CSA GCR cloud security alliance
GREATER CHINA REGION

数据分类分级实践指南2.0

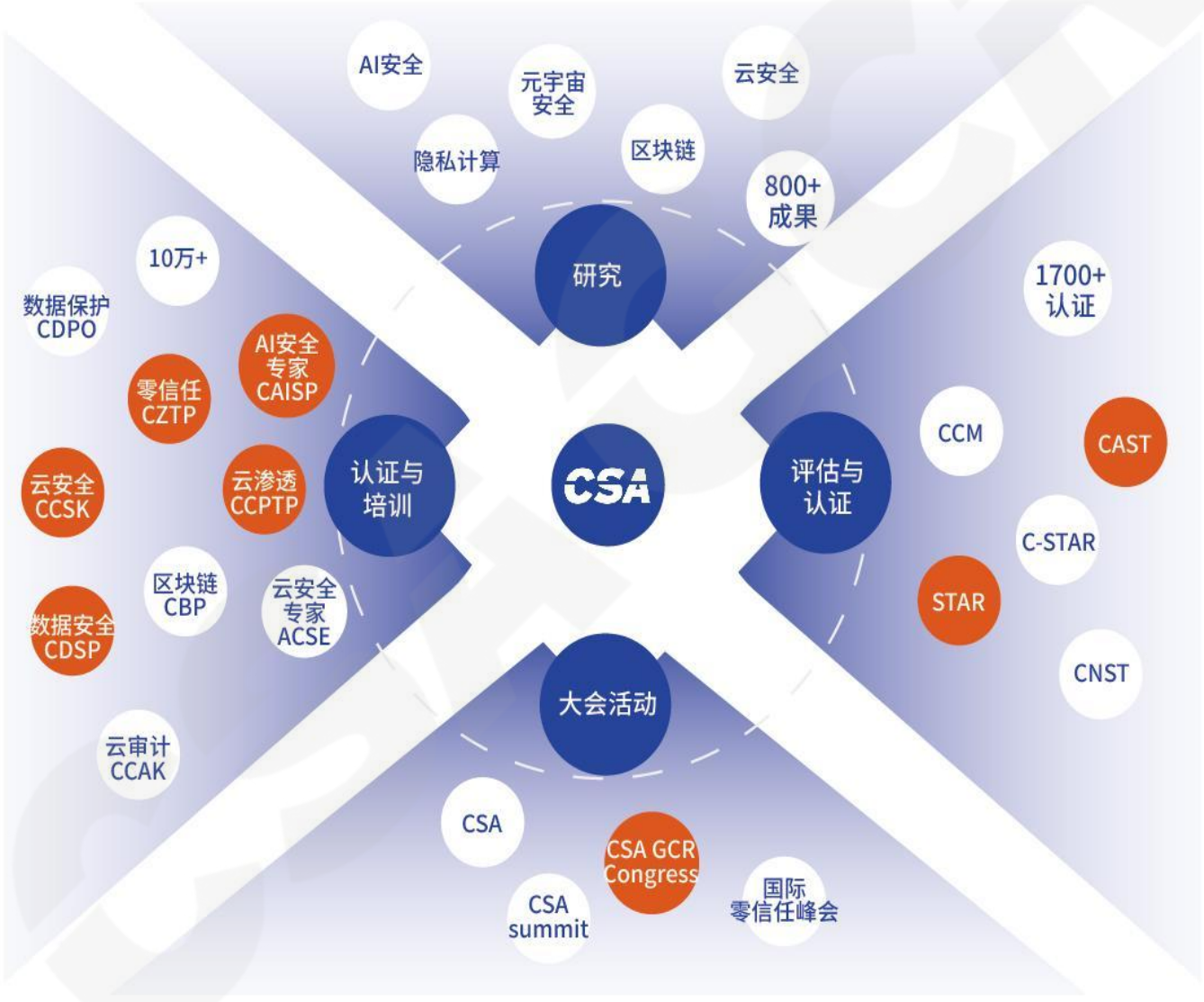
联盟简介

云安全联盟(Cloud Security Alliance,CSA)是中立、权威的全球性非营利产业组织,于2009年正式成立,致力于定义和提高业界对云计算和下一代数字技术安全最佳实践的认识,推动数字安全产业全面发展。

云安全联盟大中华区(Cloud Security Alliance Greater China Region, CSA GCR)作为CSA全球四大区之一,2016年在香港独立注册,于2021年在中国登记注册,是网络安全领域首家在中国境内注册备案的国际NGO,旨在立足中国,连接全球,推动大中华区数字安全技术标准与产业的发展及国际合作。

我们的工作

联盟会刊下载地址
了解联盟更多信息



加入我们

CSA大中华区官网
(https://c-csa.cn)

点击会员

加入联盟

填写相关申请信息

成为CSA会员

JOIN US

致谢

《数据分类分级实践指南 2.0》由 CSA 大中华区数据安全工作组内数据分类分级项目组专家撰写，感谢以下专家和单位的贡献：

组长：

艾龙

专家组：

卜宋博	贺志生	何伊圣	罗智杰
潘万鹏	唐宇	王彪	王玮
谢雄	叶柱	杨岁立	杨天识
于海南			

编委会：

王安宇	陈宏伟	黄鹏华	黄圣超
胡峰	姚凯	李安伦	谢江
廖聪城	鹿淑煜	马兆铭	刘楚楚
刘玉红	刘永亮	李腊梅	李敏波
王亮	王曦光	胡志辉	仇蓉蓉
叶红星	王兴	王贵宗	袁荣婷

研究协调员：

黄家栋	梁嘉荣	易利杰

贡献单位：

天融信科技集团股份有限公司	北京启明星辰信息安全技术有限公司
中兴通讯股份有限公司	上海观安信息技术股份有限公司
中国电信集团有限公司	数篷科技（深圳）有限公司
北京数安行科技有限公司	新华三技术有限公司

（以上排名不分先后）

关于研究工作组的更多介绍，请在 CSA 大中华区官网（https://c- csa.cn/research/）上查看。

在此感谢以上专家及单位。如此文有不妥当之处，敬请读者联系 CSA GCR 秘书处给予雅正！联系邮箱 research@c-csa.cn；云安全联盟 CSA 公众号。

序言

事物分类的管理哲学源于人类对复杂事物进行认知和有效管理的需求。享有“全球第一 CEO”美誉的杰克·韦尔奇曾说过：“管理就是把复杂的问题简单化，把混乱的事情规范化。”在分类管理的过程中，确定管理对象类型化的重要因素就是提炼其共同特征，对事物进行类型化的作用，不仅可以帮助我们对于单个事物的深入认知，还可以促进我们明晰辨别单个类型与整体集群之间的关联关系，让每一个类型在整体中都能进行归类。分级则可以理解为是能够体现数据重要性的一种特有分类，至于为何将分级从其他类别属性中凸显出来，我们可以认为是为了更好地平衡数据的利用与保护，这也正符合我国《数据安全法》的立法原则。

数据作为数字经济时代的核心要素，是国家基础性战略资源，其规模庞大、种类繁多、状态多变。我们在面对数据这一复杂性、抽象性的事物时，深入理解不同的数据属性，对其进行类型化分析，做好数据的分类与分级，是实现精细化数据安全治理的基础，是平衡数据安全保护和流通利用的必经之路。本指南在 1.0 版本基础上扩大知识半径、加深知识理解，从国内外数据分类分级管理现状、数据分类分级概述、数据分类分级能力建设、数据分类分级方法、数据分类分级实施方案、数据分类分级应用等六个方面展开探讨与分析，并提供了国内典型数据分类分级产品介绍、数据分类分级模板工具、数据分类分级参考资料、数据分类分级关键技术与方法、典型行业标准解读、数据分类分级词典示例、文件识别规则示例等参考性、资料性附录。在数字化与智能化交相辉映的时代，我们希望为数据治理、数据安全等领域的从业者及研究者提供有效参考和切实帮助，点亮探索之灯。

李雨航 Yale Li

CSA 大中华区主席兼研究院长

目录

一、国内外政策和现状

1.1 国内外政策现状

随着数字化、智能化时代的到来，数据已成为推进组织数智化转型的关键生产资料，数据安全的重要性日益彰显，数据分类分级已成为数据治理非常重要的环节。2021 年 9 月 1 日，《中华人民共和国数据安全法》正式施行，明确规定“国家建立数据分类分级保护制度”，提出“根据数据在经济社会发展中的重要程度，以及一旦遭到篡改、损毁、泄露或者非法获取、非法使用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，对数据实行分类分级保护”。2024 年 9 月 24 日发布的《网络数据安全管理条例》，第二十九条提出“国家数据安全工作协调机制统筹协调有关部门制定重要数据目录，加强对重要数据的保护。各地区、各部门应当按照数据分类分级保护制度，确定本地区、本部门以及相关行业、领域的重要数据具体目录，对列入目录的网络数据进行重点保护。”

国务院总理李强在第十四届全国人民代表大会第二次会议上做的政府工作报告中，共提到 19 次“数字”和 5 次“数据”，充分体现了我国对数字经济及数据安全的重视程度。中共中央、国务院印发的《关于构建数据基础制度 更好发挥数据要素作用的意见》（以下简称《数据二十条》）提出了建立公共数据、企业数据、个人数据的分类分级确权授权制度。《数据二十条》强调，对于公共数据，要加强汇聚共享和开放开发，强化集中授权使用和管理，推进互联互通，打破“数据孤岛”，以实现数据要素市场化配置的基础，同时也是数据安全高效流通的前提。这些措施将为我国的数据治理提供重要的制度保障。

在标准规范方面，2024 年 3 月 15 日国家标准化管理委员会发布的《数据安全技术 数据分类分级规则（GB/T 43697-2024）》国家标准中，对数据分类规则、数据分级规则、数据分类分级流程等方面给出了清晰的指导。近年来我国各行业各领域在数据分类分级方面持续探索，结合行业特点和业务属性，相继出台了多项数据分类分级标准和规范。金融、工业等行业的监管部门制定了相关配套标准指引，如《金融数据安全 数据安全分级指南》和《工业数据分类分级指南（试行）》；在地方层面，上海、浙江、贵州及武汉等地也相继发布了公共数据开放分类分级的试行指南，加强对政务数据的保护，为落实数据分类分级管理提供了指导性参考。同时，这些地方指南为公共数据治理

提供了良好的保障，有助于提高数据的共享和开放水平。截至目前，我国已有三十余项数据分类分级标准，涵盖金融、证券期货、医疗、电信、互联网、民航、工业、海洋、卷烟制造、能源、媒体、高校及政务等十三个行业（见图 1 数据分类分级标准及归类）。这些标准的出台和落实标志着我国数据分类分级工作已从理论指导阶段全面进入实践阶段，逐步形成了完整的数据分类分级体系。

国际上，数据分类分级一般统称为 Data Classification，根据需要对分类的级别（Classification Levels）和种类（Classification Categories）分别描述。例如，在云安全联盟（CSA）发布的《CSA 数据安全词汇表》文档中，数据分类（Data Classification）被定义为一种安全策略及其实施方法，目的是将信息分为若干类别，每个类别都有相应的安全策略。这些策略可适用于服务器、端点等其他资产，并且某些数据只能在具有相同分类级别的计算机上处理或存储。这一定义反映了数据分类的重要性，尤其是分类后的数据在不同级别上需要采取相应的安全策略。

与我国国家标准《GB/T 37988-2019 信息安全技术数据安全能力成熟度模型》相似，国际数据分类的实践同样强调对不同级别数据的全生命周期保护，要求根据数据的重要性采取不同的保护措施。具体来说，重要数据和核心数据在访问控制、数据加密等方面有着更高的要求，确保数据的机密性、完整性和可用性。这些措施有助于提升数据的安全性，并减少因数据泄露或滥用而带来的风险，为数字经济的发展奠定坚实的基础。

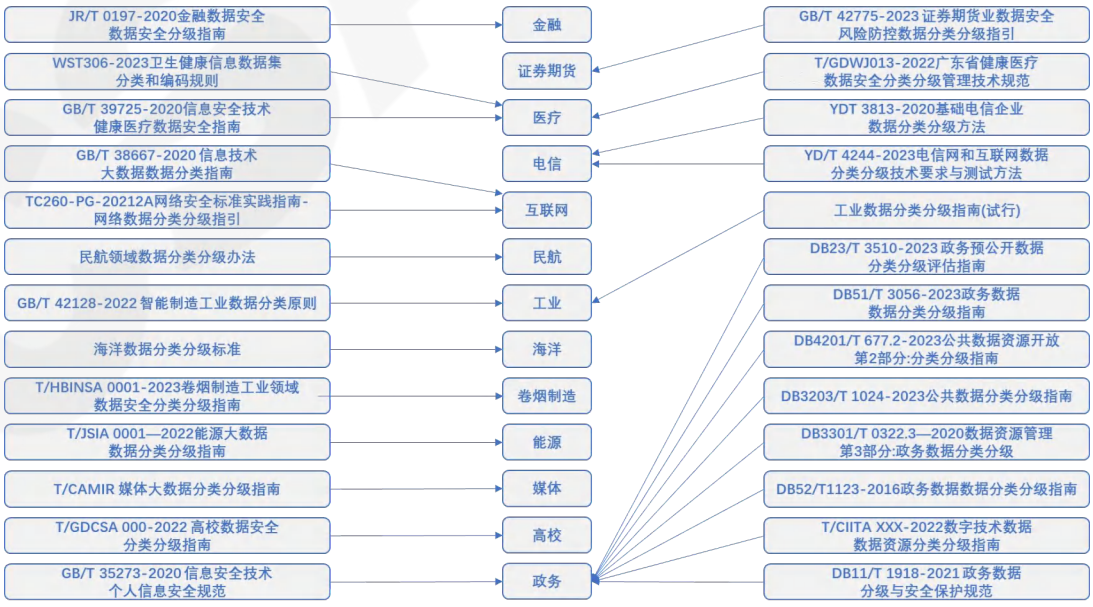


图 1 数据分类分级标准及归类

1.2 国内外技术发展现状

根据 Gartner®最新发布的《2024 年中国数据、分析和人工智能技术成熟度曲线》（见图 2 2024 年中国数据、分析和人工智能技术成熟度曲线），数据和人工智能领域在未来两到五年内将迎来一系列具有颠覆性或高影响力的创新技术的主流应用。这些技术包括复合型 AI、决策智能、国产 AI 芯片、大语言模型（LLM）以及多模态生成式 AI（GenAI）。这些新兴技术的普及将显著推动数据分析与人工智能的创新与发展。

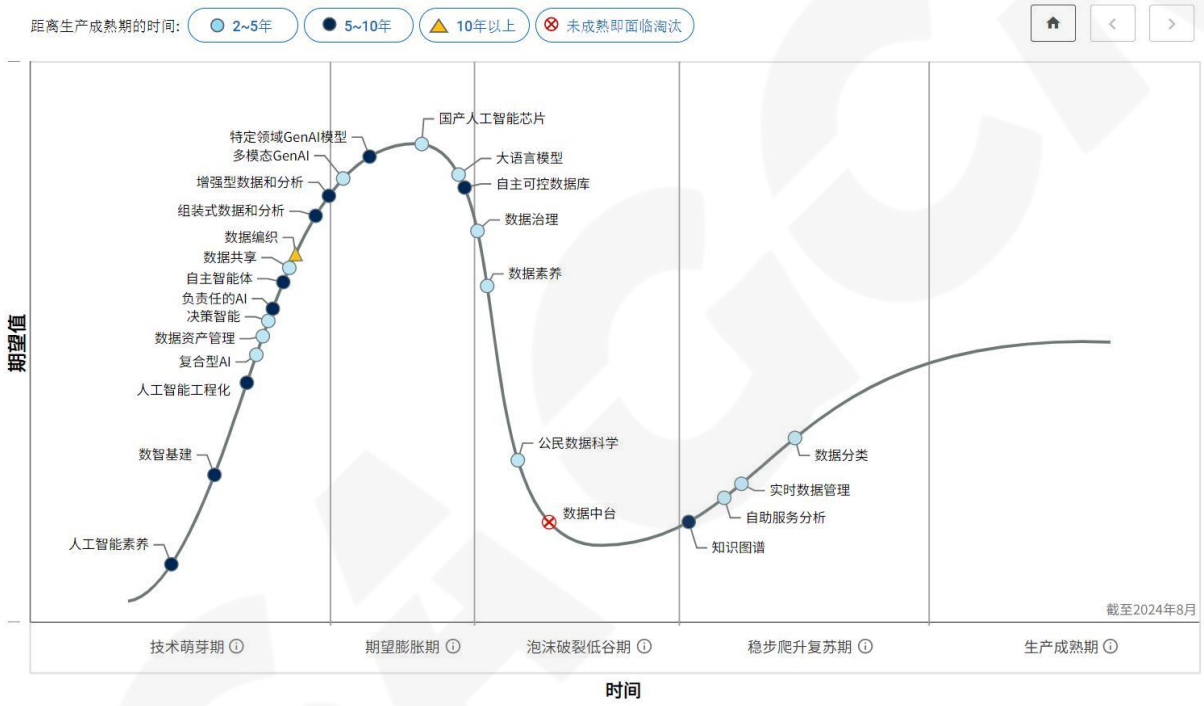


图 2 2024 年中国数据、分析和人工智能技术成熟度曲线

数据分类分级的实现方式，一般分为三种模式。一是人工执行。这是最传统且最常用的数据分类分级方式，实施人员通过类似 Excel 工具表的方式收集供方资产的元数据信息，结合分类分级标准策略进行人工手动执行。该方式要求执行人员熟悉标准策略，完全依赖人工的执行方式虽然灵活性高、可控性强，适用于数据量较少且敏感性较高的情况，但因耗时较长，其分类准确性容易受到人员经验和责任心的影响，整体效率较低。二是人工与工具相结合的方法。人工干预为数据分类提供上下文理解，而技术工具则实现高效的策略执行。这种方式通常在对数据进行初步分类后，通过智能工具自动分级，再由人工进行校验和补充。这种方式的优点在于精准度高，充分利用了人机协作的优势，是目前分类分级准确度最高的一种方式，但实施成本较高，且执行时间较长。三是工具

自动执行。该方式通过标签体系、知识图谱和人工智能等技术，工具与供方资产进行对接扫描，以获取元数据并自动执行分类分级。自动执行速度较快，能够基于已有的标准策略快速实施分类分级，同时还能识别出第一种方式未发现的数据资产。这种方式提高了业务效率，也能帮助供方更全面地识别已知及未知的资产信息，是目前国内主流的分类分级方式，受到主流安全厂家和互联网厂家的广泛推崇。其优点是速度快、覆盖面广，但缺点是对技术工具和系统资源要求较高。

国际上，Gartner 在其《2024 年数据管理成熟度曲线》（见图 3 2024 年数据管理成熟度曲线）中指出，数据分类技术正处于稳步爬升复苏期，但国内的技术发展紧随其后，仅比国际领先水平滞后 1 至 3 年。在政府的大力推动下，国内的数据分类分级能力逐步提升，正在赶超国际水平。例如，国际上的数据管理技术在数据治理、数据共享和数据资产管理等方面已取得显著进展，这为国内的技术发展提供了良好的借鉴和对标参考。

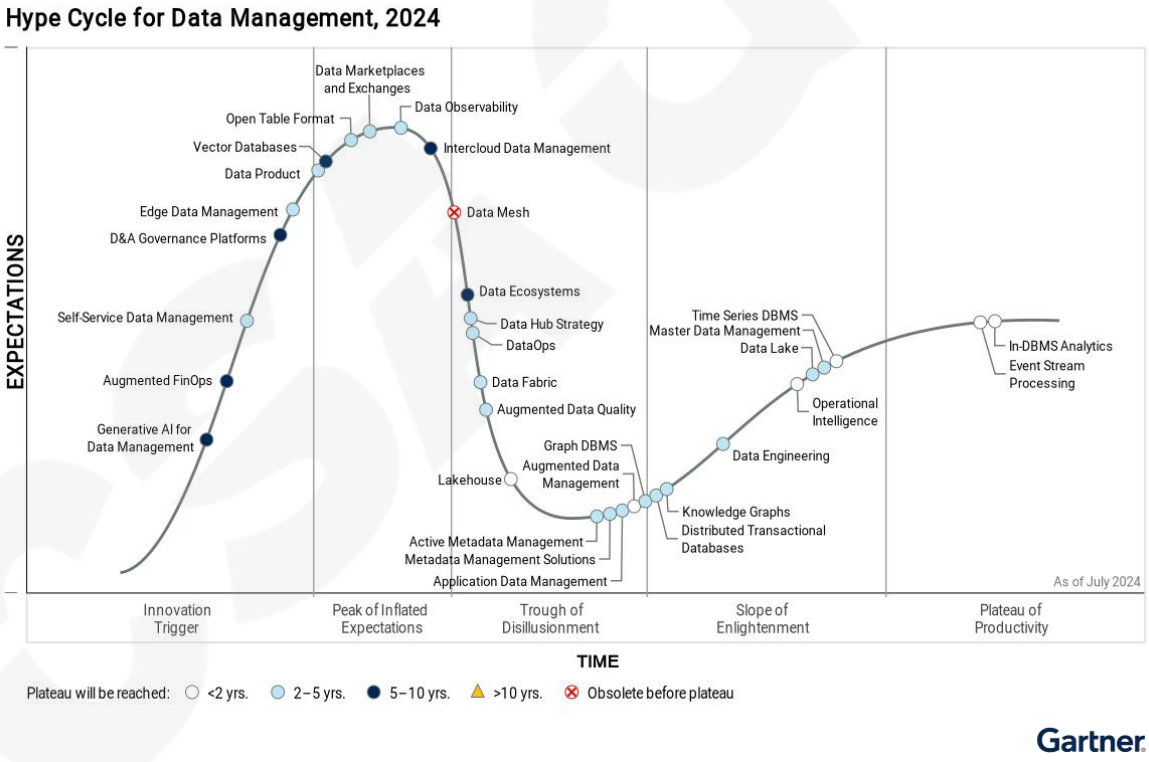


图 3 2024 年数据管理成熟度曲线

1.3 数据分类分级的目标和意义

数据分类分级的主要目标是帮助企业合理分配安全资源，提高数据安全管理的有效性，并确保数据在其全生命周期内得到适当的保护和利用。通过明确数据的价值和敏感性，企业可以制定针对性的数据保护措施，从而降低数据泄露的风险，提升数据管理效率。

数据分类分级有助于推动企业实现以下目标：

1. 提高数据管理效率：数据分类分级将数据按照主题、类型或其他相关属性进行组织，使用户能够更快速地使用所需的数据，提高数据的使用效率。通过对数据进行分类分级，组织能够完成数据资产的梳理，建立数据资产目录，从而有助于组织恰当地管理数据，使数据的存储和使用更加高效和便捷。
2. 支撑数据分析和决策：通过对数据进行分类分级，可以帮助企业更好地理解数据的含义和价值，从而支持数据分析和决策。不同级别的数据可能具有不同的分析需求和应用场景，因此分类分级可以提供更精确的数据选择和使用指南。
3. 有效控制风险：根据数据的分类分级，有针对性地进行风险评估和控制，对数据采取适宜的安全措施，如访问控制、脱敏、加密等，减少数据泄露和安全事故的发生，确保数据的机密性（Confidentiality）、完整性（Integrity）、可用性（Availability）等安全保障的同时，促进数据的开发利用。
4. 合规要求的满足：相关法律法规、国家标准、行业标准对数据的分类分级有明确的规定。组织需遵循相关合规要求，根据规定对数据进行分类分级，并根据合规要求采取相应的安全措施，确保符合法律法规和行业标准，避免违规行为和相关的法律风险。
5. 促进数据开发利用：通过对数据进行分类分级，可以促进数据共享和协作。不同级别的数据可以在不同的范围内共享，使得不同用户或组织可以根据其权限和需求访问和使用数据，促进数据的开发利用。
6. 降低成本：通过对数据进行分类分级，可以根据不同级别数据的特点和需求，合理分配资源和投入，避免资源的浪费和无效使用。

综上所述，数据分类分级的目标是为了更好地管理和保护数据，确保数据的安全。通过合理的分类和分级，提高数据管理效率，支撑数据分析和决策，有效控制风险，保

护数据安全，在满足合规要求的前提下，促进数据开发利用，降低成本，为组织的可持续发展提供支撑。

二、数据分类分级概述

2.1 数据分类分级概念

数据安全法提出“国家建立数据分类分级保护制度”，确定分类分级是国家治理数据安全的重要且基础的工作。在组织实务中，分类分级也是数据管理的重要基础性工作，但没有定义什么是分类什么是分级，不同的法律法规、标准对分类和分级有不同的表述，在实际工作中可能会引起工作结果不一致的情况。为了更好地理解分类分级的概念，本文从相关的法律法规、标准指南分析出发，深入分析数据分类和分级的内涵和概念。

国家标准《GB/T 38667-2020 信息技术 大数据 数据分类指南》 中将大数据分类定义为：“根据大数据的属性或特征，将其按一定的原则和方法进行区分和归类，并建立起一定的分类体系和排列顺序的过程。”贵州省地方标准《DB52T 1123-2016 政府数据 数据分类分级指南》中将政务数据分类定义为：“根据政府数据的属性或特征，将其按照一定的原则和方法进行区分和归类，并建立起一定的分类体系和排列顺序，以便更好地管理和使用政府数据的过程。”将政务数据分级定义为：“按照一定的分级原则对分类后的政府数据进行定级，从而为政府数据的开放和共享安全策略制定提供支撑的过程。”电信行业标准《YDT3813-2020 基础电信企业数据分类分级方法》中将数据分类定义为：“根据基础电信企业业务运营和企业自身管理特点，按照树形结构，建立数据资源分类目录树。并将整理后的数据资源列表对应到目录树，确定数据资源列表中每个数据项在目录树中所在的位置，即确定该数据项的数据类型。”将数据分级定义为：“根据基础电信企业数据重要程度和敏感程度，确定数据资源的安全等级。”浙江省地方标准《DB33/T 2351—2021 数字化改革 公共数据分类分级指南》中将数据分类定义为：“按照公共数据具有的某种共同属性或特征（包括数据对象、重要程度、共享属性、开放属性、应用 场景等），采用一定的原则和方法进行区分和归类，以便于管理和使用公共数据。”将数据分级定义为：“按照公共数据遭到破坏（包括攻击、泄露、篡改、非法使用等）后对国家安全、社会秩序、公共利 益以及个人、法人和其他组织的合法权益（受侵害客体）的危害程度对公共数据进行定级，为数据全生命周期管理的安全策

略制定提供支撑。”国家标准《数据安全技术 数据分类分级规则》（GB/T 43697-2024）中也强调和说明了数据分类分级的相关原则和方法。

综合以上我们认为，数据分类的概念可以归纳为：根据数据的属性或特征，按照规定的原则和方法区分和归类，并建立起一定的分类体系和排列顺序，以便更好地管理和使用数据的过程。数据分级的概念可以归纳为：根据数据的重要程度，以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度按不同等级加以区分，从而确定保护程度的过程。数据分类是建立统一、准确、完善的数据架构的基础，是实现集中化、标准化管理的基础，而数据分级则是数据管控和保护的前提条件。

2.2 数据分类分级面临的挑战

2.2.1 数据量大且复杂度高

随着数字化社会的发展，组织在业务运营过程中通常拥有大量数据，包括结构化数据、半结构化数据（如日志文件、XML 文件）或非结构化数据（如文档、图像、视频等）。不同行业数据的格式和规范往往大不相同，甚至同一行业的不同企业在设计数据目录的结构和元数据的规范都不相同，给企业分类分级规则的制定带来了巨大挑战。

随着时间推移，数据量不断增长，并且数据产生的速度也在不断地加快。现有技术只能做到存量数据的分类分级，而对于实时生成，不断流动的数据分类分级的工作，在选择合适的技术解决方案并将其整合到现有系统中则可能会面临技术挑战，需要庞大的系统资源和计算能力解决数据格式不一致、数据存储在多个地点等的问题。

2.2.2 人为因素引起的判断误差

对数据分类和定级的判断比较依赖人员的经验和对业务的理解。不同的人员的理解和操作方式可能存在差异，可能导致数据分类分级的结果的不一致性。实施人员有意或无意的行为可能导致数据分类和分级的标记错误。例如，某些员工可能会有意或无意将数据错误地分类为非敏感数据，从而造成相关安全控制措施无法及时、准确地实施，为数据分类分级的审核工作带来了挑战。

2.2.3 如何选择分类维度的问题

数据分类具有多维属性，不同的维度有着不同的应用目标和价值导向。首次开展数据分类时，如何选择最合适的维度对数据分类，以达到短期应用效益的最大化，是困扰组织开展数据分类的基础问题。很多时候需要通过一个分类维度实现多个目标，或者将两个分类维度混合以便分类，选择维度过于单一无法满足合规要求和应用需求，选择维度过多，便会造成工作量和资源投入的倍增。组织面对此类局面难以选择。同时，分类维度的不清晰会导致后续基于分类的很多操作都存在问题。

2.2.4 数据等级无法定量判断

针对数据的分级，需要根据数据内容确定其影响对象、影响范围、损害程度等。目前尚无科学的方法和公式支撑构建数据内容的数学模型，因此很难准确定量地对数据内容描述判断。

2.2.5 数据级别数量选择问题

目前数据到底应该分几个级别尚未形成统一，在数据分级时，核心是需要识别核心数据、重要数据和一般数据，在此基础之上，组织需要结合自身业务需求，将一般数据适当划分为更多的子级别，找到合适的级别数量，使得在使用过程中达到效率和安全管控的平衡。过多的分级会给组织管理和实际使用带来不必要的成本和资源投入，级别数量过少又会使得部分场景管控难以得到精细化的管控，出现管控过度或管控不足的情况。

2.2.6 分类分级落地实施困难

数据分类分级如果要做到全闭环管理，需要包括数据资产梳理、数据确权、数据分类分级制度规范制定、元数据标识及管控权限对接，数据分类分级最终目的是要实现数据分级管控，现实情况是大部分组织只能基于国家标准制定数据分类分级制度，并不能完全地理解和掌握其后的逻辑。一方面缺乏相应的数据资产识别和分类分级标识工具技术，另一方面因为分类分级结果无法和系统平台关联，导致数据分类分级结果无法运用，数据分类分级仅停留在制度规范中或纸面上。

三、数据分类分级能力建设

数据分类分级是一项需要多角色协同的、持续的、复杂的、系统性的工程项目，建立成熟的数据分类分级能力体系是保障数据分类分级和安全分级管控工作能够常态化有效运行的前提。

一个成熟的数据分类分级体系需要将组织的业务需求、法律法规、数据的重要性和价值等多种因素结合起来。通过全员广泛参与、构建持续性的工作流程、明确的监督机制，以及对动态变化的适应能力，构成了确保数据分类分级工作顺利实施和长效保障的基石，实现数据安全的精细化管理，从而降低数据安全风险，提升数据利用的效率，促进数据共享和业务创新。

数据分类分级能力建设包含数据分类分级职能架构、管理制度和流程、技术工具建设、持续运营机制等四个主要环节，如下图所示。

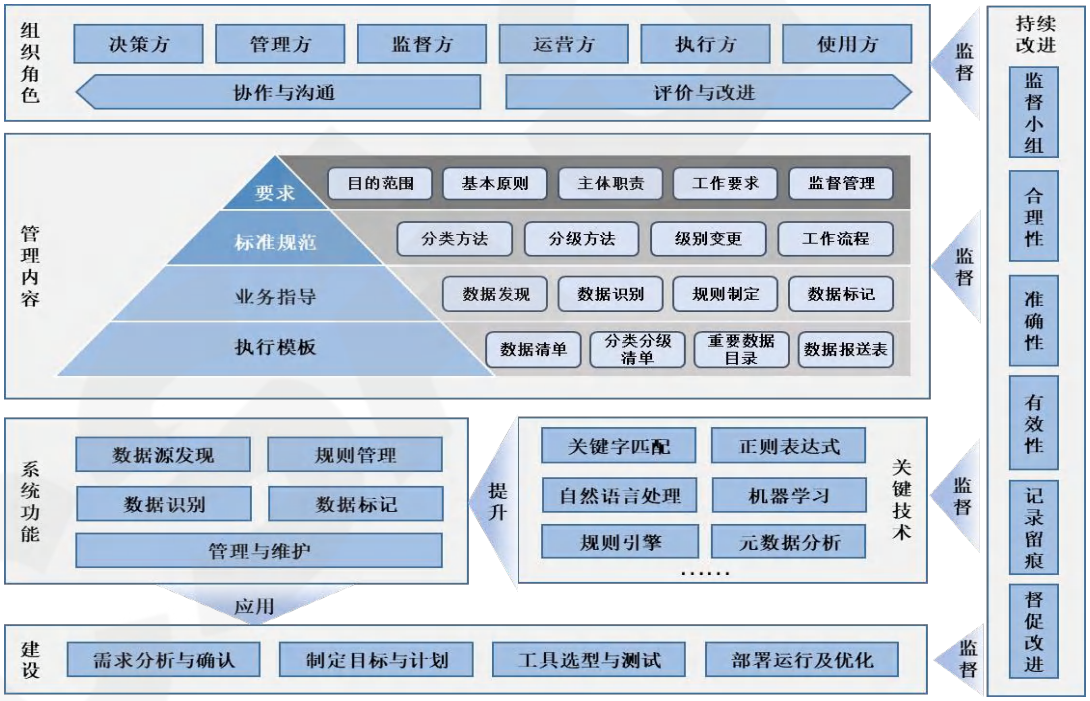


图 4 数据分类分级能力框架

3.1 数据分类分级职能架构

3.1.1 职能划分

数据分类分级职能划分是组织开展数据分类分级的首要步骤，组织决策团队通过理

解数据分类分级的目标、意义和必要性，明确组织内部各角色、各部门在数据分类分级工作中的职责和协作机制，是确保数据分类分级工作能够高效、有序、可持续运营的前提。在开展数据分类分级工作前，组织应当详细说明各部门角色的工作职能，明确数据分类分级决策方、数据分类分级管理方、数据分类分级监督方、数据分类分级运营方、数据分类分级执行方、数据使用方等角色职能。各角色的职能需要更加具体化和详细化，以便小组成员明确自己的工作范围和期望结果。

数据分类分级决策方。在组织首次开展数据分类分级工作时，若业务条线足够复杂、数据种类和规模足够庞大的情况下，建议组织成立数据分类分级专项工作组作为数据分类分级决策方，由高层领导担任组长、各部门主管领导担任成员，确保数据分类分级工作得到充分的重视和足够的资源投入。负责确定数据分类分级工作目标和管理要求，负责确保数据分类分级工作的总体方向与组织的数据治理策略一致，定期听取数据分类分级工作汇报，并指导、监督数据分类分级管理工作。

数据分类分级管理方。由数据分类分级专项工作组指派中高层管理人员作为管理负责人。负责制定和维护数据分类分级管理制度、技术规范和业务指导，定期与业务部门沟通，确保标准的实用性和适应性，组织各部门有序执行数据分类分级工作。

数据分类分级监督方。负责监督、指导数据分类分级管理制度建设、技术工具建设，明确监督检查方式和方法，负责定期监督检查各部门数据分类分级考核指标达成情况，定期提出改进建议。

数据分类分级运营方。负责建设数据分类分级技术工具，维护数据分类分级成果，监测数据资产分布与流转态势，监测数据分类分级的效果。

数据分类分级执行方。各业务系统的归口管理部门负责本领域业务数据的分类分级工作，并依据数据分类分级结果，在数据全生命周期执行数据安全管理要求。

数据使用方。内部各业务部门和外部合作组织作为数据使用方，应依据数据安全级别在数据使用过程中严格执行数据安全防护工作。

在数据分类分级过程中，职责的划分不是绝对的，不同组织可以有不同的划分方式。重要的是需要建立一个明确的数据分类分级责任体系，确保数据的正确分类分级和处理。同时，需要加强跨职能部门的沟通和协作，确保数据的分类分级工作与组织的业务需求和战略目标保持一致。

3.1.2 协作与沟通

各角色之间的协作和沟通机制也十分关键。组织需要建立一个协作平台，比如定期的工作会议、工作报告制度，以及在线协作工具，以促进信息共享和解决问题。特别是在跨部门协作方面，需要明确接口和沟通频次，确保数据分类分级工作不受部门壁垒的影响。

3.1.3 评价与改进

评价与改进包括定期的绩效评价、目标设定以及激励措施，确保各角色能够高效地完成既定任务。同时，应当设立机制鼓励创新和提出改进建议，持续优化分类分级的工作流程和标准。在开展数据分类分级管理时，全员持续参与至关重要。从最高决策层到日常操作人员，每个人都需对数据分类分级的重要性有所认识，并将之融入其日常职责中。这种全面的参与确保了分类分级的原则和流程能够在组织中得到一致的理解和执行，从而增强了数据安全治理的整体效力。同时，这也意味着数据分类分级工作不仅需要顶层的推动，还需要形成一种贯穿组织文化的自下而上的支持力量。

3.2 数据分类分级管理体系

3.2.1 数据分类分级制度规范

数据分类分级管理方应当制定并维护数据分类分级管理制度，明确数据分类分级的要求和规范，确保组织管理要求与国家法律法规、组织业务战略、数字化转型战略目标相一致。一般情况下，组织的数据分类分级管理制度可以由《数据分类分级管理办法》《数据分类分级规范》《数据分类分级业务指导》及对应的执行模板等几个部分组成。

《数据分类分级管理办法》可以包括数据分类分级管理工作的目的依据、适用范围、基本管理原则、管理主体、执行主体、数据分类分级各项工作要求、监督管理、罚则等主要内容。其中数据分类分级各项工作要求应当重点明确数据分类、数据分级、级别变更、数据汇聚等重点场景下的要求。

《数据分类分级规范》作为管理办法的下级文件。应当针对数据分类分级的原则、数据分类方法、数据分级方法、数据分类分级变更、数据分类分级流程等内容进行规范说明，用于形成组织内部一致的方法论和技术规范。

《数据分类分级业务指导》应当从数据分类分级实施执行的视角，明确业务活动识别、数据发现、数据识别、分类分级规则制定、分类分级标记等操作方法和过程，为各部门按照规范、使用技术工具和模板开展数据分类分级提供实操指导。

《数据分类分级执行模板》应当包括但不限于业务活动识别模板、数据清单模板、数据分类分级模板、重要数据目录模板、数据报送表模板等相关执行文件。

3.2.2 数据分类分级工作流程

数据分类分级工作是一项持续管理的工作，主要涉及两个场景：全量数据分类分级流程和增量数据分类分级流程。

全量数据分类分级是数据分类分级执行方依据组织《数据分类分级规范》对各自责任范围内的全部数据数据分级，形成《数据分类分级清单》，并提交数据分类分级管理方审核。数据分类分级管理方审核提交的《数据分类分级清单》，确定结果符合数据分类分级原则和规范，并应指出分类分级不合理情况，并协助数据分类分级执行方调整定级结果，由数据分类分级管理方汇总合理的定级结果。

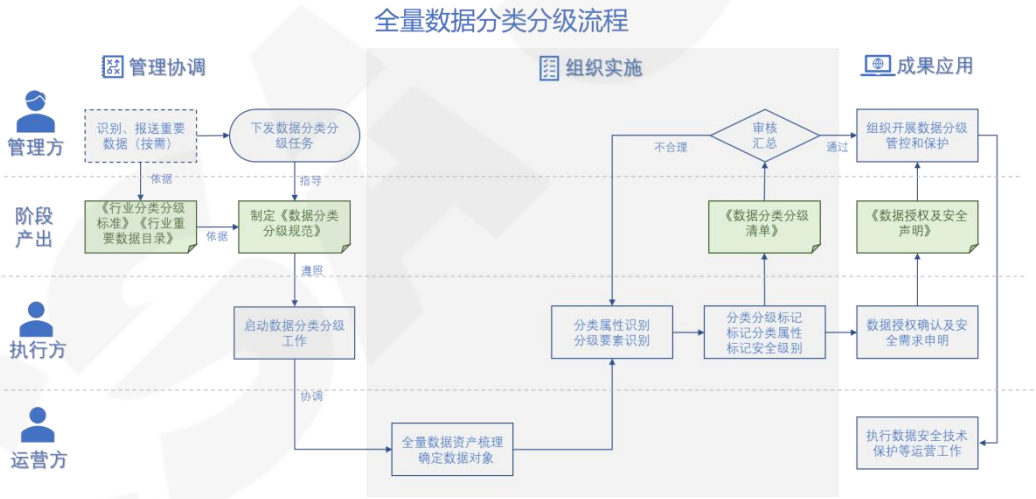


图 5 全量数据分类分级流程图

- 职责分工

数据分类分级执行方，作为数据分类分级主体责任方，负责数据梳理、数据类别识别、定级要素识别及数据安全级别判定，形成《数据分类分级清单》。

数据分类分级管理方审核确认数据分类分级执行方的分类分级结果，对分类分级不合理结果则协助数据分类分级执行方调整定级。对合理结果汇总并组织开展数据安全保护工作。

数据分类分级运营方负责将分类分级结果录入数据分类分级工具。

- 流程步骤

表 1 全量数据分类分级流程步骤

编号	责任方	过程	描述	产出
1	数据分类分级管理方	下发任务	下发数据全面分类分级任务	《数据分类分级清单》 《数据分类分级结果声明及授权》
2	数据分类分级执行方	接收任务	接收定级任务，获取《数据分类分级清单》模板	
3	数据分类分级执行方	数据梳理	对业务系统的全量数据进行全面梳理。	
4		数据盘点	定期对全部存量数据进行盘点。	
5		分类属性识别 分级要素识别	根据《数据分类分级规范》进行分类分级要素识别。	
6		数据分类分级标记	系统上线前完成数据分类分级工作，结果提交数据分类分级管理方。	
7	数据分类分级管理方	定级结果审核汇总	对分类分级结果进行审核，不合理结果协助定级方调整。	《数据分类分级清单》汇总
8			合理结果汇总	
9	数据分类分级运营方	固化分类分级结果	将《数据分类分级清单》录入数据分类分级工具。	
10	数据分类分级管理方	数据安全管理及保护	根据分类分级结果，组织开展数据安全管理及保护。	

增量数据是指信息系统在已完成数据分类分级的基础上，由于系统升级、数据采集或导入等操作导致数据库结构发生变化而新增的数据。一旦出现增量数据，数据分类分级执行方应再向数据分类分级管理方重新提出分类分级申请。增量数据分类分级场景中，应基于数据的提供方、生产方、采集方提供的数据资产清单对采集、导入的增量数据加以梳理和分类分级，并在数据进入存储环境前完成数据分类分级。

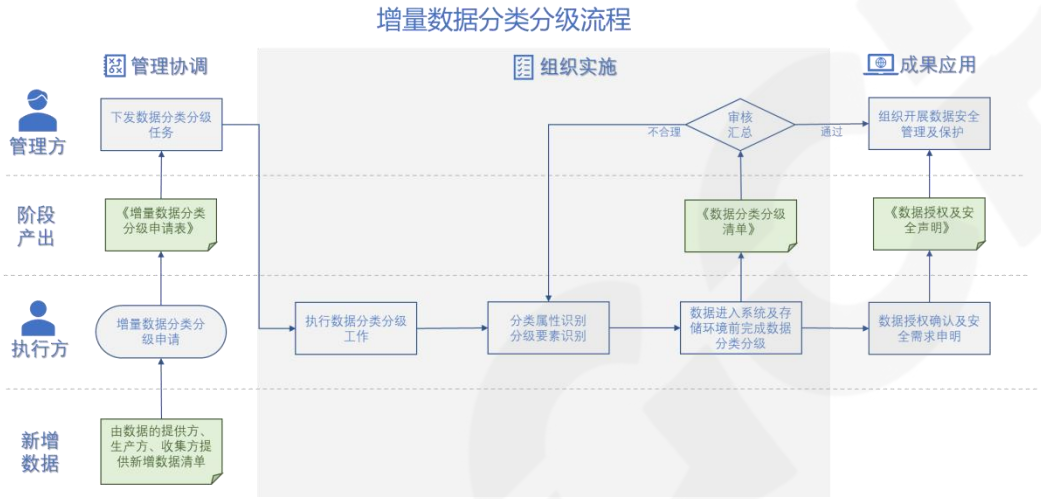


图 6 增量数据分类分级流程图

- 职责分工

数据分类分级执行方作为数据分类分级主体责任方，负责数据资产梳理、分类分级要素识别及数据分类分级判定，形成《数据分类分级清单》。

数据分类分级管理方审核确认数据分类分级执行方分类分级结果，对定级不合理结果则协助数据分类分级执行方调整定级。对合理结果进行汇总，并组织开展数据安全保护工作。

数据分类分级运营方负责将分类分级结果录入数据分类分级系统。

- 流程步骤

表 2 增量数据分类分级流程步骤

编号	责任方	过程	描述	产出
1	数据分类分级执行方	提出申请	信息系统出现增量数据时，提出增量数据分类分级申请。	《增量数据分类分级申请模板》《数据分类分级清单》
2	数据分类分级管理方	接收申请	接收增量分类分级申请，并下发增量分类分级任务	

3	数据分类分级执行方	接收任务	接收分类分级任务，获取《数据分类分级清单》模板	
4	数据分类分级执行方	数据增量清单梳理	对采集、导入的数据增量进行清单梳理。	
5		定级要素识别	根据《数据分类分级规范》进行分类分级要素识别。	
6		数据安全定级	系统上线前完成数据分类分级工作，结果提交数据分类分级管理方	
7	数据分类分级管理方	定级结果审核汇总	对定级结果进行审核，不合理结果协助分类分级方调整。	《数据分类分级清单》汇总
8			合理结果汇总	
9	数据分类分级运营方	固化分级结果	将《数据分类分级清单》录入数据分类分级工具。	
10	数据分类分级管理方	数据安全管理及保护	根据分类分级结果，组织开展数据安全管理及保护。	

3.3 数据分类分级系统建设

数据分级分类系统是一种能够发现、识别、分类分级标记数据的自动化工具，目前在各行业各领域已有广泛的应用。随着数据的快速增长和数字化转型的加速，个人信息保护和数据合规成为企业和组织面临的重要挑战，通过建立科学合理的数据分类分级系统，企业可以更好地管理和利用自身的数据资源，提高运营效率和竞争力，同时满足合规需求。

3.3.1 系统功能概述

数据分类分级系统应当是能够充分利用关键字匹配、正则表达式匹配、特征训练模型、自然语言处理、机器学习、大模型等技术，实现自动或半自动发现数据源、识别数据源、标记数据类别与级别，以提升数据分类分级效率与准确性为目标的技术工具。

(1) 数据源发现

数据源发现功能应能实现主动嗅探发现和被动监测发现两种模式。主动嗅探发现可以实现探测指定网络范围，识别数据载体 IP、端口、版本等数据源信息。应能识别关系型数据库、非关系型数据库、数据仓库、FTP/TFTP/SMB 文件系统等各类数据载体，支持周期性自动触发、事件触发、手工触发或其他启动方式。采用被动监测发现模式时，

应当能够监测和分析数据源所在网络的流量，应在有效识别各类数据源基本信息的基础上，实现数据接口的发现，并正确识别数据接口地址、类型等数据源信息。

(2) 数据分类分级规则管理

数据分类分级规则管理功能可以建立和维护数据分类分级规则库，为自动化识别数据、标记数据提供知识积累。可以根据已有的分类分级方法制定好分类和分级特征描述和映射关系，数据分类分级的识别标记过程可以采用模板中的分类规则、分类规则等作为识别和标记的标准。

(3) 数据识别

数据的识别是数据分类分级系统的核心能力，数据分类分级系统可以识别各类数据载体中的数据，在对结构化数据识别时，应根据数据特点采用相适应的自动化算法，可包括关键字匹配算法、正则表达式匹配算法、特征训练模型算法、自然语言处理、机器学习、大模型等技术。特征库可以定期进行维护和更新。在对非结构化数据识别时，应根据数据的特点支持文本、图片、办公文档、XML、HTML、各类报表、音频、视频、压缩文件等各类格式的数据。数据识别的执行方式包括：定时任务的方式、增量执行方式等。

(4) 数据标记

数据的标记是数据分类分级实施工作的主要目的，是形成数据分类分级清单的关键操作。数据标记是在数据发现、识别、标记规则设计的基础之上开展的，对结构化的数据标记包括但不限于：库名、表名、字段名、级别、所属大类、子类等信息，对非结构化数据对象的标记应包含但不限于：文件服务器地址、文件目录、文件类型、文件名、级别、所属大类、子类等信息。最后，通过人工对数据标记结果进行修订和确认。

(5) 管理与维护

数据分类分级系统可以从各级别分布、各类别分布等维度充分展示数据在数据载体中的分布情况，可以通过过滤数据源名称、数据地址、数据库名、表名、文件服务器、文件、分级、所属分类、子类等各类条件通过可视化表格、统计图等形式查询和展示。可以通过对数据安全管理平台或安全组件提供 API 接口实现数据分类分级结果在数据安全风险监测、数据安全态势感知、数据安全策略联动等方面的应用。

3.3.2 系统建设过程

3.3.2.1 需求分析与确认

根据企业的目标和业务流程的特点，分析企业在数据分类分级方面的具体需求，如需要对哪些数据分类分级、分类分级的标准是什么、分类分级后的数据如何予以安全防护、是否涉及安全工具的联动对接等等。

数据需求：明确需要对哪些存储系统内的哪些数据分类分级，包括数据的来源、格式、类型等，或者优先对哪些应用系统的数据进行梳理，数据量大概在什么级别。

数据处理需求：确定如何处理和分析数据，包括数据的清洗、整合、转换等操作。例如，是否需要对数据进行去重、填充缺失值、转换数据格式等操作，以满足后续的分类分级需求。

系统功能需求：明确数据分类分级系统的功能和性能要求，包括支持的数据源种类、数据智能识别准确度、可视化展示的多样性、系统对接能力等。考虑系统的可扩展性，例如数据量巨大时，如何保证分类分级的处理效率。

3.3.2.2 制定目标与计划

根据需求分析结果，制定详细的数据分类分级系统建设计划，包括项目范围、项目时间规划、资源分配等。组建评审小组，可以包含数据分类分级决策方、数据分类分级管理方、数据分类分级监督方、数据分类分级运营方、数据分类分级执行方、数据使用方。

3.3.2.3 工具选型与测试

- 功能评估：根据需求分析结果，评估不同数据分类分级产品的功能，如数据资产发现能力、敏感数据发现能力、数据类别级别标记能力、可视化展现等。例如系统是否支持主动探测发现数据源，是否支持非结构化文本、图片、视频等数据的处理，数据标记的准确度等。
- 性能评估：评估分类分级系统的性能，如数据扫描的速度、系统稳定性、是否支持高可用架构等，确保能够满足企业的实际需求。
- 服务评估：综合对比厂商的行业项目经验、平均交付时间和产品迭代频率等，通过对这些关键指标的综合对比，企业能够更准确地评估厂商的服务质量和能力，最终选择最佳的系统供应商。

- 产品测试：结合组织实际业务场景，围绕数据分类分级系统各项功能，制定场景化测试用例，开展全面的功能和性能测试，包括但不限于：基本功能执行效果、数据源识别准确率、数据识别准确率、数据识别任务吞吐率等内容。

3.3.2.4 部署运行及优化

- 部署运行：部署数据分类分级系统，对接各业务系统，充分结合工具技术能力，常态化开展数据分类分级全流程工作。
- 审核发布：人工审核和完善系统自动标记的结果，审批通过后批准发布，形成数据资产分类分级清单。此过程需要评审小组审核分类分级结果，审核通过后即可作为最终的分类分级结果清单。
- 持续优化：数据分类分级是一个持续的、动态的过程，随着业务需求的变化和数据量的不断增加，需要定期评估和完善数据分类分级结果，同时定期优化系统的分类分级策略，确保其持续有效。

3.4 数据分类分级监督

数据分类分级工作的监督是组织内部控制体系中的核心组成部分，它承担着确保数据管理符合既定流程和政策的重要职责。通过持续开展监督指导，组织可确保数据分类分级策略的严密性和准确性，同时也能够在发现不符合预定策略或法规要求的做法时，迅速采取纠正措施。组织应当成立专向监督小组，围绕数据分类分级的全过程开展细致的监督和评估，确保从数据的收集、处理到最终的销毁，每一步都遵循规定的标准，且有明确的责任归属和记录留痕。监督团队不仅关注流程的规范性和记录的完整性，还着眼于评估流程的实际执行情况，确保数据管理的准确性和有效性，以及组织对数据资产的全面掌控。

随着组织环境和业务需求的变化，数据分类分级监督的内容也在不断拓展。未来的监督将更加注重评估数据分类分级的合理性，以及分类分级结果的准确性和有效性。这意味着监督不仅仅局限于合规性检查，更包含了对分类分级标准是否与业务需求相匹配、是否能够适应数据安全要求的评估。此外，监督还将评估分类分级结果在实际数据保护和使用中的应用情况，包括分类分级是否有助于提高数据处理的效率、是否能够增强数据安全治理的效果。监督团队需要持续更新监督方案，以应对技术进步和市场变化，确

保监督活动能够持续为组织提供价值。

监督活动的重要性不容小觑，它不仅保障了数据分类分级管理的合规性，更是组织评估和改进数据管理实践的关键工具。通过系统的监督流程，组织可以及时发现流程的缺陷和风险点，进而采取措施加以改进。这种持续的自我完善机制不仅提高了数据的安全性和管理质量，也为组织在快速变化的商业环境中保持灵活性和竞争力提供了支持。

四、数据分类分级方法

数据分类分级方法可参考 2024 年 3 月 15 日发布的国家标准《数据安全技术 数据分类分级规则（GB/T 43697-2024）》，结合组织所属行业要求、地区要求和指导规范，根据组织实际情况开展工作。

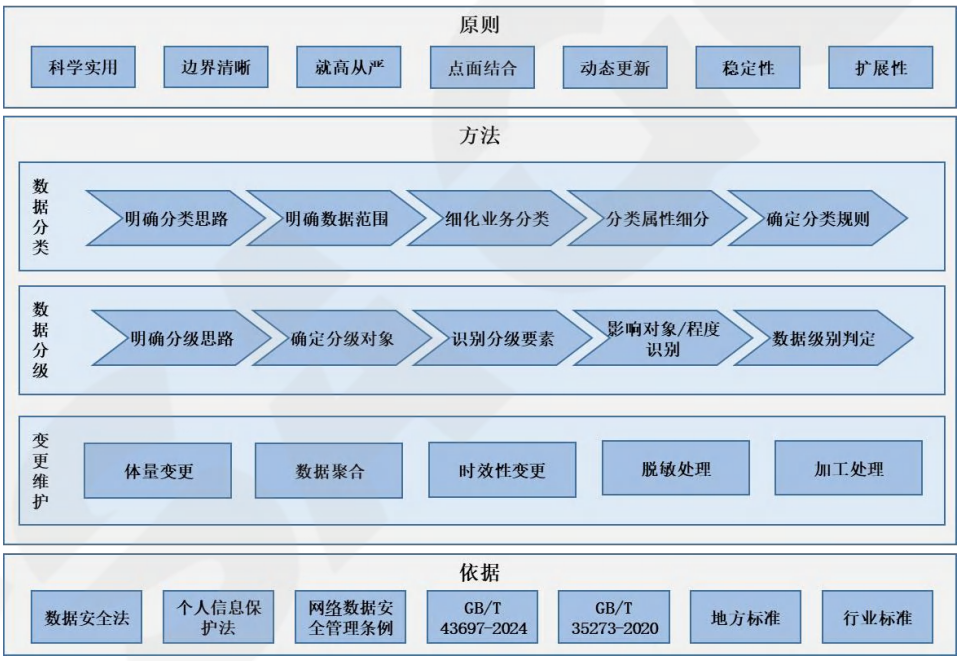


图 7 数据分类分级方法框架

4.1 数据分类分级原则

数据分类分级工作是一项涉及数据全生命周期、业务各个方面以及全员参与的持续性工作。这项工作的核心在于其动态性，要求组织在数据的价值、使用模式和业务需求发生变化时，能够灵活地调整分类分级策略。管理原则强调了数据分类分级不是一个静态的框架，而是一个能够适应组织演变的动态系统。这意味着管理策略须随着组织的发

展和外部环境的变化而发展，确保数据分类分级始终保持相关性和有效性。

组织数据分类分级可参照《数据安全技术 数据分类分级规则（GB/T 43697-2024）》中的科学实用、边界清晰，就高从严、点面结合、动态更新原则，同时，根据自身业务实际可参考扩展下述原则。

数据分类应充分考量组织数据特性，基于组织业务特点，构建层层划分、层层隶属、从总到分的分类体系，每一次划分应有单一、明确的依据。数据类目排列应依据数据类目主题之间的内在联系，遵循最大效用原则，将全部类目系统地组织起来，形成具有隶属和并列关系的分类体系，能揭示不同数据在业务归类之间的联系和区别。同时，还需要考虑稳定性、扩展性的原则：

(1) 稳定性原则：数据分类应选择分类对象最稳定的本质特性作为数据分类的基础和依据。

(2) 扩展性原则：在数据类目的设置或层级的划分上，应当保留适当余地，利于分类数据增加时可扩展。

数据分级应以数据的重要性、敏感性和遭受破坏后的损害程度为依据，遵循分级层次合理、界限清晰、数据安全防护策略合理的原则。组织在开展数据分类分级工作时应在此基础上坚持合理性、实用性、扩展性的原则：

(1) 合理性原则：数据分级应合理，不能将所有数据全部划分到某一个级别中。

(2) 实用性原则：应根据组织数据的多维特征及其相互间客观存在的逻辑关联进行科学和系统化的数据分级，数据分级结果能够为数据的应用、共享、开放过程中的数据安全策略制定提供有效决策和依据。

(3) 扩展性原则：数据分级方法应具有概括性和包容性，能实现各种类型、场景的数据安全分级，满足组织各类信息系统、数字化基础平台和业务平台中可能出现的数据类型和安全需求。

4.2 数据分类分级依据

组织在针对数据资产开展数据分类分级工作时，首先应根据《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》《网络数据安全管理条例》《数据安全技术 数据分类分级规则（GB/T 43697-2024）》《信息安全技术 个人信息安全规范（GB/T 35273-2020）》等数据分类分级政策和法律法规，再根据组织所在行业监管单位发布

的关于本行业本领域的数据分类分级标准、规范等，最后结合组织自身数据资产特点和数据安全管理的需要，按照数据的所属类型划分，对数据应用过程中涉及的数据分类，并根据该数据类型一旦遭到泄露、丢失、篡改、非法获取或非法使用后对客体的危害程度、范围等，对该数据类型根据数据安全级别划分。

4.3 数据分类方法

4.3.1 明确总体思路

组织开展数据分类工作时，应实行先按照行业领域分类、再按照业务属性分类的总体数据分类思路开展相关工作。数据分类方式包含线分类法、面分类法、混合分类法。本文推荐采用混合分类法，以线分类法为主，并结合不同数据维度的面分类法。

详细的分类分级要素说明请参考附录 C-数据分类分级参考资料。不同行业的典型标准和实践案例，请参阅附录 E-典型行业数据分类分级标准解读。

数据分类应当根据组织所属行业数据使用及管理需求，结合业务属性细化数据分类，形成从总到分的树形分类逻辑体系结构。数据分类维度多样，可根据数据的结构化状态、数据加工状态、数据的业务来源、数据是否包含个人信息、使用场景等情况来选取数据分类维度。部分维度示例如下：

(1) 数据结构化维度：结构化数据、非结构化数据、半结构化数据

(2) 数据加工维度：原始数据、加工数据

(3) 个人信息维度：普通个人信息数据、个人敏感信息数据、非个人信息数据

(4) 业务维度：生产业务、管理运营业务

(5) 使用场景：数据分析、数据交易

本指南重点以业务维度展开数据分类描述，在实际分类工作中，组织应尽量选取业务分类维度应根据管理及实际使用需求选取维度分类，或同时结合多个分类维度混合分类。

4.3.2 明确数据范围

此环节应当梳理组织的所有数据资源，可根据实际情况明确组织所管理的数据范围，并开展数据分类分级工作。

4.3.3 细化业务分类

根据组织所属行业数据使用及管理需求，使用线分类方法，一般可将数据大类分为生产业务类、管理运营类的一层类别。生产业务类下一般包含工程建设、产业生产、设备运维、营销、监督等；管理运营类下一般包含：战略规划管理、人力资源管理、供应链管理、财务管理等。

各组织可结合实际业务情况将二级类别下的数据细化形成三级、四级类别，最终形成从总到分的树形逻辑体系结构。

4.3.4 分类属性细分

基于业务属性进行一、二级类别的划分后，根据实际情况，宜依据业务的不同属性来对三级及更低层级的类别进行细化分类。细化分类可进一步参考责任部门、描述对象、数据主体、数据用途、数据处理、数据来源等属性。

如果组织的数据资产涉及法律法规中有专门管理要求的数据类别（如个人信息）时，应按照《信息安全技术 个人信息安全规范（GB/T 35273-2020）》等有关规定或标准，对业务涉及的个人信息进行识别和标记，以便在个人信息保护、数据跨境等专项场景下能够快速、准确地识别数据类型、履行合规义务。

4.3.5 确定分类规则

梳理各二级分类下的业务情况，根据不同行业的数据管理和使用情况，进一步明确组织的数据分类细则。

(1) 应采取“业务条线—关键业务—业务属性分类”的方式给出数据分类规则；

(2) 应对关键业务的数据分类结果进行归类分析，将具有相似主题的数据子类进行归类。

4.4 数据分级方法

4.4.1 明确总体思路

按照重要性和安全风险程度，组织可将数据分为一般数据、重要数据和核心数据三个基础等级。根据行业业务及数据特征，数据级别可进行细分，根据数据重要程度由低到高一般可以分为 3～5 个级别。其中，一般数据中根据数据的重要程度由低到高分为 1～3 个级别，重要数据和核心数据分别设置为一个级别。开展数据分级可实现组织对

数据的分级管控、分域存储，保证组织结构化数据资产在使用过程中达到效率和安全管控的平衡。

4.4.2 确定分级对象

开展数据分级时应先确定数据分级的对象，以结构化数据为例，常见的数据分级的对象为数据的库、表、字段或部分表、字段的集合。可根据数据体量大小确定数据分级对象的精度，如数据体量过大，无法在单次数据分类分级工作中细分到字段级别，可先以数据表或字段集合作为分级对象，在后续分类分级工作中逐步细化到字段。

4.4.3 识别分级要素

通过识别数据的各类分级要素，来确定数据的影响对象和数据的影响程度。数据分级要素广泛，包含数据的领域、群体、区域、精度、规模、深度、覆盖度、重要性、敏感性等，数据分级要素识别应符合 GB/T 43697 的规定，数据分级要素内容见附录 C。

4.4.4 影响对象识别

此环节需要分析发生数据安全风险事件时所受到损害的对象。划分可以按照群体影响数据、个体影响数据、单位自身影响数据。

数据影响对象识别因素的相关内容见附录 C。

(1) 群体影响数据：影响国家安全、公共利益、社会秩序、经济运行等数据；

(2) 个体影响数据：影响个人及法人的人身权、财产权、隐私权、个人信息权益等数据；

(3) 单位自身影响数据：影响组织自身或其他组织的生产运营、声誉形象、公信力、知识产权等数据。

4.4.5 影响程度识别

数据安全影响程度是指数据安全属性遭到破坏后，直接或间接造成的全部影响或损害的程度，从低到高划分为：无影响、轻微影响、一般影响、严重影响、特别严重影响 5 个层级。

为准确判断数据级别，应参照表 3 **数据安全影响程度定义**判断数据安全影响程度。

表 3 数据安全影响程度定义

序号	程度	定义
1	无影响	对数据资产价值、依赖数据的业务、数据主体（个人、企业、组织及单位自身等）、国家安全、社会秩序及公众利益等完全无任何影响。 例如：依照法律规定进行数据公开发布。
2	轻微影响	对数据资产价值、依赖数据的业务、数据主体（个人、企业、组织及单位自身等）、国家安全、社会秩序及公众利益等仅造成一定干扰，其造成结果能自行恢复或容易补救。例如：业务效率短时间内下降、任务进度可接受程度的推迟、造成 1W 以下敏感客户数据泄露、业务恢复时间超过 2 小时等。
3	一般影响	对数据资产价值、依赖数据的业务、数据主体（个人、企业、组织及单位自身等）、国家安全、社会秩序及公众利益等造成一定损害，其造成结果不可逆，但能采取一些措施降低损失、消除影响。 例如：企业或个人财产损失、单位形象损失、造成 1 万以上 10 万以下敏感客户数据泄露、业务恢复时间超过 6 小时等。
4	严重影响	对数据资产价值、依赖数据的业务、数据主体（个人、企业、组织及单位自身等）、国家安全、社会秩序及公众利益等造成较严重破坏，其造成结果不可逆，虽能采取一些措施挽救，但难度较大、成本较高。 例如：人身伤害、企业破产、单位严重损失、造成 10 万以上 100 万以下敏感客户数据泄露、业务恢复时间超过 12 小时等。
5	特别严重影响	对数据资产价值、依赖数据的业务、数据主体（个人、企业、组织及单位自身等）、国家安全、社会秩序及公众利益等造成特别严重破坏，其造成结果不可逆且破坏性巨大，其影响是全局性、战略性的。例如：危害人民生命安全、造成单位特别严重损失、国家政治经济利益等巨大损失、造成 100 万以上敏感客户数据泄露、业务恢复时间超过 24 小时等。

4.4.6 数据级别判定规则

结合数据影响对象及数据影响程度，数据级别判定规则的相关信息见表 4 数据分级规则，以 5 个级别的划分方法举例：

(1) 针对各对象但无影响程度的数据可视为公开数据，一般可作为 1 级数据处理；

(2) 针对单位自身的数据，造成轻微、一般影响的为 2 级数据，造成严重的为 3 级数据，特别严重影响的为 4 级数据；

(3) 针对公民个人的数据，造成轻微影响的为 2 级数据，造成一般影响的为 3 级数据，造成严重影响及特别严重影响的为 4 级数据；

(4) 针对公共利益的数据，造成轻微影响的为 3 级数据，造成一般影响的为 4 级数据，造成严重、特别严重影响的为 5 级；

(5) 针对国家安全的数据，造成轻微影响的为 4 级数据，造成一般、严重、特别严重影响的为 5 级数据。

表 4 数据分级规则

影响对象		影响程度				
		无影响	轻微影响	一般影响	严重影响	特别严重影响
群体影响	国家安全	1 级	4 级	5 级	5 级	5 级
	公共利益	1 级	3 级	4 级	5 级	5 级
个体影响	公民个人	1 级	2 级	3 级	4 级	4 级
自身影响	单位自身	1 级	2 级	2 级	3 级	4 级

4.5 数据分类分级变更

4.5.1 数据分类分级变更情形

数据在类别与级别确定后不是确定不变的，当出现数据体量变更、数据聚合、数据时效性变更、数据脱敏、数据加工等情形时，数据的业务属性、重要程度、可能造成的危害程度也会发生变化，此时应根据实际情况更新数据的类别和级别。

4.5.2 数据分类分级变更场景

(1) 数据体量明显增加或减少时（如超过 20%），考虑到发生数据风险事件时，影响对象及程度可能会提高，应根据实际情况，提高或降低数据级别；

(2) 将来自不同途径或不同系统的数据汇聚在一起，数据的原始用途或所在系统发生改变，需对数据进行重新确定类别并分级，汇聚后的数据级别应不低于原始级别。

(3) 当数据超过使用时效后，应根据实际情况，降低数据级别。

(4) 当数据经过删除、去标识化、匿名化处理等脱敏手段后，应根据实际情况，降低数据级别。

(5) 当发生数据安全事件，导致数据受到泄露、损害等情况时，应根据实际情况，提高数据级别。

五、数据分类分级实施方案

5.1 数据分类分级实施过程

数据分类分级实施阶段主要包含：业务活动识别、数据资产发现、数据资产识别、分类分级策略制定和数据标识标记。各阶段的流程如下图所示：

图 8 数据分类分级实施流程

5.2 业务活动识别

本阶段需识别和梳理全部业务活动（或增量业务活动），采取的主要方式为人工访谈、资料查阅和系统查验。

- 人工访谈：与相关业务负责人进行访谈，了解当前的业务领域和细分业务条线，整理形成业务清单。
- 资料查阅：查阅系统相关的开发设计文档、操作手册等说明性资料，了解系统承载的业务情况。
- 系统查验：对相关业务系统的功能模块、承载的业务情况进行查验，梳理各业务活动，整理形成业务清单。

业务活动识别的主要内容包含：业务领域、业务大类、业务子类、业务条线、业务操作描述、关联结构化数据集描述、关联非结构化数据集描述。

5.3 数据资产发现

数据资产发现阶段，首先发现和识别数据载体，进而发现不同数据载体中包含的数据内容。详细的系统功能和典型案例，请参考附录 A-典型数据分类分级系统介绍。相关关键技术的详细描述见附录 D-数据分类分级关键技术与方法。

5.3.1 数据载体发现

数据载体发现一般从两方面开展，一是针对已知的数据载体进行盘点和梳理，二是对未知的数据载体进行探测和发现。

(1) 已知数据载体

结构化数据通常的载体为包含 Oracle、MySQL、MSSQL、DB2、达梦、人大金仓、南大通用、ES、Hive、PostgreSQL 等在内的数据库系统，需完整地进行盘点和梳理，包含网络地址、端口、用户名、口令等信息。

非结构化数据通常的载体为文件服务器、移动存储介质、终端等。

(2) 未知数据载体

针对未知的数据载体，需要通过数据资产探测技术探测和发现。通过主动向目标网络发送探测数据包，对存有数据资产的存活主机、服务器执行探测扫描。主动探测前，需要确认扫描的网段、所使用的数据库等，然后再有针对性地扫描。主动探测能够扫描设定范围内，网络可达的所有数据资产。通过主动探测扫描技术，能够自动发现数据库的基本信息，包括：端口号、数据库类型、数据库实例名、数据库服务器 IP 地址等。

针对结构化数据，主流的可探测的数据包协议包括：TCP/IP 协议族，HTTP 等应用层协议。同时，可对当前主流的数据库及其默认端口扫描，如 MySQL（3306）、Oracle（1521）、SQL Server（1433）、DB2 数据库（5000）、PostgreSQL（5432）、DM 达梦（5236）等。针对非结构化数据，主流的可探测的数据包协议包括：TCP/IP 协议簇，HTTP、HTTPS、FTP、SFTP、TFTP 等应用层协议。

5.3.2 数据内容发现

完成数据载体发现后，分别针对结构化数据载体和非结构化数据载体中的数据执行内容发现。

(1) 结构化数据载体

针对结构化数据载体中的数据，需要识别出数据库名称、数据表名称、字段名称，这些名称通常是英文的。

(2) 非结构化数据载体

针对非结构化数据载体中的数据，需要识别出数据文件存储路径、数据文件名称、数据文件格式、数据文件量级、数据文件属性。数据文件存储路径一般指绝对路径；数据文件量级是指文件的字节大小。

5.4 数据资产识别

5.4.1 结构化数据

在结构化数据内容发现的基础上，进一步明确各数据库、数据表、数据字段的英文名称所对应的中文名称，字段说明（字段说明是指字段中存储的真实值的取值示例，如“name”字段对应的字段说明为“张三或李四”）、字段长度以及数据字段对应的数据所有方、数据使用方和数据管理方，形成标准的结构化数据资产清单，包含：应用系统名称、数据库名称、数据库中文名、数据表名称、数据表中文名、字段名称、字段中文名、字段说明、字段长度、数据所有方、数据使用方、数据管理方。标准模板的具体内容可参考附录 B-数据分类分级参考模板。

5.4.2 非结构化数据

在非结构化数据内容发现的基础上，进一步明确各数据文件的内容描述、所有方、使用方和管理方，形成标准的非结构化数据资产清单，包含：应用系统名称、数据文件存储路径、数据文件名称、数据文件格式、数据文件量级、数据文件内容描述、数据文件所有方、数据文件使用方、数据文件管理方。如附录 B 所示。

数据文件内容描述是指文件内容的摘要，可通过技术工具，结合工具的识别算法和规则、特征库，去识别已知格式文件的内容；未知格式文件的内容需要通过人工调研、核查识别。当前常用的识别技术有 NLP、OCR、视频文件处理、音频文件处理等技术，如下图所示：

图 9 常用数据识别技术

NLP，即自然语言处理，主要用于文档类型文件的处理，如 doc、txt、pdf 等格式，需要识别其中的关键内容。NLP 的基本步骤为：断句、分词、标记词性、识别命名实体、去停用词。

OCR，即光学字符识别，主要用于识别图片中的数据，如 bmp、jpg、png、tif、gif 等格式。OCR 的基本步骤为：图像预处理、字符分割、字符识别、字符组合、文字处理。

视频文件处理，主要用于识别视频文件中的数据，如 mp4、avi、flv 等格式。视频文件处理的基本步骤为：提取帧、过滤帧、图像内容识别、文字处理。

音频文件处理，主要用于识别音频文件中的数据，如 mp3、wav、flac 等格式。音频文件处理的基本步骤为：音频转文字、文字处理。

有关非结构化数据识别的规则示例，请参见附录 G。

5.5 分类分级规则制定

本阶段工作在实施过程中，项目组需要考虑企业所在行业的监管单位是否已经发布了数据分类分级相关的标准、规范和指南等文件，并根据企业所在行业是否存在数据分类分级标准的情况，参考第四章节方法，按照以下两种方式完成本次项目数据分类分级规则的设计。

(1) 所在行业有数据分类分级标准

若组织所属行业的监管单位已经发布了数据分类分级标准、规范、指南等文件，则应在行业已经发布的数据分类分级标准的基础上，结合梳理得到的《数据资产清单》等，并根据组织的数据安全管理和业务发展的需要，对行业监管单位发布的数据分类分级标准进行优化，形成组织的《数据分类分级规则》，应确保优化后的数据类别可以覆盖组织业务中涉及的所有数据内容。

(2) 所在行业无数据分类分级标准

若组织所属行业暂未发布数据分类分级相关的标准、规范、指南等，则应根据梳理得到的《数据资产清单》等，以及组织的数据安全管理和业务发展的需要，依据第四章节的方法或者参考其他行业成熟的数据分类分级规则，进行深度的分类设计。数据分类规则设计过程中，应逐层对数据分类分级策略进行细化，直至细化后的每一个数据类别

均对应到业务中最小一类的数据资产，并且应确保细化后每一个子类之间不存在交叉重复情况，以及所有子类汇总后可以覆盖业务中涉及的所有数据资产内容；在进行数据分级规则设计时，应通过综合考虑定性与定量的思想，充分考虑该类型的数据一旦遭到泄露、篡改、破坏或者非法获取、非法利用、非法共享后可能影响的对象、影响范围、影响程度等因素，从而设计分级策略。

5.6 数据标识标记

本阶段需要依据数据分类分级规则，为数据资产进行标识标记。

数据分类方面，应当结合识别的业务活动和确定的数据分类规则确定分类标识，再开展标记工作；数据分级方面，应当先确定数据分级对象（如数据项、数据集、衍生数据、跨行业领域数据等），然后识别分级要素（如数据的领域、群体、区域、精度、规模、深度、覆盖度、重要性等），再根据分级要素识别情况分析数据一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享可能影响的对象和影响程度，综合确定分级标识，最后开展标记工作。

分类分级标记过程中，可通过工具预定义的规则库实现自动标记，再通过人工审核和纠错的方式，核查数据标记的正确性、完整性，最终形成数据分类分级清单，如附录B所示。

数据标记方法主要有元数据标记和原始数据标记两种。

元数据标记法是指通过数据资产梳理建立结构化和非结构化数据的资产清单，然后标记数据字段、表、库、文件等表结构名称，然后通过建立映射表寻找相关数据内容的一种方法。映射表则主要用来存放键值对，如果提供相应的键，就能查到相应的值。

原始数据标记法在实现方式上可分为“字段标记法”和“数据水印法”。字段标记法是指把标记“打”在某一个公共访问的字段上，这个字段往往代表数据的一种属性，从而能够作为区分数据资产方式的一个重要依据。数据水印法是指通过数据水印（数据水印是一种将标识信息通过一定的规则与算法隐藏在结构化数据和非结构化数据中的技术）的方式，为数据资产打标。需要注意的是，采取原始数据标记法会涉及数据内容的变更，需在使用前评估。

六、数据分类分级的应用

图 10 数据分类分级成果应用

6.1 满足合规监管要求

自《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》发布以来，针对数据分类分级、重要数据目录报送等要求，各部委、行业监管单位相继出台了一系列政策要求、行业要求，如工信部下发的《工业和信息化领域数据安全管理办法》《工业数据分类分级指南》，国家金融监督管理总局下发的《银行保险机构数据安全管理办法》中都明确要求了所管辖行业的数据分类分级的要求，细化到各个行业，如电信、金融、能源、政务、医疗等各行业都相继发布了行业标准、团体标准。从整体政策发布趋势来看，对数据分类分级的政策要求、行业要求越来越完善，越来越严格。完成数据分类分级工作后，可梳理分类分级结果与识别的重要数据，并按规定模板填写及报送主管单位。同时在数据跨境场景中，企业应首先识别自身主体身份，根据分类分级结果，判定是否含有重要数据或个人信息，并统计重要数据和个人信息的体量，决定是否需要执行数据出境安全评估，满足数据出境的安全合规要求。

6.2 优化数据资产监测

数据作为关键生产要素，为实现充分发展数据要素的各项政策接连推出。对各企业来说，要充分释放数据要素价值，首先就需要梳理自身数据资产情况，完成数据分类分级。通过数据分类分级工作，企业可以更有效地了解和展示企业数据资产的类别与级别情况。在完成数据资产梳理工作后，可以以数据的类别、级别，数据的体量，数据在不同系统的分布情况，数据的使用情况等维度，进行关联展示与分析，根据不同关联的维度在各安全场景下进行安全预警，为后续的数据安全保护做准备。

6.3 开展数据处理活动管控

在数据的收集、存储、使用、加工、传输、提供、公开、销毁等数据处理活动中，可结合数据分类分级的结果执行安全管控，如在数据存储过程，可结合数据的级别执行分域分层存储；在数据使用环节中，可对高级别敏感数据的应用和流转执行安全监测；在数据的共享、交换环节，针对不同级别的数据设置不同的审批流程；在数据的销毁环节，对高级别的数据设立更严格的数据销毁程序等等。

6.4 细化数据安全风险及事件管理

随着数据安全威胁的日益加剧，开展数据安全风险评估，确保安全风险可控成为企业数据安全建设的工作重心。开展数据安全风险评估时，数据的重要性是确定风险危害程度分析的关键要素，数据分类分级可以作为精细化风险评估工作的前置条件，数据级别越高代表数据重要性越高，个人信息规模和数据敏感程度可以作为判断数据重要性的衡量因素。

在开展数据安全事件管理时，应根据数据分类分级的结果，将事件涉及的数据类别、级别、规模等要素与事件预警级别进行有效关联，并根据不同的事件预警级别选择对应的处理策略，在事前、事中、事后同步提升数据安全管理能力。

6.5 实现数据安全保护联动

企业为提高数据安全保护能力，可充分应用数据分类分级结果，联动各项数据安全保护措施，包括不限于：数据加密、数据脱敏、数据防泄露、访问控制、数据水印等。设立具有针对性的安全保护策略，提升数据安全保护能力，降低数据安全保护成本，针对一般数据、重要数据、核心数据的安全保护策略示例如下：

表 5 数据分级保护策略示例

级别	保护方法（示例）						
	加密	脱敏	防泄露	标识标签	备份容灾	鉴别认证	记录审计
一般数据	按需	按需	按需	需要，标识数据的类别、级别、责任人	需要，定期备份	按需	需要，接入及行为记录，定期审计
重要数据	需要，对存储数据进行分级加密	需要，且具备泛化、抑制、干扰等脱敏技术或算法	需要，且内容级检测及阻断	需要，标识数据的类别、级别、责任人、使用场景	需要，定时备份，并异地容灾	需要，进行双因素鉴别	需要，接入及行为实时监控及时审计，数据操作行为异常监测督促整改
核心数据	需要，传输加密，采用有 128 位或以上强度的国产密码技术	需要，且具备泛化、抑制、干扰等脱敏技术或算法	需要，且内容级检测及阻断	需要，标识数据的类别、级别、责任人、使用场景	需要，实时备份，并异地容灾	需要，进行双因素鉴别，并授权访问控制，限制访问范围、频率	需要，接入及行为实时监控及时审计，数据操作行为异常监测督促整改，数据安全风险监测

附录 A-典型数据分类分级系统介绍

A1 天融信数据安全分类分级系统

- 产品功能介绍

图 11 天融信数据安全分类分级系统

天融信数据安全分类分级系统（TopDCS），致力于为企业级用户提供一体化数据安全感知中心、决策中心和指挥中心。围绕着国家监管、行业监管等保合规、安全运营等多种场景需求而设计研发。系统以数据安全治理为基础，数据安全防护为手段，数据安全生命周期为管理核心，数据安全服务为支撑，基于以数据为中心的安全架构 Digital Container Shipping Association（简称“DCSA”）技术架构，利用大数据、机器学习技术，对全网数据安全信息进行采集与分析，实现数据资产全面识别与统一管理、快速定级与备案管理、组件备案与精准防护、风险发现与监测告警，实现全面合规、协同智能、统一集中的数据安全管控。

天融信数据安全分类分级系统采用 DCSA 技术架构，目标是解决跨多个产品应用分类、保护、访问控制、权限、监控、警报和审计。数据分类分级系统能够灵活应对逻辑隔离、物理隔离、跨地域等多种网络隔离场景下的数据分类分级需求。通常情况下客户使用系统进行分类分级工作，但在网络存在隔离的情况下，比如逻辑隔离、物理隔离和网络不通，数据量超大情况下，仅靠系统难以满足需求，因此需要数据安全分类分级

系统对网络隔离/不联通的情况下完成数据分类分级工作，并将结果同步至系统。

同时采用多种方式自动识别多样性数据资产，实现数据分类分级以及敏感数据目录标定；实现呈现数据安全总体态势，包括数据资产态势、告警风险态势等，从用户侧各业务的数据全生命周期提供安全防护服务。按照一体化、标准化、智能化、可视化要求，快速掌握全面数据安全态势、掌握数据面临的安全风险、隐患、及时发现、实现对数据安全告警进行及时处置，围绕数据安全生命周期建立数据安全防护机制和技术支撑体系，为各业务数据提供安全防护能力，减少数据安全事件造成的损失、提升数据安全防护能力。实现数据生命周期的可信、可管、可控、可追溯的目标。

(1) 态势监测

分析数据安全管理整体情况，持续监测分析数据资产的态势信息并进行展示。从资产分布、重要数据分布等多个角度可视化呈现组织内数据安全情况，展示包括数据资产多级别分布、数据资产类别分布以及重要数据的级别分布，帮助运营人员全面掌握数据资产态势。

(2) 数据资产管理

管理数据库资产，为组织提供高效、可靠的数据库资产维护和检索机制，便于维护数据库资产信息。对纳管的数据库资产通过列表清单形式展示和维护管理，清晰了解网络中的数据库资产及其变化的过程，方便对数据库资产进行记录维护。

(3) 数据资产管理

管理数据资产，集中、高效、安全地管理数据资产。对纳管的文件服务器资产通过列表清单的形式维护管理，包括新增、查看、编辑、删除、导入、导出，详细了解文件服务器资产及其变化过程。

(4) 数据资产探测

全面、高效、自动化识别网络中存在的数据资产和资产的元数据结构。创建针对目标 IP 范围的数据资产探测任务，自定义任务名称、扫描类型、IP 范围、任务时间，采用混合扫描技术探测识别数据库、文件服务器、设备资产等类型的资产，管理控制探测任务，包括启动、停止、查询、删除、导入、导出等，查看扫描结果信息，包括 IP、端口、推荐资产类型、主库添加状态，并可对其进行主库添加。

(1) 备库管理

精准、全面地管理数据资产探测任务中待确认的备库资产。对探测结果中的元数据

进一步划分，实现探测数据的主备库管理和备库回收站管理，支持但不限于删除、导出、主库添加、误删恢复等操作，支持以列表清单的形式查看备库资产信息，展示 IP、端口、资产类型、来源、主备库状态、时间等信息，形成多信息要素的备库资产展示。

(2) 数据资产扫描

全面、自动化、自定义扫描网络中的数据资产并展示数据资产的详细信息，帮助运营人员更加快速、直观了解数据资产状况。支持自定义扫描模式、数据特征、执行方式等信息创建资产扫描任务进行数据资产扫描，并通过列表清单的形式展示数据资产扫描任务，展示信息包括资产名称、IP、端口、资产类型、来源方式、执行方式等，可对任务进行编辑、启停、离线特征分析、查看结果等操作。

(3) 元数据管理

提供数据库、文件服务器、应用接口的综合性、集中化的元数据管理功能。支持多种数据类型混合扫描结果综合显示，提供集中化、综合性的元数据管理手段，包括数据扫描、元数据扫描、离线数据导入得到的数据库元数据、文件服务器元数据应用接口元数据，通过列表清单集中呈现，展示项包括数据库名称、资产 IP、字段名称、数据特征、业务系统、文件路径等信息。

(4) 分类分级查看

精细化、高效化实现分类分级查看功能。对数据资产进行分类分级管理，对数据资产的编目、标准化、管理有指导作用，分类分级查看支持业务系统、数据库、文件服务器多个维度对分类分级结果进行查看，显示的信息包括资产 IP、业务系统、文件总数、已分级数量、敏感数量、重要数量、个人信息数量等，便于运营人员查询、识别、管理、保护和使用数据。

(5) 分类分级检索

提供分类分级检索功能，有效避免在海量信息中迷失方向。支持数据库检索、文件系统检索、应用接口检索，支持 IP、业务系统、文件服务器多维度进行分类分级检索，并以列表清单的形式呈现检索结果，呈现的信息包括且不限于资产 IP、业务系统、字段名称、数据特征、分类分级信息、数据标签等信息，便于运营人员捕捉数据资产的关键信息。

(6) 分类分级任务

通过创建分类分级任务对目标资产实现准确、可靠的分类分级。支持对已扫描的数

据资产进行分类分级，也可以先创建数据扫描任务、再对其进行分类分级任务，并通过列表清单的形式展示已经创建的数据分类分级任务，呈现的信息包括任务名称、策略名称、目标数据资产、任务状态等，可对分类分级任务进行查看、增加、编辑、删除、执行等操作，实现分类分级任务的全流程管理，帮助运营人员提升分类分级任务管理效率。

(7) 分类分级策略

通过创建分类分级策略帮助实现数据的精细化管理，提高数据的使用效率和安全性。系统支持自定义分类分级策略，能够根据行/企业标准定义分类分级策略，包括自定义分级信息和数据标签，支持模板导入功能，能够快速实现分类分级策略制定，并支持对已有的分类分级策略进行查看、编辑、删除、复制等操作。

(8) 数据识别规则管理

提供数据识别规则管理功能，自动化匹配数据特征。识别规则域包含多种类型识别规则，对数据库、文件服务器、应用接口等多种识别对象使用关键词、正则表达式、词典等多种识别方式自动化匹配，可根据数据识别的分级分类结果以列表清单形式进行呈现，支持自定义维护数据标签，可手动对数据特征进行编辑修改，实现分级分类打标自由化，同时提供批量导入和导出功能，为运营人员提供高效的数据特征管理手段。

- **产品特点介绍**

(1) 数据资产自动发现

可根据预定任务自动进行数据资产扫描、识别整合各类数据资源，并基于行业分类分级标准形成敏感数据特征库，从而减轻人工识别工作量，降低人工操作的复杂度，实现敏感数据识别自动化，为后续的数据分析、决策支持等提供坚实的数据基础。

(2) 数据资产识别知识丰富

可识别的特征库、数据库类别多，识别知识丰富。系统内部集成了大量的数据资产识别算法和模型，能够全面、准确地识别各种类型的数据资源，包括结构化数据、非结构化数据、关系型数据、非关系型数据等。

(3) 数据特征识别智能化

采用自然语言处理（NLP）、机器学习、正则语法相结合的人工智能特征匹配技术，实现数据资产特征精准识别，智能化识别数据特征，采用多层次综合识别技术，从数据资产位置、名称特征、数据特征、关键词规律等多个维度，综合化识别数据特征。

- 产品展示

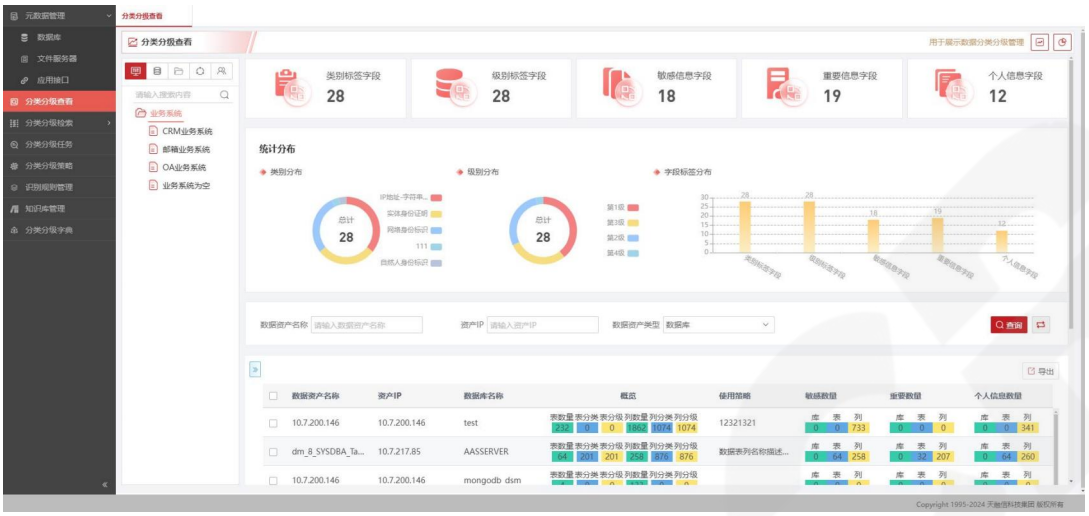


图 12 数据安全分类分级系统界面展示

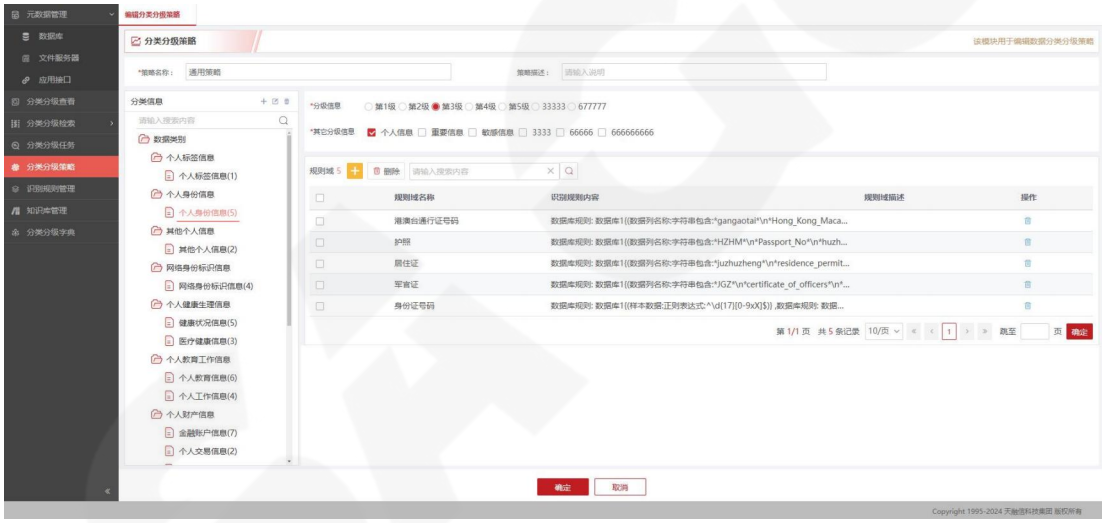


图 13 数据安全分类分级系统界面展示

- 产品应用情况

天融信数据安全分类分级系统（TopDCS）适用于拥有大量数据资产，注重数据资产价值的行业客户，帮助用户解决在数据安全管理中所面临的数据安全合规、业务数据分类分级梳理、数据资产认责及管控、数据结构及流向可视化监控、数据安全威胁检测、数据安全策略集中管控、数据安全态势可视化、数据安全运营支撑等一系列问题。通过数据分类分级建设，结合天融信网络 DLP、终端 DLP、数据脱敏、数据库网关、数据库审计等数据安全组件设备，实现数据全生命周期的可信、可管、可控。目前，本产品已成功应用于运营商、海关、能源、政府、交通、医疗、教育等 10 多个行业。

A2 观安观智敏感数据发现软件

- 产品简介功能介绍

图 14 观安观智敏数据发现软件功能介绍

观安观智敏感数据发现软件基于隐私保护与合规的数据安全治理技术框架，根据各行业的业务数据特征和分类分级规范，提供行业模板，通过自主创新研发的敏感数据识别技术全面、快速、准确发现和定位敏感数据，构建持续更新的企业敏感数据分类分级目录。内置 GDPR、网络安全法、PCI 等合规知识库，结合敏感数据目录识别和量化数据安全风险，生成统计报告，驱动数据安全策略的落地，为数据安全工作的推进提供抓手。

- **产品特点介绍**

(1) 数据分级分类

敏感数据发现系统根据行业相关规范要求及行业通用数据，既提供内置的数据分类分级行业模板，也可以基于企业实际情况新增自定义模板，多个模板间和独立并存。贴合企业自身业务数据管理需要，满足企业数据安全分类分级建设要求。

支持用户通过资源或表粒度自定义分类分级任务、导出分类分级结果。根据数据的来源、内容和用途对数据进行分类；按照数据的价值、内容敏感程度、影响和分发范围不同结合智能向量引擎、智能反馈引擎、多维指标引擎对数据进行敏感级别划分，不同敏感级别的数据有着不同的管控原则和数据开放要求。在完成任务后可通过多个视角对结果进行核查，形成多个用于安全业务的清单，如分类分级清单、重要数据清单等，分类分级清单和重要数据清单以字段粒度展示全部数据资产/重要数据资产的分类分级标注结果以及数据资产的详细信息，并且生成对应分类分级报告。同时提供全量导出功能，供用户数据上报或迎检。

(2) 敏感数据发现

明确了数据发现的对象和策略后，要对敏感数据进行识别和定位。基于规则匹配、语义算法、特征提取等技术手段，根据过滤条件对抽取的数据子集进行自动识别。内置丰富敏感数据发现算法，支持对某列数据按照自定义的数据内容字典进行匹配，对于发现无法归纳出数据特征，可使用枚举方式发现的敏感信息字段和其他用户自定义算法。敏感数据梳理任务支持指定数据表或文件，支持定时、周期性进行，支持增量个性化发现，以适应业务上定期更新数据库等业务行为，提供全方位、高效、贴心化的敏感数据梳理。

根据对上位法解读结合系统自有敏感数据识别技术，系统支持通过对不同数据模型关联，提供合规检查项管理功能。而根据合规检查项进行合规检查，生成合规风险检查报告，帮助企业对自身数据安全问题点和现状提供依据。

(3) 数据资源管理

支持通过自动资源扫描、手动导入已整理好的数据资产表格或手动添加数据源及其连接信息的方式汇总数据资源，支持基于 IP 段创建任务进行自动扫描，全方位发现数据资源，提高资源梳理效率与发现能力，减少人工整理成本。可视化展示各数据库类型、文件类型等数据资源的分布情况，可导出数据资源清单和图表。支持查看数据源的用户

信息、敏感数据统计信息和主外键关联信息，页面上展示了数据源下各表或文件的元数据，包括字段的数据类型、是否主键、注释信息等。

- **产品展示**

图 15 观安观智敏数据发现软件产品展示

- **产品应用情况**

产品广泛应用于运营商、金融、政府、能源、制造等行业。

A3 山石网科数据安全综合治理平台

- 产品功能介绍

图 16 山石网科数据安全综合治理平台功能介绍

产品采用 B/S 结构和大数据底层技术框架，搭载数据资产自动发现、数据架构智能扫描，敏感资产自动识别等先进技术引擎。能够帮助企业快速定位其内部网络中的数据服务，以及各类数据资产的分布等情况，协助用户清晰地掌握敏感数据分布、流转和使用情况，对数据资产进行不同类别和密级的划分，以便实现对敏感数据进行针对性防护。

(1) 数据源类型

支持数据库、文件系统、大数据组件多种数据源类型。

数据库类型支持国际主流数据库，如 Oracle、SQL Server、MySQL、DB2、Informix、Sybase、PostgreSQL、Cache 等，同时支持国产数据库，如达梦、人大金仓、南大通用、高斯 DB 等，文件系统支持 FTP、NFS、Samba 等类型，大数据组件支持 Hive、MongoDB、GreenPlum、Hbase、ES 等。

(2) 数据源管理

通过数据服务扫描引擎，能够自动扫描、判断、识别网络内常见的数据库和文件系统服务。数据源自动发现模块无需数据源服务端的授权，即可发现常见的大多数数据库

和文件系统类型。

支持快速扫描和精准扫描两种模式。扫描结果以可视化的方式展示出企业内部已存在的数据库或文件系统相关信息。数据资源管理模块可通过自动化方式扫描用户的数据库表，建立全局的数据资源目录，并内置了敏感数据识别规则，可以有效识别敏感数据在系统内的分布情况。同时数据资产管理模块支持对数据库、表、字段的备注定义和分类打标，可根据数据价值和特征，梳理出核心数据资产，对其进行分级分类，从而落实对数据更为精细的安全管理措施。

(3) 数据资产发现

平台提供数据资产发现和识别服务，通过对分布在数据源中的各类业务数据提供自动化梳理服务自动化扫描和采集数据所在的物理位置、逻辑位置、存储格式、状态等多维属性信息，平台利用数据资产分析模型，进行分析、归纳、统计获得完整的数据资产分布拓扑、数据资产目录、敏感数据分布地图、敏感数据访问热点分布、数据资产访问权限管理实现数据资产安全防护动态管理。

支持敏感数据自动识别，能从海量数据中精准定位不同类型敏感信息的数量和分布等信息。

内置超过 40 种的敏感特征库，涵盖社会特征、个人证件、通讯信息、财产信息、企业信息、车辆信息等常见敏感信息项，如姓名、地址、电话、身份证、统一社会信用代码、银行卡号、日期、Email 等敏感信息。

用户也可以通过自定义的方式构建企业所属行业的敏感信息特征模板。

根据敏感数据的定义，使用内容识别手段（关键字、正则表达式、内容指纹、字典等），对于数据进行内容识别和检测。

敏感数据的识别，是实现数据的索引、标记及数据发现、审计等工作的前提条件，通过技术手段区分普通数据和敏感数据。

(4) 数据架构扫描

支持数据库或文件系统的架构信息扫描，数据库获取的信息包括模式（库）、表、列（字段）等元数据信息，以及主键、外键、索引、约束、函数、存储过程、触发器、序列、同义词等数据库对象信息，文件系统包括的信息主要有目录、文件、列（字段），通过这些架构信息，能够绘制出完整的数据结构关系图表，帮助用户快速发现并理解现有的数据结构。

(5) 账号权限扫描

能够快速地对数据库的账户进行扫描，获取其角色、对象等相关权限信息。可以直观地展示账户所拥有的权限，定位账户所拥有的高风险权限，便于对数据库账户进行管理。

(6) 数据分类分级

数据分类分级主要用于维护数据类别标签及数据风险等级。支持自定义数据类别标签及数据风险等级，支持分类分级标签的自动标记或手动标记。

数据资产分级是指按照数据资产的敏感等级（密级），对数据进行划分归类的操作，企业或部门可以使用其所属行业的数据分类分级标准，也可以根据所管理的数据资产的特点，定制分级标签。

数据资产分类是指用户可以给数据资产打上分类标签，让原本无法直接理解的数据，变得可阅读、易理解。进一步帮助企业了解资源使用状况，提供业务调整决策依据与数据支撑，合理利用数据资源。

根据规则库或者数据敏感特征项对数据自动分类分级，并自动标识，形成敏感数据资产可视化和数据分布管理清单，将数据分为不同的业务类别和安全级别实施差别保护，为各项数据安全防护策略的制定提供基础，通过与数据安全防护设备进行联动可以实现数据安全防护策略自动部署和应用，避免对所有数据进行一致高强度防护成本增加和效率下降。

通过对识别的数据资产进行分类分级，针对不同级别的数据进行策略设置，以实现敏感数据的识别和跟踪管理。同时，根据差异化分类分级管控需求，可自定义数据分类分级模型。

(7) 内置分类分级行业模板

内置电信行业与金融行业分类分级标准模板，用户也可根据自己所存储的关键数据的敏感程度与价值程度，将数据分为不同的业务类别和安全级别实施差别保护。

分类分级除了可以满足《中华人民共和国数据安全法》的合规需求，在有“觉悟”的企业看来，更是提升自身信息化水平和运营能力的良方。基于业务的分类可以更好地将数据资产化，持续性为企业提供精准的数据服务；同时数据分级可以在安全角度为企业保驾护航，哪些数据可以使用、哪些不可以使用、哪些能对外开放、哪些不能开放、不同等级的数据在不同场景使用哪种安全策略，一目了然。

(8) 数据资产台账

支持以直观的方式展示系统在指定环境中扫描到的数据资产的所有信息，全面掌握数据资产的存储位置、资产目录结构、分级标签、分类标签，敏感特性项等信息。支持快速检索查看或者导出关注的资产信息。

(9) 数据流向管理

支持对数据资产的流向进行直观展示，可以清晰洞察数据在不同资产模块间的流转情况，准确定位数据资产的位置，持续跟踪数据资产动向。

(10) 数据资产地图

通过地图的形式展现可视化、全方位、多维度的企业数据资产分布情况。为企业构建知识图谱、账务数据资产全貌等，让数据生动形象起来，连接数据和应用场景，辅助企业的管理决策。

- **产品特点介绍**

山石网科数据安全综合治理平台具有支持范围广、识别速度快、易用性高、通用性强等特点，帮助企业快速发现和梳理数据资产状况，辅助企业对数据分类分级建设，洞察数据资产流向和用户权限，同时能够满足各种监管检测等场景，替代了传统的数据资产管理和梳理工作模式，极大地提高数据梳理的工作质量，进而降低企业管理成本，为企业的数据安全建设保驾护航。

- **产品展示**

图 17 山石网科数据安全综合治理平台产品展示

- **产品应用情况**

平台目前数据库类型支持国际主流数据库，如 Oracle、SQL Server、MySQL、DB2、Informix、Sybase、PostgreSQL、Cache 等，同时支持国产数据库，如达梦、人大金仓、南大通用、高斯 DB 等，文件系统支持 FTP、NFS、Samba 等类型，大数据组件支持 Hive、MongoDB、GreenPlum、Hbase、ES 等。

通过数据安全治理平台，以“数据为中心”对政府、电信、金融、医疗、交通、高校、互联网的核心应用系统的数据开展安全治理工作。首先，以数据分类分级服务的方式对数据资产进行全面的梳理；其次，依据相关标准在数据安全治理平台中对数据资产进行标签批注，以便于精准数据安全管控。

A4 大道云隐密数万象数据资产管理系统

- 产品简介功能介绍

图 18 大道云隐密数万象数据资产管理系统功能介绍

- 产品特点介绍

(1) 非结构化数据分类

通过人工智能算法，企业能够实现对非结构化数据（如文件）的高效分类，这一过程极大地提升了数据处理的速度和准确性。传统的人工分类方式受限于人力资源的有限性和人为错误的可能性，而人工智能算法则能够自动化地处理大量数据，并通过机器学习和深度学习的技术，不断优化分类的精准度。

(2) 持久化保护

在文件发送后，可精准设定仅特定 IP 地址或指定机器能够开启文件，有效限制文

件访问范围，防止文件被未授权的设备获取。并且，文件还能依据预设时间自动删除，彻底杜绝文件在超出授权时效后仍留存于外部环境而引发的安全隐患，全方位保障文件在传输与使用过程中的安全性与保密性。

(3) 资产管理

软件强大的数据灾备和同步功能，确保了企业数据的完整性和安全性。在遭遇突发错误、人为错误或技术故障等突发情况时，企业能够迅速恢复数据，避免业务中断和损失。同时，数据的实时同步功能，确保了不同部门和员工之间信息的及时共享，提高了工作效率和协作能力。

(4) 模型版权

通过区块链等先进技术，为模型赋予独一无二、不可篡改的数字标识，精准记录其创作信息、版本演变及归属详情。在模型分发与使用过程中，严格监测每一次的访问、调用与传播行为，借助加密算法与授权机制，确保只有合法授权的用户在许可范围内才能使用模型。一旦发现侵权行为，能够迅速溯源追踪，并提供具有法律效力的证据链，有力维护模型创作者的合法权益

(5) 知识遗忘

可依设定条件，如知识领域、来源、时间等精准定位要遗忘的知识。能灵活选择遗忘方式，或渐进或一次性清除。过程中严守数据安全，保障模型架构与其他知识稳定，避免误删，适配不同场景需求，有效管理模型知识数据。

- 产品展示

(1) 自动分类

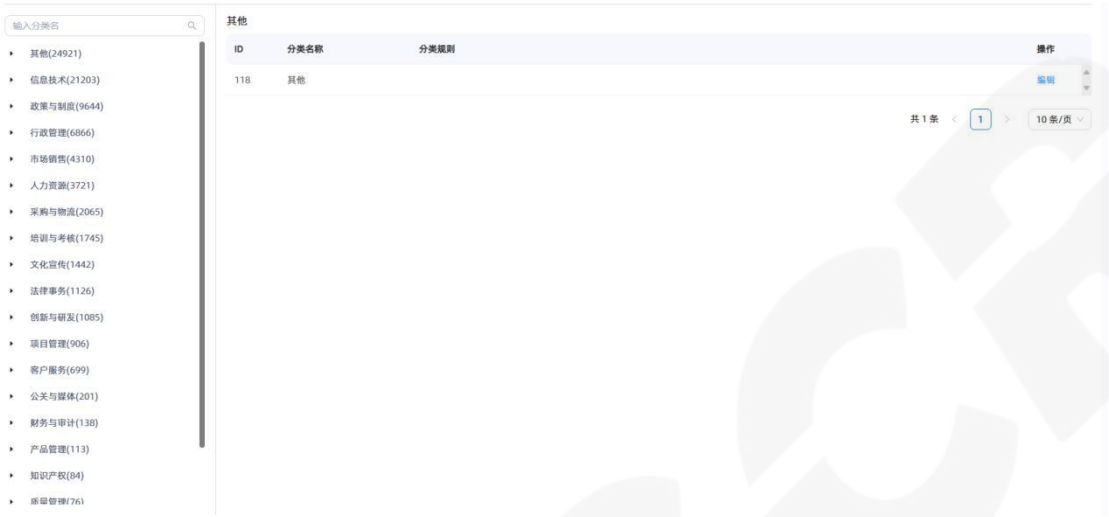


图 19 大道云隐密数云资产保护系统自动分类

(2) 态势感知-渠道管控

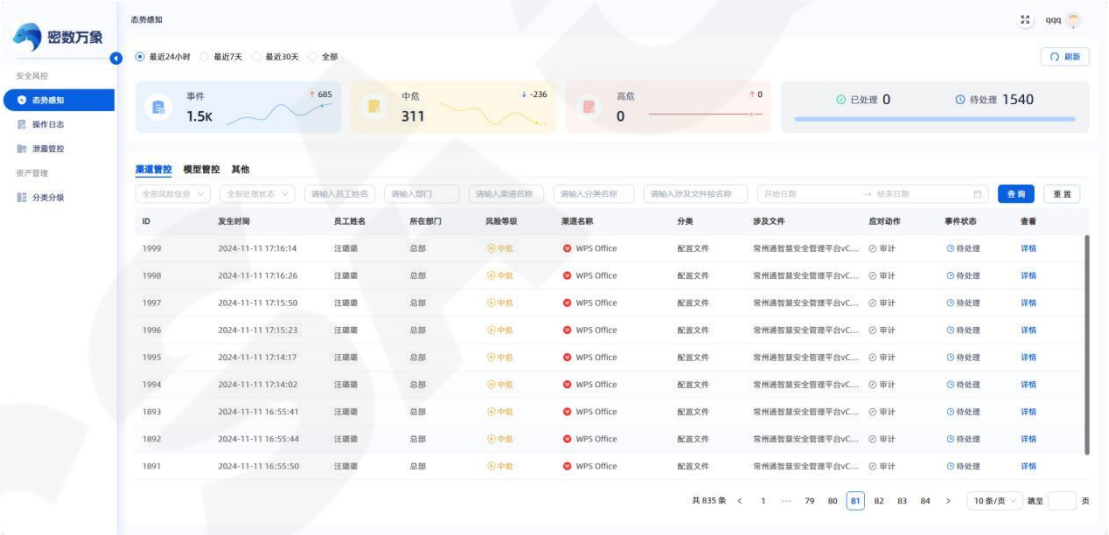


图 20 大道云隐密数云资产保护系统态势感知-渠道管控

(3) 态势感知-模型管控

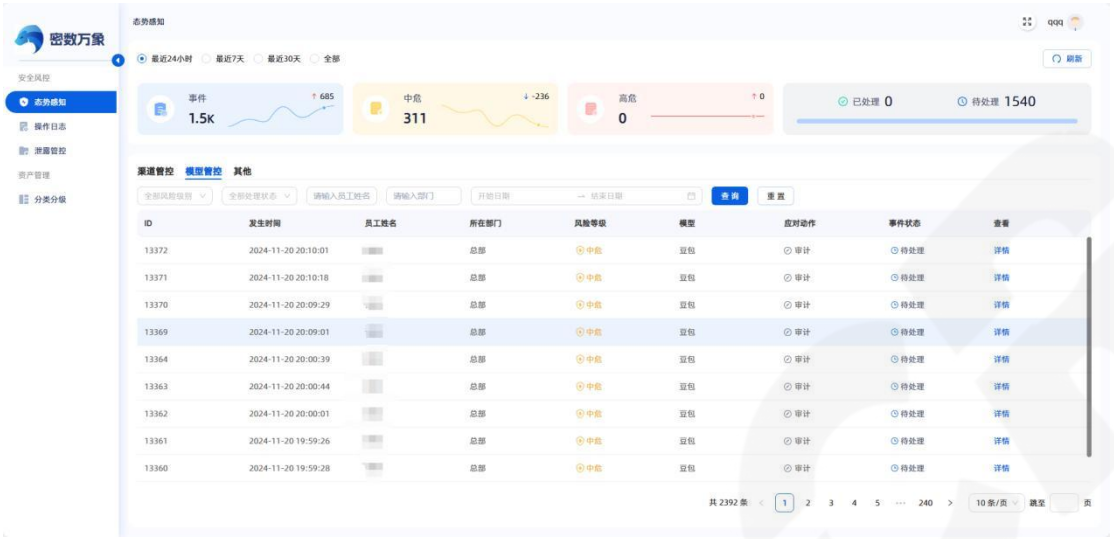


图 21 大道云隐密数云资产保护系统态势感知-模型管控

- **产品应用情况**

密数云产品以其卓越的安全性能与强大的数据处理能力，广泛应用于多个领域，特别是在制造业、金融行业以及政府机构中发挥着至关重要的作用。

A5 神州数码数据分类分级系统

- 产品功能介绍

图 22 神州数码数据分类分级系统功能介绍

神州数码数据分类分级系统是在行业或者企业分类分级标准的基础之上，给企业提供数据分类分级的管理平台，便于企业根据分类分级的结果采取恰当的安全防护措施，保证数据安全。系统支持丰富的数据源，包括传统数据库、大数据平台，以及文本文件，全面兼容国产数据库；支持自动扫描发现、标记敏感数据，如个人身份信息、银行账号等。帮助用户快速识别和梳理企业的数据资产，完成对数据的自动分类分级处理，为用户的安全管控策略提供基础和依据。

系统支持直接对数据库或者数据源进行数据采集，同时也支持第三方的数据通过 API，或者后台数据库的调用去进行数据对接，通过匹配分类分级规则、策略对数据进行自动标记。梳理成数据分类分级清单后可以通过后台数据库或者 API 的方式提供给其他的业务应用系统，实现分类分级系统与其他安全系统的联动。

- 产品特点介绍

(1) 数据源自动发现

基于流量识别技术，自动发现网络中所有的数据库服务器，进而发现数据库实例，系统支持自动批量添加自动扫描发现的数据源，解决手动添加繁琐、网络环境复杂导致的数据资产不清晰等问题。

(2) 敏感数据自动识别

结合关键字、正则表达式匹配规则和 NLP 自然语言处理引擎，识别和标记数据源中的敏感信息，如个人身份信息、资产信息等。

(3) 自动分类分级处理

使用规则引擎和算法来应用分类和分级规则，可以根据组织的需求和策略来定义，例如基于关键词、模式匹配、机器学习等，自动将数据分类到合适的级别。

(4) 数据访问权限控制

系统基于数据分类和分级规则，自动管理数据的访问权限。为数据设置适当的访问级别、角色权限和审批流程，确保只有经过授权的用户可以访问敏感数据。

(5) 数据安全策略应用

自动匹配数据安全防护策略，如加密、脱敏、添加水印等。这可确保在数据传输、处理和共享过程中得到适当的保护和安全控制。

(6) 分类分级及数据资产可视化

将分类分级结果以多个维度进行可视化呈现，以曲线图、柱状图、饼图等方式进行展示，更加直观、易懂。

● 产品展示

(1) 数据源自动扫描发现

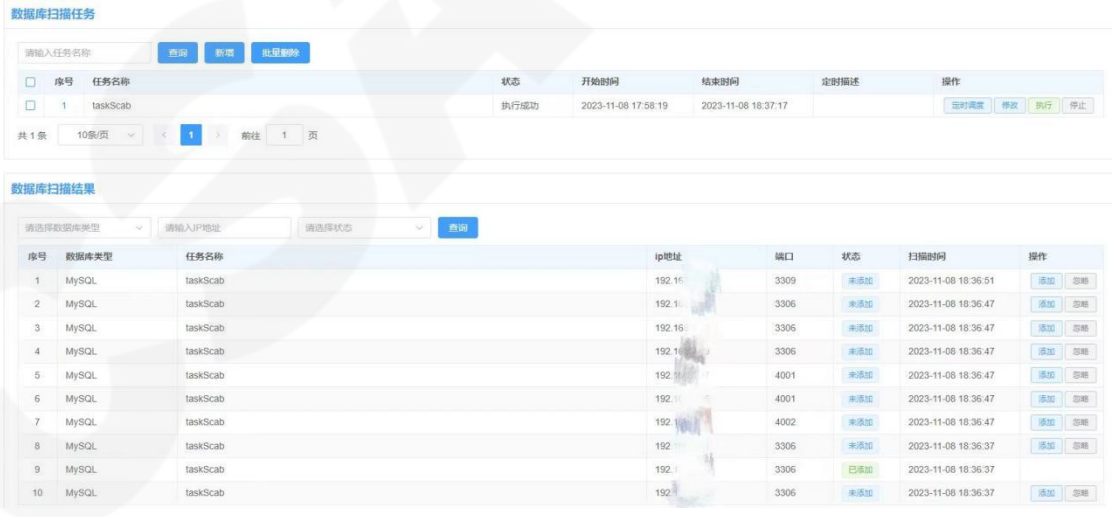


图 23 数据源自动扫描发现

(2) 分类分级策略样例展示

图 24 分类分级策略样例展示

(3) 税务数据分类分级清单样例

模式名(或用户名)*	表名*	字段名	字段备注	一级分类	二级分类	三级分类	安全级别	业务标签	数据标签
		nsrsbh	纳税人识别号	基础数据	法人数据	单位基本信息	重要	纳税人识别号	纳税人识别号
		nsrmc	纳税人名称	基础数据	法人数据	单位基本信息	重要	纳税人名称	纳税人名称
		sqrsfzjhm	申请人身份证件号码	税务数据	其他数据	人员基本信息	重要	申请人身份证件	证件号码
		sqrxm	申请人姓名	税务数据	申报	纳税人基本信息	重要	申请人姓名	中文个人名称
		nsrsbh	纳税人识别号	基础数据	法人数据	单位基本信息	重要	纳税人识别号	纳税人识别号
		nsrmc	纳税人名称	基础数据	法人数据	单位基本信息	重要	纳税人名称	纳税人名称
		sqrlxdh	申请人联系电话	税务数据	登记	纳税人联系信息	重要	申请人联系电话	电话号码
		smrsfzjhm	扫码人身份证件号码|扫	税务数据	其他数据	人员基本信息	重要	扫码人身份证件	证件号码
		smrxm_1	扫码人姓名|扫码人姓名	基础数据	税务机关数据	税务人员数据	重要	扫码人姓名	中文个人名称
		smfnsrsbh	扫码方纳税人识别号|扫	税务数据	征收	纳税人基本信息	重要	扫码方纳税人识	纳税人识别号
		smfnsrmc	扫码方纳税人名称|扫	税务数据	征收	纳税人基本信息	重要	扫码方纳税人名	纳税人名称

图 25 税务数据分类分级清单样例

- **产品应用情况**

神州数码数据分类分级系统支持丰富的数据源，包括主流的关系型数据库、键值数据库、列存储数据库、文档数据库、分布式数据库，支持通用大数据平台包括原生 Hadoop、华为 FI、CDH、CirroData、TDH 等，支持各类文件包括非结构化文件、结构化文件、本地文件、远程文件等数据源。目前，神州数码分类分级系统在金融、税务、部级组织等行业有了实际项目落地。

A6 昂楷数据安全分类分级系统

- 产品功能介绍

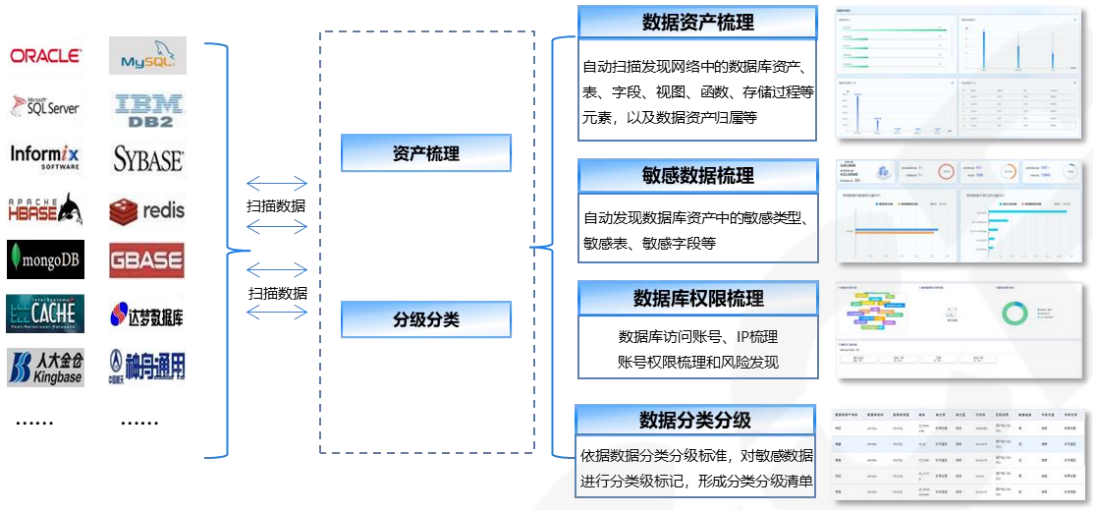


图 26 昂楷数据安全分类分级系统功能介绍

昂楷数据安全分类分级系统是一款数据资产自动发现及数据分类分级管理的安全产品，系统通过自动发现技术智能地发现数据资产信息，进行敏感数据识别，结合数据分类分级标准，快速准确地完成数据资产梳理与数据的分类分级。

- 产品特点介绍

(1) 数据资产梳理

通过 IP 和端口自动扫描的方式自动识别客户网络环境中的所有数据资产，帮助客户清晰了解数据资产分布和使用情况、数据量级、敏感数据分布等数据信息，最终形成数据资产清单。

(2) 账号权限梳理

支持对数据库资产的访问权限进行自动梳理和风险评估，支持设置和维护数据库资产的归属。

(3) 敏感数据识别

内置并支持自定义敏感数据类型与敏感特征库，自动识别发现敏感数据类型并支持人工校正，高效快速生成敏感数据清单。

(4) 内置多套分类分级标准

内置通用、医疗、金融、政务、工业等多套行业分类分级标准并支持标准自定义，

基于 AI 算法的智能标签功能与人工辅助相结合方式，帮助企业快速完成数据的分类分级。

(5) 分类分级结果联动共享

产品提供多种对外接口，可与昂楷科技现有的安全产品以及第三方安全管理平台进行联动，达到联防联控的效果，形成数据安全的防护闭环。

(6) 高可靠性

产品可靠性高，在政府、金融、医疗等多行业具备成功案例，专业的分类分级服务与实施团队，帮助用户快速完成数据分类分级。

- **产品展示**

图 27 昂楷数据安全分类分级系统产品展示

- **产品应用情况**

昂楷数据安全分类分级系统主要应用在政府、金融、医疗等行业，产品支持常见的 Oracle、MySQL、SQL Server、Sybase 等关系型数据库；支持南大通用、达梦、人大金仓、神舟等国产化数据库；支持常用的大数据平台，如：Hive、Hbase 等。

A7 美创暗数据发现和分类分级系统（DDAC）

- 产品功能介绍

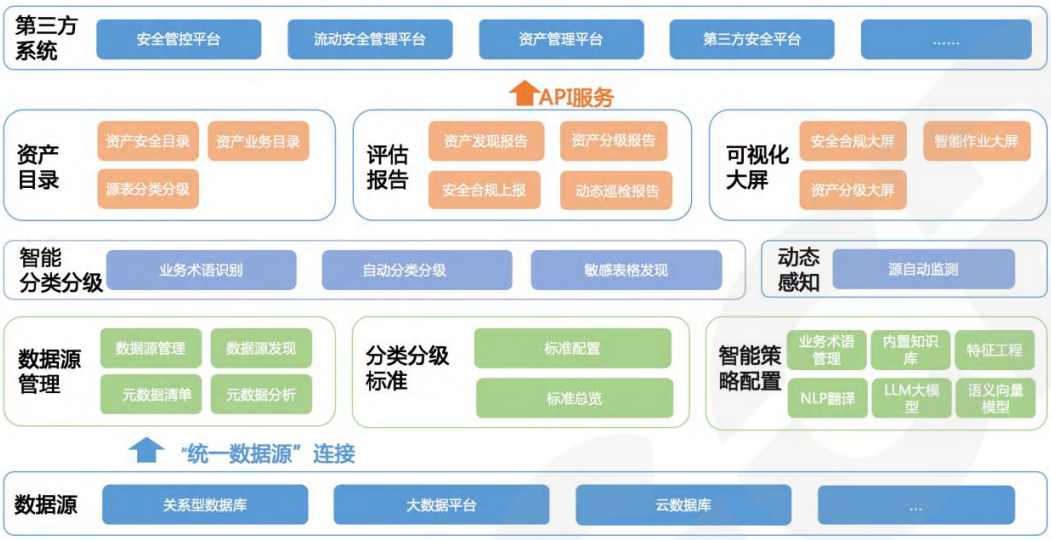


图 28 美创暗数据发现和分类分级系统（DDAC）功能介绍

美创暗数据发现和分类分级系统（DDAC）是一款致力于帮助用户快速发现并识别国家安全、个人安全和组织安全所必需的重要/个人/机密数据的产品。产品通过采集元数据并识别业务含义，并依据内置的分类分级标准进行智能化标签处理，最终形成符合行业特色的敏感资产目录，满足企业安全合规需求，以及数据流动应用场景下的安全防护需求。

(1) 工作台

快捷管理数据分类分级任务的工作平台，确保资产发现及分类分级作业高效开展。

(2) 分类分级标准

从合规角度定义安全数据和判别策略，内置分类分级标准，可支持数据重要数据、个人信息、商业机密的分类分级策略。

(3) 数据资产目录

通过目录形式提供全局安全合规分类分级标准，展示重要数据、个人信息、商业机密分类的多层级目录，为用户提供了安全合规标准目录及目录自定义扩充，在安全合规标准目录基础上，支持不同行业分类分级标准目录内置。

(4) 元数据管理

收集和维护和数据资产相关的元数据，包括数据的来源、格式、结构等信息，围绕

数据源扫描数据库表、Schema 及字段，从而更好地进行数据治理、合规性管理及数据安全风险防护。

(5) 资产发现管理

通过多维度的数据特征解析数据语义，对用户数据进行识别作业，开展数据资产盘点工作，形成资产发现清单，确保用户全面了解其拥有和使用的数据资产，从而更好地进行数据治理、合规性管理和风险管理。

(6) 分类分级管理

基于分类分级标准，利用规则匹配、NLP、智能推荐、AI 大模型等能力的自动识别技术，以提高数据分类分级识别率及准确性为目标，对资产发现清单进行智能分类分级，形成数据分类分级清单。

(7) 数据分析报告

基于数据分类分级结果，提供全局视角、数据库视角、作业视角等多维视角的数据分析统计，提供丰富的图表可视化报告，帮助用户更好地组织管理数据分类分级工作，提升数据安全性及合规性。

(8) 智能策略配置

全局化的安全级别策略配置，支持采用就高原则，就低原则，就多原则等多种安全级别配置策略。

(9) 开放平台

通过标准化的接口提供数据资产发现和分类分级的能力和结果，实现数据分类分级的赋能百态，同时可有效管理已完成对接的第三方应用。

(10) 可视化驾驶舱

以数字化为手段，从资产分级、安全合规、智能作业等不同的场景和需求维度，对分类分级工作过程和结果进行可视化呈现，形成可视化驾驶舱。

- **产品特点介绍**

(1) 分类分级标准

美创历经多年深入行业的研究，在金融、医疗、人社等各行业积累了大量的实践经验，通过对多个行业的分类分级标准的解读、整理，内置为贴合国家法规和行业业务管理需求的数据分类分级发现模板，能够实现对多个行业的自动分类分级梳理。

(2) 动态感知能力

以实时监测能力为基础，通过感知数据源、DDL、DML 等业务源的实时变动，灵活应对不同数据特征和变化的环境，更准确地对数据进行分类和分级，并缩短数据流动中的安全风险窗口。

(3) 智能分类分级

在数据字典、业务模型及规则的海量数据积累基础上，规则匹配、智能推荐、NLP、大模型、机器学习等多种智能能力加持，推动分类分级智能化落地。

(4) 安全数据可视化

构建两套目录 N 张屏，达到资产管理可视化、重要目录上报、数据出境监测等数据合规流动的目标。

(5) 赋能百态

定义和完善现有接口，打造标准对接模式，以接口为窗口，通过接口与其他产品互通信息，减少重复输入和分散管理，赋能各行业各客户需求。

- 产品展示

图 29 美创暗数据发现和分类分级系统（DDAC）产品展示

- 产品应用情况

(1) 应用的行业

美创暗数据发现和数据分类分级系统目前已成功覆盖医疗、金融、人社、高校、电信、证券、燃气、汽车、烟草、农商、大数据行业，也可在其他各行业推广应用。

(2) 支持的数据服务器类型

支持当前主流关系型数据库、国产数据库、大数据平台、云数据库和多类型格式文件。

A8 明朝万达 Chinasec（安元）智能数据治理平台

- 产品功能介绍

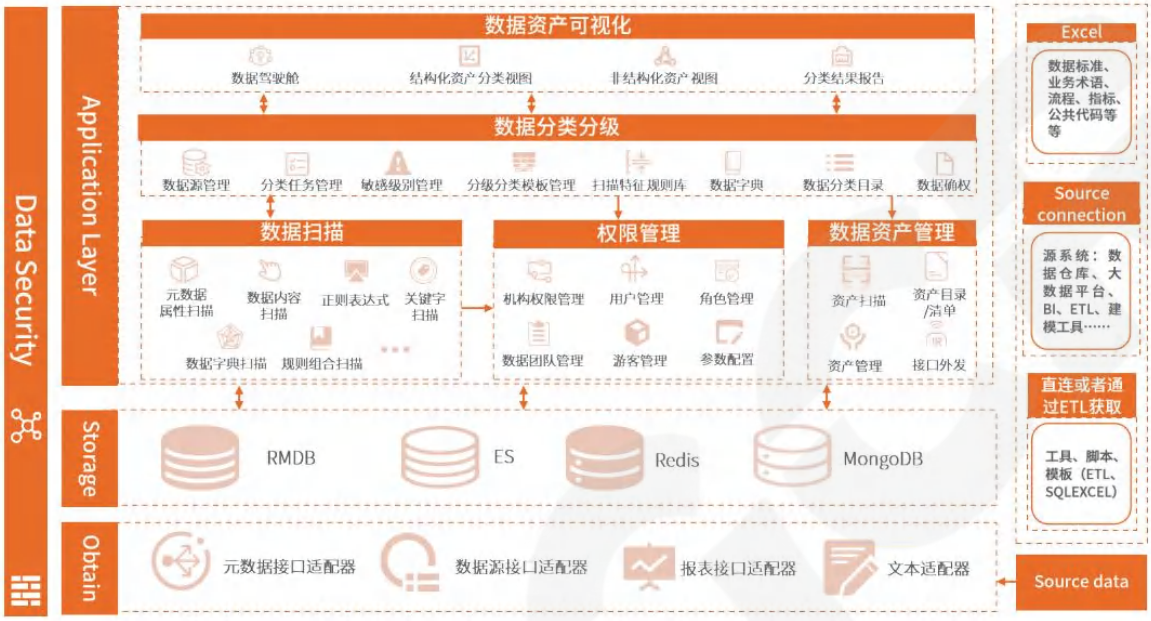


图 30 明朝万达 Chinasec（安元）智能数据治理平台功能介绍

明朝万达 Chinasec（安元）智能数据治理平台以数据分类分级作为数据治理切入点，通过自动嗅探资产可对企业内部数据资产进行敏感数据识别、梳理，以及对数据的有效理解和分析，完成对数据进行不同类别和密级的划分，形成数据资产清单、形成数据分类分级报告，全局掌握企业数据资产，有效解决了企业对数据资产的摸底以及安全保护管理工作。

- 产品特点介绍

帮助企业梳理数据资产，使企业可以快速、高效的了解企业现有情况。

平台内置企业内部数据安全分类分级的行业标准模版，方便对企业内部数据进行分类分级。

通过数据采集和分析以及分类分级的匹配，最终生成数据安全策略，解决企业使用了安全产品策略却无从下手的尴尬问题，以及策略配置不完整，导致安全漏洞的问题。

随着企业内部业务以及系统的调整，定时更新业务元数据和技术元数据，然后实时生成最新的数据安全策略，推送给企业内部的安全产品，形成自动化实时更新的安全防护体系，保护用户的数据安全。

多维度的安全防护，可以通过采集用户业务条线，组织架构，人员等多维护结合数据安全基础元数据，使得数据安全可以落实到每个业务，部门或者个人，真正实现数据安全的无死角。

- **应用场景**

目前主要应用于金融、政企、公安等领域，为企业整体的安全部署提供数据安全策略。部署灵活，过程简单，与各安全系统之间可以很好的整合和接入，从而为企业提供完整的安全防护体系，保护企业的数据安全。

- **产品展示**

图 31 明朝万达 Chinasec（安元）智能数据治理平台产品展示

- **产品应用情况**

在国家和行业规范标准的基础上，通过多年的实际分类分级业务沉淀，智能数据治理平台内置构建了适合金融、医疗行业、能源行业、个人信息、证券行业等不同行业的基础分类分级模版，并可根据企业实际的业务需要快速动态调整，以构建适合企业的数据分类分级模版。智能数据治理平台支持多种数据库与文件的自动发现，主动嗅探网内数据库、文件，支持指定 IP 段和端口的范围进行搜索资产，自动探查数据存储位置。

支持的数据服务器类型：关系型数据库、国产数据库、大数据库。

非结构化数据获取支持多种协议：FTP、FTPS、SFTP、HDFS、NFS、SMB、MINIO 等。

附录 B-数据分类分级参考模板

数据分类分级过程中的各类清单模板不限于固定形式，组织应当根据各行业各领域要求，结合自身业务实际，参考以下各类模板设计符合自身的执行结果要求。本文从结构化数据、非结构化数据、数据对象（数据集、数据项）为视角设计了如下参考模板。

B1《数据资产清单模板》

B1.1 结构化数据清单

表 5 结构化数据清单模板

序号	应用系统名称	数据库名称	数据库中文名	数据表名称	数据表中文名	字段名称	字段中文名	字段说明	数据量级	数据所有方	数据使用方	数据管理方
1												
2												

B1.2 非结构化数据清单

表 6 非结构化数据清单模板

序号	应用系统名称	数据文件存储路径	数据文件名称	数据文件内容描述	数据文件格式	数据文件量级	数据文件所有方	数据文件使用方	数据文件管理方
1									
2									

B1.3 数据对象清单

数据对象识别工作，可以在结构化数据识别和非结构化数据识别基础上开展，通过调研业务活动中具有上下文含义的数据对象（例如：简历信息、交易信息等数据集或姓名、身份证、地址等数据项），分析数据集所关联数据项的集合信息，从而支撑对业务活动中的数据对象集进行定级。

表 7 数据对象清单模板

序号	数据对象名称	所包含数据表	所包含数据项	数据形态	数据规模	所属业务	数据源	数据所有方	数据使用方	数据管理方
1										
2										

B2《数据分类分级标记模板》

B2.1 结构化数据分类分级清单

表 8 结构化数据分类分级清单模板

序号	应用系统名称	数据资源服务器	数据库名称	数据库中文名	数据表名称	数据表中文名	字段名称	字段中文名	字段类型	字段长度	数据类别	安全级别	影响对象	影响程度
1														
2														

B2.2 非结构化数据分类分级清单

表 9 非结构化数据分类分级清单模板

序号	应用系统名称	数据资源服务器	文件数据库名称	文件路径	文件目录描述	文件名	文件中文名称	文件类型	文件大小	数据类别	安全级别	影响对象	影响程度
1													
2													

B2.3 数据对象分类分级清单

表 10 数据对象分类分级清单模板

序号	数据对象名称	所包含数据表	所包含数据项	数据形态	数据规模	所属业务	数据源	数据所有方	数据使用方	数据管理方	数据类别	数据级别
1												
2												

附录 C-数据分类分级参考资料

本文引用、归纳、总结了如下数据分类分级参考资料，供组织在开展个人信息识别、重要数据识别、数据分级要素识别的过程进行快速参考。

C1 个人信息识别参考

个人信息可参考国家标准《信息安全技术 个人信息安全规范》（GB/T 35273—2020）进行识别。

个人信息是指以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息，如姓名、出生日期、身份证件号码、个人生物识别信息、住址、通信通讯联系方式、通信记录和内容、账号密码、财产信息、征信信息、行踪轨迹、住宿信息、健康生理信息、交易信息等。

个人敏感信息是指一旦泄露、非法提供或滥用可能危害人身和财产安全，极易导致个人名誉、身心健康受到损害或歧视性待遇等的个人信息。通常情况下，14 岁以下（含）儿童的个人信息和涉及自然人隐私的信息属于个人敏感信息。

表 11 个人信息示例

信息类别	信息内容概述
个人基本资料	个人姓名、生日、性别、民族、国籍、家庭关系、住址、个人电话号码、电子邮件地址等。
个人身份证信息	身份证、军官证、护照、驾驶证、工作证、出入证、社保卡、居住证等。
个人生物识别信息	个人基因、指纹、声纹、掌纹、虹膜、面部识别特征等。
网络身份标识信息	个人信息主体账号、IP 地址、个人数字证书等。
个人健康生理信息	个人历史病历等产生的相关记录，如病历、住院记录、医嘱单、检验报告、手术及麻醉记录、护理记录、用药记录、药物食物过敏信息、生育信息、门诊病史、诊治情况、家族病史、现病史、传染病等；以及与个人身体健康状况相关的信息，如体重、身高、肺部音等。
个人教育工作信息	个人职业、职位、工作单位、学历、学位、教育经历、工作经历、培训记录、成绩单等。
个人财产信息	银行账户、鉴别信息（口令）、存款信息（包括金额、支付收款记录等）、房产信息、信贷记录、征信信息、交易和消费记录、流水记录等，以及虚拟货币、虚拟兑换及兑奖码等虚拟财产信息。

个人通信信息	通信记录和内容、短信、彩信、电子邮件，以及描述个人通信的数据（通常称为元数据）等。
联系人信息	通讯录、好友列表、群列表、电子邮件地址列表等。
个人上网记录	通过日志储存的个人信息主体操作记录，包括网站浏览记录、软件使用记录、点击记录、收藏夹记录等。
个人常用设备信息	指包括硬件序列号、设备 MAC 地址、软件列表、唯一设备识别码（如 IMEI/Android ID/IDFA/OpenUDID/GUID/SIM 卡 IMSI 信息等）等在内的描述个人常用设备基本状况的信息。
个人位置信息	包括行踪轨迹、精准定位信息、住宿信息、经纬度等。
其他信息	婚史、宗教信仰、性取向、未公开的违法犯罪记录等。

表 12 敏感个人信息示例

个人财产信息	银行账户、鉴别信息(口令)、存款信息（包括资金数量、支付收款记录等）、房产信息、信贷记录、征信信息、交易和消费记录、流水记录等，以及虚拟货币、虚拟交易、游戏类兑换码等虚拟财产信息
个人健康生理信息	个人因生病医治等产生的相关记录，如病症、住院志、医嘱单、检验报告、手术及麻醉记录、护理记录、用药记录、药物食物过敏信息、生育信息、以往病史、诊治情况、家族病史、现病史、传染病史等
个人生物识别信息	个人基因、指纹、声纹、掌纹、耳廓、虹膜、面部识别特征等
个人身份信息	身份证、军官证、护照、驾驶证、工作证、社保卡、居住证等
其他信息	性取向、婚史、宗教信仰、未公开的违法犯罪记录、通信记录和内容、通讯录、好友列表、群组列表、行踪轨迹、网页浏览记录、住宿信息、精准定位信息等

C2 重要数据识别参考

参考国家标准《数据安全技术 数据分类分级规则》（GB/T 43697——2024），重要数据是一旦遭到泄露、篡改、损毁或者非法获取、非法使用、非法共享，直接对国家安全造成一般及以上危害、对经济运行造成严重及以上危害、对社会秩序造成严重及以上危害、对公共利益造成严重及以上危害的数据，以及直接或达到一定精度、规模、深度或重要性后关系国家安全、经济运行、社会稳定、公共健康和安全的特定领域、特定群体或特定区域的数据。同时还需考虑如下因素：

(1) 直接影响领土安全和国家统一，或反映国家自然资源基础情况，如未公开的领陆、领水、领空数据；

(2) 可被其他国家或组织利用发起对我国的军事打击，或反映我国战略储备、应急

动员、作战等能力，如满足一定精度指标的地理数据或与战略物资产能、储备量有关的数据；

(3) 直接影响市场经济秩序，如支撑关键信息基础设施所在行业、领域核心业务运行或重要经济领域生产的数据；

(4) 反映我国语言文字、历史、风俗习惯、民族价值观念等特质，如记录历史文化遗产的数据；

(5) 反映重点目标、重要场所物理安全保护情况或未公开地理目标的位置，可被恐怖分子、犯罪分子利用实施破坏，如描述重点安保单位、重要生产企业、国家重要资产（如铁路、输油管道）的施工图、内部结构、安防情况的数据；

(6) 关系我国科技实力、影响我国国际竞争力，或关系出口管制物项，如反映国家科技创新重大成果，或描述我国禁止出口限制出口物项的设计原理、工艺流程、制作方法的数据，以及涉及源代码、集成电路布图、技术方案、重要参数、实验数据、检测报告的数据；

(7) 反映关键信息基础设施总体运行、发展和安全保护情况及其核心软硬件资产信息和供应链管理情况，可被利用实施对关键信息基础设施的网络攻击，如涉及关键信息基础设施系统配置信息、系统拓扑、应急预案、测评、运行维护、审计日志的数据；

(8) 涉及未公开的攻击方法、攻击工具制作方法或攻击辅助信息，可被用来对重点目标发起供应链攻击、社会工程学攻击等网络攻击，如政府、军工单位等敏感客户清单，以及涉及未公开的产品和服务采购情况、未公开重大漏洞情况的数据；

(9) 反映自然环境、生产生活环境基础情况，或可被利用造成环境安全事件，如未公开的与土壤、气象观测、环保监测有关的数据；

(10) 反映水资源、能源资源、土地资源、矿产资源等资源储备和开发、供给情况，未公开的描述水文观测结果、耕地面积或质量变化情况的数据；

(11) 反映核材料、核设施、核活动情况，或可被利用造成核破坏或其他核安全事件，如涉及核电站设计图、核电站运行情况的数据；

(12) 关系海外能源资源安全、海上战略通道安全、海外公民和法人安全，或可被利用实施对我国参与国际经贸、文化交流活动的破坏或对我国实施歧视性禁止、限制或其他类似措施，如描述国际贸易中特殊物项生产交易以及特殊装备配备、使用和维修情况的数据；

(13) 关系我国在太空、深海、极地等战略新疆域的现实或潜在利益，如未公开的涉及对太空、深海、极地进行科学考察、开发利用的数据，以及影响人员在上述领域安全进出的数据；

(14) 反映生物技术研究、开发和应用情况，反映族群特征、遗传信息，关系重大突发传染病、动植物疫情，关系生物实验室安全，或可能被利用制造生物武器、实施生物恐怖袭击，关系外来物种入侵和生物多样性，如重要生物资源数据、微生物耐药基础研究数据；

(15) 反映全局性或重点领域经济运行、金融活动状况，关系产业竞争力，可造成公共安全事故或影响公民生命安全，可引发群体性活动或影响群体情感与认知，如未公开的统计数据、重点企业商业秘密；

(16) 反映国家或地区群体健康生理状况，关系疾病传播与防治，关系食品药品安全，如涉及健康医疗资源、批量人口诊疗与健康管理、疾控防疫、健康救援保障、特定药品实验、食品安全溯源的数据；

(17) 其他可能影响国土、军事、经济、文化、社会、科技、电磁空间、网络、生态、资源、核、海外利益、太空、极地、深海、生物、人工智能等安全的数据；

注：若需更多指导，请参考《数据安全技术 数据分类分级规则》（GB/T 43697-2024）原文。

C3 数据分级要素参考示例

表 13 部分数据分级要素参考示例

要素	要素说明	要素示例
领域	数据描述的业务或内容范畴。	如行业领域、业务条线、业务类目、生产经营活动、流程环节、内容主题、与国家安全、经济运行、社会秩序、公共利益相关的领域等。
群体	数据主体或描述对象集合。	1.企业员工数据，如员工的姓名、身份证、手机号、职位等； 2.个人用户群体数据，如用户的年龄、性别、身份、历史用电量等； 3.组织用户群体数据，如组织的名称、法人信息、地址、历史用电量等。
区域	数据涉及的地区范围。	如行政区域、特定区域、军事禁区、军事管理区、国防科工单位以及县级以上党政机关区域等；
规模	数据描述对象覆盖的群体、区域的范围或数量大小。	1.群体数量，如去重后的 10 万人个人信息数据； 2.数据存储量，同类数据存储量达到一定规模，如 1T 系统监测信息； 3.记录条数，如记录数量达到一定规模，100 万条。
精度	数据的精确程度。数据精度越高表示采集数据和真实数据的误差越小。	1.空间精度，如经纬度坐标，精度和比例尺优于开放标准的地图数据； 2.时间精度，如时分秒信息； 3.图像精度，如高分辨率影像； 4.定位精度，如 1～10 米范围级位置定位数据。
深度	对数据描述对象的隐含信息挖掘的触达程度，或多维度细节信息的刻画程度。	如通过数据统计、关联、挖掘或融合等加工处理，对数据描述对象的隐含信息或多维度细节信息的刻画程度。
覆盖度	数据对领域、群体、区域、时段等的覆盖分布或疏密程度。	1.员工数量占行业从业人员的百分比； 2.用户数量占总市人口的百分比。

C4 数据影响对象识别参考示例

(1) 国家安全

判断数据是否可能影响国家安全，常见考虑因素包括但不限于：

a) 影响国家政权安全、政治制度安全、意识形态安全、民族和宗教政策安全；

b) 影响领土安全、国家统一、边疆安全和国家海洋权益；

c) 影响基本经济制度安全、供给侧结构性改革、粮食安全、能源安全、重要资源安全、系统性金融风险、国际开放合作安全；

d) 影响国家科技实力、科技自主创新、关键核心技术、国际科技竞争力、科技

伦理风险、出口管制物项；

e) 影响社会主义核心价值观、文化软实力、中华优秀传统文化等；

f) 影响国家社会治理体系、社会治安防控体系、应急管理体系等；

g) 影响生态环境安全、绿色生态发展、污染防治、生态系统质量和稳定性、生态环境领域国家治理体系等；

h) 影响国防和军队现代化建设等，或者可被其他国家或组织利用发起对我国的军事打击；

i) 影响电磁空间、网络空间安全、关键信息基础设施安全、人工智能安全，或者可能被利用实施对关键信息基础设施、核心技术设备等的网络攻击，可能导致特别重大或重大网络安全和数据安全事件；

j) 影响核材料、核设施、核活动情况，或可被利用造成核破坏或其他核安全事件；

k) 影响国家生物安全治理体系、生物资源和人类遗传资源安全、生命安全和生物安全领域的重大科技成果、疾病防控和公共卫生应急体系安全，或者可能导致重大传染病、重大生物安全风险；

l) 影响在太空、深海、极地等领域的国家利益和国际合作安全；

m) 影响海外重大项目和人员机构安全、海外能源资源安全、海上战略通道安全等。

(2) 公共利益

判断数据是否可能影响公共利益，常见考虑因素包括但不限于：

a) 影响对重大疾病（尤其是传染病）的预防、监控和治疗，或者可能引发突发公共卫生事件、 造成社会公众健康危害；

b) 影响社会成员使用公共设施；

c) 影响社会成员获取公开数据资源；

d) 影响社会成员接受公共服务等方面；

e) 其他影响公共利益、社会秩序的数据。

(3) 个人权益

判断数据是否可能影响个人权益，常见考虑因素包括但不限于：

a) 影响个人私人活动、私有领域、私密部位等个人隐私；

b) 影响自然人的人格尊严；

c) 影响自然人的人身安全；

d) 影响自然人的财产安全；

e) 影响个人在个人信息处理活动中的权利，如选择权、知情权、拒绝权等；

f) 其他影响个人权益的数据。

(4) 组织自身

判断数据是否可能影响组织自身，常见考虑因素包括但不限于：

a) 导致组织遭到监管部门处罚、安全事件或法律诉讼；

b) 影响组织的重要或关键业务生产经营；

c) 造成组织经济损失；

d) 破坏组织声誉形象、公信力等；

e) 影响组织的知识产权、商业秘密、技术损失等；

f) 影响组织的公平竞争利益；

g) 其他影响法人、非法人组织合法权益的数据。

C5 数据分类分级变更参考示例

数据分类分级变更情形	数据级别变更建议
数据体量增加到特定规模导致影响对象发生变化或影响程度变大	升级
数据精度发生变化达到相关部门规定阈值	升级
数据汇聚后	升级
数据涉及的主体增多	升级
特定事件发生后数据敏感程度增加	升级
数据被公开或披露后	降级
数据脱敏后	降级
数据去标识化、匿名化后	降级

附录 D-数据分类分级关键技术与方法

D1 关键字匹配

基于关键字的敏感数据识别通常是通过读取数据库的表和字段的描述，匹配固定关键字、关键字对、关键字组以识别敏感数据。通常需要配合关键字权重、顺序、组合形式等多种参数使用，这种方法和语义识别方法结合可更好地提升识别准确率和效率。

常见的语义识别中文分词技术常见的有两大类：

1.机械分词技术、基于统计的序列标注技术。机械分词技术操作简单、方便，比较省心，但是对于歧义词以及未登录词的效果并不是很好；

2.统计模型的序列标注方法，对于识别未登录词拥有较好的识别能力，而且分词精度也比较大，同时这个方法可以不分中文、英语，着重看在语言前后顺序。

图 31 中文分词技术介绍

中文分词技术可应用于数据库敏感数据的关键词识别判定场景下。首先，判断数据库表名称和字段名称的描述信息中是否含有中文字符，对于含有中文字符的字段值进行分词。然后，将其分解为子单词和单词属性（名词、动词等）。针对分解后的名词在敏感数据特征词词典中进行匹配。最后判定是否包含有敏感数据。

D2 正则表达式检测

正则表达式是由普通字符（例如字符 a 到 z）以及特殊字符（称为"元字符"）组成的文字模式。可以用来检查一个串是否含有某种子串、将匹配的子串替换或者从某个串中取出符合某个条件的子串等。

正则表达式检测法是当前业内最常用的识别方法，通过对数据内容进行特征提取和

抽象，形成正则表达式，对数据内容进行正则匹配。

例如：

身份证号正则表达式为r'^[1-9]\d{5}(18|19|20)\d{2}((0[1-9])|(1[0-2]))(([0-2][1-9])|10|20|30|31)\d{3}[0-9Xx]$';

正则表达式还可以对数据标识符进行检测。例如：身份证是 18 位数字，而 18 位数字不一定是身份证。身份证的后四位数字用来对身份证进行校验以检验给定的 18 位数字是否是正确的身份证。同样有很多类似的数据，比如银行卡号，这类带有验证信息并通过公开/私密的算法能够完成数据验证的数据，为数据标识符。传统的数据识别方法便是匹配这些数据标识符。看是否满足规则要求，大部分数据在数据满足这些规则的时候认为它是特定标识的数据。

D3 指纹匹配

指纹匹配技术是基于数据或文件指纹，从样本数据中提取并生成指纹特征库，然后以同样的方法从待检测文档或内容中提取指纹，将得到的指纹与指纹库进行匹配，对比其相似度，进行指纹检测和识别，从而实现敏感数据匹配。此方法的难点在于数据或文件指纹的提取与学习，首先要提供含有企业想要保护的特定内容的文档集作为训练数据。然后对这些文档生成指纹，形成指纹库，并配置数据检测规则用于检测受保护的文档。

指纹匹配的过程包括指纹提取、指纹生成、指纹存储、指纹匹配、四个部分，如下图所示：

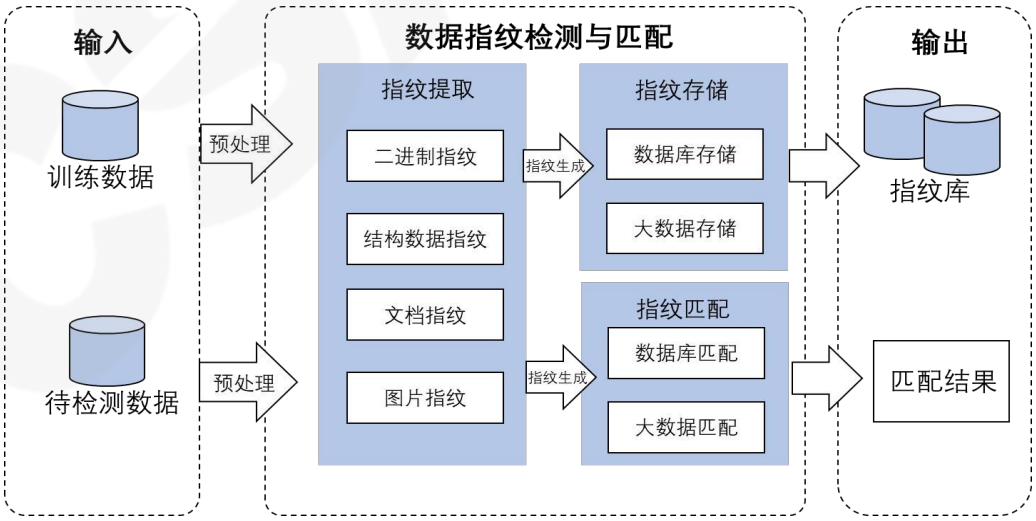


图 32 指纹匹配技术

D4 自然语言处理（NLP）

自然语言处理（NLP）驱动计算机程序将文本从一种语言翻译成另一种语言，响应语音命令，以及快速甚至实时总结大量文本。在日常生活中，您有可能与多种形式的自然语言处理（NLP）互动，包括 GPS 语音系统、数字助手、语音文本听写软件、客服聊天机器人以及其他为消费者提供便利的系统。此外，自然语言处理（NLP）在企业解决方案中也发挥着越来越大的作用，有助于精简业务运营，提高员工生产力以及简化任务关键型业务流程。

自然语言处理（NLP）任务通过分解人类文本和语音数据，帮助计算机理解所采集的内容。这些任务包括：

- 语音识别，也称为语音转文本：用于将语音数据以可靠的方式转换为文本数据。任何遵循语音命令或回答口述问题的应用都需要语音识别功能。语音识别的挑战性在于人们的说话方式—语速快，含糊不清，各种重音、语调和口音，以及语法常常不正确。
- 词性标注，也称语法标注：这个过程按照用法和上下文确定特定单词或文本片段的词性。"Icanmakeapaperplane" 中 "make" 的词性为动词，"Whatmakeofcardoyouown?"中"make"为名词。
- 词义消歧：用于对多义单词选择含义，通过语义分析过程确定单词在特定上下文中最准确的意思。例如，词义消歧可帮助区分动词。"make"在"makethegrade"（达到）和"makeabet"（做出）中的含义。
- 命名实体识别，简称 NEM：用于将单词或短语识别为有意义的实体。NEM 将"Kentucky"识别为地点，将"Fred"识别为男性的名字。
- 指代消解：用于确定两个单词是否以及何时指代同一实体。最常见的例子是确定某个代词所指的人或物体（例如，"她"指玛丽），但也可能涉及识别文本中的隐喻或习语（例如，"熊"有时并不表示动物，而是指体形魁梧、体毛较多的人）。
- 情绪分析：尝试从文本中提取主观特质，例如，态度、情绪、讽刺、困惑和怀疑。
- 自然语言生成：有时被视为语音识别或语音转文本的逆操作；该任务用于将结

构化信息转化为人类语言。

D5 机器学习

目前大多数文本分类研究工作都基于深层学习模型，DNN 是数据驱动的方法，具有很高的计算复杂性。

少量的文本数据可以通过人工的方式进行分类分级，但是随着产业数字化进程的不断深入，各企业和组织逐渐沉淀了大量的数据，同时大量数据并没有遵循数据分类分级的标准。进行数据治理过程中，这些海量的文本数据都依赖人工分类分级几乎无法实现。为了解决这一问题，大量科学家以及从业者们开始探索通过机器学习算法进行文档的自动分类。这些算法主要可以分为两类，分别是有监督学习算法（又称分类算法）和无监督学习算法（又称聚类算法），主要有两种方式：浅层机器学习和深层机器学习。具体如下所示：

D5.1 浅层机器学习

常用的浅层机器学习算法包括朴素贝叶斯（NB)、K 近邻分类（KNN)、决策树（DT)、支持向量机（SVM）等，下面简单介绍各种算法的原理。

(1) 朴素贝叶斯（NB)

朴素贝叶斯（Naïve Bayes）是一种最简单的概率模型。它基于如下假设：文本中的特征是相互独立的。根据贝叶斯方法，对于给定的文本 Dj 和类别集合 C={C1, C2, … , Cm}， 文本 Dj 属于类别 Ci 的概率可由下列公式给出：

$$P(C_i \mid D_j) = \frac{P(D_j \mid C_i)P(C_i)}{P(D_j)}$$

对给定的文本 Dj 进行分类，也就是计算出文本 Dj 属于每个类别的概率，选取概率最大的作为文本 Dj 的最终类别。训练及判决过程如下：

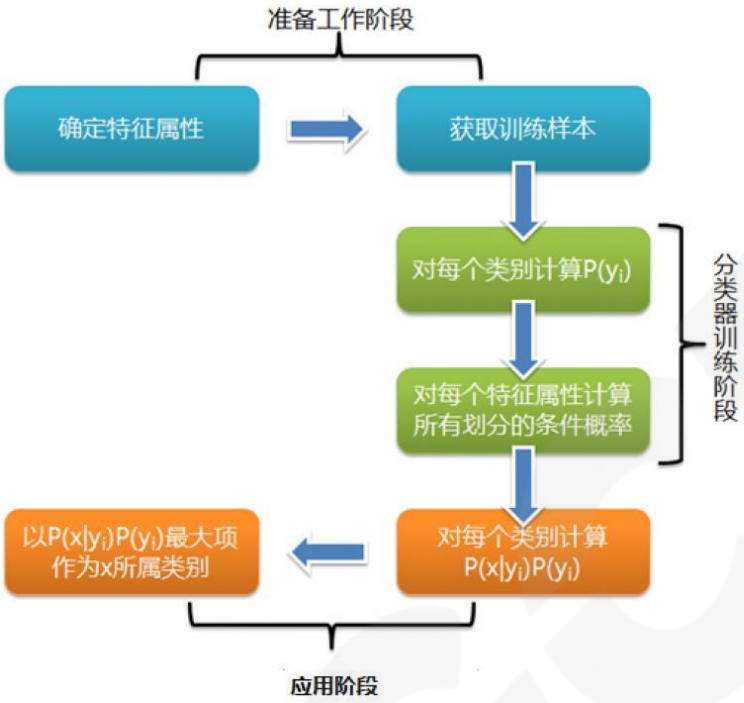


图 33 朴素贝叶斯（Naïve Bayes）概率模型

(2) K 近邻分类（KNN)

KNN 算法的核心思想是通过判断未知数据与已知数据的相似性进行分类。具体而言，首先选择与未知数据最近的 k 个已知类别的数据点，然后统计这 k 个数据点的类别分布，并以多数类别或平均距离最小的类别作为该未知数据的分类结果。

(3) 决策树（DT)

决策树算法的核心是构建一棵二叉树，其中每个非叶节点代表对一个特征的测试，每个分支对应该特征的一个取值范围，而每个叶节点则表示一个分类结果。决策树的关键在于选择最佳特征，以便将数据分为两部分，每部分的数据“纯度”越高越好。常用的纯度指标有 Gini 系数和信息熵。

(4) 支持向量机（SVM)

支持向量机（SVM，Support Vector Machine）是一种适用于文本分类的二分类方法，其目标是找到一个最大间隔超平面，将不同类别的训练数据分开。空间中可能存在多个可以分隔数据的超平面，为提升分类器的泛化能力，SVM 选择具有最大间隔的超平面作为分类边界，因此也称为最大间隔分类器。尽管 SVM 本质上是二分类算法，通过适当的策略也可以扩展到多分类任务。

D5.2 深层机器学习

深层学习模型（DNN）由人工神经网络组成，通过模拟人脑的方式从数据中自动学习高级特征。深度学习由多个隐藏层构成，具有较高的复杂性，能够在非结构化数据上进行训练。深度学习架构可以直接从输入中学习特征表示，减少了对人工干预和先验知识的依赖。实践表明，深度学习模型在文本分类任务中的准确性优于传统浅层模型。这类模型通过捕捉词语之间的上下文关系，以更深层次理解文本含义，典型代表包括 TextCNN、TextRNN 和 BERT 语言模型。

D6 规则引擎

在数据量少标注少的冷启动情况下，由于可用的训练数据非常有限，基于统计的机器学习或深度学习模型往往难以取得理想的效果。此时，规则引擎成为了一个有效的替代方案。利用正则表达式、先验知识、全文匹配、模糊匹配和黑白名单等技术，可以基于已知的信息和模式快速构建初步的文本分类系统。

即使数据量较大，但标注数据不足的低资源情况下，基于统计的方法也会受到限制。此时，可以结合规则引擎和半监督学习方法，利用少量标注数据引导模型学习，同时利用大量未标注数据进行迭代优化。此外，还可以考虑迁移学习，将在其他相关任务上学到的知识迁移到当前任务中。

在完全没有数据的极端情况下，规则引擎几乎成为唯一可行的解决方案。此时，需要依靠领域专家的知识和经验，通过编写详细的规则来覆盖可能的文本类型和分类需求。

规则引擎作为一种强大的工具，为企业数字化转型提供了重要支持。其优点主要体现在以下几个方面：首先，规则引擎通过将业务逻辑与系统代码分离，简化了业务逻辑的修改和维护，降低了开发成本。其次，规则引擎具有高度的灵活性和可扩展性，可以根据业务需求快速定制和调整规则，提高系统的响应速度和准确性。此外，规则引擎还支持可视化管理，使得规则的管理和监控变得更加直观和便捷。

然而，规则引擎也存在一些缺点。一方面，由于规则引擎通常涉及大量的业务规则，因此规则的调试和维护成本较高。另一方面，当规则数量庞大时，规则引擎的性能可能会受到影响，导致系统响应速度变慢。此外，规则引擎的使用还需要一定的技术门槛，需要开发人员具备一定的专业知识和经验。

D7 元数据分析

元数据分析技术包括元数据挖掘、语义分析和机器学习算法。元数据挖掘涉及从与每个数据集关联的属性中提取有价值的模式和趋势。语义分析深入研究每个属性代表什么，以及它们之间的关系，其中，可以借助机器学习算法帮助根据大型数据集中发现的模式自动分类和识别元数据属性。元数据分析功能主要实现针对元数据的基本分析功能，包括血缘分析（血统分析）、影响分析、实体关联分析、实体影响分析、主机拓扑分析、指标一致性分析等，它能够帮助组织单位合理分配数据保护资源和成本，是组织单位建立全生命周期数据保护框架的基础，也是有的放矢地实施数据安全管理的前提条件，统一的数据分类分级管理制度，能够促进数据在机构间、行业间的安全共享，有利于数据价值的挖掘与实现。

附录 E-典型行业数据分类分级标准解读

E1 政务

随着政务数字化进程的持续推进，电子政务领域蓬勃发展，大量涉及政务、民生、国家安全、公共利益的数据沉积在线上，为政务数据的保护带来巨大挑战。另一方面，《中华人民共和国数据安全法》中将数据分类分级保护制度与重要数据目录直接对应，并要求各地区、各部门按照数据分类分级保护制度，确定本地区、本部门以及相关行业、领域的重要数据具体目录。截至目前，政务领域涉及数据分类分级的相关标准研究如下：

E1.1GB/T39477—2020《信息安全技术政务信息共享数据安全技术要求》

2020 年 11 月，国家市场监督管理总局和国家标准化管理委员会联合发布 GB/T 39477——2020《信息安全技术政务信息共享数据安全技术要求》标准。

该标准是“政务数据开放共享标准体系”的重要组成部分，标准在国内首次提出了政务信息资源数据流转全过程中的身份认证、数据脱敏、数据加密等数据安全技术要求。通过制定并实施政务信息共享数据安全标准，推进各级政务部门电子政务领域数据应用，有效提升政务信息资源采集、共享、使用过程的数据安全防护能力，全面保障政务信息资源共享交换的数据安全。该标准开创性地提出数据全生命周期过程中的多种角色（共享数据提供方、共享数据交换服务方及共享数据使用方）在共享数据准备阶段、共享数据交换阶段及共享数据使用阶段的权利和责任义务。

该标准较粗力度地阐述了共享数据提供方对政务数据分类分级的总体要求，具体要求如下：a）应按照政务信息资源分级分类相关要求对共享数据分级分类并进行标记，根据标记可对数据安全等级进行识别，并保留标记记录；b）应按照数据级别确定并实施所必要的安全管理策略和保障措施；c）应对共享数据分级分类的变更进行记录，并通知相关数据使用方；d）应按照数据级别明确使用方对共享数据的使用权限。

E1.2 地方数据分类分级相关标准规范

截至目前，全国各地已有 16 个省市区制定有地域特色的政务数据分类分级相关规范要求，具体如下所示（按照发布的先后顺序排列）：

- 《DB52/T1123—2016 政府数据数据分类分级指南》，实施日期 2016 年 9 月 28 日
- 《上海市公共数据开放分级分类指南（试行）》，发布日期：2019 年 11 月 1 日
- 《坪山区政务数据分级分类管理办法（试行）》，发布日期：2020 年 10 月 20 日
- 《DB3301/T0322.3—2020 数据资源管理第 3 部分：政务数据分类分级》，实施日期：2020 年 11 月 30 日
- 《烟台市公共数据开放分级分类指南（试行）》，发布日期：2021 年 3 月 31 日
- 《DB33/T2351—2021 数字化改革公共数据分类分级指南》，实施日期：2021 年 8 月 5 日
- 《重庆市公共数据分类分级指南（试行）》，发布日期：2021 年 10 月 11 日
- 《DB2201/T17—2022 政务数据安全分类分级指南》，实施日期：2022 年 1 月 30 日
- 《DB14/T2442—2022 政务数据分类分级要求》，实施日期：2022 年 6 月 30 日
- 《DB3212/T1116—2022 政务数据安全分类分级指南》，实施日期：2022 年 12 月 28 日
- 《DB3202/T1049—2023 无锡市公共数据分类分级实施指南》,实施日期:2023 年 5 月 21 日
- 《DB3203/T1024—2023 公共数据分类分级指南》，实施日期：2023 年 5 月 30 日
- 《DB51/T3056—2023 政务数据数据分类分级指南》，实施日期：2023 年 6 月 1 日
- 《DB4201/T677.2—2023 公共数据资源开放第 2 部分：分类分级指南》，实施日期：2023 年 7 月 9 日
- 《DB23/T3510—2023 政务预公开数据分类分级评估指南》，实施日期：2023 年 8 月 4 日

E2 金融

自《中华人民共和国网络安全法》《中华人民共和国数据安全法》等数据保护相关法律、行政法规施行以来，我国数据监管框架已初具雏形。由于金融行业业务领域数据规模大、价值和敏感程度高的特点，维护相关金融数据的稳定性和可用性有助于维护个人、企业、金融行业乃至国家利益的稳定。同时，随着实践中频发的银行业等金融领域的数据安全事件，金融数据的保护需求愈发迫切，金融数据安全保护的首要任务是形成有金融行业属性的数据分类分级规范。近年来，金融行业陆续出台数据安全相关制度规范中均有规范数据分类分级的要求，具体如下所示：

E2.1 中国人民银行业务领域数据安全管理办法（征求意见稿）要求

E2.1.1 数据分类分级保护总体规划

中国人民银行负责组织制定数据分类分级相关行业标准，指导数据处理者开展数据分类分级各项工作，统筹确定重要数据具体目录并实施动态管理。

E2.1.2 数据分类分级制度规程

数据处理者应当建立健全本单位数据分类分级实施制度，规范分类分级工作操作规程。数据分类分级过程实施和结果审批，应当严格遵循操作规程。

E2.1.3 数据分类要求

数据处理者应当参考行业标准，根据业务开展情况建立业务分类，梳理细化数据资源目录，标识各数据项是否为个人信息、数据来源（生产经营加工产生、外部收集产生等）、存储该数据项的信息系统清单和应用的业务类别。

E2.1.4 数据分级要求

数据按照精度、规模和对国家安全的影响程度，分为一般、重要、核心三级。在中国人民银行组织下，数据处理者应当准确识别判定本单位信息系统存储的全量数据是否属于重要数据、核心数据，并填写报送重要数据目录内容，由中国人民银行汇总后确定重要数据具体目录。数据处理活动中，数据处理者还应当及时准确识别判定所涉及数据是否属于重要数据、核心数据。

E2.1.5 数据敏感性分层级

在数据分级基础上，数据处理者应当参考行业标准，根据数据遭到泄露或者被非法获取、非法利用时，可能对个人、组织合法权益或者公共利益等造成的危害程度，将数据项的敏感性从低至高进一步分为一至五共五个层级。结构化数据项应当逐一标识层级；非结构化数据项应当优先按照可拆分的各结构化数据项所对应最高层级，标识其层级。

E2.1.6 数据可用性分层级

数据可用性分层级工作纳入信息系统业务连续性分级保障体系统一考虑。数据处理者应当评估信息系统存储数据遭到篡改、破坏后可能对业务连续性造成的影响程度，明确恢复的目标要求。恢复点目标越严格，数据的可用性层级越高。在此基础上，鼓励数据处理者识别用于支撑最基本业务运转、无法承受彻底灭失风险、需要进一步进行容灾备份的数据。

E2.1.7 动态更新要求

数据处理者应当根据数据和信息系统变化情况，每年组织更新数据资源目录，避免信息系统所涉及数据项未在数据资源目录中记录、数据项标识信息不完整等情形发生。

E2.2 JR/T0197-2020《金融数据安全 数据安全分级指南要求》

金融业机构典型数据的定级规则在实际应用过程中，各金融业机构宜根据其所管辖数据的类型、特性、规模以及机构特性等因素，综合考虑本机构数据安全管理的总体目标和安全策略要求，按照一定的颗粒度对数据资产进行合理的梳理、归类和细分，最终确定数据的安全级别划分清单。此外，金融业机构所承载重要数据的安全级别宜不低于本标准确定的 5 级。重要数据的识别、认定及保护工作依据国家及行业主管部门有关规定和要求执行。

E2.2.1 定级通用规则

金融数据安全级别划分的通用规则包括但不限于；“重要数据”的安全等级不可低于本标准所述 5 级。

——个人金融信息相关数据参照 JR/T0171—2020 进行定级，并在数据安全定级过程中从高考虑。

——对于数据体量大，涉及的客户（包含个人客户和单位客户）多、涉及客户（包含个人客户和单位客户）资金量大、涉及多行业及多机构客户的情况，影响程度宜从高确定。

E2.2.2 定级规则

本标准根据金融业机构数据安全性遭受破坏后的影响对象和所造成的影响程度，将数据安全级别从高到低划分为 5 级、4 级、3 级、2 级、1 级，一般具有如下特征：

——5 级数据特征如下：

- 重要数据，通常主要用于金融业大型或特大型机构、金融交易过程中重要核心节点类机构的关键业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。
- 数据安全性遭到破坏后，对国家安全造成影响，或对公众权益造成严重影响。

注：“必须知悉”是指对数据确定知悉范围，只有对数据知悉有明确的必要性时，该对象才能对数据知悉。一般情况下遵循工作需要原则和最小化原则，前者指因工作必须才可知悉，后者指知悉的范围满足最小够用即可。

——4 级数据特征如下：

- 数据通常主要用于金融业大型或特大型机构、金融交易过程中重要核心节点类机构的重要业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。
- 个人金融信息中的 C3 类信息。
- 数据安全性遭到破坏后，对公众权益造成一般影响，或对个人隐私或企业合法权益造成严重影响，但不影响国家安全。

——3 级数据特征如下：

数据用于金融业机构关键或重要业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。

- 个人金融信息中的 C2 类信息。
- 数据的安全性遭到破坏后，对公众权益造成轻微影响，或对个人隐私或企业合法权益造成一般影响，但不影响国家安全。

——2 级数据特征如下：

- 数据用于金融业机构一般业务使用，一般针对受限对象公开，通常为内部管理且不宜广泛公开的数据。
- 个人金融信息中的 C1 类信息。
- 数据的安全性遭到破坏后，对个人隐私或企业合法权益造成轻微影响，但不影响国家安全、公众权益。

——1 级数据特征如下：

- 数据一般可被公开或可被公众获知、使用。
- 个人金融信息主体主动公开的信息。
- 数据的安全性遭到破坏后，可能对个人隐私或企业合法权益不造成影响，或仅造成微弱影响但不影响国家安全、公众权益。

E2.3 GB/T42775-2023《证券期货业数据安全风险防控数据分类分级指引》

证券期货业业务种类繁多，数据呈现出复杂性高、多样性强的特点。采用规范的数据分类、分级方法，有助于行业机构厘清数据、确定数据重要性或敏感度，并针对性地采取适当、合理的管理措施和安全防护措施，形成一套科学、规范的数据管理与保护机制，从而在保证数据安全的基础上促进数据开放共享。数据分类是数据保护工作中的一个关键部分，是建立统一、准确、完善的数据架构的基础，是实现集中化、专业化、标准化数据管理的基础。行业机构按照统一的数据分类方法，依据自身业务特点对产生采集、加工、使用或管理的数据进行分类，有利于清晰地厘清数据资产，对数据实现规范化管理，为数据的维护和扩充提供支持。数据分级是以数据分类为基础，采用规范、明确的方法区分数据的重要性和敏感度差异，并确定数据级别。数据分级有助于行业机构根据数据不同级别，确定数据在其生命周期的各个环节应采取的数据安全防护策略和管控措施，进而提高行业机构的数据管理和安全防护水平，确保数据的完整性、保密性和可用性。根据数据影响对象及数据影响程度可将证券期货业数据分为 4 级，分级方式如下：

表 14 证券期货业数据分类

影响范围	影响对象	影响程度	数据一般特征	数据重要程度标识	数据级别标识
多个子行业	子行业	严重	数据主要用于大型或特大型行业机构中的重要业务,一般针对特定人员公开,且仅为必要知悉的对象访问或使用。	极高	4
子行业内多行业机构	行业机构	严重		极高	4
子行业内多行业机构	客户	严重		极高	4
单一行业机构	行业机构	严重	数据用于重要业务,针对特定人员公开,且仅为必要知悉的对象访问或使用。	高	3
单一行业机构	客户	严重		高	3
单一行业机构	行业机构	中等,轻微	数据用于一般业务,针对受限对象公开;一般指内部管理、办公类且不宜广泛公开的数据。	中	2
单一行业机构	客户	中等		中	2
单一行业机构	行业机构	轻微、无	数据可被公开或被公众获知、使用。	低	1
单一行业机构	客户	轻微、无		低	1

E3 电信

E3.1 YD/T3813-2020《基础电信企业数据分类分级方法》

2020 年 12 月，中国通信标准化协会发布了 YD/T3813-2020《基础电信企业数据分类分级方法》。该标准规定了基础电信企业数据分类分级原则、数据分类分级工作流程和数据分类分级方法，并给出基础电信企业数据分类分级示例。

该标准根据基础电信企业生产经营管理现状和企业自身管理特点，将数据的类别分为两大类：用户相关数据、企业自身相关数据；根据基础电信企业数据重要程度以及泄露后对国家安全、社会秩序、企业经营管理和公众利益造成的影响和危害程度，将数据的级别分为四个级别，由高到低分别为：第四级数据、第三级数据、第二级数据、第一级数据。

E3.1.1 数据分类

用户相关数据

表 15 用户相关数据分类表

类别	大类	子类	范围	对应数据
用户相关数据	用户身份相关数据	用户身份相关数据	自然人身份标识	（客户、员工）姓名、证件类型、证件号码、驾照编号、银行账户、客户实体编号、集团客户编号、集团客户名称等
			网络身份标识	联系电话、手机号、座机号码、邮箱、网络客户编号、即时通信账号、网络社交用户账号、微信号、QQ 号等
			用户基本资料	（客户）职业、工作单位、年龄、性别、籍贯、兴趣爱好等；集团客户所在省市、所在行业等
			实体身份证明	身份证影印件、护照影印件、驾照影印件、营业执照影印件等；指纹、声纹、虹膜等
			用户私密资料	个人种族、家属信息、居住地址、宗教信仰、健康状况等
		用户网络身份鉴权信息	用户密码及关联信息	密码、手机客服密码，密码保护答案等
	用户服务内容数据	服务内容和资料数据	服务内容数据	短信、彩信、话音等
				即时通信内容、群内发布内容、数据文件、邮件内容、用户上网访问内容等
			联系人信息	通讯录、好友列表、群组列表等
	用户服务衍生数据	用户服务使用数据	业务订购关系	品牌、套餐定制等
				邮箱、通讯录等增值业务的注册、修改、注销等
			服务记录和日志	语音详单、短信详单、彩信详单等
				Cookie、上网日志等
			消费信息和账单	停开机、入网时间、在网时间、积分、预存款、信用等级等
				账单、固定费用、通信费用等
			位置数据	小区代码、基站号、基站经纬度坐标等
				地区代码、城市代码等
			违规记录数据	黑名单、灰名单等
		设备信息	终端设备标识	唯一设备识别码、IMEI、MAC 地址、SIM 卡、IMSI 信息等
			终端设备资料	终端型号、品牌、厂商等
	用户统计分析类数据	用户使用习惯和行为	--	用户偏好、消费习惯，通话、短信频次、上网等数量与频次等。

		分析数据		
		用户上网行为相关统计分析数据	--	用户网络行为、用户画像等

企业自身相关数据

表 16 企业自身相关数据分类表

类别	大类	子类	范围	对应数据
企业自身相关数据	网络与系统的建设与运行维护类数据	网络与系统资源类数据	公共资源类数据	DDM（数字诊断监视功能模块）、DDF（数字配线架）、ODM（光配线架连接模块）、ODF（光纤配线架）
			承载网资源	板卡、物理端口、逻辑端口、物理链路、逻辑链路、网段、IP 地址
			云资源	资源池、虚拟机 VM、存储设备、负载均衡
			设备监测、告警	设备监测、告警
			运维日志	事件、地点、时间、操作、成功与否
			后台管理账号信息	后台管理人员角色 ID、创建修改时间、账号状态
			角色信息	人员角色、编辑时间
	业务运营类数据	业务运营服务数据	产品信息	产品 ID、套餐设置、销售品 ID

E3.1.2 数据分级

各级别数据特征

- 第四级数据

一旦丢失、泄露、被篡改、被损毁会对国家安全、社会公共利益或企业利益或用户利益造成特别严重影响的数据，安全管控要求最高。

- 第三级数据

一旦丢失、泄露、被篡改、被损毁会对国家安全、社会公共利益或企业利益或用户利益造成严重影响的数据，应实施较强的安全管控。

- 第二级数据

一旦丢失、泄露、被篡改、被损毁会对国家安全、社会公共利益或企业利益或用户

利益造成一定程度影响的数据，执行基本的安全管控。

- 第一级数据

一旦丢失、泄露、被篡改、被损毁对国家安全、社会公共利益或企业利益或用户利益造成影响较小或无影响的数据，对安全管控不作要求。

数据分级示例

- 第四级数据

实体身份证明、用户私密资料、用户密码及关联信息、联系人信息、规划建设类（发布前）、网络与系统资源类、网络与系统运维类、网络安全管理类。

- 第三级数据

自然人身份标识、详单、位置数据、用户使用习惯分析数据、用户上网行为相关统计分析数据、用户使用行为分析数据、用户上网日志信息、企业发展战略、业务发展、技术研发类、统计分析类数据、招投标数据（公开前）。

- 第二级数据

网络身份标识、用户基本资料、服务内容数据、服务记录和日志、设备信息、业务订购关系、消费信息、账单、规划建设类（发布后）、渠道信息、客服数据、营销信息、招投标数据（公开后）、物资数据、业务合作类数据、合作方提供数据。

- 第一级数据

违规记录数据、产品信息、公开业务运营服务数据。

E3.2 YD/T4244-2023《电信网和互联网数据分类分级技术要求与测试方法》

本文件规定了电信网和互联网数据脱敏的技术要求与测试方法。 本文件适用于电信网和互联网数据的脱敏工作，脱敏技术能力的设计、研发、测试、评估和验收等，包括数据脱敏的提供商、用户、测评机构和监管机构等。

E4 医疗

国家标准化管理委员会于 2020 年 12 月 14 日正式发布了 GB/T39725-2020《信息

安全技术健康医疗数据安全指南》（以下简称“《健康医疗数据安全指南》”），同时，《健康医疗数据安全指南》于 2021 年 7 月 1 日正式实施。《健康医疗数据安全指南》由信安标委负责提出并归口牵头编写，健康医疗数据生命周期内可能涉及的各方主体都参与起草、审核。此指南作为推荐性国家标准，其并不具有强制法律效力，并不需要完全遵从。各健康医疗企事业单位在数据合规过程中，可将《健康医疗数据安全指南》作为参考，与此同时，仍需结合当前国家发布《中华人民共和国网络安全法》《个人信息安全规范》等法规及标准以保证安全的全面性。

E4.1 GB/T39725-2020《信息安全技术健康医疗数据安全指南》

E4.1.1 概览解读

《健康医疗数据安全指南》该指南全文共 11 个章节，包括健康医疗数据的安全目标、分类体系、使用披露要点、安全措施要点、安全管理指南、安全技术指南、典型场景数据安全等。

《健康医疗数据安全指南》主要适用于健康医疗数据控制者，即“能够决定健康医疗数据处理的目的、方式及范围等的组织或个人”。如果存在需要两方或多方共同决定数据使用处理的目、方式及范围，则视为共同控制者。同时，常见的处理者包括：健康医疗信息系统供应商、健康医疗数据分析公司、辅助诊疗解决方案供应商等。

《健康医疗数据安全指南》针对“健康医疗数据”进行了明确定义，即为“个人健康医疗数据以及由个人健康医疗数据加工处理之后得到的健康医疗相关电子数据”。例如：经过对群体健康医疗数据处理后得到的群体总体分析结果、趋势预测、疾病防治统计数据等。

《健康医疗数据安全指南》针对“个人健康医疗数据”进行了明确定义，即为“单独或者与其他信息结合后能够识别特定自然人或者反映特定自然人生理或心理健康的相关电子数据”。示例如：向个人提供医疗服务过程中采集的有关个人的既往病史、社会史、家族史等病历及数据记载，可穿戴设备采集的与个人健康相关的数据，关于个人的支付或医保数据等。

E4.1.2 健康医疗数据分类和分级

《健康医疗数据安全指南》结合数据本身的性质及其所反映的内容，将健康医疗数

据分为 6 个类别，并且在每个类别下给出了具体范围，对于医疗机构、企事业单位进行数据分类提供了参考。具体如下图所示：

表 17 健康医疗数据分类和分级表

数据类别	范围
个人属性数据	人口统计信息,包括姓名、出生日期、性别、民族、国籍、职业、住址、工作单位、家庭成员信息、联系人信息、收入、婚姻状态等； 个人身份信息,包括姓名、身份证、工作证、居住证、社保卡、可识别个人的影像图像、健康卡号、住院号、各类检查检验相关单号等； 个人通讯信息,包括个人电话号码、邮箱、账号及关联信息等； 个人生物识别信息,包括基因、指纹、声纹、掌纹、耳廓、虹膜、面部特征等； 个人健康监测传感设备 ID 等
健康状况数据	主诉、现病史、既往病史、体格检查(体征)、家族史、症状、检验检查数据、遗传咨询数据、可穿戴设备采集的健康相关数据、生活方式、基因测序、转录产物测序、蛋白质分析测定、代谢小分子检测、人体微生物检测等。
医疗应用数据	门(急)诊病历、住院医嘱、检查检验报告、用药信息、病程记录、手术记录、麻醉记录、输血记录、护理记录、入院记录、出院小结、转诊(院)记录、知情告知信息等。
医疗支付数据	医疗交易信息,包括医保支付信息、交易金额、交易记录等； 保险信息,包括保险状态、保险金额等。
卫生资源数据	医院基本数据、医院运营数据等。
公共卫生数据	环境卫生数据、传染病疫情数据、疾病监测数据、疾病预防数据、出生死亡数据等。

《健康医疗数据安全指南》结合数据重要程度、风险级别以及对个人健康医疗数据主体可能造成的损害和影响的级别进行分级，该指南将健康医疗数据划分为 5 个级别：

- 第 1 级：可完全公开使用的数据。包括可以通过公开途径获取的数据，例如医院名称、地址、电话等，这类数据可直接在互联网上面向公众公开。
- 第 2 级：可在较大范围内供访问使用的数据。例如不能标识个人身份的数据，各科室医生经过申请审批可以用于研究分析。
- 第 3 级：可在中等范围内供访问使用的数据。如果未经授权披露，可能对个人健康医疗数据主体造成中等程度的损害。例如经过部分去标识化处理，但仍可能重标识的数据，仅限于获得授权的项目组范围内使用。
- 第 4 级：在较小范围内供访问使用的数据。如果未经授权披露，可能会对个人健康医疗数据主体造成较高程度的损害。例如可以直接标识/定位到个人身份的

数据，仅限于参与诊疗活动的医护人员访问使用。

- 第 5 级：仅在极小范围内且在严格限制条件下供访问使用的数据。如果未经授权披露，可能会对个人健康医疗数据主体造成严重程度的损害。例如特殊病种（例如艾滋病、性病）的详细资料，仅限于主治医护人员访问且需要进行严格管控。

E4.1.3 健康医疗数据的使用披露原则

数据使用授权的要求

数据控制者使用或披露个人健康医疗数据时，应获得个人授权。使用的过程中，如果超出个人授权范围的，还需要再次征得主体同意。获取授权时，要明确授权使用的数据内容，数据使用用途，数据的使用期限，以及控制者的保护措施等。

获得使用授权的例外情况

在授权同意方面，存在以下几种不获得主体授权的例外情形：“向主体提供其本人健康医疗数据”“治疗、支付或保健护理时”“涉及公共利益或法律法规要求时”“受限制数据集用于科学研究、医学/健康教育、公共卫生目的时”。但是，数据控制者仍然需要谨慎地评估适用例外的场景。

回溯查询的主体权利

主体有权对数据控制者或其处理者使用或披露数据的情况进行历史回溯查询，最短回溯期为六年。最短回溯期的时间也不完全等同于数据存储的时间。根据《电子病历应用管理规范（试行）》规定门（急）诊电子病历保存时间自患者最后一次就诊之日起不少于 15 年，住院电子病历保存时间自患者最后一次出院之日起不少去标识化建议

《健康医疗数据安全指南》规定，根据使用目的尽可能地去标识化。因此，除了建议按照已发布的《个人信息安全规范》，进行相应的去标识化外，对于应用于临床研究和医药医疗研发的数据，同样要求去标识化，不得识别或定位到个人。

E4.2 WST306-2023《卫生健康信息数据集分类和编码规则》

此标准规定了卫生健康信息数据集分类与编码需遵循的基本原则、技术方法以及应用规则。

E5 教育

E5.1 T/GDCSA 000-2022《高校数据分类分级指南》

本文件规定了高校数据分类的一般要求、维度与流程，数据分级的一般要求、要素、维度、流程与变更。高校数据分类维度可从部门维度、人员维度、资产维度、资产维度、资产维度开展数据分类。同时根据数据影响对象及影响程度可将数据分为核心、重要、一般。

E6 工业

E6.1 工业数据分类分级指南（试行）

工业数据分类维度包括但不限于研发数据域（研发设计数据、开发测试数据等）、生产数据域（控制信息、工况状态、工艺参数、系统日志等）、运维数据域（物流数据、产品售后服务数据等）、管理数据域（系统设备资产信息、客户与产品信息、产品供应链数据、业务统计数据等）、外部数据域（与其他主体共享的数据等）。工业数据分级分为一级、二级、三级等 3 个级别。

E6.2 GB/T 42128-2022《智能制造工业数据分类原则》

本文件给出了智能制造工业数据的分类要求、分类依据以及分类维度。

本文从分类维度可以分为按系统层级分类、按生命周期分类、按智能特征分类。

E7 国际标准

E7.1《DATA CLASSIFICATION PRACTICES》

此标准是美国国家网络安全卓越中心（NCCoE）作为零信任保护的一部分项目，本文侧重于通过数据分类和标签来传达和维护数据保护要求，以金融、政务、制造业、科技行业、医疗等五个场景来开展数据分类工作，实现数据发现、入库、分析、分类、标记等内容的整体方案。

附录 F-典型行业数据分类分级词典示例

F1 政务

F1.1 政务数据分类示例

（GB/T 21063.4 《政务信息资源目录体系第 4 部分：政务信息资源分类》）

表 18 政务数据分类表

名　称	描　述　说　明
综合政务	关于政治方面的事务和国家的管理工作
政务综合类	
方针政策	政府制订的、宏观的、指导各个领域发展的方针政策
中共党务	关于中国共产党的规章制度、组织机构建设和发展，以及工作职责等相关信息
政府工作	关于政府的规章制度、组织机构建设和发展，以及工作职责等相关信息
人大	关于人民代表大会的规章制度、组织机构建设和发展，以及工作职责的相关信息
政协	政治协商会议的规章制度、组织机构建设和发展，以及工作职责等相关信息
法院	法院系统的规章制度、组织机构建设和发展，以及工作职责等相关信息
检察院	检察院系统的规章制度、组织机构建设和发展，以及工作职责等相关信息
机构编制	关于机构编制的管理、机构体系的当前概况和远景规划
领导人	关于领导人的简历、工作岗位、工作活动、作品等相关信息
会议、会务	会议产生的报告等相关信息，以及会议组织、管理的相关信息
重大事件	有深远影响的事件的相关信息
经济管理	关于经济的管理、规划、发展概况
经济管理综合类	
经济发展计划	关于经济的宏观的发展规划
经济管理	关于经济的宏观管理现状
经济体制改革	关于经济体制改革的管理和规划、发展情况
经贸管理	关于经济贸易的宏观管理和发展调查报告、统计资料
统计	关于统计工作的管理和发展情况
物价	关于物价的管理和调查报告、统计资料，以及物价体系规划
工商	关于市场监督管理和维护公平竞争的市场秩序
国土资源、能源	土地、水、海洋、石油、天然气、矿藏、电力资源的管理、规划和发展
国土资源与能源综合类	
土地	关于土地资源的管理、规划和开发利用情况
矿藏	关于矿产资源的管理、　规划和开发利用情况
水资源	关于水资源的管理、规划和开发利用情况
海洋	关于海洋资源的管理、规划和开发利用情况

煤炭	关于煤炭资源的管理、规划和开发利用情况
石油	关于石油、天然气行业生产、经营的规章制度、法律法规、规划和管理
燃料、燃气	关于燃料、燃气行业生产、经营的规章制度、法律法规、规划和管理
电力	关于电力生产和市场经营的规章制度、法律法规、管理机构，以及设施建设和技术发展
工业、交通	工业、企业、交通运输、邮政及相关领域
工交综合类	
工业	关于工业的规章制度、法律法规、管理机构、体系结构规划和发展
企业	关于企业创建、经营的规章制度、法律法规、管理机构和政府对企业提供的相关服务
交通运输	关于交通运输业的规章制度、法律法规、管理机构、功能服务和设施建设的规划和发展情况
信息产业	与信息产业内以及相关的行业领域
信息产业综合类	
通信	与通信相关的规章制度、法律法规、设施建设、功能服务以及机构组织等概况
计算机	与计算机相关的研发、生产、销售相关的规章制度、法律法规、技术和标准，以及该行业的市场行情等概况
软件	与软件业相关的规章制度、法律法规、技术标准和发展概况
网络	关于网络的规章制度、法律法规、设施建设规划和发展情况，以及网络的功能服务
信息技术、信息系统	关于信息系统建设的规划和发展情况，信息系统功能服务的开发利用，以及相关的规章制度、法律法规，信息技术的研发
邮政	关于邮政业的规章制度、法律法规、设施建设的规划和发展情况、功能服务的开发利用
环境污染、监测	关于环境监测的统计报告和技术标准
农业、水利	关于农业、畜牧业、林业、副业、渔业、水利的规章制度、法律法规、发展规划、统计资料、科学知识、组织机构等
农业水利综合类	
农业	关于农业的规章制度、法律法规、农产品市场、农业组织、农业科技、农业基础设施建设
林业	关于林业的规划、建设情况、规章制度、法律法规、科学技术、组织机构和相关领域经济情况
畜牧业	关于畜牧业的规章制度、法律法规、设施建设、科学技术和相关领域经济发展情况
副业	关于副业发展的规章制度、法律法规、设施建设、科学技术和相关领域的经济发展情况
渔业	关于渔业发展的规章制度、法律法规、科学技术、设施建设和相关领域的经济发展

水利	关于水利建设的规划和发展、规章制度、法律法规、科学技术和组织机构
财政	财政、会计、金融、保险、税务、审计等
财政综合类	
财政	关于财政的规章制度、法律法规、业务管理、统计资料
税务	关于税务的规章制度、法律法规、理论知识、税务体系建设
金融	关于金融的规章制度、法律法规、理论知识、组织机构、金融经济统计资料
保险	关于保险的规章制度、法律法规、理论知识、组织机构、保险经济统计资料
审计	关于审计的规章制度、法律法规、理论知识、组织机构
会计	关于会计的规章制度、法律法规、组织机构、理论知识和行业发展情况
商业、贸易	商业指以货币为媒介进行交换从而实现商品流通的经济活动，贸易指买卖的通称
商贸综合类	
国内贸易	在国内进行的买卖
对外贸易	在国际间进行的买卖
物流、仓储	物资流动、仓库储备
海关	交纳关税和货船报关结关的地方
检验、检疫	检查并验证或为防止传染病蔓延，对可能成为传染源的人员、交通工具、物资等采取隔离观察、消毒等措施
旅游、服务业	外出游览或为他人提供服侍、帮助的行业
旅游、服务业综合类	
旅游	外出游览、观光
服务业	为他人提供服侍这一职业
气象、水文、测绘、地震	大气现象，自然界中水的变化、运动等的各种现象，测量和地图制图，地壳的天然振动
气象水文综合类	
气象	大气现象
水文	自然界中水的变化、运动等的各种现象
测绘	测量和地图制图
地震	地壳的天然震动
对外事务	一个国家在国际关系方面的活动的行政事务、总务
对外事务综合类	
外交	一个国家在国际关系方面的活动
外事活动	外交事务、涉外事务方面的活动
国际关系	一个国家与其他各国及其公民之间的关系
国际组织	各国及其公民之间按照一定的目的、任务和形式组成的国际性团体机构
国际会议	各国之间共同议事，并遵循一定的议程，所举行的一种集会
政法、监察	政治和法律，监督和考察

政法综合类	
公安	社会的公共治安
国家安全	国家的独立、主权、统一和领土完整不受侵犯，国家的政权、政治制度、经济制度和社会秩序不被破坏，国家的荣誉和利益不遭受到损害或不受损害
监察	监督考察
司法	行使法律权力
科技、教育	科技就是科学技术，教育是指培养人才、传播知识的工作
科教综合类	
科技管理	对科学技术研究工作进行管理
科研工作	科学研究工作
知识产权	知识产权是指公民、法人或其他组织对其在科学技术和文学艺术等领域内，主要基于脑力劳动创造完成的智力成果所依法享有的专有权利
技术监督	对劳动生产方面的经验、知识和技巧有权进行检查管理
教育	培养人才、传播知识的工作，主要指学校教育
院校管理	对各类学院、学校进行管理
文化、卫生、体育	
文体综合类	
语言文字	语言是用以表达情意的声音，是人类最重要的交际工具，它跟思想有密切关系，是人类区别于其他动物的本质特征；文字是记录语言的符号，如汉字、拉丁字母
文学艺术	
文物、考古	文物是指历代遗留下来的具有历史、艺术价值的东西，考古是指根据古代的遗迹、遗物和文献，研究古代事物
新闻出版	图书、期刊、音像制品、电子出版物、网络出版物、投影片(含缩微制品)等出版业务
广播、电影、电视	广播是指利用无线电或电视信号对大众传播； 电影是指表达一个完整主题的、具有连贯的、活动感觉的影像； 电视是指利用电子设备传送活动图像的技术，是重要的广播和通信方式，即电视接收机，很多时候也指电视节目
医药卫生管理	对药品生产、流通、医疗器械企业进行管理
医疗保健	治疗人体疾病，保护和增进人体健康
计划生育	通过有效地控制生育的方法来执行制订子女人数和生育间隔时间
体育	一种娱乐消遣活动或需体力、智慧与技巧的比赛或竞技，它要求用或多或少的体力，按照传统的形式或一组规则进行，有时还作为一种职业在户外或室内进行
军事、国防	军事是指与军队或战争有关的事情，国防是指保卫国家的主权，领土、领海和领空的完整和安全，防御外来的武装侵略和颠覆
军事国防综合类	
国防建设	指为保卫国家的主权，领土、领海和领空的完整和安全，防御外来的武装侵略和颠覆而进行的建设活动

军事工作	与军队或战争有关事宜的工作
军队政治工作	军队政治工作就是要保证革命政党对军队的领导，使军队成为革命政党实现革命目的的工具
军事后勤工作	与军队或战争有关的后方对前方的一切供应工作
军事装备工作	与军队或战争有关的配备工作
军事技术	与军队或战争有关的经验、知识和技巧
武警	武装警察
劳动、人事	劳动特指劳动就业和社会保障，人事是指关于工作人员的录用、培养、调配、奖罚等工作
劳动人事综合类	
人事工作	人事是指关于工作人员的录用、培养、调配、奖罚等工作
劳动就业	主要指就业的基本政策、措施、方针、标准、劳动关系调整的基本规则
社会保障	主要指社会保险
工资职称	付给劳动者劳动报酬和科学技术人员等级
福利待遇	生活上的利益和享有的权利及物质报酬
民政、社区	民政指政府处理的有关人民的行政事务，社区指同一地区人所组成的整体
民政综合类	
民政	政府处理的有关人民的行政事务
社区	同一地区人所组成的整体
文秘、行政	文秘指处理文书文件类秘书 行政指企事业、各种社会团体等的内部管理
文秘行政综合类	
文秘工作	处理文书、文件类的工作
文种	能概括地表明每一种公文性质、用途的统一规范称谓
机要、保密	重要而机密、保守秘密
档案	国家机构、社会组织以及个人从事政治、军事、经济、科学、技术、文化、宗教等活动直接形成的对国家和社会有保存价值的各种文字、图表、声像等不同形式的历史记录
信访工作	接受和处理群众通过信函或面谈方式反映问题的工作
行政事务	政府机关、企事业、各种社会团体等的内部执行管理的杂务
综合党团	宗教、民族、侨民、党派团体、特别行政区的综合事务
党团综合类	
党派团体	关于各党派团体的概况
民族事务	关于少数民族事业的发展概况
宗教	关于宗教事务的管理和发展情况
侨务工作	关于侨务工作的管理和发展概况，以及侨民概况
港澳台工作	关于香港、澳门、台湾的政策法规、政治、经济、科技交流往来等概况

F1.2 政务数据分级示例

（DB11/T 1918-2021《政务数据分级与安全保护规范》）

表 19 政务数据分级表

分类	数据项	影响对象	影响程度	影响规模	可控程度	数据级别
办证申请信息	申请用户姓名	自然人	一般影响	较大范围	强可控	二级
	电话	自然人	一般影响	较大范围	强可控	二级
	办证类型	自然人	一般影响	较大范围	强可控	二级
	小区名	自然人	严重影响	较大范围	强可控	三级
	楼号	自然人	严重影响	较大范围	强可控	三级
	家庭住址	自然人	严重影响	较大范围	强可控	三级
	车牌	自然人	严重影响	较大范围	强可控	三级
	旧车牌	自然人	严重影响	较大范围	强可控	三级
	车牌颜色	自然人	一般影响	较大范围	强可控	二级
	牌照类型	自然人	一般影响	较大范围	强可控	二级
	车主与户主关系	自然人	严重影响	较大范围	强可控	三级
	费用	自然人	一般影响	较大范围	强可控	二级
	操作员	自然人	一般影响	较大范围	强可控	二级
	身份证件编号	自然人	严重影响	较大范围	强可控	三级
	身份证姓名	自然人	严重影响	较大范围	强可控	三级
	房屋证明编号	自然人	严重影响	较大范围	强可控	三级
	房产证明姓名	自然人	严重影响	较大范围	强可控	三级
	户口本地址	自然人	严重影响	较大范围	强可控	三级
	户口本姓名	自然人	严重影响	较大范围	强可控	三级
	驾驶证编号	自然人	严重影响	较大范围	强可控	三级
	驾驶证姓名	自然人	严重影响	较大范围	强可控	三级
	行驶证编号	自然人	严重影响	较大范围	强可控	三级
	行驶证姓名	自然人	严重影响	较大范围	强可控	三级
	结婚证姓名	自然人	严重影响	较大范围	强可控	三级
	结婚证编号	自然人	严重影响	较大范围	强可控	三级
	租赁合同姓名	自然人	严重影响	较大范围	强可控	三级
	租赁合同编号	自然人	严重影响	较大范围	强可控	三级
	居委会	公共服务机构	一般影响	较大范围	强可控	二级
	办证申请信息数据项集合	自然人	特别严重影响	较大范围	强可控	四级
办证规则信息	车场名称	公共服务机构	一般影响	较大范围	强可控	二级
	免费时长	自然人	一般影响	较大范围	强可控	二级
	基础分钟数	自然人	一般影响	较大范围	强可控	二级
	基础分钟金额	自然人	一般影响	较大范围	强可控	二级
	累计金额	自然人	一般影响	较大范围	强可控	二级

	每日封顶	自然人	一般影响	较大范围	强可控	二级
	包月金额	自然人	一般影响	较大范围	强可控	二级
	更新时间	自然人	一般影响	较大范围	强可控	二级
	创建时间	自然人	一般影响	较大范围	强可控	二级
办证规则信息	操作员 id	自然人	一般影响	较大范围	强可控	二级
	证件类型	自然人	一般影响	较大范围	强可控	二级
	办证规则信息数据项集合	自然人	一般影响	较大范围	强可控	二级
停车场信息	车场编号	公共服务机构	一般影响	较大范围	强可控	二级
	车场名称	公共服务机构	一般影响	较大范围	强可控	二级
	居委会编号	公共服务机构	一般影响	较大范围	强可控	二级
	更新时间	自然人	一般影响	较大范围	强可控	二级
	创建时间	自然人	一般影响	较大范围	强可控	二级
	停车场信息数据项集合	公共服务机构	一般影响	较大范围	强可控	二级
站内通知信息	标题	自然人	一般影响	较大范围	强可控	二级
	类型	自然人	一般影响	较大范围	强可控	二级
	发布时间	自然人	一般影响	较大范围	强可控	二级
	内容	自然人	一般影响	较大范围	强可控	二级
	创建人	自然人	一般影响	较大范围	强可控	二级
	创建时间	自然人	一般影响	较大范围	强可控	二级
	修改人	自然人	一般影响	较大范围	强可控	二级
	修改时间	自然人	一般影响	较大范围	强可控	二级
	站内通知信息数据项集合	自然人	一般影响	较大范围	强可控	二级
公示登记信息	项目名称	自然人	一般影响	较大范围	强可控	二级
	公示图像文件	自然人	一般影响	较大范围	强可控	二级
	操作员编号	自然人	一般影响	较大范围	强可控	二级
	删除状态	自然人	一般影响	较大范围	强可控	二级
	更新时间	自然人	一般影响	较大范围	强可控	二级
	创建时间	自然人	一般影响	较大范围	强可控	二级
	公示状态	自然人	一般影响	较大范围	强可控	二级
公示登记信息	公示登记信息数据项集合	自然人	一般影响	较大范围	强可控	二级

F2 金融

F2.1 金融数据分类分级

F2.1.1 金融数据分类示例

表 20 金融数据分类示例表

分类	子分类	描述
客户数据	个人数据	指个人的自然属性信息，包括个人自然信息、个人身份鉴别信息、个人资讯信息等信息。
	单位数据	指单位的自然属性数据，包含单位基本信息、单位身份鉴别信息、单位标签信息等单位信息。
业务数据	账户信息	指账户相关数据，如账户的基本信息、计息信息、冻结信息、介质信息和核算信息等。
	法定数字货币钱包信息	指法定数字货币钱包相关属性信息。
	合同协议信息	指合同或协议所包含的所有属性信息，如合同法以及商业银行法所规定的基本信息。
	金融监管和服务	包括反洗钱业务信息、贷款业务信息、货币金银业务信息、贷款保险业务信息等。
	交易信息	交易通用信息、保险收付费信息等。
经营管理	营销服务	产品信息、渠道信息、营销信息。
	运营管理	安防管理信息、业务运维信息、客户服务信息、单证管理信息等。
	风险管理信息	风险偏好信息、风险管控信息。
	技术管理	项目管理信息、系统管理信息。
	综合管理	战略规划信息、招聘信息、员工信息、机构信息等。
监管	数据报送	监管报送信息。
	数据收取	评级、处罚与违规信息、外部审计信息等。

F2.1.2 金融数据分级示例

表 21 金融数据分类示例表

最低安全级别	数据定级要素		数据一般特征
	影响对象	影响程度	
5	国家安全	严重损害/一般损害/轻微损害	(一)重要数据，通常主要用于金融业大型或特大型机构金融交易过程中 重要核心节点类机构的关键业务使用，一般针对特定人员公开，且仅为必须 知悉的对象访问或使用。 (二)数据安全性遭到破坏后，对国家安全造成影响，或对公众权益造成严重影响。
5	公众权益	严重损害	
4	公众权益	一般损害	(一)数据通常主要用于金融业大型或特大型机构、金融交易过程中重要核 心节点类机构的重要业务使用，一般针对特定人员公开，且仅为

4	个人隐私	严重损害	必须知悉的 对象访问或使用。 (二)个人金融信息中的 C3 类信息。 (三)数据安全性遭到破坏后，对公众权益造成一般影响，或对个人隐私或 企业合法权益造成严重影响，但不影响国家安全。
4	企业合法权益	严重损害	
3	公众权益	轻微损害	(一)数据用于金融业机构关键或重要业务使用，一般针对特定人员公开， 且仅为必须知悉的对象访问或使用。 (二)个人金融信息中的 C2 类信息。 (三)数据的安全性遭到破坏后，对公众权益造成轻微影响，或对个人隐私 或企业合法权益造成一般影响，但不影响国家安全。
3	个人隐私	一般损害	
3	企业合法权益	一般损害	
2	个人隐私	轻微损害	(一)数据用于金融业机构一般业务使用，一般针对受限对象公开，通常为 内部管理且不宜广泛公开的数据。 (二)个人金融信息中的 C1 类信息。 (三)数据的安全性遭到破坏后，对个人隐私或企业合法权益造成轻微影 响，但不影响国家安全、公众权益。
2	企业合法权益	轻微损害	
1	国家安全	无损害	(一)数据一般可被公开或可被公众获知、使用。 (二)个人金融信息主体主动公开的信息。 (三)数据的安全性遭到破坏后，可能对个人隐私或企业合法权益不造成影 响，或仅造成微弱影响但不影响国家安全、公众权益。
1	公众权益	无损害	
1	个人隐私	无损害	
1	企业合法权益	无损害	

FE2.2 证券期货业数据分类分级

F2.2.1 证券期货业数据分类

表 22 证券期货业数据分类表

业务条线		数据	
一级子类	二级子类	一级子类	二级子类
交易	交易管理	成交信息	
		委托信息	
		交易业务参数信息	
		交易日志信息	订单日志
			成交日志
监管	监察与评价管理	监察参考信息	
		监察统计及预警信息	监管统计分析结果
			监管预警信息
		评价、处罚与违规信息	
信息披露	信息披露管理	产品发行信息（公开）	
		产品发行信息（未公开）	
其他	业务管理	统计信息	
		其他业务管理	
	技术管理	规划类数据	
		运行管理	配置信息数据
			信息资产管理
			数据字典类
			日志类数据

F2.2.2 证券期货业数据分级

表 23 证券期货业数据分级表

数据级别标识	数据重要程度标识	数据特征
4	极高	1、数据的安全属性(完整性、保密性、可用性)遭到破坏后数 据损失后，影响范围大(跨行业或跨机构)影响程度一般是“严重”； 2、一般特征：数据主要用于行业内大型或特大型机构中的重要业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。
3	高	1、数据的安全属性(完整性、保密性、可用性)遭到破坏后数 据损失后，影响范围中等(一般局限在本机构)，影响程度一般 是“严重”； 2、一般特征：数据用于重要业务使用，一般针对特定人员公开，且仅为必须知悉的对象访问或使用。
2	中	1、数据的安全属性(完整性、保密性、可用性)遭到破坏后数 据损失后，影响范围较小(一般局限在本机构)，影响程度一般 是“中等”或“轻微”； 2、一般特征：数据用于一般业务使用，一般针对受限对象公开；一般指内部管理且不宜广泛公开的数据。
1	低	1、数据的安全属性(完整性、保密性、可用性)遭到破坏后数 据损失后，影响范围较小(一般局限在本机构)，影响程度一般 是“轻微”或“无”； 2、一般特征：数据可被公开或可被公众获知、使用。

F3 电信

参考 YD/T 3813-2020《基础电信企业数据分类分级方法》

F3.1 电信数据分类示例

表 24 电信数据分类示例表

一级子类	二级子类	三级子类	四级子类
用户相关数据	用户身份相关数据	用户身份相关数据	个人身份标识、网络身份标识、用户基本资料、实名认证、用户风险资料
	用户服务内容数据	服务内容和资料数据	服务内容、个人信息
	用户服务行为数据	服务使用数据	业务打点数据、服务记录和日志、消费信息和账单、位置信息、通讯记录数据
		设备信息	终端设备标识、终端设备数据

	用户统计分析类数据	用户使用习惯和行为分析数据	/
		用户上网行为相关统计分析数据	/
企业自身相关数据	网络与系统的建设与运行维护类数据	规划建设类数据（分发前后）	网络规划类、投资计划类、项目管理类
		网络基础资源类数据	公共资源数据、传输资源类数据、承载网资源、核心网资源、接入网资源
		网络运营类数据	信令、网络信标、网络、网络 ID、VLAN、划分、设备监测等
		网络安全管理类数据	安全审计记录、网络安全应急预案、连带者信息、核心区域监测、核心区域护航
	业务运营类数据	业务运营服务数据	产品信息、渠道信息、客户服务信息、营销信息
		公共业务运营服务数据	/
	企业管理数据	发展战略与重大决策	发展战略、重大决策、重要会议
		业务发展类	市场策略、营销管理、资源管理等
		技术研发类	技术管理、技术研究、专利工作
		运行管理类	/
		生产经营类	财务预算、业绩监控、考核信息等
		综合管理类	人力资源、财务信息、办公信息化、采购
	其他数据	合作方提供数据	/

F3.2 电信数据分级示例

按照数据对象的重要敏感程度，可以将基础电信企业网络数据资源分为四个安全级别，其对应的安全要求逐级递减，分别为第四级、第三级、第二级和第一级。

第四级数据：一旦丢失、泄露、被篡改、被损毁会对国家安全、社会公共利益或企业利益或用户利益造成特别严重影响的数据，安全管控要求最高；

第三级数据：一旦丢失、泄露、被篡改、被损毁会对国家安全、社会公共利益或企业利益或用户利益造成严重影响的数据，应实施较强的安全管控；

第二级数据：一旦丢失、泄露、被篡改、被损毁会对国家安全、社会公共利益或企业利益或用户利益造成一定程度影响的数据，执行基本的安全管控；

第一级数据：一旦丢失、泄露、被篡改、被损毁对国家安全、社会公共利益或企业利益或用户利益造成影响较小或无影响的数据，对安全管控不作要求。

F4 医疗

F4.1 医疗健康数据分类示例

表 25 医疗健康数据分类示例表

数据分类		
数据大类	子类	内容
个人属性数据	人口统计信息	姓名、出生日期、性别、民族、国籍、职业、住址、工作单位、家庭成员信息、联系方式、收入、婚姻状态等
	个人身份信息	姓名、身份证号、工作证、居住证、社保卡、可识别个人的影像信息、健康卡号、住院号、各类检测检验相关编号等
	个人通讯信息	个人电话号码、邮箱、账号及关联信息
	个人生物识别信息	基因、指纹、声纹、掌纹、耳廓、虹膜、面部特征等
	个人信用记录信息	个人信用档案、个人信用评分、个人信用报告等
个人健康状况数据		主诉、现病史、既往病史、体格检查（体征）、家族史、症状、起始检测数据、遗传代码数据、可导致敏感数据的健康状态数据、生活方式、基因序列、特殊产物特征、重组序列等
医疗应用数据		门（急）诊病历、住院病历、检查检验报告、用药信息、病程记录、手术记录、麻醉记录、输血记录、护理记录、入院记录、出院小结、结论（诊断）记录、知情告知信息等
医疗资金和支付数据	医疗交易信息	支付信息、消费金额、交易记录等
	保险信息	保险账号、保险状态、保险金额等
卫生资源数据	医疗基本数据	医疗机构名称、医疗机构类别、医院分类代码、床位数、医院地址、电话号码

	医院运营数据	人力资源、财务数据、物资数据、物流数据、基础运行数据等
公共卫生数据	传染病疫情数据	病名、发病人数、发病率、死亡人数、死亡率、发病数据排名、死亡数据排名等
	疾病监测数据	传染病监测信息、非传染病监测信息等
	疾病预防数据	疫苗、应接种人数、实接种人数等
	出生死亡数据	出生人数、出生率、死亡率、自然增长率等

F4.2 医疗健康数据分级示例

表 26 医疗健康数据分级示例表

数据大类	子类	内容	数据级别
个人属性数据	人口统计信息	姓名、出生日期、性别、民族、国籍、职业、住址、工作单位、家庭成员信息、联系方式、收入、婚姻状态等	4
	个人身份信息	姓名、身份证号、工作证、居住证、社保卡、可识别个人的影像信息、健康卡号、住院号、各类检测检验相关编号等	4
	个人通讯信息	个人电话号码、邮箱、账号及关联信息	4
	个人生物识别信息	基因、指纹、声纹、掌纹、耳廓、虹膜、面部特征等	4
	个人信用记录信息	个人信用档案、个人信用评分、个人信用报告等	4
个人健康状况数据	主诉	主诉、症状代码和编码体系名称、症状代码、症状起始日期时间、症状停止日期时间、症状描述、症状性程度代码、严重度代码、初诊标志等	4
	现病史	起病时间、起病符号代码、起病符号描述、起病缓急描述、起病性质原因/诱因、起病特点、伴随症状、本系统症状代码和病名代码、起病后观察情况等	4
	既往病史	既往家族病史项目名称、既往家族病史信息	4
	家族史	家族史数据项项目名称、家族史疾病类代码等	4
	过敏史	过敏史、过敏项目信息、过敏特征信息	4
	……	……	

医疗应用数据	门（急）诊病历	医疗机构组织机构代码、门（急）诊号、科室名称、患者姓名、性别、出生日期、年龄、体格检查等相关信息、既往史、遗传信息、疾病诊断等	4
	住院医嘱	住院医嘱、科室名称、医院名称、入院记录等	3
	检查检验报告	检查记录；门(急)诊号、住院号、检查报告单编号、电子申请单编号、患者姓名、性别、年龄、联系人电话号码、科室名称、病区名称、病床号、病房号、主诉、症状开始日期时间、症状停止日期时间、症状描述、操作标志、操作代码等检验记录；门(急)诊号、住院号、检验报告单编号、电子申请单编号、患者姓名、患者类型、性别年龄、科室名称、病区名称、病房号、病床号、检验申请机构、检验申请科室、检验方法名称、检验项目名称、检验类别、检验项目代码、检验结果代码、检验定量结果等	3
	用药信息	药物用法、药物使用-频率、药物使用-剂量单位、药物使用-次剂量、药物使用-总剂量、药物使用-途径代码、药物名称、药物剂型代码、中药类别代码、药物类型、药物名称代码、中药煎煮法代码等	2
	病程记录	病程记录分类，病程记录、治疗类别代码等	2
	手术记录	门(急)诊号、住院号、电子申请单编号、科室名称、病区名称、病房号、病床号、手术间编号、患者姓名、性别、年龄、术前诊断、术后诊断、手术开始日期时间、手术结束日期时间、手术/操作代码、手术名称、手术级别、手术目标部位名称、手术日期时间、介入物名称、手术体位代码、手术过程描述、手术标志、手术切口描述等	2
	……	……	
医疗资金和支付数据	医疗交易信息	支付信息、消费金额、交易记录等	3
	保险信息	保险账号、保险状态、保险金额等	4
卫生资源数据	医院基本数据	医疗机构名称、医疗机构类别、医院分类代码、床位数、医院地址、电话号码	3
	医院运营数据	人力资源、财务数据、物资数据、物流数据、基础运营数据等	3
公共卫生数据	传染病疫情数据	病名、发病人数、发病率、死亡人数、死亡率、发病数据排名、死亡数据排名等	4

	疾病监测数据	传染病流行病学监测； (1)人口学资料； (2)传染病发病和死亡及其分布； (3)病原体型别、毒力、耐药性变异情况； (4)人群免疫水平的测定； (5)动物宿主和媒介昆虫种群分布及病原体携带状 况； (6)传播动力学及其影响因素的调查；(7)防治措施效果的评价； (8)疫情预测； (9)专题调查(如暴发调查、漏报调查等)等。	4
		非传染病流行病学监测； (1)人口学资料； (2)非传染病发病和死亡及其分布； (3)人群生活方式和行为危险因素监测； (4)地理、环境和社会人文(包括经济)因素的监测； (5)饮食、营养因素的调查； (6)基因型及遗传背景因素的监测； (7)高危人群的确定； (8)预防和干预措施效果的评价	4
	疾病预防数据	疫苗、应接种人数、实接种人数等	1
	出生死亡数据	出生人数、出生率、死亡人数、死亡率、自然增长数、自然增长率等	1

F5 教育

F5.1 教育数据分类示例

表 27 教育数据分类示例表

数据一级类别	数据二级类别	数据示例
教育基础数据	人员基础数据	姓名、性别、民族、籍贯、政治面貌、户口所在地、身份证件号等个人信息。
		个人健康生理信息、个人生物识别信息、不满十四周岁未成年人的个人信息等敏感个人信息。
	学校（机构）基础数据	学校（机构）名称、学校（机构）地址、统一社会信息用代码、学校（机构）邮政编码、校区基本信息、联系电话、电子邮箱等。
教育业务数据	学生管理数据	学籍数据、学历学位数据、资助数据、就业数据、体质健康数据等。
	教职工管理数据	职务职称数据、培养数据、考核数据、聘用数据、离职数据等

	数据二级类别	数据示例
	教育教学管理数据	课程管理数据、教材管理数据、学科专业建设数据、办学条件数据、心理健康数据、学校教育质量评价数据、教师考核评价数据、学生综合素质评价数据、考试管理数据等。
	招生录取数据	招生计划、考生数据、成绩数据、录取数据等。
	科研管理数据	项目管理数据、经费管理数据、成果管理数据、考核管理数据等。
	教育信息化数据	统一身份认证数据、数字教育资源数据、教育数字证书信息等。
	国际交流与合作数据	来华留学数据、出国（境）留学工作数据、外籍教师管理数据、国际交流数据等。
	教育督导数据	义务教育督导数据、学校督导数据、督学管理数据、评估监测数据等。
教育行政管理数据	综合办公数据	公文数据、档案管理数据、会议管理数据、教育信访数据、教育舆情数据、政务公开数据等。
	政策规划数据	法规、规章、规划等文件在制定修订过程中产生的数据。
	财务与审计数据	财务管理数据、项目经费管理数据、审计数据等。
	资产管理与后勤服务数据	资产管理数据、后勤管理数据等。
	干部人事数据	编制管理数据、招录数据、职称评审数据、职务任免数据、人事考核数据、人事档案数据、教育培训数据、离退休管理数据等。
	信息系统运行与安全数据	网络拓扑数据、资产清单数据、监测预警数据、威胁情报数据、安全事件数据、日志数据等。

F6 工业

F6.1 工业数据分级示例

（DB50/T 1453-2023 工业数据分类分级导则）

表 28 工业数据分级示例表

行业类别	企业名称	数据业务类型	一级子类	二级子类	数据名称	数据产生部门	存储方式	存储位置	更新频率	数据用途	数据级别	数据描述
汽车零配件	xxx有限公司	生产制造类	生产制造数据	生产监控数据	机加数据采集	大数据中心	本地	MES系统	热数据	统计分析	L2级	机械加工场景的运行数据

汽车零配件	xxx有限公司	生产制造类	生产制造数据	工业控制信息	SPC分析	大数据中心	本地	MES系统	热数据	统计分析	L2级	过程控制数据
汽车零配件	xxx有限公司	经营管理类	经营管理数据	资源管理数据	采购订单	大数据中心	本地	ERP系统	热数据	记账分析	L2级	采购类数据
电气机械行业	xxx有限公司	生产制造类	生产制造数据	工艺参数	工艺文件	技术研发部	数据库	本地服务器	温数据	生产制造使用	L2级	生产工艺参数
电气机械行业	xxx有限公司	生产制造类	生产制造数据	X	质量数据	检测中心	本地	SAP系统	温数据	生产制造使用	L2级	过程检查、质量数据
XXX	xxx研究院	经营管理类	经营管理数据	资源管理数据	生产核算数据	财务资产部	本地	SAP系统	热数据	企业内部使用	L2级	业务统计数据(订单、仓储、排产等)
汽车零配件	xxx有限公司	生产制造类	生产制造数据	生产监控数据	机加数据采集	大数据中心	本地	MES系统	热数据	统计分析	L2级	机械加工场景的运行数据
汽车零配件	xxx有限公司	生产制造类	生产制造数据	工业控制信息	SPC分析	大数据中心	本地	MES系统	热数据	统计分析	L2级	过程控制数据
汽车零配件	xxx有限公司	经营管理类	经营管理数据	资源管理数据	采购订单	大数据中心	本地	ERP系统	热数据	记账分析	L2级	采购类数据
电气机械行业	xxx有限公司	生产制造类	生产制造数据	工艺参数	工艺文件	技术研发部	数据库	本地服务器	温数据	生产制造使用	L2级	生产工艺参数
电气机械行业	xxx有限公司	生产制造类	生产制造数据	X	质量数据	检测中心	本地	SAP系统	温数据	生产制造使用	L2级	过程检查、质量数据
XXX	xxx研究院	经营管理类	经营管理数据	资源管理数据	生产核算数据	财务资产部	本地	SAP系统	热数据	企业内部使用	L2级	业务统计数据(订单、仓储、排产等)

F7 烟草

烟草行业数据分类分级词典示例				
一级类别	二级类别	三级类别	包含内容	级别
用户相关数据	用户身份相关数据	自然人身份标识	个人基因、指纹、声纹、掌纹、耳廓、虹膜、面部特征等。	四级

用户相关数据	用户身份相关数据	网络身份标识	系统账号、IP 地址、邮箱地址及与前述有关的密码、口令、口令保护答案、用户个人数字证书等。	三级
用户相关数据	用户身份相关数据	用户基本资料	个人姓名、生日、性别、民族、国籍、家庭关系、住址、个人电话号码、电子邮箱等。	三级
用户相关数据	用户身份相关数据	用户私密资料	银行账号、鉴别信息（口令）、存款信息（包括资金数量、支付收款记录等）、房产信息、信贷记录、征信信息、交易和消费记录、流水记录等。	四级
用户相关数据	用户服务内容数据	服务内容数据	烟草产品交易内容信息。	四级
用户相关数据	用户服务内容数据	业务订购管理	烟草产品订购信息、订单信息、物流订单信息、渠道与订单信息等。	三级
用户相关数据	用户服务内容数据	服务记录和日志	服务器日志的典型例子是网页服务器的日志，其中包含页面请求的历史记录、客户端 IP 地址、请求日期/时间、请求的网页、HTTP 代码、提供的字节数、用户代理、引用地址等。	三级
企业经营数据	系统运维数据	网络规划	网络地址、掩码数据、身份标识、URL 地址数据、分域信息、网络子网、网络端口、分区地址数据。	二级
企业经营数据	系统运维数据	系统日志	各软硬件系统日志。	二级
企业经营数据	企业管理数据	生产经营数据	企业销售收入、销售回款、利润水平、经营活动现金流量、期间费用、生产成本、产品质量等。	三级
企业经营数据	企业管理数据	标准化数据	企业标准库、法律法规库、企业标准历史库等。	二级
企业经营数据	企业管理数据	企业制度数据	企业外来制度和来自企业内部制度等。	二级
企业业务数据	产品信息	业务产品信息	如卷烟基本信息、烟包信息、卷烟质检等数据。	三级
企业业务数据	产品信息	渠道信息	烟草品牌分销渠道、烟草渠道建设、渠道的定义、产品渠道，生产渠道、消费领域途径。	三级
企业业务数据	销售信息	客户服务信息	客情维护、经营指导、消费推荐、活动和品牌信息推广、策划设计、会场租赁、市场发展、客户管理维护服务等。	三级

企业业务数据	物流信息	烟草物流	烟草物流是指烟草及其制品、原辅料从生产、收购、储存、运输、加工到销售服务整个过程中物质实体运动以及流通环节的所有附加增值活动，包含烟草零售户、订单品种、仓储、进货、出货、进库、出库、入库等。	三级
企业业务数据	物流信息	烟草仓储	如统计报表、排号、扫码点件、二检排包、磅码、成件赋码、烟叶入库、烟叶出库等。	三级

附录 G-非结构化文件识别规则示例

在非结构化数据中识别文档类数据通常涉及多种技术和策略。

表 29 非结构化文件识别规则示例表

文件类型	规则类型	识别点	识别方法示例
文档	文档框架识别规则示例	标题和小标题识别	使用文本大小、粗体或特定格式（如冒号后跟文本）来识别标题和小标题。
		段落边界识别	基于文本格式（如缩进、换行）来确定段落的开始和结束。
		列表和子列表识别	通过项目符号、数字或字母序列来识别列表和子列表的结构。
		多级标题识别	通过字体大小、粗细和编号来区分不同级别的标题。
		页眉页脚识别	检测重复出现的文本元素，如页码、文档标题等，来识别页眉和页脚。
		表格和图表识别	利用线条、格子或特定布局来识别文档中的表格和图表。
	关键词和词组识别规则示例	词频统计	统计文档中最常出现的词汇，高频词可能是关键词。
		领域特定术语	识别特定行业或领域的专业术语。
		标题和小标题分析	标题和小标题中的词汇往往是关键词。
		词性标注	通过词性标注识别名词短语或动词短语。
		固定搭配识别	识别常用的词汇组合，如“经济增长”“市场调研”等。
		同义词和近义词识别	识别文档中的同义词或近义词短语，以理解上下文意义。
		上下文分析	分析词汇的上下文使用，以更准确地确定其作为关键词的重要性。
		多词表达式识别	识别固定词组或习语，如“举足轻重”“一石二鸟”等。
		主题建模	使用如 LDA（LatentDirichletAllocation）等算法来识别与文档主题相关的关键词和短语。
	语句识别规则示例	语法结构分析	分析句子的语法结构，如主谓宾结构，以识别完整的语句。
		标点符号使用	利用句号、问号、感叹号等标点符号来确定句子的界限。
		连接词和过渡词分析	识别连接词和过渡词，如“因此”“然而”，来理解句子之间的逻辑关系。
		从句和嵌套结构识别	识别复杂句子中的从句和嵌套语句结构。
		语言风格和语调分析	识别文档中的语言风格和语调，如正式、口语、幽默等。
		引用和直接引语识别	通过引号和报告动词来识别引用和直接引语。

视频	视频文字内容识别规则示例	文本区域检测	在视频帧中检测文本的位置，可能涉及复杂的背景和不同的字体。
		光学字符识别	识别视频帧中的文字内容，这可能需要处理模糊或动态变化的图像。
		上下文关联分析	将文本内容与视频中的其他视觉元素和情境进行关联分析。
	人脸识别和表情分析	人脸检测与追踪	在视频序列中检测和追踪人脸
		表情识别	分析视频中的面部表情，用于情感分析或身份验证。
		人脸识别	对视频中的人脸进行身份识别，可能需要处理不同的角度和光照条件。
	物体和场景识别	物体检测	识别视频中的物体，如车辆、行人、动物等。
		场景理解	分析视频中的场景，理解其背景、环境和上下文。
		动作识别	识别视频中人物的动作和活动，如跑步、跳舞等。
	视频行为分析	行为识别	分析视频中人物的特定行为，如走路、打电话等。
		事件检测	识别视频中的特定事件，如事故、争论等。
		轨迹分析	追踪视频中物体或人物的移动轨迹。
	音视频融合分析	口型与语音同步	检查视频中人物的口型是否与音频中的语音同步。
		场景与音效关联	分析视频场景与背景音乐或音效之间的关联。
		多通道分析	结合视频的视觉内容和音频内容进行全面分析。
	交互式视频分析	用户反馈学习	根据用户的反馈对视频分析模型进行调整和优化。
		标注和元数据分析	使用视频标注和元数据来提高视频内容的理解和分类。
图片	人脸识别规则示例	面部特征检测	通过检测人脸的关键特征点，如眼睛、鼻子、嘴巴等，来识别人脸。
		面部特征向量化	将面部特征转换为数值向量，以便于比较和识别。
		深度学习模型应用	使用卷积神经网络（CNN）等深度学习模型进行面部特征的学习和识别。
		生物特征比对	比对面部特征数据与数据库中的已知面部数据，以识别个体。
	文字内容识别规则示例（光学字符识别，OCR）	文本区域定位	定位图片中的文本区域。
		字符分割	将文本区域中的字符进行分割。
		字符识别	识别分割后的字符。
		上下文校正	使用上下文信息校正识别出的文字，提高准确性。
	物体和场景	特征提取	提取图片中的关键特征，如颜色、形状、纹理等。

	识别规则示例	模式识别	使用模式识别技术识别图片中的特定物体或场景。
		深度学习分类器	应用深度学习分类器，如 CNN，对物体或场景进行分类。
	运动和行为识别规则示例	序列帧分析	分析视频序列中连续的帧，以识别运动或行为。
		时空特征识别	识别影像中随时间变化的空间特征。
		动作模式识别	使用机器学习算法来识别特定的动作模式。
	图像分割和标注规则示例	区域分割	将图片分割成多个区域，每个区域包含不同的对象或特征。
		边缘检测	通过检测图像中的边缘来分割对象。
		语义分割	将图像中的每个像素分类到不同的类别，以识别不同的物体。
	色彩和纹理分析规则示例	色彩直方图分析	分析图像的色彩分布。
		纹理特征识别	识别图像中的纹理特征，如光滑、粗糙等。
		色彩空间变换	将图像从一个色彩空间转换到另一个，如从 RGB 到 HSV，以便更好地分析和处理图像。
音频	语音识别规则示例（自动语音识别，ASR）	声波特征提取	提取语音信号的基本特征，如频率、节奏和音量。
		声音分割和归类	将语音信号分割成较小的单位，如音素或字。
		模型训练	使用深度学习模型（如循环神经网络，RNN）训练系统以识别和转换语音数据为文字。
		上下文分析	分析语言的上下文和语法结构以提高识别准确性。
	语音情感分析规则示例	音调和节奏分析	分析语音的音调、强度和节奏来判断说话者的情感状态。
		声音特性识别	识别语音中的特定声音特性，如颤音、哭腔，以确定情感。
		机器学习分析	利用机器学习算法，如支持向量机（SVM）或神经网络，来分类和识别不同的情感状态。
	音乐和声音分析规则示例	音调和节拍检测	分析音乐中的音调和节拍，识别旋律和节奏模式。
		声音分类	将声音文件分类为不同的类别，如语音、音乐、环境声音等。
		乐器识别	识别音乐中使用的不同乐器。
	环境声音识别规则示例	声音事件检测	检测和识别特定的声音事件，如玻璃破碎、汽车喇叭等。
		声源定位	使用麦克风阵列等技术来定位声音的来源。
		背景噪声分析	分析和识别背景中的噪声类型，如交通噪声、人群聊天等。
	多通道和多模态数据融合	视频与音频同步分析	结合视频和音频数据进行更全面的分析，如在电影或监控场景中的应用。
		跨媒体内容分析	结合音频、视频、文本等不同媒体形式的数据进行综合分析。
其他非结构化文件	社交媒体数据分析规则	情感分析	分析社交媒体帖子的文本内容，以识别情感倾向，如积极、消极或中性。

		话题检测和跟踪	通过关键词、标签和话题聚类来监测和追踪热门话题。
		用户行为分析	分析用户的互动模式，如点赞、评论和分享，以了解内容的影响力和受欢迎程度。
	传感器数据分析规则	时间序列分析	分析来自传感器的时间序列数据，识别模式和异常。
		环境监测	使用气象、温度、度等传感器数据来监测和预测环境变化。
		物体追踪和定位	使用 GPS 和其他传感器数据来追踪和定位移动物体。
	网络和日志数据分析规则	异常检测	分析网络流量或系统日志以识别异常行为，如网络攻击、系统故障等。
		使用模式识别	识别使用模式或趋势，如网站访问量的高峰时段、操作系统的常见错误等。
		数据聚合和关联分析	来自不同来源的日志数据进行聚合和关联分析，以获得更全面的洞察。
	医疗和生物信息数据分析规则	基因序列分析	分析基因序列来识别遗传变异和相关疾病。
		医疗影像分析	使用计算机视觉技术分析 X 光片、CT 扫描等医疗影像，以辅助诊断。
		患者数据分析	分析患者的电子健康记录（EHR）来识别健康风险和治疗反应。
	4IoT（物联网）数据分析规则	设备性能监测	监测 IoT 设备的性能和状态，识别维护和升级的需求。
		智能家居行为模式识别	分析智能家居设备的使用数据，识别居住者的行为模式和偏好。
		能源消耗分析	分析来自智能电表的数据，以优化能源使用和成本。

参考文献

[1] 中华人民共和国数据安全法（2021 年 6 月 10 日中华人民共和国第十三届全国人民代表大会常务委员会第二十九次会议通过）.
[2] 《关于构建数据基础制度 更好发挥数据要素作用的意见》（数据二十条）.
[3] 国家标准化管理委员会.《数据安全技术 数据分类分级规则》（GB/T 43697-2024）.
[4] 《金融数据安全 数据安全分级指南》（2020 年 9 月 23 日中国人民银行发布）.
[5] 《工业数据分类分级指南（试行）》（2020 年 2 月 27 日工业和信息化部办公厅印发）.
[6] 云安全联盟 (CSA).《CSA 数据安全词汇表》.
[7] 国家标准化管理委员会.《数据安全管理能力成熟度模型》（GB/T 37988-2019）.
[8] 国家标准化管理委员会.《信息技术-大数据-数据分类指南》（GB/T 38667-2020）.
[9] 贵州省地方标准.《政府数据 数据分类分级指南》（DB52/T 1123-2016）.
[10] 工业和信息化部.《基础电信企业数据分类分级方法》（YD/T3813-2020）.
[11] 浙江省地方标准.《数字化改革 公共数据分类分级指南》（DB33/T 2351—2021）.
[12] 中华人民共和国个人信息保护法（2021 年 8 月 20 日中华人民共和国第十三届全国人民代表大会常务委员会第三十次会议通过）
[13] 网络安全审查办法（2021 年 11 月 16 日国家互联网信息办公室 2021 年第 20 次室务会议审议通过）.
[14] 中华人民共和国网络安全法（2016 年 11 月 7 日中华人民共和国第十二届全国人民代表大会常务委员会第二十四次会议通过）
[15] 《重要数据识别指南（征求意见稿）》（2022 年 1 月 13 日全国信息安全标准化技术委员会审议通过）.
[16] 国家标准化管理委员会.《信息安全技术 个人信息安全规范》（GB/T 35273-2020）.
[17] Gartner®.《2024 年中国数据、分析和人工智能技术成熟度曲线》.
[18] Gartner®.《Hype Cycle for Data management,2024》.
[19] 《网络数据安全管理条例》（2024 年 8 月 30 日国务院第 40 次常务会议通过）.
[20] 国家标准化管理委员会.《信息安全技术 网络数据分类分级要求（征求意见稿）》.
[21] 上海市地方标准.《上海市公共数据开放分级分类指南（试行）》.
[22] 坪山区.《坪山区政务数据分级分类管理办法(试行)》.
[23] 浙江省地方标准.《数据资源管理第 3 部分：政务数据分类分级》（DB3301/T0322.3—2020）.
[24] 烟台市地方标准.《烟台市公共数据开放分级分类指南（试行）》.
[25] 重庆市地方标准.《重庆市公共数据分类分级指南（试行）》.
[26] 吉林省地方标准.《政务数据安全分类分级指南》（DB2201/T17—2022）.
[27] 河北省地方标准.《政务数据分类分级要求》（DB14/T2442—2022）.
[28] 江苏省地方标准.《政务数据安全分类分级指南》（DB3212/T1116—2022）.
[29] 无锡市地方标准.《无锡市公共数据分类分级实施指南》（DB3202/T1049—2023）.
[30] 江苏省地方标准.《公共数据分类分级指南》（DB3203/T1024—2023）.
[31] 四川省地方标准.《政务数据 数据分类分级指南》（DB51/T3056—2023）.
[32] 湖北省地方标准.《公共数据资源开放第 2 部分：分类分级指南》

（DB4201/T677.2—2023）.
[33]黑龙江省地方标准.《政务预公开数据分类分级评估指南》（DB23/T3510—2023）.
[34]国家标准化管理委员会.《证券期货业数据安全风险防控数据分类分级指引》（GB/T 42775-2023）.
[35]工业和信息化部.《基础电信企业数据分类分级方法》（YD/T3813-2020）.
[36]工业和信息化部. 《电信网和互联网数据分类分级技术要求与测试方法》（YD/T4244-2023）.
[37]国家标准化管理委员会.《信息安全技术 健康医疗数据安全指南》（GB/T 39725-2020）.
[38]《健康医疗数据安全指南》.
[39]广东省地方标准.《高校数据安全分类分级指南》（T/GDCSA 000-2022）.
[40]卫生健康委员会.《卫生健康信息数据集分类和编码规则》（WS/T306-2023）.
[41]国家标准化管理委员会.《智能制造工业数据分类原则》（GB/T 42128-2022）.
[42]《DATA CLASSIFICATION PRACTICES》.
[43]国家标准化管理委员会.《政务信息资源目录体系第 4 部分:政务信息资源分类》（GB/T 21063.4）.
[44]北京市地方标准.《政务数据分级与安全保护规范》（DB11/T 1918-2021）.
[45]重庆市地方标准.《工业数据分类分级导则》（DB50/T 1453-2023）.
[46]《观安观智敏感数据发现软件-产品手册》.
[47]《天融信数据分类分级系统-产品手册》.
[48]《山石网科数据安全综合治理平台-产品手册》.
[49]《大道云隐密数云资产保护系统-产品手册》.
[50]《神州数码数据分类分级系统-产品手册》.
[51]《昂楷数据安全分类分级系统-产品手册》.
[52]《美创暗数据发现和分类分级系统（DDAC）-产品手册》.
[53]《明朝万达 Chinasec（安元）智能数据治理平台-产品手册》.
[54]https://www.zhongfu.net/news/techinfo/1089.html
[55]https://en.wikipedia.org/wiki/Naive_Bayes_classifier

CSA GCR cloud security alliance
GREATER CHINA REGION

Cloud Security Alliance Greater China Region

扫码获取更多报告

云安全联盟大中华区
官网：https://c-csa.cn 电话：0755-86548359
地址：上海市浦东新区东育路255弄5号3FB30（前滩国际经济组织集聚区）/
广东省深圳市南山区粤海街道北京航空航天大厦1号楼905