ICS 35.240
CCSL70

DB11

北　京　市　地　方　标　准

DB11/T XXXX—202X

数据资源治理通用技术要求

General technical requirements for data resource governance

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。

XXXX-XX-XX 发布　　XXXX-XX-XX 实施

北京市市场监督管理局　发布

目　　次

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。

本文件由北京市经济和信息化局提出。

本文件由北京市经济和信息化局归口。

本文件由北京市经济和信息化局组织实施。

本文件起草单位： XXXX 。

本文件主要起草人： XXXX 。

数据资源治理通用技术要求

1 范围

本文件规定了数据资源治理所涉及的框架、数据架构、数据标准、数据质量、元数据和数据全生命周期的管理技术要求。

本文件适用于数据要素资源治理的规划、组织和实施。

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 18030 信息技术 中文编码字符集

GB/T 18391.1 信息技术 元数据注册系统(MDR) 第1部分：框架

GB/T 35295 信息技术 大数据 术语

GB/T 36073 数据管理能力成熟度评估模型

3 术语和定义

GB/T 18391.1、GB/T 22240和GB/T 35295界定的以及下列术语和定义适用于本文件。

3.1

数据资源 data resources

具有或预期具有价值的数据集合。

注：数据资源多以电子形式存在。

3.2

元数据 metadata

关于数据或数据元素的数据（可能包括其数据描述），以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

3.3

数据生存周期 data lifecycle

将原始数据转化为可用于行动的知识的一组过程。

4 缩略语

下列缩略语适用于本文件。

CRUD:创建、读取、更新和删除（Creat, Read, Upadte and Delete）

ODS:操作数据存储（Operation Data Store）

DWS:数据仓库服务（Data Warehouse Service）

DIM:维度（dimension）

DWT:数据仓库主题（Data Warehouse Topic）
ADS:应用数据存储（Application Data Store）
SLA:服务级别协议(Service Level Agreement)
NOSQL:非关系型数据库（Not Only SQL）
URL:统一资源定位符(Uniform Resource Locator)
JSON:JavaScript 对象表示法(JavaScript Object Notation)

5 治理框架

数据资源治理框架包括数据架构管理、数据全生命周期管理和数据治理组织构建三部分。

数据架构是数据管理的基础，从业务需求出发，盘点需要接入的数据源并形成数据源清单，做好数据分层和数据分布规划等。

数据全生命周期应包括但不限于：数据接入、数据探查、数据清洗转换、数据整合、数据存储、数据变更、数据运维和数据服务等八个阶段。数据标准管理、数据质量管理、数据安全管理和元数据管理贯穿与数据全生命周期的各个阶段，全过程保障数据质量和安全，如图 1 所示。

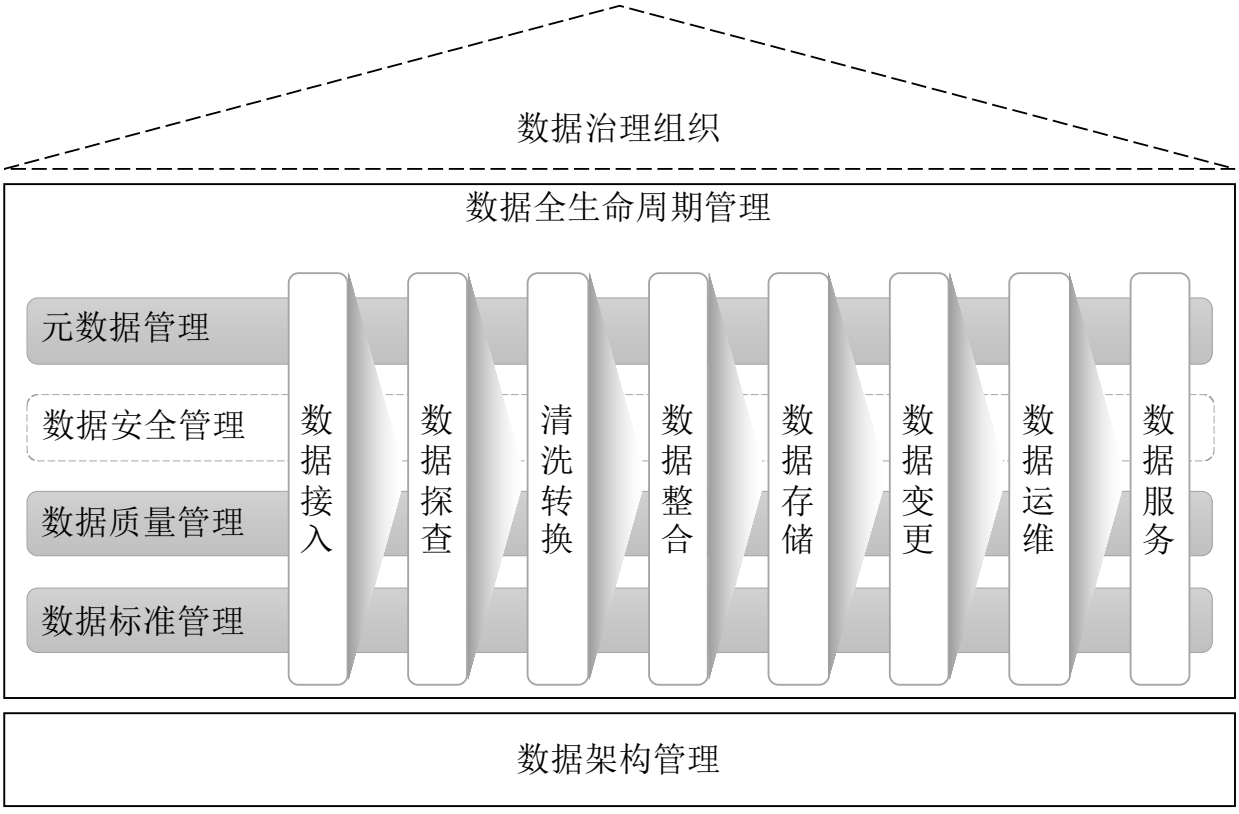

图 1 数据治理技术框架

注 1：数据安全包括数据分类分级、数据访问控制、数据加密和解密、数据备份和恢复、数据传输和传送、敏感数据保护、数据安全监控、数据清除和销毁、安全合作伙伴管理、应急响应和恢复等内容。但本文件对数据安全部分内容不做具体要求，相关要求参见《信息安全技术 大数据安全管理指南》（GB／T 37973-2019 ）、《信息安全技术 云计算服务安全指南》（GB/T 31167-2014 ）、《信息安全技术 大数据服务安全能力要求》（GB/T 35274-2017 ）、《信息安全技术　云计算服务安全能力要求》（GB/T 31168-2014 ）等标准。

注 2：本文件不涉及数据治理组织内容。

6 数据架构管理

6.1 总体要求

数据架构设计时应保证：

a) 明确识别各组织的数据需求，基于数据资源盘点的结果构建数据资源目录设计和维护数据架构；

b) 使用数据架构来指导数据集成和控制，并使数据资源汇聚与需求保持一致；

c) 开展数据模型设计、数据流设计，并明确数据分布情况，管理数据的模型和策略以及规则。

6.2 数据资源盘点

数据源盘点应包括但不限于：

a) 通过梳理本组织数据资源，形成统一标准、统一管理和统一服务的数据源清单，掌握全局数据资源的现状和特点，为数据资源的治理提供基础支撑；

b) 以国家、行业现有的基础信息库为基础，梳理形成业务相关的基础库；将数据资源基础库进行数据目录的分级分类，对应数据资源进行编码和标识。将整理、编码标识后的数据进行数据资源注册、入库等操作。

6.3 数据分布

应针对数据模型中的数据定义，明确数据在组织、流程、系统等方面的分布关系，制定 CRUD 规划，确保数据的安全及权属关系，见 GB/T 36073。如图 2 所示。

主题域	实体类别	实体名称	项目管理	供应商管理	采购管理	物流管理	合同管理	财务MIS	内审内控	综合办公	协同工作	决策支持
采购与供应商	供应商	供应商基本信息	R	R	R	R	R	CRUD			R	R
		潜在供应商信息			CRUD							R
		供应商绩效			CRUD			R				R
		供应商认证信息			CRUD							R
		供应商评级			CRUD			R				R
	合同	合同模板					CRUD					R
		合同基本信息	R		R	R	CRUD	R	R	R		R
		合同审批信息					CRUD					R
		合同立项审核信息	CRUD				R					R
		合同财务审核信息					R	CRUD				R
		合同法律审核信息					R			CRUD		R
		合同违约与争议					CRUD			R		R
		合同变更	R		R	R	CRUD	R	R	R		R
		合同解除	R		R	R	CRUD	R	R	R		R
		合同跟踪信息	R		R		CRUD					R

图 2 CRUD 矩阵示例

6.4 数据分层结构

6.4.1 分层要求

应设计数据结构、减少重复开发、屏蔽源数据的影响等信息，实现数据血缘追踪。数据架构应至少包括以下四层：数据来源层、数据支撑层、数据存储层和数据分析层，如图 3 所示。

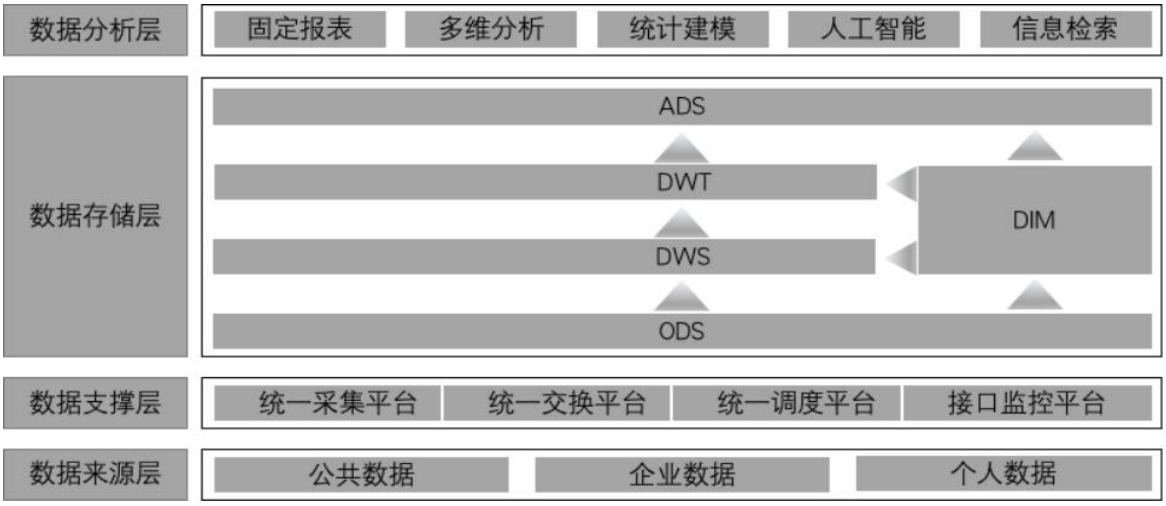

图 3　数据分层结构

6.4.2　数据来源层

本层宜包括：公共数据、企业数据和个人数据，涵盖传统的数据库，数据仓库，分布式数据库，NOSQL 数据库，半结构化数据，无结构化数据，爬虫，日志系统等。

6.4.3　数据支撑层

本层宜包括：数据清洗、数据转换、数据加工、数据关联、数据标注、数据预处理、数据加载、数据抽取等工作。

6.4.4　数据存储层

本层存储经过清洗处理后的可用于生产系统的数据，宜包括元数据，业务数据库，模型数据库等。数据存储宜划分为五层：

a)　ODS：保存最原始数据，按业务概念组织细节数据，并进行名称、代码等标准化处理，保留最完整的历史；
b)　DWS：存储整合后的明细数据，在本层应进行指标与维度的标准化，包括数据清洗，脱敏，维度退化等，保证指标数据的唯一性；
c)　DIM：公共维度表，用于建立一致性维度数据，规范化维度属性，降低数据计算口径和算法不一致风险；
d)　DWT：存储汇总数据，关于各个主题的加工和使用，是共性聚合值；
e)　ADS：面向业务定制的应用数据，根据不同的业务需求采用星型或雪花型模型设计方法构建的数据集市。

6.4.5　数据分析层

本层实现对数据的深加工，宜建立满足需求的数据统计分析模型，建立大数据运行处理平台。

6.5　数据流向

数据流向：

a)　应按照标准的数据流向进行开发，即 DWS －>DWT －> ADS；
b)　不应同层依赖；
c)　ADS 不应直接使用 DWS 的表；
d)　不应出现反向依赖；
e)　应避免数据链路成环。

6.6 主题域划分

应进行数据主题域划分，在较高层次上将数据进行综合、归类和分析利用：
——按照业务或者业务过程划分；
——根据需求方划分；
——按照功能或者应用划分；
——按照部门划分。

7 数据规范管理

7.1 梳理制度资源

梳理制度资源应包括：
a) 通过梳理国家、行业、地方法律法规及标准文献，形成管理标准及资源库；
b) 结合业务需求，确定标准化工作对象，明确标准化具体内容；
c) 整理业务术语、数据标准、代码集、核心元数据。

7.2 业务词汇管理

业务词汇管理包括：
a) 应支持业务词汇表管理权限配置；
b) 应建立通用公开业务词汇表；
c) 业务词汇管理内容应包括：
   1) 标准文档库管理；
   2) 限定词、同义词、术语等信息库管理；
   3) 标准字典管理；
   4) 数据源字典管理。

7.3 参考数据和主数据

参考数据和主数据要求包括：
a) 应能够识别数据值域，识别参考数据和主数据取值范围；
b) 应支持参考数据、主数据和应用系统的集成。

7.4 数据项要求

数据项应包括但不限于：
a) 数据命名；
b) 数据类型；
c) 长度；
d) 业务含义；
e) 计算口径；
f) 归属部门。

8 数据质量管理

8.1 质量核验

8.1.1 质量核验要求

质量核验要求应包括：
a) 明确数据质量监控的数据指标项；

b） 评估使用的数据质量度量维度及其权重值；
c） 定义表示标准质量和质量差数据的值和范围；
d） 对不同的度量规则，执行不同的数据质量评估；
e） 查看并确认数据质量是否可被接受；
f） 在适当数据流转中采取纠正措施；
g） 定期重复上述步骤。

8.1.2 规范性检核内容

规范性检核内容应包括：
a） 数据标准；
b） 数据模型；
c） 元数据；
d） 业务规则；
e） 权威参考数据。
具体检核方法与描述参见附录 A 表 A.1。

8.1.3 完整性检核内容

完整性检核内容应包括：
a） 数据元素完整性；
b） 数据记录完整性。
具体检核方法与描述参见附录 A 表 A.2。

8.1.4 准确性检核内容

准确性检核内容应包括：
a） 数据内容正确性；
b） 数据格式合规性；
c） 数据重复率；
d） 数据唯一性；
e） 脏数据出现率。
具体检核方法与描述参见附录 A 表 A.3。

8.1.5 一致性检核内容

准确性检核内容应包括：
a） 相同数据一致性；
b） 关联数据一致性。
具体检核方法与描述参见附录 A 表 A.4。

8.1.6 时效性检核内容

准确性检核内容应包括：
a） 基于时间段的正确性；
b） 基于时间点及时性；
c） 时序性。
具体检核方法与描述参见附录 A 表 A.5。

8.1.7 可访问性检核内容

可访问性检核内容应包括：

a) 可访问性；
b) 可用性。

具体检核方法与描述参见附录 A 表 A.6。

8.2 质量分析

8.2.1 定性分析

数据质量定性分析可采用第三方评测法、用户反馈法，专家评议等方法。质量子元素评分根据定性评价进行。

8.2.2 定量分析

数据质量定量分析可采用回归分析、因子分析、鱼骨图分析、帕累托分析、矩阵数据分析等方法。

8.2.3 综合分析

宜定性和定量分析相结合的方法对数据质量进行分析。

8.2.4 质量问题原因分类

影响数据质量的问题主要包括技术、业务、管理三个方面，参见附录 B。

9 元数据管理

9.1 需求分析

应明确元数据类型和详细级别，分析内容包括但不限于：

a) 更新频次：元数据属性和属性集更新的频率；
b) 同步情况：数据源头变化后的更新时间；
c) 历史信息：是否需要保留元数据的历史版本；
d) 访问权限：谁可以访问元数据，如何访问；
e) 存储结构：元数据如何通过建模来存储；
f) 集成要求：元数据从不同数据源的整合程度、整合的规则；
g) 运维要求：更新元数据的处理过程和规则（记录日志和提交申请）；
h) 管理要求：管理元数据的角色和职责；
i) 质量要求：元数据的质量需求；
j) 安全要求：元数据的安全需求，是否可以公开等。

9.2 元模型管理

元模型的类型应包括但不限于：

a) 业务类元模型：如指标、KPI、报表等元模型；
b) 技术类元模型：如关系型数据库、OLAP、接口、ETL 等元模型；
c) 管理类元模型：包括系统资源、人员管理、任务管理等元模型。

9.3 编制元数据规范

组织应根据元数据需求分析结果，结合国家标准、行业标准等，形成自己的元数据标准。组织对内的元数据标准包括命名规范、自定义属性、安全、可见性和处理过程文档，组织对外的元数据标准包括数据交换格式、应用程序接口设计等。

9.4 存储

应建立元数据存储库，实现元模型以及元数据的存储，可采用不同的架构方法存储元数据，包括但不限于集中式、分布式、混合式等。

a) 集中式元数据存储由单一的元数据存储库组成，不支持将请求从用户直接传递给各种工具，适用于寻求高度一致性的组织；。
b) 分布式元数据存储架构，元数据应分散存储在各自的源系统中，通过实时从源系统检索数据来响应用户请求；。
c) 混合式架构应结合集中式和分布式架构的特性。

9.5 创建与采集

应基于相对应的元模型，通过自动或手动的方式，获取到组织定义的元模型中所需要管理的元数据信息。自动采集包括但不限于使用适配器、扫描仪、网桥应用程序。

9.6 集成与变更

应对不同类型、不同来源的元数据进行集成，包括从组织外部获取的数据中的元数据，并将技术元数据与相关的业务、流程和管理元数据集成在一起，形成对数据描述的统一视图，并基于规范的流程对元数据的变更进行及时更新和管理。

9.7 应用

应根据组织业务需求实现基于元数据的共享服务与应用，包括但不限于元数据的查询、统计、基于元数据的血缘分析、影响分析等。

9.8 管理机制与评估

应建立元数据管理机制，明确元数据的管理过程及角色、职责；建立元数据管理的质量标准和评估指标，开展元数据绩效评估并持续改进。

10 数据全生命周期管理

10.1 数据接入

10.1.1 接入要求

数据接入应满足以下要求：

a) 数据质量：数据准确、完整、唯一性，避免重复数据和不一致数据，确保数据质量；
b) 数据安全：保障数据的机密性、完整性、可用性，确保数据在传输和存储过程中不被泄露、丢失或被篡改；
c) 数据格式:按照一定的数据模型和数据字典定义数据结构和格式,确保数据的统一性和标准化；
d) 数据加工：在入库前对数据进行清洗、转换、集成等加工处理，以满足数据仓库的需求；
e) 数据可追溯：记录数据来源和处理过程，保留原始数据和处理日志，方便数据审计和追溯；
f) 数据量控制：控制数据入库的频率和数据量，避免过度入库导致数据仓库不稳定或占用过多存储资源。

10.1.2 明确接入数据源

接入源类别包括但不限于：

a) 关系型数据库；
b) 非关系型数据库；
c) 接口服务；
d) 实时数据库日志；
e) 消息队列服务；

f） 文本文件；
g） 压缩包、图片等二进制文件。

10.1.3 制定接入方案

根据明确的数据源类型指定相应接入方案,包括但不限于：
a） 接口推送/拉取数据；
b） 数据库源采集；
c） 消息队列生产/消费数据；
d） FTP 文件推送/拉取。

10.1.4 数据接入格式

数据接入格式：
a） 应支持 JSON、XML、CSV 等数据格式；
b） 应规定数据格式的具体规范，例如字段名称、字段类型、字段长度等。

10.1.5 接口要求

应规定数据接口的要求，例如接口名称、参数、返回值等。

10.1.6 数据接入流程

应制定数据接入流程，明确数据接入的流程和责任。流程应包括数据接入申请、审核、测试、上线等环节。

10.2 数据探查

数据探查应包括单表数据内容分析、多表间数据关系分析的指标定义，准入标准等。

10.3 数据清洗转化

数据清洗应进行以下操作：
a） 非空检核：若字段应为非空时，对字段数据进行非空检核；
b） 主键重复检核：多个业务系统中同类数据经过清洗后，在统一保存时，为保证主键唯一，进行检核工作；
c） 非法代码清洗：对非法代码、代码与数据标准不一致等情况进行校核及修正；
d） 非法值清洗：对取值错误、格式错误、多余字符、乱码等情况进行校核及修正；
e） 数据格式检核：通过属性值的格式检核来衡量数据准确性，包括时间格式、币种格式、多余字符和乱码等；
f） 记录数检核:对各个系统相关数据之间的数据总数检核,或者数据表中每日数据量的波动检核。

10.3.1 数据清洗流程

数据清洗从数据的准确性、完整性、一致性、唯一性、时效性和有效性方面处理数据的缺失值、越界值、不一致代码和重复数据等问题，数据清洗流程如图 4 所示：

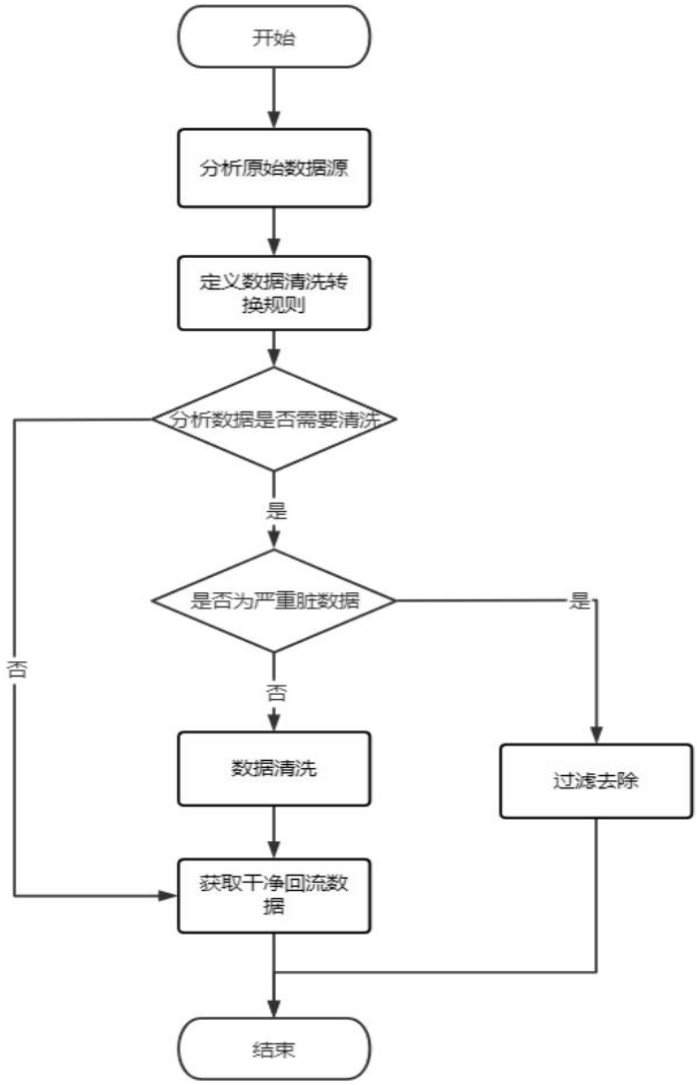

图 4　数据清洗流程

10.3.2　ETL 设计要求

10.3.2.1　ETL 数据映射

应包括源数据集属性、目标数据集属性和 ETL 规则：

a)　源数据集属性和目标数据集属性应包括：
   1)　实体名称；
   2)　字段名称；
   3)　字段简述；
   4)　类型；
   5)　非空属性。

b)　ETL 规则：
   1)　应能够建立数据源过滤规则；
   2)　应描述从源数据集获取数据过程中过滤掉记录的规则；
   3)　应设置关联规则；
   4)　应定义列转换规则；
   5)　应具备目标数据集更新规则；
   6)　应建立 ETL 作业列表。

10.3.2.2 ETL 抽取方式

ETL 抽取方式应包括但不限于：

a) 全量抽取；
b) 增量抽取；
c) 触发器方案；
d) 时间戳方案；
e) 日志方案；
f) 消息队列方案。

10.3.3 ETL 开发要求

10.3.3.1 ETL 开发原则

ETL 开发应遵循以下基本原则：

a) 可观赏性：代码要求结构应清晰、整齐、整体层次分明；
b) 可读性：代码在合适的位置应添加必要的注释；
c) 风格统一性：代码上下文应风格统一，不应多种风格杂糅；
d) 命名规范性：作业命名应清晰易懂，便于更新维护。

10.3.3.2 作业命名

作业命名规则应按顺序至少包括如下三部分：

a) JOBTYPE：作业类型；
b) DESCRIPTION：有效的描述信息；
c) SEQNUM：作业编号，当应拆分作业数据时，作业编号用于区分多次加载数据的情况。

10.3.3.3 编码

编码要求应包括：

a) 使用 with 语句替代嵌套子查询；
b) 不应使用 SELECT * ；
c) SQL 关键字应统一大写或小写，不应大小写混用；
d) 使用缩进，使代码结构化，缩进应默认使用 4 个空格；可将 tab 键设置为 4 个空格；
e) 两表关联时宜使用 JOIN 关键字，不宜使用直连写法。

10.3.3.4 脚本

脚本要求应包括但不限于：

a) ETL 脚本独占一个文件，以.sql 结尾；
b) 在关键逻辑处添加必要注释；
c) 一个脚本文件对应一张数据表的生成；
d) 脚本结尾应增加分号；
e) 若脚本有参数，参数名称宜大写。

10.3.3.5 流程

开发流程要求应包括但不限于：

a) 建立完善的 ETL 开发流程，每个环节都应严格管控。ETL 开发流程应包括但不限于调研应求、评审、开发、测试和上线；
b) 建立 ETL 变更流程；
c) ETL 变更流程与需求变更紧密结合；

d) 修改 ETL 映射文件和业务逻辑文件应从文档开始，并有统一的入口；
e) 修改文档应包括版本号、修改原因描述、修改过程、修改时间和修改影响范围。

10.3.4 ETL 维护要求

10.3.4.1 预警机制

ETL 维护预警机制内容如下：
a) ETL 日志应分为 3 类：
   1) 执行过程日志；
   2) 错误日志；
   3) 总体日志。
b) 警告发送；
c) 重跑机制；
d) 数据一致性稽核；
e) 业务数据稽核。

10.3.4.2 维护管理机制

应建立维护管理机制，包括但不限于：
a) 建立快速访问的远程登录机制；
b) 维护人员被告知错误后，应快速分析定位问题类型并处理；
c) 应日常收集问题日志，定期进行问题复盘；
d) 开发不规范导致的维护问题，应定期进行开发规范培训，视情况加入考核机制。

10.3.5 数据修正处理

10.3.5.1 空值处理

按照缺失比例和字段重要性分别制定处理策略，应包括直接删除和填充内容：
a) 对于无用字段应删除；
b) 对于缺失的内容，应按规则进行填充，填充方法包括：
   1) 以业务知识或经验推测填充缺失值；
   2) 以同一指标的计算结果（均值、中位数、众数等）填充缺失值；
   3) 以不同指标的计算结果填充缺失值。

10.3.5.2 格式内容处理

格式内容处理包括 3 种情况：
a) 应将时间、日期、数值、全半角等显示格式统一；
b) 若有不该存在的字符，可以半自动校验和半人工方式来找出潜在问题，并去除错误的字符；
c) 对于内容与该字段应有内容不符时，应详细识别问题产生原因，不应简单删除。

10.3.5.3 逻辑错误处理

逻辑错误处理应包括去重、去除异常值和修正矛盾内容：
a) 去重处理应在格式内容处理之后；
b) 异常值处理首先应识别异常值，然后由操作人员按照经验和业务流程判断其值的合理性；
c) 修正矛盾内容应通过字段相互验证方式实现。

10.3.5.4 其他情况处理

其他情况包括但不限于：
a) 敏感数据处理；

b) 非需求数据处理；
c) 枚举值处理；
d) 关联性验证。

10.4 数据整合

10.4.1 数据模型设计

10.4.1.1 模型设计要求

模型设计要求应包括逻辑模型与物理模型规范：

a) 逻辑模型要求：
   1) 应具有统一的数据结构、统一的视图，满足第三范式的规范化模型；
   2) 应具有灵活性和可扩展性；
   3) 应确定主题域；
   4) 应确定数据存储规划；
   5) 应定义关系表模式。
b) 物理模型要求：
   1) 应确定数据的存储结构；
   2) 应确定索引策略；
   3) 应确定数据与索引存储位置；
   4) 应确定存储存分配参数；
   5) 模型命名要求；
   6) 模型关系要求；
   7) 字段命名要求；
   8) 约束要求；
   9) 建模要求。

10.4.1.2 命名要求

数据模型设计的命名要求应包括：

a) 通用命名要求

   通用命名应：
   1) 表和字段名应以英文字母开头；
   2) 表和字段长度应不要超过 64 个英文字符；
   3) 表和字段名应使用小写英文单词，命名需满足信息描述的定义；
   4) 表和字段命名应采用下划线来分隔词根，优先使用词根中已有关键字；
   5) 表和字段名禁止使用非标准的缩写，禁止使用 SQL 中的关键字；
   6) 数据域命名要求：应使用与业务系统相关的、常用的命名方式或缩写，如日志域（log）、广告域（adv）、位置域（loc）、用户域（usr）等；
b) 表命名要求

   表命名应使用单数名词、复数名词、前缀等：
   1) 约定缩写；
   2) 常量命名要求；
   3) 文件命名要求；
   4) 规范代码编写习惯。
c) 维度字段命名要求：

   维度字段命名应：
   1) 与原系统业务字段保持一定的关联，根据业务特色沉淀公共命名属性和专有名词；

2) 日期维度按定义的统一分区格式存储；

d) 指标字段命名要求：

指标字段命名应：

1) 指标命名方式可包括业务主题（修饰词）、量化词（词根）、周期限定词等；
2) 基础指标；
3) 普通指标命名要求；
4) 日期类型指标命名要求；
5) 聚合类型指标。

10.4.2 数据模型评审流程

数据模型评审流程包括：

a) 应在 mapping 设计结束之后统一发起模型评审，评审对象包含数据模型与 mapping 设计；
b) 参与数据模型评审的评委成员中，应至少包含一名业务需求人员；
c) 评委成员应建立模型评分规则对模型各项指标评分，总分为满分的模型方可进入开发阶段。

10.4.3 脚本开发要求

数据脚本开发要求应包括但不限于以下内容：

a) 建表要求；
b) 数据格式要求；
c) 作业流要求；
d) 数据字典要求；
e) 维度要求；
f) 指标来源要求；
g) 指标一致性建设要求；
h) 迭代要求；
i) 数据要求；
j) ETL 要求；
k) 代码风格要求；
l) 错误处理要求；
m) 文档要求；
n) 版本控制要求。

10.5 数据存储

10.5.1 数据存储功能要求

数据存储的功能应包括但不限于与：

a) 应支持多服务器分布式集群部署；
b) 平台应提供压缩和解压缩算法；
c) 应提供多租户隔离功能；
d) 应支持数据备份与恢复。

10.5.2 存储治理

数据存储治理应包括但不限于：

a) 数据应设置合理的生命周期；
b) 僵冷数据应进行及时处理，处理方式应包括：

1) EC 转换；
2) 归档保全；

3) 删除销毁。

10.6 数据变更

10.6.1 变更类型

数据资源库的变更类型应包括但不限于：

a) 数据模型变更；
b) 数据记录变更；
c) 数据管理信息变更。

10.6.2 变更审批流程

变更审批流程至少应包括：

a) 应识别变更类型，并限制“不建议变更类型清单”中的数据变更；
b) 应充分识别影响范围，通知受影响内容的负责人并由负责人进行审批；
c) 应定期进行变更审批流程的执行审计和评估，进一步改进和督促数据变更的标准化。

10.6.3 监控与协同处理

应构建数据变更的监控、变更通知能力，完成上下游变更协同：

a) 应主动识别到各类数据变更；
b) 应具备自动变更通知的能力，如通过邮件、短信、即时通讯工具等方式进行数据变更信息的传递；
c) 应建立上下游变更协同机制。

10.7 数据运维

10.7.1 服务等级协议 SLA 管理

服务等级协议 SLA 管理的内容应包括：

a) 建立和维护统一的服务等级协议（SLA），明确故障分级和服务指标；
b) 确保服务等级达到 SLA 的要求，并提供稳定的服务。

10.7.2 建立运维机制

建立运维机制的内容应包括：

a) 制定值班方式、值班流程，并提出明确的值班要求；
b) 运维人员的配备应根据运维管理目的或 SLA 来拟定，运维团队的核心岗位应有人员备份和储藏；
c) 建立问题升级机制，包括问题级别定义和问题升级的条件；
d) 建立复盘机制，定期进行总结。

10.7.3 监控功能要求

监控功能应包括但不限于：

a) 数据库监控：
   1) 基础信息采集；
   2) 活动性能监视；
   3) 系统持续运行的稳定性。
b) 运维平台监控：
   1) 资源使用情况监控；
   2) 作业运行情况监控；
   3) 失败作业监控；
   4) 基线监控；

5） SLA 高级别任务的预警提醒。

10.7.4 异常处理流程

异常处理流程应包括：

a） 确定故障现象并初判问题影响；
b） 建立应急恢复机制；
c） 根据客户授权，处理异常。

10.8 数据服务

10.8.1 API 接口

10.8.1.1 基本要求

数据服务 API 接口的传输协议和消息封装格式基本要求如下：

a） 应提供服务请求成功、失败等各种情况的接口返回状态码；
b） 传输的数据应采用加密和防篡改技术保证数据的完整有效性；
c） 应支持跨操作系统、跨语言平台调用；
d） 宜采用 HTTP/HTTPS 作为传输协议；
e） 接口的请求和返回结果宜采用 JSON 格式封装。

10.8.1.2 命名规则

命名应遵循以下规则：

a） 由半角格式的英文或数字组成；
b） 命名在本规则范围内应唯一；
c） 虚拟名称命名为半角格式的英文无缝连写。

10.8.1.3 基础参数

应包含请求基础参数和返回基础参数，接口参数应遵循以下规则：

a） 各参数命名由半角格式的英文、数字或“_”符号组成；
b） 各参数首词汇采用小写字母；
c） 所有的响应数据编码为国家标准要求格式。

10.8.1.4 接口使用

访问数据接口时，其地址为 URL 格式，URL 地址参数说明：

a） URL 地址中各参数应由半角格式的英文、数字或“_”符号组成；
b） URL 地址中各参数首词汇应采用小写字母；
c） URL 里的所有请求参数名和参数值的数据编码应符合 GB 18030 格式。

10.8.2 库表接口

10.8.2.1 基本要求

通过数据平台进行数据服务库表接口进行数据交换、数据桥接，应确保资源的发布、审核、申请流程畅通。

10.8.2.2 操作流程

库表接口的操作流程应分为数据资源提供方与数据需求方：

a） 数据资源提供方应在数据共享交换平台对数据资源进行注册和发布，内容应包括但不限于：
1） 数据桥接；

2) 数据源创建与注册；
3) 资源注册与发布；
4) 资源下线。

b) 数据需求方在数据共享交换平台进行数据资源的申请和订阅访问，内容应包括但不限于：
1) 资源申请；
2) 申请审核；
3) 资源订阅。

10.8.3 文件接口

10.8.3.1 基本要求

数据服务文件接口应包括数据文件、校验文件：

a) 数据文件是接口单元的实例，每个数据文件应且只对应一个接口单元；
b) 数据文件的校验信息应且只应被其接口单元对应的校验文件所包含。

10.8.3.2 接口协议

数据服务文件接口的协议，应包括但不限：

a) HTTP 下载；
b) FTP 下载。

10.8.3.3 文件命名和格式

数据服务文件接口的命名应遵循以下规则：

a) 应只由半角格式的英文、下划线和数字组成；
b) 不应包括任何对数据内容的描述信息；
c) 一个完整的数据文件命名定义可参考数据文件命名信息表；
d) 校验文件名称可为完整的数据文件命名加“.MD5”后缀。

附 录 A
（规范性）
数据质量检核内容与方法举例

规范性检核内容与方法见表 A.1：

表 A.1 规范性检核内容与方法

质量元素（权重）	质量子元素	质量子元素权重	指标描述	单位成果 分析方法	单位成果质量 子元素分值
规范性（0.2）	数据标准	根据实际需求定义	数据符合数据标准的度量。 注 1：检核数据质量时应收集数据在命名、创建、定义、更新和归档时遵循的标准，包括国际标准、国家标准、行业标准、地方标准或相关规定等。 注 2：和数据归档一样甚至更重要，在一个完整地数据规划中旧数据的销毁一般也有一个比较详细且具有可执行性的规定。	定性分析	符合=100%，不符合=0
	数据模型	根据实际需求定义	数据符合数据模型的度量。 注 1：数据模型是一种直观描述组织数据结构的手段，是数据表达的规范。 注 2：检核数据质量时应检查是否存在清晰可理解的数据模型定义以及这些数据的组织形式。	定量分析	质量子元素分值 ×100% 式中： 满足数据模型要求的数据集中元素的个数； 被检核的数据集中元素的个数。
	元数据	根据实际需求定义	数据符合元数据的度量。 注：元数据标注、描述或刻画其他数据，以使检索、或使用信息更容易，检核数据质量时应检查是否提供可解读的元数据文档。 示例：包含各字段名称、描述、类型值域等内容的数据字典为一种元数据文档。	定量分析	质量子元素分值 ×100% 式中： 满足元数据定义数据项的个数； 元数据项数。

	业务规则	根据实际需求定义	数据符合业务规则的度量。 注1：业务规则是一种权威性原则或指导方针，用来描述业务交互，并建立行动和数据行为结果及完整性的规则。 注2：检核数据质量时应检查是否存在良好归档的业务规则。	定量分析	质量子元素分值 ×100% 式中： 满足业务规则的数据集中元素的个数； 被检核的数据集中元素的个数。
	权威参考数据	根据实际需求定义	参考数据是系统、应用软件、数据库、流程、报告及交易记录和主记录用来参考的数值集合或分类表。 注：检核数据质量时应收集参考数据列表。 示例：一张用于一个特定字段的有效值列表为一种参考数据类型。	定量分析	质量子元素分值 ×100% 式中： 满足参考数据规则的数据集中元素的个数； 被检核的数据集中元素的个数。
	安全规范	根据实际需求定义	安全规范是安全和隐私方面的规则，包括数据权限管理，数据脱敏处理等。	综合分析	符合=100%，不符合=0

完整性检核内容与方法见表A.2：

表 A.2　完整性检核内容与方法

质量元素	质量子元素	权重值	指标描述	单位成果 分析方法	单位成果质量 子元素分值
完整性（0.2）	数据元素完整性	根据实际需求定义	按照业务规则要求，数据集中应被赋值的数据元素的赋值程度。	定量分析	×100% 式中： 被赋值的数据集中元素的个数； 预期被赋值的数据集中元素的个数。
	数据记录完整性	根据实际需求定义	按照业务规划要求，数据集中应被赋值的数据记录的赋值程度。	定量分析	×100% 式中： 被赋值的数据集中元素的个数； 预期被赋值的数据集中元素的个数。

准确性检核内容与方法见表A.3：

表 A.3　准确性检核内容与方法

质量元素	质量子元素	权重值	指标描述	单位成果 分析方法	单位成果质量 子元素分值
准确性(0.3)	数据内容正确性	根据实际需求定义	数据内容是否是预期数据。	定量分析	×100% 式中： 满足数据正确性要求的数据集中元素的个数； 被检核的数据集中元素的个数。
	数据格式合规性	根据实际需求定义	数据格式（包括数据类型、数值范围、数据长度、精度等）是否满足预期要求。 示例：性别一栏不能出现男/女以外的内容；身份证号不能出现标点符号；以及对字符编码的一些限制，都应通过规定内容的格式来实现。	综合分析	×100% 式中： 满足格式要求的数据集中元素的个数； 被检核的数据集中元素的个数。
	数据重复率	根据实际需求定义	特定字段、记录、文件或数据集意外重复的度量。	定量分析	×100% 式中： 重复的数据集中元素的个数； 被检核的数据集中元素的个数。
	数据唯一性	根据实际需求定义	特定字段、记录、文件或数据集唯一性的度量。	综合分析	×100% 式中： 满足唯一性要求的数据集中元素的个数； 被检核的数据集中元素的个数。求的项数。
	脏数据出现率	根据实际需求定义	正确字段、记录、文件或数据集之外无效数据的度量。 示例：事务发生回滚时由于回滚机制不健全或不完善导致可能出现脏数据。	定量分析	×100% 式中： 有脏数据出现的数据集中元素的个数； 被检核的数据集中元素的个数。

一致性检核内容与方法见表 A.4：

表 A.4　一致性检核内容与方法

质量元素	质量子元素	权重值	指标描述	单位成果 分析方法	单位成果质量 子元素分值
一致性(0.2)	相同数据一致性	根据实际需求定义	同一数据在不同位置存储或被不同应用或用户使用时，数据的一致性；数据发生变化时，存储在不同位置的统一数据被同步修改。	定量分析	×100% 式中： 满足相同数据一致性要求的数据集中元素的个数； 被检核的数据集中元素的个数。

	关联数据一致性	根据实际需求定义	根据一致性约束规则检查关联数据的一致性。	定量分析	×100% 式中： 满足关联数据一致性要求的数据集中元素的个数； 被检核的数据集中元素的个数。

时效性检核内容与方法见表 A.5：

表 A.5　时效性检核内容与方法

质量元素	质量子元素	权重值	指标描述	单位成果分析方法	单位成果质量子元素分值
时效性(0.1)	基于时间段的正确性	根据实际需求定义	基于日期范围的记录数或频率分布符合业务应求的程度。	定量分析	×100% 式中： 满有效性要求的数据集中元素的个数； 被检核的数据集中元素的个数。
	基于时间点及时性	根据实际需求定义	基于时间戳的记载数、频率分布或延迟时间符合业务应求的程度。	定量分析	×100% 式中： 满足及时性要求的数据集中元素的个数； 被检核的数据集中元素的个数。
	时序性	根据实际需求定义	数据集中同一对象的数据元素之间的相对时序关系。	定量分析	×100% 式中： 满足时序性要求的数据集中元素的个数； 被检核的数据集中元素的个数。

可访问性检核内容与方法见表 A.6：

表 A.6　可访问性检核内容与方法

质量元素	质量子元素	权重值	指标描述	单位成果分析方法	单位成果质量子元素分值
可访问性(0.1)	可访问性	根据实际需求定义	数据在应时的可获取性。	综合分析	符合=100%，不符合=0
	可用性	根据实际需求定义	数据在设定有效生存周期内的可使用性。	定量分析	式中： 满足可用性要求的数据集中元素的个数； 被检核的数据集中元素的个数。

附 录 B
（资料性）
数据质量问题分类

影响数据质量的技术成因，见表 B.1：

表 B.1 影响数据质量技术成因

序号	因素	内容
1	数据库表设计	表 1 数据库表结构、数据库约束条件、数据校验规则的设计开发不合理，造成数据录入无法校验或校验不当，引起数据重复、不完整、不准确
2	数据源的数据质量控制	数据源存在数据质量问题
3	数据采集	数据采集过程质量问题， 例如：采集点、采集频率、采集内容、映射关系等采集参数和流程设置的不正确，数据采集接口效率低，导致的数据采集失败、数据丢失、数据映射和转换失败
4	数据传输	数据传输过程的问题，例如：数据接口本身存在问题、数据接口参数配置错误、网络不可靠等都会造成数据传输过程中的发生数据质量问题
5	数据装载	数据装载过程的问题，例如：数据清洗规则、数据转换规则、数据装载规则配置有问题
6	数据存储	数据存储的质量问题，例如：数据存储设计不合理，数据的存储能力有限，人为后台调整数据，引起的数据丢失、数据无效、数据失真、记录重复。

影响数据质量的业务成因，见表 B.2：

表 B.2 影响数据质量业务成因

序号	因素	内容
1	业务应求不明确	业务应求不清晰，例如：数据的业务描述、业务规则不清晰，导致技术无法构建出合理、正确的数据模型
2	业务应求频繁变更	业务应求的变更，应求一变，数据模型设计、数据录入、数据采集、数据传输、数据装载、数据存储等环节都会受到影响，稍有不慎就会导致数据质量问题的发生。
3	业务端数据输入不规范	业务端数据输入不规范
4	数据作假	数据作假

影响数据质量的管理成因，见表 B.3：

表 B.3 影响数据质量管理成因

序号	因素	内容
1	缺乏数据思维	没有认识到数据质量的重要性，重系统而轻数据
2	缺乏数据认责机制	没有明确数据归口管理部门或岗位，缺乏数据认责机制
3	缺乏统一的管理机制	没有明确数据归口管理部门或岗位，缺乏数据认责机制
4	缺乏统一的数据输入规范	数据输入规范不统一，不同的业务部门、不同的时间、甚至在处理相同业务的时候，由于数据输入规范不同，造成数据冲突或矛盾。
5	缺乏有效的数据质量控制措施	数据质量问题从发现、指派、处理、优化没有一个统一的流程和制度支撑，数据质量问题无法闭环。

6	缺乏数据质量问题闭环管理制度	数据质量问题从发现、指派、处理、优化没有一个统一的流程和制度支撑，数据质量问题无法闭环。