

在政务信息化中构建

XML 与关系数据库双重数据管理机制

文 / 王静 孙宏伟 · 国防科学技术工业委员会信息中心

数据集成是政务信息化建设过程中的关键。本文针对数据集成过程所面临的困难, 根据当前政务信息大都存储于关系型数据库而数据需要面向网络实现共享使用的现状, 提出构建基于 XML 与关系型数据库双重数据管理机制, 并给出构建该机制的可行方案。

一、引言

政务信息化, 是政府部门执行力和监管力的决定因素之一, 也是促进政务公开, 建设服务型政府的重要推动力之一。网络技术的发展极大地推动了政务信息化的进程, 同时也对传统的政务模式提出了挑战: 信息资源不再是某一个部门的内部资料; 对同一管理对象, 人事、政策、投资决策等各个部门需要具有相同的信息口径, 这就要求部门间的信息具有互通性, 也就是实现数据集成。

二、数据现状

政务建设在过去长时间的手工操作、分系统各自为政的过程中, 形成了大量的内部资源, 这些资源大多以人工管理方法、文件系统、各种数据库系统等形式存在, 它们共同构成了一个异构数据源。要将这些异构数据源集成在一起, 需要解决以下几方面的问题:

1. 异构性

异构性是政务信息化建设过程中实现数据集成所必须解决的首要问题。异构性主要表现为系统异构、模式异构和结构异构。

系统异构主要指数据源所依赖的应用系统、数据库管理系统乃至操作系统之间

的不同导致的数据异构。

模式异构主要指数据在存储模式上的不同, 一般的存储模式包括关系模式、对象模式、对象关系模式和文档嵌套模式等几种。其中关系模式为现在的主流存储模式, 需要注意的是, 即便是同一类存储模式, 它们的模式结构可能也存在着差异, 例如 Oracle 所采用的数据类型与 MS SQL Server 所采用的数据类型并不是完全一致的。

结构异构主要指应用系统采取的内部数据格式、数据接口标准的不同。

系统异构、模式异构以及结构异构使得系统间数据离线交换与在线共享无法直接进行。

2. 完整性

数据的完整性是数据集成的前提, 它可以方便数据的处理过程, 提高效率。数据集成的目的是为应用间数据交换与数据共享提供统一访问的支持。

集成后的数据所具有的完整性, 主要包括数据结构正确和数据约束条件完整两方面含义。数据结构正确是指正确地提取数据本身及其结构; 数据约束条件是指数据与数据之间的约束依赖关系, 用来表征数据间逻辑关联的特征, 在数据交换、数据集成、数据有效性验证等方面具有重要的意义。

3. 语义冲突

信息资源之间存在着语义上的区别, 这些语义上的不同可能引起各种矛盾。语义冲突有简单的名字语义冲突 (不同的名字代表相同的概念, 相同的名字代表不同的概念), 也有复杂的结构语义冲突 (不同的模型表达同样的信息)。语义冲突会带来数据集成结果的冗余, 干扰数据处理、发布和交换, 甚至导致错误的出现。所以如何正确处理语义冲突、减少语义冲突引起的错误也是数据集成需要解决的一个问题。

4. 权限瓶颈

由于数据资源归属不同的部门, 有些资源还具有较高的涉密性和敏感性, 所以如何在访问异构数据源数据时具有合适的、稳定的权限, 实现对数据源访问权限的控制, 就成为连接异构数据源必须解决的问题。

5. 附加约束

集成两个或多个数据源的时候, 数据源的数据之间可能存在着某种逻辑关系, 那么, 把这种逻辑关系附加到集成结果中的过程就称为附加约束。

6. 内容限定

多个数据源之间的数据集成, 并不是要将所有的数据进行集成, 那么, 如何定义要集成的内容范围, 就构成了集成内容的

限定问题。

三、构建双重数据管理机制

在现有的各种数据管理方式中，关系数据库 (RDB) 凭借其多级的安全机制、稳健的并发访问控制和高速的处理效率，在政府资源管理中占据了主导地位。此外，关系数据库具有的技术成熟度及应用广泛性也将使关系数据作为主导的数据存储手段而长时间存在。但是，关系数据库存在语义描述能力差的缺点，很难实现数据语义的持久性和传递性，而数据交换和信息共享都是在“语义一致”的前提下进行的。关系数据库客观上需要一种灵活的数据应用方式，以便将数据有效地发布到网络环境，消除平台差异、实现更大范围的数据交换与信息共享。

在网络技术发展过程中，XML (Extensible Markup Language, 可扩展标记语言) 作为一种支持网络环境的数据表达标准，引起了人们的极大关注。XML 的优点在于采用简单、柔性的标准化格式表达数据和在应用程序间交换数据方面；XML 的强

大之处在于将用户界面和结构化数据相分离，允许不同来源的数据无缝集成以及对同一数据的多种处理。上述优点和特点，使得 XML 可以为结构化数据、半结构化数据、关系数据库、对象数据库等多种数据源的数据内容加入标记，适于作为一种统一的数据描述工具，扮演异构应用间数据交换载体或多源异构数据集成全局模式的角色。事实上，XML 已经成为网络环境下数据表达的事实标准。

XML 在数据交换方面的自描述、可扩展性等优势使其适于充当网络环境下数据描述标准与异构应用间数据交换载体的角色，但 XML 缺乏像关系数据库那样有效的数据管理手段，因此存在着数据管理困难、操作效率低、安全性差的隐患。所以，客观上需要寻求一种有效的底层数据管理机制来代替或完善现行的文件系统。

综上所述，XML 和 RDB 在数据应用和数据管理方面各有优势，它们之间不能互相代替，应是一种相辅相成、优势互补的关系。在政务信息化建设中应当以关系数据库为底层数据存储手段，建立底层数据的

XML 视图作为应用间数据交换及直接网络应用通道，这就是双重数据管理机制的构想。

如图 1 所示，双重数据管理机制是指：作为数据存储与管理载体的底层关系数据库、作为系统间数据操作与直接网络应用通道的 XML 视图以及两者之间进行数据处理与协调的机制。在此构想中，政府的资源数据仍主要存放在 RDB 中，向下兼容并保留了安全成熟的 RDB 技术，不影响已有和局部电子政务系统的正常应用；而通过建立底层关系数据库的 XML 视图，使得实现基于 XML 的各种数据应用与通过 XML 的数据直接网络应用成为可能。外部访问用户（可以是人或程序）面向的是 XML 视图，从用户的角度来看，XML 视图和数据库之间的区别是透明的，而对用户而言，重要的不是数据存在何处，而是 XML 视图保证的数据逻辑结构合理性和操作界面方便性。这种双重数据管理机制实现了向下兼容传统的关系数据库技术，保护既有投资；向上采用 XML 在网络环境下进行数据交换，符合政务信息化发展要求。

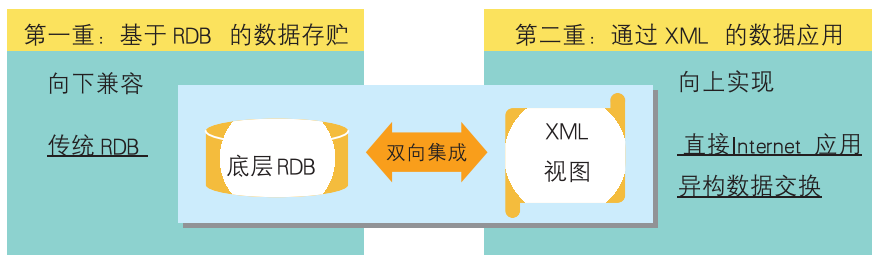


图1 双重数据管理机制示意图

四、双重数据管理机制的构建方式

建立双重数据管理机制，首先需要解决 XML 与关系数据库间的协调问题，这种协调包括三方面内容：

- 静态转换。通过关系模式与 XML Schema 之间的双向模式映射，实现 XML 与关系数据库的静态双向数据转换。

- 动态检索。处理用户通过 XML 视图对底层关系数据库的动态检索操作，将用户对 XML 视图的动态检索操作转换为对底层关系数据库的 SQL 检索操作，标记 SQL 执行结果，以 XML 格式返回给用户。

- 同步修改。处理用户通过 XML 视图对底层关系数据的修改，将用户对 XML 视图的修改操作转换为对底层关系数据库的修改，并建立这两个修改的同步执行机制。

在以上 XML 与关系数据库集成的基础上，可以进一步实现政务信息资源的数据集成，包括以下三个步骤：

- 建立一个基于 XML—Schema 的全局数据模式。该模式作为政务信息的全貌性逻辑描述，客户端需要共享其它系统数据时将访问该全局模式，通过检索与修改等数据操作，实现“可读可写”的数据共享。

- 建立全局模式与子模式的逻辑关系。全局数据模式是各子模式的逻辑组成，子模式与本地数据库具有一一对应关系。由于各子模式之间存在着数据重叠与冗余，如关于基层企事业单位，投资管理系统与财务系统采用的基础数据是不一致的，所

以全局模式不能是子模式的简单相加，需要对子模式进行分解，获得元子模式，相同部分进行合并，然后再整合成全局模式。同时，要建立全局模式与子模式之间的对应关系。

- 子模式与本地数据库之间以 XML 与关系数据库集成中间件。作为数据协调中间件，这种中间件不但可处理子模式与本地数据库之间的模式映射与数据静态转换，还能处理客户端通过全局模式对底层数据的检索、修改等操作，从而实现“可读可写”的数据共享。

五、结束语

本文论述的“以 RDB 为底层数据存储载体，通过 XML 视图进行数据交换”的双重数据管理机制，为信息化建设中的数据集成提供一种参考解决方案，并已成功应用于国防科工委的政务信息建设中。^{[3][4]}

参考文献：

- 1 Sun Hongwei, Zhang ShuSheng, Zhou Jingtao et al. Mapping XML—Schema to Relational Schema. Lecture Note in Computer Science, 2002, (2510): .322—329
- 2 孙宏伟. XML 与 RDB 的多层次双向数据集成技术研究与应用.[学位论文]. 西安: 西北工业大学, 2003

SinoEGov 资讯

河南：许昌启用电子政务内网应用系统

从9月1日起，许昌市电子政务内网应用系统将全面启用。许昌市政务系统将采取电子文件和纸质文件双轨运行的方式办公。从2007年1月1日起，许昌市将实现无纸化办公。

电子政务内网应用系统包括内网门户网站、公文交换系统、信息简报上报系统、OA（办公用软件）办公系统和公务邮件系统。该系统9月1日启用后，许昌市政务系统将实现公文网上及时传递和信息共享。从2007年1月1日起，许昌市将不再普发纸质文件、信息和简报等能够在政务内网上传递共享的纸质材料，实现无纸化办公。同时，许昌市将不断扩大电子政务应用的范围，力争两年内100%的行政审批项目实现网上申请、网上查询结果，50%以上的行政许可项目能够在线处理，100%的政府行政许可和非许可审批项目纳入电子监察范围，使电子政务真正成为便民、利民、服务于民的平台，成为党和政府与人民群众沟通互动的平台。

（来源：许昌日报）