

# 一种多自治域层次互操作模型的系统框架\*

李瑞轩 卢正鼎 吴炜 文珠穆

(华中科技大学计算机科学与技术学院 武汉 430074)

E-mail: [rxli@public.wh.hb.cn](mailto:rxli@public.wh.hb.cn)

**摘要** 数据互操作是信息集成领域需要解决的关键问题。联邦数据库和多数据库是解决分布式异构环境中多个数据源的集成与互操作的两种方法,但它们各有优缺点。完全的集中不能保证局部系统的自治,全部的分散又不能保证系统之间的共享和互操作。在分析联邦数据库与多数据库特点与差异的基础上,提出了一种多自治域的层次互操作模型 MDHI,通过对 Panorama 多数据库系统进行扩展,给出了一种基于 Web Service 的 Panorama Web One 系统框架。这种框架既满足了局域范围内的集成和处理效率,又提供了一种集成广域范围内多种异构数据源的方法,更加符合当前实际应用的需要。

**关键词** 多数据库系统 联邦数据库系统 数据集成 互操作性 自治性

文章编号 BA8685

文献标识码 A

中图分类号 TP311.13

## An System Framework for Multi-Domain Hierarchical Interoperability Model

Li Ruixuan, Lu Zhengding, Wu Wei, Wen Zhumu

(College of Computer Science & Technology, Huazhong University of Science & Technology, Wuhan 430074)

**Abstract:** Data interoperability is one of the key issues of information integration. Federated database and multidatabase are the two methods to solve the problem of integration and interoperation among multiple data sources in distributed and heterogeneous environment. But they both have different relative merits. It can't ensure the autonomy of local systems when hiring complete centralization of management, while data sharing and interoperation can't be achieved when using total separation. This paper firstly analyzes the characteristic and differences of federated database and multidatabase and presents a multi-domain hierarchical interoperability (MDHI) model. It also gives a framework for Panorama Web One system based on Web Service by extending Panorama multidatabase system. This framework cannot only integrate and process data in local area efficiently, but also provide a method to integrate multiple heterogeneous data sources in wide area. It will satisfy the present real application requirements better.

**Keywords:** Multidatabase systems, Federated database systems, Data integration, Interoperability, Autonomy

### 1 引言

目前我们所处的计算环境,无论从硬件还是软件角度看,均是异构的。不同的硬件平台,不同的操作系统,各种各样的数据源分布在网络的各个角落。数据库技术的发展已导致了許多数据库系

---

基金项目: 国家“十五”科技攻关计划项目(项目编号: 2002BA103A04)

作者简介: 李瑞轩, 博士研究生, 主要研究方向为分布式异构系统集成、分布式系统安全。卢正鼎, 教授, 博导, 主要研究方向为分布式系统、数据库、信息安全。吴炜, 博士研究生, 主要研究方向为分布式异构系统集成。文珠穆, 博士研究生, 主要研究方向为信息系统安全。

统的出现，在一台大型计算机上可能存在着多种数据库系统，在计算机网络上也常有数以千计的数据库可供访问。多个数据库同时并存，甚至文件系统和数据库系统并存，而且各成员系统均有一定的自治性。计算机技术的快速发展，使得各种各样的软件层出不穷，基于这些软件的应用渗透到各个领域，它们在数据格式、数据处理、数据显示等方面各具特色。

随着应用需求的不断增加，越来越多的用户希望能够同时访问和处理来自多个软件系统中的数据，实现多个软件系统与不同数据源之间的互操作。一个理想的互操作系统是这样的，在分布异构环境中，不同的系统能彼此更安全地获取和处理对方的信息；用户能方便地搜索和查询所需的信息，并能方便地使用各种不同类型和格式的数据；信息管理者能很好地管理信息，为用户服务，并将资源充分地提供给用户<sup>[1]</sup>。通常，互操作性（Interoperability）是指不同计算机系统、网络、操作系统和应用程序一起工作并共享信息的能力，它强调将具有不同数据结构和数据格式的软件系统集成在一起协同工作。实际上，互操作在不同的情况下具有不同的侧重点，强调软件功能模块之间相互调用时称为软件互操作；强调数据集之间相互透明的访问时称为数据互操作；强调信息的共享在一定语义约束下的互操作称为语义互操作等。

数据互操作是信息集中需要解决的关键问题，它是指在分布异构环境中实现不同数据格式和数据结构的数据的共享与相互访问，包括不同文件系统相互访问、不同数据库中数据的访问以及 Web 文档的访问等。联邦数据库系统和多数据库系统是解决分布式异构环境中多个数据源的集成与互操作问题的两种方法，但它们各有优缺点<sup>[2]</sup>。本文在分析联邦数据库系统与多数据库系统特点和差异的基础上，提出了一种多自治域的层次互操作模型，通过对我们自行研制的多数据库系统 Panorama 进行扩展，给出了基于 Web Service 的 Panorama Web One 系统实现框架。

## 2 分布异构环境下的互操作系统

### 2.1 多数据库系统

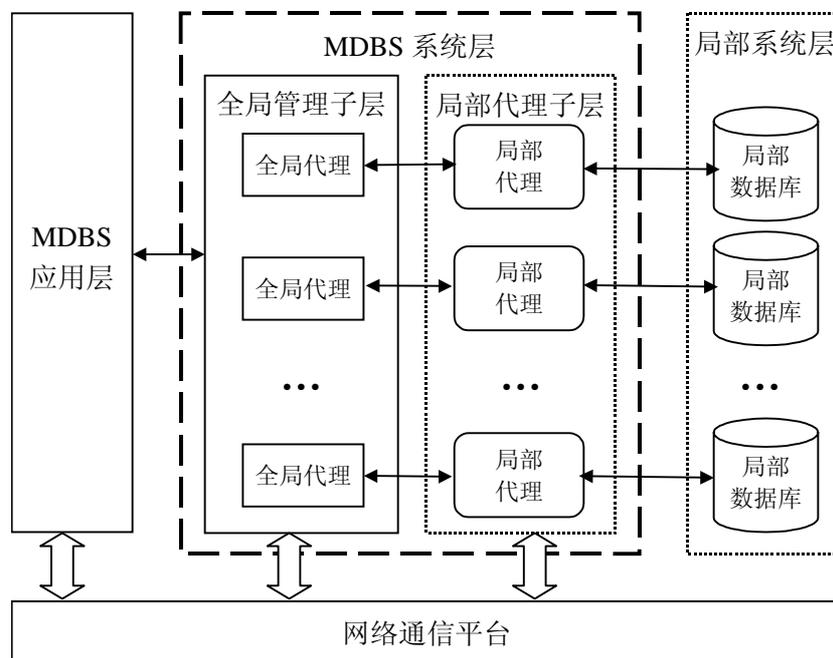


图 1 多数据库系统体系结构

多数据库系统是指对已经存在的多个异构数据库，在不影响其局部自治性的基础上，构造用户所需要的某种透明性的分布式管理系统，以支持对物理上分布的多个数据库的全局访问和数据库之间的互操作。一个多数据库系统（MDBS）是预先存在的、分布的、异构的、自治的多个数据库系统的联合<sup>[3]</sup>。通常，将组成一个多数据库系统的各个数据库称为成员数据库（CDBS）或局部数据库（LDBS）。各个局部数据库可以全部存在于同一场地，也可以分布在多个不同场地，它们可以是同

构的亦可以是异构的。多数据库系统的体系结构如图 1 所示。该体系结构共分为三层：多数据库应用层、多数据库系统层和局部系统层，其中多数据库系统层又分为全局管理子层和局部代理子层。多数据库应用层与全局管理子层之间、全局管理子层与局部代理子层之间通过网络通信平台或分布计算平台进行交互。局部代理与局部系统之间则通过数据库系统提供的各种接口进行通信，如调用级接口 CLI、ODBC 等。

多数据库系统提供了一种从系统和规模上来解决异种数据库集成与互操作的方法。它能够接受和容纳多个分布异构数据库系统，对外呈现出一种集成结构和统一视图，而对内又允许各个异构数据库的“自治性”。多数据库系统具有全局的统一管理，因此具有统一的全局模式。与传统的分布式数据库不同的是，传统的分布式数据库（DDB）属于全局逻辑集成方式，全局概念模式给出了整个数据库的概念视图；而多数据库系统属于松散集成方式，全局模式只表示每个局部数据库系统所共享的那部分数据的集合，各局部数据库的“私有”部分不参与多数据库系统全局管理。

## 2.2 联邦数据库系统

联邦数据库系统（FDBS）提供了另外一种组织、访问和更新共享信息的逻辑方式。和多数据库系统一样，联邦数据库系统也是多个自治的成员数据库系统（CDBS）的集合，同时实现它们之间的数据共享与互操作<sup>[4]</sup>。与多数据库系统不同的是，联邦数据库中不存在全局的中央控制中心，这样能从根本上保证数据的自治性，提高数据的私有和安全特性。在大多数情况下，FDBS 是异构的，各成员数据库系统可能有不同的数据模型、查询语言、并发控制、事务处理机制、安全策略等，另外，联邦数据库可能有自己的联邦数据模型，联邦数据库不能直接操作成员数据库。成员本身也可以是联邦数据库，一个数据库可以是多个联邦数据库系统的成员数据库。图 2 给出了一种联邦数据库系统的体系结构。图中共包括两个联邦数据库：FDBS1 和 FDBS2，其中 FDBS1 包含成员 CDBS1.1 和 CDBS1.2，FDBS2 包含成员 CDBS1.2 和 CDBS2.1，同时 CDBS2.1 又是由成员数据库 CDBS2.1.1 和 CDBS2.1.2 组成的联邦数据库。

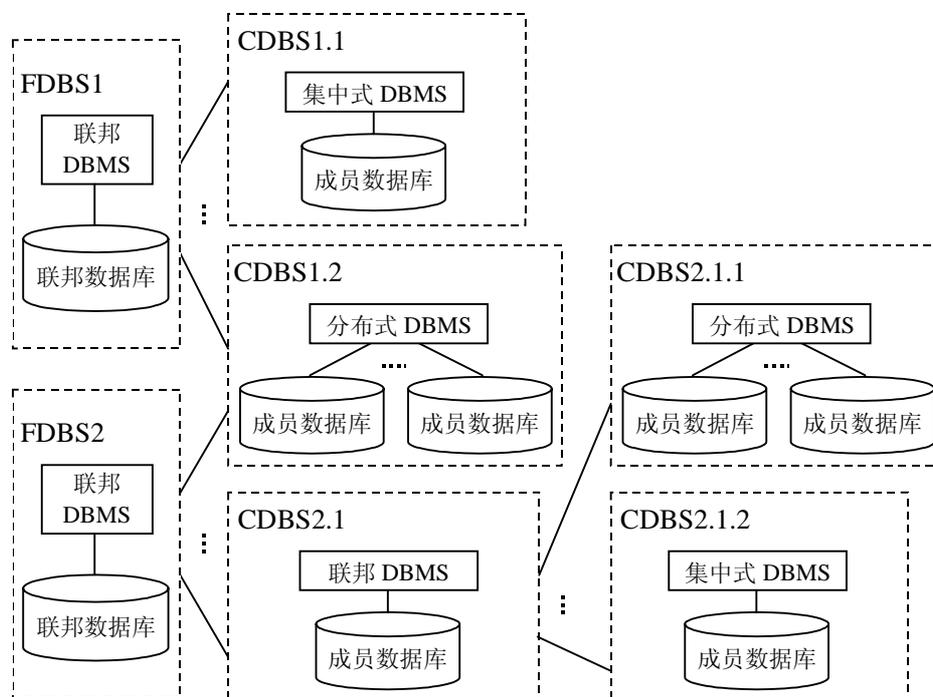


图 2 联邦数据库系统体系结构

## 3 多自治域的层次互操作模型

从前面的讨论中可以看出，多数据库系统和联邦数据库系统在组成结构和实现方法上没有本质

意义上的不同，主要的差别在于联邦数据库系统没有全局模式和中央控制中心，各组成数据库系统间的耦合更加松散，真正实用的联邦数据库管理系统在实现上更加困难，保证全局的安全事务处理几乎不可能。在某些情况下，如只包含查询语句、实现领导决策辅助功能，进行这方面的研究是很有必要和有意义的。全部采用多数据库的方式需要存在一个全局的系统管理层，各局部系统与全局交换信息，局部依附与全局，但同时又有很强的自治性，数据源之间不能直接交换信息，需要通过多数据库全局系统层这个中间桥梁，但在实际情况下，数据源之间的关系较为松散，有时不可能构造这样一个全局控制中心。因此，完全的集中，不能保证局部的自由；全部的分散，不能保证相互的共享和互操作。

这里给出的一种分布异构环境下的多自治域层次互操作（MDHI）模型，既结合多数据库系统的构架，在局部可控制的范围内实施集中；又兼顾联邦数据库系统的性质，在全局采用联邦的方式进行数据互操作，是一种混合互操作模型。MDHI 模型的相关定义如下。

**定义 1** 自治域  $AD_i = (LDBS, LA, GM, GDM, TCP)$ ，其中 LDBS 为自治域内成员系统的集合；LA 为局部代理的集合，负责全局与局部的信息交换；GM 为自治域的全局管理器；GDM 为自治域的全局数据模型；TCP 为自治域的事务提交协议。

自治域中的局部代理、全局管理器、全局数据模型和事务提交协议等按多数据库系统的方式实现，完成在自治域范围内的数据库互操作。

**定义 2** 互操作模型  $MDHI = (AD, FSA, FIEP, FAC)$ ，其中 AD 为所有自治域的集合；FSA 为安全代理集合，是自治域间交换信息的桥梁；FIEP 为联邦信息安全交换协议，在多代理协商语言的基础上定义自治域间交换信息的规则；FAC 为联邦安全访问控制协议，定义联邦系统的安全访问控制机制，解决各自治域安全协议的不一致。

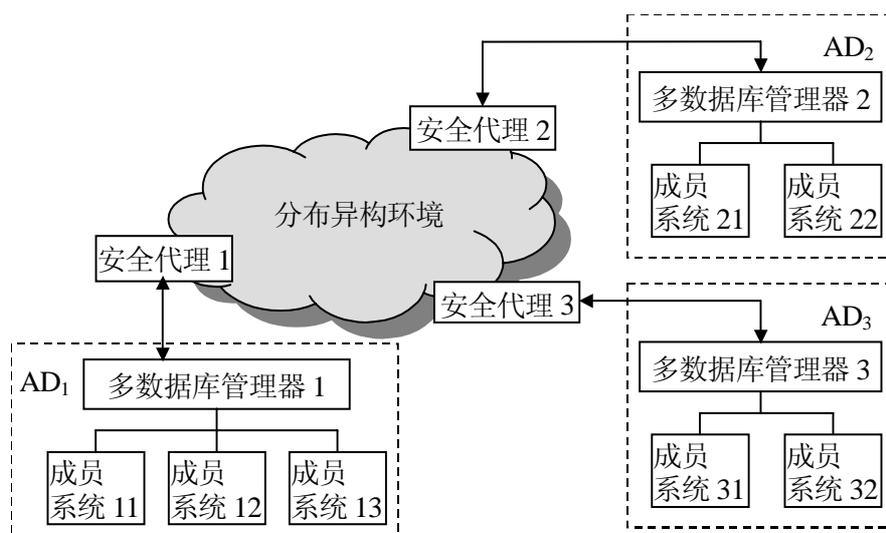


图 3 多自治域层次互操作模型结构图

MDHI 模型的体系结构如图 3 所示。其主要思想是将地理位置相近的成员系统（如同在一个单位或部门的多个系统）按多数据库的方式集中管理，建立全局统一的身份认证和访问控制机制；而将不能集中的成员系统按联邦数据库的方式进行管理，没有统一的全局模式，仅有公共的数据交换协议（如采用 SOAP 协议）。成员系统按域的方式进行组织。能集中管理的系统构成一个自治域，自治域内采用多数据库的方式进行数据的存取和互操作。自治域间则进行松耦合的管理，跨域的互操作可能需要进行数据模型、查询语言和安全机制等的转换及相互映射。

#### 4 基于 Web Services 的 Panorama 系统框架

我们自行研制的多数据库系统 Panorama 2003 实现了多种数据源（如 ORACLE、SYBASE、DB2、文件数据源等）的集成，为用户提供了一个透明访问多个数据源的公共接口<sup>[5][6]</sup>。通过在多个局部数据库系统的基础上构建一个虚拟数据库管理系统，解决不同数据源的异构性和数据的分布性。用户与

访问传统数据库系统一样，使用多数据库查询语言提交对全局模式的查询。Panorama 2003 采用 CORBA 作为底层通信平台，局部代理、查询管理、查询分解及结果合并等模块均以 CORBA 服务的方式封装，实现了网络访问的透明性。全局模式以 XML 方式来定义，提供一个模式管理工具来辅助解决不同模式间的冲突。除此之外，Panorama 2003 还将安全机制融入多数据源的集成中，实现了基于角色的访问控制和基于公钥的身份认证机制。Panorama 2003 实质上是位于应用系统和数据管理系统（如数据库系统、文件及 Web 系统等）之间的一个中间件系统，实现了分布异构系统的数据集成，为用户透明访问多种数据源的数据提供了一个统一的访问界面。

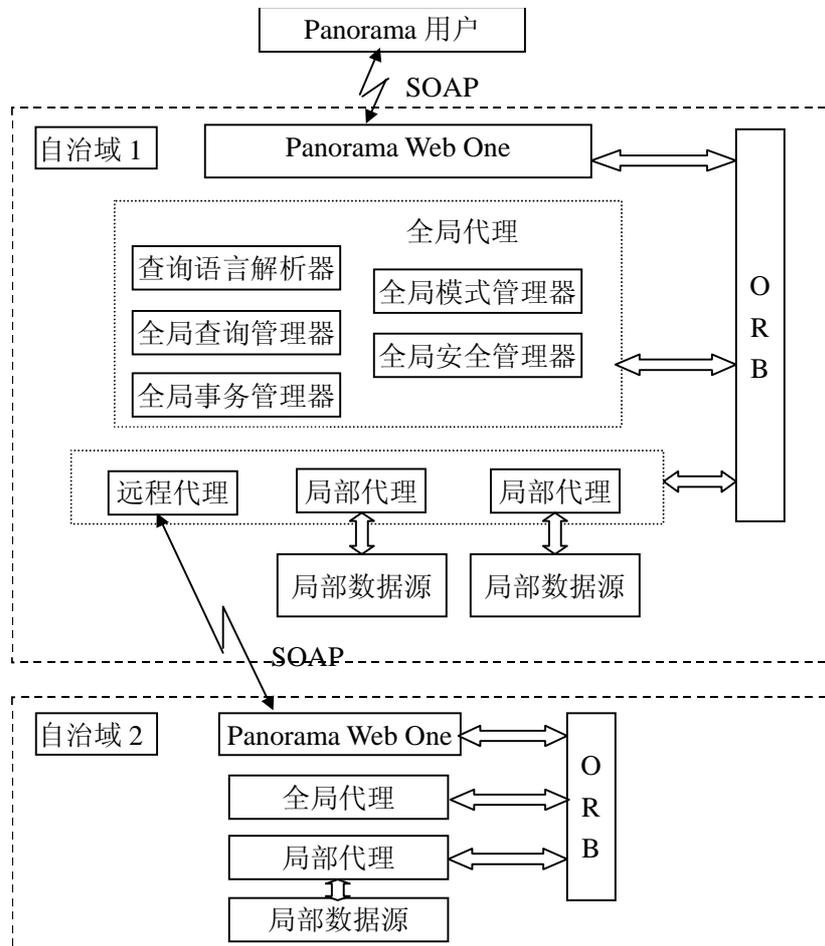


图 4 基于 Web Services 的 Panorama Web One 体系结构

Panorama 2003 系统主要对局域范围内的数据源实现集成，由于目前网络安全问题并没有得到很好的解决，Panorama 2003 系统对位于防火墙之后的广域范围的数据源和应用系统很难进行集成。而基于 HTTP 的简单对象访问协议 SOAP 则能很容易地透过局域网的防火墙，实现广域范围内的对象互操作。同时，Web Service 能对应用系统以组件的方式封装后提供服务，通过 Web Service 能实现应用系统的集成。因此，我们对 Panorama 2003 系统进行了扩展，提出了一种基于 SOAP 和 Web Service 的多自治域互操作系统——Panorama Web One。图 4 给出了 Panorama Web One 的体系结构。为了简洁起见，自治域 2 采用了简化的图示。

在 Panorama Web One 系统中，对局域范围内的数据源，由于集成和处理效率考虑，仍然按多数据库方式对共享数据进行集成，使用 CORBA 作为通信平台，并建立局域范围内的全局模式，进行统一管理。对于跨多个局域范围的数据共享与互操作，则由 Panorama Web One 将共享数据按多数据库系统服务的形式包装成 Web Service，为全局用户提供更好的服务方式，全局用户的服务请求及结果返回采用 SOAP 协议。同时，远程的共享数据源也包装成 Web Service，本地的远程代理通过 SOAP 协议与远程多数据库系统的 Web Service 进行交互，请求对远程数据源的访问。

## 5 总结

数据互操作是信息集成领域需要解决的关键问题。联邦数据库系统和多数据库系统是解决分布式异构环境中多个数据源的集成与互操作的两种方法，但这两者各有优缺点。完全的集中不能保证局部系统的自治，全部的分散又不能保证系统之间的相互共享和互操作。本文在分析联邦数据库系统与多数据库系统差异的基础上，提出了一种多自治域的层次互操作模型，通过对我们自行研制的多数据库系统 Panorama 进行扩展，给出了一种基于 Web Service 的 Panorama Web One 系统框架。这种框架既满足了局域范围内的集成和处理效率，又提供了一种集成广域范围内多种异构数据源的方法，更加符合当前实际应用的需要。在未来的工作中，我们将对多自治域层次互操作模型中的模式集成、数据包装、查询处理等方面进行更为深入的研究。

## 参 考 文 献

- [1] S. Dawson, S. Qian, P. Samarati. Providing Security and Interoperation of Heterogeneous Systems. Distributed and Parallel Databases, 2000, 8(1): 119~145
- [2] Ruixuan Li, Zhengding Lu, Weijun Xiao, Bing Li, Wei Wu. Schema Mapping for Interoperability in XML-based Multidatabase Systems. In Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA 2003), Prague, Czech Republic, published by IEEE Computer Society, September, 2003, 235~240
- [3] A. Elmagarmid, M. Rusinkiewicz, and A. Sheth. Management of Heterogeneous and Autonomous Database Systems. San Francisco: Morgan Kufmann publishers, 1999
- [4] A. P. Sheth, J. A. Larson. Federated Database System for Managing Distributed, Heterogeneous, and Autonomous Database. ACM Computing Surveys, 1990, 22(3): 183~236
- [5] 卢正鼎, 李兵, 肖卫军, 李瑞轩. 基于 CORBA/XML 的多数据库系统研究与实现. 计算机研究与发展, 2002, 39(4):443~449
- [6] Mudar Sarem, Ruixuan Li, Weijun Xiao, Zhengding Lu. The Interoperability of Different DBMSs Through a CORBA-based Multidatabase System. Wuhan University Journal of Natural Sciences, 2002, 7(4): 408~414