专家洞察

分条析理
安全有道

分类分级成为企业数据
安全差异化管理的先驱者

IBM 商业价值研究院

IBM

主题专家

张宁
IBM GBS CBDS 团队
数据治理解决方案
首席业务咨询顾问
znznbj@cn.ibm.com

张骐微
IBM GBS CBDS 团队
数据治理高级顾问
zqweiz@cn.ibm.com

王莉
IBM 商业价值研究院
高级咨询经理
gbswangl@cn.ibm.com

扫码关注 IBM 商业价值研究院

官网　微博　微信公众号　微信小程序

谈话要点

数据分类分级既是合规要求，也有现实意义

数据分类是数据安全保护的重要基础性工作之一，其最终目的是实现数据的安全共享。

建设分类分级清单需要四大抓手

要做好数据分类分级清单建设，可以从权威保障、逻辑梳理、物理映射、技术工具四个方面着手。

数据分类分级不是终点，而是安全管理的起点

数据分类分级是一个从整体规划、建设分类分级清单、到采取保护措施的闭环管理过程。

—

数据安全，刻不容缓

近年来，国内外数据泄露的事件屡屡发生，泄露事故严重影响了客户的满意度和企业的声誉，进而对企业效益产生负面影响。在 IBM Security 发布的《2020 年数据泄露成本报告》[1] 中显示，数据泄露事件给企业造成的平均成本为 386 万美元。其中，客户个人身份信息是造成企业耗费成本最高的一项。

金融行业也存在数据泄露事件。2017 年底，广州某证券公司报案称，该公司多名客户投诉，刚开户不久就有冒充该证券公司的人员打电话或发微信推荐股票，让客户跟单操作，怀疑客户个人信息泄露。广州市公安机关在北京、湛江、深圳、珠海等地同步收网，抓获犯罪嫌疑人 40 余人，源头“内鬼”2 人，缴获公民个人信息 230G。

林某为某券商的证券经纪人，利用自学黑客技术，攻破了多家政府网站和证券或期货公司内部系统。林某非法获取公民个人信息共计 400 余万条，情节特别严重，有期徒刑 5 年 2 个月，并处罚金人民币 2 万元。[2]

上述案例提醒我们，证券期货行业应严格落实信息安全保密措施、加强管理，严防信息泄露。

整体来看，证券期货行业数据空间可以划分为三个维度：一是业务空间，即金融机构在开展交易中介、资本中介、融资中介、投资研究、自营投资、OTC 市场等业务活动时产生的数据。二是管理空间，即金融机构在进行人力、合规、风控、财务等对内经营管理活动时产生的数据。三是服务空间，即金融机构在和外部的个人投资者、机构投资者、融资机构等相关服务对象进行交互时产生的数据。

从以上三个维度的数据不难看出，证券期货行业的数据具有体量大、敏感程度高、重要程度高的特点：

体量大：特别是交易所或结算公司这类大型机构，汇聚了行业内多家机构的数据。

敏感程度高：大量数据涉及投资者的基本信息、账户信息、财务信息等。

重要程度高：涉及到交易数据、风控数据、经营管理数据等。

因此，金融机构数据基于安全的差异化管理，以及合规使用尤为重要。意识到数据安全重要性的金融机构，纷纷在安全保护领域发力。

分类分级是数据安全保护体系中一项重要的基础性工作，然而在开展数据安全分类分级工作时，很多企业面临着挑战：

- *缺乏整体解决方案*：面对企业自身众多的数据，分类要如何划分、级别设置几个层级、分类分级结果要如何落地？这些问题导致企业不知如何着手数据分类分级工作。
- *缺少数据安全差异化管理*：很多企业都有建立数据使用的审批流程，但审批流于形式，因为并不清楚哪些属于敏感数据或者敏感数据存在哪里。导致管理成本虽高，但安全保护方面收效甚微。
- *欠缺专业人才和能力*：数据安全的分类分级对专业能力要求较高，一方面需要有金融业务的知识储备，并且熟悉对应系统中的数据；另一方面需要具备数据治理能力，特别是数据安全领域的专业能力。

随着监管的要求越来越严格，如何切实做好数据的分类分级，保障数据安全？ IBM 结合多年的项目实施经验，总结了关键的应对策略。

警钟长鸣，监管亮剑

为了维护市场安全运行和维护投资者合法权益，监管机构针对数据分类分级工作，从不同层面明确了证券期货经营机构的合规责任，成为业内机构满足法规依从性的重要依据。

在法律层面，在全国人大 2016 年 11 月发布的《网络安全法》[3] 和 2020 年 10 月发布的《个人信息保护法（草案）》[4] 中均有相关规定。《网络安全法》中要求网络运营者应当按照网络安全等级保护制度的要求，采取数据分类、重要数据备份和加密等措施。《个人信息保护法（草案）》中要求个人信息处理者需对个人信息实行分类分级管理。

在行业标准层面，证监会于 2018 年 9 月发布了金融行业标准《证券期货业数据分类分级指引》[5]（以下简称“《指引》”）。该指引详细阐明了适用范围、数据分类分级的前提条件、数据分类及分级方法、数据分类分级中的关键问题处理。

我们梳理数据分类分级的监管脉络，可以发现《网络安全法》提出了数据分类分级要求，为后续出台的证券期货行业的具体指引提供了法律渊源。而《指引》虽然是推荐性行业标准，但金融机构开展数据分类分级工作，不仅是满足法规依从性的要求，也具有极强的现实意义，有以下几点尤其值得关注：

适用数据范围广

除了经营和管理活动中常用的产生、采集、加工、使用或管理的网络数据或非网络数据之外，甚至包括了通过购买或数据共享等方式获得的外部数据，可谓是基本涵盖了数据获得的所有可能方式。

因地制宜采取安全保护措施

在考虑数据定级和数据的安全保护环节时，要根据不同的安全属性侧重，采取相适应的保护措施。并且综合考虑数据体量的大小，在实际使用中，采取适当的、合理的安全防护措施。例如同级别的单个客户信息和批量客户信息，查询批量客户信息时应增加更严格的访问控制手段，且对查询行为进行审计。

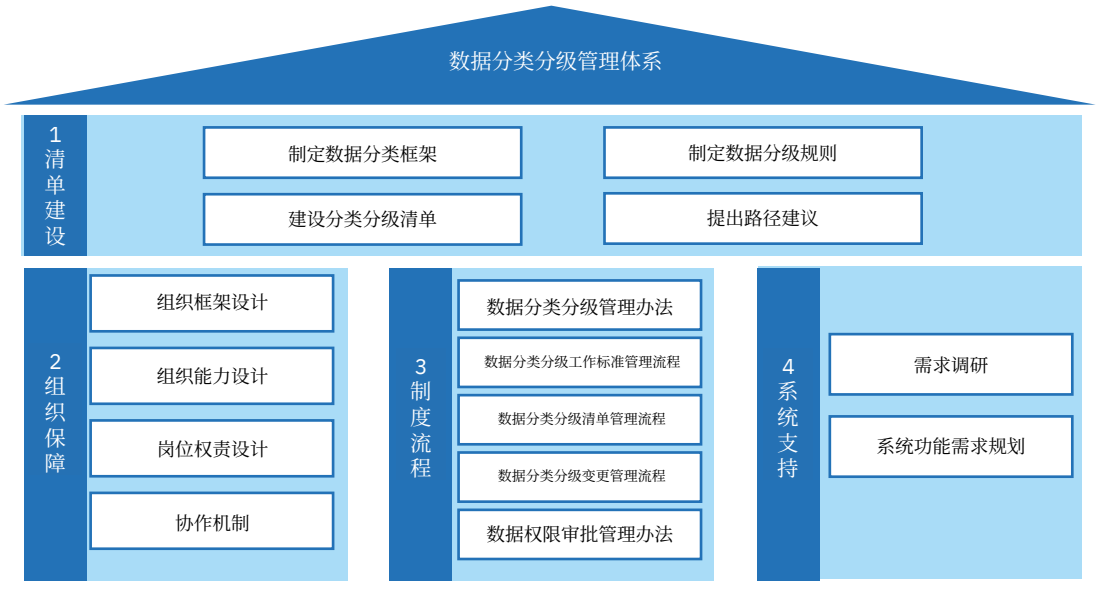
分类分级的最终目的是实现数据的安全共享

数据分类分级是数据安全保护的重要基础性工作之一，基于分类分级从而实现对重要、敏感数据的进一步安全保护，使数据能够在安全合规的情况下被共享使用，发挥更大的价值。

数据分类分级和个人信息保护、网络安全等问题密切相关。证券期货行业的经营机构在未能满足监管合规要求的情况下，可能面临中国证监会采取的从责令改正、责令增加合规检查次数，到暂停业务等的行政监管措施。需要提醒的是，除了证监会外，主要执法机构还有国家网信办、工信部、公安部等。多重执法主体间不仅可能存在着权责边界模糊、交叉执法的问题，在监管“九龙治水”现象存在的情况下，证券期货行业的经营机构需要结合自身情况，进行全面的合规建设，避免监管风险和行政处罚。

图 1

数据分类分级管理体系



分类分级，从容应对

如何开展数据分类分级工作呢？从数据安全保护的落地性、有效性和持续性等角度考虑，首先要建设数据分类分级清单，这是数据分类分级的核心工作内容；其次要有健全的保障机制，即在组织架构、制度流程、技术工具方面为数据分类分级工作保驾护航（见图 1）。

IBM 认为，要做好数据分类分级清单建设，可以从以下四个方面着手：

权威保障：充分研读和理解行业权威机构的发文，保障数据分类分级工作的正确性、合理性以及业务覆盖的完整性。

逻辑梳理：在逻辑层面构建数据分类分级清单，包括数据分类框架、数据定级规则和数据项清单三部分内容。

物理映射：完成数据分类分级的落地实现，即把数据分类分级清单中逻辑层面的数据项，与 IT 系统的字段关联起来。

技术工具：金融机构的数据量庞大，借助技术工具有效、持续的开展数据分类分级工作。

下面，我们将分别进行阐述。

权威保障

以证券行业为例，《证券期货业数据分类分级指引》在工作方法、工作原则等方面提供了理论指导；《证券期货业业务标准规划》从业务条线、业务过程的角度提供了三级的业务分类参考；《证券期货业逻辑数据模型》从数据角度提供了数据的分类参考。这些都可作为数据分类分级工作的重要输入。

逻辑梳理

分类框架和定级规则是数据项分类分级清单的前提，有了分类的框架和定级的规则，才可以为数据项确定其归属的分类和级别。数据的逻辑梳理可以按以下三个步骤开展：

智能化的数据资产发现和分类 [6]

某金融机构与 IBM 合作，进行数据治理工作，其中一项工作是对客户重要信息实现在系统中的定位，即客户的重要信息，需要被映射到各个系统数据库表的具体字段。

客户当前的数据字典和元数据描述存储于 Excel 中，包括 10 个系统、1 万余张表、20 多万个字段。其中字典数据质量欠佳，有一半字段没有中文名，而且还有一部分字段的英文名是由拼音首字母组成（例如，资金流向和证件类型都是 ZJLX）。如果按照英文到中文的关键字进行匹配和定位具体字段就很容易出错。

如果采用人工方式锁定和分析个人客户重要信息的分布情况，则至少需要一名有经验的数据分析师 15 个工作日以上的时间。针对相关法律规定、监管文档及附件进行个人客户相关的重要信息进行搜集得到的大约 50 种信息项，IBM 使用词向量技术，泛化为 12 类客户重要信息，包括涉及联系电话的手机号、座机号在内的词库扩展，并使用泛化后的信息和业务系统数据库进行匹配。

借助人工智能技术，IBM 帮助客户在 3 个工作日内完成了客户重要信息定位的全部工作，而且准确性优于人工分析的结果。

海通证券：数据分类分级，促进数据共享与价值发挥[7]

海通证券作为国内头部券商，积累了大量的业务数据，也十分注重数据资产的安全保护。在《证券期货业数据分类分级指引》发布后，与 IBM 合作开展数据分类分级与业务数据安全保护项目，旨在通过开展数据分类分级工作，建立数据分类分级清单及管理规范，完善公司数据访问控制手段。

该项目结合海通证券的数据现状、监管要求和行业标准，以及 IBM 的领先实践，为海通证券构建了 2 大类 5 层级的分类框架，4 层级的数据定级规则和定级调整规则，近 1200 项的经纪业务条线数据分类分级清单。在分类框架的设计上，考虑到逻辑数据模型的主题划分在金融领域是一套成熟的分类结构并被广泛地应用，且预定义了丰富的业务属性信息，因此引入了逻辑数据模型作为参考，以保证分类结构的稳定性和实用性。如“交易”业务分类下的二级分类由“主体、账户、品种、合同、事件、资产、渠道、营销、行情资讯”组成。数据级别由低到高分为四级：1 级最低，表示可被公开或可被公众获知使用的数据，如公开渠道的行情资讯；4 级最高，表示可以识别触达到个人信息主体的数据，一旦泄露、滥用可能危害人身和财产安全，如个人手机号码。定级调整规则包括公开披露、脱敏处理、数据时效等情况下数据级别的调整，如披露信息未公开时的数据级别是 2 级，公开后满足数据 1 级的定级条件，则其数据级别下调至 1 级。数据分类分级清单包括逻辑数据项、逻辑数据项对应的业务一级分类、业务二级分类、数据一级分类、数据二级分类、数据三级分类、数据级别、定级说明。

海通证券数据分类分级清单是在行业监管、国家及行业标准规范下进行建设的，充分满足合规要求。作为数据安全管理工作的重要抓手，对数据实行差异化管理和保护，在保证数据安全的基础上促进了数据的共享和价值发挥。

- 搭建数据分类框架

首先要进行业务细分。银行、证券等金融机构的业务条线众多，业务细分通常是逻辑划分，不与细节的、具体的数据对应。业务细分根据业务条线的复杂程度，一般划分 1 至 2 个层级，层级过多不利于分类的维护和管理。

其次是数据归类。在业务细分的基础上，按数据性质、管理需要、使用需要、重要程度进行数据归类。如，按数据性质，证券行业的账户数据可分为交易账户数据、资金账户数据和银行账户数据；按管理需要，风控数据可分为市场风险数据、信用风险数据、操作风险数据等；按使用需要，客户数据可分为个人客户数据、机构客户数据；按重要程度，信息披露数据可分为公开信息和未公开信息。

- 确定数据定级规则

数据的安全属性包括数据的完整性、保密性和可用性。分析数据的安全属性遭到破坏后的影响对象、影响范围和影响程度，以此确定数据的安全保护级别，数据安全级别通常分为 4 级。在完整性和可用性要求基本一致的情况下，则以保密性为主要定级依据。

由于数据在传输、使用等过程中，因各类业务需要，在数据体量或敏感程度等方面会发生变化，因此数据的级别也要随之调整。如个人客户的手机号码是高敏感字段，但经过脱敏处理后，其数据级别可向下调整一级。

- 制定数据项分类分级清单

梳理各业务条线的重点业务及其产生的数据，然后分析数据项所属的分类，与数据分类框架进行映射。再根据数据定级规则，确定数据项的安全级别。最终形成数据项分类分级清单。

物理映射

数据分类分级清单中的一个数据项可能对应多个系统中的多个字段，例如清单中的“个人手机号码”，它可以表示个人客户的手机号码，也可以表示本机构员工的手机号码。因此逻辑层面的数据分类分级清单建设完成后，要想切实发挥作用，还需要进行物理映射，即数据分类分级清单与业务系统进行映射关联。建立映射有两种方式，一种基于数据字典进行映射，一种在系统中的物理表上新增分类分级的属性列。

技术工具

数据管理者实现数据分类分级信息的统一管理和日常维护，需要借助相应的技术手段来提升工作的效率，降低人工成本。搭建数据分类分级系统，作为工作平台以有效支撑数据分类框架、数据定级规则、数据分类分级清单、业务系统物理映射等的建设和日常维护工作。特别是物理映射环节，运用 AI 技术以最小的投入来实现数据分类分级的落地。同时通过这个平台，还可以解决数据分类分级的展示、查询和管理需求。

下面，我们结合具体的应用场景，来对比一下引入数据分类分级前后的审批流程。前文提到，很多企业内部的数据审批由于缺乏审批依据而流于形式，未能起到有效的数据保护作用。比如，一些企业在数据使用的审批环节设置了很多审批节点，看似对数据安全保护有利，但每个审批者对自己审批的重点不清晰，往往变成因为业务部门需要数据，就只能提供数据的局面。这样的流程，既不能起到安全保护的审批效果，又因冗长的审批节点增加了工作成本。

针对此痛点，建议企业结合自身具体情况，设计可落地的数据审批流程（见图 2）。在数据使用审批环节，重点关注所需数据是否涉及高等级数据；如果涉及，业务部门是否能接受脱敏或加密等物理处理。通过基于数据级别的差异化管理，使得数据分类分级和安全保护工作更具抓手。

在审批流程的设计中，有两个原则需要强调：

- 差异化原则：不同级别的数据使用申请要遵循不同的审批流程，目前业内普遍以数据级别越高，越需要高层管理者审批为原则。例如业务部门申请的数据存在不同的风险类别，较低级别的数据、个人身份识别信息、以及一旦泄露会造成巨大风险的高级别数据，所对应的审批部门和流程应该存在差异化设计。
- 从严保护原则：建议企业对数据贯彻从严保护原则，若申请的数据集包含四级 / 三级数据，数据集的整体安全级别应当整体升级，从严保护。

图 2

引入数据分类分级后的审批流程