DOI: 10.53104/xdkxts.2026.02.03.004

大数据环境下数据全生命周期安全治理研究

闫磊[1]

1. 长春工业大学，吉林 长春 130012

摘要：大数据应用改变了数据的产生、存储和使用方式，数据不再长期停留在单一信息系统中，而是在采集、传输、存储、处理、共享、归档和销毁等环节持续流动。数据流动范围扩大后，传统以网络边界、系统边界为核心的安全防护方式暴露出明显局限，单独部署防火墙、访问控制或数据加密等安全措施已经难以覆盖复杂的数据使用过程。本文采用文献研究和归纳分析方法，从数据全生命周期角度分析大数据环境中的安全风险。研究认为，当前数据安全治理中的主要问题并不是缺少单项安全技术，而是生命周期各阶段之间缺少连续的安全控制，数据分类分级与具体控制策略衔接不足，不同安全工具之间相互割裂，数据流转后的责任边界不够清楚，安全运营也存在重建设、轻持续治理的问题。针对这些问题，本文提出以数据资产识别和分类分级为基础，将身份认证、权限控制、加密、脱敏、安全审计、数据防泄漏和安全态势感知等技术嵌入数据生命周期，同时建立贯穿采集、传输、存储、处理、共享和销毁过程的责任体系。治理重点应由单一系统安全转向以数据为中心的动态治理，使安全策略能够跟随数据流动并根据数据敏感程度进行调整。在此基础上，本文进一步提出分阶段建设和持续评价机制，为大数据环境下数据安全治理体系的实际建设提供较为简单、可执行的思路。

关键字：大数据；数据安全；数据全生命周期；安全治理；数据分类分级

Research on Data Security Governance Throughout the Full Data Lifecycle in the Big Data Environment

YAN Lei[1]

1. Changchun University of Technology, Changchun 130012, P. R. China

Correspondence to: YAN Lei

Abstract: Big data applications have changed the way data are generated, stored, processed, and used. Data no longer remain within a single information system but continuously flow through collection, transmission, storage, processing, sharing, archiving, and destruction. As the scope of data circulation expands, traditional security approaches centered on network and system boundaries have shown clear limitations. Individual measures such as firewalls, access control, and encryption can no longer provide sufficient protection for complex data activities. Based on literature review and analytical induction, this paper examines data security risks from the perspective of the full data lifecycle. It argues that the major difficulty in current data security governance is not simply the lack of security technologies, but the absence of continuous controls across different lifecycle stages. Other problems include weak connections between data classification and concrete protection measures, fragmented security tools, unclear responsibility after data circulation, and insufficient continuous security operation. To address these problems, this paper proposes a governance framework based on data asset

引用格式： 闫磊. 大数据环境下数据全生命周期安全治理研究[J]. 现代科学探索, 2026, 2(3):33-44.

identification and classification. Technologies including identity authentication, access control, encryption, data masking, security auditing, data loss prevention, and security situation awareness should be embedded into the entire data lifecycle. Meanwhile, responsibilities should cover data collection, transmission, storage, processing, sharing, and destruction. Data security governance should gradually shift from system-centered protection to data-centered dynamic governance so that security policies can follow data flows and adjust according to data sensitivity. A phased implementation and continuous evaluation mechanism is also proposed to provide a practical and understandable approach for full-lifecycle data security governance in big data environments.

Key words: big data; data security; full data lifecycle; security governance; data classification

1 引言

大数据应用使数据逐渐成为信息系统运行、业务决策和智能分析的重要基础资源。与传统数据库环境相比，大数据具有来源多、规模大、类型复杂、处理速度快和流转频繁等特点。企业或机构获取的数据可能来自业务系统、移动终端、传感器、互联网平台和第三方接口，经过清洗、整合和分析以后又可能被多个部门、多个应用系统甚至外部合作机构重复使用。数据所处的位置、访问主体和使用目的不断变化，使安全问题从过去较为明确的系统安全问题转变为贯穿数据流转过程的综合治理问题。

陈东丰等指出，大数据与云计算的融合虽然提升了数据处理能力，但开放和分布式架构也扩大了安全风险的暴露范围[1]。这种变化说明，只关注服务器、数据库或网络出口已经不能完整解释大数据环境中的安全问题。例如，一份敏感数据在数据库中可能得到较好的访问控制，但被导出为文件后可能进入办公终端；经过脱敏后可能进入测试环境；经由接口共享后又可能进入第三方系统。如果安全措施只存在于原来的数据库中，数据离开系统边界后就容易出现保护失效。

数据全生命周期视角能够较好地解决这一问题。吕晓婷、王伟认为，在多源异构、高频流动和跨域计算环境下，传统静态安全边界正在逐渐模糊，需要从数据全流程出发识别安全风险[2]。较早的研究也已经将大数据生命周期划分为采集、存储、分析、交互、应用直至销毁等环节，并指出不同阶段均存在不同形式的安全隐患[3]。从这一角度看，数据安全治理不能理解为某个技术部门部署若干安全产品，而应理解为数据从进入组织到最终退出组织控制范围期间的一组连续管理活动。

现有研究已经从数据加密、隐私保护、风险管理、合规治理以及数据生命周期等角度形成了较多成果，但在实际应用中仍容易出现两个问题。一个问题是研究经常分别讨论某种安全技术，缺少对不同技术在数据生命周期中的位置和关系进行整理；另一个问题是部分治理研究强调制度建设，却没有充分说明制度如何转化为具体技术控制。

本文以数据全生命周期作为分析主线，对大数据环境下数据采集、传输、存储、处理使用、共享和销毁阶段的风险进行梳理，并分析当前安全治理中存在的割裂问题。在此基础上，从数据分类分级、身份权限、加密脱敏、安全审计、风险监测和组织治理等方面构建治理体系，再根据生命周期各阶段提出具体优化措施。本文关注的重点不是增加安全技术的数量，而是解决安全技术和管理措施如何伴随数据持续运行的问题。

2 数据全生命周期安全治理的基本认识

2.1 数据生命周期的范围

数据生命周期并没有绝对统一的划分方式。不同业务系统的数据处理过程存在差异，但从安全治理角度看，可以将其概括为数据采集、传输、存储、处理与使用、共享与开放、归档与销毁几个主要阶段。王红心、龙文佳从大数据获取、存储、分析和使用等阶段构建生命

周期保护模型，强调每一个处理阶段都需要配置相应的安全措施[4]。这种划分方式的价值不在于严格确定阶段数量，而在于将数据安全控制与数据实际运行过程联系起来。数据采集是生命周期的起点，包括业务录入、设备采集、日志采集、网络获取以及第三方数据接入。数据传输是数据在终端、服务器、数据库、云平台和外部系统之间发生移动的过程。存储阶段包括结构化数据库、分布式存储、云存储、文件系统以及备份系统中的数据保存。处理和使用阶段涉及数据查询、修改、清洗、计算、分析和模型训练。共享阶段涉及部门交换、接口调用、文件传输以及对外开放。归档和销毁则解决长期保存价值降低或者保存期限结束后的数据处置问题。需要注意的是，现实中的生命周期并不是严格的直线。数据经过分析后可能产生新的数据，新数据又可能重新进入存储、共享和分析过程。一个用户行为记录经过计算可以形成用户标签，多个标签进一步形成分析报告，分析结果又可能作为其他业务系统的输入。安全治理实际上面对的是不断循环的数据流，而不是简单的“产生—使用—删除”流程。

2.2 数据安全治理目标

大数据环境下的数据安全目标仍然包括保密性、完整性和可用性，但仅依靠这三个传统目标已经不足以描述全部治理要求。沈汀等将企业数据安全与合规治理联系起来，说明数据保护还需要考虑数据处理活动是否符合相应的制度要求以及企业内部责任是否得到落实[5]。换言之，数据即使没有发生泄露，如果采集范围明显超过业务需要，或者数据已经失去使用目的却长期保留，同样说明治理存在问题。本文认为，全生命周期安全治理至少需要实现四个目标。

其一是保证数据在不同阶段具有与风险相匹配的保护强度。普通公开数据和高度敏感数据不应使用完全相同的安全策略，否则容易出现重要数据保护不足或者普通数据保护成本过高的问题。

其二是保证数据流动过程可控。数据安全不能只判断“谁可以访问”，还应回答“访问什么数据、在什么场景下访问、访问以后能否下载、能否继续共享以及是否留下记录”。

其三是保证数据活动可以追溯。发生数据异常后，应能够确定数据来源、访问人员、操作时间、处理方式和流向，为安全事件调查提供依据。

其四是在安全和使用之间保持合理平衡。数据安全的目的不是禁止数据使用。如果将所有数据全部禁止下载、禁止共享，虽然能够减少部分风险，却会降低数据价值。真正有效的治理应根据风险提供不同强度的使用方式。

2.3 从系统安全向数据中心化治理转变

传统信息安全建设经常以系统为单位。一个业务系统上线时配置防火墙、账号权限、数据库备份和日志审计，系统通过安全检查后即认为基本完成安全建设。这种方法在数据相对封闭的情况下能够发挥作用，但大数据环境中的数据经常跨系统流动。张美鸥等提出，数据安全体系可以从组织与人员、制度与流程和技术能力等方面进行建设，并通过风险评估不断验证和完善[6]。马威风、冯波则提出围绕数据全生命周期建设“以数据为中心”的治理体系[7]。两者实际上反映出数据安全治理思路的变化，即安全控制对象应从某台服务器或某个业务系统逐渐转向数据本身。所谓以数据为中心，并不是脱离系统。数据必须存在于具体系统、数据库和网络之中。其核心含义是安全策略应根据数据属性而变化。某类敏感数据无论进入数据库、文件服务器还是云平台，都应继承相应的保护要求。当数据从内部系统共享到外部环境时，安全等级不能因为系统发生变化而自动失效。

基于这一思路，本文将全生命周期治理概括为“识别数据—判断风险—匹配控制—记录流动—持续调整”。其中数据资产识别是基础，分类分级决定控制强度，身份权限和加密脱敏负责具体防护，日志审计和数据流转记录解决追溯问题，风险监测和安全运营负责根据环境变化调整策略。

3 数据全生命周期中的主要安全风险

3.1 数据采集阶段的风险

数据采集决定了后续数据处理的基础。大数据系统常常通过接口、终端、传感设备、日志和第三方平台采集数据，由于采集入口较多，容易产生数据来源不可信、采集范围过大以及终端接口暴露等问题。叶建平在分析大数据安全问题时指出，大量数据集中和应用会同时带来个人隐私和信息安全方面的风险[8]。

采集风险中比较容易被忽视的是“多采数据”。部分系统在建设时倾向于认为数据越多越好，于是将暂时没有明确用途的数据一并采集。数据数量增加后不仅扩大存储成本，也扩大了未来可能泄露的数据范围。如果某业务只需要用户所在城市，却直接采集精确位置，就形成了额外风险。因此，采集安全并不只是防止攻击者伪造数据，还需要控制数据进入系统之前的必要性。

采集接口也是重要风险点。自动化采集通常依赖 API、SDK 或者设备接口，如果接口缺乏身份认证、调用频率限制和参数检查，攻击者可能利用接口批量获取数据。物联网场景中部分终端自身计算和安全能力较弱，攻击者控制终端以后还可能向平台上传伪造数据，破坏数据完整性。

3.2 数据传输阶段的风险

数据完成采集后需要在不同节点之间传输。传输过程中的典型风险包括窃听、数据包篡改、中间人攻击、认证信息泄露以及错误发送。大数据和人工智能环境下，数据在系统之间的连接越来越多，网络攻击、数据篡改和数据泄漏已经成为常见风险类型[9]。

传统网络环境下，企业可能重点保护互联网出口，而内部网络中的数据传输采用相对宽松的控制方式。但是云计算、远程办公和跨部门协同使“内部网络天然可信”的假设逐渐失效。一名合法进入内部网络的攻击者可能利用未加密的数据传输进一步获取敏感信息。

此外，传输风险并不只来自网络攻击。工作人员将包含敏感数据的文件通过普通邮件、即时通信软件发送，或者因为收件人选择错误而将文件发送给无关人员，也属于典型的数据传输安全问题。因此，治理措施需要同时覆盖网络协议和人员操作。

3.3 数据存储阶段的风险

存储阶段通常是数据最集中的阶段，也是攻击者最有可能获取大量数据的位置。大数据环境中的存储方式已经由单一关系数据库扩展到分布式数据库、数据湖、对象存储、云数据库以及离线备份等多种形式。廖芳等在讨论大数据安全保障技术时，将数据加密、访问控制、脱敏和安全审计列为重要安全技术[10]。

存储风险可以分为三个层面。第一个层面是未授权访问。如果数据库权限配置过大，普通业务账号可能访问超出职责范围的数据。第二个层面是存储配置错误。例如云存储桶设置为公开访问，虽然系统本身没有受到攻击，却可能直接暴露大量数据。第三个层面是备份风险。实际安全建设经常重点保护生产数据库，却忽视备份文件。攻击者如果获取未加密的完整备份，同样能够获得全部业务数据。

吴琼在安全存储与检索系统研究中，将数据采集、存储、安全模块和检索模块放在同一系统架构中，并将加解密与检索效率同时考虑[11]。这说明存储安全不能简单理解为“把数据加密”。数据仍然需要被正常查询和调用，安全措施必须与数据库性能和业务使用要求配合。

3.4 数据处理与使用阶段的风险

数据进入分析和应用阶段后，访问人员数量通常增加。开发人员、数据分析人员、运营人员和业务人员可能根据不同任务调用数据。在这一阶段，最主要的问题往往不是外部攻击，而是权限使用超出业务需要。例如，数据分析人员为了完成统计任务可能只需要用户年龄段和消费金额，却直接获得完整用户信息；测试人员为了测试程序功能，将生产数据库复制到测试环境；开发人员为了排查错误，长期保存从生产环境导出的数据文件。这些行为并不一定具有恶意，但都可能扩大敏感数据暴露范围。数据处理还会产生大量中间数据。数据清洗后的临时文件、查询结果、缓存、导出文件和算法训练集都可能成为新的数据副本。如果安全资产清单只记录原始数据库，就无法发现这些副本。大数据环境中真正需要治理的对象不是一份固定数据，而是数据经过复制、加工后形成

的一组关联数据。

3.5 数据共享与开放阶段的风险

共享是数据价值释放的重要方式，同时也是安全治理难度最大的阶段之一。数据在组织内部时，管理者能够直接控制账号、设备和网络环境；一旦数据交付合作单位或者外部平台，原组织对数据的控制能力会明显降低。

区块链等技术已经被用于提高共享过程中的可信程度。例如夏心锋等针对电力营销数据提出结合加密和区块链的数据共享方法，以提高共享安全性和可追溯能力[12]。这类技术能够改善传输可信和操作记录问题，但不能解决所有共享风险。如果本来不应该共享的数据被批准共享，即使整个传输过程使用高级加密，治理行为仍然存在问题。

共享阶段需要同时判断共享对象、共享目的、共享范围和共享时间。某个合作单位获得数据并不意味着可以永久保存，也不意味着可以转交其他机构。数据共享安全的难点正是从“是否允许共享”进一步转变为“允许什么人在什么条件下以什么方式使用”。

API 已经成为数据共享的重要形式。与一次性文件交付相比，API 能够减少大规模文件复制，却产生新的风险。如果接口调用缺乏身份认证、速率限制和异常行为识别，合法接口也可能成为批量获取数据的通道。

3.6 数据归档与销毁阶段的风险

生命周期末端往往是安全治理最薄弱的部分。许多组织非常重视数据采集和使用，却缺少明确的数据保留时间。数据只要进入系统便长期存在，旧服务器、历史数据库、备份磁盘和工作人员终端中积累大量已经没有实际业务价值的数据。

郭子铮提出利用大数据构建网络安全态势感知能力，对安全威胁进行识别和响应[13]。但风险监测如果没有覆盖历史资产，同样可能形成盲区。旧数据库虽然业务访问频率较低，却可能仍然连接网络。一旦旧系统长期不更新补丁，其风险甚至高于正在运行的新系统。

销毁数据也不是简单执行删除命令。普通文件删除经常只删除文件索引，数据内容仍然可能被恢复。云平台和分布式系统还存在多个副本，生产环境删除后，备份和容灾系统可能继续保存。真正的数据销毁应能够确认主要副本、缓存、备份以及相关临时数据是否按照要求完成处理。

表 1 根据上述分析对数据生命周期各阶段的主要风险进行归纳。

表 1 数据全生命周期主要风险及安全控制

生命周期阶段	主要安全风险	典型控制措施
数据采集	过度采集、来源不可信、接口滥用、终端被控制	最小化采集、来源认证、API 鉴权、数据完整性检查
数据传输	窃听、篡改、中间人攻击、错误发送	加密传输、双向认证、文件外发控制、传输日志
数据存储	越权访问、配置错误、备份泄露、密钥泄露	存储加密、权限控制、密钥管理、备份保护
数据处理与使用	权限过大、内部滥用、临时文件泄露、生产数据进入测试环境	最小权限、动态授权、脱敏、操作审计
数据共享与开放	二次传播、接口爬取、第三方管理失控	数据脱敏、API 网关、共享审批、数字水印、第三方安全管理
数据归档与销毁	长期保存、遗留数据暴露、删除不彻底	保留期限管理、安全归档、数据擦除、销毁验证

4 当前数据安全治理存在的主要问题

4.1 生命周期安全控制存在割裂

刘永凤在数据安全风险分析中将人员、信息、技术和管理等因素同时纳入治理范围[14]。现实中的突出问题是，这些安全措施虽然可能已经存在，却分别由不同部门负责。例如网络团队负责防火墙，数据库团队负责数据库权限，业务部门决定数据共享，审计团队检查操作日志。各措施单独来看可能都在运行，但缺少围绕数据对象的统一联系。

这种治理方式容易出现控制断点。敏感数据在数据库内受到严格权限控制，但业务人员能够直接下载为 Excel 文件；文件离开数据库后，原有数据库审计和权限策略立即失效。数据进入测试系统后，也可能因为测试系统安全级别较低而失去原来的保护措施。

全生命周期治理需要解决的不是增加一个新的安全产品，而是让安全规则跨越不同系统保持一致。数据发生位置变化以后，安全策略仍应能够根据分类等级继续执行。

4.2 数据分类分级与技术控制衔接不足

数据分类分级是安全治理的基础，但实践中容易停留在文档层面。企业可能已经制定“普通数据、敏感数据、重要数据”等分类规则，却没有进一步规定每个等级具体使用何种技术措施。如果数据等级仅存在于制度文件中，安全设备并不知道某个数据库字段属于什么等级，也无法自动决定是否允许导出。这样一来，分类分级就只能依靠工作人员记忆执行。真正具有操作性的分类分级至少需要完成两个转换：从管理语言转换成可识别的数据标签，再从数据标签转换成技术控制策略。例如，将身份证号码识别为敏感数据以后，系统应自动限制普通人员查看完整号码；进入测试环境时自动脱敏；通过接口输出时保留调用记录；批量下载时触发审批或风险告警。

4.3 安全工具数量增加但缺少协同

当前数据安全技术已经比较丰富，包括数据库审计、堡垒机、数据防泄漏、加密系统、脱敏平台、API 安全和安全态势感知平台等。李庆丰从技术、业务流程、组织文化和跨界协作等维度讨论数据安全，强调安全问题具有明显的系统性[15]。在实际建设中，机构可能不断购买安全工具，却没有形成统一策略。例如数据库审计发现某账号短时间查询大量敏感数据，但这一异常信息没有传递给身份权限系统，账号仍然可以继续访问；数据防泄漏设备发现文件外发，却无法获得文件所包含数据的分类等级，于是只能根据关键词进行粗略阻断。这种“工具很多、协同很少”的建设模式会增加管理成本。全生命周期安全治理应建立统一的数据资产和风险信息基础，让不同安全工具围绕同一个数据对象形成联动。

4.4 数据流转后的责任边界不清楚

数据共享以后最容易出现“数据已经交出去，原部门管不了；接收部门认为数据来源部门应该负责”的情况。包兴隆、赵倩认为，数据安全管理问题不仅与技术能力有关，也与管理机制和人员安全意识不足有关[16]。从生命周期角度看，责任应跟随数据处理活动变化。数据提供方需要确认是否具备共享必要性以及共享范围是否合理，数据接收方则需要按照约定用途使用和保护数据。共享过程还需要保留审批、交付和调用记录。内部数据流转同样需要责任划分。数据从生产部门进入数据分析平台后，不能简单认为生产部门继续承担全部责任。数据平台管理者应负责平台权限和存储安全，分析人员应对具体使用行为负责，业务部门则负责判断使用目的是否合法合理。

4.5 安全建设重部署、轻持续运营

部分单位将数据安全项目理解为一次性建设任务。安全设备部署、制度文件编制和检查验收完成以后，项目即宣告结束。但数据资产每天都在变化，新数据库、新接口、新应用和新员工持续进入业务环境。如果安全治理不能同步更新，最初建立的安全体系很快就会与实际业务脱节。

数据安全更接近持续运营，而不是一次性建设。安全部门需要定期发现新增数据资产，检查高权限账号，分析异常访问，更新数据等

级，处理安全告警，并跟踪问题整改情况。

医院信息系统研究提供了一个较直观的案例。周士华构建智能运维与数据全生命周期防护体系后，报告的应用结果包括运维效率提升40%、医疗数据泄露事件下降 65%、系统可用性达到 99.9%[17]。这些数据属于特定研究场景，不能直接推广到所有单位，但能够说明一个问题：数据安全治理最终应通过可以观察和比较的指标评价，而不能只以“是否购买安全产品”作为建设成果。

5 数据全生命周期安全治理体系的构建

5.1 建立统一的数据资产清单和分类分级体系

数据资产不清楚，后续安全控制就无法准确落地。很多单位能够统计服务器和网络设备，却不能回答“有哪些敏感数据、存在哪里、由谁负责、流向哪些系统”。全生命周期治理的第一步应从设备清单进一步转向数据资产清单。数据资产清单至少应记录数据名称、来源、所属系统、责任部门、存储位置、数据类型、敏感等级、使用人员、共享对象和保存期限。对于规模较小的单位，可以通过人工登记和数据库扫描相结合的方式建立；数据规模较大时，可以利用自动发现工具识别数据库字段和文件内容。刘春雷在医疗大数据保护实践中提出，数据保护需要与大数据分析研究同步规划和同步建设[18]。这一观点对数据分类同样适用。分类分级不应等到系统上线以后补做，而应在数据进入系统时即确定基本属性。

分类分级不宜设计得过于复杂。普通应用场景可以采用三级或四级结构。公开信息保护要求最低；内部业务数据限制在组织内部使用；敏感数据需要加强访问控制和脱敏；高度敏感或者关键数据则需要审批、强加密和严格审计。

5.2 构建统一身份与最小权限控制机制

数据安全的核心问题之一是“谁能够对什么数据执行什么操作”。传统账号权限往往长期固定，员工岗位发生变化后旧权限未及时回收，导致权限不断累积。全生命周期权限管理需要建立“身份—角色—数据—操作”的对应关系。普通用户可以查看汇总结果，业务主管可以查看所在部门明细，数据管理员可以维护数据库，但不一定需要查看全部明文内容。权限应根据工作需要配置，而不是根据职位高低简单扩大。铁路大数据安全研究将访问安全、数据安全、传输安全和综合安全纳入全生命周期技术体系[19]。访问安全之所以应处于基础位置，是因为大部分数据操作最终都与身份有关。在条件允许时，可以引入动态访问控制。系统除了判断账号身份，还可以结合访问时间、设备、地点、数据敏感等级和行为特征决定是否授权。例如，员工在办公设备上查询少量客户资料属于正常行为，但在非工作时间通过陌生设备批量查询数据时，可以要求二次认证或者直接阻断。

5.3 将加密、脱敏和隐私保护措施嵌入数据流转过程

加密是数据安全中最常见的技术之一，但如果只在数据库存储阶段加密，其保护范围仍然有限。全生命周期治理需要分别考虑传输加密、存储加密和密钥管理。传输阶段可以使用成熟的加密通信协议，降低数据被窃听或篡改的风险。存储阶段可以根据数据敏感程度选择字段加密、文件加密或数据库透明加密。备份文件也应纳入加密范围。密钥不能与加密数据保存在完全相同的位置，否则攻击者同时获得数据和密钥以后，加密作用会明显下降。脱敏主要用于“数据需要使用，但不需要完整真实内容”的场景。例如软件测试只需要验证身份证号格式，可以用随机生成的数据替代真实号码；数据分析只关心年龄区间，可以不提供准确出生日期。

5.4 建立数据流转追踪和安全审计机制

数据安全治理不能只做事前控制，还需要记录实际发生的操作。日志审计应覆盖登录、查询、修改、导出、共享和删除等关键行为。传统日志往往以系统为单位保存，数据跨系统流动后难以形成完整路径。因此，全生命周期治理可以围绕重要数据建立流转记录。例如某数据由系统 A 采集，经数据平台处理后进入系统 B，再通过接口提供给系统 C，整个过程应尽可能保留主体、时间、目的和操作记录。对于敏感文件，可以通过数字水印增加追踪能力。水印

不一定需要直接显示，可以将用户、时间或文件编号等信息嵌入文件。当文件发生异常传播时，可以辅助确认来源。审计机制还需要解决“日志很多但没人看”的问题。大型系统每天可能产生大量访问日志，单纯保存并不能形成安全能力。应根据数据敏感等级设置异常规则，例如短时间批量查询、连续下载、非工作时间访问、首次访问高敏数据等行为可以提高风险等级。

5.5 利用态势感知实现动态风险治理

传统安全策略往往是静态的，一个账号获得权限以后，只要管理员不修改，权限就长期有效。但现实风险会发生变化。账号可能被盗，终端可能感染恶意程序，员工也可能出现异常操作。

网络安全态势感知的思路可以扩展到数据安全领域。系统通过汇总身份信息、访问日志、数据库审计、API 调用和终端行为，对数据活动进行连续观察。安全治理由“事后查看日志”逐渐转向“发现异常后及时调整控制”。

这里并不要求构建复杂人工智能模型。对于普通单位，简单规则同样可以产生效果。例如一分钟内读取记录数超过日常平均水平较多时告警；某员工首次访问不属于本部门的数据时提高风险等级；同一账号短时间从不同地区登录时要求重新认证。

技术重点是形成反馈机制。风险监测发现异常以后，不能只产生一条告警，而应根据风险程度采取限制下载、冻结账号、二次验证或者人工复核等措施。

5.6 建立组织、制度和技术协同机制

数据全生命周期治理不是安全部门单独能够完成的工作。业务部门最清楚数据用途，技术部门负责系统运行，安全部门负责风险控制，管理部门负责制度和责任划分。

比较合理的方式是确定数据所有者、数据管理者和数据使用者。数据所有者主要负责确认数据业务属性、敏感程度和共享范围；数据管理者负责数据库、平台和技术控制；数据使用者需要在授权范围内处理数据并承担相应责任。

制度建设也需要与技术控制对应。例如制度规定敏感数据外发需要审批，技术平台就应提供外发审批功能；制度规定离职人员及时回收权限，身份管理系统就需要与人员状态建立关联。只有管理规则能够转化为系统配置，制度才具有真正的执行能力。

王佑镁等在数据全生命周期治理研究中引入 PDCA 循环思路，通过持续识别风险、实施措施、检查结果和调整治理方式形成动态治理[20]。虽然其研究场景主要涉及教育数据伦理，但这种持续改进方式同样适用于一般数据安全治理。数据安全环境持续变化，治理体系也需要定期重新评估。

6 数据生命周期各阶段的安全优化路径

6.1 采集阶段实施最小化和可信接入

采集阶段应解决“该不该采、从哪里采、采来的数据是否可信”三个问题。业务系统在设计数据字段时，应要求业务部门说明数据用途，不能简单以“以后可能有用”为理由无限增加采集范围。

对于 API 和自动采集程序，需要建立身份认证和接口权限控制。不同系统调用同一数据接口时应使用独立身份，避免多个应用共同使用一个永久账号。接口调用需要配置调用次数限制和异常访问检测，减少批量爬取风险。

设备采集数据还要考虑设备身份。如果传感器或者终端能够随意接入平台，攻击者可以伪造设备上传错误数据。可以通过设备证书、唯一标识和数据校验机制提高来源可信度。

数据进入平台以后立即进行初始分类，比后期统一整理更加有效。系统可以根据字段名称和内容特征自动识别身份证号、手机号、地址等常见敏感字段，再由工作人员确认。

6.2 传输和存储阶段强化加密与密钥管理

数据跨网络传输时，应尽量避免明文协议。内部系统之间的数据传输也不能因为处于内部网络就完全取消加密。文件传输需要建立受控渠道。敏感文件通过普通即时通信软件和个人

邮箱传播，很难继续追踪和撤回。可以使用内部文件交换平台，在平台中控制下载次数、有效期和访问对象。存储阶段应重点保护生产数据库、数据湖和备份系统。不同系统不宜直接共享高权限数据库账号，应用程序应使用独立账号访问所需的数据表。数据库管理员账号需要重点审计。密钥管理是加密体系中容易被忽略的部分。密钥应与业务数据分离保存，并设置合理的权限和更新机制。否则即使数据库已经加密，只要密钥以明文配置在应用服务器中，攻击者仍然可能直接获得。备份系统需要与生产系统使用相近的保护要求。备份的目的是提高可用性，但完整备份通常包含大量数据，本身就是高价值攻击目标。

6.3 处理和使用阶段实行精细权限与环境隔离

数据处理阶段需要严格区分生产、开发、测试和分析环境。真实生产数据不应未经处理直接复制到测试环境。如果测试必须使用类似数据，可以优先采用脱敏或模拟数据。查询权限应与任务匹配。业务人员只需要查看统计结果时，应通过报表系统提供结果，而不是开放底层数据库查询权限。分析人员只需要部分字段时，也应限制不相关敏感字段。批量数据操作需要比普通单条查询设置更高的控制强度。查看一条用户记录可能属于日常业务，但一次下载十万条记录已经形成明显不同的风险。系统可以设置查询量和下载量阈值，超过阈值后要求审批。临时文件也应纳入生命周期管理。数据处理完成以后，临时表、缓存文件和导出结果不能无限期保留。可以配置自动清理周期，减少长期积累的数据副本。

6.4 共享与开放阶段建立“审批—处理—交付—追踪”机制

数据共享前需要判断共享目的、数据范围和接收主体。共享审批不能只记录“同意”两个字，还应记录具体字段、使用期限以及是否允许继续转交。对于不需要真实身份的数据，可以在共享前进行脱敏。对于仅需要统计结果的合作场景，可以直接提供聚合结果，而不是提供原始明细数据。API 共享需要通过统一网关控制。网关能够完成身份认证、调用频率限制、日志记录和异常阻断，使接口不再直接暴露内部业务系统。第三方数据安全责任也需要纳入治理。外部合作单位获得数据以后，应明确保存环境、使用范围、保存时间和删除要求。合作结束后，需要确认数据是否完成删除，而不能默认合同结束就意味着数据自动消失。数据水印可以用于高敏文件共享。在发生异常传播后，通过水印信息能够辅助确认文件的原始接收者。

6.5 归档和销毁阶段建立数据退出机制

数据生命周期管理必须明确“什么时候不再需要数据”。不同数据应设置不同保存期限。业务长期需要的数据可以进入归档环境，不需要继续在线保存在生产数据库。归档数据虽然访问频率较低，但并不意味着安全要求降低。相反，历史数据通常数量大，可能包含多年的用户信息，需要采用独立访问权限和加密措施。达到保存期限的数据应进入销毁流程。数据库记录、文件、缓存和备份需要分别确认。对于存储介质淘汰，还需要防止通过数据恢复工具重新获取内容。销毁过程最好形成记录，包括销毁对象、执行时间、操作人员和验证结果。这样才能证明数据已经真正退出生命周期。

7 数据安全治理体系的实施与评价

7.1 按风险分阶段推进建设

全生命周期治理涉及系统和部门较多，如果一次性要求所有数据达到最高安全标准，实施成本会非常高。普通企业和机构更适合从高风险数据开始。

第一阶段可以进行数据资产盘点，识别重要数据库、敏感字段和主要共享接口。在这一阶段不需要部署大量新系统，重点是解决“有什么数据”的问题。

第二阶段围绕高敏数据落实访问控制、加密、脱敏和日志审计，优先解决权限过大、数据明文传输、测试环境使用真实数据等比较明显的问题。

第三阶段逐步连接身份系统、数据库审计、数据防泄漏和态势感知平台，使分散控制形成协同。

第四阶段建立常态化运营机制，根据新增业务、风险事件和审计结果持续修改数据资产

清单与安全策略。

这种实施方式能够避免安全建设停留在宏观规划，也降低一次性投入过大的问题。

7.2 建立可以量化的安全评价指标

数据安全治理是否有效，应通过实际指标判断。指标不需要非常复杂，可以根据生命周期阶段建立简单评价体系。

表 2 数据全生命周期安全治理评价指标示例

评价维度	指标示例	评价含义
数据资产	数据资产识别覆盖率	判断是否掌握主要数据资源
分类分级	敏感数据分类完成率	判断分类规则是否真正落实
权限安全	超期权限回收率	判断权限是否及时清理
数据使用	高敏数据异常访问数量	判断异常行为变化情况
数据共享	对外共享审批覆盖率	判断共享行为是否进入统一管理
安全审计	关键数据操作日志覆盖率	判断关键行为能否追溯
风险处置	高风险告警平均处置时间	判断安全运营效率
数据销毁	到期数据处置完成率	判断生命周期末端是否受到管理

指标的作用不是追求数字越高越好。例如安全告警数量下降有时意味着风险减少，也可能意味着监测能力下降。因此，应结合多个指标进行判断。在实际应用中，还可以比较安全体系建设前后的数据泄露事件数量、权限违规次数、问题整改周期和系统可用性。前文医院信息系统案例中出现的运维效率、安全事件和系统可用性数据，就属于这种效果评价方式[17]。

7.3 避免安全治理过度化

数据安全治理还需要防止另一个极端，即所有数据都按照最高标准保护。这样虽然表面上提高了安全等级，却可能使普通业务审批复杂、数据共享困难和计算成本上升。合理方式是让保护强度与数据风险匹配。公开数据不需要复杂加密审批，内部普通数据可以通过基本权限保护，敏感数据再增加脱敏和审计，高敏数据才需要强认证、严格审批和持续监控。

技术选择也应考虑组织规模。大型互联网平台可能需要自动数据发现和复杂行为分析系统，中小机构则可以通过数据库权限、加密传输、文件审批和日志检查解决主要问题。安全治理的效果并不完全取决于技术复杂度，而取决于风险和措施是否对应。

7.4 加强人员和流程管理

很多数据安全事件并非复杂攻击造成，而是账号共享、密码管理不规范、文件误发以及权限未及时回收等基本问题。技术工具能够减少人为错误，但无法完全替代人员管理。员工进入岗位时，应明确其数据访问范围；岗位变化后调整权限；离职时立即关闭账号并回收设备。对能够批量访问敏感数据的岗位，可以设置更严格的审计要求。

安全培训也需要与具体业务结合。单纯讲解抽象安全规定效果有限。对普通员工，应重点说明文件外发、账号密码和个人终端使用；开发人员应重点关注生产数据和测试数据隔离；数据库管理员则应重点掌握高权限账号和备份安全。这样可以把全生命周期治理从安全部门的任务转变为参与数据处理的各岗位共同执行的工作。

8 结论

大数据环境下的数据安全问题已经从单一

系统内部的保密问题转变为覆盖数据持续流动过程的综合治理问题。数据从采集开始，会经历传输、存储、处理、共享、归档和销毁等环节，每个阶段均存在不同形式的安全风险。如果仍然按照网络设备、服务器和业务系统分别部署安全措施，容易在数据跨系统流动时出现控制断点。

本文认为，全生命周期安全治理的核心并不是简单增加安全产品，而是以数据作为持续管理对象。数据资产识别和分类分级应作为治理基础，身份认证和最小权限决定谁能够访问数据，加密和脱敏降低数据暴露后的风险，审计和流转追踪保证数据活动能够追溯，态势感知则根据访问行为和风险变化动态调整安全策略。组织制度需要与这些技术措施对应，使数据从进入组织开始直到最终销毁都能够找到明确的责任主体和控制方式。

针对生命周期不同阶段，采集环节应控制数据范围并保证来源可信；传输环节重点采用加密和受控文件交换；存储环节需要强化数据库权限、密钥和备份管理；处理阶段需要控制生产数据使用和临时数据副本；共享阶段建立审批、脱敏、接口控制和追踪机制；归档销毁阶段明确保存期限并验证实际删除结果。不同阶段并不是独立运行，而应通过统一数据标签、访问策略和审计信息连接起来。

对于一般单位而言，没有必要一次建设非常复杂的数据安全体系。可以先从数据资产盘点和高敏数据识别开始，再逐步补充权限、加密、脱敏和审计措施，最后形成跨系统的持续安全运营。通过数据资产覆盖率、异常访问数量、权限回收率、共享审批覆盖率和到期数据销毁率等简单指标，可以判断治理措施是否真正发挥作用。

数据价值来源于合理使用，数据安全的目的也不是限制所有流动。全生命周期安全治理真正需要解决的问题，是使数据在可以使用的前提下保持可控，使不同风险等级的数据承担不同的安全成本，并使数据每一次重要流动都具有明确的权限、责任和记录。这样的治理方式更符合大数据环境下数据持续流动和反复利用的实际特点。

参考文献：

[1]陈东丰，陈扬烯，蔡学泓. 大数据云计算环境下的数据安全探究[J]. 网络安全技术与应用，2026，(7)：62-64.

[2]吕晓婷，王伟. 数据全生命周期安全风险与防护技术分析[J]. 计算，2026，2(4)：40-46.

[3]郑昌年，潘竹. 大数据全生命周期安全隐患分析[J]. 内蒙古科技与经济，2021，(14)：86-87.

[4]王红心，龙文佳. 面向大数据全生命周期保护模型及方法[J]. 网络空间安全，2020，11(2)：1-7.

[5]沈汀，江志杨，姜凌艳，王娟娟. 大数据时代企业数据安全合规治理的关键问题研究[J]. 军民两用技术与产品，2026，(5)：51-53+62.

[6]张美鸥，张旭俊，张丽娅. 大数据安全治理体系实践探索[J]. 网络安全技术与应用，2025，(7)：74-76.

[7]马威风，冯波. 大数据背景下数据安全治理策略研究与实践[J]. 办公自动化，2025，30(7)：124-128.

[8]叶建平. 大数据视域下数据安全问题探究[J]. 科技创新与应用，2025，15(19)：140-143.

[9]王鹏，张弘，张玉，刘爽，吴睿. 大数据与人工智能时代下数据安全的风险及应对措施[J]. 科学技术创新，2025，(15)：75-78.

[10]廖芳，李雪松，杨力，李香冬，罗磊. 大数据应用中的数据安全保障技术研究[J]. 中国新通信，2025，27(3)：34-36.

[11]吴琼. 基于大数据的网络数据安全存储检索系统的设计研究[J]. 中国新通信，2024，26(23)：38-40.

[12]夏心锋，陈泽，周于超，刘莉莉. 面向区块链的电力营销大数据安全共享研究[J]. 电气技术与经

济, 2025, (2) :245-248.

[13]郭子铮. 基于大数据的网络安全态势感知在公安数据安全中的应用[J]. 网络安全技术与应用, 2025, (8) :82-84.

[14]刘永凤. 大数据时代档案数据安全治理探究[J]. 兰台内外, 2025, (24) :9-11.

[15]李庆丰. 大数据时代档案数据安全的多维度分析与保障策略[J]. 机电兵船档案, 2025, (6) :46-48.

[16]包兴隆, 赵倩. 大数据背景下数据安全管理对策[J]. 电子元器件与信息技术, 2025, 9(3) :174-176+180.

[17]周士华. 基于大数据驱动的医院信息系统智能运维与数据全生命周期安全防护研究[J]. 电脑知识与技术, 2026, 22(3) :71-73.

[18]刘春雷. 健康医疗大数据安全保护管理实践与思考[J]. 网络安全技术与应用, 2025, (11) :70-72.

[19]李平, 赵冰, 刘艺飞. 面向全生命周期的铁路大数据安全保障技术体系研究[J]. 中国铁路, 2018, (2) :32-36.

[20]王佑镁, 郭府宁, 王旦, 柳晨晨. 数据全生命周期：教育数据伦理风险防范新视角[J]. 中国教育信息化, 2026, 32(1) :74-83.

版权声明

出版声明