数据科学与大数据技术实验室建设探索及研究

张雯莉

桂林旅游学院

DOI:10.32629/mef.v3i1.532

[摘要] 社会的飞速发展促进了各个行业的数据量呈现出爆发式增长。通过利用数据可以对资源进行合理配置、辅助决策,是创造价值的一种方法,同时也是创建新兴产业的一种依据。数据的爆发式增长所带来的是大量的数据储存和对数据中有效信息的提取。数据也已然成为研究各事物之间存在的抽象关系并做出决策的重要依据。而大数据就是其中的媒介,通过大数据可以快速精准的做出预测和判断。本文首先介绍了大数据技术,其次对大数据技术实验室的建设进行了相关探讨。

[关键词] 大数据; 实验室; 探索

Exploration and Research on the Construction of Data Science and Big Data Technology Labs

Zhang Wenli

Guilin Tourism University

[Abstract] The rapid development of society has promoted the explosive growth of data volume in various industries. The use of data to reasonably allocate resources and assist decision-making is a method of creating value and a basis for creating emerging industries. The explosive growth of data has brought about a large amount of data storage and extraction of effective information from the data. Data has also become an important basis for studying the abstract relationships between things and making decisions. And big data is the medium among them, through big data, you can make predictions and judgments quickly and accurately. This article first introduces big data technology, and then discusses the construction of big data technology laboratory.

[Keywords] big data; laboratory; exploration

在当前科学技术不断革新的背景下,迅速上升的数据量促进了大数据时代的到来,而我国已经将大数据作为一种战略资源,因此大数据成为了一项重点研究工作。基于大数据进行资源的合理配置、辅助决策已经成为大数据的特点。而大数据同时还是一种数据处理技术,可以在数据中进行搜索,在得到所需信息之后,可以为决策提供指导建议。而高校在开展教学的过程中,通常会存在实验条件较差或是缺乏实验基地等相关问题,导致师生在实验方面的需求无法满足。所以,建设数据科学与大数据技术实验室具有十分积极的意义。

1 大数据技术

大数据可以说是一种信息资产,并且无法采用常规软件来对其进行分析、储存、收集,同时大数据还具有海量、多样化等众多特点,是当前时代的一种基础型战略资源。而大数据除此之外还具有速度快、高数量、高价值、多样性等众多特点,通过对于大数据进行整理和分析,可以获得新需求,因此具有十分重要的价值。大数据技术可以挖掘数据中所包含的特点,同时通过建立模型、迭代验证、确立模型等环节来实现一个完整的流程。大数据技术可以对社会中的各个行业开展预测性分析,并且能够提出可行性的建议,具有良好的辅助决策作用。而大数据技术还可以分为计算分析技术、数据储存技术。其中,数据计算分析技术又包括预测模型、自然语言处理以及模式识别技术等,而数据储存技术则含有

MPP架构新型数据库集群、非结构化数据收集架构等。

2 建设大数据技术实验室的意义

大数据技术作为当前最为前卫的一种技术,所以建设大数据技术实验室,能够为广大学师生搭建一个大数据技术的研究平台,使师生可以在实验室内开展课题研究、教学实训等活动,同时大数据技术实验室还可以培养学生掌握大数据技术应用理论和能力,从而提高学生的大数据处理能力以及分析能力等,同时还可以改善理论研究所存在的局限性,从而为开展大数据技术的相关研究提供帮助。

3 建设数据科学与大数据技术的实验室

3.1 硬件与软件建设。在可以满足教学与科研需求的前提下,数据技术实验室必须要遵循国家相关标准,并采用当前国内外有效的安全技术手段和先进的大数据技术,基于此来建立实用性、可拓展性的大数据实验平台。大数据技术实验室中,关于系统硬件平台的构建:在系统中,硬件部分主要包括计算中心数据储存、可视化模块与终端等。在这其中,数据储存、计算中心在整个系统中属于关键核心,可以为系统提供底层的网络资源、计算机资源以及存储资源等,并且具有并行计算、虚拟化、云存储等多方面的功能。而许多公司在当前都已经开始提供硬件相应的解决方案,常见的主要有曙光、Dell以及HP等。大数据及时实验室软件平台的建设:大数据技术软件平台是指平台内的一系列软件架构和不熟,

例如,多种储存模型管理系统。大数据平台可以借鉴其他建设大数据的经验,并通过整体规划、分步实施的建设思路来建立大数据管理平台,能够为大数据应用提供底层的基础数据支撑,同时建立具有典型性与示范性意义的提大数据平台。

而大数据技术实验室中,建立信息安全运行平台也非常的重要,基于此能够为数据信息系统构建安全支撑体系,以此来保障各种业务应用的运行安全,通过利用大数据技术还能够对信息系统实现安全管理,将安全保护策略贯彻到信息系统中的每一个层面。

3.2建设实验室总体框架。实验室的总体框架构建,是将软件体系结合硬件设备。而实验室需要具有应用层、基础设施层、虚拟资源层以及云服务平台。其中,基础设施层是由数据储存、传递储存设备以及网络设备所构成,主要是负责开展基础物理性的网络建设、网络维护等工作;应用层则是教学的交互式界面的建设;云服务平台是由数据维护管理的相关设备组合而成;而虚拟资源层则是为开展数据计算、用户信息管理的平台建设。

3.3大数据技术实验室管理平台的建设。大数据技术实验室管理平台中,需要具备协同开发桌面云平台、知识资源库管理平台以及协同开发服务端平台等。首先,建设知识资源库管理平台,一方面建设源码库、类库等支持型的构建,另一方面需要建设实际素材、开发计划以及案例库等文档,以此来提高项目研发的效率。另外,每一个高校都具有人才培养目标,所以在大数据技术实验室的建设中,还需要具有相关的知识资源案例,并可以提供出文档撰写示范、编码案例示范以及项目案例示范等。其次,协同开发服务端平台,在平台中导入系统,而说明书也需要作为注册的初始化文档,在系统中导入说明书。最后,微系统开发桌面云平台。在开发过程中,在云计算数据中心机房中需要部署桌面计算机的内存硬件、计算资源以及CPU等,然后对物力资源实施虚拟化转变,基于用户所提出的实际需求,为用户提供虚拟桌面服务。

3.4数据源。大数据实验室中还需要必须要具备大量的数据,而普遍常见的大数据类型主要有视频、文字、图像、声音以及传感器数据等。而对于大数据技术实验室而言,数据的来源是一项非常重要的问题。对于实验室中的大数据,其安全与隐私保护也是实验室必须要保障的。

3.4.1行业数据。企业在实际的运行发展过程中,将会产生大量的数据信息,而通过与企业之间开展合作,能够获得其中的部分数据。随着信息化技术的发展,企业业务的扩大,企业在研发、生产以及销售、储存等环节中都会产生并积累大量的数据信息。

3.4.2网上数据。互联网中也具备海量数据,因此在互联网中同样可以获得数据。通过网络爬虫技术、第三方公司提供的资源等途径,都可以获得所需的数据。而网络爬虫是一种能够自动下载网页的程序,基于既定的抓取目标,能够选择性访问互联网上的网页与链接,从获取所需数据。而部分的第三方公司将会专门提供数据收费服务,此方式也同样可以获得数据。

3.5 Spark大数据实验室建设分析。基于Spark大数据实验室建设为案例开展分析,该实验室是在OpenStack云平台的基础上建设的,OpenStack是开展虚拟机管理,也就意味着对于虚拟机进行挂起、建设、重启以及销毁等操作进行管理。而该实验室的主要集群平台中,是由六台或更多的虚拟机配置,其中,有四台虚拟机是工作节点,一台作为主节点,最后一台是主备份节点,通过虚拟机可以组成 Spark 服务器集群。除此之外,还可以将 hadoop2.X 配置合理添加进去,但前提是需要保证其部署结构与Spark 部署结构之间相同,该配置可以作为Spark on yarn的模式环境,以此来保障Spark集群能够提供Spark on yarn、Spark standalone这两种模式。基于该配置,Spark 大数据实验室的建设既具有轻配置、易扩展的特点,同时还具有易安装、重资源等特点。既可以满足课堂研究与教学需求,也可以满足实践方面的需求。

3.6大数据技术实验室队伍的建设。当前大数据技术开发队伍还十分缺乏,所以建设实验师队伍的最佳途径是选择多方参与合作的形式。通过利用社会上知名企业内的大数据技术,开展校企合作模式,一起参与到培养学生实践能力的建设中。企业可以为高校提供企业教师、技术工程师等作为学生的兼职教师,从而为学生专门开展具有较强实践性的教学内容,以此来弥补学校教师所存在的不足。而学校教师则可以为开展理论方面的课程,起到相互弥补的效果。

4 结语

综上所述,大数据是当前时代下的一种产物,但同时也是一种战略资源。而我国的大数据技术中的信息技术与国外之间相比较,仍旧还存在一定的差距。因此,加强我国大数据科学技术的建设力度,已经成为当前高校的发展与剑和目标之一,而培养相关技术人才也已经是一项紧急任务。对于高校来说,利用自身的多学科特点,在大数据技术实验室建设中将技术和管理体系进行结合,并基于专业优势的条件背景下,以此来提高高校的科研水平,并为社会培养出更多社会所需要的技术型人才。

[参考文献]

- [1]杨谨羽.大数据技术实验室建设探索与研究[J].科技经济导刊,2018(19):165.
- [2]谭忠理.基于云计算和大数据教学实训平台的设计研究[J].智能建筑与城市信息,2014(8):86-90.
- [3]王振华,洪泓,陈春丽.大数据技术实验室建设探索与研究[J].电脑知识与技术,2017(12):132-133.
- [4]黄叶超,郑庚.基于大数据技术实验室建设探索与研究[J].电子测试,2017(24):56+53.
- [5]李晓丹,刘云翔,王浩,等.应用型计算机专业大数据分析实验室建设[J].实验技术与管理,2017(09):236-238.
- [6]莫潇晓.大数据背景下计算机专业实验室建设及管理研究[J].电子技术与软件工程,2017(09):151-153.

作者简介:

张雯莉(1983--),女,汉族,广西桂林人,讲师,硕士,研究方向:教育管理。