

基于 SOA 架构的高校数据资源整合研究

李学俭,何文华

(广东女子职业技术学院 网络中心,广东 广州 511450)

摘要:根据高校校园数字化建设的现状,分析了应用整合的重心在于数据资源的整合,提出了常见的数据整合解决方案的不足,结合实际应用建构了基于 SOA 架构数据整合方案。采用对比的方法,阐述了采用 SOA 架构实现数据资源整合的体系结构,将系统划分为业务系统层、共享服务层和整合业务应用层,其中共享服务层是数据资源整合方案的重点,采用共享库和存储转发两种模式实现数据资源的交换,通过企业服务总线(ESB)进行 Web 服务的注册与发现和 Web 服务之间的消息传输和安全管理等功能。通过数据资源的整合,可以统一学校全局信息标准,创建基于业务的数据主题,实现跨部门、跨应用系统的数据交换,同时集中存放的共享数据可以为学校后续应用系统的建设和领导决策提供数据支持。

关键词:数据整合;SOA 架构;企业服务总线;Web 服务

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2010)01-0078-04

Research of Campus Data Integration Based on SOA

LI Xue-jian, HE Wen-hua

(Network Center, Guangdong Women's Polytechnic College, Guangzhou 511450, China)

Abstract: According to the conditions of campus computing, data integration (DI) should come to be the emphasis in campus computing project. The faults of some familiar schemas on DI are put forward, and then a viable solution based on service oriented architecture (SOA) is discussed and implemented. By the method of antitheses, the implemented architecture is explained: the system includes three parts, i. e. business systems layer, shared services layer and integrated business application layer, and the second one is the emphasis on which data exchange is carried out by either of shared database mode and store-and-forward mode, and enterprise service bus (ESB) is used for registering web services, discovering web services, passing messages among web services and managing them. By the integration of data resources, information standards can be uniformed, data subjects based on business can be created, data exchange can be carried out, and moreover shared data can provide data to support the following applications and decision-making.

Key words: data integration; SOA; ESB; Web services

0 引言

高校数字化建设的目标是以数字化信息和网络为基础,架构一所“虚拟校园”,对教学、科研、管理、生活服务等信息进行整合和应用,拓展现实校园的时间和空间维度,提高传统校园的运行效率,扩展业务功能,实现校园的全面信息化。基于该目标,伴随着网络技术、存储技术及信息技术的发展和单位成本的下降,自 90 年代末以来各高校紧跟国外高校数字化建设的步伐,改造与升级了网络工程,建造了 G 级骨干网络,逐步构建了各种管理信息系统,如教务管理系统、学习管理系统(LMS)、人力资源管理系统(HR)、资产管理

系统,基本涵盖了高校各职能部门的业务范围,为简化应用系统增多给用户账号管理带来的混乱,都采用了统一身份验证(SSO)和门户系统(Portal)将各应用系统进行集成,这些都为校园数字化建设奠定了基础,也是校园数字化建设的第一阶段,即应用建设阶段。

随着 Web2.0 的产生和基于 B/S 开发模式的应用,新旧业务系统之间存在许多应用交叉,跨部门业务协作的需求越来越强烈,这给数字化建设带来了挑战。如何处理业务部门之间业务集成以及新旧系统业务衔接成为了校园数字化建设的新课题,是校园数字化建设的第二个阶段^[1,2],即应用系统整合阶段。美国“大学信息化校园”(Campus Computing Project)项目组近几年的调研报告^[3,4]以及国家教育部组织的创新与发展论坛,反映出应用整合是伴随着应用系统的建设和信息化本身的发展而发展,在这一阶段,出现了各种解决方案,如 API 接口, EAI, SOA 等。虽然实现技术差

收稿日期:2009-04-29;修回日期:2009-07-12

基金项目:广东省科技计划项目(2007B010200036)

作者简介:李学俭(1978-),男,江西九江人,讲师,硕士,研究方向为数字化校园建设、并行计算;何文华,硕士,副教授,研究方向为管理信息系统、高等教育管理。

异较大,但思路却非常一致,即通过这些技术的实施完成校园内应用系统之间的数据交换,使业务系统间形成基于底层数据的互联,达到业务集成的目的,即数据整合。

1 数据资源整合的目标与框架

数据资源的整合并不是简单地将一些数据从一个系统复制或转换地存储到另一个系统,而是需要从学校全局出发,采用统一信息标准,建设全局的共享数据中心,为业务系统提供全局的共享数据以及数据交换服务,并为基于业务系统的综合决策分析和扩展应用提供基础和可延伸的空间。

1.1 统一数据标准

在进行数据资源整合时,需要统一学校全局应用系统的信息标准。应依据国家相关信息标准以及教育部信息管理规范,将全局信息分为多个信息标准集(如教学管理),信息标准集下分信息子集(如教学管理信息标准集下分基本子集和教务子集),信息子集下分信息类(如基本子集中的学籍类)。显然,在进行实际的划分时,各子集之间以及新子集与现有业务信息表示之间肯定会存在不一致甚至矛盾的地方,此时可以遵照“下级从属上级”的原则予以解决,如在学校内部,有些部门内部用一些应用系统,当这些系统的信息描述与权威标准(包括国家相关信息标准、教育部信息标准以及学校自定义标准)冲突的时候,应以权威标准为准。

1.2 划分数据主题

统一信息标准是全局数据资源整合的指导原则和行动纲要,直接关系到具体实施时信息标准统一的成效。实施时,采用“分而治之”的思路,将基于校园现有应用系统的全局共享数据按主题的方式进行划分,形成与业务相关的信息主题。例如对学生而言可以形成学生的数据主题,教务系统中有学生的基本信息、学籍信息及成绩信息,财务系统有该学生的消费信息,应将这些信息归入到学生主题之中,以此类推,应将教师相关的基本信息、工资信息、异动信息、科研信息进行集成形成教师主题。为便于基于数据主题的应用,应根据应用的访问权限对相关表的操作进行封装,建立相应的主题对象^[5,6]。

1.3 数据整合的框架

进行数据资源整合需要实现业务系统的数据共享以及数据交换服务,通过创建共享数据中心实现,共享数据中心包含基于全局信息的共享数据库以及共享数

据交换组件,实现共享数据的集中存放以及转发,实现框架如图 1 所示。

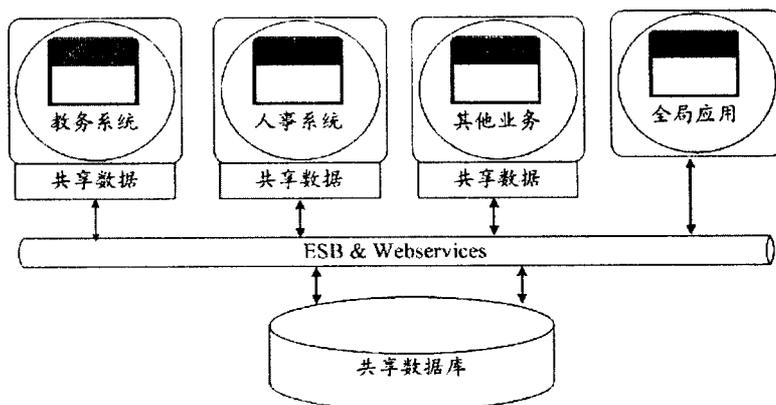


图 1 数据整合框架

所有整合的业务系统与共享数据中心进行关联,对于非关键的可共享业务系统,共享数据中心将共享的数据从源系统抽取,进行转换、清洗等操作后存贮到共享库中,通过触发共享数据中心的相关逻辑将该数据推向其他订阅该数据的业务系统;而对于一些关键业务系统(如财务系统、人事档案系统),可以通过共享数据中心的数据交换组件实现业务集成,将交换的数据通过数据通道(如图 1 所示的 ESB, Web services)在各系统间传递与交换^[7]。

2 数据资源整合的 SOA 架构

2.1 传统的校园数据整合技术

早期的数据整合主要解决的问题是点对点的应用对接问题,如新生入校报到时需要涉及到多个部门应用系统之间的信息交互,系统要由招生办的录取名单生成报到数据,财务和后勤管理系统要根据迎新报到情况进行收费和资源分配,教务系统要将报到数据进行学籍编排,图书管理系统要通过学籍数据制作借书证。为解决这类数据整合问题,传统的应用技术包括以下二种^[8-10]:

(1) API 接口。

原有系统有预留的 API 接口,或者采用集成开发商定制 API 接口,通过 API 接口,实现业务数据的导入/导出,解决点对点的系统集成问题。这种方案的不足主要表现在 API 的点对点集成难于管理,随着应用交互的增多 API 接口数的增加将成几何级增长,而原有系统的升级与改造往往导致“牵一发而动全身”,需要开发新的接口。

(2) EAI 集成。

EAI 是将基于异构平台下的业务应用系统集成在一起的一种技术。EAI 通过中间件作为粘合剂来连接

各种业务相关的异构系统、应用以及数据源,从而满足应用系统之间信息共享的需要。EAI 的基本原理在于源系统使用专有适配器转换消息格式,并将消息发送到消息总线上,进而流动到已集成的其他应用系统的专有适配器中,专有适配器进行比较以决定是否将消息翻译成符合其应用系统要求的格式。由于专有的适配器、中间层的消息转换规则和消息路由都是紧耦合的,所以在应用时不够灵活,可扩展性较差。

2.2 数据资源整合的 SOA 架构

SOA 应用于高校数据整合的基本思想是,将校园中各个应用程序的不同功能单元抽象为服务,或将数据封装成服务,通过服务之间定义良好的接口和契约联系起来,而不依赖于具体的硬件平台、操作系统和编程语言,使各个系统能够通过统一和通用的方式进行交互,实现灵活的可扩展性、重用性。

2.2.1 体系结构

图 2 是基于 SOA 解决方案的体系结构,共分为三个层次,即业务系统层、共享服务层和整合业务应用层,其中共享服务层是整个体系的核心层。

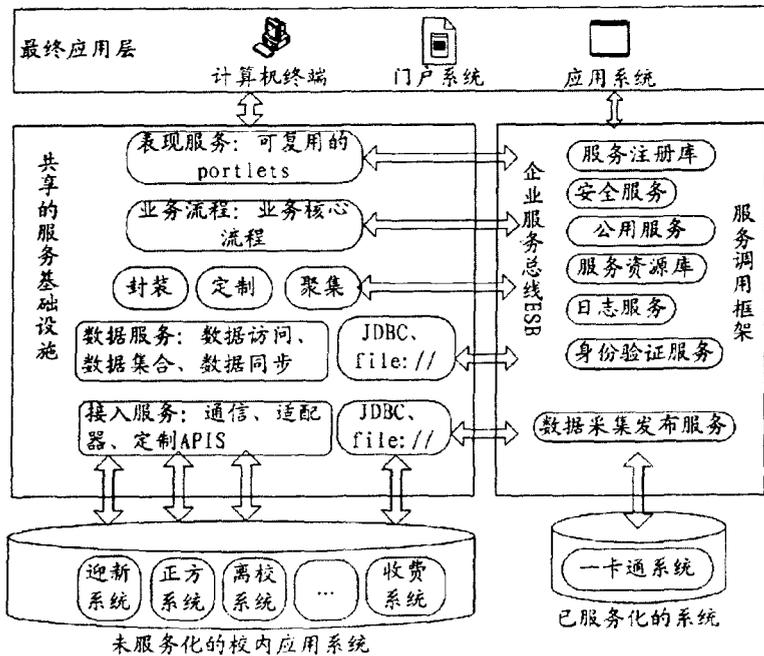


图 2 基于 SOA 数据整合解决方案的体系结构

(1) 业务系统层。

业务系统层处于整个体系的最底层,是校园内应用系统的集合,提供原始的共享数据来源,并有与其他应用系统交换数据的需求。

(2) 共享服务层。

共享服务层处于体系结构的中间层,是解决方案的核心层,包含接入服务、数据调用服务、数据封装服务业务流程以及企业服务总线(ESB)。在共享服务层

中,所有的服务调用均通过 ESB 进行。运行时,ESB 动态地通过服务目录的查询接口,通过查询服务的路由信息、服务的访问权限、服务的优先级、服务的版本信息等,决定如何进行服务调用。将共享服务层按服务管理、安全等因素细分为以下若干个模块:

* 服务接入模块。

服务接入模块采用如消息、适配器、API、数据连接等多种机制,连接到源数据库系统或文件系统,进行数据的输入与输出操作。

* 数据采集与发布模块。

系统将采集过程涉及的数据进行缓存,对于存储转发型的数据只做临时存储,而对于各系统共享的基础数据则保存至相应的共享库。

* 身份验证与权限管理模块。

所有服务统一部署并需要进行身份验证方可正常调用,为了更有效地控制数据的精确性,系统可以对相应的数据管理进行授权,只有取得授权后方可对相应数据进行维护。

* 错误处理与日志服务。

需要具有较强的容错机制,在调用服务过程中出错可以进行回滚,以恢复数据的有效性,通过系统日志记录操作过程中的重要运行信息,便于对各系统的数据情况进行统一布控。

* 其他模块。

包括消息模块、服务监控模块等,采用消息机制进行统一规范的数据交换协议并控制数据交换的整个过程,服务监控模块可以保证服务之间的安全性及完整性。

(3) 整合业务应用层。

整合业务应用层是基于数据整合后的综合应用,可以进行跨应用系统间的业务整合、综合信息查询以及全校的决策分析。

2.2.2 数据交换组件

数据交换组件处于共享服务层,实现共享数据的存储与转发功能,其运行机制如图 3 所示。

数据交换组件连接业务系统 1、业务系统 n 等,类似于交换机,实现应用系统间的数据交换时,有两种模式,一种是采集与交换的时候以共享库为中介进行保存,另一种是仅转发数据。

(1) 基于共享库的数据交换。

基于共享数据库的数据交换双方,采用 ETL(数据抽取、转换与抽取)方式实现。当源数据变化时,触

发 ETL 进行相关操作,将数据抽取到共享库进行保存。然后再触发共享库的执行逻辑,将变更的数据传递到受影响的目标系统中。

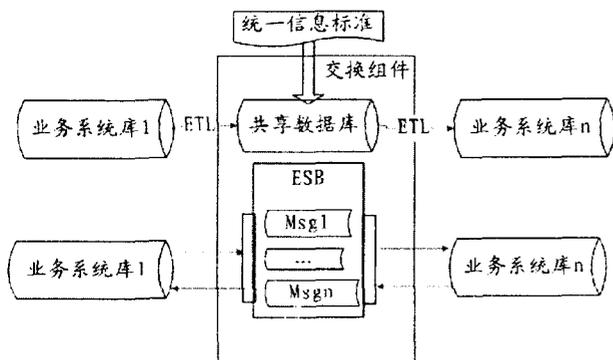


图3 数据交换组件

(2) 基于转发模式的数据交换。

基于转发模式的数据交换双方采用消息机制实现。当源数据变化时,触发消息通知服务执行,ESB 将消息进行传递与转发,到达目标系统后反馈消息给源系统。源系统将变更的数据取出并封装成消息或 XML 文件,形成数据流,数据流经同样的步骤流向目标系统,从而完成数据的传递与交换过程。

在这种模式下,数据交换组件须完成过程管理的辅助功能^[11-12],包括消息流与数据流的管理、消息的安全性与完整性管理、服务的回滚等。

3 分析

采用 SOA 架构进行数据资源整合时,需要解决好两方面的问题:一是整合过程中实施细节问题,二是整合后的应用拓展问题。只有这两方面的问题解决了,数据资源整合整体解决方案才能在高校数字化建设中发挥真正作用。

3.1 整合过程的技术细节

为实现数据资源整合的目标,在实施基于 SOA 的解决方案时候要管理好元数据以及对数据服务的封装。元数据的管理是全局统一信息标准的集中体现,对全局信息标准数据字典和代码集进行维护,对表描述或表所含数据项描述及数据结构可以进行增加、删除、修改、导出存档和查询,是对于校园所有信息系统元数据的归并与整理,为后续新建系统提供指导与规范;为便于对源数据的访问以及附带基于数据流之上的业务逻辑,需要将数据封装成数据服务,在实现时可以采用定制 Web service,也可以采用专用 ETL 工具将源数据进行抽取后通过专用接口发布成服务。

3.2 整合后的应用拓展

数据资源整合后,学校信息中心要与其他业务部

门通力合作,制定基于共享数据中心的全局应用,如共享数据中心在解决新生报到的数据交互问题时,可以通过制定新生报到主题库,实时将新生在业务系统中的数据变更情况进行通知与发布,其他相关部门便可联动实现完整的报到流程。除此以外,数据资源整合后形成的共享数据库,涵盖了各种业务数据,基于这些数据可以进行数据挖掘,从多维度进行数据分析,为学校决策提供支持。

参考文献:

- [1] 2007 年中国教育信息化创新与发展论坛. 教育信息化发展现状与展望[EB/OL]. 2007-06. <http://www.e-chinaedu.cn/zhuanti/index.html>.
- [2] 2008 年中国教育信息化创新与发展论坛. 上海交通大学信息资源综合服务平台的建设与应用[EB/OL]. 2008-10. <http://www.e-chinaedu.cn/zhuanti/2008/luntan/gaodengluntan.asp>.
- [3] The Campus Computing Project. The 2008 National Survey of Information Technology in U.S. Higher Education: Campuses Invest in Emergency Notification [EB/OL]. 2008-10. <http://www.campuscomputing.net/content-item/new-2008-campus-computing-survey-executive-summary-29-oct-2008>.
- [4] The Campus Computing Project. The 2007 National Survey of Information Technology in U.S. Higher Education: IT Security and Crisis Management Pose Continuing Challenges[EB/OL]. 2007-10. <http://www.campuscomputing.net/survey-summary/2007-campus-computing-survey>.
- [5] 许鑫,苏新宁,吴乃冈. 高校共享数据中心平台的设计与实现[J]. 现代图书情报技术, 2005(6): 48-53.
- [6] 韩海航,张永智. 数据交换与共享技术在交通行业数据资源整合中的应用研究[J]. 计算机应用与软件, 2007, 24(9): 109-112.
- [7] 李卫忠. IBM 详谈 ESB 部署方法论并发布 ESB 实用新品. ESB: SOA 的信息传输高速公路[J]. 微型机与应用, 2008(4): 92-93.
- [8] 刘翔. 中小企业数据交换平台的研究与设计[D]. 武汉: 华中科技大学, 2006.
- [9] 张钦德. 一个面向 EAI 的数据交换平台的设计与实现[D]. 上海: 上海交通大学, 2007.
- [10] 曾一,袁纲,张元平,等. 基于 web 服务的电子政务数据交换中心的设计与实现[J]. 计算机科学, 2007, 34(11): 98-101.
- [11] 梅立军,付小龙,刘启新,等. 基于 SOA 的数据交换平台研究与实现[J]. 计算机工程与设计, 2006, 27(19): 3601-3603.
- [12] 谢晓祥. 基于 SOA 架构的学校应用整合[D]. 上海: 华东师范大学, 2007.