

智能数据构建与管理 Dataphin 数据萃取

ALIBABA CLOUD

文档版本: 20220519



法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。 如果您阅读或使用本文档,您的阅读或使用行为将被视为对本声明全部内容的认可。

- 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档,且仅能用 于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息,您应当严格 遵守保密义务;未经阿里云事先书面同意,您不得向任何第三方披露本手册内容或 提供给任何第三方使用。
- 未经阿里云事先书面许可,任何单位、公司或个人不得擅自摘抄、翻译、复制本文 档内容的部分或全部,不得以任何方式或途径进行传播和宣传。
- 由于产品版本升级、调整或其他原因,本文档内容有可能变更。阿里云保留在没有 任何通知或者提示下对本文档的内容进行修改的权利,并在阿里云授权通道中不时 发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠 道下载、获取最新版的用户文档。
- 4. 本文档仅作为用户使用阿里云产品及服务的参考性指引,阿里云以产品及服务的"现状"、"有缺陷"和"当前功能"的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引,但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的,阿里云不承担任何法律责任。在任何情况下,阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害,包括用户使用或信赖本文档而遭受的利润损失,承担责任(即使阿里云已被告知该等损失的可能性)。
- 5. 阿里云网站上所有内容,包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计,均由阿里云和/或其关联公司依法拥有其知识产权,包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意,任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外,未经阿里云事先书面同意,任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称(包括但不限于单独为或以组合形式包含"阿里云"、"Aliyun"、"万网"等阿里云和/或其关联公司品牌,上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司)。
- 6. 如若发现本文档存在任何错误,请与阿里云取得直接联系。

通用约定

格式	说明	样例
⚠ 危险	该类警示信息将导致系统重大变更甚至故 障,或者导致人身伤害等结果。	
▲ 警告	该类警示信息可能会导致系统重大变更甚 至故障,或者导致人身伤害等结果。	警告 重启操作将导致业务中断,恢复业务 时间约十分钟。
〔) 注意	用于警示信息、补充说明等,是用户必须 了解的内容。	大意 权重设置为0,该服务器不会再接受新 请求。
? 说明	用于补充说明、最佳实践、窍门等,不是 用户必须了解的内容。	⑦ 说明 您也可以通过按Ctrl+A选中全部文件。
>	多级菜单递进。	单击设置> 网络> 设置网络类型。
粗体	表示按键、菜单、页面名称等UI元素。	在 结果确认 页面,单击 确定 。
Courier字体	命令或代码。	执行 cd /d C:/window 命令,进入 Windows系统文件夹。
斜体	表示参数、变量。	bae log listinstanceid
[] 或者 [alb]	表示可选项,至多选择一个。	ipconfig [-all -t]
{} 或者 {a b}	表示必选项,至多选择一个。	switch {act ive st and}

目录

1.概述	05
2.基本概念	07
3.进入数据萃取	13
4.初始化萃取项目	15
5.平台管理	18
5.1. 创建并管理ID类目	18
5.2. 创建并管理标签类目	21
5.3. 创建行为域和业务线	26
6.ID中心	29
6.1. 创建并管理实体ID	29
6.2. 引入ID映射表及开启OnelD归一化	32
7.行为中心	37
7.1. 创建行为元素	37
7.2. 创建并配置行为规则	40
7.3. 查看行为数据及下载行为定义表	55
8.标签中心	58
8.1. 创建规则统计标签	58
8.1.1. 场景引导	58
8.1.2. 步骤一: 创建规则统计标签	58
8.1.3. 步骤二: 配置并提交规则统计标签	60
8.2. 创建行为偏好标签	76
8.2.1. 步骤一: 创建行为偏好标签	77
8.2.2. 步骤二: 配置并提交行为偏好标签	79
8.3. 创建注册上挂标签	88
8.4. 创建标签逻辑表	92

1.概述

面向客户数据价值深度挖掘的诉求,萃取支持以实体(如消费者、设备)为中心的数据融通和深度挖掘,提 供标签可视化配置及分类管理能力,帮助企业降低开发成本、快速构建标签体系、沉淀可直接应用的高价值 数据,以助力业务精准营销和决策分析。

功能介绍

基于Dat aphin数据建模研发沉淀的数据或简单加工清洗后的明细数据,数据萃取提供实体识别与连接、行为数据清洗与整合、标签可视化配置及自动化调度的能力,包括ID中心、行为中心、标签中心和平台管理。

≡	Dataphin · 研发	萃取 ☆ 运维 权限 ———	Q、 全局搜索	ter 🕹 🖉 🖉
Distill B	Data_distill • V			
IC	中心 行为中心 标签中心	欢迎使用Dataphin数据萃取!		
4		载源苹取为您提供D定义,行为载弱规范化结构化聚集。 标签可现化配置的能力,结合onelD体系,您可以统一实		
Ø	• •	体信息,高速、有效构建可直接应用的核心标签,以消除 数定观岛,提升数定价值。		
⇆	• • •			
Q	 (2) (3) 	S 10#43	行为中心	● 标签中心
)			
	· • (1)	支持架体 ID 的过去人族 OnellD 規則表的引入、通过 ID mapping 架构即 一实体识 別以深度地理数据价值		
) (2)			
	• 10 million - 10		 創建实体ID 由 	④ 引入ID 映射表
	• 1 M			
	· • • (1)			
		▲ 行为中心		\mathbf{r}
		通过行为构建及行为规则定义,实现对行为数据的规范化结构化聚合。 以支持标签 生产及用户营销等场景		A
			 ● 定义行为元素 団 	❷ 创建行为规则 玊
	(2)			
	No. of Concession, Name	% 标签中心		
		支持基于萃取平台的标签目助化生产及原生标签引入,提供统一的标签 管理平台,并通过标签逻辑表统一对外提供服务		
	(2)			-
	· Manager and Party and Pa		 ● 創建标签 ÷ 	❷ 创建标签逻辑表 →
	• No.		<u> </u>	•
		為 平台管理		
	(0)	支持定义全局通用的ID类目、行为域及标签类目,以实现集中管理		
	行为规则对象列表		0 行为城隍义 []	2 英目煌义 日
◙	平台管理			

- 创建并管理实体ID:实体ID定义和ID间mapping关系的管理,是后续标签计算来源数据连关联的核心。实体ID和ID映射表的详细说明如下:
 - 实体ID:您可定义从不同视角对实体进行描述的ID,如消费者实体:会员ID、手机号;商品实体:价格、产地。
 - ID映射表:您可引入以OneID为主键的ID映射表,以实现对实体ID的识别与连接,并应用于下游标签生产中。
- 创建行为元素:行为数据规范化、结构化聚集。其中,行为元素用于统一名称规范,避免数据关联遗漏; 行为规则用于定义从不同的来源表中按照一定的规则抽取数据并统一存储,保证数据可用性。行为元素和 行为规则的详细说明如下。
 - 行为元素:行为元素是行为的组成元素,行为由行为域、业务线、动作、对象构成,不同行为元素之间
 可自由组合。
 - 行为规则:行为规则定义每个行为从不同来源表的取数规则,可指定ID、对象、对象属性等对应的来源表字段。
- 场景引导:标签可视化配置与自动化调度。支持多种类型标签的配置与统一管理,并通过标签逻辑表统一 对接下游提供服务。详细说明如下:
 - 标签生产:支持规则统计标签、行为偏好标签、注册上挂标签的可视化配置与自动化调度。

- 标签消费:您可以基于业务目标自由组合需要的ID及标签生成标签逻辑表,并通过配置逻辑表集成任务
 导出到应用数据库向上层提供服务。
- 创建并管理ID类目: 支持定义全局通用的ID类目、行为域及标签类目, 以实现集中管理。

操作指引



相关参考

基于电商销售场景构建偏好标签

2.基本概念

本文为您介绍数据萃取的基本概念,包含实体ID、行为、行为属性、行为规则、标签和标签逻辑表,以便于 您更好地理解数据萃取。

实体ID

实体ID是某个体系中相对唯一的编码,类似于身份标识,用于识别某个实体。在某一具体的事物中,实体ID 一般是不变的,例如员工工号、身份证号码、计算机网址。

例如,顾客在电商平台淘宝网站浏览商品的行为中,对应的行为主体、行为元素和实体ID,如下图所示。



行为

行为是用户(行为主体)有目的的活动,一个完整的行为包括**行为主体、行为元素**以及相关的行为属性信息。

行为元素包括行为域、业务线、动作和对象,Dataphin将定义的行为元素写入至行为定义表中。行为元素的 详细说明如下表所示。

行为元素	描述
行为域	聚合业务含义一致的行为数据。例如, 电商域、文娱域。
业务线	基于行为域将行为数据进一步细分,行为域下的各业务线相互独立。例如,淘宝业务 线、天猫业务线。
动作	行为主体发出的操作。例如,购买、浏览。

行为元素	描述
对象	行为主体操作的具体事物。例如,商品、电影。
对象属性	对象的描述性信息。例如,名称、品牌、年份。

例如,顾客在电商平台淘宝网站购买美妆类商品的行为中,对应的行为主体和行为元素,如下图所示。



行为属性

行为属性是对行为的描述信息,包括发生次数、金额、时间和频次。详细说明见下表。

行为属性	描述	示例
发生次数	行为的发生次数。	某顾客在电商平台淘宝网站购买了4次美妆类商品。
金额	行为执行过程中需 要支付的金额。	某顾客在电商平台淘宝网站购买美妆类商品的订单中支付200元。
时间	行为发生的时间。	2021年6月1日某顾客在电商平台淘宝网站购买美妆类商品。
频次	行为中动作发生的 频次。	一个月内,某顾客在电商平台淘宝网站购买了4次美妆类商品。频次 为4次/月。

行为规则

从业务数据表中筛选出符合行为的数据的规则,称为行为规则。行为规则筛选出的行为数据写入至行为规则 产出表(oi_behavior_det ail_d)中。

定义行为规则及将行为数据写入至行为规则产出表中,如下图所示。

指定待分析的来源数据表			源数据表 筛选出符合行为的数据			筛选出的数据写入至行为规则产出表			
			行为规则	U			行为规则产出表		
	业务数据表			配置筛选规则					
column	type	comment		筛选的字段	行为		字段	字段名称	
name	string	用户姓名		name	实体ID	\rightarrow	person_ids	ID原始值	
item_id	string	商品ID		item_id	对象标识		object_id	对象标识	
cate_level	string	商品分类		cate_level	对象属性		object_properties	对象属性	
stat_count	string	行为发生的次 数		stat_count	发生次数	\rightarrow	stat_count	发生次数	
							ds	日期	
							source_table	来源主表	
							behavior_id	行为ID	

⑦ 说明 一个行为可以从多个业务数据表中筛选数据,因此基于一个行为可以生成多个行为规则。

行为看板

行为看板用于可视化展示行为规则产出表(oi_behavior_detail_d)中的数据,同时支持下载行为定义表。

行为看板		统计数据最近更新时间: 2021-06-17 21:10:03
行为中心产出表: oi_statistics_behavior_detail_d 已接入行为规则:	18 最近90天累计采集用户行为: 929	
 最示比例: 100% + - 里賀 1) 可点击単个減、业务线、动作対象名称或图标以查看详情 	し 下載行为定义表	行为数据抽样:所有行为 抽样业务日期: 2021-07-22
 2)可hover单个域、业务线、动作对象图标获取描述信息 域 ① 点击单个域查看详情 	山务线 动作对象	
	浏览的_简品	暫无统计数据
No_NBTA	2 收藏的_简品	
	购买店_商品	

行为规则产出表的更多信息,请参见行为规则。

标签

标签用于刻画用户画像、描述商品、描述企业、描述位置等。例如,文娱爱好消费者、美妆商品、金融企业、上海区域。

标签包括规则统计类标签和注册上挂标签,详细说明如下:

- 规则统计类标签:基于行为数据或标签数据筛选需要打标的对象,并为对象配置标签值,最后产出的标签数据写入至标签结果表。行为数据和标签数据的详细说明如下:
 - 行为数据即行为规则产出表的数据。例如,基于最近30天内购买商品、收藏商品和浏览商品的行为数据,筛选出购买、收藏和浏览商品的类目为动漫的用户,并设置用户的标签为文娱。

标签数据包括所有已提交的标签。例如,基于购买食品种类的标签,进一步生成购买零食分类的标签。
 如筛选出购买食品种类的标签值为果脯、薯片、坚果的数据,并指定新的标签值为零食爱好者。

标签结果表的结构如下图所示。

字段	字段类型
ID_type	string
ID_value	string
label_value	string
ds	string

注册上挂标签:将Dataphin已有数据表(汇总逻辑表或物理表)与实体ID进行关联,并为数据表的某些字段添加标签。上挂标签创建完成后,仅生成实体ID、字段和标签的映射关系,映射关系的示例如下图所示。

来源表				实体
column	comment		标签	标签主键
字段01	字段01的描述	\longrightarrow	标签01	实体ID
字段02	字段02的描述	\longrightarrow	标签02	
字段03	字段03的描述	\longrightarrow	标签03	

标签逻辑表

标签逻辑表用于将单个实体ID下多个标签进行关联,或将有关联的多个实体ID及其标签进行关联,关联后将 实体ID字段、标签字段及时间分区字段写入至标签逻辑表。例如,基于某顾客ID已生成身高、地域和爱好的 标签,将身高、地域和爱好标签关联后,生成对应的标签逻辑表,如下图所示。



标签逻辑表的典型应用场景说明如下。



● **营销推送场景**:同步标签逻辑表至业务数据库后,即可基于Quick Audience等应用进行营销数据分析及圈 选投放。

例如,某企业新推出一款高端美白系列护肤品,企业需要筛选出潜在顾客进行试用。假设已有顾客的年龄、价格偏好和护肤偏好的标签,您就可以将这些标签关联生成标签逻辑表。您可以基于Quick Audience 平台对标签逻辑表进行圈选年龄在20~35岁之间、护肤偏好为美白、价格偏好为800~1500元的顾客群体。

数据分析场景:同步标签逻辑表至业务数据库后,即可通过Quick BI等智能服务平台进行可视化数据分析。Quick BI更多信息,请参见通过仪表板分析数据。

例如,基于顾客的居住地和护肤偏好的标签生成标签逻辑表。同步标签逻辑表至业务数据库中,即可引入 至Quick BI平台,分析并展示上海、深圳、广州和北京地域顾客的护肤偏好分布。

3.进入数据萃取

在您开始使用数据萃取前,首先需要进入数据萃取控制台。本文为您介绍如何进入数据萃取的各模块。

前提条件

已完成萃取项目的初始化,请参见初始化萃取项目。

场景说明

≡	Dataphin · 研发	萃取 ☆ 运维 权限 Q 全局搜索		
Distill B ID	莽取数据中心 Data_distill ◆ ✓ ■ 中心 行为中心 标签中心	欢迎使用Dataphin数据萃取!		
	实体ID じ E Q 输入关键字 ▶ 产品 (1) Ifs_test1 (0) ▶ 目を使用 (1)	行力域 近し、 込み 対象 ごの ごの		
	, ● 頁回D (1) > ■ ffs週試 (3) > ■ 週試 (1)	行为规则 (行为看板) (行为看板) (行为看板) (行为看板) (行为看板) (行为看板) (行为有板) (行) (行为有板) (行) (行) (行) (行) (行) (行) (行) (行) (行) (行		
	实体ID对象列表 :			
功	能模块	使用场景		
进	入行为中心	 用于定义用户行为和行为规则,实现对行为数据的规范化、机构化的聚合,以支持标签 生产。行为中心包括行为元素和行为规则: 行为元素是行为的组成元素,用于标准化定义与归类业务数据。行为元素由动作、对 象和对象属性组成。行为由行为域、业务线、动作、对象和对象属性组成。 行为规则定义行为从来源表取数规则,包括选择行为、选择来源表、指定实体ID、对 象、对象属性和行为属性。 		
进	入旧中心	支持定义实体ID,您可以从不同视角定义对实体的描述。例如,消费者实体包括会员ID 和手机号,商品实体包括价格和产地。		
进入标签中心 用于生产规则统计类标签和消费标签,同时支持将物理表字段和逻辑表指标 本地标签。				
进入平台管理 用于定义和分类管理全局通用的ID类目、行为域、业务线和标签类目,帮 业务视角进行核心资产生产和管理。				

进入行为中心

- 1. 登录Dataphin控制台。
- 2. 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。

Dataphin工作区				▲ Access Key配置
**	灾迎来到Dataphin!	智能数据仓库构建	高效数据资产管理	
选择工作区地域				开通其他Region 续费
(梁圳) (北京) 华东2 (上海) 华南1 (深圳)	华东1 (杭州)			
进入Dataphin >>				
 日开通: 基础研发版(MarCompute): 数/編成量-200月回 数/编算取,示点版 资产)出理-标准版 资产)出现-标准版 数/编成务-最大500QPS 	200数攝处理单元 开退日順 升级 当前状态	: : :: 还有228天到期。	续费	
快速开始相关工作				
数据研发 项目管理 数据萃取	调度运维 资产地图	8 资产质量	资产治理	
数据服务 成员管理				

您也可以单击快速开始相关工作区域的数据服务,快速进入数据萃取。

在Dataphin首页,单击研发后,鼠标悬停至开发上,单击萃取。
 系统默认进入数据萃取的行为中心。

在行为中心定义行为元素和行为规则。行为中心的更多信息,请参见创建行为元素、创建并配置行为规则。

进入ID中心

- 1. 请参见进入行为中心,进入数据萃取模块。
- 2. 在**行为中心**页签,单击ID中心,进入ID中心。

在ID中心,即可创建实体ID。创建实体ID的更多信息,请参见创建并管理实体ID。

进入标签中心

- 1. 请参见进入行为中心,进入数据萃取模块。
- 2. 在行为中心页签,单击标签中心,进入标签中心。
 在标签中心开发统计标签、注册上挂标签和标签逻辑表。标签的更多信息,请参见步骤一:创建规则统 计标签、创建注册上挂标签和创建标签逻辑表。

进入平台管理

- 1. 请参见进入ID中心、进入行为中心或进入标签中心,进入ID中心、行为中心或标签中心。
- 在ID中心、行为中心或标签中心页面,单击左侧列表底部的平台管理,进入平台管理。
 在平台管理新建行为域、业务线、ID类目和标签类目。平台管理的更多信息,请参见创建行为域和业务 线、创建并管理ID类目、创建并管理标签类目。

4.初始化萃取项目

在正式开始萃取数据研发工作之前,需要超级管理员先初始化萃取项目的工作空间。本文为您介绍如何初始 化萃取项目的工作空间。

前提条件

创建萃取项目单独的计算引源,请参见新建MaxCompute计算源。

使用限制

系统仅支持超级管理员角色的账号进行萃取项目的初始化配置。

注意事项

- 萃取项目初始化完成后,计算源不能修改。
- 请勿在计算源列表中删除萃取项目引用的计算源,否则萃取项目中的任务无法恢复。

初始化配置

- 1. 使用阿里云账号,登录Dataphin控制台。
- 2. 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
- 3. 在Dataphin产品首页,单击顶部菜单栏中的研发。
- 4. 鼠标悬停至开发上后,单击萃取。
- 5. 在初始化配置页面,选择计算源。

Dataphin · 研发 ^{萃取} ≠	运维	权限	Q 全局搜索
初始化配置在使用的服萃取功能相绑定的计算引擎源名称,数据	萃取相关的计算的	由此承担,绑定完成后,系统会自动帮您建立"数据萃取"项目	
研发模式:	Basic 🛈		
项目名称:	数据萃取		
项目英文名:	Data_distill		
描述:	数据萃取专用项	目,系统自动创建,不可修改。与ID中心、行为中心、标签中心相关的产出表、论	+算节点、资源消耗,皆位于此项目内。
* 计算引擎源	distill	◇ ① 测试连接性	
业务板块名称:	萃取数据中心		
业务板块英文名:	LD_distill		
空间类型:	应用层		
		初始化失败,请点击按钮重新初始化	
		确定并开始初始化	
$\gamma \rightarrow \phi$			

<!>↓ 注意!

- 。 萃取项目初始化完成后, 计算源不能修改。
- 。 请勿在计算源列表中删除萃取项目引用的计算源,否则萃取项目中的任务无法恢复。

6. 单击测试连接性。

7. 测试连接通过后,单击确定并开始初始化。初始化成功后,即可进入萃取项目的工作空间进行研发数

据。

已初始化的萃取项目说明

萃取项目初始化完成后,系统自动生成了默认的基础参数,详情见下表。

项目	说明
业务板块	 英文名称固定为LD_distill。 名称固定为萃取数据中心。
项目	 英文名称固定为data_distill。 名称固定为萃取数据。
数据域	 英文名称固定为data_distill_domain。 名称缩写固定为distilldom。 名称固定为数据萃取域。
其他ID	 名称固定为其他ID。 英文名固定为other_IDs。 描述固定为系统类目,不可编辑和删除,未分组的ID默认归属此类目。
其他标签	 名称固定为其他标签。 英文名固定为other_labels。 描述固定为系统类目,不可编辑和删除,未分组的标签默认归属此类目。

② 说明 如果您购买的数据萃取是2.9.5之前版本,则初始化后萃取项目有两个虚拟维度(ID_type和 ID_value)和一张标签逻辑表(dws_label_id)。

初始化后萃取项目的配额限制

Dataphin对单租户的资源数量配额限制说明。

模块	规格描述	默认值
	行为域的个数。	50
	单个行为域下业务线的个数。	10
	动作的个数。	100
	对象的个数。	100
	对象属性的个数。	300
数据萃取	行为的个数。	100
	工厂标签的个数。	200

模块	规格描述	默认值
	标签类目的个数。	50
	手工标签的个数。	300
	每条行为下规则的个数。	10

后续使用萃取项目的注意事项

在您使用数据萃取项目过程中,请勿删除萃取模块中的系统表,系统表如下表所示。如果您删除了系统表,则有可能导致任务运行失败。

表名	名称	描述
oi_behavior_detail_d	行为中心源数据接入	抽取源数据的行为大宽表。
oi_behavior_detail_d_sample	行为中心抽样表	oi_behavior_detail_d的抽样表,用 于行为看板查看数据。
oi_behavior_label_value_d	行为中心打标结果	行为打标的中间表。
oi_behavior_label_value_all	行为中心打标结果	行为打标的累积表。
oi_person_label_value_all	行为中心打标结果	行为中心打标结果的表。
oi_label_value_rule_d	标签打标规则表	标签打标的中间表。
oi_dim_behavior_id	行为定义维表	行为定义的表,从源数据同步而来。
oi_behavior_stat_d	行为数据统计天表	行为数据统计表。
oi_behavior_stat_sync	行为数据统计90天表	行为统计90天的数据。
oi_behavior_object_all	行为数据对象累积年表	行为对象,对象属性的累积表。
oi_behavior_property_value	行为数据属性表	对象值表,存储行为、对象以及对象 值的表,供您在标签打标时参考。
oi_statistics_behavior_detail_d	行为中心源数据统计表	包含行为属性的来源行为数据大宽 表。

5.平台管理

5.1. 创建并管理ID类目

为了更便捷高效地管理实体ID, Dataphin支持创建ID类目。ID类目包括一级ID类目和二级ID类目。一级ID类目 用于分类管理实体ID, 二级ID类目用于对一级ID类目更细粒度的划分。本文为您介绍如何创建并管理一级ID类 目和二级ID类目。

使用限制

- 系统支持最多创建20个一级ID类目。
- 所有层级的ID类目总数不能超过200个。
- Dataphin仅支持超级管理员和项目管理员新建、编辑和删除ID类目。

系统内置ID类目

Dataphin内置了其他ID类目,责任人默认为超级管理员,并且不支持编辑和删除。

☑ 平台管理 X						: 8
行为域 ID类目 标签类目						
ID类目列表		请输入类目行	名称或英文名	Q + 新建	ID类目	С
类目名称	ID数量	更新人	描述		ł	操作
	1	head-sector.			e v	
	0	Squid-to(b).			e d	Ī
•	1	lapeddorche.			e v	
	3	lapse behavior la su			e d	
其他ID others	1	Suj	系统类目, 不可	编辑	e d	Ť.

创建并管理一级ID类目

↓ 注意 系统支持最多创建20个一级ID类目。

1. 进入平台管理。

- 2. 在平台管理页面,单击ID类目页签。
- 3. 在ID类目页签,单击新建ID类目。
- 4. 在新建一级ID类目对话框,填写一级类目名称和一级类目英文名。

新建一级ID类目	C. Verandra ago a	×
* 一级类目名称	请填写类目名称	
* 一级类目英文名	请填写类目英文名	
描述	请填写描述	
	0/128	
	取消血痛症	
	取 消 确 定	

参数	描述		
一级类目名称	一级ID类目名称由中文、字母、数字、下划线(_)或短划线(-)组合组成。		
一级举日益文名	一级ID类目英文名由字母、数字、下划线(_)组合,且以字母开头。		
*****	↓ 注意 创建后不支持修改英文名。		
描述	填写对一级ID类目简单的描述。		

5. 单击**确定**。

对于已添加的一级ID类目,您还可以执行以下操作。

功能	操作步骤	
	i. 在I D类目列表 页面,单击操作列下的☑图标。 ⅲ. 在 编辑一级ID类目 对话框,修改 一级类目名称 和描述。	
编辑一级ID类目	注意 一级类目英文名不支持修改。	
	iii. 单击确定,完成一级ID类目名称的修改。	

功能	操作步骤
删除一级ID类目	 注意 系统仅支持删除没有二级ID类目和没有被实体ID引用的一级 类目。 删除一级ID类目并提交成功后,即可更新至生产环境,请您 谨慎操作。
	i. 在ID类目列表页面,单击操作列下的 <u>画</u> 图标。 ii. 在删除类目对话框,单击确定。

完成一级ID类目创建后,即可基于一级ID类目创建实体ID,请参见创建并管理实体ID。

创建并管理二级ID类目

↓ 注意 新建二级ID类目前,需要新建一级ID类目。

- 1. 请参见创建并管理一级ID类目,进入ID类目页签。
- 2. 在ID类目页签,单击需要创建二级ID类目的一级类目的操作列下的国图标。
- 3. 在新建二级ID类目对话框, 配置参数。

新建子ID类目		×
uperAdminiCal Supe		
* 所属一级类目	GU员会	
* 二级类目名称	请填写类目名称	
* 二级类目英文名	请填写类目英文名	
描述	请填写描述	
apel AdminiSu Sape	0/128	
uper AdminiSu Supe	现消 · · · · · · · · · · · · · · · · · · ·	
参数	描述	
所属一级类目	默认为所属的一级目录名称,不支持修改.	
二级类目名称	二级类目名称由中文、字母、数字、下划线(_)或短划线(-)组合组成。	

参数	描述	
	二级类目英文名由字母、数字、下划线(_)组合,且以字母开头。	
二级类目英文名	注意 创建后不支持修改英文名。	
描述	填写对二级ID类目的简单描述。	

4. 单击确定。

对于已添加的二级ID类目,您还可以执行以下操作。

功能	操作描述
编辑二级ID类目	 i. 在ID类目列表页面,单击一级目录前的●图标,展开当前一级ID类目 下的子级ID类目列表。 ii. 单击操作列下的☑图标,修改二级类目名称和描述。
	注意 二级类目英文名不支持修改。
	iii. 单击 确定 ,完成二级ID类目名称的修改。
删除二级ID类目	 ↓ 注意 ○ 系统仅支持删除没有被实体ID引用的子级ID类目。 ○ 删除子级ID类目并提交成功后,即可更新至生产环境,请您 谨慎操作。
	 i. 在ID类目列表页面,单击一级ID类目前的●图标,展开当前一级ID类目下的子级ID类目列表。 ii. 单击操作列下的直图标。 iii. 在删除子类目对话框,单击确定。

完成二级ID类目创建后,即可基于一级ID类目及其二级ID类目创建实体ID,请参见创建并管理实体ID。

5.2. 创建并管理标签类目

为了更便捷高效地管理标签,Dataphin支持创建标签类目。标签类目包括一级标签类目和子标签类目。一级标签类目用于分类管理标签,子标签类目用于对一级标签类目更细粒度的划分。本文为您介绍如何创建一级标签类目和子标签类目。

使用限制

- 最多创建50个一级标签类目。
- 所有层级的标签类目总量不超过500个。

- 创建子级标签类目的限制:
 - 当前, Dataphin仅支持创建5级标签类目。
 - 新建子标签类目前,至少需要新建一个一级标签类目。

Dataphin内置标签类目

Dataphin内置了其他标签类目,责任人默认为超级管理员,并且不支持编辑、删除及添加子标签类目。

◎ 平台管理	>								
行为域 IC	D类目 标签	类目							
		and states and a state of the state							
类目名称		更新人	总标签数	公开标签数	私有标签数	描述			
95	用 na	Santaniti.	1	1	0			Ę (Z
2	Hc ho	handshirth.	5	5	0			EM	Z
95	Wé Wé	fage for the state	4	4	0			Ę. (Z
• 🔀	تد te:	Superior (Se	4	4	0			Ę (Z
• 🔀	lfs lfs	Sectors: Su	52	50	2	saminiS ^a Suma sa	on (Su Supe	e, (Z
panna 🐹	其他标签 other_labels	(Su.	25	24	1	系统类目, 不	可		NE

创建并管理一级标签类目

- 1. 进入平台管理。
- 2. 按照下图指引,进入新建类目对话框。

☑ 平台管理 ×						: 8
行为域 ID类目 标签类目					2	
标签类目列表			请输	认类目名称或英文名 Q	+ 新建一级类目	С
类目名称	更新人 总	总标签数 公开	开标签数 私有	标签数 描述	損	宇
ma 用。	Namhdrindh. 1	1	0		e d	
Ha ho	5	5	0		e d	
wa wa	4	4	0		e d	

3. 在新建类目对话框,填写类目名称和类目英文名。

新建类目				×
* 类目名称	请填写类目名称			
* 类目英文名	请填写类目英文名		î	
描述	请填写描述			
			0/128	
		取消	确	定

参数	描述
类目名称	填写类目名称。类目名称的命名规则如下: • 只能包含中文、字母、数字、下划线(_)或短划线(-)。 • 全局唯一。 • 长度不能超过64字符。
类目英文名	填写类目英文名。类目创建后,其英文名无法修改。类目英文名的命名规则如下: • 只能包含字母、数字、下划线(_)。 • 以字母开头。 • 全局唯一。 • 长度不能超过64字符。
描述	填写对一级标签类目简单的描述。

4. 单击**确定**。

对于已添加的一级标签类目,您还可以执行以下操作。

功能	操作描述
编辑一级标签类目	i. 在标签类目列表页面,单击操作列下的应图标,修改类目名称和描述。 述。
	↓ 注意 类目英文名不支持修改。
	ii. 单击 确定 ,完成一级标签类目名称的修改。

功能	操作描述			
删除一级标签类目	 注意 o Dataphin仅支持删除没有被标签引用的一级标签类目。 o 删除一级标签类目并提交成功后,即可更新至生产环境,请您谨慎操作。 			
	i. 在 标签类目列表 页面,单击操作列下的 <u>画</u> 图标。 ii. 在 提示 对话框,单击 确定并提交 ,完成一级标签类目的删除。			

创建并管理子标签类目

↓ 注意

- 当前, Dataphin仅支持创建5级标签类目。
- 新建子标签类目前,需要新建一级标签类目。
- 1. 请参见创建并管理一级标签类目,进入标签类目页签。
- 2. 在标签类目页签,单击需要创建子标签类目的一级类目操作列下的国图标。

♀ 平台管理	х						: \$
行为域 ID类	长日 标签类日						
标签类目列表	表				请输入类目名称可	英文名 Q	+ 新建一级类目 C
类目名称		更新人	总标签数	公开标签数	私有标签数	描述	操作
apniny 💦	用,————————————————————————————————————	konteredu.	1	1	0		
20	Ho	Reprint and a second se	5	5	0		

3. 在新建类目对话框,配置参数。

新建类目					X
* 所属上级类目	服饰a				
* 类目名称	请填写类目名称				
* 类目英文名	请填写类目英文名	1			()
描述	请填写描述				
				0/128	
			取消		确定

参数	描述
类目名称	填写类目名称。类目名称的命名规则如下: • 只能包含中文、字母、数字、下划线(_)或短划线(-)。 • 全局唯一。 • 长度不能超过64字符。
类目英文名	填写类目英文名。类目创建后,其英文名无法修改。类目英文名的命名规则如下: • 只能包含字母、数字、下划线(_)。 • 以字母开头。 • 全局唯一。 • 长度不能超过64字符。
描述	填写对子标签类目简单的描述。

4. 单击**确定**。

对于已添加的子标签类目,您还可以执行以下操作。

功能	操作描述
----	------

功能	操作描述
编辑子标签类目	 i. 在标签类目列表页面,单击目录前的●图标,展开当前标签类目下的子级标签类目列表。 ii. 单击操作列下的应图标,修改类目名称和描述。 ↓注意 类目英文名不支持修改。
	iii. 单击 确定 ,完成子标签类目名称的修改。
删除子标签类目	↓ 注意 Dataphin仅支持删除没有被标签引用的子级标签类目。
	 i. 在标签类目列表页面,单击一级标签类目前的●图标,展开当前标 签类目下的子级标签类目列表。 ii. 单击操作列下的画图标。 iii. 在提示对话框,单击确定并提交,完成子标签类目的删除。

后续步骤

建议按照业务场景构建企业标签体系并创建对应的标签类目,再进行标签的开发和生产。创建标签,请参见步骤一:创建规则统计标签或创建注册上挂标签。

5.3. 创建行为域和业务线

行为域和业务线用于从业务场景视角对行为数据进行分类,帮助开发者更好的理解相关业务流程。后续您可 以基于已经定义的行为域和业务线创建行为,本文为您介绍如何创建行为域和业务线。

基本概念

本文涉及到相关概念的解释如下表所示。

概念名称	相关概念说明	使用说明
行为域	聚合业务含义一致的行为数据。例如,电商域、 文娱域。	用于定义业务分类。
业务线	基于行为域将行为数据进一步细分,行为域下的 各业务线相互独立。例如,淘宝业务线、天猫业 务线。	用于细分业务分类。

使用限制

Dataphin仅支持超级管理员和项目管理员创建行为域和业务线。

步骤一: 创建行为域

- 1. 进入平台管理。
- 2. 按照下图指引,进入新建行为域对话框。

Q 平台管理	×						: 8
行为域 ID类目	标签类目						
17为域&业务线				Q 请输入域或业务线谷	3	+ 新建行为	ы́ С
行为域		描述				2	操作
S zy zy zy							e d
业务线		描述					操作

3. 在新建行为域对话框,配置参数。

参数	描述
行为域英文名	行为域英文名的命名规则如下: 包含字母、数字或下划线(_)。 长度不能超过64个字符。
行为域名称	行为域名称的命名规则如下: 包含中文、数字、字母、下划线(_)或短划线(-)。 长度不能超过64个字符。
描述	填写对行为域的简单描述。

4. 单击**保存**,完成行为域的创建。

步骤二: 创建业务线

- 1. 进入平台管理。
- 2. 按照下图指引,进入新建业务线对话框。

176管理	×					: \$
行为域 ID类目	标签类目					
行为域&业务线				Q、请输入域或业务线名	+ 新建行为域	С
行为域		描述				操作
zy_电商 zy_test					2	I. C.

3. 在新建业务线对话框,配置参数。

参数	描述
行为域	选择业务线所属的行为域。

参数	描述
业务线英文名	业务线英文名的命名规则如下: • 包含字母、数字或下划线(_)。 • 长度不能超过64个字符。
业务线名称	业务线名称的命名规则如下: • 包含中文、数字、字母、下划线(_)或短划线(-)。 • 长度不能超过64个字符。
描述	填写对业务线的简单描述。

4. 单击保存,完成业务线的创建。

后续步骤

完成行为域和业务线创建后,即可创建行为元素,具体操作请参见<mark>创建行为元素。</mark>行为元素创建后,即可创 建行为规则,具体操作请参见<mark>创建并配置行为规则</mark>。

6.ID中心

6.1. 创建并管理实体ID

为了更精准、更全面地刻画实体对象,Dataphin支持从不同视角定义实体的描述ID。例如,定义消费者的实体ID为会员ID、手机号,定义商品的实体ID为商品名称、商品编码。本文为您介绍如何创建实体ID。

前提条件

完成实体ID类目的创建或确认Dataphin内已创建实体ID类目:

- 如果您是项目管理员或超级管理员在创建实体ID前需要完成ID类目的创建。具体操作,请参见创建并管理ID类目。
- 如果您是开发者在创建实体ID前需要确认ID类目已完成创建。

背景信息

实体ID是某个体系中相对唯一的编码,类似于身份标识,用于识别某个实体。在某一具体的事物中,实体ID 一般是不变的,例如员工工号、身份证号码、计算机网址。

操作步骤

- 1. 进入ID中心。
- 2. 进入ID中心后,系统默认进去实体ID页面。在实体ID页面,单击实体ID后的图图标。

≡	D	ataphin	・研发	萃取≠	运维	权限	Q 全局捜索		
Distill B ID	萃取数据中 Data_dis 中心	心 till 行为中心	 ♀∨ + 	欢迎使用D	ataphin	数据萃取!			
Ø	实体ID Q 输〉) 产品	关键字 計 (1)	ĊÐ	●行为中心 >>	行为元素	行为域	业务线	动作	
	→ 」 其代 → 」 「fs冽 → 」 『fs冽	lest1 (0) 則口 (1) 則甙 (3) ť (1)				\rightarrow	行为规则	- 54 - 540 ⁻⁶¹ -55	

您也可以通过以下方式,进入新建实体ID对话框:

○ 单击项目名称(Data_distill)后的 配图标后,选择ID中心 > 实体ID。

○ 单击左侧列表下方的实体ID对象列表。在萃取对象列表页下的实体ID页签, 单击新建ID。

	Dataphin	·研发	萃取≠ ì	运维 权限		Q、全局搜索	Ö.	ଚ ୧୦ 💽
Distill B		∾ ∨ ∎	 菜取对象列表页 动作 对象 对象 	× 就属性 行为规则	实体ID 规则统计标签	行为偏好标签 注册上挂标签	标签逻辑表 标签物	: 宮 四表 标签报告 ⁽¹⁾
10	冲心 行为中心	标签中心	实体ID		Saber Managori - Saber Man		2	+ 新建ID C
	头体D Q 输入关键字	O ⊞	Q 请输入ID名称e	成英				
	▶ ■ 企业ID (1)	l l	最近更新人	请选择	V ata	×		
	> ■ HAHAHAH22 (1) > ■ 会员ID (0)		ID名称	所属类目	最后更新人	最 近更新时间 描述		操作
	▶ 商品ID234 (0) ▶ ID类目 (4)		wendang wendan	HAHAHAH22	Rear Advecting and the	2021-03-16 12:04:11		区面
	▶ ■ 手机号码 (0)		leimu_2	ID类目	Nanteringanes	2021-03-15 15:46:21		
	▶ 具他ID (15) ▶ 测试dy123 (1)		笑 类目_1 LEIMU_3	ID类目-leimu	has been by a second	2021-03-15 15:44:50		
0	实体ID对象列表 1	> >	形式 xingshi	其他ID	Nantonchantes	2021-03-15 10:07:41 123		C ii

3. 在新建实体ID对话框,配置参数。

新建		Х
* ID名称	请输入ID名称	
* ID英文名	请输入ID英文名	(j)
	ID英文名将作为标签逻辑表的列名,默认存储为小写	
∗ 所属ID类目	其他ID	\sim
描述	请输入描述	
		- //
	取消 确定	

参数	描述				
ID名称	ID名称命名规则如下: • 包含中文、字母、数字、下划线(_)或短划线(-)。 • 不能超过64个字符。				
	实体ID创建成功后, ID英文名无法修改。命名规则如下: • 字母、数字、下划线(_)组合,且以字母开头。				
ID英文名	⑦ 说明 D英文名将作为标签逻辑表的列名,字母默认存储为小写字母。				
	• 已字母开头。• 不能超过64个字符。				

参数	描述
所属ID类目	选择实体ID所属的ID类目。 如果您还没有创建ID类目,则请您先完成业务相关ID类目的创建。如何创建ID类目, 请参见 <mark>创建并管理ID类目</mark> 。
描述	填写对实体ID的简单描述。

4. 单击确定,完成实体ID的创建。

对于已添加的实体ID,您还可以执行以下操作。

功能	操作描述
编辑实体ID	 i. 在实体ID页面的左侧列表区域,鼠标悬停至图标,单击编辑。 您也可以通过以下步骤,进入编辑对话框: a. 在实体ID页面的左侧列表区域,单击页面下方的实体ID对象列表。 b. 在萃取对象列表页的实体ID页签,单击操作列下的应图标。 ii. 在编辑对话框,修改ID名称、所属ID类目或描述。 ① 注意 D英文名不支持修改。 iii. 单击确定,完成实体ID的修改。
删除实体ID	 注意 系统仅支持删除没有被行为规则或标签引用的实体ID。 i. 在实体ID页面的左侧列表区域,鼠标悬停至 图标,单击删除。 您也可以通过以下步骤,进入提示对话框: a. 在实体ID页面的左侧列表区域,单击页面下方的实体ID对象列表。 b. 在萃取对象列表页的实体ID页签,单击操作列下的面图标。 ii. 在提示对话框,单击确定。

同时,您也可以按照下图操作指引,查看基于该实体ID生成的行为规则和标签逻辑表。

≡	Dataphin · 研发	萃取 ⇒ 运维 权限	Q、全局住在	凿 🖉 &
Data 1 学校355 B Data c ID中心 変体II マ マ マ マ マ マ マ マ マ マ マ マ マ	BHOD の 日本 日本 日本 日本 日本 日本 日本 日本	次迎使用Dataphin数据萃取! 8章和为是用DOEX、行为重要局线和KI和E #空节和心理和DEX、行为重要局线和KI和E #空节和心理和DEX、行为重要局线和KI和E #空节和心理和DEX、行为重要的是一般是一切用的 #空节和。		 基本信息 第二日 第三日 第三日 第二日 第二日
	o mobile			行为名称 规则ID
	e) 手机号 29 userid		🔒 行为中心	电商数据验证-xxh_浏选-商品 154
6	🖸 userid1			电度 ===- 数据验证-xxh_收藏-商品 227
6	3 用户名 ·			电商数编验证-xxh_收藏-商品 223
> 🖿 2	zhongying_id_leimu (0)	ID中心	12 m	电商其他数据验证-收量-商品2 269
		支持实体 印 的是以及 OweiD 种材质的引入,通过 ID mapping 实现性 一实在ID 别以用做把描述或Him 值		电离 動力-数据验证-收藏-商品 15
			4	电商。自己·其他数据验证·测试·多规则 426
			 ● 创建筑(な)D ⊡ 	电离_==-其他数据验证-测试-多规则 429
				电商_1==-数据验证-标签异常-多规则 470
				电商_曲边数描绘证-异常改藏-商品 598
		■ 行力中40 通过行力均識及行力和規則型义、实现对行力数据的规范化结构化聚合。 以交转标签 生产及用产雪等等场景		行法逻辑表
			 ● 定义行为元業 □ 	名称
\$3000	对象列表			logic_test3
	理 >			test_wuzhuo_1230

后续步骤

您可以基于实体ID、用户行为数据及标签数据创建行为规则、规则统计类标签、注册类上挂标签或标签逻辑 表。

6.2. 引入ID映射表及开启OneID归一化

ID映射表可以唯一识别实体对象的OneID为主键,记录了实体对象的多个实体ID之间的关联映射关系。您可以 将已经开发好的ID映射表引入至萃取模块,并开启OneID归一化功能,实现基于OneID聚合行为数据并计算标 签值的目的。本文为您介绍如何引入ID映射表及开启OneID归一化。

背景信息

OnelD即实体唯一标识。基于当前已有的、丰富的实体ID数据,通过HOB-GN算法模型等,计算各系统和域中的多个独立的ID之间的关联关系,识别出核心ID和非核心ID,识别并聚合映像为自然界中的唯一实体,赋予该实体一个唯一的OnelD,并最终产出ID映射表。

ID映射表记录了能唯一识别实体对象的OneID及其相关的多个实体ID之间的映射关系。通过这个映射关系,您可以基于某一个实体ID快速地查询到描述该实体对应的其他实体ID值,并将不同实体ID对应的所有行为数据关联起来,更全面地刻画实体。

例如,某用户的一个实体ID为Mobile(13900001234),该用户对应的OneID为001;OneID为001的用户具 有另一个实体ID为Email(username@example.com),基于该映射关系,您就可以轻松的查询到Mobile为 13900001234的用户的Email为username@example.com。ID映射表如下所示。

oneid	key_type	key_id
001	Mobile	13900001234
001	Email	username@example.com

前提条件

在开始执行操作前,请确认您已准备好ID映射表。您可以通过新建资源创建ID映射表或通过配置离线单条管道引入ID映射表至Dataphin。ID映射表必须包含oneid、key_type、key_id 3个字段,否则可能会导致实体ID无法归一。ID映射表的字段说明如下。

字段名	描述	参考值
oneid	唯一识别实体对象的ID值。根据OnelD找到 对应key_type值的key_id。	001、002、003等。
key_type	实体ID名称。	UserID、Mobile、Email等。
key_id	实体ID对应的取值。	13900001234、 username@example.com等。

下图为某ID映射表的数据。

key_type 🕎	key_id 🛛	oneid 🕎	ds 🛛
userid	1001	001	20220119
userid	1002	002	20220119
userid	1003	003	20220119
userid	1004	004	20220119
userid	1005	005	20220119
email	bochao1@com	001	20220119
email	ziud0212@com	002	20220119

使用限制

当前, 仅支持超级管理员及项目管理员角色的账号创建并管理ID映射表。

操作步骤

- 1. 进入ID中心。
- 2. 在ID中心,按照下图操作指引,进入ID映射表配置对话框。

≡	Dataphin · 研发	萃収 ≓ 运维 权限	Q、 全局複変	ස <i>ද</i> ද
	Data_destition QrV Image: Constraint of the constraint of t	次迎使用Dataphin数据萃取! 新品花为368年00-25、行为意味能和4564%差集 6名年1842年28年2月、结合中国新名。第1915年-6 48.8、美生,有名中国专家新闻986-64-8、13989 新品系。进行影响传。		
			日本 行为中心	🌘 标签中心
		● ID中心 見中学にには か加ええ ContO 時間時時時に入、構成 IO mapping 気況時 一気は2月 知以の現代的意味的情報	• effectio D	• 31A0 RATA E
		▲ 行为中心 進行方均能量及行力局能定义、实现对行为数编的规范化线性化(聚会、 以因为标签 生产及用产管核等容量	 主义行为元素 ① 	 ・ ・ はほ行为成员 ・ ・ ・
			_	0
	ID缺封表对象列表 >	● 标签中心		
ø	平台管理	文持基于学家/半台的场密目前化生产从原生标签引入,提供统一的标签 管理平台,并通过标签逻辑乘纯一对外提供服务		
_				

3. 在ID映射表配置对话框, 配置参数。

ID映射表配置				Х
* 选择ID映射表	请选择物理表 如选择的表加工方式不符合 ID {	到排表定义可能会解析失败或者	者 ID 无法归一, 请参考	V
选择映射表调度节点	请输入节点ID或名称 建议自行配置OnelD映射表产出	(任务对应的调度节点,若此处	未配置,可能无法正常依赖,影响	▽数据准确性
* 是否分区表	○ 是 ⑧ 否			
ID归—规则	预览			
包含	ID原始值 ① d_type,id_person_ids	ID_Type ① 如Tel	ID值 ① 如 abc	行为规则结构
	OnelD 如 0001	四一天系: & ID值 ID_Type 如 Tel	join on Key_id Key_id 如 abc	ID倒排表结构
· · · · · · · · · · · · · ·	一化功能开启后,对于归一化成功 测进行行为规则及标签值的计算; 1个UserID和1个mobile映射到那 1.0};只基于mobile算出来的结 : {美白:0.7,保湿0.3}	b的 ID (ID 倒排表能映射到对) 旧一化失败的 ID, 仅基于原始 司1个OneID, 则标签结果表会 课: {保湿: 1.0}; 基于OneID;	应的 ID_type和ID_value),将基于 剖D进行计算。 存储三条记录:只基于UserID算出 计算出来的结果(汇总UserID和mo	F原始ID和OneID 来的结果:{美 obile对应的行为数
			取消	提交

参数	描述
选择ID映射表	选择已准备好的ID映射表。Dataphin中所有项目中的生产环境物理表。
选择映射表调度节点	选择ID映射表的产出任务或者同步任务对应的节点,支持选择所有生产环境的脚本节 点(包括代码任务节点、同步节点和集成节点)。若此处未配置,可能无法正常依 赖,影响数据准确性。 ID映射表的依赖说明: • 配置了调度节点,则按照调度节点上挂依赖。 • 未配置调度节点(或异常情况下找不到已配置的调度节点),则Dataphin自动根 据 项目名.表名 自动解析调度节点,并上挂依赖解析到的调度节点。 • 未配置且自动解析失败,则不上挂依赖。
是否分区表	Dataphin根据您选择的ID映射表是否为分区表,自动配置,无需您手动配置。

参数	描述
分区字段	 分区字段说明: ○ 已选ID映射表是分区表,则展示分区字段配置的编辑区,默认填充 ds=\${bizda te} ,您也可以根据业务情况手动填写。 ○ 已选ID映射表不是分区表,则不展示。

4. 单击提交,完成ID映射表的创建。

对于已添加的ID映射表,您还可以执行以下操作。

≡	Dataphin · 研发						
Distill B	Distill 萃取数据中心 B Data_distill ♀∨ ■						
10	0中心	行为中心	标签	中心			
പ്പ	ID映射表	Ę		Ċ 🖻			
	Q 崳)						
▣	Data_distill.xxh_id_mapping			Ø:			
	🖻 Dat	a_distill.test_id_ma	p (已尹	☑ 编辑			
				直 删除			

功能	操作描述				
编辑ID映射表	 i. 在ID映射表页面的左侧列表区域,鼠标悬停至 图标,单击编辑。 ii. 在ID映射表配置对话框,修改选择映射表调度节点、是否分区表及 分区字段。 iii. 单击提交,完成ID映射表的修改。 				
	注意 系统仅支持删除没有被标签逻辑表引用的ID映射表。				
删除ID映射表	i.在ID映射表页面的左侧列表区域,鼠标悬停至,图标,单击删除。 ii.在提示对话框,单击确定并提交。				

开启ID映射表归一化

○ 注意

- 开启ID映射表归一化后,对全局所有标签计算生效。
- 开启之前标签值不会重新计算,如果需要针对历史分区的数据基于oneID重新计算,可以通过补数据实现。

开启后对于归一化成功的实体ID(即在ID映射表中能成功匹配成对的key_type和key_id),Dataphin将基于 该实体ID值和OneID值分别进行行为规则及标签值的计算。例如,1个UseID值和1个Mobile值映射到同一个 OneID值,则标签结果表会存储三条记录,包括只基于UserID值算出来的结果 美白:1.0 ;只基于Mobile值 算出来的结果 保湿:1.0 ;基于OneID值计算出来的结果(即汇总UserID值和Mobile值对应的行为数 据) 美白:0.7,保湿:0.3 。

1. 在ID映射表页面,单击页面底部的ID映射表对象列表。

≡	Dataphin · 研发	萃取〓	运维 权	限			Q、 全局拨案	- # 🖓 Q [_]
Distil	Data_distil ♀∨ ■	部取対象列表页	×	label_test ×				: 8
	and the second	动作对象对	象尾性行为顺	[3] 规则统计标签 行为庸好标注	签 注册上挂标签 案体ID	ID映射表		
	HO (7/140 (8/2440	ID映射表 OneID	旧一化配置说明				请输入ID映射表名称 Q	+ 引入ID鉄射泰 C
₽		映射表名称			是否分区表	久害人	开启OnelD归一化	· · · · · · · · · · · · · · · · · · ·
D	Q fasAA482≯	Data_distil	l.xxh_id_mapping		8	toendre boendersplatelies		
	Data_distill_xoh_id_mapping_ds	Data_distil	Looh_id_mapping_	ds	<u>R</u>	Association (Section (Section 1))		
		1						
	ID除封表对象列表 >							
Ø	平台管理							

在ID映射表对象列表页面,打开开启OneID归一化开关。
 开启后,萃取模块就可以基于ID映射表聚合行为数据并计算标签值。
7.行为中心

7.1. 创建行为元素

行为元素用于标准化定义和归类行为数据,同时也是后续创建行为规则的基础元素。行为元素包括动作、对 象和对象属性,本文为您介绍如何创建行为元素。

背景信息

行为元素间可以自由组合,不同的组合定义了不同的行为。行为由行为域、业务线、动作和对象组成。例如,顾客在电商领域的淘宝业务线中购买商品、顾客在电商领域的淘宝业务线浏览商品、顾客在电商领域的 天猫业务线收藏商品。

行为元素用于定义业务含义,帮助开发者更好的理解相关业务流程,同时用于统一数据规范以实现数据聚合。您可以通过定义行为元素对商品对象进行统一的定义,规避了后续标签无法计算的问题。

概念名称	相关概念说明	使用说明
步骤一:创 建行为域	聚合业务含义一致的行为数据。例如 <i>,</i> 电商域、 文娱域。	用于定义业务分类。
步骤二:创 建业务线	基于行为域将行为数据进一步细分,行为域下的 各业务线相互独立。例如,淘宝业务线、天猫业 务线。	用于细分业务分类。
动作	行为主体发出的操作。例如,购买、浏览。	用于定义用户行为中的动作。
对象	行为主体操作的具体事物。例如,商品、电影。	用于定义用户行为中的对象。
对象属性	对象的描述性信息。例如,名称、品牌、年份	用于定义对象的属性。后续对象属性可以作为标 签生产过程中来源行为数据的过滤条件。

行为涉及到相关概念的解释如下表所示。

使用限制

Dataphin仅支持超级管理员、项目管理员和开发者创建行为元素。更多权限信息,请参见数据萃取权限列表。

操作流程

操作步骤	描述
步骤一: 创建动作	定义了行为中行为主体发出的操作。例如购买、浏览。
步骤二:创建对象	定义了行为主体操作的具体事物。例如商品、电影。
步骤三: 创建对象属性	定义了行为中对象的描述性信息。例如名称、品牌、年份。

步骤一:创建动作

1. 进入行为中心。

2. 按照下图指引,进入新建动作对话框。

≡	D	ataphiı	℩ ∙研发	
Distill B	萃取数据中 Data_dis	迎込 till	⇔∨	Ð
IC)中心	行为中心	标签	中心
	动作		८⊡ ⊙	看板
	Q 输)	(关键字		
Θ	🔁 收薪			
÷	• 浏5	5		
	🔁 统计	t		
Q	日购到	Ę		

3. 在新建动作对话框, 配置参数。

参数	描述
动作英文名	动作英文名的命名规则如下: • 包含字母、数字或下划线(_)。 • 长度不能超过64个字符。
动作名称	动作名称的命名规则如下: • 包含中文、数字、字母、下划线(_)或短划线(-)。 • 长度不能超过64个字符。
描述	填写对动作的简单描述。

- 4. 单击**提交**。
- 5. 在提交备注对话框,填写备注信息,单击确定并提交,即可将动作提交至生产环境。

步骤二:创建对象

- 1. 进入行为中心。
- 2. 按照下图指引,进入新建对象对话框。

≡	D	ataphiı	n·研发	
Distill B	萃取数据中 Data_dis	سکہ ' till	\$∨	
IC)中心	行为中心	标签 2	中心
=	对象		¢⊡ o	看板
Θ	n (√ ⊯) ■} lfs	商品		
	₽a lfs_			
	唱人			
	いたい。 いたい いたい いたい に の に の に の に の に の に の に い の に の い の に の い の い	1		

3. 在**新建对象**对话框,配置参数。

参数	描述
对象英文名	对象英文名的命名规则如下: • 包含字母、数字或下划线(_)。 • 长度不能超过64个字符。
对象名称	对象名称的命名规则如下: • 包含中文、数字、字母、下划线(_)或短划线(-)。 • 长度不能超过64个字符。
描述	填写对对象的简单描述。

4. 单击**提交**。

5. 在**提交备注**对话框,填写备注信息后,单击确定并提交,即可将创建对象提交至生产环境。

步骤三: 创建对象属性

- 1. 进入行为中心。
- 2. 按照下图指引,进入新建对象属性对话框。



3. 在新建对象属性对话框, 配置参数。

参数	描述
对象属性英文名	对象属性英文名的命名规则如下: • 包含字母、数字或下划线(_)。 • 长度不能超过64个字符。
对象属性名称	对象属性名称的命名规则如下: • 包含中文、数字、字母、下划线(_) 或短划线(-)。 • 长度不能超过64个字符。
描述	填写对对象属性的简单描述。

4. 单击提交。

5. 在提交备注对话框,填写备注信息后,单击确定并提交,即可将对象属性提交至生产环境。

后续操作

完成行为元素定义后,即可基于行为元素创建行为规则。如何创建行为规则,请参见创建并配置行为规则。

7.2. 创建并配置行为规则

Dataphin支持将明细或轻度汇总的行为数据进行筛选和配置,生成行为规则。行为规则用于将原始数据进行规范化、结构化的聚集。后续行为规则可以用作标签生产的来源数据,简化了标签产出的流程,同时也可以用作下游营销投放用户圈选的来源信息。本文为您介绍如何创建和配置行为规则。

前提条件

在开始执行前,请确认以下信息:

• 已创建行为元素。具体操作,请参见创建行为元素。

- 已完成行为域和业务线的创建。具体操作,请参见创建行为域和业务线。
- 已创建实体ID,具体操作,请参见创建并管理实体ID。

注意事项

来源表必须是分区增量表,目前不支持非分区表和分区全量表。如果来源表为非分区表和分区全量表,则导 致行为中心产出表无法产出数据。

操作流程

操作	描述
步骤一:创建行为规则	配置行为规则的行为数据和业务来源表,以定义行为规则在业务来源表中的读数据的规 则。
步骤二:配置行为规则	配置行为规则即指定了从不同的来源表里面抽取行为数据的规则。同时,配置行为规则 的调度参数和行为中心产出表的存储生命周期。
步骤三:预览并提交行为 规则	您可以基于生产环境的行为规则,对特定行为的用户进行打标归类,从而实现用户数据 标签化管理。

步骤一: 创建行为规则

- 1. 进入行为中心。
- 2. 按照下图指引,进入新建行为规则对话框。

≡	Da	ataphi	n · 研发	Ì
Distill B	萃取数据中 Data_disti	心 ill	∾∨	Ð
IC)中心	行为中心	标签	中心
4	行为规则	I	d 🗄 🖸	看板
	Q 输入	关键字		
Ŷ	► ► lfs_F	电商-lfs_淘宝-购到	买-lfs_商品 (1)
	► b zy_F	电商-拼夕夕-浏览	-商品 (1)	
	► ▶ zy_⊧	电商-拼クタ-收藏	-商品 (1)	
N	► ▶ zy_₽	ーーーー 电商-拼夕夕-购买	-商品 (1)	
	→ 🕨 lfs_⊧	电商-lfs_淘宝-浏	览-测试下线(1)

3. 在新建行为规则对话框, 配置如下信息, 完成后单击确定。

i. 配置行为规则的基本信息。

新建行为规则				Х
基本信息 业务板块 萃取数据中心 空间类型 基础层	所属项目 数据萃取			
* 行为域&业务线 电商	> 淘宝 >			
* 动作 浏览	∨ * 対象 商品	×		
来源表设置	superhalten superhalten superh	6 ⁰⁰⁰⁰⁰⁰ 500 ⁰⁸	ad whether	
来源表类型 💿 物理表 🕢 逻辑表				
* 来源主表	✓ ① 缺少数据源?前往同步任务或代码任务加工			
*存储类型 🧿 増量 🔊 💿 全量				
erAun 筛选条件 ①				
 □ 格式化 □ 规范性校验 ② 参考示例 1 ds='\${bizdate}' 				
		54 547 - 5472	a.dooin(Su	0 X
		取消		确定

参数	说明
行为域和业务线	选择行为数据归属的行为域和业务线。行为域指聚合业务含义一致的行 为数据,例如电商域和文娱域。业务线用于分类管理行为域的行为数 据。例如淘宝业务线和天猫业务线。
动作	选择行为规则统计的行为动作。动作指行为主体发出的操作,例如购买和浏览。
对象	选择行为规则统计的对象。对象指行为主体操作的具体事物,例如商品 和电影。

行为规则是由行为(行为域、业务线、动作、对象)和来源表唯一确定。选择了行为元素组成行为 后,Dataphin会自动归属到这个行为的文件夹。后续配置标签时,选择了某个行为,Dataphin就把 这个行为下所有行为规则对应的来源表数据聚合到一起进行标签计算。

ii. 配置行为规则的数据来源。

新建行为规则	X
其木信自	
业务板块 萃取数据中心 空间类型 基	融层 所属项目 数据萃取
* 行为域&业务线 电商	→ 満宝 →
* 动作 浏览	◇ * 対象 商品 ◇
来源表设置	
来源表类型 💿 物理表 🔵 逻辑表	
* 来源主表	✓ ① 缺少数据源?前往 同步任务或代码任务 加工
*存储类型 💿 增量 🖉 全量	Super ed minites Super
" ^{Aunn} 筛选条件 ①	
A式化 C 规范性校验 ② 参考	示例
1 ds='\${bizdate}'	
	a M
	0 /
	取消 确定
参数	说明
	选择来源表的类型。Dataphin支持选择 物理表 或 逻辑表 。适用场景说 明如下:
	 物理表是非逻辑表(维度逻辑表、事实逻辑表或汇总逻辑)的表。
	如果需要基于物理数据表产出行为规则,则选择 物理表 。
	■ 逻辑表即Dataphin生成的维度逻辑表、事实逻辑表和汇总逻辑表。
	或采需安全了Datapini已的建的逻辑表(金度逻辑表、李奕逻辑表 或汇总逻辑)产出行为规则,则选择 逻辑表 。逻辑表名的格式说明
来源表类型	如下· ■ 维度逻辑表名格式为 dim 自定义英文名 。
	■ 事实逻辑表名格式为 fct_ 业务过程英文名_自定义名称 _di
	或 fct_ 业务过程英文名_自定义名称_ df 。
	■ 汇总逻辑表名格式为 dws_自定义英文名 。
	您需要提前创建事实逻辑表、维度表或汇总逻辑表:
	■ 如何创建事实逻辑表,请参见 <mark>创建并配置事实逻辑表</mark> 。
	■ 如何创建维度逻辑表,请参见 <mark>新建维度</mark> 。
	■ 如何创建汇忌逻辑表,请参见 <mark>新建汇总逻辑表</mark> 。

参数	说明				
来源主表	后续规则配置中需要从来源主表中选择字段。Dataphin支持选择跨项目的数据表(Prod项目和Dev项目的数据表)作为来源主表。 建议选择生产环境的数据表,以避免以下两种情况的发生: 避免行为规则参与调度时,因该数据表在生产环境不存在而导致调度失败。 避免配置行为规则调度参数时,自动解析依赖关系失败。 生产环境数据表名称的格式说明如下: \${Prod项目英文名}.物理表名 \${Basic项目英文名}.物理表名				
	■ \${Basic 业务板块英文名 }.逻辑表名 存储类型即来源表的存储类型。因Dataphin仅支持处理存储类型为增				
存储类型	量表的数据,因此来源表的存储类型仅支持选择 增量 。				
筛选条件	基于来源主表定义数据的筛选条件。例如,需要筛选业务日期为昨日的 数据,则筛选条件填写为 ds=\${bizdate} 。 如果选择生产环境的来源表,建议您在筛选条件时需要明确数据表的项 目或业板块,避免规范性校验不通过。 生产环境数据表名称的格式说明如下: \${Prod项目英文名}.物理表名 \${Prod业务板块英文名}.逻辑表名 \${Basic项目英文名}.物理表名 \${Basic项目英文名}.物理表名				

步骤二:配置行为规则

1. 完成步骤一: 创建行为规则后, 系统默认进入规则配置页面。在规则配置页面, 配置行为规则值。

规则配置				Q	请输入搜索关键字	Ó
行为规则信息	行为规则值		数据类型		操作	
* 行为主体 多ID配置说明					ID类型管理	+ 新增行为主体
D 契体ID 包含id_type,id_person_ids	ID_Type ID_value 坚果 ∨ : name		string			
* 对象 ① 至少填写1个,可选择表示对象ID或对	象名称的来源字段以唯一标识对象					
oid 对象标识 object_id	标识值 name		string			
		≫下拉展示更多				
* 对象属性 ①					对象属性管理	+ 新增对象属性
Attr 对象属性 object_properties	歴性		string			
行为属性 ① 用户行为的描述信息						
Cn 发生次数 stat_count	屬性價 ① id		double			Ť.
		≫下拉展示更多				
系統						
A ds	业务日期分区,格式yyyymmdd		string			

行为规则信息	行为规则值	数据类型
行为主体	 行为主体即进行某一行为的主体,例如购买商品行为中的行为主体就是客户。 Dataphin中行为主体由ID_Type和ID_value两个参数确定。 ID_Type:需配置为此前已创建的、行为主体对应的实体ID。如果您还没有实体ID,则需要创建实体ID。具体操作,请参见创建并管理实体ID。 ID_value:需配置为业务数据来源表中,行为主体对应的字段名。 在ID_value下拉列表中展示来源表的所有字段,选择某一字段为ID_Type对应的ID_value。 您可以通过新增行为主体新增多个实体ID,后续可以作为行为偏好标签、规则统计标签的主键,但当前标签仅支持单主键。因此如果您配置了多个实体ID,则Dataphin优先取第一个值不为空的实体ID进行存储,建议您优先配置主实体ID。 	Dataphin自动将已 选择 的 ID_value 字段 的数据类型转化为 String。

行为规则信息	行为规则值	数据类型
对象	对象即进行某一行为的对象,例如购买商品的行为中的对象就是商品。 Dataphin中对象由对象标识或对象名称参数确定。 • 对象标识:即对象ID。对象标识值需配置为业务数据来源表 中,对象ID对应的字段名。 标识值的下拉列表中展示来源表的所有字段。 • 对象名称:需配置为业务数据来源表中,对象名称对应的字段 名。 对象名称的下拉列表中展示来源表的所有字段。 添加对象标识值和对象名称的操作步骤如下: i. 在标识值下拉列表中,选择来源表中的某一字段作为对象的 ID。 ii. 单击对象名称右侧的 ◎图标后,在下拉列表中选择某一字段 作为对象名称。	Dataphin自动将已 选择的 标识值 和 对 象名称 字段的数据 类型转化为 String。
对象属性	 対象属性即对象的事实性描述信息,例如视频对象的名称、年份、 导演等。对象属性需要配置的参数包括属性和属性值: 。属性:选择已创建的对象属性。 如果您还没有对象属性,则需要创建对象属性。如何创建对象 属性,请参见步骤二:创建对象。 。属性值:指定来源表中属性的字段。 在属性下拉列表中展示来源表的所有字段,选择某一字段为属 性对应的属性值。 ↓注意 请尽量保持属性值(来源字段)的稳定性,避免 对统计类标签计算结果的影响。通常,同一个对象的同一属性 值稳定且唯一。 	Dataphin自动将已 选择的 属性值 字段 的数据类型转化为 String。

行为规则信息	行为规则值	数据类型
行为属性	 17万规则值 可选配置参数,通常来源表为事实逻辑表时,才需要配置行为属性参数。行为属性即用户行为的描述信息。 行为属性用于后续开发统计类标签的统计字段的权重分配依据。行为属性需要配置的参数包括发生次数、金额、时间和频次: 次生次数即行为的发生次数。 在发生次数后的属性值下拉列表中选择某一字段为发生次数的属性值。 ① 注意 如果您未设置发生次数字段,则系统将以每条行为数据记录为来源表的一条记录数据。 金额即行为对应的金额,例如购买商品行为中对应的支付金额。 在金额后的属性值下拉列表中选择某一字段为金额的属性值。 时间即行为发生的时间,例如购买商品行为发生的时间。 在时间后的属性值下拉列表中选择某一字段为时间的属性值。 频次即为行为中动作发生的频次,例如购买商品中购买发生的频次。 在频次后的属性值下拉列表中选择某一字段为频次的属性值。 	 Dataphin自动 将已选择的发生 次数、金额字 段的数据类型转 化为Double。 Dataphin自动 将已选择的时 间、频次字段 的数据类型转化 为String。

2. 在规则配置页面,单击页面上方的调度配置。在调度配置面板,配置参数。

i. 配置基本信息区域的参数。

基本信息		
节点名称	oi_behavior_detail_d_143	
U点芽		
节点类型	Maxcompute_SQL	
负责人	e — — — — — — — — — — — — — — — — — — —	
描述	行为规则143的调度配置	
优先级	中等优先级	
参数配置	对筛选条件部分填写的系统无法自动解析的参数进行赋值,以保证调度任务正常运行 节点参数配置说明	3

⑦ 说明 Dataphin自动生成名称、节点ID和节点类型的参数,不支持修改。

参数	描述
负责人	支持转交负责人。当任务运行报错时,Dataphin会及时通知到该任务的负责 人,便于您及时处理异常。 负责人其他的操作权限取决于负责人的角色。关于角色对应的操作权限的信息, 请参见数据萃取权限列表。
优先级	优先级定义了同一时间同一批待调度任务的优先级。系统默认选择为 中等优先 级。您可以根据业务需求修改任务调度的优先级。系统支持的优先级包括最低优 先级、低优先级、中等优先级、高优先级和最高优先级。
参数配置	参数配置用于定义逻辑表任务调度的参数。例如,定义了 \${bizdate}=20210610,则系统调度任务时业务日期为20210610。

ii. 配置调度配置区域的参数。

湖度配置 时间属性 ● 正常调度 调度周期 日∨ 0	○ 暫停调度 0:00 ③
参数	描述
时间属性	 选择任务在生产环境调度的时间属性。时间属性包括正常调度和空跑调度: 正常调度:按照调度周期的时间配置调度,并正常执行,通常任务默认选中该项。 暂停调度:即可暂停该任务及该任务的下游任务的调度,一旦调度到该任务会直接返回失败,不会执行。 适用场景说明如下: 正常调度:适用于任务需要正常调度以产出数据的场景。 暂停调度:适用于某个任务暂时不用执行,但后面还会继续使用的场景。
调度周期	 调度周期可选择日、周和月,您可以根据业务需要,指定任务运行的具体时间点。调度周期说明如下: 日调度,即调度任务每天自动运行一次。新建周期任务时,默认的时间周期为每天0点运行一次。 周调度,即调度任务每周的特定几天,在特定时间点自动运行一次。 月调度,即调度任务在每月的特定几天,在特定时间点自动运行一次。
	⑦ 说明 周调度和月调度中,如果您没有指定任务的运行日期,为保证 下游实例正常运行,Dataphin会每天生成实例后直接设置为运行成功,而 不会真正执行任何逻辑,也不会占用资源。

iii. 配置依赖关系区域的参数,并单击确定。

依赖关系 ① 依赖解析			
上游依赖 物理节点 逻辑表节点 ①			+ 新增上游依赖
父节点输出名称 节点名	节点ID(实例ID)	负责人	操作
oi_schedule_root oi_schedule_root_n	n_2	10000000000	Ť
ld_kongyiyewuba dim_gd01	n_;	Appendix of Appendix	Ť
当前节点			
輸出名称 节点名	节点ID (实例ID)	负责人	掘作
oi_behavior_detail oi_behavior_detail	282	Su	G
			取消 确定

参数	描述
	您可以通过自动解析和手动添加两种方式,为行为规则节点添加上游依赖的节 点:
	 单击依赖解析,依赖解析根据行为规则选择的来源主表类型进行解析上游节点:
	 如果来源主表的类型为物理表,则Dataphin仅支持自动解析出上游依赖的 物理节点。能成功解析出依赖的物理节点需要满以下两个条件:
	■ Dataphin内已有基于来源主表生成的周期任务。
	■ 周期任务的节点输出名称格式为 项目名.表名 。
	如果来源主表的类型为逻辑表,则Dataphin仅支持自动解析出上游依赖的 逻辑表节点。只要该逻辑表被引用,则会解析出来源逻辑表对应的逻辑表 节点。
	例如,来源表类型为 逻辑表 ,来源主表 为LD_kongyiyanwu****.DIM_GD01。
	行为规则信息
	*动作 浏览
	* XA象 Ifs_商品 V
	* 対象 Ifs_商品 V 来源表设置
	* 对象 Ib_商品 * 求源表设置 来源表关型 物理表 ● 逻辑表
	* 対象 lb_商品 ↓ * 対象 lb_商品 ↓ 来源表设置 来源表类型 ○ 物理表 ● 逻辑表 * 来源主表 LD_kongylyewdev ↓ LD_kongylyewDIM_GD01 ↓
	* 对象 Is_商品 ↓ * 対象 Is_商品 ↓ 来源表送置 * 求源主表 LD_kongylyew ↓ * 水源主表 LD_kongylyew ↓ ↓ K< 軟解析后, 上游依赖的逻辑表节点就是来源主表节点 ↓ (ld_kongyiyanwu****.dim_gd01) 。
	* 対象 is_mail ✓ * 対象 is_mail ✓ * 対象 is_mail ✓ * 就表表受置 ※ * 求源表送型 物理表 ⑧ 愛祖表 * 求源主表 LD_kongylyew _dev ✓ LD_kongylyew · 水源主表 LD_kongylyew _dev ✓ LD_kongylyew _DIM_GD01 · 水源主表 LD_kongylyew _dev ✓ LD_kongylyew _DIM_GD01 ✓ · 水源主表 LD_kongylyew _dev ✓ LD_kongylyew _DIM_GD01 ✓ · 水源主表 LD_kongylyanwu****.dim_gd01) . · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · ·
	* 対象 Is_商品 * 対象 Is_商品 * 対象 Is_商品 * 就要表送望 物理表 * 求源表类型 物理表 ● 逻辑表 * 求源主表 LD_kongylyew _ dev LD_kongylyew · 水源主表 LD_kongylyew · 水源主表 LD_kongylyew · 水源主表 LD_kongylyew · 小蘭 .DIM_GD01 · 水源主表 .DIM_GD01 · 水源主表 .DIM_GD01 · 水源主表 .DIM_GD01 · 水源主表 .DIM_GD01 · 小蘭
	* 対象 Is_BBB * 対象 Is_BBB * 就家表送留 * 速源表送型 物理表 ● 逻辑表 * 来源主表 LD_kongylyew

参数	■ 基于业务场景,如果需要添加其他节点作为当前行为规则节点的上游节点, 描述,则需要手动添加上游依赖的物理节点。
	↓ 注意 Dataphin不支持手动添加逻辑表节点。
	单击 新增上游依赖 ,在 新建上游依赖 对话框中,输入所依赖节点的输出名称 的关键字进行搜索节点,搜索到后单击 确定新增 。
当前节点	当前节点为您展示行为规则的数据输出节点。行为规则的输出节点名称固定为oi_behavior_detail_d,不支持修改。 如果行为规则已提交且被其他任务所依赖,则单击操作列下的③图标,可以查 看输出节点的下游节点。

3. 配置行为中心产出表的存储生命周期后,单击确定。

物理化配置		×		
 物理化配置 * 存储生命周期 快速选择 18 	30 365 3600 天			
* 选择分区字段 11 ds	behavior_id source_table			
备注:存储生命周期对全局行为规则生效,默认36000天,无最大限制,仅超级管理员可修改; 建议您设定的存储生命周期不小于下游标签选择的的来源数据时间跨度,否则标签数据可能出错 取消 确定				
参数	描述			
存储生命周期	存储生命周期用于定义行为中心产出表在系统的生命周期。默认36000天,无最大的制。 建议您设定的存储生命周期大于或等于下游标签设定的来源数据时间跨度,否则标 签数据可能出错。	限		
选择分区字段	Dataphin默认展示 ds、behavior_id 和 source_table ,不支持修改。			

⑦ 说明 如果用户配置了多个实体ID,每条记录仅存储第一个值不为空的实体ID的ID_value,且 值存储在person_ids字段中。

后续您也可以按照下图操作指引,修改来源表信息。

行为规则信息 调度配置 物理化配置 行为	规则版本		行为规则信息
2_8_x_test_basic_1car 行为规则ID:- 行为名称:	: lfs_test-下线-xxh_浏览-赛车 行为中心产出表:-		基本信息
行为规则 业务板块:萃取数据中心 马	空间类型:基础层 所属项目:数据萃取 行为域&业务线:Ifs_test-下线	动作 : xxh	业务板块 萃取数据中心
规则配置			空间关型 基础层
行为和即使自	行为规则通		所履项目 数据萃取
13 23/06/318-02	137370010		◆行为域&业务线 Ifs test ✓
* 行为主体 多时配置说明			- ・ 动作 如何の · · · · · · · · · · · · · · · · · · ·
空体ID 包含id_type,id_person_ids	ID_Type ID_value 邮箱 V : benz		 ▼対象 審主
			中海港沿海
* 対象 ① 至少填写1个,可选择表示对象ID或对	1歲名称的来源字段以唯一标识对象		
Oid 对象标识 object_id	标识值 porsche		
		※下拉屋示面	
No. Ett. O		1 12000 100	
* 刈擦鳩性 ①			(高达张件 U) I ds='\${bizdate}'
对象压性 object_properties	IIII性 IIIII IIIIIIIIIIIIIIIIIIIIIIIIII		
17万属1主 (1) 用户行为的描述信息			
Cn 发生次数 stat_count	犀性值 ① audi		
			1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

切换来源表之前请您确认以下信息:

如果行为规则对应的节点已经在生产环境调度并生成数据,需要通过代码任务或即席查询,清空行为
 中心表的历史分区相关数据,以保证下游标签运算不受历史数据影响

在Data_distill项目中,行为中心产出表oi_statistics_behavior_detail_d、oi_behavior_detail_d有3个 分区,分别是ds(业务日期),source_table(来源表),behavior_id(行为ID),两个表均需要手 动清除历史分区的数据。下文代码示例以oi_behavior_detail_d为例进行介绍。

```
#1. 执行以下命令,查看物理表的分区。
show partitions oi_behavior_detail_d;
#回显例如:
ds=20191020/source_table=data_distill.data_distill_movie_info/behavior_id=20
ds=20191021/source_table=data_distill.data_distill_movie_info/behavior_id=20
#2. 执行以下命令,删除该规则相关的所有分区。
ALTER TABLE oi_behavior_detail_d
DROP IF EXISTS PARTITION(ds = '请输入业务日期', source_table ='请输入来源表名称', behavio
r_id = '请输入行为ID');
#3. 执行以下命令,查看是否有遗留分区。如果有遗留分区,重复上述两步骤,直至删除完所有分区。
show partitions oi_behavior_detail_d;
```

 切换来源表后,原有来源字段及计算逻辑可能失效,且调度配置部分已解析的上游依赖可能被清空, 需重新选择字段并配置上游依赖。

步骤三:预览并提交行为规则

1. 在规则配置页面,按照下图操作指引,预览已配置的行为规则数据。

行为规则信息 调度配置 物理4	化配置 行为规则版本					ť	- E e 🕫 📀
							1
C C C C Data_distill.behavio 行为规则D: 445 业务板块: 華取数	or_film_01 行为名称: 电离数据验证- 雷中心 空间类型: 基础层 所)	 行为中心产出表: oi 重项目: 数据萃取 行为域&业务线: 	_statistics_behavior_detail_d 电商数据验证 动作: #	金不同 対象:多规则			Ø∃开启ID归一化
规则配置					Q	请输入搜索关键字	
行为规则信息		行为规则值			数据类型	操作	
* 行为主体 多ID配置说明						ID类型管理	+ 新增行为主体
□D 实体ID 包含id_type,id_person_i	ID_Type ds userid	D_value ∨ : id			string		
★ 対象 ① 至少填写1个,可选择表	示对象ID或对象名称的来源字段以唯一被	词形对象					
vp起 ママ 対象标识 vp起 マロ object_id	标识值 films_name		v	2	string		
数据预范							×
person_ids	object_id	object_properties	behavior_properties	ds	source_table	behavior_id	
{"userid":["1001"]}	八	("视角":["2D"],"票房":["311076"]}	00200202020123	20220123	Data_distill.behavior_film_01	450	
{"userid":["1002"]}	我	{"视角"`["2D"],"重房"`["282972"]}	00190202020123	20220123	Data_distill.behavior_film_01	450	
{"userid":["1003"]}	∰	{"视角":["3D"],"震房":["160996"]}	001902020123	20220123	Data_distill.behavior_film_01	450	
{"userid":["1004"]}	±	{"视角":["3D"],"票房":["112254"]}	008020220123	20220123	Data_distill.behavior_film_01	450	
{"userid":["1005"]}		{"视角":["2D"],"票房":["83649"]}	00110202020123	20220123	Data_distill.behavior_film_01	450	
{"userid":["1006"]}	拆卸==。	{"视角":["3D"],"票房":["60215"]}	00170202020123	20220123	Data_distill.behavior_film_01	450	

2. 按照下图操作指引,提交已创建的行为规则。

				_
行为规则信息 调度配置 物理化配置 行为规	规则版本			
			•	
2.8_x_test [lfs_test-下线-xxt_浏览-覆车 行为中心产出质:- 间类型:基础层 所属项目:数度萃取 行为城&业务线: lfs_test ———————————————————————————————————	动作:xon_测宽 对象:赛车	建交管注 2 诺瑜入此次继交的相关指注说明,如变更信息	Ł
抑则配署				
MUCHUE.			0/128 /	
行为规则信息	行为规则值		器 提交前请确认您已经点击过调度配置的自动解析按钮,否则逻辑表依赖可能 可达。14	
* 行为主体 多ID配 置说明			だよ上生 項写着注以说明此次提交的交更内容 当地项目、現本成功に、即再新客中产环境、法密使退作()	
实体ID 包含id_type,id_person_ids	ID_Type ID_value 俞晓敏 V : benz		st	
* 対象 ① 至少填写1个,可选择表示对象ID或对象	象名称的来源字段以唯一标识对象			
Oid 对象标识	标识值 porsche		string	
	Foreign (1997)			
		≫下拉展示更多		
★ 対象属性 ^①			对象属性管理 + 新增对象属性	
Attr 对象厘性	属性 属性值		string	
- object_propetities	menu · polacite			
行为属性 ① 用户行为的描述信息				
发生次数	屬性值 ①			
stat_count	audi		double	

您也可以在菜单树直接对草稿、开发中、已提交状态的行为规则进行管理。

≡	Dataphin · 研发								
Distill B	萃取数据 Data_dis	え till	∾∨	Ð					
IC)中心	行为中心	标签中心						
	行为规	U (Ċ 🖻 🕒	看板					
	Q 輸)	∖关键字							
\heartsuit	- • 电	6_ 		(2) 🚺					
	Data_distill.behavior_								
	Q	Data_distill.behavio	☑ 编辑						
Q	⊢	_test-	□ 克隆						
	⊢ Ifs_	_test-	🕑 行为规	则任务					
	▶ ▶ 电	商的 	♪ 补数据						
		的c	⊎ 下线						
	ير الم	≭ Data_distill.ui_beha	靣 下线并	删除					

执行**下线**或**下线并删除**行为规则操作时,如果行为规则对应的节点已经在生产环境调度并生成数据, 需要通过代码任务或即席查询,清空行为中心表的历史分区相关数据,以保证下游标签运算不受历史数 据影响。

在Data_distill项目中,行为中心产出表oi_statistics_behavior_detail_d、oi_behavior_detail_d有3个分区,分别是ds(业务日期),source_table(来源表),behavior_id(行为ID),两个表均需要手动清除历史分区的数据。下文代码示例以oi_behavior_detail_d为例进行介绍。

#1. 执行以下命令,查看物理表的分区。
show partitions oi_behavior_detail_d;
#回显例如:
ds=20191020/source_table=data_distill.data_distill_movie_info/behavior_id=20
ds=20191021/source_table=data_distill.data_distill_movie_info/behavior_id=20
#2. 执行以下命令,删除该规则相关的所有分区。
ALTER TABLE oi_behavior_detail_d
DROP IF EXISTS PARTITION(ds = '请输入业务日期', source_table ='请输入来源表名称', behavior_
id = '请输入行为ID');
#3. 执行以下命令,查看是否有遗留分区。如果有遗留分区,重复上述两步骤,直至删除完所有分区。
show partitions oi behavior detail d;

后续步骤

完成行为规则创建后,即可基于行为规则创建规则统计标签和行为偏好标签:

- 基于行为规则数据圈创建规则统计标签。具体操作请参见步骤一:创建规则统计标签。
- 基于行为规则数据创建行为偏好标签。具体操作请参见步骤一:创建行为偏好标签。

7.3. 查看行为数据及下载行为定义表

行为看板用于可视化展示行为规则产出表(oi_behavior_detail_d)中的数据,同时支持下载行为定义表。 当您完成行为元素的定义后,您可以通过行为看板查看行为元素的分布。本文为您介绍如何查看行为数据及 下载行为定义表。

背景信息

本文涉及到的术语解释,请参见基本概念。

查看行为数据

- 1. 进入数据萃取。
- 2. 在数据萃取页面,单击左侧区域的行为中心页签。
- 3. 在行为中心页面,单击行为看板。
- 4. 在行为看板页面,查看统计数据和行为抽样数据。

行为看板			统计数据最近更新时间: 2021-03-26 10:35:3
行为中心产出表: ol_statistics_behavior_detail_d	已接入行为规则: 14 最近90天累计采集用户行为: 1,508	1	
显示比例: 100% + - 重豐	androllow Surper public dellow	2 山 下銀行为定义表 行为遗	u鼎抽样:所有行为 3
 可点击单个域、业务线、动作对象名称或图标以 可hover单个域、业务线、动作对象图标荐取描述 	查看详情 Y信息	抽样业	务日期: 2021-03-25 高小
域 ① 点击单个城查看详情	业务线	动作对象): IPAddress:17109872784
			1) AN 2010 (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)
		浏览商品): UserID:6
			が行対象: 浏览商品
		查看人	380세1는: { 베리머와린曰 :[속이용], 베리머인(1日 :[429.0])
投资_1	a state and		D: UserID:4
		査査商品 ア	所行政論: 浏览商品 1線居性: ("商品类目":["稽句"], "商品价格":["4999.0"]}
金融>			
	收藏商品): UsenD:1002)/F对象: 查看商品
		7	象 居性:{"商品价格":["429.0"]}
		购买商品	1 Juni 19 (1993
	理财 1	対策人): UsenD: 1003)作対象: 查看商品
Service Se	and a second statement of the second seco	χ.	a)鼎性: ("商品价格":["429.0"]]
and and a set	纳税_1	统计商品	ן. בימדור 20
		0.8) 「100-112)作[2]象: 统计商品
	supervised states supervised states	2	#象屬性: {"性别":{"男"]}
	浏览版 商品	I	D: UserID:王老二
	TADDAN _ HANN		2月17日 - 2月19日 - 2月1900 - 2月1000 - 2月1000 - 2月10000 - 2月1000000000000000000000000000000000000

行为看板页面右上角,为您展示当前页面展示的统计数据最近的更新时间。

区域	描述
0	 。展示行为看板的统计数据来源于oi_statistics_behavior_detail_d 。展示已接入行为看板的行为规则数量。 。展示近90天累计采集用户行为数量。
	从左到右依次为行为域、业务线和动作对象。同时您可以显示比例区域调整下方图 标展示的比例。 • 鼠标悬停至 域 下方的柱,查看使用某个行为域的 行为数 及该行为数占所有域行为 总数百分比。 选择某个行为域后,页面仅展示该行为域下的所有业务线和所有的动作对象的数 据。
	 鼠标悬停至业务线下方的柱,查看使用某个业务线的行为数及该行为数占所有域行为总数百分比。 选择某个行为域下的业务线后,页面右侧为您展示该业务线下的所有行为数据抽样的结果。

数据萃取·行为中心

國域		AREUと登録学情 1929年2月	占 下载行力定义表 行为 前年3	数据抽样: 金融 - 投资_1 295日期: 2021-02-24
		1046210-0-		D: IPAddress
	XX ① 杰士苹小SZEE1499	129936	NUTEX18R	动作对象:购买商品
				对象履性: ("商品英目":["动麦"])
			浏览商品	
				ID: UserID:6
				动作对象: 刘克商品
			主要を	对象层性: ("商品英目":["零食"],"商品价格":["429.0"])
		投资_1		D: UserID:5
			适营商品	2011月8日: 月2月1日 11年日日: 「第三天日」「第三人体」(12000.011)
	金融			CHARGE C A MANAGEM C AN COLD MODERATINE C ADDR. O II
				D: UserID:5
			收藏商品	10. 00000000000000000000000000000000000
				对象屈性: ("简品英目":["積包"],"简品价格":["2000.0"])
			购买商品	
				ID: UserID:1005
		理财_1	浏览人	动作对象:查看商品
				对象屈性: ("商品价格":["2000.0"])
		納稅_1	统计商品	
	二日本 1 401 F 人 1 三元比例: 100% (一度目 1) 可点由単小域、业务域、の7078名称 2) 可からの単小域、生活成、2017年3月3日		د ر الال م ارية دم الرود دم الرود. <u>د ک</u> هر م شرکته	- メヘ JiLi JLi IT o 行力数据抽样: 金融 - 投資 1 - 购买商品 抽样业务日期: 2021-03-25 ID: IPAddress
	Let			? 动作对象:购买商品
	域	11/19/132 动作对1	家 対象属性	对象屈住:("商品类目":("積包"))
				ID: IPAddress:
		浏览商品		ID: IPAddress: 动作对象: 购买商品
		浏览商品		 ID: IPAddress: 动作对象:购买商品 对象届性:("商品类目":("动漫");
		河道南岳		ID: IPAddress: 动作对象: 购天商品 对象履性: ("商品类目"("动爱")
		前流電品		ID: IPAddress: 动作对象: 购买需品 对象层性:("南品英目-("动み")) ID: IPAddress
		刘治南县		10: IPAddess: 前行対象: 時天海岳 対象環境: (南級英目*(1功度1) ID: IPAddress 功作対象: 時天海岳
		刘凤南岳 帝事人		U2: IPAddess: が作対象: 购天商品 対象層性:(荷根英目*(切象引) U2: IPAddess 均作対象: 购天商品 対合語:(「夜風英目*(本員1)
		的风寒品. 章语人		 D: IPAdres: D: IPAdres: D: State: (「電風供用」:切倒口) (D: IPAdres: (D: IPAd
		東京電告 金融人		 D: IPAdres: D: IPAdres: N支索法:(南京法(市政委)) IPAdres: N方法: 和支索法 N支索法:(東京法) D: IPAdres: D: IPAdres:
		如流增压 登着人 段度_1		 D: IPAdres: D: IPAdres: 가송지상: (靑風英目:(功濟) D: IPAdres: D: IPAdres: 가가장: 정국표표 가송지상: (靑風英目:(下裏口) D: IPAdres: D: IPAdres: D: IPAdres: D: IPAdres: D: IPAdres: D: IPAdres:
		和成增益 意意人 政策_1		 D: IPAdres: D: IPAdres: 対象相性:(電磁共同:(功能) ID: IPAdres: 以下対点:原気構成 ID: IPAdres: U: IPAdres:
		東流電法 重要人 段度_1 重要項品		 D: IPAdres: D: IPAdres: 助(7)法, 秋天委長 対象周後:(南島英昌(下京貴)) D: IPAdres: 功(7)法, 秋天委長 対象周後:(南島英昌)(下書力) D: IPAdres: 助(7)法, 秋天委長 対象周後:(南島英昌)(下書力)
	±10	和成年品 章要人 段变_1 重要有品		 D: IPAdres: D: IPAdres: 対象相性:(電磁共同:(功能)) D: IPAdres: 初かざ油: 所支商局 ジング目:(電磁共同:(家県)) D: IPAdres: 初かざ油: 所支商局 ジン型:(電磁共同:(零音)) D: IPAdres: ジン型:(電磁共同:(零音)) D: IPAdres:
	金融	東京長 現美_1 王章東品		 D: IPAddes: D: IPAddes: 정, 제품: (학교, 여리, 가 3종) D: IPAddes: 정, 가 2종: (학교, 여리, 각종) D: IPAddes: 정, 가 3종: (학교, 여리, 각종) D: IPAddes: D: IPAddes: D: IPAddes:
	±20	東京考査		 D: IPAdres:
	±20	和成有品 服用人 服用人	東級共民	 D: IPAdres: D: IPAdres: (●面供用:「●面供用:「
	±8	東京第長 東京人 安定1 安定5日 の原売品	東政府日	 D: IPAdres: D: IPAdres: 如: IPAdres: 可念面性:(電面供目(可高面)) D: IPAdres: 功:方法:防天委员 对余语性:(電面供目(可高面)) D: IPAdres: 动:方法:防天委员 对余语性:(電面供目(可容)) D: IPAdres: 动:方法:防天委员 动:方法:防天委员 对余语性:(電面供目(了容))
	±8	東京氏	東西興日	 D: IPAdres: D: IPAdres: D: IPAdres: R: 「衛島県日"(功楽) ID: IPAdres: R: R: R
	±10	東京第長 東京 1.東京 日 東京 長 東京 東京 長 東京 東京 東京 東京 東京 東京 東京 東京 東京 東京 東京 東京 東京	東西府目	 D: IPAdres:
	±8	東京派	東島供信	 D: IPAdres:
	金融	東京第長 重要人 設置1 変変電品 の面電品 取実発品	東政府目	 D: IPAdres:
	28	東京長	第五件目	 D: IPAddes:
		 2月前前 2月前 3月前 3 3 3 3 3 4 4<!--</td--><td></td><td> D: IPAdres: D: IPA</td>		 D: IPAdres: D: IPA
	◎●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●	2008年5 2011 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018日 2018 2018 2018 2018 2018 2018 2018 2018	^{東国展目} 昨日的抽样数据,您	0.1 PAddes: 0.1734/BE: ("商品與目"("助賣") 0.1 (PAddes: 0.1
3	■■ 展示右侧选中的行	(2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018) (2018)	昨日的抽样数据,您	 D: PAdres: D: PAdres: D: PAdres: PAdres: PAdres
3	▲ 一 展示右侧选中的行 个月的抽样业务日	REREA REC_1 EREA CEREA REREA REREA REREA DO MH样数据。默认展示I 明。	^{東國共同} 昨日的抽样数据,您	 D: PAddes: D: PAddes: D: PAddes: PARE: (電風母()(30()) D: PAddes:

下载行为定义表

在行为看板页面,单击下载行为定义表,即可下载行为中心的所有行为元素。



8.标签中心

8.1. 创建规则统计标签

8.1.1. 场景引导

在您开始创建规则统计标签前,需要了解创建规则统计标签的需求场景,以引导您快速完成规则统计类标签 的创建。本文为您介绍创建规则类统计标签的场景,及各场景下创建标签的操作引导。

场景说明

规则统计标签支持基于用户行为数据筛选特定用户,并对用户添加标签。同时支持基于已有标签生产衍生标 签。典型的应用场景说明如下:

• 场景1: 基于用户行为数据, 筛选出需要添加标签的用户, 并对用户自定义标签

例如,基于最近30天内购买商品、收藏商品和浏览商品的行为数据,筛选出购买、收藏和浏览商品的种类 为动漫的用户,并指定用户的标签为文娱。

场景2:基于用户行为数据,筛选需要添加标签的用户,并配置用户行为属性的统计值为用户的标
 签

例如,基于最近30天内购买商品、收藏商品和浏览商品的行为数据,筛选出购买、收藏和浏览商品的价格 为100~800元,且商品的种类为女装的用户,并配置商品价格的平均值为用户的标签。

• 场景3:基于已打过标签的对象,筛选出需要再次添加标签的对象,并为对象自定义新的标签

例如,最近30天内注册商品分类的标签包括女装、坚果、汽车、笔记本,筛选出已打上女装、坚果的商品,再为这些商品指定新的标签为生活用品。

操作引导

各场景下创建规则类统计标签的流程相同。操作引导如下:

- 1. 步骤一: 创建规则统计标签
- 2. 步骤二: 配置并提交规则统计标签

8.1.2. 步骤一: 创建规则统计标签

您可以对同一类数据进行打标归类,从而实现数据分类管理(标签化管理)。例如,对键盘和显示器打标归 类为办公用品。本文为您介绍如何创建规则统计标签。

前提条件

在开始创建规则统计标签前,请确认您已满足以下条件:

- 已完成实体ID的创建,具体操作请参见创建并管理实体ID。
- 如果需要基于行为数据创建规则统计标签,则需要完成行为规则的创建,具体操作请参见创建并配置行为规则。

注意事项

在您开始创建规则统计类标签前,需要了解创建规则统计标签的场景,以引导您快速完成规则统计类标签的 创建。创建规则统计标签的场景说明,请参见场景引导。

操作步骤

- 1. 进入标签中心。
- 2. 默认进入规则统计标签页面。单击规则统计标签后的图图标。
- 3. 在新建标签对话框,配置参数后,单击确定。

新建标签						×
★ 标签英文名	请填写标签英文名		(j)	* 标签名称	请填写标签名称	
	标签英文名将作为标签	逻辑表的列名,默认	人存储为/	19		
描述	请填写描述					0//00
						0/128
* 所属类目	请选择标签类目			\sim	()	
* 公开状态	公开			\sim		
* 标签类型	规则统计					
* 打标方式	指定枚举值			\sim	(i)	
	您可以基于生产环境的	的行为或标签数据圈	l选需要打	T标的实体ID,	并直接指定标签值,	创建后不可修改
* 结果表名	Data_Distill_ 请填写	结果表名				
					取消	确定

参数	描述					
	填写标签英文名。创建标签后,其英文名无法修改。命名规则如下: • 只能包含字母、数字或下划线(_)组合组成。					
标签英文名	⑦ 说明 标签英文名将作为标签逻辑表的列名,字母均默认存储为小写字母。					
	○ 项目空间内唯一。○ 长度不能超过64字符。					
标签名称	填写标签名称。命名规则如下: 。 只能包含中文、字母、数字、下划线(_)、短划线(-)组合组成。 。 项目空间内唯一。 。 长度不能超过64字符。					
描述	填写对标签的简单描述。					

参数	描述
所属类目	即标签的类目,选择此前已创建的标签类目。
公开状态	即标签的状态,用于定义项目空间内成员是否具有查看该标签的权限。 公开状态包括 公开 和隐藏: • 公开状态的标签: Data_distill项目空间内所有成员都可以查看。 • 隐藏状态的标签: Data_distill项目空间内仅项目管理员和标签负责人可以查看。 更多权限信息,请参见用户角色和权限。
标签类型	默认为 规则统计 ,不支持修改。
打标方式	 即为待打标对象配置标签值的方式。打标方式包括指定枚举值和指定统计值。详细 说明如下: 指定枚举值:自定义待打标对象的标签值。适用于基于行为数据或生产环境已有 标签圈选打标对象的场景。 如果您的业务场景是下述举例的标签开发场景,建议您选择指定枚举值。 场景1:基于用户行为数据,筛选出需要打标的用户,并自定义用户的标、场景 3:基于已打过标签的数据,筛选出需要再次打标的对象,并自定义对象新的标 签。 指定统计值:基于已配置的行为属性进行统计计算,且统计值作为待打标对象的 标签值。适用于基于行为数据圈选打标对象的场景。 如果您的业务场景是下述举例的标签开发场景,建议您选择指定统计值。 场景2:基于用户行为数据,筛选出需要打标的用户,并配置行为属性的统计值为 用户的标签 注意 标签创建后,打标方式不支持修改。
结果表名	 填写标签产出数据写入的表名。命名规则如下: 只能包含字母、数字、下划线(_)组合组成。 项目空间内唯一。 长度不能超过64字符。 ✓ 注意 规则统计标签提交后,其名称无法修改。 建议结果表名命名为标签名称,便于其他节点通过搜索结果表名依赖该结果表。

后续步骤

完成上述基础信息的配置后,即可进行打标规则和标签值的配置。具体操作,请参见<mark>步骤二:配置并提交规</mark>则统计标签。

8.1.3. 步骤二: 配置并提交规则统计标签

完成规则统计标签基本信息配置后,您需要为规则统计标签配置数据来源、标签值,及配置规则统计标签的 调度周期和存储生命周期,后续可以测试规则统计标签并查看标签报告。本文为您介绍如何配置、测试规则 统计标签、查看标签报告及提交规则统计标签至生产环境。

前提条件

如果您的业务场景为场景3:基于已打过标签的数据,筛选出需要再次打标的对象,并自定义对象新的标 签,则在您配置规则统计标签前需要完成标签的创建。标签包括规则统计标签或注册上挂标签:

- 如何创建规则统计标签,请参见步骤一:创建规则统计标签。
- 如何创建注册上挂类标签,请参见创建注册上挂标签。

操作流程

- Step 1: 配置规则统计标签的取数逻辑及标签值的计算规则
- Step 2: 配置规则统计标签的调度参数及标签输出表的存储生命周期
- (可选) Step 3: 测试规则统计标签并查看标签报告
- Step 4: 提交规则统计标签至生产环境

配置规则统计标签的取数逻辑及标签值的计算规则

- 完成步骤一:创建规则统计标签后,默认进入新建标签页面。在新建标签页面,配置筛选待打标对象的规则,并为待打标对象配置标签值。
 - i. 在STEP 1区域,选择筛选来源数据的方式。
 圈选方式指Dataphin筛选待打标对象的方式,包括按行为圈选和按标签圈选:
 - 按行为圈选:基于用户行为数据,筛选需要打标的用户。

如果业务场景是下述举例的标签开发场景,建议您选择按行为圈选。

场景1:基于用户行为数据,筛选出需要打标的用户,并自定义用户的标签、场景2:基于用户行为数据,筛选 出需要打标的用户,并配置行为属性的统计值为用户的标签。

- 按标签圈选:基于已打过标签的数据,筛选需要再次打标的对象。 如果业务场景是下述举例的标签开发场景,建议您选择按标签圈选。 场景说明。
- ii. 在STEP 1区域,单击开始配置。
- iii. 在新建标签的页面右侧的数据来源配置页签,配置来源数据后,单击下一步。
 不同场景的数据来源配置参数不同,说明如下:
 - 场景1:基于用户行为数据,筛选出需要打标的用户,并自定义用户的标签、场景2:基于用户行 为数据,筛选出需要打标的用户,并配置行为属性的统计值为用户的标签。

	en gopen	 数据来源配 國选用户行为 	置 1数据						标签值配置 基于圈选的用户行法	为数据定义标签值		
数据考	末 源配置											
圈选7	方式筛选 按行为圈选											
* 数据	来源-行为筛选											
Į.	* 行为筛选											
	行为域		业务线			动作对象						
	lfs_电商		lfs_淘宝			购买-lfs_商品	浏览-lfs_商品	收藏-lfs_商品	A			
2	+ 新增行为 按共有行为属性过滤	① (1/3)									Ę	副置引导
	金额	\sim	介于	V	0	- 800						
	+ 新增行为属性											
3	按共有对象属性筛选	^(1/3)									Ę	8 配置引导
	商品类目	V.	精确匹配(in)	\sim	女装,							
	+ 新增对象属性											

参数	描述
行为筛选	行为筛选用于筛选行为数据统计的范围,即筛选到具体的行为规则。行为筛选 需要依次选择行为域、业务线和动作和对象。 选择某个行为域后,只能选择该行为域下的业务线。选择了业务线后,只能选 择该业务线下的单个或多个行为规则。 例如,选择电商行为域下的淘宝业务线,选择淘宝业务线下的浏览商品、购买 商品和收藏商品的行为规则。 如果您还没有行为规则,则需要提前完成行为规则的创建。具体操作,请参 见创建并配置行为规则。
	 基于已筛选的行为规则,配置行为规则的共有行为属性,以进一步筛选行为主体。Dataphin最多支持配置3个共有行为属性规则。每个行为属性规则均需要配置行为属性、筛选条件和筛选值。 配置单个行为属性过滤规则: 高、筛选行为属性。 行为属性的下拉列表中,展示了已筛选的行为规则配置的行为属性。例如,已筛选的行为规则配置了金额、发生次数的行为属性,则在共有属性下拉列表中仅展示金额和发生次数。 选择筛选条件。筛选条件包括=、!=、>、>=、<、<=、介于、非介于、非介于、精确匹配、精确排除、模糊匹配和不为空。 筛选条件对应的SQL语句,请参见筛选条件对应的SQL语法。 化据行为属性和筛选条件配置筛选值。筛选值说明如下: >、>=、<、<=、介于、非介于的筛选值必须配置为数值类型。 !=、=、精确匹配、精确排除和模糊匹配的筛选条件对应的筛选值的数据类型没有限制,Dataphin会根据您配置筛选值的数据类型,自动转换行为规则输出表(oi_behavior_detail_d)中对应字段的数据类型,以实现筛选值与行为规则输出表的数据进行对比,筛选出符合行为属性过滤规则的行为主体。 不为空的筛选条件,无需配置筛选值。 例如,共有行为属性的金额介于0~800。
	例如,共有行为属性的金额介于0~800。

参数	配置多个行为属性过滤规则:	
2		
按共有行为属性过滤	b. 行为属性过滤规则间的逻辑关系默认为且。单击且,选择行为 的运算逻辑,并完成新增行为属性过滤规则的配置。	属性间
	运算逻辑包括且、或和排除。如果是3个共有行为属性过滤规则 两个行为属性过滤规则先计算,运算结果再与下一条共有行为原 滤规则进行计算。	l, 则前 属性过
	例如,在时间精确匹配2020年5月21日的行为数据范围内,筛注 额小于800的行为数据为下图最终筛选到的行为数据。	先到金
	2 按共有行为属性过滤 ① (2/3) 字面	置引导
	Ξ 时间 ∨ 精确匹 ∨ 2020-05-21	A
	三 金额 ∨ < ∨ 800	<u></u>
	例如,下图的最终筛选到的行为数据包括时间精确匹配2020年日的行为数据和金额小于800的行为数据。 2 按共有行为属性过滤 ① (2/3) 8 雷	5月21 記置引导
	三 时间 ∨ 精确匹 ∨ 2020-05-21	或
	三 金额 ∨ < ∨ 800	
	例如,在时间精确匹配2020年5月21日的行为数据范围内,排队小于800的行为数据为下图最终筛选到的行为数据。	余金额
		6直51守
	三 四间	排除
	三 金额 ∨ < ∨ 800	

参数	描述
	基于已筛选的行为规则,配置行为规则的共有对象属性,以进一步筛选行为主体。Dataphin最多支持配置3个共有对象属性规则。每个对象属性规则均需要配置对象属性、筛选条件和筛选值:
	■ 配置单个对象属性筛选规则:
	a. 筛选对象属性。
	对象属性的下拉列表中,展示了已筛选的行为规则配置的对象属性。 例如,已筛选的行为规则配置了价格和类目的对象属性,则在共有属 性下拉列表中仅展示价格和类目。
	b.选择筛选条件。筛选条件包括=、!=、>、>=、<、<=、介于、非介 于、精确匹配、精确排除、模糊匹配和不为空。
	筛选条件对应的SQL语句,请参见 <mark>筛选条件对应的SQL语法</mark> 。
	c. 根据对象属性和筛选条件配置筛选值。筛选值说明如下:
	■ >、>=、<、<=、介于、非介于的筛选值必须配置为数值类型。
按共有对象属性筛选	!=、=、精确匹配、精确排除和模糊匹配的筛选条件对应的筛选值 的数据类型没有限制,Dataphin会根据您配置筛选值的数据类型, 自动转换行为规则输出表(oi_behavior_detail_d)中对应字段 的数据类型,以实现筛选值与行为规则输出表的数据进行对比,筛 选出符合对象属性筛选规则的行为主体。
	■ 不为空的筛选条件,无需配置筛选值。
	例如,共有对象属性的商品类目精确匹配为女装。
	■ 配置多个对象属性筛选规则:
	a. 完成单个对象属性筛选后,单击 新增对象属性 。
	b. 对象属性筛选规则间的逻辑关系默认为且。单击且,选择对象属性间的运算逻辑,并完成新增对象属性筛选规则的配置。
	运算逻辑包括且、或和排除。如果配置了3个共有对象属性筛选规则, 则系统的运算前两个对象属性筛选规则先计算,运算结果再与下一条 对象属性筛选规则进行计算。

■ 场景说明

数据来源 國 选用户行 3 3	2 标签值配置 数据 基于圈选的用户行为数据定义标签值	
数据来源配置		
圈选方式筛选 按标签圈选		
* 数据来源-按标签圈选 ①		8 配置引导
二 01 商日公務		
	稿件ULIQ(In) V 又微了	
〒 新埠标益		
	取消	下一步
参数	描述	
	按标签数据筛选出再次需要打标的对象。Dataphin最多支持配置3个领则。每个筛选规则均需要配置标签、筛选条件和筛选值。	^{策选规}
	■ 配置单个筛选规则:	
	a. 筛选标签。	
	标签的下拉列表中,展示了Dataphin内所有的规则统计标签和 挂标签。	1注册上
	b.选择筛选条件。筛选条件包括=、!=、>、>=、<、<=、介于、 于、精确匹配、精确排除、模糊匹配和不为空。	非介
	筛选条件对应的SQL语句,请参见 <mark>筛选条件对应的SQL语法</mark> 。	
	c. 根据标签和筛选条件配置筛选值。筛选值说明如下:	
	■ >、>=、<、<=、介于、非介于的筛选值必须配置为数值类	型。
	 !=、=、精确匹配、精确排除和模糊匹配的筛选条件对应的 的数据类型没有限制,Dataphin会根据您配置筛选值的数据 自动转换标签结果表中对应字段的数据类型,以实现筛选值 结果表的数据进行对比,筛选出符合筛选规则的对象。 	筛选值 居类型 <i>,</i> 重与标签
	■ 不为空的筛选条件,无需配置筛选值。	
	例如,注册商品分类标签精确匹配到女装。	
	■ 配置多个标签筛选规则:	
	a. 完成单个标签筛选规则配置后,单击 新增标签 。	
	b. 标签筛选规则间的逻辑关系默认为且。单击且,选择标签间的 辑,并完成新增标签筛选规则的配置。]运算逻
	运算逻辑包括且、或和排除。如果是3个标签筛选规则,则前 筛选规则先计算.运算结果再与下一条共有标签筛选规则讲行	两个标签 计算。
그는 그는 서는 대가 나는		

రౖ గా చె ඕ টు 参数	<u>例如</u> ,下图最终筛选到的标签数据,包括商品分类标签精确匹配到坚果的标签 数据和注册商品分类标签精准匹配到女装的标签数据。
	* 数据来源·按标签圈选 ①
	三 注册商品分类 ∨ 精确匹配(In) ∨ 坚果∡
	三 注册商品分类 ∨ 精确匹配(in) ∨ 女装
	例如,在注册商品分类标签精确匹配到生活用品的标签数据范围内,筛选到商品分类标签精准匹配到女装的标签数据为下图最终筛选到的标签数据。
	三 注册商品分类 ∨ 精确匹配(in) ∨ 生活用品
	三 注册商品分类 ∨ 精确匹配(in) ∨ 女装⊿
	例如,下图最终筛选到的标签数据为在注册商品分类标签精确匹配到生活用品的标签数据范围内,排除注册商品分类为女装的标签数据。
	● 近期末期不均定間近 ∪
	三 注册商品分类 ∨ 精确匹配(in) ∨ 生活用品▲
	三 注册商品分类 ∨ 精确匹配(in) ∨ 女装⊿

筛选条件对应的SQL语法

筛选条件	对应SQL语句	示例
=	= "	= '80'
!=	!= ''	!= '80'
>	> "	> '80'
>=	>= ''	>= '80'
<	< "	< '80'
<=	<= ''	<= '80'
精确匹配	in ''	in 'name'
模糊匹配	like '%%'	like '%name%'
精确排除	not in ''	not in 'name'
介于	between '' and ''	between '1' and '100'

筛选条件	对应SQL语句	示例
非介于	not between '' and ''	not between '1' and '100'
不为空	is not null	无

iv. 在标签值配置页签,为筛选到的待打标对象配置标签值,并单击确定。

不同场景的标签值配置参数不同,说明如下:

场景1:基于用户行为数据,筛选出需要打标的用户,并自定义用户的标签与场景3:基于已打过标签的数据,筛选出需要再次打标的对象,并自定义对象新的标签。

自定义待打标对象的标签值。例如,自定义用户商品偏好的标签值为女装。

	•	数据来源配置	2 标签值配置 基于圈选的用户行为数据定义标签值
* 标签值配置			
标签值定义	女装		

■ 场景2: 基于用户行为数据,筛选出需要打标的用户,并配置行为属性的统计值为用户的标签。

配置行为数据的统计字段及其计算方式,且统计值作为用户的标签值。例如,统计字段选择为**金**额,计算方式选择为**求平均值**,则金额的平均值为用户价格偏好的标签值。

		数据来源面 圏选用户行 ³	置 の数据		2	标签值配置 基于圈选的用户	行为数据定义标签值	1
* 标签值	配置							
统计	字段	金额		\sim				
计算刀	方式 🧏	求平均值		\vee				

- 2. 在新建标签页面,配置筛选行为或标签数据的时间范围。
 - i. 在STEP 2区域, 单击 **∨**图标。
 - ii. 在STEP 2区域, Dataphin默认时间跨度为30天。您也可以单击时间跨度后应图标,修改时间跨度。
 - iii. 在新建标签页面右侧的时间跨度配置区域,配置时间跨度,单击确定。
 Dataphin支持配置的时间跨度为30天、60天、90天、180天、365天。

配置规则统计标签的调度参数及标签输出表的存储生命周期

1. 在新建标签页面,单击页面上方的调度配置。在调度配置面板,配置参数。

i. 配置基本信息区域的参数。

• 基本信息				
名称	ĸ			
节点ic	d 暂无, 请提交调度			
节点类型	별 MAX_COMPUTE_SQL			
负责人		×		
描述	情输入50个字符内的描述文案			

⑦ 说明 Dataphin自动生成名称、节点ID和节点类型的参数,不支持修改。

参数	描述
负责人	支持转交负责人。当任务运行报错时,Dataphin会及时通知到该任务的负责 人,便于您及时处理异常。 负责人其他的操作权限取决于负责人的角色。关于角色对应的操作权限的信息, 请参见数据萃取权限列表。
描述	支持修改规则统计标签已有的描述,也支持为标签添加描述。

ii. 配置调度配置区域的参数。

• 调度配置					
时间属性	● 正常调度 ◎ 空	跑调度			
暂停调度	• 否 是				
调度周期	H v	05:00 🕓			
cron 表达式	005**?				
最近的生成时间	2021-05-20 05:00:00 2021-05-24 05:00:00	2021-05-21 05:00:00	2021-05-22 05:00:00	2021-05-23 05:00:00	
优先级	最高优先级	~			
参数	描述				

参数	描述
时间属性	 选择任务在生产环境调度的时间属性。时间属性包括正常调度和空跑调度: 正常调度:按照调度周期的时间配置调度,并正常执行,通常任务默认选中该顶。 空跑调度:按照调度周期的时间配置调度,但都是空跑执行,即一调度到该任务便直接返回成功,没有真正的执行任务。 适用场景说明如下: 正常调度:适用于任务需要正常调度以产出数据的场景。 空跑调度:适用于任务有段时间不需要产出数据,但不能阻塞下游任务执行的场景。
暂停调度	即可暂停该任务及该任务的下游任务的调度,一旦调度到该任务会直接返回失 败,不会执行。通常用于某个任务暂时不用执行,但后面还会继续使用的场景。
调度周期	 调度周期可选择日、周和月:,您可以根据业务需要,指定任务运行的具体时间点。调度周期说明如下: 日调度,即调度任务每天自动运行一次。新建周期任务时,默认的时间周期为每天0点运行一次。 周调度,即调度任务每周的特定几天,在特定时间点自动运行一次。 月调度,即调度任务在每月的特定几天,在特定时间点自动运行一次。 ⑦说明 周调度和月调度中,如果您没有指定任务的运行日期,为保证下游实例正常运行,Dataphin会每天生成实例后直接设置为运行成功,而不会真正执行任何逻辑,也不会占用资源。
优先级	优先级定义了同一时间待调度任务的优先级。默认选择为 中等优先级 。您可以 根据业务需求修改任务调度的优先级。Dataphin支持的优先级包括最低优先 级、低优先级、中等优先级、高优先级和最高优先级。

iii. 配置依赖关系区域的参数,并单击确定。

• 依赖关系	自动解析				
上游依赖	物理节点	逻辑表节点 ①			手动添加上游
输出名称		节点名	节点ID (实例ID)	负责人	操作
oi_beh _26		oi_beh _26	n_ 14	Supervision of the second	Ť.
oi_beh _28		oi_beh _28	n_ 87	Spectra scharette mighterette	iii iii
oi_beh _30		oi_beh _30	n_ 62	Sector Card	Ť.
pg_hac rule		pg_hac	n_ 10	Sector Contractor	
当前节点					
输出名称		节点名	节点ID (实例ID)	负责人	操作
Data		oi_statisti		Anna Anna Anna An	
-54/5M		14753		10 Sa 1950	取消 确定
参数		描述			

参数	描述					
	 您可以通过自动解析和手动添加两种方式,为规则统计标签节点添加上游依赖的 节点: 单击自动解析,自动解析根据标签选择的圈选方式不同,自动解析到的节点 也不同: 圈选方式选择为按行为圈选,则Dataphin仅支持自动解析出标签中已选择 行为规的输出物理节点。输出名称格式为					
	oi_behavior_detail_d oi_behavior_detail_d n_2 oi_behavior_detail_d oi_behavior_detail_d n_2 oi_behavior_detail_d oi_behavior_detail_d n_2 876					
上游依赖	 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>					
	输出名称 节点名 节点ID (实例ID) 负责人					
	Data_Distill oi_statistics_label_31 n el 97					
	Data_Distill oi_statistics_label_66 n34					
	Data_Distilloi_statistics_label_31 n el 97					
	 基于业务场景,如果需要添加其他节点作为当前规则统计标签的上游节点,则需要手动添加上游依赖的物理节点。 注意 Dataphin不支持手动添加逻辑表节点。 单击新增上游依赖,在新建上游依赖对话框中,输入所依赖节点的输出名称的关键字进行搜索节点,搜索到后单击确定新增。 					
	当前节点为您展示规则统计标签的数据输出的物理节点。当前节点名称即规则统 计标签结果表名称,规则统计标签名称的格式为Data_Distill_结果表名,不支 持修改。					
当前节点	关于标签结果表的结构、数据预览等信息,请参见物理表详情。					
	如果规则统计标签已提交且被其他任务所依赖,则单击操作列下的 <mark></mark> 图标,可 以查看输出节点的下游节点。					

2. 在新建标签页面,单击页面上方的物理化配置。在物理化配置面板,配置规则统计标签的存储生命周期后,单击确认。

物理化配置	
•存储配置	
*存储生命	周期 3650
	快速选择 90 180 365 3650 天
* 选择分区	· 分子 · 小 · 小 · 小 · 小 · 小 · 小 · 小 · 小 · 小 ·
	取消 确认
参数	描述
	存储生命周期用于定义规则统计标签在Dataphin的生命周期。存储生命周期对所有 的规则统计标签生效,默认3650天,无最大限制。
存储生命周期	注意 仅超级管理员支持修改存储生命周期。
选择分区字段	Dataphin默认展示 ds ,不支持修改。

测试规则统计标签并查看标签报告

1. 在新建规则统计标签页面,按照下图操作,测试运行规则统计标签。

标签信息 调度配置 物理化配置 标签版本	윤 🕮 🖾 🛷 🧿
● 規則統計类标签=指定被率值 ☑ 标签名称: 61.員品分类02 标签名称: 61.員品分类02 你还就上菜: 0.日本日本市場 ⑦ 经结束数项 ③ 标签摘签 ● 经结束数项 ③ 标签摘签 ● 新聞 ○ ● 「日本市場」 ○ ● 「日本市場」 ●	
* 选择圈选方式 按行为圈选 🗸	23
电度_勿助-数据验还浏览.褐品 郎 区 电商_勿助-数据验证收器,商品 或 打标方式:指定枚举值 来源1 按行为网路	您可以納增國得致 Step 1 的配置项,也可以直接进行 Step 2 的配置
电度 勿助 数据检证 浏览 標品 武 电商 勿助 数据检证 收露 商品 或 … 打标方式: 描定枚举值 来源2 按行为网选	
电度、勿动数据给还浏览、商品 或 回 电度、勿动数据检还收益、商品 或 打玩示式: 推定收举语 来源于操行力调选	

2. 测试成功后,单击查看标签报告。
智能数据构建与管理 Dat aphin

数据萃取·标签中心



- 在标签报告页面,查看标签的基本信息、数据来源、标签分布及结果表数据抽样。
 根据打标方式不同,生成的标签报告也不同,详细说明如下:
 - 指定枚举值



○ 指定统计值



提交规则统计标签至生产环境

在新建规则统计标签页面,按照下图操作指引,提交规则统计标签至生产环境。

标签信息 调度配置 物理化配置 标签版本	n 🛛 🖉 🖾 🖉 🖉 🖉
2 規學统計类标签::maskififi ☑ 杨运车称::02,最新时间 杨运车标:03,000 杨运车 杨运车后:000,000 Ball Ball Ball 公开技法: 公开 Ball Ball Data ● 杨运客管理场: ● 杨运客管场: ● 杨运客管场: ● 杨运客管场: ● 杨运客管场: ● 杨运客管理场: ● 周辺市场: ● 周辺市场: ● ● ● 周辺市场: ● 周辺市场: ● ● ●	・提文智注 2 適応 乱が現後の相关報注時期、如交更偽思 0/128 現文前講体USPE46年並ば褒変配置的自該解析接任、否則認確率故能可認 方法上述 0/128 現文前講体USPE46年並ば褒変配置的自該解析接任、否則認確率故能可認 方法上述 0/128 現文前式の法、即更新至止一环境、清潔成現作 3 取消 1 取消 1 取消 1 取消 1 取消 1 取消 1 取消 1
电度、勿补数据验证例法有品 彩 [2] 电度、勿补数据验证小器有品 彩 打玩访玩: 描述统计值 来源1] 除行为确认	您可以希腊或佛致 Step 1 的配置项,也可以直接进行 Step 2 的配置
STEP 2 助同時度低高度 ・ ・力放気中, 波風気引用効用 * 时間時度:30天 区	

后续步骤

- 1. 在运维中心查看并运维规则统计标签的萃取任务。如何查看并运维萃取任务,请参见萃取任务。
- 2. 查看并运维规则统计标签的萃取实例。如何查看并运维萃取实例,请参见萃取实例。
- 3. 当规则统计标签的萃取实例的运行状态为成功时,您可以通过即席查询,查询业务数据(即规则统计标签结果表)的产出情况,如下图所示。

	1 sele	ct * from Data	a_Distill_	, in a line	
С	onsole	Result	1997		
	result_0	X resul	t_1 ×		
		id_type 🛛 🖫	id_value 🛛	label_value 🛛 🖫	ds 🗉
	1	Name	孙	文娱	20210407
	2	Name	Ξŧ	女装	20210407
	3	Name	胡	女装	20210407
	4	Name	赵	女装	20210407
	_	N 1	<i>μ</i> ι≥ππ	ᆝᄮᆉᄮᆣᅑᇚᇊ	00040407

如何通过即席查询查看业务数据的产出情况,请参见查询并下载数据。

8.2. 创建行为偏好标签

8.2.1. 步骤一: 创建行为偏好标签

您可以聚集多种用户行为数据,结合时间、重要性等因素数据,通过界面配置偏好标签值的计算规则,从而 获取用户商品价格偏好、商品类目偏好等数据,助力企业寻找业务发力点。本文为您介绍如何创建行为偏好 标签。

前提条件

在开始执行操作前,请确认您已满足以下条件:

- 已完成实体ID的创建,具体操作请参见创建并管理实体ID。
- 已完成行为规则的创建,具体操作请参见创建并配置行为规则。

操作步骤

- 1. 进入标签中心。
- 2. 在标签中心页面,按照下图操作指引,进入新建标签对话框。



3. 在新建标签对话框,配置参数。

新建标签				×
★ 标签蓝文文	法指写标本监文文	 	法估写标签文称	
- 10022/C/CH	标签英文名将作为标签逻辑表的列名,默	认存储为小写	BW DRUCHD.	
描述	请填写描述			
				0/128
* 所属类目	请选择标签类目	\sim	()	
* 公开状态	公开	\sim		
* 标签类型	行为偏好			
* 结果表名	Data_Distill_ 请填写结果表名			
			取消	确定

参数	描述		
	填写标签英文名。创建标签后,其英文名无法修改。命名规则如下: • 只能包含字母、数字或下划线(_)组合组成。		
标签英文名	⑦ 说明 标签英文名将作为标签逻辑表的列名,字母均默认存储为小写字母。		
	◎ 项目空间内唯一。◎ 长度不能超过64字符。		
标签名称	填写标签名称。命名规则如下: • 支持任何字符。 • 长度不能超过64字符。		
描述	填写对标签的简单描述。		
所属类目	即标签的类目,选择此前已创建或系统内置的标签类目。		
公开状态	即标签的状态,用于定义项目空间内成员是否具有查看该标签的权限。 公开状态包括 公开 和隐藏: • 公开状态的标签: Data_distill项目空间内所有成员都可以查看。 • 隐藏状态的标签: Data_distill项目空间内仅项目管理员和标签负责人可以查看。		
标签类型	默认为 行为偏好 ,不支持修改。		

参数	描述
	填写标签产出数据写入的表名。命名规则如下: • 只能包含字母、数字、下划线(_)组合组成。 • 项目空间内唯一。 • 长度不能超过64字符。
结果表名	注意 行为偏好标签提交后,其名称无法修改。
	建议结果表名命名为标签名称,便于其他节点通过搜索结果表名依赖该结果表。

4. 单击确定,完成行为偏好标签的创建。

后续步骤

完成上述基础信息的配置后,即可配置行为偏好标签。具体操作,请参见步骤二:配置并提交行为偏好标 签。

8.2.2. 步骤二: 配置并提交行为偏好标签

完成行为偏好标签的基本信息配置后,您需要为行为偏好标签配置数据来源、标签值,及配置行为偏好标签的存储生命周期,后续可以测试行为偏好标签并查看标签报告。本文为您介绍如何配置、测试行为偏好标签、查看标签报告及提交行为偏好标签至生产环境。

前提条件

在开始执行操作前,请确认您已满足以下条件:

- 已完成行为偏好标签基本信息的配置,具体操作请参见步骤一:创建行为偏好标签。
- 已完成行为规则的创建,具体操作请参见创建并配置行为规则。

操作流程

- Step 1: 配置行为偏好标签的取数逻辑及标签值的计算规则
- Step 2: 配置行为偏好标签的取数逻辑及标签值的计算规则
- (可选) Step 3: 测试行为偏好标签并查看标签报告
- Step 4: 提交行为偏好标签至生产环境

配置行为偏好标签的取数逻辑及标签值的计算规则

1. 在新建标签页面, 配置筛选待打标对象的规则, 并为待打标对象配置标签值。

i. 在STEP 1区域,单击开始配置。

🍤 新建行为偏好标签 🛛 🗙
标签信息 物理化配置 标签版本
💫 行为偏好类标签 🛛 🖸
标签名称: -
标签类目: -
公开状态: 限廠 标签输出表: -
STEP 1 打标对象圈选及标签值定义 へ -基于行为数据圈选ID
开始配置 >
+ 新增配置 1/10
STEP 2 标签值计算及返回规则配置 へ -每个ID对应的标签值配置
开始配置 >

ii. 在新建标签的页面右侧的数据来源配置页签,配置来源数据后,单击下一步。

🎙 新建行为偏好标签 🛛 🗙					: 8
标签信息 物理化配置 标签版本					£ 2 0 1 0
🔺 行为编好类标签 🛛 🗹	1 චූ	7.据来源配置 选用户行为数据		② 标签值配置 基于需适的用户行为数据定义标签值	
标签名称: - 标签类目: - 公开状态: 隐藏	* 数据来源-行为筛选				
标签输出表: - ○ 标签配置预选	1 * 行为筛选 行为域	业务线	动作对象		
🚙 STEP 1	请选择行为域	∨ 请选择业务线	> 请选择动作对象		
打标对象圈选及标签值定义 个 -基于行为数据圈选ID	+ 新增行为	(3)			8 配置引导
と 配置中	+ 新增对象属性				
+ 新措配置 1/10					
STEP 2 标签值计算及返回规则配置					
开始配置>					
				取消	下─∌

参数	描述
	行为筛选用于筛选行为数据统计的范围,即筛选到具体的行为规则。行为筛选需 要依次选择行为域、业务线和动作和对象。
行为筛选	选择某个行为域后,只能选择该行为域下的业务线。选择了业务线后,只能选择 该业务线下的单个或多个行为对象。
	例如,选择电商行为域下的淘宝业务线,选择淘宝业务线下的浏览商品、购买商 品和收藏商品的行为规则。
	基于已筛选的行为规则,配置行为规则的共有对象属性,以进一步筛选行为主 体。Dataphin最多支持配置3个共有对象属性规则。每个对象属性规则均需要配 置对象属性、筛选条件和筛选值:
	■ 配置单个对象属性筛选规则:
	a. 筛选对象属性。
	对象属性的下拉列表中,展示了已筛选的行为规则配置的对象属性。例 如,已筛选的行为规则配置了价格和类目的对象属性,则在共有属性下 拉列表中仅展示价格和类目。
	b.选择筛选条件。筛选条件包括=、!=、>、>=、<、<=、介于、非介于、 精确匹配、精确排除、模糊匹配和不为空。
	筛选条件对应的SQL语句,请参见 <mark>筛选条件对应的SQL语法</mark> 。
	c. 根据对象属性和筛选条件配置筛选值。筛选值说明如下:
	■ >、>=、<、<=、介于、非介于 <i>,</i> 建议筛选值配置为数值类型,如果 筛选值不是数值类型则可能导致运行失败或比较结果不符合预期。
	!=、=、精确匹配、精确排除和模糊匹配的筛选条件对应的筛选值的 数据类型没有限制,Dataphin会根据您配置筛选值的数据类型,自动 转换行为规则输出表中对应字段的数据类型,以实现筛选值与行为规 则输出表的数据进行对比,筛选出符合对象属性筛选规则的行为主 体
	■ 不为空的筛选条件,无需配置筛选值。
按共有对象属性筛选	

		描述	禹住师远为	也见り:			
		a. 完成单个	对象属性颌	帘选后,	单击	新增对象属性。	
		b. 对象属性 运算逻辑	筛选规则间 或 或 排除	目的逻辑 ,并完成	关系 新增	默认为 且 。单击且, 对象属性筛选规则的	选择对象属性间的 的配置。
		运算逻辑包括且、或和排除。如果配置了3个共有对象属性筛选规则,则					
		属性筛选	规则进行证	十算。			
		例如,在时间 [;] 800的行为数排	精确匹配2 居为下图最	020年5月 终筛选到	月21日 到的行	日的行为数据范围内 F为数据。], 筛选到金额小于
		2 按共有行为属	生过滤 🛈 (2/3)			8 配置引导
		☰ 时间	\vee	精确匹	. ~	2020-05-21	E
		三 金额	\vee	<	\vee	800	
		例如,下图的: 为数据和金额,	最终筛选至 小于800的	l的行为 行为数排	数据包 居。	回括时间精确匹配20	020年5月21日的行
		2 按共有行为属	生过滤 🛈 (2/3)			8 配置引导
		三 时间	\sim	精确匹	. ∨	2020-05-21	或
		三 金额	\vee	<	\vee	800	
=	= ''	2 按共有行为属	生过滤 🛈 (2/3)	= '8	0'	╞ 配置引导
!=		目 时间	\checkmark	精确匹	. ∨	2020-05-21	
	!= ``				!= '	80'	排即
>	> ''	∃ 金额	\vee	<	!= ':	80' ⁸⁰⁰ 80'	44E
> >=	>= ''	三金额	V	<	!= ' > ' >= '	80' ⁸⁰⁰ 80' '80'	14E
> >= <	i= " > " >= " < "	金额	V	<	!= '	80' 800 80' '80' 80'	
> >= < <=	i= " > " >= " < "	金额	V	<	!= '{ > '{ >= ' < '{ <= '	80' 800 80' '80' 80' '80'	41E
> >= < <= 精确匹配	!= " > " >= " < "	∃ 金额	V	<	!= 'a	80' 800 80' '80' 80' '80' hame'	34E
> >= < <= 精确匹配 模糊匹配	!= " > '' >= '' < ''	Ξ 金额	V		!= ': > ': >= ' < ': <= ' in 'r like	80' 800 80' '80' 80' '80' hame' '%name%'	
> >= < <= 精确匹配 模糊匹配 精确排除	!= " > '' >= '' < ''				!= '' > '' >= '' < '' in 'r like not	80' 800 80' '80' '80' '80' '80' '80' 'ame' '%name%' in 'name'	
> >= < <= 精确匹配 模糊匹配 精确排除 介于	!= " > '' >= '' < ''	Ξ 金额	V		!= 'i > 'i >= ' < 'i in 'r like not	80' 800 80' '80' '80' '80' '80' 'ame' '%name%' in 'name' ween '1' and '100	
> >= < <= 精确匹配 精确排除 介于 非介于	!= " > '' >= '' < ''	三 金额 and '' en '' and ''			!= 'i > 'i >= ' < 'i in 'r like not bet'	80' 800 80' '80' '80' '80' '80' '80' 'ame' '%name%' in 'name' ween '1' and '100 between '1' and	' '100'

iii. 在标签值配置页签,为筛选到的待打标对象选择打标方式及配置标签值后,并单击确定。
 打标方式包括直接打标和对象属性映射打标,详细说明如下:

■ 直接打标: 自定义待打标对象的标签值。例如自定义用户商品偏好的标签值为女装。

	数据来源配置 國选用户行为数据	2 标签值配置 基于局部的用户行为数据定义标签值
* 标签值-打标方式 ①		
打标方式 直接打标		
* 标签值配置 ①		
标签值定义 请输入标签值	I	

■ 对象属性映射打标:基于共有对象属性值,通过设置属性值映射规则生成对应的标签值。

	✓ 数据来源配置 图念用户行为数据				(2 标签值配置 基于意志的用户行为数据定义标签值	
∗标签值-打标方式 ①							
打标方式 对象属性映射打标							
■ 标签值配置①							
·· · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · <	标签值 9时,优先采用靠前的规则,因此属性3	型準順字及映射规则原字將數術每条行为返	简的标签值				
1 * 选择共有对象属性映射打	标						
共有对象属性 (最多选3个)							
价格 美目							
属性值分布预选 三价格							
对象分布百分比							
5%							
3%							
2% · · · · · · · · · · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·					· · · · · · · · · · · · · · · · · · ·
0%	789.45 789.0 429.77 1000.45	123.22 123.23 123.24 123.33 2000	.0 2000.55 2889.53 2889.88 30	0.23 367.99 368.66 399.0 3	399.44 42.55 429.0 429.55	4999.0 4999.58 728.45 789.66 799.0	889.0 889.51 999.48
0							居住但
◎ 柱状圈为所遗行为下预路	谢据,与特定陈遗条件无关						
○ ● 法提供财方式 价格							
○ 属性值应用为标签值							
② 设计映射规则 ① 3/10							
< V 30	0	» 标签值 低价女装					
≡ > ∨ 20	00	» 标签值 高价女装	Ê.				
三 剩余所有 ∨		» 标签值 普通女装	ā				
					म्र	湖 返回上一步 确知	2 标签生成现场置

↓ 注意 配置的共有对象属性及映射规则顺序会影响最终的标签值。如果配置了多个共有 对象属性,会按照配置的顺序将其对应的映射规则进行排序和计算,优先返回第一个命中的 映射规则对应的结果作为标签值进行存储。

参数	描述
打标方式	选择 对象属性映射打标 。
共有对象属性	可选的共有对象属性为 数据来源配置 区域配置的共有对象属性。您可以根据 业务需求,选择共有对象属性。
选择映射方式	 映射方式包括属性值应用为标签值和设计映射规则: 属性值应用为标签值:选中对象属性的来源字段对应的字段值直接作为标签值存储。 设计映射规则:基于属性来源字段的原始值,通过设置映射规则,计算生成新的标签值,以达到规范化结果数据的目的。映射规则支持最多配置10个。例如,共有对象属性为品牌,原始属性值为香奈奈或XiangNaiNai,对应的标签值为xiangnainai。 筛选条件请参见筛选条件对应的SQL语法;

- 2. 在新建标签页面, 配置筛选行为数据的时间范围。
 - i. 在STEP 2区域,单击<mark>、</mark>图标后,单击开始配置。
 - ii. 在标签值计算及返回规则配置页面,配置参数后,单击确定。

标签信息 物理化距量 标签版本		A II II 🕫 🖉 🎯
▲ 行为偏好类标签	标签 德计算及 透明 顺利 配置 ①	
标签名称: 行为庸好标签2 标签关目: 高忙标签	= 邻签值计释时间按查整置 ◎	
公开状态: 公开 标签输出表: Data_Distil_x_test4	第7時の間かび時度 30 天 作品担当 30 00 180 395	
	▲ 标签值计算 受减配置 ①	
STEP 1 打阪7線圏造及标落值定义 へ 基刊行力地和認知の	■ - 出4会社が10月間時 単位10月前日 7 · ·	
电電(勿助天勝)()路,衛品 10 区 电電(勿助天勝)()路,衛品 10 打伤力式:刀象魔性映射打标 [鼓[1]]	■ - 資品資本有27尚無法	
+ #1922 1/10		
◆ STEP 2 标签值计算及返回编制管理 ▲ -每个0%应前标签组图量		
<u>د</u> ۲۵۹+	● 鄂盗德计算 经重配量 ◎	
	国利 02 第100万条48 02 国利 03 第10万条48 03 第10万条48 03 第10万条48 03 第10万条48 03 第10万条48 03	0.2 0 0.1 0 0.7 0
	 - 标签适应值数量 	
	夏田谷街市公2 0N ~ 100N	
	· 透照時容量个数 (長歩) ③ 3	
	展开設時>	
		2014 340
		196 AL

区域	参数	描述
标签值计算-时间 跨度配置	来源数据时间跨 度	Dataphin将基于所选时间跨度抽取对应的行为数据进行标签值 计算。来源数据时间跨度默认为30天,您也可以根据业务情况 自定义设置,同时Dataphin提供30天、90天、180天、365天 的快捷选项。
	选择衰减时间周 期	选择标签值的衰减时间周期。 不同时间的行为对标签值的影响程度不同,一般认为,时间越 近影响程度越大;同一周期内行为的影响力相同,不同周期行 为的影响力按照设置的衰减周期和曲线随时间递减。 对于标签值会随着行为快速变化的,建议选择较短的时间周期 (如纸尿裤型号偏好);对于标签值相对稳定,不随行为快速 变化的,建议选择较长的时间周期(如预测个人肤质)。
标签值计算-衰减 配置	请选择衰减时间 曲线	 衰减时间曲线包括平滑衰减、线性衰减、指数衰减: 平滑衰减:行为对标签值的影响不随时间衰减。例如购买护肤品类型(个人肤质基本不变)。 线性衰减:行为对标签值的影响随时间线性衰减。例如偏好音乐风格。 指数衰减:行为对标签值的影响随时间指数衰减。例如购买纸尿裤类型。
	标签值计算-权 重配置	要求最多两位小数,且各项权重加和为1;权重越大,表示该行 为对最后标签值的影响越大。

区域	参数	描述			
枟恷汳冋伖巸罟	覆盖范围设定	即覆盖用户百分比,行为发生次数在某个值以下的用户占所有用户的百分比。 (仅对覆盖范围内的用户打标,以降低行为次数过少或过多等偶然情况对标签准确度的影响。例如选中行为是购买保湿类护脱品,10%的用户购买次数在3次以下,90%的用户购买次数在100次以下,选择范围为10%-90%,则只对购买次数在3次以100次以下的用户打标,防止购买次数过少(3次以下)或过多(100次以上)对准确度的影响。			
	返回标签值个数 (最多)	某个用户对应不同的行为记录,则可能被打上多个标签值,可 通过配置返回标签值个数收拢用户画像,优先返回偏好度高的 标签值。仅支持填写1~100间的正整数。 例如,用户001的护肤品牌偏好标签值分别为 A品 牌:0.7 、 B品牌:0.2 、 C品牌:0.1 ,返回标签值个数 为2,则结果表中记录的该用户的标签值为 A品牌:0.7 、 B品牌:0.2 。			

配置行为偏好标签输出表的存储生命周期

- 1. 在新建行为偏好标签页面,单击页面上方的物理化配置。
- 2. 在物理化配置面板, 配置行为偏好标签产出表的存储生命周期。

物理化配置		×
• 物理化配置		
* 存储生命周期 1000		
快速选择	释 180 365 3600 36000 天	
* 选择分区字段 ↓	ds behavior_id source_table	
备注:存储生命周期对全 建议您设定的存储	局行为规则生效,默认36000天,无最大限制,仅超级管理员可修改; 生命周期不小于下游标签选择的的来源数据时间跨度,否则标签数据可能出错	
	取消 确定	

参数	描述
存储生命周期	存储生命周期用于定义行为偏好标签产出表在系统的生命周期。默认36000天,无最大限制。 大限制。 建议您设定的存储生命周期大于或等于下游标签设定的来源数据时间跨度,否则标 签数据可能出错。
	↓ 注意 仅超级管理员支持修改存储生命周期。

参数	描述
选择分区字段	Dataphin默认展示 ds 、 behavior_id 和 source_table ,不支持修改。

3. 单击确定。

如果您希望在提交生产环境前对标签数据进行试跑,可以通过**测试运行**功能获取标签报告,查看标签结 果抽样及标签值分布图,具体操作请参见<mark>测试行为偏好标签并查看标签报告</mark>。

(可选)

测试行为偏好标签并查看标签报告

1. 在新建行为偏好标签页面,按照下图操作,测试运行行为偏好标签。

标签信息 物理化配置 标签版本	6 🗵 🗐 🗸 💿
	認可以納權國學致 Step 1 的配置項、由可以直接进行 Step 2 的配置
STEP 2 标签值计算及返回规则配置 ヘ イロ対应的标签值配置	
最终返回标签值个数:3 衰減周期:7天 衰減曲线:平滑衰減	

2. 测试成功后,单击查看标签报告。

>>>> 編編标签: 日期●				:	\$
<mark>标弦</mark> 信息 物理化配置 标签版本		6 I	ē. 🖸	\$	0
	で 単二、 数it 状行成功 大行日	<u>A</u>		4	
STEP 2 転型信計算及返回規則電置 年心の対応対応運動空間 へ 希心の対応対応運動空間 へ 最後信計算及返回規則電置 へ 最後道路均差 ご 最後現時:7天 東京曲城:平府最減 ご					

 在标签报告页面,查看标签的基本信息、数据来源、标签分布及结果表数据抽样。 如果试跑结果符合您的预期,您可以通过提交操作将行为偏好标签配置更新至生产环境,具体操作请参 见提交行为偏好标签至生产环境。

提交行为偏好标签至生产环境

在新建行为偏好标签页面,按照下图操作指引,提交行为偏好标签至生产环境。

标签信息 物理化配置 标签版本	
図 行力施設実际 図	
STEP 2 販売値付け算込返回規則配置 -每个10対应的运道面流過 へ 最終返回時空程へ数:3 衰減策略:7済衰減 ご	

↓ 注意 提交后,必须对标签进行补数据,否则后续标签计算的结果可能不准确。

后续步骤

- 1. 在运维中心查看并运维行为偏好标签的萃取任务。如何查看并运维萃取任务,请参见萃取任务。
- 2. 查看并运维行为偏好标签的萃取实例。如何查看并运维萃取实例,请参见萃取实例。
- 当行为偏好标签的萃取实例的运行状态为成功时,您可以通过即席查询,查询业务数据(即行为偏好标 签结果表)的产出情况,如下图所示。

	1 sele	ct * from Da	ata_Distill_	and the second sec		
				-Admin@dataprov		
С	onsole	Result				
	result 0	x re	sult 1 X			
	_		_			_
		id_type	a id_value	🛯 label_value	₹.	ds 🗉
	1	Name	孙	文娱		20210407
	2	Name	王津	女装		20210407
	3	Name	胡	女装		20210407
	4	Name	赵			20210407
	_	N.1		/\\+		00040407

如何通过即席查询查看业务数据的产出情况,请参见查询并下载数据。

8.3. 创建注册上挂标签

您可以将已开发好的标签(例如无需额外加工的来源表字段、代码任务加工生成的物理表字段、规范建模加 工的派生指标等)上挂至萃取标签中心进行统一管理。后续可以基于注册上挂标签和规则统计标签构建标签 逻辑表并为向下游应用提供服务。本文为您介绍如何创建注册上挂标签。

应用场景

注册上挂标签适典型应用场景:

- 统一管理标签场景:您可以直接将通过其他方式已开发好的标签(例如无需额外加工的来源表字段、代码任务加工生成的物理表字段、规范建模加工的派生指标等)上挂至萃取标签中心进行统一管理。
- 二次加工场景:基于注册上挂标签可以生成步骤一:创建规则统计标签或创建标签逻辑表,以实现为下 游应用提供服务。

例如, Dataphin已有年龄和性别标签, 您可以基于已有标签创建20岁以下的年轻女性的衍生标签。

原理介绍

注册上挂标签用于将Dat aphin已有数据表(汇总逻辑表或物理表)与实体ID进行关联,并为数据表的某些字段添加标签。上挂标签创建完成后,仅生成实体ID、字段和标签的映射关系,映射关系的示例如下图所示。

来源表				实体
column	comment		标签	标签主键
字段01	字段01的描述	\longrightarrow	标签01	实体ID
字段02	字段02的描述	\longrightarrow	标签02	
字段03	字段03的描述	\longrightarrow	标签03	

相关概念

本文涉及到相关概念的解释如下表所示。

概念名称	相关概念说明
实体ID	实体ID是对实体从不同视角的描述信息,实体包括用户、商品、企业、位置等。 例如,用户的实体ID为会员ID和手机号码、商品的实体ID为商品ID和商品产地、企业的实体ID 为国有企业和金融企业。
标签	标签用于刻画用户画像、描述商品、描述企业、描述位置等。 例如,文娱爱好消费者、美妆商品、金融企业、上海区域。

前提条件

- 已完成标签类目的创建。如何创建标签类目,请参见创建并管理标签类目。
- 已完成注册上挂标签需要关联实体ID的创建。如何创建实体ID,请参见创建并管理实体ID。

• 已确认Dataphin中有您需要的汇总逻辑表和物理表。如何搜索物理表和汇总逻辑表,请参见搜索数据。

步骤一: 配置实体和来源表的关联逻辑

- 1. 进入标签中心。
- 2. 在标签中心页面,按照下图指引,进入新建标签页面。



3. 在新建标签页面,选择来源主表。

1 * 来源主表		
来源表类型	○ 汇总逻辑表 ○ 物理表	
来源主表	请选择物理表。	
参数	描述	
来源表类型	选择需要上挂的标签所属的来源表,您可以是物理表,也可以是汇总逻辑表。	
来源主表	根据来源表类型,选择类型下生产环境的汇总逻辑表或物理表。	

4. 配置关联逻辑。

1	*关联逻辑						
	标签主键	ID_type =	坚果	\sim	ID_value =	test_p_01_ch	\sim
	时间分区	ds =	ds	\sim			

参数	描述
	选择此前您创建的实体ID,包括ID_type和ID_value: ID type可选范围是实体ID。 ID value可选范围是来源表字段。
标签主键	 注意 如果已选中的来源主表已经有了注册上挂标签,则不支持修改关联逻辑。 汇总逻辑表的关联逻辑及统计粒度,系统自动识别不支持修改。
	如果您选的实体ID存在重复值导致无法确定实体的唯一性,您可单击 新增 ,增加多 个实体ID进行关联,确保实体ID的唯一性。
	↓ 注意 如果您添加多个ID,将生成联合主键。提交后不可修改注册上挂标 签的关联逻辑。
时间分区	选择时间分区,仅支持选择ds。

步骤二:为汇总逻辑表或物理表的字段添加标签

1. 在新建标签页面的添加标签区域,按照下图指引,选择来源主表中需要添加标签的字段。

添加标签						
Q、请输入搜索关键字		新増列表(2) 历史已添加列表(0)				
- 字段及描述		- 标签名称	标签英文名 ①	所属类目	公开状态	攝作
Iand Be∉		宝马	bmw	其他标签 🗸 🗸	公开 🗸	C Ó
volkswagen 大众	2	真迪	audi	其他标签 🗸	公开 🗸	区前
□ mazda 马首达	»					

↓ 注意 关联主键字段不支持为其添加标签。

2. 在新增列表页签, 配置参数。

新增列票	新增列表(2) 历史已添加列表(0)						
	标签名称	标签英文名 ①	所属类目		公开状态		操作
	宝马	bmw	其他标签		公开		区前
	奥迪	audi	其他标签		公开		区面
参数		描述					
标签名称		标签名称的命名规则如下: 。 支持任何字符。 。 长度不能超过64个字符。					

参数	描述			
	注册上挂标签创建后, 其英文名无法修改。命名规则如下: • 包含字母、数字、下划线(_)。			
标签英文名	⑦ 说明 标签英文名将作为标签逻辑表的列名,字母默认存储为小写字母。			
	◎ 长度不能超过64个字符。			
所属类目	选择此前已创建的标签类目。			
公开状态	即标签的状态,用于定义项目空间内成员是否具有查看该标签的权限。 公开状态包括公开和隐藏: • 公开状态的标签: Data_distill项目空间内所有成员都可以查看。 • 隐藏状态的标签: Data_distill项目空间内仅项目管理员和标签负责人可以查看。 更多权限信息,请参见数据萃取权限列表。			

您可以在历史已添加列表页签,查看该来源主表中已添加的标签。

新增列表(0)	历史已添加列表(2)			
标签名称		标签英文名	所属类目	公开状态
公司名称		c name	其他标签	公开
ID		c id	其他标签	公开

步骤三:提交标签至生产环境

在新建注册上挂标签页面,按照下图操作指引,提交注册上挂标签至生产环境。

标签信息								6 🗖 🛛
关联配置								- 1
 * 朱海主表 朱海主表 「江急運爆表 ④ 物理表 朱海主表 2.8_xtest_basic. t_car 					* 提交备注 请输入此次提	交的相关备注说明	2 , 如交更信息	
■ * 关联逻辑 ① 标弦主接 ID_type = usend1 时间分区 ds = ds		ID_vatue = benz	< ◆ 新譜		填写备注以说明) 当前项目,提交/	比次提交的变更内 成功后,即更新至	容 生产环境,请谨慎摄作 取 消	0/128 El 3] 确定并提交
添加标签								
Q、清编入搜索关键字		新增列表(2) 历史已添加列表(0)						
李段及描述			标签英文名 ①	所屬类目		公开状态		摄作
☑ land 路虎		宝马	bmw	其他标签		公开		区面
□ volkswagen 大众		魔迪	audi	其他标签		公开		ī ē
mazda 马自达	»							

后续步骤

基于您的业务场景,选择后续操作步骤:

 基于注册上挂标签可以生成衍生标签。例如,Dataphin已有年龄和性别标签,您可以基于已有标签创建20 岁以下的年轻女性的衍生标签。

操作指导:基于已打过标签的数据,筛选出需要再次打标的对象,并自定义对象新的标签。

• 基于注册上挂标签可以生成标签逻辑表,以实现为下游应用提供服务。

操作指导:创建标签逻辑表。

8.4. 创建标签逻辑表

标签逻辑表用于将单个实体ID下的多个标签,或将有关联的多个实体ID及其标签进行关联。您可以基于特定 业务场景下的实体ID和标签创建标签逻辑表,同步标签逻辑表至业务数据库,后续引入至应用进行营销投放 圈选或报表分析等。本文为您介绍如何创建、配置及提交标签逻辑表。

应用场景

标签逻辑表的典型应用场景说明如下。



营销推送场景:同步标签逻辑表至业务数据库后,即可基于Quick Audience等应用进行营销数据分析及圈选投放。

例如,某企业新推出一款高端美白系列护肤品,企业需要筛选出潜在顾客进行试用。假设已有顾客的年龄、价格偏好和护肤偏好的标签,您就可以将这些标签关联生成标签逻辑表。您可以基于Quick Audience 平台对标签逻辑表进行圈选年龄在20~35岁之间、护肤偏好为美白、价格偏好为800~1500元的顾客群体。

数据分析场景:同步标签逻辑表至业务数据库后,即可通过Quick BI等智能服务平台进行可视化数据分析。Quick BI更多信息,请参见通过仪表板分析数据。

例如,基于顾客的居住地和护肤偏好的标签生成标签逻辑表。同步标签逻辑表至业务数据库中,即可引入至Quick BI平台,分析并展示上海、深圳、广州和北京地域顾客的护肤偏好分布。

原理介绍

标签逻辑表用于将单个实体ID下多个标签进行关联,或将有关联的多个实体ID及其标签进行关联,关联后将 实体ID字段、标签字段及时间分区字段写入至标签逻辑表。例如,基于某顾客ID已生成身高、地域和爱好的 标签,将身高、地域和爱好标签关联后,生成对应的标签逻辑表,如下图所示。



相关概念

本文涉及到相关概念的解释如下表所示。

概念名称	相关概念说明
实体ID	实体ID是对实体从不同视角的描述信息,实体包括用户、商品、企业、位置等。 例如,用户的实体ID为会员ID和手机号码、商品的实体ID为商品ID和商品产地、企业的实体ID 为国有企业和金融企业。
标签	标签用于刻画用户画像、描述商品、描述企业、描述位置等。 例如,文娱爱好消费者、美妆商品、金融企业、上海区域。
标签逻辑表	标签逻辑表用于将某实体下的多个标签进行关联。关联后将实体ID字段、标签字段和时间分区 字段写入至标签逻辑表。

前提条件

标签逻辑表是基于实体ID和标签进行创建,因此在您开始创建标签逻辑前需要完成实体ID和标签的创建:

- 如何创建实体ID, 请参见创建并管理实体ID。
- 标签包括规则类统计标签和注册上挂标签:
 - 如何创建规则类统计标签,请参见步骤一:创建规则统计标签。
 - 如何创建注册上挂类标签,请参见创建注册上挂标签。

操作流程

操作流程	描述
步骤一: 创建标签逻辑表	完成标签的基本参数的配置后,Dataphin自动为标签逻辑表生成ID字段、标签字段和分 区字段。
步骤二:配置标签逻辑表	完成标签逻辑表的调度配置和存储生命周期的配置。
步骤三:提交标签逻辑表	提交至生产环境的标签逻辑表,即可参与生产环境的调度。

步骤一: 创建标签逻辑表

- 1. 进入标签中心。
- 2. 按照下图指引,进入新建标签逻辑表对话框。



^{3.} 在新建标签逻辑表对话框, 配置参数后, 单击保存。

新	建标签逻辑表		X
	* 标签逻辑表名称		
	*标签逻辑表英文名	label_	
	* 选择ID	请选择实体ID	×
	标签取值	○ 基于Oneld ● 基于实体ID	
	*选择标签	请选择标签	~
	描述		
			11
		取消保存	

参数	是否为必填参数	描述
标签逻辑表名称	是	标签逻辑表创建后,其名称无法修改。命名规则如下: • 只能包括汉字、字母、数字、下划线和短划线(-)。 • 项目空间内唯一。 • 长度为64字符以内。
标签逻辑表英文名	是	标签逻辑表创建后,其英文名无法修改。命名规则如下: 只能包括字母、数字、下划线和短划线(-)。 项目空间内唯一。 长度为64字符以内。 Dataphin默认为标签逻辑表英文名前添加 label_,以标识标 签逻辑表。
选择ID	是	选择此前已创建的实体ID。同时,Dataphin支持选择某实体下的多个实体ID。
标签取值	是	Dataphin当前仅支持选择基于实体ID。

参数	是否为必填参数	描述
选择标签	是	基于已选择的实体ID,选择此前已创建并需要关联的标签: 只选择了一个实体ID,则标签能选到基于这个ID创建的规则统计标签、注册上挂标签。 选择多个实体ID,则标签能选到基于所选实体ID组成的复合主键关联的注册上挂标签。 关于注册标签的更多信息,请参见创建注册上挂标签。
描述	否	填写对标签逻辑表的简单描述。

4. 保存成功后, Dat aphin自动进入标签逻辑表详情页面。在标签逻辑表详情页面, 查看标签逻辑表的字段。

标签逻辑表的字段包括ID字段(已选择实体ID)、标签字段(已选择的标签)和分区字段(ds),如下 图所示。

▶ 标签逻辑表	label_test_ test_w 基于实体ID取值	规则统计类标签:1	行为情好类标签:0 注册上挂类标签:0				
						添加时间 >	+ 新建字段
名称		最近更新人	所属类目	属性	描述	最近添加时间	操作
ID字段(1)							
	zm zmtest	Supe	Angleson 20	实体ID	-	2021-05-19 15:45:13	Ť
标签字段(1)							
	test007 test007	Supe	december (1998)	统计类标签	-	2021-05-19 15:45:13	i a se
分区字段(1)							
	时间分区 ds	Supe	And a second	string	-	2021-05-19 15:45:13	Ť

步骤二:配置标签逻辑表

1. 在标签逻辑表详情页面,单击页面上方的调度配置。在调度配置面板,配置参数。

i. 配置基本信息区域的参数。

• 基本信息		
名称	label_lfs_test_	
bi点节		
节点类型	MAX_COMPUTE_SQL	
负责人	· · · · ·	
描述	请输入50个字符内的描述文案	

⑦ 说明 Dataphin自动生成名称、节点ID和节点类型的参数,不支持修改。

参数	描述
负责人	支持转交负责人。如果任务运行过程中触发告警,且监控配置指定责任人为告警 信息的接收对象,则Dataphin会及时通知到该任务的负责人,便于您及时处理 异常。
	负责人其他的操作权限取决于负责人的角色。关于角色对应的操作权限的信息, 请参见 <mark>用户角色和权限</mark> 。
描述	支持修改标签逻辑表已有的描述,也支持为标签逻辑表添加描述。

ii. 配置调度配置区域参数。

• 调度配置				
时间属性	● 正常调度 ○ 空跑调度			
暂停调度	● 否 是			
调度周期	⊟ ∨ 00:00 ⊙			
cron 表达式	000**?			
最近的生成时间	2021-06-11 00:00:00 2021-06-12 00:00:00	2021-06-13 00:00:00	2021-06-14 00:00:0	00
	2021-06-15 00:00:00			
优先级	中等优先级			
参数	描述			

参数	描述
时间属性	 选择任务在生产环境调度的时间属性。时间属性包括正常调度和空跑调度: 正常调度:按照调度周期的时间配置调度,并正常执行,通常任务默认选中该项。 空跑调度:按照调度周期的时间配置调度,但都是空跑执行,即一调度到该任务便直接返回成功,没有真正的执行任务。 适用场景说明如下: 正常调度:适用于任务需要正常调度以产出数据的场景。 空跑调度:适用于任务有段时间不需要产出数据,但不能阻塞下游任务执行的场景。
暂停调度	即可暂停该任务及该任务的下游任务的调度,一旦调度到该任务会直接返回失 败,不会执行。通常用于某个任务暂时不用执行,但后面还会继续使用的场景。
调度周期	 调度周期可选择日、周和月:,您可以根据业务需要,指定任务运行的具体时间点。调度周期说明如下: 日调度,即调度任务每天自动运行一次。新建周期任务时,默认的时间周期为每天0点运行一次。 周调度,即调度任务每周的特定几天,在特定时间点自动运行一次。 月调度,即调度任务在每月的特定几天,在特定时间点自动运行一次。 ⑦说明 周调度和月调度中,如果您没有指定任务的运行日期,为保证下游实例正常运行,Dataphin会每天生成实例后直接设置为运行成功,而不会真正执行任何逻辑,也不会占用资源。
优先级	优先级定义了同一时间待调度任务的优先级。默认选择为 中等优先级 。您可以 根据业务需求修改任务调度的优先级。Dataphin支持的优先级包括最低优先 级、低优先级、中等优先级、高优先级和最高优先级。

iii. 配置依赖关系区域的参数后,单击确定。

依赖关系 自动解析]	anin'S ¹ aanin'		. Antiol Sta
上游依赖 物理节点	逻辑表节点 ①			手动添加上游
输出名称'	节点名	节点ID (实例ID)	负责人	操作
Data_Distill _test9	ger hammilie	12	_ SuperAutor Sup	1
Data_Distilljb	oi_sta 3	1,000,000	Transfer Strategies (1) Traggering (1)	Ţ.
当前节点				
输出名称	节点名	节点ID (实例ID)	负责人	操作
ld_distill.label_lfs_test	label_lfs_	-	And States	
				取消 确;
参数	描述			
上游依赖	 単击自动解词 赖节点的规则 选择的标: 物理表节, 物理表节, ⑦ 注 如果需要添加 如果需要添加 如果需要添加 的物理节点。 	新, Dataphin根据标则不同。说明如下: 签为注册上挂标签,「 点或汇总逻辑表节点; 意如果来源物理表 hin可能解析不出来, 签为规则统计标签,「 节点。 加其他节点作为当前标 Dataphin不支持到	签逻辑表选择标签的类都 Dataphin根据注册上挂机 为上游依赖的物理节点或 影的输出名称非 项目名 需要您手动添加该来源4 自动解析出标签结果表书 示签的上游节点,则需要	型,目动解析上游体 示签的来源表解析出 就逻辑表节点。 ·表名 样式,则 物理表为上游依赖 5点为上游依赖的标

参数	描述
当前节点	当前节点为您展示标签逻辑表数据输出的物理节点。当前节点名称即标签逻辑表 名称,标签逻辑表名称的格式为LD_distill.label_标签逻辑表英文名,不支持 修改。关于标签逻辑表的结构、数据预览等信息,请参见物理表详情。 如果标签逻辑表已提交且被其他任务所依赖,则单击操作列下的 图标,可以 查看输出节点的下游节点。

2. 在**标签逻辑表**页面,单击页面上方的**物理化配置**。在**物理化配置**面板,配置标签逻辑表的存储生命周期后,单击确认。

•物理化配置	
*存储生命周期 36000	
快速选择 18	30 365 3600 <mark>36000</mark> 天
* 选择分区字段 11 ds	
super Laminson Super Laminson Super Laminson Super Laminson	
	取消通知
参数	描述
存储生命周期	存储生命周期用于定义标签逻辑表在Dataphin的生命周期。存储生命周期对所有的标签逻辑表生效,默认36000天。您可以自定义标签逻辑表存储的生命周期,Dataphin对存储生命周期没有最大限制。 Dataphin支持超级管理员、项目管理员和开发者角色的用户设置标签逻辑表的存储 生命周期。

步骤三:提交标签逻辑表

选择分区字段

- 1. 在标签逻辑表详情页面,单击页面右上方的 🛛 图标。
- 2. 在提交备注对话框,填写备注信息,单击确定并提交,即可将标签逻辑表提交至生产环境。

Dataphin默认展示ds,不支持修改。

后续步骤

基于您的业务诉求,选择后续的操作步骤:

- 在运维中心查看并运维标签逻辑表的萃取任务,保证任务的正常运行。具体操作,请参见萃取任务。
- 查询标签逻辑表的业务数据产出情况,如下图所示。如何查询业务数据,请参见查询并下载数据。

userid 🗉	high 🖪	town 🖫	prefer 🗔	ds 🖫
1	170	徐州	女装	20210406
2	172	武汉	女装	20210406
3	175	杭州	女装	20210406
4	178	上海	女装	20210406
5	179	北京	女装	20210406
6	170	大同	女装	20210406
7	180	南京	女装	20210406
8	185	南昌	女装	20210406

通过逻辑表输入组件和业务数据库对应的输出组件构建同步任务,同步标签逻辑表至业务数据库。具体操作,请参见配置离线单条管道。

基于标签逻辑表进行二次数据分析,包括数据建模或代码任务的研发。具体操作,请参见新建维度或新建资源。