

# 面向企业信息基础设施集成的商务智能系统框架<sup>\*</sup>

倪 旻 徐晓飞 邓胜春

(哈尔滨工业大学计算机科学与工程系 哈尔滨150001)

**摘要** 商务智能系统是多种技术的综合体,提出了面向企业信息基础设施集成的闭环商务智能系统框架,商务智能系统的生命周期和信息、知识在其中的作用。以及通过现有的企业信息基础设施构建商务智能系统,通过商务智能系统达到持续优化的方法,使用包含企业本体论、商务过程模型、元数据仓库的企业元模型和两级映射模式对集成过程中的语义一致性进行控制。

**关键词** 商务智能,数据仓库,知识发现,信息基础设施,基于语义的集成

## An Enterprise Information Infrastructure Integration Oriented Business Intelligence Framework

NI Min XU Xiao-Fei DENG Sheng-Chun

(Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001)

**Abstract** Business intelligence (BI) system is a syntactic system based on many technologies. A closed-loop BI system framework is presented, along with the study of BI system's life-cycle, and the methodology of using current existing enterprise information infrastructure to build a BI system to achieve the state of continuous optimization. Meta-enterprise model, which contains enterprise ontology, business process model, and meta-data repository, and two layers mapping schema are used to control the semantic coherence in the integration process.

**Keywords** Business intelligence, Data warehouse, Knowledge discovery, Information infrastructure, Semantic integration

## 1 绪论

在信息时代,企业管理者如何运用企业运营过程中产生的大量业务数据迅速做出正确决策,已成为影响企业生存的关键问题。商务智能<sup>[1,2]</sup>通过把顾客和运营信息转化成知识,提供对未经处理的数据进行巩固和分析并将其加工为可执行的决策信息的能力,使企业能利用不同来源的顾客、运营和市场信息增强竞争力。

商务智能的本质是多种技术的综合体<sup>[3]</sup>,所包含的技术有:数据挖掘(Data Mining)、数据仓库(Data warehousing)、在线数据分析(OLAP)、决策支持系统(DSS)、知识管理(Knowledge Management)、人工智能(Artificial Intelligence)等。对数据的有效利用,发现其中的决策依据是商务智能的核心内容。商务智能为企业比单纯的数据挖掘或OLAP的结果具有更高可用性和商业价值的知识,在现阶段的数据挖掘技术研究中,大部分工作侧重于挖掘算法的研究,而对数据挖掘的结果给出商业解释,限于采用的具体技术(如神经网络等)及缺乏对领域知识的运用而远远滞后于数据挖掘算法的发展,而这正是商务智能所致力解决的问题。现有的OLAP应用提供上钻、下钻、旋转、切片和其它操作,也可以通过以交叉表、图表及其它方式给出分析结果,但OLAP的汇总和比较是交互进行的,在实际应用中往往需要专门的统计人员进行分析模型的选择和结果的解释,而商务智能面向企业的管理人员,通过商务智能软件集成和分析企业数据,发现潜在的业务机会,进而制定更好的业务决策,并提供竞争优势<sup>[4,5]</sup>。本

文对企业商务智能系统的相关问题进行了研究,研究了企业信息基础设施在商务智能系统中的作用,提出了基于数据仓库的商务智能系统构架,以及商务智能系统的生命周期和构造方法。

## 2 商务智能系统中的信息基础设施介绍

商务智能为企业的商务过程提供决策支持,而决策本身是一个具有系统性、综合性的过程<sup>[6]</sup>,因此,商务智能的系统框架应该在企业现有的信息基础设施(包括ERP、SCM、CRM、其它业务系统和数据仓库等)之上构造,以充分利用现有系统的数据基础,进行商务分析、知识发现、商务问题优化,并保证商务知识在企业内部的一致性。下面,我们先对企业现有信息基础设施中的关键模块进行介绍,然后在此基础上提出面向制造业企业的商务智能系统框架。企业现有信息基础设施通常由下列系统中的一个或多个组成。

### 2.1 数据仓库

商务智能的作用是通过智能数据分析、知识发现、数据展现等为企业的商务活动提供支持,而上述这些活动必须基于企业商务活动产生的大量数据。从目前企业信息化的程度来看,数据仓库已经被广泛地采用。尽管数据挖掘等技术并不一定要有数据仓库的支持,但它们仍然经常被看作是对数据仓库中大量数据的处理,因为数据仓库为数据挖掘和分析提供了丰富的数据资源,而数据仓库的应用也使企业对基于数据仓库的分析工具产生更多的需求。

根据数据仓库创始人 W. H. Inmon 在其《建立数据仓库》

<sup>\*</sup>本研究由863计划(2002AA413310)和国家自然科学基金(60084004)资助。倪 旻 博士研究生,主要从事商务智能、数据挖掘等研究。徐晓飞 教授,博导,研究领域为CIMS,数据库,信息管理系统,敏捷制造。邓胜春 博士,副教授,研究领域为供应链管理,分布式信息处理,多数据库。

一书中对数据仓库所下的定义,数据仓库就是面向主题的、集成的、稳定的、随时间变化的数据集合,用以支持经营管理中的决策制定过程<sup>[7]</sup>。数据仓库中的面向主题与传统数据库的面向应用相对应。主题是一个在较高层次将数据归类的标准,每一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,必须进行数据加工和集成,这是建立数据仓库的关键步骤,首先要统一原始数据之中的矛盾之处,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的内容,而不是未加工的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很少修改或根本不修改的;数据仓库是不同时间的数据集合,它要求数据仓库中的数据保存时限能满足进行决策分析的需要,而且数据仓库中的数据都要表明该数据的历史日期(时间戳)。

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,而是来源于其它数据库,它要建立一个较全面和完善的信息服务应用的基础上,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,数据仓库查询系统有以下一些特点。

数据仓库和生产信息系统不同,它保存的是历史数据,一般不作修改,因此用户针对数据仓库的交易主要是查询。数据仓库的查询和业务处理系统的查询有很大的区别<sup>[8]</sup>。

业务处理系统的查询一般都很简单,通常较少使用表联接操作,每次返回的数据量都很小,这类问题的特点是“知道自己要什么”,例如根据银行账号查余额。这类操作都是基于索引进行的,如账号可以作为储蓄系统的索引。由于这些特点,数据库大小对系统性能影响不大。

数据仓库系统的查询通常都非常复杂,主要有两种查询方式:一种以报表为主,从数据库中产生各种形式的业务报表,这种查询是预先计划好的(Pre-Defined Query);另一种则是随机的、动态的查询(Ad-Hoc Query),对查询的结果也是不能预料的。

数据仓库中的查询由于其复杂性,一般会经常使用多表的联接、累计、分类、排序等操作,这些大都要对整个表进行搜索。每次查询返回的数据量一般很大,对于 Ad-Hoc 查询而言,经常需要根据上次查询的结果进行进一步的搜索,这个过程常称为钻取。

根据这些特点,数据库大小对于数据仓库系统的性能影响很大。当数据仓库投入使用后,各业务部门的要求会越来越多,使得数据仓库中数据量的平均增长速度很快,因此,设计数据仓库时,系统的可扩展能力是必须考虑的重要因素之一。

系统的并行能力是另一个要考虑的重要因素。因为查询的复杂性,每个查询将占用很多的系统资源,如果并行处理能力不强,多个用户同时发出查询请求时,响应时间会显著延长。

## 2.2 企业资源计划(ERP)

ERP 系统侧重于整合企业内部的资源,将各业务环节贯穿起来,实现内部效率的最大化。使用 ERP 系统,所有信息实时地在业务部门间传递,保证了企业在任何时间点都是一个整合的实体在运作,对于采购、库存、生产、成本及监控提供了坚实高效的基础。

现代的 ERP 系统,又可理解为扩展 ERP 系统,其在强调提高企业内部效率的同时,也注重对企业外部资源,如供应

商、客户和营运商的协调管理。它更加符合现代供应链的历年,即强调合作和提高渠道效率。

## 2.3 供应链管理系统(SCM)

供应链(Supply Chain)是指:围绕核心企业,通过对信息、物流、资金流的控制,从采购原材料开始,制成中间产品以及最终产品,最后由销售网络把产品送到消费者手中的将供应商、制造商、分销商、零售商、直到最终用户连成一个整体的基于功能的网链结构模式。

供应链管理(Supply Chain Management)是一种集成的管理思想和方法,它执行供应链中从供应商到最终用户的物流计划和控制等职能。早期人们把供应链管理的重点放在管理库存上,作为平衡有限的生产能力和适应用户需求变化的缓冲手段,它通过各种协调手段,寻求把产品迅速、可靠地送到用户手里所需要的费用与生产、库存管理费用之间的平衡点,从而确定最佳的库存投资额。因此其主要的工作任务是管理库存和运输。现在的供应链管理则把供应链上的各个企业作为一个不可分割的整体,使供应链上各企业分担的采购、生产、分销和销售的职能成为一个协调发展的有机体。

供应链管理以同步化、集成化生产计划为指导,以各种技术为支持,尤其以计算机网络为依托,围绕供应、生产作业、物流(主要指制造过程)、满足需求来实施。供应链管理主要包括计划、合作、控制从供应商到用户的物料(零部件和成品等)和信息。供应链管理的目标在于提高用户服务水平 and 降低总的交易成本,并且寻求两个目标之间的平衡。

供应链管理的背后动机是增加渠道的竞争力。其包含两个基本的信念:合作行为将减少风险,提高整个物流过程的效率;排除浪费和重复努力。美国营销协会(AMA)把渠道定义为:“公司内部的组织单位和公司外部的代理与经销商、批发商与零售商的结构。通过这种结构,进行商品、产品或服务的交易”。优越的渠道结构能够带来竞争优势。在传统的渠道中配置的大量存货构成了极大的风险。分享信息和共同计划可以排除或减少与存货相关的风险。

## 2.4 客户关系管理系统(CRM)

客户关系管理(CRM, Customer Relationship Management)由 Gartner Group 提出,其定义是企业与顾客之间建立的管理双方接触活动的信息系统。在网络时代的客户关系管理应该是利用现代信息技术手段,在企业与顾客之间建立一种数字的、实时的、互动的交流管理系统。它既不是一种纯粹的概念,也不单纯是一种技术。CRM 是一套体系,在这套体系中,人是主导因素,技术是积极有效的辅助手段。一般说来,CRM 的主要内容包括三个方面:营销自动化(MA)、销售过程自动化(SFA)和客户服务(Customer Service)。这三个方面是影响商业流通的重要因素,对 CRM 项目的成功起着至关重要的作用。

## 2.5 数据仓库和各业务系统在商务智能体系中的作用

除上面提到的 ERP、CRM、SCM 之外,企业中还存在着其它业务系统,如 OLTP 应用和其它外部业务处理系统等。除支持各自的业务处理过程外,这些系统还为企业的数据仓库提供了丰富的数据资源,而且通过各业务系统的数据库或数据仓库可以构建综合分析应用,对企业的各方面的绩效和整体表现进行评估。

## 3 基于数据仓库的商务智能框架

在商务智能体系中,我们强调数据仓库的作用,由于以下

原因:

- a) 数据仓库为分析应用提供了可以通过统一方式访问的数据来源;
- b) 数据仓库是面向主题的,通过建造数据仓库,可以消除原有数据源之间可能存在的数据冗余和不一致性;
- c) 通过定义企业商务问题的元数据构建的数据仓库,可以提供在一个统一的语义环境下进行数据分析和其它业务系

统的开发和集成。

然而,建立数据仓库对企业来说只是拥有了构建商务智能应用的基础,从业务系统或外部数据源产生的大量原始数据,经过抽取、转换、装载进入数据仓库,具有了全局一致的语义信息。这些信息还必须进行知识发现,以获得可以应用于决策的商务知识。基于这种认识,我们在图1中给出了商务智能的框架结构。

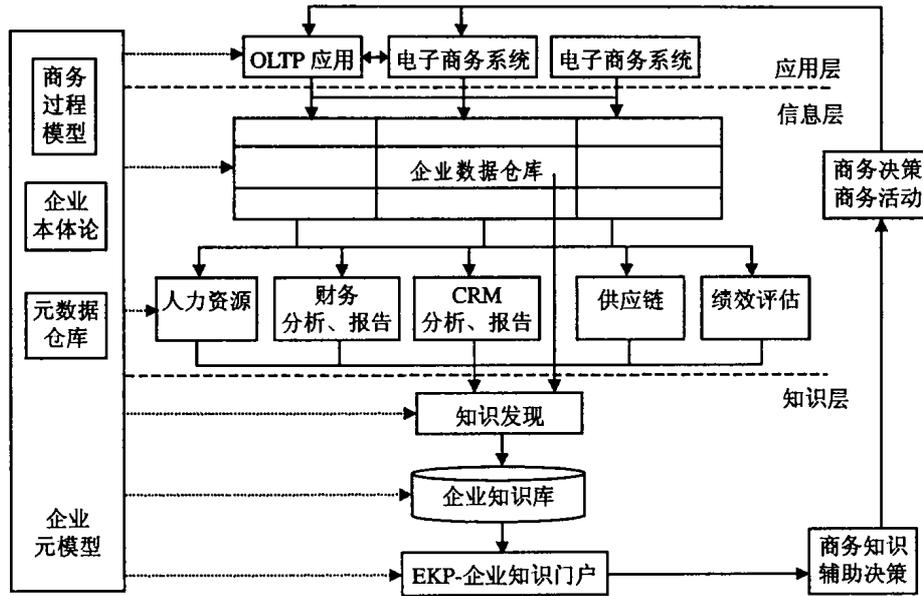


图1 商务智能框架

如图1所示,我们将企业的商务智能体系划分为三个层次,分别为数据层、信息层和知识层,在这三个层次中,数据层完成业务数据的收集和处理功能,信息层完成数据的抽取、转换、装载和语义一致化功能,而知识层的主要作用是知识发现,管理和表示。

企业元模型由三部分组成,分别为企业本体论、商务过程模型、元数据仓库。

企业本体论:为企业的基本概念,如过程、实体、约束关系、状态、角色等提供一个显式的、共享的、一致的定义(元本体论),以避免语义的歧义。企业本体论中的元本体论应该尽可能地与领域无关,以保证这些元本体论可以被用来方便地构建各类不同本体论术语。企业本体论作为构建稳定的、可重用的企业模型基础是非常重要的。为了实现企业模型的稳定性,进一步实现其柔性,企业模型应反映企业中最重要要素以及这些要素的约束关系,以支持不断变化的企业需求。由于稳定的企业模型可以在整个组织内共享,因而企业本体论就成为支持基于企业模型开发的应用系统通信的媒介,而对于已开发的系统则需通过向企业本体论的转换而与其它应用系统进行通信,这样企业本体论作为一种交换格式,支持整个企业集成系统的通信,也实现了整个企业各集成系统的互操作性。

商务过程模型:商务过程模型是用来对企业基本过程进行可视化,并使用业务功能对相关的业务进行控制。即表示了基本的商务过程及其控制功能。为了完成这些,每个商务过程模型中都至少包括几种不同的元素:基本流程(如物流、信息流、资金流等基本过程);基本活动(基本流程上的基本活动代表了如原材料的入库、最终产品装配及最终产品的发送等);业务功能;触发器等。

元数据仓库:元数据仓库存储企业信息系统中各模块的

元数据及其与全局一致的企业本体论中本体之间的对应关系。通过建立这种映射关系,使得各子系统之间的信息集成成为可能。

商务智能系统的应用层由企业原有的 OLAP 应用、电子商务系统和外部信息源等组成。信息层由数据仓库和基于数据仓库的各种分析、报告工具组成,商务智能系统构建的重点是知识层,其中包括:知识发现部分,企业知识库的管理与维护,企业知识门户以及商务知识和辅助决策信息的生成。下面我们对知识层的各个部分的特点和功能作出说明。

知识发现:商务智能中的知识发现意味着在一些事实或观察数据的集合中寻找模式的决策支持过程。所能发现的知识有以下几种:广义型知识,反映同类事物共同性质的知识;特征型知识,反映事物各方面的特征知识;差异型知识,反映不同事物之间属性差别的知识;关联型知识,反映事物之间关联或依赖的知识;预测型知识,根据历史和当前的数据推测未来数据;偏离型知识,揭示事物偏离常规的异常现象。所有这些知识都可以在不同的概念层次上被发现,随着概念树的提升,从微观到宏观,以满足不同用户、不同层次决策的需要。知识发现的方法包括分类、聚类、模式识别、决策树、关联规则、遗传算法、不确定性处理、可视化等。

企业知识库:知识库是数据库在知识处理领域的拓广和延伸。知识库在知识处理中也是一个至关重要和必不可少的概念。在知识库中可存放各种数据,组织、管理和维护数据库的方法对于知识库来说可以继续使用,至少可供参考借鉴。但是,知识库的主要任务还是存储大量的知识,因此,可将知识库定义为经过分类组织的“知识的一个集合”。

企业知识门户(EKP):企业知识门户作为知识库的界面,提供给企业工作人员(知识工人)和相关企业,如供应商,经销商等,管理,查询和维护知识的工具。企业知识门户应该具有

访问的便利性、可定制、和提供多种展现方式等特性。

商务知识和辅助决策生成:商务知识可以由知识发现模块自动地从数据仓库中提取,或者由知识工人通过企业知识门户从企业知识库中识别出来。通过数据挖掘获得的规则通常必须经过后处理,以提炼出在商务意义上的有趣模式,而不是仅仅具有统计显著性的模式。商务问题的辅助决策过程,往往是企业针对某些商务目标寻找优化方案的过程,通常可以看作一个多目标决策过程,使用企业元模型中对于商务问题的描述,以及综合数据仓库中的相关信息,可以对优化方案进行搜索,并将结果反馈给知识工人,作为决策的参考。

#### 4 商务智能系统的生命周期和构造方法

##### 4.1 商务智能系统的生命周期

基于上面给出的企业商务智能框架,我们提出的商务智能系统生命周期如图2所示。

商务智能系统的生命周期图由右下方开始,沿顺时针方向分为四个象限,在第一象限,是处理企业各种具体业务的企业信息系统,并为数据仓库提供数据来源;在第二象限中,使用知识发现工具对数据仓库中的数据进行挖掘和智能数据分析,得到在商务语义下有趣的知识;第三象限侧重于对发现知识的管理和进一步的分析;在第四象限,关键的商务知识被利用,为企业的商务活动和过程优化提供决策支持。整个商务智能系统以数据仓库为核心,各象限中的功能模块围绕数据仓库进行构建和集成,在现有制造业企业中,在操作知识的处理方面已经开发了大量的系统,而在商务数据、挖掘结果的处理、决策知识的生成方面也缺少统一的方法论指导和适用的工具,而这正是商务智能研究力图解决的问题。

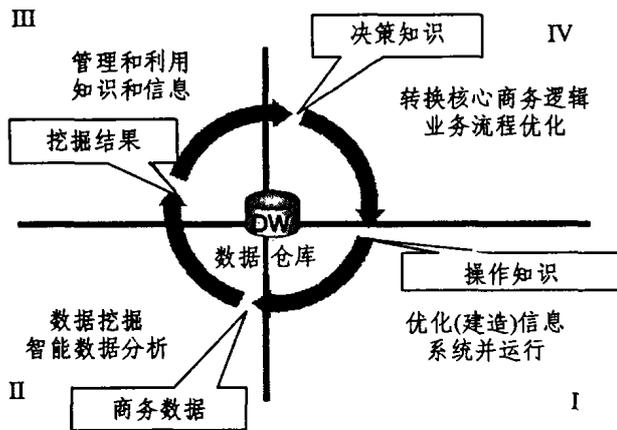


图2 商务智能系统的生命周期

生命周期中的四个阶段构成一个循环,说明商务智能系统的构建和应用是一个迭代的过程,商务智能系统帮助企业从现有的业务中发现问题,寻找可能的优化方案,用以改进商务过程,然后对业务系统作出相应的调整,并进入下一阶段的循环。

##### 4.2 集成化商务智能系统的构造方法

在商务智能系统中,一般要求所有的数据和数据展现方式都是语义一致的。但是由于参与构成商务智能系统的各个信息基础设施通常处于异构分布的环境中,因而可能导致各种各样的异构性问题,即发生各种冲突问题。我们主要考虑业务数据的集成,由于不同的业务系统是独立设计和开发的,在参与集成的业务系统之间会产生各种语义冲突,这种语义的冲突包括:①命名不一致:不同的业务对象在不同的业务系统

中的名称不同,于是引发了同形异义与同义异形的问题;②格式不一致:包括数据的类型、域、范围、精确性以及数据项的合成等方面的不同;③结构不一致:取决于在一个业务系统中的如何使用,一个业务对象在不同的数据库中可能会有不同的结构。

不同业务系统的集成通过两级映射实现:第一级映射是从统一的企业本体库到商务过程模型之间的映射,第二级映射是从商务过程模型到元数据仓库的映射。两级映射的过程不仅体现在数据仓库的 ETL 过程,而且应用在整个商务智能框架中,实现语义的一致性。

商务智能系统的前期构造可分为两种基本途径:集成到数据仓库和从数据仓库开始的构造方法。前者是从现存的各个业务系统中通过知识工程的方法提取业务本体,构造企业本体库,然后进行元数据仓库的设计;后者是对于数据仓库构建以后,对于新加入商务智能系统的业务模块,进行业务建模和本体提取,在必要时对现有的商务过程模型库和企业本体库进行扩充或删改,定义二级映射模式,将新的业务模块集成到系统中。

在数据仓库构造之后,就可以进行商务智能应用的开发,如针对人力资源、财务、供应链、绩效评估等方面的分析应用。以使企业各部门可以在他们熟悉的语义环境中进行分析和作出决策。可以进一步应用知识发现技术,将发现的知识加入企业知识库,并对企业知识库进行动态的维护,企业知识库中的知识可以通过企业知识门户(EKP, Enterprise Knowledge Portal)发布给企业内部、外部的相关人员和合作伙伴等。EKP 可以提供决策者比 OLAP 更高层次的决策支持。对涉及企业经营过程重组的决策,通常会引起企业信息基础设施的调整,这些调整过的业务模块再次被集成到商务智能系统中,形成商务智能系统的闭环结构。

结论 商务智能系统可以有效地增强企业的敏捷度和竞争力。通过集成化的商务智能系统的构建,企业可以有效地利用现有的信息基础设施中的数据,并且在一个一致的语义下对数据进行访问和操纵。通过应用层、信息层和知识层的划分,知识发现方法的应用可以从大量信息中提取企业决策所需的相关知识,并且在企业知识库中进行维护,通过企业知识门户发布,更好地支持企业的有效决策。闭环的商务智能框架和商务智能生命周期可以保证企业商务智能系统和商务过程的持续优化。

#### 参考文献

- 1 Anderson M, Whitten D, Lett B. OIS Scenario: The Emerging Electronic Workplace. Gartner Group, R-OIS-126, 1996
- 2 Brethenoux E, et al. Data Warehouse, Data Mining and Business Intelligence: The Hype Stops Here. Gartner Group, R-300-105, 1996
- 3 Dresner H. Business Intelligence Imperative. Gartner Group, AV-13-8785, 2001
- 4 Porter, Michael E. Competitive Advantage. The Free Press, New York, 1985. 11~15
- 5 Christensen C M. The past and future of competitive advantage. MIT Sloan Management Review, Winter, 2001. 105~109
- 6 Chaffee E E. Three models of strategy. Academy of Management Review, 1985, 10 (1): 89~98
- 7 Inmon W H (美)著,王志海等译. 数据仓库 (Building the Data Warehouse). 机械工业出版社, 2000. 3~25
- 8 高洪深. 决策支持系统(DSS)——理论·方法·案例(第二版). 清华大学出版社, 广西科学技术出版社, 2000. 124~146