

T/XJSIA

新疆维吾尔自治区软件行业协会团体标准

T/XJSIA 023—2024

碳达峰碳中和 大数据系统 第3部分：数据治理与服务技术要求

Peak Carbon and Carbon Neutral Big Data Systems
Part3: Data Governance and Services Technical Requirements

2024 - 12 - 13 发布

2024 - 12 - 13 实施

新疆维吾尔自治区软件行业协会 发布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 2

5 总体技术要求 2

6 数据治理技术要求 3

 6.1 数据资源目录 3

 6.2 数据分级分类 3

 6.3 数据血缘管理 3

 6.4 数据模型管理 4

 6.5 数据质量管理 5

 6.6 数据运维管理 5

7 数据服务技术要求 6

 7.1 查询检索 6

 7.2 比对订阅 7

 7.3 数据推送 7

 7.4 模型分析 7

 7.5 数据操作 8

 7.6 数据服务接口 9

参考文献 9

前 言

本文件按照GB/T1.1—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。

《碳达峰碳中和 大数据系统》系列标准分为四个部分：

- 第1部分：技术参考模型；
- 第2部分：数据处理技术要求；
- 第3部分：数据治理和服务技术要求；
- 第4部分：业务应用技术要求。

本文件为第3部分。

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这些专利的责任。

本文件由新疆维吾尔自治区软件行业协会提出并归口。

本文件起草单位：红有软件股份有限公司、新疆云计算应用重点实验室、新疆大学、克拉玛依众城石油装备研究院股份有限公司、新疆维吾尔自治区计量测试研究院、金风科技股份有限公司、克拉玛依碳和网络科技有限公司、新疆天链遥感科技有限公司、广汇能源股份有限公司、新疆维吾尔自治区软件行业协会、新疆维吾尔自治区数字经济联合会、国家计算机网络与信息安全管理中心新疆分中心、新疆丝路六合科技股份有限公司、网新疆电力有限公司信息通信公司、乌鲁木齐大数据产业发展投资有限公司、新疆航天信息有限公司、新疆石文科技发展有限公司、新疆丝路智汇信息科技有限公司、新疆万维网络工程有限公司、新疆永兴同辉信息技术有限公司、新疆云天科技工程技术有限公司。

本文件主要起草人：何芳、陈瑞军、孙素平、唐虎强、刘永杰、刘明艳、张伟、陆勤华、汤杰、秦继伟、石飞、马君刚、刘雪松、冯雪峰、柳青、王涛、谢远东、周斌、于春伟、廖振良、杨晓、董伟伟、刘睿、葛磊、刘靓、吕莲花、曹永光、姚荣斌、张敏、郭江涛、李雅洁、杨柳、赵阳、陈勇、邱兰、魏利红、雷虎、赵欣善、李忠海。

引 言

为了促进大数据技术在碳达峰碳中和领域的广泛应用和创新发展,我们特制定碳达峰碳中和大数据系统系列标准,拟由以下四部分构成:

——第1部分:技术参考模型。提出了碳达峰碳中和大数据系统的技术参考模型。明确了系统的整体结构、数据流向、各部分的功能要求以及各组成部分之间的关系;

——第2部分:数据处理技术要求。规定了碳达峰碳中和大数据系统在数据接入阶段、数据处理阶段以及数据组织存储过程应具备的功能和技术要求;

——第3部分:数据治理和服务技术要求。规定了碳达峰碳中和大数据全生命周期数据治理活动和数据服务活动应具备的功能和技术要求;

——第4部分:业务应用技术要求。规定了碳达峰碳中和大数据系统应具备的业务应用功能。

本文件是拟建系列标准第3部分。旨在通过碳达峰碳中和大数据全生命周期数据治理和数据服务实现数据资源的透明、可管、可控、可用,提升数据资源的管理和利用能力。数据治理涵盖了数据资源目录管理、数据分级分类、数据血缘管理、数据质量管理以及数据运维管理。数据服务则提供了对数据组织及数据治理过程各类数据资源的访问和管理,满足政府决策、企业减排、公众参与等多方面的需求。本部分可与其他部分联合使用。

碳达峰碳中和 大数据系统

第 3 部分：数据治理与服务技术要求

1 范围

本文件规定了碳达峰碳中和大数据系统的数据治理与服务功能要求。数据治理部分详细规定了数据资源全生命周期的管理要求，包括数据资源目录的构建、数据分级分类管理、数据血缘管理、数据模型管理、数据质量管理以及数据运维管理等关键环节。数据服务部分则明确了数据资源查询检索、比对订阅、推送、模型分析以及数据操作、服务接口等具体功能和技术要求。

本文件适用于碳达峰碳中和大数据系统的设计、开发、管理和应用。

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 35295-2017 信息技术 大数据 术语
GB/T 36344-2018 信息技术 数据质量评价指标
GB/T 38667-2020 信息技术 大数据 数据分类指南
GB/T 38672-2020 信息技术 大数据 接口基本要求
GB/T 44109-2024 信息技术 大数据 数据治理实施指南

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据治理 data governance

对数据资源管理行使权力和控制的活动集合（计划、监督和执行）。

[来源：GB/T 44109-2024，术语和定义3.1]

3.2

大数据系统 big data system

实现大数据参考体系结构的全部或部分功能的系统。

[来源：GB/T 35295—2017，定义2.1.14]

3.3

数据资源目录 data resources catalog

与组织业务活动相适应的、便于检索的多层次分类、分级的数据资源编排形式。

[来源：GB/T 42450-2023，术语和定义3.4]

3.4

数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

[来源：GB/T 36344-2018，术语和定义2.3]

3.5

原始数据 raw data

终端用户所存储使用的各种未经过处理或简化的数据。

注：原始数据有多种存在形式，如文本数据，图像数据，音频数据或者几种数据混合存在。

[来源：GB/T 36344-2018，术语和定义2.4]

3.6

日志数据 log data

记录在系统运行中发生的事件的数据。

[来源：GB/T 38672-2020，术语和定义3.3]

4 缩略语

下列缩略语适用于本文件。

API：应用程序接口（Application Programming Interface）

CSV：逗号分隔值（Comma-Separated Values）

AUC：曲线下面积（Area Under the Curve）

5 总体技术要求

依据《碳达峰碳中和 大数据系统 第1部分：技术参考模型》，碳达峰碳中和大数据提供者将新的数据或信息引入大数据系统。数据消费者使用大数据应用提供者提供的的应用。大数据计算框架提供者建立一种计算框架，在此框架中执行数据处理转换以及大数据全生命周期的数据治理和数据服务，以满足系统协调者定义的需求。

在数据提供者、数据消费者、大数据应用提供者、大数据框架提供者之间存在丰富的数据调用需求，需要支持模块间以及与外部系统的信息传递和互操作，数据访问接口通过大数据框架提供者提供的数据组织活动统一提供。

碳达峰碳中和大数据处理技术要求应至少符合GB/T35589-2017、GB/T 38667-2020界定的要求，同时应具备数据治理以及数据服务两方面功能。总体功能架构如图1所示。

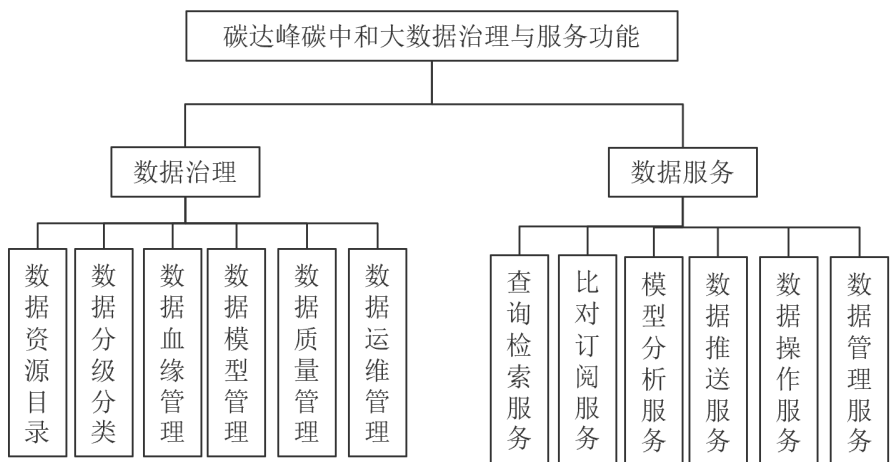


图 1 碳达峰碳中和大数据处理技术功能框架

数据治理与服务总体技术要求包括：

数据治理：系统应具备完善的数据资源目录构建、数据分级分类管理、数据血缘追踪与优化、数据模型管理、数据质量监控与提升以及数据运维管理等能力，共同确保数据资源的透明、可管、可控，并促进数据的流通与价值提炼。

数据服务：系统应提供高效、灵活的数据查询检索、比对订阅、推送、模型分析以及数据操作等功能。这些功能应满足用户对数据的全面、准确、及时获取需求，同时确保数据的安全性与隐私保护。

6 数据治理技术要求

6.1 数据资源目录

6.1.1 目录构建原则与目的

数据资源目录的构建应遵循全面性、准确性、时效性和可追溯性。其目的在于提供一个清晰、有序的数据资源框架，便于数据的管理、查找和使用。通过数据资源目录，用户可以迅速了解系统中存在哪些数据资源，以及这些数据的来源、格式、存储位置、更新频率等关键信息，确保数据的时效性和鲜活度。

6.1.2 目录内容

- 1) 数据资源名称：清晰、准确地描述数据资源的名称，便于用户识别和查找。
- 2) 数据资源描述：对数据资源的详细描述，包括数据的来源、用途、范围等，帮助用户了解数据的背景和用途。
- 3) 数据格式与类型：说明数据的存储格式（如CSV、Excel、JSON等）和数据类型（如数值型、字符型等），确保用户能够正确解析和使用数据。
- 4) 数据存储位置：指出数据在系统中的存储路径或位置，便于用户获取数据。
- 5) 数据更新频率：说明数据的更新周期和时间，确保用户能够获取到最新的数据。
- 6) 数据访问权限：根据数据的敏感度和重要性，设定数据的访问权限，确保数据的安全性和合规性。

6.1.3 目录更新与维护

- 1) 定期更新：根据数据的更新频率，定期对数据资源目录进行更新，确保目录中的数据与系统中的实际数据保持一致。
- 2) 数据审核：在更新数据资源目录之前，对数据进行审核和校验，确保数据的准确性和完整性。
- 3) 权限管理：对数据资源目录的访问权限进行严格管理，确保只有授权用户才能访问和修改目录。
- 4) 文档记录：对数据资源目录的更新和维护过程进行文档记录，便于追踪和审计。

6.2 数据分级分类

6.2.1 数据敏感度与重要性评估

- 1) 敏感度划分：根据碳达峰碳中和业务中数据的敏感程度，将数据划分为不同等级，如公开级、普通级、敏感级、重要级等。公开级数据可对外公开，普通级数据在组织间共享、敏感级仅限组织内部使用，重要级数据需严格控制访问权限。
- 2) 重要性评估：根据数据在碳达峰碳中和业务中的关键作用，评估数据的重要性。关键数据需重点保护，确保数据的完整性和准确性；非关键数据可适当放宽管理要求。

6.2.2 数据分类与标签管理

- 1) 数据分类：根据数据的性质、来源、用途等因素，将数据划分为不同的类别，如能源数据、排放数据、政策数据等。每类数据需明确其特点和应用场景，便于后续的数据管理和使用。
- 2) 标签管理：为每类数据添加相应的标签，如数据类型、数据级别、数据来源等。标签管理有助于快速识别数据的属性和归属，提高数据管理的效率和准确性。同时，标签管理也为数据的安全访问和合规使用提供有力支持。

6.3 数据血缘管理

6.3.1 数据血缘追踪与记录

1) 数据生成路径：详细追踪数据从接入、处理、存储、治理到消费服务的整个生命周期，记录每个数据项的产生源头、转换过程及最终去向，确保数据的可追溯性。

2) 数据依赖关系分析：分析数据之间的依赖关系，包括数据项之间的父子关系、关联关系等，为数据变更影响分析和数据治理提供依据。

6.3.2 数据血缘维护与优化

1) 血缘图谱更新：随着业务的发展和数据的增加，定期更新数据血缘图谱，确保血缘信息的准确性和完整性。

2) 血缘优化：基于数据血缘分析，提出数据流程优化建议，减少数据冗余，提高数据处理效率。

6.4 数据模型管理

6.4.1 模型构建

模型构建流程：明确模型构建的步骤和方法，包括数据预处理、特征选择、模型训练等，确保模型构建的科学性和规范性。

6.4.2 训练数据管理

1) 数据质量监控：在模型训练过程中，持续监控数据质量，包括数据的完整性、准确性、一致性等，确保训练数据的可靠性。

2) 数据预处理与增强：对训练数据进行预处理，如缺失值填充、异常值处理、数据归一化等，同时采用数据增强技术提高模型的泛化能力。

3) 数据集划分：合理划分训练集、验证集和测试集，确保模型在训练过程中不会过拟合，同时能够准确评估模型的性能。

6.4.3 算法管理

1) 算法库管理：对预置的通用模型（如自然语言处理、多媒体信息处理等各类模型）以及新构建的模型进行统一注册和发布管理。支持模型共享。支持遵循标准的第三方模型导入；建立和维护算法库，包含用于数据分析的各类算法，如预测模型、分类算法、聚类算法等，确保算法的有效性和适用性。

2) 算法选择与评估：根据项目需求和数据特点，选择合适的算法进行模型构建，并通过交叉验证、A/B测试等方法评估算法性能。

6.4.4 模型评价

1) 评价指标体系：建立模型评价指标体系，包括准确率、召回率、F1分数、AUC值等，全面评估模型的性能。

2) 模型验证与测试：通过独立的数据集对模型进行验证和测试，确保模型在实际应用中的稳定性和可靠性。

6.4.5 模型优化与迭代

1) 模型迭代：根据业务需求和数据分析结果，不断优化和迭代模型，提高模型的预测精度和泛化能力。

2) 模型调优策略：采用网格搜索、随机搜索、贝叶斯优化等调优方法，对模型参数进行调整，以提高模型的预测精度和泛化能力。

6.4.6 模型版本管理

1) 版本记录与追踪：记录数据模型的版本信息，包括版本号、发布时间、变更内容等，便于追踪和回溯。

2) 版本切换与兼容：提供版本切换功能，确保在模型升级过程中不影响现有业务的正常运行，同时保证新旧版本的兼容性。

6.5 数据质量管理

6.5.1 质量检测数据采集

1) 数据源监控：实时或定期从各个数据源采集数据，包括数据库、文件、API接口等，确保数据的全面性和及时性。

2) 数据样本抽取：在数据采集过程中，根据预设的规则和策略，抽取具有代表性的数据样本，用于后续的质量检测和分析。

6.5.2 多维度质量检核

1) 完整性检核：检查数据是否完整，包括字段是否齐全、记录是否缺失等，确保数据的全面性和完整性。

2) 准确性检核：通过对比、校验等方法，检查数据的准确性，如数值范围是否合理、日期格式是否正确、数据间关系是否一致等。

3) 一致性检核：检查不同数据源或不同时间段的数据是否一致，避免数据冲突和矛盾。

4) 时效性检核：评估数据的更新频率和时效性，确保数据能够及时反映实际情况。

5) 可解释性检核：对于机器学习模型输出的数据，检查其解释性是否足够清晰，能否为业务决策提供有力支持。

6.5.3 问题发现及跟踪

1) 质量检测报告：根据质量检测的结果，生成质量检测报告，详细列出存在的问题、问题的严重程度、影响范围等，为后续的问题处理提供依据。

2) 问题跟踪系统：建立问题跟踪系统，记录问题的发现时间、处理状态、责任人等信息，确保问题能够得到及时跟踪和处理。

3) 闭环处理流程：对于发现的问题，制定解决方案并实施，同时验证解决方案的有效性，确保问题得到彻底解决。在问题解决后，更新质量检测标准和规范，防止类似问题再次发生。

6.6 数据运维管理

6.6.1 运维规则管理

1) 规则定义与更新：根据业务需求和技术规范，定义数据接入、处理、组织、治理和服务等任务的运维规则，包括数据处理流程、数据存储结构、服务调用接口等，并定期更新以适应业务变化和技术升级。

2) 阈值设定：为各项任务设定性能、质量和安全等方面的阈值，如处理时间、错误率、资源占用率等，作为监控和告警的依据。

3) 策略配置：根据规则和阈值，配置自动化运维策略，如自动重启失败的任务、自动调整资源分配等，以提高运维效率和响应速度。

6.6.2 运行状态采集和监控

1) 实时监控：通过数据采集工具和技术，实时采集各项任务的运行状态信息，包括任务进度、资源使用情况、处理结果等。

2) 可视化展示：将采集到的运行状态信息以图表、仪表盘等形式进行可视化展示，便于运维人员直观了解系统运行状态。

3) 异常检测：通过算法和模型对运行状态信息进行分析，及时发现异常状态，如任务失败、性能下降等。

6.6.3 数据运维报表

1) 定期生成报表：根据运维需求，定期生成数据运维报表，包括任务运行统计、资源使用情况分析、错误日志数据汇总等。

2) 报表分析：对报表中的数据进行分析，识别潜在的问题和趋势，为优化运维策略提供依据。

3) 报表分享：将报表分享给相关业务部门和领导，增强他们对数据运维工作的了解和支持。

6.6.4 告警管理

1) 告警触发：当检测到异常状态时，根据预设的告警规则触发告警，通过邮件、短信、即时通讯等方式通知相关人员。

2) 告警处理：运维人员接收告警后，及时分析原因并采取措施进行处理，确保任务恢复正常运行。

3) 告警记录：记录告警的触发时间、处理过程、处理结果等信息，便于后续分析和总结。

6.6.5 运维日志审计

1) 日志数据收集：全面收集各项任务的运行日志，包括系统日志、应用日志、错误日志等。

2) 日志数据分析：对收集到的日志进行分析，识别潜在的问题和风险，为优化系统性能和提升服务质量提供依据。

3) 日志数据审计：定期对日志进行审计，检查运维人员是否按照规范进行操作，确保运维工作的合规性和安全性。

7 数据服务技术要求

7.1 查询检索

7.1.1 数据资源情况查询检索

1) 全面覆盖：提供对业务相关的原始库、资源库、主题库、业务库、知识库以及数据资源目录的全面查询检索，确保用户能够访问到所有相关数据资源。

2) 分类查询：支持按照数据类型（如环境监测数据、能源消耗数据、碳排放数据等）、地域（如国家、地区、城市等）、时间（如年度、季度、月度等）以及行业（如能源、交通、建筑等）进行分类查询，帮助用户快速定位所需数据。

3) 元数据检索：提供对数据资源的元数据（如数据描述、来源、更新时间等）的查询检索功能，帮助用户了解数据的基本情况和质量。

7.1.2 结构化和非结构化数据查询检索

1) 结构化数据查询：支持对业务中的结构化数据（如数据库中的表格数据）进行精确查询、模糊查询、组合查询和批量查询。精确查询要求能够准确返回用户指定的数据；模糊查询允许用户输入关键词或模糊匹配条件，返回相关数据；组合查询允许用户将多个查询条件进行逻辑组合，实现复杂查询；批量查询支持一次性查询多个数据项。

2) 非结构化数据查询：提供对非结构化数据（如文本、图像、视频等）的查询检索功能。对于文本数据，支持关键词检索、全文检索等；对于图像和视频数据，支持基于内容的检索（如颜色、形状、纹理等特征检索）。

7.1.3 查询结果处理

1) 数据统计汇总信息：提供对查询结果的统计汇总功能，如数据总量、平均值、最大值、最小值等，帮助用户快速了解数据概况。

2) 判定查询关键词是否命中：对于用户输入的查询关键词或实体，提供是否命中（存在）的判定信息，确保用户能够准确了解查询结果。

3) 数据摘要或明细信息：根据用户需求，提供数据的摘要信息（如数据概览、主要指标等）或明细信息（如详细数据记录、原始数据等），满足不同层次的数据需求。

7.1.4 性能与优化

1) 查询效率：确保查询检索服务的响应速度在合理范围内，满足用户对数据实时性的需求。

2) 负载均衡：在查询请求量较大时，通过负载均衡技术确保系统的稳定运行和高效查询。

3) 可扩展性：支持查询检索服务的水平扩展和垂直扩展，以适应业务数据的不断增长和查询需求的不断变化。

7.2 比对订阅

7.2.1 数据源整合与差异监测

1) 即时数据校验：系统能即时对比来自不同数据源的同类数据（例如碳排放、能耗、可再生能源产量），迅速识别并标记数据间的差异与异常。

2) 历史趋势追踪：提供深度历史数据比对服务，用户可回溯过往数据变动，洞悉数据长期趋势，为策略规划提供坚实依据。

7.2.2 个性化订阅与警报机制

1) 定制订阅：用户可按自身需求，设定数据比对项目（如时间范围、地理区域、行业分类），并配置通知偏好。

2) 智能预警系统：一旦数据比对结果触及预设阈值或条件，系统将自动触发警报，通过邮件、短信或应用内消息即时通知用户。

3) 通知内容个性化：用户可选择接收详尽的比对报告、简明摘要或仅关注异常通知，确保信息接收既高效又精准。

7.2.3 数据洞察与可视化呈现

1) 可视化工具：集成先进的可视化工具，将复杂数据比对结果转化为直观图表、地图，助力用户快速把握数据动态。

2) 专业解读：结合行业知识与业务需求，提供数据解读服务，揭示数据背后的深层含义，为决策提供智力支持。

7.3 数据推送

7.3.1 实时数据更新与推送

即时性保障：确保业务关键数据（如碳排放、能耗、政策动态）的实时更新与推送，让用户始终掌握最新信息。

推送频率灵活设置：用户可根据实际需求，自定义数据推送频率，从实时更新到定期汇总，满足不同场景需求。

7.3.2 内容与格式定制服务

内容精选：用户可自主挑选推送内容，涵盖数据更新、分析报告、预警信息等，确保信息接收的相关性。

格式多样化：提供邮件、短信、应用内通知及API接口等多种推送方式，满足用户在不同场景下的信息接收偏好。

7.3.3 智能解读与可视化推送

可视化嵌入：在推送的信息中直接嵌入图表、地图等可视化元素，使数据更加直观易懂。

深度解读与建议：基于数据背后的业务逻辑与行业知识，提供深度解读与行动建议，助力用户做出更加精准的决策。。

7.4 模型分析

7.4.1 数据统计与分析

1) 多维度统计：支持对业务数据进行多维度统计，包括时间维度（如年度、季度、月度）、地域维度（如国家、地区、城市）、行业维度（如能源、交通、建筑）等，以揭示数据的分布特征和变化趋势。

2) 深度分析：运用统计学方法和数据挖掘技术，对数据进行深度分析，发现数据之间的关联性和潜在规律，为业务决策提供依据。

7.4.2 规律性探索

1) 时间序列分析：利用时间序列分析方法，对业务数据的时间序列特征进行建模和预测，揭示数据随时间变化的规律和趋势。

2) 周期性分析：识别数据中的周期性成分，如季节性变化、周期性波动等，为业务规划和策略调整提供参考。

7.4.3 预测模型

1) 碳排放预测：建立碳排放预测模型，基于历史数据和当前趋势，预测未来碳排放量的变化情况，为碳中和目标的实现提供数据支持。

2) 能源消耗预测：构建能源消耗预测模型，分析能源消耗的影响因素，预测未来能源消耗量的变化趋势，为节能减排策略的制定提供依据。

3) 政策效果评估：利用模型分析政策实施对业务的影响，评估政策的可行性和有效性，为政策制定和调整提供参考。

7.4.4 结果可视化与解读

1) 可视化展示：提供直观、易用的数据可视化工具，将模型分析结果以图表、地图等形式展示给用户，帮助用户快速理解分析结果。

2) 结果解读：结合业务背景和专业知识，对模型分析结果进行解读和分析，提供有价值的洞察和建议，帮助用户做出更明智的决策。

7.5 数据操作

7.5.1 数据增加操作

1) 数据录入：支持用户或系统批量导入数据，如碳排放量、能源消耗量、减排项目信息等。

2) 数据校验机制：在数据录入过程中，实施严格的数据校验机制，确保数据的准确性、完整性和一致性。

3) 数据归档：对新增加的数据进行归档处理，确保数据的可追溯性和可查询性。

7.5.2 数据删除操作

1) 安全删除机制：提供安全的数据删除机制，确保在删除数据时不会破坏数据的完整性和关联性。

2) 数据备份：在删除数据前，自动进行数据备份，以防误删或数据丢失。

3) 删除日志记录：记录数据删除操作的日志，以便在需要进行数据恢复或审计。

7.5.3 数据修改操作

1) 数据更新：提供便捷的数据更新功能，支持用户或系统对数据进行实时或批量更新。

2) 版本控制：对数据修改进行版本控制，确保每次修改都有记录可查，便于数据回溯和审计。

3) 冲突检测与解决：在数据修改过程中，实施冲突检测机制，及时发现并解决数据冲突问题。

7.5.4 数据操作权限管理

1) 权限分配：根据用户角色和业务需求，合理分配数据操作权限，确保数据的安全性和合规性。

2) 权限监控：对数据操作权限进行实时监控，及时发现并处理权限滥用或异常操作行为。

7.5.5 数据操作性能优化

1) 高效算法：采用高效的数据操作算法，提高数据增加、删除、修改等操作的执行效率。

2) 负载均衡：在数据操作过程中，实施负载均衡策略，确保系统在高并发情况下依然能够稳定运行。

7.6 数据服务接口

7.6.1 服务接口要求

- 1) 接口功能：按需将数据治理和数据服务的能力进行接口封装，为内部模块调用、内部其他子系统调用，以及外部应用提供服务。
- 2) 高度集成与协同：服务接口应能够无缝集成7.1至7.6章节所述的各项功能，实现数据在各模块间的流畅传递与高效协同。
- 3) 灵活扩展与可配置：随着业务需求的不断变化和技术的发展，服务接口应具备灵活扩展和可配置的能力，以满足未来的需求变化。
- 4) 高性能与低延迟：针对大规模数据处理和实时分析的需求，服务接口应提供高性能和低延迟的服务，确保数据的及时性和准确性。
- 5) 安全性与隐私保护：在提供数据服务的同时，服务接口应严格遵守数据安全与隐私保护的相关规定，确保数据的安全性和用户的隐私。

7.6.2 数据接口类型

- 1) 数据采集接口：支持从多种数据源（如物联网设备、业务系统、第三方平台等）采集数据，确保数据的全面性和准确性。提供数据清洗、转换、校验等功能，确保数据的质量和一致性。
- 2) 数据存储接口：支持分布式存储、关系型数据库、非关系型数据库等多种存储方式，满足不同类型数据的存储需求。提供数据备份、恢复、容灾等功能，确保数据的可靠性和安全性。
- 3) 数据处理接口：支持数据清洗、转换、整合、聚合等多种数据处理操作，满足复杂的数据处理需求。提供流处理、批处理等多种处理模式，支持实时和离线数据处理。
- 4) 数据分析接口：提供丰富的数据分析功能，如趋势分析、对比分析、预测分析等，支持用户自定义分析模型和算法。支持多维分析、数据挖掘、机器学习等多种分析技术，提供可视化分析结果。
- 5) 数据共享接口：提供API接口、数据服务总线等多种共享方式，满足多样化的数据共享需求。
- 6) 安全管理接口：提供用户认证、权限管理、数据加密、日志审计等安全管理功能，确保数据的安全性和用户的隐私。支持与第三方安全系统（如防火墙、入侵检测系统、数据加密系统等）的集成和协同。

7.6.3 技术实现与要求

- 1) 高性能与可扩展性：采用分布式架构、负载均衡、缓存等技术手段，提高服务接口的性能和可扩展性。同时，支持自动扩容和缩容，以应对业务量的变化。
- 2) 安全性与隐私保护：采用HTTPS协议、身份验证、权限控制、数据加密等技术手段，确保数据在传输和存储过程中的安全性。同时，遵守相关法律法规和隐私政策，保护用户的隐私和数据安全。
- 3) 高可用性与容错性：采用容错机制、故障转移、数据备份等技术手段，确保服务接口的高可用性和容错性。在发生故障时，能够自动切换至备用节点或进行故障恢复，确保服务的连续性和稳定性。
- 4) 监控与运维：建立完善的监控和运维体系，对服务接口的性能、可用性、安全性等方面进行实时监控和预警。同时，提供日志分析、故障排查、性能优化等功能，便于运维人员及时发现和解决问题。
- 5) 接口文档与测试：提供详尽的接口文档和测试用例，方便开发者和用户理解和使用服务接口。同时，采用自动化测试框架和持续集成技术，对服务接口进行充分的测试和验证，确保其正确性和稳定性。

参 考 文 献

- [1]GB/T5271.17-2010 信息技术词汇第17部分：数据库
- [2]GB/T 42450-2023 信息技术 大数据 数据资源规划
- [3]GB/T36073-2018 数据管理能力成熟度评估模型
- [4]GB/T 38633-2020 信息技术 大数据 系统运维和管理功能要求

[5]GB/T 44109-2024 信息技术 大数据 数据治理实施指南

[6]GB/T32151（所有部分）温室气体排放核算与报告要求

[7]GB/T33760 基于项目的温室气体减排量评估技术规范通用要求

[8]科技部等九部门：科技部等九部门关于印发《科技支撑碳达峰碳中和实施方案（2022—2030 年）》的通知：国科发社〔2022〕157 号
