大数据时代数据的分类分级管理及安全防护

洪斌[1]　康涛[2]

1. 金华市金东大数据技术有限公司；2. 金华市金东区大数据发展中心

摘　要：随着大数据时代的到来，数据已经成为驱动科技和商业发展的关键资源。数据的分类、分级管理和安全防护变得尤为重要。本论文旨在探讨大数据时代数据的分类分级管理及安全防护策略，介绍了数据分类分级的重要性，分析了不同维度的分类方法。讨论了数据安全的威胁，并提出了多层次的安全防护措施。探讨了数据管理的策略，包括访问控制、数据生命周期管理和合规性要求。

关键词：大数据；数据分类；数据安全

【DOI】10.12252/j.issn.2096-627X.2022.02.255

引言

随着信息技术的飞速发展，大数据已经渗透到我们生活的方方面面。大数据的涌现为科学研究、商业运营和社会管理带来了巨大机遇，然而，随之而来的是数据的海量和多样性，给数据的分类、分级管理和安全防护带来了前所未有的挑战。在大数据时代，如何科学合理地对数据进行分类分级，同时确保数据的安全性和隐私性，已经成为亟待解决的问题。

一、数据分类分级的重要性

（一）分类分级的意义

在大数据时代，数据的快速增长和多样性已经成为一个显著的特征。在这个背景下，数据分类分级作为一个重要的数据管理策略，具有深远的意义。数据分类分级是将海量的数据按照其特征、属性和用途进行有序的划分和归类的过程。这一过程不仅是对数据进行组织和整理的手段，更是为了更好地满足不同领域需求，提高数据利用效率，加速数据分析和挖掘的速度。

合理的数据分类分级可以带来多方面的益处。首先，它能够帮助组织和企业更好地理解其数据资产，清楚地了解每类数据的特点和价值，从而有针对性地进行数据管理和决策[1]。其次，通过对数据进行分类分级，可以更加有效地组织和存储数据，提高数据的存取效率，减少冗余和重复存储。此外，数据分类分级也为数据的分析和挖掘提供了更有力的支持。在数据量巨大的情况下，通过对特定类别或级别的数据进行重点分析，可以加快分析的速度，从中快速获取有价值的信息。

不同领域对数据的需求各异，因此合理的数据分类分级可以更好地满足这些需求。例如，在医疗领域，对患者的病历数据进行分类分级可以帮助医生更好地了解患者的病情，制定更精准的治疗方案。在金融领域，对交易数据进行分类分级可以帮助银行和金融机构更好地监测风险，预测市场趋势。因此，数据分类分级不仅仅是一种数据管理手段，更是满足不同领域需求、推动创新和发展的重要工具。

（二）数据分类的维度

数据的结构是数据分类的一个重要维度。根据数据的结构特点，可以将数据分为结构化数据、半结构化数据和非结构化数据。结构化数据是指以表格形式存储的数据，如关系型数据库中的数据，这种数据具有清晰的字段和记录，易于管理和查询。半结构化数据则具有一定的结构，但不同记录之间的结构可能会有所不同，如XML和JSON格式的数据。非结构化数据则是指没有固定格式的数据，如文本、图像、音频和视频等，这种数据的特点是信息量大，但处理较为困难。通过对数据的结构进行分类，可以更好地采用适合的工具和方法对数据进行处理和分析。

数据的内容也是数据分类的一个重要标准。根据数据的内容特点，可以将数据分为文本数据、图像数据、音频数据等。不同类型的数据有着不同的特点和处理方法[2]。例如，文本数据可以通过自然语言处理技术进行分析，图像数据可以应用计算机视觉技术进行识别和分析，音频数据则可以使用音频处理算法进行处理。将数据按照内容进行分类，可以更加针对性地应用不同领域的数据分析技术。

数据的来源也是数据分类的一个重要维度。数据可以来自各种渠道和来源，包括传感器数据、社交媒体数据、企业内部数据等。不同来源的数据可能具有不同的特点和质量，因此需要根据数据来源进行分类和管理，以保证数据的准确性和可靠性。

数据的用途也是一个关键的分类维度。数据在不同领域和业务中可能有不同的用途，如市场分析、风险评估、医疗诊断等。通过将数据按照用途进行分类，可以

更好地满足不同业务需求，提高数据的应用价值。

二、数据的安全防护

（一）数据安全威胁

数据泄漏是指未经授权的情况下，敏感数据被泄漏给未授权的人员或组织。随着越来越多的数据存储在云端和网络中，数据泄漏的风险也不断增加。个人隐私数据、金融信息、医疗记录等可能被不法分子获取，导致个人隐私权受到侵犯，甚至可能导致金融欺诈和身份盗窃等问题。

数据篡改是指数据在传输或存储过程中被篡改、修改，从而导致数据的真实性和完整性受到破坏[3]。例如，黑客可能篡改电子商务平台的订单数据，导致支付金额被修改，从而实施盗窃行为。数据篡改不仅可能损害个人用户的利益，还可能影响商业合作伙伴的信任关系。

数据被盗用是指不法分子窃取他人的数据资源，然后用于非法活动。大数据时代的数据价值日益显著，因此黑客和犯罪分子倾向于获取数据以牟取暴利。这种情况可能导致商业机密被披露，公司的核心竞争力受损，甚至国家安全受到威胁。

（二）多层次的安全防护

网络层面的安全防护：在数据传输的过程中，网络是最容易受到攻击的层面之一。为了保护数据在网络传输中的安全性，可以采用防火墙、入侵检测系统（IDS）和入侵防御系统（IPS）等技术[4]。防火墙可以监控网络流量，阻止未经授权的访问和恶意攻击。IDS和IPS可以及时检测异常行为，防范入侵行为。

数据加密技术：数据加密是保护数据机密性的重要手段。通过对数据进行加密，即使数据被盗取，攻击者也无法直接获得明文数据。加密技术可以应用于数据传输和数据存储过程中，确保数据在传输和存储中不被窃取或篡改。

访问控制：访问控制是管理数据访问权限的关键。通过为不同的用户和角色分配不同的权限，可以确保只有授权人员才能访问特定的数据。细粒度的访问控制策略可以减少数据泄漏和滥用的风险。

应用层面的安全防护：应用层面的安全防护涵盖了数据在应用程序中的使用和处理过程。这包括对应用程序的安全漏洞进行修补，确保应用程序不受到恶意攻击。此外，对数据的输入和输出进行有效的检测和过滤也是重要的安全措施，以防止数据注入和恶意代码的攻击。

安全培训和意识提升：在实施多层次安全防护措施的同时，培养员工和用户的安全意识同样重要。通过定期的安全培训和教育，可以帮助员工了解安全风险，学会识别威胁，并采取适当的防范措施。

（三）数据隐私保护

数据匿名化：数据匿名化是一种常见的数据隐私保护方法。通过去除或替换数据中的个人身份信息，使得数据无法与特定个人关联起来，从而保护个人隐私。常见的匿名化方法包括脱敏处理、数据泛化和数据扰动等。匿名化可以在数据共享和数据分析中起到重要作用，使得数据可以被更广泛地应用而不泄漏个人隐私。

脱敏技术：脱敏技术是一种保护数据隐私的有效手段。通过对敏感数据进行处理，如将具体数值替换为范围值或符号，可以减少数据被识别和恢复的可能性。例如，将出生日期脱敏为年龄范围，可以保护个体的具体生日信息。这种方法可以在数据分析中保持数据的有用性，同时保护隐私。

隐私协议和用户授权：在数据收集和使用过程中，明确的隐私协议和用户授权是保护数据隐私的重要手段。在收集个人数据时，需要事先告知数据主体数据的用途、范围和保护措施，并获得其明确的同意。用户可以根据自己的需求选择是否分享数据，从而保护自己的隐私。

差分隐私：差分隐私是一种高级的隐私保护技术，通过向数据添加一定的噪音，使得攻击者难以从数据中识别个体信息。这种方法可以在一定程度上保护数据的隐私，同时保持数据分析的有效性。

三、数据管理策略

（一）访问控制与权限管理

访问控制的重要性：访问控制是一种防御性的安全措施，通过限制对数据的访问来防止未经授权的访问。在大数据环境中，访问控制至关重要，因为数据存储在云端和网络中，很容易受到黑客和恶意用户的攻击。访问控制可以确保只有经过身份验证和授权的人员才能进入系统，并限制其可以访问的数据和操作。

权限管理的实施：权限管理是访问控制的一部分，它涉及对已经获得访问权限的用户授予合适的操作权限。在数据处理和分析过程中，不同用户需要不同的权限来执行特定的任务。权限可以根据用户的角色、职责和需要进行分配，确保用户只能访问他们需要的数据，防止数据被误用或滥用。

角色基础的访问控制（RBAC）：角色基础的访问控

制是一种常用的权限管理方法。通过将用户分配到不同的角色，每个角色都有一组特定的权限，可以简化访问控制的管理。例如，一个金融公司可以设定不同的角色，如管理员、分析师和客户，每个角色都有不同级别的访问权限。

强化身份验证：在访问控制中，强化身份验证是防止未经授权访问的关键。除了用户名和密码外，多因素身份验证、生物特征识别和硬件令牌等方式可以增强用户的身份验证过程，提高系统的安全性。

审计与监控：访问控制与权限管理并不仅仅是建立起来就可以了。定期的审计和监控是确保机制有效性的一部分。通过监控用户的访问行为和权限使用情况，可以及时发现异常活动，并采取适当的措施应对。

（二）数据生命周期管理

数据创建阶段：数据的生命周期始于其被创建的瞬间。在数据创建阶段，需要确保数据的准确性和完整性。数据应该按照事先定义的标准进行采集和录入，以确保数据的质量和一致性。同时，数据的元数据（描述数据的数据）也应该被记录下来，以方便后续的管理和使用。

数据存储与使用阶段：数据存储与使用阶段是数据生命周期的主要部分。在这个阶段，需要根据数据的重要性和访问频率，将数据存储在适当的存储介质中，如硬盘、云存储等。同时，为了确保数据的安全性和隐私性，访问控制和权限管理机制应该得以实施，只有经过授权的人员可以访问和使用数据。

数据归档与备份阶段：随着时间的推移，某些数据可能不再频繁使用，但出于合规性、法律要求或业务需求，仍需要保留。数据归档是将这些数据从主要存储中迁移到较低成本的存储介质中的过程。此外，定期的数据备份是保障数据安全的关键手段，以防数据丢失或损坏。

数据销毁与清除阶段：数据生命周期的终点是数据销毁与清除阶段。在某些情况下，为了保护隐私和防止数据泄漏，数据可能需要被彻底销毁。在销毁数据时，应该采用安全的数据销毁方法，如数据清除、物理破坏等，以确保数据无法被恢复。

数据生命周期管理的关键在于制定明确的策略和规程，以确保数据在不同阶段得到适当的处理和管理。这有助于合理地利用数据资源，避免数据的无谓存储和滥用，减少数据管理的成本和风险。

（三）合规性要求

医疗行业的合规性要求：医疗领域对数据的隐私和保护提出了高标准。HIPAA（美国医疗保险可移植性与责任法案）标准要求医疗机构和相关组织保护患者的健康信息，确保这些信息不被滥用或泄漏[5]。在数据管理中，医疗机构需要采取相应的措施，如数据加密、权限管理等，以保障患者隐私和合规性。

金融行业的合规性要求：金融领域涉及大量敏感金融信息，因此合规性要求尤为严格。GDPR（通用数据保护条例）是欧盟颁布的法规，要求金融机构保护客户数据，并明确规定了数据处理、存储和传输的要求。金融机构需要制定详细的数据保护政策，以确保数据在合规性范围内得到安全处理。

行业标准的遵循：不同行业还可能有各种各样的行业标准和规范。例如，在科研领域，可能需要遵循特定的研究伦理和数据共享原则。无论是金融、医疗还是其他领域，都需要充分了解并遵循相关的合规性标准和法规，确保数据的合法性和安全性。

数据管理策略的制定：在数据管理过程中，制定适应于不同行业合规性要求的策略至关重要。这可能涉及数据保护、隐私政策、用户授权等方面的规定。数据管理策略应该考虑到数据的生命周期，确保数据在不同阶段都能符合合规性要求。

结语

综上所述，合理的数据分类分级有助于提高数据利用效率，而多层次的安全防护措施能够保障数据的安全性。同时，科学的数据管理策略也是数据管理的重要方面，包括访问控制、数据生命周期管理和合规性要求等。随着技术的不断发展，大数据管理与安全领域仍然面临诸多挑战和机遇，需要不断探索创新解决方案。

参考文献

[1]李姝文.数据分级存储与管理研究[J].智能城市，2017，3（01）：64.

[2]张萌.大数据应用中的数据分级管理探究[J].数字通信世界，2021（06）：176-178.

[3]陈安，王翀.基于分类分级的电网企业运营数据全过程保密管理[J].中外企业家，2018（11）：91-92.

[4]张芬.大数据时代数据的分类分级管理及安全防护[J].计算机产品与流通，2019（01）：129.

[5]宄颜邓邓，陶成煦.美国政府数据分类分级管理的实践及启示[J].情报理论与实践，2020，43（12）：172-177+155.