

## 异构数据库应用系统集成的简化方法\*

李春强<sup>1,2</sup>, 柴玮岩<sup>1</sup>, 刘健<sup>2</sup>, 陈立南<sup>3</sup>

(1. 北京理工大学计算机学院, 北京 100081; 2. 北京信息科技大学信息管理学院, 北京 100101;

3. 北京邮电大学网络信息中心, 北京 100876)

**摘要:**针对现有集成方法主要面向数据库系统层面的不足,重点讨论面向应用系统层面的集成方法。通过引入数据库应用系统的访问控制新视角,将访问控制层从应用层中划分出来,继而提出融入式访问控制技术,并基于该访问控制技术提出一种新的异构数据库应用系统简化集成方法。通过实例展示了如何应用简化集成方法实现异构数据库应用系统的集成,并通过对比分析验证了新方法的优势。

**关键词:**数据库应用系统;异构数据库;信息集成;访问控制;简化集成

**中图分类号:**TP311 **文献标志码:**A **文章编号:**1001-2486(2016)02-061-07

## Simplified method for integration of heterogeneous database application system

LI Chunqiang<sup>1,2</sup>, CHAI Weiyan<sup>1</sup>, LIU Jian<sup>2</sup>, CHEN Linan<sup>3</sup>

(1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China;

2. School of Information Management, Beijing Information Science & Technology University, Beijing 100101, China;

3. Network Information Center, Beijing University of Post and Telecommunications, Beijing 100876, China)

**Abstract:** The existing integrated methods are mainly aimed at the layer of database system. In view of the shortage of those methods, the integration method in the application layer was focused on. According to a new perspective of access control for the database application system, the access control layer was separated from the application layer. A new method called as "fusion" access control was put forward. Based on the method, a simplified integration method of heterogeneous database application system was proposed. An example was shown about how to apply the simplified integration method to achieve the integration of heterogeneous database application system, and the advantages of the new method were verified by a comparison analysis.

**Key words:** database application system; heterogeneous database; information integration; access control; simplified integration

自数据库(尤其是关系数据库)技术产生以来,数据管理技术取得了革命性进步。关系数据库技术经历了自20世纪70年代以来四十多年的发展,围绕关系数据库管理系统形成了一个完整的生态系统(厂家、技术、产品、服务等),提供了包括数据采集、数据管理、数据查询和分析、数据展现(可视化)等技术和产品,创造了巨大的数据库产业<sup>[1-2]</sup>。数据库应用系统(Database Application System, DBAS)即为其中之一,它普遍地应用在大多数企业信息化建设之中。然而,由于企业的不断发展及其组织内、外部的调整与变化,例如企业间的兼并或业务功能的扩展等,致使异构DBAS的集成问题出现。随着企业信息化进程的进一步加速,每个业务系统数据库的异构问

题越来越突出,而这些已存在的业务系统通过多年的积累已产生了大量的业务数据,如何有效地利用这些数据,即集成异构的DBAS,成为集成领域的一个热点<sup>[3]</sup>。

解决这一问题需要集成不同的数据管理系统,对存储在不同系统中的信息提供统一访问。DBAS是针对某个特定目标,建立在数据库管理系统(Database Management System, DBMS)之上的计算机应用系统,它包含两个核心的部分,即数据库系统和应用系统。现有的集成方法主要面向数据库系统层面,用户在访问数据源时,越过了应用的显示层和业务逻辑层,直接对应用的数据结构或数据库进行操作。此外,这种集成通常需要建立一些全局统一模式,如一个全局的数据库模

\* 收稿日期:2015-03-18

基金项目:“核高基”国家科技重大专项基金资助项目(2012ZX01039-004-48)

作者简介:李春强(1973—),男,吉林敦化人,副研究员,博士研究生,E-mail:tsiang@126.com

式,便于异构数据库集成的公共数据模型,以及向用户提供统一模式和访问接口全局应用。因此,这种集成方式通常比较复杂,其利用率低而成本高,且周期长,不能充分满足当前异构 DBAS 集成需求的多样性。这种集成方式通常针对需要全局联合查询数据来提供相关服务的情形(如决策支持),称之为紧密集成。

正如文献[4]指出的那样,信息集成存在于不同的领域,每一领域有各自特定的需求,这些需求需要新方法进行信息集成。因此,李春强针对异构 DBAS 的集成问题,重点讨论面向应用系统层面的集成,试图寻求一种成本低而利用率高的快速集成解决方案,它主要针对系统功能扩展或业务集成之类的问题,称之为松散集成。

DBAS 是由数据库系统、应用程序系统、用户组成的,可以很容易地看出传统的紧密集成和欲求解决的松散集成方式之间的区别:松散集成方式的各集成部分在数据库模型、数据模式等方面还是同构的,仅仅是应用系统的功能不同而已,而紧密集成后各集成部分将具有统一的应用系统功能。

## 1 相关工作

信息集成是利用若干数据库或其他的信息来源<sup>[5]</sup>,使得来自这些来源的数据协同工作的过程。集成的数据库可能是物理的(如:仓库)或虚拟的(mediator 或中间件,即使不是物理上存在也可以被查询到)。来源可能是常规的数据库或者其他的信息类型,比如网页的集合。一般有两种最常用的使数据库或其他分布式信息源协同工作的方法:一种是建立数据仓库,将众多遗留数据库中的信息进行适当的翻译,周期性地拷贝到一个中心数据库中,如文献[6];另一种是实现一个 Mediator 或中间件,它是一种软件组织,不存储任何自己的数据,它是支持各个不同数据库中数据的一个集成模型,并在这个模型和每个数据库所使用的实际模型之间进行翻译,如文献[3]和文献[7]。

目前,绝大多数的研究都是基于后一种方法,且解决方案几乎都是通过寻求一种通用的公共数据模型,即以局部数据库模式为基础,建立全局的数据模式或全局外视图,如文献[8-17]。其中,文献[11]从本体的角度出发,认为基于本体语义集成异构信息来源的系统结构一般分为三层:底层是各个分布的异构数据源;中间层存在一个描述特定用户群关心领域的全局本体,它对每个用

户具有统一的语义,同时它与底层的异构数据源有着特定的对应关系;上层是用户查询访问层。在基于本体的信息集成中,如何表示用户对本体概念实例的查询、如何建立本体到异构数据源的映射关系、如何根据这些关系将全局查询划分为若干个局域查询、如何集成局域查询结果,这些问题都需要解决。文献[16]提出了一种网格环境下模式复用的异构数据库访问和集成方法,该方法首先提出了一种模式复用的虚拟数据库构建策略,定义了异构数据库资源的全局统一视图,在此基础上给出了数据库资源注册方式和虚拟数据库查询映射方法,为资源使用者提供了访问数据库的一致访问接口。还有一些研究则借助一些流行的架构(如 CORBA, XML 和 Java 等),如文献[12-15]。

然而,现有研究存在着若干问题。第一,这种数据模型既要自然地描述各种数据源的数据,又要方便地表示来自各种异构数据源的异构数据,因而必须具有很强的描述能力<sup>[8]</sup>。第二,这些研究所针对的数据源不仅包含数据库系统,还包含其他一些数据源,如文件系统、电子邮件、电子表格、WWW 上的 HTML 文件等。实验证明<sup>[9]</sup>,简单领域本体构建对于领域专家来说工作量比较小,一般很容易完成,而且对模式匹配成功率的提高有很大的作用。包含的功能越多,为了求同存异,实现的方式必然越复杂,尤其是在科学技术日新月异的今天,分支越来越细,大而全的集成方案可能会带来诸如集成的高复杂度、高成本以及低效率等问题。第三,现有的集成方法主要是针对数据库系统层面,几乎摒弃了其上的应用系统。数据需要专门的业务逻辑处理才能成为有利用价值的信息,而应用系统正是为此目的开发的,它能充分发掘和应用数据库中的数据。所以,数据库系统之上的应用系统是很有利用价值的。第四,集成需求是有差别的。企业信息化建设是一项复杂的系统工程,不同的行业、企业规模和建设阶段都会带来不同的建设需求,应该详细分析、因地制宜和有针对性地选择符合企业所能承载的资金成本以及技术含量的集成方案。第五,现有相关研究几乎都将集成系统的访问控制与集成方法分开考虑,如文献[18]和文献[19]。

综合以上分析可以看出,虽然有大量的研究信息集成的文献,但是信息集成所面临的挑战远未解决,特别是对于企业环境,企业真正需要的系统应该允许以一种透明的方式集成信息,而不需要整个的结构化和非结构化的数据源的模式信

息<sup>[4]</sup>。因此,李春强等针对应用系统层面的集成需求,寻求一种低复杂度、低成本以及高效率的集成解决方案。

### 2 DBAS 的访问控制视角

DBAS 是由数据库系统、应用程序系统、用户组成的,具体包括:数据库、DBMS、数据库管理员、硬件平台、软件平台、应用软件和应用界面。这几部分以一定的逻辑层次结构方式组成一个有机的整体,它们之间的层次结构关系如图 1 所示。

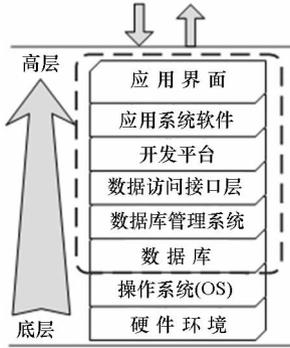


图 1 DBAS 层次结构示意图

Fig. 1 Schematic diagram of DBAS hierarchy

图 1 中虚线框所示部分为组成 DBAS 的要素:数据库系统(包括数据库、DBMS 和接口)和应用程序系统。此外,由于 DBAS 会有多个用户使用,为了使数据能被任何有权使用的用户合法使用,DBAS 还有一个很重要的组成部分,即访问控制。访问控制技术就是为了限制访问主体对访问客体的访问权限——能访问系统的何种资源以及如何使用这些资源,阻止未经允许的用户有意或无意地获取数据的技术。访问控制的手段包括用户识别代码、口令、登录控制、资源授权(例如,用户配置文件、资源配置文件和控制列表)、授权核查、日志和审计。从访问控制的新视角重新分析 DBAS,如图 2 所示。

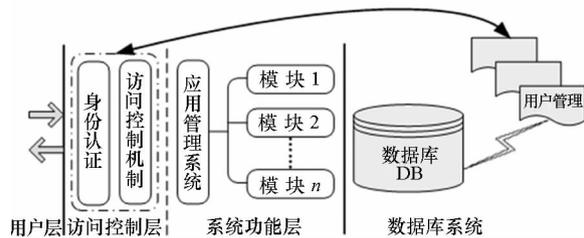


图 2 DBAS 的访问控制视角

Fig. 2 DBAS access control perspective

图 2 中将访问控制功能从应用系统中划分出来,记为访问控制层;划分出访问控制层后,原应

用系统的保留部分记为系统功能层;用户在访问控制层的监控之下访问 DBAS。一般地,访问控制所需要的信息(如用户账号、密码和授予的权限等)被存储在数据库特定的一系列表中,如图 2 中所示的“用户管理”,用户访问 DBAS 时需首先通过访问控制层从数据库的“用户管理”中检索用户信息,在通过身份验证后用户才能按所授予的权限来访问 DBAS。换言之,在 DBAS 中数据是通过应用程序提取、组织和显示的,这不仅需要一定的介质来对数据进行存储、加工和整理,并进行开发和利用,还需要一定的访问控制方式。因此,在信息集成时,需要综合考虑信息的存储、组织、管理和访问控制。

基于上述,从 DBAS 的访问控制视角,可以进行如下推导:

1) 将 DBAS 形式化表示为  $R = (U, AC, AS, DB)$ , 其中:  $U$  表示用户,  $AC$  表示访问控制,  $AS$  表示应用管理系统,  $DB$  表示数据库系统;

2) 若设  $U$  为定义域,  $DB$  为值域,  $AS$  为映射函数,  $AC$  为限制定义域的取值以及整个函数映射关系的规则。则可将  $R = (U, AC, AS, DB)$  中各参数之间的关系表示为如图 3 所示的函数关系,即  $DB = AS(U)$  (附加  $AC$  规则)。

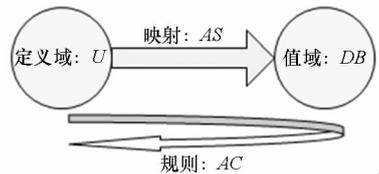


图 3 DBAS 的函数映射关系示意图

Fig. 3 Function map of DBAS

对 DBAS 而言,只要拥有合法的用户账号,就可以在授权范围内访问它,这一点从访问控制的视角能更清晰地看到。此外,从访问控制的视角来看,可以说以往的大多数研究是通过寻求函数映射关系的方式来提供集成解决方案的;然而,找符合定义域的值比寻求函数映射关系要容易得多。在面向应用系统层面的异构 DBAS 集成中,正是通过各数据源的“定义域”来寻求解决方案的。

### 3 嵌入式访问控制

数据库系统通常采用三级模式结构,并提供两级映像功能,这是数据库管理系统内部的系统结构,如图 4<sup>[20]</sup>所示。虽然实际的数据库管理系统产品种类很多,它们支持不同的数据模型、使用

不同的数据库语言、建立在不同的操作系统之上、数据的存储结构也各不相同,但它们在体系结构上通常都具有相同的特征。数据库系统的三级模式是对数据的三个抽象级别,它把数据的具体组织留给 DBMS 管理,使用户能逻辑地、抽象地处理数据,而不必关心数据在计算机中的具体表示方式与存储方式。在这三级模式之间,数据库管理系统提供了两层映像,它保证了数据库系统中的数据能够具有较高的逻辑独立性和物理独立性。在加入集成系统之前,每个 DBAS 就已经存在,并拥有专属的 DBMS,即集成系统中的各数据源都享有各自独立的自治性;在实现数据共享的同时,每个数据源仍保留有各自独立的应用特性、完整性控制和安全性控制。

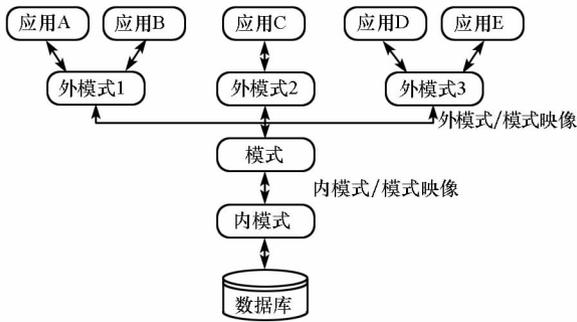


图 4 数据库系统的三级模式结构

Fig. 4 Three levels pattern structure of database system

可以通过下述方式实现集成系统的访问控制。

首先,将整个集成系统的访问控制域(用 D 表示)划分为以下三种类型(如图 5 中虚线框所示):

- 1)本地域(用 LD 表示):本地访问控制域。集成系统中,各数据源自身原有的访问控制域。
- 2)协作域(用 PD 表示):协作源访问控制域。对某一个数据源来说,集成系统中其他数据源的原有访问控制域。
- 3)协调域(用 RD 表示):协调系统访问控制域。集成系统中,用于集成各数据源的那部分系统(称为协调系统)对应的访问控制域。

相应地,将系统中的用户(用 U 表示)分为以下三种类型:

- 1)本地用户(用 LU 表示):集成系统中,各数据源的用户。
- 2)协作用户(用 PU 表示):对某一个数据源来说,来自集成系统中其他数据源的用户。
- 3)协调用户(用 RU 表示):协调系统访问控制域中的用户。

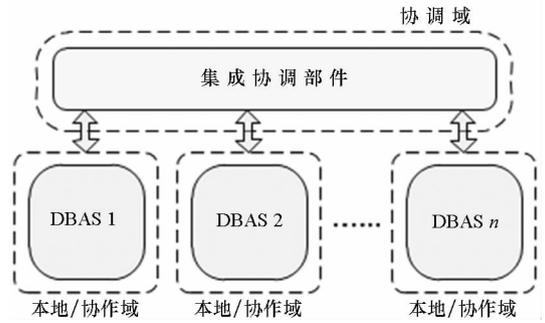


图 5 集成系统的访问控制区域划分示意图

Fig. 5 Schematic diagram of the access control area of integrated system

最后,拟订如下规则:

- 1)各数据源申请加入集成系统时,系统会给每一个数据源(包括协调系统在内)分配一个唯一的系统标识符(System Identifier, SID)。
- 2)各域中所建用户的用户名不允许以 SID 开头。
- 3)各域采用各自的访问控制机制,可以相同也可以不同,且互不干扰。在本地域 LD(或协作域 PD)中,访问控制方式仍然保持各数据源系统原来的访问控制机制。所以,在集成系统中,只需为协调域 RD 选用适宜的访问控制机制即可。
- 4)各域管理本域的用户,各用户在其对应的域中活动,且只需在本域内保证用户名唯一性约束即可。

5)通过增加一些功能机制,可实现各用户跨域访问。

本地用户 LU 向协调系统申请,实现跨域访问,有两种方式:跨域到协作域,称为用户平移;跨域到协调域,称为用户升级(这种方式只有在协调系统有全局应用时才适用,且不提倡,可根据实际需要进行取舍)。跨域时,该用户名前加其所在数据源的 SID 作为前缀,并以这个“新”用户名作为所跨域的用户,接受所跨域的管理,并按所跨域的授权在该域中活动。

协调用户 RU 向协调系统申请,可以跨域到各数据源所在域中活动,称为用户降级。同样地,跨域后该用户名也要进行变更,即用户名前加协调系统的 SID 作为前缀,其余同上种跨域访问方式所述。

6)申请跨域的用户仅限定为各域所建的用户,而不包括“跨域”用户(即以 SID 开头的用户)。

整个集成系统的访问控制域可简单地表示为: $D = LD + PD + RD$ ;用户可表示为: $U = LU +$

PU + RU。集成系统的访问控制是以(域,域用户)进行的,即:(LD,LU),(PD,PU),(RD,RU)。这样划分的好处在于,将整个系统的访问控制“封装”到各数据源和集成部件中,使得原已实现的访问控制功能在集成系统中还能继续正常运转而无须任何改变。同一个用户,经协调系统进行跨域处理后,可以用“同一个”用户名在各域中活动。对于用户权限的授予,可以采用先默认授权,再根据需要进行调整的方式。对用户而言,这一切都是透明的。这样一来,可以大大降低集成系统访问控制机制的复杂度,所需要的只是增加一些协调各数据源“定义域”的功能机制;另外,可能需要增加用于调用各 DBAS 的入口链接等,而这只需简单的一个应用界面和一些后台链接处理即可。所增加的功能机制有很多实现方式,特别是在数据库系统中可通过一些简单的方法来实现,下文实例部分将会介绍一种实现方式。

这种方法是在所要访问的系统中生成一个真正的“本地”用户,它完全融入被访问系统的内部之中。因此,将其称为融入式访问控制技术。

## 4 简化集成方法

现有的集成方案几乎都是将访问控制与集成方法分离开来讨论,即先定集成,再定访问控制机制。简化集成方法(Simplified Integration Method, SIM)是以融入式访问控制为基础的,即将访问控制与集成方法结合起来实现集成的方法。

图6为SIM方法的系统框架,各部件功能如下。

1)注册中心:负责对集成系统中各数据源的注册信息进行维护管理,包括各数据源的入口地址、系统功能简介等,并分配标识SID。

2)用户中心:负责对协调域中的用户进行维护管理,包括用户名和所授予的权限。这里的用户包括协调域所建的用户和来自各数据源申请跨域访问的用户。

3)申请/审批中心:负责用户跨域访问申请的提交和审批。

4)协调中心:负责用户跨域访问处理。

5)访问控制:协调域的访问控制机制。

6)全局应用系统:负责为用户提供访问集成系统的应用界面,主要包括各数据源名称列表,并附有所包含系统功能和数据资源的简介以及各DBAS应用系统的入口地址链接等。

一个新的数据源DBAS在注册中心进行注册,获得自己的标识SID后,即可加入到集成系统

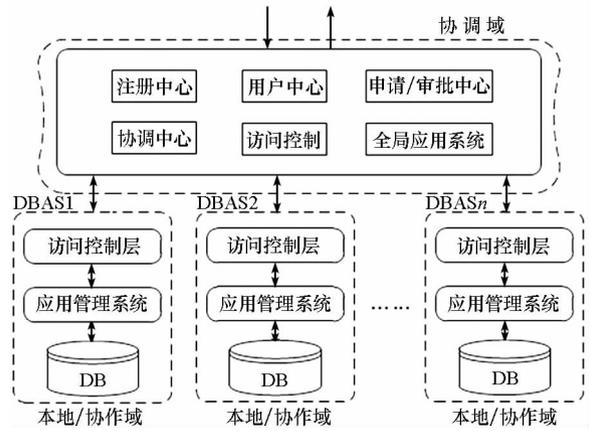


图6 简化集成的系统框架

Fig.6 System framework for simplifying integration

中。集成系统中的各数据源,只要拥有合法的本地域用户账号,即可申请跨域访问。在全局应用系统界面,用户可以浏览到所有集成的数据源(附有所包含系统功能和数据资源的简介),点击欲访问数据源的链接,即可申请跨域访问该数据源。当有跨域访问时,协调中心会首先验证该用户是否具有此访问权限,如果有则为用户进行跨域处理并提供转接访问资源的服务;否则,返回相应的提示信息给用户,并引导用户转向申请/审批中心,请用户先提交跨域申请,待审批通过后,再进行跨域访问。

以上所述为简化集成的通用框架。这种集成方式较适合于企业间的兼并、整合之类的情形,各数据源在集成系统中不仅能保留自己的系统(即在总部的授权下实现独立自主运营),还能实现互相访问(即各部门之间实现信息资源共享)。特别地,如果是作为一个子模块以进行功能扩展的方式来集成,则上述框架可以更为简化。

## 5 实例与分析

### 5.1 实例

实例描述:有一政府办公自动化系统,它是基于浏览器/服务器(Browser/Server, B/S)结构的DBAS,并已投入运行了一段时间,现客户要求增加一个“公文流转”功能。目前,已有一个开发完成的具有唯一公文流转功能的DBAS,它很符合客户的要求。问题在于,这两个DBAS是完全独立的系统,虽然都是B/S结构,且后台数据库也均采用MS SQL Server 2005,但二者之间是异构的数据库系统。

其实,这就是要在一个DBAS(称为主方)中嵌入或扩展一个功能模块(称为客方)。但是,客

方拥有自身独立的应用系统和数据库系统,它与主方之间是异构的。集成的需求在于客方的整个系统功能,包括数据库系统和应用系统;因此,这是面向应用系统层面的松散集成。如图 7 所示,为该实例的简化集成示意图。

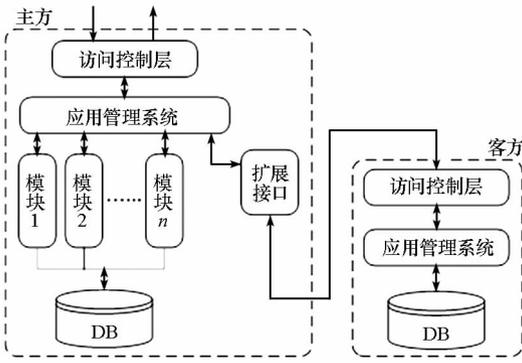


图 7 实例的简化集成示意图

Fig. 7 Example of simplified integrated schematic diagram

实现方案步骤为:

1) 由于主方(办公自动化系统)和客方(公文流转系统)的后台数据库均为 MS SQL Server 2005,因此先将客方的数据库导入主方的 DBMS 中。这样,主方的 DBMS 中就有两个数据库。

2) 编写一个触发器,对主方数据库中的权限表进行监控。一旦授权某个用户公文流转功能,就触发预先编写好的代码,即:在客方数据库的用户管理表中添加一条新记录,将该用户的用户名及密码等相关信息作为新记录相应字段的值。

3) 将主方应用系统的公文流转功能链接到客方应用系统的入口地址,并进行后台数据库转接处理。

4) 进行一些完善工作:①对主方的权限表还需要编写一个触发器,用于当收回某个用户的公文流转权限时,删除客方用户管理表中相应的用户信息;②主方中,具有公文流转权限的用户账号被“停用”或“删除”时,客方用户管理表中相应的用户信息也要做相应的处理,这也通过编写触发器的方式来实现;③系统界面相关信息的一致性,主要包括客方界面的一些标识信息要与主方保持一致,例如:系统名称、使用单位、开发商等信息标识;④界面设计风格的调整,对于界面风格要尽量在视觉上给人一种很自然的、融为一体的感觉。

这样一来,当具有公文流转权限的用户从主方系统点击公文流转功能时,就会自动地从主方系统跨域到客方系统,不仅前台界面链接到了客方的应用系统,而且后台也转接到了客方的数据库。此外,由于该用户在被授予公文流转权限时,

其用户账号及密码等相关信息已自动加入到客方数据库的用户管理表中,故而当该用户跨域进入客方 DBAS 时,身份验证通过,自然可以顺利地在客方系统中进行相关操作。反之,不具有公文流转权限的用户,因其在客方系统中没有相关身份信息,故身份验证将不会通过,从而被拒绝访问。这一切就好像是在同一个 DBAS 中发生的一样,用户完全感觉不到是在两个 DBAS 之间进行切换操作。

### 5.2 性能分析

这个实现方案只用了三个触发器、一个入口地址链接、少量的后台数据库转接处理和界面调整,就实现了两个异构 DBAS 的集成。通常情况下,这种松散集成仅需数个工作日便可完成,而传统的紧密集成方式通常需要数月才能完成。此外,新方法不仅在时间上具有绝对优势,而且在其他很多方面也都具有明显优势,见表 1。因此,与传统方法相比,新方法要远远优于传统的集成方法。

表 1 对比分析表

Tab. 1 Comparative analysis

对比项目名称	传统方法	新方法
集成工程时长	数月	数个工作日
需求分析	复杂	简易
集成模型	需建立复杂模型	无
人员需求	需求分析人员、系统架构师以及大量的开发人员	少量,且仅需一般开发人员即可
独立自主性	弱	强
扩展性	弱	强

上述集成案例是李春强等在实际工作当中遇到的真实案例,该集成系统已投入使用,且运转良好。通过这个实例可以看到,在 SIM 方法的实际应用中,根据具体情况很多功能部件都可以简化,从而最终可用比较简单的方法来实现系统集成。

### 6 结论

信息集成是一个充满活力的领域,它不仅由工程创新所驱动着,而且还有问题本身的演变<sup>[21]</sup>。针对异构 DBAS 的集成,主要面向应用系统层面,寻求其低成本、高利用率的快速集成解决方案。所采用的简化集成方法充分利用了 DBAS 的原有资源,并将访问控制与集成方法结合考虑,

无论是原数据源的应用系统,还是访问控制机制,全部充分利用,减少了浪费,降低了集成成本,大大简化了集成。这种集成方式可用于软件复用、协作开发等。当然,还有很多不足的地方尚需完善,如跨域用户的授权方式及其流程等方面。

## 参考文献 (References)

- [1] Wild C, Erdős J, Warmuth M, et al. Planned and ongoing projects (POP) database: development and results [J]. *International Journal of Technology Assessment in Health Care*, 2014, 30(5): 497–503.
- [2] 金澈清, 钱卫宁, 周敏奇, 等. 数据管理系统评测基准: 从传统数据库到新兴大数据[J]. *计算机学报*, 2015, 38(1): 18–34.  
JIN Cheqing, QIAN Weining, ZHOU Minqi, et al. Benchmarking data management systems: from traditional database to emergent big data [J]. *Chinese Journal of Computers*, 2015, 38(1): 18–34. (in Chinese)
- [3] Wei X M. Heterogeneous database integration middleware based on web services[J]. *Physics Procedia*, 2012, 24(Part B): 877–882.
- [4] Mohania M, Bhide M. New trends in information integration[C]// *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication*, ACM, 2008: 74–81.
- [5] Quix C, Jarke M. Information integration in research information systems[J]. *Procedia Computer Science*, 2014, 33(12): 18–24.
- [6] Naoual M, Abderrafiaa E. Semantic technologies applying to data warehouses federation [J]. *Journal of Emerging Technologies in Web Intelligence*, 2014, 6(1): 15–19.
- [7] Scott M, Boardman R P, Reed P A, et al. Managing heterogeneous datasets [J]. *Information Systems*, 2014, 44(8): 34–53.
- [8] Timko I, Dyreson C, Pedersen T B. A probabilistic data model and algebra for location-based data warehouses and their implementation [J]. *Geoinformatica*, 2014, 18(2): 357–403.
- [9] 王博, 郭波. 自治异构数据源聚集模型与算法研究[J]. *计算机研究与发展*, 2008, 45(9): 1546–1553.  
WANG Bo, GUO Bo. Study of aggregation process model and algorithms of autonomy heterogeneous data sources [J]. *Journal of Computer Research and Development*, 2008, 45(9): 1546–1553. (in Chinese)
- [10] Tomazela B, Hara C S, Ciferri R R, et al. Empowering integration processes with data provenance [J]. *Data & Knowledge Engineering*, 2013, 86(5): 102–123.
- [11] 李剑, 宋靖宇, 钟华. 基于本体的异构信息集成查询划分及转换[J]. *软件学报*, 2007, 18(10): 2495–2506.  
LI Jian, SONG Jingyu, ZHONG Hua. Ontology-based query division and reformulation for heterogeneous information integration [J]. *Journal of Software*, 2007, 18(10): 2495–2506. (in Chinese)
- [12] Wang Y P, Ge J H, Shao J P, et al. Research for data exchange technology of heterogeneous database based on XML [J]. *Manufacturing Automation Technology*, 2009, 392/394: 903–907.
- [13] Zhang Z Y, Wang H H, Zhang H. Research of heterogeneous database integration based on XML and JAVA technology[C]// *Proceedings of 2009 International Conference on E-Learning, E-Business, Enterprise Information Systems and E-Government*, 2009: 275–278.
- [14] Liu X H, Shi Y M, Xu Y B, et al. Heterogeneous database integration of EPR system based on OGSA-DAI [J]. *High Performance Computing and Applications*, 2010, 5938: 257–263.
- [15] Zhang Z Y, Gu W, Cao Z. Research and implementation the system integration on distributed heterogeneous database based on the Hibernate [C]// *Proceedings of 2011 International Conference on Computer Science and Network Technology (ICCSNT)*, 2012: 1537–1540.
- [16] 陈小武, 潘章晟, 赵沁平. 网格环境中模式复用的异构数据库访问和集成方法[J]. *软件学报*, 2006, 17(11): 2224–2233.  
CHEN Xiaowu, PAN Zhangsheng, ZHAO Qinpeng. A schema-reusable method on heterogenous databases access and integration in grid environment [J]. *Journal of Software*, 2006, 17(11): 2224–2233. (in Chinese)
- [17] Gao Z, Luo S W, Lin Y, et al. A grid-based integration model of heterogeneous database systems [C] // *Proceedings of International Conference on Information Technology and Computer Science*, 2009: 126–129.
- [18] Borkar V, Carey M, Engovatov D, et al. Access control in the aqualogic data services platform [C] // *Proceedings of ACM Sigmod/Pods Conference*, 2009: 939–946.
- [19] Leighton G, Barbosa D. Access control policy translation, verification, and minimization within heterogeneous data federations[J]. *ACM Transactions on Information and System Security*, 2011, 14(3): 25.
- [20] 王珊, 萨师焯. 数据库系统概论[M]. 4版. 北京: 高等教育出版社, 2006.  
WANG Shan, SA Shixuan. *Introduction to database system*[M]. 4th ed. Beijing: Higher Education Press, 2006. (in Chinese)
- [21] Bernstein P A, Haas L M. Information integration in the enterprise[J]. *Communications of the ACM*, 2008, 51(9): 72–79.