



大数据整合机制与信息共享服务实现^{*}

康瑛石, 郑子军

(浙江工商职业技术学院 宁波 315012)

摘 要: 大数据的规模效应给其组织、存储、分析和使用带来了极大的挑战。针对大数据信息服务需要, 对数据资源的组织与管理模式进行了探讨。在以服务为中心的原则上, 构建以互联网和大数据为依托的资源整合与共享服务模式, 重点介绍了大数据的资源整合、共享及互访问工作原理。最后以智慧农业大数据应用为例, 给出现代农业工厂信息资源共享服务机理及应用示范。

关键词: 大数据; 资源整合; 信息共享服务; 农业工厂

doi: 10.3969/j.issn.1000-0801.2014.12.014

Implementation of Mechanism of Big Data Resources Integration and Sharing Service

Kang Yingshi, Zheng Zijun

(Zhejiang Business Technology Institute, Ningbo 315012, China)

Abstract: The information resources organization and management model were discussed. Firstly, based on the principles of information service, a network and big data service integration model was built. Secondly, the information integration, sharing and access mechanism were introduced respectively. Finally, agriculture information sharing service and platform implementation were given, which provided the theoretical and application demonstration for intelligence agriculture factory information construction and realization.

Key words: big data, resource integration, information share service, agriculture factory

1 引言

随着信息技术的快速发展, 大数据的产生与应用密切相关。目前, 基于 Web 的互联网应用、社会计算以及基于行业应用的商务智能和海量数据管理服务成为大数据研究的 3 个重要领域^[1,2]。就 Web 应用而言, 传统的电子商务系统和搜索引擎应用以及正在兴起的社会网络和社会计算是典型的“以数据为中心”的应用。基于行业的大数据服务则是更加综合性的应用, 交互性更强, 数据的产生和来源也更多, 其成功的商业模式必然需要精细的数据处理和

有效分析。在各行业的信息化建设和发展过程中, 所采用的数据标准和传输格式不尽相同, 严重地阻碍了信息的互通和共享, 使得数据成为“僵尸”数据, 造成大大小小的信息“孤岛”^[3-5]。随着以物联网、大数据为基础的智慧信息技术的快速发展, 对数据及信息整合、服务与应用提出了迫切的要求。Nature 早在 2008 年就推出了大数据专刊, 阐述了在数据驱动的研究背景下, 解决大数据应用问题所需的技术以及面临的挑战。Science 在 2011 年 2 月推出专刊, 主要围绕科学研究中的大数据问题展开讨论。参考文献[3]指出大数据时代的数据具有多源异构、分布广泛、动态增长

^{*} 国家发展和改革委员会资助项目(No.[2008]3293), 宁波市(副省级)智慧城市建设项目(No.[2014] WT012)



的特点,导致其与传统数据管理迥然不同,数据管理面临新的挑战。尤其是大数据越来越多地散布于不同的数据管理系统中,为了便于数据分析,需要进行数据的有效集成、整合和共享^[6]。

本文结合大数据应用及共享服务需要,对其表现与组织形式、数据资源整合机制以及信息共享服务管理模式进行了探讨,强调基于以云服务为中心的原则,构建以网络和大数据为依托的信息资源整合与共享服务模式,重点介绍了信息资源整合、共享及互访问协同工作原理。最后以智慧农业大数据应用为例,给出现代农业工厂信息资源共享服务机理及应用示范。

2 大数据服务的信息整合机制

对大数据进行集成、共享及统一表示,实现大数据处理、查询、分析和共享服务是大数据发展中亟需解决的关键问题。大数据服务是一种新的数据资源使用模式,通过对各类数据操作进行封装,为信息消费者提供标准化、随需检索分析的服务。目前针对大数据服务的研究还处于概念讨论阶段,仍然面临多方面挑战,如缺乏一种规范化的大数据服务架构,以屏蔽数据资源和操作复杂性;缺乏体现通用性的非结构化数据模型和服务接口规范。为了解决以上问题,需要对大数据服务的数据模型、服务模型、实现方法进行系统研究^[7]。

2.1 大数据信息目录概念模型

由于大数据具有多源、异类、异构的特点,需对大数据

的基础资源层进行统一编目,得到公共资源目录。该公共资源目录为数据分组的统一格式规范,包含数据接口定义、元数据格式以及数据编码等^[7,8]。信息资源交换框架层需要建立资源交换体系的技术规范,包括通信协议、消息处理机制以及业务交换流程等。其概念模型如图1所示。目录体系和交换体系标准分别从概念、技术要求和相关的管理要求方面进行了规范,见表1。

表1 数据信息目录体系和交换体系

分类	概念	技术	管理
信息资源目录体系	总体框架	接口规范	技术管理要求
		核心元数据	
		资源分类	
信息资源交换体系	总体框架	资源标识编码	技术管理要求
		通信协议、消息处理、流程等	
		数据库接口规范和数据表示规范	

信息资源提供者的主要职责是确保编目信息的正确性,并执行数据录入、编目和注册;信息资源管理者的主要职责是保证目录信息的安全,并对目录进行建立、更新及维护。在此基础上,信息资源的流转和共享服务由信息处理平台自动实现和统一管理,系统平台对外提供Web方式的访问接口,并实现公共资源目录和交换服务目录的发布。

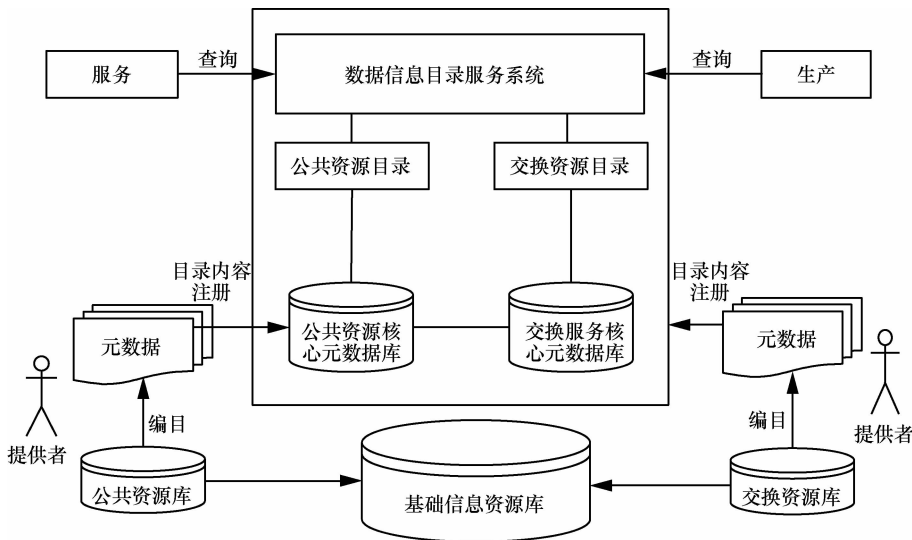


图1 数据信息目录概念模型

各级信息提供者通过公共资源元数据和交换服务元数据进行统一编目,提取特征值并通过元数据注册系统向核心数据库注册入库。管理中心发布已注册的目录内容,对外提供查询服务和信息共享服务。信息使用方依据各自权限通过信息资源目录“一站式”使用服务信息。具体流程如图 2 所示。

2.2 大数据整合机制的内核层次及外围管理

大数据资源整合机制包括基础数据建设层、信息交换框架层、应用服务平台层 3 个层次,整合管理框架部分包括农业信息资源目录体系与标准体系、信息安全保障体系、信息组织机构与管理制度 3 个方面。

(1) 基础数据建设层

主要解决哪些数据资源可以进行整合的问题,根据行业应用特点的不同而不同。以构建农业基础数据资源库群为例,包括环境温度、湿度、气体浓度、光照强度等基础信息库;执行层的业务数据库群,包括作物生长检测数据、设备检测数据、病虫害防御数据等,这些数据在控制过程中可共享,或在一定范围内共享;发布层的数据为可公开发布的共享数据信息,如生产计划、销售计划等信息数据库。

(2) 信息交换框架层

要实现信息资源的整合应用,信息交换是必经环节,应满足两方面需求:一是部门与各计算机之间的信息互联互通;二是数据使用者与数据提供者之间进行互动交流。信息资源交换体系主要由信息交换服务平台、信息交换协议标准、发布服务标准等构成,建立了一个分布式的信息交换平台,实现农业信息资源的交换与共享。信息交换

技术范式包括两个方面:一是数据转换标准;二是异构数据库的汇集整理。数据资源整合管理模型如图 3 所示。

(3) 应用服务平台层

要实现信息资源的跨平台服务,其基础是实现信息共享和服务。信息共享服务平台的建设需从前台服务模式设计、后台应用技术支撑方面进行综合考虑,从总体上加强信息服务的可获性、有效性、准确性、及时性。整合机制的外围管理框架主要包括信息安全保障体系、信息资源目录体系与标准体系、信息组织机构与管理制度体系。通过配套的管理制度将数据资源的整合管理纳入制度化的轨道。

3 大数据信息共享服务机理及实现

在各种数据资源整合的过程中,由于数据异构性、载体多样性、来源广泛性等特点,容易造成其管理和利用的障碍。因此,在整合过程中应对数据进行统一表示,建立信息标准体系及信息访问机制。重点解决数据共享请求与分析、数据并发与同步、互斥控制等问题。根据数据特点,采用基于 agent 的分布式信息共享体系结构,通过 XML-agent 技术解决信息协同技术问题。整个系统可以分为主数据库群、数据引擎、信息目录中心、信息节点以及访问 agent 等,如图 4 所示。

主中心数据库是由多个分布式数据服务器组成的负载均衡系统,通过数据库引擎中间件对外提供异构数据的多种访问接口。信息目录中心主要用于各级计算机系统或数据中心的目录信息索引、元数据、数据字典等,包括系统

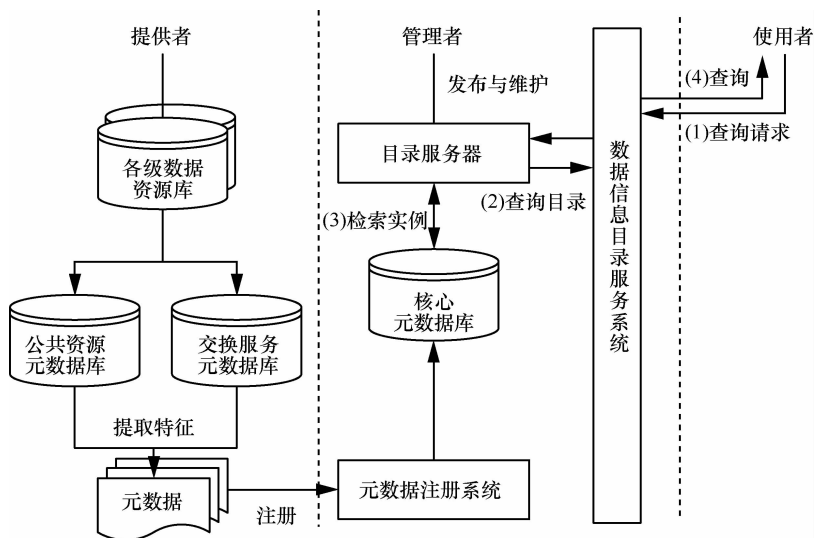


图 2 目录资源体系工作流程

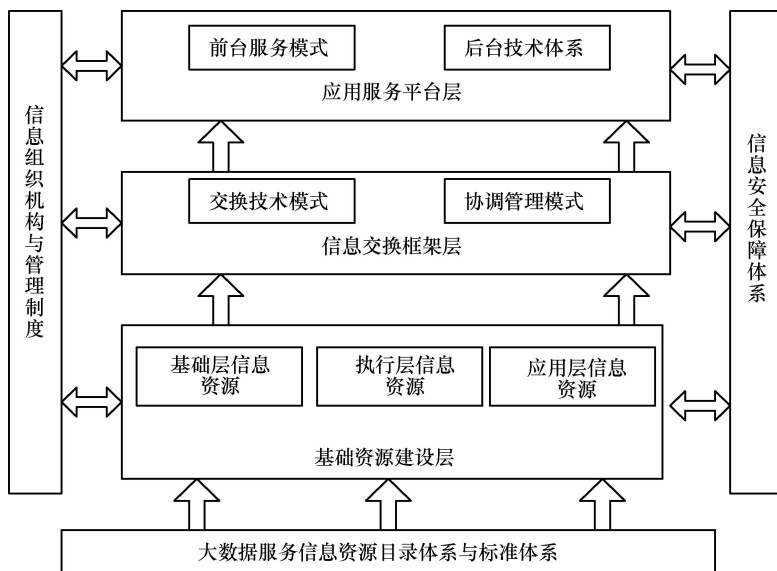


图3 数据资源整合管理模

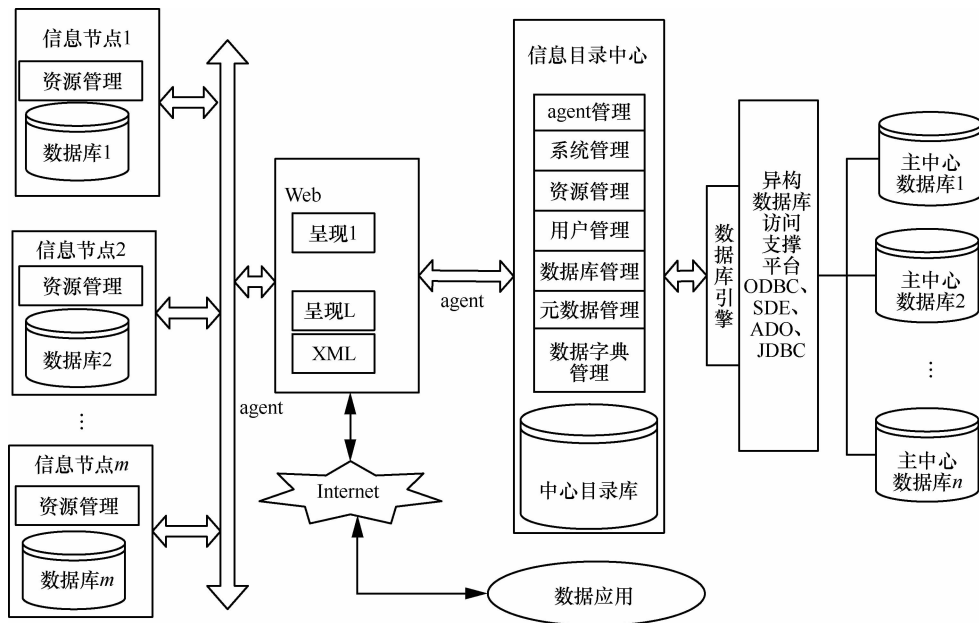


图4 基于 agent 的信息共享体系结构

管理、资源管理、用户管理、数据库管理等模块,实现与主中心数据库的访问。计算机系统通过 XML-agent 对信息节点服务器的数据进行直接或间接存储访问。系统 agent、用户 agent 以及数据 agent 遵循 IHE 规范,并以 XML 作为统一的数据描述语言,相互协作实现对信息节点、目录中心及数据库自动化数据的搜索和整合。信息节点由多台独立的数据服务器构成,具有独立的访问控制权限,并能通过信息节点 agent 向信息目录中心发起请求并同步资源。数据资源库节点分层体系如图 5 所示。

4 智慧农业大数据整合平台的构建与实现

智慧农业是现代农业工厂发展的高级阶段,是指以新兴物联网、信息技术、云计算和大数据为一体,实现农业生产环境的智能感知、智能预警、智能决策、智能分析,为农业生产提供精准化种植、可视化管理、智能化决策的先进生产模式^[9]。智慧农业所需的大数据服务,是指农业产前、产中、产后全流程动态服务的数据信息。需要对农业大数据进行统一协同和管理,依托大数据分析技术建设更高

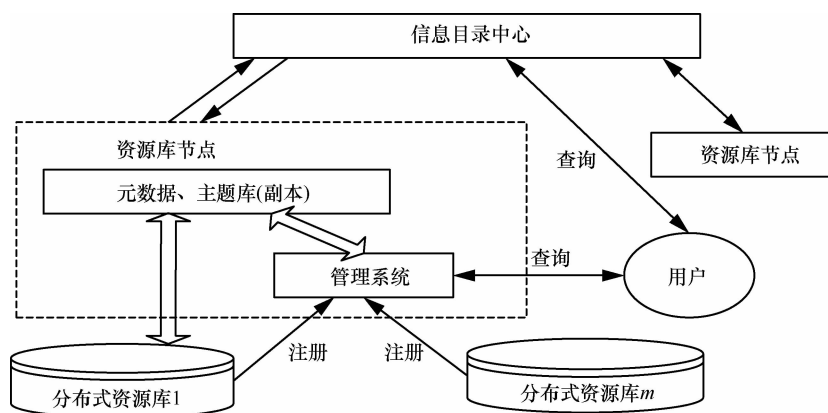


图5 信息资源库分层体系结构

效、先进、开放、整合及共享的信息服务体系。现代农业工厂的大数据服务就是要提升农业资源的合理化分配,实现农业生产控制的智能化,为农业现场生产、资料供应、生产指挥、作物控制及管理决策,提供统一、有效的决策参考^[10]。其根本意义在于能为农业生产、供应及流通提供实用的数据和服务。从农业生产方面看,现代信息技术可以指导生产者、经营者、管理者提高生产效益,运用信息技术,可以使有限的农业资源得以优化配置,提高劳动生产率;从农业资源管理的角度看,依靠资源共享和协同工作可以提高农业信息服务的工作效率、管理水平和服务质量。

基于网络的农业资源整合平台是提升服务的基础,把现代农业工厂各个分布式生产车间计算机控制信息系统的生产、管理及销售数据与服务平台系统动态集成起来,向跨工厂、跨地区、跨业务的用户提供快捷、有效的集成服务。现代农业工厂大数据整合与服务平台系统

架构如图6所示。

通过建立多级信息资源整合与服务平台系统,对信息资源整合与服务进行统一管理、统一协调,按照协作与分工的原则,各车间级、工厂级、业务级、区域级服务平台负责所属范围的资源组织与开发、利用。在此基础上,资源整合平台的实现要解决资源整合和服务整合两个问题。

资源整合可分两个步骤进行:第一步利用中间件技术建立虚拟信息整合系统,实现资源统一访问、调用模式;第二步是将一些具有保存价值的信息资源加以整合并存入信息仓库,实现系统资源的整合。

服务整合主要通过中间件平台,对外提供统一的业务服务界面及操作,对内部各子系统实现数据流透明及标准接口,对分布式、异步式等各级数据库系统信息流进行访问和控制。在早期的农业工厂信息系统中,能实现计算机系统内部的信息化管理和数据资源共享。在面向现代农业

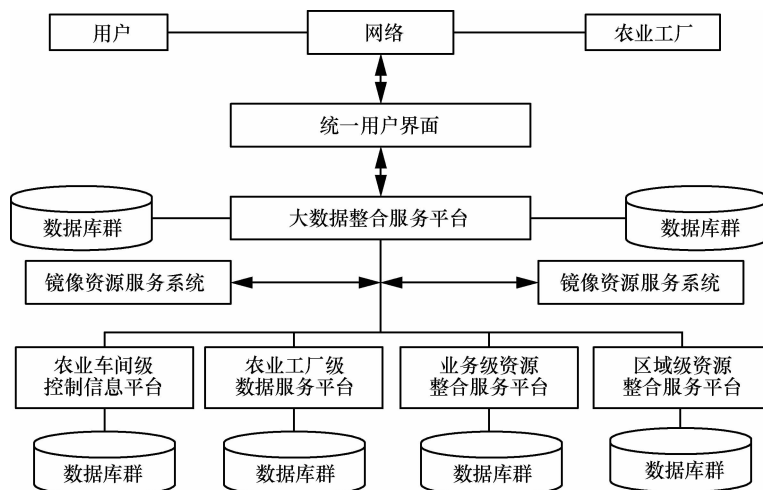


图6 农业大数据整合与服务平台系统架构



工厂的大数据服务过程中,必须建立系统信任网体系,让不同信息节点间的异构信息实现标准化,最终实现农业大数据资源的共享访问。

通过 multi-agent 技术对信息提供方和需求方创建统一的 agent 进行注册、认证以及访问控制,以实现信息节点资源的互访问^[9]。基于 agent 的数据互访问机制和流程如图7所示。

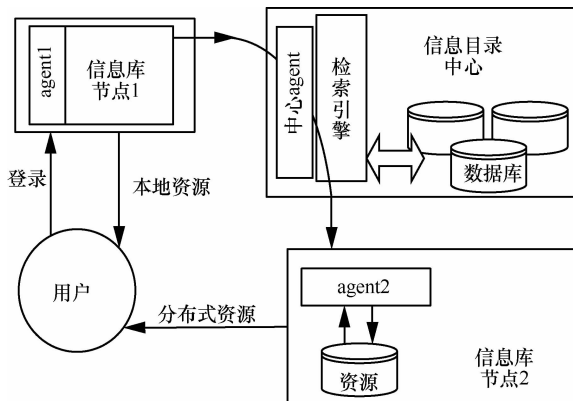


图7 信息资源节点互访问协同工作原理

5 结束语

随着云计算、物联网等技术及应用的发展,数据呈现爆炸式增长,人们正被数据洪流所包围。在此过程中,由于信息系统或者数据中心所采用的数据标准和传输格式不尽相同,严重阻碍了数据信息互通和共享,造成大大小小的信息“孤岛”。为此,大数据的有效组织、集成和服务成为重要的研究内容。本文对大数据的表现与组织形式、数据资源整合机制以及信息共享服务的管理模式进行了探讨,强调在以云服务为中心的原则上,构建以网络和大数据为依托的信息资源整合与共享服务模式,重点介绍了信息资源整合、共享及互访问协同的工作原理。最后以智慧农业大数据应用为例,给出现代农业工厂信息资源共享服务机理及应用示范。

参考文献

- 1 何清. 大数据与云计算. 科技促进与发展, 2014(1): 35~40
- 2 张引, 陈敏, 廖小飞. 大数据应用的现状与展望. 计算机研究与发展, 2013(2): 216~233
- 3 孟小峰, 慈祥. 大数据管理: 概念、技术与挑战. 计算机研究与发展, 2013(1): 146~169
- 4 Shim K. MapReduce algorithms for big data analysis, and

storage of big data. Proceedings of the VLDB Endowment, Istanbul, Turkey, 2012

- 5 Gao F, Abd-Almageed W, Hefeeda M. Distributed approximate spectral clustering for large-scale datasets. Proceedings of the 21st International ACM Symposium on High Performance Parallel and Distributed Computing, New York, USA, 2012: 223~234
- 6 王珊, 王会举, 覃雄派. 架构大数据: 挑战、现状与展望. 计算机学报, 2011, 34(10): 1471~1483
- 7 韩晶. 大数据服务若干关键技术研究. 北京邮电大学博士学位论文, 2013
- 8 商晓帆. 电子政务信息资源管理机制研究. 图书馆学研究, 2007(9): 72~74
- 9 杨其长, 魏灵玲, 刘文科等. 植物工厂系统与实践. 北京: 化学工业出版社, 2012
- 10 熊曙初. 电子政务信息资源整合与重构. 北京: 兵器工业出版社, 2008
- 11 刘晓洁, 杨剑坤. 植物工厂数字农业的革命. 文明, 2011(3): 28~49
- 12 丁杰, 李莹, 夏英成. 发展农业信息服务业推进农业信息化建设. 现代情报, 2007(11): 61~62
- 13 康瑛石, 王海宁, 虞江锋. 基于云计算虚拟化系统研究. 电信科学, 2011, 27(4): 61~67
- 14 康瑛石, 郑子军. 以中小制造业企业为中心的智慧信息服务系统构建. 电信科学, 2014, 30(8): 149~153
- 15 孙忠富, 杜克明, 郑飞翔. 大数据在智慧农业中研究与应用展望. 中国农业科技导报, 2013(6): 33~39

[作者简介]



康瑛石, 男, 浙江工商职业技术学院教授、硕士生导师, 主要研究方向为机械设计及自动化、工业设计。



郑子军, 男, 现就职于浙江工商职业技术学院, 主要研究方向为机械设计、材料成型技术。

(收稿日期: 2014-09-26)