

团体标准

T/CESA XXXX—202X

物联网操作系统平台运行规范 第2部分：数据处理

Operational specifications for IoT operating system platform

Part 2 data processing

征求意见稿

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

已授权的专利证明材料为专利证书复印件或扉页，已公开但尚未授权的专利申请证明材料为专利公开通知书复印件或扉页，未公开的专利申请的证明材料为专利申请号和申请日期。

202X-XX- XX 发布

202X-XX- XX 实施

中国电子工业标准化技术协会 发布



版权保护文件

版权所有归属于该标准的发布机构，除非有其他规定，否则未经许可，此发行物及其章节不得以其他形式或任何手段进行复制、再版或使用，包括电子版，影印件，或发布在互联网及内部网络等。使用许可可于发布机构获取。

目 次

前 言..... III

引 言..... IV

1 范围..... 1

2 规范性引用文件..... 1

3 术语和定义..... 1

4 总体要求..... 2

4.1 总体目标..... 2

5 数据处理要求..... 2

5.1 数据类型和格式..... 2

5.2 数据质量..... 6

5.3 数据压缩..... 2

5.4 数据格式转换..... 3

5.5 字符编码转换..... 7

5.6 多媒体数据处理..... 7

5.7 数据备份和恢复..... 3

5.8 数据隐私..... 3

5.9 数据安全..... 3

附 录 A..... 4

附 录 B..... 6

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》相关规定起草。

本文件是《物联网操作系统平台运行规范》系列标准的第2部分。

本文件由中国电子工业标准化技术协会归口。

本文件起草单位：。

本文件主要起草人：。



引 言

近年来，我国物联网产业蓬勃发展，其在制造、交通、医疗、民生等领域加速应用，推动各行各业数字化、智能化、绿色化、融合化发展。作为物联网产业的核心基础设施之一，物联网操作系统平台正在加快与产业发展的深度融合。

作为平台运行的基础，物联网操作系统平台运行环境标准化对于推动平台兼容适配和稳定高效运行至关重要。CESA-2024-044 至 CESA-2024-048 标准旨在规范物联网操作系统平台的运行环境，拟由下列五个部分组成：

——第 1 部分：总体要求。目的在于确立物联网操作系统平台运行规范标准化工作的总体框架。

——第 2 部分：数据处理。目的在于确立物联网操作系统平台运行环境应具备的数据处理相关能力与应满足的要求。

——第 3 部分：网络通信。目的在于确立物联网操作系统运行环境应具备的网络通信相关能力与应满足的要求。

——第 4 部分：通用操作系统功能。目的在于确立物联网操作系统平台运行环境应具备的通用操作系统核心功能与程序运行时。

——第 5 部分：IaaS 服务技术。目的在于规范 IaaS 层服务应为物联网操作系统平台提供的功能，以及 IaaS 服务应满足的性能指标与测试要求。

物联网操作系统平台运行规范 第2部分：数据处理

1 范围

本文件规定了物联网操作系统平台（以下简称“平台”）的运行环境应具备的数据处理能力，包括数据格式和类型、数据备份和恢复、数据隐私、数据安全数据质量等要求，以支撑平台数据处理相关需求。

本文件适用于平台的设计者、开发者以及操作系统环境的提供者和在物联网环境中实现数据处理相关功能的相关人员。

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 35273-2020 信息安全技术 个人信息安全规范
- GB/T 41479-2022 信息安全技术 网络数据处理安全要求
- GB/T 37964-2019 信息安全技术 个人信息去标识化指南
- GB/T 37025-2018 信息安全技术 物联网数据传输安全技术要求
- GB/T 28181-2022 公共安全视频监控联网系统信息传输、交换、控制技术要求
- GB/T 29765-2021 信息安全技术 数据备份与恢复产品技术要求与测试评价方法

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[来源：GB/T 33745-2017，2.5.1]

3.2

数据清洗 data cleaning

发现并纠正数据文件中可识别的错误的最后一道程序，包括检查数据一致性，处理无效值和缺失值等。

3.3

压缩算法 compaction algorithm

在电子与通信领域也常被称为信号编码，包括压缩和还原（或解码和编码）两个步骤。

3.4

字符编码 character encoding

将字符集中的字符映射到数字值或二进制序列的方法，以便计算机能够处理和存储文本数据。

3.5

加密算法 encryption algorithm

将明文转换为密文的算法

[来源：GB/T 25069-2022, 3.280]

3.6

元数据 metadata

描述数据及其环境的数据。

[来源：GB/T 33745-2017, 2.5.7]

4 总体要求

4.1 总体目标

平台运行时应提供数据处理相关能力，旨在保证平台数据的完整性、安全性、一致性和可靠性。

5 数据处理要求

5.1 数据类型和格式

平台运行时宜支持附录 B 中 B.1 和 B.2 所提供的常见的数据类型和格式。

5.2 数据质量

平台运行时应提供以下保证数据质量的能力：

a) 数据清洗：

应建立数据清洗机制，包括：

- 1) 识别错误、不一致或不完整的数据：应保证能够及时识别错误、不一致和不完整的数据。宜实现自动化数据验证和检查机制。
- 2) 纠正和删除不良数据：应提供数据纠正功能，自动或者手动修复数据中的错误。应提供对重复、冗余或者过时数据的删除机制。

b) 数据格式标准化：

平台运行时宜实现数据格式标准化，使用规范的字段命名、数据编码规则和单元标准。

c) 数据审计

应建立数据审计机制，记录、监视和审查数据活动，跟踪数据访问、修改该和处理活动，并生成审计日记以供审查。应确保数据的安全性、合规性、可追溯性和完整性

5.3 数据压缩

平台运行环境应满足下列数据压缩要求：

- a) 压缩算法：宜支持附录 B 中 B.3 所提供的常见压缩算法。
- b) 压缩模式：应支持无损压缩和有损压缩。
- c) 数据可用性：应保证压缩前后的数据仍保留有价值的信息，不影响用户对数据的使用。
- d) 高压缩率：宜提高压缩率，减少存储和传输成本。

- e) 数据访问性：应支持快速访问和搜索压缩后的数据。

5.4 数据格式转换

平台运行环境应支持数据格式的转换，具体要求包括：

- a) 数据机密性：在数据格式转换过程中，保护敏感信息的隐私和安全性。
- b) 数据可用性：转换后的数据保持可用性，原始数据和转换后的数据应具有相同的信息价值。
- c) 数据完整性：转换后的数据保证完整性，确保数据在转换过程中不受到篡改，或支持迅速发现篡改。
- d) 数据元数据管理：确保转化过程中元数据的等价性。
- e) 数据处理效率：确保数据转换的效率和速度，减少转换过程对系统性能的影响。
- f) 数据验证：在格式转换后，应进行数据类型检查、范围检查、唯一性检查以确保数据符合目标格式的规则 and 标准。
- g) 兼容性：确保转换后的数据能够与目标系统或应用程序相兼容。

5.5 字符编码转换

平台运行时应支持不同字符编码之间的文本数据转换，确保应用程序能够正确处理和现实来自不同地区和语言的文本，确保文本数据的正确性和一致性。

5.6 多媒体数据处理

平台运行时应支持处理多媒体数据，对多媒体数据的传输、交换和控制符合 GB/T 28181-2022 的要求。

5.7 数据备份和恢复

平台运行时应具有数据备份和数据恢复策略，以应对数据丢失、系统故障和灾难性事件。应符合 GB/T 29765-2021 的数据备份和恢复技术要求及测试方法。

5.8 数据隐私

平台运行时对于包含敏感信息的数据，须遵守数据隐私法规和规定，应符合 GB/T 37964-2019 的要求和 GB/T 41479-2022 中的数据删除和匿名化处理要求，保障数据的隐私。

5.9 数据安全

平台运行时应符合 GB/T 41479-2022 的要求，在处理个人信息时应符合 GB/T 35273-2020 的要求，以防止未经授权的访问、泄露和数据破坏，确保数据的安全可靠。

附 录 A

(规范性附录)

系统环境

A.1 libbz2

系统应提供 libbz2 库，其共享对象名应为 libbz2.so.1。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libbz2 是用于数据压缩和解压缩的 C 语言库，实现了 Bzip2 压缩算法。Bzip2 是一种常用数据压缩算法，用于文件压缩和数据存储中，以减小数据的大小。libbz2 提供用于在应用程序中进行数据压缩和解压缩的函数和工具。

A.2 libexpat

系统应提供 libexpat 库，其共享对象名应为 libexpat.so.1。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libexpat 是用于解析 XML 文档的 C 语言库，其提供了高效、轻量级的 XML 解析功能。

A.3 libiconv

系统应提供 libiconv 库，其共享对象名应为 libiconv.so.2。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libiconv 是用于字符编码转换的 C 语言库，用于处理不同字符编码之间的文本数据转换。开发人员可以使用它来确保应用程序能够正确处理和显示来自不同地区和语言的文本。

A.4 liblzma

系统应提供 liblzma 库，其共享对象名应为 liblzma.so.5。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：liblzma 是用于数据压缩和解压缩的 C 语言库，提供了数据压缩和解压缩的功能，实现了 LZMA 压缩算法。LZMA 是一种高效的压缩算法，能够实现较高的压缩比率。

A.5 libpng16

系统应提供 libpng16 库，其共享对象名应为 libpng16.so.16。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libpng16 是用于处理 PNG 图像的 C 语言库，提供了读取、写入和处理 PNG 图像的功能。

A.6 libxml2

系统应提供 libxml2 库，其共享对象名应为 libxml2.so.2。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libxml2 是用于解析、生成和操作 XML 文档的 C 语言库。它提供了处理 XML 数据的工具，使开发人员能够在各种应用程序中有效地处理和交换结构化数据。

A.7 libz

系统应提供 libz 库，其共享对象名应为 libz.so.1。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libz 是用于数据压缩和解压缩的 C 语言库。它实现了 DEFLATE 压缩算法，这是一种常用的无损数据压缩算法，通常用于文件压缩、网络传输和数据存储中。libz 提供了用于在应用程序中进行

数据压缩和解压缩的函数和工具

A.8 libcharset

系统应提供 libcharset 库，其共享对象名应为 libcharset.so.1。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libcharset 是用于字符集和字符编码处理的 C 语言库，提供了一种可移植的方法来确定文本的字符编码，以适应多种区域设置和编码要求，使跨不同系统的文本处理更加一致和可靠。

A.9 liblber-2.4

系统应提供 liblber-2.4 库，其共享对象名应为 liblber-2.4.so.2。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：liblber-2.4 实现了 ASN.1 基本编码规则的 LDAP 版本，是 OpenLDAP 的一部分。

A.10 libunistring

系统应提供 libunistring 库，其共享对象名应为 libunistring.so.2。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libunistring 是用于处理 Unicode 字符和字符串的 C 语言库。Unicode 是一种国际标准，用于表示世界上几乎所有已知的字符和符号，包括不同语言的字符、表情符号、技术符号等等。libunistring 提供一组在 C 程序中处理和操作 Unicode 的函数和工具，在处理不同字符编码和字符串操作时提供一致性。

A.11 libuuid

系统应提供 libuuid 库，其共享对象名应为 libuuid.so.1。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：libuuid 是用于生成和管理 UUID 的库。UUID 是一种用于标识对象或实体的唯一标识符，通常用于分布式系统、数据库、文件系统和许多其他应用中，以确保标识的全局唯一性。

A.12 Brotli 相关库

系统应提供 Brotli 相关库，包括 libbrotlicommon 和 libbrotlidec，其共享对象名应分别为 libbrotlicommon.so.1 和 libbrotlidec.so.1。函数库的存放位置是实现定义的，但在动态连接器的搜索目录中。

说明：Brotli 相关库提供 Brotli 压缩算法支持。Brotli 算法是谷歌公司开发的一种无损数据压缩算法。

附 录 B

(规范性附录)

常见数据格式和压缩算法

B.1 常见数据类型表

表 1 常见数据类型

日期时间	字符串	布尔值	整数
------	-----	-----	----

B.2 常见数据格式表

表 2 常见序列化数据格式

XML	Protocol Buffers	JSON	MessagePack
-----	------------------	------	-------------

表 3 常见文本文件数据格式

txt	CSV	TSV	INI
-----	-----	-----	-----

B.3 常见压缩算法表

表 4 常见压缩算法表

Bzip2	LZMA	DEFLATE	Brotli
-------	------	---------	--------