中国物品编码中心

数据模型设计规范

V1.0

**2024年8月**

目录

[中国物品编码中心 0](#_Toc28016)

[1 概述 1](#_Toc30426)

[2 数据模型设计概要 2](#_Toc5383)

[01. 主要内容 2](#_Toc21812)

[02. 主要目的 2](#_Toc9924)

[3 主题域设计 4](#_Toc5436)

[01. 主题域概要 4](#_Toc5129)

[02. 主题域设计规范 4](#_Toc7668)

[4 概念模型设计 6](#_Toc32366)

[01. 概念模型概要 6](#_Toc28358)

[02. 概念模型设计规范 6](#_Toc791)

[03. 概念模型检验标准 7](#_Toc1539)

[5 逻辑模型设计 8](#_Toc11733)

[01. 逻辑模型概要 8](#_Toc5134)

[02. 逻辑模型的要素 8](#_Toc14849)

[03. 逻辑模型设计流程 8](#_Toc10560)

[04. 逻辑模型设计规范 8](#_Toc14193)

[6 物理模型设计 11](#_Toc15237)

[01. 物理模型概要 11](#_Toc7173)

[02. 物理模型设计流程 11](#_Toc2622)

[03. 物理模型设计规范 11](#_Toc20117)

**文档版本历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **版本** | **变更内容** | **变更时间** | **主要参与人员** |
| V1.0 | 初稿 | 2024.08 | 栗许静 |
|  |  |  |  |

# 概述

本文是一体化开发平台建设及管理过程中，应遵循的数据建模原则的具体定义，为数据建模提供指导方针和具体规范，它还是一份规范与实例相结合的实用性参考手册。数据模型设计包括主题域设计、概念模型设计、逻辑模型设计、物理模型设计等几部分，以上部分均需要遵循《数据模型设计规范》。

本指南主要包括以下内容:

（1）主题域设计

（2）概念数据模型设计

（3）逻辑模型设计

（4）物理模型设计

# 数据模型设计概要

## 主要内容

1. 主题域设计：主题域设计是以数据为中心，对所有数据进行分类的过程。
2. 概念模型设计：基于不同主题域，从宏观角度分析和设计核心数据模型的过程。
3. 逻辑模型设计：对概念模型进行细化时，对数据的逻辑结构形态进行定义、编辑和描述的数据模型设计过程。
4. 物理模型设计：将逻辑模型中的实体、属性、关系等转换为表、字段、主外键等物理存储结构的过程。

主题域一般是由总数据架构师来确定，开发一个新系统的时候，建模人员要根据主题域以概念模型、逻辑模型、物理模型的顺序进行设计。一般情况下，不会修改主题域但如果出现原先定义的主题域中未能包含新的业务需求数据时，需要由总数据架构师进行修改。

## 主要目的

数据模型设计规范的目的是使设计的数据模型具有整合性、灵活性、可扩展性、重用性和一致性等特点。以下为各特点的注意事项：

1. 整合性

* 所有的数据都应该从企业级角度进行综合管理。
* 所有的数据都应该在生成时，确认是否存在性质相似的数据。
* 确保在不影响性能的情况下，最大程度地遵循三范式。

1. 灵活性

* 应避免在应用程序中实现复杂业务规则，尽可能将复杂的业务逻辑反映在数据中，从而使数据模型能准确完整地描述业务规则。
* 设计时应尽量避免对属性进行增加或修改的情况。

1. 可扩展性

* 确保数据模型的整合性和灵活性。
* 设计的数据结构便于备份和重构。
* 数据模型的设计应能够满足海量数据处理的需求。

1. 重用性

* 设计时尽量减少新数据的添加，应优先考虑利用现有数据。

1. 一致性

* 应严格遵守业务规则，提供业务上所需要的有效数据。

# 主题域设计

## 主题域概要

在数据模型设计中，主题域的作用是为数据结构提供蓝图，为数据重构与整合提供方向，以及提供数据管理标准。

* 为数据的构成和整合提供基本方向
* 确定数据建模的工作范围、顺序、任务
* 从宏观角度把握和理解企业数据结构
* 为数据分类和数据有效管理提供标准

## 主题域设计规范

1. 主题域分类原则：在对主题域进行分类时，必须遵循一定的原则。以此为基础构建的数据模型才能具有整合性、灵活性和可扩展性。主题域分类原则如下所示:

* 以数据为中心进行分类
* 消除数据重复
* 灵活应对业务需求
* 确保数据的关联性及使用便利性
* 确保主题域之间的均衡性

1. 主题域分类方法

以'基本’、'详细’、'关系’、'管理’、'分类’、'分析’六大结构，对数据集合进行分类。

大主题域

[不同主题域间关系]

[结构分类及关联性]

分类 主题域1

基本 关系 分析 关系

详细 管理 主题域2

主题域的分类

# 概念模型设计

## 概念模型概要

概念模型是通过重要的实体及实体之间的关系，说明业务规则的一种ERD。每个大主题域都要有一个说明大主题域数据结构的概念模型。其目的是从宏观角度了解相关业务的整体结构，从而，只使用了主要的实体、关系、属性以及子类。

1. 概念模型定义

* 从宏观角度分析和设计企业核心数据结构的数据模型。
* 定义核心数据集合及它们之间关系的模型。

1. 概念模型的作用

* 为数据结构设计提供标准、基础、参考信息
* 管理各主题域中主要数据的构成要素
* 设计应用架构时提供参考信息

## 概念模型设计规范

概念模型只对基本信息进行详细描述(例如:重要实体中存储的数据集合是什么，最重要的标识符的继承关系等)，从宏观上定义数据间的业务关系，重点关注重用性和可扩展性。

1. 概念模型构成规范

* 各大主题域都有一个概念模型。
* 由各数据主题域中作为核心的主要实体和实体之间的关系构成表达出能够代表各业务的核心数据。
* 组成的结构要能够灵活应对数据需求变化。

1. 实体设计规范

* 所有实体都应该具有定义和说明，且标识符和主要属性也都应被定义。
* 将仅代表历史信息的实体从概念模型中排除。
* 将编码实体全部删除，将其表示在共同的编码域中。

1. 标识符设计规范：为了理解实体的含义仅描述主要的标识符。
2. 子类设计规范

* 子类的名称使用实体分类名。
* 用子类表达有助于理解实体构成和功能。
* 子类的名称不允许使用‘其它’等字眼。

1. 关系设计规范：概念模型中实体之间可以用M:M关系表示。
2. 属性设计规范

* 若通过标识符就能够明确表达实体含义，则可以不标记其它属性。
* 只使用理解实体含义时需要的主要属性。

1. 数据要素（实体、属性）的命名规范：使用标准化词典中的单词。
2. 数据要素的定义描述规范

* 定义的描述要简洁明了，避免使用含义不清楚的过多修饰语句。
* 使用中文为原则，对于没有对应的中文名称或者是被广泛使用的英文，允许使用英语和外来语。
* 使用陈述型的短语或句子进行表达。
* 使用统一的单词，避免使用相同含义的同义词。

1. 定义数据要素(实体、属性、关系)时的建议

* 准确描述数据要素的含义。
* 必须正确且明确数据含义。
* 能够独立进行表达。
* 避免循环性的推导。

## 概念模型检验标准

|  |  |  |  |
| --- | --- | --- | --- |
| 构成要素 | 检验基准 | 构成要素 | 检验基准 |
| 核心实体 | 集合性 | 核心关系 | 可选性 |
| 识别性 | 关系的度 |
| 永久性 | 业务指向性 |
| 使用性 | 核心属性 | 原子性 |
| 关系性 | 含义性 |

# 逻辑模型设计

## 逻辑模型概要

逻辑模型要以数据为中心进行设计，并且必须在充分理解业务需求的基础上进行设计。设计的逻辑模型要具有扩展性强、一致性高、共享性强等特点。设计逻辑模型不仅仅是为了构建数据库，更重要的是起到对业务进行细致分析、准确定义的作用。

1. 逻辑模型定义

利用实体及相互之间的关系，准确描述业务规则的实体关系图。逻辑模型要为保证业务所需数据结构的正确性及一致性，使用一系列标准的规则将各种对象的特征体现出来，并对各对象之间的关系进行准确定义

1. 逻辑模型设计时的考虑事项

* 不只针对当前业务现状，同时要将业务发展计划包含在内。
* 必须根据已定义好的标准命名方法及域创建方法进行设计，以保证逻辑模型的致性。
* 必须有熟知业务的人员参与建模，将实际业务所需内容充分反映在模型中。
* 必须要考虑设计的逻辑模型在向物理模型转换时具有较高的效率。

## 逻辑模型的要素

逻辑模型中所涉及到的重要要素有实体、标识符、关系、属性。

## 逻辑模型设计流程

逻辑模型整合

属性定义及范式化

实体选定及关系定义

逻辑模型设计流程中包括实体选定及关系定义、属性定义及范式化、数据模型整合三部分内容。

## 逻辑模型设计规范

逻辑模型以业务的详细说明书、当前系统分析结果、需求事项、概念数据模型为依据进行设计。

1. 实体及关系设计规范

* 实体设计规范
* 实体的验证标准：确认是需要管理的对象；确认实体中应存在两个以上的属性及两个以上的实例；确认是实际的数据集合。
* 实体的定义描述标准：实体的名称要能够体现实体特性；将实体是由哪些实例组成的构成层面、实体的功能层面、数据生成规则等内容全方位的进行详细的描述。
* 子类的定义标准：是否能够通过定义子类来明确体现实体的特性(实例的构成或功能)；子类间是否存在明确的不同属性；通过子类与其它实体指定关系，使业务规则变得更加清晰。
* 关系设计规范
* 实体与实体间关系的定义，要求尽量详细化，且严格遵循业务规则。
* 利用合适语言准确描述实体与实体之间关系。
* 关系描述语言尽量使用“拥有”、“所属”、“从属”、“包含”、“被包含”等关键词。
* 标识符设计规范
* 主标识符的设计规范如下：
* 在属性名称的前面加“#”符号，以表示主标识符。
* 主标识符必须是能够确保实体中各个实例唯一性的最小集合单位。
* 构成主标识符的属性必须拥有具体值，不能为空(NULL)值。
* 必须将业务中使用频度相对比较高的属性指定为主标识符。
* 构成主标识符的属性值长度要一致。
* 辅助标识符设计规范如下：
* 在属性名称的前面加“(#)”符号，表示辅助标识。
* 辅助标识符必须是能够唯一识别实体中的一个实例。
* 构成辅助标识符各属性必须拥有具体值，不能为空(NULL)值。
* 必须将业务中使用频度相对比较高的属性指定为辅助标识符。
* 设计数据库时应基于辅助标识符创建唯一索引。

1. 属性设计及范式化规范

* 属性设计规范
* 属性设计时应遵循的四大规范
* 规范一：属性必须是不能再被分解的最小单位集合；
* 规范二：属性必须具有单一值；
* 规范三：避免使用衍生属性；
* 规范四：属性设计尽量详细。
* 属性详细设计规范
* 明确定义属性值是否为空；
* 根据业务规则指定属性默认值；
* 准确定义枚举类型属性值取值范围。
* 范式化规范：遵循不同范式，能够提升数据质量和确保数据模型的合理性。

|  |  |
| --- | --- |
| 范式类型 | 定义 |
| 第一范式 | 数据库表的每一字段都是不可分割的原子数据项 |
| 第二范式 | 须满足第一范式，且要求实体的属性完全依赖于主标识符 |
| 第三范式 | 在第二范式基础上，任何非标识符属性不依赖于其它非标识符属性 |

1. 逻辑模型整合规范

逻辑模型整合是整合相似实体、关系、属性，并解决业务流程不一致的过程。逻辑模型整合的规范如下所示：

* 两个实体的主标识符性质相同
* 两个实体的主标识符相同并且具有相同值域
* 关系整合

1. 实体和属性的标准化规范

实体和属性的中文命名方法相同，应使用《数据标准化词典》来命名。

# 物理模型设计

## 物理模型概要

在物理模型设计时建模人员或数据库设计人员要将逻辑模型按照关系型数据库的结构和特征，转换为关系型数据库所能够支持的物理形式。

1. 物理模型定义

按照一定规则和方法，将逻辑模型中所定义的实体、属性、属性约束关系等要素转换为数据库软件所能够识别的表关系图的一种物理描述。

1. 物理模型设计时的考虑事项

* 要确保业务需求及业务规则所要求的功能得到满足，性能得到保障。
* 物理模型要确保数据的一致性及数据的质量。
* 新业务或新功能增加时能够以较少的改动或不改动就能够满足需求的扩展。

## 物理模型设计流程

反范式化

定义完整性约束

关系表转换

物理数据模型管理

数据库优化

前三部分是物理模型设计流程中必须要经历的阶段，与具体数据库无关，将在本规范中进行详细定义；数据库优化和物理模型管理依赖于具体数据库软件，将不在本规范中进行定义。

## 物理模型设计规范

在实际项目中，通常利用建模工具来实现逻辑模型向物理模型的转换。在建模工具中，只要指定转换规则，基本上都可以实现自动化转换。因此，在没有利用建模工具实现逻辑模型向物理模型转换过程中，本章节中所定义的物理模型设计规范将会有非常大的指定意义。

关系表转换步骤：

* 将实体转换为表
* 将主标识符转换为主键
* 将属性转换为字段
* 为字段指定域名(包括数据类型和长度)
* 转换超类与子类之间的关系

1. 关系表转换规范

* 表转换方法及规范
* 独立实体将被转换为独立的表：主标识符都会被作为表的主键使用。
* 部分依赖实体将会被转换为部分依赖的表：应考虑业务中的数据规则。
* 主键转换方法及规范
* 逻辑模型中每个实体的主标识符应被转换为物理模型中每张表的主键。
* 确保主键能够维持不同表之间的参照完整性。
* 确保主键能够唯一识别表中的每一条记录。
* 确保主键字段的值不能为NULL。
* 主键不允许被变更。
* 字段转换方法及规范
* 必须使用已定义的标准用语和标准域。
* 应按照非空字段、固定长度字段、可变长度字段、允许为空字段先后顺序分配空间。
* 应根据不同的数据库软件特征，选择能够在最大程度上合理利用磁盘空间的字段创建顺序。
* 域应用方法及规范
* 通过分析单词的通用性，为具有相同或相似的属性指定相同的域。

1. 反范式化规范

为了解决性能问题或减少表连接的深度而允许字段或数据行重复的数据模型设计方法。

1. 参照完整性约束定义规范

* 插入参照完整性约束定义方法及规范
* 依赖：参照的父表中主键值已经存在时，才能向子表中插入一条记录。
* 自动：如发现父表中主键值不存在，则数据库会自动在父表中插入一条主键值，然后再完成子表中新记录的插入操作。
* 设置为默认值：如发现父表中主键值不存在，则将父表的主键值设置为父表主键的默认值后，向子表中的插入记录。
* 指定条件：满足用户提前设定的条件时，完成一条新记录的插入操作。
* 设置为空置：向子表中插入一条记录时，参照的父表主键值不存在则对子表中的FK设置为空。
* 修改参照完整性约束定义方法及规范
* 限制：对PK进行修改时，首先检查是否有其它表与该表有继承关系
* 级联：继承该表的所有子表中的FK也会被同时修改；在实际项目中，PK通常是不会被修改的。
* 删除参照完整性约束定义方法及规范
* 限制：检查是否有其它表与该表有继承关系。
* 级联：将表中的记录与参照该条记录的所有子表中的记录同时删除。
* 设置为默认值：将参照该条记录的所有子表中的FK设置为默认值。
* 指定条件：满足用户指定的条件时执行删除操作。
* 设置为空值：将参照该条记录的所有子表中的 FK设置为空值。

1. 表和字段的标准化规范

* 对表命名时，以逻辑模型中相应的实体中文名为依据，关联查找出该用语的英文名，并将其作为该表的表名。
* 对字段命名时，以逻辑模型中相应的属性中文名为依据，关联查找出该用语的英文名，并将其作为该字段的字段名。
* 字段的数据类型和数据长度,通过该字段对应的属性找出相应的域,指定该字段的数据类型和数据长度。

1. 其他设计规范

* 应充分了解和考虑性能要素（性能要素：数据库软件特征、网络环境等）。
* 充分考虑表关联的深度(越深越容易影响性能)对性能的影响。
* 为子类的转换选定合适的方法。
* 对将要存储大量数据的表，设计合理的分区以提升处理效率。
* 以表为单位，通过对全部数据读取类型的分析，设计战略索引。

体系结构设计

数据库与元数据

数据抽取转换

与加载

逻辑模型

设计

中间件开发

数据仓库系统的

开发周期

分析

概念模型

实施

系统的安装

与测试

规范与需求分析

维护

数据填充与调试

用户评价与反馈

系统上线应用

系统维护

面向用户需求

业务需求

反复的逐步设计过程

收集分析确认

常用方法:信息包图法

关键任务:分析和理解数据仓库的主题

概念模型设计

将需求模型转

换为关系模型

常用方法:星型图法

关键任务:事实表和维度表的设计，包括事实、粒度、聚合与分割等问题的确定

逻辑模型设计

由关系模型转

为存储模型

常用方法:关系表(通用)

关键任务:物理数据库表及其存储结构设计

物理模型设计

详细技术细节

