文章编号：1005-8451（2016）09-0088-04

数据仓库安全模型研究与设计

焦怡博，刘湘黔，崔亦博
（北京交通大学 计算机与信息技术学院，北京 100044）

摘 要：随着数据仓库的发展，安全问题日益突出，一方面数据仓库本身需要足够的开放性，另一方面数据仓库的数据价值很大，要求对数据仓库用户加强数据访问的限制。开放与限制，两者的矛盾会严重制约数据仓库使用。论述了数据仓库安全问题产生背景及安全模型的重要性，针对数据仓库的安全问题提出安全模型。该模型能够减少数据仓库会出现的安全问题，为企业进行数据仓库设计提供参考。

关键词：数据仓库；安全模型；元数据；角色
中图分类号：U29-39 **文献标识码：**A

Security model for data warehouse

JIAO Yibo, LIU Xiangqian, CUI Yibo
(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract: With the development of data warehouse, the security problem is becoming more and more serious. On the one hand, the data warehouse itself needs enough openness. On the other hand, the data warehouse is of great value, which is required to strengthen the restriction of data access for the users. Opening and restriction, the contradiction between the two will seriously restrict the use of data warehouse. This article discussed the background of data warehouse security and the importance of security model, put forward a security model for the security of data warehouse. The security model could reduce security problems and provide a reference for the design of data warehouse to the enterprise.

Key words: data warehouse; security model; metadata; role

由于全国各个城市轨道交通运营规模不断扩大，各类数据的记录和存储而产生海量数据的分析、挖掘与应用将成为轨道交通智能化数据管理的核心内容。目前，国内的轨道交通运营管理企业基本都拥有自动售检票系统（AFC，Auto Fare Collection）等智能化的设备管理系统，其中存储着大量历史数据。

传统的数据库已经不适合存储如此大量的数据，体现在如下几个方面：

（1）传统数据库系统仅能够满足数据的增、删、改、查以及报表生成等功能不能满足大量复杂查询和即席查询的需求。（2）传统的数据库系统已经不能满足用户对海量数据进行分析以及处理的需求，目前现有的数据库系统是联机事务处理系统（OLTP），而并不是联机分析处理系。（3）传统数据库系统无法满足对分析的数据对象进行灵活并多维的查询。（4）传统数据库生产系统不能够满足城市轨道交通运营信息快速膨胀以及数据一体化整合的需求。

如何存储并利用如此巨量数据，成为当今国内轨道交通运营管理企业亟待解决的一个重要课题。

1 数据仓库

数据仓库实质上是一个数据集合，它是面向主题的、稳定的、集成的并且随时间不断变化。 数据仓库发展至今已成为相当重要的数据重组技术，它可以方便地构建联机分析处理（OLAP）、数据挖掘等高级数据分析应用，数据仓库系统为企业经营决策提供了良好支持，已成为企业运营的一个重要环节[1]。

由于数据仓库从多种数据源中提取数据，并经过分析、处理，其数据价值已经提升；数据仓库中数据共享的形式、范围都在呈现出新的变化，再加上数据仓库本身不是一个管理系统，只是位于数据库和分析系统应用之间，主要用来组织数据，并不知道用户要访问的内容。也就是说数据仓库更加侧重于数据组织与重组，没有固定的实现形式。

收稿日期：2016-06-15
作者简介：焦怡博，在读硕士研究生；刘湘黔，副教授。

数据仓库的本质是开放的，数据仓库的主旨之一是可以使用户轻松访问大量数据。因此任何安全模型都会限制实现该主旨，当然也会成为数据仓库设计的一大障碍。如果数据仓库的分析员在使用数据仓库时因为安全模型而受到限制，就会大大影响其分析的效率和正确率[2]。但是，基于某些商业需求，有时候数据仓库会需要保持很高的安全性，因为一个潜在的数据漏洞有可能导致不可挽回的损失。因此，好的数据仓库安全模型设计就显得非常有必要。

国内对于数据仓库安全的研究还处于边缘阶段，国外对这方面的研究取得了很多成果。例如：

Weipple 等人提出了一种基于数据仓库予以环境的授权模型，论文中描述了数据仓库中多维数据模型的基本要素、主题以及 OLAP 操作。但是，该模型还缺乏对概要数据操作的访问控制表述以及访问权限派生的问题[3]。

学者 Remzi 在文献 [4] 中提出一种基于角色的数据仓库安全访问控制模型，该模型将安全对象建立在数据仓库多维数据模型上，并对 OLAP 分析性操作进行访问控制。

Indu Singh 在论文 [5] 中总结了近几年国外数据仓库安全性方面研究，这些成果有很大的借鉴意义。

本文根据 RBAC 模型提出了一种基于角色的元数据访问安全模型。

2 模型介绍

现今，数据仓库安全模型已出现众多安全机制。

Katic 在论文中提到一种基于元数据的访问模型，并在其 WWW-DIS-DWH 项目中根据用户的权限大小，为不同的用户组构造不同的元数据，然后将这些元数据分别赋予相应用户组访问的权限。这并不影响用户正常使用数据仓库，因为只需要有限内容的元数据，用户便可以得到自己想要的数据[6]。

Katic 的安全模型依靠结构化的描述语言（MQL）。用户组的访问控制以 MQL 语句存储。当接到用户发送地请求后，先将它转化为 MQL 查询，再与相应的用户组的访问权限对比。由于依靠 MQL，使得 Katic 的安全模型有一定的专有性，不宜推广。

本模型在借鉴 Katic 对于元数据访问控制方法的同时，不使用 MQL 语言而是使用一种实时生成元数据控制表的机制，使得本模型相对于 Katic 安全模型有很大通用性。

不同用户访问数据仓库中的数据时，有的用户可以从高度集成的数据开始，然后不断深入下钻至详细的数据中。有的用户要求在某程度集成的层面上进行操作。这样就很难管理不同用户对数据的访问。因此加入角色访问控制机制，隔离用户和元数据，产生一个代理层，解耦了用户和元数据的关系，方便管理者对用户进行管理[7]。

将基于角色访问控制和对元数据进行访问限制两方面结合起来能有效解决数据仓库安全问题。

2.1 整体设计

安全模型整体设计如图 1 所示。

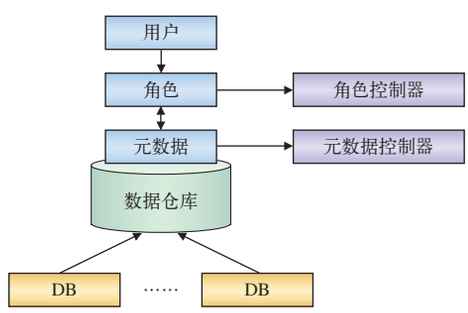

图1 安全模型整体设计

数据库（DB，Database）：存储大量的历史数据。

数据仓库：从数据库中经过抽取（Extract）、转换（Transform）、加载（Load），即 ETL 操作，将数据存储到数据仓库中。

元数据：主要是描述数据属性的信息，用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。元数据是数据仓库环境的一个重要组成部分。元数据就是关于数据的数据。自从有了程序和数据，元数据就是信息处理环境的一部分。但是在数据仓库中，元数据扮演一个重要角色。正因为有了元数据，数据仓库才被有效地利用。元数据使最终用户或决策支持系统（DSS）分析员能够探索各种分析主题的可能情况[1]。

角色：给每个用户分配不一样的角色，不同的角色有不一样的权限，可以访问不同元数据。

角色管理器：可以通过角色管理器来管理角色的激活与去活、将角色分配给用户、修改角色对应

的元数据等，角色管理器具体细节会在下文提到。

元数据控制器：通过管理 1 张元数据控制表来管理元数据访问规则，具体细节会在下文提到。

元数据分配给相应角色，同时为系统创建角色，并将角色分配给用户。

整个系统流程如下：

（1）用户登录。进行登录身份验证，通过的用户可以进入系统。(2)角色登录。用户申请角色登录，通过角色身份验证的用户就拥有该角色访问某些元数据的权限。(3）整理信息。用户申请数据，综合整理用户想要访问的信息。如 A 用户使用 B 角色，想要在 C 时间访问 D 数据的 E 维度。(4）元数据控制器。为用户生成元数据控制表，通过元数据控制表判断该用户是否具有访问元数据的权限，如有则向用户返回数据。

2.2 结构设计

2.2.1 角色控制器设计

数据仓库的安全大多借助其他传统系统，在传统系统中最直接的安全措施就是访问控制。访问控制大多可以分为：基于对象的访问控制、基于任务的访问控制和基于角色的访问控制。基于角色的访问控制（RBAC，Role-based Access Control）模型主要思想是用户在系统中扮演不同角色，各角色在系统中拥有不同权限。通过对角色的管理来控制用户对数据的访问，角色成为模型中链接访问主题和受控对象之间的一座桥梁 [7]。图 2 是对 RBAC 模型进行改进之后的模型。

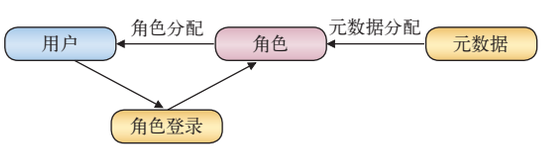

图2 角色控制模型

角色是角色控制器的核心。系统分配通过元数据分配使每个角色能访问有限的元数据，通过角色分配使每个用户获得一个或多个角色；用户通过角色登录进行身份验证，通过后才具有该角色的权限；角色为用户和元数据搭起了一座桥梁的同时也将元数据和用户隔离开来，达到了访问控制的目的。

用户和一个或多个角色相关联，角色同一个或多个元数据相关联；角色可以根据实际的工作需要添加、删除以及修改；已经通过用户身份验证的用户，在进入系统后如果需要使用角色访问数据，仍要进行角色身份验证来激活相应的角色。除此之外，角色之间可以定义一些关系，比如：继承关系、排斥关系等，也可以按需要添加约束条件；比如：定义采购和财务两个角色为互斥角色，则这两个角色不能分配给一个用户。

2.2.2 元数据控制器的设计

为方便理解元数据控制器的结构，这里构造了 6 个表：用户表、角色表、时间表、IP 表、数据立方体表、维度表。这 6 个表用来模仿角色在访问元数据时的 6 个维度：用户、角色、时间、IP、维度、立方体。如：用户张三使用角色系统管理员在时间 2009.05.04 访问票务立方体的客流分析维度，IP 为 192.0.0.168。

如表 1 所示，用户表记录了用户的 ID 和名字。

表1 用户表

USER_ID	USER_NAME
00001	张三
00002	李四
…	…

如表 2 所示，角色表记录了系统所划分的多种角色，如系统管理员、电工、数据分析组等等。

表2 角色表

ROLE_ID	ROLE_NAME
10001	系统管理员
10002	数据分析组
…	…

时间表如表 3 所示，记录了某角色什么时间按访问哪些立方体或元数据。如某角色只被允许在月末的时候访问票务分析立方体。

表3 时间表

TIME_ID	年/月/日
20001	2009.02.04
20002	2015.04.29
…	…

如表 4 所示，记录了系统允许使用的 IP 地址，非正常 IP 系统可以进行限制，禁止其访问重要数据。

数据立方体表如表 5 所示，描述系统中有哪些

表4 IP表

LOCATION_ID	IP
30001	192.0.0.168
30002	211.177.22.54
…	…

可以访问的数据立方体，如票务分析、行车记录等。

表5 数据立方体表

CUBE_ID	CUBE_ARRT1	CUBE_ATTR2
30001	票务分析	…
30002	行车记录	…
…	…	…

维度表如表 6 所示，记录了各个立方体中可供用户分析的维度，如票务立方体可供分析的维度有客流分析、OD 分析等。

表6 维度表

CUBE_ID	DIMENSION_ID	DIMENSION_1	DIMENSION_2
30001	40001	客流分析	OD分析
30002	40002	收益分析	票卡分析
…	…	…	…

通过以上 6 个表数据构建一个元数据访问控制表，如表 7 所示。每次有访问请求时，系统会根据用户生成 1 张元数据访问控制表，将用户所扮演的角色，各角色可以访问的立方体，各立方体拥有的维度等信息都保存到元数据访问控制表中，通过查询元数据访问控制表，就能得到该用户是否具有访问数据权限，根据元数据访问控制表返回数据。

表7 元数据访问控制表

元数据访问控制表
ROLE_ID
CUBE_ID
DIMENSION_ID
TIME_ID
LOCTION_ID
…

元数据访问控制表具有以下特性：

（1）易用性。系统管理者可以随时根据需求对上述维度（角色、时间、IP、维度、立方体）进行增加、修改等操作，通过调整用户组对应的元数据定义，就可以实现安全访问控制策略的变换。（2）通用性。元数据访问控制表可以在任何数据仓库模型中进行设置。

2.3 模型优势

（1）便于数据系统的集成。用户访问控制都是集成在一起，只要修改角色控制器和元数据控制器就可以控制整个系统的访问。（2）从角色以及元数据两方面控制了用户的访问，安全性增加。（3）用户在进行数据访问时，元数据控制器会根据元数据访问表返回其有权访问的数据，其他数据被屏蔽了，这就减少了用户非法访问的企图，增加了系统的安全性。（4）用户只需要进行一次用户登录身份验证和一次角色身份验证，就可以正常访问其需要的数据。在保证数据安全性的前提下，保证用户访问数据仓库的易用性。做到了安全性与易用性的兼顾。（5）本模型具有良好的通用性，可以移植到大部分数据仓库系统中。

3 结束语

针对数据仓库安全性不足的状况，本文基于角色控制和元数据控制的方法提出了一种数据仓库安全模型。该模型通过角色管理，限制访问元数据达到访问控制的目的，有效提高了数据仓库的安全性，为数据仓库系统的设计提供了参考。

参考文献：

[1] Willianm H. Lnmon.Building the Data Warehouse[M]. 王志海，译 . 北京 : 机械工业出版社，2000.

[2] 孙 涛 . 浅析数据仓库安全问题及措施 [J]. 电子技术与软件工程，2014(13).

[3] 余文霞 . 浅谈数据仓库安全问题及安全措施 [J]. Science & Technology Information，2013（4）.

[4] 张大刚 . 数据仓库基于角色强制访问控制研究 [J]. 计算机与现代化，2011（5）.

[5] Indu Singh, Manoj Kumar. Evaluation of approaches for designing secure data warehouse[C]. International Conference on Advances in Computing, 2012.

[6] 马艳锋，谭立彦 . 基于元数据的数据仓库安全模型实现研究 [J]. Microcomputer & Its Applications，2012，31（9）.

[7] 李 毅，刘 海，王岳斌 . 基于角色的数据仓库分级安全模型的形式化研究 [J]. 计算机应用于软件，2010，27（5）.

责任编辑 徐侃春