

大数据视野下高校创新分析系统开发与设计

王彤宇

(济南职业学院计算机学院, 山东 济南 250103)

摘 要: 当前各类高校的创新创业教育呈燎原之势, 此教育改革产生海量的教师教学创新、学生创新活动等数据, 以及频繁档案更新任务, 需要高效数据存储与处理能力的系统与之相匹配。本文重点探讨以大数据技术为核心的高校创新分析系统开发设计理念, 提出了解决信息处理速度和系统评价机制的软件设计方向, 并运用Succinct、Spark、Mesos、SQL和Operator等技术实现设计, 还给出了海量异构数据存储硬件资源方案, 为专业教育与创新创业教育融合质量评价、人才培养模式创新评价构建了高效分析环境。

关键词: 大数据分析; 高校管理; 创新分析系统

中图分类号: TP309 **文献标识码:** A

Development and Design of the University Innovation Analysis System from the Perspective of Big Data

WANG Tongyu

(Department of Computer Science, Jinan Vocational College, Jinan 250103, China)

Abstract: The development and design of the innovative analysis system in universities from the perspective of big data needs to reflect the high efficiency of information processing. The paper explores the development and design thoughts of the innovation analysis system with the big data technology as the core, proposes and conducts the software design to improve information processing efficiency and system evaluation mechanism with the application of Succinct, Spark, Mesos, SQL, Operator and other technologies. Additionally, some schemes are also proposed for the massive heterogeneous data storage hardware resource, constructing the efficient analysis platform for the quality assessment of the integration of professional education and innovation education, and for the innovation evaluation of the talent cultivation mode as well.

Keywords: big data analysis; university management; innovation analysis system

1 引言(Introduction)

信息化时代发展背景下, 高校创新分析系统开发主要从提升信息处理速度方面进行^[1]。在高校信息化管理系统中, 能够应用大数据分析技术中的多种计算方式, 对当前所确定的系统开发方向做出控制, 并通过创新分析系统的开发管理, 形成高校信息化管理之间的相互结合^[2]。信息处理速度提升, 直接关系到高校信息系统运行是否切实有效, 面对海量的教学改革、学生创新活动等数据, 以及频繁档案更新任务, 不仅信息处理速度要保证达到预期标准, 除此之外还需要在信息处理中, 引入数据自动更新功能^[3]。

2 高校创新分析系统开发设计的主要方向(The main development and design direction of the university innovation analysis system)

2.1 加快信息处理速度

发现数据之间彼此干扰, 能够通过有效信息分析处理技

术做出隔离, 保障信息处理过程中各个模块处于独立状态^[4]。同时在信息管理中也能根据请求有效联系, 对信息处理范围进行放大, 大数据分析中的信息整合能力, 以及海量信息提取处理能力, 均是构建高校创新管理信息系统所需要的^[5]。目前所应用的主流技术包括Succinct、SparkSQL等。

2.2 优化系统的评价机制

高校构建创新分析系统开发机制过程中, 需要对系统运行可行性, 以及其中所存在的问题做出评价。因此未来创新分析系统开发设计主流方向之一对于系统评价机制的优化与完善^[6]。能够根据使用一段时间的数据反馈, 对接下来的数据信息处理方向做出调整。优化系统评价机制, 这需要应用高效计算技术来完成, 在抽象数据集合中, 能够快速确定信息提取方向, 并通过信息技术之间的整合, 对数据分析整体速度做出提升。完善系统评价机制, 需要建立在网络数据提取基础上, 尤其是针对当前的主流新技术, 更应该考虑其运

行环境对自身数据安全性的威胁。系统构建中会有平台内部的调节器来实时反应信息变化,在不断评价过程中寻求新的数据信息接口,避免接口选择错误而影响到存储速度与安全性。在这一新型设计开发方向背景下,大数据分析成为主流技术支撑,主要应用在基础框架构建与内部程序汇编完善方面。大数据分析基本框架构建见图1。

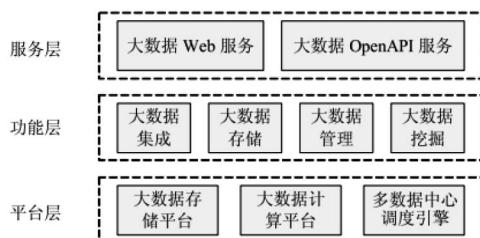


图1 大数据分析平台基本框架分层

Fig.1 The basic framework and hierarchy of the big data analysis platform

3 高校创新分析系统开发设计思路(Development and design thoughts of the university innovative analysis system)

3.1 构建多元评价体系

基于大数据分析基础所构建的创新分析系统,首先是对多元化评价体系的构建,可以利用资源管理Mesos技术,构建出用户访问的等级框架任务。这样当用户需要对大数据分析系统进行访问时,在系统内部能够通过分析运算来自动分类,对所接收到的信息请求自动分层,能够形成初步的多元评价体系。评价体系还需具备大数据查询功能,在所构建功能层的最基层利用Spark技术作为搜索功能实现引擎,启动SQL与Operator技术,形成评价搜索中的高速度类型划分。Spark Streaming技术能够实现评价时间的整合与区分,当请求之间存在彼此联系时,能够通过这一技术方法进行更为高效的内部控制。多元评价体系构建还需要形成类似树状结构的分析基础,这便要求在系统分析过程中,启动近似查询引擎,通过账户查询请求,能够将数据的精密程度控制在允许误差内,通过GraphX基于BSP建立基本模型。常用技术方法对最终平台评价系统构建也发挥着基础控制的功能。

3.2 建立评价互动机制

利用算法公式建立评价互动机制,基于大数据分析基础上评价互动机制的形成,需要在程序汇编中设置反馈系统。在创新分析系统投入使用后,所构建的多元评价机制能够根据最终数据运算结果做出反馈,这样便能够形成完善的评价互动机制。评价互动不仅需要了解并获取信息,还需要根据分析结果做出自动调整,发出相应的功能请求,在系统程序汇编中构建出如下的运算程序公式:

```
Var X=load(“some_data”.2 to 10)
Var Y=load(“some_data”.1)
Xar(fn=model,summery)=doClassify(X,Y)
```

公式中形成了评价互动反馈机制,并在系统基本框架构建出以xy为轴点的位置区域,这样进入到系统评价活动中,对于数据信息获取最终位置的确定也更加合理。评价互动机制的形成,主要体现在功能层面之间的相互配合,互动评价中的功能测试数据规模结构如图2所示。

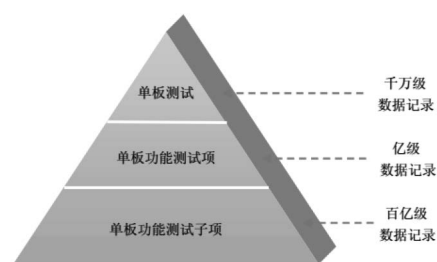


图2 系统评价测试数据规模

Fig.2 System evaluation test data scale

3.3 完善配套维护制度

高校创新分析系统开发构建过程中,需要对所应用的配套设施进行完善,建立起完善的配套管理制度,并对所构建的数据信息库定期更新审核。观察在数据系统中,是否受外界网络接入点环境影响,存在影响数据信息安全性的隐患,定期对软件程序部分自动化检测。检测任务同样可以建立在信息系统中,由信息系统在数据后台内部完成。预留给操作者的界面十分简洁,大部分分析计算任务都是在后台系统中自动完成的,这样既保证了速度但不会造成数据信息丢失问题。使用Mahout/Spark完成维护过程中的程序分布查找,了解不同类型信息所处的具体位置,这样在运行使用过程中,可以最大程度确保信息化技术的高效实现,并通过彼此之间的交互转换来完成更高层次的功能任务。维护制度中还要体现信息技术之间的整合要求,高校创新分析系统设计构建中,软件程序汇编达到预期标准后,便需要进入到程序内部控制中,通过几种控制方法之间的相互融合,达到更为高效的控制效果。

3.4 开展指标适用性分析

指标适用性分析建立在数据获取基础上,高校创新分析系统构建后,随着创新项目计划开展会产生大批数据,将这些数据整合在统一的框架结构内,并根据分析运算结果判断软件控制程序是否符合当前的信息化管理需求。根据数据信息指标反馈结果,对系统的适用性做出判断。适用性也具有统一的衡量标准,如果能够达到标准值上,便表示所确定的软件汇编程序是正确的,反之则存在安全隐患。需要通过数据库强化更新,保证稳定性。系统内部的各项控制管理计划开展,同样需要通过各项管理功能之间整合联系。高校创新分析系统开发模块划分如图3所示。



图3 高校创新分析系统功能模块划分

Fig.3 Function module division of the university innovation analysis system

4 大数据分析技术在创新分析系统开发设计中的具体应用(The application of big data analysis technology in the development and design of the innovation analysis system)

4.1 数据分布排列运算

大数据分析中的分布排列运算,是高校创新分析系统开发设计中常用的功能部分,通过对数据排列方式进行设计,实现类似数据库的整合。这样系统运算过程中,便能够减少数据库内部审核所浪费的时间,大数据分析需要建立在数据

库监控基础上,使用数据挖掘技术创建不同开放接口,通过内部控制流程开放,营造出分布式排列中的内部控制核心系统,Open API作为主要的大数据挖掘技术,进入到系统内部控制核心部分,更需要体现出系统的核心功能稳定性。尤其是在系统挖掘设计中,各个数据库之间的相互整合联系情况。区域智能数据中心提供基于云计算的大规模数据存储及数据挖掘平台,通过平台服务器对外接口提供数据存储、分析与挖掘服务。用户使用Web浏览器或智能终端应用程序提出数据存储和分析的服务请求,经Web服务器通过互联网将服务请求发送给数据中心平台服务器,平台服务器对服务请求进行解析,发送给工作流引擎调度执行,执行结果通过互联网发送给用户终端。对于数据排列与运算分析控制,系统内部重点体现在管理功能层方面。数据分布排列运算,通过不同排列方式整合来达到功能突破的效果,高校创新分析系统设计应用中应该考虑多种客户端的开发构建,包括移动手机客户端。教师与学生使用移动手机设备访问网络的需求不断增多,通过数据分布排列形式变化,实现多种功能之间的融合,移动手机客户端功能开发,可增进使用中的客户体验感。

4.2 基于误差补偿的原始数据恢复

程序汇编可以使用函数式递增方法来进行,这样更能区分出不同功能层之间的独立性,所配备的核心硬件资源在大数据分析过程中要注重平台的拓展性,尤其是针对当前比较常见的原始数据缺失问题,更应该从缺失数据值的发现时间段开始记录,对误差进行有效补偿,避免误差继续严重影响到最终的使用。数据缺失值发现如图4所示。

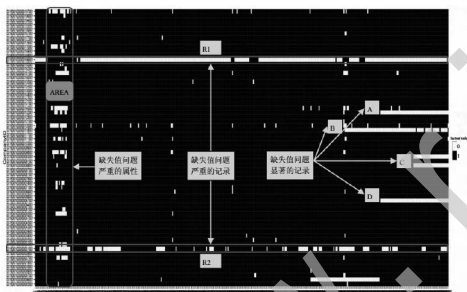


图4 原始数据缺失值发现

Fig.4 Missing values of original data

原始数据一旦缺失在记录中会有明显异常,观察异常情况便能够了解到原始数据的变化问题,将其标录在系统中,能够构建出的系统应用功能,与常规数据之间的结合效果十分突出,在分析运算中也要重点注意这一标识。针对传统分析方法交互性和可理解性不足的问题,研究启发式、人机交互、可视化数据挖掘新技术,实现大数据挖掘的高度人机交互功能。针对大数据挖掘分布式调度的挑战,研究基于云计算的分布式工作流调度、负载均衡技术,构建高效分布式工作流执行引擎。通过这些技术方法来减少系统运行中的原始数据缺失。

4.3 海量异构数据存储

海量异构数据存储能够帮助扩大信息化系统的分析能力,首先计算资源要达到标准,高校在创新分析系统优化设计中配备的计算硬件资源参考如下:30台RH2288运算处理服务器,CPU核心处理器的内核共应该达到860核,总共内核达到存储量8TB。硬件部分的核心总存储量达到600TB。这样的硬件存储,在运行过程中可以达到预期的控制效果,能够与

软件部分充分配合,从而达到核心处理效果。进入到核心运算处理环节中,海量异构数据存储数量十分大,能够在存储中达到数据中心点的整合,确定中心点后,即使通过数据配合,也能避免数据基础中出现相互干扰。海量异构数据的整合与控制能力,是当前大部分数据分析功能不能达到的。高校创新分析系统中,侧重于创新创业信息汇编整合,帮助学生在入校学习期间能够更深入的了解市场需求变化,市场竞争信息变化速度十分快,要求将变化中的信息特征体现在分析系统中,大数据分析中的海量异构数据处理,重点体现在这一方面,对于数据信息的构建,以及相互处理十分重要。

4.4 多源异构数据融合

构建异构处理控制功能层,通过系统之间的相互交换来提升稳定性,并避免系统感染,尤其是针对当前的异构数据源问题,处于功能相互控制系统中,两者之间相互交换共同控制干扰,提升管理效率。数据之间的融合,与数据库构建有直接联系,通常情况下是将数据库作为基础来进行深层次的构建,体现出数据之间的融合性,并与管理计划相互结合,共同实现管理效率的提升进步。多源异构数据的来源不同,通过大数据分析技术应用,将其结合在有效的网络平台中,进而实现更加高效的控制与内部整合,数据库之间的联系性是变化的,同时也能在完整的结构系统中进行相互转换,达到更理想的控制效果。将大数据分析技术应用在高校创新分析系统设计应用中,设计理念创新时还需要考察实际情况,评估是否具有可行性。

5 结论(Conclusion)

大数据已经成为全社会的共识,大数据所蕴含的经济价值和创新价值已经引起社会各界的高度关注。我国拥有海量的数据资源,广阔多样的应用场景,潜力巨大的消费市场,当务之急是如何快速有效突破数据价值挖掘的瓶颈。大数据分析与应用平台,是大数据时代必备的基础设施,也是突破当前技术瓶颈的有效突破口。开发和建设大数据分析与应用平台将带来三个方面的价值:第一,有助于不断汇集大数据技术创新成果,并用最先进的技术为用户提供一站式的应用服务;第二,有助于降低用户技术门槛,为应用开发提供共性基础设施与服务,从而加快应用创新;第三,有助于形成大数据技术产品和行业解决方案,促进我国大数据产业加快形成。将大数据分析技术应用在高校创新分析系统设计中,具有广阔的开发前景。

参考文献(References)

- [1] 孙艳.基于互联网大数据的高校网络关注度分析——以无锡地区为例[J].三门峡职业技术学院学报,2016(4):52-56.
- [2] 牛雨.基于大数据背景的高校学生管理信息化分析[J].中国新通信,2017,19(5):27.
- [3] 杨居义.基于高校教学督导共享平台大数据分析研究[J].课程教育研究,2016(10):33-36.
- [4] 沈蕾.大数据分析平台支持下高校智慧校园建设研究[J].教育现代化,2018(7):035-036.
- [5] 刘攀,王倩.大数据背景下的高校《GNSS定位测量》的教学设计研究[J].中国包装,2018(3):045-046.
- [6] 刘震.基于大数据分析的高校教学工作创新研究[J].晋城职业技术学院学报,2017,10(4):45-47.

作者简介:

王彤宇(1967-),男,硕士,教授.研究领域:数据分析,计算机专业教育。