

关于大数据的数据处理探讨

霍福华

(山西省财政税务专科学校, 山西 太原 030024)

摘要: 随着信息技术和计算机科技的进一步发展, 互联网应用普及, 以往人们忽视的数据资源价值更加凸显, 由此发展的大数据技术更是将这些海量的数据信息资源充分挖掘出来, 实现数据资源转化为经济来源, 为学校优化管理方式和优化课程设置、为企业制定生产和发展决策、为政府制定相关管理政策等都提供了有效的参考。大数据在数据处理中, 用到的相关技术较多, 掌握这些技术应用, 对于进一步推动大数据发展具有重要意义。本文分析了大数据的数据特点, 分析目前大数据数据处理中的关键技术, 并针对大数据数据处理技术的进一步优化和应用, 提出几点建议。

关键词: 大数据; 数据处理; 关键技术; 应用

中图分类号: TP399 **文献标识码:** A

Discussion on Data Processing for Big Data

HUO Fuhua

(Shanxi Finance & Taxation College, Taiyuan 030024, China)

Abstract: with the further development of information technology and computer technology as well as the popularization of the Internet, the value of data resources neglected by people in the past has become more prominent. Big data, which is developed from this, fully excavates these massive data and information resources, realizes the conversion of data resources into economic sources, optimizes management methods and curriculum settings for schools, formulates production and development decisions for enterprises, and provides effective references in making relevant administration policies for the government. Big data involves many related technologies in data processing, so mastering the application of these technologies is of great significance for promoting the development of big data. This paper analyzes the data characteristics of big data, analyzes the existing key technologies in the process of big data, and puts forward some suggestions for the further optimization and application of big data processing technology.

Keywords: big data; data processing; key technology; application

1 引言(Introduction)

目前, 大数据已经在人们的生活中普遍应用, 人们在消费、生产、工作, 以及多种领域中都会留下一定的数据信息, 为大数据提供了充足的数据资源, 而大数据的发展也在一定程度上便利了人们的生产和生活, 强化大数据处理技术应用, 优化大数据技术系统, 对于发挥大数据更大的价值具

有重要意义。

2 大数据的数据特点(The data characteristics of big data)

就大数据中的数据总体特点来看, 主要体现在以下几个方面:

第一, 数据的庞大性, 大数据收集和处理的的数据是海量

的, 这些数据是动态变化的, 在数据处理的过程中还会不断增长和变化, 因此, 大数据处理的数据还在不断扩展, 数据量不断增长。大数据的数据量庞大要求进行大数据处理和设备的, 以及软件、硬件等自身的存储能力和信息处理能力必须要十分强大, 才能应对时刻变化增长的数据^[1]。

第二, 数据的多样性, 大数据收集来的海量数据中, 数据的存在形式可能是多种多样的, 除了一些常规的数据, 还包括图片、文字、视频等非结构化的数据资料, 对于这些资料必须要进行有效识别和处理, 才能转化为有价值的信息, 这需要大数据数据处理系统进一步优化系统识别能力, 提升数据转化能力。

第三, 数据的快速性, 大数据对于海量数据的处理主要是针对快速变化的数据开展分析处理的, 因此, 需要通过快速的系统处理和计算快速得出处理结果, 提供有效的信息。

第四, 数据的准确性, 大数据处理的数据信息结果必须要准确的, 包括搜集信息来源、数据储存和数据计算等, 这一系列的过程都要是准确无误的, 才能确保最终的数据分子结果的准确性。

3 大数据的数据处理关键技术(Key technology of data processing for big data)

3.1 大数据采集技术

大数据进行数据处理的前提是, 大数据系统已经获取了大量的数据信息, 这就需要用到大数据的数据采集技术, 在大数据数据采集技术中, 包含了传感器、射频识别技术、计算机转换软件技术等的应用, 对于互联网中某一领域出现的数据, 包括结构化的、半结构化的, 以及非结构化的数据信息进行广泛收集和识别, 这是大数据技术的前提和基础。要想实现对于互联网中海量信息资源的准确有效的获取, 需要借助分布式高速高可靠数据爬取或采集、高速数据全映像数据采集技术的应用, 能够实现高速的数据解析和转换目标^[2]。

就大数据采集技术系统来说, 其中包含的数据采集功能区分为以下几点:

第一, 数据智能感知层, 这一功能区中包含了对于数据传输技术、通信技术、智能识别技术等的应用, 能够对于互联网平台中某一通道或者是领域中的各类数据实施有效的跟踪和接入, 完成数据信息的初步处理, 为采集到的数据整合和传输奠定基础。

第二, 基础支撑层, 这一功能区主要是为系统提供虚拟服务器, 对于各类数据信息创造合理的分析环境, 这一数据处理功能分区中需要重点对于数据的获取和存储, 以及整理和分析等的可视化技术接入提供有效支持。

3.2 大数据预处理技术

大数据系统要进行分析 and 挖掘, 就必须为它提供干净、准确、简洁的数据。然而通过数据采集技术收集到的原始数据一般是“脏”的、不完全的、冗余的和模糊的, 通常存在杂乱性、重复性、不完整性问题, 很少能直接满足数据挖掘算法的要求^[3]。所以, 大数据数据采集后, 需要通过数据集成(Data Integration)、数据清洗(Data Cleaning)、数据变换(Data Transformation)、数据简化(Data Reduction)等操作^[4]摒弃一些与挖掘目标不相关的属性, 为数据挖掘内核算法提供干净、准确、更有针对性的数据, 将其中多余的或者是无效信息进行删除, 去粗取精, 实现初步的数据筛选, 从而减少挖掘内核的数据处理量, 提高了挖掘效率, 提高了知识发现的起点和知识的准确度。

3.3 大数据存储及管理技术

大数据采取的数据信息量十分庞大, 对于海量的数据信息大数据系统必须要能有效地存储和管理。针对采集的数据信息, 系统通过构建数据库, 进行数据的集中存储和管理。在进行数据信息进行存储和管理过程中, 需要确保数据库中的数据信息形式尽可能保持一致, 这就需要做对于不同种类数据信息的有效转化和处理。要突破分布式非关系型大数据管理与处理技术, 异构数据的数据融合技术, 数据组织技术, 研究大数据建模技术; 突破大数据索引技术; 突破大数据移动、备份、复制等技术^[5]。

在数据的存储和管理中, 还需要注重信息的有效分类, 建立对应的数据库, 实现信息的有效调用, 此外, 还要针对数据的安全问题进行有效保护, 通过使用加密技术、访问限制、数据销毁等技术应用, 提升数据库的数据安全级别, 避免数据存在遗失和被窃取的问题^[6]。

3.4 大数据分析和挖掘技术

大数据的分析和挖掘技术主要是对于数据进行深度挖掘, 通过多样化的数据分析技术, 挖掘数据内潜在的有价值的信息。这种数据分析和挖掘技术主要是通过对于数据进行随机抽样, 缩小数据规模, 进一步提升数据分析效率, 挖掘出更多的有价值的信息, 为最终的数据分析结果提供参考^[7]。数据分析和挖掘技术还在进一步优化中, 也在越来越多的行业中发挥了积极作用, 是目前大数据技术重点开发的对象, 为大数据的价值挖掘, 实现更好的分析结果优化提供了技术支持。

在大数据系统中, 海量的数据采集最终的目的是为了分析出数据隐藏的规律和关系, 通过分析得出结果, 提供有效的信息参考和发展建议等。这就需要针对海量的数据实施

数据挖掘,挖掘数据中潜在的信息资源,发挥数据的价值。现阶段的大数据分析和挖掘技术还只是比较浅层次的分析,能够针对大数据的数据规律和发展趋势等进行有效分析,但是实际上,一组大数据包含的价值往往是多方面,仅仅从一两个面进行挖掘往往不能发挥大数据的最大使用价值^[8]。因此,在进行大数据系统的开发中们还需要做好关键部分的技术升级和改造,提升大数据分析和挖掘能力。

4 大数据数据处理技术有效应用的对策(Measures for the effective application of big data processing technology)

4.1 不断优化和升级技术系统

针对目前的大数据处理技术系统,要进一步研究和优化设计,不断提升各个技术系统的数据处理能力,实现在更短的时间内获取更全面的数据信息,实现快速处理得出分析结果,针对大数据系统中的数据采集技术、数据预处理技术、数据存储和管理技术、数据分析和挖掘技术等进行不断优化升级,促进系统数据处理效能不断提升。

4.2 开展技术培训,提升技术应用水平

新时期,要强化大数据技术的应用,要不断加强技术的普及和学习,积极开展技术培训工作,例如,以电子商务、金融、连锁零售、电信、互联网、产品研发生产、传统实体经济等方面在数据分析与挖掘实战经验为案例,使学员在拓展大数据应用及其发展趋势视野的基础上,提升对分析、挖掘经验的领悟,做到学以致用、举一反三,解决企业实际数据挖掘的现实问题,从而提升数据挖掘与治理能力,推动企业竞争力提升和产业创新发展。相关的大数据人才培养机构也要积极围绕“大数据在教育教学中的应用探讨”,从大数据及关键技术、教育大数据、大数据技术在教育中的应用、应用案例等四个方面深入细致的探究^[9]。从大数据的产生及其意义、构建大数据的基本流程、身边大数据的典型案例等三个方面,深入浅出地介绍了大数据产生的背景和发展历程。要认真学习贯彻落实党的十九大精神,充分认识大数据的重要意义,牢牢把握大数据时代带来的重大机遇,自觉将大数据技术应用于教育教学、管理服务之中,不断提高教育教学质量,提升管理水平,源源不断地为社会培养一批又一批的大数据技术创新应用人才。

5 结论(Conclusion)

现阶段,大数据技术在社会生产和生活中的应用对于社会进步和发展意义重大,必须要进一步完善大数据技术应用,开发升级技术系统,把握关键技术改造,促进大数据更

好的挖掘数据价值,提升系统技术应用水平。

参考文献(References)

- [1] Valerio Persico, Antonio Pescapé, Antonio Picariello, Giancarlo Sperli. Benchmarking big data architectures for social networks data processing using public cloud platforms[J]. Future Generation Computer Systems, 2018: 89.
- [2] Weiwei Wang. Research on Computer Information Processing Technology under the Background of Big Data[A]. International Information and Engineering Association. Proceedings of 2018 International Conference on Computational Science and Engineering (ICCSE 2018)[C]. International Information and Engineering Association: Computer Science and Electronic Technology International Society, 2018: 5.
- [3] Lin Mao. Improvement on Exploration Data Processing of Cluster Architecture in Big Data Era[A]. CPS, SEG, CPS/SEG Beijing 2018 International Geophysical Conference & Exposition Electronic papers[C]. CPS, SEG, 2018: 7.
- [4] Yixiang Jiang. Analysis of Computer Information Processing Technology Under the Background of Big Data[A]. Wuhan Zhicheng Times Cultural Development Co., Ltd. Proceedings of the 2018 International Symposium on Communication Engineering & Computer Science (CECS 2018)[C]. Wuhan Zhicheng Times Cultural Development Co., Ltd, 2018: 4.
- [5] 曲冬梅. 大数据背景下信息处理技术探索[J]. 现代信息技术, 2018, 2(03): 18-19.
- [6] 谭保华, 任志鹏, 王鹏, 等. 基于大数据技术的高校校友信息整合分析平台开发及应用[J]. 长春理工大学学报(自然科学版), 2018, 41(06): 132-136.
- [7] 郑芬芳. 基于数据分析的学科服务营销平台研究——以福建省高校数字图书馆为例[J]. 情报探索, 2018(12): 34-38.
- [8] 刘哲, 刘军, 张朔. 电子商务物流服务中大数据应用的驱动因素与创新发展策略[J]. 山东师范大学学报(自然科学版), 2018(04): 454-457; 464.
- [9] 刘冬, 刘更新, 黄祖一. 大数据出版之数值型数据的应用探索——以中国口岸数据库和中国海关统计数据库为例[J]. 科技与出版, 2018(12): 90-95.

作者简介:

霍福华(1984-), 男, 硕士, 讲师. 研究领域: WEB应用开发, 数据分析与处理.