

数据设计开发实验指南

版本说明

实验背景

本实验将通过电商平台中常见的用户表和订单表，介绍如何在数开平台上完成从标准到建模到开发的整个开发治理流程，帮助用户更好的了解平台的使用。

实验目的

希望通过本次实验，您可以掌握：

1. 数开平台涉及与开发的流程；
2. 数据标准产品的使用方式；
3. 指标、模型、开发等产品模块的使用方式；
4. 各产品在使用过程中，相关权限的设置。

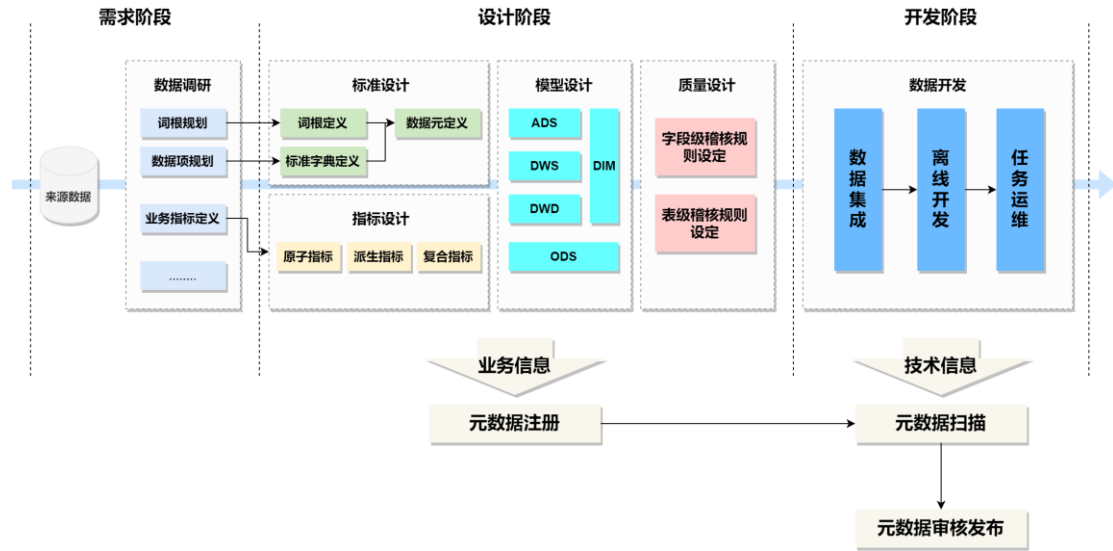
实验内容

1. 准备

1.1 流程图

数开平台遵循**先设计、后开发**的理念，因此在整个使用的过程中会先标准、指标、模型等相关的设计，设计完成后数据开发模块遵循设计的标准进行数据的集成、开发以及运维。

流程图如下：



本实验共分为三个阶段：**需求阶段**、**设计阶段**、**开发阶段**。需求阶段主要对业务数据进行调研、拆解数据、确定词根、数据项以及业务指标；设计阶段基于调研的内容进行标准和指标的设计并应用于模型和质量，设计完成后进行元数据的注册并完成业务信息的录入；开发阶段根据设计阶段的规范进行数据开发、约束开发流程，通过元数据扫描完成元数据技术信息的录入，最后将元数据进行审核并发布，完成整体业务的设计开发闭环，构建出基于**标准元数据**的数据体系。

1.2 实验数据

本实验将涉及两张表，分别为用户表和订单表，需事先在 MySQL 中进行创建（使用 MySQL 模拟外部数据源）。

用户信息表 (best_practice_user_table):

user_id	phone_number	gender	edu	email_address
2022010100001	15810001234	1	1	11111aa@163.com
2022010100002	15810001235	1	1	22111aa@163.com
2022010100003	15810001236	2	2	33111aa@163.com

2022010100004	15810001237	1	2	44111aa@163.com
2022010100005	15810001238	2	3	55111aa@163.com
2022010100006	15810001239	1	3	66111aa@163.com
2022010100007	15810001240	2	4	77111aa@163.com
2022010100008	15810001241	2	5	88111aa@163.com
2022010100009	15810001242	1	3	99111aa@163.com
2022010100010	15810001243	1	4	15111aa@163.com

SQL 代码:

```
CREATE TABLE `best_practice_user_table` (
  `user_id` varchar(64) COMMENT'用户 ID',
  `phone_number` bigint COMMENT'手机号',
  `gender` int COMMENT'性别',
  `edu` int COMMENT'教育程度',
  `email_address` varchar(64) COMMENT'邮箱地址'
)
```

数据链接: [best practice user table](#)

订单表 (best_practice_order_table):

order_id	user_id	pay_time	payment	pt_d
1234	2022010100001	2022-05-01 12:10:00	5332	2022-05-01
1235	2022010100002	2022-05-01	1179	2022-05-

		12:10:01		01
1236	2022010100003	2022-05-02 12:10:02	199	2022-05-02
1237	2022010100004	2022-05-02 12:10:03	599	2022-05-02
1238	2022010100005	2022-05-03 12:10:04	2144	2022-05-03
1239	2022010100006	2022-05-03 12:10:05	199	2022-05-03
1240	2022010100007	2022-05-04 12:10:06	1179	2022-05-04
1241	2022010100008	2022-05-05 12:10:07	998	2022-05-05
1242	2022010100009	2022-05-05 12:10:08	1179	2022-05-05
1243	2022010100010	2022-05-05 12:10:09	599	2022-05-05

SQL 代码:

```
CREATE TABLE `best_practice_order_table` (
  `order_id` bigint COMMENT '订单 ID',
  `user_id` varchar(64) COMMENT '用户 ID',
  `pay_time` TIMESTAMP COMMENT '付款时间',
  `payment` double COMMENT '实付金额',
  `pt_d` DATE COMMENT '付款日期, 可做分区'
```

)

数据链接: [best practice order table](#)

2. 需求阶段

2.1 标准规划

2.2.1 词根拆解

字段名称	字段中文名称	词根拆解
phone_number	手机号	手机, 号码
gender	性别	性别
edu	学历	学历
email_address	邮箱地址	邮箱, 地址

说明: 此处的词根拆解举例只是选取了比较有典型性的字段作为例子, 在实际业务过程中可将所有字段进行词根拆解作为标准化设计的准备, 当然在定义数据元的时候并不是一定要先做词根的定义, 可跳过词根这个阶段直接定义数据元。

2.2.2 数据项规划

字段名称	字段中文名称	数据项定义	是否需要标准
phone_number	手机号	手机号码	是
gender	性别	性别	是
edu	学历	学历	是
email_address	邮箱地址	邮箱地址	是

2.2 业务指标规划

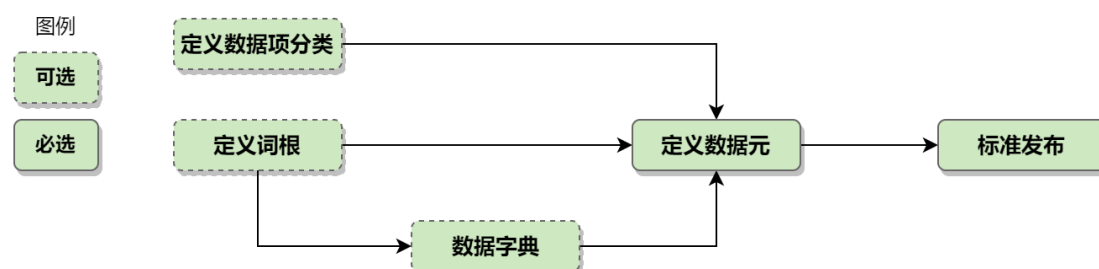
2.2.1 指标规划

字段名称	指标类型	指标名称
payment	原子指标	消费金额

3. 设计阶段

3.1 标准设计

标准定义流程：



角色分配：

角色	说明
数据标准制定人员	具有数据字典、数据目录、数据元、数据项分类、词根、标准发布功能权限的角色
数据标准审批人员	具有数据字典、数据目录、数据元、数据项分类、词根、标准发布、流程审核功能权限的角色

3.1.1 词根设计

根据 2.2.1 词根拆解结果，进行词根的设计，如下：

拆解后的词根	英文全称	英文简称	中文全称
手机	PHONE	PHONE	手机
号码	NUMBER	NO	号码，数字，编号

性别	GENDER	GNDR	性别
学历	EDUCATION	EDU	学历
邮箱	EMAIL	EMAIL	邮箱, 电子邮箱
地址	ADDRESS	ADD	地址

步骤一：在数据标准页面，点击词根进入词根页面，点击添加词根按钮



步骤二：根据表格内的设计完成词根的填写

以手机为例：

添加词根

X

* 英文简称 ⓘ	<input type="text" value="PHONE"/>
* 英文全称 ⓘ	<input type="text" value="PHONE"/>
* 中文全称 ⓘ	<input type="text" value="手机 x"/>
数据来源	<input type="text" value="请输入词根来源"/>
备注	<input type="text" value="请输入备注"/>
提交人	

取消

确定

完成后词根列表中呈现如下：

<input type="checkbox"/> 英文简称	英文全称	中文全称	词根编码 ⓘ
<input type="checkbox"/> ADD	ADDRESS	地址	ETM0015
<input type="checkbox"/> EMAIL	EMAIL	邮箱; 电子邮箱	ETM0014
<input type="checkbox"/> EDU	EDUCATION	学历	ETM0013
<input type="checkbox"/> GNDR	GENDER	性别	ETM0012
<input type="checkbox"/> NO	NUMBER	号码; 数字; 编号	ETM0011
<input type="checkbox"/> PHONE	PHONE	手机	ETM0010

3.1.2 数据字典设计

根据数据项规划，性别和学历需要定义数据标准，而性别和学历的数据值又是枚举类型，因此可以定义数据字典，如下：

字典名称	字典值	代码描述
性别	1	男
	2	女
学历	1	小学
	2	初中
	3	高中
	4	专科
	5	本科
	6	硕士
	7	博士

步骤一：在数据字典 - 标准字典页面，点击添加分类按钮，创建字典分类项



步骤二：点击分类名称右侧的快捷按钮选择添加字典



步骤三：根据字典的设计，完成数据字典的添加

以性别为例，配置完字典描述后，需要录入字典值，最后点击**保存**。

< 数据字典 / 添加字典

字典描述

* 字典名称 * 英文名称

英文全称 * 内部标识符

* 分类 * 类型

* 状态 * 版本

* 提交人

字典值 [批量导入添加](#) [添加字典项](#)

代码	代码描述	操作
1	男	编辑 删除
2	女	编辑 删除

共 2 条记录 < 1 > 25 条/页

学历：

字典描述

字典名称: 学历 英文名称: EDU 英文全称: EDUCATION

内部标识符: CD-S00000023 分类: 用户标准信息 类型: 标准字典

状态: 新增 版本: 1.0.0 提交人: 周正

字典值 批量导入添加 添加字典值

代码	代码描述	操作
1	小学	编辑 删除
2	初中	编辑 删除
3	高中	编辑 删除
4	专科	编辑 删除
5	本科	编辑 删除
6	硕士	编辑 删除
7	博士	编辑 删除

步骤四：将配置完成的数据字典进行提交审核，具有审核权限的用户在流程审核中进行审批

提交审核，目标状态选择**标准**：

数据字典 编辑 **提交审核** 删除 导出

原始字典 标准字典

输入中/英文...

字典值

提交审核

字典学历

当前状态【新增】，请选择审核的目标状态 **标准** ▾

进行审批：

数据标准 标准审核 发布审核

数据标准管理 未办结 已办结

标准类型: 全部 当前状态: 全部 目标状态: 全部 申请人: 全部

名称	唯一标识	标准类型	版本	当前状态	目标状态	申请人	申请时间	操作
<input checked="" type="checkbox"/> 性别	CD-S00000022	字典	1.0.0	新增	标准		2022-07-09 00:58:10	通过 拒绝
<input checked="" type="checkbox"/> 学历	CD-S00000023	字典	1.0.0	新增	标准		2022-07-09 00:58:05	通过 拒绝

3.1.3 数据元设计

根据 2.2.2 的数据项规划对数据元进行设计，如下：

中文名称	数据类型	数据格式	值域	自定义值域	安全等级
手机号码	字符型	n11	自定义	无	高
性别	字符型	n1	标准字典	性别	低
学历	字符型	n1	标准字典	学历	低
邮箱地址	字符型	c..20	自定义	无	高

说明：

1. 数据元的配置项可在标准配置里面去设置，此处以非必要最简配置的原则进行举例，实际应用中视情况而定。
2. 手机号码设定为字符型而非数字型的原因是脱敏功能当前只支持识别字符型，因此此处为后续实验做准备。
3. 数据元规则配置中的**数据安全要求**此处只配置安全等级，介绍脱敏时再进行介绍。

步骤一：在数据元页面，根据数据元设计添加数据元

以手机号码为例，属性填写页面：

标识类属性

* 中文名称	手机号码	* 中文全拼	shou-ji-hao-ma
* 英文名称	PHONE_NO	英文全称	PHONE_NUMBER
* 内部标识符	DE-00000072	* 版本	1.0.0

定义类属性

定义	数据元定义描述	对象类词	数据所描述的对象
特性词	描述对象属性名称		

表示类属性

表示词	请选择表示词	* 数据类型	字符型
* 数据格式	n11 修改	计量单位	数据元计量单位
* 值域	自定义	* 自定义值域	无

规则配置页面:

1 属性填写 ————— 2 规则配置

数据元信息

中文名称	手机号码	英文名称	PHONE_NO
内部标识符	DE-00000073	数据类型	字符型

质量稽核要求

来源依据	质量类型	稽核要求	建议规则
数据格式: n11	格式有效性	格式类型: 数字字符 (n) 长度类型: 固定长度 整数位数: 11 小数点: 无	通用型字符校验、长度校验、精度校验
补充格式校验	格式有效性	<input type="text" value="请输入检验要求"/>	<input type="text" value="填写正则表达式"/>

数据安全要求

敏感类型	* 安全等级	脱敏规则	脱敏算法	脱敏配置
<input type="text" value="请选择敏感类型"/>	<input type="text" value="高"/>	<input type="text" value="请选择脱敏规则"/>	-	-

步骤二：将添加完成的数据元，批量提交审核，并选择目标状态为标准



步骤三：在流程审核页面，完成审批



3.1.4 标准发布

步骤一：在标准发布页面，标准类型选择数据元，点击提交发布。



输入发布说明，提交发布。

发布内容：本次新增 4，修改 0，废止 0

标准类型：数据元

当前版本：1.0.4

* 发布版本：1.0.5

* 发布说明：新增4个数据元，分别为邮箱地址、学历、性别、手机号码

26 / 200

取消

确定

将标准类型切换至字典，同样点击提交发布。

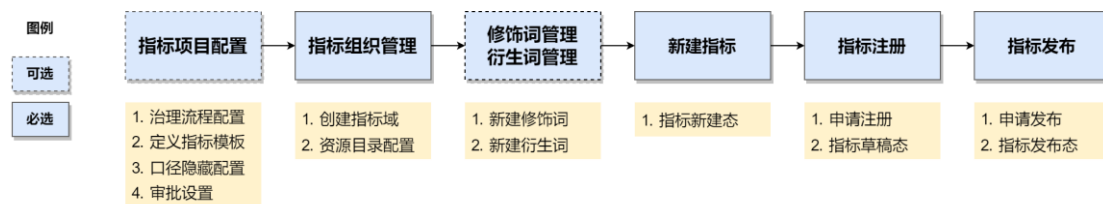
步骤二：在流程审核页面，切换至发布审核页签，进行审批。

数据标准	标准审核	发布审核					
	未办结	已办结					
数据项标准管理	标准类型：全部 申请人：全部						
数据元	批量拒绝 批量通过						
数据字典	标准类型：全部 申请人：全部						
词根							
数据项分类							
综合规范管理							
资源目录配置							
标准流程管理							
流程审核							
	<input checked="" type="checkbox"/>	标准类型	当前版本	发布版本	申请人	申请时间	操作
	<input checked="" type="checkbox"/>	字典	1.0.1	1.0.2		2022-07-09 01:44:23	通过 拒绝
	<input checked="" type="checkbox"/>	数据元	1.0.4	1.0.5		2022-07-09 01:43:31	通过 拒绝

审批通过后，标准设计这部分便完成了。

3.2 指标设计

指标定义流程：



关于指标整个治理的内容可查看：待补充

说明：

1. 指标项目配置是对整个指标的全局性配置，并非每次指标新建时都需要进行操作，因此这里不多做介绍，想了解该功能可查看产品用户手册。
2. 如果创建的是原子指标，则修饰词和衍生词不需要进行创建。
3. 理论上来说指标域也不是每一次新建指标时都需要进行的环节。
4. 此实验中关闭衍生指标，即不对原子指标做进一步区分，实际情况因业务而异。

指标内容规划：

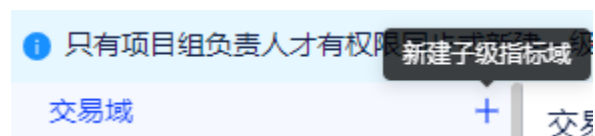
一级指标域	一级域标识	二级指标域	二级域标识	指标类型	指标名称	指标标识	业务口径
交易域	trade	订单	order	原子指标	消费金额	payment	消费金额

角色分配：

角色	说明
指标开发人员	具有指标字典、字段属性、修饰词管理员、衍生词管理员功能权限的角色
指标审批人员	指标域的负责人、指标审批者、数据治理管理员
数据治理专员	审核指标确权、治理工单的角色

步骤三：创建二级指标域

点击一级指标域右侧加号新建子级指标域。



步骤四：完成二级指标域的创建

新建子级主题域 X

* 域名称

* 域标识

* 域负责人 同父级主题负责人 自定义

* 指标审批者

平台审批

全部角色可选

3.2.2 创建指标

步骤一：在指标字典页面，点击新建指标按钮



步骤二：根据指标设计新建原子指标

< 指标字典 / 新建指标

基本信息

* 指标类型 原子指标 派生指标 复合指标

* 指标名称

* 指标标识

* 指标域

是否核心指标

可分析维度

安全级别

口径定义

* 业务口径

* 业务口径负责人

技术口径

技术口径负责人

保存

相似性检测

配置完成后建议做**相似性检测**，避免当前的指标名称、指标标识以及业务口径被重复构建，保证指标无二义性。

步骤三：注册指标

新建指标保存完毕后，将在指标列表中出现，此时指标处于**新建态**，需要通过申请注册变为**草稿态**。

#	指标名称	指标类型	指标标识	指标版本	指标状态	指标域	查看次数	操作
1	消费金额	主原子指标	payment	1.0.0	新建	交易域-订单	0	编辑 删除 流程操作 申请注册

申请注册需要域负责人、审批人或数据治理管理员进行审批。

< 待我处理 / 工单详情

指标注册- 工单编号: 483eaaa3-ff91-11ec-8a5f-6c92bf5cbad6

申请人信息

员工姓名: 申请日期: 2022-07-09

申请信息

元数据类型: 指标 指标编码: payment 指标名称: 消费金额 指标版本: 1.0.0 指标状态: 新建

系统提示: 审批人通过后, 指标将由新建态转为草稿态

步骤四：发布指标

注册完的指标需要发布才能被使用，因此需要申请发布。

#	指标名称	指标类型	指标标识	指标版本	指标状态	指标域	查看次数	操作
1	消费金额	主原子指标	payment	1.0.0	草稿	交易域-订单	0	编辑 删除 流程操作 申请确权 申请变更安全等级 申请治理 申请发布

申请发布需要域负责人、审批人或数据治理管理员进行审批。

待我处理 / 工单详情

指标发布-...-2022-07-09 工单编号: dc43f811-f93-11ec-b043-7e146b63aedd

申请人信息

员工姓名: ... 申请日期: 2022-07-09

申请信息

元数据类型: 指标 指标编码: payment 指标名称: 消费金额 指标版本: 1.0.0 指标状态: 草稿

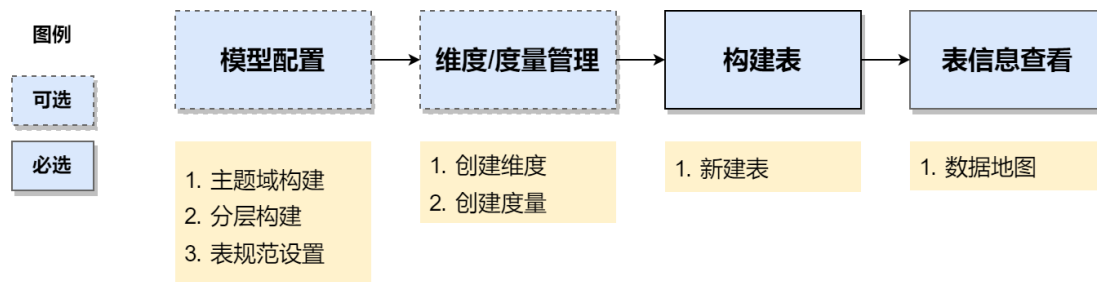
系统提示: 审批人通过后, 指标将由草稿态转为发布态

审批链路

指标发布后，指标设计的整个环节便也完成。

3.3 模型设计

模型构建流程图：



说明：

1. 模型配置一般会在项目初期由数据团队负责人进行统一的规范定义，因此并非每次建表时都要操作。
2. 本实验以 **ods 层和 dwd 层**表作为例子，因此在做表名的规范设计时只考虑该两层，且会使用无度量模式即 dwd 层可关联指标而非度量，实际情况根据客户使用习惯决定。

3. 这里的权限主要为数据权限，即有某一 hive 库的建表权限。

构建 ods 层及 dwd 层表信息如下：

序号	MySQL 中表名	Hive 中表名
1	best_practice_user_table	ods_主题域名称_usertable
2	best_practice_order_table	ods_主题域名称_ordertable
3	-	dwd_主题域名称_user_order

3.3.1 数仓规范定义

步骤一：设置主题域

主题域名称同指标域名称：

主题域名称	英文缩写	域负责人	域审批人	主题选择控制
交易域	trade	自行填写	自行填写	可不填

进入数仓配置页面，在主题域配置中点击新建主题域按钮进行创建。

新建主题域

* 主题域名称

* 英文缩写

* 域负责人

* 域审批人

主题选择控制  ! 当前主题域负责人默认可选择当前主题下的所有未级主题。

步骤二：设置子主题

在**主题域**页面，点击交易域名称旁的“+”创建子主题，子主题域名称同二级指标域。



子主题域名称	英文缩写	域负责人	域审批人	主题选择控制
订单	order	同父级主题负责人	同父级主题审批人	可不填

主题归属	交易域
* 主题名称	<input type="text" value="订单"/>
* 英文缩写	<input type="text" value="order"/>
主题负责人	<input checked="" type="radio"/> 同父级主题负责人 <input type="radio"/> 自定义
主题审批人	<input checked="" type="radio"/> 同父级主题审批人 <input type="radio"/> 自定义
主题选择控制 ?	<input type="text" value="点击设置"/>

步骤三：分层配置

由于 ods 和 dwd 都是预置分层，因此不需要创建，但是需要注意的是在**允许建表的库**一栏中要保证用户自己有建表的库在这个范围里，否则将无法在该层建表。

步骤四：表设计规则配置

此处主要是对表名规范的设置，ods 层表名的格式为 ods+主题域英文缩写+自定义表名，dwd 层表名的格式为 dwd+主题域英文缩写+自定义表名。

ods 表名规范设置如下：

- 1) 表名拼接规则为: f1_f2_f3_f4_f5_表名主体_s1_s2_s3。
- 2) 表名仅支持小写, 对于正则表达式请按小写设置。
- 3) 字典集仅支持选择“简单字典集”。
- 4) 常见的配置可参考 [配置建议](#)

前缀	前缀	控制逻辑	
f1	固定字符串	ods	X
f2	主题域缩写	表名会按照用户选择的主题自动替换	X

+ 添加

后缀	后缀	控制逻辑
		无数据

+ 添加

建表说明 用于展示给建表人员的说明性文字, 比如未采用基础字典配置的后缀枚举值说明、表名主体的命名建议等, 从而指导建表人员设置合理的表名

0/256

确定 取消

dwd 层表名规范设置如下:

- 1) 表名拼接规则为: f1_f2_f3_f4_f5_表名主体_s1_s2_s3。
- 2) 表名仅支持小写, 对于正则表达式请按小写设置。
- 3) 字典集仅支持选择“简单字典集”。
- 4) 常见的配置可参考 [配置建议](#)

前缀

前缀 控制逻辑

f1

固定字符串

dwd

X

f2

主题域缩写

表名会按照用户选择的主题自动替换

X

+ 添加

后缀

后缀 控制逻辑

无数据

+ 添加

建表说明

用于展示给建表人员的说明性文字, 比如未采用基础字典配置的后缀枚举值说明、表名主体的命名建议等, 从而指导建表人员设置合理的表名

0/256

确定

取消

此外, 还需要确认是否强制分区设置关闭 (本实验中希望自行选择是否配置分区)。

表分层	表名规则	表属性默认项配置	强制分区表
维表 - dim	表名拼接规则: [规则] 全量后缀 建表说明: - 编辑 删除	Hive MySQL 编辑	Hive <input checked="" type="checkbox"/> MySQL <input type="checkbox"/>
桥接层 - map	无 设置	Hive 编辑	Hive <input checked="" type="checkbox"/>
明细层 - dwd	表名拼接规则: dwd_一级主题域缩写_表名主体 建表说明: - 编辑 删除	Hive MySQL 编辑	Hive <input type="checkbox"/> MySQL <input type="checkbox"/>
汇总层 - dws	表名拼接规则: [规则] 建表说明: - 编辑 删除	Hive 编辑	Hive <input checked="" type="checkbox"/>
应用层 - ads	表名拼接规则: [规则] 建表说明: - 编辑 删除	Hive 编辑	Hive <input checked="" type="checkbox"/>
原始数据层 - ods	表名拼接规则: ods_一级主题域缩写_表名主体 建表说明: - 编辑 删除	Hive MySQL Vertica Greenplum Oracle 编辑	Hive <input type="checkbox"/> MySQL <input type="checkbox"/> Vertica <input type="checkbox"/> 更多

步骤五：开启无度量模式（可选）

如果已经开启无度量模式，那么在模型设计中心页面看不到度量。如果未开启，可在默认项配置页面打开。

模型设计中心
数仓配置

- 默认项配置
- 主题域配置
- 分层配置
- 字典集配置
- 扩展信息配置
- 表设计规则管理

请按照当前项目建设需要，设置默认的数据源。如有需要，可在【项目中心(新)】进行数据源登记后再选

Hive [规则] [规则] [规则] [规则] [规则]
MySQL [规则] [规则]
Vertica [规则]
Greenplum [规则]
Oracle [规则]

设置
设置
设置
设置

展开全部23个 设置

建模模式配置
无度量模式

3.3.2 表设计

由于 ods 表是直接来自 MySQL 同步至 Hive，因此表结构无变化。

ods_trade_usertable:

字段名称	主键	类型	数据元	数据字典	枚举值
user_id	是	string	-	-	
phone_no	否	string	手机号码	-	
gndr	否	string	性别	性别	1:男 2:女
edu	否	string	学历	学历	1:小学 2:初中 3:高中 4:大专 5:本科 6:硕士 7:博士
email_add	否	string	邮箱地址	-	

ods_trade_ordertable:

字段名称	主键	类型
order_id	是	string
user_id	否	string
pay_time	否	string
payment	否	string
ds (分区字段)	否	string

dwd 表为两张 ods 表做 join 操作所得，因此 dwd_trade_user_order 表结构如下：

字段名称	主键	类型	标准化 标签	数据元	数据字典	枚举值
------	----	----	-----------	-----	------	-----

order_id	是	string	-	-	-	-
user_id	否	string	-	-	-	-
phone_no	否	string	-	手机号码	-	-
gndr	否	string	-	性别	性别	1:男 2:女
edu	否	string	-	学历	学历	1:小学 2:初中 3:高中 4:大专 5:本科 6:硕士 7:博士
email_add	否	string	-	邮箱地址	-	-
pay_time	否	string	-	-	-	-
payment	否	string	指标	-	-	-
ds (分区字段)	否	string	-	-	-	-

步骤一：创建建表工单

在**与我相关**页面，点击**创建建表工单**按钮，进入新建页面。



步骤二：配置表名

在表名配置阶段，根据之前设计进行内容填写。

* 表分层 [查看分层和选库说明](#)

* 数据源类型

* 数据源

* 库

* 主题归属

表基本属性配置

您选择的表分层下设置了表设计规则，请按照规则设置表名及其他信息

表名拼接规则 ods_一级主题域缩写_表名主体

建表说明 无

表名配置

表名称 ods_trade_usertable

配置完成后点击下一步。

步骤三：选择表单模式，进入基础属性及字段配置页面进行配置

此处配置分几种情况：

第一种，如果是手动填入字段可直接输入。

第二种，如果是关联数据元，则选择数据元及数据字典，如手机号码。

字段配置

基础信息 — 标准化信息 — 扩展属性

#	* 字段名称	数据元	数据字典	枚举值	操作
1	<input type="text" value="user_id"/>	<input type="text"/>	<input type="text" value="请选择数据字典"/>	<input type="text" value="每行一个枚举值，参考格式:枚举值,说"/>	上移 下移 删除
2	<input type="text" value="phone_no"/>	<input type="text" value="手机号码"/>	<input type="text" value="请选择数据字典"/>	<input type="text" value="每行一个枚举值，参考格式:枚举值,说"/>	上移 下移 删除

字段的信息会根据数据元自动填入。

字段配置

基础信息 — 标准化信息 — 扩展属性

#	* 字段名称	数据元	数据字典	枚举值	操作
1	user_id		请选择数据字典	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除
2	phone_no	手机号码	请选择数据字典	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除
3	gndr	性别	性别	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除
4	edu	学历	学历	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除
5	email_add	邮箱地址	请选择数据字典	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除

此处还可以对有枚举值的字段进行补充说明，将枚举项填入枚举值这列中。

#	* 字段名称	数据字典	枚举值	操作
1	user_id	请选择数据字典	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除
2	phone_no	码	每行一个枚举值, 参考格式'枚举值.说明'	上移 下移 删除
3	gndr	性别	1.男 2.女	上移 下移 删除
4	edu	学历	1.小学 2.初中 3.高中	上移 下移 删除

第三种，如果是有标准化标签，比如说指标、度量、维度这种，那么可直接引用它们，因为 ods 不涉及这些，放在 dwd 中介绍。

步骤四：其它配置使用默认项，点击提交后进入工单审批页面，完成审批即创建完成

字段配置

基础信息 — 标准化信息 — 扩展属性

#	* 字段名称	数据元	数据字典	枚举值	操作
1	user_id		请选择数据字典	每行一个枚举值	上移 下移 删除
2	phone_no	手机号码	请选择数据字典	每行一个枚举值	上移 下移 删除
3	gndr	性别	性别	1.男 2.女	上移 下移 删除
4	edu	学历	学历	1.小学 2.初中 3.高中	上移 下移 删除
5	email_add	邮箱地址	请选择数据字典	每行一个枚举值	上移 下移 删除

+ 添加字段

基础信息 — 标准化信息 — 扩展属性

#	* 分区字段名称	* 类型	主键	描述	数据元	操作
---	----------	------	----	----	-----	----

上一步 提交 保存

审批建表工单：

模型设计中心 未提交 0 待审批 7 被驳回 2 待我审批 9 已通过 29 不合理命名表清单 批量导入 创建建表工单

表名称 请输入工单编号、表名称

工单编号	工单类型	工单操作对象	对象配置模式	工单状态	工单信息	操作
100004489	新建表	training1_ods_trade_user table Hive	表单模式	待审批	创建人： 表分层：ods-原数据层 主题归属：交易域订 审批人： 提交备注： -	通过 驳回

[与我相关](#)

步骤五：重复上述步骤完成表 ods_trade_ordertable 和表 dwd_trade_user_order 的创建

ods_trade_ordertable:

字段配置

基础信息 — 标准化信息 — 扩展属性

#	* 字段名称	* 类型	主键	描述	数据元	操作
1	order_id	string	<input checked="" type="checkbox"/>			上移 下移 删除
2	user_id	string	<input type="checkbox"/>			上移 下移 删除
3	pay_time	timestamp	<input type="checkbox"/>			上移 下移 删除
4	payment	string	<input type="checkbox"/>			上移 下移 删除

[+ 添加字段](#)

基础信息 — 标准化信息 — 扩展属性

#	* 分区字段名称	* 类型	主键	描述	数据元	操作
1	ds	string	<input type="checkbox"/>			上移 下移 删除

dwd_trade_user_order:

基础信息 — 标准化信息 — 扩展属性

#	* 字段名称	描述	标准化标签	数据元	数据字典
1	order_id		维度 指标 请先选择左侧类型		请选择数据字典
2	user_id		维度 指标 请先选择左侧类型		请选择数据字典
3	phone_no		维度 指标 请先选择左侧类型	手机号码	请选择数据字典
4	gndr		维度 指标 请先选择左侧类型	性别	性别
5	edu		维度 指标 请先选择左侧类型	学历	学历
6	email_add		维度 指标 请先选择左侧类型	邮箱地址	请选择数据字典
7	pay_time		维度 指标 请先选择左侧类型		请选择数据字典
8	payment	消费金额	维度 指标 消费金额		请选择数据字典

+ 添加字段

基础信息 — 标准化信息 — 扩展属性

#	* 分区字段名称	* 类型	主键	描述	标准化标签	数据字典
1	ds	string	<input type="checkbox"/>		维度 请先选择左侧类型	

需要注意的是 payment 这个字段可选择**标准化标签**中的指标。

8	payment	消费金额	维度 指标 消费金额			
---	---------	------	------------	--	--	--

步骤六：配置完成后，在主题域中可查看域下的三张表，点击表名可跳转至数据地图

查看表详情

模型设计中心 主题域(32)

交易域 - trade

数据源类型: Hive MySQL Vertica Greenplum Oracle

表分区: 全部 dim map dwd dws ads ods tmp

二级主题: 全部 订单

#	库表名称	负责人	表描述	关键信息	表分区	主题归属	操作
1	training1_dwd_trade_user_order			主键: 已配置 指标: 1个 维度: 0个	dwd	交易域/订单	查看核心属性 变更主题/分区 取消当前域关联
2	training1_ods_trade_ordertable			主键: 已配置	ods	交易域/订单	变更主题/分区 取消当前域关联
3	training1_ods_trade_usertable			主键: 已配置	ods	交易域/订单	变更主题/分区 取消当前域关联

ods_trade_usertable 表详情页面:

ods_trade_usertable

内部表 · 关联表 · 查看DDL · 查看Select语句 · 查看数据权限 · 自动分析 · 订阅通知

描述信息 明细信息 数据预览 产出信息 数据血缘 使用说明

字段信息 DDL变更 读写记录

主题: user_id

非分区字段

序号	字段名称	类型	描述	标准化标签	数据元	数据字典	操作
1	user_id	string	-	-	-	-	血缘 编辑
2	phone_no	string	-	-	手机号码	-	血缘 编辑
3	gnbr	string	-	-	性别	性别	血缘 编辑
4	edu	string	-	-	学历	学历	血缘 编辑
5	email_addr	string	-	-	邮箱地址	-	血缘 编辑

ods_trade_ordertable 表详情页面:

ods_trade_ordertable

内部表 · 关联表 · 查看DDL · 查看Select语句 · 查看数据权限 · 自动分析 · 订阅通知

描述信息 明细信息 数据预览 产出信息 数据血缘 使用说明

字段信息 分区信息 DDL变更 读写记录

主题: order_id

非分区字段

序号	字段名称	类型	描述	标准化标签	数据元	数据字典	操作
1	order_id	string	-	-	-	-	血缘 编辑
2	user_id	string	-	-	-	-	血缘 编辑
3	pay_time	string	-	-	-	-	血缘 编辑
4	payment	string	-	-	-	-	血缘 编辑

分区字段

序号	字段名称	类型	描述	标准化标签	数据元	数据字典	操作
1	ds	string	-	-	-	-	血缘 编辑

dwd_trade_user_order 表详情页面:

dwd_trade_user_order

内部表 · 关联表 · 查看DDL · 查看Select语句 · 查看数据权限 · 自动分析 · 订阅通知

描述信息 明细信息 数据预览 产出信息 数据血缘 使用说明

字段信息 分区信息 DDL变更 读写记录

主题: order_id

关联关联图标: 详情

非分区字段

序号	字段名称	类型	描述	标准化标签	数据元	数据字典	操作
1	order_id	string	-	-	-	-	血缘 编辑
2	user_id	string	-	-	-	-	血缘 编辑
3	phone_no	string	-	-	手机号码	-	血缘 编辑
4	gnbr	枚举	-	-	性别	性别	血缘 编辑
5	edu	枚举	-	-	学历	学历	血缘 编辑
6	email_addr	string	-	-	邮箱地址	-	血缘 编辑
7	pay_time	string	-	-	-	-	血缘 编辑
8	payment	string	消费金额	消费金额	消费金额	-	血缘 编辑

分区字段

序号	字段名称	类型	描述	标准化标签	数据元	数据字典	操作
1	ds	string	-	-	-	-	血缘 编辑

3.4 质量设计 (可选)

质量监控任务的配置在整个流程中较为灵活，可先与开发配置也可在开发过程中进行配置，本身又属于治理的环节所以配置的时机并不固定。

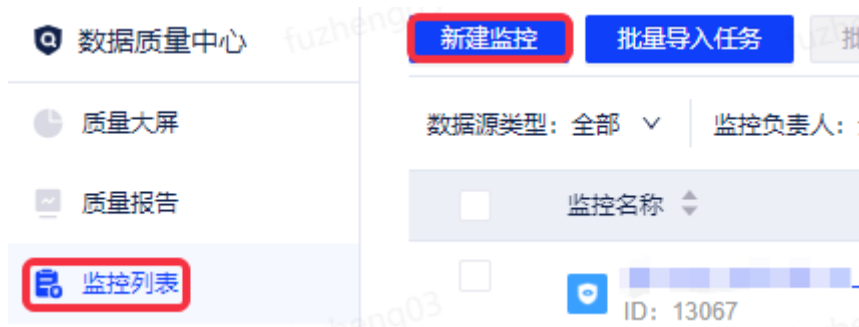
质量设计这个环节重点介绍下数据标准和质量稽核任务的联动，非数据元相关的字段不在

此处进行介绍，可根据实际需要自定义字段或表的质量规则。

此外，当前质量稽核任务的创建权限为所有用户因此权限这块不多做介绍。

3.4.1 质量稽核任务的创建

步骤一：在监控列表页面，点击新建监控按钮



步骤二：设置监控对象

以 ods_trade_usertable 为例，选择对应的库和表，监控名称自动生成。



如果有分区字段，可在分区筛选中选择分区字段进行设置，如：

▼ 监控对象

* 数据源类型 Hive

* 库 training1

* 表 ods_trade_ordertable

分区字段	分区值	是否监控
ds	\${azkaban.flow.1.days.ago}	<input checked="" type="checkbox"/>

▼ 基础信息

* 监控名称 dqc_training1-ods_trade_ordertable

描述 请输入监控描述

0/120

点击**下一步**，完成监控对象配置。

步骤三：设置规则

点击**新建规则**按钮，选择**标准推荐规则**。



选择各个字段进行规则配置，以 **phone_no** 为例：

新建标准推荐规则

字段

- phone_no string 手机号码
 - 格式有效性
 - 默认规则
- gndr string 性别
 - 格式有效性
 - 默认规则
 - 值域有效性
 - 默认规则
- edu string 学历
- email_add string 邮箱地址

phone_no_格式有效性_默认规则设置

规则类型: 字段级 表级(自定义SQL)

* 校验类型: 有效性校验

* 规则名称: phone_no_格式有效性_默认规则

* 规则代码: phone_no rlike "[0-9]{11}\$"

按逆向方式执行规则

采样方式	计算规则	基准值	期望范围
规则命中行数	默认	-	=
规则命中行数	与基准值计算比率	总行数	= 100%

规则控制: 强 弱

描述: 请输入描述

配置完成后，规则列表如下：

新建规则	模板规则(0)	自定义规则(0)	标准推荐规则(0)			
<input checked="" type="checkbox"/>						
规则控制/规则名称	规则类型	校验类型	监控字段	期望范围	状态	操作
<input checked="" type="checkbox"/> phone_no_格式有效性_默认规则	字段级	有效性校验	phone_no 查看代码	与基准值(总行数)计算比率: = 100%	<input checked="" type="checkbox"/> 已启用	编辑 删除 禁用
<input checked="" type="checkbox"/> gndr_格式有效性_默认规则	字段级	有效性校验	gndr 查看代码	与基准值(总行数)计算比率: = 100%	<input checked="" type="checkbox"/> 已启用	编辑 删除 禁用
<input checked="" type="checkbox"/> gndr_值域有效性_默认规则	字段级	有效性校验	gndr 查看代码	与基准值(总行数)计算比率: = 100%	<input checked="" type="checkbox"/> 已启用	编辑 删除 禁用
<input checked="" type="checkbox"/> edu_格式有效性_默认规则	字段级	有效性校验	edu 查看代码	与基准值(总行数)计算比率: = 100%	<input checked="" type="checkbox"/> 已启用	编辑 删除 禁用
<input checked="" type="checkbox"/> edu_值域有效性_默认规则	字段级	有效性校验	edu 查看代码	与基准值(总行数)计算比率: = 100%	<input checked="" type="checkbox"/> 已启用	编辑 删除 禁用
<input checked="" type="checkbox"/> email_add_格式有效性_默认规则	字段级	有效性校验	email_add 查看代码	与基准值(总行数)计算比率: = 100%	<input checked="" type="checkbox"/> 已启用	编辑 删除 禁用

步骤四：点击完成按钮，完成配置

这里需要说明下，因为现在没有数据，如果是先进行数据同步再进行质量的设置，那么可进行试跑操作，这里只需要完成就行。

配置完成后，将出现在监控任务列表里面。

数据源类型	全部	监控负责人	全部	表负责人	全部	全部	请输入监控名称、表名称
监控名称	监控对象	监控负责人	已启用/总规则数	创建时间	修改时间	操作	
<input checked="" type="checkbox"/> ods_training1-ods_trade_usertable (ID: 13070)	数据源类型: hive 表: training1_ods_trade_usertable 表负责人: *	曹	6 / 6	2022-07-10 01:08:37	2022-07-10 01:08:37	复制 删除 试跑 引用详情 质量报告 执行任务	

步骤五：根据上述步骤完成表 ods_trade_ordertable 和表 dwd_trade_user_order 的监控稽核任务（可选）

此处如果大家想尝试可自行进行这两表以及其它质量规则的设置。

3.5 业务元数据

事实上在完成 3.3 模型设计的时候就可以进行元数据的采集和录入工作。

角色分配：

角色	说明
元数据登记人员	具有元数据、采集任务、采集任务实例的角色
数据治理管理员	元数据注册、发布等审批人员
数据治理专员	审核指标确权、治理工单的角色

3.5.1 采集元数据

步骤一：在数据地图-元数据管理-元数据采集页面，点击新建采集任务按钮



步骤二：配置采集任务基本信息

新建采集任务

1 基本信息 2 选择采集范围 3 配置采集周期

* 采集任务名称: 实验表元数据采集

任务描述: 请输入采集任务描述

任务分组: 请选择任务分组 [设置分组](#)

* 负责人: [模糊处理]

步骤三：选择采集范围

选择 ods_trade_usertable、ods_trade_ordertable 和 dwd_trade_user_order 三张表。

1 基本信息 2 选择采集范围 3 配置采集周期

* 数据源类型: Hive

* 数据源名称: hz10-hive

* 集群: hz10

* 库/Schema: training1

* 表范围:

52 项待添加 一键全选

请输入搜索内容

- mysql2hive_stat...
- dwd_common_ssss...

> 加入右侧
< 加入左侧

< 6 / 6 >

3 项已添加 一键清空

请输入搜索内容

- ods_trade_order...
- ods_trade_usert...
- dwd_trade_user_...

< 1 / 1 >

步骤四：配置采集周期

此处选择手工采集立即触发。（可根据实际情况进行周期性采集的设置）

1 基本信息 ———— 2 选择采集范围 ———— 3 配置采集周期

* 采集类型： 定时采集 手工采集

* 手工采集时间： 立即触发 自定义时间

点击**确认**后，采集任务会出现在采集列表里面，如果采集状态显示成功，表示采集完成。

采集任务名称	任务分组	数据源类型	数据源名称	负责人	采集类型	最近采集状态	创建时间	操作
实验表元数据采集	-	Hive	hz10-hive		手工采集	成功	2022-07-10 01:29:56	编辑 删除

3.5.2 编辑元数据

步骤一：在元数据注册页面查看采集上来的元数据

采集上来的元数据处于**新建态**，支持直接编辑。

<input type="checkbox"/>	序号	资产编码	状态	最新发布版本	库/Schema	表物理名称	表中文名称	数据源类型	操作
<input type="checkbox"/>	1	TB-00000055	新建	V1	training1	ods_trade_ordertable	-	Hive	编辑 申请注册 更多 ▾
<input type="checkbox"/>	2	TB-00000056	新建	V1	training1	ods_trade_usertable	-	Hive	编辑 申请注册 更多 ▾
<input type="checkbox"/>	3	TB-00000057	新建	V1	training1	dwd_trade_user_order	-	Hive	编辑 申请注册 更多 ▾

步骤二：对元数据进行编辑，补充业务元数据

ods_trade_usertable 表信息阶段：

1 表信息 ———— 2 字段信息

基础信息

表编码：

* 表中文名称：

表物理名称：

表安全等级：

* 技术负责人：

表安全标签：

来源系统：

* 主管部门：

* 业务负责人：

表质量评分：

备注：

表敏感标签：

业务信息

主题域：

* 业务描述：

ods_trade_usertable 字段信息阶段:

user_id:

字段列表	字段设置
1. user_id	基础信息
2. phone_no	字段编码: FD-0000056-B001
3. gndr	* 中文名称: 用户ID
4. edu	* 物理字段名称: user_id
5. email_add	* 字段类型: string
+ 添加字段	* 英文全称: user_id
	主键标识: <input checked="" type="checkbox"/>
	* 安全级别: 1-低 自动推荐 原始字典: 请选择原始字典 添加 查看
	标准字典: 请选择标准字典 添加 查看 数据元: 请选择数据元 添加
	数据质量管理...: 请选择数据质量管理信息
	* 元数据类型: 基础元数据
	备注: 请输入备注
	业务信息
	* 业务定义: 用户ID
	关联指标: 请选择关联指标 添加

phone_no:

字段列表	字段设置
1. user_id	基础信息
2. phone_no	字段编码: FD-0000056-B002
3. gndr	* 中文名称: 手机号码
4. edu	* 物理字段名称: phone_no
5. email_add	* 字段类型: string
+ 添加字段	* 英文全称: phone_number
	主键标识: <input type="checkbox"/>
	* 安全级别: 3-高 自动推荐 原始字典: 请选择原始字典 添加 查看
	标准字典: 请选择标准字典 添加 查看 数据元: 手机号码 添加
	数据质量管理...: 请选择数据质量管理信息
	* 元数据类型: 基础元数据
	备注: 请输入备注
	业务信息
	* 业务定义: 手机号码
	关联指标: 请选择关联指标 添加

gndr:

字段列表	字段设置
1. user_id	
2. phone_no	
3. gndr	<p>基础信息</p> <p>字段编码: FD-00000056-B003</p> <p>* 中文名称: 性别</p> <p>* 物理字段名称: gndr</p> <p>* 字段类型: string</p> <p>* 英文全称: gender</p> <p>主键标识: <input type="checkbox"/></p> <p>* 安全级别: 1-低 自动推荐 原始字典: 请选择原始字典 添加 查看</p> <p>标准字典: 性别 添加 查看 数据元: 性别 添加</p> <p>数据质量管理...: 请选择数据质量管理信息</p> <p>* 元数据类型: 基础元数据</p> <p>备注: 请输入备注</p> <p>业务信息</p> <p>* 业务定义: 性别 关联指标: 请选择关联指标 添加</p>
4. edu	
5. email_add	
+ 添加字段	

edu:

字段列表	字段设置
1. user_id	
2. phone_no	
3. gndr	
4. edu	<p>基础信息</p> <p>字段编码: FD-00000056-B004</p> <p>* 中文名称: 学历</p> <p>* 物理字段名称: edu</p> <p>* 字段类型: string</p> <p>* 英文全称: education</p> <p>主键标识: <input type="checkbox"/></p> <p>* 安全级别: 1-低 自动推荐 原始字典: 请选择原始字典 添加 查看</p> <p>标准字典: 学历 添加 查看 数据元: 学历 添加</p> <p>数据质量管理...: 请选择数据质量管理信息</p> <p>* 元数据类型: 基础元数据</p> <p>备注: 请输入备注</p> <p>业务信息</p> <p>* 业务定义: 学历 关联指标: 请选择关联指标 添加</p>
5. email_add	
+ 添加字段	

email_add:

字段列表	字段设置
1. user_id	<div style="border: 1px solid #ccc; padding: 5px;"> <p>基础信息</p> <p>字段编码: <input type="text" value="FD-0000056-B005"/> * 中文名称: <input type="text" value="邮箱地址"/></p> <p>* 物理字段名称: <input type="text" value="email_add"/> * 字段类型: <input type="text" value="string"/></p> <p>* 英文全称: <input type="text" value="email_address"/> 主键标识: <input type="checkbox"/></p> <p>* 安全级别: <input type="text" value="3-高"/> 自动推荐 原始字典: <input type="text" value="请选择原始字典"/> 添加 查看</p> <p>标准字典: <input type="text" value="请选择标准字典"/> 添加 查看 数据元: <input type="text" value="邮箱地址"/> 添加</p> <p>数据质量管理... : <input type="text" value="请选择数据质量管理信息"/> * 元数据类型: <input type="text" value="基础元数据"/></p> <p>备注: <input type="text" value="请输入备注"/></p> </div>
2. phone_no	
3. gndr	
4. edu	
5. email_add	
+ 添加字段	<div style="border: 1px solid #ccc; padding: 5px;"> <p>业务信息</p> <p>* 业务定义: <input type="text" value="邮箱地址"/> 关联指标: <input type="text" value="请选择关联指标"/> 添加</p> </div>

ods_trade_ordertable 表信息阶段:

1 表信息
 —————
2 字段信息

基础信息	
表编码:	<input type="text" value="TB-0000055"/>
* 表中文名称:	<input type="text" value="ods_trade_ordertable"/>
表物理名称:	<input type="text" value="ods_trade_ordertable"/>
表安全等级:	<input type="text" value="请选择表安全等级"/>
* 技术负责人:	<input type="text" value=""/>
表安全标签:	<input type="text" value="请选择表安全标签"/>
业务信息	
主题域:	<input type="text" value="交易域 / 订单"/>
来源系统:	<input type="text" value="请选择来源系统"/>
* 主管部门:	<input type="text" value=""/>
* 业务负责人:	<input type="text" value=""/>
表质量评分:	<input type="text" value="仅支持输入0-100数字"/>
备注:	<input type="text" value="请输入备注"/>
表敏感标签:	<input type="text" value="请选择表敏感标签"/>
* 业务描述:	<input type="text" value="订单表"/>

ods_trade_ordertable 字段信息阶段:

order_id:

字段列表	字段设置
1. order_id	
2. user_id	
3. pay_time	
4. payment	
5. ds	
+ 添加字段	
	<div> <p>基础信息</p> <p>字段编码: <input type="text" value="FD-0000055-B001"/></p> <p>* 中文名称: <input type="text" value="订单ID"/></p> <p>* 物理字段名称: <input type="text" value="order_id"/></p> <p>* 字段类型: <input type="text" value="string"/></p> <p>* 英文全称: <input type="text" value="order_id"/></p> <p>主键标识: <input type="checkbox"/></p> <p>* 安全级别: <input type="text" value="1-低"/> 自动推荐 原始字典: <input type="text" value="请选择原始字典"/> 添加 查看</p> <p>标准字典: <input type="text" value="请选择标准字典"/> 添加 查看 数据元: <input type="text" value="请选择数据元"/> 添加</p> <p>数据质量管理... : <input type="text" value="请选择数据质量管理信息"/></p> <p>* 元数据类型: <input type="text" value="基础元数据"/></p> <p>备注: <input type="text" value="请输入备注"/></p> </div> <div> <p>业务信息</p> <p>* 业务定义: <input type="text" value="订单ID"/></p> <p>关联指标: <input type="text" value="请选择关联指标"/> 添加</p> </div>

user_id:

字段列表	字段设置
1. order_id	
2. user_id	
3. pay_time	
4. payment	
5. ds	
+ 添加字段	
	<div> <p>基础信息</p> <p>字段编码: <input type="text" value="FD-0000055-B002"/></p> <p>* 中文名称: <input type="text" value="用户ID"/></p> <p>* 物理字段名称: <input type="text" value="user_id"/></p> <p>* 字段类型: <input type="text" value="string"/></p> <p>* 英文全称: <input type="text" value="user_id"/></p> <p>主键标识: <input type="checkbox"/></p> <p>* 安全级别: <input type="text" value="1-低"/> 自动推荐 原始字典: <input type="text" value="请选择原始字典"/> 添加 查看</p> <p>标准字典: <input type="text" value="请选择标准字典"/> 添加 查看 数据元: <input type="text" value="请选择数据元"/> 添加</p> <p>数据质量管理... : <input type="text" value="请选择数据质量管理信息"/></p> <p>* 元数据类型: <input type="text" value="基础元数据"/></p> <p>备注: <input type="text" value="请输入备注"/></p> </div> <div> <p>业务信息</p> <p>* 业务定义: <input type="text" value="用户ID"/></p> <p>关联指标: <input type="text" value="请选择关联指标"/> 添加</p> </div>

pay_time:

字段列表	字段设置
1. order_id	
2. user_id	
3. pay_time	
4. payment	
5. ds X	
+ 添加字段	
	<p>基础信息</p> <p>字段编码: FD-00000055-B003</p> <p>* 中文名称: 支付时间</p> <p>* 物理字段名称: pay_time</p> <p>* 字段类型: string</p> <p>* 英文全称: pay_time</p> <p>主键标识: <input type="checkbox"/></p> <p>* 安全级别: 1-低 自动推荐 原始字典: 请选择原始字典 添加 查看</p> <p>标准字典: 请选择标准字典 添加 查看 数据元: 请选择数据元 添加</p> <p>数据质量管理...: 请选择数据质量管理信息</p> <p>* 元数据类型: 基础元数据</p> <p>备注: 请输入备注</p> <p>业务信息</p> <p>* 业务定义: 支付时间</p> <p>关联指标: 请选择关联指标 添加</p>

payment:

字段列表	字段设置
1. order_id X	
2. user_id	
3. pay_time	
4. payment	
5. ds	
+ 添加字段	
	<p>基础信息</p> <p>字段编码: FD-00000055-B004</p> <p>* 中文名称: 消费金额</p> <p>* 物理字段名称: payment</p> <p>* 字段类型: string</p> <p>* 英文全称: payment</p> <p>主键标识: <input type="checkbox"/></p> <p>* 安全级别: 1-低 自动推荐 原始字典: 请选择原始字典 添加 查看</p> <p>标准字典: 请选择标准字典 添加 查看 数据元: 请选择数据元 添加</p> <p>数据质量管理...: 请选择数据质量管理信息</p> <p>* 元数据类型: 基础元数据</p> <p>备注: 请输入备注</p> <p>业务信息</p> <p>* 业务定义: 消费金额</p> <p>关联指标: 请选择关联指标 添加</p>

说明: 如果 payment 在 dwd_trade_user_order 表中, 则此处还需关联指标。

ds:

字段列表	字段设置
1. order_id	
2. user_id	
3. pay_time	
4. payment	
5. ds	<div style="border: 1px solid #ccc; padding: 5px;"> <p>基础信息</p> <p>字段编码: FD-00000055-B005 * 中文名称: 分区字段</p> <p>* 物理字段名称: ds * 字段类型: string</p> <p>* 英文全称: ds 主键标识: <input type="checkbox"/></p> <p>* 安全级别: 1-低 自动推荐 原始字典: 请选择原始字典 添加 查看</p> <p>标准字典: 请选择标准字典 添加 查看 数据元: 请选择数据元 添加</p> <p>数据质量管理...: 请选择数据质量管理信息 * 元数据类型: 基础元数据</p> <p>备注: 请输入备注</p> <p>业务信息</p> <p>* 业务定义: 分区字段 关联指标: 请选择关联指标 添加</p> </div>
+ 添加字段	

dwd_trade_user_order 表为上述两表 join 之后所得因此设置内容基本一样不再举例。设置完成点击**保存**即可，元数据的发布等技术信息补上后再进行。

4. 开发阶段

开发阶段将重点介绍数据集成和数据开发。

4.1 数据集成

本实验是将 Mysql 表中数据同步到 Hive 库中，因此作为用户需要有 mysql 数据源的读表权限以及数据传输新建、编辑任务等相关权限。

4.1.1 新建同步任务

首先以用户表为例：

步骤一：在数据传输 - 离线同步任务页面，点击新建任务



步骤二：完成基本信息配置

基本信息

* 任务名称	mysql2hiveusertable
负责人	<input type="text"/>
* 任务类型	向导模式
引用参数组 [?]	请选择参数组
描述	输入同步任务描述，最长不超过128个字符

步骤三：配置数据来源与去向

数据来源与去向

数据来源	数据去向
* 数据源 ^与	* 数据源
MySQL Training_test2	Hive <input type="text"/>
* 库表选择 ^与	* 库表选择
training2 best_practice_user_table X	training1 ods_trade_usertable
+ 跨库表	快速创建表
* 全量初始化 ^与	* 写入规则
<input type="radio"/> 关	Insert overwrite
数据过滤 ^与	
<input checked="" type="radio"/> 条件 <input type="radio"/> 流水型 <input type="radio"/> 自定义	
列 操作符 条件	
<input type="text"/> <input type="text"/> <input type="text"/>	
+ 添加	
特殊字符替换 ^与	
<input type="radio"/> 关	
并发读取 ^与	
<input type="radio"/> 关	
* 流量控制	
<input checked="" type="radio"/> 不限速 <input type="radio"/> 限速	

步骤四：配置字段映射

字段映射 ^{重置}

序号	来源表字段 ^与	类型	描述	去向表字段	类型	描述
1	user_id	VARCHAR(64)	用户ID	user_id	string	-
2	phone_number	BIGINT(19, 0)	手机号	phone_no	string	-
3	gender	INT(10, 0)	性别	gndr	string	-
4	edu	INT(10, 0)	教育程度	edu	string	-
5	email_address	VARCHAR(64)	邮箱地址	email_add	string	-

步骤五：数据脱敏配置（选填）

数据传输支持静态脱敏，需要事先在安全中心配置脱敏规则，此处以预置规则邮箱地址举

例。

配置如下，点击扫描后，系统会对识别表中符合邮箱地址格式的字段，并进行脱敏样例的展示。

来源表字段	脱敏规则名称	敏感类型	脱敏算法	算法配置	脱敏样例
email_address	邮箱地址遮盖脱敏	邮箱	mask	@前遮盖, 遮盖字符*	*****@163.com

配置完成，点击确定完成用户表的同步。

订单表同步：

对于订单表和用户表稍有不同的是，订单表在 mysql 中并未分区，但是有一个可以用来分区的字段，同步到 hive 后会需要通过这个字段进行分区，因此此处配置稍有不同。

来源表字段	脱敏规则名称	敏感类型	脱敏算法	算法配置	脱敏样例
email_address	邮箱地址遮盖脱敏	邮箱	mask	@前遮盖, 遮盖字符*	*****@163.com

其它项的配置无特殊性，配置完成点击确定即可。

4.1.2 数据同步任务创建

步骤一：试运行已创建完成的两个同步任务

任务名称	任务状态	数据来源类型	来源信息	操作
mysql2hiveordertable	未提交	MySQL	training2:best_practice_order_table	编辑 运行 日志 更多
mysql2hiveusertable	未提交	MySQL	training2:best_practice_user_table	编辑 运行 日志 更多

点击**运行**后，需要选择队列，如果没有队列可选需要找管理员添加队列的权限。

步骤二：将运行成功的任务，提交上线

任务名称	任务状态	数据来源类型	来源信息	操作
<input checked="" type="checkbox"/> mysql2hiveordertable	未提交	MySQL	training2.best_practice_order_table	编辑 运行 日志 更多
<input checked="" type="checkbox"/> mysql2hiveusertable	未提交	MySQL	training2.best_practice_user_table	编辑 运行 日志
<input type="checkbox"/> mysql2hive...	未提交	MySQL	training2.best_practice...	编辑 运行 日志

步骤三：对已提交的同步任务创建离线开发任务

任务名称

任务名称	任务状态	数据来源类型	来源信息	操作
<input checked="" type="checkbox"/> mysql2hiveordertable	已提交	MySQL	training2.best_practice_order_table	编辑 运行 日志 更多
<input checked="" type="checkbox"/> mysql2hiveusertable	已提交	MySQL	training2.best_practice_user_table	编辑 运行 日志 更多

在弹框中，选择保存位置及输入离线开发任务的名称。

创建离线开发任务

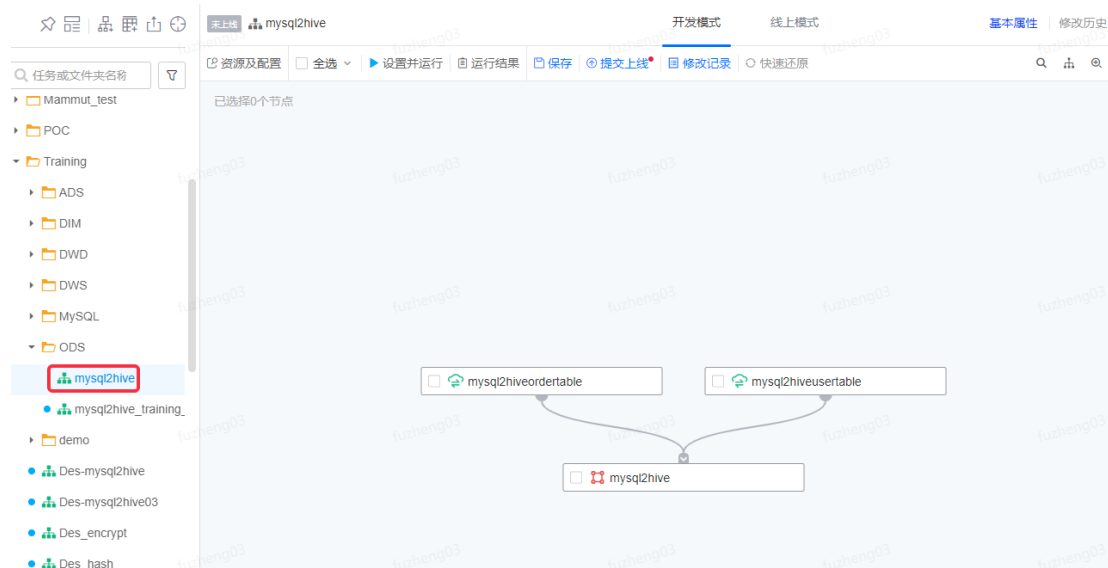
* 保存位置

* 离线开发任务名称

数据同步节点名称

* 分组数量

步骤四：创建完成后可在离线开发中查看到此任务

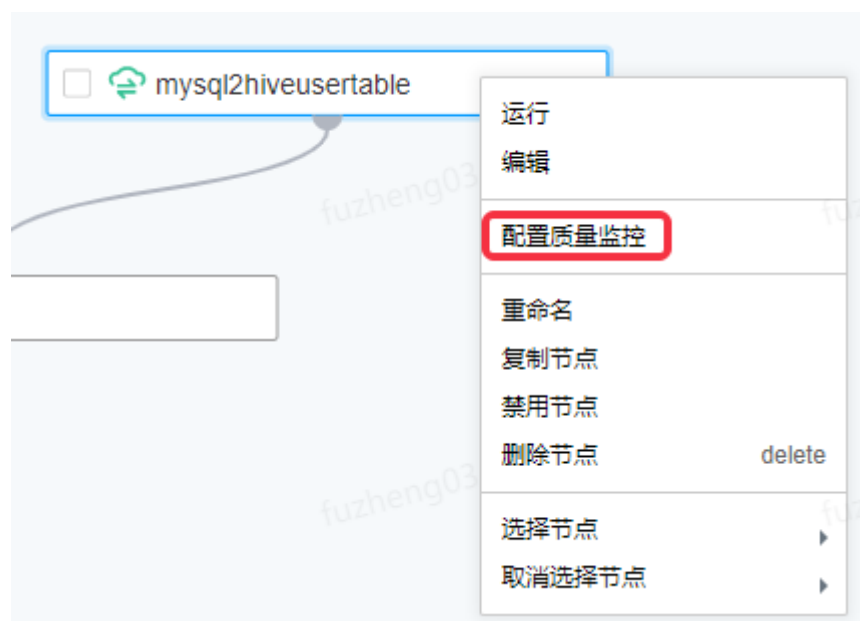


此时的任务处于开发模式，还未提交至线上模式配置调度。

步骤五：绑定质量监控任务

由于之前已配置了 `ods_trade_usertable` 表的质量监控任务，因此此处可将其绑定在任务上。

右键 `mysql2hiveusertable` 节点，选择配置质量监控。



在弹框中的监控名称选择监控任务，如果忘记了，可前往数据质量中心查看，执行失败策略选择默认即可。

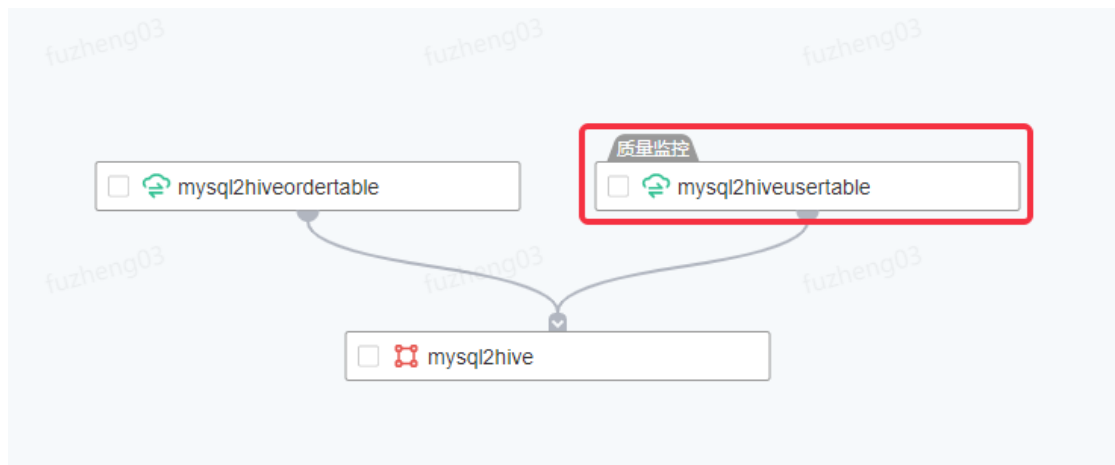
配置质量监控-mysql2hiveusertable

监控名称* [前往数据质量中心](#)

执行失败策略 ? 终止本任务 继续下游调度

set 属性	属性 ?	值
		无数据

配置完成后，点击**临时保存**。

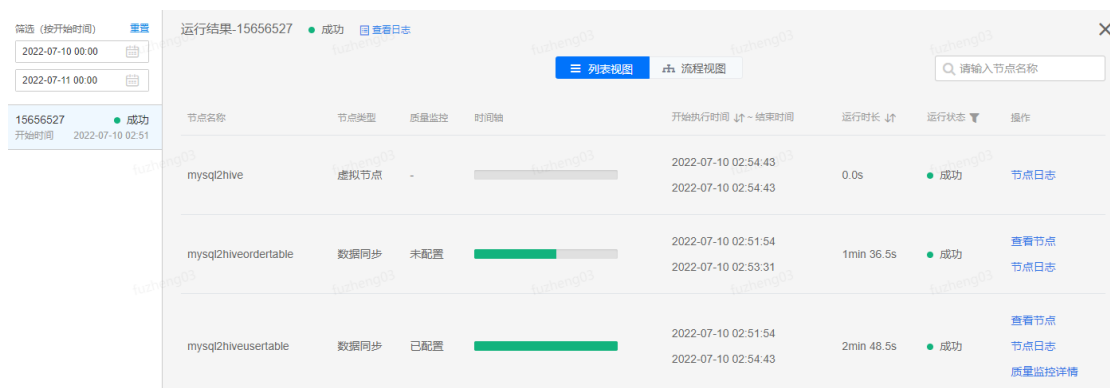


步骤六：节点试运行

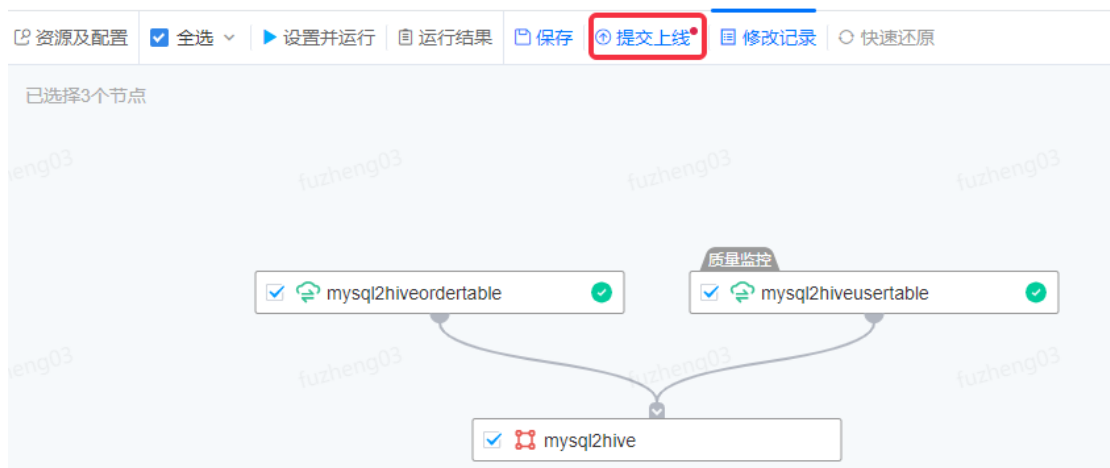
为了保证开发的质量，此时建议将节点和质量监控都进行试跑，全选所有节点并点击**设置并运行**。



在运行结果页面中查看结果，如果成功则可进行提交上线。



步骤七：将任务提交上线并配置调度



在线上模式下编辑调度，点击**编辑调度**按钮，在页面中设置调度周期。

编辑调度

调度设置

调度设置类型 常规设置 crontab设置

调度周期

调度日历 [查看调度日历](#)

首次执行时间

调度失效时间 不设置（永久执行） 自定义

跨流依赖设置

跨周期自依赖

依赖对象 | 任务依赖

| 节点依赖

执行设置

队列

高优先级

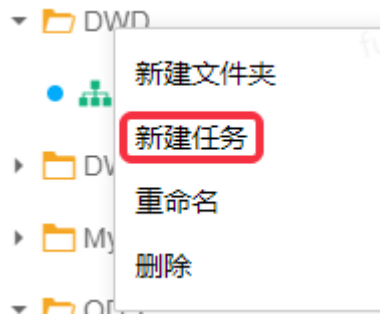
自动重试 ⚠️ 若运行的节点开启了自动重试，按照节点设置的规则优先。

除了修改调度周期外其它用默认配置，点击确定并提交调度。

4.2 数据开发

步骤一：在离线开发页面，新建任务

右键文件夹点击**新建任务**。



填写任务名称，此处不涉及资源组和参数组引用因此不配置。

新建任务 ×

任务名称 *

任务类型 多节点任务 单节点任务 任务组

保存位置 *

[+ 新建文件夹](#)

引用资源组 (?)

引用参数组 (?)

默认外部数据源 (?) [+ 添加](#)

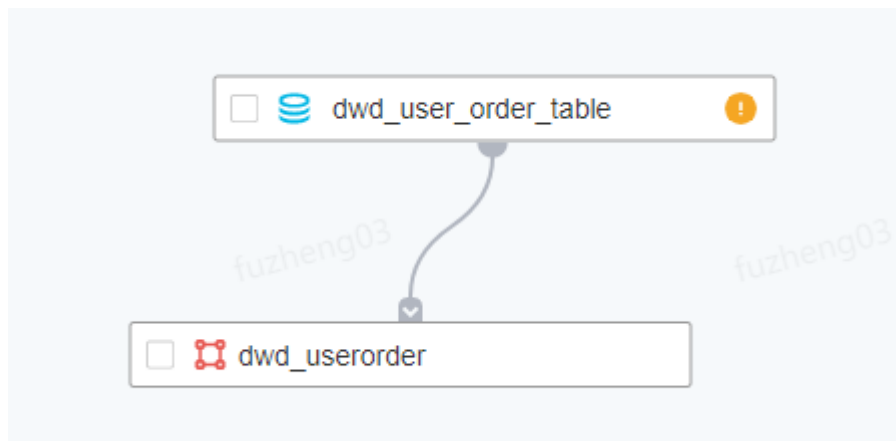
任务描述

0/120

步骤二：拖入右侧 SQL 节点



对节点进行重命名，并将 SQL 节点和虚拟节点进行连接。



步骤三：编写 SQL 代码

双击 SQL 节点，输入如下代码：

```
INSERT OVERWRITE TABLE training1.dwd_trade_user_order
PARTITION (ds='${azkaban.flow.1.days.ago}')
SELECT t1.user_id AS user_id -- string
      , t1.phone_no AS phone_no -- string
      , t1.gndr AS gndr -- string
      , t1.edu AS edu -- string
      , t1.email_add AS email_add -- string
```

```
, t2.order_id AS order_id -- string
, t2.pay_time AS pay_time -- string
, t2.payment AS payment -- string
FROM (
SELECT user_id -- string
, phone_no -- string
, gndr -- string
, edu -- string
, email_add -- string
FROM training1.ods_trade_usertable
) t1
LEFT OUTER JOIN (
SELECT order_id -- string
, pay_time -- string
, payment -- string
, ds -- string
, user_id -- string
FROM training1.ods_trade_ordertable
) t2
ON t1.user_id=t2.user_id
;
```

输入完成后，点击**临时保存**。

步骤四：对节点进行试运行并查看运行结果



运行参数用默认配置即可。

节点名称	节点类型	质量监控	时间轴	开始执行时间	结束时间	运行时长	运行状态	操作
dwd_user_order_table	SQL	未配置		2022-07-10 03:24:18	2022-07-10 03:26:25	2min 7.0s	成功	查看节点 节点日志 结果预览
dwd_userorder	虚拟节点	-		2022-07-10 03:26:25	2022-07-10 03:26:25	0.0s	成功	节点日志

运行成功后即可提交上线。

步骤五：配置调度

选中需要配置的节点，点击**设置并提交**按钮进入编辑调度页面。

调度设置如下：

编辑调度

调度设置

调度设置类型 常规设置 crontab设置

调度周期

调度日历 [查看调度日历](#)

首次执行时间

调度失效时间 不设置 (永久执行) 自定义

步骤六：配置依赖

首先，点击**智能推荐**按钮。

跨流依赖设置

跨周期自依赖

依赖对象 | 任务依赖

| 节点依赖

在**任务依赖的智能推荐**页面，选择需要依赖的任务或节点，点击下一步。

任务依赖的智能推荐

- 1. A.本功能会基于“SQL”节点的输入表进行“任务依赖”推荐，推荐数据来源于任务的运行实例，因此请确保上游表的产出任务至少在线上调度运行过一次；
- B.非SQL节点当前无法给出推荐，需返回到“编辑调度”中自行配置依赖。

1 节点解析及输入表选择 ————— 2 确定依赖

本任务节点名称	输出表	输入表	是否选择 ?
dwd_user_order_table	training1.dwd_trade_user_order	training1.ods_trade_ordertable	<input checked="" type="checkbox"/>
		training1.ods_trade_usertable	<input checked="" type="checkbox"/>

下一步 取消

说明：如果此时无法自动添加依赖关系，可能的原因是当前需要依赖的节点并未线上调度过，此时可手动添加。

将依赖添加后如下：

跨流依赖设置

跨周期自依赖

依赖对象 | 任务依赖

添加任务依赖节点 智能推荐

当前任务: dwd_userorder 编辑

项目名称	任务名称	已选节点
poc	mysql2hive	mysql2hiveordertable × mysql2hiveusertable ×

节点依赖

添加节点依赖节点

步骤七：完成配置

其它配置项可使用默认配置，点击确定并提交调度。

之后所有的任务可在任务运维处进行管理。

4.3 元数据技术信息的采集

在数据地图页面，对 ods_trade_usertable 进行编辑，在表信息页面的技术信息项点击扫描技术信息。

技术信息 **扫描技术信息**

所属分层:	请选择所属分层	Schema:	training1
数据源类型:	Hive	数据源:	hz10-hive
所属集群:	请输入所属集群	存储方式:	请输入存储方式
序列化格式:	请输入序列化格式	存储大小:	请输入存储大小
存储路径:	请输入存储路径	建表语句:	请输入建表语句
分区信息:	请输入分区信息		

此时，系统会自动将表的技术信息扫描出来，可自行选择忽略或覆盖。

技术信息 扫描技术信息

所属分层:	请选择所属分层	Schema:	training1
数据源类型:	Hive	数据源:	hz10-hive
所属集群:	请输入所属集群	存储方式:	请输入存储方式
序列化格式:	请输入序列化格式	存储大小:	请输入存储大小
存储路径:	请输入存储路径	建表语句:	请输入建表语句
分区信息:	请输入分区信息		

扫描结果: org.apache.hadoop.hive.q1.io.parquet.serde.ParquetHiveSerDe [忽略](#) [覆盖](#)

扫描结果: hdfs://hz-cluster10/user/poc/hive_db/training1.db/ods_trade_usertable [忽略](#) [覆盖](#)

扫描结果: 990.0B [忽略](#) [覆盖](#)

将所有信息覆盖后点击下一步，查看字段信息是否发生变化，如有变化自行选择忽略或覆盖，全部完成后点击保存即可。其它两张表重复上述操作。

4.3.1 元数据注册

将编辑完成的元数据申请注册。

<input type="checkbox"/>	序号	资产编码	状态	最新发布版本	库/Schema	表物理名称	操作
<input type="checkbox"/>	1	TB-00000055	新建	V1	training1	ods_trade_ordertable	编辑 申请注册 更多 ▾
<input type="checkbox"/>	2	TB-00000056	新建	V1	training1	ods_trade_usertable	编辑 申请注册 更多 ▾
<input type="checkbox"/>	3	TB-00000057	新建	V1	training1	dwd_trade_user_order	编辑 申请注册 更多 ▾

选择审核人审批通过后，元数据为草稿态。

<input type="checkbox"/>	序号	资产编码	状态	最新发布版本	库/Schema	表物理名称	操作
<input type="checkbox"/>	1	TB-00000055	草稿	V1	training1	ods_trade_ordertable	申请确权 申请变更安全等级 更多 ▾
<input type="checkbox"/>	2	TB-00000056	草稿	V1	training1	ods_trade_usertable	申请确权 申请变更安全等级 更多 ▾
<input type="checkbox"/>	3	TB-00000057	草稿	V1	training1	dwd_trade_user_order	申请确权 申请变更安全等级 更多 ▾

4.3.2 元数据发布

将处于草稿态的元数据进行发布。

<input type="checkbox"/>	序号	资产编码	状态	最新发布版本	库/Schema	表物理名称	操作
<input type="checkbox"/>	1	TB-00000055	草稿	V1	training1	ods_trade_ordertable	申请确权 申请变更安全等级 更多 ▾
<input type="checkbox"/>	2	TB-00000056	草稿	V1	training1	ods_trade_usertable	申请确权 申请变更安全等级 更多 ▾
<input type="checkbox"/>	3	TB-00000057	草稿	V1	training1	dwd_trade_user_order	申请确权 申请变更安全等级 更多 ▾

同样需要选择审核人进行审批，审批通过后元数据变为已发布，此时元数据的注册完成。

<input type="checkbox"/>	序号	资产编码	状态	最新发布版本	库/Schema	表物理名称	表中文名称	操作
<input type="checkbox"/>	1	TB-00000055	已发布	V2	training1	ods_trade_ordertable	ods_trade	申请治理 申请下线 更多 ▾
<input type="checkbox"/>	2	TB-00000056	已发布	V2	training1	ods_trade_usertable	ods_trade	申请治理 申请下线 更多 ▾
<input type="checkbox"/>	3	TB-00000057	已发布	V2	training1	dwd_trade_user_order	dwd_trade	申请治理 申请下线 更多 ▾