掘，利用挖掘方法找出潜在销售对象，根据会员行为特征进行产品和服务匹配，实行精确化营销策略，结合销售业务流程，可以把该分析环节做成分析系统嵌入销售业务之中，这样可以给销售人员提供智能化的信息量，让销售人员在短时间掌握销售对象信息，做出销售策略，提升销售的准确性。通过销售分析型流程的嵌入可以把销售业务监控、潜在用户分析、销售策略、执行实施一体化，提升销售业务的效率与质量。

69

3) 客服分析型流程

客服的主要工作是做好会员维系，增强会员的体验效果，减少会员流失。客服人员要做好维系工作，就需要对维系对象有个全面的了解和认识，借助 BI 技术，可以对维系对象量身定制一份体检表，全方位分析会员的特征与存在的缺陷，帮助客服人员及时发现问题，做出准确的维系策略，增强会员的体验效果，降低会员的流失率。因此客服维系体系可以借助 BI 技术，对会员进行体检分析，把分析结果融入到客服运营系统，使得用户管理、分析一体化。在分析性系统的帮助下，客服人员就能全面把握会员特征，在短时间内做出维系策略，解决会员存在的潜在问题，提升会员的体验效果。使得客服的业务监控、会员体检分析、维系策略、执行实施流程一体化，从而提升客服维系工作的效率与质量。

4) 平台监控分析型流程

电子商务平台每天都有大量的会员进行登录访问，难免会出现一些异常情况，为保障电子商务平台的正常运营，就需要有个智能监控系统，时时反映平台的运用

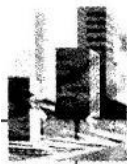
效果及特征，在出现异常情况时能及时做出反应，快速定位症结，达到快速解决问题的目的。平台运营涉及的面较广，异常种类和异常现象众多，如何才能满足监控需要，这就需要借助 BI 技术，针对各种监控要求定制分析型流程，一方面可以提升平台监控的智能水平，在出现异常状况时能及时分析出异常原因，帮助监控人员及时解决问题，另一方面 BI 分析形式灵活，可以根据监控要求进行变动，能较好地满足监控需求。在分析型流程定制的帮助下，监控工作也就变得流程化、体系化，便于工作管理与维护，真正提高操作人员的工作效率与质量。

4.5 小结

BI 对于高层的支持主要体现在高层仪表盘和企业级绩效分析，目的是帮助决策层了解企业各个环节的整体的、宏观的情况，提供管理和决策上的帮助。高层仪表盘的指标设计必须把握“多就是少”的原则，直观地呈现给客户最关心的数据。

BI 部门级 OLAP 分析能够更好地为其他业务部门的工作提供助力，有效地发现问题，改进问题，并对效果进行监督，更合理有效地安排各员工与整体部门的工作计划，提升员工工作效率，有效地激励员工工作积极性，用数据提供有理有据的部门级的监控、分析、管理与决策。

分析型工作流程定制能使工作人员的工作变得流程化、系统化，能大幅度提升操作人员的工作效率与工作质量，电子商务领域需要依托信息进行快速的响应，这些都给分析型流程定制提供了广阔的应用平台，分析型流程定制能极大地满足电子商务的运营特征，推动运营工作向体系化、标准化方向发展，能大大提升公司运营管理效率，极大的增强公司的核心竞争力。



第五章 BI 对外支持网站用户

电子商务企业的数据库中储存了大量的数据，从这些数据中能够提取出支持企业自身运营决策分析的知识。对于电子商务网站的用户来说，这些数据涵盖了用户自身、用户的客户以及其他用户的信息，基于这些数据的分析，能够让用户了解自己、了解自己的客户和了解同一平台上的其他用户(竞争对手)。焦点科技公司推出的数据罗盘，就是依托于中国制造网上的海量数据结合BI 数据分析和数据挖掘技术，打造出来的用以支持网站会员的在线数据分析平台。BI 对于网站用户分析的支持，是网站提升客户满意度和网站价值的重要途径。下面从四个角度讨论 BI对于网站用户分析的支持。

5.1 网站推广效果及优化

71

5.1.1 用户网站推广目标

如图5.1所示，用户在电子商务网站进行推广，是想通过网络平台提升产品的曝光率，让更多的买家通过访问网站关注到自己的产品，并且对自己的产品产生兴趣，从而带来询盘并最终达成交易。

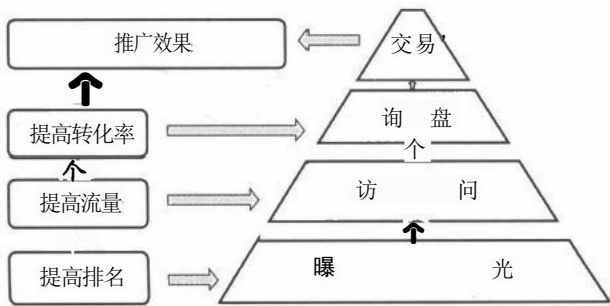


图5.1 用户网站推广过程

因此，企业在网站上的推广目标包括下列三方面：

提高排名：该目标是为了获得更多的曝光量和展现机会，以得到更多的流量。

提高流量：该目标是提高访问量，获得更多关注。

提高转化率：该目标是提高排名和流量的最终目的，也就是在访问量的基础上提高询盘量，达成交易的意向。

5.1.2 推广效果影响因素

客户在网站上的推广效果主要有以下影响因素(如图5.2)：

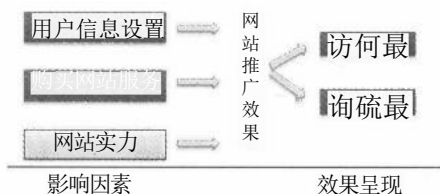


图5.2 网站推广效果影响因素

(1) 用户信息设置，用户信息设置的完整性、正确性和丰富性是吸引客户关注的根本。

(2) 购买的网站服务，客户通过购买网站服务可以获得相对更好、更多的网站推广机会，如购买广告、关键词等服务。

(3) 网站自身实力，电子商务网站的实力是决定用户推广效果的关键，是用户选择推广平台的主要考虑因素。

前面两个因素是可以通过用户的主观努力进行优化和改善的。那么，如何优化推广效果，应该从哪些方面进行优化？

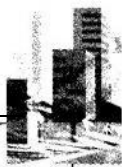
首先要了解推广效果，然后进行进一步的分析。

用户可以通过从网站上获得的询盘情况了解到自己的推广效果，但是却不能进行对比、分析和深入了解推广情况，以及可以从哪些方面进行推广优化。BI 可以通过对用户的网页访问、搜索等情况进行综合分析，全面清晰地体现用户的网站推广状况和问题，并且帮助用户调整推广策略和改进推广手段。用户网站推广效果的提升，不仅可以减轻网站销售和客服的压力，而且增强了客户对于网站的黏性。

5.1.3 BI 推广效果分析和优化支持

网站推广效果如何，是否达到预期目标，用户可能会期待知道这几个问题：

- (1) 哪些用户对我比较关注？
- (2) 用户是怎样访问到我的？
- (3) 我的访问情况如何？
- (4) 我的询盘情况如何？
- (5) 我该从哪些方面优化推广效果？



围绕这几个问题，BI 可以从下列几个方面进行解答：

1) 流量来源分析

访问来源分析即分析买家是通过什么渠道访问到我的。

用户的访问途径包括：搜索引擎访问、站内搜索访问、行业目录访问等，如图5.3所示。通过分析访问日志数据，可以使用户更详细地掌握网站流量主要是来自于哪些页面、哪个搜索引擎等。若用户在网站上投入一定广告，当推广一段时间后，再查看一下由广告页带来的流量所占的比重，就可以知道自己的推广效果如何。

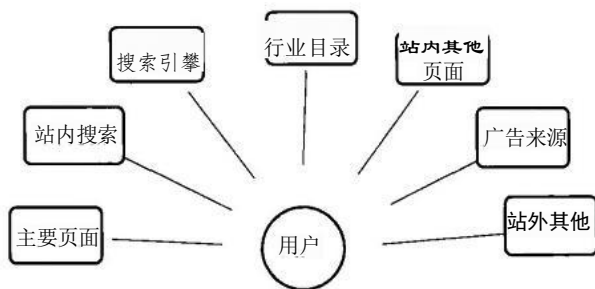


图 5.3 流量来源

流量来源分析可以以时间、流量来源为维度，访问量、访问量占比为指标进行分析，让用户了解自己的访问来源情况。

73

2) 访问情况分析

访问情况是网站推广效果的直观反映，是指用户在网站上获得的所有访问信息。

访问情况分析包括下列几个方面：

(1) 总访问量分析：分析用户在网站上得到的总流量。以时间为维度(如月份)，访问量和访问人数为指标呈现客户在网站上获得的访问量趋势，帮助用户横向了解推广效果，并且可以比较历史和当前的效果。

(2) 访问时段分析：分析流量的访问时间分布。以时间段为维度(如时)，访问量和访问人数以及各自的占比为指标，让用户了解买家的访问行为特征。

(3) 搜索关键词分析：分析关键词的流量分布状况。通过站内搜索进行访问是网站用户的主要访问方式之一，用户产品关键词的设置是决定能否被搜索中的关键。对搜索关键词带来的访问进行深入分析，以时间、搜索关键词为维度，访问量、访问量占比为指标进行分析，可以让用户了解关键词带来的流量情况，为用户优化关键词提供依据。

(4) 访问客户分析：分析买家的地区分布。以时间(如月份)、地区为维度，访问量占比为指标，可以让用户同时从横向和纵向上对比买家地区分布，了解不同地

区买家的需求偏好，了解自己的产品在哪些地区较受欢迎以及在不同地区的推广效果，从而可以有方向性地改变推广策略。

3) 询盘转化分析

询盘转化分析主要分析每一个产品的访问转化为询盘的情况，以产品询盘量与该产品访问量的比值衡量产品的询盘转化率。但是对于转化率的分析，还需要结合访问量进行，因此转化率分析指标设为访问量和询盘转化率。

如果某个产品的访问量高，但是询盘转化率低，则说明大量买家访问到了产品，却对产品没有很大的兴趣，可能是因为产品的介绍信息设置得不够吸引买家，用户应该丰富产品信息，并且及时更新和维护。

如果某个产品的访问量低，询盘转化率也低，则说明产品推广力度不够，或者该产品缺乏市场。用户可以考虑加强推广，购买网站推广服务(比如广告、搜索排名等)，或者将资源重组分配给其他比较热门的产品，提高热门产品的推广力度。

如果某个产品的访问量高，询盘转化率也高，则说明目前该产品的推广效果良好。

如果某个产品的访问量低，询盘转化率高，则说明该产品受到市场的欢迎，但是依然需要加强推广力度和曝光率，使其增加访问量并且获得更多的询盘。

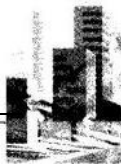
4) 关键词分析

关键词分析通过分析买家在整个网站上进行搜索所用的关键词，让用户了解买家的搜索习惯，为用户优化关键词设置提供数据支持。关键词的选择，核心还是通过流行度和竞争度，即判断哪些关键词是国外买家经常使用的热词，并且使用得相对较多，为产品判断最有效的关键词，最终目的是提升较热关键词的搜索排名。

因此关键词分析以关键词为维度，关键词流行度和竞争度为指标。其中，关键词流行度用买家搜索量衡量；竞争度用关键词使用量(即目前有多少用户的产品设置了该关键词)进行度量。

5.2 市场分析

一个完整的市场由各类专业市场组成，如商品服务市场、金融市场、技术市场、旅游市场、房地产市场等。在此我们提出的市场分析，从一定程度上是把市场限定在行业的角度去分析。市场受经济、政治、进出口、自然等多重因素的影响而灵活多变，面临市场不景气或竞争激烈等环境，买卖物资需求和供给会出现不平衡，长此以往双方都将蒙受经济损失，因此，对市场行情趋势的超前获知和应对措施的选择非常重要，供应商的销售策略更是不容小觑，不同的解决方式将会产生截然不同的效果。



面对市场变化给买卖双方带来的供需不平衡等一系列问题,需要及时地对未来市场的发展做出预测及判断,并给供应商提供好的建议及解决策略,这就是商业智能(BI)要展开的两个大的方面的探索和研究。如图5.4所示,考虑从宏观市场和电子商务平台的具体行为两方面来分析市场趋势的变化,通过经济、政治及贸易进出口数据特征来预测未来市场的发展,通过平台买家的操作行为来总结市场行业的热销产品、潜力产品、热搜索词,为供应商的营销策略提供根据。

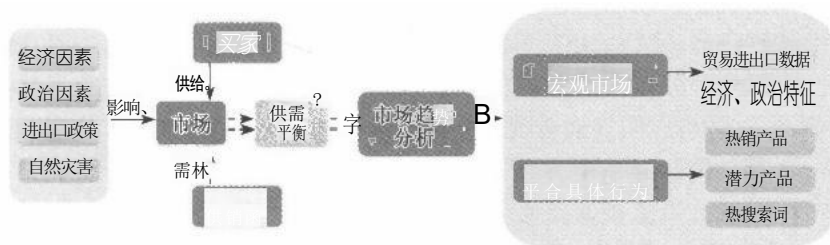


图5.4 市场分析的整体思路

5.2.1 平台行为分析

对于供应商和买家来说,都希望能够洞悉市场,从而把握时机形成交易。电子商务网站是帮助买卖双方实现交易的平台,基于买卖双方在平台上的行为数据进行分析,可以了解买卖的趋势即市场的趋势。

电子商务平台中的供应商在得知市场趋势的大致变化后,可以了解自身所处大行业和子行业的未来发展趋势,必要时可通过查询热门子行业来适当调整自身行业。而子行业下面还有各种子产品,此时子产品及其属性的设置可以参考子行业下的热销产品、潜力产品、热搜索词,这些都是通过分析平台买家的操作行为得来,具有一定的客观性。

1) 热销产品分析

行业热销产品指某行业下哪些产品最能够得到客户的认可与购买,产品的热销度用一段时间内产品的访问量来进行衡量,访问量越高越热销。

通过行业热销产品分析,供应商可以知道当前买家更加青睐于哪些产品,并且可以将自己所属行业下的产品与同级别同行业中的热销产品相比较,如图5.5例子所示,从而分析自己的产品与热销产品的差距,所谓知己知彼才会提高;或者及时调整自己的产品,投买家所好,争取获得更多的买家关注度。

2) 潜力产品分析

行业潜力产品指某行业下哪些产品在未来最有可能得到客户的认可与购买,产品潜力指数用一段时间内产品的访问量增长率衡量,增长率越高越具有潜力。

根据潜力产品分析, 供应商可以提前获知哪些产品即将热销, 供应商可以据此调整未来的销售策略和推广策略, 加大对于潜力产品的销售和宣传力度, 争取在竞争对手之前吸引买家的注意。买家也可以根据潜力产品分析, 把握市场先机, 及时出手, 降低采购成本。

3) 热搜关键词分析

热搜关键词指在网站上被用户搜索频率高、非常流行的一类词, 各行各业都有热搜关键词, 如使用百度指数输入几个相关的关键词就可以查看这些词的用户和媒体的关注度趋势, 还会给出上升最快的一些相关搜索词, 同时还可查看这些词在全国的分布及人群分布。图5. 5为照明市场某公司产品搜索词与该市场热搜词比对的例子, 这样可以明显地看到本公司搜索词的热度, 又可以借鉴甚至适当地调整购买一些效果较好的关键词, 为自己争取更多的盈利机会。

热销产品	热销度	潜力产品	潜力值	热搜词	搜索热度
Light1		Light1		LED	
Light2		Light2		Light	
Light3		Light3		LED Light	
本公司产品		本公司产品		本公司产品	
Light4		Light4		Lighting	
Light5		Light5		Lights	

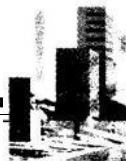
图5. 5 照明市场本公司产品与市场热销、潜力产品、热搜词的对

通过这一系列的功能描述, 可见热搜关键词对各行各业都很重要, 会较大幅度地影响客户的判断和行为。所以好多商家会不惜血本定期购买类似于百度竞价排名这类的产品, 为的是保证商家企业的排名尽量靠前, 保证公司及产品有一定的曝光率, 尽量为自己创造盈利的机会。但要知道, 不是人人都买得起、承担得起这类的耗资, 毕竟这不是一投入就一定能保证有成效的, 回报率仍然是个未知数。

5. 2. 2 宏观市场分析及预测

买卖双方的行为固然是影响微观市场的主要因素。但是, 宏观因素比如经济政策、政治因素、进出口政策等却是影响整个宏观市场走势的关键, 有时甚至能起到确定性的作用。比如说2012年5月7日央行突然决定降息, 这是从2008年12月23日以来首次降息, 同时央行首次调整存贷款利率浮动区间, 其对于市场的意义不可低估, 它意味着国内货币政策将出现新转向, 这种减息方式不仅在一定程度上保护存贷款人的利益, 也为银行利率市场化创造条件, 利率下调对宏观是整体的利好, 有利于疲弱的经济复苏, 这说明国内经济将会逐渐好转, 行业发展会稳步向前, 贸易进出口量会有更大的提升空间。

但是, 经济、政治等因素不能够利用具体的指标进行度量和预测, 因此BI在进



行市场分析的时候需要结合当前和未来的国家宏观环境进行分析，甚至需要人为调整分析和预测的结果。

5.3 行业及竞争对手分析

在各个行业网站泛滥的互联网领域里，如何才能真正把握住行业的命脉，怎样才能在整个行业竞争群体中逐步向前发展。下面对行业经济特性和吸引力的分析，将帮助我们深入地了解整个行业，而竞争对手的分析将帮助我们及时把握竞争对手的动向，掌握优劣势，逐步向前发展。

5.3.1 行业经济特性分析

一个行业是由一组企业构成的，它们生产的产品非常相似。我们可以从以下的角度来描述行业的经济特性：

- 市场规模和增长速度
- 市场竞争的地理区域
- 竞争者数量，相对规模，垂直一体化程度
- 顾客数量和相对规模
- 企业进入和退出行业的难易程度
- 核心技术变化速度
- 规模经济和经验曲线效应的程度
- 商品是标准化还是差异化产品
- 整个行业的盈利水平

77

我们可以选取合适的指标来衡量或近似描述上述经济特性，然后应用聚类模型可以将各个行业进行分群，根据相应指标可以描述各个群体的特征。如果某个群体表现为竞争企业众多，产品同质化，企业进入或退出的成本低，那么可以认为该群体的行业为完全竞争的。而随着竞争企业数量的减少，产品的异质化，进入或退出的成本变高时，竞争就由完全竞争逐渐转向垄断。下面我们将具体分析某一行业的竞争环境。

5.3.2 行业竞争环境分析

行业竞争环境分析主要是从多方面对行业的机会与威胁做一系列的分析，认为机会大于威胁则认为行业具有吸引力。这里可以运用波特的五种竞争力量模型，将大量不同的因素汇集在一个简便的模型中，依次分析行业的基本竞争态势。“五力模型”[11]确定了竞争的五种主要来源：供应商的讨价还价能力、购买者的讨

价还价能力、新进入者的进入障碍、替代品的威胁以及来自目前在同一行业的公司间的竞争。图5.6是“五力模型”分析的主要因素：

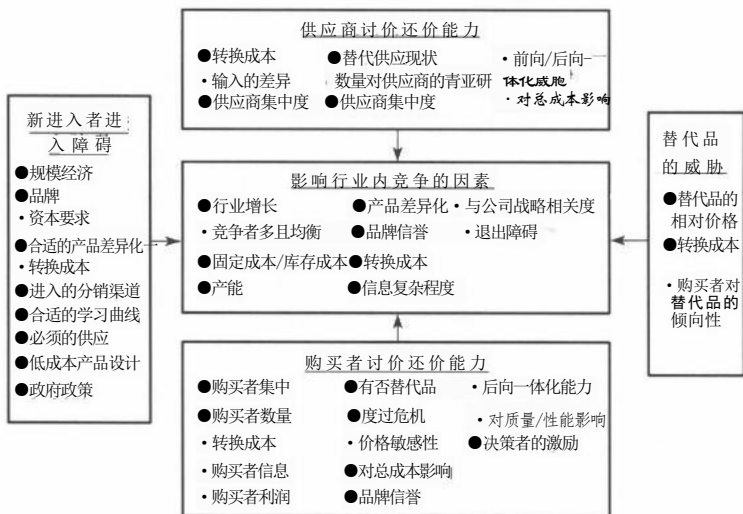


图5.6 “五力模型”分析

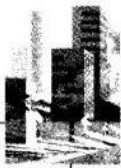
由于行业间的差异性，应用该模型首先要识别适用于行业与公司的五种力量的因素；其次根据相关指标对各个力量的因素进行评分，既要当前的现状进行评分，也要对未来的可能变化作出相应的预测；然后根据恰当的权重将各个因素加以综合，得到五种力量当前和未来的评分值；最后结合评分列出该行业的机会以及所面临的威胁。一般的，若企业间竞争大、进入障碍低、有替代产品威胁、供应商和购买者有相当大的讨价还价能力，那么可以认为该行业缺少吸引力，意味着五种力量的组合会降低行业整体利润水平，该行业中的企业利润率会很低，甚至趋近于零。

对行业经济特性及竞争环境的分析，将帮助我们更好地了解某个行业当前的发展状况，并且能够预测未来的发展趋势，这样对于企业来说可以选择一个更适合发展的行业，这样至少在方向上就具有一定的优势，下面将简单介绍一下行业竞争对手的分析。

5.3.3 竞争对手分析

竞争对手分析主要目的在于通过了解竞争对手的信息，获知竞争对手的发展策略以及行动，取长补短，从而有效地制定自己的战略方向与措施。在进行竞争对手分析时，既要考虑到现有的竞争者，尤其是那些与自己同速增长或比自己增长快的竞争对手，注意发现任何竞争优势的来源，也要考虑潜在的竞争者。

既然要分析，那么对竞争对手的信息进行收集是非常重要的基础工作，对于电



子商务来说，主要的就是网站上各种信息的搜集。以 B2B 行业为例，我们可以利用数据抓取工具定期抓取竞争对手网站上的会员及产品信息，这样就可以定期监控竞争对手的整体规模变化趋势、会员的新增与流失、会员行业分布、区域分布等，结合自身情况，分析与竞争对手各自用户群体特征及变化趋势，分析与竞争对手间的相对优势与劣势，充分利用自己的行业资源及行业客户，保持并扩大竞争优势，弥补竞争劣势。

5.4 增值服务效果及推荐

电子商务竞争日趋激烈，要想做好电子商务平台，必须重视增值服务与推荐，满足客户变化多样的需求，这样才能吸引客户，提升其满意度和忠诚度，也能够创造价值。

5.4.1 增值服务

增值服务核心内容是根据客户需要，在基本服务基础上提供的定制化服务，以提高用户满意度、满足用户个性化服务要求，并且能够从中赚取一定比例的收益。

电子商务已经走进我们的生活，也有越来越多的人开始通过网络进行贸易销售。互联网已经不仅仅是资讯的媒体，更成为交易的渠道。那么，在众多电子商务平台中，中国制造网、阿里巴巴、淘宝等为什么可以成功？因为他们成就了网上价值链，为买家和卖家提供了必要的辅助服务，例如：信用、支付、物流、品牌等。这些中间件由于成本太高是单一企业无法完成或很难完成的。而中国制造网、阿里巴巴、淘宝等通过聚合交易的平台和统一推广的方式，完成了这个网络贸易集市所需要的平台、服务和推广。对于企业而言，这就是增值服务，实实在在的可以带来买家的增值服务，因此他们愿意为这些增值服务付费。

以淘宝为例，其增值服务非常之多，比如店铺管理增值服务，消费者保障增值服务，卖家推广服务。除此以外，淘宝还有自身与一些第三方公司利用数据分析挖掘等相关商业智能知识，将网站上的各种访问交易等信息全面整合，涉及流量分析、成交分析、客服效率分析、“宝贝”分析、营销效果分析等，这些通过数据转化成的知识对于卖家是尤为重要，能够帮助卖家时刻了解店铺的经营状况，能够实实在在地帮助卖家提升经营效益和效率(图5.7)。

B2B、B2C 等其他网站也是类似的，那些能够提供更为全面的服务且运营成本可控的电子商务平台，对于用户而言更有吸引力。因此，为用户提供更为专业、深入、完善的服务是电子商务平台所应追求的终极目标之一。

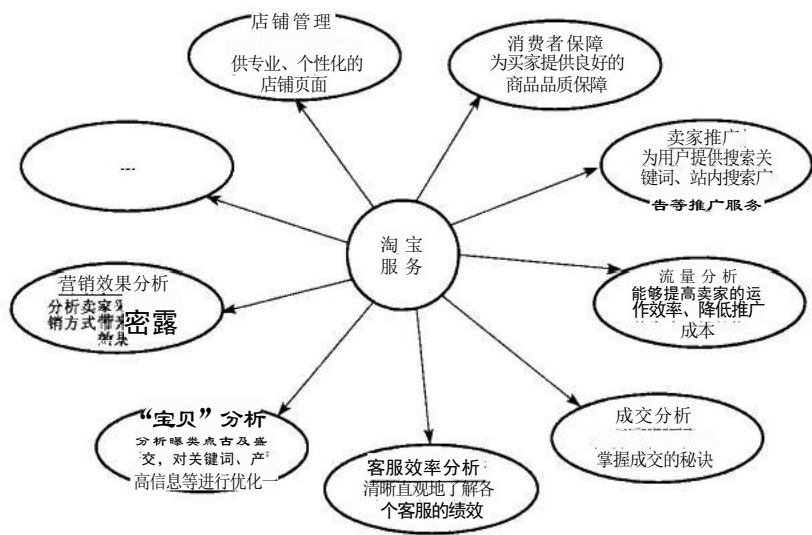


图 5.7 淘宝服务

5.4.2 电子商务推荐系统

电子商务推荐系统就是通过了解和学习客户的需求与爱好，为用户提供商品信息和建议，推荐其可能感兴趣的商品，模拟销售人员帮助客户完成购买过程。

根据用户获得推荐的自动化程度和持久性程度，可以将电子商务推荐系统分为非个性化电子商务推荐系统、基于属性的电子商务推荐系统、商品相关性推荐系统和用户相关性推荐系统。

1) 非个性化电子商务推荐系统

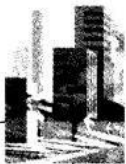
向当前用户提供的推荐结果可能基于其他用户对商品的平均评价，或者基于电子商务系统的销售排行，或者基于电子商务系统的编辑推荐。这种推荐技术属于自动化方式推荐，独立于各个用户，每个用户得到的推荐都是相同的。如亚马逊的销售排行、根据顾客平均评分推荐。

2) 基于属性的电子商务推荐系统

根据商品的属性特征向用户产生推荐列表，这种推荐系统类似于搜索引擎，用户需要手工输入所需商品的属性特征。由于这种推荐方式需要用户显式输入商品的属性特征，因此属于手工方式推荐。如亚马逊提供的搜索(筛选)推荐。

3) 商品相关性推荐系统

商品相关性推荐系统主要应用关联规则或序列规则，依据网站原有的记录为用户提供推荐，这些推荐不但可以满足用户的个性化偏好，还可以在在一定程度上预测用户的购买行为。如亚马逊的“经常一起购买的商品”、“购买此商品的顾客也同



时购买”。

4) 用户相关性推荐系统

用户相关性推荐系统又称为协同过滤推荐系统，这种推荐系统一般采用最近邻技术，利用用户的历史喜好信息计算用户之间的距离，然后利用目标用户的最近邻居对商品评价的加权评价来预测目标用户对特定商品的喜好程度，系统从而根据这一喜好程度来对目标用户进行推荐。如亚马逊的“购买了您最近浏览过的商品的顾客同时购买了”。

在实际应用中，我们可以选择其中的一种推荐方式，或者将多种推荐结合起来。推荐系统的关键思想是根据用户在访问时的交互性，在用户与站点之间增加的一个推荐过程，商品相关推荐主要工作在于对用户的当前访问序列进行预处理，将访问序列提交到推荐引擎中，根据商品相关库中的知识推荐用户感兴趣的商品，其他形式推荐的交互过程相对较为简单，整个推荐系统主要由离线处理部分和在线推荐部分组成，模型架构如图5.8所示。

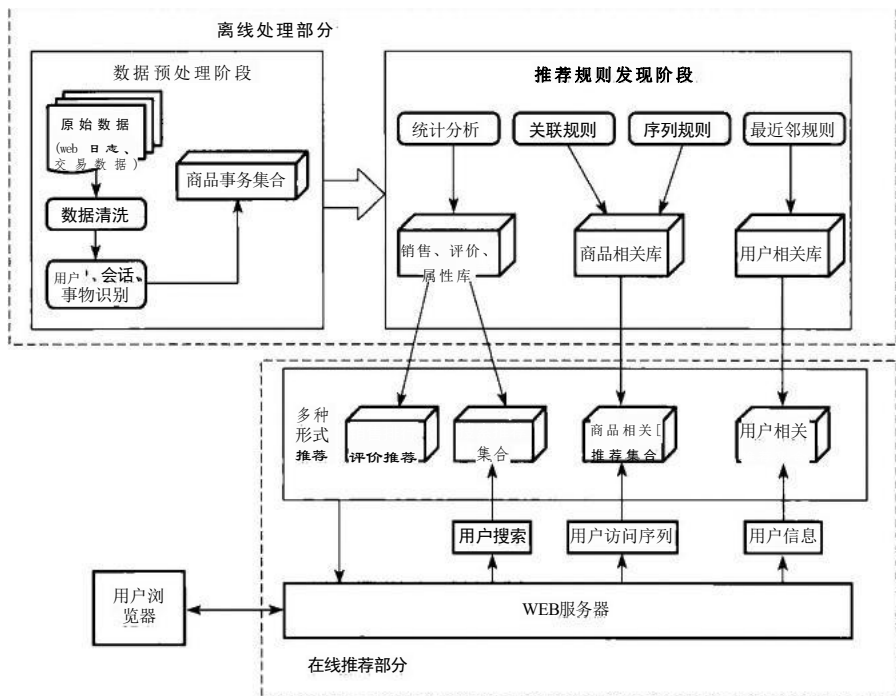


图5.8 电子商务推荐系统模型架构

和增值服务一样，推荐系统在电子商务中也发挥了其重要的作用，它能够将网站的浏览者转变为购买者，能够提高网站的交叉销售能力，提高客户对网站的忠诚



度。亚马逊的推荐系统做得比较成功,据说在美国,84%的网购消费者登录亚马逊时并不确定自己想买什么,而亚马逊结合了各种形式的推荐系统,向用户推荐其可能感兴趣的物品,从而成功地为消费者“创造”并满足其需求。

5.5 小结

BI 对于网站用户的支持主要体现在下列四个方面。

网站推广效果及优化: BI 通过对用户的网页访问、搜索和询盘等情况进行综合分析(比如分析买家的访问来源、常用搜索关键词等),全面清晰地体现用户的网站推广状况和问题,并且帮助用户调整推广策略和改进推广手段。用户网站推广效果的提升,不仅可以减轻网站销售和客服的压力,而且增强了客户对于网站的黏性。

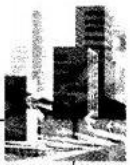
市场分析: 市场分析包括对买家的意向分析和对宏观市场的分析和预测。买家意向分析通过分析买家的行为动态,比如点击或询盘较多的热门产品、点击或询盘提升较大的潜力产品和买家常用的热搜关键词等,从而帮助客户调整自己的销售策略,取得市场先机。而宏观因素比如经济政策、政治因素、进出口政策等却是影响整个宏观市场走势的关键,有时甚至能起到决定性的作用,因此对于宏观市场的分析也很重要。但是,宏观因素不便于利用具体的指标进行度量和预测,或者数据不易于获取,因此,BI 在进行市场分析的时候需要结合当前和未来的国家宏观环境进行分析,需要人为调整分析和预测结果。

行业及竞争对手分析:

(1) 行业经济特性分析是对市场中的所有行业的分析,根据各行业的经济特性(如市场规模和增长速度等),利用聚类模型将各行业分群,根据各个群体特征可以分析出该群体中的行业经济特性,如果某个群体表现为竞争企业众多,产品同质化,企业进入或退出的成本低,那么可以认为该群体的行业为完全竞争的。

(2) 行业竞争环境分析是对具体某个行业的分析,是从多方面对行业的机会与威胁做一系列的分析,如果机会大于威胁则认为行业具有吸引力;可以运用波特的五种竞争力量模型分析行业的基本竞争态势。

(3) 竞争对手分析主要目的在于通过了解竞争对手的信息,获知竞争对手的发展策略以及行动,取长补短,从而有效地制定自己的战略方向与措施。以 B2B 行业为例,我们可以利用数据抓取工具定期抓取竞争对手网站上的会员及产品信息,这样就可以定期监控竞争对手的整体规模变化趋势、会员的新增与流失、会员行业分布、区域分布等,结合自身情况,分析与竞争对手各自用户群体特征及变化趋势,分析与竞争对手间的相对优势与劣势,充分利用自己的行业资源及行业客



户，保持并扩大竞争优势，弥补竞争劣势。

增值服务效果及推荐：增值服务核心内容是根据客户需要，在基本服务基础上提供的定制化服务，以提高用户满意度、满足用户个性化服务要求，并且能够从中赚取一定比例的收益。电子商务推荐系统就是通过了解和学习客户的需求与爱好，为用户提供商品信息和建议，推荐其可能感兴趣的商品，模拟销售人员帮助客户完成购买过程。根据用户获得推荐的自动化程度和持久性程度，可以将电子商务推荐系统分为非个性化电子商务推荐系统、基于属性的电子商务推荐系统、商品相关性推荐系统和用户相关性推荐系统。

第六章 B2B 电子商务 BI 成功实施的关键点

BI项目成功的关键点在于技术、业务、项目管理等问题，而电子商务行业还必须面临两个问题：大数据和对外用户分析支持。这里笔者列了九个最为重要的关键点，对每个关键点力求给出解决方案。

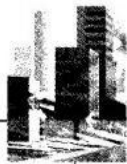
6.1 是否有强大的 BI 技术架构？

既能满足自身当前需求，同时又能长远立足的强大技术架构才是我们需要的。当前 BI 技术已经较为成熟，技术架构也已经较为成熟，一个电子商务公司的企业级 BI 技术架构最起码要包括标准数据源接口、ODS、数据仓库、数据集市、业务系统接口这样五层，尽管这五层会有很多细节，并且不同人设计都会有些差异，但是每一层在目标上是基本一致的，这里我们从目标和定位来理解现有的技术架构。

(1)标准数据源接口：解决系统取数据的问题，可以是直接数据接口，也可以是文本、HTML 等格式，不管哪种方式都必须满足BI 取数据的基本要求，特别是大数据表能够支持增量抽取数据，必须要有字段能够识别新老数据；另外数据接口要求稳定，数据库接口要求顺畅，文本等数据要求能够及时抵达。

(2)ODS：这个层是辅助作用，首先是减轻对数据源抽取数据的压力，一般设计采用和源完全一样的数据结构，实现表与表之间的一一映射；其次是可以为转换做些预处理、数据清洗等工作，这样便于ETL 进数据仓库的时候简单高效，不用为繁杂的算法及异常处理考虑太多；最后，ODS 还是将来实时报表与分析的直接数据入口，因为ODS 的设计简单这样既可以对数据源进行无压力实时抽取，又能让分析压力转移到 ODS 上来，因为数据从 ODS 转换加载到数据仓库往往时间较长，所以基于ODS 实时分析较为可行。

(3)数据仓库：主题分割、合理粒度和维度建模是数据仓库设计面临的主要问题；一般建议数据仓库需要按照主题适当分割进行设计和实施，特别是在表设计上要保持独立业务的表结构独立性，事实表尽量单一，尽量不要有跨主题业务数据间联合表出现，否则很容易在数据仓库层面陷入设计混乱，造成数据不一致；粒度选择一般还是尽可能的明细，尽可能的存放更长时间的数据，这样理论上可以满足将



来尽可能多的需求，但事实上电子商务行业数据非常巨大，所以粒度选择是必然的问题，存放时间也不得不做出选择；最新的分布式技术有望采用低成本的 PC 机实现大数据在数据仓库的存储，这方面技术有望在三年内进入成熟的企业级应用，现在 Greenplum 的云计算在数据仓库的成功就是该技术方向的一个标杆案例；笔者比较推崇 Ralph Kimball 的维度建模思想，其中星形结构模型、维度代理关键字、扁平维度、维度字段冗余等理念具有非常高的实用价值。

(4) 数据集市：数据集市的设计就是为了解决具体项目的数据存放问题，最大挑战还是系统的性能，为了保证性能有三个事情可以做：第一是去掉不相关数据，不要的字段、不要的记录就过滤掉；第二是把仓库明细数据汇总到合适粒度；最后就是根据分析需要建立起跨主题业务数据的联合表，解决跨主题事实大数据表之间的关联问题。

(5) 业务系统：BI 除了自身能提供分析报表之外，很多时候还要对业务系统提供有分析价值的数据，相当于来自业务系统的数据经过BI 系统分析之后反馈结果给业务系统，这样 BI 就反过来支持业务系统操作人员的分析需求。

这五层在BI 技术架构中都有明确的价值定位，只要确实理解了每个层的价值定位，即使在开发过程中方法会有区别，但最终走向还是殊途同归。

85

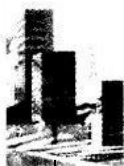
6.2 是否有大数据的存储设计能力？

大数据是电子商务企业不得不面对的问题，大数据带来最大挑战就是 ETL 性能差，查询慢，解决办法还就是分布式部署及数据模型合理设计。

分布式部署及云计算是解决大数据的有效方法，传统数据库比如 ORACLE、DB2 在大数据查询的时候 I/O 存在巨大瓶颈，简单增加CUP、内存等方式不仅成本较大，而且效果有限，而分布式或者云计算则可以让数据的I/O 及运算压力分摊到不同机器上面，从而实现性能线性提升。

分布式开发当前还刚刚起步，开发要求较高，在这个阶段建议的是对数据量巨大的表(比如访问日志、搜索日志等大表)进行分布式部署，而小数据量的用户信息、产品信息等则可以继续存放在传统数据库里面；集市数据因为过滤、汇总等操作从而数据量得到了大大的削减，一般还是建议存放在传统数据仓库里面，这样便于前端应用的开发，当然像 Hadoop、Greenplum 等分布式技术也支持前端应用的开发，只不过对大部分人来讲会陌生些。

不管出现什么技术，合理的设计永远是需要，其中数据模型设计尤为关键。ODS、数据仓库、数据集市表结构设计思想完全不一样，因为大家的价值定位也是不一样的。哪怕是同为数据仓库也是分为几层的，每一层设计也是有区别的。如图6.1



所示,我们认为可以把 BI数据模型设计分为九层,这九层定位各不相同,但层层递进。

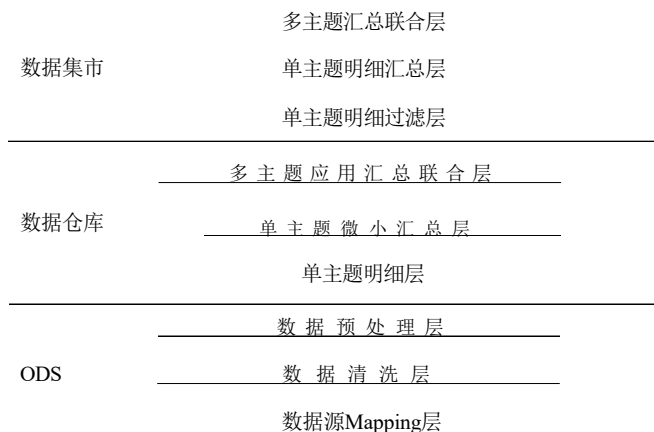


图6.1 BI体系九层数据模型

下面是每一层的定位简单介绍。

(1)ODS:ETL 的辅助层,提升ETL 能力,减轻对源及DW 的压力,简化ETL ,复杂程度。

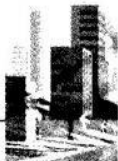
① 数据源 Mapping 层:目标是快速获得源数据到本地;Mapping 层表结构和源尽量保持一致,这样更新的时候可以把对源的压力降低到最低点。

② 数据清洗层:源数据过来之后还是要做一些检验,比如说唯一性检验、一致性检验、格式检验等,并进行校正,对于检验和矫正都要不断完善规则。这里的数据清洗和数据挖掘清洗意义不一样,数据挖掘清洗的往往是一些异常值,而这里的清洗就是清除掉不规范的数据,至于这个值为零或是很大都可以不管,只要符合业务规则就可以;当然也可以加入人工编辑来进行审核及纠正;这个层级对后面ETL 不出错有很大帮助,可以保障数据仓库数据准时更新,另外对后面数据质量也是一个保证。数据清洗放在前面主要是因为这里的数据相对比较原始,有利于保证后面数据的统一性和正确性,并有利于追踪问题。

③ 数据预处理层:BI 里面公认 ETL 比较复杂,很费心,调优难度大,为此很有必要在复杂处理之前做一些预处理,这层的价值定位就是为后面ODS 数据 ETL 进入数据仓库做些处理,主要包括生成增量数据临时表、派生字段、派生联合表等;这些处理将有助于ETL 数据到数据仓库变得简洁快速。

(2)数据仓库:定位为分析型需求的公共数据源,数据统一,粒度明细;下面我们看看里面各层在设计上怎么来提升数据仓库的能力。

① 单主题明细层:这层是数据仓库最基本的数据层,数据分主题,优先采用星



形结构的数据模型，这层的价值定位为**基础数据层**；不管任何需求都可以从这里找到需要的数据；这里要求数据全而准，并且结构化设计，应用性能其次考虑。

② **单主题微小汇总层**：明细数据直接拿来用的可能性一般不大，往往都需要汇总到一定粒度，一般来讲各种维度交叉汇总会比较多，这样我们可以把集市应用需求比较大的交叉汇总在数据仓库层面进行预备，这件事情就是这层要做的，通过一段时间BI项目实施后我们可以总结出部分这种微小汇总层，因为这个层次的存在，后面我们的数据处理中80%的数据可以从这里来，因为汇总对粒度的提升，数据的削减，从而带来了系统性能的大幅提升；不过要注意这里不同的汇总表会有一定冗余，要特别注意信息定义的一致性，如果信息定义发生变化要改一起改；这层的数据表需要根据应用不断丰富及完善。

③ **多主题应用汇总联合层**：上面在仓库层面会进行一些单主题微小汇总，这个汇总是对应用支撑的初步准备，除了表太大影响性能之外，还有个问题就是跨表关联，特别是大表关联更是致命的，为此在仓库层面我们还可以设计一批跨主题表的数据联合，存放到新的表中，这部分数据应用更为彻底；典型的如数据挖掘中的客户视图，产品生命周期等宽表都是应用对象；这些表对象有大量应用，这样可以为很多集市提供统一数据，有利于各个集市数据的统一，减轻后期应用开发的难度，并且提高系统性能。

(3) **数据集市**：集市定位在具体项目及领域的应用，以完成项目最终需求为目的进行设计，不管是数据定义还是数据模型都需要保持足够的灵活性，这样便于修改，毕竟需求是多变的，里面也可以从三层来进行设计。

① **单主题明细过滤层**：面向未来，很多公司数据仓库和BI应用是相互独立的，这个思路值得肯定，所以对于每个项目来讲首先考虑的是自己项目集市需要保留一份相对明细的数据，以便将来需求的扩展及变动；如果空间不是问题，这层数据基本保持和仓库表数据一样粒度，但可以进行记录过滤，比如只取近两年制造业数据等；如果数据没必要太明细，并且数据量有点大，那么我们可以进行一定程度的汇总，将数据量汇总到可以接受的范围。

② **单主题明细汇总层**：虽然上一个层也可以进行汇总，但大家目标不一样，上一个层次汇总是为了减小数据量，这个层的汇总是为了更贴近应用，直接把数据汇总到适合应用的粒度；也可能是为后面的多主题汇总数据联合进行单主题数据的准备，这样通过减小数据量，有利于后面的数据联合。

③ **多主题汇总联合层**：数据分析一般都需要多主题联合分析，比如说供应链就是个多主题的联合问题，客户生命周期也是多主题联合问题；为此需要为前端联合分析提供对应的多主题数据联合表，并且汇总到应用需要粒度即可，没必要太过明细。

纵观上面九层，我们可以层层推进地设计数据模型，也可以根据数据量和应用的复杂程度有选择地对某些层次进行设计，其中灵活性比较大，但如果能理解每个层次的价值定位，这九个层次的设计会是一个整体，能够为 BI 系统模型提供整体而又独立的设计架构。

6.3 是否能构建适合公司自身的软硬件平台？

在BI 产品市场中免费的产品较少，客观地说免费产品和收费产品还是有比较大的差距，所以选择软硬件平台意味着投资，难于随便更改，所以要非常慎重。当然也见过不少财大气粗的公司几乎把所有主流 BI 产品买了个遍，这种公司虽然钱不是问题，但最后在具体使用的时候一样需要做出选择。有效的做法还是从最终的效果要求来剖析我们到底需要什么软硬件配置，再结合当前的产品发展趋势来对软硬件进行选择。

(1) 数据仓库及 ETL 选择：在电子商务行业，大数据带来的 I/O 问题始终制约着数据仓库访问和 ETL 处理性能，表结构的设计一方面可以提升整体性能，另一方面也是为了解决性能问题，但就表结构再怎么设计充其量也是优化，瓶颈最终还是软硬件配置和设计。这里我们先了解一下我们 ETL 每天更新的要求。

- 优秀：2个小时以内。
- 可以接受：不超过4个小时。
- 不可接受：超过4个小时，需要考虑优化。

如果性能不可能接受，有下面一些办法进行调优。

- 提升数据仓库服务器硬件配置：磁盘、CPU、内存、网络都可能是瓶颈，特别是硬盘，现在 SAS 盘价格还是比较贵，这个投资对电子商务行业这种大数据量来讲还是不小的投资。

- 传统 ETL 工具的并行处理也可以进行优化，但其解决的问题始终还停留在 ETL 的抽取和计算过程，对于目标数据仓库的加载和应用查询还是无能为力，也就是说并不能彻底解决数据仓库整体性能问题。

- 最好的办法就是分布式部署，云计算概念，典型代表就是 EMC 公司的 Greenplum 产品，且已经在数据仓库行业获得了成功，并且在2012年进入Gartner 最具魔力象限，当然这套方案是要付费的；免费的 Hadoop 技术当前也流行，通过几十上百的普通 PC 分布式部署，把数据仓库大数据存储和计算压力分不到这些便宜的机器上，从而实现廉价解决方案；不过要注意当前云计算还是刚起步，在设计过程中对人的要求还是比较高，能否找到合适的人是问题。

(2) 报表分析：要评价一个报表分析好不好除了看数据准不准这类开发问题，



但更重要的就是要像 Google 和 iPhone 一样重视用户体验，报表分析除了功能强大、开发简单之外，速度、美观、简单同样必不可少；很多厂商在性能上有自己解决办法，但大部分对开发者设计能力要求较高，要求开发者对产品非常熟悉才可行，最终落实到项目上当然很难出来其标榜的性能；前端报表分析的内存化技术应该引起大家注意，现在 Gartner 前端报表工具最具魔力象限连续几年纳入了 Qlik-view 这个产品，这个产品代表了基于内存的报表分析技术，用户体验极佳。

(3) 数据挖掘：数据挖掘工具毫无疑问 SAS EM 和 SPSS Clementine 最为流行，不过工具需要有，但仅有工具远远不够，做好数据挖掘最起码还得具备下面三种能力，否则再好的软硬件配置也无能为力，这点相比较其他方面更为明显。

① 重视数据挖掘基础：数据挖掘技术最为重要的是算法，应用算法作为应用最长远的是优化，具备一定数据挖掘算法基础是必备的；不可否认挖掘算法要理解起来还是有一定难度，并且每个方向都有那么几个经典算法，全部要熟悉还是要潜心研究两年以上才可行，但对于有志长期从事数据挖掘工作的人来讲还是有必要的；另外数据挖掘技术研究较热，对应用开发来讲还是要经常学习一些新的算法，多参加一些数据挖掘学术会议会掌握最新研究方向和算法很有必要。

② 数据挖掘思想的灵活应用：数据挖掘的核心是算法，但灵魂是思想，要做到灵活应用还是要对挖掘几个方向的思想有深入理解。比如说用户分组，把同一个用户的账号全部找出来，这个技术可以理解为聚类，但一般聚类只有几个类，而这个结果显然是和用户数同一个数量级的组，也可以认为是关联度，但关联度分析往往是找最大关联组合，而我们这里是要找出每一个的关联用户，最后采用的是相似度函数进行定义和计算，思想还是聚类和关联。

③ 数据挖掘算法的灵活实现：挖掘算法实现不见得非要用挖掘工具，其实算法不复杂，我们知道.net,Java 都是很不错的实现工具，另外值得注意的是存储过程借助数据库技术在大数据批量处理方面具有明显的优势，如果算法能够借助存储过程来进行实现，在性能方面将非常有优势；数据库和存储过程对于数据挖掘人员来讲是必备技术，做数据挖掘的人如果不懂得怎么存放和整理数据是很被动的。

6.4 是否懂得合理利用数据挖掘技术？

在电子商务领域充斥着各种数据信息：文本信息、数据信息、视频信息、图片信息等，既有结构化信息又有非结构化信息，信息类型和信息量都非常巨大，业务面和业务逻辑也都比较复杂，涉及很多复杂的问题，如网络欺诈预警问题，市场趋势预测问题等，仅依靠传统技术方法不能得到较好的解决方案，因此这时就需要借助数据挖掘技术，而数据挖掘技术分支较多，每种分支的算法技术也较多。如何才能



充分发挥数据挖掘技术的作用去更好地解决电子商务领域的疑难杂症，这就需要我们懂得数据挖掘技术的应用环境及应用特征，如果一个挖掘问题没有找到合适的挖掘方法，那么不但不能解决问题，反而会使得问题的解决方案僵化，会很被动，因此是否懂得合理利用挖掘技术就显得尤为重要。

1) 数据挖掘技术的应用环境

挖掘技术种类繁多，每种技术都有它自身的特征，有它的适应环境，应用不当往往会取得适得其反的效果，所以在应用挖掘技术之前首先要分析问题的背景环境，然后配对挖掘技术环境，挖掘技术的环境常常包括挖掘问题类型环境(预测、分类、聚类分群、关联)、数据信息环境、算法环境、评估环境等。在做挖掘课题时，首先需要确定挖掘问题类型，找准问题方向，这就是需要确定挖掘技术类型，比如对未来未知类型的信息进行估计时往往使用预测型挖掘，对现有群体特征进行总结时往往是聚类挖掘，对产品之间的关联特征进行分析时往往属于关联型挖掘等，然后确定数据信息环境，这里主要是了解数据结构特征，数据挖掘技术根据算法逻辑也会有些环境要求，比如硬件要求、数据量要求、数据格式要求等，这就需要根据现有的数据信息结合算法要求去选择算法，因为算法繁多，往往需要选择多个算法，这时就需要熟悉算法评估规则，不同的算法也有不同的评估要求及评估规范，这些也都是在项目可行性分析前需要考虑的。总体而言，在应用数据挖掘技术时需要全面考虑挖掘技术的各种应用环境。

2) 数据挖掘技术的应用特征

数据挖掘技术在应用过程中有它自身的特征，要做好挖掘项目，就需要了解所选择挖掘技术的应用特征，这样才能达到好的应用效果，比如在做聚类分析时往往需要注意数据结构特征和数据的异常处理，有些聚类算法偏好数值型信息，且对异常数值特别敏感；有些算法偏重概率模型，对数据结构没有什么特殊要求，有些时候还需要确定样本的权重因子来进行差异处理。而在做预测问题时，比如时间序列分析就需要注重序列的平稳性、周期性、季节性影响因素及噪音的分布假设等，对异常信息处也比较敏感，分类问题往往对数据格式有特殊要求，比如因变量往往是字符型，有些算法需要自变量之间相关性较弱比如 Logistic 模型，有些算法需要自变量是数值型，有些学习型模型还要预防过度学习问题等。关联问题往往对数据格式有较强的要求，事前需要一些参数设置，比如提升度、置信度等，对关联结果也需要设置一定过滤机制等。总体而言，各种算法有它自身的优点，但在应用过程中也有它独特的要求，如果不注重这些潜在的要求，即使选对了挖掘方法，挖掘结果的应用效果往往也会大打折扣，达不到预期的效果。因此在应用挖掘技术时，需要考察该算法的应用特征，在应用过程中需要注重这些特征细节，才能更好地发挥挖掘技术的作用。

3) 数据挖掘技术的评估要求

挖掘技术种类较多, 往往一种技术难以满足要求, 所以在应用的过程中往往也会选择多种技术以便后续的比较分析工作的进行, 因此这就涉及技术抉择问题, 如何从所选的技术中选择出最佳的技术, 这就需要熟悉挖掘技术的评估要求及评估规范。不同类型的挖掘技术评估标准往往差异较大, 比如聚类技术, 一般而言以群体之间差异越大、群体内部差异越小为宜; 分类技术, 常常考察分类结果的准确率、覆盖率及提升度等指标; 关联技术常常分析事件之间关联的支持度与置信度水平等。有了这些评估标准和规范后就可以应用其对模型效果进行综合评估, 评估效果最好的为首选, 次之的可以作为备选, 有时根据评估结果也可以对多种挖掘技术进行综合处理, 即把挖掘方法进行组合形成新的挖掘模式, 这样可以综合多种挖掘方法的特征效果, 来满足应用需求。

综上所述, 要应用好挖掘技术, 就需要了解挖掘技术, 熟悉数据挖掘技术特征。目前市场上有很多挖掘软件(SAS/SPSS/CLEMENTINE/R 等), 集成了主流的挖掘技术, 包括预测技术、分类技术、聚类技术、关联技术等。基本常用的挖掘软件都能找到, 而且形成了很多标准化操作流程, 这些软件操作界面友好, 操作形式简单, 使得挖掘技术应用简单易行, 但使用时不能生搬硬套, 挖掘技术有它自身的特征, 是挖掘工具难以具备的, 要使用好挖掘技术就需要去了解这些技术的特征, 在应用的过程中也必须满足挖掘技术的应用要求, 这样才能获得良好的挖掘效果, 切实解决碰到的挖掘问题。

91

6.5 数据质量怎么解决?

商业智能的实施是建立在企业多年积累下来的丰富数据基础之上, 数据质量问题是制约商业智能项目成功的因素之一, 数据质量不高最终会对商业智能项目的效果产生影响, 甚至可能导致项目失败。数据质量的定义有多种, 但是目前并没有一个大家都公认的定义。我们可以认为符合我们分析要求不会影响结果的数据就是有质量的数据, 数据符合一致性, 完整性和正确性要求。

造成数据质量不高的原因有多种, 一部分是业务系统自身设计的原因, 比如一些数据业务逻辑存在问题, 数据缺失等; 有的是数据录入过程中的错误造成的, 电子商务行业很多数据都是需要用户自己输入的, 数据录入的时候由于录入人员的主观或客观原因可能会录入一些错误数据或者不合规范的数据; 还有一部分是在商业智能实施数据使用过程中造成的数据质量问题, 例如来自多个业务系统的数据一致性问题、ETL 程序设计存在漏洞、数据更新策略存在问题、数据仓库开发的 bug 等。

我们可以从图6.2所示的数据循环过程中看出数据从数据源到 Data mining

的过程是一个循环的过程，解决数据质量问题我们应该从这几个数据层面上层管理，在每一个层面需要解决的数据质量问题是不同的。如果前面一层出现数据质量问题就会对后面的数据质量产生影响，我们应该针对每层可能出现的情况进行处理，拿出有针对性的解决办法。

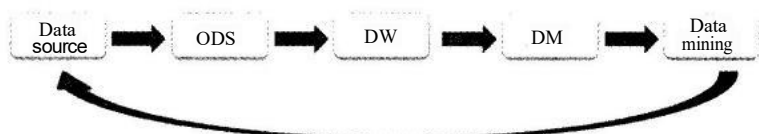


图6.2 数据循环过程

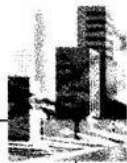
(1)数据源：电子商务企业数据多种多样，有网站数据、后台系统数据等各种数据源，不管数据源是什么方式，都需要提供满足BI基本要求的数据接口。各个数据源的数据要有标准的定义，大数据表需要提供可用于增量抽取数据的字段，可以识别新老数据。

(2)ODS：这一层一般和数据源表对表直接映射及对数据进行一些预处理工作，在 ODS 主要处理的数据质量问题是要保证数据记录与数据源的一致性，以及清洗一些对分析会产生影响的“脏数据”。ODS 层我们有一个 mapping 层，这一层的任务是要保证数据与数据源一一映射，这部分数据要处理好数据一致性的问题，不能在抽取的时候丢失掉记录也不能生产重复的记录。mapping 层之后我们会有一个数据清洗层，在这里我们处理掉一些“脏数据”，把一些记录进行清洗以符合我们下一步处理的要求。

(3)数据仓库：数据仓库是分析型需求的公共数据源，存放着用于分析和数据挖掘的各种明细的和汇总的数据，数据仓库的数据质量尤为重要。在数据仓库层，维度数据的唯一性是很重要的，必须保证每条维度记录是唯一的，相同的代理关键字只能对应相同的维度，维表里不能出现两条相同的记录。事实表设计的时候要保证独立业务表结构的独立性，事实表尽量单一主题，尽量不要有跨主题业务数据间联合表出现，避免数据仓库层面陷入设计混乱，造成数据不一致。

(4)DM：数据集市用于存放各种单主题汇总数据或者多主题联合汇总数据，在数据集市层主要可能出现的数据质量问题是指标定义不清晰、指标二义性问题。解决这一问题的方法是建立专门的指标集，明确每个指标的定义，保证每个指标名称与定义的一致性、指标名称的唯一性和指标统计计算方法的统一性。

(5)Data mining:数据挖掘在数据质量管理中发挥重要的作用，我们通过数据挖掘的帮助诊断数据质量问题，追踪数据质量问题的源头，通过灵敏度分析数据质量问题的影响因素，通过分析提出改进数据质量的方案。数据挖掘分析帮助改进数据源的数据质量，从源头上解决数据质量问题。



BI系统必须要有专用的模块对数据质量进行监控,对数据进行全过程的监控,保证数据正确性、数据一致性、数据规范性、数据重复性等。在BI项目实施中加入一些关键指标的监控是保证数据质量一个重要方法,通过对关键指标的监控可以很容易地检查数据质量是否存在问题。

总之,数据质量在商业智能项目中举足轻重,保证数据质量是商业智能项目成功的关键点之一。每个BI从业人员应该重视数据质量问题,并且应该培养整个企业重视数据质量的意识。通过技术手段和相关的标准提升数据质量,将数据质量问题对项目成功造成的负面影响控制到最低。

6.6 B2B行业的数据增值服务?

BI往往都是间接的为企业创造价值,而数据增值服务是BI直接为企业创造价值的主要途径之一,数据增值服务做得好与坏直接决定了BI为企业创造价值的多与少。数据增值服务不仅是电子商务行业的需求,同时也是社会信息化发展到一定程度的必然需求。

(1)行业需求: B2B电子商务行业与传统行业不同,节约成本不能从根本上解决企业的收益问题,企业发展的关键在于产品的创新,在电子商务行业一个有价值的新产品的开发成本与产品所带来的巨大收益相比几乎可以忽略,这就决定了电子商务企业BI的成败关键就看数据增值服务做得好与不好。

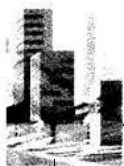
(2)客观市场发展趋势需求: 目前电子商务行业企业收入来源已经开始由传统的单一收取会员会费向多元化收费渠道转变,增值服务渠道带来的收入比重在逐年增加,而绝大部分增值服务都源于BI数据分析的支持,因此BI数据增值服务做得好与坏决定了公司整体增值服务的好坏,备受电子商务企业的关注。

(3)客户需求: 电子商务行业所面对的客户大部分是企业,而随着全社会信息化进程的推进,企业的信息化程度越来越高,企业的运作已经不再单纯的以管理层人为的主观意识去管理,而是更多地依赖于信息辅助, B2B行业的客户对信息增值服务渴求度越来越高,因此数据增值服务就成了电子商务行业BI今后的努力的重要方向之一。

做好数据增值服务需要做好以下几个方面:

(1)正确的价值定位: 数据增值服务也是服务的一种,满足客户需求才是最重要的,只有我们的服务能够真正的为客户带来价值,我们的服务才能得到用户的认可,数据增值服务自身的价值才能都得到体现。

(2)敏锐的直觉: 电子商务是一个发展非常迅速的行业,经常会有一些前所未有的用户需求产生,在电子商务竞争如此激烈的今天,谁能第一时间把握住机会,



谁就能够占得先机。

(3)快速实施：发现机会难，把握机会更难，一个有价值的数据增值服务如何快速的加以实现，优秀的实施策略、优秀的实施团队等方面缺一不可。

(4)循序渐进的推进模式：任何的数据增值服务不可能一上来就非常完美，尤其是受到基础数据的限制较多，因此这是一个不断优化与完善的过程，如果从开始就要求产品非常完美，那么这个产品将永远不可能与用户见面。所以产品的持续推进非常重要，只有不断地持续推进，产品才能不断完善。

(5)与时俱进：电子商务是一个变化非常快的行业，当前的服务能够满足用户的需求不代表以后一直能够满足用户的需求，随着市场变化与竞争对手的变化，我们的服务也要随之而变化，这样才能始终保持服务的价值所在，否则将会被淘汰。

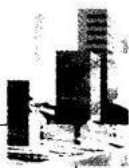
6.7 怎样获得持续的投入和支持？

企业的投入除了看公司有没有充足的预算，另外就是看我们BI本身能否有产出，前者只能说争取，后者作为BI人员来讲则可以主导。为此，下面我们从BI价值点，分步实施、稳步出成果，什么地方可以节省等这样三个层面来谈BI在钱上面的事情。

1)BI 价值点

要出成果先要知道BI价值点在哪里？我们从BI特点和业务这两个方向来进行理解。

(1)从BI特点来看：首先，BI提升分析效率，在报表分析这方面可以把日常OLAP分析和报表进行系统化定制，报表和OLAP分析不再困难，原来分析人员80%的做报表时间可以降到零；其次，BI提升分析能力，简单手工报表及EXCEL分析除了数据较少，汇总粒度过高，同时分析维度太少，其直接结果就是只能做到局部、片面、宏观的分析，并且这种分析难于升级，从而造成分析能力低下，而系统化的BI系统则可以不断迭代，不断丰富完善分析功能，分析数据全面，既可以看宏观有高度的数据，也可以深入看明细，操作起来灵活而又方便，再加上分析人员可以利用由原来20%的时间提高到100%的时间来进行分析，其分析效果、分析能力必将得到提升；第三，BI里面的利器——数据挖掘的潜在价值更是不可估量，B2B电子商务行业是典型的大用户量、高访问量、多产品线的行业，潜在机会和危机处处都是，了解用户喜好进行针对性的服务和支持是行业深入发展的必然需求，快速锁定欺诈行为和用户等潜在风险性预警也关系到企业命脉，这些都需要数据挖掘技术来实现。从上面可知，只要是能提升企业效率、效益，帮助企业智能化分析和决策的需求都是BI需要去做的，并且价值明显。



(2)按照现在流行的企业管理的平衡计分卡的要求, BI 在业务应用上也划分为内部运营、客户关系、学习与成长和财务状况等这四部分, 这种划分很全面。但是, 事实上当前没有几个公司能做到这种领先的程度, 相比而言, 较现实的做法还是BI 先侧重于企业核心竞争力的塑造, 比如说零售行业注重的是库存管理、供应链能力, B2B电子商务行业重视的是产品创新和用户规模, 品牌商重视的是产品和推广能力等, 不同行业的核心竞争力是不一样的, 哪怕同一行业中的不同企业的定位也是会有区别, 特别是在当前细分市场的年代。所以, BI 咨询人员务必理解企业核心竞争力, 并且设计出相应的需求。

2)分步实施、稳步出成果

有价值不一定有效果, 最终我们还是要拿出成绩来, 这也是我们提倡的整体规划、分步实施的切实定位, 目标就是稳步出成果。一般来讲每半年有成果产出比较可行, 涉及的方方面面风险的前文中已经详细介绍过, 在此另外强调几点:

(1)软件要选好: BI软件不要选太复杂的, 一定要能快速开发、部署的。

(2)需求不要一味求全: 只要对用户有较大的价值, 只有几个零散的需求点也可以先做; 不同于其他类型的 IT 项目, BI 项目先培养起使用客户是需求能够完善的前提条件。

(3)长期重视核心业务, 短期内兼顾其他需求: 从长远来看, BI 的实施也需要注重企业核心业务、企业核心竞争力的塑造方面, 但短期来说, 很多一般的需求也得兼顾, 这样有利于出阶段性的成果, 并且能够全方面了解公司业务, 这种积累对核心业务的 BI 应用非常有帮助, 比如说 B2B行业对内的客服及销售人员的分析需求, 就可以转换提升为客户需求, 可以直接开发到网站上支持用户自己来分析。

95

3)什么地方可以节省

清晰了BI 价值点, 并且逐步有成果之后应该有利于获得资源, 但不得不说BI 开发人员要注重节省, 并且循序渐进地进行项目的投入, 盲目投入不可取, 该投不投、缩手缩脚有时候也是致命的, 下面我们从三个方面来谈投入与节省。

(1)人力: IT 开发人力成本是持续成本, 也是主要成本; 要做好一个企业级的BI 系统必须要具备架构师和咨询专家角色的人才来掌控方向, 这个非常关键, 这个钱不能省; 开发人员可以大部分是新人, BI 上手快, 一般实施难度不大, 这样可以节省不少人力成本, 新人有高手带一般有半年时间都可以具备一定独立开发能力; 因为BI 的入门较为容易, 所以招聘方式可以采用资深加有潜质的新人, 资深可以引领方向, 新人能够快速入门, 快速成长, 成为性价比很高的中坚开发力量。

(2)软硬件投入: 软硬件投入必不可少, 但像很多企业把BI 产品都买个遍是不可取的, 很多企业完全用免费的也是有问题的, 软硬件的采购还是要立足企业的



真实需求及各个BI产品的特征来进行选择；软件方面包括数据仓库及ETL工具、前端OLAP工具、数据挖掘工具，这三者必不可少。总体来讲，ETL可以考虑数据库存储过程和一些免费工具（比如 Kettle），其他方面的免费工具里具体的应用要求还有较大的差距，建议购买付费软件。BI产品商家虽然都标榜自己的产品怎么好，别家怎么差，事实上架构相同的产品最终实施效果都差不多，相比较核心技术及架构不一样的产品间实施效果差距还是很明显的，并且贵的还不一定是好的。在拿不准的情况下，建议参考 Gartner 的评价，笔者与各种产品接触下来的经验与 Gartner 的评价较为接近。

6.8 小结

笔者根据在电子商务实施BI的实践经验，总结了下面几个在电子商务行业成功实施BI的关键点。

1) 强大的 BI技术架构

一个电子商务公司的企业级 BI 技术架构最起码要包括标准数据源接口、ODS、数据仓库、数据集市、业务系统接口。这五层在BI技术架构中都有明确的价值定位，只要确实理解了每个层的价值定位，即使不同的开发团队在开发过程中方法会有区别，但最终走向还是殊途同归。

2) 大数据的存储设计能力

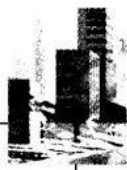
大数据是电子商务企业不得不面对的问题，大数据带来的最大挑战就是 ETL 性能差、查询慢，解决办法还是分布式部署及数据模型合理设计。分布式部署及云计算是解决大数据的有效方法，分布式或者云计算可以让数据的I/O 及运算压力分摊到不同机器上面，从而实现性能线性提升。在技术之外，合理的模型设计也是应对大数据处理的关键。ODS、数据仓库、数据集市表结构设计思想完全不一样，因为大家的价值定位也是不一样的。笔者将BI数据模型设计分为九层，建议层层推进的设计数据模型，或者根据数据量和应用的复杂程度有选择地对某些层次进行设计。

3) 合适的软硬件平台

如何选择软硬件，有效的做法还是从最终的效果要求来剖析到底需要什么软硬件配置，再结合当前的产品发展趋势来进行选择。

4) 做好需求

BI的需求开发其实就是培养需求、培养用户的过程，笔者对培养BI需求和用户提出几点建议：普及BI知识、用户决定需求、不要等完善的需求、BI在于帮助客户、统一业务规则、选择合适的项目经理。



5) 合理利用数据挖掘技术

首先，在应用数据挖掘技术时需要全面考虑挖掘技术的各种应用环境；其次，了解所选择挖掘技术的应用特征；第三，熟悉挖掘技术的评估要求和评估规范，应用其对模型效果进行综合评估；最后，市场上有很多挖掘软件以及集成主流的挖掘技术，使得挖掘技术应用简单易行，但是软件只是帮助我们的工具，挖掘人员必须清楚明白挖掘技术的原理和特征，避免生搬硬套，才能达到良好的挖掘效果。

6) 保证数据质量

笔者认为符合分析要求且不会影响结果的数据就是有质量的数据，数据符合一致性、完整性和正确性要求。解决数据质量问题我们应该从数据源到数据集市或数据挖掘中间的几个数据层面上层层管理，且每一个层面需要解决的数据质量问题是不同的。BI 系统必须要有专用的模块对数据质量进行监控，并且在BI 项目实施中加入一些关键指标的监控。

7)BI 提供数据增值服务

数据增值服务是 BI 直接为企业创造价值的主要途径之一。BI 能够帮助企业进行产品创新，扩展企业的收费渠道；BI 数据分析直接支持网站客户的需求已经成为行业中的现实，同时也是电子商务行业 BI 今后努力的重要方向之一。

8) 持续的投入与支持

获得资源有两种方式，一是等待公司分配，二是主动争取。BI 的价值点是 BI 争取更多投入与支持的筹码，让价值转变成可视化的成绩与效果才能使我们手中的筹码得到别人的认可。笔者提倡分步实施、稳步出成果的项目开发方式，这样有利于循序渐进的培养客户和完善整体需求架构，客户的认可和需求是BI 争取投入与支持的保障。另外，如果BI 在人力与软硬件配置上面有良好的规划和认识，可以在保证项目效果的同时节约不必要的支出。

第七章 焦点科技 BI 应用价值及展望

焦点科技作为一个以电子商务平台为主体的一个互联网企业，既具备一般企业的基础特点，同时又具备 B2B 网站运营的特点。

7.1 应用价值

焦点科技作为一个以电子商务平台为主体的一个互联网企业，既具备一般企业的基础特点，同时又具备 B2B 网站运营的特点。所以 BI 在焦点既担负着企业运营分析支撑又担负着网站分析支撑的责任。

1) 价值一，企业运营分析支撑

企业运营分析支撑见图 7.1 所示：

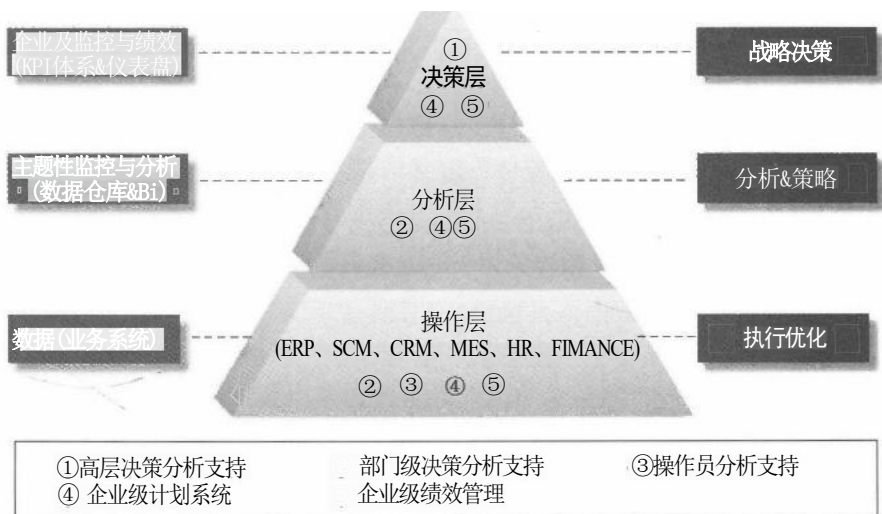


图 7.1 企业运营分析支撑

(1) 高层决策支持

通过 BI 的分析，使公司高层对公司的经营状况有一个清晰的认识，及时发现公司存在的问题与不足，结合公司所处市场环境的变化，对公司未来的发展方向做出正确的决策；通过完善的企业级计划系统将工作任务下发给各级部门，并通过



BI 的企业级绩效管理的系统时间监控工作任务完成情况。

(2) 部门级决策支持

通过BI 的分析，部门级领导能够更加深刻地理解公司高层下达的工作任务，并且对部门内部的工作能力有一个合理的认识，从而将任务合理地分解到操作层，并实时监控部门内工作任务的完成情况，便于发现问题及时调整。目前在焦点已经有超过70%的忠诚领导都在使用 BI 系统，工作效率提升了30%以上。

(3) 操作员分析支持

• 提高工作效率

操作员的工作往往是一些纯手工处理的事务，效率比较低下；通过BI 的数据分析与挖掘，我们完全可以把纯手工处理的事务，变成半人工，甚至是完全由计算机自动来完成。例如我们的用户信息审核，以前完全需要人工来判断，通过自动化编辑系统，利用文本挖掘技术，绝大部分的用户信息审核都可以通过计算机来完成，只有少量的信息需要人工判断，大大提升了审核的效率，节约了人力成本。

• 提高工作质量

BI通过数据分析，为操作员提供分析支持，使他们的工作更加具有针对性，工作质量有显著提升。例如精确化销售系统，把最有可能销售成功的用户提供给销售人员，并且告知用户相关的分析结果，例如什么样产品可能是用户最喜欢的，用户大概能接受什么样的价格等。有了这些分析的支撑使我们的销售能够知己知彼，自然他们的工作质量就能有很大的提升。

99

2) 价值二，网站分析支撑

(1) 个性化服务提升网站服务水平

电子商务站点面对的是不同的顾客群体，利用商业智能来识别顾客提供个性化服务，将会大大提升网站的服务质量，提高用户对网站的认可程度，增强用户的忠诚度。这些个性化服务主要有四种类型：问候、用户定制、实时消息和特色推荐。当顾客访问站点时，站点将自动辨认出顾客的身份，用个性化服务为顾客提供量身定做的促销商品和市场推广活动。通过对反馈信息进行商业智能分析，可以得知实行个性化服务后，站点的盈利水平是否得到提高。

(2) 网站功能效果监控与优化

通过网站的流量分析与用户行为分析，监控网站各功能的效果，为合理优化网站功能模块提供依据。例如通过用户注册流程的数据分析，发现网站现有注册流程中哪些环节导致用户注册时流失比较严重，从而改进用户注册流程，提高用户注册成功率。

(3) 网站资源合理利用

网站资源是电子商务公司盈利的重要来源之一，因此网站资源是否能够得到

最合理的利用就决定了公司收入的多少。利用BI的数据分析，对资源的价值作出合理的评估，避免资源浪费，同时也能发现新的有价值资源。例如通过数据分析，发现用户搜索相对较热但是又没有用户购买的关键词资源，并把它们推荐给用户，既提高了用户的价值，又能为公司带来收益。

(4) 实现数据增值与服务

目前焦点科技已经在开发数据分析类信息增值服务产品“数据罗盘”，这是BI在焦点科技的又一大突破。数据分析类产品要以增加目标客户价值为导向，将数据转化为能够帮助用户的知识，用户通过这些知识实现盈利。在这个过程中焦点实现收益。信息增值服务产品开创了电子商务一个新的盈利渠道，使公司在竞争激烈的电子商务市场中更具竞争力。

7.2 BI 应用展望

BI在初级阶段的应用主要是基本的数据分析和报表，进一步的BI应用将会产生更大的价值(图7.2)：

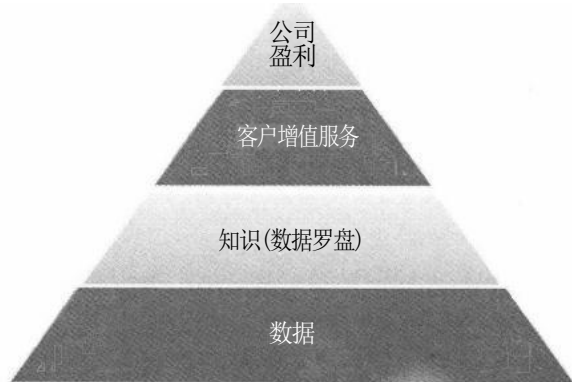


图7.2 BI的应用展望

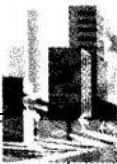
笔者认为BI的应用将会朝以下几个方向发展：

1) 从传统功能向增强型功能转变

目前大多数的BI系统都已经能够实现数据分析层的功能，而相对价值较大的数据挖掘、企业建模是BI系统应该加强的应用，这部分应用相对来讲难度比较大，无论是人力还是财力要求都要比初级的BI系统要高得多，但是它能够为企业带来的价值也是巨大的，是将来BI的主要发展方向之一。

2) 从单独的商业智能向嵌入式商业智能发展

从单独的商业智能向嵌入式商业智能发展是目前商业智能应用的一大趋势，即在企业现有的应用系统中，如财务、人力、销售等系统中嵌入商业智能组件，使普



遍意义上的事务处理系统具有商业智能的特性。但是嵌入应用系统应该是BI系统的某个组件而不是整个BI系统，一个相对完整的商业智能开发过程，如企业问题分析、方案设计、原型系统开发、系统应用等过程是不可缺少的，因此不可能把整个BI系统集成到多个应用系统中，而嵌入的一般都是BI的一些应用层分析结果，这样既不会对BI的整个开发流程有影响，又能对业务系统起到直接的帮助，无论是成本还是效率都是目前最好的选择。

3) 从间接价值向直接价值发展

企业通过智能外网把企业商业智能所形成的数据、增值信息和知识像日用品(Commoditization)一样地销售给新类型的客户，实现数据的市场化，可以把企业的数据金矿变成企业新的利润增长点，发现新的市场机会和新的商务模式，增强企业的核心业务，为企业提供了持续的核心竞争力。

7.3 小结

BI在焦点既担负着企业运营分析支撑又担负着网站分析支撑的责任。

企业运营分析支撑包括高层决策支持、部门级决策支持和操作员分析支持。BI帮助高层对公司的经营状况有一个清晰的认识，从而制订完善的企业级计划并且下发到各级部门，通过BI的企业级绩效管理系统监控任务完成情况；BI帮助部门级领导合理认识部门内部的工作能力，从而合理地分配任务到操作层，并监控任务完成情况；BI能够帮助操作员将日常的一些手工事务变成半人工半自动或全自动，能够为操作员提供分析支持，使他们的工作更加具有针对性，从而提高操作员的工作效率和工作质量。

网站分析支持是BI对平台和平台客户的分析支持。利用商业智能来识别顾客提供个性化服务，将会大大提升网站的服务质量，提高用户对网站的认可程度，增强用户的忠诚度；通过网站的流量分析与用户行为分析，监控网站各功能的效果，为合理优化网站功能模块提供依据；通过网站的流量分析与用户行为分析，监控网站各功能的效果，为合理优化网站功能模块实现网站资源合理利用提供支持；利用BI技术开发的数据产品为客户提供数据增值与服务，如焦点科技开发的数据分析类信息增值服务产品“数据罗盘”。

BI应用的发展将会实现BI更大的价值。BI的应用将从传统的数据分析层功能增强至价值更大的数据挖掘和企业级建模；从独立的BI系统发展到在企业现有的应用系统中，如在财务、人力、销售等系统中嵌入商业智能组件，使普遍意义上的事务处理系统具有商业智能的特性；商业智能将直接为企业创收，实现数据的市场化，可以把企业的数据金矿变成企业新的利润增长点。

第八章 商业智能最新发展方向

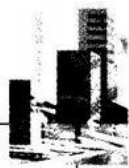
未来 BI 的发展在信息融合方面将更集中；在用户体验方面将更人性化、操作更简单；在应用领域方面将更广泛，移动 BI 将成为新战场；在系统部署方面将更智能，云中部署或成主流方向；在商用价值方面将更广阔，应用门槛将变低，中小型企业也能定制各种应用 BI；在技术方面，可视化与交互式分析将成为新亮点。整体而言，未来 BI 的发展将会更加智能，全面集成，技术更成熟，应用更广泛。

8.1 信息集成度更高

未来的商业智能发展趋势将是基于全面信息集成的服务，是一种企业级跨部门运作的基础信息系统，可以联结企业各个岗位上的工作人员，可以联结企业各类信息系统和信息资源，真正实现跨平台，最后演变成门户化。在基于企业战略和流程的大前提下，BI 可通过类似“门户”的技术对各个业务系统进行整合，使得 BI 与 OA、CRM、ERP、SCM 以及其他系统之间能实现融合集成，系统之间的结构化数据能通过门户管理平台互相调用、展现，全面提供决策支持、知识挖掘、商业智能等一体化服务，实现企业数字化、知识化、虚拟化。这时 BI 或已远远超出传统 BI 的范畴，演变成为企业综合性强大管理支撑平台，换为“企业知识门户EKP”、“管理支撑平台 MSS”等名称或更能体现其价值。

8.2 操作更简单、体验更人性化

未来 BI 的门户更加强调人性化，功能日趋“傻瓜”，强调易用性、稳定性、开放性，强化人与人沟通、协作的便捷性，重视对于众多信息来源的整合，并进一步完善可以拓展的管理支撑平台框架，从而改变以往“人去找系统”的现状，实现“系统找人”的全新理念与功能。今后的 BI 系统能让合适的角色在合适的场景、合适的时间里获取合适的知识、数据，充分发掘和释放人的潜能，并真正让企业的数据、信息转变为一种能够指导人行为的意念、能力。其实“人性化”也是一种“自动化”，充分体现管理系统的最大价值与作用。



8.3 速度更快、性能更优：BI 充分融合云技术

云计算概念方兴未艾，云计算功能极其强大，以云为基础的商业智能在线服务将成为全新的商业智能部署的主流方向。目前，云计算的重要性已经能够影响到各个 BI 厂商未来的生存线。从某种意义上而言，只有产品是面向云规模架构设计并符合云运营模式的 BI 软件才能获得用户企业的青睐，才能在今后持续取得成功。尽管 BI 向云迁移的过程中仍然面临许多的挑战，但随着越来越多的企业将其业务应用置于云端，“在云中部署 BI”已不是一个可望而不可即的理想目标。据悉，BI 专业厂商 Informatica 已经进行了向用户交付云服务的尝试。此前该公司发布的 Informatica BI 数据集成平台已经能同时部署在预装系统或“云”网络之中，为用户提供云端集成服务。

目前市场上已经有一些云集成 BI 服务，SaaS BI 日渐雄起，受到中小企业的青睐。IDC 机构预测，SaaS BI 将成为今后一两年 BI 市场的一大热点，也是各大 BI 厂商角逐的焦点。因为传统的 BI 工具过于昂贵，且建设过程较为复杂，从中小企业的自身状况和当前经济形势分析，低费用高效的 SaaS 租用模式无疑已成为中小企业的重要选择，将给企业带来良好的发展平台。Gartner 预计到 2012 年，25%~30% 的企业将使用“软件即服务”SaaS 这种业务模式，15%~20% 的中小企业组织将通过 SaaS 获得具行业特色的分析应用软件，SaaS BI 将成为 BI 投资组合应用的标配。然而基于现阶段的 SaaS 的发展状况看，该商业模式的应用尚不成熟，还需大力培育、推广，只有到相对比较成熟、完善的情形之下，BI 才能借助该模式的力量达到星火燎原的目标。时下，BI 应用还是会以传统模式为主。

103

8.4 可视化技术及交互式分析成技术发展亮点

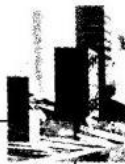
目前，已有越来越多的企业用户不再满足于传统较为单一、呆板的图像展现，因为更多的分析需求需要借助图像 OLAP 这种多维分析模式来完成，比如噪音数据、数据集趋势等。据称，Oracle 已经率先开展了数据可视化分析的先河，为用户提供可视化数据探索服务。这种类型的分析手段也将是未来 BI 领域的又一亮点、一大趋势。同时，BI 传统的过滤、上钻、下钻、比较等功能也难于满足一些特殊企业用户的分析要求。因此新一轮的 BI 高级分析需求已经在不同行业悄然升温。在未来的 BI 平台中，预测、分摊、假设模拟、数据挖掘等交互式技术将成为新一轮的 BI 系统升级改造方向。据称，集成整合 Essbase 服务技术的 Oracle 商业智能平台已能为用户提供面向策略级、未知信息的分析预测能力和个性化自助式定制等。

未来BI将融合协同、智能、门户、无线等精髓，并日趋经济、简捷、高效，将不断绽放“新花”，焕发出迷人的光彩。BI厂商只有抓住这些发展趋势和技术潮流，加以研发推广，才能更好地为国内企业决策管理服务，并获得企业的青睐，抢得先机。

8.5 小结

未来BI的发展将会更加智能，全面集成，技术更成熟，应用更广泛。

首先，未来的商业智能发展趋势将是基于全面信息集成的服务，将是一种企业级跨部门运作的基础信息系统。其次，未来BI的门户更加强调人性化，功能日趋“傻瓜”，强调易用性、稳定性、开放性，强化人与人沟通、协作的便捷性，重视对于众多信息来源的整合。第三，以云为基础的商业智能在线服务将成为全新的商业智能部署的主流方向，面向云规模架构设计并符合云运营模式的BI软件才能获得用户企业的青睐。第四，可视化技术及交互式分析将成为BI技术发展的亮点。



参 考 文 献

- [1] 戈尔夫雪利, 等. 数据仓库设计: 现代原理与方法[M]. 战晓苏, 吴云浩, 皮人杰, 译. 北京: 清华大学出版社, 2010
- [2] Codd EF, Codd S B, Salley C T. Providing OLAP(On-line Analytical Processing)[J]. User-Analysts: An IT Mandate, 1993(9)
- [3] Jiawei Han, 等. 数据挖掘: 概念与技术[M]. 范明, 孟小峰, 等, 译. 北京: 机械工业出版社, 2007
- [4] 陈封能, 斯坦巴克, 库马尔. 数据挖掘导论[M]. 范明, 范宏建, 等, 译. 北京: 人民邮电出版社, 2011
- [5] WHInmon. 数据仓库[M]. 王志海, 林友芳, 等, 译. 北京: 机械工业出版社, 2009
- [6] Tom White. Hadoop权威指南[M]. 周敏奇, 等, 译. 北京: 清华大学出版社, 2011
- [7] Ralph Kimball, 等. 数据仓库生命周期工具箱[M]. 唐富华, 孙媛媛, 译. 北京: 清华大学出版社, 2009
- [8] 郑人杰. 计算机软件测试技术[M]. 北京: 清华大学出版社, 1992
- [9] Gartner. 2012 年数据仓库魔力象限报告[R]. 2012
- [10] David Parmenter. Key Performance Indicators[M]. Hoboken, NJ: John Wiley & Sons, 2007
- [11] Michael Porter. How competitive forces shape strategy[J]. Harvard Business Review, 1979(3)

文章简介

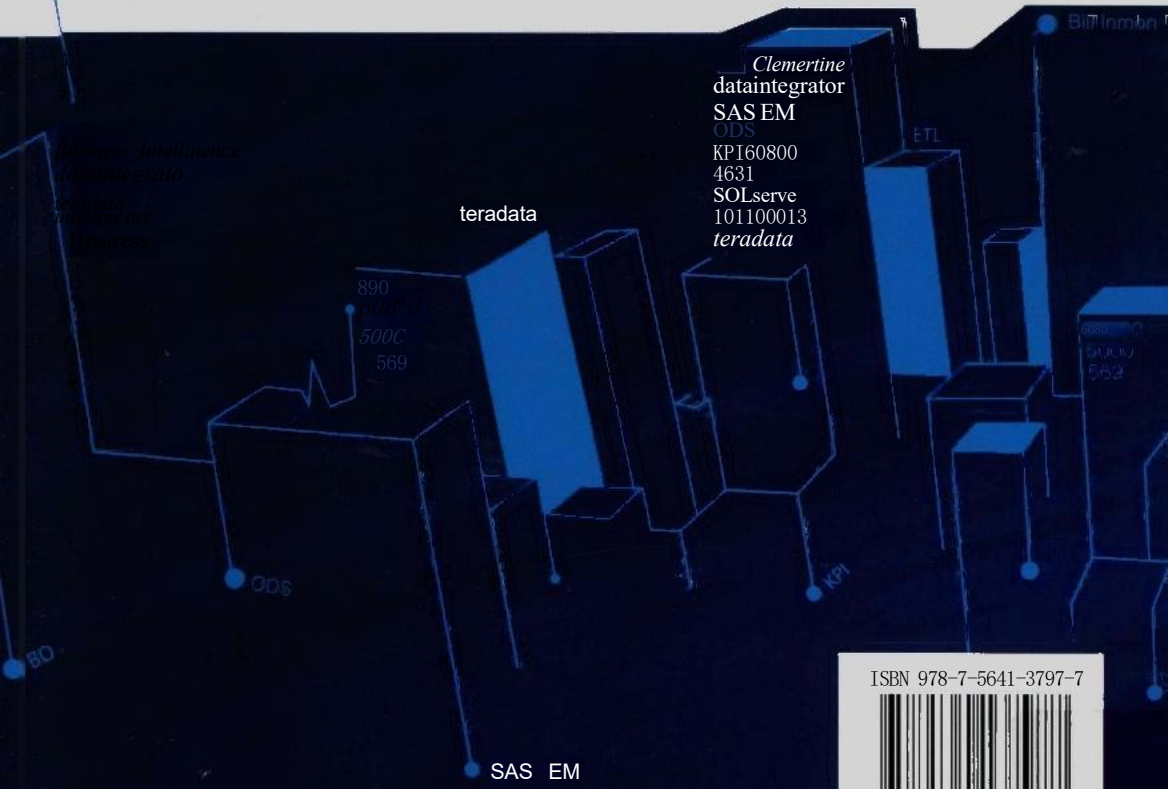
电子商务是信息化时代的产物。相比于传统行业，电商行业在信息构成方面有其独特的特点，适用于电子商务的商业智能实施方案必须能够解决复杂数据源、数据分析实时性要求高和大数据处理等问题。

本书根据实际的商业智能实施经验，从解决方案、设计理念以及技术等角度详尽介绍了电子商务中商业智能的应用以及发展趋势；并结合云计算技术探讨了同时数据仓库、OLAP 分析、数据挖掘等技术进行了深入价值和电子商务行业成功实施BI的经验。

本书内容深入浅出，是企业领导决策层、商业智能开发人：

本书特点

- 全面实用地介绍商业智能，涵括数据仓库、ETL、OLAP 分w、 蚁据亿地，以及无验的正业级B 内容。
- 不拘泥于软件和产品，注重BI的本质和最佳实践的追求。
- 既有通用的方案和技术标准，又有相应的技术实现。
- 注重BI实施策略及理念，统一规划、逐步实施BI的经验具有较高的现实性。



ISBN 978-7-5641-3797-7

9 178756411379771 >

定价：30.00元