

团 体 标 准

T/ZAII 035—2022

数据中台 元数据规范

Data middle platform — Metadata specification

2022 - 06 - 10 发布

2022 - 06 - 10 实施

浙江省物联网产业协会 发布

目 次

前 言 II

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 元模型 2

6 元数据描述 2

 6.1 概述 2

 6.2 描述约定 2

 6.3 技术属性 3

 6.4 业务属性 4

 6.5 操作属性 6

7 元数据扩展 7

 7.1 概述 7

 7.2 扩展类型 7

 7.3 扩展原则 7

 7.4 扩展步骤 7

8 元数据校验 8

 8.1 概述 8

 8.2 校验内容 8

 8.3 校验步骤 9

9 元数据表示 9

附 录 A （资料性） 数据中台元模型示例..... 10

参 考 文 献 11

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。

请注意本文件中的某些内容可能涉及专利。本标准的发布机构不承担识别专利的责任。

本文件由浙江省物联网产业协会提出并归口。

本文件起草单位：杭州玳数科技有限公司、杭州云掣科技有限公司、杭州易知微科技有限公司、杭州云聚数智科技有限公司、浙江省标准化研究院、杭州云象网络技术有限公司、浙江省物联网产业协会、吉利汽车集团有限公司、温州大学、浙江蓝卓工业互联网信息技术有限公司、浙江安厨大数据技术有限公司、杭州中威电子股份有限公司、杭州天舰信息技术股份有限公司。

本文件主要起草人：陈吉平、徐进挺、费翔、宁海元、阚海明、林丹丹、黄昊哲、徐杰、朱旭丽、于俊、黄步添、刘振广、郑金伟、张笑钦、黄辉、邵黎勋、董天祥、曹玉龙、李鹏、纪卫平、罗联上、聂世元、叶颖哲。

数据中台 元数据规范

1 范围

本文件规定了数据中台元数据的元模型、描述方法、扩展要求、校验及表示。
本文件适用于数据中台的数据集编目、建库，以及数据交换、数据集成等。

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 18391.1-2009 信息技术 元数据注册系统（MDR） 第1部分：框架

GB/T 18793 信息技术 可扩展置标语言（XML）1.0

GB/T 36350-2018 信息技术 学习、教育和培训 数字化学习资源语义描述

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据中台 data middle platform

一套通过产品技术、解决方案、规范标准、团队组织的整合，实现数据汇聚、治理、运营的架构。

3.2

元数据 metadata

定义和描述其他数据的数据。

[来源：GB/T 18391.1-2009, 3.2.16]

3.3

数据项 data item

用于描述元数据的基本数据单元

[来源：GB/T 36350-2018, 2.5]

3.4

元模型 metamodel

规定一个或多个其他数据模型的数据模型。

[来源：GB/T 18391.1-2009, 3.2.20]

3.5

属性 attribute

一个对象或实体的特征。

[来源：GB/T 18391.1-2009, 3.1.1]

3.6

技术属性 technical attribute

从源库同步的技术类元数据。

3.7

业务属性 business attribute

通过平台进行维护的业务类元数据。

3.8

操作属性 operational attribute

描述处理和访问数据的细节的元数据。

4 缩略语

下列缩略语适用于本文件。

DDL 数据库模式定义语言 (Data Definition Language)

XML 可扩展标记语言 (Extensible Markup Language)

5 元模型

数据中台元模型由技术属性、业务属性、操作属性构成，可根据需求增加新的属性，模型示例图如图A.1所示。各属性应包含以下数据项：

- a) 技术属性：表名、所属数据库、创建时间、表结构信息、主键、分区、外键；
- b) 业务属性：负责单位、负责部门、负责人、表中文名、主题域、数仓层级、业务域；
- c) 操作属性：创建日期、变更频率、DDL 最后变更时间、储存位置、储存大小、最近同步时间。

6 元数据描述

6.1 概述

本文件采用摘要表示的方式定义和描述元数据，摘要内容包括中文名称、定义、英文名称、数据类型、值域、缩写名、约束/条件、最大出现次数和备注等。

6.2 描述约定

6.2.1 中文名称

元数据的中文名称，用第6.3中各条的标题来表达。

6.2.2 英文名称

元数据的英文名称。所有组成词汇的首字母应大写，且用空格连接。

6.2.3 缩写名

元数据的缩写名。宜采用英文名称的首字母大写连写。

6.2.4 定义

元数据的基本内容、概念和说明。

6.2.5 数据类型

元数据的数据类型，如：字符串、日期型、时间型、布尔型、整型、浮点型等。

6.2.6 值域

元数据可取值的范围。值域中的日期型取值统一为北京时间。

6.2.7 约束/条件

该元数据是否可选。该说明符分别为：

- a) M：必选，表明该元数据应选；
- b) C：一定条件下必选，当满足约束条件中所定义的条件时必选，条件必选用于以下三种可能性之一：
 - 1) 当在多个选项中进行选择时，至少有一个选项为必选，且应使用；
 - 2) 当一个元数据已经使用时，选用另一个元数据；
 - 3) 当一个元数据已经选择了一个特定值时，选用另一个元数据。
- c) 0：可选，根据实际应用可选择也可不选的元数据。已经定义的可选元数据，可指导部门元数据标准制定人员充分说明其信息。

6.2.8 最大出现次数

说明该元数据可以出现的最大次数，只出现一次的用“1”表示，多次重复出现的用“N”表示，允许不为1的固定出现次数用相应的数字表示，例如“2”“3”“4”等。

6.2.9 备注

与元数据有关的注释。

6.3 技术属性

6.3.1 表名

英文名称：Table Name

缩写词：TN

定义：数据库或者数据源的数据表的名称

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：

6.3.2 所属数据库

英文名称：Data Base

缩写词：DB

定义：元数据所描述的数据表所在的原始数据库名称

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：

6.3.3 创建时间

英文名称：Create Time

缩写词：CT

定义：元数据所描述的数据表的创建时间

数据类型：日期型

值域：自由文本

可选/必选：M

最大出现次数：1

备注：按照“yyyy-mm-dd hh:mm:ss”格式表示。

6.3.4 表结构信息

英文名称：Table Structure

缩写词：TS

定义：元数据所描述的数据表的表结构信息，包括：字段名、字段中文名、数据类型、长度、精度、允许空、默认值、字段说明

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：

6.3.5 主键

英文名称: Primary Key

缩写词: PK

定义: 数据表的主键, 能够唯一标识一行数据的主关键字段名称

数据类型: 字符串

值域: 自由文本

可选/必选: M

最大出现次数: 1

备注: 多个主键用半角逗号“, ”分割。

6.3.6 分区

英文名称: Partition Key

缩写词: PTK

定义: 数据表的分区, 用于将数据按照不同的分区维度进行切分管理

数据类型: 字符串

值域: 自由文本

可选/必选: M

最大出现次数: 1

备注: 多个分区用半角逗号“, ”分割。

6.3.7 外键

英文名称: Foreign Key

缩写词: FK

定义: 数据表的外键, 用于指向另一张数据表的主键, 帮助两张表做数据关联

数据类型: 字符串

值域: 自由文本

可选/必选: M

最大出现次数: 1

备注: 多个外键用半角逗号“, ”分割。

6.4 业务属性

6.4.1 负责单位

英文名称: Responsible Company

缩写词: RC

定义: 数据表的责任单位

数据类型: 字符串

值域: 自由文本

可选/必选: M

最大出现次数: 1

备注:

6.4.2 负责部门

英文名称: Responsible Department

缩写词: RD

定义: 数据表的责任部门

数据类型: 字符串

值域: 自由文本

可选/必选: M

最大出现次数: 1

备注：

6.4.3 负责人

英文名称：Table Manager

缩写词：TM

定义：数据表的责任人，责任人一般为该数据表的创建者

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：采用英文描述。

6.4.4 表中文名

英文名称：Table Name Cn

缩写词：TNC

定义：数据表的中文描述名称

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：采用中文描述。

6.4.5 主题域

英文名称：Subject Domain

缩写词：SD

定义：数据表所属的业务主题域，如销售主题域、物流主题域等

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：采用英文描述。

6.4.6 数仓层级

英文名称：Data Warehouse Layer

缩写词：DWL

定义：数据表在数据的分层架构中所属的层级

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：采用英文描述，一般为ODS、DIM、DWD、DWS、ADS。

6.4.7 业务域

英文名称：Business Domain

缩写词：BD

定义：与业务流程相关的数据分域

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：

6.5 操作属性

6.5.1 创建日期

英文名称：Metadata Create Date

缩写词：MCD

定义：元数据的创建日期。

数据类型：日期型

值域：自由文本

可选/必选：M

最大出现次数：1

备注：按照“yyyy-mm-dd”格式表示。

6.5.2 变更频率

英文名称：Change Frequency

缩写词：CF

定义：数据表的数据更新频率，变更频率决定了表的数据时效性

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：一般以秒、分钟、小时、天、周、月为单位。

6.5.3 DDL 最后变更时间

英文名称：DDL Update Time

缩写词：DUT

定义：元数据所描述的数据表的DDL最后更新时间

数据类型：日期型

值域：自由文本

可选/必选：M

最大出现次数：1

备注：按照“yyyy-mm-dd hh:mm:ss”格式表示。

6.5.4 存储位置

英文名称：Storage Location

缩写词：SL

定义：元数据所描述的数据表的访问地址

数据类型：字符串

值域：自由文本

可选/必选：可选

最大出现次数：1

备注：路径用半角正斜杠符号‘/’分割。

6.5.5 存储大小

英文名称：Storage Size

缩写词：SZ

定义：元数据所描述的数据表的物理存储空间大小

数据类型：字符串

值域：自由文本

可选/必选：M

最大出现次数：1

备注：单位为KB、MB、GB、TB等。

6.5.6 最近同步时间

英文名称：Recent Sync Time

缩写词：RST

定义：元数据所描述的数据表的最近一次执行元数据同步的时间

数据类型：日期型

值域：自由文本

可选/必选：可选

最大出现次数：1

备注：按照“yyyy-mm-dd hh:mm:ss”格式表示。

7 元数据扩展

7.1 概述

7.1.1 当本文件规定的元数据不能满足数据实际使用时，可根据实际需求对本文件规定的元数据进行扩展。

7.1.2 元数据的扩展可在元数据的获取过程中，通过对内容进行修改或新增等方式实现。

7.2 扩展类型

数据中台元数据允许进行以下类型的扩展：

- a) 属性扩展：增加元数据的属性；
- b) 内容扩展：对描述元数据的摘要内容进行修改、新增；
- c) 增加新的数据项。

7.3 扩展原则

元数据的扩展应遵循以下基本原则：

- a) 技术属性只作内容扩展，不增加新的数据项；
- b) 业务属性可作内容扩展，亦可增加新的数据项；
- c) 操作属性可作内容扩展，亦可增加新的数据项；
- d) 新增的元数据属性、摘要内容及数据项不应与本文件定义的命名、定义相冲突。

7.4 扩展步骤

7.4.1 对数据中台元数据进行扩展时，主要有：分析现有元数据、定义元数据属性代码表、扩展新元数据属性代码表、定义新元数据属性、扩展新元数据属性、元数据扩展验证、完成元数据扩展文档等7个步骤。

7.4.2 元数据扩展流程如图1所示。

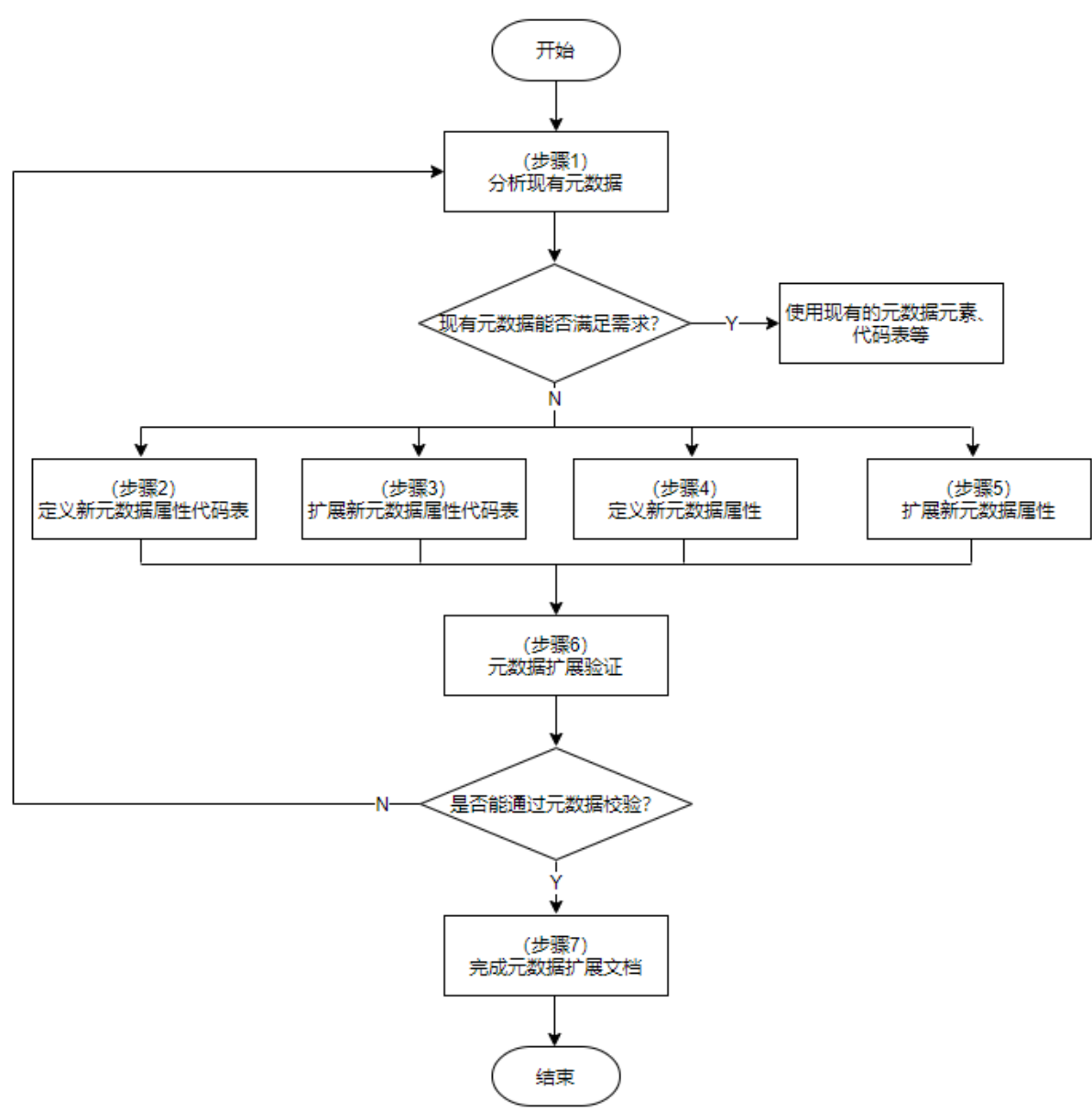


图1 元数据扩展流程图

8 元数据校验

8.1 概述

- 8.1.1 元数据校验包括一致性校验和完整性校验。
- 8.1.2 元数据校验的启动类型包括手动触发式校验和周期性自动校验。
 - 8.1.2.1 手动触发式校验一般指元数据管理员根据需求触发元数据校验流程。
 - 8.1.2.2 周期性自动校验一般指按分钟、小时、天、周、月等周期启动自动校验流程。

8.2 校验内容

8.2.1 一致性校验

校验获取到的元数据的信息是否跟源数据库的信息保持一致。通常包括：
a) 数据是否一致；

- b) 数据类型是否一致;
- c) 数据描述是否一致。

8.2.2 完整性校验

校验元数据的信息是否完整，不存在数据缺失。通常包括：

- a) 属性是够齐全;
- b) 必填信息是否齐全;
- c) 数据是否完整。

8.3 校验步骤

8.3.1 对数据中台元数据进行校验时，主要有：启动元数据校验、元数据一致性校验、元数据完整性校验、校验问题定位修改、完成元数据校验等 5 个步骤。

8.3.2 元数据校验流程如图 2 所示。

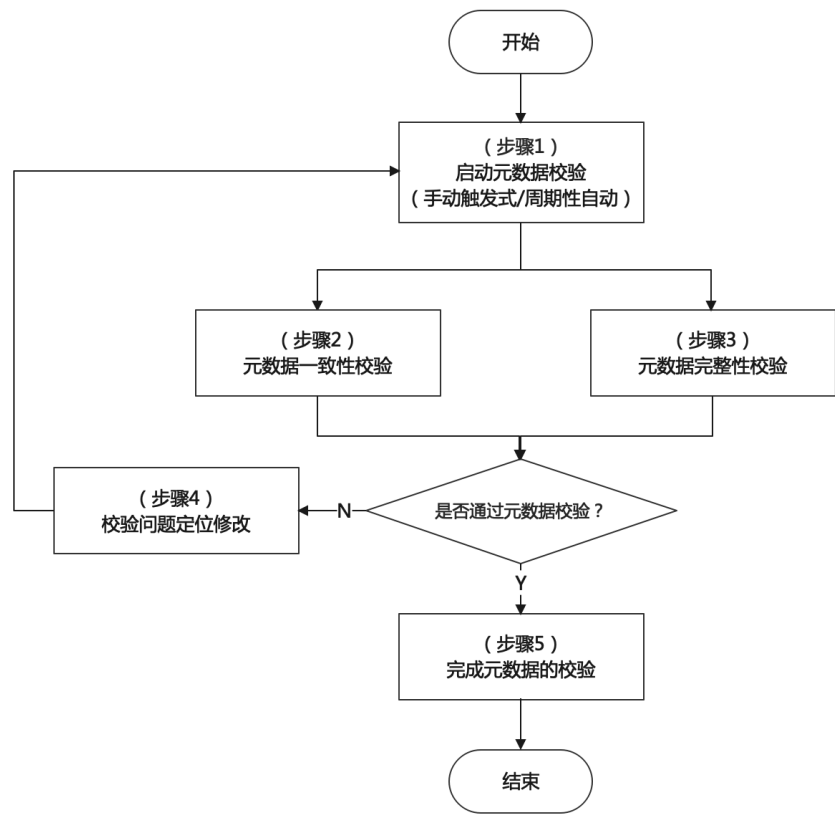


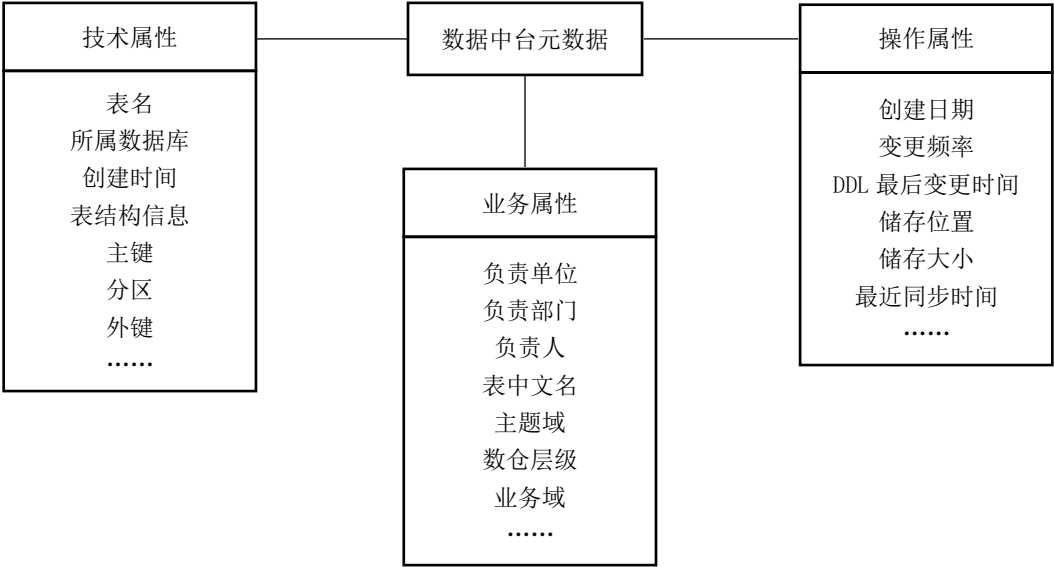
图2 元数据校验流程图

9 元数据表示

数据中台元数据以可扩展置标语言（XML）格式进行表示，以DTD或XML Schema对XML文档的结构进行定义，可扩展置标语言（XML）应遵照GB/T 18793执行。

附录 A
(资料性)
数据中台元模型示例

数据中台元模型示例如图A. 1所示。



图A. 1 数据中台元模型

参 考 文 献

- [1] 项阳. 数据中台元年[J]. 中国教育网络, 2020, 0 (1) :25-26.
 - [2] DAMA国际. DAMA数据管理知识体系指南[M]. 机械工业出版社, 2020.
-