ICS 35.240.01
CCS L7

团　　体　　标　　准

T/SZAS XX-2022
T/CSTE XXXX-XXXX

质量分级及“领跑者”评价要求 数据质量增强系统

Assessment requirements for quality grading and forerunner - data quality enhancer system

（征求意见稿）

2022-XX-XX 发布　　2022-XX-XX 实施

深圳市标准化协会

中国技术经济学会

发布

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》和T/CAQP 015-2020、T/ESF 0001-2020《“领跑者”标准编制通则》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。

本文件由深圳市标准化协会和企业标准“领跑者”工作委员会提出

本文件由深圳市标准化协会和中国技术经济学会联合归口。

本文件起草单位：深圳市华傲数据技术有限公司、XXXXXX、深圳市标准化协会

本文件主要起草人：曾新科、何旭珩、陈瑶、龚健、姚晓锋、陈立、贾西贝、但丹

本文件为首次发布。

质量分级及“领跑者”评价要求　数据质量增强系统

1　范围

本文件规定了数据质量增强系统质量及企业标准水平评价的术语和定义、评价指标体系、评价方法及等级划分。

本文件适用于软件和信息技术服务中数据质量增强系统相关质量及企业标准水平评价，相关机构开展质量分级和企业标准水平评估、“领跑者”评价以及相关认证时可参照使用，企业在制定企业标准时也可参照本文件。

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 19001 质量管理体系　要求

GB/T 24001 环境管理体系 要求及使用指南

GB/T 45001 职业健康安全管理体系 要求及使用指南

YD/T 3760-2020 大数据 数据管理平台技术要求与测试方法

3　术语和定义

YD/T 3760-2020界定的以及下列术语和定义适用于本文件。

3.1

数据质量增强系统 data quality enhancer system

提供支持的软件产品，一般包括数据标准管理、质量分析、数据清洗、数据融合、规范化建模和数据质量知识库管理等。

3.2

数据质量 data quality

数据的完整性、规范性、一致性、唯一性和关联性，是进行大数据信息挖掘的前提和保障。

[来源：YD/T 3760-2020，定义2.2.5]

3.3

数据标准data standard

保障数据的内外部使用和交换的一致性和准确性的规范性约束，通常可分为基础类数据标准和指标类数据标准。

注：其中基础类数据标准一般包括参考数据和主数据标准、逻辑数据模型标准、物理数据模型标准、元数据标准、公共代码和编码标准等。指标类数据标准一般分为基础指标标准和计算指标（又称组合指标）标准。基础指标一般不含维度信息，且具有特定业务和经济含义，计算指标通常由两个以上基础指标计算得出。

[来源：YD/T 3760-2020，定义2.2.4]

4 缩略语

下列缩略语适用于本文件。

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

5 评价指标体系

5.1 基本要求

5.1.1 近三年，生产企业无较大及以上环境、安全、质量事故。

5.1.2 企业应未列入国家信用信息严重失信主体相关名录。

5.1.3 企业可根据 GB/T 19001 、GB/T 24001、 GB/T 45001 建立并运行相应质量、环境和职业健康安全，鼓励企业根据自身运营情况建立更高水平的相关管理体系。

5.1.4 产品应为量产产品（/服务应为规模化提供的服务），数据质量增强系统领跑标准应满足国家强制性标准及相关质量增强系统(产品标准)规定的要求。

5.2 评价指标分类

5.2.1 数据质量增强系统质量分级及“领跑者”评价指标体系包括基础指标、核心指标和创新性指标。

5.2.2 基础指标为数据标准管理能力。

5.2.3 核心指标包括质量分析能力、数据清洗能力及数据融合能力;核心指标分为三个等级，包括先进水平，相当于企标排行榜中 5 星级水平；平均水平，相当于企标排行榜中 4 星级水平；基准水平，相当于企标排行榜中 3 星级水平。

5.2.4 创新性指标为规范化建模能力和数据质量知识库管理能力，划分成平均水平和先进水平两个等级，其中先进水平相当于企标排行榜中的 5 星级水平，平均水平相当于企标排行榜中 4 星级水平; 鼓励根据条件成熟情况适时增加与产品性能和消费者关注的相关创新性指标。

5.3 评价指标体系

5.3.1 数据质量增强系统“领跑者”标准评价指标体系框架见表 1。

表 1 数据质量增强系统评价指标体系框架

序号	指标类型	评价指标	指标水平分级			指标来源（判定依据和方法）
			基准水平	平均水平	先进水平	
1	基础指标	数据标准管理能力	以人工管理方式支持数据标准管理，能力描述如下： 1)应明确数据质量相关的数据标准，形成相关文件制度；	利用技术工具提供数据标准管理能力，以保障数据质量管理的规范性，能力描述如下： 1）应明确数据质量相关的数据标准，形成相关文件制度； 2、应具备标准文档、标准数据元、标准术语、标准代码集、版本管理等管理功能。	利用技术工具提供数据标准管理能力，以保障数据质量管理的规范性，能力描述如下： 1）应明确数据质量相关的数据标准，形成相关文件制度； 2）应具备标准文档、标准数据元、标准术语、标准代码集、版本管理等管理功能； 3）应基于技术工具执行标准符合性检测任务，针对每个字段配置字段格式、代码集和规则表达式等检测规则，并自动生成标准符合性检测报告，包含不符合标准的字段等。	市场需求

表 1 数据质量增强系统评价指标体系框架（续）

序号	指标类型	评价指标	指标水平分级			指标来源（判定依据和方法）
			基准水平	平均水平	先进水平	
2	核心指标	质量分析能力	以人工管理方式支持数据质量管理，能力描述如下： 1）支持基于手动编辑规则脚本/SQL/正则表达式等方式，执行数据质量分析任务； 2）应支持数据质量报告的生成，并明确报告的框架，包括内容结构和相关必要元素。	利用技术工具提供数据质量分析能力，应100%达到以下能力描述： 1)应基于技术工具配置数据质量分析的定时任务，支持质量检测字段的选择，以及配置每个字段的质量检测规则； 2)应利用技术工具自动生成数据质量分析报告，包含问题数据明细清单，可根据数据质量分析结果，对数据质量进行评分。	利用技术工具提供数据质量分析能力，应100%达到以下能力描述： 1）应基于技术工具配置数据质量分析的定时任务，支持质量检测字段的选择，以及配置每个字段的质量检测规则； 2）应利用技术工具自动生成数据质量分析报告，包含问题数据明细清单，可根据数据质量分析结果，对数据质量进行评分。 3）应支持对数据质量分析规则的管理与维护。数据质量分析规则包括完整性、唯一性、有效性、一致性、准确性和及时性等通用规则，以及可自定义数据质量分析规则。 4）应支持对数据实体进行质量分析任务配置，可对数据源、表、字段等不同对象进行智能探查，包括： -列分析：对表内字段空值、重复值、长度、频率、结构等信息进行分析，发现数据规律，以及数据质量问题； -比对分析：对不同来源的数据的关联性与一致性进行分析； 5）应基于技术工具，执行数据质量规则智能识别任务。应内置不少3种常用的数据识别规则（包括但不限于值域规则、字典规则和函数依赖规则）。	市场需求

表 1 数据质量增强系统评价指标体系框架（续）

序号	指标类型	评价指标	指标水平分级			指标来源（判定依据和方法）
			基准水平	平均水平	先进水平	
3	核心指标	数据清洗能力	以人工管理方式支持数据清洗任务，能力描述如下： 1）支持基于手动编辑规则脚本/SQL/函数等方式，执行数据质量清洗任务；	利用技术工具提供数据清洗能力，应100%达到以下能力描述： 1）应基于技术工具，添加标准字段列，引入数据清洗所需的标准内容； 2）应将有质量问题的数据，以系统工单形式反馈至相关部门进行处理，并支持问题数据工单的全过程跟踪。 3)应基于技术工具执行数据清洗定时任务，支持标准字段列与数据源字段建立映射配置，同时支持页面配置的方式以实现将标准化清洗规则引入数据清洗任务中；	利用技术工具提供数据清洗能力，应100%达到以下能力描述： 1）应基于技术工具，添加标准字段列，引入数据清洗所需的标准内容； 2）应将有质量问题的数据，以系统工单形式反馈至相关部门进行处理，并支持问题数据工单的全过程跟踪。 3）应基于技术工具执行数据清洗定时任务，支持标准字段列与数据源字段建立映射配置，包括配置源与标准数据元间代码映射、数据清洗规则和多函数算法包等。同时支持页面配置的方式以实现将标准化清洗规则引入数据清洗任务中。 4）应提供混合规则配置能力，支持混合搭配代码映射、清洗规则、多函数算法包能力，提供标准表配置标准列、字段映射、业务时间配置、质量稽查规则后，自动生成SQL脚本，支持对自动生成的SQL脚本自定义修改功能； 5）应具备低代码式的数据集成开发能力，支持以页面设置方式，实现开启表拉链能力；同时支持在不同数据分层之间，进行数据表快速复制的能力；通过可视化，自动展现源表与目标表间的关联关系，包括表关联、字段关联等。	市场需求

表 1 数据质量增强系统评价指标体系框架（续）

序号	指标类型	评价指标	指标水平分级			指标来源（判定依据和方法）
			基准水平	平均水平	先进水平	
4	核心指标	数据融合能力	以人工管理方式支持数据融合任务，能力描述如下： 1)支持基于手动编辑规则脚本/SQL/函数等方式，执行多源数据融合任务。	利用技术工具提供多源数据融合能力，应 100%达到以下能力描述： 1）应基于技术工具，提供数据模型管理的能力，包括数据模型的设计、展示、导出建表 SQL 以及逆向工程等。 2）应通过可视化页面设置方式，支持源字段与目标字段建立映射配置； 3、应内置不少于 3 种常用的数据融合策略，宜支持无需编写 SQL 代码方式，通过可视化操作实现数据融合策略的配置。	利用技术工具提供多源数据能力，应 100%达到以下能力描述： 1）应基于技术工具，提供数据模型管理的能力，包括数据模型的设计、展示、导出建表 SQL 以及逆向工程等。 2）应通过可视化页面设置方式，支持源字段与目标字段建立映射配置； 3）应内置不少于 6 种常用的数据融合策略（包括但不限于一数一源、来源优先、数据新鲜度、最大值、最小值、最高频），宜支持无需编写 SQL 代码方式，通过可视化操作实现数据融合策略配置。 4）应支持基于技术工具，自动化一键生成数据融合任务的工作流，包括自动生成数据融合任务间上下游任务的依赖关系。支持数据融合定时任务配置。 5）应提供以页面配置方式实现数据融合任务开发，包括表关联、字段映射等常规操作，自动根据可视化页面生成 SQL 脚本，并具备自定义修改功能； 6）应具备低代码式的数据集成开发能力，支持以页面设置方式，实现开启表拉链能力；同时支持在不同数据分层之间，进行数据表快速复制的能力；通过可视化，自动展现源表与目标表间的关联关系，包括表关联、字段关联等。	市场需求

表 1 数据质量增强系统评价指标体系框架（续）

序号	指标类型	评价指标	指标水平分级			指标来源（判定依据和方法）
			基准水平	平均水平	先进水平	
5	创新性指标	规范化数据建模能力	以人工管理方式支持数据建模任务，能力描述如下： 1）应支持用户以人工方式按国内外现行标准文件要求，执行数据模型，表结构和数据元字段设计任务。	利用技术工具提供规范化数据建模能力，应 100%达到以下能力描述： 1）基于技术工具，支持数据模型实体及属性的标准创建、修改、删除和查询；支持数据模型标准与数据标准的映射； 2）应具备标准字段库管理功能，包括但不限于： -对标准表管理，并实现标准表与标准字段关联； -对标准字段管理，并实现标准字段与标准代码集关联； -对标准代码集管理；	利用技术工具提供规范化数据建模能力，应 100%达到以下能力描述： 1）应提供规范化建模管理能力，支持概念模型、逻辑模型、物理模型的建模设计；支持将标准表快速引入，以完成概念模型、逻辑模型、物理模型的数据表创建；支持将标准字段快速引入到数据表中创建字段；支持手工创建字段时，自动联想并提供可选标准字段能力。 2）应具备标准字段库管理功能，包括但不限于： -对标准表管理，并实现标准表与标准字段关联； -对标准字段管理，并实现标准字段与标准代码集关联； -对标准代码集管理； 3）应提供规范化建模检查能力，检查数据模型中各表、字段的标准化符合率。	市场需求
6		数据质量知识库管理能力	以人工方式管理知识库，能力描述如下： 1）应支持以人工方式建立数据质量知识库，对标准文档、标准数据元、标准代码集、标准术语、规则表达式等知识进行管理。	利用技术工具提供数据质量知识库管理能力，应 100%达到以下能力描述： 1）应基于技术工具执行知识库管理工作，包括标准文档、标准数据元、标准代码集、标准术语、规则表达式等管理； 2）支持用户对知识库执行检索、归纳分类以及知识分发等操作。	利用技术工具提供数据质量知识库管理能力，应 100%达到以下能力描述： 1）应基于技术工具执行知识库管理工作，包括标准文档、标准数据元、标准代码集、标准术语、规则表达式等管理； 2）支持对知识库执行增加、删除、修改、归纳分类、知识分发等操作。 3）支持在数据质量增强系统的其他功能模块中将相关知识手动同步更新到知识库。 （1）质量分析规则在知识库积累：新配置的数据质量分析规则，可以手动更新到知识库。 （2）制定的标准内容在知识库积累：新配置的标准文档、标准数据元、标准代码集、标准术语，可以手动更新到知识库。	市场需求

6 评价方法及等级划分

评价结果划分为一级、二级和三级，各等级所对应的划分依据见表2。达到三级要求及以上的企业标准并按照有关要求进行自我声明公开后均可进入数据质量增强系统企业标准排行榜。达到一级要求的企业标准，且按照有关要求进行自我声明公开后，其标准和符合标准的产品或服务可以直接进入数据质量增强系统企业标准“领跑者”候选名单。

表 2 指标评价要求及等级划分

评价等级	满足条件			
一级应同时满足	基本要求	基础指标要求	核心指标先进水平要求	创新性指标至少有 2 项达到先进水平要求
二级应同时满足	基本要求	基础指标要求	核心指标平均水平要求	创新性指标至少有 1 项达到平均水平要求
三级应同时满足	基本要求	基础指标要求	核心指标基准水平要求	—