ICS 35.240.99
CCS L67

T/CCIASC

团 体 标 准

T/CCIASC 0006—2024

数据分类分级产品技术要求

Technical requirements for data classification and grading product

2024－05－09 发布　　2024－05－16 实施

中国计算机行业协会　　发 布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国计算机行业协会提出并归口。

本文件起草单位：北京天融信网络安全技术有限公司、中国软件评测中心（工业和信息化部软件与集成电路促进中心）、北京市政务信息安全保障中心（北京信息安全测评中心）、国网吉林省电力有限公司、吉林烟草工业有限责任公司、北京卓识网安技术股份有限公司、腾讯云计算（北京）有限责任公司、江苏保旺达软件技术有限公司

本文件主要起草人：李雪莹、寇增杰、晋钢、万可、包英明、张静、唐刚、张德馨、李琨、安健、杨志、刘思思、李媛、祝英杰、丁雪、刘长风、张昊、杨侨思、刘韧、石爱民、王学清、袁立、李滨、任萌、乐元、姬生利、刘险峰、钟丹晔

数据分类分级产品技术要求

1 范围

本文件规定了数据分类分级产品的技术要求，包括功能要求、性能要求、自身安全要求和安全保障要求。

本文件适用于指导组织开展数据分类分级产品的设计和开发，也适用于指导第三方对数据分类分级产品进行采购和选型。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 25069-2022 信息安全技术 术语

3 术语和定义

GB/T 25069-2022界定的以及下列术语和定义适用于本文件。

3.1 数据 data

任何以电子方式对信息的记录。

[来源：GB/T 43697-2024 ,3.1]

3.2 数据分类 data classification

根据数据的属性或特征，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，以便更好的管理和使用数据的过程。

注：数据分类的对象通常是数据项、数据集。数据项是数据库表的某一列字段。数据集是由多个数据项组成的集合，如数据库表、数据文件等。

3.3 数据分级 data grading

依据数据的重要性、影响程度等将数据划分为不同级别，以便对不同级别的数据实行有针对性的保护。

3.4 元数据 metadata

描述数据属性的信息，用来支持如指示存储位置、历史数据、资源查找、文件记录等功能。

4 缩略语

下列缩略语适用于本文件。

FTP：文件传输协议（File Transfer Protocol）

IP：网际协议（Internet Protocol）

NFS：网络文件系统（Network File System）
SFTP：安全文件传送协议（Secret File System）

5 数据分类分级产品技术要求框架

数据分类分级产品技术要求框架如图1所示，其中，功能要求包括数据源探测与管理、数据资产发现与梳理、分类分级策略管理、分类分级结果管理方面的要求；性能要求包括数据源识别准确率、数据分类识别准确率、数据分类分级任务吞吐率方面的要求；自身安全要求包括用户管理、身份鉴别、安全审计、系统安全、通信安全和数据安全方面的要求；安全保障要求包括供应链安全、设计与开发、生产和交付、运维服务保障、用户信息保护方面的要求。

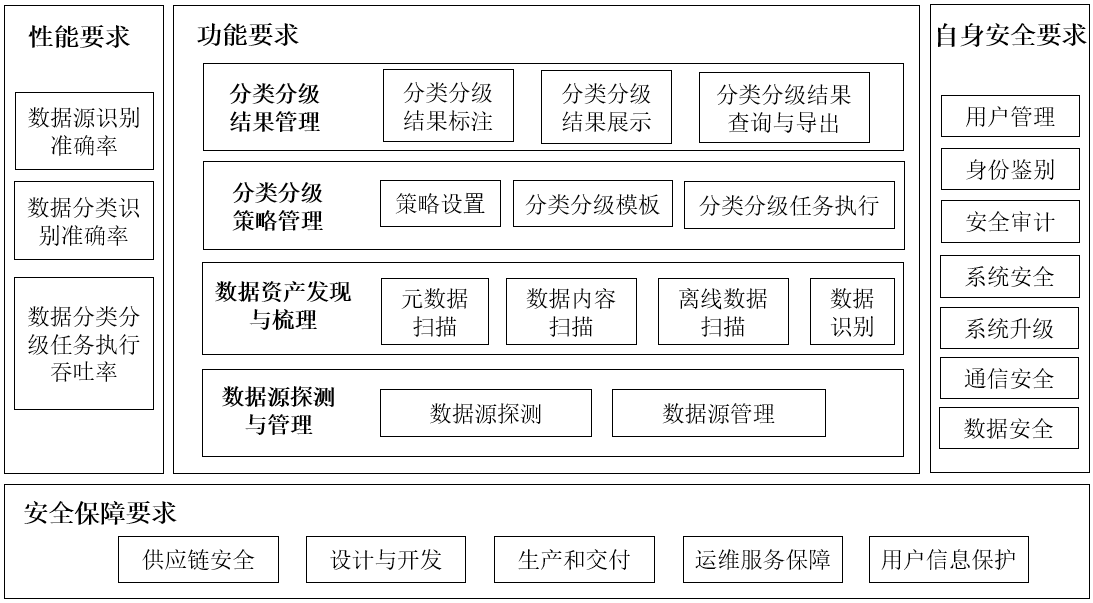

图1 数据分类分级产品技术要求框架

6 功能要求

6.1 数据源探测与管理

6.1.1 数据源探测

数据分类分级产品应支持多种类型的数据源接入，广泛的适配各种数据源，从而对数据分布、规模、种类进行全面的掌握。数据源发现与识别功能包括：

a) 应支持向给定的 IP 范围地址段/端口/端口范围发送探测消息进行数据资产的自动探测，且自动探测应支持周期性自动触发、事件触发、手工触发等启动方式；
b) 应支持动态数据源（如 HTTP 接口）的发现，并正确识别动态数据源地址等数据源信息；
c) 应支持关系型数据库（如，Oracle、MySQL 等）的发现，并正确识别该数据库的 IP 地址、端口、版本信息等数据源信息；
d) 应支持非关系性数据库（如，MongoDB、Elastic Search、Hive 等）的发现，并正确识别该数据库的 IP 地址、端口、版本信息等数据源信息；
e) 应支持文件存储系统（例如 FTP 服务器、SFTP 服务器、NFS 服务器等）的发现，并正确识别该文件存储系统的 IP 地址、端口、版本信息等数据源信息。

6.1.2 数据源管理

数据分类分级产品应支持手工录入或批量导入数据源信息，并支持将数据源识别的结果进行展示，展示形式宜为列表或图标方式。

6.2 数据资产发现与梳理

6.2.1 元数据扫描

数据分类分级产品应支持数据结构扫描，可以扫描到库、表、列等维度的数据结构，最小细粒度到字段。

6.2.2 数据内容扫描

数据分类分级产品应支持数据内容扫描，即数据库表结构扫描和数据扫描，扫描表结构的同时扫描字段的数据内容，可根据内容确定字段属性。

6.2.3 离线数据扫描

针对离线数据扫描的技术要求包括：

a) 应支持数据库离线文件或本地文件数据扫描，以识别数据资产信息；
b) 应支持上传文件进行元数据扫描。

6.2.4 数据识别

针对不同类型的数据源进行扫描识别的技术要求包括：

a) 在识别结构化数据时，应能支持识别元数据、数据内容，以及元数据和数据内容的组合识别，并根据数据特点采用相适应的算法，如匹配算法、模型算法等；
b) 在识别非结构化数据时，应根据数据特点至少支持以下几种格式的数据：txt、xml、Excel、mp3、avi 等；
c) 数据识别的执行方式应采用以下一种或几种：定时全量执行（采用定时任务的方式，对全量数据进行识别）、增量执行（对相较上次数据识别任务执行时增加或变化的数据，进行数据识别）等。

6.3 分类分级策略管理

6.3.1 策略设置

为实现数据分类分级策略的快速定义，产品应具备：

a) 支持批量导入数据分类分级策略；
b) 支持手动调整数据分类分级策略，包括增加、删除和修改策略。

6.3.2 分类分级模板

数据分类分级产品宜内置多种数据分类分级模板，数据分类分级模板宜参考GB/T 43697-2024，用户也可根据数据属性、数据规模等在模板中进行自定义添加分类分级策略。

6.3.3 分类分级任务执行

数据分类分级产品应根据已设置的数据分类分级策略或选定的数据分类分级模板执行分类分级任务。

6.4 分类分级结果管理

6.4.1 分类分级结果标记

分类分级结果标记功能的技术要求包括：

a) 应遵循数据分类分级策略实现对数据类型和级别的标记；
b) 标记过程不应对数据源数据造成修改或删除等破坏性影响；
c) 标记结果应支持人工修改。

6.4.2 分类分级结果展示

分类分级结果展示功能的技术要求包括：

a) 应提供多维度地数据资产统计分析和图形化、图像化呈现，实现对数据资产按照类别、级别、系统等维度进行可视化展示；
b) 敏感数据资产呈现出分类分级属性。可支持基于类别、级别的统计、查询，并以柱图、饼图等形式进行展示；
c) 应支持单个或多个数据源汇总的展示。结果展示内容包括但不限于数据源名称、数据源地址、数据库名、表名、字段名、文件类型、文件名、分级、所属分类等信息。

6.4.3 分类分级结果查询与导出

分类分级结果查询与导出的技术要求包括：
a) 应支持分类分级结果的可视化查询，宜支持匹配样本查询；
b) 宜支持导出查询结果，导出文件的格式宜是以下一种或多种：Excel、TXT、压缩文件等。

7 性能要求

7.1 数据源识别准确率

数据分类分级产品在对数据源进行扫描探测的过程中，数据源的识别正确率应不低于90%。

7.2 数据分类识别准确率

数据分类分级产品在对已识别的数据源进行数据分类时，所能准确识别出的字段数目占总字段数目的比例应不低于95%。

7.3 数据分类分级任务吞吐率

数据分类分级产品根据策略执行数据分类分级任务时，单个任务在单位时间内处理的字段数量应满足其声称值。

8 自身安全要求

8.1 用户管理

数据分类分级产品应支持对不同用户角色进行权限管理，内置系统管理员、安全管理员、审计管理员3种角色。基于三权分立原则，对特权用户进行有效的权限分离。
a) 安全管理员进行数据源管理、策略管理、分类分级任务配置等；
b) 系统管理员进行账户、角色管理和系统安全配置等；
c) 审计管理员进行各类操作日志信息审计等。

8.2 身份鉴别

数据分类分级产品应对用户身份进行标识与鉴别，技术要求包括：
a) 应对登录用户进行身份标识和鉴别，且保证用户标识唯一；
b) 应在通信过程中对登录的敏感信息进行加密传输；
c) 应支持双因素验证策略；
d) 应定期提示用户更换登录口令，并要求口令具有一定复杂度；
e) 应对用户身份鉴别信息进行安全保护，保障用户鉴别信息存储的保密性。

8.3 安全审计

数据分类分级产品应具备对自身进行安全审计的能力，包括：
a) 应对系统运行状况、用户行为，包括账号登录、分类分级策略添加、修改、删除、查看以及分类分级结果导出、查询、修改等操作、事件的时间等内容进行日志记录，生成审计日志；
b) 应提供机制保护审计日志不被非授权访问和修改；
c) 审计日志保存时间应不少于 6 个月。

8.4 系统安全

数据分类分级产品不应存在已公开的中、高风险漏洞。

8.5 系统升级

数据分类分级产品应具备保障升级安全机制，避免得到错误的、伪造的升级包和补丁程序。

8.6 通信安全

数据分类分级产品应保障系统在远程管理过程中的通信传输安全。

8.7 数据安全

数据分类分级产品应对已识别和导入的数据进行保密性和完整性保护。

9 安全保障要求

9.1 供应链安全

数据分类分级产品开发者应满足以下安全保障要求：

a） 制定供应商选择、评定和日常管理的程序，对供应商的开发环境、规范和人员、开发工具、安全测试和安全验证机制等提出管理要求，以确保其提供的关键部件满足安全要求，并保存对供应商选择、评价和日常管理的记录；

b） 建立供应链各环节核心要素的追溯能力，保障核心要素供应稳定；

注：核心要素包括核心技术知识产权、工具及部件等。核心技术知识产权如源代码、软硬件设计图等；工具如开发软件、编译软件、测试软件、测试仪表、管理软件、拷机软件等；部件如硬件机箱、操作系统等。

c） 持续开展安全意识和技能培训。

9.2 设计与开发

数据分类分级产品开发者应满足以下安全保障要求：

a） 识别数据分类分级产品设计和开发环节的安全风险，进行威胁建模，制定安全策略，保障开发环境安全，制定安全开发制度和流程；

注 1：安全开发制度和流程包括但不限于代码编写规范、研发环境安全管理制度、研发人员安全管理制度、研发交付制度等。

b） 制定数据分类分级产品安全功能和自身安全功能的设计文档，该文档描述与安全功能和自身安全功能一致；

c） 为数据分类分级产品确定唯一的版本号，为配置项确定唯一标识，建立并维护配置项列表；

注 2：配置项包括但不限于源代码、工具、文档、组件、配置信息等。

d） 不在产品中设置恶意程序、隐蔽接口或未明示功能模块等，并通过用户协议、产品说明书等途径将所有功能模块、接口等告知用户；

e） 对数据分类分级产品进行安全性测试。

9.3 生产和交付

数据分类分级产品开发者应满足以下安全保障要求：

a） 建立和执行规范的产品完整性检测流程，采取措施防范自制或采购的组件被篡改、伪造等风险；

b） 建立内部交付和外部交付的控制程序，确保数据分类分级产品在交付过程中不被破坏或篡改；

c） 向用户明示包含在产品中的所有功能模块、外部接口和私有协议，告知用户产品中预置的所有账户和默认口令。

9.4 运维服务保障

数据分类分级产品开发者应满足以下安全保障要求：

a） 在法律法规规定或与用户约定的期限内，为产品提供持续的安全维护，不单方面中断或终止安全维护；

b） 保护用户对软件（包含固件）安装和升级等的知情权和选择权，安装和升级软件时明示用户并获得用户同意；

c） 建立和执行针对产品安全缺陷、漏洞的应急响应机制和流程，对发现的产品安全缺陷和漏洞采取修复或替代方案等补救措施，及时告知用户安全风险和可用的补救措施，并向有关主管部门报告。

9.5 用户信息保护

数据分类分级产品开发者应满足以下安全保障要求：

a） 明示收集用户信息的目的、方式、范围、种类、存储位置和处理方式；

b） 建立和执行用户信息管理制度和流程，在产品设计、生产、升级等各阶段保障用户信息的安全，不超范围使用用户信息。

参 考 文 献

[1] GB/T 35273-2020 信息安全技术 个人信息安全规范
[2] GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型
[3] GB/T 38667-2020 信息技术 大数据 数据分类指南
[4] GB 42250-2022 信息安全技术 网络安全专用产品安全技术要求
[5] YD/T 3813-2020 基础电信企业数据分类分级方法
[6] JR/T 0158-2018 证券期货行业分类分级指引
[7] JR/T 0171-2020 个人金融信息保护技术规范
[8] JR/T 0197-2020 金融数据安全数据安全分级指南
[9] JR/T 0218-2021 金融业数据能力建设指引
[10]DB52/T 1123-2016 政府数据分类分级指南
[11]2020年，工业和信息化部办公厅印发《工业数据分类分级指南（试行）》