ICS 35.240.01
CCS L 70

T/CS

团 体 标 准

T/CS 318—2026

全量数据质量管理平台技术规范

Technical specifications for full-volume data quality management platform

2026－06－12 发布 2026－06－12 实施

中国商品学会 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 1
5　运行环境 1
6　技术架构 2
7　功能要求 4
8　非功能要求 9
9　测试 11
10　运维 12
附录 A（资料性）　Q 值计算 13
附录 B（资料性）　元规则列表 17
参考文献 19

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。

本文件由成都康赛信息技术有限公司提出。

本文件由中国商品学会归口。

本文件起草单位：成都康赛信息技术有限公司、成都奥明特软件技术有限公司、舍法科技（四川）有限公司。

本文件主要起草人：马晨曦、唐雪飞、唐雪峰、王春明、唐彪。

全量数据质量管理平台技术规范

1 范围

本文件规定了全量数据质量管理平台的运行环境、技术架构、功能要求、非功能要求、验收及运维，描述了相应的测试方法。

本文件适用于全量数据质量管理平台的设计、验收及运维。

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 28827.1 信息技术服务 运行维护 第 1 部分：通用要求

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求程度。

[来源：GB/T 36344—2018,2.3]

3.2

数据质量管理平台 data quality management platform

基于公司数据实施人员和客户方相关数据质量管理员提供数据质量测量、监控、报告功能的平台。

3.3

脏数据 dirty data

不符合预设业务规则或质量标准的数据记录。

3.4

质量总分（Q 值） quality score

基于多维数据质量指标计算得出的综合评分，用于量化数据整体质量水平。

3.5

元规则 meta-rule

平台内置的、预定义的、参数化的数据质量检测模板，用于描述一类特定的数据质量问题的检测逻辑。

4 缩略语

下列缩略语适用于本文件。

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

CPU：中央处理器（Central Processing Unit）

EDQAF：企业数据质量评估框架（Enterprise Data Quality Assessment Framework）

ID：身份标识（Identity Document）

IP：互联网协议（Internet Protocol）

SQL：结构化查询语言（Structured Query Language）

5 运行环境

5.1 物理部署

系统采用分布式架构部署，主要物理节点包括质量管理平台应用服务器、质量引擎服务器、系统数据库服务器及缓存服务器等，可与中台内部服务器协同部署。

5.2 硬件环境

5.2.1 系统服务器要求

系统各服务器节点应具备满足业务处理和数据存储需求的硬件资源，并支持集群化部署以实现高可用性与水平扩展。核心要求如下：

a) 应用服务器：配置充足的运算核心和内存资源，以保障平台服务的高并发处理能力和响应速度；
b) 质量引擎服务器：作为核心计算节点，应具备较高的 CPU 处理能力和内存容量，以满足复杂质量规则检测和分析的计算需求，并支持部署多台以实现负载分担；
c) 数据库服务器：配置足够的 CPU、内存及大容量高速存储，以保障系统元数据、质量结果等数据的可靠存储和高效访问；
d) 缓存服务器：应配备适量内存资源，用于提升热点数据访问性能和系统整体吞吐量。

5.2.2 系统服务器配置

各服务器具体配置应依据实际数据规模、检测规则复杂度、用户并发量及性能目标进行规划与弹性调整。

5.3 软件环境

5.3.1 服务器端

部署主流稳定版本的操作系统，支持 Java 运行环境、Web 应用、关系型数据库及主流缓存数据库。

5.3.2 客户端

应能使用主流浏览器的最新稳定版本进行访问。

6 技术架构

全量数据质量管理平台采用分层、模块化及分布式架构设计，整体技术架构自下而上分为基础平台层、数据资源层、服务支持层、应用层和用户层，各层协同工作，共同构成完整的数据质量管理体系。平台技术架构图具体见图 1。

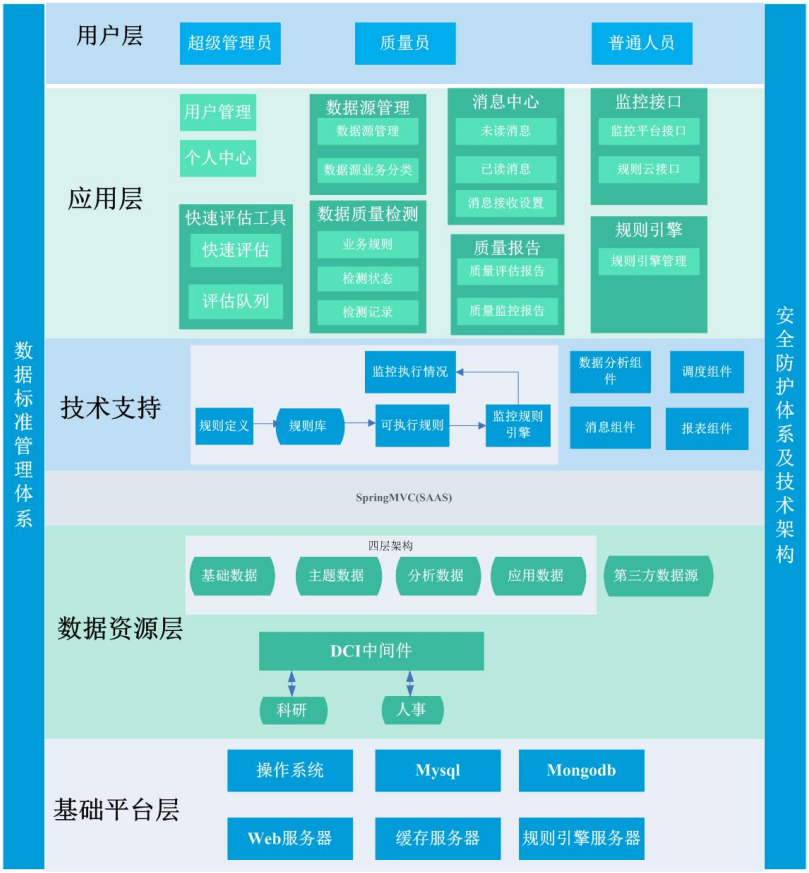


图 1　全量数据质量管理平台技术架构图

6.1　基础平台层

基础平台层为系统提供底层硬件和基础软件支撑，包括：

a)　服务器：支持应用服务器、规则引擎服务器、数据库服务器及缓存服务器的集群化部署，以满足高并发和高可用性需求；

b)　操作系统：支持主流稳定的服务器操作系统；

c)　数据库：采用关系型数据库存储系统元数据、规则配置、监控结果等结构化数据；利用非关系型数据库存储质量检测过程中产生的海量错误明细数据及过程数据，以支持高效查询和灵活扩展；

d)　中间件和运行环境：基于 Java 平台，采用标准的 Web 应用服务器部署应用服务，并利用消息中间件、调度组件等支持异步处理与任务调度。

6.2　数据资源层

数据资源层是平台实施质量管理的对象，负责接入和管理各类数据源，包括：

a)　多源数据接入：支持对接关系型数据库及第三方数据源，通过统一的数据源管理模块进行连接配置、状态监控和权限控制；

b)　数据逻辑组织：支持基于业务分类对数据表进行多级逻辑分组，便于按业务域进行质量规则的集中配置和管理。

6.3　服务支持层

服务支持层是平台的核心能力层，封装了各类可复用的技术组件和服务，包括：

a)　规则引擎：作为核心计算组件，提供可插拔、分布式部署的规则执行能力。通过规则描述语言将业务规则模板转化为可执行任务，支持对多类质量维度（有效性、一致性、完备性、及时性、完整性）规则的高效、并发执行；

b) 调度组件：提供灵活的任务调度机制，支持手动、定时及周期性地触发数据质量评估和监控任务；
c) 消息组件：实现系统内部事件和通知的异步通信，支持向用户推送质量告警、任务状态等信息；
d) 数据分析组件：提供数据快速评估功能，对字符、数值、日期等字段类型进行基础统计分析，为规则配置提供数据洞察。

6.4 应用层

应用层面向用户提供具体的功能模块，通过 Web 界面进行交互，主要包括：
a) 系统管理：涵盖用户管理、角色权限控制、个人中心等功能；
b) 数据源管理：实现数据源的增删改查、连接测试及业务分类管理；
c) 规则管理：提供基于五大维度的业务规则定义、配置、试运行及生命周期管理；
d) 质量评估：提供单字段的快速评估工具及评估队列管理；
e) 质量监控：实现质量规则的批量执行、监控过程跟踪、结果报告生成及问题数据查看；
f) 消息中心：统一管理用户接收到的各类系统消息和质量告警。

6.5 用户层

用户层定义了平台的不同使用者角色及其权限，主要包括：
a) 超级管理员：拥有系统全部管理权限；
b) 系统管理员：负责数据源、规则、用户等日常运维管理；
c) 质量员/普通用户：根据授权，使用质量评估、监控、报告查看等功能，并可接收相关质量消息。

7 功能要求

7.1 用户登录

平台与数据中台应实现统一身份认证，用户可依托中台身份体系跳转访问本平台。

7.2 数据质量监控

7.2.1 平台应提供数据源维度的全局质量概览仪表板，至少包括以下组件和功能：
a) 监控指标统计：实时展示监控表数、监控字段数、业务规则总数、累计检测数据量等核心指标；
b) 阈值指标轮播展示：动态展示附录 B 所列 34 种元规则对应的核心质量指标（如单字段填充率、唯一性比率、格式合规率等）的计算结果及达标状态，具体指标清单见附录 B；
c) Q 值趋势图：以折线图形式展示选定数据源或业务分类的 Q 值随时间的变化趋势；
d) 脏数据量/Q 值双 Y 轴图：在同一图表中以双 Y 轴形式展示各检测时间点的 Q 值及对应的脏数据总量；
e) 五大维度分布雷达图：展示“完备性、一致性、及时性、有效性、完整性”五个维度的得分占比，支持鼠标悬停查看维度占比、维度得分及维度贡献值；
f) 数据质量排名：按业务分类展示其当前 Q 值的排名及较上次排名的变化（上升/下降）；
g) 待检测业务规则列表：列表展示状态为“等待检测”的业务规则，包含规则名称、预设阈值等信息；
h) 历史运行记录：表格展示历次检测的执行时间、总体 Q 值及各维度得分；
i) 检测时间信息：明确展示上次成功检测的时间和距离下一次计划检测的剩余时间。

7.2.2 平台应支持以下交互操作：
a) 选择数据源：支持用户切换查看不同数据源的质量监控仪表板；
b) 过程回放：支持回放最近一次检测的全过程动画；
c) 手动立即检测：支持对选定数据源或业务分类触发一次即时质量检测。

注：Q 值计算见附录 A。

7.3 快速评估工具

平台应具备免规则配置的字段即时质量分析能力。

7.3.1 选择评估对象

7.3.1.1 一键评估全表

应支持对单张数据表全部字段发起批量质量评估。

7.3.1.2 单字段评估

应支持针对数据表内单个指定字段发起质量评估。

7.3.1.3 更改字段类型

支持对字符串类型字段临时转换为数值型、日期型（支持多种日期格式）开展评估，并可恢复字段原始类型。

7.3.2 评估任务监控

7.3.2.1 队列视图

列表应支持统一展示运行中、待执行两类评估任务队列，展示内容包含数据源、业务分类、数据表、评估字段、任务创建时间、任务状态等信息。

7.3.2.2 任务控制

应支持对处于待执行状态的评估任务进行终止处理。

7.3.3 评估结果管理

7.3.3.1 结果分类视图

应支持按照字符类、数值类、日期类对历史评估结果进行分类展示。

7.3.3.2 结果查询和筛选

应支持按字段名称、创建时间等条件筛选历史评估记录。

7.3.3.3 详细分析

针对单条评估记录，具备以下分析能力：

a) 统计并展示字段最大值、最小值、平均值、汇总值、空值数量等基础数据指标；
b) 针对字符型字段，通过饼图及列表展示内容模式聚类结果，包含模式类型、数据量、占比；
c) 支持追溯查看指定字段的全量历史评估记录。

7.3.3.4 结果管理

应支持对评估结果记录进行单条删除或批量删除。

7.4 业务规则管理

平台应提供对数据质量业务规则的全生命周期精细化管理和智能配置功能。

7.4.1 规则概览和导航

7.4.1.1 树状导航

应支持按“数据源-业务分类-表/视图”的层级结构浏览，支持中英文切换，并展示各节点下的规则数量统计。

7.4.1.2 全局统计

应支持在数据源、业务分类层级，统计展示五大质量维度下已启用规则数量、规则总量及维度分布占比。

7.4.2 规则列表和操作

7.4.2.1 多维度筛选

应支持按质量维度、启禁用状态、是否异常规则（如表缺失、字段缺失）及规则名称进行组合筛选。

7.4.2.2 批量操作

应支持对选中的规则进行批量启用、禁用、删除、导出（仅限支持导出的元规则，见附录 B）。

7.4.2.3 规则试运行

应支持对选定的一条或多条规则进行即时检测，并在检测结果管理中查看详情。

7.4.2.4 规则详情

应支持展示规则基础信息，包含规则名称、所属质量维度、元规则编码与名称、权重、阈值、启用状态、更新时间；支持对单条规则进行查看、编辑、删除。

7.4.3 规则配置

7.4.3.1 新增规则

应采用向导式配置，分为“选择元规则”和“配置规则参数”两步。

7.4.3.2 元规则库

平台应内置覆盖五大质量维度的数据质量元规则库。元规则库至少应包含附录 B 所列的 34 种元规则。每种元规则应有唯一的代码标识、明确的中文名称、详细的功能描述、适用场景说明以及建议的阈值和权重配置。

7.4.3.3 智能辅助

7.4.3.3.1 规则推荐：可依据表字段的元数据（如字段名、类型、注释）智能推荐适用的元规则。

7.4.3.3.2 快速模板：应支持一键为当前表生成“单字段填充率”、“唯一性检测”等常用规则模板。

7.4.3.4 参数配置

应支持根据所选元规则，可视化配置检测字段、阈值、权重、数据检测范围、跨表关联对象及关联字段等参数。

7.5 检测调度管理

平台应提供灵活、可配置的质量检测任务调度机制。

7.5.1 全局视图

列表应展示所有已配置检测任务的数据源或业务分类，包含其有效规则总数、维度分配比例、检测周期、下次检测时间、检测状态（已完成/检测中/暂停）。

7.5.2 任务配置

7.5.2.1 周期检测

应支持为数据源或业务分类配置手动、每小时、每天、每周、每月等不同粒度的周期性自动检测任务。

7.5.2.2 维度权重设置

应支持自定义五大质量维度在计算该数据源或业务分类 Q 值时的权重分配比例，各维度权重之和须为 100%。

7.5.3 任务控制

应支持立即触发手动检测、暂停或恢复定时检测任务。

7.6 检测结果管理

平台应具备全量检测结果查询、分析、处置能力。

7.6.1 检测记录查询

7.6.1.1 多维度筛选

应支持按被检测数据源/业务分类、检测范围（整体/试运行）、检测类型（定时/即时）、提交人、检测状态、检测时间范围进行查询。

7.6.1.2 记录列表

应展示每次检测的运行 ID、被检测对象、涉及业务分类数、检测规则条数、检测类型、提交人、开始时间、耗时、状态、Q 值、脏数据条数、脏数据清空状态等信息。

7.6.2 检测结果详情

7.6.2.1 摘要信息

应支持展示单次检测整体信息，包含任务状态、总耗时、检测对象、覆盖业务分类、检测规则数量、维度权重分配比例。

7.6.2.2 规则级结果

应支持展示单条业务规则的检测详情，包含数据表名称、规则名称、元规则信息、实测值与阈值、权重、规则得分、检测耗时、运行状态；支持追溯查看对应脏数据明细。

7.6.2.3 脏数据操作

应支持对检测完成的记录，清空本次产生的所有脏数据；支持预览、导出、删除具体规则下的脏数据。

7.6.3 实时任务监控

应支持实时监控正在运行的检测任务，展示任务进度、已检测规则数、总规则数，并支持查看任务详情、暂停或终止运行中的检测任务。

7.7 数据质量报告

平台应支持自动生成和查看多维度、结构化的数据质量报告。

7.7.1 综合质量报告

7.7.1.1 报告概览

应以“报告封面”形式汇总展示各数据源在最新一次检测中的整体质量情况，包括数据源名称、Q 值、最新检测时间。

7.7.1.2 报告详情

应包含被检测数据源详情、Q 值趋势图、五大维度雷达图、各业务分类 Q 值排名和统计、以及下钻至业务分类报告的能力。

7.7.2 业务分类报告

7.7.2.1 报告概览

应集中展示各业务分类报告基础信息，包含业务系统名称、检测时间。

7.7.2.2 报告详情

应包含该业务分类的检测详情（表数量、字段数量、规则数量、数据量、脏数据量）、Q 值、五大维度分布、脏数据量/Q 值双 Y 轴图、脏数据占比 TOP10 表分析，以及检测表和检测字段的明细列表。

7.7.3 历史报告分析

7.7.3.1 趋势分析

应支持选择数据源和多个时间点，以折线图对比展示其下各业务分类 Q 值的历史变化趋势。

7.7.3.2 报告对比

应支持对比不同时间点的质量报告，分析质量变化。

7.8 智能专题报告

平台具备基于算法模型的深度数据质量分析及专题报告生成能力，报告可依托附录 B 元规则或专用分析模型生成。

7.8.1 专题报告分类

7.8.1.1 模式聚类专题报告

基于字符型字段内容，利用模式聚类算法自动分析数据模式分布，识别异常模式，生成分析报告。

7.8.1.2 波动分布专题报告

针对数值型字段，分析其在不同时间间隔内的统计量（如计数、求和、平均值）波动情况，生成趋势和异常分析报告。

7.8.1.3 列关系专题报告

分析表内或跨表的多列字段间的统计关系（如相关性），发现潜在的逻辑约束问题，生成关系分析报告。

7.8.2 专题报告管理

应支持配置生成报告的规则、立即检测生成报告、查看历史报告时间轴、下载及删除报告。

7.9 数据源管理

平台应提供对各类数据源的统一接入和管理视图。

7.9.1 树状管理视图

应以树形目录展示所有已接入的数据源、其下的业务分类和具体数据表/视图，支持关键字搜索和表数量统计。

7.9.2 数据源操作

7.9.2.1 新增/编辑

应支持添加或编辑“其他数据源”（如 Oracle, MySQL），配置连接信息、测试连接、设置元数据同步周期。

7.9.2.2 集成中台数据源

应支持直接从数据中台选择“全量数据中心”的数据层（原始层、主题层等）作为数据源，并自动同步其业务分类。

7.9.2.3 状态管理

应支持查看数据源连接状态，执行立即同步元数据、删除数据源（无规则时可删除）等操作。

7.9.3 业务分类和表管理

7.9.3.1 业务分类管理

应支持为“其他数据源”新增、编辑业务分类，配置数据负责人、时间范围起点、是否在报告中显示。

7.9.3.2 表信息预览

应支持预览任意数据表的表结构（字段名、类型、注释）和样本数据。

7.9.3.3 中英文切换

应支持全局切换树状导航及列表中表名、字段名的中英文显示。

7.10 系统管理和配置

平台应提供必要的系统后台配置和运维支持功能。

7.10.1 引擎管理

应监控所有数据质量检测引擎节点的运行状态（IP、端口、CPU、内存、负载、心跳），支持启用或禁用引擎。

7.10.2 配置管理

7.10.2.1 五大维度方案配置

应内置多套五大维度权重分配方案，并支持自定义新增、编辑权重方案。

7.10.2.2 业务规则缺省值配置

应支持为附录 B 所列 34 项元规则统一配置全局默认阈值、判定条件及权重。

7.10.3 消息中心

应统一管理用户接收到的系统消息和质量告警，支持订阅检测完成、运行异常、阈值超限等通知。

7.10.4 关于信息

应支持展示平台版本、版权信息及相关术语释义内容。

8 非功能要求

8.1 性能

平台在单引擎检测模式下，应符合以下性能要求：

a) 对 1 000 万行数据的单表进行单字段填充率检测，响应时间不应超过 30 s；
b) 对 1 亿行数据的分区表进行 10 条规则并发检测，总耗时不应超过 10 min；
c) 在每秒 100 次并发 API 请求下，核心页面加载时间不应超过 3 s。

注：具体性能指标可根据实际部署环境调整，但应在产品文档中明确标称值。

8.2 易用性

平台易用性要求如下：

a) 应提供图形界面的系统安装配置工具，以便于系统部署；
b) 应提供完整的产品文档，包括但不限于安装部署手册、管理员使用手册、应用开发指南、用户操作手册等。

8.3 可靠性

8.3.1 高可用

高可用要求如下：

a) 应提供系统自动故障探测及管理功能；
b) 系统组件应无单点故障风险；
c) 集群任意节点发生故障时，不应出现服务中断、数据丢失或数据不一致的情况；
d) 集群任意单元发生故障时，系统操作应不受影响；
e) 系统应能长期无故障不间断运行。

8.3.2 数据冗余存储和分布

数据冗余存储和分布要求如下：

a) 应提供元数据多副本存储功能，任意节点发生故障时不影响系统继续提供服务的能力；
b) 应提供基于分区容错的主副本规划功能，具有提前规划各副本数据物理分布的能力。

8.3.3 数据备份和恢复

数据备份和恢复要求如下：

a) 应提供分布式文件存储备份和恢复功能；
b) 应提供分布式结构化数据存储备份和恢复功能；
c) 应提供分布式列式存储备份和恢复功能；
d) 应支持数据全量备份和增量备份；
e) 应支持数据自动备份和手动备份。

8.3.4 故障恢复和迁移

故障恢复和迁移要求如下：

a) 任意节点发生故障后，系统应提供将修复后的节点接回系统的能力；
b) 故障恢复和迁移过程不应影响系统用户数据的完整性和一致性；
c) 故障恢复和迁移过程不应影响系统整体服务能力。

8.4 兼容性

平台应具备良好的环境适应性，支持在主流的操作系统、数据库及浏览器环境中部署和运行。

8.5 可维护性

平台应便于系统的部署、维护和升级，具体要求如下：

a) 系统安装和初始化：支持系统初次安装和初始化，内置管理员账户。
b) 日志和监控：提供系统运行日志、操作审计日志，便于问题排查和系统维护。

8.6 安全性

8.6.1 用户管理

用户管理要求如下：

a) 应对登录用户进行身份标识和鉴别；
b) 用户身份鉴别信息应满足一定的复杂度要求，并定期更换；
c) 应提供登录失败处理措施，如结束会话、限制非法登录次数、登录连接超时自动退出等措施。

8.6.2 权限管理

权限管理要求如下：

a) 应以系统组件为单位配置角色和用户；
b) 应按照权限最小化的原则为用户配置权限；
c) 应支持按照数据表级、数据列级的粒度为用户分配权限；
d) 应支持按照不同操作类型（如增、删、改、查、执行等）为用户分配权限。

8.6.3 日志管理

日志管理要求如下：

a) 应提供记录系统操作日志功能，记录用户的重要操作；
b) 应保证系统操作日志的完整性和不可篡改性，已记录的日志不应被修改或删除；当日志达到存储上限时，应按照预设策略（如按时间轮转、归档）进行清理，但清理前应导出备份；
c) 操作日志应包括日期、时间、操作者信息、操作类型、操作描述和操作结果等；
d) 应提供对系统操作日志进行统计、查询、分析及生成报表的功能。

8.6.4 数据安全

数据安全应符合下列要求：

a） 应提供数据存储加解密功能，支持数据库级数据加密；

b） 应提供系统敏感数据加密传输功能，并且加密密钥可被替换；

c） 应支持数据列级的数据加密。

9 测试

9.1 功能测试

9.1.1 用户登录

输入正确用户名及密码，点击登录，验证能否正常进入系统首页。

9.1.2 数据质量监控

查看各类统计图表、执行手动检测、选择数据源、回放检测过程。

9.1.3 快速评估工具

选择评估对象、执行一键评估全表、查看评估结果、删除或取消评估任务。

9.1.4 业务规则管理

通过树状导航切换数据源/业务分类/表层级；执行查询、新增、编辑、启用/禁用、批量导出、试运行等操作。

9.1.5 检测调度管理

配置检测周期、设置维度权重、启停检测任务、手动触发检测。

9.1.6 检测结果管理

按多维度查询检测结果、查看脏数据明细、导出数据、监控进行中的检测任务。

9.1.7 数据质量报告

生成并查看综合报告、业务分类报告、历史报告；进行报告对比和趋势分析。

9.1.8 智能专题报告

立即检测、配置规则、查看并下载专题报告。

9.1.9 数据源管理

新增/删除数据源；同步元数据；切换中英文显示；预览表结构和样本数据。

9.1.10 系统管理

启用/禁用引擎；新增/编辑/删除系统配置；查看关于信息。

9.2 非功能测试

9.2.1 性能

9.2.1.1 基准测试：

a） 选取单表 1 000 万行测试数据，执行单字段填充率检测，记录任务响应时间，验证时长不超过 30 s。

b） 选取分区表 1 亿行测试数据，配置 10 条业务规则并启动并发检测，记录整体执行总耗时，验证时长不超过 10 min。

9.2.1.2 负载测：逐步递增数据量、规则数量与任务并发数，持续监测任务执行时长、页面响应时长及服务器 CPU、内存、网络等资源使用率，确认系统运行平稳，各项指标处于正常区间。

9.2.1.3 压力测试：模拟每秒 100 次并发 API 请求的访问压力，持续访问平台核心页面，记录页面加载时长，验证加载时间不超过 3 s；测试过程中检查页面无加载失败、数据错乱、服务中断等异常现象。

9.2.2 易用性

9.2.2.1 安装配置

使用图形化安装工具完成系统部署，验证安装过程是否顺畅、配置是否简单。

9.2.2.2 文档验证

对照用户手册执行关键操作，验证文档内容准确性和步骤清晰度。

9.2.3 可靠性

9.2.3.1 故障容错：模拟集群节点故障，验证系统是否自动切换且服务无中断、数据无丢失。
9.2.3.2 备份恢复：执行数据备份并模拟数据损坏场景，验证恢复后数据完整性和系统可用性。
9.2.3.3 长时间运行：连续运行系统 72 h，监测其稳定性。

9.2.4 兼容性

9.2.4.1 环境兼容：在主流操作系统和浏览器上安装运行系统，验证基本功能正常可用。
9.2.4.2 数据源兼容：连接不同类型数据库进行数据读写，验证接口兼容性。

9.2.5 可维护性

9.2.5.1 日志监控：检查系统日志是否完整记录操作和错误信息，验证日志查询功能正常。
9.2.5.2 升级维护：执行系统版本升级，验证过程平滑且不影响现有数据和服务。

9.2.6 安全性

9.2.6.1 权限控制：创建不同角色用户，验证权限分配准确且无法越权访问。
9.2.6.2 安全机制：验证密码策略、登录控制、数据加密等安全功能是否按设计正常工作。

9.3 测试交付

9.4 测试交付

测试完成后应提供以下交付物：

a) 系统测试方案；
b) 测试用例；
c) 阶段测试报告；
d) 缺陷清单；
e) 系统测试报告；
f) 测试用例执行结果；
g) 系统操作手册；
h) 安全测试报告。

10 运维

10.1 系统在运行维护方面应符合 GB/T 28827.1 的规定。
10.2 应建立日常维护制度，定期对硬件设备及网络系统进行检查、维护，保障系统的兼容性和开放性。

附 录 A
（资料性）
Q 值计算

A.1 相关术语和定义

见表 A.1。

表 A.1 Q 值计算相关术语

术语	定义
质量维度	评价数据质量的五个基本方面： 完备性：数据拥有所有必须或适当的部分，第一条件是存在性； 一致性：数据在时间序列上不存在非预期的变异或变更，强调稳定性和连贯性； 及时性：数据表示现实世界的及时程度，与数据的可用性相关； 有效性：数据对一组业务规则、标准或已定义值域的符合程度，强调空间符合性； 完整性：数据对数据关系规则的符合程度，数据能完备、一致和有效地表示业务概念
规则	基于某一元规则创建的、应用于具体表字段的质量检测实例
权重	分配给单条规则的重要性值，范围为 1～10 的整数，值越大代表该规则越重要
权重占比	单条规则的权重在其所属质量维度内的相对比重。计算公式为：本规则权重/所属维度下所有规则的总权重
阈值	规则检测的期望值或目标值，可分为比率型或数量型
维度占比	五大质量维度在计算上级（业务分类或数据源）Q 值时的预设权重分配比例，各维度占比之和为 100%
规则得分	单条规则基于实际检测结果计算出的分数，范围为 0 分～100 分
规则权重得分	单条规则对其所属维度质量分数的贡献值，计算公式为：规则得分×该规则在本维度的权重占比
Q 值	表征某一数据范围（如表、业务分类、数据源）整体数据质量的综合得分，基于各维度质量分数及其维度占比计算得出

A.2 计算模型和场景说明

A.2.1 基本场景设定

A.2.1.1 假设一个数据源下包含 n 个业务分类，每个业务分类的五大维度占比可独立配置。例如：

——业务分类 1 的维度占比：1C1，1C2，1T，1V，1I（和为 100%）；

——业务分类 2 的维度占比：2C1，2C2，2T，2V，2I（和为 100%）；

……

——业务分类 n 的维度占比：nC1，nC2，nT，nV，nI（和为 100%）。

A.2.1.2 每个业务分类下，每个维度包含若干条规则。例如：

——业务分类 1 的完备性（C1）维度有 m1 条规则；

——业务分类 2 的一致性（C2）维度有 p2 条规则；

——以此类推，具体见表 A.2。

表 A.2 不同业务分类的规则

业务分类	五大维度				
	完备性（C1）	一致性（C2）	及时性（T）	有效性（V）	完整性（I）
1	1C1 m1 条规则	1C2 p1 条规则	1T q1 条规则	1V s1 条规则	1I w1 条规则
2	2C1 m2 条规则	2C2 p2 条规则	2T q2 条规则	2V s2 条规则	2I w2 条规则
3	3C1 m3 条规则	3C2 p3 条规则	3T q3 条规则	3V s3 条规则	3I w3 条规则
…	…	…	…	…	…
n	nC1 mn 条规则	nC2 pn 条规则	nT qn 条规则	nV sn 条规则	nI wn 条规则

A.2.2 数据源的维度占比计算

A.2.2.1 数据源的完备性占比按公式（A.1）进行计算：

$$C1=\frac{\sum_{\lambda=1}^{n}\lambda C1}{n} \quad \text{(A.1)}$$

式中：

$C1$——数据源的完备性占比；

n——业务分类的总数；

λ——求和索引变量，代表第 λ 个业务分类。

A.2.2.2 数据源的一致性占比按公式（A.2）进行计算：

$$C2=\frac{\sum_{\lambda=1}^{n}\lambda C2}{n} \quad \text{(A.2)}$$

式中：

$C2$——数据源的一致性占比；

n——同式（A.1）；

λ——同式（A.1）。

A.2.2.3 数据源的及时性占比按公式（A.3）进行计算：

$$T=\frac{\sum_{\lambda=1}^{n}\lambda T}{n} \quad \text{(A.3)}$$

式中：

T——数据源的及时性占比；

n——同式（A.1）；

λ——同式（A.1）。

A.2.2.4 数据源的有效性占比按公式（A.4）进行计算：

$$V=\frac{\sum_{\lambda=1}^{n}\lambda V}{n} \quad \text{(A.4)}$$

式中：

V——数据源的有效性占比；

n——同式（A.1）；

λ——同式（A.1）。

A.2.2.5 数据源的完整性性占比按公式（A.5）进行计算：

$$\mathrm{I}=\frac{\sum_{\lambda=1}^{n}\lambda I}{n} \quad \text{(A.5)}$$

式中：

I——数据源的完整性占比；

n——同式（A.1）；

λ——同式（A.1）。

A.3 计算

A.3.1 单条规则得分（R）计算

A.3.1.1 条件为等于阈值，当 $A=THR$，则 R=100 分；当 $A<THR$ 或 $A>THR$，则按公式（A.6）进行计算：

$$R=\left[1-\frac{|THR-A|}{THR+|A-THR|}\right]\times 100\% \quad \text{(A.6)}$$

式中：

R——本规则得分；

THR——阈值（小于 1 的百分数适用）；

A——实际值（小于 1 的百分数适用）。

A.3.1.2 条件为不小于阈值，当 $A\geqslant THR$，则 R =100 分；当 $A<THR$，则按公式（A.7）进行计算：

$$R = [1 - (THR - A)/THR] \times 100\% \cdots\cdots (A.7)$$

式中：

R——同式（A.6）；

THR——同式（A.6）；

A——同式（A.6）。

A.3.1.3 条件为不大于阈值，当 $A \leqslant THR$，则 R =100 分；当 $A > THR$，则按公式（A.8）进行计算：

$$R = [1 - (A - THR)/(1 - THR)] \times 100\% \cdots\cdots (A.8)$$

式中：

R——同式（A.6）；

THR——同式（A.6）；

A——同式（A.6）。

A.3.2 业务分类的 Q 值计算

A.3.2.1 业务分类 1 的 Q 值按公式（A.9）进行计算：

$$Q1 = \left[\sum_{i=1}^{m1}(RwscC1i)\right] \times 1C1 + \left[\sum_{i=1}^{p1}(RwscC2i)\right] \times 1C2 + \ldots + \left[\sum_{i=1}^{w1}(RwscIi)\right] \times 1I \cdots\cdots (A.9)$$

式中：

$Q1$——业务分类 1 的 Q 值；

R——同式（A.6）；

$Rwsc$——单个规则业务分类权重得分，即本规则得分*本规则业务分类权重占比。

A.3.2.2 业务分类 2 的 Q 值记为 Q2，计算同上。

……

A.3.2.3 业务分类 n 的 Q 值记为 Qn，计算同上。

A.3.2.4 总 Q 值按公式（A.10）进行计算：

$$Q_{total} = (Q1 + \mathrm{Q2} + \ldots + \mathrm{Q}n)/n \cdots\cdots (A.10)$$

式中：

Q_{total}——总 Q 值；

$Q1$、$Q2 \ldots Qn$——各业务分类的 Q 值；

n——业务分类的总数。

A.3.3 数据源 Q 值计算

A.3.3.1 完备性（C1）的 Q 值按公式（A.11）进行计算：

$$QC1 = \left[\sum_{i=1}^{m1}(RwscC1i)\right] \times \left[1C1/\sum_{j=1}^{n}(jC1)\right] + \left[\sum_{i=1}^{m2}(RwscC1i)\right] \times \left[2C1/\sum_{j=1}^{n}(jC1)\right] + \ldots + \left[\sum_{i=1}^{mn}(RwscC1i)\right] \times \left[nC1/\sum_{j=1}^{n}(jC1)\right] \quad (A.11)$$

式中：

$QC1$——完备性（C1）的 Q 值；

n——业务分类的总数。

A.3.3.2 一致性（C2）的 Q 值记为 QC2，计算同上。

……

A.3.3.3 完整性（V）的 Q 值记为 QCV，计算同上。

A.3.3.4 总 Q 值按公式（A.12）进行计算：

$$Q_{total} = QC1 \times \frac{\sum_{\lambda=1}^{n}\lambda C1}{n} + QC2 \times \frac{\sum_{\lambda=1}^{n}\lambda C2}{n} + \ldots + \mathrm{QCI} \times \frac{\sum_{\lambda=1}^{n}\lambda I}{n} \cdots\cdots (A.12)$$

式中：

Q_{total}——总 Q 值；

$QC1$、$QC2$...QCI——各维度 Q 值；
n——业务分类的总数。

附 录 B
（资料性）
元规则列表

见表 B.1。

表 B.1 元规则列表

质量维度	元规则代码	元规则中文名称	核心功能描述	阈值设置			是否支持导入/导出	是否可以自动推荐规则	是否可以生成到表的快速模板
				阈值	条件	默认值			
完备性	EDQAF-12	单字段填充率	检测指定字段非空记录的比率	单字段填充率	≥	95%	是	是	是
	EDQAF-14	唯一性检测	检测指定字段重复记录的比率	重复率	≤	5%	是	否	是
	EDQAF-18	字段内容完备性	检测多个字段同时为非空的记录比率	多字段填充率	≥	80%	否	否	否
	EDQAF-01	参照数据充分性	检查码表是否充分覆盖数据表字段值	缺失代码值数量	=	0	否	否	否
	EDQAF-15	单字段默认完备性	检测字段值等于特定默认值的记录比率	单字段默认完备率	≥	80%	否	否	否
	EDQAF-16	日期完备性	检测日期字段值在指定最小/最大范围内的记录比率	单字段日期完备率	≥	90%	否	否	否
	EDQAF-17	日期合理性	检测日期字段最大最小值之差在合理间隔内	日期时间间隔	≥	空	否	否	否
一致性	EDQAF-36	历史数值比较	数值字段统计量（和/均值）与历史值的波动比率	波动比率	≤	5%	是	否	否
	EDQAF-02	字段模式聚类	检测字符字段内容符合指定模式（如身份证、电话）的比率	符合模式比率	≥	99%	否	否	否
	EDQAF-34	时序和业务规则的一致性	检测两个时间字段符合指定时序关系（如借书还书）的记录比率	业务时序一致率	=	100%	是	否	否
	EDQAF-44	跨表数额列计算结果一致性	比较两表间数值字段统计量（和/均值）的一致性	跨表不一致性	≤	10%	否	否	否
	EDQAF-31	记录数一致性	检测字段不重复值数量与预期值的差异比率	记录数一致性比率	≤	10%	否	否	否
	EDQAF-32	记录比率一致性	检测两字段不重复值数量的比率与预期值的差异	记录数比率一致性比率	≤	10%	否	否	否
	EDQAF-37	聚合日期记录数一致性	比较不同时间段内记录数的一致性	聚合日期记录数一致性比率	≤	10%	否	否	否
	EDQAF-38	聚合日期数额一致性	比较不同时间段内数值字段统计量的一致性	聚合日期数额一致性比率	≤	10%	否	否	否
及时性	EDQAF-26	数据更新频率	检测在指定周期内数据表有更新的记录数量	更新数据量	≥	1 000	否	否	否
	EDQAF-25	数据处理用时	检测数据处理作业的实际用时是否超限	处理用时	≤	无	否	否	否

表B.1　元规则列表（续）

质量维度	元规则代码	元规则中文名称	核心功能描述	阈值设置			是否支持导入/导出	是否可以自动推荐规则	是否可以生成到表的快速模板
				阈值	条件	默认值			
有效性	EDQAF-27	数据规范约束	检测字符字段内容是否符合自定义的起止/包含字符串约束	符合数据约束比率	≥	80%	否	否	否
	EDQAF-28	长度规范约束	检测字段内容长度是否在最小/最大长度范围内	符合长度约束比率	≥	80%	是	是	否
	EDQAF-06	自定义正则约束	检测字段内容是否符合自定义正则表达式	符合正则约束比率	≥	80%	否	否	否
	EDQAF-07	电子邮箱规范约束	检测字段内容是否符合邮箱格式	符合电子邮箱约束比率	≥	80%	否	否	否
	EDQAF-08	手机号规范约束	检测字段内容是否符合手机号格式	符合手机号约束比率	≥	80%	否	否	否
	EDQAF-09	身份证号规范约束	检测字段内容是否符合身份证号格式	符合身份证号约束比率	≥	80%	否	否	否
	EDQAF-10	邮编规范约束	检测字段内容是否符合邮编格式	符合邮编约束比率	≥	80%	否	否	否
	EDQAF-20	自定义值范围约束	检测字段内容是否在自定义的值列表内	符合值范围约束比率	≥	80%	否	否	否
	EDQAF-21	码表约束	检测字段内容是否在指定码表的代码值范围内	符合码表约束比率	≥	80%	是	是	否
	EDQAF-22	最大最小值约束	检测数值字段内容是否在最小/最大值范围内	符合最大最小值约束比率	≥	80%	否	否	否
	EDQAF-23	小数位数约束	检测数值字段内容的小数位数是否符合约束	符合小数位数约束比率	≥	80%	是	否	否
	EDQAF-29	表内多列有效性	检测表内多字段间是否符合自定义逻辑关系	符合多列有效约束比率	≥	80%	否	否	否
完整性	EDQAF-41	跨表的数据值比较	检测两表通过关联字段映射后，指定字段值的一致性比率	跨表数据值一致率	≥	90%	否	否	否
	EDQAF-03	跨表的数据格式比较	检测两表间同名字段数据格式（如日期、数值格式）的一致性比率	跨表数据格式一致率	≥	90%	否	否	否
	EDQAF-40	父子参照完整性	检测父表中存在于表无关联子记录的记录比率	父子参照完整率	≥	95%	否	否	否
	EDQAF-39	子父参照完整性	检测子表中存在于父表无关联父记录的记录比率	子父参照完整率	≥	95%	否	否	否
	EDQAF-05	跨表默认值一致性	检测两表间同名字段默认值的一致性差异比率	默认值一致性差异率	≤	10%	否	否	否

参 考 文 献

[1] GB/T 36344—2018 信息技术 数据质量评价指标