

国家开放大学公共数据平台构建思考

孙雨 鲍赞力

(国家开放大学 信息化部 北京 100039 国家开放大学 教务部 北京 100039)

摘要: 国家开放大学不断探索技术在教育中的应用,经过多年的信息化建设,信息化在其教育教学中起到重要作用。但由于在建设过程中历史的局限性,逐步建立起来的各应用系统在数据编码、共享、统一等方面缺乏长远考虑和统一规划,造成应用系统之间数据难以互联互通,出现“信息孤岛”现象,在一定程度上阻碍学校信息化应用教育的效果。针对国家开放大学信息化建设现状及问题,梳理数据共享的需求,设计公共数据库建设技术架构,力图从数据层打通各应用系统,实现应用系统之间的互联互通。

关键词: 国家开放大学;数据编码;公共数据平台;数据共享

为贯彻落实《国家中长期教育改革和发展规划纲要(2010-2020年)》提出的“办好开放大学”,2010年10月,国务院印发《关于开展国家教育体制改革试点的通知》(国办发〔2010〕48号)。2012年7月31日,国家开放大学在人民大会堂举行成立揭牌仪式,国务委员刘延东同志出席会议并做重要讲话,指出开放大学在教育服务国家发展、提升国际竞争力方面的重要作用,是我国实现终身教育体系、形成学习型社会的重要举措,是实现满足多样化的学习需求、促进教育公平的重要途径^[1]。

国家开放大学(以下简称国开)是一所以技术和教育深度融合的新型大学,其自从成立之初不断致力于信息化的建设,力图以信息化引领教育现代化,探索构建了以网络学习空间、网络核心课程、网络教学团队、网络学习支持、网络考试测评、网络教学管理为主的“六网融通”教学模式,以及“云、路、端”的信息化顶层架构^[2]。

一、国家开放大学信息化现状与问题

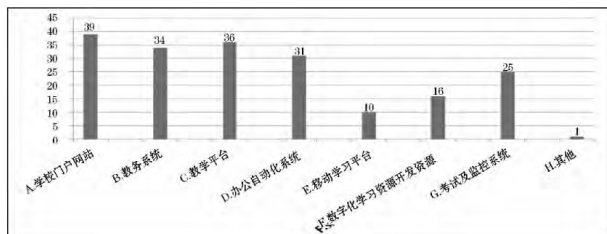
国开由总部、分部、学院和学习中心组成,实施总部和分部两级办学,信息系统建设、管理和技术运维服务主要集中在总部和分部,学院和学习中心主要是应用总部和分部的信息化平台^[3]。为了解开放大学系统教育信息化建设现状,配合高等教育信息化发展报告,针对开放大学分部以网站问卷调查的形式进行调研,调查问卷主要涉及信息化规划、信息化基础设施、应用系统、教学应用、保障体系等方面,共发放52份“开放大学教育信息化现状调查问卷”,收到39份分部的问卷。调查问卷按照填报问卷地区划分,其中东部沿海分部20家,西部边远学校10家,中部学校9家^[4]。

被调研的39家分部均建立了学校门户网站,绝大多数基于国开总部提供的教务平台和教学平台二次开发形成各自相应的平台,极少数分部直接使用国开总部的教学和教务平台(如图一所示)。各

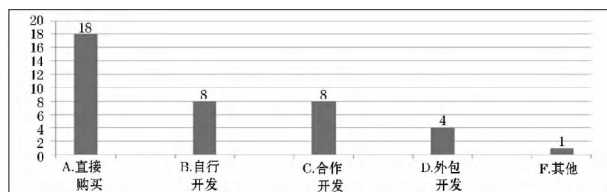
收稿日期 2017-05-03

作者简介 孙雨,女(汉族),硕士,助理研究员,研究方向 教育信息化,信息与信息处理。

分部各种业务系统研发模式主要的建设方式是直接购买,占被调查分部的46%(如图二所示)。其中,64%分部的数据库建设采取分散数据库、分散管理的模式,主要采取SqlServer数据库,其次是Mysql,最后是Oracle。



图一 应用系统建设情况



图二 应用系统建设模式

(一)系统之间难以互联互通 信息孤岛现象严重

应用系统由不同业务部门在不同历史时期根据不同需求的背景下构建,各系统采用不同的技术路线和数据库,系统之间的数据很难实现互联互通,随着多年的演变、迭代和系统重构,在客观上形成了一个信息孤岛,各系统之间的数据共享和业务协同难以自动化实现共享,需要人工进行数据的导入导出,数据容易出错。

(二)数据重复存储,一致性难以维护

信息化发展是一个不断完善和发展的过程,国开2000年左右开始探索信息化建设,起初应用系统的建设多采用业务部门自行规划、单独建设的模式,在考虑与其他应用系统的对接方面较为欠缺,导致数据的重复采集和存储,一方面造成投资浪费。另一方面,同一份数据在物理上存在于多个数据库之中,从而造成数据的不一致,给业务带来严重风险。

(三)数据利用不足,大数据没有真正带来价值

大数据应用于教育不仅驱动开放教育创新发展,而且可以成为教育与信息技术融合的抓手^[5]。国

开在开展开放教育过程中,多年来积累了海量的开放教育业务数据,有近千万的学生数据、上万门课程数据、PB级的远程教育资源数据,但这些数据分散在各类业务部门、各系统,难以实现对历史业务数据进行深度分析和挖掘。

(四)缺乏统一的数据规范

由于应用系统建设学校各业务部门主导完成需求和设计工作,在数据设计过程中缺乏学校专业、课程、教材、资源等方面的统一的编码标准规范,导致数据在不同系统中存在不同的定义和表示,给系统的使用和数据共享交换造成了障碍,也给新旧系统之间的数据迁移和割接带来极大的挑战。

二、信息化数据共享需求分析

目前,国开各级学校、各类业务部门对基础数据平台的建设提出了以下几方面的需求:

(一)信息共享需求

信息共享是业务协同的基础,通过调研发现,教务、资源等业务部门对以下信息提出了迫切的共享需求:专业信息、课程设置及开设信息、资源建设信息、学生信息、考试信息、教师信息、考试形式信息、各分部联系方式信息、教师联系信息、教材使用情况信息等,尤其是学生信息数据信息的共享需求尤为迫切。

(二)数据交换需求

教务、教学、学习支持、考试、资源建设与管理等各业务部门之间存在较强的业务关联和业务协同的需求,系统之间物理隔离,自动的数据共享与交换,不仅可以提高工作效率,更重要的是保证数据的一致性和准确性。

(三)信息安全保障需求

对信息使用安全方面的保护,主要通过人工限制信息的使用对象、限制信息数据项等方式进行,缺乏数据的安全审计。在制度层面,需要建立信息集中审核发布流程,确保信息使用的安全性。公共数据平台的数据采集来源于不同部门的应用系统

数据库,数据的修改、删除等操作权限不仅公共数据平台管理人员具有,而且原有系统数据库管理人员仍然具有相应的权限。因此,需要规避和考虑多人具有数据的修改、删除权限导致数据失真、误删等安全风险的问题。

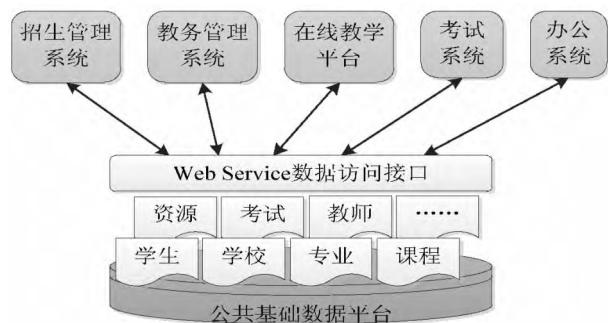
(四)数据深度分析利用需求

通过建立数据分析模型,对历年积累形成的海量开放教育数据进行深层次的分析挖掘,从而发现数据背后的问题和规律,为远程教育科研、教育战略规划与调整、专业设置、课程配置等提供决策依据。为决策层提供辅助的信息化支持。

三、公共数据平台设计思路

(一)系统架构规划

建设一体化的教务教学业务应用系统,功能涵盖远程在线教学、教务管理、在线考试、教育资源管理、日常办公等国家开放大学的核心业务。业务应用系统的建设将基于公共数据平台,国开分部、学院、学习中心、教学点以及广大的学生用户通过网络远程登录一体化业务系统,使用系统提供的服务。基于公共数据平台的学历教育业务应用系统架构如图三所示。公共数据平台统一管理学校、学生、专业、课程等业务数据,数据对象在公共数据平台上的存储具有唯一性,从而保证了数据在各业务系统之间的一致性。此外,公共数据平台向应用提供Web Service访问接口,通过将数据管理与数据操作相分离,使得系统具有良好的扩展性。



图三 公共数据平台关系图

(二)数据编码规范

建立统一的数据编码标准规范、编码管理流程

以及编码管理辅助平台,保证国开数据一致性,实现广泛的资源共享和数据交换。编码规范应遵循以下原则:

1.识别唯一性:同一个数据编码要具有全局唯一性,以明确区分各业务对象实体。

2.特征一致性:随着远程教育业务不断发展和扩大,数据编码在不同的系统中重叠、交叉现象将极为常见。数据编码的特征经常被用作业务流程的判断条件和数据分析的具体维度层次,保证数据编码的关键特征在不同应用、不同系统中的高度一致性是实现应用整合、数据整合及数据综合利用的必要条件^[6]。

3.编码稳定性:数据编码在业务处理和流转过程中其识别信息和关键的特征会被处理过程中产生的数据所继承、引用、复制,除非数据编码本身的特征发生改变,其本身属性通常不会随着业务处理的过程而修改。

4.长期有效性:数据编码有效性贯穿业务对象在业务中的整个生命周期,有时根据实际需要会更长。在数据编码失去效用时,系统通常是标记无效或标记删除而非直接物理删除。只有在对数据进行归档时,才会将数据编码信息从系统中彻底清除。

(三)平台技术需求

公共数据平台将承载国开在学历继续教育方面的全部业务数据,以及体系内各级各类教学机构在非学历继续教育方面的共享业务数据,应实现对类型多样的海量数据的有效管理、高效增删改查、关联分析与深度挖掘等功能支持,其设计应紧密结合数据与应用的特征:

1.平台需要具有较高的扩展性和并发访问支持能力:由于数据量大,用户访问量,要求公共数据平台具有较高的扩展性。在学历继续教育方面,公共数据平台将支撑全国44个分部、上千所学院、上万个教学点以及360万名学生对教务、教学、考试、招生、毕业、教育资源等PB级的海量数据的存储管理与十万级别的并发数据访问需求;

2.多种类型的数据并存:系统中既有传统的结构化数据,如专业规则、课程计划、学生、学校等信息;也有非结构化数据,如教学资源包括文本、音频、视频、图片等等。非结构化数据具有增长快、存储需求大的特征,因此要求平台能够有效地管理异构的数据。

3.数据之间强关联性:很多业务数据之间往往不是孤立存在而是有一定的关联甚至是依存关系。对于具有关联关系的数据,为满足数据对存储容量和访问性能的要求,采用分布式关系型数据库管理系统。

4.数据综合分析:随着数据的逐步整合和汇聚,平台将积累海量的学生信息、学习过程记录、教育资源等数据,通过对数据进行多维度的综合分析与挖掘,掌握远程开放教育的规律,进一步创新教学模式,把握全国开放教育的业务发展状况,为国开总部制定专业规则、课程设计、资源制作提供依据,为领导决策提供数据支持。平台应采用数据仓库技术,提供对海量异构数据的高效分析挖掘功能与接口,促进基于大数据的应用系统的开发,真正体现和利用大数据的价值。

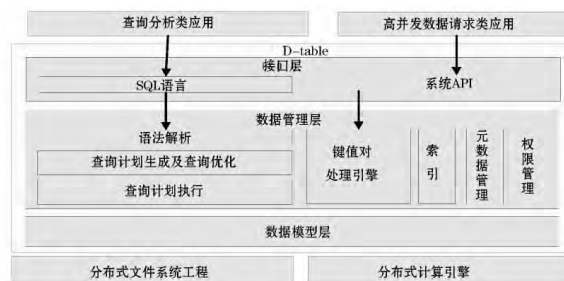
5.数据共享交换:国开各级各类教学机构根据自身的业务需要,建设大量的特色教育资源,促进国开教育事业的发展,在各级教学机构之间实现资源共享和数据交换,以促进优质教育资源共享和教育公平。为此,建设统一的数据共享交换系统,向各级各类教学机构提供信息资源目录服务和数据交换服务。数据共享交换系统应提供基于 Web Service 的访问接口,便于实现远程的数据交换,在分部、学院等下级教学机构部署数据交换代理,用以实现数据格式转换和基于 Web Service 的数据推送与同步。

6.应用协同:各应用系统之间的协同主要体现在各种数据的共享和交换上,如教务管理系统需要将注册并选课学生的信息及时同步给学习平台,以保证学习者可以访问相应的学习资源。保证应用系统之间数据共享和交换的实时性与一致性是实现

高效应用协同的关键。

(四)平台总体架构

公共数据平台的总体架构如下图所示:



图四 公共数据平台架构图

平台自底向上,共由三层模块构成,包括数据模型层、数据管理层以及接口层。其中,每个层次所负责的功能如下所述:

1.数据模型层

数据模型层负责数据存储和访问服务。将国开各种结构化数据、半结构化数据、非结构化数据等多种形式数据经过建模后,保存在分布式文件系统上。利用分布式文件系统高扩展性和容错能力,解决数据存储的规模以及数据可用性的问题。同时,系统还通过多维度的索引机制,向用户提供针对结构化数据或者非结构化数据在多个维度上的快速访问能力。

2.数据管理层

数据管理层负责将查询、索引等数据请求以及数据的权限管理,提供完整的数据服务。高并发数据请求由“键值对数据处理引擎”进行处理,处理后的数据直接存储到结构化存储中。这样,查询分析类应用就可以直接对存储在结构化存储中的数据进行分析类查询,避免了数据从在线系统导入离线系统的额外开销。

3.接口层

接口层负责将数据管理层的数据服务,以接口的方式提供给外部模块使用。对于查询分析类的应用,系统提供类 SQL 的查询语言,以方便外部用户的使用,而对于高并发数据请求类应用,则以系统 API 的方式,提供给外部用户对数据进行增/删/改/

查的操作接口。

四、技术难题思考

(一)系统选型与定制开发

根据国开对业务数据的分类情况,原有系统尤其是学习平台和教务系统,分别采用的是 SqlServer 和 oracle 数据库,对系统平台的数据共享拟采用的数据共享平台进行选型,首先调研及论证现有的产品或开源软件无法满足需求,则需要进行数据管理系统的定制开发,分别对应用系统进行梳理及建立对应关系。对具有强关系型的结构化业务数据采用主流关系型数据库管理系统,教学视频、文档等非结构化数据应采用 NoSQL 数据库,如 MongoDB。

(二)数据库组织设计

关系型数据需要进行数据库表结构的组织与设计,应满足第三范式,提高存储效率和减少数据冗余。对于非结构化数据而言,应根据数据的访问特征,将数据进行合理的组织。在数据组织的设计中,应着重考虑平台的可扩充性和扩展性,以方便添加新的数据对象和数据关系。

(三)自动化数据提取和同步程序

开发用于从原有数据库中自动提取数据并导入公共数据平台的软件工具,该工具可以定期地从原有数据库中将最新修改的数据同步到公共数据平台。公共数据库与原有应用系统数据库之间数据如何进行同步以及同步的频率等相关同步程序需要严谨的考虑、论证以及不断根据运行实际进行调整,以满足各应用系统数据及时共享,又能满足数据的准确性。

(四)基于Web Service的数据访问接口

基于 Web Service 的数据访问接口将公共数据平台的数据管理和应用系统的数据操作相互分离,实现应用系统与公共数据平台之间的松耦合,新建应用系统无须了解数据库的表结构、数据表之间的参照关系等底层设计,从而使应用系统具有良好的扩展性和兼容性。此外,应制定公共数据平台

数据访问接口规范,以统一不同开发商开发的应用系统对数据的访问,同时也使得应用系统与公共数据平台的扩展和升级更加灵活方便^[7]。

五、总结

公共数据平台通过对国开已有数据梳理,采用通用数据访问接口,提供对一体化存储数据资源的访问服务,并采取海量结构化数据和非结构化数据一体化存储组织,使其具有高可扩展能力,以应对国开不断增加的业务应用需求。在具体公共数据平台研发的过程中重点做好系统的选型、研发工具,更重要的是考虑与已有各种应用系统之间的对接,同步数据的传输和数据访问接口技术难题。公共数据平台从底层解决各系统和平台的自动传输和共享,减少各系统之间人工干预带来的数据失真,提高工作效率。

参考文献:

- [1]中国远程教育编辑部.适应新要求承担新使命 努力办好中国特色开放大学[J].中国远程教育,2012(15):5-6.
- [2]杨志坚.泛在学习:在理想与现实之间[J].开放教育研究,2014(4):19-23.
- [3]张兴良,李慈章.英国开放大学与国家开放大学教师构成比较与启示[J].中国校外教育:下旬,2015(1):19-20.
- [4]蒋国珍,孙雨.国内开放大学教育信息化现状调查[J].中国教育网络,2015(9):69-72.
- [5]李玉顺.信息技术与教育教学深度融合的发展需求与趋势[J].中国教育信息化,2014(6):3-8.
- [6]宋兴彬,王耀波.基于EAI的税务数据交换平台的建立[J].计算机工程与设计,2006(12):4581-4584.
- [7]孙胜利.大型分布式信息系统数据同步技术研究[D].石家庄:石家庄铁道大学,2014.