

数据仓库在我国企业的应用现状及实施策略分析

On the Application of Data Warehouses in Chinese Enterprises

何朝红

HE Chao-hong

(广西物资集团总公司, 广西南宁 530022)

(Guangxi Materials Group Corp., Nanning, Guangxi, 530022, China)

摘要:分析数据仓库在我国企业中的应用现状。认为目前我国企业数据仓库的应用主要存在对数据仓库的概念理解不深,对项目的长期性、艰巨性认识不足,数据库系统中数据的积累不够,用户没有参与数据仓库建设的需求分析,项目实施过程中的管理混乱,具体实现中的技术问题尚未完善等问题。企业应从确定建立数据仓库的合适时机,选择切合实际的实施方法,取得最高管理层的支持和认可,确定基本目标、量化预期收益,选择正确的设计思路,选择合适的合作伙伴等6个方面应用数据仓库技术。

关键词:数据仓库 应用 问题 策略

中图分类号:TP311.13 文献标识码:A 文章编号:1002-7378(2006)04-0375-03

Abstract: The application of data warehouses in Chinese enterprises is discussed. The problems in the application are revealed as follows. The data warehouse can not be fully understood. The long term and hardness of the project running are neglected. There are less data accumulated in the data warehouse. There is lack of users participating in the demand analysis of the data warehouse establishment, and poor management in the implementation of the project. The technical problems have not been improved in the operation of the data warehouse. The relevant suggestions are presented. For the project to be supported, it needs to choose a correct time and a pragmatic execution way to construct a project, set up an essential goal, measure the expecting benefits, and select a suitable partner.

Key words: data warehouse, application, problems, strategy

近年来,计算机网络和数据库技术的迅速发展和广泛应用,使得企业管理进入了一个崭新的时代,企业的许多业务得到了联机事务处理(OLTP)信息系统的支持。然而,随着市场竞争的加剧,正确及时的决策成为企业生存和发展的重要环节。特别是随着数据库系统的逐日运行,产生了大量的业务数据,如何安全有效地存储和管理这些数据,并从大量繁杂的数据中获取其中有用的决策信息,为高层管理人员提供快速、准确和方便的决策支持,成为目前企业提高管理水平和竞争优势必须解决的问题。数据仓库(Data Warehouse)技术的产生和发展,为这个问题的解决提供了有效的理论和方法指导。

数据仓库是20世纪90年代初出现的数据管理

的新技术,到20世纪90年代中期已在发达国家的许多行业中得到应用^[1]。建立数据仓库的目的,就是把来自企业内部和外部的大量异构数据按辅助决策主题的要求进行加工、集成,为高层管理人员提供各种类型的、有效的数据分析,起到决策支持的作用。因此,数据仓库已经成为继Internet之后的又一技术热点。

1 数据仓库在我国企业的应用现状

随着各种计算机技术,如数据模型、数据库技术和应用开发技术的不断发展,数据仓库技术也在不断完善,并在实际应用中发挥了越来越大的作用。据美国国际数据公司(IDC)调查,使用数据仓库技术的投资回报率(ROI)平均超过400%,平均回报时间为2~3年;若从部门级数据仓库(也称“数据集市”)开始实施,则投资回报率平均超过500%^[2]。

目前,在美国,30%到40%的公司已经或正在

建造数据仓库^[3]。在经济发达国家,大型企业几乎都建立了自己的数据仓库。数据仓库的应用已遍及银行、证券、保险、税务、电信、医疗保健、商业零售业等各个行业和部门。这几年,我国越来越多的企业也逐步认识到利用数据仓库技术获得决策支持的重要性,许多大型企业开始建立数据仓库系统。例如,中国银行广州分行“八五”期间就开始建设数据仓库,包括从数据采集、数据处理到数据存储的全过程;1999年4月,深圳国信证券的数据仓库系统一期工程完成;上海集装箱码头项目、深圳华为、上海强生、上海日立、广东省长话计费系统、新飞集团、浙江省移动通讯计费系统等都已成功地采用了数据仓库技术;上海宝钢、深圳招商银行目前也在创建自己的数据仓库,开展数据分析工作^[4]。

2 数据仓库在我国企业的应用中存在的问题

应该指出的是,虽然近几年数据仓库在我国的应用已开始形成热潮,但与发达国家相比还存在着很大的差距。由于我国的生产力水平普遍较低,企业信息化基础建设起步较晚,用于信息化建设的投资力度也较低,因此,要使数据管理上升到数据仓库水平会遇到一些困难。虽然很多企业都在实施数据仓库项目,但成功的却为数不多,问题主要集中在以下6个方面。

2.1 对数据仓库的概念理解不深

数据仓库中的数据不是简单的数据拷贝,而是按决策主题的要求从 OLTP 数据库或其他外部数据源中经过复杂的提取、清洗、转换、汇总而得来的;而且,数据仓库中存放的都是历史数据,一旦装入,一般是不会修改的。因此,不能对仓库中的数据进行实时更新,只能周期性地刷新,也就是追加新的数据。正是由于对数据仓库的定义理解不深,一些企业在开发中采用复制服务器直接从 OLTP 系统向数据仓库传播数据,或者允许数据进行实时更新,从而对数据仓库造成灾难性的后果。

2.2 对项目的长期性、艰巨性认识不足

有些用户以为今天建立了数据仓库系统,明天就能够解决商业竞争中的所有问题,能取得很大的效益。这种急功近利的思想没有充分认识到项目建设的长期性和艰巨性也会影响企业有效应用数据仓库技术。实际上,数据仓库不一定马上就能得到很大的效益,有时候数据仓库中挖掘出来的东西并不一定是很有用的。数据仓库系统的建立是一个不断完

善的过程。

2.3 数据库系统中数据的积累不够

数据仓库中的数据主要是从 OLTP 数据库中抽取的,而只有在大量数据的基础上进行联机分析处理才可能得到有价值的决策信息。但由于国情所限,目前我国企业在数据库的规模上,与国外企业相比还存在很大的差距;更有甚者,有些企业的数据库由于没有资金维护而变成了“死库”。建设数据仓库系统,没有数据积累是没有任何意义的。

2.4 用户没有参与数据仓库建设的需求分析

数据仓库的最终用户是企业的中、高层领导或其助手,这与 OLTP 系统直接面对具体的操作人员有明显的不同。对许多企业来说,要使最终用户与信息技术人员真正在一起合作是较困难的。而且在项目的初始阶段,领导们也往往提不出具体的目标,只是希望信息技术人员对出现的问题能提供快速的解决方案。这就必然导致最后开发出来的系统难以满足用户的决策要求。

2.5 项目实施过程中的管理混乱

建立数据仓库的目的是大力提高经济效益。然而,执行有效的数据仓库策略的最大障碍常常不是技术方面的。由于集中式数据仓库具有可控性和可靠性,但灵活性不够好,而分布式数据集市具有很大的灵活性但可能会导致各部门之间不能有效地协同工作。因此,在确定数据仓库的功能规模问题上,经常暴露出存在于机构内部关系的复杂性和行政上的种种障碍,从而增加了技术实现上的难度。

2.6 具体实现中的技术问题尚未完善

数据仓库中的数据可能来自不同的数据源,它们分布在不同的硬件、数据库、网络环境中,为不同的业务部门服务。因此,对这些数据的整合可能会遇到很多技术困难,而如果不能保证数据仓库中数据的高质量,也就很难获得有价值的决策信息。另外,数据分析工具的选择问题,目前仍没有一种大众化的、简单易用的集成工具。

3 我国企业应用数据仓库技术应采取的策略

随着现代社会商业模式变革的进一步普及和深入,数据仓库将成为企业获得竞争优势的关键技术。针对应用中存在的上述问题,我们认为,企业要有效应用数据仓库技术,应采用以下策略。

3.1 确定建立数据仓库的合适时机

每个企业应根据自己的实际情况,决定要不要

建立数据仓库以及何时应用等问题。对于如金融、邮电等行业的企业而言,以前本就有大型数据库的坚实基础,现在利用数据仓库为其开拓更有竞争力的市场就很有必要。而对于国内基础设施尚待完善的中小企业来说,立即上马数据仓库未必是一项明智的选择,因为没有几十万,乃至上百万的资金,是没有办法建立一套数据仓库系统的。而且,如果企业规模比较小,所要处理的信息量比较少,普通的 MIS 系统所提供的功能管理已很满意,领导特殊的需求可通过其他手段实现,当然也就没有必要实施或可以暂缓实施数据仓库,而应将有限的资金用在数据库系统的维护上,为今后建立数据仓库奠定基础。

3.2 选择切合实际的实施方法

数据仓库的建立有两种方法:一是先建中央数据仓库,再根据需求建立相应的数据集市。这种方法可控制好,便于各部门协调工作,但建设周期长,投资较大,有时数据仓库建好后,企业数据模型已发生变化,所以一般要建立良好的迭代过程来保障系统的增长与完善。二是自下而上先建数据集市,再建中央数据仓库。这种方法建设周期短,费用较低,但建设时必须要有全局观点,要保持数据的一致性,便于今后其他数据集市和中央数据仓库的实现。

一般来说,在整个企业对数据仓库建设缺乏成功经验的情况下,从个别数据比较规整且有一定规模的业务子系统入手,先建立扩展性好、部门级应用的数据集市,在成功使用的基础上,再在多个数据集市上构筑面向企业应用的集中式数据仓库,也许是最为现实的方案,而且收效快,风险小。

3.3 取得最高管理层的支持和认可

对一个企业来说,建设数据仓库是一项投资大、周期长、见效慢的系统工程。如果得不到高层领导在行政上、尤其是财政上的支持,项目的实施就无法进行。而要获得领导的支持,一方面必须在项目的开始阶段就使他们参与进来,了解其对系统功能的需求,因为他们是企业的决策者,是系统的最终用户;另一方面,要让他们尽早了解项目实施的艰巨性和长期性,使其明白数据仓库不是万能的,未来的系统也不可能是十全十美的。只有决策者在一开始就对可能出现的困难有充分的思想准备,才有可能对项目有自始至终的大力支持。此外,在数据所有权和数据存放等问题上的内部纷争,很容易给数据仓库带来进程上的滞延和失败,最高管理层的支持也有助于打破各个部门之间日益严重的数据保护壁垒,为项目的顺利实施扫清障碍。

3.4 确定基本目标,量化预期收益

在项目开始实施以前,用户必须明确建立数据仓库的目的、要解决的问题及应达到的目标;同时,也应该认清哪些是关键性的、影响成功的因素,以便在方案的实施过程中进行跟踪。这些基本目标确立以后,下一个基本要求是对来自数据仓库的预期收益进行量化。量化的目标不一定是数字或金融表达式,而只需要具体明确、有意义即可。目前,许多企业都采用金融衡量标准,比如投资回报率,来对收益进行量化。其他类型的收益衡量标准还包括成本节约程度以及可获得的能够进行衡量的效率等。只有对项目定义目标和量化收益之后,管理层才能有据可依地判断一个数据仓库的成功与否。

3.5 选择正确的设计思路

由于 OLTP 系统为功能驱动,而数据仓库系统由数据驱动,因而反映在项目的设计阶段(概要设计和详细设计),前者一般开始于过程分析,后者则从数据分析入手。数据分析又有两个方向,一是从数据提炼需求;另一是从需求寻找数据。所谓从数据提炼需求,就是通过考察企业已有业务系统所提供的数据,分析从这些数据中能提炼出哪些有用的决策信息。这种做法适合于项目开始的原型设计阶段。而从需求寻找数据,则是针对用户的某种应用需求,考察其所需的信息能否从业务系统中抽取以及怎样抽取。在项目的进行过程中,也就是在用户根据原型完善系统的时候,这种做法使用较普遍。

3.6 选择合适的合作伙伴

在建立数据仓库时,选择合适的合作伙伴非常关键。这些合作伙伴,包括顾问、分销商以及软件和硬件的厂家,对他们的选择一是基于其能力,二是基于其是否具有长期同舟共济的愿望。要充分考虑这些合作伙伴在初始目标实现以后是否还会积极主动地帮助自己解决面临的问题,随着项目的继续推进会是什么情况,等等。企业只有与所有的合作伙伴进行充分有效的协作和交流,才能取得数据仓库技术应用的最后成功。

参考文献:

- [1] 刘东波. 数据仓库技术的现状与未来[J]. 微型机与应用, 2000(7): 4-7.
- [2] 张凤荔, 葛晓峰, 周明天. 数据仓库研究及其在现代企业中的应用[J]. 计算机科学, 1999, 26(6): 71-74.
- [3] INMON, WILLIAM H. Building the data warehouse [M]. 2nd ed. New York: Wiley, 1996.
- [4] 张伟民, 董萌, 章韵, 等. 电信网络管理的数据仓库解决方案[J]. 计算机应用研究, 2000(7): 36-38.

(责任编辑: 韦廷宗)