

T/BSIA

团 体 标 准

T/BSIA XXXX—XXXX

智慧城市数据中台建设规范

Code for the construction of smart city data element platform

（征求意见稿）

（本草案完成时间：2023年2月24日）

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

北京市软件和信息服务业协会 发布

目 次

前言	II
引言	III
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 体系架构	1
6 功能要求	2
6.1 数据集成	2
6.2 数据开发	3
6.3 数据治理	3
6.4 资产管理	7
6.5 数据开放	8
6.6 数据分析	9
6.7 运维监控	9
7 数据安全	10
8 运维要求	10
8.1 人员要求	10
8.2 运维内容要求	10

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××××提出。

本文件由××××归口。

本文件起草单位：

本文件主要起草人：

引 言

本文件的发布机构提请注意，声明符合本文件时，可能涉及到本文件[5.1条]与[一种视频监控集群的控制方法、装置、设备及存储介质]、本文件[5.2条]与[一种多关系型数据引擎的SQL合并执行系统及方法]、本文件[5.3条]与[跨集群数据处理方法、装置、电子设备以及存储介质]、本文件[5.4条]与[一种异常血缘关系的确定方法、装置、设备及介质]相关的专利的使用。

本文件的发布机构对于该专利的真实性、有效性和范围无任何立场。

该专利持有人已向本文件的发布机构承诺，他愿意同任何申请人在合理且无歧视的条款和条件下，就专利授权许可进行谈判。该专利持有人的声明已在本文件的发布机构备案。相关信息可以通过以下联系方式获得：

专利持有人姓名：云粒智慧科技有限公司。

地址：北京市西城区阜成门外大街甲28号西楼13层06室。

请注意除上述专利外，本文件的某些内容仍可能涉及专利。本文件的发布机构不承担识别专利的责任。

智慧城市数据中台建设规范

1 范围

本文件规定了智慧城市数据中台建设的体系架构、功能要求、数据安全、运维要求等内容。本文件适用于智慧城市数据中台整体规划及具体项目的规划、设计、建设与运维。

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 25069 信息安全技术 术语
- GB/T 37043 智慧城市 术语
- GB/T 37973 信息安全技术 大数据安全管理指南
- GB/Z 38649 信息安全技术 智慧城市建设信息安全保障指南
- GB/T 39477 信息安全技术 政务信息共享 数据安全技术要求

3 术语和定义

GB/T 25069、GB/T 37043界定的以及下列术语和定义适用于本文件。

智慧城市数据中台 smart city data element platform

智慧城市数据中台是一套可持续的让数据用起来机制，一种战略选择和组织形式，是依据特有的业务模式和组织架构，通过有形的产品和实施方法论支撑，构建一套持续不断把数据变成资产并服务于业务的机制。

注：数据中台具备异构数据统一计算、存储的能力，同时让分散杂乱的数据通过规范化的方式管理起来。数据中台借鉴了传统数仓面向主题域的数据组织方式，基于维度建模理论，构建统一的数据公共层和应用层。数据中台依赖于大数据平台完成数据研发全流程，同时增加了数据治理和数据服务化以及数据资产内容，是数据要素服务的基础能力平台。

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）

SSH：专为远程登录会话和其他网络服务提供安全性的协议（Secure Shell）

VM：虚拟机主机（Virtual Manufacturing）

Topic：类似于一个频道，A在里面发消息，B订阅这个频道，然后从topic中获取A发送的信息

5 体系架构

智慧城市以连接、算法、分析驱动数据要素流转，深刻重塑城市经济生产组织方式和社会活动运行流程，推动要素更加高效配置，带动城市产业、治理服务全面升级。智慧城市数据中台体系架构应包括数据集成平台、数据治理平台、资产管理平台、数据开放平台、数据开发平台、数据分析平台、数据安全平台与运维监控平台等。智慧城市数据中台体系架构见图1。

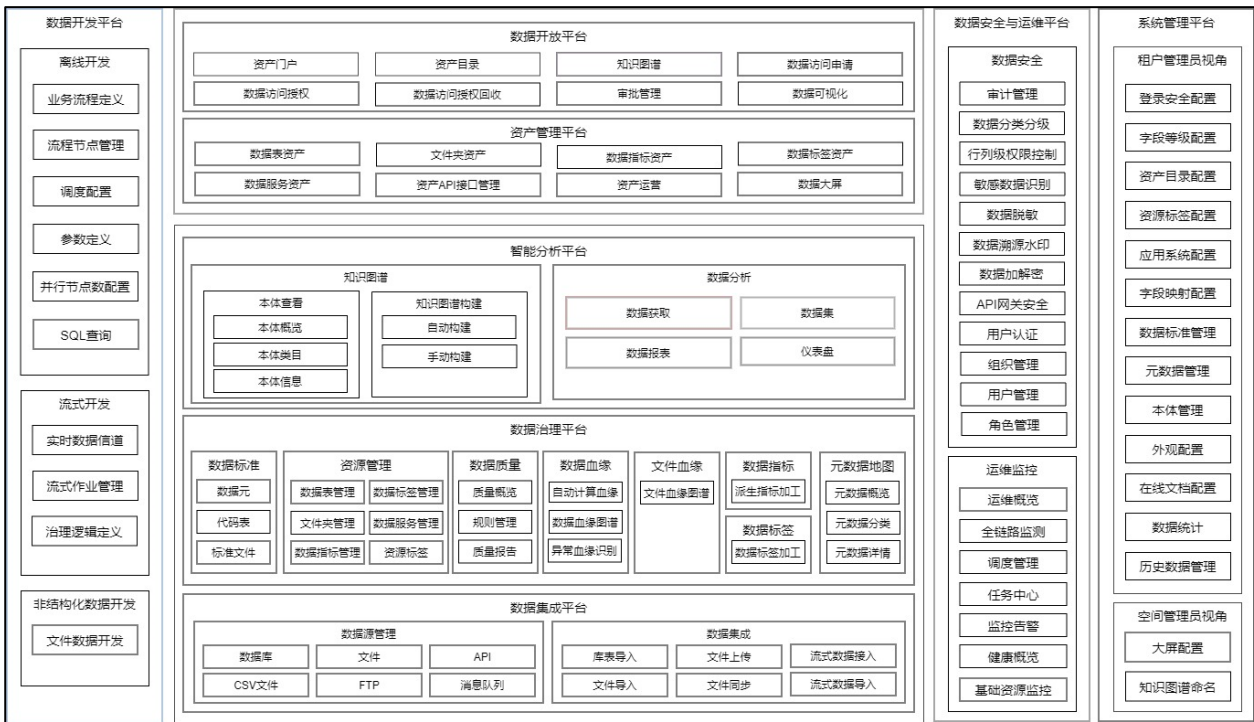


图1 智慧城市数据中台体系架构

6 功能要求

6.1 数据集成

6.1.1 概述

数据集成是指利用数据抽取、消息服务、数据交换、填报采集等技术手段，把各种异构网络、异构数据源、不同类型的数据采集到数据中台进行集中存储。数据集成是数据中台数据接入的入口，数据集成的方式有结构化数据接入、非结构化数据接入、API数据接入、流式数据接入、数据源接入、实时信道数据接入等。

6.1.2 结构化数据接入

6.1.2.1 库表导入

数据开发者通过定义库表导入作业完成结构化数据导入中台。库表导入为周期性作业，宜支持通过选择来源数据库的表，将数据同步至目标数据表。

6.1.2.2 CSV 文件导入

数据开发者通过定义文件导入作业完成数据入中台。CSV文件导入为一次性作业，应支持通过上传CSV、Excel文件，将数据导入至目标数据表。

6.1.3 半结构化数据导入

数据开发者通过定义文件导入作业完成数据入中台。半结构化数据导入为一次性作业，应支持通过上传XML、HTML、JSON文档，将数据导入至目标数据表。

6.1.4 非结构化数据接入

6.1.4.1 文件上传

数据开发者通过定义文件上传作业完成数据入中台。文件上传为一次性作业，应支持通过上传文件夹或批量上传多个文件，将文件数据同步至目标文件夹。

6.1.4.2 文件拉取

数据开发者通过定义文件拉取作业完成数据入中台。文件拉取为周期性作业，应支持通过选择多级文件夹下的文件做增量同步、覆盖同步。

6.1.5 API 数据接入

数据开发者通过定义API接入作业完成数据入中台。应支持http协议、json格式数据的接入，并将API获取的数据同步至目标topic。

6.1.6 流式数据接入

数据开发者通过定义流式数据接入作业完成流式数据离线导入中台。流式数据接入为周期性作业，支持通过选择来源实时数据信道将数据同步至目标表。

6.1.7 数据源

智慧城市数据中台数据库连接配置应支持MySQL、SQLServer、PostgreSQL、Oracle、达梦、人大金仓、瀚高等常用数据库，数据源连接信息的配置管理，包括增加、删除、查找、修改、测试连通性。

6.1.8 实时信道数据

智慧城市数据中台实时信道数据是管理流式数据的载体和通道，应能对实时数据信道进行管理，包括对Topic的增加、删除、修改、查找。

6.2 数据开发

6.2.1 概述

数据开发是一套数据加工并对加工过程管控的工具，数据开发人员可以利用数据开发模块提供的功能，把数据加工成对业务有价值的形式并供给业务使用。

智慧城市数据中台数据开发模块面向不同的数据处理场景，便捷地完成数据开发，实现离线数据、实时数据的快速融合处理，结合原生的SQL探索能力，大幅提高数据开发效率并降低数据开发成本。

6.2.2 离线开发

6.2.2.1 业务流程

业务流程指针对业务实体，抽象出业务流程的概念，从业务视角组织数据的开发治理，提高任务管理效率。

6.2.2.2 SQL 查询

基于SQL语言开发模型是数仓数据治理的核心，模型开发的过程中，数据建模人员需要先探查数仓中有哪些数据再进行建模。同时，当有些数据导入数仓后，发现有误，数据开发者无需配置业务流程，即可快速完成数据修正。

SQL查询应可支持查询数据、插入数据、清空表、添加/删除/查询分区。

6.2.3 流式开发

流式开发基于实时流计算引擎，对无边界数据集进行连续不断的处理、聚合和分析，获取结果具备低延迟性，可有效控制在毫秒或秒级别。

流式数据开发应支持基于模型、规则，计算生成事件，并将处理结果输出到指定目标源。

6.2.4 非结构化数据开发

非结构化数据开发宜支持以可视化方式呈现非结构化数据治理的全链路，便于文件数据的问题定位与溯源管理。

6.3 数据治理

6.3.1 概述

数据治理是从根本上建立数据标准和规范，统一数据资产管理，统一数据调度，保证数据可控、可用、可信的工程；同时，以数据运营的方式，从政务数据和社会数据中提升社会效益和经济效益，为数字政府和智慧城市的建设提供可信、可用的数据支撑。数据治理指数据中台依托标准规范体系，对归集的数据进行清洗整合、主题建模、指标定义、标签生产、质量检核、血缘计算等操作，以提升数据质量、挖掘数据价值。

数据治理的主要内容包括元数据管理、数据质量、数据血缘、数据指标。

6.3.2 元数据管理

6.3.2.1 元数据采集

元数据采集宜支持自动扫描数据中台各数据源中的数据表结构信息，并识别表结构变更信息，完成数据源元数据信息的采集汇聚。

6.3.2.2 元数据版本

元数据版本宜支持查询、以列表形式展示元数据版本信息，每个来源系统对应列表中的一条版本信息、查看该采集作业对应的元数据历史版本。

6.3.2.3 元数据类目

元数据采集完成后，宜自动生成元数据类目。

6.3.2.4 元数据地图

元数据地图宜有序视图呈现系统所有已采集元数据与数据中台数据仓的所有数据表，进行信息整合，支持展示元数据的全域信息。包括但不限于元数据列表、数据表分布、数据量分布。

元数据地图宜根据元数据采集作业自动生成元数据目录，目录名称宜为被采集元数据的数据源名称。每个数据表宜可显示以下内容：

- 数据表中英文名；
- 共享次数；
- 参考依据；
- 字段总量；
- 数据总量；
- 存储容量；
- 更新时间。

6.3.3 数据质量

6.3.3.1 概述

在实际生产中，数据计算任务没有告警，但不代表数据就是正确的，比如源数据异常、代码逻辑修改等原因都会造成结果数据错误。数据质量聚焦规则管理、数据扫描、发现问题，实现对组织数据质量的有效监管，为数据正确性保驾护航。

6.3.3.2 质量概览

智慧城市数据中台应支持数据提供质量概览，便于从全局了解组织数据质量全貌，具体内容如下：

——重点关注：

- 数据表总数；
- 已质检数据表数；
- 质检覆盖率；
- 质检不合格表数；
- 质检不合格率。

——规则应用趋势：

统计数据表总数、质检数据表数、质检不合格表数的变化趋势，直观呈现质检工作的覆盖情况以及数据质量的变化趋势。

——部门质检覆盖率：

对各部门数据的质检覆盖率进行统计，以督促质检覆盖率低的部门开展数据质检工作。

——部门质检不合格率：

对各部门数据的质检不合格率进行统计，以督促质检不合格率高的部门提升数据质量。

6.3.3.3 规则管理

6.3.3.3.1 概述

6.3.3.3.1.1 质量规则是数据质量管理的基础。规则管理包括：告警设置、分区表达式配置、质量规则配置、调度配置。

6.3.3.3.1.2 质量管理列表应能展示中台所有的数据表，以对数据表配置质量规则。质量管理列表包括但不限于以下内容：

——数据表英文名；

——数据表中文名；

——责任人；

——规则更新时间；

——有效质量规则数；

——校验结果；

——最新校验时间；

——配置规则；

——质量报告；

——其它。

6.3.3.3.2 告警设置

如果质量规则检验的结果出现异常，应有质量告警信息通知相关人员。

6.3.3.3.3 质量规则配置

6.3.3.3.3.1 新建质量规则

新建质量规则应支持内置规则与自定义规则两种方式，具体如下：

——内置规则：

- 规则名称；
- 规则类型；
- 对象类型；
- 分区表达式；
- 比较方式；
- 期望值；
- 告警等级；
- 内置规则类型。

——自定义规则：

- 规则名称；
- 规则类型；
- 对象类型；
- 分区表达式；
- 采样方式；
- 过滤条件；
- 比较方式；
- 期望值；
- 告警等级。

6.3.3.3.3.2 编辑质量规则

针对质量规则不同的状态，应可支持如下操作：

- 查看任务；
- 修改规则；
- 删除规则；
- 试运行；
- 上线；
- 下线；
- 调度配置。

6.3.3.3.4 调度配置

显示生成该数据表的业务流程的调度信息，作为定义质检调度周期的参考。

6.3.3.4 数据扫描

数据扫描是按既定质检规则对数据进行扫描质检。支持周期调度触发质检。调度周期指数据表质检执行的频率，如：分钟、小时、日、周、月。支持对同一数据表的不同质检对象定义按不同的调度周期进行数据扫描。

6.3.3.5 质量报告

智慧城市数据中台应支持生成数据质量报告，数据质量报告应由以下三部分构成：

- 数据表基础信息；
- 质量分析结果；
- 报告详情。

6.3.4 数据血缘

6.3.4.1 概述

数据血缘指数据产生的链路关系，即数据的产生、加工融合、流转流通等到最终消亡，数据之间自然会形成的一种关系。建立在数据融合之上，自动计算血缘关系，实现以数据流向为主线的血缘追溯，为数据的影响分析、溯源分析奠定坚实基础。

6.3.4.2 血缘图谱

智慧城市数据中台应能自动计算形成全局的数据血缘图谱，血缘图谱旨在形象、直观地将数据血缘关系进行展示，方便数据开发者探索数据问题本质，及时定位问题，辅助数据治理。

6.3.4.3 影响分析

一个数据源头的核心业务的改动，下游各分析应用应保持同步。当系统进行升级改造时，如果修改了数据结构、ETL程序等元数据信息，依赖数据血缘的影响分析，可以快速定位出元数据修改会影响到哪些下游数据，从而减少系统升级改造带来的风险。

6.3.4.4 溯源分析

数据的血缘关系，应能体现数据的来龙去脉，帮助追踪数据的来源和处理过程。当数据发生异常，数据血缘提供追溯数据的来源和完整的数据处理链路，辅助定位异常发生的原因，把风险控制在适当的水平。

6.3.5 数据指标

6.3.5.1 概述

数据指标是业务运营过程中，按照既定业务逻辑对数据进行加工处理，使之成为描述、衡量、分析、预测业务结果的基础原材料。

智慧城市数据中台数据指标应能提供统一的指标元数据管理，维护指标类目，定义指标名称、计算口径等，以有效满足跨业务应用对指标的使用诉求。

6.3.5.2 指标类目

指标类目指通过对指标打标签，抽象标签集的名称、设置标签之间的关系来形成复杂的指标类目结构。

6.3.5.3 指标管理

定义指标元数据对指标进行管理维护，包括但不限于：增加、删除、查找、修改、变更、发布/批量发布、下架/批量下架、关联类目。

6.4 资产管理

6.4.1 资源目录

6.4.1.1 概述

数据资产的最大特点是具有重复利用性和增值性，有效的数据能为组织创造巨大价值。数据中台提供资产目录，以资产目录的形式提供了数据资产展示的窗口，同时支持不同部门间数据的共享交换，让数据在不同的部门中流通，从而创造出更大的价值。能够支持构建不同类目的数据资产，包括基础数据、主题数据等。

6.4.1.2 资产搜索

智慧城市数据中台应有数据资产搜索功能。

6.4.1.3 资产详情

智慧城市数据中台应有数据资产详情展示功能。

6.4.2 数据大屏

智慧城市数据中台应具备数据可视化功能，为用户提供可视化的资产视图，具体功能如下：

——数据看板：

- 数据总量；
- 数据接入总量；
- 治理产生的数据量；
- 质检合格率；
- 服务部门数；
- 数据被消费次数。

——数据接入：

- 接入对象统计；
- 数据接入量变化趋势。

——数据治理：

- 以血缘链路的形式呈现。

——数据消费：

- 访问热度排名；
- 资产目录。

6.4.3 类目管理

通过对数据资产打标签，抽象标签集的名称、设置标签之间的关系来形成复杂的类目结构。按照不同的业务需求，可形成的多套类目视图。类目管理包括：对类目进行增加、删除、查找、修改、发布、下架。

类目由数据中台管理员管理维护，支持数据开发者以类目为核心批量关联数据资产，具体内容如下：

- 类目新建：由管理员创建类目，包括：类目名称、上级类目；
- 类目修改：由管理员对类目进行修改，包括：类目名称、上级类目；
- 类目删除：由管理员对未发布的类目进行删除；
- 类目发布：由管理员对未发布的类目进行类目发布，支持批量发布类目；

- 类目下架：由管理员对已发布的类目进行下架，支持批量下架类目；
- 类目关联资源：由数据开发者通过某个已发布的类目进行关联资源的操作，支持批量关联资源；
- 类目移动：在左侧类目树上选择某类目，出现上移和下移的图标，直接點選操作；
- 类目查询：通过类目名称、状态、是否关联资源进行组合查询，查询结果显示在结果表中；
- 类目查看：查看类目详情，包括：类目基础信息、类目关联数据资源信息。

6.4.4 资源管理

6.4.4.1 数据表管理

智慧城市数据中台资产管理应支持对数据表进行管理，包括数据表增加、删除、修改、发布、变更、下架、关联类目、文件导入、查询、查看等操作。

6.4.4.2 文件夹管理

智慧城市数据中台资产管理应支持对数据文件夹进行管理，包括文件夹增加、删除、查找、修改、变更、发布、下架、关联类目的管理。

6.4.5 资源标签

6.4.5.1 概述

智慧城市数据中台资产管理应支持资源标签的配置，资源标签通过对标签类型和标签进行维护管理，以满足不同项目在创建资源所需属性字段不同的需求。

6.4.5.2 标签类型

6.4.5.2.1 主要功能

应支持对标签类型进行新建、修改、删除、查询。

6.4.5.2.2 标签类型新建

应支持创建标签类型，填写标签类型基础信息，包括：标签类型名称、描述等；标签类型名称不应重复。

6.4.5.2.3 标签类型修改

应支持对标签类型进行修改，包括：标签类型名称、描述。

6.4.5.2.4 标签类型删除

应支持对未关联标签的标签类型进行删除。

6.4.5.2.5 标签类型查询

应支持通过标签类型名称、是否关联标签进行组合查询，查询结果显示在结果列表中。

6.4.5.3 标签

智慧城市数据中台资产管理应支持对数据标签进行新建、修改、删除、查询等操作，具体内容如下：

- 标签新建：创建新的标签，同一标签类型下的标签不应重复；
- 标签修改：对标签进行修改；
- 标签删除：对未关联资源的标签进行删除；
- 标签查询：应支持通过标签名称、标签类型进行组合查询。

6.5 数据开放

6.5.1 概述

智慧城市数据中台数据开放是把数据变成一种服务能力，通过数据服务让数据资产参与到业务，激活整合数据中台，是数据中台价值重要体现之一。通过安全、可控、有序的措施保障向数据消费者开放、共享数据，充分挖掘数据资源“宝矿”，发挥最大效益，提升政府治理能力和公众服务水平。

6.5.2 数据资产目录

数据资产的最大特点是具有重复利用性和增值性，有效的数据能为组织创造巨大价值。数据中台提供资产目录，以资产目录的形式提供了数据资产展示的窗口，同时支持不同部门间数据的共享交换，让数据在不同的部门中流通，从而创造出更大的价值。

资产目录是实现信息资源共享、业务协同和数据开放的基础，是各部门之间信息共享及数据向社会开放的依据。资产目录便于业务人员快速定位所需数据资产，多元化的资源共享方式，助力组织实现不同类型数据资产的共享。

6.5.3 资产访问授权

6.5.3.1 数据表访问授权

应支持对数据表的访问申请进行审批授权。

6.5.3.2 文件夹访问授权

应支持对文件夹的访问申请进行审批授权。

6.5.4 资产访问授权回收

6.5.4.1 数据表访问授权回收

应支持对已授权的数据表进行授权回收。

6.5.4.2 文件夹访问授权回收

应支持对已授权的文件夹进行授权回收。

6.5.5 审批管理

智慧城市数据中台数据开放平台应具备数据审批管理功能。

注：数据访问申请、访问授权、访问授权回收都涉及在线工作流审批，通过工作流完成数据资产访问申请、访问授权以及访问授权回收的审批管理，提供个人工作台以统一查看待审批的、已审批的、已申请的等相关信息。

6.6 数据分析

智慧城市数据中台应具备数据分析功能，应支持但不限于以下操作：

- 数据获取：数据集作为数据源和可视化展示的中间环节，承接数据源的输入，并为可视化展示输出数据表；
- 数据报表：支持在线定义报表，含明细表、分组表、交叉表等；
- 可视化分析：通过拖拉拽的形式，将图表（折线图、柱状图等）拖到仪表板布局里，通过自定义布局构建数据分析仪表板，支持开展可视化数据分析。

6.7 运维监控

6.7.1 概述

运维监控是数据中台得以健康、持续运转的基础。智慧城市数据中台运维监控平台应能对数据全生命周期的监控、告警，便于迅速定位和排除故障，保障业务正常稳定运行。

6.7.2 运维概览

智慧城市数据中台运维监控平台应具备运维概览功能，运维概览宜支持展示包括但不限于以下内容：

- 需要重点关注的运维指标；
- 任务运行状态分布；

- 任务完成情况；
- 运行时长排行；
- 任务出错排行；
- 调度任务数量趋势；
- 监控日志管理；
- 任务类型分布。

6.7.3 调度管理

智慧城市数据中台运维监控平台应具备调度管理功能，对数据处理任务进行监控、告警，便于迅速定位和排除故障，保障业务正常稳定运行。

调度管理信息列表应包括但不限于以下内容：

- 作业名称；
- 作业类型；
- 状态；
- 调度周期；
- 近 1 周期任务数。

6.7.4 任务中心

任务中心集中纳管中台的所有任务，包括数据集成、SQL开发以及数据质量，对任务做统一呈现，便于查看执行失败的任务并定位任务失败原因。其中任务状态宜分为以下阶段：

- 等资源；
- 运行中；
- 运行成功；
- 运行失败；
- 运行超时；
- 运行过期。

7 数据安全

智慧城市数据中台完善的数据安全能力，提供数据分类分级、敏感数据识别、数据加解密、API安全网关、危险操作认证、数据水印溯源等能力，保证数据的隐私合规、可审计、可回溯，多重安全保护机制，全方位保障整数据安全。

平台宜具备隐私计算统一接入能力，支持以主流隐私计算技术完成高价值敏感数据服务，保障数据要素安全有序流通。

智慧城市数据中台数据安全平台应符合GB/T 37973、GB/T 39477、GB/Z 38649的要求。

8 运维要求

8.1 人员要求

智慧城市数据中台运维人员应符合以下要求：

- 熟悉 Linux 系统相关的管理工具（如 grep、sed、awk 等）及 shell、Python 等计算机语言；
- 应具备 MySQL、postgresql、国产信创数据库等关系型数据库操作技能；
- 计算机网络、操作系统等基础；
- 熟悉开源数据平台及相关技术，包括但不限于 Hadoop、MapReduce、redis、Hive、Hbase。

8.2 运维内容要求

智慧城市数据中台上线试运行完成后，应进行业务应用情况监控检查；当发现业务或数据库读写不正常时应及时处理。主要检查内容如下：

——系统检查：

- 检查系统日志是否有告警；
- 检查CPU、内存、网络等系统性能是否异常。

——业务检查：

- 定期巡检，检查数据应用业务是否正常。
-