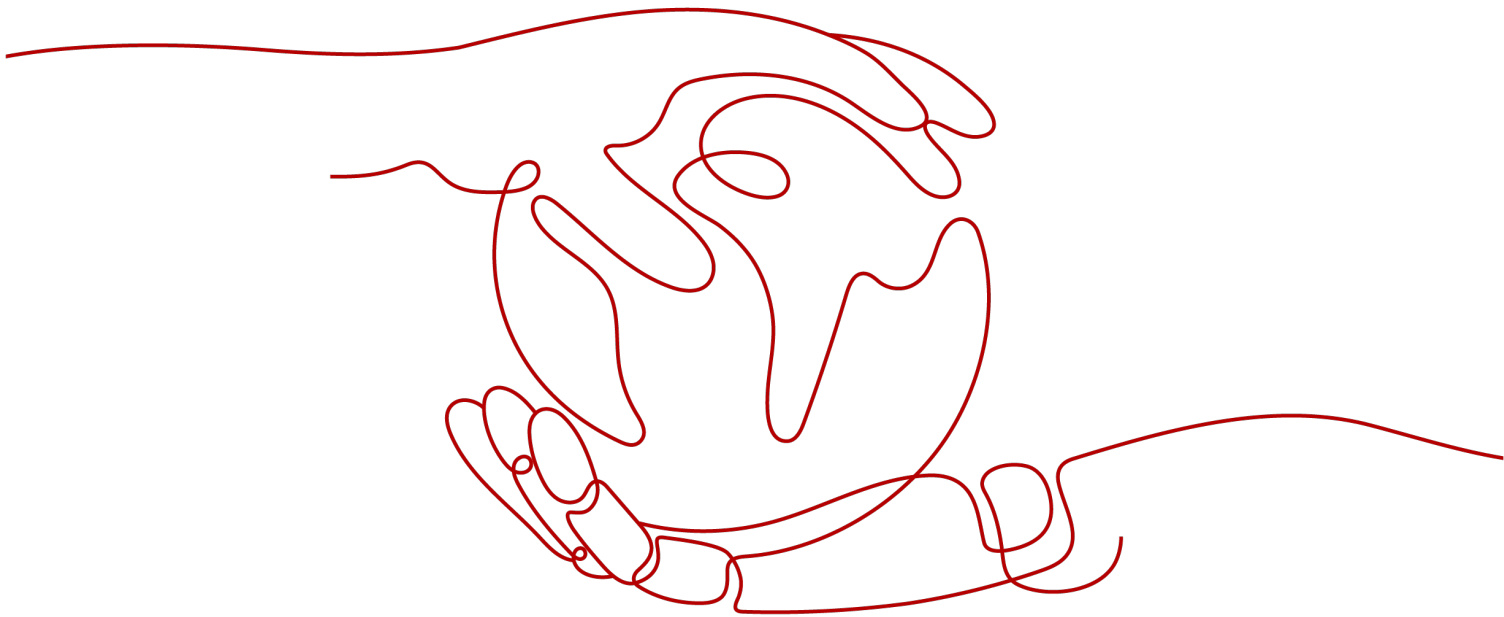


解决方案实践

八六三智能数据平台解决方案实践

文档版本 1.0
发布日期 2024-08-07



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.....	4
3 实施步骤.....	6
4 修订记录.....	17

1 方案概述

应用场景

在全球技术进步、市场环境变化以及政策引导下，产业数字化转型是必然趋势，但企业在数字化转型过程中存在诸多挑战：

- 数据整合与共享难题：在设备制造生产管控中，各部门之间往往存在数据壁垒，导致客户数据难以实现整合和共享。
- 数据不透明与不一致：在设备制造过程中，由于数据分散在多个系统和部门中，客户往往面临数据不透明和不一致的问题。
- 数据安全性与隐私保护：设备制造行业涉及大量敏感数据，如设备参数、客户信息等。由于数据治理缺失或不足，可能面临泄露和滥用的风险。
- 数据质量与可靠性问题：由于数据采集、存储、处理等环节的不规范，数据可能存在质量问题，如数据错误、缺失、重复等。

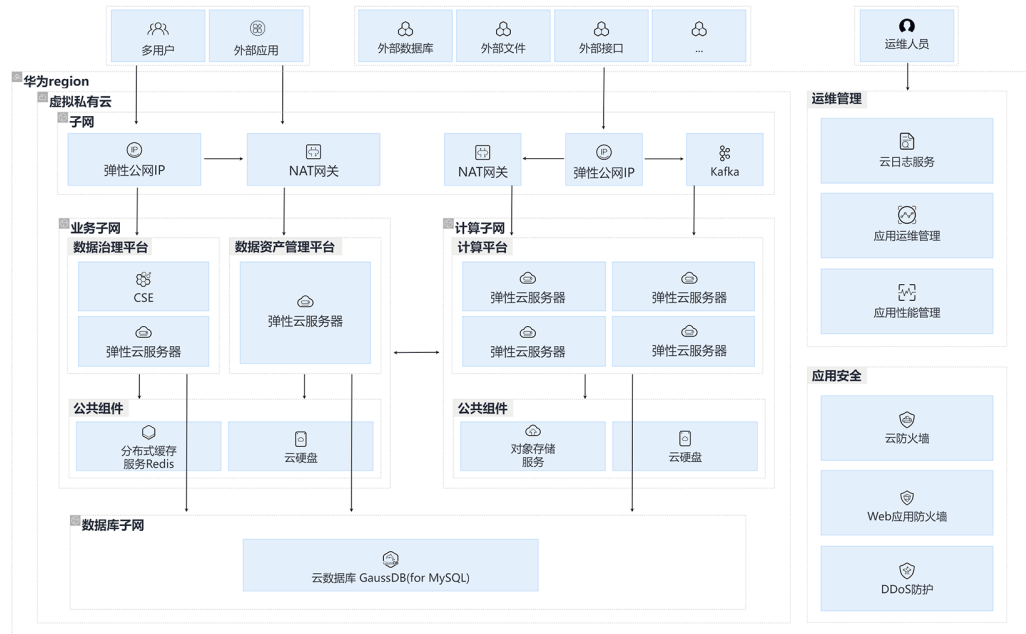
本章节介绍如何通过智能数据平台对企业数据资源进行全面管理和治理。以数据之治，铸就企业数字基石，助力企业数字化转型。

方案架构

图 1-1 业务架构



图 1-2 部署架构图



架构描述:

- 通过将整体服务部署在华为云，实现销售行为管理解决的整套方案；
- 通过ECS实现服务部署和运行，实现高性能计算；
- 通过VPC，EIP，NAT等部署简洁高效低成本的网络架构；
- 通过EVS，OBS等实现稳定可靠灵活的存储；
- 通过CSE、AOM、APM、LTS等实现服务管理&运维手段，保障服务稳定可靠运行。
- 通过云防火墙、web应用防火墙、Ddos防护等网络安全能力，提供云上互联网边界和VPC边界的防护，抵御外部Ddos等安全攻击，保障数据资产安全平台的安全运行。

方案优势

竞争力:

- **全面性:** 涵盖数据全生命周期各环节，支持各种数据源的接入，能够满足客户提供一站式数据治理服务；
- **合规性:** 内置行业模板，保障数据治理过程符合法律法规要求。
- **可拓展性:** 模块化设计，可根据企业需求进行灵活配置以及拓展功能；结合华为云服务，可快速进行适配和部署；
- **易用性:** 直观操作界面和友好用户体验，可快速上手，降低学习和使用成本；
- **安全可靠:** 结合华为云计算、存储、网络以及LTS/APM运维能力，实现解决方案稳定、安全、灵活扩容的对外服务能力。借助华为云CSE、redis、kafka、GaussDB for MySQL的灵活弹性，保障数据治理流程中突发场景下的解决方案的可靠性和高效数据处理能力。

价值:

通过数据梳理、采集、存储、管理、应用帮助企业实现数据资产的最大化利用，提升数据质量和安全性，确保数据合规性，从而推动业务决策的科学化、标准化和智能化。

- **确保数据质量和准确性：**通过数据清洗、标准化、验证等手段，提高数据的准确性和一致性，减少因数据质量问题导致的决策失误。
- **强化数据安全和隐私保护：**采取先进的数据加密、访问控制等技术，确保数据在存储、传输和使用过程中的安全性，防止数据泄露和滥用。
- **优化数据访问与共享：**简化数据访问流程，提高数据共享效率，打破数据孤岛，实现数据资源的充分利用和协同工作。
- **提升数据价值与利用率：**通过数据挖掘、分析等技术，发现数据中的潜在价值，为业务创新和发展提供有力支持。

2 资源和成本规划

表 2-1 资源和成本规划

云资源	规格	数量	每年费用 (元)
NAT	中型	1	5,865.00
EIP	包年包月, 100M	1	77,150.00
ECS	节点资源规格: 32vCPU&64G内存&系统盘100G 通用计算增强型 c7.8xlarge.2 32vCPUs 64GiB	4	134,320.00
ECS	节点资源规格: 16vCPU&32G内存&系统盘60G 通用计算增强型 c7t.4xlarge.2 16vCPUs 32GiB	1	18,684.00
ECS	节点资源规格: 16vCPU&16G内存&系统盘60G 通用入门型 t6.4xlarge.1 16vCPUs 16GiB	1	6,720.00
OBS	10T存储包	1	11,529.00
EVS	通用型SSD100 GiB IOPS上限3,000 IOPS突发上限8,000	6	4,200.00
ELB(内网)	实例类型: 独享型 10M	1	13,151.00
Redis	8G 主备	1	6,353.00
Kafka	kafka.8u16g.cluster 单个代理TPS 150000 单个代理最大分区数 1000 单个代理消费组数 150 单个代理流量规格 (MB/s) 375 代理数量 3 单个代理存储空间 高IO 200G	1	78,900.00
GaussDB(for MySQL)	独享型 x86 8 vCPUs 32 GB 存储空间 300G	1	54,080.00
CSE	实例数 500	1	10,020.00

云资源	规格	数量	每年费用 (元)
以下为可选			
APM	企业版初级Agent包	1	700.00
CFW	标准版	1	30,500.00
WAF	独享模式 WI100 ¥13.10/小时	1	990.00
DDoS	免费版	1	0
总计: 453,161.00 (必选: 420,971.00 可选: 32,190.00)			

3 实施步骤

数据集成

- **数据源：**数据源管理主要进行数据源连接配置&管理。

图 3-1 连接池管理



序号	数据源名称	源类型	分组类型	参数	描述	创建时间	更新时间	操作
1		StarRocks	来源	查看				编辑 删除
2	Hive	Hive	应用	查看	Hive			编辑 删除
3		Hive	其他	查看				编辑 删除
4		MySQL	数仓	查看				编辑 删除
5		MySQL	其他	查看				编辑 删除

在连接池管理页面，可以新建、编辑、查询、查看数据源配置信息。

图 3-2 新建数据源

The screenshot shows a '新建数据源' (New Data Source) dialog box. It contains the following fields and controls:

- * 名称: Text input field.
- * 分组类型: Dropdown menu with '请选择' (Please select).
- * 用户名: Text input field.
- * 密码: Text input field.
- * 源类型: Dropdown menu with 'MySQL' selected.
- * IP地址: Text input field.
- * 端口: Text input field.
- * 库名: Text input field.
- url: Text input field.
- 描述: Text area with placeholder '请输入描述' (Please enter description) and a character count '0 / 300'.
- Buttons: '取消' (Cancel) and '确定' (Confirm).

在连接池管理页面，单击“新建”，可以新建数据源配置信息，内容包含数据源名称、分类类型、用户名、密码、源类型、IP、端口、库名、数据源描述等。

- **数据集：**用于数据的主题域和主题管理，并支持主题下创建数据存储的库表模型。

图 3-3 主题域管理

The screenshot shows the '主题域管理' (Theme Domain Management) interface. It includes a search bar with the following fields and controls:

- 主题域名称: Search input field.
- 开始日期: Date input field.
- 结束日期: Date input field.
- 查询: Search button.
- 重置: Reset button.
- Buttons: '+ 新建' (New), '删除' (Delete), and search icons.

Below the search bar is a table with the following columns:

序号	主题域名称	英文名称	描述	创建人	创建时间	更新时间	操作
1			针对近日来福...	admin			编辑 删除
2		SimpleAttendan...	随着企业的规模...	admin			编辑 删除

At the bottom right, there is a pagination control showing '1' and '10条/页' (10 items/page).

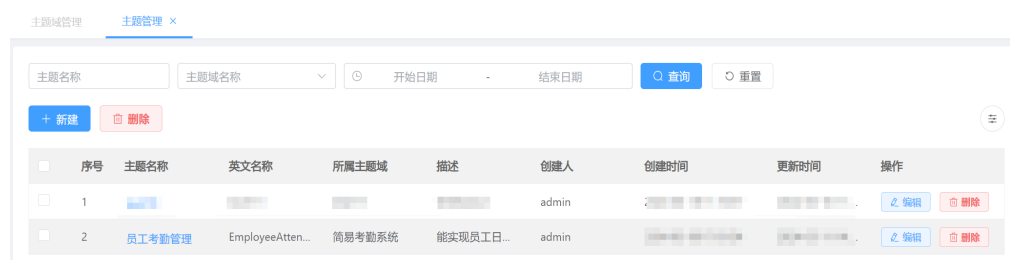
主题域管理，用于数据的主题域管理，支持新增、编辑、删除等操作。

图 3-4 新建主题域



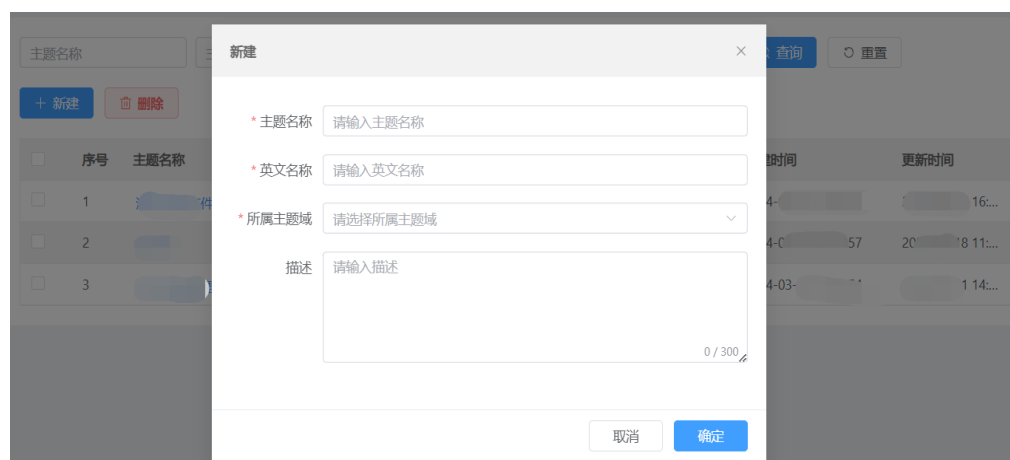
在主题域管理页面，单击“新建”，可以新建主题域信息。

图 3-5 主题管理



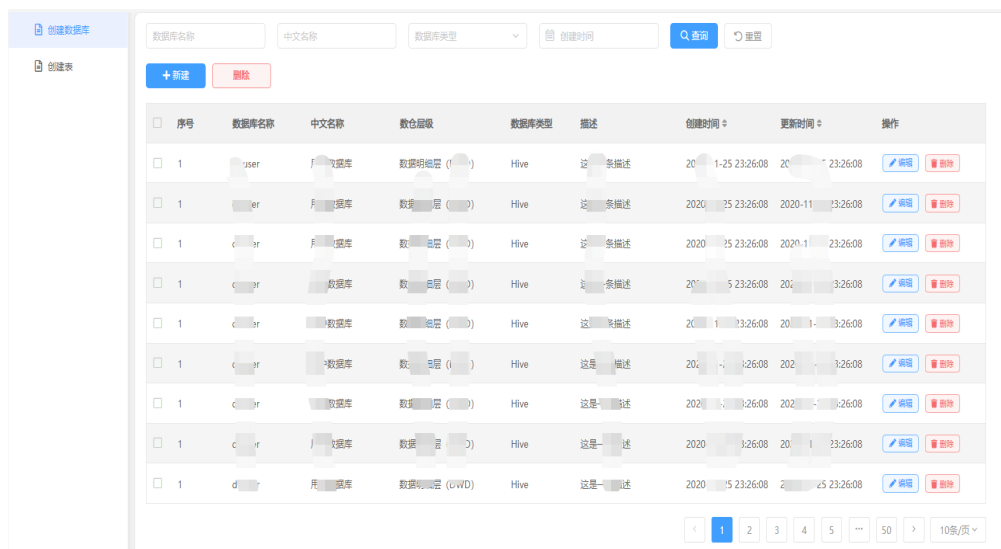
主题管理，用于数据的主题管理，可以新建、编辑、查询、查看主题信息。

图 3-6 新建主题



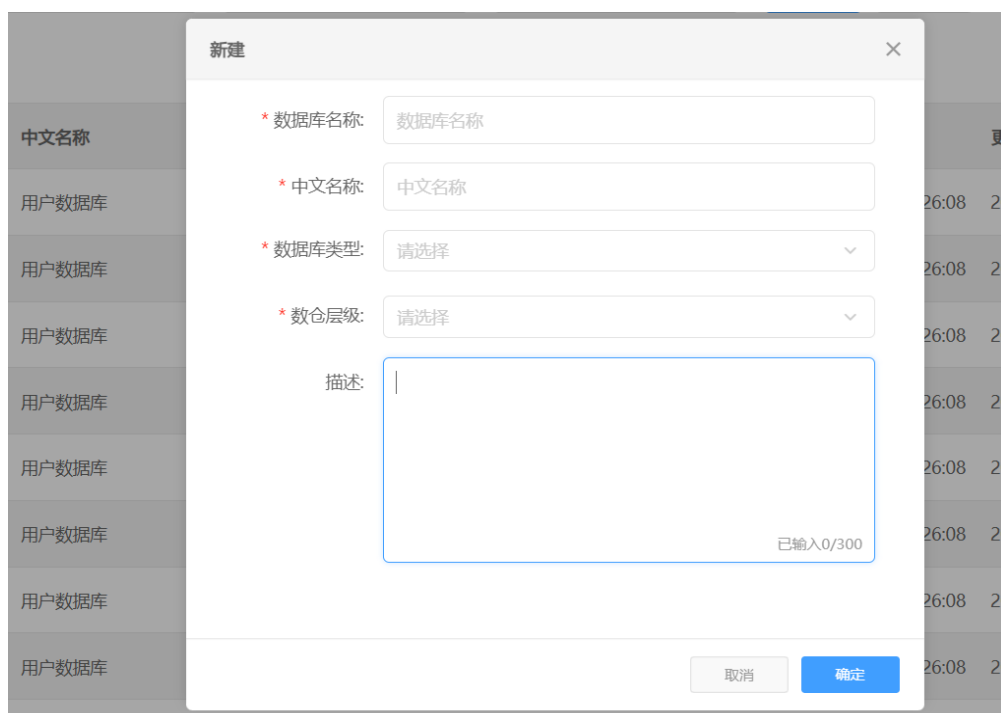
在主题管理页面，单击“新建”，可以新建主题信息。

图 3-7 数据库管理



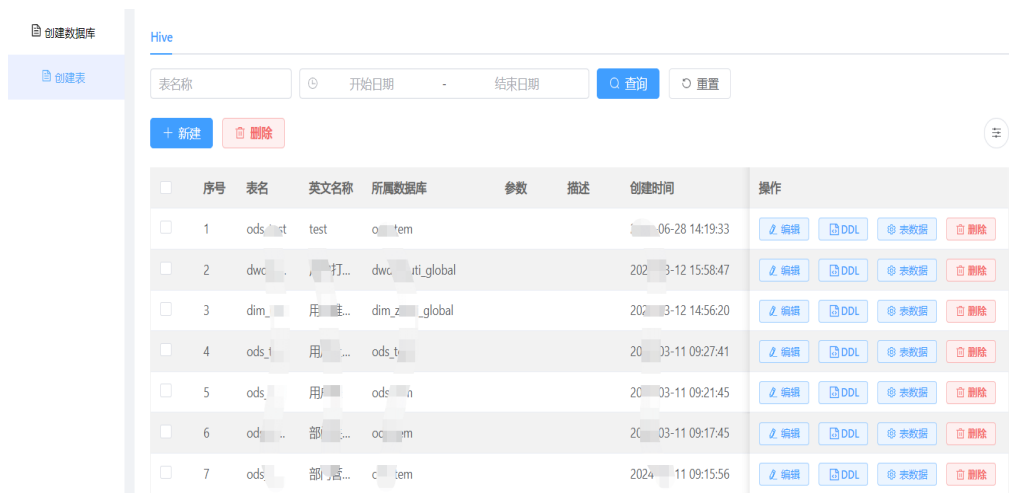
数据库管理，用于数据存储的数据库管理，支持新增、编辑、删除等操作。

图 3-8 新建数据库



在数据库管理页面，单击“新建”，可以新建数据库信息。

图 3-9 表管理



表管理，用于数据存储的表模型管理，支持新增、编辑、删除等操作。

图 3-10 新建表



在表管理页面，单击“新建”，可以新建表模型信息。

- **项目管理：**项目管理是数据处理集中管理的基本单元，完成具体数据处理任务的集合。该模块主要用于数据采集、处理的可视化编排和执行。

图 3-11 项目管理

序号	项目名称	所属用户	工作流定义数	正在运行的流...	描述	创建时间	更新时间	操作
1	数据	admin	0	0		2024-06-28 15:24:45	2024-	编辑 删除
2	人理	admin	3	0		2024-03-19 11:15:29	2024-	编辑 删除
3	人理	admin	7	0		2024-03-12 09:23:46	2024-	编辑 删除
4	数据治理	admin	2	0		2024-03-12 09:09:47	2024-	编辑 删除

项目管理，用于数据处理任务的项目管理，支持新增、编辑、删除等操作。

图 3-12 新建项目

创建项目

项目名称 *

所属用户 *

项目描述

取消 确定

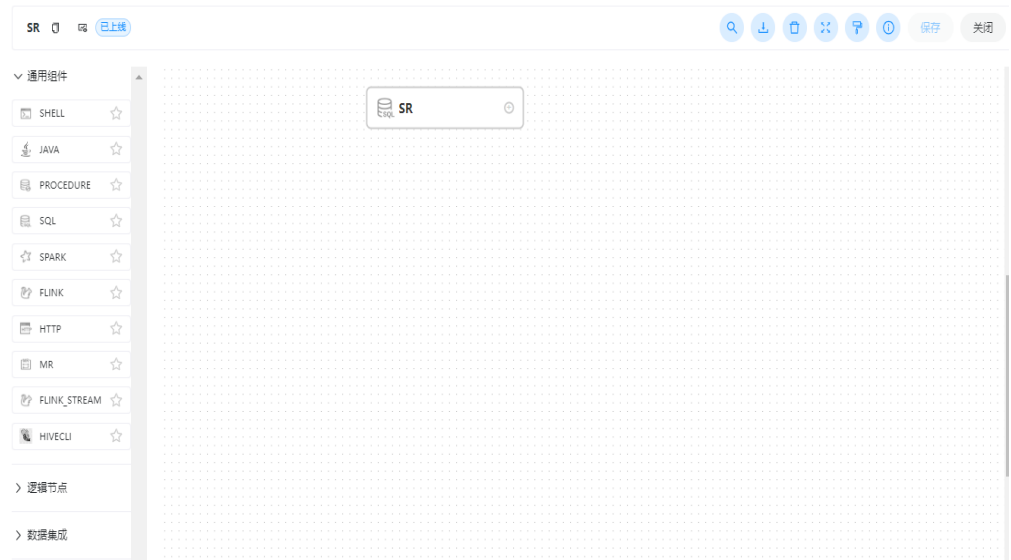
在项目管理页面，单击“新建”，可以新建项目信息。

图 3-13 workflow 管理

序号	工作流名称	状态	定时状态	创建时间	更新时间	操作
1		下线	下线	2024-06-24 16:02:34	2024-06-27 14:11	编辑 运行 定时 上线 更多
2		上线	上线	2024-04-07 16:21:45	2024-07-16:21	编辑 运行 定时 下线 更多
3	mysql-backup	上线	上线	2024-02-14 14:44:23	2024-03-16:11	编辑 运行 定时 下线 更多

工作流管理，用于数据采集、处理和流转的流程配置管理，支持新增、编辑、删除、上线、运行等操作。

图 3-14 新建 workflow



在工作流管理页面，单击“新建”，可以可视化配置 workflow 信息。

图 3-15 任务管理

批量任务 实时任务

任务名称 任务类型

序号	任务名称	工作流名称	工作流状态	任务类型	版本	上游任务	创建时间	操作
1	W_date Data	HIMS	已下线	SHELL	v1		2024-06-26 10:13:52	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>
2	ods_t_user_clock_in	用户打卡记录明细表 (D...	已上线	DEPENDENT	v5		2024-11-12 10:22:12	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>
3	ods_t_user_clock_in	用户打卡记录明细表 (D...	已上线	SHELL	v15	[1] ods_t_user_clock_in	2024-12-16 07:29	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>
4	ods_t_user	用户打卡记录明细表 (D...	已上线	SPARK	v1	[1] ods_t_user_clock_in	2024-10-07 15:58:51	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>
5	ods_t_user	用户打卡记录采集 (mysq...	已上线	SQL	v3	[1] ods_t_user	2024-10-18 16:01:40	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>
6	ods_t_user_clock_in_sum			SQL	v2		2024-10-18 15:41:20	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>
7	ods_t_user			SQL	v2		2024-10-18 15:47:09	<input type="button" value="编辑"/> <input type="button" value="版本"/> <input type="button" value="删除"/>

任务管理，用于数据处理任务的管理，支持新增、编辑、删除等操作。

图 3-16 新建任务

任务名称 *

请输入名称(必填)

任务类型 *

SHELL

工作流名称 *

请选择

运行标志 缓存执行

正常 禁止执行

描述

请输入描述

任务优先级 *

MEDIUM

Worker分组 * 环境名称

default 请选择

任务组名称 组内优先级

请选择 请输入 - +

在任务管理页面，单击“新建”，可以新建任务信息。

数据治理

- **元模型管理：**元数据存储库的数据模型，定义元数据的属性、关系的模型叫做元模型，每类元数据都对应一个元模型。

图 3-17 元模型管理

序号	元模型名称	描述	创建人	创建时间	更新时间	操作
1	字段元模型	这是对表字段属性信息的补充。	admin	2024-03-12 14:09:50	2024-03-28 09:24:01	编辑
2	表元模型	这是对数据表属性信息的补充。	admin	2024-03-12 14:09:05	2024-03-25 18:07:37	编辑
3	库元模型	这是对数据库属性信息的补充。	admin	2024-03-10 14:08:35	2024-03-25 17:32:03	编辑

在元模型管理页面，可以编辑维护库、表、字段元模型信息。

图 3-18 编辑元模型

元模型名称: 库元模型

描述: 这是对数据库属性信息的补充

基础属性

序号	字段名称	中文名	数据类型	小数位	必填	默认值	操作
1	res	正则	字符串		<input type="checkbox"/>		下移 删除
2	remarks	备注	字符串		<input type="checkbox"/>		上移 删除

业务属性

序号	字段名称	中文名	数据类型	小数位	必填	默认值	操作
1	name	名字	字符串		<input type="checkbox"/>		下移 删除
2	remarks	描述	字符串		<input type="checkbox"/>		上移 删除

[提交](#) [取消](#)

在元模型管理页面，单击“编辑”，可以编辑元模型信息。

- **元数据管理：**元数据管理主要用于组织、存储、维护和利用描述数据的数据（即元数据）。

图 3-19 元数据管理

元数据管理

请输入内容

[编辑](#)

基础信息

正则: [模糊]

备注: [模糊]

业务信息

名字: 业务数据

描述: 描述信息

技术信息

名称: baseinfo

中文名称: 基础信息表

创建时间: [模糊]

描述: [模糊]

在元数据管理页面，可以编辑维护库、表、字段元数据信息。

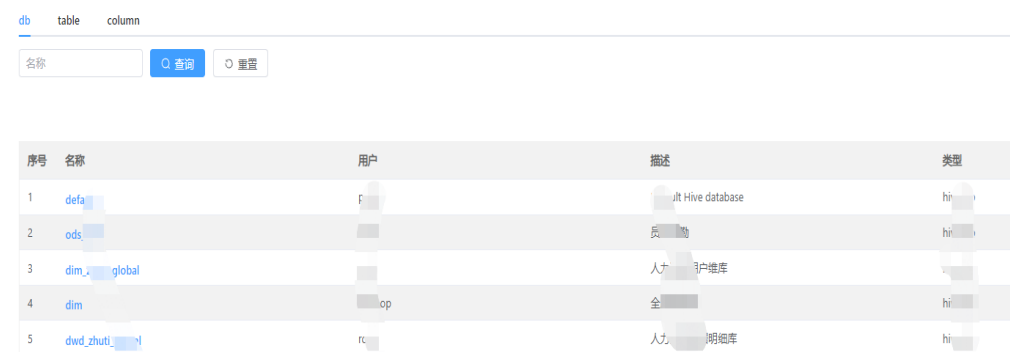
图 3-20 编辑元数据



在元数据管理页面，单击“编辑”，可以编辑元数据信息。

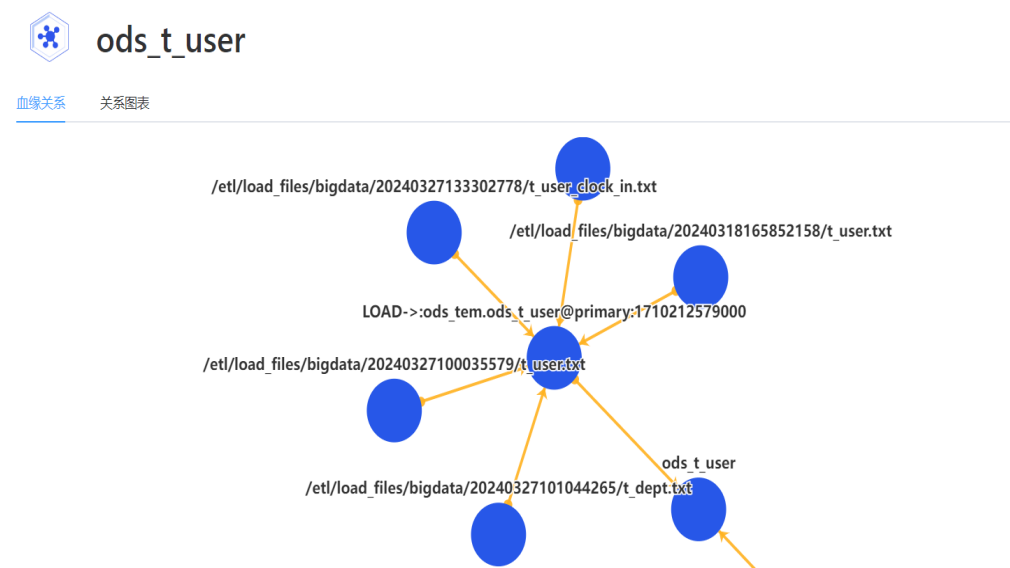
- **元数据分析：**元数据分析的主要目的是深入了解数据的各个方面，包括数据的结构、质量、来源、使用方式等，支持查看元数据血缘分析、影响分析等。

图 3-21 元数据分析管理



在元数据分析页面，支持查看库、表、字段的元数据分析。

图 3-22 血缘关系



血缘分析主要用于表示数据与数据之间的关系。主要包括表级血缘关系和字段血缘关系。

4 修订记录

表 4-1 修订记录

发布日期	修订记录
2024-08-07	第一次正式发布。