基于元数据的数据仓库安全模型实现研究

马艳锋[1],谭立彦[2]

(1.中国航空结算有限责任公司,北京 100028;
2.北京首信科技股份有限公司,北京 100015)

摘 要:元数据是数据仓库的关键技术之一,也是数据仓库安全性的解决途径之一。对基于元数据的数据仓库安全模型进行了研究,对原模型进行了改进,并提出了在 Oracle9i 数据仓库平台上的解决方案。

关键词:数据仓库;元数据;安全模型

中图分类号:TP311.138 **文献标识码**:A **文章编号**:1674-7720(2012)09-0014-04

Research of implementation for data warehouse security models based on metadata

Ma Yanfeng[1], Tan Liyan[2]

(1.Accounting Centre of China Aviation System & Communication Div, Beijing 100028, China;
2.Beijing Capitek Corp, Ltd., Beijing 100015, China)

Abstract: Metadata is one of the key technologies in the data warehouse, and it is also an important way to solve the security problems. This paper gives a deep survey on the data warehouse security models based on metadata and extends the fuction of the original model, and gives a solution on Oracle9i.

Key words: data warehouse; metadata; security model

客运数据分析系统——SMARTRIX 数据仓库,拥有一系列深层数据分析产品,可以进行不同维度的分析和挖掘,为航空公司的决策支持提供量化依据。该系统目前的运维模式是运行支持人员使用专用的程序进行数据仓库数据的维护,客户通过统一的访问界面 BO 来访问报表数据,不同客户可以访问的报表是在 BO 中进行的设置,报表数据与数据库的连接是通过专用的 ID 来访问数据库数据的,这种架构目前可以满足安全需求。但随着业务的发展,数据仓库的数据源越来越多样化,在维护中需要更多的技术支持人员访问数据库,同时客户的系统支持人员也要求参与到数据仓库数据的管理中。这样,目前的数据仓库安全性就不能满足未来业务发展的需要,因此加强数据仓库的安全性研究变得非常迫切。由于在 SMARTRIX 数据仓库设计时的元数据建立得非常完备,而业界在通过元数据控制安全方面也有不少研究,为此本文主要针对数据仓库的安全性进行了研究,提出了一种基于元数据的数据仓库访问控制模型,并研究了它在 oracle 平台下的实现方式,为下一步加强数据仓库的安全性做些探索。

1 基于元数据的数据仓库安全模型及其研究

元数据是关于数据的数据,是描述数据仓库内数据的结构和建立方法的数据。元数据不仅描述了数据仓库所包含的内容,而且能够帮助用户识别数据仓库中的数据质量[1]。由于元数据贯穿数据仓库实施的整个过程,技术人员要想很好地开发数据仓库,业务人员要想很好地使用数据仓库必须透彻地理解元数据,从而元数据成为数据仓库访问的一个必经之路,没有元数据就无法访问数据仓库中的数据,访问控制环境可以通过元数据引导[2]。

参考文献[3]把元数据分为“结构元数据”和“访问元数据”。结构元数据主要用来创建和维护数据仓库,描述了数据仓库的结构和内容、数据主体、数据特征及其之间的相互关系。“访问元数据”描述了数据仓库和终端用户应用程序之间的动态关系[1],包括企业、用户的价值

衡量，描述了数据仓库服务器、数据库、表以及从数据源到数据仓库传送数据所执行的操作。“访问元数据”针对维度视图和表层次建立了“Drill-Down”和“Roll-Up”规则，这些数据也可以包含用户定义的访问规则与查询策略来限制用户对数据仓库数据的访问。

1.1 安全模型原理

在数据仓库原系统的基础上添加了两个部件：“安全管理者”和“信息查询管理层(SQML)”。通过“安全管理者”系统管理员定义用户组可访问的数据，可以指定用户组可以访问哪些事实表、维度等，添加新的用户时需要指定此用户属于哪个用户组，加入组的用户具有这个用户组中数据的访问权限。“信息查询管理层”的功能是在信息服务器上实现的，其主要的功能是从 Internet 上接收用户请求，并检查查询内容是否在用户的权限之内。通过添加的这两个组件，达到了间接控制数据仓库数据访问的目的。其原理如图 1 所示。

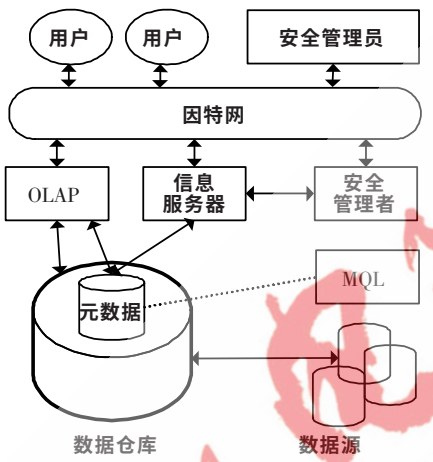


图 1 基于元数据的数据仓库安全模型原理图

1.2 存在问题

上述基于元数据的安全模型需要一种结构化的描述语言(MQL)的支持，使用此安全模型的数据仓库中的数据是用 MQL 语言以树形的结构表示出来的。首先，“安全管理者”定义一个新的用户组，存储下这个用户组的访问权限，用户组的访问规则以 MQL 语句存储下来。SQML 接收到用户的请求并把它转化为 MQL 查询，再与存储的用户组的访问权限对比，查看用户查询的内容是否在访问权限之内。由于需要 MQL 语言的支持，使得此模型具有一定的私有性，不易在通用的数据仓库中进行推广。

1.3 借鉴与推广

但是，借助“访问元数据”来控制数据仓库访问的思想是可以借鉴的。CWM“公共数据仓库元模型”在 2000 年由 OMG 提出，是专门为数据仓库元数据而制定的一套标准。该规范提供了一个描述数据源、数据目标、转化、分析、处理、操作等与建设和管理数据仓库相关信息的元数据基础框架，并为在多个厂商的产品之间进行元数据的通信和共享提供了一个切实可行的标准。MDC 组织的 OIM“开放信息模型”与 OMG 组织已经合并，今后所有的数据仓库工具都将遵循统一的 CWM 标准。CWM 的标准化使得数据仓库元数据变得易于访问。

鉴于上述原因，利用元数据标准 CWM 的标准化，本文从实践的角度提出一个基于元数据的数据仓库安全模型的实施方案。此模型是对原模型的扩展与推广，相对于原模型，所构造的模型具有结构可控、使用方便、易于在商业产品上实施等特点。

2 安全模型原型系统设计

2.1 系统框架

通过把系统中的元数据扩充，加入“访问元数据”，“访问元数据”中存储用户的访问规则。终端用户通过“访问元数据”进行数据仓库信息的访问，访问元数据中设定了用户的访问规则，如果访问控制系统中用户的权限满足用户查询的内容需求，则允许用户继续进行访问，否则拒绝访问。同时用户的访问规则可以通过访问元数据控制器进行调整，访问元数据控制器为管理员进行“访问元数据”的管理提供了友好的界面。系统框架如图 2 所示。

首先，由于数据仓库系统的用户对于数据仓库中存储不同的多维数据具有不同的访问权限，用户、立方体、立方体维度以及用户的访问权限这每一个要素都代表了一个控制角度，它们之间可以组成一个多维的访问控制模型，这个模型的数据就是“访问元数据”的主体。多维数据完全可以采用数据仓库技术把该数据模型展现出来，通过多维数据来建立用户的访问规则使其能够充分利用数据仓库组织多维数据的优势，强化和改善访问控制的过程。

该模型的核心部件是访问元数据引擎，主要具有以下功能：

(1)通过“访问元数据”可以查看用户要查询的内容是否在他的权限之内。

(2)能够自动从事实表和维表中获得新的控制信息。

(3)通过“访问元数据”控制平台可以方便地对“访问元数据”及其相关的事实表和维表进行管理。

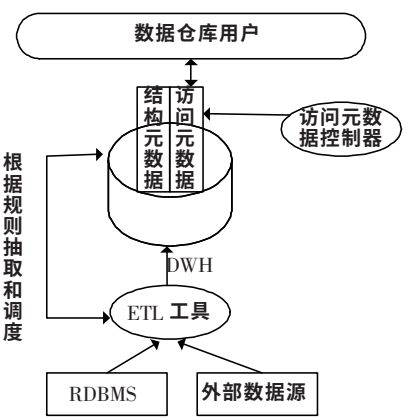


图 2 基于元数据的数据仓库安全模型

2.2 访问元数据设计

2.2.1 访问控制维的设计

为简便起见，在这里构造了4个表：用户表User_table、立方体表Cube_table、维度表Dimension_table和时间表Time_table。抽象为“访问元数据”的4个访问控制维度（用户、立方体、维度与时间），可以控制某个用户在什么时间内是否具有对某个立方体某些维度的访问权限。表的构造分别如表1~表4所示（表中数据仅做示例）。

表1 用户表

U_id	U_name	U_pass
1001	张华	123
1002	刘明	456
1003	王莉	789
…	…	…

表2 立方体表

Cube_id	Cube_attr1	Cube_attr2
2001	收益分析	…
2002	商务分析	
2003	促销分析	
…	…	…

表3 维度表

Cube_id	Dimension 1	Dimension 2
2001	航线收益分析	代理人运输贡献分析
2002	航线网络分析	营销运价分析维
2003	现金返点分析	返还机票分析
…	…	…

表4 时间表

Time_id	年	月	周
4001	2009	01	1
4002	2010	01	1
4003	2011	01	1
…	…	…	…

用户表可以构成“访问元数据”的一个访问控制维度，如用户张华登录系统时可以选择“张华”这个维度来查看用户的访问权限。

立方体表描述了当前系统中用户可以访问的数据立方体。如当前系统中存在3个数据立方体（收益分析、商务分析和促销分析），在用户访问时可以获得用户想访问哪个立方体的数据，是收益分析的数据还是商务分析的数据。

维度表描述了系统中每个用户可访问的立方体分别由哪些维度构成。如立方体2001（即收益分析立方体）中含有航线收益分析维和代理人运输贡献分析维等。

时间表描述了系统中的用户可以在什么时间内是否可以访问数据仓库中的数据。例如，用户只可以在周一访问收益分析立方体的数据等。

除了构造维表之外还需要用户的访问事实表（Access_fact），访问事实表中存储了用户的访问权限及其相关数据。

2.2.2 “访问元数据”星型结构

维表和事实表之间的星型关系如图3所示，通过这些数据可以构造进行安全访问控制的“访问元数据”模型。通过模型可以控制某个用户对于某个立方体的某个维度在什么时间内是否具有访问的权限。

基于这种结构的访问控制系统，可以方便地添加、删除和修改控制维度，能够控制更多的用户信息，例如可以再添加“销售收入分析”控制维度，使得系统可以控制是否让某些地区的用户进行数据仓库的访问。维度越多，所构造的“访问元数据”越复杂。

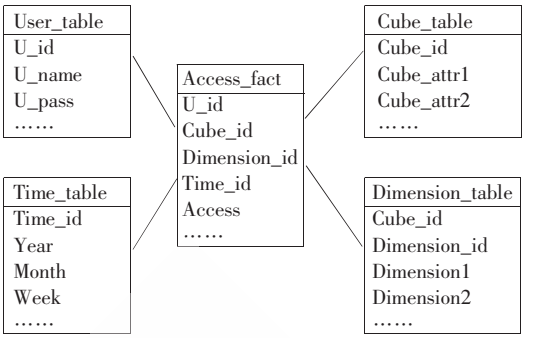


图3 访问元数据星型结构图

2.3 访问流程设计

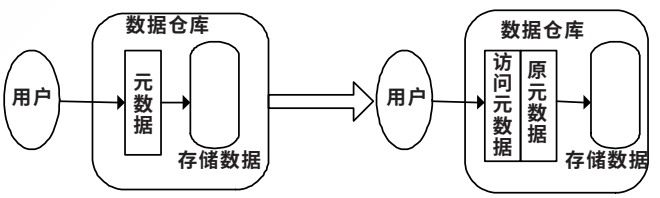
通常用户访问数据仓库中的数据时，首先是用户提出访问请求，这些访问请求被处理后传送到数据仓库的元数据处，再根据元数据的处理从存储关系表或者多维数据库中获得用户需要的数据，如图4（a）所示。在这里，元数据起着极其重要的作用，它是到后台数据库中寻找数据的组织者，是用户访问数据的必经之路。鉴于此，把数据仓库中的元数据扩展为两部分，即本来的元数据（此处称之为原元数据）和“访问元数据”，在“访问元数据”中加入用户的访问安全规则，通过规则控制用户对数据仓库的访问，达到安全访问数据仓库数据的目的。这样，加入访问元数据之后的数据仓库系统访问如图4（b）所示。

这样，用户登录的过程可以细化为如下步骤：

（1）首先用户登录需要经过身份验证，通过身份验证的用户才可以进入系统。

（2）获取用户要访问什么信息，如张三在周一提出访问促销分析立方体的现金返点分析维度。

（3）通过“访问元数据”进行用户权限鉴别，查看用户是否具有相应的权限，如果有则可继续数据仓库数据的访问，否则返回用户无权访问等信息。



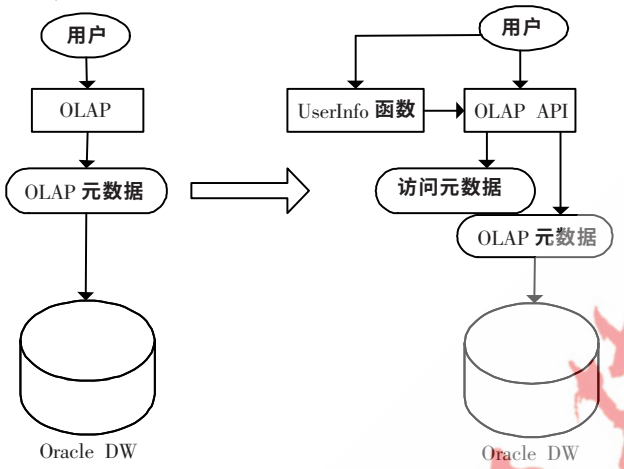
(a)通用数据仓库访问　(b)基于元数据的访问

图4 通用数据仓库访问与基于元数据的访问比较示意图

2.4 可行性分析

在当今数据仓库相关商业产品中，Oracle占据重要地位，Oracle 9i将数据库、OLAP、Data Mining、J2EE以及Web集成于一体，通过适当的剪裁，可以灵活地适应多种网络数据应用环境，非常便于实施开放的、跨平台的应用逻辑，是一个优秀的数据仓库平台解决方案。更重要的是Oracle 9i支持CWM元数据规范，提供了丰富的元数据操作接口。Oracle 9i的这些功能为基于元数据的安全模型的实施提供了方便。

图5是通过“访问元数据”的方式来实现对Oracle数据仓库环境中关系数据和多维数据访问的原理图，创建“访问元数据”以支持安全访问控制，并且以API的方式提供给用户使用。用户只有通过增加的这套API才能访问“访问元数据”，达到控制安全的目的。这种方式可以无缝地应用到SMARTRIX数据仓库中，因为这套系统恰恰是使用Oracle 9i做的开发，在所有需要进行系统支持的用户访问和维护数据仓库数据之前添加一个元数据访问控制模块后，只允许用户访问自己业务相关的模块，即可大大增强原有系统的安全性。



(a)原系统访问　　(b)基于元数据访问

图5　基于元数据的安全模型实现原理图

支持人员进行数据访问时，UserInfo函数将收集到支持人员的访问需求，并把这些信息以参数的形式传送到OLAP API，通过OLAP API对“访问元数据”进行访问。如果支持人员访问的数据在他的权限之内，则允许支持人员继续对存储在数据仓库中的数据进一步地访问；否则，返回非法操作。CWM2是Oracle 9i中引入的标准，目前只能通过CWM2 PL/SQL packages来操作[4]。利用CWM2 API可以方便地构建访问控制维和构建“访问元数据”模型。

本文根据企业现有的SMARTRIX数据仓库系统可能面临的安全问题进行了分析，深入研究了数据仓库分析决策支持系统的安全性需求以及元数据在数据仓库中的重要作用，讨论了基于元数据的数据仓库安全模型。参考文献[3]中提出一个基于元数据的数据仓库安全模型，但是由于它需要MQL语言的支持，不易于在商业数据仓库平台上实施。针对这一问题，本文利用CWM元数据标准的规范化，对其模型进行改进，并提出了在商业数据仓库平台Oracle 9i平台上的实现方案，有效解决了作者企业决策支持系统下一步可能面临的安全访问控制问题，确保了数据仓库的安全性。相对原模型，本模型具有操作简便、结构可控同时可以控制多个要素的优点，不需要MQL语言的支持，当前通用的支持CWM规范的数据仓库系统都可使用，这是对原模型功能的完善与增强，可以很好地解决现有系统面临的安全威胁，这无论在理论上还是实践上都具有重要的意义。

参考文献

[1] KATIC N, QUIRCHMAYR G, SCHIEFER J, et al. A prototype model for data warehouse security based on metadata [C]. Proceeding of the 9th International Workshop on Database and Expert Systems Applications. 1998:300-308.

[2] PIATTINI M, RODEROJ A. Auditing data warehouse security [C]. Proceeding of IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology. 1999: 255-261.

[3] STADUDT M, VADUVA A, VETTERLI T. The role of metadat for data warehouse[EB/OL][2000-05-18] http://www.informatik.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/SVV99.pdf.

[4] WU N. Oracle9i OLAP CWM2 API使用指南. [2007-07-05] http://www.dwway.com/article-3084-1.html.

(收稿日期：2011-12-02)

作者简介：

马艳锋，女，1980年生，硕士研究生，工程师，主要研究方向：数据仓库与数据挖掘，网络软件。