

企业数据仓库的体系结构和建设

On the System Structure and Construction of Enterprise Data Warehouse

叶 嘉

Ye Jia

(广西广播电视大学,广西南宁 530022)

(Guangxi Broadcast and Television Univ., Nanning, Guangxi, 530022, China)

摘要:在分析数据仓库与传统数据库系统的区别的基础上,阐述企业数据仓库的体系结构,提出企业数据仓库系统的建设阶段,为企业信息化的进一步发展奠定基础。

关键词:数据仓库 数据库 企业 体系结构 建设

中图法分类号:TP311.13 文献标识码:A 文章编号:1002-7378(2005)04-0306-03

Abstract: On the basis of analyzing the difference between the Data Warehouse and the traditional database system, the system structure of Enterprise Data Warehouse is elaborated. The construction of Enterprise Data Warehouse is the foundation of informational enterprise's further development.

Key words: data warehouse, database, enterprise, system structure, construction

信息作为现代企业的资源,占据着越来越重要的地位,已经成为现代企业科学管理的基础和正确决策的前提,以及有效调控的手段。能否拥有及时、准确、全面的信息已经成为衡量一个企业是否具有发展潜力的重要指标^[1]。

面对纷繁复杂的市场竞争,众多企业立足于多年积累的数据和自身核心业务,大量的日常业务数据需要管理。为此,本文在分析数据仓库与传统数据库系统的区别的基础上,阐述企业数据仓库的体系结构和企业数据仓库系统的建设阶段,为企业信息化的进一步发展奠定基础。

1 数据仓库系统与传统数据库系统的区别

传统的数据库系统面向以事务处理为主的OLTP应用,它主要用于日常事务处理工作,存放在数据库中的数据大体是操作型数据。在事务处理环境中,用户的行为特点是数据的存取操作频率高而每次操作处理的时间短,因此,传统数据库系统可以允许多个用户按分时方式使用系统资源,同时保持

较短的响应时间^[2]。

分析处理和事务处理具有极不相同的性质,在分析处理环境中,为适应数据分析处理,数据库中存放的数据要求是分析型的数据,某个DSS应用程序也可能需要连续运行几个小时,从而消耗大量的系统资源。如果在事务处理型应用环境中直接构建分析处理型应用明显是不可行的,必须把分析决策型数据从事务处理系统应用环境中分离出来,建立单独的分析决策型应用环境。数据仓库正是为了构建这种新的分析决策型应用环境而出现的一种数据存储和组织技术,并成为研究数据管理技术的新课题。

数据仓库是面向主题的、集成的、稳定的、随着时间变化的数据集,用以支持经营管理中的决策制定过程^[3]。数据仓库的数据是面向主题的、集成的、不可更新的,并且是随时间不断变化的。

将数据库系统和数据仓库系统结构的各个组成部分作一个简单的比较(表1)可以看出,数据仓库管理系统比数据库管理系统简单很多。数据库管理系统中许多技术难点,如完整性保护、并发控制等等,在数据仓库的管理中几乎可以省去。但是由于数据仓库的查询数据量很大,所以就对数据查询提出了更高的要求,它要求采用各种复杂的索引技术;同时由于数据仓库面向的是商业企业的高层管理者,他们会对数据查询的界面友好性和数据表示提出更高

收稿日期:2005-05-10

修回日期:2005-09-01

作者简介:叶 嘉(1965-),女,广东南海人,讲师,主要从事数据库、管理信息系统的研究。

的要求。

表1 数据库系统与数据仓库系统的比较

项目	数据库系统	数据仓库系统
类型	操作型数据,增、删、改操作频繁	分析型数据,极少有更新操作
核心	功能强大,面向OLTP应用	因极少有更新操作,故功能简单
工具	以查询工具为主	以分析工具为主

2 企业数据仓库的体系结构

一个典型的企业数据仓库系统通常包含数据源、数据存储与管理、OLAP服务器以及前端工具与应用四个部分。图1为一个企业数据仓库系统的体系结构。

数据源是数据仓库系统的基础,是整个系统的数据源泉。通常包括企业内部信息和外部信息。内部信息包括存放于企业操作型数据库中(通常存放在RDBMS中)的各种业务数据和办公自动化(OA)系统包含的各类文档数据。外部信息包括各类法律法规、市场信息、竞争对手的信息以及各类外部统计数据及各类文档等。

数据的存储与管理是整个数据仓库系统的核心。在现有各业务系统的基础上,对数据进行抽取、清理,并有效集成,按照主题进行重新组织,最终确定数据仓库的物理存储结构,同时组织存储数据仓库元数据(具体包括数据仓库的数据字典、记录系统定义、数据转换规则、数据加载频率以及业务规则等信息)。按照数据的覆盖范围,数据仓库存储可以分为企业级数据仓库和部门级数据仓库(通常称为“数据集市”)。数据仓库的管理包括数据的安全、归档、

备份、维护、恢复等工作。这些功能与目前的DBMS基本一致。

OLAP服务器是对分析需要的数据按照多维数据模型进行再次重组,以支持用户多角度、多层次的分析,发现数据趋势^[4]。

前端工具主要包括各种数据分析工具、报表工具、查询工具、数据挖掘工具以及各种基于数据仓库或数据集市开发的应用。其中数据分析工具主要针对OLAP服务器,报表工具、数据挖掘工具既针对数据仓库,同时也针对OLAP服务器。

3 企业数据仓库系统的建设

把信息加以整理、归纳和重组,并及时提供给相应的管理决策人员,是数据仓库的根本任务。因此,从产业界的角度看,数据仓库建设是一个工程,也是一个过程。

企业数据仓库的建设通常按照快速原型法予以实施,主要包括:确定范围、环境评估、分析、设计、开发、测试和运行等几个阶段。同时企业数据仓库又是一个在原型的基础上进行不断迭代的过程。

3.1 确定范围

确定范围的主要任务包括了解方向性分析处理需求,确定信息需求,确定数据覆盖范围。方向性需求包括:决策类型、决策者感兴趣的问题(或对象)等。在确定范围时应该重视用户驱动和数据驱动相结合,同时可以借鉴国内外已有的成功经验。

3.2 环境评估

环境评估是对企业数据仓库系统建设的硬件环境和软件环境进行选型和准备。

在硬件平台选择中需要选择与数据仓库系统规模相适应的核心服务器,同时建议将数据仓库系统平台与业务处理平台分离开。

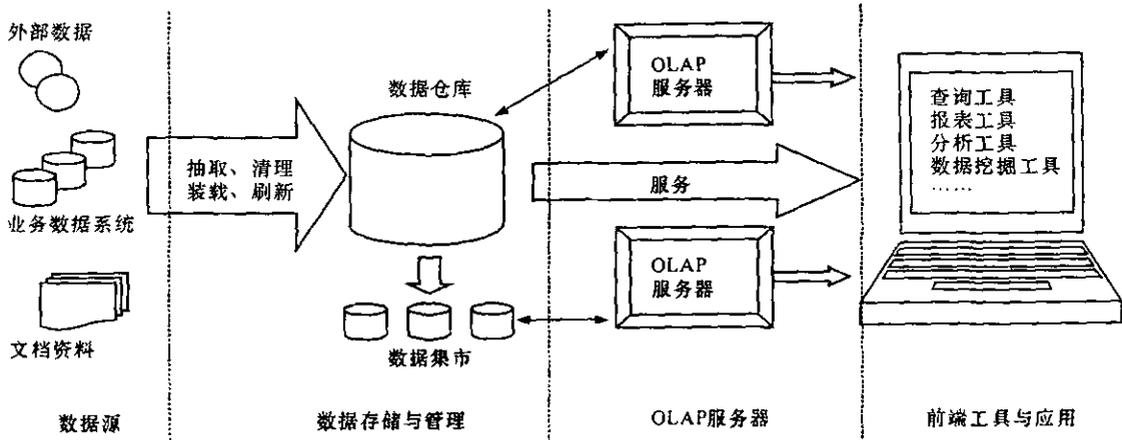


图1 企业数据仓库的体系结构

软件平台的选择主要包括数据仓库引擎、OLAP引擎和前端分析展现工具的选择。产品进行测试是软件选型的一种有效方法,各个企业可以根据自身的数据状况对各类产品进行测试。

3.3 分析

分析阶段主要包括深入了解数据源和分析数据仓库系统所包含的主题域及其相互之间的关系。分析阶段必须坚持用户参与,并且与原有系统开发或维护人员进行深入的沟通。

3.4 设计

数据仓库设计的主要任务包括与操作型系统接口的设计和数据仓库本身的设计两个部分的内容。其中与操作型系统接口的设计主要是指数据抽取、清理、转换和刷新策略的设计。从多个不同的数据源中抽取数据,需要解决数据的不一致性,以保证数据的质量。数据的不一致性主要包含模式冲突和语义冲突。从操作型数据库模型到数据仓库模型的转变需要大量细致的工作,例如:

消除纯粹是操作型的数据;将包含在多个表中的有关数据进行合理合并;适当增加部分导出数据;在码值中增加时间关键字;按照合适的数据粒度进行综合。

数据仓库本身的设计包括数据仓库逻辑数据模型的设计、数据仓库物理数据模型的设计。由于目前数据仓库产品尚未形成一套统一的标准,因此在数据仓库设计阶段必须有数据仓库专家和数据仓库系统产品提供商的共同参与。

3.5 开发

开发阶段所要完成的主要内容包括数据仓库建模、数据抽取和加载模块、数据访问模块以及开发实际应用模块。这阶段应该有行业专家和数据仓库专家的参与。

3.6 测试

测试是保证系统可靠性的重要手段。数据仓库测试与一般软件系统测试不同的是数据仓库的测试

不仅包括对软件系统的测试,同时包括对数据的测试。在测试阶段必须保证测试的充分性,同时注意测试数据的覆盖范围。

3.7 运行

系统运行主要包括用户培训、数据加载、数据访问及应用等。在数据仓库系统的运行过程中,不断收集用户新的需求,不断进行完善。

数据仓库系统的建设是一个不断完善、健全的过程。这个过程是随着业务量、业务范围和客户的不断发展而发展的,同时随着业务的发展,数据仓库的价值也将随之得到增长。

4 结束语

总的来讲,数据仓库是一种面向数据应用的数据管理技术,它提供了集成化的、历史化的数据管理功能,支持综合性的数据分析,特别是战略分析。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进步,数据仓库技术也不断发展,并在金融业、制造业、商贸业和社会服务等方面都有着广泛的应用前景,并在实际应用中发挥了巨大的作用。数据仓库所产生的巨大效益同时又刺激了对数据仓库技术的需求,数据仓库市场正以迅猛势头向前发展。

参考文献:

- [1] Sperley. 企业数据仓库规划建立与实现[M]. 北京:人民邮电出版社,2000.
- [2] 武 森,高学东,巴斯蒂安. 数据仓库与数据挖掘[M]. 北京:冶金工业出版社,2003.
- [3] Inmon W H. 构建数据仓库[M]. 王志海,林有芳译. 北京:机械工业出版社,2003.
- [4] 王 珊. 数据仓库技术与联机分析处理[M]. 北京:科学出版社,1998.

(责任编辑:邓大玉 韦廷宗)

(上接第300页)

的为计算机网络教学服务,使教师和学生减少学习使用NS的时间,把精力集中在课程内容教学和学习上。

参考文献:

- [1] 徐雷鸣,庞 博,赵 耀. NS与网络模拟[M]. 北京:人

民邮电出版社,2003.

- [2] Kevin Fall,kannan varadhan. The NS Manual [EB/OL]. <http://www.isi.edu/nsnam/ns/ns-documentation.html>,2004.

(责任编辑:邓大玉)