

大数据环境下的电子数据审计:机遇、挑战与方法

陈 伟¹ SMIELIAUSKAS Wally²

(南京审计大学管理科学与工程学院 南京 211815)¹ (多伦多大学罗特曼管理学院 多伦多 M5S 3E6)²

摘 要 电子数据审计的研究与应用是近年来审计领域的热点问题。大数据时代的到来给电子数据审计带来了机遇与挑战。首先阐述了研究大数据环境下电子数据审计的重要性;然后分析了电子数据审计的内涵及原理;在此基础上,重点研究了大数据环境下电子数据审计面临的机遇与挑战,并结合大数据的特点以及目前已有的大数据分析技术与工具,探讨了大数据环境下开展电子数据审计的方法;最后给出了大数据环境下开展电子数据审计的相关建议。研究结果为今后大数据环境下开展电子数据审计提供了理论基础。

关键词 电子数据审计,数据密集型科学,大数据,云计算,计算机辅助审计技术

中图法分类号 C931.6,F239.1 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.1.002

Opportunities, Challenges and Methods of Electric Data Auditing in Big Data Environments

CHEN Wei¹ SMIELIAUSKAS Wally²

(School of Management Science and Engineering, Nanjing Audit University, Nanjing 211815, China)¹

(Rotman School of Management, University of Toronto, Toronto M5S 3E6, Canada)²

Abstract The research and application of electronic data auditing are a hot topic in audit area. The arrival of the era of big data is creating opportunities and challenges for electric data auditing practice and research. However, there are few studies on this issue. In this paper, the importance of researching electric data auditing in big data environments was analyzed firstly. Then, the concept and principle of electric data auditing were analyzed. Then, opportunities and challenges of electric data auditing in big data environments were studied. With the characteristics of big data existing and big data analysis technologies and tools, methods of electric data auditing in big data environments were discussed. Finally, advices for implementing electric data auditing in big data environments were given. Research results in this paper can provide theory for implementing electric data auditing in big data environments.

Keywords Electric data auditing, Data-intensive science, Big data, Cloud computing, Computer assisted audit technologies

1 引言

审计工作一直得到国内外政府和社会的重视。传统手工审计是通过对纸质账簿的检查来实现这一职责的,20世纪80年代,以查账为主要手段的审计职业遇到了信息技术的挑战。传统审计面临着“打不开账,进不了门,审不了数”的困境。随着被审计单位信息化趋向普及,审计对象的信息化使得审计信息化成为必然。审计信息化对审计人员和审计工作的开展提出了更高的要求。我国高度重视审计信息化工作,为了满足审计信息化建设的需要,国家审计署已经成功开展了“金审工程”一期和二期的建设工作^[1-3]。为了探索适合我国国情的联网审计实施方案以及一些数据的采集与分析方法,国家审计署还成功开展了两期国家“863”计划项目,并依托国家“十二五”科技支撑计划项目开展审计信息化的研发及应用示范

工作。另外,“金审工程”三期也即将启动建设。

对我国来说,在信息化环境下如何审计被审计单位的电子数据,发现大案、要案,是政府审计的一项重要任务;国际内部审计师协会(Institute of Internal Auditors, IIA)也高度关注电子数据分析技术^[4]。可见,电子数据审计是目前国内外审计领域关注的重点。2014年12月,国家审计署机构调整,增设了电子数据审计司^[5],其主要职责为:审计电子数据的归口管理;审计电子数据的采集、验收和整理工作;组织开展跨行业、跨部门、跨地区的数据分析工作,对电子数据进行综合分析和利用等。电子数据审计司的增设充分说明电子数据审计在我国目前审计工作中的重要性。

随着信息技术的发展,大数据(Big Data)时代的到来为电子数据审计提供了机遇和挑战。国家审计署刘家义审计长在2013年12月27日的全国审计工作会议上指出:积极跟踪

到稿日期:2014-11-13 返修日期:2015-02-13 本文受国家自然科学基金(71572080),教育部人文社会科学研究规划基金(14YJAZH006),江苏省社会科学基金(13GLC016),江苏省“六大人才高峰”高层次人才项目(2014-XXRJ-015)资助。

陈 伟(1976—),男,博士,教授,主要研究方向为IT审计;Wally Smieliauskas 男,博士,教授,主要研究方向为审计理论。

国内外大数据分析技术的新进展、新动态,探索在审计实践中运用大数据技术的途径,为推动大数据背景下的审计信息化建设做好准备。

综上所述,研究大数据环境下的电子数据审计问题具有重要的理论意义和应用价值。本文结合目前大数据的研究与应用现状,研究了大数据环境下电子数据审计面临的机遇、挑战以及大数据环境下的电子数据审计方法。

2 电子数据审计的研究与应用背景分析

2.1 电子数据审计的起源及内涵

随着信息技术的发展,组织的运行越来越依赖于信息技术(Information Technology, IT)。因此,一方面,在信息化环境下信息技术成为审计的对象,即如何对被审计单位应用的信息技术进行审计,一般情况下多称为信息系统审计(Information Systems Auditing, ISA);另一方面,在审计信息化环境下,信息技术成为审计的工具,即审计人员如何应用信息技术帮助他们开展审计工作,也即计算机辅助审计技术(Computer Assisted Audit Technologies, CAATs)^[6,7]。概括来说,常用的计算机辅助审计技术可以分成两类:一类是用于验证程序/系统的计算机辅助审计技术,即面向系统的计算机辅助审计技术;另一类是用于分析电子数据的计算机辅助审计技术,即面向数据的计算机辅助审计技术,也可以称之为电子数据审计技术。电子数据审计是我国目前开展审计信息化的重点。国际上也高度关注电子数据审计问题,国际内部审计师协会于2011年发布的全球技术审计指南《数据分析技术》中重点分析了面向数据的CAATs在审计数据分析中的应用^[4]。

2.2 电子数据审计的原理

如前文所述,电子数据审计是目前审计工作的重点。在实际的审计工作中,为了避免影响被审计单位信息系统的正常运行,并保持审计的独立性,规避审计风险,审计人员在开展电子数据审计时,一般不直接使用被审计单位的信息系统进行查询分析和检查,而是将所需的被审计单位的电子数据采集到审计人员的计算机中,利用相关软件进行分析,其原理如图1所示。

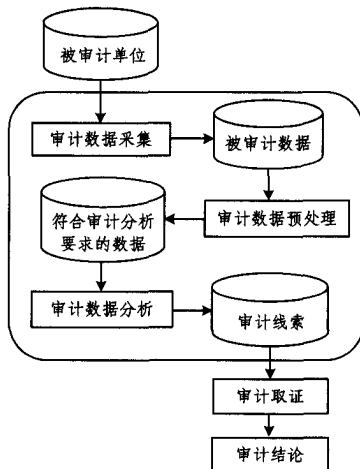


图1 电子数据审计的原理

对于电子数据审计,目前还没有给出明确的定义。根据目前对该术语的使用情况,电子数据审计一般可以被理解为“对被审计单位的电子数据进行采集、预处理以及分析,从而发现审计线索,获得审计证据的过程”。如前文所述,审计署机构调整,增设了电子数据审计司,这也充分说明电子数据审计在目前审计工作中的重要性。

2.3 远程联网电子数据审计——持续审计(联网审计)

审计人员根据审计任务的需要,到被审计单位现场采集电子数据,然后对这些电子数据进行预处理并完成数据分析,获得审计证据,这种开展电子数据审计的方式可称为现场电子数据审计,这是目前电子数据审计的主要方式。近年信息技术的发展使得审计信息化向持续、动态、实时的方向发展,持续审计(或连续审计)(Continuous Auditing, CA)成为审计信息化的一个重要发展方向^[4,8-15]。我国正在研究与应用的联网审计也是分离式持续审计的一种方式,其原理如图2所示^[1-3]。不难发现,相对于现场电子数据审计,我国正在研究与实施的联网审计也可以看成是远程联网电子数据审计,其原理可以看成是一个采用远程联网方式从被审计单位采集电子数据,并对其进行分析,获取审计证据的过程。联网审计技术的应用为审计单位积累了大量的电子数据,这为开展审计大数据分析提供了条件。

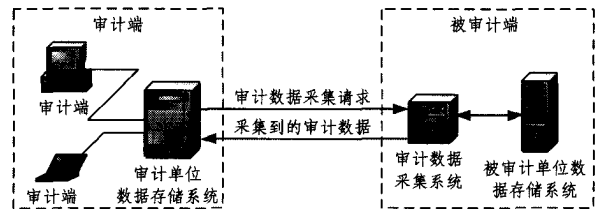


图2 联网审计的实现原理

3 大数据的相关研究分析

3.1 大数据的内涵

如前文所述,目前大数据的研究与应用已经成为国内外的热点。《科学》杂志2011年的专刊讨论了如何管理大数据^[16],Gartner把大数据技术列入全球未来5年10大关键技术趋势之一^[17],我国学者也密切关注大数据领域^[18],大数据技术将会给科学研究、商业、公共管理等领域带来重大变革^[19]。不同的文献给出了大数据的不同定义,最常见的是世界著名咨询机构麦肯锡公司对大数据的定义:大数据指的是大小超出常规数据库工具获取、存储、管理和分析能力的数据集^[20]。Gartner把大数据定义为:大数据是具有大容量、快速、和(或)多样性等特点的信息资产,为了能提高决策、洞察发现和流程优化,这种信息资产需要新形式的处理方法^[17]。

3.2 大数据的特点

概括来说,大数据主要具有以下4个特点^[16,17,19]:

(1)大量(Volume)

数据量大,非结构化数据的超大规模和增长比结构化数据增长快10~50倍;另一方面,计算量较大。

(2)多样性(Variety)

大数据的形式多样,有很多种不同形式,如文本、图像、视频、机器数据等。

(3)快速(Velocity)

一方面,数据量增长速度快;另一方面,大数据要求实时分析,处理速度要快。

(4)真实性(Veracity)

数据必须是准确的、可靠的、一致的,具有可追溯性。

3.3 大数据分析技术与工具

大数据分析需要一些能在有限的时间内对大量数据进行有效分析的技术。为了充分从大数据中挖掘有用的信息,不同种类的大数据技术被研究出来,这些技术覆盖了计算机科学、统计学、经济学等学科。同时,一些用于分析大数据的工具也被开发出来。不同的大数据分析工具有不同的专长,一些主要为批处理数据设计,一些擅长实时数据分析。另外,一个大数据分析工具也有自己特有的功能。概括来说,这些工具可分为3类^[19,21-24]。

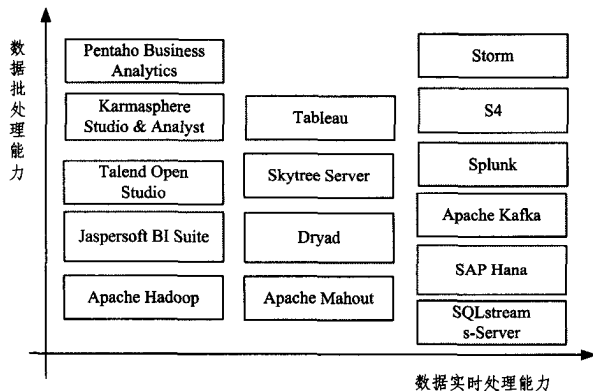


图3 3类工具主要产品的分类比较

(1)批处理工具

这类工具多是基于 Apache Hadoop 框架,常见的包括 Mahout、Dryad 等。以 Mahout 为例,它基于 Apache Hadoop 框架,使用 Map/Reduce 模式,集成了一批包括聚类、分类、频繁模式挖掘等的典型的大数据挖掘和机器学习算法。

(2)流处理工具

这类工具多用于流数据(如日志文件、工业传感器数据等)的实时分析,常见的有 Storm 和 StreamCloud^[23]、SQL-stream^[24]、S4 等。淘宝等企业通过采用流数据挖掘技术分析用户的网上浏览记录来获得有价值的信息,以支持企业决策。

(3)交互式分析工具

用户可以使用这类工具以图表或表格的形式查看、比较和分析数据,这类工具包括 Google's Dremel^[22]、Apache Drill^[21]等。

根据大数据分析工具的数据批处理能力和实时处理能力,3类工具主要产品的分类比较如图3所示^[19,21-24]。

4 大数据环境下电子数据审计的发展机遇

科学研究在经历了实验科学(Empirical Science)、理论科

学(Theoretical Science)、计算科学(Computational Science)这3个阶段后,进入了数据密集型科学阶段(Data-intensive Science)^[18,19,25,26],与之相伴的是大数据(Big Data)时代的到来。大数据时代的到来为各行业提供了机遇和挑战^[27-29]。目前,大数据的研究和应用已经成为国内外的热点。世界各国均高度重视大数据相关问题的研究与探索,并从国家战略层面推出研究规划以应对大数据带来的机遇和挑战。2012年3月,美国奥巴马政府公布了“大数据研究与发展计划”,未来的十年将是一个“大数据”引领的智慧科技时代。正如2012年2月《纽约时报》的一篇专栏文章所称,“大数据”时代已经降临,在商业、经济及其他领域中,决策将日益基于数据和分析做出,而非基于经验和直觉。在公共卫生、经济发展和经济预测等领域,“大数据”的预见能力已崭露头角。亚马逊公司前任首席科学家 Andreas Weigend 认为“数据是新的石油”。IBM 执行总裁罗睿兰认为“数据将成为一切行业当中决定胜负的根本因素,数据最终将成为人类至关重要的自然资源”。麦肯锡认为:大数据可以发挥重要的经济作用,不但有利于私人商业活动,也有利于国民经济和公民。数据可以为世界经济创造重要价值,提高企业和公共部门的生产率和竞争力,并为消费者创造大量的经济剩余^[20]。只要具有适当的政策推动,大数据的使用就将成为未来提高竞争力、生产力、创新能力以及创造消费者盈余的关键要素。同样,大数据时代的到来也为电子数据审计提供了机遇,主要表现在以下方面。

(1)审计取证更充分

如前文所述,随着被审计单位财务和业务数据的电子化,获取被审计单位的电子数据以开展电子数据审计已经成为审计的重要方式。联网审计技术也使得审计单位获取被审计单位电子数据的范围和频率大大增加。大数据环境下,被审计单位提供更多、更全面的数据,审计单位可以充分利用采集来的各方面数据建立集中统一的被审计单位数据中心。在此基础上,借助不同于传统 SQL 关系数据库的新的大数据分析技术,构建审计大数据分析平台和使用更智能的大数据分析技术,通过分析“从数据入口到数据库平台”的更大范围的数据来源,对被审计单位的电子数据进行系统、全面以及跨部门的综合分析,从而解决目前数据分析局限于查找单个问题的缺陷,获得更充分的审计证据,更大地发挥审计的威力。

(2)更多的数据分析技术可供选择

麦肯锡认为:传统的数据分析技术,如关联规则挖掘、分类、数据聚类、遗传算法、机器学习、自然语言处理、神经网络、预测模型等,也可用于目前的大数据分析^[20]。但大数据环境下,开展大数据审计需要更多的智能技术,比如能够针对不同的数据类型进行交叉分析的智能技术、语义分析技术、图文转换技术、模式识别技术、地理信息技术、多媒体数据(图像、音频、视频等)处理技术、社交关系数据处理技术等。

目前,为了满足大数据环境下数据分析的需要,一些专门用于处理大数据的关键技术也被研究出来,如 Big Table、云

计算、分布式系统、Hadoop、HBase、Map/Reduce^[30]、可视化技术等,这为开展电子数据审计提供了机遇。因此,大数据环境下开展电子数据审计时,可以尝试采用这些技术来完成审计数据分析。

(3) 大数据的可视化技术更有助于审计数据的分析

可视化技术是大数据应用的重点之一^[20]。目前,针对可视化分析的需要,一些企业已成功推出一些可视化数据分析软件,如 tableau^[31]、IBM Many eyes 等。可视化审计分析方式能够帮助审计人员快速有效地交互分析大量的数据,所提供的洞察力有助于审计人员更快、更准确地从复杂的被审计数据中发现审计线索。

(4) 审计大数据的实时和快速分析将得以实现

随着云计算、流处理等技术的应用,以及采用粒计算(Granular computing)、量子计算(Quantum computing)等来解决大数据的大计算量技术的研究不断发展,审计大数据的实时处理和快速决策将得以实现。另一方面,大数据环境下的审计数据实时分析能够使联网审计更好地实现实时的审计,真正达到持续审计的目的。

(5) 审计结论更科学

大数据环境下,审计证据的获取、审计报告的形成、审计意见的决策等都可以基于对审计大数据的分析,只要数据可靠,审计结论必然可靠,这使得审计结果更科学。

5 大数据环境下电子数据审计面临的挑战

尽管大数据技术给审计信息化带来了机遇,但在大数据时代开展电子数据审计将面临一些挑战,主要表现在如下方面。

(1) 审计大数据的真实性

大数据环境下,影响数据真实性的因素很多。为了能得到正确、可靠的审计证据,防止大数据环境下的“假账真审”,必须保证被审计的数据是真实的,防范与控制大数据环境带来的审计风险非常重要,其中审计大数据质量控制是关键问题。

(2) 审计大数据的控制和保护

大数据环境下,为了获得全面、可靠的审计证据,需要从众多的被审计单位采集大量敏感和重要的数据来进行分析,这些审计大数据常常会含有一些详细的、潜在的能够反映被审计单位机密的信息,如银行客户的用户名、密码等。这些采集来的数据一般以分布的方式集中存储在审计单位的数据中心,如采用云计算平台方式存储。来自网络的攻击会影响审计大数据的安全,一些对审计数据中心的恶意进攻也会造成更严重的后果,这就需要审计大数据拥有合适的、贯穿审计数据采集、审计数据传输、审计数据存储、审计数据维护、审计数据分析等整个数据生命周期的控制和保护,以降低审计风险。

(3) 审计大数据分析风险

审计大数据的复杂性给数据分析带来了一定困难。大数

据环境下,数据信息全面,隐藏的或未知的信息较多,采集到的大量数据为审计数据分析提供了基础。为了能做到事中审计或者是实时审计,需要强大、高效、实时的审计数据分析方法。另一方面,大数据环境下,数据复杂性也急剧增长,其多样性(多源、异构、多模态、不连贯语法或语义等)、低价值密度(大量不相关信息、知识“提纯”难度高)、实时性(数据需实时生成、存储、处理和分析)等复杂特征日益显著。审计单位现有的计算机系统和审计软件不能应对急剧增长、种类众多的被审计数据,审计大数据的复杂性给数据分析带来了一定困难,造成了审计数据分析的风险。因此,大数据环境下,如果不采用大数据技术实现从传统的审计数据分析向审计大数据分析的过渡,必将影响审计数据分析结果和分析效率,造成一定的审计风险。

(4) 大数据分析平台的选择风险

大数据环境下,开展审计大数据分析需要借助大数据分析平台与工具。目前,大数据分析平台供应商的数目较多,因此,审计单位在采用大数据技术开展电子数据审计时如何选择合适的大数据分析平台非常重要,审计单位和被审计单位应该根据自己的需求,尝试使用多个大数据分析平台来测试应用程序,以选择最佳的大数据分析平台。如果选择的大数据平台不合适,则会造成一定的审计风险。

(5) 审计大数据的全面性尚不够

大数据环境下,审计单位需要访问第三方数据来源并将自己的信息与外部信息进行集成,以充分发挥大数据的潜力。然而,目前尚未建立起数据访问与数据共享机制,这为充分获得大数据的价值带来了障碍,影响了审计取证的查全率,造成了一定的审计风险。

(6) 审计大数据的存储

大数据环境下,已有的审计数据存储技术将不能完全满足大数据环境的需要,被审计单位的大数据为审计数据的存储提出了挑战。研究适合大数据环境的审计数据存储技术成为开展电子数据审计的一项重要任务。

6 大数据环境下的电子数据审计方法探析

6.1 大数据环境下的电子数据审计方法与现有电子数据审计方法的比较

传统环境下,审计人员常采用审阅法、复算法、盘存法、函证法、鉴定法等方法开展审计工作。如前文所述,信息化环境下,审计的对象是电子数据,因此审计证据的获取多是通过采用信息技术对被审计数据进行分析来完成的。随着大数据时代的到来,被审计单位的大数据环境为电子数据审计提出了挑战。由大数据的特点可知,目前信息化环境下现有的电子数据审计方法将不能完全满足大数据环境下电子数据审计的需要,因此大数据环境下需要新的电子数据审计方法。

综合现有文献的分析,大数据环境下的电子数据审计方法与现有电子数据审计方法的比较如表 1 所列^[19,32,33]。

表1 大数据环境下的电子数据审计方法与现有电子数据审计方法的比较

比较内容	现有电子数据审计方法	大数据环境下的电子数据审计方法
被审计数据来源	审计所需要的被审计单位的部分数据,主要是审计交易数据。	审计所需要的被审计单位的各种类型数据,不仅仅包括审计大交易数据,还包括审计大交互数据。
审计数据采集方法	常用的审计数据采集方法主要有直接复制、通过中间文件采集、通过 ODBC 接口采集、通过专用模板采集以及远程联网数据采集等。	审计大数据大环境下,除了采用现有审计数据采集方法之外,一些专门针对大数据的采集方法也可用于审计大数据采集之中。如对于非结构化数据的采集,可以采用网络数据采集方法,这种方法通过网络爬虫等方式从网站上获取审计数据信息;对于系统日志数据采集,可以采用 Hadoop 的 Chukwa、Facebook 的 Scribe 等。
审计数据预处理方法	针对采集来的结构化数据进行审计数据预处理,主要是解决不完整数据、不一致的数据、不正确的数据、重复的数据等问题,其中,名称转换、数据类型转换、代码转换、横向合并、纵向合并、空值处理等是目前电子数据审计数据预处理过程中的常用方法。	不仅仅包括对结构化数据的预处理,还包括如何把非结构化和半结构化的数据通过预处理转化成结构化数据。
审计数据分析方法	常用的审计数据分析方法主要包括:账表分析、数据查询、审计抽样、统计分析、数值分析等。	传统的数据分析技术也可用于目前的审计大数据分析。但为了满足大数据环境下数据分析的需要,一些专门用于处理大数据的关键技术如 Big Table、云计算、分布式系统、Hadoop、HBase、Map/Reduce、可视化技术等被用于审计大数据分析。另外,借助以上技术,审计数据实时分析成为可能。
审计数据存储方法	一般情况下,采用一般服务器来存储数据;联网审计环境下,可以在审计机构构建联网审计的海量数据存储系统,或建立审计数据中心系统。	目前的数据存储技术不能满足审计大数据环境的需要,审计大数据的存储方法发生改变,包括存储设施、存储架构、数据访问机制等。可借助云计算平台进行审计数据存储,但这同时又带来了审计大数据的存储安全问题。

6.2 大数据环境下的电子数据审计方法原理

根据前文分析,大数据环境下的电子数据审计方法如图4所示。

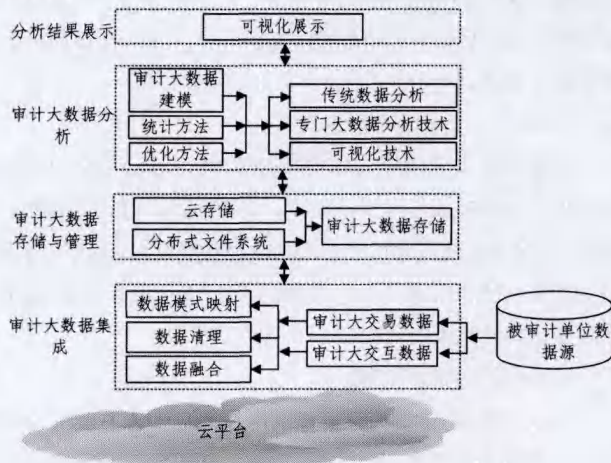


图4 大数据环境下的电子数据审计方法原理

其原理简述如下。

(1) 审计大数据来源

大数据环境下审计数据来源多种多样,主要包括:

① 审计大交易数据。这些数据是传统的结构化的、通过关系数据库进行管理和访问的静态历史数据。

② 审计大交互数据。这些数据源于社交媒体,包括传感器信息、海量图像文件、Web 文本、电子邮件等。

(2) 审计大数据集成

审计大数据集成的目的是把从不同被审计单位或同一被审计单位不同数据源中的各种不同数据整合在一起,这些数据往往涉及诸多数据源,并且它们的数据模式也可能不一样。因此,大数据环境下,在对审计大数据进行分析时,首先需要对这些被审计单位的审计大交易数据和审计大交互数据进行集成和数据预处理,以满足审计大数据分析的需要。

(3) 审计大数据存储与管理

目前的数据存储技术不能满足审计大数据环境的需要,

大数据环境下,审计数据的存储方法发生改变,包括存储设施、存储架构、数据访问机制等。因此,可借助云计算平台或分布式文件系统进行审计大数据的存储与管理。

(4) 审计大数据分析与结果展示

如前文所述,传统的的分析技术如关联规则挖掘、分类、数据聚类、遗传算法、机器学习、自然语言处理、神经网络、预测模型等,也可用于目前的大数据分析。但大数据环境下,开展大数据审计需要更多的智能技术。目前,为了满足大数据环境下数据分析的需要,一些专门用于处理大数据的关键技术也被研究出来,如 Big Table、云计算、分布式系统、Hadoop、HBase、Map/Reduce、可视化技术等。因此,可借助以上技术进行审计大数据的分析与结果展示。

(5) 审计大数据可视化分析技术

信息技术的发展促进了可视化技术^[34]的发展,通过可视化界面,有助于对海量复杂数据的分析。大数据环境下,采集来的审计数据在数量和复杂度上都给审计数据分析带来巨大挑战,数据可视化分析方式有助于审计人员探索、分析和解释复杂的海量数据。可视化提供了审计人员对审计数据的直观分析,审计人员可以通过交互界面对审计数据进行了分析和了解。可视化审计分析的最终目的是呈现审计大数据的内涵,并挖掘出隐藏在审计大数据内部的规律。可视化审计分析是手段,挖掘知识或规律才是目的。需要指出的是:由于大数据的4V特点,相比传统小数据集的可视化,大数据的可视化要相对复杂一些^[35-37]。一般来说,采用可视化手段进行审计数据分析的流程如下:

通过某种可视化软件将被审计数据转化为审计人员可以分析观察的图形和图像。审计人员结合自己的审计背景知识,发挥人类视觉系统高通量的特性,通过视觉系统对可视化的图形和图像进行分析、观察和认知,从而从总体上系统地理解和分析被审计数据的内涵和特征。另外,审计人员交互地改变可视化软件的设置,改变输出的可视化图形和图像,从不同的方面获得对被审计数据的理解,从而全面地分析被审计数据。概括来说,审计大数据可视化分析流程如图5所示。

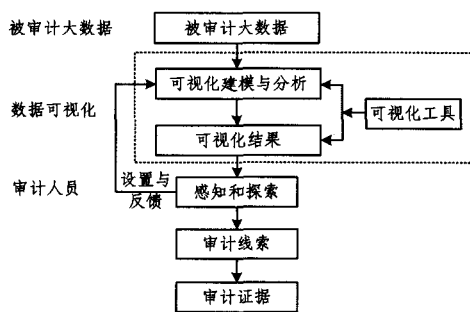


图5 审计大数据可视化分析流程

(6) 云计算与审计大数据存储和分析

如前文所述,大数据时代的到来为电子数据审计提供了机遇和挑战。其中被审计单位的大数据为审计数据存储和分析提出了挑战,已有的审计数据存储和分析技术将不能满足大数据环境的需要,而云计算则为解决大数据时代的电子数据审计面临的挑战提供了一种有效的解决办法。

云计算(Cloud Computing)是一种数据分析技术。云计算3个层次的服务模式包括:软件服务(Software as a Service, SaaS)、平台服务(Platform as a Service, PaaS)和设施服务(Infrastructure as a Service, IaaS)^[38,39]。云计算能够充分利用物理设施的弹性来处理快速增长的数据,被认为是解决大数据问题最有效的技术^[40]。大数据本身就是一个问题集,而云计算则为大数据提供存储、访问和计算。大数据必须有云作为基础架构,才能得以顺畅运营。云计算提供了基础架构平台,审计大数据分析等应用运行于该平台。没有云计算,就没有数据集中采集和存储的商业基础,而云计算为大数据提供了存储空间和访问渠道,大数据则是云计算的灵魂和必然的升级方向^[32]。因此,如何利用云计算技术开展审计大数据存储和分析、获得审计证据、支撑审计报告,是大数据环境下开展电子数据审计将要面临的重要问题。

结束语 总之,大数据时代的到来使得审计工作将不得不面临被审计单位的大数据环境,如果不研究大数据环境下如何开展审计工作,审计人员将再次面临无法胜任审计工作的挑战。本文结合目前大数据的研究与应用现状,研究了大数据环境下电子数据审计面临的机遇、挑战以及大数据环境下的电子数据审计方法。由本文的研究可知,相对于传统的审计数据分析,审计大数据分析更复杂,大数据分析技术还处于发展的初始阶段,现有的大数据技术和工具还不能完全解决目前所有的大数据问题,还需要政府、企业、学术界的共同努力,从软硬件方面出发来研究解决大数据问题。因此,如何设计高效、实时的审计大数据分析系统是今后开展电子数据审计的一项重要任务,这需要我们不断地创新审计技术和审计思路。

综合以上分析,在大数据环境下实施电子数据审计时,应充分利用大数据技术带来的优势,回避大数据技术带来的风险,根据需求和实际情况选择最佳实施方案。另外,根据大数据系统设计的原理^[19,41,42],今后在大数据环境下开展电子数据审计时,设计审计大数据分析系统的建议如下。

(1) 针对不同的需要设计好的系统架构

传统的信息系统架构设计不能满足大数据系统设计的需要,好的大数据系统要有高层次的系统架构设计;但另一方面,即使是一个好的大数据系统架构,也不能适合所有的大数据应用。因此,在大数据系统设计初期,针对不同的大数据应

用系统,需要设计不同的、合适的系统架构。比如,为了满足大数据实时分析的需要,需要采用 NoSQL 技术;为了满足大数据批处理分析的需要,需要采用 Map/Reduce 技术。

(2) 合理使用不同的大数据分析平台

每一种大数据分析平台都有自己的优势和局限性,没有一种大数据分析平台能够解决所有审计问题。如果审计人员能通过使用不同的、合适的大数据分析平台来解决不同的审计数据分析问题,便可充分发现审计证据,降低审计风险。

(3) 能够支持多种审计大数据分析方法

目前审计任务多种多样,如海关审计、银行审计、税务审计、社保审计、固定资产投资审计等,不同的审计任务需要不同的审计数据分析方法。因此,为了满足不种行业、不同种类电子数据审计的需要,所设计的审计大数据分析平台应该能支持多种审计大数据分析方法。

参考文献

- [1] 国家 863 计划审计署课题组. 计算机审计数据采集与处理技术研究报告[R]. 北京:清华大学出版社,2006
- [2] Chen W, Liu S F, Smieliauskas W, et al. Influence factors analysis of online auditing performance assessment: a combined use between AHP and GIA[J]. Kybernetes; The International Journal of Cybernetics, Systems and Management Sciences, 2012, 41 (5/6): 587-598
- [3] 陈伟. 联网审计技术与绩效评价[M]. 北京:清华大学出版社,2012
- [4] Lambrechts A J, Lourens J E, Millar P B, et al. Global technology audit guide (GTAG): Data analysis technologies[M]. The Institute of Internal Auditors, 2011
- [5] State Auditing Administration [OL]. <http://www.audit.gov.cn>, 24 (in Chinese)
国家审计署[OL]. <http://www.audit.gov.cn>
- [6] Chen Wei. A Performance Assessment Method of Online Auditing Based on APP[J]. Journal of Audit & Economics, 2011, 26 (5): 47-52
陈伟. 一种基于 AHP 的联网审计绩效评价方法[J]. 审计与经济研究, 2011, 26(5): 47-52
- [7] IIA. Global technology audit guide (GTAG): Information technology risk and controls(2nd Edition)[M]. The Institute of Internal Auditors, 2012
- [8] Chou C L, Du T, Lai V S. Continuous auditing with a multi-agent system[J]. Decision Support Systems, 2007, 42(4): 2274-2292
- [9] CICA/AICPA. Continuous auditing research report [R]. The Canadian Institute of Chartered Accountants, Toronto, Ontario, 1999
- [10] Alali A F, Pan F. Use of audit software: review and survey[J]. Internal Auditing, 2011, 26(5): 29-36
- [11] Marco A, Giuseppe D, Rob M, et al. What's next for internal auditing? [R]. The Institute of Internal Auditors, 2011
- [12] Vasarhelyi M, Alles M, Kuenkaikaew S, et al. The acceptance and adoption of continuous auditing by internal auditors: A micro analysis [J]. International Journal of Accounting Information Systems, 2012, 13(3): 267-281
- [13] Alles M G, Kogan A, Vasarhelyi M A. Collaborative design research: Lessons from continuous auditing[J]. International Journal of Accounting Information Systems, 2013, 14(2): 104-112

(下转第 34 页)

- Automatica Sinica, 2012, 38(11): 1721-1734 (in Chinese)
- 程强, 陈峰, 董建武, 等. 概率图模型中的变分近似推理方法[J]. 自动化学报, 2012, 38(11): 1721-1734
- [12] Xiao Qin-kun, Gao Song. Application of Bayesian Network in Intelligent Information Processing[M]. Beijing: National Defense Industry Press, 2012 (in Chinese)
- 肖秦琨, 高嵩. 贝叶斯网络在智能信息处理中的应用[M]. 北京: 国防工业出版社, 2012
- [13] Liu Dun, Li Tian-rui, Liang De-cui. Based on rough set three policy decision analysis[M]//Jia Xiu-yi, Shang Lin, Zhou Xian-zhong, et al. Three-way Decision theory and application. Nanjing: Nanjing University Press, 2012: 49-60 (in Chinese)
- 刘盾, 李天瑞, 梁德翠. 基于决策粗糙集的三支政策决策分析[M]//贾修一, 商琳, 周献中, 等. 三支决策理论与应用. 南京: 南京大学出版社, 2012: 49-60
- [14] Jia Xiu-yi, Shang Lin, Zhou Xian-zhong, et al. Three-way Decision theory and application [M]. Nanjing: Nanjing University Press, 2012 (in Chinese)
- 贾修一, 商琳, 周献中, 等. 三支决策理论与应用[M]. 南京: 南京大学出版社, 2012
-
- (上接第 13 页)
- [14] Gonzalez G C, Sharma P N, Galletta D F. The antecedents of the use of continuous auditing in the internal auditing context [J]. International Journal of Accounting Information Systems, 2012, 13(3): 248-262
- [15] Rutgers Accounting Web[OL]. 2014. <http://raw.rutgers.edu>
- [16] Overpeck J T, Meehl G A, Beny S, et al. Dealing with data [J]. Science, 2011, 331(6018): 639-806
- [17] Gartner E S. 10 Critical Tech Trends for the Next Five Years [EB/OL]. <http://www.forbes.com/sites/ericavitz/2012/10/22/gartner-10-critical-tech-trends-for-the-next-five-years>
- [18] Gong Xue-qing, Jin Che-qing, Wang Xiao-ling, et al. Data-Intensive Science and Engineering: Requirements and Challenges[J]. Chinese Journal of Computers, 2012, 35(8): 1563-1578 (in Chinese)
- 宫学庆, 金澈清, 王晓玲, 等. 数据密集型科学与工程: 需求和挑战[J]. 计算机学报, 2012, 35(8): 1563-1578
- [19] Chen C L P, Zhang C Y. Data-intensive applications, challenges, techniques and technologies A survey on Big Data[J]. Information Sciences, 2014, 275: 314-347
- [20] Manyika J, Chui M, Brown B, et al. Big data: The Next Frontier for Innovation, Competition, and Productivity [R]. McKinsey Global Institute, 2011
- [21] Kelly J. Apache drill brings sql-like, ad hoc query capabilities to big data [EB/OL]. <http://wikibon.org/wiki/v/Apache-Drill-Brings-SQL-Like-Ad-Hoc-Query-Capabilities-to-Big-Data>
- [22] Melnik S, Gubarev A, Long J J, et al. Dremel: interactive analysis of webscale datasets[J]. Proceeding of the 36th International Conference on Very Large Data Bases, 2010, 3(1): 330-339
- [23] Gulisano V, Ricardo J P, Marta P M, et al. Streamcloud: an elastic and scalable data streaming system[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(12): 2351-2365
- [24] Sqlstream [OL]. <http://www.sqlstream.com/products/server>
- [25] Bell G, Hey T, Szalay A. Beyond the data deluge[J]. Science, 2009, 323(5919): 1297-1298
- [26] Hey T, Tansley S, Tolle K. The fourth paradigm: data-intensive scientific discovery[R]. Microsoft Research, 2009
- [27] James P A, Bruce H, Gabrielle L, et al. Data-intensive science in the us doe: case studies and future challenges[J]. Computing in Science and Engineering, 2011, 13(6): 14-24
- [28] Lynch C. Big data: how do your data grow? [J]. Nature, 2008, 455(7209): 28-29
- [29] Divyakant A, Philip B, Elisa B, et al. Challenges and Opportunities with Big Data[R]. Cyber Center Technical Reports, Purdue University, 2011
- [30] Deam J, Ghemawat S. Mapreduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1): 107-113
- [31] Tableau [OL]. <http://www.tableausoftware.com>, 2014
- [32] CCF 大数据专家委员会. 中国大数据技术与产业发展白皮书 [R]. <http://www.ccf.org.cn>, 2013
- [33] Chen Wei, Zhang Jin-cheng, Qiu R. A Survey on Computer-assisted Audit Techniques (CAATs) [J]. Computer Science, 2007, 34(10): 290-294 (in Chinese)
- 陈伟, 张金城, Qiu R. 计算机辅助审计技术 (CAATs) 研究综述 [J]. 计算机科学, 2007, 34(10): 290-294
- [34] Simeon S, Michael H B, Arturas M. Visual Data Mining: Theory, Techniques and Tools for Visual Analytics [M]. Springer, 2008
- [35] Geng B, Li Y, Tao D C, et al. Parallel lasso for large-scale video concept detection[J]. IEEE Transactions on Multimedia, 2012, 14(1): 55-65
- [36] Heer J, Mackinlay J D, Stolte C, et al. Graphical histories for visualization: supporting analysis, communication, and evaluation [J]. IEEE Transactions on Visualization and Computer Graphics, 2008, 14(6): 1189-1196
- [37] Thompson D, Levine J A, Bennett J C, et al. Analysis of large-scale scalar data using hixels [C]//IEEE Symposium on Large Data Analysis and Visualization (LDAV). 2011: 23-30
- [38] Chen Wei, Smieliauskas W. Study on Online Auditing Methods in Cloud Computing Environments [J]. Audit Research, 2012(3): 37-44 (in Chinese)
- 陈伟, Smieliauskas W. 云计算环境下的联网审计实现方法探析 [J]. 审计研究, 2012(3): 37-44
- [39] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing [J]. Communications of the ACM, 2010, 53(4): 50-58
- [40] Sakr S, Liu A, Batista D M, et al. A survey of large scale data management approaches in cloud environments [J]. IEEE Communications Surveys & Tutorials, 2011, 13(3): 311-336
- [41] Lee G. Using in-memory analytics to quickly crunch big data [J]. IEEE Computer Society, 2012, 45(10): 16-18
- [42] Nathan M, James W. Big data: principles and best practices of scalable realtime data systems [M]. Manning, 2012