

团 体 标 准

T/GDIOT 004-2024

城域窄带物联感知平台数据管理标准

The Data Management Standard for Metropolitan Area Narrowband
Internet of Things Perception Platform

2024-04-12 发布

2024-04-12 实施

广东省物联网协会 发布

目 次

前 言 2

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 数据管理架构 2

6 数据管理总体要求 3

7 数据汇聚 4

8 数据存储 5

9 数据维护 7

前 言

本文件依据 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。

本文件由广东省物联网协会归口。

本标准起草单位：广州技象科技有限公司、海珠区政务服务数据管理局、海珠区数字政府运营中心、智慧城市物联网国家重点实验室（澳门大学）、中山大学、香港科技大学（广州）、技象科技（南京）有限公司、航科广软（广州）数字科技有限公司、中电科普天科技股份有限公司、广州智慧城市投资运营有限公司、中国铁塔股份有限公司广东省分公司、科学城（广州）信息科技集团有限公司、广州智投能源科技有限公司。（排名不分先后）

本标准主要起草人：吴瑶、闵铁锐、刘丹丹、刘毅、黄津、王惠杰、孔冬冬、张嘉诚、陶勇、张艺辉、马少丹、夏明华、江明、崔颖、巩紫君、张宇、梁辉、莫峰华、柳谦、杨翼、涂梓珺、周硕源、林亚波、熊旭东、张锐群、林海。

城域窄带物联感知平台数据管理标准

1 范围

本文件提出了城域窄带物联感知平台（以下简称“平台”）数据管理整体框架，给出了面向城域窄带物联网感知终端设备数据采集、存储、治理和应用的各项技术要求，使其他系统或部门的异地、异构数据库的数据可以进行汇聚整合，形成更全面的数据库来支撑物联感知数据的挖掘、分析及共享。

本文件适用于城域窄带物联感知平台关于城域窄带物联感知终端的数据采集和存储系统的设计、开发、测试、评估、运维、管理等。

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 20273-2019 信息安全技术 数据库管理系统安全技术要求》

《GB/T 33474-2016 物联网参考体系结构》

《GBT 35295-2017 信息技术 大数据 术语》

《GB/T 38637.2-2020 物联网 感知控制设备接入》第 2 部分：数据管理要求

《GB/T 52711-2000 信息技术 词汇》第 1 部分：基本术语

《GB/T 33745-2017 物联网 术语》定义 2.5.4

《GB/T37025-2018 信息安全技术 物联网数据传输安全技术要求》

《T/GDIOT005-2022 城域窄带物联感知平台软件参考架构》

3 术语和定义

下列术语和定义适用于本文件。

3.1 数据管理 data management

在数据处理系统中，提供对数据的访问，执行或监视数据的存储，以及控制输入输出操作等功能。

[来源：GB/T 5271.1-2000. 定义 01.08.02]

3.2 数据治理 data governance

对数据进行处置、格式化和规范化的过程。

注 1：数据治理是数据和数据系统管理的基本要素；

注 2：数据治理涉及数据全生存周期管理，无论数据是处于静态、动态、未完成状态还是交易状态。

[来源：GB/T 35295-2017，定义 2.1.43]

3.3 感知数据 sensing data

通过数据采集获取的原始数据或在此基础上进行加工处理的表征对象信息的数据统称。

[来源：GB/T 33474-2016，定义 3.11]

3.4 结构化数据 structured data

一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

[来源：GB/T 38637.2-2020，定义 3.3]

3.5 非结构化数据 unstructured data

不具有预定义模型或未以预定义方式组织的数据。

[来源：GB/T 35295-2017，定义 2.1.25]

3.6 脏数据 dirty data

指不符合质量标准、不准确、不完整、重复或过时的数据。

3.7 数据分析 data analysis

为提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。

[来源：GB/T 33745-2017，定义 2.5.4]

4 缩略语

下列缩略语适用于本标准：

SDK：Software Development Kit，软件开发工具包

API：Application Programming Interface，应用程序编程接口

ODS：Operational Data Store，操作数据存储

DWD：Data Warehouse Detail，数据仓库明细层

DWS：Data Warehouse Service，数据仓库服务层

ADS：Application Data Store，应用数据存储层

5 数据管理架构

城域窄带物联感知平台数据管理分为数据汇聚、数据存储、数据维护三部分，参考架构如图1所示。

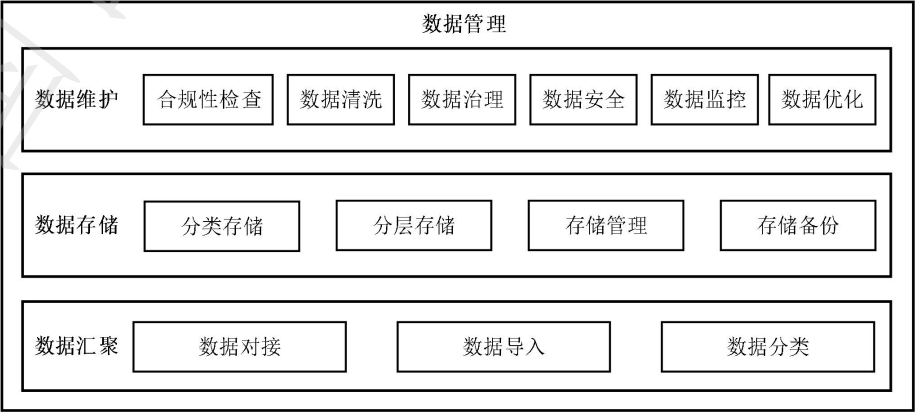


图 1 城域窄带物联感知平台数据管理参考架构

城域窄带物联感知平台数据管理实现对接入到平台的感知数据进行规范化管理，包括数据汇聚、数据存储、数据维护三个部分：

- a) 数据汇聚：实现多数据源的数据汇聚接入，功能包括数据对接、数据导入、数据分类；
- b) 数据存储：应实现对感知数据存储方式的规范管理，包括分类存储、分层存储、存储管理、存储备份等；
- c) 数据维护：应实现对平台数据进行一系列维护活动和程序，确保数据的质量和可信度，包括数据合规性检查、数据清洗、数据治理、数据安全、数据监控、数据优化等管理。

6 数据管理总体要求

6.1 功能要求

数据管理功能要求如下：

- a) 依托子标准《城域窄带物联感知平台数据汇聚接入标准》，实现数据对接、数据导入、数据分类的管理；
- b) 应提供高效、可靠的数据存储解决方案，以保存大量的时序数据；存储功能应支持数据的快速检索、查询和备份，并确保数据的安全性和完整性；
- c) 应对收集的数据进行清洗、转换和预处理，以便进一步的分析和应用；
- d) 应实施强大的安全措施，如加密、访问控制、安全审计和威胁检测，保护数据免受未经授权访问、篡改和泄露；
- e) 应有一套完整的数据治理框架，包括数据质量控制、数据生命周期管理和数据合规性监管等，以确保数据的正确使用和管理；
- f) 应能够灵活扩展，以适应不断增长的数据量和新设备及业务需求；
- g) 应具备数据备份和灾难恢复的能力，以防止数据丢失和系统故障。

6.2 性能要求

数据管理性能要求如下：

- a) 处理速度：数据管理服务应能够快速处理请求和数据，数据管理服务内部数据处理时间不超过3S；
- b) 吞吐量：数据管理服务应具备高吞吐量，能够同时处理大量的数据传输和事务，确保在高峰时段也能保持性能，对于城域级平台，数据管理服务应支持不低于500并发/秒的数据管理能力；
- c) 可伸缩性：随着接入设备数据量的增加，数据管理服务应能够通过增加资源来扩展性能，同时保持稳定性和可用性；
- d) 可靠性：数据管理服务应具有高可靠性，能够持续稳定运行，减少故障和意外停机的时间；
- e) 容错性：在出现故障时，数据管理服务应能够快速恢复并继续运行，保证数据不丢失，服务不中断；
- f) 安全性：在保证性能的同时，数据管理服务还应确保数据安全和隐私保护，防止未经授权访问和数据泄露。对于敏感数据如密码等，需采用加密算法对数据进行加密；

- g) 兼容性：数据管理服务应能够在不同的硬件和软件环境中运行，支持多种操作系统、浏览器和设备，支持在全国产化环境中运行；
- h) 维护性：数据管理服务应易于维护和升级，以便快速部署新功能和性能优化；
- i) 适应性：数据管理服务应能够适应不断变化的业务需求和技术发展，具备灵活的架构和模块化设计；
- j) 监控和诊断：数据管理服务应具备监控工具，以便实时跟踪数据管理性能指标，包括数据存储、数据治理等的响应监控等，以助于及时发现并解决问题。

7 数据汇聚

7.1 概述

数据汇聚旨在通过数据对接、数据分类、数据导入等能力，将其他系统或部门的异地、异构数据库的数据进行汇聚整合，形成更全面的数据库来支撑物联感知数据的挖掘、分析及共享。数据汇聚包括数据对接、数据导入、数据分类，具体要求请见第 7.2~7.4 节。

7.2 数据对接

数据对接是指城域窄带物联感知平台支持外部数据导入并转化为平台规范要求的数据格式的过程，要求如下：

- a) 宜支持窄带物联感知终端设备的直连接入、其它窄带物联感知平台的接入、以国标或省级团体标准规定的执行方式进行接入；
- b) 应支持 SDK、API 等多种应用接入方式；
- c) 应提供一个平台级的数据集成框架，可以实现从其他系统采集数据，对数据进行加工处理成适用的数据格式；
- d) 应支持跨地域、跨部门、跨层级的数据交换技术支撑体系的快速建设；
- e) 为参与数据对接的组织机构及其人员、设施、资源、汇聚模式、汇聚流程、汇聚安全等要素提供运行、管理的技术支撑；
- f) 应支持完全松耦合的数据汇聚架构。

7.3 数据导入

数据导入是将符合平台要求的数据导入平台的过程，要求如下：

- a) 应支持多路并行数据导入的能力；
- b) 应支持数据导入容错，如断点续传；
- c) 应支持接受终端推送数据和平台按频率主动拉取数据两种方式。

7.4 数据分类

数据资源根据业务分类可分为以下四大类：

- a) 设备基础数据：与感知设备本身相关的基础信息，通常包括设备的标识符、配置信息、固件版本等；

- b) 设备监测数据：感知设备直接收集的原始数据，如温度、湿度、光照、压力等；
- c) 设备状态数据：感知设备处于当前状态或环境状态时产生的数据，通常由设备基础数据和设备监测数据经过处理获得，如在线/离线、异常/正常等；
- d) 设备监测事件数据：感知设备或系统中发生的特定事件所记录的数据，如报警、故障等；

在四大类数据基础上，可根据具体业务属性、安全等维度，对数据类型再进行细分。数据分类依据、数据分类方法、数据分类处理流程如下。

7.4.1 数据分类依据

数据分类依据如下：

- a) 根据数据类型进行分类，例如传感器数据、日志数据、用户数据等；以便于确定数据的处理和存储方式；
- b) 根据数据来源设备或系统进行分类，例如来自温度传感器的数据、来自空气质量监测传感器的数据等；以便于追踪数据的起源并进行源头分析；
- c) 根据数据的隐私级别和敏感性进行分类，可以被分类为公开、内部、敏感或机密；以便于实施适当的安全措施和访问控制。

7.4.2 数据分类方法

数据分类方法如下：

- a) 按照层次结构进行分类，例如从大类到子类，以便于组织和管理大量数据；
- b) 按照标签分类，为数据分配标签，以实现灵活的数据分类、检索及权限控制；
- c) 对于数值型数据，可根据数值范围进行分类，例如温度数据可以按照摄氏度范围进行分类；
- d) 根据数据的时间属性，可对数据进行时间序列分类，以便于按时间维度对数据进行分析。

7.4.3 数据分类处理流程

数据分类的处理流程如下：

- a) 在分类之前，需要对数据进行清洗和整理，确保数据质量和数据一致性；
- b) 根据定义的分类规则和方法，将数据分配到相应的类别中；
- c) 对分类后的数据添加明确的标签或标识，便于后续的检索和管理；
- d) 将分类后的数据存储到适当的数据库或数据仓库中，确保数据的安全性和可访问性；
- e) 对分类后的数据进行分析，提取有价值的信息和洞察；
- f) 定期对分类体系进行审查和更新，以适应业务变化和新数据的加入；
- g) 确保分类后的数据符合相关法律法规和行业标准。

8 数据存储

8.1 概述

数据存储需要支持多种存储方式，如根据数据资源分类、分层、实时性和历史性数据的特点采用差异化存储方式，提供数据隔离机制和管理手段，并支持存储备份，防止数据丢失。

8.2 分类存储

分类存储是指依据数据资源的开放类型采用差异化存储方式，要求如下：

- a) 应支持数据隔离机制，使不同数据使用方的数据相互独立不可见；
- b) 政府数据、重要民生数据等应存放于政务云上且满足政务数据不同类型的服务使用需求；
- c) 社会开放数据应存入公有云中，并根据业务对公有数据分级分类确定存储期限。

8.3 分层存储

分层存储是指依据数据的处理阶段和颗粒度对数据进行分层存放，以提高数据处理效率。要求如下：

- a) 城域窄带物联感知平台数据分层应包括 ODS、DWD、DWS、ADS；
- b) ODS 层：作为操作数据存储层，该层用于保存接近实时的源系统数据。ODS层应保持数据的完整性和时效性，支持日常运营分析和报告的需求，宜采用短周期存储方式，以便快速响应业务变化；
- c) DWD 层：作为数据仓库详细层，该层负责对ODS层的数据进行清洗、转换和汇总。DWD层的数据保存周期应根据数据的分析价值和历史数据分析需求来确定，宜保证足够的历史数据深度；
- d) DWS 层：作为数据仓库服务层，该层进一步对DWD层的数据进行聚合和计算，生成更为精细化的业务指标和数据模型。该层数据通常具有较高的业务价值，应根据业务分析和报告需求来调整数据的保存策略，确保数据既能支持长期趋势分析，也能满足快速访问的需求；
- e) ADS 层：作为应用数据服务层，专门针对终端业务应用和用户需求，提供定制化的数据服务和数据产品。考虑到ADS层数据的高度定制化和直接的业务影响，宜采取灵活的数据更新和维护策略，确保数据的实时性和准确性。宜永久保存数据。

8.4 存储管理

存储管理是对存储对象采用差异化存储策略，进而优化存储的过程，要求如下：

- a) 应支持多种存储方式，包含但不限于对象存储、列式存储、云存储、分布式数据库存储；
- b) 应支持将设备上报的数据直接透传给应用处理，或将数据在平台内存储，并提供配置数据存储周期功能；
- c) 应支持时序型数据库存储实时性数据，可用于监测和检查设备所采集的实时数据等；
- d) 应支持关系型数据库存储历史性数据，可用于分析优化生产管理过程等。

8.5 存储备份

存储备份是防止数据丢失，增加文件安全性的方式，要求如下：

- a) 应支持备份进程并行的容错机制，备份进程的宕机不影响数据的可用性；
- b) 应支持数据备份的集中控制，应针对不同使用要求制定数据备份策略；
- c) 应支持镜像备份、冗余备份等方式提高数据存储可靠性；

- d) 应支持备份程序与应用程序的分离；
- e) 应支持对备份数据进行压缩存储。

9 数据维护

9.1 概述

数据维护是一系列活动和程序，旨在确保数据的质量和可信度，以便数据能够有效地支持业务决策和运营。

9.2 合规性检查

合规性检查是对汇聚的数据进行检查，要求如下：

- a) 应对数据进行检查，保证数据的准确性和完整性；
- b) 应提供数据校验功能，如空值检测、长度检查、数值范围检查、正则表达式校验等；
- c) 可依据数据的重要性及敏感程度将数据划分不同的级别，实行分级处理或者按照相应规则进行处理。

9.3 数据清洗

数据清洗是数据准备和预处理过程的一个重要环节，通过技术手段识别并纠正数据错误或不一致等问题，是将初始数据转化为满足数据质量要求或应用要求的过程，以提高数据的质量和准确性，确保数据在后续的分析、决策支持、数据挖掘等处理中的可用性和可靠性。要求如下：

- a) 应支持对错误数据、脏数据、重复数据等异常数据的处理；
- b) 应支持对缺失数据进行填补；
- c) 应支持对数据不一致性进行监测；
- d) 应提供清洗前后的数据比对功能；
- e) 应支持逻辑矛盾、关联性验证、不合理数据的清洗。

9.4 数据治理

数据治理是指对数据质量的管理，要求如下：

- a) 应支持数据的标准化处理，将非标准化终端的数据转化为城域窄带物联感知平台统一格式的数据，包括统一的数据类型标识、统一的数值格式；
- b) 应提供数据从产生到消亡的全周期管理，支持识别、度量、监控、预警、数据管理流程配置和数据治理审核机制；
- c) 应支持数据质量分析和预警，甄别数据优劣程度，对质量较差数据进行清理和长期不用数据进行清洗和归档；
- d) 应支持对数据属性进行标注功能，包括但不限于单个和批量标注、标注维护等；
- e) 应支持敏感数据识别、标记和脱敏，脱敏类型包括但不限于替换、截断、掩码、重排、日期偏移取整等；

- f) 应对处理后的数据进行编码，便于数据管理、数据开放共享及数据应用，数据资源编码结构及规则应按照 DB37/T 3521.1 的规定；
- g) 应支持数据资源目录动态管理，包括但不限于梳理、编制、审核、发布等；
- h) 应制定统一的数据质量管理体系，包括数据源管理、数据质量规则管理、数据质量分析、数据质量评估等，实现对问题数据发现、分析和解决的全流程管理；
- i) 通过数据治理后，平台数据错误率不超过5%、数据完整性不低于95%、一致性错误不超过总记录数的1%。

9.5 数据安全

9.5.1 数据完整性

数据完整性要求如下：

- a) 数据传输过程中的完整性应满足 GB/T 37025-2018 中 6.1和7.1的要求；
- b) 应通过封装签名、测试字验证、引用约束等机制对数据完整性进行检测，保证数据存储过程的完整性，提供非完整数据的解决措施。

9.5.2 数据加密

数据加密性要求如下：

- a) 对于特定的敏感字段或业务数据应使用加密方式存储；
- b) 应支持国家密码管理主管部门批准使用的密码算法，使用国家密码管理主管部门认证核准的密码产品，遵循相关密码国家标准和行业标准；
- c) 应按照国家相关保密部门要求采用合适的加密算法、密钥长度和密钥管理机制；
- d) 数据传输过程中的加密过程应符合 GB/T 37025-2018 中6.3 和7.3 的要求。

9.5.3 数据访问权限控制

数据访问权限控制要求如下：

- a) 应支持制定安全策略，并根据安全策略启用控制用户对数据的访问；
- b) 应实现业务数据、系统数据和数据库系统等不同级别的用户权限分离管理机制，同时对用户访问应分配最小访问权限；
- c) 数据提供给第三方访问时，应严格限制默认账户的访问权限，定期修改账户的口令。

9.5.4 数据备份

数据备份要求如下：

- a) 应制定定期备份计划，确保所有关键数据和系统状态被定期备份。可以采用全备份、增量备份或差异备份的策略；
- b) 应实施多级备份策略，包括本地备份、近线备份和离线备份，以确保在不同级别的故障或灾难中都能恢复数据；
- c) 应对备份数据进行加密，以防止数据在传输过程中或存储时被未经授权访问；

- d) 应定期验证备份数据的完整性和可恢复性，确保在需要时能够成功恢复；
- e) 应合理管理备份存储空间，确保备份数据不会因存储空间不足而失败。

9.5.5 数据恢复

数据恢复要求如下：

- a) 应建立数据恢复计划，以便在数据安全事件发生时迅速采取行动，并进行有效的数据恢复；
- b) 确保恢复的数据保持其原始的完整性和一致性，避免数据恢复后出现不一致或数据丢失的情况；
- c) 在数据恢复过程中，应保护数据不被未经授权访问或泄露；
- d) 应确保有足够的技术支持和专业人员来处理数据恢复过程中可能出现的问题；
- e) 应有更广泛的灾难恢复计划，以应对可能导致数据丢失的各种灾难性事件。

9.5.6 数据可审计性

数据可审计性要求如下：

- a) 审计范围应覆盖业务数据的用户行为，针对数据的重要性设定不同级别的行为记录；
- b) 审计记录应包括事件的日期、时间、类型、主体标识、客体标识和结果等；
- c) 应能够根据记录数据进行分析，并生成审计报表；
- d) 应保护审计记录，避免受到未预期的删除、修改或覆盖等。

10 数据监控

数据监控要求如下：

- a) 应确保能够实时地收集、传输和处理物联感知设备生成的数据；
- b) 应确保在网络故障或其他异常情况下，仍然能够持续地监控和记录物联感知设备产生的数据应验证并记录所有接收到的传感器读数，并确保没有丢失任何重要信息；
- c) 应建立机制来检测异常值、离群点或其他不正常行为，并及时发出警报；
- d) 应定义指标来评估所收集到的物联感知设备生成的数据质量，例如一致性、准确性、完整性、可用性、时效性等；
- e) 应定义指标来评估经过数据治理后的物联感知设备的数据质量，例如一致性、准确性、完整性、可用性、时效性等。

11 数据优化

数据优化要求如下：

- a) 应使用适当存储技术（如数据库系统）对大规模产生并持久保存长期历史信息进行有效管理。同时可以考虑使用压缩算法降低存储空间需求；
- b) 宜对原始传感器读数进行聚合操作（如平均值计算），以减少数据量并提高处理效率。同时，可以使用滤波算法来去除噪声和异常值；
- c) 应对采集到的物联感知数据进行清洗和校正，以消除错误、缺失或不一致的信息，并确保其可靠性；

- d) 对于大规模产生的物联感知数据，在传输过程中应考虑使用有效的压缩方法以降低带宽需求，并优化传输协议。