

## 基于 ESB 的共享数据中心的研究与实现

林怀恭<sup>1</sup> 聂瑞华<sup>1,2</sup> 罗辉琼<sup>2</sup> 黄序鑫<sup>1</sup> 马 将<sup>1</sup>

<sup>1</sup>(华南师范大学计算机学院 广东 广州 510631)

<sup>2</sup>(华南师范大学网络中心 广东 广州 510631)

**摘 要** 针对数字化校园建设过程中如何保持应用系统间数据一致性的问题,在分析了几种实现数据同步的技术的基础上,提出一种基于企业服务总线 ESB(Enterprise Service Bus)的共享数据中心实现数据同步的方案,成功实现了数字化校园中多个系统间的数据同步,对于处理分布式系统间的数据同步问题具有一定的借鉴意义。

**关键词** 数字化校园 共享数据中心 数据同步 ESB

## ON ESB BASED DATA-SHARING CENTER AND ITS IMPLEMENTATION

Lin Huaigong<sup>1</sup> Nie Ruihua<sup>1,2</sup> Luo Huiqiong<sup>2</sup> Huang Xuxin<sup>1</sup> Ma Jiang<sup>1</sup>

<sup>1</sup>(School of Computer, South China Normal University, Guangzhou 510631, Guangdong, China)

<sup>2</sup>(Network Center, South China Normal University, Guangzhou 510631, Guangdong, China)

**Abstract** Aiming at keeping data consistency among application systems when constructing digital campus and on the basis of analyzing several data synchronization methods, in this paper we put forward a solution to synchronize data through a data-sharing center based on ESB (Enterprise Service Bus), which implemented the data synchronization among different application systems in digital campus successfully. This solution has reference meaning in dealing with the problem of keeping data synchronization among distributed systems.

**Keywords** Digital campus Data-sharing center Data synchronization ESB

## 0 引 言

信息化建设已经成为高校建设的重点。而在高校信息化建设过程中,由于在建设初期缺乏统一规划,应用系统大多都是从部门、基层的角度来开发实施,缺乏全局的考虑,使得开发出来的各部门的系统之间没有统一的数据标准,数据格式各不相同,系统之间无法实现有效的数据共享,同一类型的数据存在于多个系统中,系统之间无法实现数据同步,这样这些“信息孤岛”无论在使用上还是在维护上都会因为要实现数据的统一化增加一定的工作量,造成不必要的资源浪费。所以实现各个系统之间的自动数据同步,将各个系统间的信息真正整合起来,实现系统间的信息互动并保持数据的一致性,将是数字化校园建设的一个重点要解决的问题。

保持数据同步的一种做法是传统的 EAI 实现方式,这种实现方式需要各个系统对与之交互的系统都要提供相应的数据接口,开发工作量较大,并且各个应用开发厂商需要各自协调接口规范的讨论工作,接口依赖性较高,实现周期不可控<sup>[1]</sup>。另一种做法则是建立“共享数据库”,通过共享相关基础或重要数据,相关数据的改动首先反映到共享数据中心,再由共享数据中心通知其他应用系统作相应更新。采用这种方法,原有系统的改动可以降到最低,从而可以快速实施和产生效益<sup>[2]</sup>。本文将以南南师范大学数字化校园共享数据中心的建设为例,提出一种基于 ESB 的共享数据中心来解决数据同步问题的方案,解决分布式环境下数据同步的问题。

## 1 共享数据中心的的需求分析及设计原则

### 1.1 需求分析

构建共享数据中心,首先是要建立信息标准,新建的应用系统的数据都按统一的信息标准进行表达,方便后期系统之间数据的交互,方便校内数据与行业之间数据的流通。

共享数据中心的构建大致分为两个部分,分别为共享数据中心平台构建和共享数据中心的实施。共享数据中心平台的搭建与业务关系不大,主要是完成共享数据中心所需要的底层平台支撑,这包括对各种数据资源的元数据管理、业务数据的清洗处理、各种数据接口的处理、数据流的逻辑处理等;而在共享数据中心的实施阶段,就需要根据具体的需求,利用已经建设好的平台来建立各种业务逻辑,使得各个应用系统的数据可以真正在共享数据中心中流转起来,同时又可以在共享数据中心中提供各种综合数据服务,以完善业务上的需求。

综上所述,共享数据中心的构建需要经历三个阶段:平台构建阶段、数据构建阶段以及数据服务提供阶段。每个阶段都需要进行一定的开发和部署工作,而每一个阶段也会实现各种不同粒度的功能以供客户使用。其中在平台构建阶段,主要需要完成元数据定义、数据生成器以及数据导入处理等功能,为后期的数据构建打下基础。

收稿日期:2008-10-20。林怀恭,硕士生,主研领域:计算机网络与应用技术。

## 1.2 设计原则

(1) 系统耦合性低 由于共享数据中心需要与多个系统进行数据的对接,因此在设计上必须考虑耦合性的问题。多个系统在进行系统集成过程中不可能在短期内全部部署,而只能通过规划逐步与共享数据中心衔接上线,这样就需要在设计过程中尽量降低应用系统与共享数据中心间的耦合性,减少系统间的数据依赖。

(2) 可扩展性强 由于各个应用系统是逐步加入到共享数据中心的,因此对于其扩充性也需要作为重点考虑。共享数据中心必须能支持日后其他应用系统的交互数据整合,同时也可以支持对新加入的数据进行数据服务发布。

(3) 数据完整性 共享数据中心将采用与各个应用系统完全不同的数据模型,这种数据模型可以保证进入共享数据中心的数据完整性,以保证共享数据中心中所存储的数据的权威性。

(4) 可维护性好 共享数据中心的大部分维护工作都是在浏览器下进行,提供最大化的配置功能,为日后的系统更新以及系统异常纠错奠定基础。

(5) 数据安全性高 由于共享数据中心中的数据大部分是与业务相关的核心数据,并且也会涉及到个人隐私的数据,因此对于其安全性要求会很高。通过系统的多级权限控制以实现不同权限级别的人员只能访问其权限范围内的数据,以保证数据的保密性以及安全性。

(6) 保证数据传输性能 另外,由于共享数据中心的核心是数据服务,因此对于数据访问的性能要求也相当重要。共享数据中心必须能支撑多个应用系统的并发事务访问以及承载百万级的数据查询访问。对于未来业务的发展所带来的系统压力,也需要在设计时考虑进去。

## 2 共享数据中心的设计

### 2.1 设计思路

共享数据中心主要目的是将多个系统的数据进行整合,基本上可以分为两种整合方式:

(1) 通过数据集中的方式,将所有系统的核心数据以并集的形式存放到共享数据中心中。这种方式优点是可以完成系统级的整合,数据可以统一进行管理,并且系统安全上也得以加强。但同时也存在系统改造的工作量巨大的弊端。实现并集模式的共享数据中心需要对目前应用系统的数据进行并集整理,这需要对数据冲突的部分进行一定的取舍,同时还需要对目前的应用系统进行全面改造,将原有的核心数据操作都转移到共享数据中心上,而且还需要对各个系统中人员身份及权限进行统一,以保证数据操作的安全。

(2) 采用数据交集的方式,它是通过前期的系统调研,将各个应用系统间需要交互的数据进行整理,以完成共享数据模型的构建。系统间的数据交换完全通过共享数据中心来实现,应用系统只需要对共享部分的数据操作进行相应的系统改造就可以实现,并且权限模型可以沿用旧有系统的。这种实现模式可以减少共享数据中心的构建周期,对旧有系统的改造规模较小,而且应用系统与共享数据中心之间是松散耦合的,也就是说各个应用系统脱离了共享数据中心也可以正常运行。但交集模式不是共享数据中心的一个完整实现,它主要通过实现共享数据中心中的核心功能—数据交互来实现系统间的整合,因此交集

模式是共享数据中心并集模式的一个过渡阶段。

并集模式的共享数据中心的实现可以提供更丰富的数据汇总资源,其最终目的是实现多个系统的统一实现,实现难度较大、周期长。而交集模式的共享数据中心一般是对多个应用系统进行小规模改动,以实现系统间的数据共享。采用共享数据中心交集模式可以让各个开发厂商直接对共享数据中心接口,数据的转换等工作完全由共享数据中心来完成,这样大大降低了项目实施风险。同时,基于交集模式,可以实现对各个系统数据的获取,如果对其功能进行扩展,比如在一定的权限控制下获取更多应用系统的数据,逐渐形成共享数据中心的数据积累,通过这种方式亦可以逐渐构建起并集式共享数据中心,实现周期较短。

针对华南师范大学数字校园的现状分析,确定采用并集模式的共享数据中心作为建设方案。

### 2.2 技术方案

在基于共享数据中心来实现数据同步的方案中,由于原有系统建设所采用的技术和协议不尽相同,按照传统 EAI 的方式,共享数据中心必须要先建立与各系统联通的接口,通过这种方式可以实现系统间的数据同步。然而采用这种方式,系统间的接口紧密耦合,可扩展性差,建设的成本高,且可维护性较差<sup>[3]</sup>。

而采用基于 ESB 技术的方案则可以避免上述的缺陷。ESB 是服务间的连接框架,其核心功能包括消息转换、消息机制、基于内容的路由和服务容器四部分。ESB 采用基于 XML 规范的消息格式,可支持多种标准,如 Web Service、JMS、JCA 等标准,支持同步/异步消息机制,支持服务间的消息通信。ESB 使服务实现和服务通信相分离,服务之间松散耦合,基于标准协议建立通信联系,使应用系统具有更好的开放性和可扩展性<sup>[4]</sup>。

ESB 的上述特点使得它很适用于数字校园中数据中心与各个应用系统数据整合的场景。在华南师范大学数字校园共享数据建设中,共享数据中心的 ESB 通过开源技术 Mule 来实现,主要负责控制业务数据由接口进入,经过清洗、存储最后分发到其他应用系统的整个业务编排。通过在 Mule 中按照一定的业务处理规则进行路由设置以实现灵活的组合业务处理。

## 3 共享数据中心的实现

### 3.1 应用架构

本项目中,共享数据中心的实现是基于数据库技术而又构建在 J2EE 的架构之上,具体架构如图 1 所示。

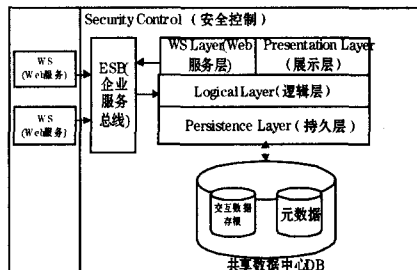


图1 共享数据中心应用架构

其中,Web 服务层为其他应用系统提供访问数据服务,包括对共享数据中心数据同步的一些操作;展示层则完成对共享数据中心的维护,这包括元数据的管理、系统日常维护等方面的功能。共享数据库中包含两部分的数据:交互数据的存根以及

共享数据中心元数据。交互数据存根主要是保证数据在各个系统间的准确性和唯一性,而共享数据中心元数据则是对共享数据中心模型的定义及管理,它保证了共享数据中心的正常运作。出于安全性的考虑,共享数据中心不提供对数据的查询修改等功能,避免造成数据异常,导致各个系统出现问题。

### 3.2 数据同步的实现

共享数据中心最主要的功能就是实现数据共享,即数据的同步和交换。共享数据中心将应用系统的数据复制到数据交换区,在对交换区的数据进行清洗和转换之后保存到相应的主题库中,然后再分发到各个应用系统,如图2所示。

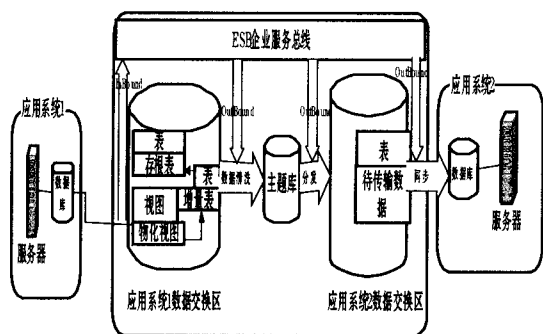


图2 共享数据中心结构图

#### 3.2.1 共享数据中心的组成

如图2所示,共享数据中心分为两个主要部分:数据交换区和主题库。

共享数据中心为每个应用系统创建一个数据库用户,这个用户所拥有的资源称为共享数据中心数据交换区。该用户下可能存在以下类型的对象:

(1) 物化视图 物化视图是远程数据的本地副本。在交互区中根据应用系统数据库需要导入的表创建相应的物化视图,并且根据需要在ESB总线中通过QuartZ的方式启动任务定时对这些视图进行刷新。

(2) 物化视图日志 如果需要进行快速刷新,则需要建立物化视图日志。物化视图日志根据不同物化视图的快速刷新的需要,可以建立为ROWID或PRIMARY KEY类型的。还可以选择是否包括SEQUENCE、INCLUDING NEW VALUES以及指定列的列表<sup>[5]</sup>。

(3) 增量表 进行数据清洗之前,会先根据物化视图日志查询出从上次清洗到当前时间为止发生变化的数据。然后将这些数据保存到增量表中等待进行清洗。

(4) 待传送数据 当应用系统的数据经过清洗后除了保存到主题库外,还要将相应的记录分发到其它应用系统对应的待传送数据表中,由传送任务将这些表的数据传送到相应的应用系统。

主题库是共享数据中心的核心区域,保存的是应用系统经过清洗后的数据,也就是最终的共享数据中心数据。

#### 3.2.2 数据导入和清洗步骤

共享数据中心的数据导入(也可称为数据复制)、清洗、分发这一系列动作是通过ESB总线来调度的,包括四个主要步骤:

(1) 复制应用系统数据 将应用系统新增的或发生了变化数据(即增量数据)导入到共享数据中心数据交换区的相应增量表中,这个步骤可以通过应用系统调用共享数据中心Web-

Service来完成,也可以通过共享数据中心的QuartZ启动定时任务从物化视图中获取。

(2) 数据清洗 数据清洗是将临时数据存储转入主题库的过程。此阶段则由ESB的UMO调用存储过程来实现。因此,该模块主要生成由临时数据存储转入中心数据存储的数据清洗存储过程。有两种方式启动数据清洗任务,一种是将清洗任务配置为OutBound,当通过WebService或QuartZ任务导入数据后,ESB总线立即激发相应的OutBound来进行数据清洗,在这个过程中数据清洗任务是被动执行的,对实时性要求比较高的数据一般采用此种方式;另外一种是将数据清洗任务配置在QuartZ中,将QuartZ任务作为InBound来定时启动。

(3) 将清洗后的数据从主题库分发到数据交换区 清洗后的数据一方面要更新到共享数据中心相应的主题库中,另外还要分发到需要这些数据的各应用系统的待传送数据表中。分发完毕后删除对应系统的增量表。

(4) 将数据交换区的数据传送到应用系统 将共享数据中心数据交换区的待传送数据更新到相应的应用系统的数据库中。

### 3.3 应用效果

基于ESB的共享数据中心满足了数据传输的实时性和共享数据中心与各应用系统之间的低耦合性和可扩展性的要求,达到了如下效果:

(1) 整理现有系统的数据,将存在于各个“信息孤岛”中的脏数据逐步清除,确保共享数据中心中数据的权威性。

(2) 通过共享数据中心完成与已有应用系统的数据对接,保证这些系统之间核心数据交互的正确性。

(3) 根据业务需求,为未来的应用系统提供数据支持服务,即新加入的系统仅需要通过数据中心提供的服务即可实现与现有系统的数据同步。

## 4 结束语

共享数据中心在数字化校园建设中占有十分重要的位置,它完成各个系统之间的数据同步,保证各个系统中所存在的数据的唯一性和有效性。本文在分析了几种数据同步方案的基础上,根据华南师范大学数字化建设的实际情况,提出了基于ESB的共享数据中心来实现系统间数据同步的方案,成功解决了新旧系统的数据同步问题,同时使得原有系统的改动最小,并具有很好的可扩展性,这对于处理类似的分布式系统间的数据同步问题有一定的参考价值。

## 参考文献

- [1] 郑曙光. EAI技术浅析(2004-10-11). <http://dev2dev.bea.com.cn/techdoc/200410115.html>.
- [2] 聂瑞华. 数字校园规划与建设方案. 华南师范大学数字校园建设办公室, 2006: 1-5.
- [3] 殷晓岚, 付远彬, 李京. 企业数据集成模式的研究[J]. 计算机工程与应用, 2002(12): 253-255.
- [4] 谢继晖, 白晓颖, 陈斌, 等. 企业服务总线研究综述[J]. 计算机科学, 2007, 34(11): 13-18.
- [5] 龚涛. Oracle 10g数据库管理[M]. 北京: 中国水利水电出版社, 2005: 285-295.