

解决方案实践

华为云政务大数据解决方案实践

文档版本 1.1
发布日期 2024-05-06



版权所有 © 华为技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 方案概述	1
2 资源和成本规划	6
3 数据工程实施流程	10
4 实施步骤	12
4.1 组建专班	12
4.2 数据架构规划设计	13
4.2.1 数据调研	13
4.2.2 业务域 L1 和主题域 L2 定义	15
4.2.3 梳理业务流程	16
4.2.4 业务流程活动	18
4.2.5 业务对象 L3 梳理和数据 owner 确定	21
4.2.6 逻辑实体 L4 设计和数据来源	27
4.2.7 属性 L5 设计和数据分级	31
4.2.8 数据标准设计	34
4.2.9 数据指标设计	36
4.3 数据集成	38
4.3.1 数据集成概述、工具介绍及入湖范围	38
4.3.2 数据集成前探查	39
4.3.3 全量数据同步	39
4.3.4 增量数据同步	44
4.4 规范设计	49
4.4.1 配置中心	49
4.4.2 流程设计	52
4.4.3 主题设计	54
4.4.4 码表管理	56
4.4.5 数据标准	59
4.4.6 关系建模	62
4.4.7 业务指标	64
4.4.8 技术指标	67
4.4.9 维度建模	69
4.5 数据开发	72
4.5.1 概述与连接管理	72

4.5.2 脚本开发.....	73
4.5.3 作业开发.....	75
4.5.4 实例管理、补数据监控及配置环境变量.....	77
4.6 数据质量.....	78
4.7 数据资产.....	86
4.8 数据服务.....	90
4.8.1 创建专享版集群.....	90
4.8.2 设置 API 分配配额.....	91
4.8.3 新建审核人.....	92
4.8.4 开发 API.....	93
4.8.5 调用 API.....	95
5 修订记录.....	98

1 方案概述

应用场景

2022年9月13日国务院办公厅印发《全国一体化政务大数据体系建设指南》，明确建设目标：2023年底前，全国一体化政务大数据体系初步形成，基本具备数据目录管理、数据归集、数据治理、大数据分析、安全防护等能力，数据共享和开放能力显著增强，政务数据管理服务水平明显提升。到2025年，全国一体化政务大数据体系更加完备，政务数据管理更加高效，政务数据资源全部纳入目录管理。并明确统筹管理一体化、数据目录一体化、数据资源一体化、共享交换一体化、数据服务一体化、算力设施一体化、标准规范一体化、安全保障一体化八大建设任务。但当前政务大数据普遍存在如下问题：

1. 体制机制缺失：尚未建立政务数据治理体制机制，各级数据资源管理部门和业务部门职责分工不明确，无法持续性开展数据治理工作。
2. 业务目标不清晰：政务数据治理工作仅围绕数据开展，并未以业务目标为导向，围绕相应的业务目标开展数据治理实施工作，导致数据治理价值无法体现
3. 缺乏统一数据标准：政务信息化建设存在“各自为政、条块分割、烟囱林立、信息孤岛”等问题，由此导致数据标准管理分散，数据标准执行力度差。
4. 数据资源底数不清：各部门仅针对国办要求进行政务信息资源目录进行梳理，并未对部门全量数据进行盘点，无法做到“心中有数”。
5. 数据质量亟待提升：虽已建成了统一的数据共享交换平台，但各部门对数据质量的重视程度不高，数据质量把控不严，导致数据错、重、漏现象频现。

通过建设一套包含大数据MRS、数据仓库DWS、数据治理DataArts Studio的数管平台，采用湖仓一体的先进技术架构，兼顾权限管控和资源隔离的能力，构建省市资源中枢和能力底座；同时引入《华为数据之道》，以业务对象为核心，进行数据治理，最终实现：

1. 理清资产：盘清全省（或市本级）政府数据家底，按需归集政务数据与公共数据；
2. 高效共享：持续提升数据质量，完善数据管理，完成国家对数据开放共享的各项要求，实现高质量数据的共享和使用，实现跨区域、跨部门、跨层级的数据融合，完善数据统筹协调的技术基础；
3. 业务创新：支撑政务服务实现“秒批秒办”“全程网办”“跨省通办”“一件事一次办”等；智能分析跨部门、跨行业、跨地域监管数据，支撑风险预警、决策分析、重点事件跟踪，联合监管等业务。实现舆情洞察，全域感知、情感分析、趋势研判等能力，支撑政府科学决策。

方案架构

本文着重介绍大数据平台技术架构和数据工程总体架构。

华为政务大数据方案在客户已建的数据交换平台的基础上，构筑5平台5体系。

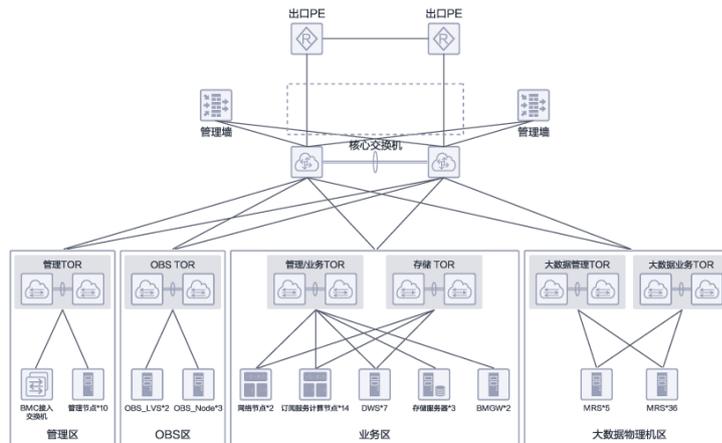
图 1-1 业务架构图



5平台包括算力设施平台、共享交换平台（数据归集平台）、数据使能平台，数据资源平台、数据服务平台。

- 算力设施平台提供统一的数据湖功能，为数据存储与计算提供使用基础。依托云底座以及MRS、DWS数据平台，提供大数据的计算分析存储能力。
- 共享交换平台在已有的数据共享交换平台基础上，新增数据集成能力，为政务场景各种类型的数据提供归集接入的功能，将数据入湖。通过DATAARTS STUDIO CDM批量同步、MRS CDL/Kafka实时同步等提供同构/异构数据源之间数据迁移服务，实现数据自由流动。
- 数据使能平台致力于管理政务数据，实现数据“找得到、管得住、用得好”。依托DATAARTS STUDIO服务平台，实现数据模型的规范设计统一、数据开发、数据质量以及数据资产管理。
- 数据资源平台将政务数据目录化，建设政务基础库。通过MRS/DWS平台批处理、流式处理能力实现归集库->中心库->基础资源库的建设。
- 数据服务平台将湖内数据开放，依托ROMA Connect平台的API接口能力，开放给上层应用使用。

图 1-2 部署架构图



架构描述：

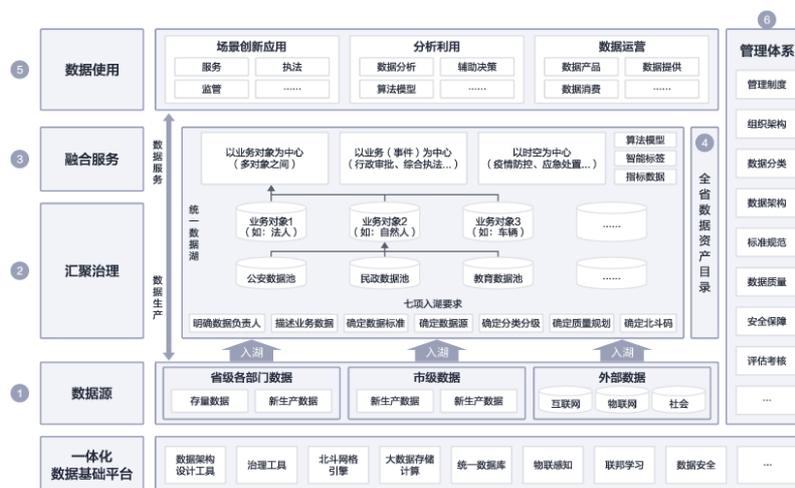
1. 整体方案交付内容包括HCS 基础云服务+MRS+DWS+DataArts Studio。
2. 大数据物理机区也可与业务区部署在一起，当前仅仅是提供了一种部署方案，各局点可根据现网情况进行部署，主要参考集成设计指导书为主。

5体系包括集成及交付验证体系，运维保障体系，业务运营体系，安全保障体系及标准规范体系。集成及交付验证体系为政务大数据平台的设计、交付、开发、验证的标准体系。可复制可实施。运维保障体系为政务大数据平台的监控能力，提供平台的运维监控以及可视化的能力。业务运营体系为业务统一的运营流程，提供基础的运营管理。安全保障体系为平台具备的各个层次的安全能力，保障平台安全不受侵。标准规范体系为平台遵循的统一规范，保障平台按照统一标准设计实施。

在5大体系和5平台的统一支撑下，构建形成以人口、法人、宏观经济、地理信息、电子证照等公共基础数据库和各类主题数据库为基础，推动各级各部门政务数据信息向的政务大数据共享开放平台迁移集聚，发挥政务大数据支撑多部门协同服务的作用，简化优化群众办事流程，尽可能利企便民，打造公共服务和社会治理的新模式，带动地方经济发展，同时支撑更加科学的政府管理研判和决策机制。

数据工程总体框架主要包括数据来源、汇聚治理、融合服务、数据资产目录、数据使用、管理体系等六个方面。

图 1-3 数据工程总体框架



- ①数据来源：包括政务部门数据和外部数据，政务部门数据分为存量数据和增量数据，外部数据包括互联网数据、物联网数据和社会数据。
- ②汇聚治理：政务部门数据按照七项要求入湖，开展数据治理。入湖要求包括明确数据负责人、确定数据源、描述业务数据、确定数据标准、确定数据分类分级、确定质量规则、有时空属性的数据补充北斗网格码。
- ③融合服务：采用以业务对象、业务（事件）、北斗时空的方式进行数据融合治理，根据需要通过算法模型、智能标签、指标数据的融合方式对外提供数据服务。
- ④数据资产目录。将入湖数据和融合数据统一映射到全省数据资产目录系统，最终形成全省数据资源一张图，并以目录的方式统一对外提供检索和服务。
- ⑤数据使用：通过数据服务调用高质量的数据湖数据，支撑各部门场景创新、分析利用、数据运营三类应用。
- ⑥管理体系：建设全省统一的数据管理机制、组织架构、数据分类、数据架构、标准规范、数据质量、安全保障、评估评价等，形成全省数据管理共识，保障数据工程有效落实。数据架构

数据架构是以业务对象为核心，将业务领域中人、事、物、地作为核心数据要素进行管理，包括L1领域分类、L2主题域、L3业务对象、L4逻辑实体、L5属性。通过数据架构让全省有共同语言，明确数据来源和管理责任单位，解决数据问题争议，实现“一数一源、一源多用”，确保数据在各业务领域间高效、准确地使用。

L1是业务域，是顶层信息分类，基于业务管理边界进行分类，通过数据视角体现最高层面关注的业务领域。

L2是主题域，是互不重叠的数据分类，管辖一组密切相关的业务对象。

L3是业务对象，是数据架构的核心层，用于定义业务中重要的人、事、物、地，数据架构建设和数据治理主要围绕业务对象开展。

L4是逻辑实体，用于描述业务对象的某方面特征的一组或多组属性信息。

L5是属性，用于描述业务对象的某方面特征的最小颗粒度的属性信息。

方案优势

- 定位明确：华为聚焦做大数据工具，以开放、合作、共赢为原则，融入当地生态；此外华为明确定位为解决方案供应商，承诺永不进行数据变现。使能生态伙伴，按照统一的架构与标准，进行预集成验证，通过华为最严苛的安全测试，整合从应用、业务到平台的完整解决方案。
- 合规性高：满足政务大数据平台安全合规的需求；数据“三不一可”；进不来，拿不走，看不懂，可追溯。数据全生命周期安全保障，包括数据采集、传输、存储、共享、使用、销毁；数据安全管控，授权、认证、防护、加密、DLP、脱敏、审计、溯源、分析多维度管控；按照管理权、执行权、审计权的分立模式，实现一套数据全生命周期的统一大数据安全体系服务。
- 高效治理：针对客户多工具、多平台带来页面多入口、账户不统一、管理成本激增等问题，DataArk服务提供符合华为云规范的N统一特性（认证、鉴权、用户、角色、服务目录、工作空间、API网关），为客户的多种数据治理工具提供统一入口，实现数据的统一管控；遵从华为数据之道，更快构建数字化转型体系，提供一站式治理，全流程可视可管，统一入口，统一体验，统一数据源管理，一次接入，全程共享
- 流程贯通：多厂商数据治理工具会带来数据断点，同一份数据很难在多套工具间同步，DataArk服务提供关键指标、数据、模型、数据标准的全流程贯通能力，保障数据规范端到端一致性；数据一次生成落地后，自动在工具间同步，治理效率

大幅提升；数据源、数据湖仓等信息共享，数据打通，减少维护工作，保障正确性；

- ISV伙伴对接预集成：服务已完成5+伙伴治理工具与华为产品的预对接工作，通过自动化部署本服务，30分钟内即可完成与客户现网ISV工具的对接；ISV伙伴快速对接适配：DataArk服务提供开放框架，规范接口，支持客户指定伙伴30天内快速灵活接入（新模块的具体开发对接时长，根据ISV能力会有浮动）；
- 行业沉淀：用户可通过DataArk数智融合集成平台的行业资产沉淀订阅能力，快速订阅华为在政府行业中沉淀的数据标准、模型、脚本、资产和指标，帮助客户快速获取行业内优秀数据资产，并且更专注于自身业务本质的治理；自动分析局点资产，分级脱敏，确保资产分享过程数据安全；订阅资产一键部署到本地工作空间，高效完成数据治理。

2 资源和成本规划

表 2-1 资源和成本规划内容说明

模块	分类	名称	关键参数	数量/台	备注
云平台	服务器节点	网络节点	1. 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 2. 配置256GB内存； 3. 配置2块480G SSD硬盘； 4. 配置6*SP333 10GE光口（满配光模块）+4*GE电口； 5. 冗余风扇，配置2块900W冗余交流电源；	2	/
		计算节点	1. 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 2. 配置512GB内存； 3. 配置2块480G SSD硬盘； 4. 配置4*SP333 10GE光口（满配光模块）+4*GE电口； 5. 冗余风扇，配置2块900W冗余交流电源；	14	通过HCS manag eone发放 DataAr ts Studio 资源

模块	分类	名称	关键参数	数量/台	备注
		裸金属网关节点	1. 2个64位处理器，主频≥2.6GHz，每个处理器≥48核； 2. ≥192G内存，数量≥6（64/48/32核）； 3. ≥2块SSD/SAS HDD，单块要求≥480G/600G； 4. ≥8个10GE/25GE网口，SP380(Mellanox CX-4 Lx)； 5. 支持RAID 0、1、10，支持JBOD，SP460C-M 4G(Avago 3516)； 6. 冗余风扇，配置2块900W冗余交流电源；	2	/
		管理节点	1. 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 2. 配置448GB内存； 3. 配置2块960G SSD硬盘+1块3.2TB NVME SSD硬盘+5块4TB SATA硬盘； 4. 配置4*SP333 10GE光口（满配光模块）+4*GE电口； 5. 配置1*9460-8i(Avago 3508) RAID卡； 6. 冗余风扇，配置2块900W冗余交流电源；	10	/
		对象存储服务器-LVS节点	1. 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 2. 配置192GB内存； 3. 配置2块480G SSD硬盘+1块960G SSD硬盘； 4. ≥6个10GE/25GE网口，≥2块网卡，SP382(Mellanox CX-5)，SP380(CX4)； 5. 支持RAID 0、1、10，支持JBOD，9440-8i(Avago 3408)； 6. 冗余风扇，配置2块900W冗余交流电源；	2	/

模块	分类	名称	关键参数	数量/台	备注
		对象存储服务器-存储节点	<ol style="list-style-type: none"> 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 10*16GB内存； 2块SSD，单块要求≥480GB + 4块3.2TB NVME SSD硬盘+12块16TB SATA硬盘，且扇区格式需为512n或512e； ≥4个10GE/25GE网口，SP382(Mellanox CX-5)，SP380(CX4)； 支持RAID 0、1、10，支持JBOD，9440-8i(Avago 3408)； 冗余风扇，配置2块900W冗余交流电源； 	3	/
	存储节点	生产存储	<ol style="list-style-type: none"> 2U 12盘位 直通机箱； CPU: 2*Kunpeng 920，单颗处理器主频≥2.6GHz，核心≥48核心； 内存: 8*16GB； 配置 2*固态硬盘-480GB-SATA-2.5-6Gb/s+ 5*7.68TB SAS SSD 硬盘； 网口：配置≥1GE*4和4*25GE光口（含2个10G多模光模块和2个25G多模光模块）； 电源：配置2块900W冗余白金交流电源； 提供82T可用存储 	3	/
大数据平台	大数据类型1 (MRS)	大数据管理/控制节点	<ol style="list-style-type: none"> 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 配置8*32GB内存； 配置2*480GB SATA SSD,4*960GB SATA SSD； 配置4*SP333 10GE（含4个多模光模块）； 配置1*9460-8i(Avago 3508) RAID 卡； 配置2*900W AC冗余电源。 	5	/

模块	分类	名称	关键参数	数量/台	备注
	大数据类型2 (MRS)	Spark&Hive节点	1. 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 2. 配置8*32GB内存； 3. 配置2*480GB SATA SSD,12*8TB SATA HDD； 4. 配置4*SP333 10GE（含4个多模光模块）； 5. 配置1*9460-8i(Avago 3508) RAID卡； 6. 配置2*900W AC冗余电源。	36	/
	数据仓库集群 (DWS)	分布式并行数据库节点	1. 配置两颗鲲鹏920处理器，单颗处理器主频≥2.6GHz，核心≥32物理核心； 2. 配置512GB内存； 3. 配置26块960G SSD硬盘； 4. 配置4*SP333 10GE（含2个多模光模块）； 5. 配置1*9460-8i(Avago 3508) RAID卡； 6. 冗余风扇，配置2块900W冗余交流电源；	7	/

 说明

按照安徽销售模式三（软硬件全买断）：硬件约400w，软件+专业服务约3300w

3 数据工程实施流程

本文档介绍如何构建端到端数据处理流程，协助政府进行数据治理，实现政务大数据一体化管理

表 3-1 构建一体化政务大数据平台数据处理端到端流程

序号	类	项目	描述
1	组建专班	首席数据官	首席数据官
		业务专员	业务专员
		数据专员	数据专员
2	数据架构规划设计	数据调研	数据调研
		业务域L1和主题域L2定义	业务域L1和主题域L2定义
		梳理业务流程	梳理业务流程
		业务流程活动	业务流程活动
		业务对象L3梳理和数据owner确定	业务对象L3梳理和数据owner确定
		逻辑实体L4设计和数据来源	逻辑实体L4设计和数据来源
		属性L5设计和数据分级	属性L5设计和数据分级
		数据标准/指标设计	数据标准/指标设计
3	数据工程实施	数据集成	数据集成前探查
			全量数据同步
			增量数据同步
		规范设计	配置中心（审核人、功能、模型管理配置）
			流程设计

序号	类	项目	描述
			主题设计
			码表管理
			数据标准
			关系建模（逻辑模型、物理模型）
			业务指标
			技术指标
			维度建模
		数据开发	连接管理
			脚本开发
			作业开发
			实例管理
			补数据监控
			配置环境变量
		数据治理	数据质量监控
		数据资产	元数据采集
			业务资产
			技术资产
		数据服务	创建专享版集群
			设置API分配配额
			新建审核人
			开发API
			调用API

4 实施步骤

- 4.1 组建专班
- 4.2 数据架构规划设计
- 4.3 数据集成
- 4.4 规范设计
- 4.5 数据开发
- 4.6 数据质量
- 4.7 数据资产
- 4.8 数据服务

4.1 组建专班

首席数据官

为落实数字安徽总体建设方案部署要求，统筹推进数据工程建设各项工作，各部门需要在本单位落实首席数据官制度，成立数据工程推进专班。并完成单位的组织及管理机制建设，构建首席数据官领导的数据治理组织，建设本单位数据工程运作机制。

首席数据官负责部门信息化项目建设规划及管理、数据标准规范体系建设、数据架构统筹管理、数据质量持续提升、数据创新应用规划、数据争议裁决、常态化指导监督等工作，首席数据官一般由单位的分管领导担任。

业务专员

负责本业务领域的数据的业务含义、术语定义、业务流程梳理优化、数据架构管理、数据标准制定，对数据的真实性、准确性、一致性等数据管理工作负责。

业务专员一般由部门各处室具体业务领域的负责同志担任。

数据专员

数据专员是首席数据官的助手，是数据管理工作的具体执行者，负责信息化项目、数据工程的建设维护，定期向首席数据官汇报信息化建设进展。

各部门数据专员一般由信息化处室的负责同志担任。

4.2 数据架构规划设计

概述

数据架构规划设计主要包含数据调研、梳理业务流程、业务流程活动、业务域L1和主题域L2定义、业务对象L3梳理和数据owner确定、逻辑实体L4梳理、数据分类、属性L5梳理和数据标准设计、数据分级设计、数据指标设计九大步骤，通过以上步骤完成数据架构规划设计。

数据架构规划设计是一体化政务大数据建设过程非常重要环节，有利于数据资源的梳理以及后续数据资源的管理。

4.2.1 数据调研

数据架构设计的基础是从业务中明确各类数据信息，保证数据治理结果支撑政府业务的开展及应用。调研开始首先要明确调研路径，以各业务处室为单位进行调研，提前对处室业务进行了解包括业务范围、应用系统功能、核心业务流程说明等信息。

步骤1 编制调研模板

业务调研：基于数据工程的需求，对本部门的业务处室，进行访谈；访谈前准备好相应的访谈提纲，向调研对象阐述访谈背景之后，围绕问题清单开展访谈工作，同时参照资料收集清单，请业务处室提供相关材料。访谈提纲编写要求如表1所示。

表 4-1 访谈提纲模板

访谈提纲模板	
访谈主题	通过业务访谈深入了解XXX业务和数据现状，收集业务人员对数据工程项目的诉求
访谈对象	各业务处室领导、骨干业务人员
访谈时间	约1小时
访谈地点	XXXX大楼XX层XXX室
访谈背景	构建一体化政务大数据，摸清数据家底
问题清单	1. 当前业务对数据要求 2. 数据质量存在哪些问题等
收集资料清单	1. 业务系统涉及的哪些方面数据 2. 与周边哪些部门系统交互

在访谈结束，形成访谈纪要。

系统调研：除业务调研外，还需对相关系统进行调研，编制系统模板，重点调研IT系统的基本信息、数据量、数据库信息、数据字典、主键、索引、大字段、码表、数据库表信息等信息。系统调研模板要求如表2和表3所示。

表 4-2 系统表调研模板

系统名称	表名称	表注释	更新周期	数据是否存在物理删除	主键	新增数据时间戳	更新时间戳	总体记录数量(条)	月均增量(条)
数据表所属业务系统	表名称	中文注释	日周月年	从数据库中彻底删除	主键字段说明	数据新增的时间戳字段	数据同步机制为增量的需填写,提供时间戳字段	总数据记录条数	根据历年数据增量计算平均值
XX业务管理系统	/	/	/	/	/	/	/	/	/

表 4-3 系统表字段调研模板

表名称	表注释	字段名称	字段注释	字段类型	字段长度	是否非空	是否主键
表名称	中文注释	字段名称	字段中文释义	字段类型(根据数据库区分)	字段长度(根据数据库区分)	是否非空	是否主键
sq_info_n	申请记录表	OAE002	性别	NUMBE R(18)	18	是	否

按照项目的整体推进计划，与相关负责人沟通，确定调研计划；双方按照约定时间及计划，有序开展调研工作。

步骤2 业务处室调研

参考调研访谈计划，与各业务处室及各应用系统建设单位进行调研。

参与成员：业务处室负责人、骨干业务人员；

目的：重点了解各业务处室业务流程、业务指标及收集相应表单，作为数据架构及业务指标梳理的参考，了解处室对数据工程的期望，作为规划设计的重要输入和参考；

建议访谈形式：以面对面调研为主；

资料收集清单：需要收集资料如表4所示。

表 4-4 资料收集清单表

序号	项目	资料说明	作用
1	三定方案	了解职能配置、内设机构和人员编制规定	加深业务理解，便于资产信息目录梳理
2	业务流程	请提供相关业务的处理流程图	通过业务流程图，熟悉业务办理流程，了解数据分布
3	业务流程过程涉及的表单	核心业务表单、下级部门上报的报表等	通过表单搜集，便于资产信息目录梳理
4	部门考核指标	来自上级部门或自己部门制定的考核指标及计算公式	便于指标的设计和提炼
5	部门业务开展过程中的分析指标	为满足业务的开展而设计的对业务场景进行分析的相关指标及计算公式	便于指标的设计和提炼
6	数据共享交互需求	请提供长期与其他业务产生交互的相关数据需求	通过需求梳理，便于后期的数据集成方案的设计
7	部门总结报告、总结报表等	提供部门月度、季度、年度等总结报告和报表相关资料	便于业务理解以及业务指标的设计和提炼

步骤3 系统调研总结

输出调研总结报告，包括业务现状，系统现状，数据现状，整理总结信息与处室确认。

----结束

4.2.2 业务域 L1 和主题域 L2 定义

数据架构L1和L2明确了业务对象、逻辑实体、属性等信息的归属，实现各部门数据的归类。

L1为业务域，是描述部门数据管理的高层级分类。目前通常有两种业务域分类方式：基于数据自身特征边界进行分类和基于业务管理边界进行分类。为了强化部门的数据管理责任，更好地推进数据资产梳理、数据治理和数据应用建设，采用业务管理边界划分方式，即将L1业务域与业务职能相匹配。

L2为主题域，是互不重叠的数据分类，管辖一组密切相关的业务对象，通常同一个主题域有相同的数据Owner。

为统一全省数据架构分类，减少编目差异，标准规范《业务域与主题域管理》中已有的L1与L2明细，选取适用于本部门的目录进行后续业务对象L3的关联。

表 4-5 业务域与主题域示例表

业务域	主题域
示例：内部运行管理	事项管理
	运行管理
	好差评

业务域：L1业务域以国办数据目录所属领域目录为基础，结合安徽省各部门业务实际情况，对业务域范围进行划分；

主题域：L2主题域参考各部门所管辖的业务范围与权责清单，对相同类别内容进行聚合，实现主题域的全省统一，方便后续业务对象的分类管理。

4.2.3 梳理业务流程

业务流程图是业务与系统的共同语言,是进行系统分析的依据。业务流程描述了业务走向，业务流程图的绘制应当按照业务的实际处理步骤和过程来进行。绘制流程图的过程是全面了解业务处理的过程，在绘制业务流程图之前，要对现行系统进行调研。流程图梳理主要步骤如下：

- 绘制主体流程：调研业务，绘制主体流，深入业务主体的逻辑关系分析主体脉络，即按照基本要素中的角色、活动、协作关系围绕业务主流程梳理；
- 补充分支流程：从角色入手，政务场景下角色关键且有限，从实际调研业务角色去分析，绘制业务分支流程图，使业务流程具象化，保证业务流程各活动节点不遗漏；
- 流程闭环：通过业务梳理，与业务负责人明确开始与结束活动，实现流程的闭环；
- 流程的分析和优化：优化已梳理的流程，并列明当前流程的当前支撑系统，方便与后续资产梳理自下而上的对应。

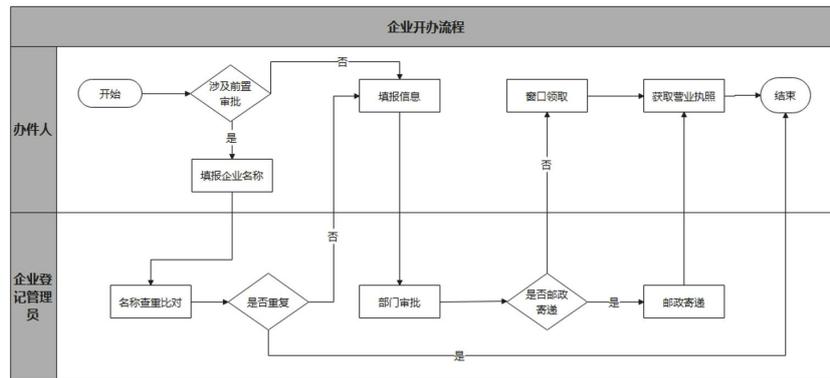
样例一：企业开办流程活动流程图

企业开办基本流程包含以下步骤：

1. 判断是否涉及前置审批：申请人需要判断开办的企业是否需要前置审批，如需要，则填写企业名称等企业开办预登记信息；
2. 名称查重比对：确定企业名称是否已申请，如未申请则继续提交企业开办申请资料；
3. 提交资料：填写企业信息、注册资金、地址信息、高管信息、经营范围等企业登记信息并递交申请材料；
4. 资料审核：企业登记部门对资料进行审核，判断是否能够发放企业营业执照；
5. 领取执照：通过准予设立登记通知书、办理人身份证到窗口领取营业执照或邮寄营业执照正、副本。

具体如图2企业开办流程图所示

图 4-1 企业开办流程活动流程图



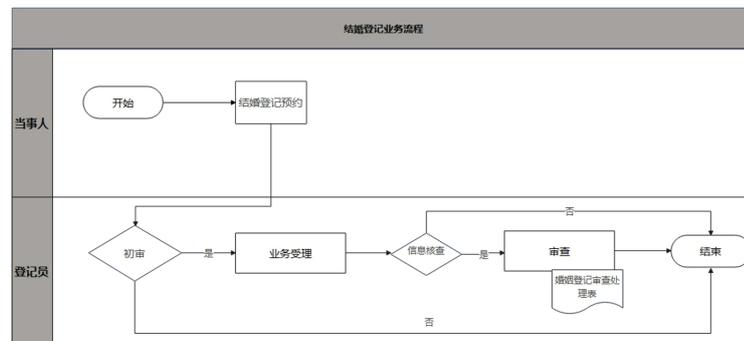
样例二：国内结婚登记流程活动流程图

结婚登记业务办理的基本流程包含以下步骤：

1. 双方当事人选择是否预约了结婚登记业务办理；
2. 国内结婚登记员审查当事人是否满足结婚登记申请要件，满足条件方可受理；
3. 国内结婚登记员受理当事人提交的结婚登记申请；
4. 国内结婚登记员对当事人提交的证件、证明、声明及国内结婚登记历史记录进行核查、审查；
5. 审查通过，颁发结婚证。

具体如图3国内结婚登记业务流程所示。

图 4-2 国内结婚登记流程活动流程图



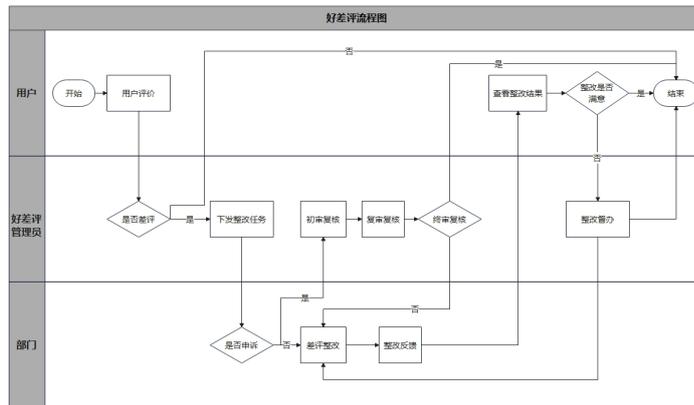
样例三：好差评流程活动流程图

好差评基本流程包含以下步骤：

1. 办件人在办件完成后对办件结果进行评价；
2. 根据用户评价类型（好评/差评），对差评涉及办件事项和办件人发送整改任务；
3. 部门业务办理人员收到整改任务单后判断是否需要申诉；
4. 政务服务平台管理人员对业务办理人员的申诉请求进行审核；
5. 政务服务平台管理人员对业务办理人员的申诉请求进行复核。

具体如图4好差评流程图所示。

图 4-3 好差评流程活动流程图



4.2.4 业务流程活动

为保障业务对象梳理的完整性，基于业务流程图梳理相应的业务步骤，通过对业务步骤的分析，为业务对象梳理提供素材。首先需要填写业务流程中的步骤名称、流程步骤描述、角色、输入、输出等内容。具体梳理模板与填写方法如表6所示。

表 4-6 业务流程说明表

编号	步骤名称	流程步骤描述	角色	输入	输出
示例：	部门审批	根据用户提供的企业登记信息，对企业登记进行审核	市场监督管理局审批人	企业登记信息，企业人员信息	企业营业执照

业务流程梳理要明确流程步骤名称，并针对各业务步骤进行描述说明，结合业务流程中相关参与角色，明确角色信息以及流程中各步骤中涉及的输入、输出等相关信息。

1. 步骤名称：通过对业务流程梳理、分析，基于业务流程的分析结果及流程图信息，明确业务流程中的各步骤的名称；
2. 流程步骤描述：结合业务现状，针对流程步骤描述各环节的具体工作，重点在于表达出此流程步骤工作的具体内容；
3. 角色：针对业务需求，对此步骤的岗位角色进行明确，对已经明确的岗位说明进行描述，此项工作重点在于将此环节的相关岗位角色表达清晰；
4. 输入：说明此流程步骤中的输入信息，能够清晰的表达出业务流程与数据信息的对应关系。针对识别到的输入数据信息的描述，重点写清楚此业务流程需要哪些表单的相关输入，基于此步骤的角色人处理哪些信息；
5. 输出：说明此流程步骤中的输出信息，能够清晰的表达出业务流程中此步骤办理完毕后可输出哪些数据信息。针对输出数据信息的描述，重点写清楚此业务流程在办理完结后需要输出哪些表单信息。

样例一：企业开办业务流程活动

通过对企业开办业务进行梳理及分析，明确企业开办业务流程中共存在8个业务活动节点，且各活动节点中均存在活动的输入、输出。业务流程活动通过对活动节点中输入、输出数据进行分析，识别相应的业务对象内容。

企业开办流程具体包含涉及前置审批、填报企业名称、名称查重比对、是否重复、填报信息、部门审批、是否邮政寄递、邮政寄递、窗口领取、获取营业执照共计8个活动节点。输入、输出信息如下表所示。

表 4-7 企业开办业务流程说明表

序号	步骤名称	流程步骤描述	角色	输入	输出
1	涉及前置审批	申请人是否需要企业进行名称去重预处理	办件人	企业预登记信息	企业预登记信息
2	填报企业名称	申请人填报申请的企业名称	办件人	企业预登记信息	企业预登记信息
3	名称查重比对	系统根据输入的企业名称，在系统中查询重复的名称	系统	企业预登记信息	企业预登记信息
4	是否重复	用户根据系统反馈的结果判断是否能够使用该名称进行申请	办件人	企业预登记信息	企业预登记信息
5	填报信息	用户在页面填报企业相关信息	办件人	企业预登记信息	企业登记信息
6	部门审批	部门对提交的信息进行评审	审核人员	企业登记信息	企业营业执照
7	是否邮政寄递	判断是否进行邮政寄递	办件人	是否邮政寄递信息	邮政寄递信息
8	获取营业执照	用户获取企业营业执照	办件人	企业营业执照	企业营业执照

样例二：国内结婚登记业务流程活动

通过对婚姻管理的结婚登记（国内）业务流程进行梳理及分析，明确业务流程中共存在5个业务活动节点，且各活动节点中均存在活动的输入、输出。业务流程活动通过对活动节点中输入、输出数据进行分析，识别相应的业务对象内容。

国内结婚登记业务流程具体包含结婚预约登记、初审、业务受理、信息核查、审查共计5个活动节点。输入、输出信息，具体如下表所示。

表 4-8 国内结婚登记业务流程说明表

序号	步骤名称	流程步骤描述	角色	输入	输出
1	结婚登记预约	结婚当事人提交结婚登记预约申请	结婚当事人	双方当事人基础信息	预约申请信息
2	初审	国内结婚登记员审查当事人是否满足结婚登记申请要件	登记机构工作人员	预约申请信息、双方当事人基础信息	当事人基础信息
3	受理	国内结婚登记员受理结婚登记申请，核验当事人的身份	登记机构工作人员	初审后的当事人基础信息	受理的结婚登记申请信息
4	信息核查	查询当事人的国内结婚登记记录信息、法院离婚信息、外交部国内结婚登记记录信息以及死亡信息等，判定是否符合办理条件	登记机构工作人员	受理的结婚登记申请信息、当事人基础信息	核查后的当事人基础信息（包含离婚信息，法院离婚信息，外交部国内结婚登记信息，死亡信息）
5	审查	符合办理条件的当事人确认《结婚登记审查处理表》和结婚证相关信息	登记机构工作人员	核查当事人基础信息	结婚证，结婚登记审查处理表

样例三：好差评业务流程活动

通过对好差评业务进行梳理及分析，明确好差评业务流程中共存在11个业务活动节点，且各活动节点中均存在活动的输入、输出。业务流程活动通过对活动节点中输入、输出数据进行分析，识别相应的业务对象内容。

好差评业务流程具体包含用户评价、下发整改任务、是否申诉、初审复核、复审复核、终审复核、差评整改、整改反馈、查看整改结果、用户追评、整改督办共计11个活动节点。输入、输出信息如下表所示。

表 4-9 好差评业务流程说明表

序号	步骤名称	流程步骤描述	角色	输入	输出
1	用户评价	办件人在办件完成后对办件结果进行评价	用户	/	评价内容信息
2	下发整改任务	根据用户评价类型（好评/差评），好差评事项涉及办件事项和办件人发送整改任务	系统	办件信息、评价信息	整改任务单

序号	步骤名称	流程步骤描述	角色	输入	输出
3	是否申诉	部门业务办理人员收到整改任务单后判断是否需要申诉	部门业务办理人员	整改任务单	判断结果
4	初审复核	政务服务平台管理人员对业务办理人员的申诉请求进行审核	政务服务平台管理人员	申诉请求	初审意见
5	复审复核	政务服务平台管理人员对业务办理人员的申诉请求进行复核	政务服务平台管理人员	申诉请求	复审意见
6	终审复核	政务服务平台管理人员对业务办理人员的申诉请求进行终审。终审通过，则停止整改，流程结束；终审不通过，则业务办理人员进行整改。	政务服务平台终审人员	申诉请求	终审意见
7	差评整改	接收整改任务单后根据任务书中的内容进行整改。	业务办理人员	整改任务单	整改结果
8	整改反馈	提交整改结果的反馈，并由系统推送给办件人	部门业务办理人员	整改结果	整改结果
9	查看结果	对整改结果满意，则流程结束，否则把反馈意见推送给政务服务平台。	办件人	整改结果	反馈意见
10	用户追评	评价整改是否满意	办件人	整改结果	反馈意见
11	整改督办	对办件人提交的不满意反馈意见进行督办，发送给部门业务人员进行整改。	政务服务平台管理人员	反馈意见	督办通知

4.2.5 业务对象 L3 梳理和数据 owner 确定

- 数据架构业务对象识别方法

基于业务流程梳理，以及对相关业务的了解，梳理出业务流程开展过程中的主要信息项即输入输出信息，并结合业务对象的识别原则，识别出此流程开展过程中的相应业务对象。

业务对象识别要遵循以下四项原则：

- a. 运行和管理中重要的人、事、物、地：如果缺少了某个业务对象，业务运作和管理将会出现严重问题，甚至业务都将无法存在；

- b. 具有唯一身份标识信息：可以通过唯一标识区分、准确检索和支持跨领域分布式共享业务对象；
- c. 相对独立：业务对象之间是关联关系，不是所属关系，拥有自己的唯一编码；
- d. 可实例化：业务对象有相应的属性，可以被记录到数据库表中，每出现一个这样的人、物，或发生一次这样的事，就会在数据库表中有相应的记录。

具体业务对象的梳理模板与填写方法如表10所示。

表 4-10 业务对象识别表

序号	步骤名称	主要信息项	业务对象
示例	用户评价	办件人信息、办件事项信	办事评价

步骤名称：结合业务流程的各流程步骤，填写具体的活动节点；

主要信息项：针对此活动节点的主要输入、输出信息，明确此活动节点的主要信息项。此环节填写过程中重点涉及到此活动节点的表单信息填写清晰，保证填写完整性；

业务对象：用于定义业务领域重要的人、事、物、地。数据架构建设和治理主要围绕业务对象开展。

- **数据owner确定方法**

结合业务现状以及各部门的数据管理现状，针对识别出的相关业务对象，明确各业务对象的管理责任方，即数据owner。数据owner明确了数据的归属责任方，为数据管理提高了数据管理能力，增强数据的准确性及应用性。

数据owner针对业务对象进行设计，数据owner确定过程中需要明确业务域、主题域、业务对象等信息，保障数据的全面性。填写方法如下表所示。

表 4-11

业务域	主题域	业务对象	业务对象数据owner部门	业务对象数据owner人员	逻辑实体	数据源	数据分类
示例：内部运行管理	好差评	办事评价	电子政务与应用处	XXX	/	/	/

业务域：业务对象所属业务域；

主题域：业务对象所属主题域；

业务对象：将数据架构中业务对象的具体内容，填写到模板中的业务对象栏中；

数据owner部门：基于业务对象明确各业务对象的管理责任部门；

数据owner人员：基于业务对象明确各业务对象的管理责任人员；

逻辑实体：结合上述规范设计的逻辑实体；

数据分类：结合业务现状以及数据需求现状，结合单位数据分类内容，明确各逻辑实体的分类；

数据来源：结合数据产生来源，录入产生系统名称。

样例一：企业开办业务对象与数据owner

企业开办业务对象案例

通过对企业开办业务系统中输入、输出信息进行分析梳理，识别主要信息项内容，结合业务对象识别原则，梳理各活动节点的业务对象，包括企业预登记信息、企业登记信息、企业人员信息三个业务对象。

表 4-12 企业开办业务对象识别表

序号	步骤名称	主要信息项	业务对象
1	涉及前置审批	/	/
2	填报企业名称	主体身份代码、企业名称、名称区划、企业字号、名称行业、组织形式。	企业预登记信息
3	名称查重比对	/	/
4	是否重复	/	/
5	填报信息	主体身份代码、企业名称、住所所在行政区划、住所、邮政编码、联系电话、姓名、国别、职务、职务产生方式、证件类型、证件号码、移动电话	企业登记信息 企业人员信息
6	部门审批	企业名称、统一社会信用代码、注册资本、类型、成立日期、法定代表人、营业期限、经营范围、住所	企业营业执照
7	是否邮政寄递	/	/
8	邮政寄递	企业名称、统一社会信用代码、注册资本、类型、成立日期、法定代表人、营业期限、经营范围、住所	企业营业执照
9	窗口领取	企业名称、统一社会信用代码、注册资本、类型、成立日期、法定代表人、营业期限、经营范围、住所	企业营业执照
10	获取企业营业执照	/	/

企业开办数据owner案例

通过对业务对象的梳理，并结合市场监督管理局组织现状，明确企业预登记信息、企业登记信息、企业人员信息、企业营业执照四个业务对象的数据owner为省市场监督管理局登记注册局。

表 4-13 企业开办 L1-L4 确认表

业务域	主题域	业务对象	业务对象数据owner部门	业务对象数据owner人员	逻辑实体	数据源	数据分类
市场监管	市场主体准入与退出	企业预登记信息	登记注册局	XX	/	/	/
市场监管	市场主体准入与退出	企业登记信息	登记注册局	XX	/	/	/
市场监管	市场主体准入与退出	企业人员信息	登记注册局	XX	/	/	/

样例二：国内结婚登记业务对象与数据owner

结婚（国内）登记业务对象案例

通过对结婚登记业务系统中输入、输出信息进行分析梳理，识别主要信息项内容，结合业务对象识别原则，梳理各活动节点的业务对象，包括婚姻预约信息、结婚登记机构、结婚登记机构工作人员、当事人、国内结婚登记信息、结婚证信息业务对象。

表 4-14 结婚登记业务对象识别表

序号	步骤名称	主要信息项	业务对象
1	结婚登记预约	双方当事人基础信息，预约申请信息	结婚当事人，婚姻预约信息，登记机构，结婚登记机构工作人员
2	初审	预约申请信息、双方当事人基础信息，初审后的当事人基础信息	结婚当事人 婚姻预约信息
3	受理	初审后的当事人基础信息，受理的国内结婚登记申请信息	结婚当事人 结婚登记信息

序号	步骤名称	主要信息项	业务对象
4	信息核查	受理的国内结婚登记申请信息、当事人基础信息，核查后的当事人基础信息	结婚当事人 结婚登记信息
5	审查	核查当事人基础信息，发证机关，结婚证，结婚登记审查处理表	结婚当事人 结婚登记信息 结婚证信息

国内结婚登记数据owner案例

通过对业务对象的梳理，并结合省民政厅组织现状，明确业务对象数据owner为社会事务处。

表 4-15 国内结婚登记数 L1-L4 确认表

业务域	主题域	业务对象	业务对象数据owner部门	业务对象数据owner人员	逻辑实体	数据源	数据分类
生活服务	婚姻登记	婚姻预约信息	社会事务处	XX	/	/	/
生活服务	婚姻登记	结婚登记机构工作人员	社会事务处	XX	/	/	/
生活服务	婚姻登记	国内结婚登记信息	社会事务处	XX	/	/	/

样例三：好差评业务对象与数据owner

好差评业务对象案例

通过对好差评业务系统中输入、输出信息进行分析梳理，识别主要信息项内容，结合业务对象识别原则，梳理各活动节点的业务对象，包括办事评价、整改任务、评价申诉三个业务对象。

表 4-16 好差评业务对象识别表

序号	步骤名称	主要信息项	业务对象
1	用户评价	办件人信息、办理结果信息（办理结果、办理开始时间、结束时间等）、评价内容信息（评价分类：好评/差评；评价内容；建议、诉求等）	办事评价
2	下发整改任务	整改单编号、时间、业务部门、业务人员、办件事项信息、办理结果信息、评价内容信息、整改要求、要求完成时间等。	整改任务
3	是否申诉	申诉单编号、整改单编号、办件事项编号、申诉原因、申诉内容等	评价申诉
4	整改反馈	整改单编号、办件事项编号、整改结果、完成时间等。	整改任务
5	查看整改结果	整改结果信息、办件人反馈信息（反馈时间、反馈内容、建议、诉求等）。	办事评价
6	用户追评	办件人信息、办理结果信息（办理结果、办理开始时间、结束时间等）、评价内容信息（评价分类：好评/差评；评价内容；建议、诉求等）	办事评价
7	整改督办	整改单编号、整改反馈信息、要求完成时间、优先级等。	整改任务

好差评数据owner案例

通过对业务对象的梳理，并结合省数据资源局组织现状，明确办事评价、整改任务、评价申诉业务对象数据owner为电子政务与应用处。

表 4-17 好差评 L1-L4 确认表

业务域	主题域	业务对象	业务对象数据owner部门	业务对象数据owner人员	逻辑实体	数据源	数据分类
内部运行管理	好差评	办事评价	电子政务与应用处	XX	/	/	/
内部运行管理	好差评	评价申诉	电子政务与应用处	XX	/	/	/
内部运行管理	好差评	整改任务	电子政务与应用处	/	/	/	/

4.2.6 逻辑实体 L4 设计和数据来源

逻辑实体是数据架构的L4级，是描述业务对象某种业务特征的属性集合，具有一定逻辑关系。

- 逻辑实体设计

逻辑实体是具有一定逻辑关系的属性集合，逻辑实体设计需遵循强制性设计规范与建议性设计规范，结合设计方法，明确单位各业务域的逻辑实体。

强制性规范：

逻辑实体不能脱离业务对象独立存在，不能归属于多个业务对象，业务对象与逻辑实体的关系是1:1或1:N，不允许N:1的情况出现；

提供数据服务或跨业务领域使用的基础数据，要单独设计逻辑实体；

两个业务对象间的关系逻辑实体，归属于业务发生时间先后顺序中后出现的业务对象；

逻辑实体设计时不考虑水平(横向)拆分，水平拆分在物理表设计中考虑。

建议性规范：

描述业务对象不同业务特征的属性集合，可以单独设计为逻辑实体；

逻辑实体中，在某特定场景下使用的属性，可以垂直(纵向)拆分成另一个逻辑实体；

- 数据源设计

数据源是指业务上首次正式发布某项数据的应用系统，经过认证，作为唯一数据源头被周边系统调用。

数据源管理原则：

部门所有系统需要经过业务专员与数据专员的认证，所有关键数据必须认证数据源；

关键数据仅能在数据源录入、修改，全流程共享，其它调用系统不能修改，下游环节发现的数据源质量问题，应当在数据源进行修正；

所有系统必须从数据源获取关键数据；

数据Owner确保数据源的数据质量，对不符合数据质量标准的数据源，必须限期整改。

数据源认证参照《数据源管理规范》，特殊业务数据源需要对属性数据源进行确认。

- 数据分类设计方法

数据分类是针对不同特性的数据采取不同的管理策略，以实现最大的治理成效。不同分类的数据，采取不同的治理方法，投入不同的资源。具体数据分类包括主数据、基础数据、事务数据、观测数据、时空数据、规则数据、统计数据等。

主数据：部门核心业务对象数据。实例：人口、法人、车辆、房屋、事项。

基础数据：可枚举的、基本确定不变数据。实例：行政区划、组织划分、经济分类。

事务数据：业务办理过程数据。实例：执法监管、行政审批。

观测数据：传感器信息、物联网数据。实例：视频监控、气象观测、水文监测、环境监测。

时空数据：北斗码数据。实例：轨迹数据、位置数据、空间数据。

规则数据：描述业务规则的数据。实例：审批规则、执法规则数据。

统计数据：依据某种规则进行计算、统计。实例：GDP指标、财政收入指标等。

结合上述所梳理的业务域、主题域、业务对象，结合业务流程的输入、输出，参考应用系统数据库表，明确各逻辑实体的数据分类。

- 逻辑实体ER图设计方法

逻辑实体ER图体现逻辑实体之间的关系，从概念模型转换而来，将实体与关系映射到关系数据模型，进行细化，包括了逻辑实体之间的关系、属性、定义、描述和范例。逻辑实体ER图将作为应用系统建设的规范，作为应用系统物理数据模型的输入。

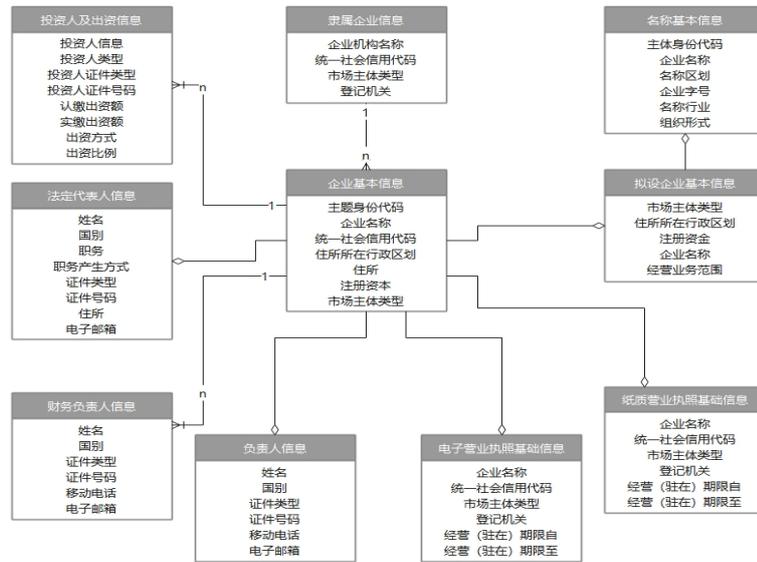
样例一：企业开办业务逻辑实体

基于企业开办业务梳理的相关业务对象，集合企业开办业务流程之间的数据关系，设计企业开办业务的逻辑数据实体为名称基本信息、拟设企业基本信息、企业基本信息、隶属企业信息、法定代表人信息、负责人信息、投资人及出资信息、财务负责人信息、纸质营业执照基础信息、电子营业执照基础信息共计10个逻辑实体。

表 4-18 企业开办 L1-L4 确认表

业务域	主题域	业务对象	业务对象数据owner部门	业务对象数据owner人员	逻辑实体	数据源	数据分类
市场监管	市场主体准入与退出	企业预登记信息	登记注册局	XX	名称基本信息	企业登记网上注册申报服务系统	事务数据
市场监管	市场主体准入与退出	企业登记信息	登记注册局	XX	拟设企业基本信息	企业登记网上注册申报服务系统	事务数据
市场监管	市场主体准入与退出	企业人员信息	登记注册局	XX	企业基本信息	综合业务系统	主数据

图 4-4 企业开办逻辑实体 ER 图



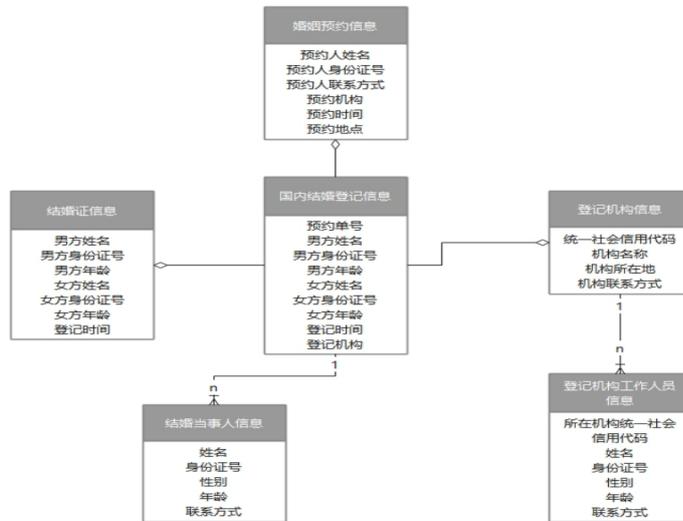
样例二：国内结婚登记逻辑实体

基于结婚登记业务梳理的相关业务对象，集合结婚登记业务流程之间的数据关系整理逻辑实体。内容如表所示。

表 4-19 结婚登记 L1-L4 确认表

业务域	主题域	业务对象	业务对象数据 owner 部门	业务对象数据 owner 人员	逻辑实体	数据源	数据分类
生活服务	婚姻登记	婚姻预约信息	社会事务处	XX	婚姻预约信息表	婚姻在线登记系统	事务数据
生活服务	婚姻登记	结婚当事人	社会事务处	XX	结婚当事人信息表	婚姻在线登记系统	主数据
生活服务	婚姻登记	登记机构	社会事务处	XX	登记机构信息	婚姻在线登记系统	事务数据
生活服务	婚姻登记	结婚登记机构工作人员	社会事务处	XX	结婚登记机构工作人员信息	婚姻在线登记系统	主数据

图 4-5 国内结婚登记逻辑实体 ER 图



样例三：好差评业务逻辑实体

基于好差评业务梳理的相关业务对象，集合好差评业务流程之间的数据关系，设计好差评业务包含人员与办事评价关联信息、评价打分项信息、评价基本信息、评价详细信息、评价模板信息、评价来源-热线信息、差评处理信息、参与办件的政务人员信息、评价来源-办件信息、评价来源-调研问卷信息、评价模板与评价选项对应关系等逻辑实体。

表 4-20 好差评 L1-L4 确认表

业务域	主题域	业务对象	业务对象数据 owner 部门	业务对象数据 owner 人员	逻辑实体	数据源	数据分类
内部运行管理	好差评业务	办事评价	电子政务与应用处	XX	人员与办事评价关联信息	好差评系统	事务数据
内部运行管理	好差评业务	办事评价	电子政务与应用处	XX	评价打分项信息	好差评系统	事务数据
内部运行管理	好差评业务	办事评价	电子政务与应用处	XX	评价基本信息	好差评系统	事务数据

图 4-6 好差评逻辑实体 ER 图



4.2.7 属性 L5 设计和数据分级

- 属性设计方法

在梳理逻辑实体的过程中充分理解、分析业务规格。在设计属性过程中，业务流程和业务规则说明文档是很有参考价值的材料之一，是业务建模必须参考的主要依据。属性设计要考虑当前业务实际规划和未来业务变化而且应尽量详细，预留充分的时间对设计的属性进行检验(交叉检查，业务检查)。属性名称尽量要具有明确含义，便于理解并避免歧义。

- 数据分级方法

根据《数据分级管理规范》中的数据敏感级别定义，将数据敏感级别划分为四级，由低至高分别为：L1级、L2级、L3级、L4级；数据分级由各业务处室自主完成，并应细化到数据字段级，法律、法规另有规定的除外；数据资源定级应根据业务定级和字段定级的最高级来定级。

数据分级是依据对不同对象的影响范围和影响程度来定义级别，具体定义如下表所示。

表 4-21 数据分级定义

影响程度	影响范围	影响对象			
		组织合法权益	个人合法权益	公共利益	国家安全
无影响	/	L1	L1	N/A	N/A
轻微	较小范围	L2	L2	N/A	N/A
	较大范围	L2	L3	L3	N/A
	超大范围	L3	L3	L3	L4
一般	较小范围	L3	L3	N/A	N/A

	较大范围	L3	L4	L4	N/A
	超大范围	L4	L4	L4	L4
严重	较小范围	L4	L4	N/A	N/A
	较大范围	L4	L4	L4	N/A
	超大范围	L4	L4	L4	L4

如：数据泄露导致个人合法权益受到较大范围的一般影响，则该数据应定义为L4级数据。

- 数据分级流程

各部门业务专员根据数据分级要求，利用数据分级方法确定数据资源安全等级，并进行相应标识；部门数据专员将审批通过的数据分级清单挂接到一体化数据基础平台相应的数据资产目录下。

属性设计及数据分级模板如下表所示。

表 4-22 业务属性表

业务域	主题域	业务对象	逻辑实体	业务属性	安全等级
-	-	-	-	-	-

业务域：业务对象所属业务域；

主题域：业务对象所属主题域；

业务对象：是业务领域重要的人、事、物、地，承载了业务运作和管理涉及的重要信息；

逻辑实体：是具有一定逻辑关系的逻辑数据实体合；

业务属性：是描述所属业务对象的性质和特征，反映信息管理最小粒度；

安全等级：描述数据安全等级，参照数据分级。

样例一：企业开办业务属性

对企业开办业务进行分析，结合企业开办业务需求，明确各逻辑数据实体的相关属性，支撑后续数据管理需求，实现数据安全等级清晰明确。

表 4-23 企业开办业务属性示例

业务域	主题域	业务对象	逻辑实体	业务属性	安全等级
市场监管	市场主体准入与退出	企业登记信息	企业基本信息	主体身份代码	L2级数据

业务域	主题域	业务对象	逻辑实体	业务属性	安全等级
市场监管	市场主体准入与退出	企业登记信息	企业基本信息	企业(机构)名称	L2级数据
市场监管	市场主体准入与退出	企业登记信息	企业基本信息	统一社会信用代码	L2级数据

样例二：国内结婚登记业务属性

对婚姻登记业务进行分析，结合婚姻登记业务需求，明确各逻辑数据实体的相关属性，支撑后续数据管理需求，实现数据安全等级清晰明确。

表 4-24 国内结婚登记业务属性示例

业务域	主题域	业务对象	逻辑实体	业务属性	安全等级
生活服务	婚姻登记	国内结婚登记机构工作人员	国内结婚登记机构工作人员信息	机关ID	L2级数据
生活服务	婚姻登记	国内结婚登记机构工作人员	国内结婚登记机构工作人员信息	姓名	L2级数据
生活服务	婚姻登记	国内结婚登记机构工作人员	国内结婚登记机构工作人员信息	人员面貌	L2级数据

样例三：好差评业务属性

对政务服务好差评业务进行分析，结合省数据资源局好差评业务需求，明确各逻辑数据实体的相关属性，支撑后续数据管理需求，实现数据安全等级清晰明确。

表 4-25 好差评业务属性示例

业务域	主题域	业务对象	逻辑实体	业务属性	安全等级
内部运行管理	好差评管理	办事评价	人员与办事评价关联信息	记录编码	L2级数据
内部运行管理	好差评管理	办事评价	人员与办事评价关联信息	姓名	L2级数据
内部运行管理	好差评管理	办事评价	人员与办事评价关联信息	评价结果编码	L2级数据

4.2.8 数据标准设计

数据标准设计方法

针对属性字段设计数据标准，数据标准用于描述业务层面需共同遵守的属性层数据含义和业务规则。其描述了对某个数据的共同理解，这些理解一旦确定下来，就应作为标准被共同遵守。数据标准设计遵从以下原则：

以业务为导向，必须满足各标准各方的迫切需求，并根据业务需求分阶段推进制定工作；

明确无歧义，标准的定义必须反映实际业务情况，在部门范围内有一致的理解；

权威性，标准的定义必须充分参考各类成熟的外部标准（国家标准、行业标准、地方标准），将遵循的外部标准出处、内容进行明确的标注和解释；

可落地性，数据标准在定义时必须考虑技术层面的实现难度，在各相关系统可以落地。数据标准设计模板如下表所示。

表 4-26 参考规范示例

*业务对象	*逻辑实体	*业务属性	*业务定义及用途	*业务规则	*数据类型	数据长度	是否非空	允许值范围	参照的标准	值域标准
示例：办事评价	人员与办事评价关联信息	记录编码	记录唯一标识	IT主键	字符型	32	Y	/	/	/

业务对象：是业务领域重要的人、事、物、地，承载了业务运作和管理涉及的重要信息；

逻辑实体：是具有一定逻辑关系的逻辑数据实体合；

业务属性：是描述所属业务对象的性质和特征，反映信息管理最小粒度；

业务定义及用途：属性的具体业务含义，表达业务的内容，以及属性对业务的反馈的业务用途；

业务规则：对属性进行规范的内容，如对办事人员的姓名要求必须与实际的办事人员一一对应；

数据类型：数据的保存类型，如字符型、日期型等；

数据长度：是指这个对象在内存中要占的空间，也决定了一个对象能存储数据的大小；

是否非空：判断该属性是否为必填属性，是否允许不填内容导致的空值，对于重要的属性多为必填属性，如身份证号等信息；

是否有允许值列表：如是，则列出可选的允许值，例如性别属性中，除了性别的信息外，不允许出现其他信息；

允许值范围：定义如果存在允许值，对允许值得范围做出规定，如性别属性中，允许值为男和女或使用代码形式代表男和女；

参照的标准：该标准参考的行业外部或者内部公开发布的标准规范，如GB/T 39046-2020 政务服务平台基础数据规范；

值域标准：与码表编码对应。

样例一：企业开办参考标准

表 4-27 企业开办参考标准示例

* 业务对象	* 逻辑实体	* 业务属性	*业务定义及用途	*业务规则	*数据类型	数据长度	非空	*是否有允许值列表	允许值范围	参照的标准	值域标准
企业登记信息	企业基本信息	企业名称	登记企业的名称	登记企业的名称	字符型	32	是	无	/	/	/
企业登记信息	企业基本信息	统一社会信用代码	登记企业的唯一标识	登记企业的唯一标准	字符型	18	是	无	/	GB 32100-2015《法人和其他组织统一社会信用代码编码规则》	/

样例二：国内结婚登记参考标准

表 4-28 国内结婚登记参考标准示例

* 业务对象	* 逻辑实体	* 业务属性	*业务定义及用途	*业务规则	*数据类型	数据长度	非空	*是否有允许值列表	允许值范围	参照的标准	值域标准
结婚预约登记	婚姻预约信息	男方姓名	在户籍管理部门正式登记注册的姓氏名称	男方当事人姓名	字符型	50	是	/	否	否	/

* 业务对象	* 逻辑实体	* 业务属性	*业务定义及用途	*业务规则	*数据类型	数据长度	非空	*是否有允许值列表	允许值范围	参照的标准	值域标准
结婚预约登记	婚姻预约信息	男方证件类型	颁发的可以证明个人身份的证件名称	男方证件类型	数值型	2	是	0、1、2、4、5	否	否	JMGC/XX 204 5.1.1 5 C001 5 身份证件类型代码

样例三：好差评参考标准

表 4-29 好差评参考标准示例

* 业务对象	* 逻辑实体	*业务属性	*业务定义及用途	*业务规则	*数据类型	数据长度	非空	*是否有允许值列表	允许值范围	参照的标准	值域标准
办事评价	关联信息	记录编码	记录唯一标识	/	字符型	32	是	无	/	/	/
办事评价	关联信息	UAAC 登录名	办事人员的 UAAC 账号	登录系统时的 UAAC 账号	字符型	32	是	无	/	GB/T 39046-2020 政务服务平台基础数据规范	/

4.2.9 数据指标设计

数据指标设计方法

指标是衡量目标总体特征的统计数值，是表征某一业务活动中业务状况的数值指示器。指标由指标名称和指标数值两部分组成，指标名称及其涵义体现了指标质的规定

性和量的规定性两个方面的特点，指标数值反映了指标在具体时间、地点、条件下的数量表现。业务指标用于指导技术指标的设计，技术指标是对业务指标的具体实现。

通过与业务处室（通常为指标解释部门的业务人员）、系统厂商沟通，从业务角度梳理当前业务流程和应用系统，确定业务关注点，梳理并确定能够反应业务状况变化的数据指标，包括指标编码、业务域、主题域、指标名称、指标定义、指标Owner等维度内容。业务指标样例模板如表30所示。

表 4-30 业务指标模板

序号	指标编码	*业务域	*主题域	指标名称	指标定义	指标Owner
-	-	-	-	-	-	-

指标编码：对数据指标进行归类编码，ZB +六位行政区划+中央业务指导部门编码（2位）+流水号（4位）；示例：340000为安徽省，11为民政部，安徽省民政厅指标为ZB340000110001，数据资源局使用编码89；

业务域：指标所属的业务域；

主题域：主表所属的主题域；

指标名称：指标的名称；

指标定义：指标的业务定义，需要明确计算方式；

指标Owner：是数据指标的责任主体，对指标的质量、定义、计算逻辑等具有解释权和决定权，通常为个人或者部门。

样例一：企业开办数据指标

表 4-31 企业开办数据指标

序号	指标编码	*业务域	*主题域	指标名称	指标定义	指标Owner
1	ZB340000 310001	市场监管	市场主体 准入与退出	企业名称 总数	目前名称库中名称 的总数	登记注册局
2	ZB340000 310002	市场监管	市场主体 准入与退出	预登记 企业成 功率	预登记企业审批成 功率，为企业登记 审批成功数/预登记 企业总数	登记注册局
3	ZB340000 310003	市场监管	市场主体 准入与退出	企业总 数	目前完成企业登记 的总数	登记注册局

样例二：婚姻管理数据指标

表 4-32 婚姻登记数据指标

序号	指标编码	*业务域	*主题域	指标名称	指标定义	指标Owner
1	ZB340000110001	生活服务	婚姻登记	年度结婚人数总数	本年度结婚的人数总数	社会事务处
2	ZB340000110002	生活服务	婚姻登记	结婚人数月同比增长数	结婚人数对比上月同比增长	社会事务处
3	ZB340000110003	生活服务	婚姻登记	结婚人数月同比增长率	本月结婚人数/上月结婚人数	社会事务处

样例三：好差评数据指标

表 4-33 好差评数据指标

序号	指标编码	*业务域	*主题域	指标名称	指标定义	指标Owner
1	ZB340000990001	内部运行管理	好差评业务	评价人数	是指对服务进行评价的人数量。	电子政务与应用处
2	ZB340000990002	内部运行管理	好差评业务	评价次数	是指服务客户对服务进行评价的总次数。	电子政务与应用处
3	ZB340000990003	内部运行管理	好差评业务	差评次数	是指服务客户评价为差评评价的总次数且经核实为有效差评。	电子政务与应用处

4.3 数据集成

4.3.1 数据集成概述、工具介绍及入湖范围

概述

数据集成是指将政府部门IT系统中的业务源数据按需通过全量或者增量的方式接入数管平台中，并确保数据的一致性、有效性及适用性。源数据种类可包括结构化数据、非结构化数据。在本次数据治理实施项目中，使用到的数据集成工具为CDM云数据迁移工具。

该过程需要注意的是保障政务数据应收尽收，并且数据已具备相关标准规范，避免出现漏数、脏数等问题。

工具介绍

云数据迁移（Cloud Data Migration，简称CDM），是一个高效、易用的批量数据迁移服务。CDM围绕大数据迁移上云和智能数据湖解决方案，提供了简单易用的迁移能力和多种数据源到数据湖的集成能力，降低了客户数据源迁移和集成的复杂性，有效地提高数据迁移和集成的效率。

关于数据集成工具的具体使用方法可以参考用户手册。

入湖范围

在咨询阶段盘点的数据资产中，政务领域的有效表和业务指标。治理实施项目将聚焦政务领域，结合选取的业务指标，从咨询项目输出的数据资产目录中选取表作为集成入湖范围，并完成数据集入湖。

选取出来的表及时做好标记，入湖后也要做好记录，避免出现重复入湖动作。

4.3.2 数据集成前探查

数据探查是指用适当的统计、分析等方法对原始数据进行分析，对其特征加以汇总和理解，提取有用信息和形成结论而对数据加以详细研究和概括总结的过程，以求在数据入湖前最大化地了解数据现状，为数据集成方案的设计提供数据依据。

步骤1 数据探查分析流程

依据数据集入湖清单展开数据探查，重点探查与分析表与字段注释、数据类型、数据长度、数据量、数据主键、时间戳字段以及数据空值及分布情况等内容。在获取数据接入授权与数据源端连接信息后，确定数据探查方法并实施探查，参照数据标准与数据集需求对探查结果进行分析，最终输出数据探查分析结果，形成问题清单，指导数据集集成方案设计。

步骤2 数据探查实施

数据探查接入采用CDM工具将业务源端数据迁移至DWS数据库的数据接入方式，开展数据的探查与分析活动，该方法具备安全稳定与灵活复用的特性。

数据探查实施采用平台采集与手动分析相结合的方式，首先在数管平台的“数据资产”模块中，创建元数据采集作业，勾选“数据概要”配置，对数据概要信息进行采集，可以基于不同的探查目的进行单表探查、多表探查、全量探查、采样探查，同时支持手动单次采集与周期调度采集。采集完成后可在资产管理界面进行查看与刷新。

步骤3 数据探查结果分析

针对数据探查结果，对待入湖的数据进行数据量、主键以及时间戳等内容的分析，判断数据质量问题，制定数据集集成方案。

----结束

4.3.3 全量数据同步

针对数据增长缓慢、非核心业务相关、无时间戳、无增删改标识、数据量低于十万条的明细表数据，以及码表、维度表等数据，采用全量同步的方式接入数管平台，并存

放在贴源层的全量表内。数据初始化全量同步是指第一次从业务源库中采用全量同步的方式获取一次数据。

步骤1 确定数据源

通过调研获取IT系统以及数据字典信息，从调研结果中确认数据来源系统。

步骤2 目标库手动建表

根据需要的表和字段，在MRS-Hive数据库中建立相应的库和表

表及字段的命名规范为：ods_源系统简写_源系统表名_{TYPE简码},全量表的type简码为f,所有的表名称及字段名称小写

标准字段：ext_etl_dt（ETL时间），ext_src_sys_id（源系统标识），dt、ph（分区字段）

表的其他要求：采用ORC存储格式和SNAPPY压缩；

在数管平台的“数据开发”模块，选择MRS_Spark_Agent数据连接和ods_mes_bc数据库，执行对应的DDL,创建以_f结尾的全量表。

步骤3 新建CDM连接

入口路径：一体化数据基础平台控制台 > 数据集成

在“集群管理”标签页面，单击“作业管理”进入集成作业作业管理页面

图 4-7 新建 CDM 连接



单击“连接管理”标签，进入新建和编辑连接的界面，这里配置CDM产品访问源库的权限，主要包括源端数据库和目的端数据库的类型，地址、端口、账号、密码等。

第一步：创建源端（Oracle）的数据连接（如果已创建，则跳过）

第二步：创建目的端（MRS-Hive）的数据连接（如果已创建，则跳过）

数据库连接配置成功后测试连通性，测试成功即配置完成。

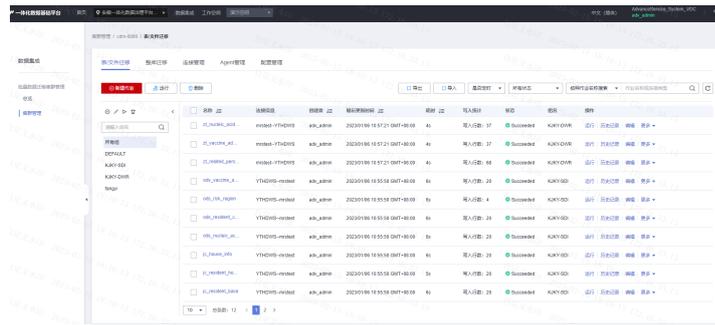
第一步：源端数据连接的连通性测试

第二步：目的端数据连接的连通性测试

步骤4 新建CDM作业

打开“表/文件迁移”标签页，为便于作业管理，需要为作业创建分组，当前平台中已创建有MES_BC组用于存放所有来源于MES_BC数据源的数据迁移作业，MES数据源的迁移作业隶属于MES_BC作业组，单击MES_BC作业组后在页面右上角检索框中查询作业名称，若无返回结果则单击页面左上角“新建作业”。

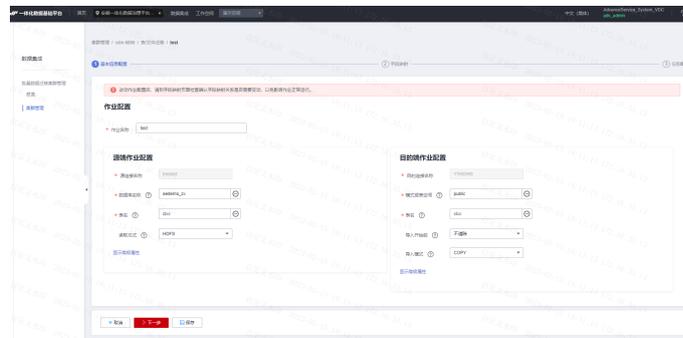
图 4-8 新建 CDM 作业 1



选择源端的库、表，并且选择目的端手动创建的表。

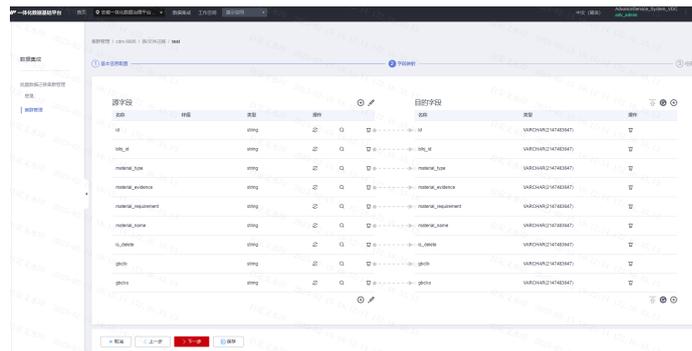
单击“下一步”，选择需要同步的表字段

图 4-9 新建 CDM 作业 2



添加标准字段

图 4-10 新建 CDM 作业 3



添加的标准字段的具体值如下：

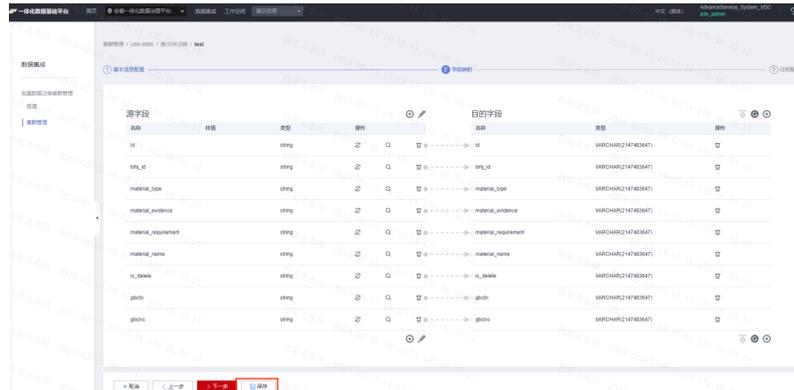
表 4-34 标准字段值

ext_etl_dt	$\{\$\{dateformat (yyyy-MM-dd HH:mm:ss) \}$
ext_src_sys_id	mes_bc

ext_etl_dt	<code>\${dateformat (yyyy-MM-dd HH:mm:ss) }</code>
dt	<code>to_char (sysdate,'yyyy-mm-dd')</code>

选择作业分组，单击“保存”将保存作业；单击“保存并运行”将保存并启动作业。

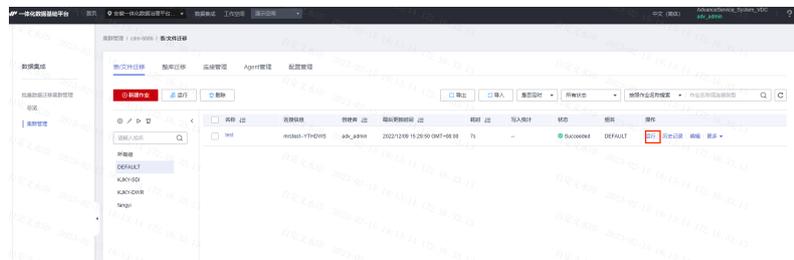
图 4-11 新建 CDM 作业 4



步骤5 CDM作业手动执行

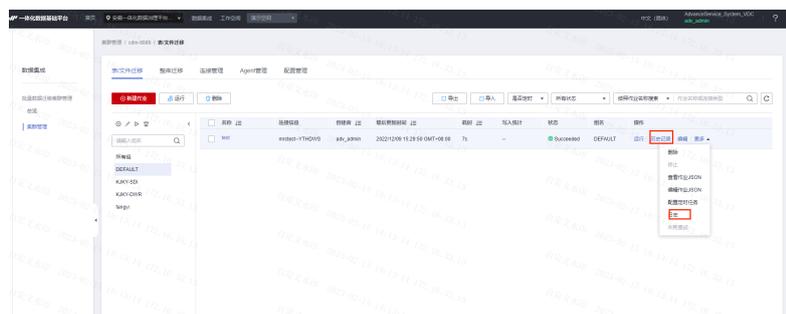
单击作业右侧的“运行”，单次运行作业。

图 4-12 CDM 作业手动执行



单击“历史记录”，可以看到历史运行记录及日志

图 4-13 CDM 作业手动执行 2

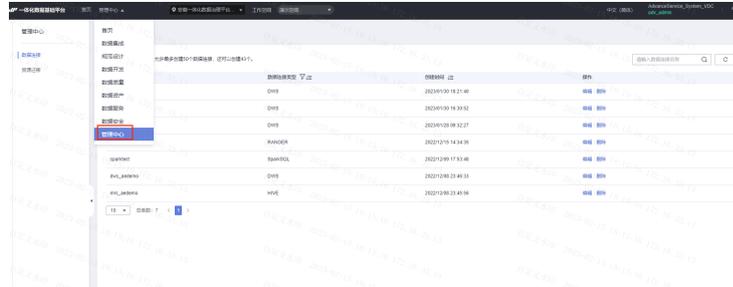


步骤6 全量同步ETL脚本开发

步骤7 创建数据开发连接

入口路径：一体化数据基础平台控制台 > 管理中心

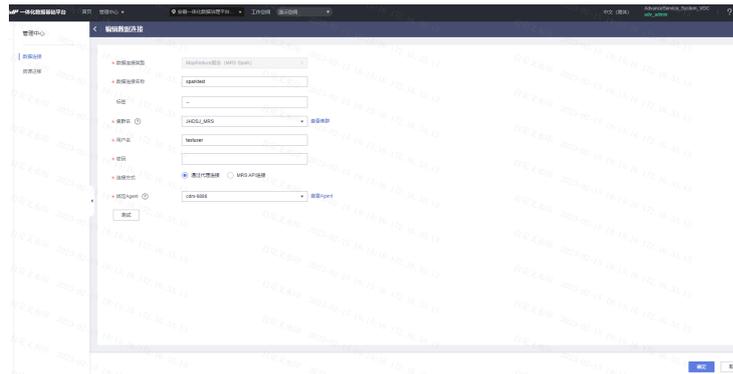
图 4-14 创建数据开发连接 1



登录一体化数据基础平台控制台，单击“管理中心”，创建数据开发的连接（如果已创建，则跳过）

单击“创建数据连接”，创建spark数据连接（MRS_Spark_Agent）

图 4-15 创建数据开发连接 2



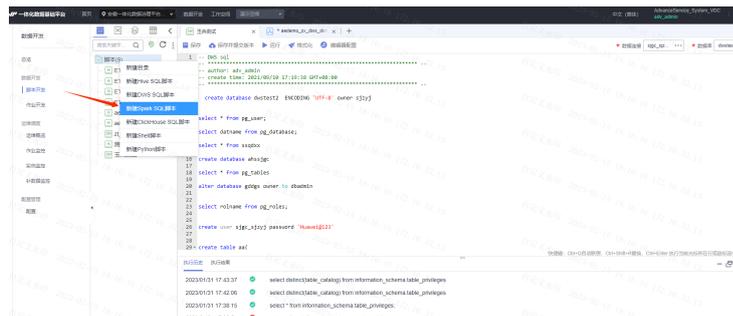
测试连接的连通性

步骤8 创建脚本

入口路径：一体化数据基础平台控制台 > 数据开发

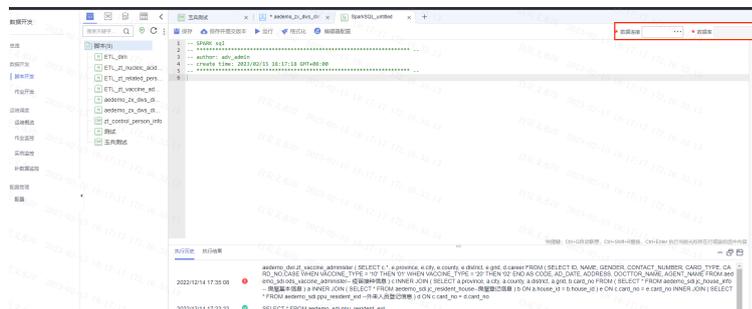
单击“数据开发”，进入数据开发的界面，新建Spark SQL脚本

图 4-16 创建脚本 1



选择数据连接和数据库

图 4-17 创建脚本 2



添加表头注释

脚本开发

脚本测试（测试阶段可以把参数值换成具体的值进行测试，这里使用参数为了方便以后的作业调度）

测试完成后，单击“保存并提交版本”

按照脚本的命名规则命名，并选择对应的目录，单击“确定”

----结束

4.3.4 增量数据同步

CDM工具同样支持数据的增量同步，经过与业务方确认，考虑到可能存在业务方修改180天内历史数据的场景，因此在增量数据同步中以180天的时间范围获取数据。根据源表时间戳，获取源表内最新180天内数据存放在贴源层的增量分区表内，并通过ETL脚本将增量同步的数据进行增转全处理后存放至贴源层的全量分区表中。需要注意的是在增量同步数据前，需要进行一次初始化全量同步。增量分区表内历史数据的存放周期为7日，对应的全量分区表内历史数据的存放周期为8日，用于数据的历史回溯，且历史数据的存放周期可依据后期数据实际使用需求进行相应的调整。以“订单进程量管理详细表”为例，说明数据增量同步流程。

步骤1 确定数据源

通过调研获取数据源以及数据字典信息，从调研结果中确认数据来源系统。

步骤2 目标库手动建表

根据需要的表和字段，在MRS-Hive数据库中建立相应的库和表；

表及字段的命名规范为：ods_源系统简写_源系统表名_{TYPE简码},增量表的type简码为i,全量表的type简码为f,所有的表名称及字段名称小写

标准字段：ext_etl_dt（ETL时间），ext_src_sys_id（源系统标识），dt、ph（分区字段）

表的其他要求：采用ORC存储格式，并采用SNAPPY压缩

在数管平台的“数据开发”模块，选择MRS_Spark_Agent数据连接和ods_mes_bc数据库，执行对应的DDL,创建以_i结尾的增量表和以_f结尾的全量表。

步骤3 新建CDM连接

入口路径：一体化数据基础平台控制台 > 数据集成

在“集群管理”标签页面，单击“作业管理”进入集成作业作业管理页面

图 4-18 新建 CDM 连接



单击“连接管理”标签，进入新建和编辑连接的界面，这里配置CDM产品访问源库的权限，主要包括源端数据库和目的端数据库类型，地址、端口、账号、密码等。

第一步：创建源端（Oracle）的数据连接（如果已创建，则跳过）

第二步：创建目的端（MRS-Hive）的数据连接（如果已创建，则跳过）

数据库连接配置成功后测试连通性，测试成功即配置完成。

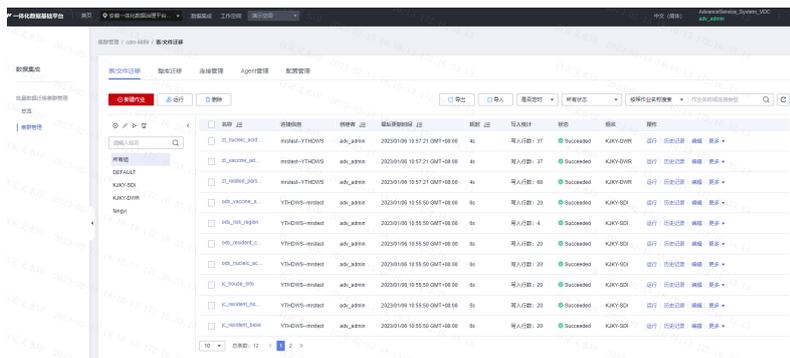
第一步：源端数据连接的连通性测试

第二步：目的端数据连接的连通性测试

步骤4 新建CDM作业

打开“表/文件迁移”标签页，为便于作业管理，需要为作业创建分组，当前平台中已创建有MES_BC组用于存放所有来源于MES_BC数据源的数据迁移作业，MES数据源的迁移作业隶属于MES_BC作业组，单击MES_BC作业组后在页面右上角检索框中查询作业名称，如果无返回结果则单击页面左上角“新建作业”。

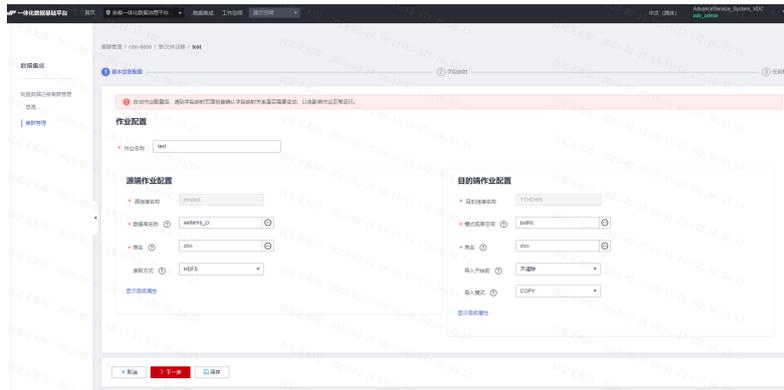
图 4-19 新建 CDM 作业 1



选择源端的库、表，并且选择目的端手动创建的表，然后加上增量条件。

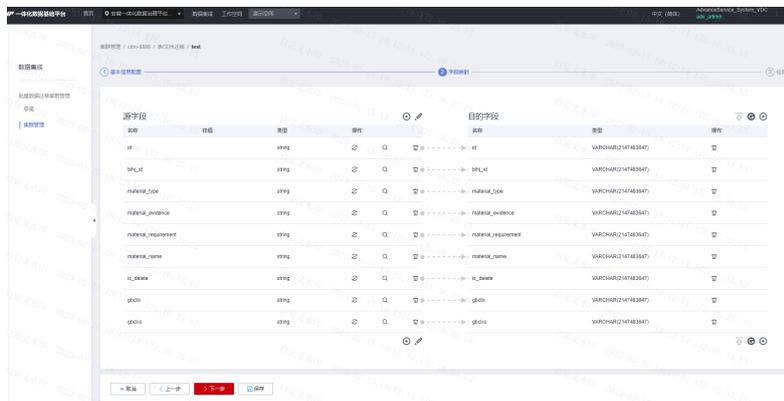
单击“下一步”，选择需要同步的表字段

图 4-20 新建 CDM 作业 2



添加标准字段

图 4-21 新建 CDM 作业 3



添加的标准字段的具体值如下：

表 4-35 标准字段具体值

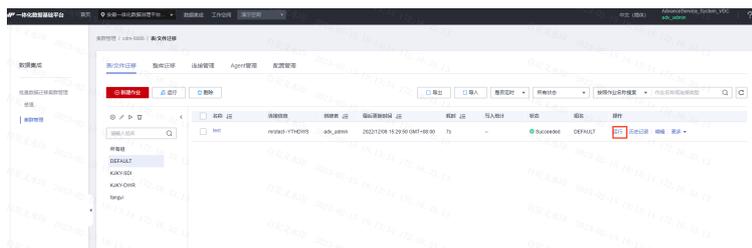
ext_etl_dt	\${dateformat (yyyy-MM-dd HH:mm:ss) }
ext_src_sys_id	mes_bc
dt	to_char (sysdate,'yyyy-mm-dd')

选择作业分组，单击“保存”将保存作业；单击“保存并运行”将保存并启动作业。

步骤5 CDM作业手动执行

单击作业右侧的“运行”，单次运行作业。

图 4-22 CDM 作业手动执行 1



单击“历史记录”，可以看到历史运行记录及日志

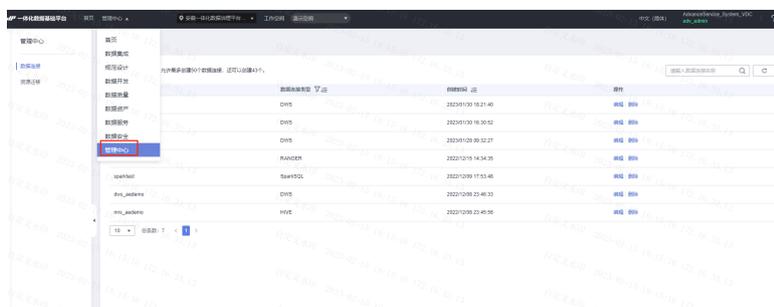
图 4-23 CDM 作业手动执行 2



步骤6 增量同步ETL脚本开发

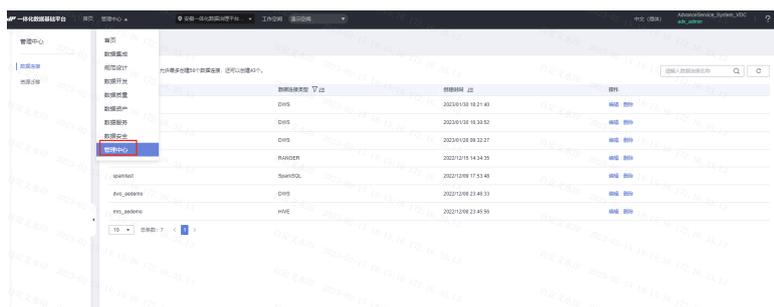
创建数据开发连接，入口路径：控制台 > 管理中心

图 4-24 管理中心 1



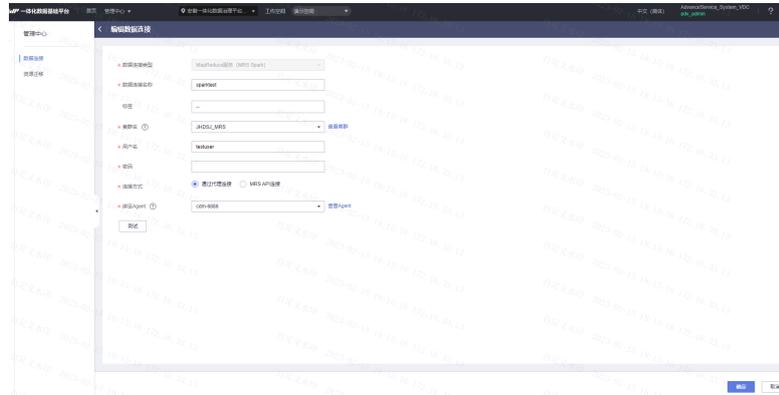
登录DataArts Studio控制台，单击“管理中心”，创建数据开发的连接（如果已创建，则跳过）

图 4-25 管理中心 2



单击“创建数据连接”，创建spark数据连接（MRS_Spark_Agent）

图 4-26 管理中心 3



测试连接的连通性

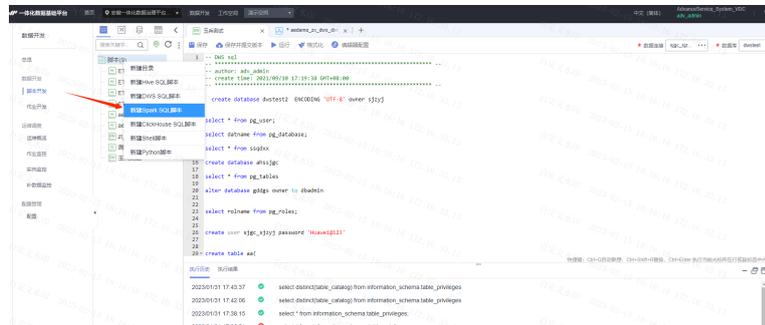
创建脚本

入口路径：控制台 > 数据开发

单击“数据开发”，进入数据开发的界面

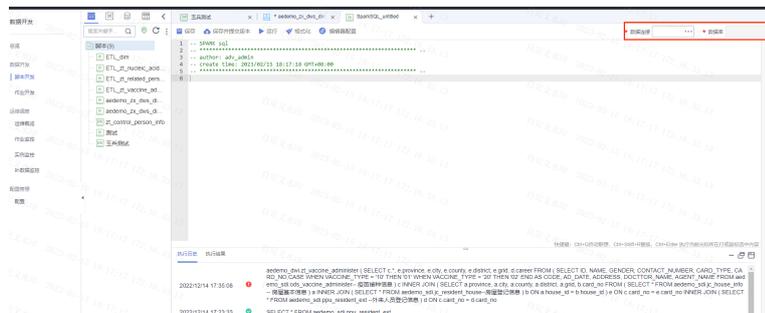
新建Spark SQL脚本

图 4-27 数据开发 1



选择数据连接和数据库

图 4-28 数据开发 2



添加表头注释

设置脚本参数

开发增转全脚本

增转全规则：通过增量时间戳获取180天的增量数据，然后将增量数据和180天前的数据进行整合，存放在新生成的分区里。

删除分区

保留7天的历史数据，用于数据回溯和脚本重跑

脚本测试（测试阶段可以把参数值换成具体的值进行测试，这里使用参数为了方便以后的作业调度）

测试完成后，单击“保存并提交版本”

按照脚本的命名规则命名，并选择对应的目录，单击“确定”

----结束

4.4 规范设计

4.4.1 配置中心

- 审核人管理

审核人必须是当前工作空间下具有审核权限的成员，需要先在“首页-空间管理”的工作空间内编辑并添加空间成员

图 4-29 审核人管理 1



图 4-30 审核人管理 2



在控制台首页，选择实例，单击“进入控制台”，选择对应工作空间的“规范设计”模块，进入规范设计页面

图 4-31 审核人管理 3



在规范设计控制台，单击左侧导航树中的“配置中心”，进入相应页面后，选择“审核人管理”页签。

图 4-32 审核人管理 4



在“审核人管理”页面，单击“添加”按钮，在弹出的添加对话框中，选择要添加的审核人即可

图 4-33 审核人管理 5



根据需要，可以添加多个审核人。

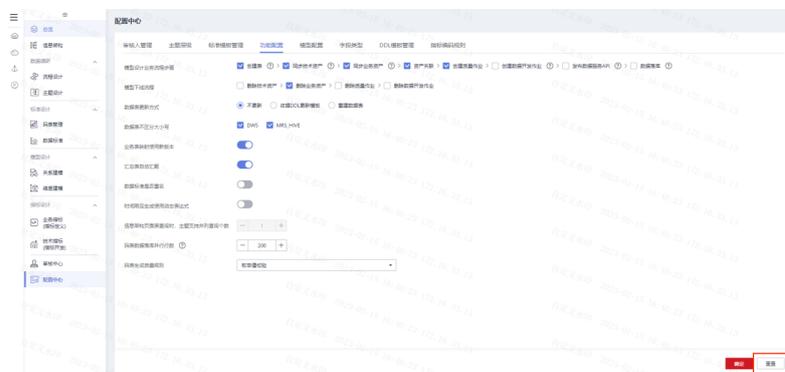
图 4-34 审核人管理 6



- 功能配置

在功能配置页面，可根据用户具体的功能需求配置参数，然后单击“确定”。如果单击“重置”可恢复默认设置。

图 4-35 功能配置



模型设计业务流程步骤：此处勾选的流程，在关系建模或维度建模的对象发布上线时，系统会依次自动执行。

创建表：当规范设计中的表发布并通过审核后，系统将自动在对应的数据源中创建相应的物理表。在表删除时，系统也会自动删除物理表。

同步技术资产：关系建模或维度建模中的表发布后，同步表到数据资产模块作为技术资产，同时同步标签到对应技术资产。

同步业务资产：同步逻辑模型到数据资产，作为业务资产，同时同步标签到对应业务资产。

资产关联：实现业务资产与技术资产的关联。业务资产与技术资产同步完成后，在数据资产模块中查看对应的业务资产或技术资产详情时，可以看到相关联的技术资产或业务资产。该功能要求表信息中含有数据源信息。

创建质量作业：当关系建模或维度建模中的表发布并通过审核后，对于关联数据标准（包含数据长度或允许值）或关联质量规则的表，系统将自动在数据质量模块中创建一个质量作业。

创建数据开发作业：汇总表发布后，自动生成端到端的全流程数据开发作业。

发布数据服务API：汇总表发布后，自动生成数据服务API，此功能仅当数据服务支持汇总表的数据连接时生效。

数据落库：码表维度发布后，会自动将码表的数值填入维度表中。

模型下线流程：选择当模型下线时，是否同步删除技术资产、业务资产、质量作业、数据开发作业。

数据表更新方式：当规范设计中的表在发布后进行了修改，是否同时更新数据库中的表。默认为“不更新”，但在配置中心可以依据自己的需求设置更新动作。依据DDL模板，在模板里面配置对应的更新语句即可。

不更新：不更新数据库中的表。

依据DDL更新模板：依据DDL模板管理中配置的DDL更新模板，更新数据库中的表，但能否更新成功是由底层数仓引擎的支持情况决定的。由于不同类型的数仓支持的更新表的能力不同，在规范设计控制台中所做的表更新操作，如果数仓不支持，则无法确保数据库中的表和规范设计中的表是一致的。

重建数据表：先删除数据库中已有的表，再重新创建表。选择该选项可以确保数据库中的表和规范设计中的表是一致的，但是由于会先删除表，因此一般建议只在开发设计阶段或测试阶段使用该选项，产品上线后不推荐使用该选项。

数据表不区分大小写：对于选中的连接类型，在发布相应类型的表时，同步技术资产时名称将不区分大小写，找到相同的即认为已存在。

业务表映射使用新版本：系统默认为新版本映射。新版本映射功能支持join等操作，推荐使用新版本映射。

汇总表自动汇聚：发布衍生指标或复合指标时，系统支持自动生成汇总表，一个统计维度对应一个汇总表。自动生成的汇总表可在汇总表页面下选择“自动汇聚”页签查看。

数据标准是否重名：默认关闭，打开后数据标准可以重名。

信息架构页面表查询时，主题支持并列查询个数：默认为1个，暂不支持设置。

码表数据落库并行行数：码表维度发布后，设置将码表的数值填入维度表中时的并行操作行数。当码表数值较多时，会导致落库失败，可以适当调小该参数。

码表生成质量规则：下拉选择即可。当码表的数据量较小时，选择“枚举值校验”即可；否则选择字段一致性校验。

⚠ 注意

当多人或多个团队在同时使用这个功能时，要协调好功能配置项，否则会造成个别的操作失败。

- 模型配置

在模型设计过程中，设置维度（维度表）、事实表、汇总表的默认表编码前缀，通过以下页面进行配置。

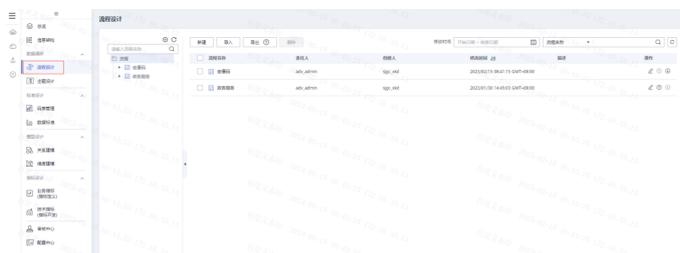
4.4.2 流程设计

宏观层面：重点是组织架构与流程架构匹配，组织设置必须支撑流程价值实现，组织职责必须清晰具体，不能有重叠、空白、过多、过少；

微观层面：看流程角色职责是否通过岗位职责的设置有效落实，岗位职责与其匹配的流程角色职责相符。

步骤1 单击左侧导航栏“流程设计”并进入流程设计页面。

图 4-36 流程设计 1



步骤2 单击导入，单击下载流程模板。

图 4-37 流程设计 2



步骤3 下载模板如下

表 4-36 模板

上级流程	*名称	*责任人	描述
-	-	-	-
-	-	-	-

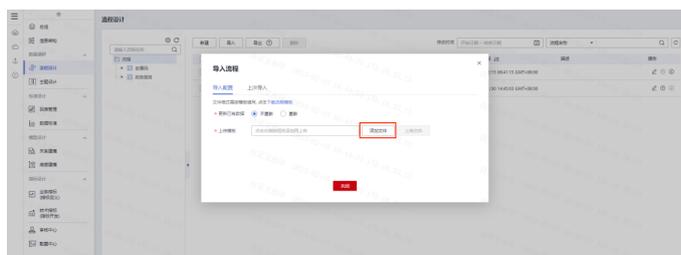
模板参数说明如下：

表 4-37 模板参数说明

参数名	说明
上级流程	第一层的流程，其上级流程为空，不用填。 非第一层的流程，其上级流程不能为空。上级流程为多级流程时，流程之间以“/”分隔。
*名称	流程名称。
*责任人	流程的责任人，可以手动输入名字或直接选择已有的责任人。
描述	流程的描述信息。

步骤4 上传模板，单击添加文件，选择填写后的流程模板，然后上传文件即可。

图 4-38 流程设计 3



导入参数配置说明

表 4-38 参数配置说明

参数名	说明
更新已有数据	如果所要导入的流程，在规范设计中已经存在，是否更新已有的流程。支持以下选项： 不更新：当流程已存在时，将直接跳过，不处理。 更新：当流程已存在时，更新已有的流程信息。 在导入流程时，只有创建或更新操作，不会删除已有的流程。
上传模板	选择所需导入的流程设计文件。

导入完成。

----结束

4.4.3 主题设计

主题设计是通过分层架构表达对数据的分类和定义，帮助厘清数据资产，明确业务领域和业务对象的关联关系。

步骤1 单击左侧导航栏“主题设计”并进入主题设计页面。

图 4-39 主题设计 1



步骤2 单击导入，单击下载主题导入模板。

图 4-40 主题设计 2



下载模板如下：

表 4-39 下载模板

上级流程	*名称	*编码	别名	描述	数据owner部门	*数据owner人员
-	-	-	-	-	-	-
-	-	-	-	-	-	-

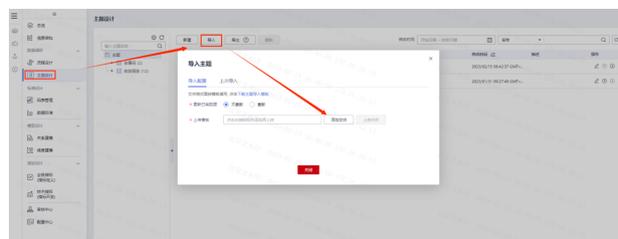
模板参数说明如下：

表 4-40 模板参数说明

参数名	说明
上级主题	上层主题的编码路径，以/分隔。
*名称	中文名称。只允许除/、\、<、>以外的字符。
*编码	英文名称。只允许英文字母、数字、空格、下划线、中划线、左右括号以及&符号。
别名	主题对象的别名。
描述	主题对象的描述信息。对于最低层级主题，此项参数为必选。在导入文件中应补充最低层级主题的描述信息。
数据owner部门	数据的拥有者所在部门。对于最低层级主题，此项参数为必选。在导入文件中应补充最低层级主题的数据owner部门信息。
*数据owner人员	数据的拥有者，支持填写多个，中间以逗号分隔。

上传模板，单击添加文件，选择填写后的主题模板，然后上传文件即可。

图 4-41 主题设计 3



导入参数配置说明

表 4-41 参数配置说明

参数名	说明
更新已有数据	在导入时是否更新已有的主题信息（主题域分组、主题域或业务对象）。在导入时，系统将按编码判断将要导入的主题信息在系统中是否已存在。 不更新：当主题信息已存在时，将直接跳过，不更新。 更新：当主题信息已存在时，更新已有的主题信息。 在导入主题信息时，只有创建或更新操作，不会删除已有的主题信息。
上传模板	选择所需导入的主题设计文件。

导入完成。

----结束

4.4.4 码表管理

码表管理是对业务调研阶段得到的码表进行落地，由可枚举数据构成，存储枚举数据名称与编码的映射关系。码表的作用主要有：

在数据清洗中用于标准化业务数据以及补充映射字段

在质量监控中用于监控业务数据的值域范围

在维度建模中可以引申为枚举维度

步骤1 单击左侧导航栏的码表管理并进入码表管理页面。

图 4-42 码表管理 1



图 4-43 码表管理 2

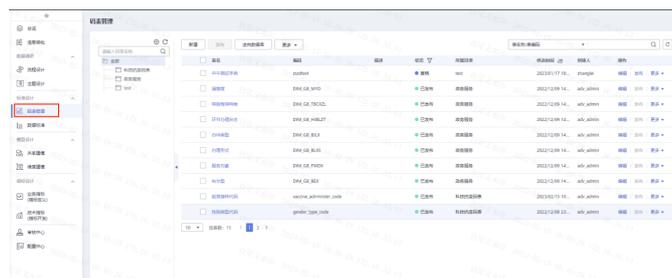
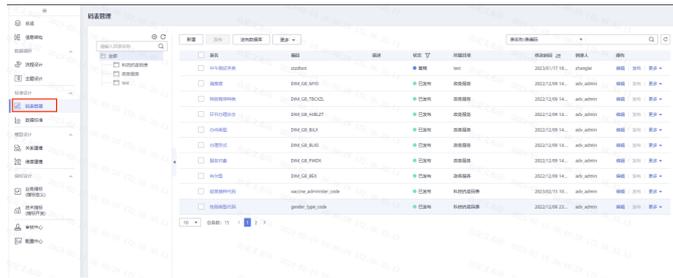


图 4-44 码表管理 3



步骤2 在添加码表之前需要为码表创建一个目录分类，用于存放相应的码表。单击左侧的加号，输入目录名称，单击保存。

图 4-45 码表管理 4



步骤3 在左侧的目录树中，选择一个目录，再单击“更多 > 导入”。您也可以在所选择的码表目录上单击鼠标右键，然后选择菜单“导入”。

图 4-46 码表管理 5

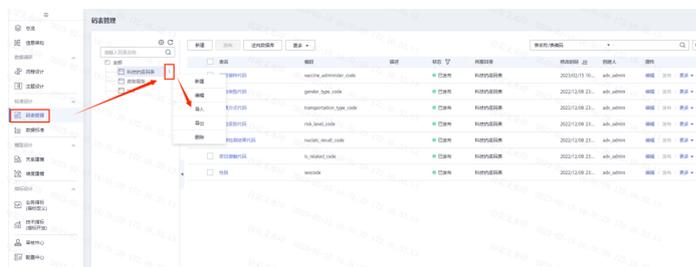
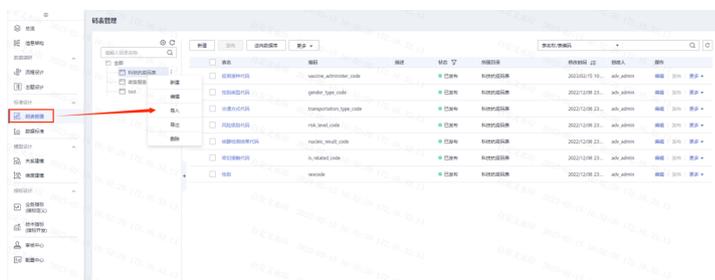
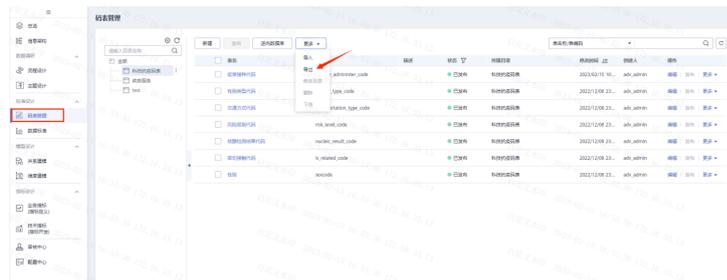


图 4-47 码表管理 6



步骤4 下载码表管理模板

图 4-48 码表管理 7



下载模板如下

表 4-42 下载模板

目录	*表名称	*表编码	表描述	*字段名称	*字段编码	*字段数据类型	字段描述	是否生成标准
-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-

模板参数说明如下：

表 4-43 模板参数说明

参数名称	说明
目录	码表所属的目录。多级目录以“/”分隔。
*表名称	码表名称。只能包含中文、英文字母、数字、左右括号、中划线和下划线，且以中文或英文字母开头。
*表编码	码表的英文名称。只能包含英文字母、数字、下划线，且以英文字母开头。
表描述	码表的描述信息。支持的长度0~600个字符。
*字段名称	字段名称。只能包含中文、英文字母、数字、左右括号、空格、中划线和下划线，且以中文或英文字母开头。
*字段编码	字段编码。只能包含英文字母、数字、下划线，且以英文字母开头。
*字段数据类型	支持的数据类型有：STRING、BIGINT、DOUBLE、TIMESTAMP、DATE、BOOLEAN、DECIMAL。
字段描述	字段的描述信息。支持的长度0~600个字符。

参数名称	说明
是否生成标准	true：生成数据标准。 false：不生成数据标准。默认为false。 注意：如果要自动生成数据标准，还需在“配置中心 > 标准模板管理”中勾选上“引用码表”选项。

步骤5 在“导入码表”对话框中，根据页面提示“添加文件”，然后单击“上传文件”。

图 4-49 上传文件



表 4-44 导入参数配置说明

参数名	说明
更新已有表	在导入时是否更新已有的码表信息。在导入时，系统将按编码进行判断将要导入的码表在系统中是否存在。支持以下选项： 不更新：当码表已存在时，将直接跳过，不更新。 更新：当码表已存在时，更新已有的码表信息。如果码表处于“已发布”状态，码表更新后，您需要重新发布码表，才能使更新后的表生效。 在导入码表时，只有创建或更新操作，不会删除已有的码表。
上传模板	选择所需导入的码表文件。

步骤6 导入后发布。

---结束

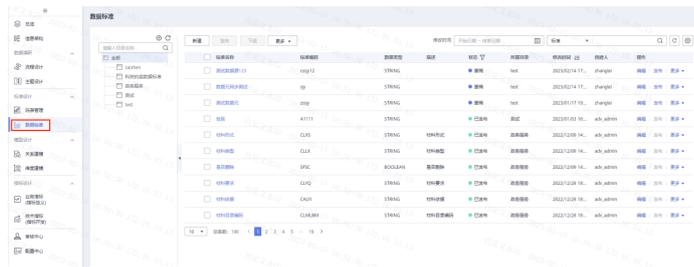
4.4.5 数据标准

数据标准是用于描述政务层面需共同遵守的数据含义和业务规则，它描述了政务层面对某个数据的共同理解，这些理解一旦确定下来，就应作为政府层面的标准在政府内被共同遵守。

数据标准也称数据元，由一组属性规定其定义、标识、表示和允许值的数据单元，是不可再分的最小数据单元。您可以将数据标准关联到各个业务上的数据库中。其中，标识符、数据类型、表示格式、值域是数据交换的基础，它们用于描述表的字段元信息，规范字段所存储的数据信息。

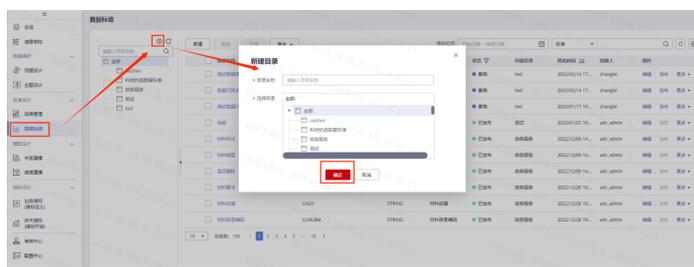
步骤1 进入规范设计单击左侧数据标准菜单，进入数据标准。

图 4-50 数据标准 1



步骤2 为数据标准建立目录分类，用于存放相应的数据标准。单击左侧的加号，输入目录名称，单击保存即可。

图 4-51 数据标准 2



步骤3 在数据标准的目录结构中，选择一个指定的目录名称，然后单击上方“更多 > 导入”，单击下载数据标准导入模板。

图 4-52 数据标准 3

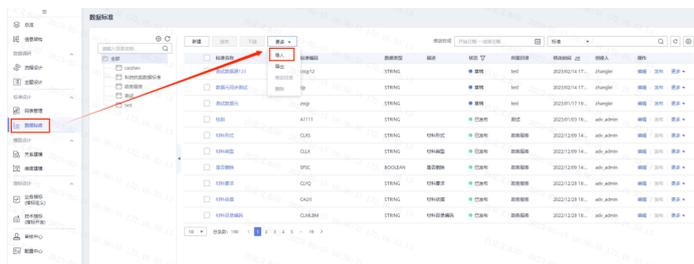


图 4-53 数据标准 4



步骤4 下载模板如下

表 4-45 下载模板

目录	* 标准名称	* 标准编码	* 数据类型	* 数据长度	* 是否有允许值	允许值	引用码表	* 业务规则责任人	数据监控责任人	* 标准层级	同义词	* 版本号	* 可信数据源	* 业务规则	数据示例	* 主题域分组	业务定义	主题域	业务对象	* 特征类型
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

模板参数说明如下：

表 4-46 模板参数说明

参数名称	说明
目录	导入的数据标准所属的目录。
*标准名称	数据标准的中文名称。只能包含中文、英文字母、数字、左右括号、空格、中划线和下划线，且以中文或英文字母开头。
*标准编码	支持自动生成和自定义两种方式。
*数据类型	数据类型有：STRING、BIGINT、DOUBLE、TIMESTAMP、DATE、BOOLEAN、DECIMAL。
数据长度	可以为空，或者输入1~10000之间的数值。数据长度为空时，对数据长度不做限制。 如果输入了数据长度标准，当数据类型为STRING时，会为关联该标准的属性创建数据质量作业，其他类型暂不支持创建质量作业。
是否有允许值	true表示有允许值，false表示没有允许值。
允许值	当参数“是否有允许值”为true时，必须设置“允许值”。支持添加多个允许值，最多支持20个。 多个允许值之间以逗号分隔。
引用码表	填写已创建的码表名称。
码表字段	当“引用码表”不为空时，请设置该引用码表中的“码表字段”，这样就可以将码表字段和数据标准相关联。
业务规则责任人	填写业务规则责任人，可以手动输入名字或直接选择已有的责任人。
数据监控责任人	填写数据监控责任人，可以手动输入名字或直接选择已有的责任人。

参数名称	说明
标准层级	global: 全局级别。domain: 非全局级别。
描述	描述信息。支持的长度0~600字符。
用户自定义字段（可选）	如果在定制数据标准模板时，您添加了一个或多个自定义字段，则在导入模板中也需要填写相应的字段，如果未添加自定义字段，则无需填写

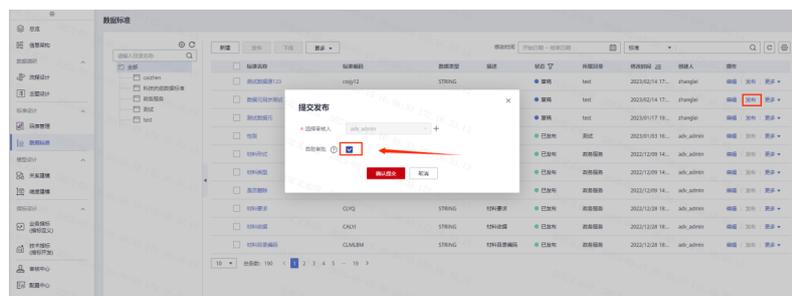
在“导入数据标准”对话框，选择上一步配置好的数据标准模板文件，然后单击“上传文件”。如果上传的模板文件校验不通过，请修改正确后，再重新上传。

图 4-54 重新上传



单击发布，发布的时候勾选自助审批。

图 4-55 发布



标准发布完成。

---结束

4.4.6 关系建模

逻辑模型是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，并对各实体之间的关系进行准确定义。同时，逻辑模型也为构建物理模型提供了有力的参考依据，是最终成功设计数据库过程中必不可少的一个阶段。

以人口逻辑模型为例：

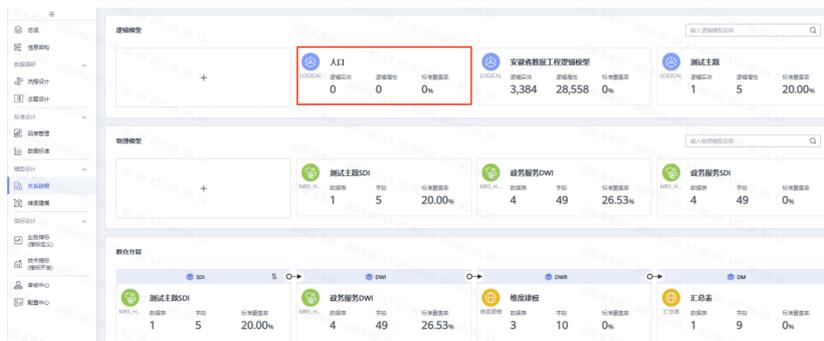
步骤1 单击逻辑模型下面加号，进入新建页面，输入模型名称为“人口”，单击确定。

图 4-56 关系建模 1



步骤2 单击逻辑模型“人口”，进入逻辑模型“人口”的管理界面。

图 4-57 关系建模 2



步骤3 单击更多，选择“导入”，单击“下载关系建模导入模板”

图 4-58 关系建模 3



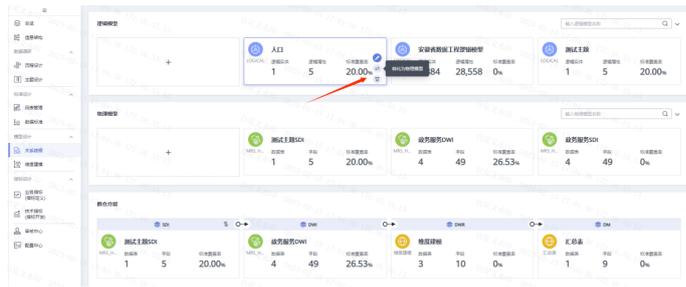
----结束

物理模型是指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（Table Relationship Diagram）的一种物理描述。物理模型建模通常是由逻辑模型创建之后，统一转换成物理模型。

以人口物理模型为例：

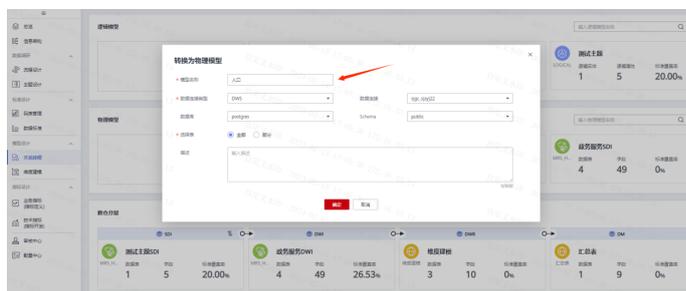
步骤1 选中“人口”逻辑模型，单击转换为物理模型按钮，进入编写页面。

图 4-59 关系建模 4



步骤2 物理模型名称输入“人口”，更新已有表选择更新，最后选择相应的数据连接类型、数据连接、数据库，最后单击确定即可。

图 4-60 关系建模 5



步骤3 发布物理模型表结构。转换过去的物理模型表结构为草稿状态，勾选之后单击发布，就会将逻辑模型创建好表信息及结构创建到数据库中形成物理模型。

图 4-61 关系建模 6



---结束

4.4.7 业务指标

经过数据调研和需求分析之后，根据需求落地指标。指标是衡量目标总体特征的统计数值，是能表征某一业务活动中业务状况的数值指示器。指标一般由指标名称和指标数值两部分组成，指标名称及其涵义体现了指标质的规定性和量的规定性两个方面的特点，指标数值反映了指标在具体时间、地点、条件下的数量表现。业务指标用于指导技术指标，而技术指标是对业务指标的具体实现。

● 指标确认

指标数据是对基础类数据加工而产生的指标数据标准化规范，通过业务属性、技术属性和管理属性来描述指标数据规范化要求。

业务属性：对指标应遵循的业务属性的统一定义与解释，如：业务定义、业务口径、指标维度、业务流程等

技术属性：业务应用对指标技术属性的统一要求与定义，如：数据类型、取数范围、计量单位等

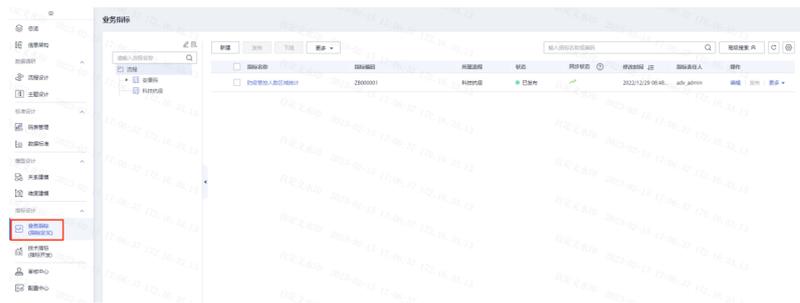
管理属性：对数据进行管理的统一要求与定义，如：管理部门等

实施团队收到咨询的业务指标，每个指标包括了业务属性、技术属性、管理属性。实施团队需要对相应的指标进行检查，主要审核指标的定义和口径是否清晰、相关的信息项是否填写完整、源数据是否满足指标需求等。

- 指标导入

单击左侧导航栏的业务指标并进入业务指标页面。

图 4-62 指标导入 1



单击“更多 > 导入”，下载业务指标管理模板。

图 4-63 指标导入 2



模板如下：

表 4-47 模板

* 流程架构	* 指标名称	指标编码	指标别名	* 设置目的	* 指标定义	* 计算公式	数据来源	计量单位	* 统计周期	统计维度	* 刷新频率	统计口径 & 修饰词	指标应用场景	备注	度量对象	度量对象	* 指标管理部门	* 指标责任人	业务对象	关联技术指标
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

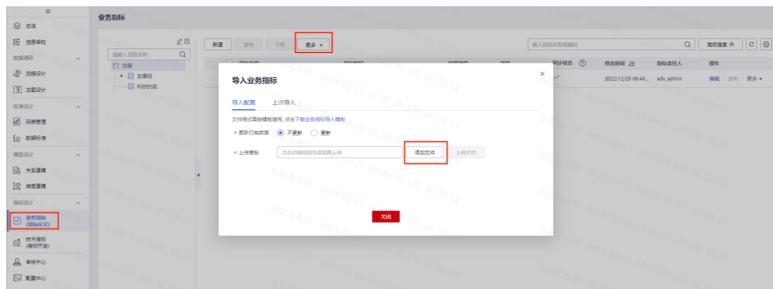
模板参数说明如下：

表 4-48 模板参数说明

参数说明	说明
指标名称	业务指标的名称。
指标编码	指标编码是自动生成的，生成规则可以在规范设计的“配置中心”页面进行配置。
指标别名	可选参数。
所属流程	选择指标所属的业务流程
设置目的	描述设置该指标的目的。
指标定义	需准确描述指标的定义。
备注	备注信息。
计算公式	定义业务指标的计算逻辑，以便指导开发者根据计算公式设计原子指标、衍生指标。业务指标是为了指导技术指标的落地，实际并不做运算。
统计周期	指定指标的统计周期，以便指导开发者根据统计周期设计时间限定。
统计维度	可以在下拉列表中选择已经创建的维度。
统计口径和修饰词	限定是对业务场景限定抽象，用于度量范围的圈定。
刷新频率	指标数据的刷新频率。开发者或运维者可以依据指标的刷新频率，合理设置衍生指标的调度频率。
指标应用场景	描述指标的应用场景。
关联技术指标类型	下拉选择与业务指标关联的技术指标类型，包含衍生指标和复合指标。
关联技术指标	下拉选择与业务指标关联的技术指标。
度量对象	衡量该指标的度量字段。
计量单位	指标的计量单位。

在“导入业务指标”对话框中，根据页面提示“添加文件”，然后单击“上传文件”。

图 4-64 指标导入 3



导入后发布。

- 指标拆解

根据业务指标拆分为技术指标，指标的分类主要分为原子指标、衍生指标、复合指标。

原子指标约定为业务、业务过程、度量的组合。基于某一业务事件行为下的度量，是业务定义中不可再拆分的指标，具有明确业务含义的名称。如重量，库存量等。

衍生指标约定为修饰词、业务规则、原子指标的组合。衍生指标 = 1个原子指标 + 多个修饰词（可选）+ 时间周期，可以理解为原子指标业务统计范围的圈定，时间周期用于明确统计的时间范围或时间点。

复合指标，又有多种规则：

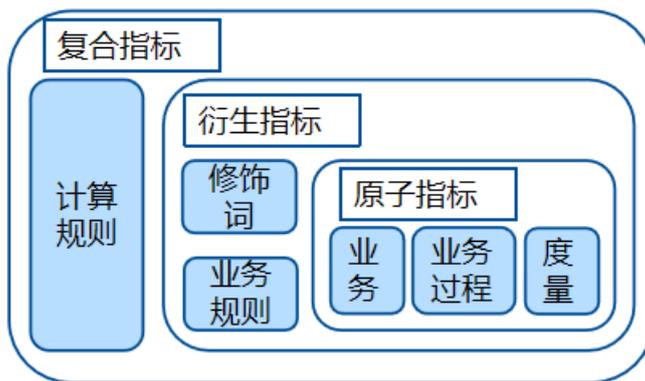
计算型，指多个原子指标或衍生指标根据加减乘除计算得来；

比率型，指两个业务指标的比值；

比例型，指某业务的占比、百分比；

原子指标、衍生指标、复合指标三类指标的关系如下：

图 4-65 指标拆解



根据以上对指标进行拆解，模版参考《政府部门指标拆解-模板》。

4.4.8 技术指标

技术指标是为了支撑业务指标的具体实现。

- 原子指标

原子指标是对指标统计逻辑、具体算法的一个抽象。为了从根源上解决定义、开发不一致的问题，指标定义明确设计统计逻辑（即计算逻辑），不需要重复开发，从而提升了开发效率，也保证了统计结果的一致性。

- 衍生指标

衍生指标是原子指标通过添加限定、维度卷积而成，限定、维度均来源于原子指标中的属性。

衍生指标=原子指标+统计维度+时间限定+通用限定。

- 复合指标

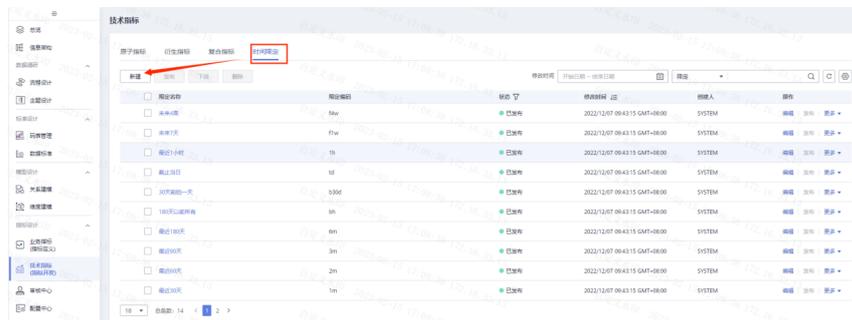
复合指标是由一个或多个衍生指标叠加计算而成，其中的维度、限定均继承于衍生指标。注意，不能脱离衍生指标、维度和限定的范围，去产生新的维度和限定。

- 时间限定

原子指标是计算逻辑的标准化定义，时间限定则是条件限制的标准化定义。为保障所有统计指标统一、标准、规范地构建，时间限定在业务板块内唯一，并唯一归属于一个来源逻辑表，计算逻辑也以该来源逻辑表模型的字段为基础进行定义。由于一个时间限定的定义可能来自于归属不同数据域的多个逻辑表，因此一个时间限定可能归属于多个数据域。以时间限定“当前年”为例。

在规范设计控制台，单击左侧导航树中的“技术指标”，选择“时间限定”页签，单击“新建”按钮。

图 4-66 时间限定 1



在新建时间限定页面，配置参数，然后单击“发布”并勾选自助审核。

图 4-67 时间限定 2

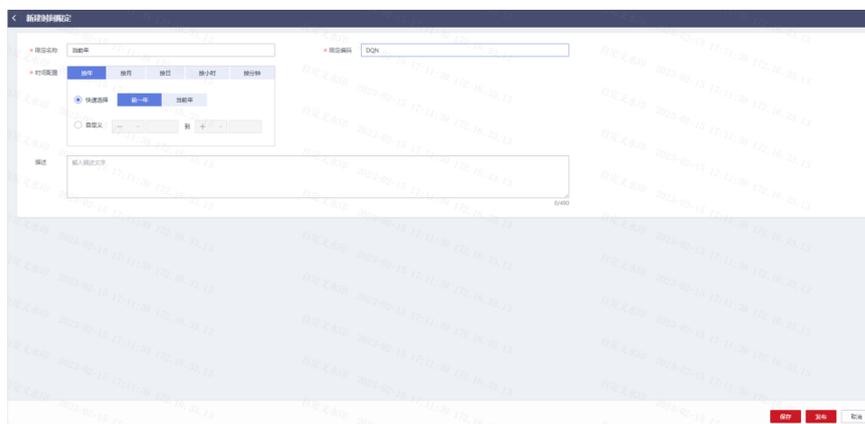
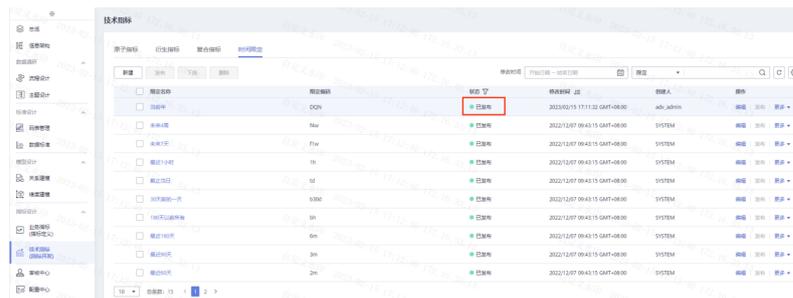


表 4-49 参数说明

参数名称	说明
*限定名称	只能包含中文、英文字母、数字和下划线，且必须以中文或英文字母开头。
*限定英文名称	只能包含英文字母、数字和下划线。
*时间配置	可选择“按年”、“按月”、“按日”、“按小时”或“按分钟”，然后根据需要选择“快速选择”或“自定义”进行时间条件的设置。自定义时，“-”表示从当前时间向前的时间段，“+”表示从当前时间向后的时间段。例如，过去一年到未来三年，可以按年自定义为“-1到+3”或“+3到-1”。
描述	描述信息。支持的长度0~490字符。

查看当前年，状态为发布。

图 4-68 时间限定 3

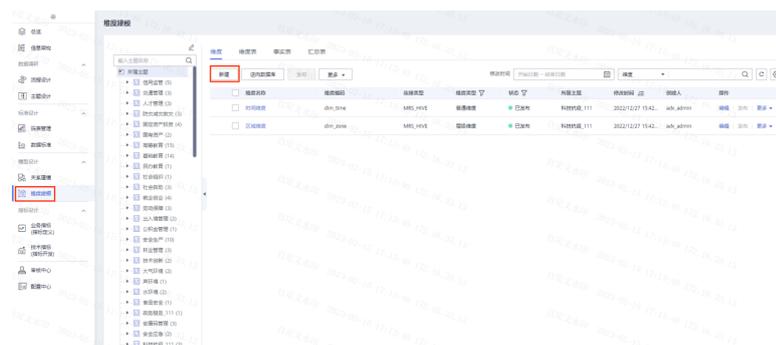


4.4.9 维度建模

维度是用于观察和分析业务数据的视角，支撑对数据汇聚、钻取、切片分析，用于SQL中的GROUP BY条件。维度多数具有层级结构，如：时间维度（其中包括年度、季度、月度等级别的内容）。

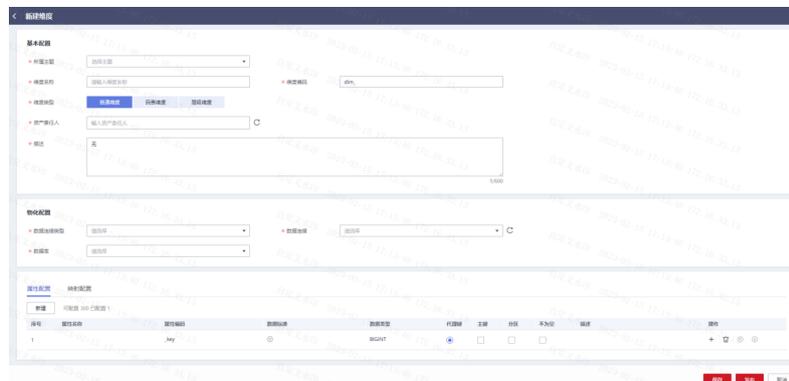
步骤1 单击左侧导航栏维度建模，进入维度建模页面，选择右边的新增，进入新增维度编辑页面。

图 4-69 维度建模 1



步骤2 选择所属主题，填入基本信息，包括：维度名称、维度编码、维度类型、资产责任人。填入物化配置，包括：数据库连接类型、选择相应数据连接、数据库、Schema。最后添加的属性配置。单击发布，勾选自助审批。

图 4-70 维度建模 2



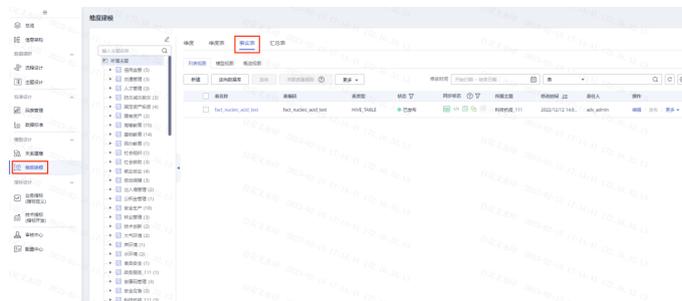
步骤3 发布之后会将维度物化到数据库中，在维度表里可以查看。

----结束

事实表： 归属于某个业务过程的事实逻辑表，可以丰富具体业务过程所对应事务的详细信息。创建事实逻辑表即完成公共的事务明细数据沉淀，从而便于提取业务中事务相关的明细数据。

步骤1 单击“维度建模”，右侧选择事实表切换到事实表管理界面。

图 4-71 事实表 1



步骤2 单击新建，进入事实表新建页面，选择所属主题，填入表名称、表编码、数据连接类型、数据库连接、Schema、资产责任人、存储方式及描述等；配置事实表添加维度字段。最后单击发布，勾选自助审批。

图 4-72 事实表 2



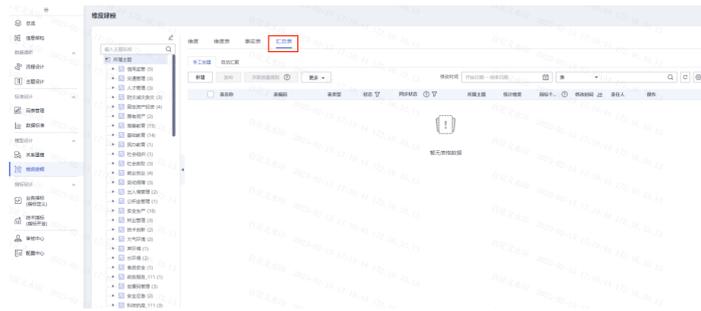
布后，在事实表中查看

----结束

汇总表的创建必须依赖衍生指标和复合指标，因此需要在技术指标添加完成之后才能进行汇总表的新建，将复合指标计算结果写入汇总表中。

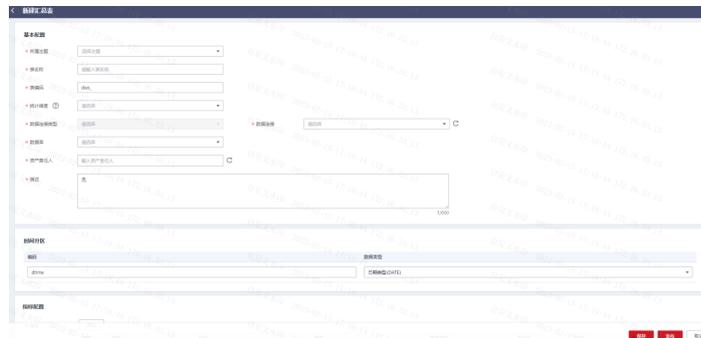
步骤1 单击规范设计左侧“维度建模”，切换到汇总表页面，单击新建，进入新增汇总表编辑页面。

图 4-73 汇总表 1



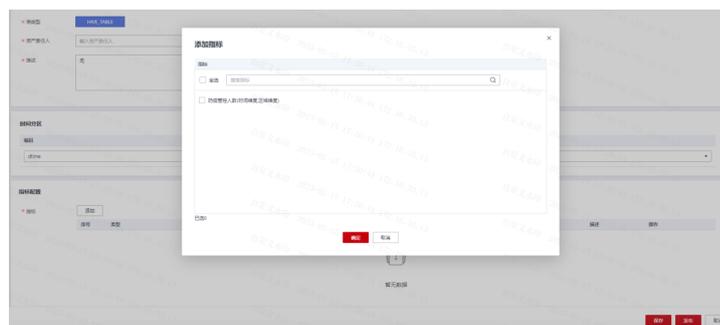
步骤2 进入新增汇总表编辑页面。填入所属主题、表名称、表编码、统计维度选择客户类型、数据连接类型、数据源、数据库和、Schema、资产责任人。

图 4-74 汇总表 2



步骤3 指标配置，单击添加，选择之前创建的衍生指标和复合指标。最后单击右下角的发布，勾选自助审批。

图 4-75 汇总表 3



查看汇总表，状态为发布状态。

----结束

数据集市是面向数据应用为出发点，一个数据集市可以支持多个相关的数据应用。数据集市的作用与指标汇总层类似，但服务的目标系统只是指标汇总层的一部分，因此集市比指标汇总层更面向应用和用户。

集市层与指标汇总层类似，也是采用维度建模方式，常见的有星座模型、雪花模型等，只需要按实际应用需要设计、开发即可。

步骤1 设计步骤

图 4-76 设计步骤



应用需求分析：这部分主要针对数据应用系统所需要的数据范围及指标来确定，因此是先分析数据应用系统的需求，再提炼数据集市的数据接口需求

目标表设计：即数据集市结果表设计，根据数据应用的需求进行数据表设计

数据源分析：主要分析数据仓库的数据如何满足集市的需要，即确定需要数据仓库的那些表数据，如果数据仓库没有，则需要数据仓库进行数据补充

数据仓库表与目标表字段进行映射，确定加工规则

验证数据并调整加工规则

步骤2 实施步骤

根据设计好的目标表，创建DM层物理表

基于两张汇总表进行逻辑开发和字段映射

验证数据并配置调度作业，调度作业可以加在已有的作业中，也可以单独创建集市层作业，本次打样单独创建集市层作业。

----结束

4.5 数据开发

4.5.1 概述与连接管理

概述

数据开发包含脚本开发和作业调度，脚本开发主要是指SQL的ETL脚本，数据从贴源层到集市层的加工和流动，作业是对数据入湖、ETL脚本、数据质量等进行时间编排和按照计划周期性执行。

连接管理

登录管理中心 > 单击右侧数据连接 > 单击创建数据连接 > 选择对应的数据库类型 > 进行用户密码验证 > 单击测试通过

图 4-77 连接管理 1

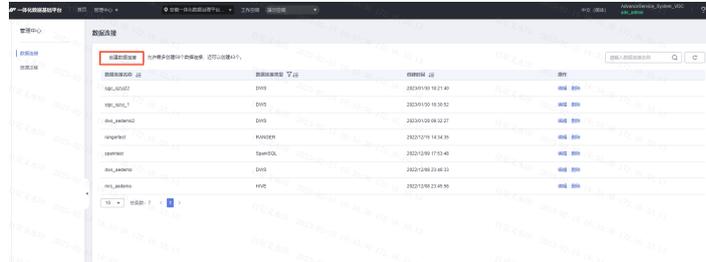
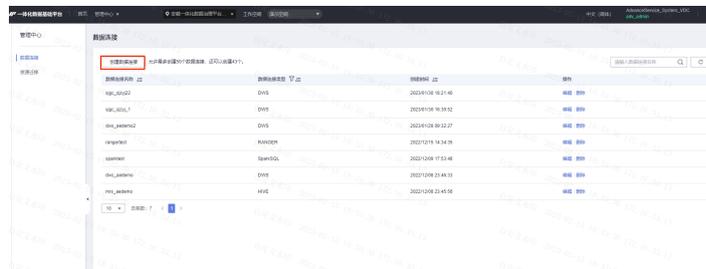


图 4-78 连接管理 2



4.5.2 脚本开发

数据连接创建成功后，进行脚本开发，按照数仓分层规范和《数据湖开发规范》进行关系建模开发和维度建模开发。

步骤1 创建spark sql脚本

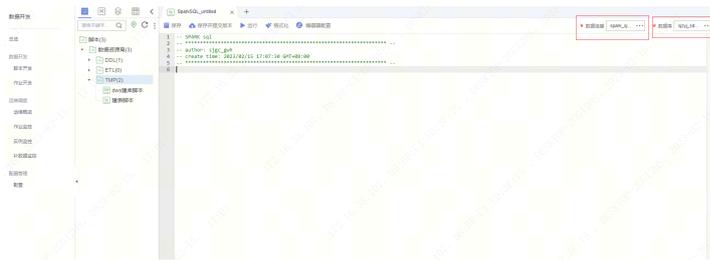
在“脚本开发”功能下，在对应的脚本文件夹下创建spark sql脚本。

图 4-79 脚本开发 1



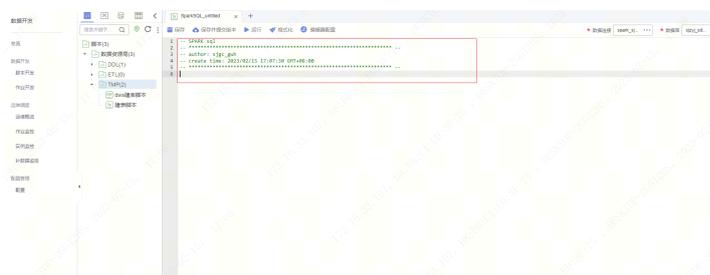
SQL脚本右上角配置数据连接和数据库

图 4-80 脚本开发 2



SQL脚本需要顶部添加作者和日期等脚本备注信息

图 4-81 脚本开发 3



脚本测试成功后，保存并提交版本，这样才能在调度作业中生效。

图 4-82 脚本开发 4



步骤2 创建dws sql脚本

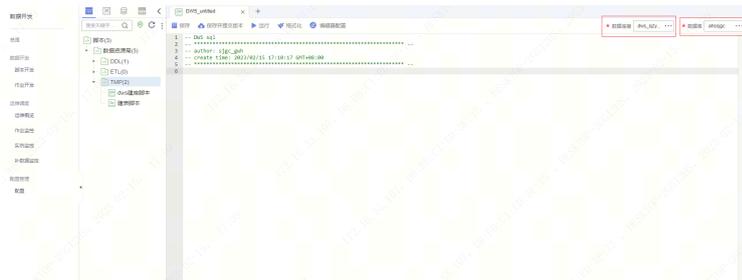
在“脚本开发”功能下，在对应的脚本文件夹下创建dws sql脚本。

图 4-83 创建 dws sql 脚本 1



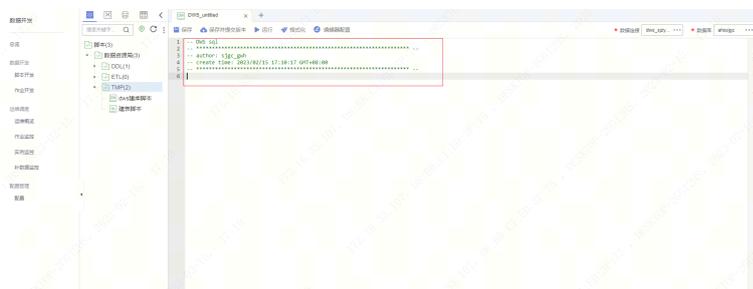
SQL脚本右上角配置数据连接和数据库

图 4-84 创建 dws sql 脚本 2



SQL脚本需要顶部添加作者和日期等脚本备注信息

图 4-85 创建 dws sql 脚本 3



脚本测试成功后，保存并提交版本，这样才能在调度作业中生效。

图 4-86 创建 dws sql 脚本 4



----结束

4.5.3 作业开发

作业是指将SQL脚本或者数据集成、数据质量作业等进行编排并进行周期性的调度。

步骤1 脚本作业配置

为了精细化管理和便于运维，ETL作业配置规则如下：

- 按逻辑数据实体创建调度作业画布。
业务指标用到的逻辑数据实体，将每个逻辑数据实体放在一个画布中，包括ODS层、DWI层、DWR层基础数据，然后按照ODS层作业 > DWI层作业 > DWR层基础数据作业顺序排列。
业务指标未用到的逻辑数据实体，将5个逻辑数据实体放在一个画布中，包括ODS层、DWI层，然后按照ODS层作业 > DWI层作业顺序排列，注意使用Dummy空算子进行作业间的关联。

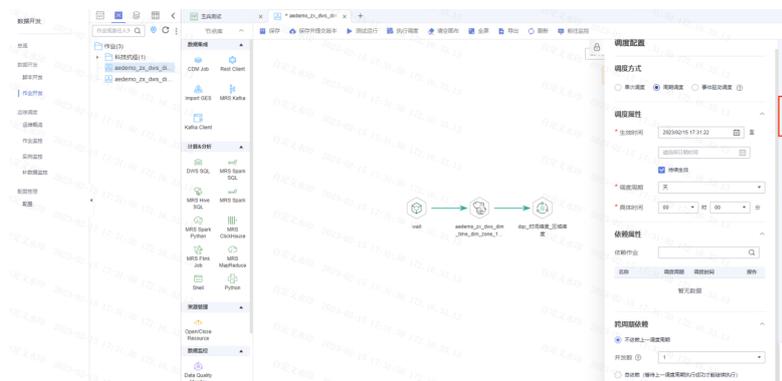
每个作业使用作业开发算子，每个算子关联对应的ETL脚本。
每个开发算子的失败策略为“终止后续节点执行计划”
建议每个画布中的作业算子不超过20个。

- 按事实表创建调度作业画布
按照DWR层事实表作业 > DWR层汇总表作业顺序排列
每个作业使用作业开发算子，每个算子关联对应的ETL脚本
每个开发算子的失败策略为“终止后续节点执行计划”
建议每个画布中的作业算子不超过20个
- 按质量作业创建调度作业画布
质量作业普遍执行耗时比较长，为了不影响整体的作业，故单独创建质量作业。
将20个质量作业放在一个画布上，使用Dummy空算子进行作业间的关联。
每个作业使用作业质量算子，每个算子关联对应的质量作业。
建议每个画布中的作业算子不超过20个。

步骤2 作业调度配置

步骤3 单击画布右侧“调度配置”，选择周期调度，并单击“执行调度”，调度会根据配置的调度时间定时调度。

图 4-87 调度



逻辑数据实体作业调度配置

事实表作业调度配置

事实表作业需配置依赖作业

失败策略为“挂起”

质量作业调度配置

质量作业需配置依赖作业

失败策略为“终止执行”

----结束

4.5.4 实例管理、补数据监控及配置环境变量

实例管理

作业手动调度或者周期性启动调度都会创建一个新的实例，在实例管理中可以查看每次运行的作业实例，可以查看日志，节点运行时间。

补数据监控

周期性的调度如果有执行失败，可以针对作业创建补数据任务，制定日期范围和作业的指定节点。

配置环境变量

配置作业参数，当某参数隶属于多个作业，可将此参数提取出来作为环境变量，环境变量支持导入和导出。

表 4-50 环境变量列表

序号	变量名	变量配置
1	ETL_TODAY	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,0),"yyyy-MM-dd") }</code>
2	ETL_YESTERDAY	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-1),"yyyy-MM-dd") }</code>
3	ETL_HOUR	<code>#{DateUtil.format (DateUtil.addHours (Job.planTime,0),"HH") }</code>
4	ETL_TODAY_MINUS7	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-7),"yyyy-MM-dd") }</code>
5	ETL_TODAY_MINUS8	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-8),"yyyy-MM-dd") }</code>
6	ETL_TODAY_MINUS1	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-1),"yyyy-MM-dd") }</code>
7	ETL_TODAY_MINUS2	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-2),"yyyy-MM-dd") }</code>
8	ETL_TODAY_MINUS180	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-180),"yyyyMMdd") }</code>
9	ETL_CURR_YEAR	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,0),"yyyy") }</code>
10	ETL_YEAR_ONEDAY	<code>#{DateUtil.getYear (DateUtil.addYears (Job.planTime,-2)) }-01-01</code>
11	ETL_YEAR_ONEDAY_NO_	<code>#{DateUtil.getYear (DateUtil.addYears (Job.planTime,-2)) }0101</code>

序号	变量名	变量配置
12	ETL_TODAY_NO_	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,0), "yyyyMMdd") }</code>
13	ETL_YESTERDAY_NO_	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-1), "yyyyMMdd") }</code>
14	ETL_TODAY_MINUS180_YES_	<code>#{DateUtil.format (DateUtil.addDays (Job.planTime,-180), "yyyy-MM-dd") }</code>

4.6 数据质量

数据质量监控

数据质量管理围绕着数据从产生到销毁的全生命周期，共分为事前防范、事中监控与事后治理三个阶段。事前防范即依照数据标准对业务源端数据产生的过程进行约束，规避问题数据的产生，同时对已发现的存量数据问题进行修正，从而规避由于数据质量问题导致的后续数据使用异常。在数据质量管理的实施过程中在推动业务侧做好事前防范的同时，需要重点关注数据质量的监控以及数据质量的整体提升。

数据质量监控模块是对数据库里的数据质量进行质量管理的工具。可从完整性、有效性、及时性、一致性、准确性、唯一性六个维度进行单列、跨列、跨行和跨表的分析。并能够根据数据标准自动生成质量作业。需要注意的是，数据质量的提升是一个持续的过程，需要贯穿在数据创建，转换和数据传输的整个生命周期。

- 规则模板

创建规则是数据质量的核心，数据质量支持对离线数据的监控，您可根据表的实际需要创建规则。目前创建规则的方式包括模板规则和自定义规则。

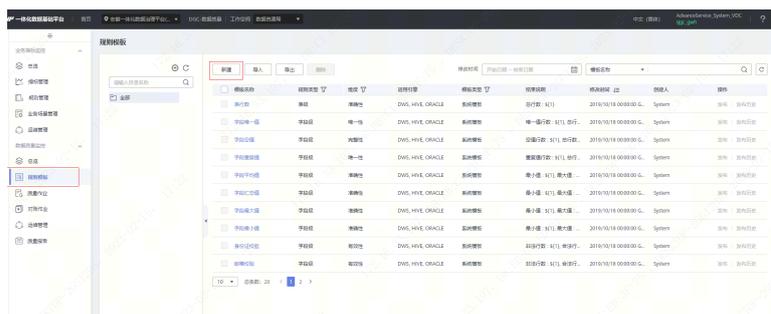
系统内置的模板规则共计20种，分为库级规则、表级规则、字段级规则和跨字段级规则四种类型，常见的有表行数、字段唯一值、字段空值、字段重复值、正则表达式校验、日期格式校验、字段一致性校验、合法性校验、枚举值校验等。当系统内置规则模板不足以满足您的需求，可通过自定义的方式新建规则模板。

新建规则模板

以常用的“长时间_XXX域格式校验”为例：

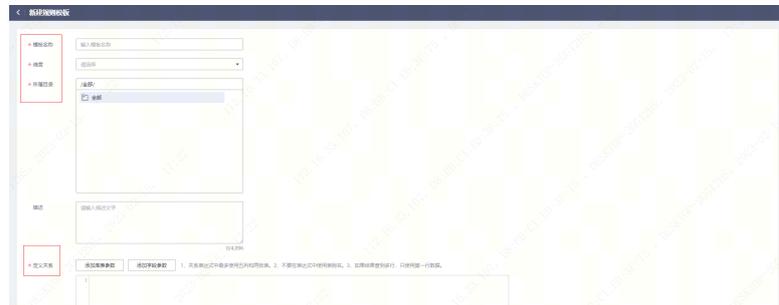
选择控制台 > 数据质量 > 规则模板 单击“新建”，在弹出的新建规则模板页面中进行配置

图 4-88 规则模板 1



在弹出的新建规则模板页面中输入规则名称，选择规则匹配的维度，定义SQL模板并对输出结果进行说明。

图 4-89 规则模板 2



维度：数据质量支持从完整性、有效性、及时性、一致性、准确性、唯一性六个维度进行单列、跨列、跨行和跨表的分析。自定义质量规则时，请对此规则进行维度匹配。

定义关系：输入SQL语句，实现对数据的查找。

样例1：统计表行数，输入SQL语句

```
select count ( ${Column1} ) from ${Schema_Table1}
```

其中\${Column1}通过单击“添加字段参数”生成，\${Schema_Table1}通过单击“添加库表参数”生成。

样例2：“长时间_XXX格式校验”，输入SQL语句

```
select ifnull ( b - a,0 ) ,ifnull ( a,0 ) ,ifnull ( b,0 ) ,ifnull ( a/b, 1 ) from ( select ( select count ( 1 ) as a from ${Schema_Table1} where ${Column1} regexp ' ( ^ ( ( 1[8-9][0-9]{2} ) | ( 2[0-9]{3} ) ) - ( ( ( 01|03|05|07|08|10|12 ) - ( ( 0[1-9] ) | ( [1-2][0-9] ) | ( 3[0-1] ) ) ) ) | ( 04|06|09|11 ) - ( ( 0[1-9] ) | ( [1-2][0-9] ) | ( 30 ) ) ) ) | ( ( 02 ) - ( ( 0[1-9] ) | ( [1-2][0-9] ) ) ) ) ) ( ( 0[0-9] ) | ( 1[0-9] ) | ( 2[0,1,2,3] ) ) : ( [0-5][0-9] ) : ( [0-5][0-9] ) $ ' ) a , ( select count ( 1 ) as b from ${Schema_Table1} ) b )
```

输出结果说明：对SQL获得结果的每一列进行说明，列说明之间用逗号进行分隔。

样例：“长时间_XXX格式校验”，输出结果说明为：

非法行数,合法行数,总行数,合法比率

结果说明应该与关系定义的输出结果顺序一一对应。

描述：用于通过文字阐述数据质量监控规则的依据及规则内容

样例：“长时间_XXX格式校验”，描述为：

用于数据标准【长时间_XXX】的格式校验，YYYY-MM-DD 时（2位）:分（2位）:秒（2位）

单击“确定”后，系统默认发布此规则模板，版本名称默认为V1.0。

图 4-90 规则模板 3

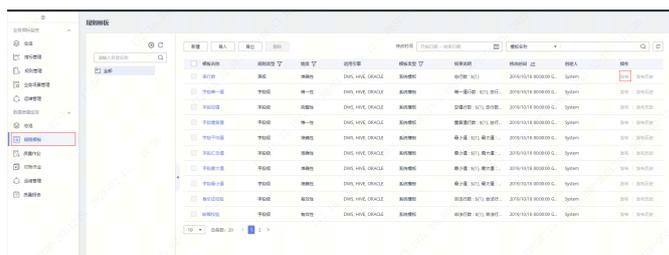


管理规则模板

自定义规则模板不支持直接修改已发布的历史版本。当您有修改需求，可以通过发布新版本以修改规则模板，并可以选择下线历史版本且将待下线历史版本关联的作业迁移到新版本上。具体请参见如下操作。

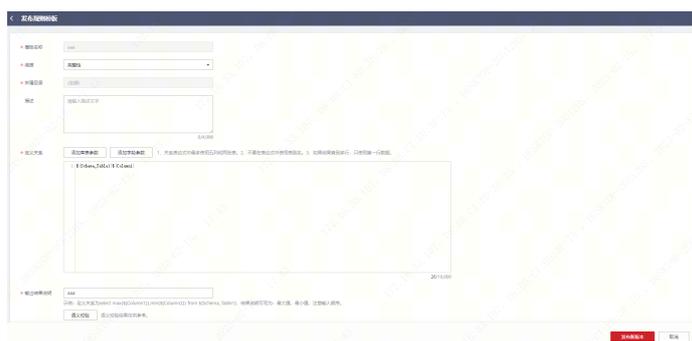
选择“数据质量监控 > 规则模板”，在规则模板列表中找到待修改的规则模板，单击操作列的“发布”。

图 4-91 规则模板 4



支持修改维度，修改输出结果说明和重新定义关系。

图 4-92 规则模板 5



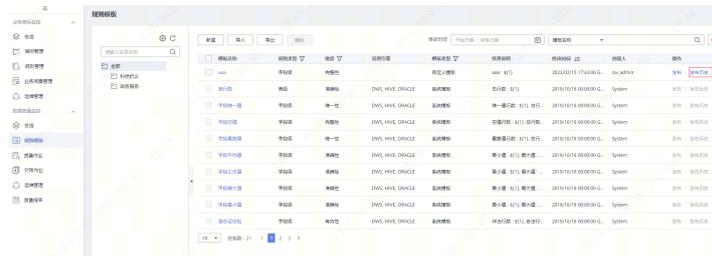
单击“发布新版本”，在提交发布对话框中，重新设置版本名称，并确认发布。

图 4-93 规则模板 6



提交发布后，单击操作列的“发布历史”，可以查看该规则模板的发布记录，支持查看变化信息、修改版本名称、下线对应版本等。

图 4-94 规则模板 7



如需下线历史版本，单击历史版本最右侧的“下线”按钮。
如果该版本没有关联作业，单击确认即可下线。
如果该版本存在关联作业，需要选择迁移版本，将新版本与作业关联后，单击确认才能完成下线。

图 4-95 规则模板 8

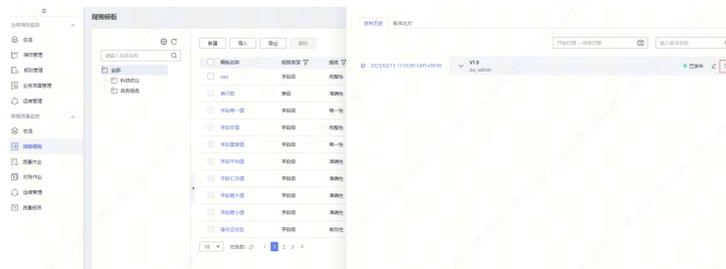
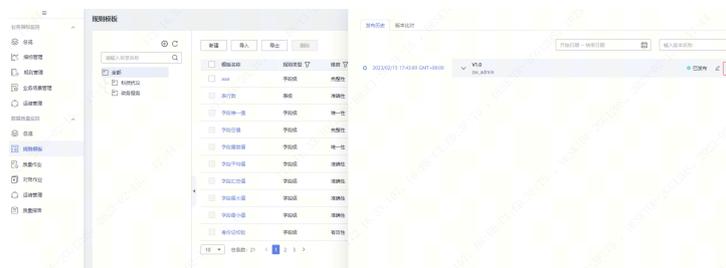
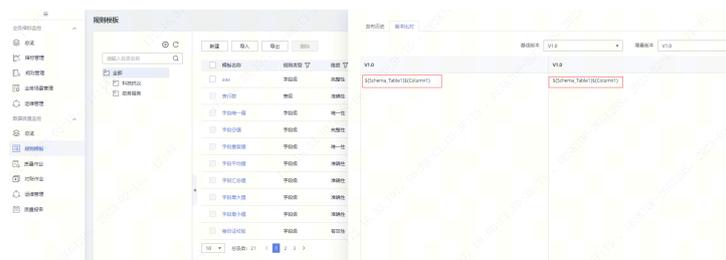


图 4-96 规则模板 9



发布历史处支持进行版本比对，直观展示修改点。

图 4-97 规则模板 10



- **质量作业**
在规范设计中，基于数据模型，提供了自动创建质量作业的能力，自动生成数据质量作业需要5步：

- a. 创建数据标准
- b. 在已经创建好的数据标准关联质量规则或者引用码表
- c. 新建模型
- d. 模型关联数据标准
- e. 模型发布

创建标准并关联质量作业

进入规范设计单击左侧“数据标准”菜单，进入数据标准界面，单击“新建”按钮，进入数据标准创建界面。

图 4-98 质量作业 1



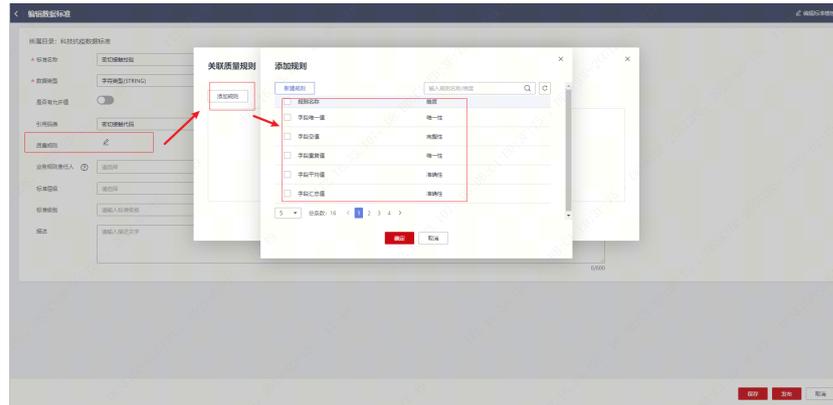
引用码表数据标准：填写标准名称、标准编码、数据类型、描述信息等，并选择要引用的码表及码表字段，完成配置后发布。

图 4-99 质量作业 2



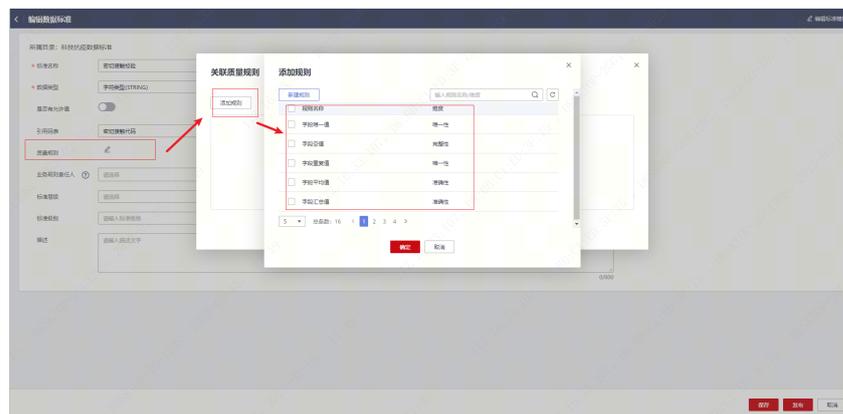
关联质量规则数据标准：填写标准名称、标准编码、数据类型、描述信息等，并配置质量规则，内置20个质量规则，包括行数、字段唯一值、空值、重复值、身份证、枚举值、邮箱、正则表达式等，完成配置后发布。

图 4-100 质量作业 3



关联质量规则数据标准：填写标准名称、标准编码、数据类型、描述信息等，并通过自定义正则表达式配置质量规则，完成配置后发布。

图 4-101 质量作业 4



模型关联质量规则并发布

如下在建模过程中，编辑时可以针对每一个字段关联数据标准从而生成。选择需要生成数据质量作业的逻辑模型，单击“编辑”按钮。

图 4-102 质量作业 5



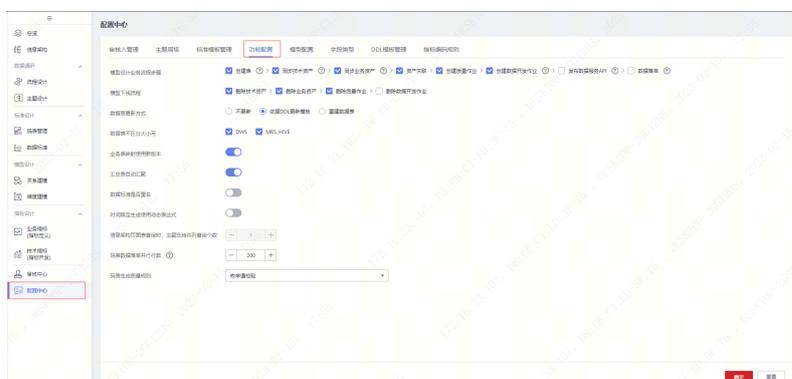
通过将数据模型中具体字段与数据标准已定义的数据质量规则、码表、枚举值、长度等内容，来基于统一数据标准监控表中的数据质量。

图 4-103 质量作业 6



发布模型，配置了质量规则后，需要发布模型才能生成质量作业。
首先，需要确保质量作业开关是开着的，在配置中心选择功能配置，点选创建质量作业并单击确定即可。
审核并发布后，即可自动生成质量作业：

图 4-104 质量作业 7



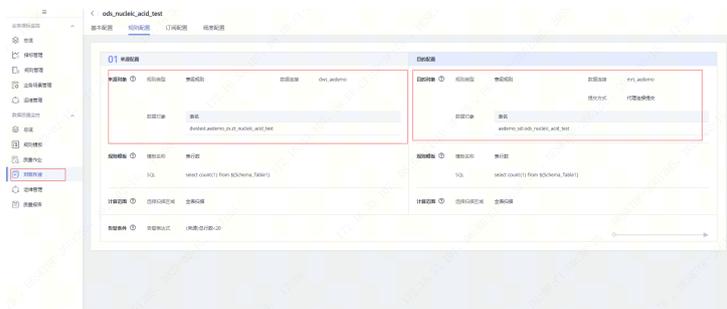
● 对账作业

数据对账对于数据开发和数据迁移流程中的数据一致性至关重要，而跨源数据对账的能力是检验数据迁移或数据加工前后是否一致的关键指标。数据质量监控中的对账作业支持跨源数据对账能力，可将创建的规则应用到两张表中进行质量监控，并输出对账结果。

对账作业的功能是检测两张表的一致性的，一般用源数据经过数据集成后进行一致性和完整性验证，用到最多的就是表行数和数据表空值扫描功能

创建对账作业如下图所示：主要功能是选择源库表和目的表进行对比，可以多表一起对比，计算范围建议第一次是全表扫描，后续是增量扫描。

图 4-105 对账作业



● 运维管理

用于查看质量作业的运行记录和运行结果，支持按时间和名称、运行状态筛选运行记录，其中运行结果链接可以查看规则的运行结果详细信息，如果出现异常还

可以通过查看执行日志查看异常原因，也可以在处理问题中转发给指定负责人进行处理。

图 4-106 运维管理 1

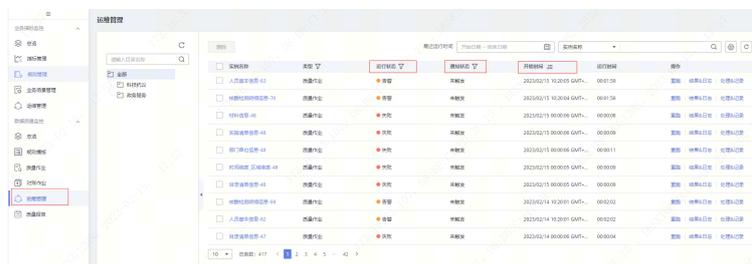


图 4-107 运维管理 2



- **质量报告**

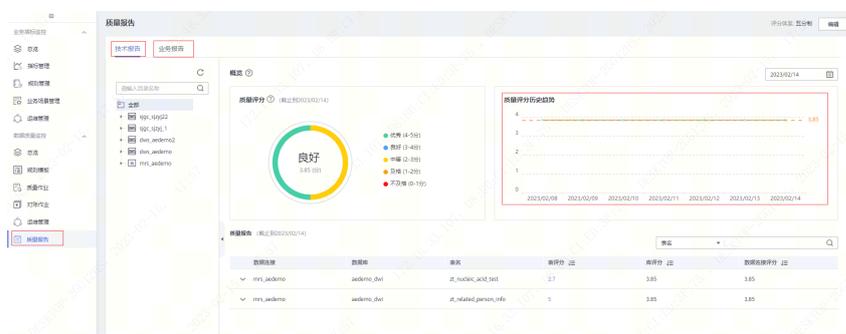
质量报告功能分为两个角度：业务质量报告和技术质量报告。

业务质量报告：是从规范设计中的业务分层角度统计不同分层的业务质量运行结果数据，并根据汇总数据结果，对不同业务分层进行质量打分，评估数据质量状况。

技术质量报告：是从数据本身的技术分层角度（即连接、库、表、字段）统计不同数据分层的数据质量运行结果数据，并根据统计数据结果，对不同的库、表进行质量评分。

对于以上两种不同的质量报告，均支持按天查看质量评分，查看最近7天的质量评分历史变化趋势，以及逐层下钻查看质量评分来源及关联的数据质量规则。

图 4-108 质量报告



数据质量提升是一个发现问题、分析问题、处理问题的过程，首先是基于质量作业的监控发现并跟踪数据质量问题，其次是对数据质量问题的异常情况与产生原因进行分析，待分析出结果后再制定相应的解决方案解决数据质量问题，从而达到数据质量逐步提升的最终效果。

数据质量问题分析：针对数据质量作业中发现的数据问题，通过数据质量规则，结合对异常数据的探查统计，分析问题类型与问题产生原因，用以制定问题解决方案。

数据质量问题处理：经由数据质量问题分析后，常见异常数据产生原因与处理方式如下：

表 4-51 常见异常数据产生原因与处理方式

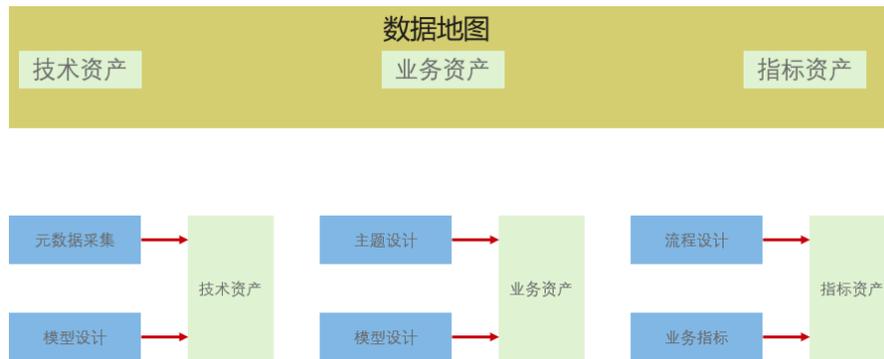
产生原因	问题举例	处理方式
数据标准定义有误	例如数据标准中“客户代码”并没有相应的编码或业务标准，无需进行格式、长度相关的质量检查	在质量作业中取消相应质量规则
数据标准关联有误	例如将数据标准“长时间_XX域”关联到实际数据内容为日期的字段，从而导致质检结果显示该字段数据内容全部异常	核实数据实际内容与业务含义，在质量作业中取消相应质量规则，或关联上相匹配的数据标准
质量规则有误	例如数据标准“XXX”定义数据为001~999，但在设计质量规则时将000视为正常数据	调整质量规则，使其与数据标准逻辑一致
数据系统建设阶段没有统一的格式标准，以及数据录入过程中未进行格式约束，更或是系统升级或切换过程中未对存量数据进行格式同化处理，导致数据质量较差	例如时间或日期数据存在上百种异常情况，在数据使用过程中，影响基于时间的范围限定以及基于时间的数据分析，无法有效的发挥数据价值	将数据异常情况反馈至业务部门，建议其对存留数据进行统一格式处理，并对录入过程进行格式约束在出现数据质量问题影响到数据使用，但数据部门尚无能力整改的情况下，可针对性的对数据进行清洗转换处理，待业务端完成整改后再取消清洗转换动作

4.7 数据资产

概述

数据统一资产视图包括技术资产，业务资产和指标资产。通过元数据采集，将各种数据源元数据信息采集到数据资产模块，作为技术资产呈现。同时支持采集规范设计创建的数据模型，作为业务资产呈现，并与技术资产自动关联。规范设计中创建的指标也会采集并作为指标资产呈现。

图 4-109 数据地图



元数据采集

元数据（Metadata）是关于数据的数据。在数据仓库系统中，元数据可以帮助数据仓库管理员和开发人员非常方便地找到目标数据，用于指导其进行数据管理和开发工作，提高工作效率。元数据管理模块是数据治理的基石，支持创建自定义策略的采集任务，采集数据源中的技术元数据、全链路的血缘管理。

- 采集任务

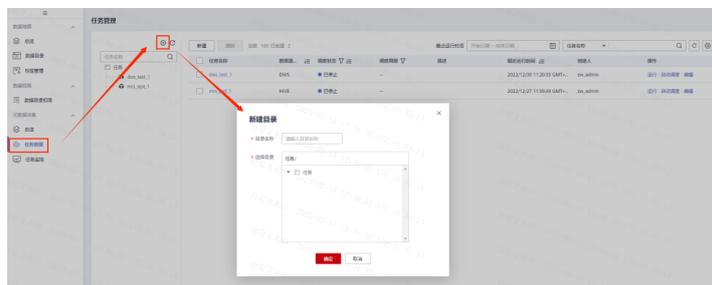
单击左侧导航树上“元数据采集>任务管理”，右侧区域显示任务管理界面。

图 4-110 采集任务 1



单击“新建目录”进入新建目录界面，输入相关参数值，完成目录创建。

图 4-111 采集任务 2



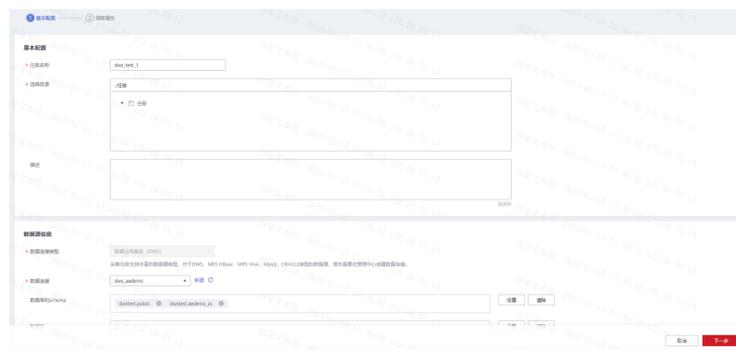
单击任务管理界面的“新建”按钮，进入“任务创建”界面。

图 4-112 采集任务 3



“任务创建”界面中填入连接类型、数据连接、数据库、表、采集策略等配置，单击“下一步”保存。

图 4-113 采集任务 4



任务创建完成后，“任务管理”界面中显示刚才创建的任务。

图 4-114 采集任务 5



单击操作栏中“运行”按钮运行任务元数据采集。

图 4-115 采集任务 6



DWI层、DWR层元数据采集，采用发布数据模型的时候采集的方式。
DM层元数据按需采集。

- **任务监控**
单击左侧导航树上“元数据采集>任务监控”，右侧区域显示任务监控界面，展示任务运行状态、日志等信息。

图 4-116 任务监控

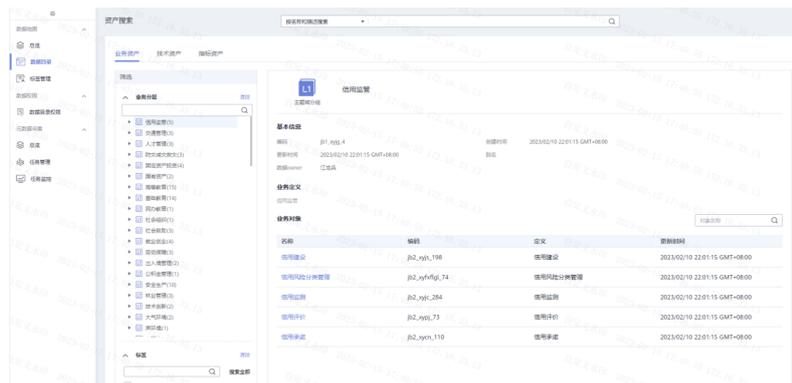


业务资产

规范设计中主题设计作为业务资产目录呈现，规范设计创建的数据模型作为业务资产呈现，支持发布模型时采集，并与技术资产自动关联。

单击左侧导航树上“数据地图>数据目录”，右侧区域单击“业务资产”并根据需要选择要查看的业务资产数据。

图 4-117 业务资产



技术资产

通过元数据采集，将各种数据源元数据信息采集到数据资产，作为技术资产呈现。

单击左侧导航树上“数据地图>数据目录”，右侧区域单击“技术资产”并根据需要筛选数据库连接、table、database等来查看技术资产数据。

通过执行调度作业，自动生成血缘关系图。

图 4-118 技术资产 1

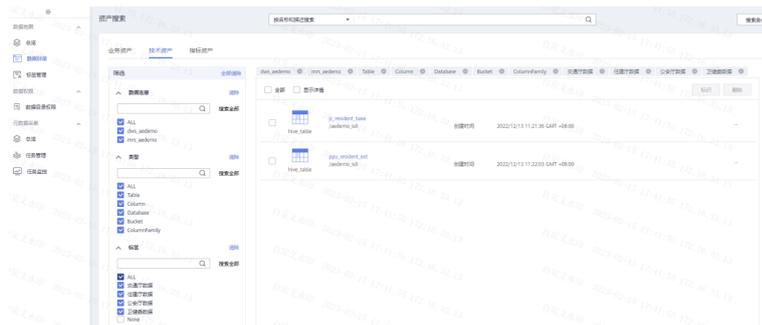


图 4-119 技术资产 2



规范设计中创建的指标采集并作为指标资产呈现。

单击左侧导航树上“数据地图>数据目录”，右侧区域单击“指标资产”并根据需要选择要查看的指标数据。

图 4-120 技术资产 3



4.8 数据服务

4.8.1 创建专享版集群

步骤1 创建数据服务专享集群增量包，系统会按照所选规格自动创建一个数据服务专享集群。单击工作空间“XXXX”的“数据服务”，单击左侧的“专享版”。

图 4-121 创建专享版集群 1

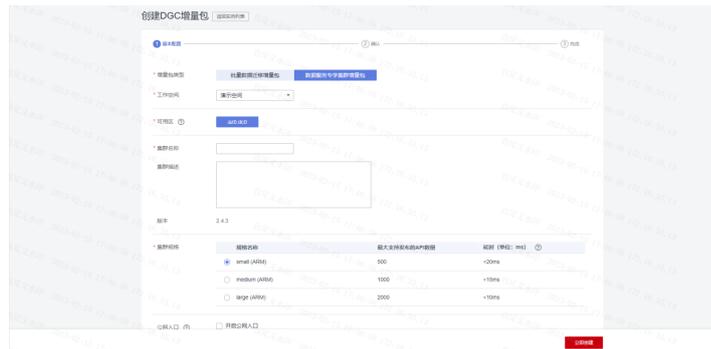


步骤2 单击左侧的“集群”，单击“新建集群”，进入创建增量包页面进行配置，单击“立即创建”，确认规格后提交。

图 4-122 创建专享版集群 2



图 4-123 创建专享版集群 3



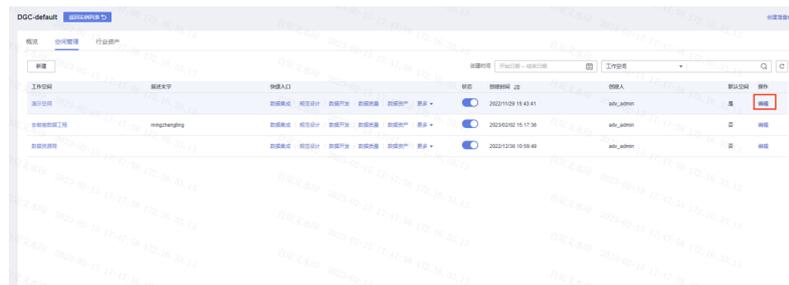
----结束

4.8.2 设置 API 分配配额

专享版集群创建成功后，需要设置API分配配额，当分配配额之后，才能创建相应的API，配额设置参考如下步骤。

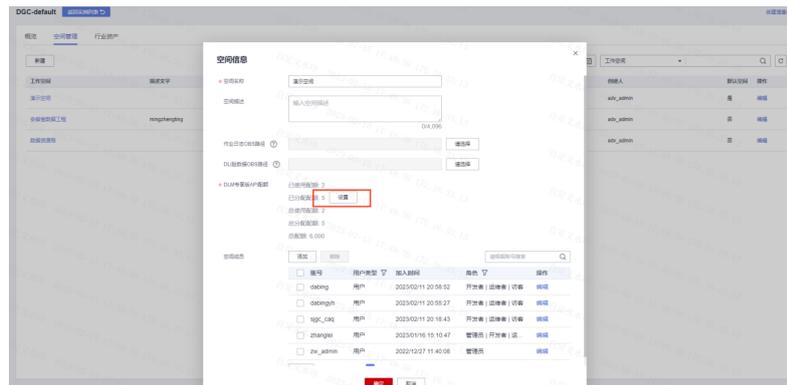
步骤1 在“空间管理”页签中，单击工作空间操作列“编辑”链接。

图 4-124 设置 API 分配配额 1



步骤2 在“空间信息”中，单击“设置”按钮对已分配配额进行配置。

图 4-125 设置 API 分配配额 2



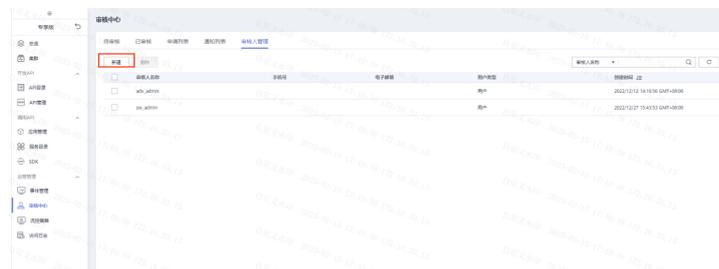
----结束

4.8.3 新建审核人

如果数据服务模块当前还未添加审核人，请先添加审核人，因为创建API需要提交审核人审核。只有工作空间管理员角色的用户才具有添加审核人的权限。

步骤1 单击工作空间“XXX”的“数据服务”，单击左侧的“专享版”。单击左侧导航栏中的“审核中心”，进入相应页面后，选择“审核人管理”页签，然后单击“新建”按钮。

图 4-126 新建审核人 1



步骤2 选择审核人（此处的账户列表来自于工作空间成员），单击“确认”完成审核人的添加。根据需要，可以添加多个审核人。

图 4-127 新建审核人 2



----结束

4.8.4 开发 API

- 创建API

- a. 进入“数据服务>API目录”页面，单击加号，输入新建API目录名称，可新建API目录。对应已成功创建的API目录的操作列，可重新编辑API目录或者管理API。

图 4-128 创建 API1



- b. 单击“编辑”，可修改API目录名称信息。仅当API处于已创建、已驳回、已下线、已停用的情况下才能进行API修改。

图 4-129 创建 API2



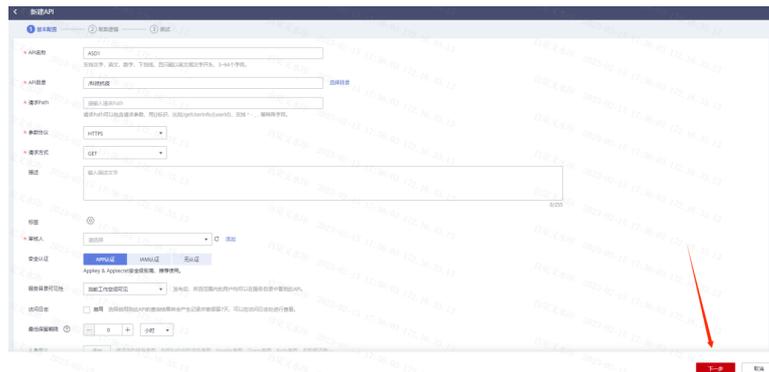
- c. 进入“API管理”页面，单击“新建”，填写API基本信息。

图 4-130 创建 API3



- d. 配置好API基本信息后，单击“下一步”，即可进入API取数逻辑页面。

图 4-131 创建 API4



- e. 选择数据源、数据连接、数据库和数据表，获取到需要配置的表。

图 4-132 创建 API5



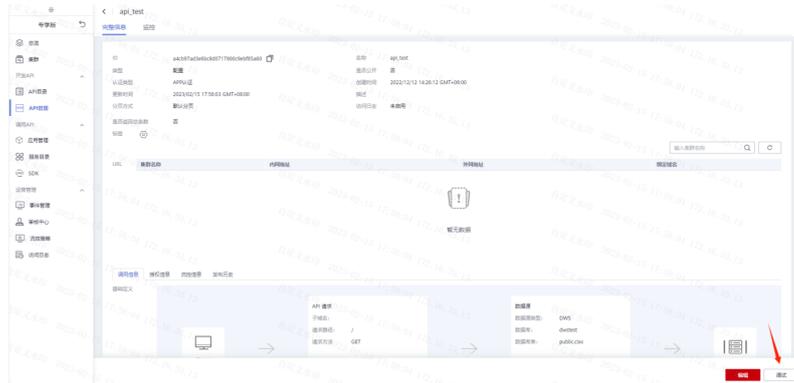
- f. 配置参数字段选择好数据表之后，单击“参数设置”后的“添加”，添加参数页面自动列出这个表的所有字段，分别勾选需要设置为请求参数、返回参数和排序参数的字段，分别添加到请求参数、返回参数和排序参数列表当中。
- 调试API
 - a. 单击“开发API > API管理”，进入到API管理信息页面。

图 4-133 调试 API1



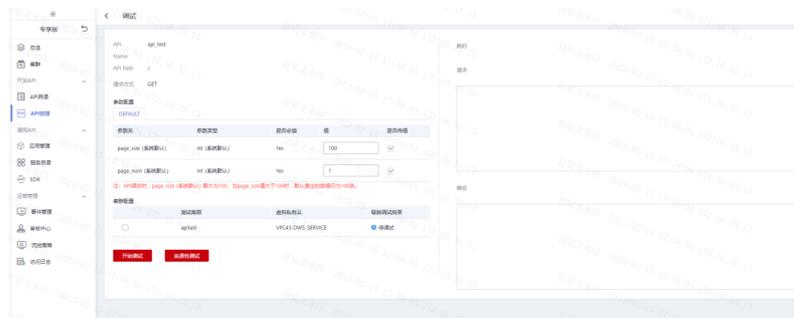
- b. 单击“API名称”，进入API详情页面，单击“调试”。或者在待调试的API所在行，单击“更多 > 调试”。

图 4-134 调试 API2



- c. 添加请求参数后，单击“开始测试”。右侧返回结果回显区域打印API调用的Response信息。调用成功时，返回HTTP状态码为“200”和Response信息。调试失败时，返回HTTP状态码为4xx或5xx。

图 4-135 调试 API3



- d. 可以通过调整请求参数与参数值，发送不同的请求，验证API服务。
- 发布API
进入“数据服务>总览>开发API>API管理”页面，在API服务列表操作列中，选择“更多>发布”。

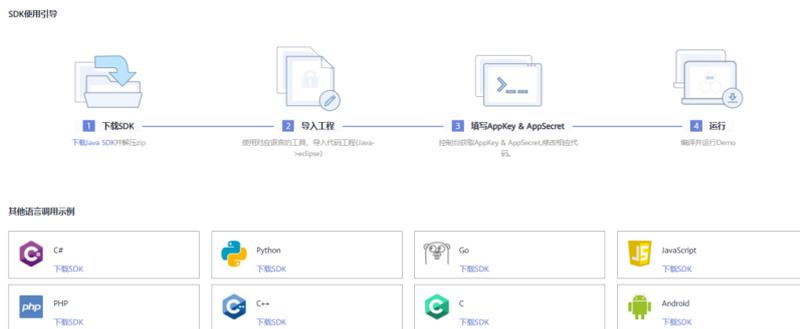
图 4-136 发布 API



4.8.5 调用 API

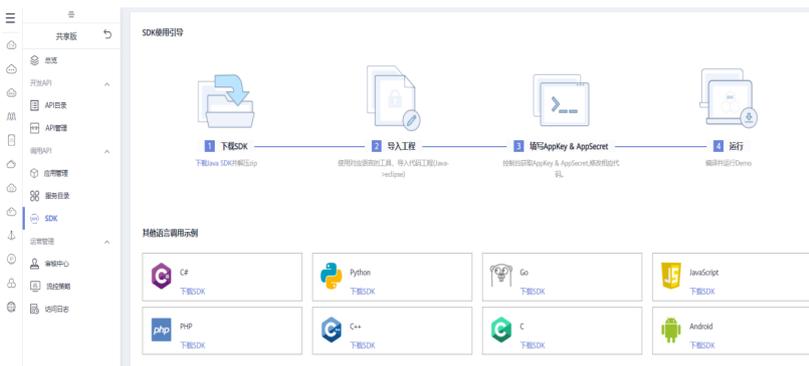
- 步骤1** 数据服务提供APP认证方式，API调用者通过APP认证方式调用API。使用APP认证时，需要通过SDK访问。数据服务提供了基于Java、Go、Python、JavaScript、C#、PHP、C++、C、Android等多种语言的SDK包。

图 4-137 调用 API1



- 步骤2** 访问服务前，首先需要得到API的ID、请求URL和请求方法，在数据服务的“服务市场”页面，单击API名称，在“完整信息”页面查看API的ID、请求URL和请求方法。
- 步骤3** 对于APP认证的API，必须提供有效的AppKey、AppSecret才能够生成认证签名，在“应用管理”中生成一个APP，并将APP绑定到API，就可以使用APP对应的AppKey和AppSecret访问该API。
- 步骤4** 从console页面下载相应语言的SDK。

图 4-138 调用 API2



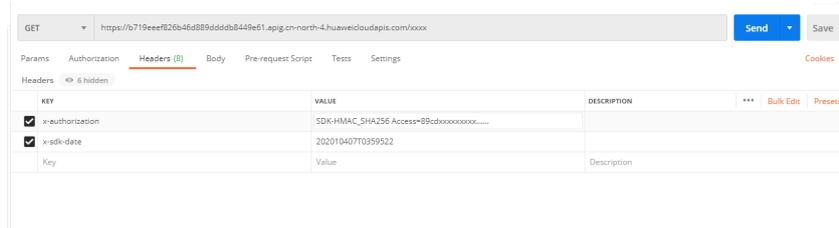
- 步骤5** 运行sdk获取签名。

图 4-139 调用 API3



- 步骤6** 将sdk获取到值x-authorization、x-sdk-date，填入到postman的header里，然后“send”获取相应的数据。

图 4-140 调用 API4



----结束

5 修订记录

表 5-1 修订记录

发布日期	修订记录
2024-05-06	规范词、敏感词专项处理，章节优化
2023-08-07	第一次正式发布。