

大数据审计中的数据质量控制与保障机制研究

任依依

陕西国际商贸学院, 中国·陕西 西安 712046

【摘要】在信息技术飞速发展的今天,大数据审计是审计领域的重要发展方向。而数据质量是大数据审计的落脚点,数据质量控制保障机制有效与否直接影响审计工作的质量和效率。本文深入研究大数据审计中数据质量控制与保障机制,研究数据质量的内涵,特征以及重要性,探讨当前的数据质量存在的问题,并从多维度构建相应的控制与保障机制,为提升大数据审计数据质量提供理论与实践参考。

【关键词】大数据审计;数据质量;控制机制;保障机制

引言

随着大数据技术的广泛应用,审计行业迎来了大变革,大数据审计逐渐成为监管领域的主流。在大数据审计过程中,数据是审计工作开展的根本依据,数据质量的高低直接决定着审计结果的准确性和可靠性。但在实际的大数据审计中,数据的质量面临着很多问题,比如数据的完整性,准确性,一致性存在一定的问题,这些问题严重影响了大数据审计的效果。因此,深入研究大数据审计中的数据质量控制与保障机制具有重要的理论和实践意义。

1 大数据审计中数据质量的内涵与特征

1.1 数据质量的内涵

大数据审计的数据质量是指数据满足审计业务需求的程度,包括数据的准确性,完整性,一致性,时效性,可用性等多个方面。准确性,是指数据能够真实,准确地反映被审计对象的实际情况;完整性,是指数据不存在缺失,包括所有必要信息;一致性,是指数据在不同的来源和处理过程中,保持一致;时效性,是指数据能够及时反映当前的情况;可用性,是指数据能够被方便的获取和使用。

1.2 数据质量的特征

数据质量具有多维度的特征,每个维度都影响数据的使用价值。比如数据质量的基本要求是准确性,完整性,而时效性和可用性则影响到数据在审计中的应用效果。在大数据环境下,数据处于不断的更新和变化之中,数据质量也会随着数据的更新而发生变化。因此,需要对数据质量进行持续控制,以保证数据满足审计需求。大数据审计的数据来源广泛,包括内部数据和外部数据,数据类型多样,包括结构化数据,半结构化数据和非结构化数据。这就使得数据质量的

管理变得更加复杂,需要考虑诸多因素的影响。

2 大数据审计中数据质量面临的问题

2.1 数据来源复杂,质量参差不齐

大数据审计数据来源包括被审计单位内部信息系统数据,包括财务系统,业务管理系统产生的数据,以及互联网数据,第三方机构数据等外部数据。内部信息系统数据存在因人工录入产生的字段值错误,像金额数据小数点错位、字符型数据包含非法字符等情况;也有系统接口传输导致的格式不规范问题,例如日期字段存在多种格式表述。外部数据方面,互联网抓取的数据可能因网页结构变化出现字段缺失,第三方机构提供的数据由于采集标准不同,同一指标在定义和计算方法上存在差异,且其真实性和可靠性缺乏统一的验证机制。

2.2 数据处理过程中的质量损失

数据采集环节,采用接口采集时,若接口协议变更未及时更新采集程序,会导致数据采集不完整;网络爬虫采集受网站反爬机制影响,可能出现数据抓取失败或内容错误。数据清洗环节,在进行数据去重操作时,因关键字段选择不当,可能误删有效数据;缺失值处理若采用简单的均值填充,会破坏数据的原始分布特征;异常值处理中,对异常值的界定标准不明确,可能误判正常数据为异常值并进行处理。数据转换环节,不同数据库间的数据类型转换,如将数值型数据转换为字符型数据时,可能因精度丢失导致数据错误;字段映射关系建立不准确,会使转换后的数据与原始数据含义不符。数据存储环节,选择不恰当的存储方式,如将非结构化的日志数据存储 in 关系型数据库中,会导致数据存储效率低,且难以查询,并且存储介

质故障可能导致数据丢失。

2.3 数据质量受数据安全风险的影响

数据在大数据环境下的风险。数据泄露可以通过网络攻击、内部人员违规操作等方式出现,导致审计数据被非法获取、篡改。数据篡改可能是黑客利用系统漏洞对审计数据进行修改,也可能是内部人员未经授权对数据进行变更,降低数据的真实性和可靠性。病毒攻击会破坏数据文件,不能读取数据,或病毒程序修改数据内容影响数据质量。这些安全风险破坏了数据的完整性、可用性和真实性,从而给大数据审计的结果带来负面影响。

2.4 数据质量缺乏统一的标准和规范

当前在大数据审计中,各地区,各部门,各单位在数据采集,处理,存储等方面都缺乏统一的标准。例如,在数据采集阶段,不同单位对同一业务数据的采集字段、采集频率各异;在数据处理过程中,数据清洗的规则、转换的标准不一致,导致数据处理结果的不统一;在数据存储时,数据的命名规则、存储结构不统一规范,导致数据难以共享、难以整合。例如,对于“营业收入”这一指标,不同单位可能采用不同的会计核算方法计算,使得数据无法直接进行比较和分析,给数据质量的评估和控制带来极大困难。

3 大数据审计中数据质量控制机制的构建

3.1 数据采集环节的质量控制

数据采集前,审计人员会同被审计单位相关人员充分沟通,了解被审计单位业务流程、信息系统架构,明确审计目标和具体的数据需求,制定详细的数据采集计划,包括采集的数据范围、采集时间、采集方式等。根据数据源的特点选择采集方式:对于被审计单位内部结构化数据,优先采用接口采集方式,保证数据的实时性和准确性;对于互联网上的非结构化数据,采用网络爬虫采集,针对不同的网站结构特点编制相应的爬虫程序;对于少量的特殊数据,采用人工录入方式,但应制定严格的录入规范。对重要的数据,如财务数据,关键业务数据等采集至少两种方式,并将两种方式采集的数据进行比对和验证,确保数据的准确性和完整性。在数据采集过程中,建立数据采集审核机制,实时对采集的数据进行审核。审核内容包括数据的格式是否符合要求,如日期字段是否为指定格式;数据的逻辑是否合理,如金额数据是否为正数、数量数据是否大于零等,及时发现并纠正数据采集过程中的错误。

3.2 数据清洗环节的质量控制

根据数据的类型、特点和审计需求,制定详细的数据清洗规则。数据去重规则明确关键字段的组合方式,如对于客户信息数据,以“客户编号”和“身份证号码”作为关键字段进行去重;缺失值处理规则根据不同字段的性质和重要程度,采用不同的处理方法,如对于数值型字段,可采用回归预测的方法填充缺失值,对于字符型字段,可采用众数填充或根据业务逻辑进行填充;异常值处理规则确定异常值的判断标准,如利用箱线图法确定数值型字段的异常值范围,并明确异常值的处理方式,如修正或删除。利用大数据技术和人工智能技术,选择先进的数据清洗算法和工具。对于缺失值处理,采用机器学习算法,如随机森林算法,根据其他相关字段的值对缺失值进行预测和填充;对于异常值发现,利用数据挖掘技术中的聚类算法,将偏离聚类中心的数据识别为异常值。建立数据清洗质量评估体系,设定明确的评估指标。数据完整性指标是用来评估数据清洗后数据中缺失值的比例是否降低到可接受范围;数据准确性指标是通过与原始数据或权威数据进行比对,评估数据清洗后数据的准确性是否提高;数据一致性指标是检查同一数据在不同字段中的表述是否一致,评估数据清洗后一致性是否得到改善,并根据评估结果及时调整数据清洗规则和方法。

3.3 数据转换环节的质量控制

制定数据转换前的统一格式和标准。明确数据的字段名称,以保证不同来源的同一种含义的数据使用同一种字段名称;规定数据类型如数值型、字符型、日期型等,明确每一种数据类型的具体格式要求;设定取值范围,如金额字段大于等于零,日期字段要在合理的时间范围内等。依据制定的数据格式和标准,建立数据转换映射关系。对每一个需要转换的字段,明确其在原始数据中的来源字段、转换规则和目标字段。数据转换映射关系需要经过审计人员和数据技术人员共同审核和验证,保证映射关系的准确性和合理性。的数据转换过程中,进行数据转换的全面测试。功能测试检查数据转换是否实现了预期的功能,如字段的转换、计算等是否正确;性能测试评估数据转换的效率,确保在大数据量情况下数据转换能够及时完成;兼容性测试验证转换后的数据在不同的系统和环境中是否能够正常使用,对转换后的数据进行验证和比对,确保数据转

换的质量。

3.4 数据存储环节的质量控制

根据数据的特点和审计需求,选择合适的数据存储方式,如关系型数据库、NoSQL数据库、数据仓库等。不同的数据存储方式具有不同的特点和适用场景,应根据实际情况进行选择。为了防止数据丢失和损坏,建立数据备份和恢复机制,定期对数据进行备份,并确保备份数据的完整性和可用性。在数据出现问题时,能够及时恢复数据,保证审计工作的正常进行。加强数据存储安全管理,采取有效的安全措施,如访问控制、数据加密、防火墙等,防止数据泄露和篡改。

4 大数据审计中数据质量保障机制的构建

4.1 建立健全数据质量管理制度

制定详细的数据质量管理制度,明确数据质量的责任主体、管理流程、考核标准等,确保数据质量工作有章可循。数据质量管理制度应涵盖数据采集、处理、存储、使用等各个环节,形成完整的数据质量管理体系。建立数据质量责任追究制度,对数据质量问题的责任主体进行追究和处理,提高相关人员的数据质量意识和责任意识。

4.2 加强数据质量人才队伍建设

大数据审计需要既懂审计业务又懂数据技术的复合型人才,因此,应加强数据质量人才队伍建设,引进和培养专业的数据质量人才。可以通过招聘、培训、交流等方式,提高审计人员的数据质量意识和专业技能。建立完善的数据质量人才培养体系,定期对审计人员进行数据质量培训,更新审计人员的数据质量知识和技能,提高审计人员的数据质量控制和保障能力。培训内容可以包括大数据技术、数据质量管理方法、数据质量工具的使用等。

4.3 建立数据质量监督与评估体系

建立数据质量监督机制,对大数据审计过程中的数据质量进行实时监督和检查,及时发现和解决数据质量问题。监督可以采用内部监督与外部监督相结合的方式,其中内部监督由审计单位内部的质量控制部门负责,外部监督可邀请专家和第三方机构进行评价和监督。完善数据质量评估指标体系建立科学、合理的数据质量评估指标,对数据质量进行全面、客观的评估。评估指标应包括数据的准确性、完整性、一致性、时效性、可用性等多个方面,能反映数据质量的实际情况。要定期开展数据质量评估活动,

对大数据审计中的数据质量进行评估分析,总结数据质量工作的经验和教训,提出改进措施和建议,不断提高数据质量水平。

4.4 加强数据质量技术创新和应用

加大数据质量技术研发投入,支持数据质量相关技术研发和创新,开发出更先进、更高效的数据质量控制和保障技术和工具,提高数据质量控制和保障的自动化和智能化水平。积极推广应用先进数据质量技术和工具,将大数据技术,人工智能技术,区块链技术等运用到数据质量控制和保障中,提高数据质量控制和保障的效果和效率。比如,利用区块链技术保证数据的不可篡改性,可追溯性,增强数据的真实性和可靠性。构建数据质量技术交流与合作平台,加强数据质量技术领域的交流与合作,推进数据质量技术的共享和创新。

结论

大数据审计中的数据质量控制与保障机制大数据审计是一项复杂的系统工程,不仅需要从事数据采集、清洗,转换,存储等多个环节进行质量控制,同时还要建立健全数据质量管理制度,加强数据质量人才队伍建设,构建数据质量监督与评估体系,加强数据质量技术创新和应用等保障机制。我们只有通过多维度,全方位的努力,才能有效提升大数据审计中的数据质量,为大数据审计工作的顺利开展提供坚实的基础。大数据质量控制与保障机制的完善与创新在未来的大数据技术发展与应用过程中需要不断满足新挑战与需求。

参考文献:

- [1] 刘文玲. 大数据赋能下财务审计质量控制提升路径研究[J]. 国际商务财会, 2025, (08): 71-74.
- [2] 李湘茹. 大数据时代审计工作的质量控制与风险管理[J]. 数字经济, 2024, (07): 104-106.
- [3] 刘唯佳. 大数据背景下会计师事务所审计质量控制研究[D]. 山西财经大学, 2022.
- [4] 李君. 大数据环境下审计数据质量控制研究[D]. 南京邮电大学, 2018.
- [5] 牛艳芳, 王蕾, 孙瑜. 基于PDCA循环的大数据审计质量控制[J]. 新会计, 2019, (09): 42-44.

作者简介: 任依依(1993.11-), 女, 陕西西安, 硕士, 研究方向: 大数据审计、审计质量。