

智能数据构建与管理 Dataphin 数据开发

ALIBABA CLOUD

文档版本: 20220708



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。	⚠ 危险 重置操作将丢失用户配置数据。
⚠ 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) 注意 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	<mark>0</mark> 8
2.规范建模	12
2.1. 概述	12
2.2. 规范定义-维度	13
2.2.1. 新建维度	13
2.2.2. 查看维度逻辑表	21
2.2.3. 编辑维度	21
2.2.4. 继承维度	23
2.2.5. 下线和删除维度	23
2.3. 规范定义-业务过程	26
2.3.1. 新建业务过程	26
2.3.2. 编辑业务过程	28
2.3.3. 克隆业务过程	29
2.3.4. 创建逻辑表	30
2.3.5. 查看相关事实逻辑表	31
2.3.6. 删除业务过程	32
2.4. 逻辑表-维度逻辑表	33
2.4.1. 添加关联维度、属性和子维度	33
2.4.2. 编辑主表	38
2.4.3. 编辑模型关系	43
2.4.4. 定义公用计算逻辑	44
2.4.5. 编辑字段	47
2.4.6. 物理化配置	50
2.4.7. 调度配置	51
2.4.8. 参数配置	56
2.4.9. 查看历史信息	58

2.4.10. 下线和删除维度逻辑表	59
2.5. 逻辑表-事实逻辑表	60
2.5.1. 创建并配置事实逻辑表	60
2.5.2. 物理化配置	70
2.5.3. 修改事实逻辑表	72
2.5.4. 调度配置	73
2.5.5. 查看历史信息	77
2.5.6. 下线和删除事实逻辑表	78
2.6. 规范定义-原子指标	79
2.6.1. 新建原子指标	79
2.6.2. 修改原子指标	84
2.6.3. 查看同源原子指标	86
2.6.4. 克隆原子指标	86
2.6.5. 查看相关派生指标	88
2.6.6. 创建派生指标	88
2.6.7. 下线和删除原子指标	89
2.7. 规范定义-业务限定	93
2.7.1. 新建业务限定	93
2.7.2. 克隆业务限定	96
2.7.3. 查看同源业务限定	97
2.7.4. 查看相关派生指标	98
2.7.5. 修改业务限定	99
2.7.6. 下线和删除业务限定 1	00
2.8. 规范定义-派生指标 1	02
2.8.1. 新建派生指标 1	02
2.8.2. 修改派生指标 1	03
2.8.3. 下线和删除派生指标 1	05
2.8.4. 查看相关汇总逻辑表 1	06

	2.9. 逻辑表-汇总逻辑表	107
	2.9.1. 新建汇总逻辑表	107
	2.9.2. 编辑汇总逻辑表	108
	2.9.3. 物理化配置汇总逻辑表	115
	2.9.4. 查看汇总逻辑表信息	116
	2.9.5. 删除汇总逻辑表	118
	2.10. 规范建模-建模引擎	118
3	.编码研发	122
	3.1. 数据开发概述	122
	3.2. 新建资源	127
	3.3. 新建用户自定义函数	128
	3.4. 新建MAXC任务	131
	3.4.1. 新建SQL任务	131
	3.4.2. 新建MapReduce任务	135
	3.4.3. 新建Spark任务	137
	3.5. 开发流批一体任务	139
	3.5.1. 新建镜像表	139
	3.5.2. 新建元表	143
	3.5.3. 新建自定义源	147
	3.5.4. 新建代码模板	150
	3.5.5. 创建Flink SQL任务	154
	3.5.6. 创建Flink Template SQL任务	167
	3.5.6.1. 步骤一: 创建并调试Flink Template SQL计算任务	167
	3.5.6.2. 步骤二: 配置并提交Flink Template SQL任务	170
	3.5.7. 创建Flink DataStream计算任务	180
	3.6. 新建ADB任务	190
	3.6.1. 新建ADB for PostgreSQL任务	190
	3.6.2. ADB for PostgreSQL支持的语法	192

3	.7. 新建通用任务	205
	3.7.1. 创建、配置并提交离线计算模版	205
	3.7.2. 新建Shell任务	212
	3.7.3. 新建Virtual任务	215
	3.7.4. 创建Python计算任务	217
	3.7.5. Dataphin PyODPS	221
3	.8. 创建并配置机器学习任务	223
3	.9. 调度配置	232
3	.10. 管理计算任务	237
4.即	〕席查询	243
4	.1. 概述	243
4	.2. 查询并下载数据	243
4	.3. 管理即席查询	247
5.S	QL使用说明	250
5	.1. MaxCompute SQL与函数支持	250
	5.1.1. 常用命令	250
	5.1.2. SQL使用限制	252
	5.1.3. 数据类型	253
	5.1.4. 运算符	257
	5.1.5. DDL语句	261
	5.1.6. DML操作	262
	5.1.7. DQL操作	263
	5.1.8. SQL增强操作	265
	5.1.9. 函数操作	265

1.概述

数据开发是Dataphin的核心功能,支持规范模型、数据处理和主题式查询。同时,数据开发模块对接了 Dataphin的权限管理、运维和发布模块,致力于为您构建便捷、高效的数据开发平台。

背景信息

如果Dataphin实例为智能研发版,则数据开发支持规范建模、数据处理和即席查询。

如果Dataphin实例为基础研发版,则系统支持数据处理和即席查询,不支持规范建模功能。

功能介绍

Dataphin旨在为您提供工具化、自动化、智能化和价值化的智能数据建设与治理的平台:

 规范建模:基于已明确的业务需求,从业务顶层开始设计,规范定义数据的标准及计算口径。同时,根据 业务需求完成数据模型的设计。

规范建模为您提供可视化、标准化、规范化和自动化定义数据,而非传统手动写SQL代码的方式,这样就 避免了统计指标计算口径不一致的情况。可视化构建数据仓库的模型,逻辑表提交后即生成对应智能黑盒 调度任务,自动化生产数据。

⑦ 说明 仅绑定业务板块的项目支持规范建模功能。

- 数据处理:通过编写代码的方式构建复杂的数据模型、构建同步任务和构建不同类型的代码任务。
- 即席查询:即主题式查询,面向业务主题的数据查询,屏蔽了物理模型中技术特性带来的影响,基于逻辑 模型从业务视角出发对外提供查询服务。

Dataphin支持Basic和Dev-Prod的数据开发模式:

- Basic项目绑定了Basic业务板块,系统支持规范建模功能。Basic项目绑定了Prod业务板块,则系统不 支持规范建模功能,仅支持数据处理和即席查询功能。
- Dev项目支持规范建模、数据处理和即席查询功能。Prod项目支持规范建模和数据处理功能。

页面介绍

⑦ 说明 如果您的Dataphin实例为基础研发版,则不支持规范建模。

Dev项目

数据开发·概述

≡ 👖 Dataphin · 🗃	发 开发	t≠ 发布 运 —	维权限 3		0 Q i	局搜索	ाह्य के 🕹 🖏
Dev V2 #RESize## avatestrate	次迎信 和 2 章 动	吏用Dataphin!					
 ↔ HINES 5 C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA C SALA<th></th><th>✔ 规范定义 ≫</th><th>(4) (4) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1</th><th>正 ◎ 不会江語</th><th>▲ 原子描标 ○</th><th>正 ◎</th><th></th>		✔ 规范定义 ≫	(4) (4) (1) (1) (1) (1) (1) (1) (1) (1) (1) (1	正 ◎ 不会江語	▲ 原子描标 ○	正 ◎	
₽ ♣ □ □	0	数据建模 >>	日本の注	→ 事实逻辑表 I ③	Ĵ	Ĵ	
		智能计算		登録資産》 当行乃進於夏根東 时间正确有序地生	和武道位理产生的政强任务,以确保所有任务都有 产教编。	接頭	
				计算任务			同步任务
<u> </u>	0	□ 任务研发 》	MAXC_SQL .		FLINK_SQL		DATAX 同步
			MAXC_MR	PYTHON	FLINK_DATASTREAM		i manue isto.
	0	《文件引用》	MAXCEEst ©	FUNKKAX MEV	- - - - - - - ○	で た に 注 度 で の	-12 元录 □ ◎
	0	Q. 即席查询 》	SQL 董词	• •			

功能区编号	描述
0	当前打开的项目名称,以及项目的环境类型(Prod或Dev)。您可以单击项目名称后的 ✓图标,在下拉列表中,依次选择项目分类(Prod、Dev或Basic)及项目名称,切换 项目。同时,您也可以在搜索框中输入项目名称关键词,快速搜索项目,提高了搜索的 效率。
2	环境类型标识按钮,单击 图标,切换Dev和Prod项目。Prod项目表示生产环境的项 目,Dev项目表示开发环境项目。 例如,当前您访问的是Dev项目(编号②的环境类型标识为 Prod项目(),单击 图标后, 切换至Prod项目(编号②的环境类型标识为 Prod项目)。 ⑦ 说明 Basic项目没有环境类型标识按钮。
3	快速新建规范定义、逻辑表、数据处理和即席查询任务的入口。
4	刷新任务列表、新建文件和新建文件夹的按钮。
6	输入任务名称的关键字的方式,快速搜索任务。
6	已创建的文件夹、任务和文件列表。
\bigcirc	左侧导航栏提供的规范建模和数据处理的相关操作入口。
8	数据开发模块提供萃取、集成、发布、运维和权限模块的入口。

功能区编号	描述
	在全局搜索框中输入物理表、逻辑表、规范定义、函数或资源等名称的关键字,搜索目 标对象。您可以单击 研发、萃取和资产 页签,快速搜索目标对象。
9	同时,您也可以单击 <mark>冒</mark> 图标,在 代码搜索 框中,输入代码的关键字快速搜索代码。
	⑦ 说明 系统仅支持搜索版本升级后(2020年7月14日)提交过的代码。
0	展示了 规范定义 包含的元素及相互关系,并提供了创建 规范定义 的入口。
1	展示了 数据建模 包含的元素,同时为您提供新建 事实逻辑表 和 汇总逻辑表 的入口。
12	展示了 任务研发 包含的 计算任务、计算任务 和 同步任务 的类型,并提供了创建任务的 入口。
(3)	展示了 文件引用 包含的函数管理、自定义源、资源管理、代码模板和元表,并提供了 创建文件的入口。
(14)	即席查询 仅包含 SQL查询 ,并提供了创建SQL查询的入口。

● Baisc项目

○ Baisc(非Data_distill)与Dev项目页面功能相似, Baisc项目不支持发布功能。

≡	Dataphin ·	研发	开发≠	运维	权限							昆 谷 🗢 🔍
Basi v2_ 规		✓< ■ ■	欢迎使用	Dataphin!								
* ** &	维度 ○、输入关键字 ▶ 重用户域 (9)	Ċ ₪ ;;		✔ 规范建义 》				业务过程 1 ◎				秋 減生調時 已 ◎
 <!--</th--><th></th><th></th><th></th><th>🕏 数振建模 >></th><th></th><th> 二、方法 注意支援表 ④ </th><th></th><th> ● ● </th><th></th><th></th><th></th><th>ご (1) この注明表 ① ◎</th>				🕏 数振建模 >>		 二、方法 注意支援表 ④ 		 ● ● 				ご (1) この注明表 ① ◎
			•	WHEN-			80	→ 還推測算 运行及時 时间正計	【》 許逆環境和調 有序地生产者	其最处理产生的数据任务,以确保所有任务都有投3 其最。	e	
			■ 任务研发 ≫		SPARK MR SOL	 憲法计算任務 ・ ・ ・ 	SHELL PYTHON VIRTUAL		实时计算任务 FLINK_SQL FLINK_TEMPLATE_SQL FLINK_DATASTREAM		同步任新 DATAX 同步 Pipeline 同步	
				▲▲ 文件引用 >>		直接 或试计 得 函数 ● ◎	A A A A A A A A A A A A A A A A A A A	3		交通 発展再現 日 ②	2 60 实时计算模板 ⊡ ②	\$251计展元表 □ 0
	维度对象列表	,		Q. 即席查询 >>		SQL 查询	÷ ©					

○ Dat a_dist ill项目

Data_distill和其他Baisc项目页面功能相似,但Data_distill项目不支持规范建模功能。

● Prod项目

Prod和Dev项目页面功能相似,但Prod项目仅支持查看数据处理和规范建模模块的资源(资源在Dev项目开发并发布到Prod项目),不支持发布和即席查询功能。

SQL语法解析器的关键字

SELECT, DELETE, INSERT, UPDATE, FROM, HAVING, WHERE, ORDER, BY, GROUP, INTO, AS, CREATE, ALTER , DROP, SET, NULL, NOT, DISTINCT, TABLE, TABLESPACE, VIEW, SEQUENCE, TRIGGER, USER, INDEX, SESSION, PROCEDURE, FUNCTIO N. PRIMARY, KEY, DEFAULT, CONSTRAINT, CHECK, UNIQUE, FOREIGN, REFERENCES, EXPLAIN, FOR, IF, ADD, ALL, UNION, EXCEPT, INT ERSECT, MINUS, INNER, LEFT, RIGHT, FULL, OUTER, JOIN, ON, SCHEMA, CAST, COLUMN, USE, DATABASE, TO, AND, OR, XOR, CASE, WHEN . THEN, ELSE, ELSIF, END, EXISTS, IN, NEW, ASC, DESC, IS, LIKE, ESCAPE, BETWEEN, VALUES, INTERVAL, LOCK, SOME, ANY, TRUNC ATE, RETURN, TRUE, FALSE, LIMIT, KILL, IDENTIFIED, PASSWORD, DUAL, BINARY, SHOW, REPLACE, PERIOD, BITS, WHILE, DO, LEAVE , ITERATE, REPEAT, UNTIL, OPEN, CLOSE, OUT, INOUT, EXIT, UNDO, SQLSTATE, CONDITION, DIV, WINDOW, OFFSET, ROW, ROWS, ONLY , FIRST, NEXT, FETCH, OF, SHARE, NOWAIT, RECURSIVE, TEMPORARY, TEMP, UNLOGGED, RESTART, IDENTITY, CONTINUE, CASCADE, RES TRICT, USING, CURRENT, RETURNING, COMMENT, OVER, TYPE, ILIKE, RLIKE, FULLTEXT, START, PRIOR, CONNECT, WITH, EXTRACT, CURSOR, MER GE, MATCHED, ERRORS, REJECT, UNLIMITED, BEGIN, EXCLUSIVE, MODE, WAIT, ADVISE, SYSDATE, DECLARE, EXCEPTION, GRANT, R EVOKE, LOOP, GOTO, COMMIT, SAVEPOINT, CROSS, PCTFREE, INITRANS, MAXTRANS, INITIALLY, ENABLE, DISABLE, SEGMENT, CREATION, IMMEDIATE, DEFERRED, STORAGE, MINEXTENTS, MAXEXTENTS, MAXSIZE, PCTINCREASE, FLASH CACHE, CELL FLASH CACHE, NONE, LOB, STORE 、 CHUNK、 CACHE、 NOCACHE、 LOGGING、 NOCOMPRESS, KEEP DUPLICATES, EXCEPTIONS, PURGE, COMPUTE, ANALYZE, OPTIMIZE, TOP, ARRAY, DISTR IBUTE, EOF, ERROR, IDENTIFIER, HINT, VARIANT, LITERAL INT, LITERAL FLOAT, LITERAL HEX, LITERAL CHARS, LITERAL NCHARS, LITERA L ESCAPE, LITERAL PATH, LITERAL ALIAS, LINE COMMENT, MULTI LINE COMMENT, BINARY FLOAT, BINARY DOUBLE, PARTITION, PARTITIONED, OVERWR ITE, SORT, SEL, LOCKING, ACCESS, VOLATILE, MULTISET, POSITION, RANGE N, FORMAT, QUALIFY, MOD, CONTAINS, CONCAT, UPSERT . LPAREN, RPAREN, LBRACE, RBRACE, LBRACKET, RBRACKET, SEMI, COMMA, DOT, DOTDOT, DOTDOTDOT, EQ, GT, LT, LT SUB GT, BANG, BANGBANG, BANG TILDE, BANG TILDE STAR, BANG TILDE TILDE STAR, TILDE, TILDE STAR, TILDE STAR, TILDE EQ, QUES, QUESQUES, QUESBAR, QUESAMP, COLON, COLONCOLON, COLONEQ, EQEQ, EQGT, LTEQ, LTEQGT, LTGT, GTEQ, BANGEQ, BANGGT, BAN GLT, AMPAMP, BARBAR, BARBARSLASH, BARSLASH, PLUS, SUB, SUBGT, SUBGTGT, STAR, SLASH, AMP, BAR, CARET, CARETEQ, PERCENT, LTLT, GTG T. MONKEYS AT. MONKEYS AT AT. POUND, POUNDGT, POUNDGTGT, MONKEYS AT GT, LT MONKEYS AT, NOT EXTEND RIGHT, NOT EXTEND LEFT, D ISTRIBUTED, WITHIN, TILDE TILDE, BANG TILDE TILDE, RANDOMLY, ENCODING, STATISTICS, INCLUSIVE, SUBPARTITION, EVERY, RENAME, WIT HOUT, INVALIDATE, METADATA, IMPORT, SERVER, OPTIONS

2.规范建模

2.1. 概述

规范建模是基于数据仓库理论的工具化,致力于提供体系化和系统化的数据建模的功能。

页面介绍

执行如下操作,进入数据规范建模模块:

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>>。
- 3. 进入**维度**页面:
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。
 - iii. 在数据开发页面,单击规范建模。

在数据规范建模页面,支持数据的规范定义及建模。

≡	Dataphin · 研发	开发 ≠ 发布	运维 权限		Q 全局搜索	3	ස් 👆 ග වේ 💿
Dev v2_	→ → Prod Dev	欢迎使田Dataphin!					
规范廷	模 ① 数据处理 即席查询	Meelen Baaphin.	(
2	原子指标 🖒 🖸	0			₼		
~	Q 输入关键字 2	✓ 规范定义 ≫				-	V
	▼ 💽 公共城 (4)		維度	业务过程	原子損标 ▶ ◎	业务限定	派生描标 + ③
^	🔺 age 🛛 🖓 🗄		白叶山泽		î	1	
۲	🔺 age1		H 3/13134				
~	▲ cd_lvl1_id	۵					
~	▲ z_age	● 数据建模 ≫	2	~			V
			维度逻辑表	事实逻辑表			汇总逻辑表
			0				+ 0
<u>×</u>		智能计算		运维调度 >			
			\longrightarrow	运行及难护逻辑表机 时间正确有序地生产	I政播处埋广生的政播任务,以确保所有 「数据。	9仕労部有按照 <	
							A21.J

编号	描述
0	在左侧导航栏中,选择需要创建的规范定义和逻辑表。在对应的模块后,就可以开始创 建规范定义和逻辑表。
2	在搜索框中输入关键字进行快速搜索规范定义和逻辑表。
3	在搜索框内输入任务关键词,全局搜索任务。但仅购买了 数据萃取 功能的用户,可以查 询数据萃取对象。
4	选择对应 规范定义 下的 图标,新建规范定义。
5	选择数据建模下的。图标,新建逻辑表。

数据建模的前提

- 数据建模前,需要将外部数据引入到项目绑定的数据源中,详情请参见数据引入。
- 当您进入数据建模研发模块后,需要选择规范建模的项目空间。Dataphin支持的数据开发模式包括Basic 模式和Dev-Prod模式,详情请参见创建Basic项目和新建业务板块。
- 仅绑定业务板块的项目可以通过规范建模功能在项目空间进行开发任务。如绑定错误或未绑定业务板块, 则规范建模功能不可用。关于项目和业务板块的绑定,请参见创建Basic项目。
- Basic模式下,Basic项目绑定了Basic业务板块后,支持执行规范建模操作。如果Basic项目绑定了Prod业务 板块,则不支持执行规范建模操作,仅可以使用数据处理和即席查询功能。
- Dev-Prod模式下, Dev项目可以直接创建、编辑、删除任务操作。新建完成并提交到待发布列表中, 需要执行发布操作。发布成功后可以在对应Prod下进行自动调度。

2.2. 规范定义-维度

2.2.1. 新建维度

维度是人们观察事物的角度,是确定事物的多方位、多角度、多层次的条件和概念。本文为您介绍如何新建 维度。

前提条件

完成创建数据域,详情请参见新建数据域。

背景信息

维度即进行数据统计的对象。通常情况下,维度是实际存在、不因事件发生而存在的实体,例如时间维度、 地区维度、产品维度等。创建维度,即从顶层规范业务中实体(或称主数据)的存在性及唯一性。

新建维度的权限限制,详情请参见数据开发权限列表。

操作步骤

- 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>>。
- 3. 进入新建维度页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。

ii. 在数据开发页面,按照下图操作指引,进入新建维度页面。

≡	Dataphin · 研发	开发 ≓ 发布 运维	权限				Q、 全局提素		1 🛍 -	ෂ 🕹 ව
	17297_test02_dev	欢迎使用Dataphin!					((
~ ⊛	(修設度 ① 1日) 3 2 私人犬田子 → 11 cross_pro (2) → 11 gubh_regression (1)	4	ieg ⊡ ©		注程 ⊗	E:R I	● ● ● ● ● ● ● ● ● ●	±ekaz T		↓ 減标 ⊗
* *	→ ■ ■ new (1) → ■ leat (1) → ■ 這時重加時社 (2)	◆ 数据建模 >>	2010年20日 建築建築株英 ②	→ 	щж ⊘					
8		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		*	运續 導度 》 运行及维护逻辑表 时间正确有序地生)	Q数据处理产生的数据任 "数据。	委,以魏梁所有任务都有按照 ————————————————————————————————————			
				801	计算任务	MIDTIN			同步任务 Piceline 同時	
		□ 任侮研发 ≫		ode		SHELL			T Iplania 1925	
				MR		PYTHON				
	ignes:46-744	经文件	31用 >>		Had	2009££\$X ⊙	く 没理答理 ・ ②			

4. 在新建维度页面, 配置参数。

i. 配置维度基本信息区域参数。

↓ 新建维度 ×	: 8
维度基本信息	
业务板块	所屋项目
项目空间类型应用层	 * 数据域 请选择数据域 >>
★ 维度英文名 dim_ 请输入维	葉英文名 ★ 維度名称 请输入维度名称
维度描述 请输入维度描述	
	0/128
参数	描述
	选择该维度所属的数据域。
数据域	⑦ 说明 数据域的划分,详情请参见划分数据域。
	维度英文名由字母、数字或下划线(_)组合组成,且默认以dim_作为前缀。
维度英文名	⑦ 说明 维度英文名不能超过60个字符。
	维度名称由汉字、字母、数字、下划线(_)或短划线(-)组合组成。
维度名称	⑦ 说明 维度英文名不能超过64个字符。
维度描述	填写对维度简单的描述。

ii. 配置维度逻辑信息区域参数。

维度逻辑信息用于定义维度对象范围和对象特征,确保维度真实且唯一性存在。Dataphin为您提供四种维度类型,以满足不同维度对象建设需求。

腹逻辑信息					
	1 设置维度类型			② 主键定义	
	e		Q	€	
	普通维度	普通维度(层级)①	枚举维度	虚拟维度	

维度类型	描述
普通维度	需定义主键及维度主从关系。普通维度创建完成并已成功发布后自动生成对应的 维度逻辑表。
普通维度(层级)	需定义主键、来源逻辑及层级,支持使用多个维度进行分析,不同维度之间存在 稳定的层级关系。维度创建完成并已成功发布后自动生成一组有层级关联关系的 维度逻辑表。
枚举维度	需定义主键,支持列举及标准化枚举值以标识枚举维度对象。
虚拟维度	需定义主键,支持没有具体业务实体承载、没有可固化数据范围逻辑定义的维度 对象,也可以规范化业务建模。

■ 选择**普通维度**,配置参数信息如下。

维度逻辑信息									
				💄 普通約	主度			主键&维度主从关系定义	
* 主键英文名	请输入维度主任	建英文名		* 主键名称	请输入维度主键名称		★ 主键类型	STRING	
* 主键计算逻辑	🔓 格式化	⊘ 参考示例	同 规范性校验						
	1						Line Karles		
所屬主維度	• 否 0 是					更改维度关型			

参数	描述			
主键英文名	主键英文名由字母、数字或下划线(_)组合组成。			
	⑦ 说明 主键英文名不能超过64个字符。			
	主键名称由汉字、数字、字母、下划线(_)或短划线(-)组合组成。			
主键名称	⑦ 说明 主键名称不能超过64个字符。			
主键类型	主键类型包括STRING、BIGINT、DOUBLE、DATETIME和DECIMAL。			
主键计算逻辑	主键类型包括STRING、BIGINT、DOUBLE、DATETIME和DECIMAL。 填写维度主键计算逻辑: a. 在主键计算逻辑区域,编辑SQL代码,示例如下。 select province from dataphin_test where ds='\${bizdate}'; 您也可以单击主键计算逻辑后的参考示例,查看示例代码。 b. 单击主键计算逻辑后的规范性校验,可以校验您编写的SQL代码是否符合规范。 如果规范校验不通过,您可以单击格式化,系统自动调整SQL代码的格式。 ⑦ 说明 SQL代码中需要限制时间分区条件,否则可能出现运维报错或数据错误。			
所属主维度	选择所属的主维度。如果选择是,则需单击主维度显示框,在下拉列表中选择 所属的主维度。			

■ 选择**普通维度(层级)**,在配置向导页面,完成以下操作。

£逻辑信息	
	■ 普通修業(環境) ② 直接起来源意思规定义 ····································
 bL 及び設計工作の公式 bL (1) 	主観名称 前紙入用度主観名称 ID ・主観実型 STRING ∨ ・主衆教名 ∨ ①
参数	描述
	主键英文名由字母、数字或下划线(_)组合组成。
主键英文名	⑦ 说明 主键英文名不能超过64个字符。
	维度名称由汉字、字母、数字、下划线(_)或短划线(-)组合组成。
主键名称	⑦ 说明 主键名称 不能超过64个字符。
主键类型	主键类型包括STRING、BIGINT、DOUBLE、DATETIME和DECIMAL。
	选择主表。
主表表名	⑦ 说明 建议您选择生产环境中的表。

a. 在**主键&来源逻辑定义**页签, 配置参数。

b. 单击下一步。

c. 在**层级定义**页签, 配置参数。

進度逻辑信息		·····································	3 章级定义		
主键英文名 ddd_id	主键名称 rdlD	主婦與型 STRING	已透主责 dataphin_online_test.customer_1		
 ・子環字段 ・ 屋段定文 2 	 • 父辰字段 ✓ びあり个层段 生成計子μ度 ● 着 	* 名称字段			
 - 最級領市法定义 ● 権式化 ● 参示例 ● 単示例 ・ 約約5522年 ● 権式化 ● 参示例 ○ 規定性 	582 ①		8 9		
		预克 返回上一步			
参数	描述				
子层字段	选择层级字段中的	子层字段。			
父层字段	选择层级字段中的	父层字段。			
名称字段	选择名称字段。				
层级定义	填写层级的个数,系统支持最多9个层级。				
生成叶子维度	选择是否生成叶子 如果选择是,则 如果选择 舌 ,则	维度: 系统会自动生成D 需要您定义叶子约	├子维度。 ŧ度的生成逻辑。		
层级根节点定义	基于来源主表字段, a. 在 层级根节点 is_parent 您也可以单击 b. 单击 规范性校 如果规范校验 格式。	,设定层级根节点 京定义 区域,编辑 ='true' *参考示例,查看 * * * * * * · 和 <u>词</u> 过,您可以自	(即无父层级)的判断条件: SQL代码,示例如下。 代码示例。 编写SQL代码是否符合规范。 单击 格式化 ,系统自动调整SQL代码的		
时间分区逻辑	基于来源主表字段, a. 在 时间分区逐 ds=\${bizd 您也可以单击 b. 单击 规范性校 格式。	,设定数据更新及 逻辑区域,编辑SC ^{late}} 参 考示例 ,查看《 2验 ,可以校验您 :不通过,您可以自	获取时间周期的判断条件: LI代码,示例如下。 代码示例。 编写SQL代码是否符合规范。 单击 格式化 ,系统自动调整SQL代码的		

■ 选择**枚举维度**,配置参数信息如下。

维度逻辑信息				
+ 枚挙進度			主键定义	
设置教举信息				
① 参考示例		code (编号)	value (取值)	
F友		F	*	
			-	
	-			

在**设置枚举信息**的左侧区域中,填写枚举维度成员与对应成员的信息,并以英文逗号(,)隔开,示例如下。

F,	女
Μ,	男

在设置枚举信息的右侧区域,为您展示系统自动解析的code和value。

■ 选择**虚拟维度**,配置参数信息如下。

地反应相口心	
	⑦ 虚拟追责
◆ 主願美文名 ddfa	 ・主線名称 订单 ・主線共型 STRING 更改速度映型
参数	描述
主键英文名	主键英文名由字母、数字或下划线(_)组合组成。 ⑦ 说明 主键英文名不能超过64个字符。
主键名称	维度名称由汉字、字母、数字、下划线(_)或短划线(-)组合组成。 ⑦ 说明 主键名称不能超过64个字符。
主键类型	主键类型包括STRING、BIGINT、DOUBLE、DATETIME和DECIMAL。

⑦ **说明** 如果虚拟维度为维度逻辑表或事实逻辑表的关联维度,则虚拟维度可以作为派生指标的统计粒度。

5. 在新建维度页面,单击页面右上角的圆图标,保存维度。

- 6. 提交维度。
 - i. 在新建维度页面,单击页面右上角的 《图标。

ii. 在提交备注页面,填写备注信息。

iii. 单击确定并提交,完成维度的提交。

提交成功的**普通维度和层级普通维度(层级)**,系统自动生成对应维度逻辑表和一组具有层级关联关系的维度逻辑表。

- 7. (可选)发布维度。
 - 如果您的开发模式是Dev-Prod模式,则需要发布新建的维度,详情请参见发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度,即可进入生产环境。

2.2.2. 查看维度逻辑表

本文为您介绍如何查看基于某个维度生成的维度逻辑表。

前提条件

完成创建维度,详情请参见新建维度。

背景信息

- 系统仅支持普通维度和普通维度(层级),查看维度逻辑表。
- 系统仅支持提交成功的维度,查看维度逻辑表。
- 查看维度逻辑表的权限限制,详情请参见数据开发权限列表。

操作步骤

1. 登录Dataphin控制台。

- 2.
- 3.
- 在**维度**页面,查询需要查看维度逻辑表的维度。
 您可以在搜索框中,输入维度名称关键字,快速搜索维度。
- 5. 单击需要查看维度逻辑表的维度。
- 6. 在**维度属性**框中,鼠标悬停至 图标后,单击**维度逻辑表**,查看维度逻辑表详情。

您也可以通过以下方式,查看维度逻辑表详情:

- i. 在**维度**页面,单击左侧功能区域下方的**维度对象列表**。
- ii. 在对象列表页的维度页签,查询需要查看维度逻表的维度。
 快速查询维度:
 - 在搜索框中, 输入维度名称关键字, 快速搜索维度。
 - 选中某个**所属项目**,快速筛选。
 - 单击 図标后,选中某个数据域、状态、最近更新人、创建人或维度类型,快速筛选。

iii. 单击需要查看维度逻辑表的维度操作列下的民图标,查看维度逻辑表详情。

2.2.3. 编辑维度

本文为您介绍如何编辑维度。

前提条件

完成创建维度,详情请参见新建维度。

背景信息

编辑维度的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3.
- 4. 在维度页面,查询需要编辑的维度。

您可以在搜索框中, 输入维度名称关键字, 快速搜索维度。

5. 鼠标悬停至需要编辑的维度后,再将鼠标悬停至 图标,单击编辑。

您也可以通过以下方式,进入编辑维度页面:

- 通过属性框的方式:
 - a. 单击需要编辑的维度。
 - b. 在维度属性框中, 单击编辑。
- 。 通过对象列表页的方式:
 - a. 在维度页面,单击左侧区域下方的维度对象列表。
 - b. 在**对象列表页的维度**页签,查询需要编辑的维度。 快速查询维度:
 - 在搜索框中, 输入维度名称关键字, 快速搜索维度。
 - 选中某个**所属项目**,快速筛选。
 - 单击 <>>> 图标后,选中某个数据域、状态、最近更新人、创建人或维度类型,快速筛选。
 - c. 单击需要编辑的维度操作列下☑图标。
- 6. (可选)解锁维度。
 - 如果维度已被您自己锁定,则无需解锁维度。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上角的 圖图标,解锁维度。

当@图标变更为@图标,解锁成功。

- 7. 在编辑维度页面,编辑参数,详情请参见新建维度。
- 8. 在编辑维度页面,保存和提交维度。
 - i. 单击页面右上方的 图标, 保存维度。
 - ii. 单击页面右上方的 《图标, 提交维度。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成维度的提交。
- 9. (可选)发布维度。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度,即可进入生产环境。

2.2.4. 继承维度

本文为您介绍如何基于已有维度创建其子维度(继承维度)。

前提条件

- 完成维度的创建,详情请参见新建维度。
- 维度对应的维度逻辑表已提交。

背景信息

- 系统仅支持普通维度继承维度。
- 系统仅支持已提交维度逻辑表对应的维度继承维度。
- 继承维度的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1.
- 2.
- 3.
- 4. 在**维度**页面,查询需要继承的维度。

您可以在搜索框中,输入维度名称关键字,快速搜索维度。

- 5. 单击需要继承的维度。
- 6. 在维度属性框中, 鼠标悬停至 图标, 单击继承。

您也可以通过以下方式,进入继承维度:

- i. 在维度页面, 单击左侧区域下方的维度对象列表。
- ii. 在对象列表页的维度页签,查询需要继承的维度。
 快速查询维度:
 - 在搜索框中, 输入维度名称关键字, 搜索维度。
 - 选中某个**所属项目**,筛选维度。
 - 单击 <>> 图标后,选中某个数据域、状态、最近更新人、创建人或维度类型,筛选维度。

iii. 单击需要继承的维度操作列下。图标。

7. 在继承维度页面, 配置参数, 详情请参见新建维度。

其中,数据域不支持修改,所属主维度默认为需要继承的维度,您也可以修改所属主维度。

2.2.5. 下线和删除维度

本文为您介绍如何下线、下线并删除和删除维度。

前提条件

完成创建维度, 详情请参见新建维度。

背景信息

- 维度的状态的说明:
 - 保存成功的维度,状态为草稿。
 - 提交成功的维度, 状态为**已提交**。
 - 编辑已提交状态的维度,并成功保存,状态更改为开发中。
 - 已提交状态的维度,成功下线后,状态更改为草稿。
- 下线、删除和下线并删除维度的限制:
 - 权限限制,详情请参见数据开发权限列表。
 - 仅开发中和已提交状态的维度, 支持下线操作。
 - 仅草稿状态的维度,支持删除操作。
 - 仅开发中和已提交状态的维度, 支持下线并删除操作。

下线维度

⑦ 说明 系统仅支持下线不存在依赖的维度。

- 1. 登录Dataphin控制台。
- 2.
- 3.
- 4. 在**维度**页面,查询需要下线的维度。

您可以在搜索框中,输入维度名称关键字,快速搜索维度。

- 5. (可选)解锁维度。
 - 如果维度已被您自己锁定,则无需解锁维度。
 - 如果维度已被他人锁定,则需要解锁维度:
 - a. 鼠标悬停至需要下线的维度后,再将鼠标悬停至 图标,单击编辑。
 - b. 在编辑维度页面,单击页面右上方的 图标。

当@图标变更为@图标,解锁成功。

6. 鼠标悬停至需要下线的维度后,再将鼠标悬停至图标,单击下线。

您也可以通过以下方式,进入下线维度的提示对话框:

- 通过属性框的方式:
 - a. 单击需要下线的维度。
 - b. 在**维度属性**框中, 鼠标悬停至 图标, 单击下线。
- 通过对象列表页的方式:
 - a. 在维度页面, 单击左侧区域下方的维度对象列表。

b. 在对象列表页的维度页签, 查询需要下线的维度。

快速查询维度:

- 在搜索框中, 输入维度名称关键字, 快速搜索维度。
- 选中某个**所属项目**,快速筛选。
- 单击 <>>> 图标后,选中某个数据域、状态、最近更新人、创建人或维度类型,快速筛选。
- c. 鼠标悬停至需要下线的维度操作列下下图标,单击下线。
- 7. 在提示对话框,填写备注信息。
- 8. 单击确定并提交,完成下线维度。

删除维度

- 在**维度**页面,查询需要删除的维度。
 您可以在搜索框中,输入维度名称关键字,快速搜索维度。
- 2. (可选) 解锁维度。
 - 如果维度已被您自己锁定,则无需解锁维度。
 - 如果维度已被他人锁定,则需要解锁维度:
 - a. 鼠标悬停至需要删除的维度后,再将鼠标悬停至图标,单击编辑。
 - b. 在编辑维度页面,单击页面右上方的 ·· 图标。

当@图标变更为@图标,解锁成功。

3. 鼠标悬停至需要删除的维度后,再将鼠标悬停至图标,单击删除。

您也可以通过以下方式,进入删除维度的提示对话框:

- 通过属性框的方式:
 - a. 单击需要删除的维度。
 - b. 在维度属性框中, 鼠标悬停至 图标, 单击删除。
- 。 通过对象列表页的方式:
 - a. 在维度页面, 单击左侧区域下方的维度对象列表。
 - b. 在**对象列表页的维度**页签,查询需要删除的维度。 快速查询维度:
 - 在搜索框中, 输入维度名称关键字, 快速搜索维度。
 - 选中某个**所属项目**,快速筛选。
 - 单击 <>> 图标后,选中某个数据域、状态、最近更新人、创建人或维度类型,快速筛选。
 - c. 鼠标悬停至需要删除的维度操作列下下图标,单击删除。
- 4. 在提示对话框,填写备注信息。
- 5. 单击确定并提交,完成删除维度。

下线并删除维度

- ⑦ 说明 系统仅支持下线并删除不存在依赖的维度。
- 在**维度**页面,查询需要下线并删除的维度。
 您可以在搜索框中,输入维度名称关键字,快速搜索维度。
- 2. (可选)解锁维度。
 - 如果维度已被您自己锁定,则无需解锁维度。
 - 如果维度已被他人锁定,则需要解锁维度:
 - a. 鼠标悬停至需要下线并删除的维度后,再将鼠标悬停至图标,单击编辑。
 - b. 在编辑维度页面,单击页面右上方的 ·· 图标。

当@图标变更为@图标,解锁成功。

3. 鼠标悬停至需要下线并删除的维度后,再将鼠标悬停至图标,单击下线并删除。

您也可以通过以下方式,进入下线并删除维度的提示对话框:

- 通过属性框的方式:
 - a. 单击需要下线并删除的维度。
 - b. 在维度属性框中, 鼠标悬停至 图标, 单击下线并删除。
- 。 通过对象列表页的方式:
 - a. 在维度页面, 单击左侧区域下方的维度对象列表。
 - b. 在**对象列表页的维度**页签,查询需要下线并删除的维度。 快速查询维度:
 - 在搜索框中, 输入维度名称关键字, 快速搜索维度。
 - 选中某个所属项目,快速筛选。
 - 单击 ≥图标后,选中某个数据域、状态、最近更新人、创建人或维度类型,快速筛选。
 - c. 鼠标悬停至需要下线并删除的维度操作列下下图标,单击下线并删除。
- 4. 在提示对话框,填写备注信息。
- 5. 单击确定并提交,完成下线并删除维度。

2.3. 规范定义-业务过程

2.3.1. 新建业务过程

业务过程是业务活动中不可拆分的事件,例如下单、支付和退款。创建业务过程,即从顶层视角,规范业务 活动中事件的内容类型及唯一性。本文为您介绍如何新建业务过程。

前提条件

完成创建数据域, 详情请参见新建数据域。

背景信息

新建业务过程的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入业务过程页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,按照下图操作指引进入新建业务过程页面。



4. 在新建业务过程对话框,配置参数。

新建业务过	过程			×
业务板块	v2_7_be	所属项目 v2_7_beijing_projec	项目空间类型应用层	
* 数据域	~			
* 英文名	请输入英文名, :	允许字母、数字或下划线		
* 名称	请输入业务过程	名称		
描述				
				0/128
			取消	提交
参数		描述		
数据域		选择数据域。		

参数	描述
	英文名由字母、数字、下划线(_)组合组成。
英文名	⑦ 说明 英文名不能超过57个字符。
	名称由汉字、数字、字母、下划线(_)或短划线(-)组合组成。
名称	⑦ 说明 名称不能超过64个字符。
描述	填写简单的描述。

- 5. 提交业务过程。
 - i. 单击提交。
 - ii. 在提交备注页面,填写备注信息。

iii. 单击确定并提交,完成业务过程的提交。

- 6. (可选)发布业务过程。
 - 如果您的开发模式是Dev-Prod模式,则需要发布新建的业务过程,详情请参见发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的业务过程,即可进入生产环境。

2.3.2. 编辑业务过程

业务过程是业务活动中不可拆分的事件,例如下单、支付和退款。本文为您介绍如何编辑业务过程。

前提条件

完成创建业务过程, 详情请参见新建业务过程。

背景信息

编辑业务过程的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入**业务过程**页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目为您的数据开发空间,则不需要 选择项目。

iii. 在数据开发页面,单击规范建模。如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。

- iv. 在规范建模页签, 单击左侧导航栏 · 业务过程图标。
- 在业务过程页面,查询需要编辑的业务过程。
 您可以在搜索框中,输入业务过程名称关键字,快速搜索业务过程。
- 5. 鼠标悬停至需要编辑的业务过程后,再将鼠标悬停至图标,单击编辑。

您也可以通过以下方式,进入编辑业务过程对话框:

- 通过属性框的方式:
 - a. 单击需要编辑的业务过程。
 - b. 在业务过程属性框中, 单击编辑。
- 。 通过对象列表页的方式:
 - a. 在**业务过程**页面, 单击左侧区域下方的**业务过程对象列表**。
 - b. 在对象列表页的业务过程页签,查询需要编辑的业务过程。快速查询业务过程:
 - 在搜索框中, 输入业务过程名称关键字, 搜索业务过程。
 - 选中某个**所属项目**,筛选业务过程。
 - 单击 <>> 图标后,选中某个最近更新人、创建人或数据域,筛选业务过程。
 - c. 单击需要编辑的业务过程操作列下应图标。
- 6. 在编辑业务过程对话框, 配置参数, 详情请参见新建业务过程。
- 7. 提交业务过程。
 - i. 单击提交。
 - ii. 在提交备注对话框,填写备注信息。
 - iii. 单击确定并提交,完成业务过程的提交。
- 8. (可选)发布业务过程。
 - 如果您的开发模式是Dev-Prod模式,则需要发布业务过程,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的业务过程,即可进入生产环境。

2.3.3. 克隆业务过程

本文为您介绍如何克隆业务过程。

前提条件

完成创建业务过程,详情请参见新建业务过程。

背景信息

克隆业务过程的权限限制,详情请参见数据开发权限列表。

操作步骤

1. 登录Dataphin控制台。

2.

- 3. 进入业务过程页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
 - iii. 在数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
- 在业务过程页面,查询需要克隆的业务过程。
 您可以在搜索框中,输入业务过程名称关键字,快速搜索业务过程。
- 5. 单击需要克隆的业务过程。
- 6. 在**业务过程属性**框中,鼠标悬停至 图标,单击**克隆**。

您也可以通过以下方式,进入新建业务过程对话框:

- i. 在**业务过程**页面, 单击左侧区域下方的**业务过程对象列表**。
- ii. 在对象列表页的业务过程页签,查询需要克隆的业务过程。
 快速查询业务过程:
 - 在搜索框中, 输入业务过程名称关键字, 搜索业务过程。
 - 选中某个**所属项目**,筛选业务过程。
 - 单击

 图标后,选中某个最近更新人、创建人或数据域,筛选业务过程。
- iii. 鼠标悬停至需要克隆的业务过程操作列下下图标后,单击克隆。
- 7. 在新建业务过程对话框, 配置参数, 详情请参见新建业务过程。
- 8. 提交业务过程。
 - i. 单击提交。
 - ii. 在提交备注对话框,填写备注信息。
 - iii. 单击确定并提交,完成业务过程的提交。
- 9. (可选)发布业务过程。
 - 如果您的开发模式是Dev-Prod模式,则需要发布业务过程,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的业务过程,即可进入生产环境。

2.3.4. 创建逻辑表

本文为您介绍如何基于业务过程创建事实逻辑表。

前提条件

完成创建业务过程, 详情请参见新建业务过程。

背景信息

> 文档版本: 20220708

创建逻辑表的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入业务过程页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
 - iii. 在数据开发页面,单击规范建模。如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
 - iv. 在规范建模页签, 单击左侧导航栏 业务过程图标。
- 4. 在**业务过程**页面,查询需要创建逻辑表的业务过程。
 - 您可以在搜索框中,输入业务过程名称关键字,快速搜索业务过程。
- 5. 单击需要创建逻辑表的业务过程。
- 6. 在**业务过程属性**框中,鼠标悬停至,图标,单击**创建逻辑表**。

您也可以通过以下方式,进入新建事实逻辑表对话框:

- i. 在业务过程页面,单击左侧区域下方的业务过程对象列表。
- ii. 在对象列表页的业务过程页签,查询需要创建逻辑表的业务过程。
 快速查询业务过程:
 - 在搜索框中, 输入业务过程名称关键字, 搜索业务过程。
 - 选中某个**所属项目**,筛选业务过程。
 - 单击 ◎ 图标后,选中某个最近更新人、创建人或数据域,筛选业务过程。
- iii. 鼠标悬停至需要创建逻辑的业务过程操作列下属图标。
- 7. 在新建事实逻辑表对话框, 配置参数, 详情请参见创建并配置事实逻辑表。
- 8. 提交逻辑表。
 - i. 单击提交。
 - ii. 在提交备注对话框,填写备注信息。
 - iii. 单击确定并提交,完成逻辑表的提交。
- 9. 发布逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的逻辑表,即可进入生产环境。

2.3.5. 查看相关事实逻辑表

本文为您介绍如何查看某个业务过程的相关事实逻辑表。

前提条件

完成创建业务过程, 详情请参见新建业务过程。

背景信息

查看业务过程的相关事实逻辑表的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入业务过程页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
 - iii. 在数据开发页面,单击规范建模。如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
 - iv. 在**规范建模**页签, 单击左侧导航栏 22 业务过程图标。
- 4. 在业务过程页面,单击左侧区域下方的业务过程对象列表。
- 在对象列表页的业务过程页签,查询需要查看相关事实逻辑表的业务过程。
 您可以通过以下方式,快速查询业务过程:
 - 在搜索框中, 输入业务过程名称关键字, 搜索业务过程。
 - 选中某个**所属项目**,筛选业务过程。
 - 单击 診图标后,选中某个最近更新人、创建人或数据域,筛选业务过程。
- 6. 单击需要查看相关事实逻辑表的业务过程操作列下属图标,查看相关事实逻辑表。

在**相关事实逻辑表**对话框,您可以单击某个事实逻辑表。在该事实逻辑表的主表页面,查看详情。

2.3.6. 删除业务过程

本文为您介绍如何删除业务过程。

前提条件

完成业务过程的创建,详情请参见新建业务过程。

背景信息

- 系统仅支持删除无相关事实逻辑表的业务过程。
- 删除业务过程的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入**业务过程**页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
 - iii. 在数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
 - iv. 在规范建模页签, 单击左侧导航栏**型业务过程**图标。
- 4. 在业务过程页面,查询需要删除的业务过程。

您可以在搜索框中,输入业务过程名称关键字,快速搜索业务过程。

5. 鼠标悬停至需要删除的业务过程后,再将鼠标悬停至 图标,单击删除。

您也可以通过以下方式,进入删除业务过程的提示对话框:

- 通过属性框的方式:
 - a. 单击需要删除的业务过程。
 - b. 在业务过程属性框中, 鼠标悬停至 图标, 单击删除。
- 通过对象列表页的方式:
 - a. 在业务过程页面, 单击左侧区域下方的业务过程对象列表。
 - b. 在对象列表页的业务过程页签,查询需要删除的业务过程。快速查询业务过程:
 - 在搜索框中, 输入业务过程名称关键字, 搜索业务过程。
 - 选中某个**所属项目**,筛选业务过程。
 - 单击 <>>> 图标后,选中某个最近更新人、创建人或数据域,筛选业务过程。
 - c. 鼠标悬停至需要删除的业务过程操作列下下图标,单击删除。
- 6. 在提示对话框,填写备注信息。
- 7. 单击确定并提交,完成业务过程的删除。

2.4. 逻辑表-维度逻辑表

2.4.1. 添加关联维度、属性和子维度

维度提交成功后,系统会自动创建与维度一一对应的维度逻辑表。同时,系统支持对维度逻辑表添加属性、 关联维度和子维度。

前提条件

完成维度的提交,详情请参见新建维度。

背景信息

- 添加关联维度、属性和子维度的权限限制,详情请参见数据开发权限列表。
- 系统仅支持已提交或开发中状态的维度逻辑表,添加子维度。

添加关联维度

? 说明 添加关联维度的限制:

- 如果事实逻辑表的数据时效为天,则系统仅支持添加数据时效为天的维度。
- 如果事实逻辑表的数据时效为小时,则系统支持添加数据时效为天或小时的维度。
- 如果事实逻辑表的数据时效为分钟,则系统支持添加数据时效为天、小时或分钟的维度。

1. 登录Dataphin控制台。

- 2.
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
 - iii. 在数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
 - iv. 在规范建模页签, 单击左侧导航栏 4 度逻辑表图标。
- 4. 在维度逻辑表页面,查询需要关联维度的维度逻辑表。

您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。

- 5. 鼠标悬停至需要关联维度的维度逻辑表后,再将鼠标悬停至 图标,单击编辑。
- 6. (可选)解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上角的高图标,解锁维度逻辑表。

当@图标变更为@图标,解锁成功。

- 7. 进入新建关联维度对话框。
 - 如果该维度逻辑表未添加关联维度,则在该维度逻辑表的主表页面,单击添加关联维度。
 - 如果该维度逻辑表已添加关联维度,则在该维度逻辑表的主表页面,鼠标悬停至已关联的某个字段, 单击回图标。
 - ◎ 无论该维度逻辑表已添加还是未添加关联维度,您都可以通过以下方式,进入新建关联维度对话

框:

a. 在维度逻辑表的主表页面,单击页面上方的主表编辑。

b. 在主表编辑页面,鼠标悬停至新建字段上,单击关联维度性。

- 8. 添加关联维度。
 - i. 在新建关联维度对话框, 配置参数。

新建关联维展								\times
关联维度信息								
* 关联维度信息	请输入关键词				\sim			
维度类型		主键及主键计算逻辑	8					
编辑关联逻辑								
维度表需关联	字段				同 关联	已选关联维度		
请输入关键)	词		\sim	,	-0 //4/			
编辑维度角色	1							
* 角色英文名	dim_							
*角色名称								
默认值设定①								
∗ 默认值	-110							
						取消	确兌	l

参数	描述			
关联维度信息	选择需要关联的维度。			
编辑关联逻辑	选择维度表需要关联的字段。			
编辑维度角色	基于关联维度信息,系统为您展示 角色英文名 和 角色名称 。您也可以编辑 角色 英文名和 角色名称 。			
默认值设定	如果未关联维度时,则系统自动关联字段的补值设定为-110。			

- ii. 单击确定。
- 9. 保存、提交和发布维度逻辑表。
 - i. 单击页面右上方的 图标, 保存维度逻辑表。
 - ii. 单击页面右上方的 《图标, 提交维度逻辑表。
 - iii. 在提交备注页面,填写备注信息。

- iv. 单击确定并提交,完成维度的提交。
- v. 发布维度逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

添加属性

- 在**维度逻辑表**页面,查询需要添加属性的维度逻辑表。
 您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。
- 2. 鼠标悬停至需要添加属性的维度逻辑表后,再将鼠标悬停至图标,单击编辑。
- 3. (可选) 解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上方的高图标,解锁维度逻辑表。

当@图标变更为@图标,解锁成功。

- 4. 进入新建属性对话框。
 - 如果该维度逻辑表未添加属性,则在该维度逻辑表的主表页面,单击添加属性。
 - 如果该维度逻辑表已添加属性,则在该维度逻辑表的主表页面,鼠标悬停至已添加的某个属性字段, 单击回图标。
 - 无论该维度逻辑表已添加还是未添加属性,您都可以通过以下方式,进入新建属性对话框:
 a. 在维度逻辑表的主表页面,单击页面上方的主表编辑。
 - b. 在**主表编辑**页面,鼠标悬停至新建字段,单击属性。
- 5. 添加属性。
i. 在新建属性对话框, 配置参数。

新建属性						×
来源表 💿 引ん	入字段 OSL自定义字段					
来源表	10.00.00	v 0	新建字段			
	✓ 系統默认会将项目空间替换为变量形式使用替 览。\${cathy_basic_test}.s_order		字段英文名	字段名称	数据类型	操作
关联逻辑字段	order	v 0	来源物理表:			
选择新增字段	(注:可以从不同物理表选择字段,同一字段可以添	加多次)	order.	商 始之称。	BIGINT	. C 🖻
字段及描述	Q 请输入搜索关键字			000000 per 12 **		
orde			create 原始英文名: create_time	原始名称:	STRING #	. C 🖻
010		.				
crea		0				
pay		0				
suc		0				
		-				
					Re 204	周右并结论
					取 消	採仔州校验

- 如果来源表选中引入字段,则执行以下操作,添加属性:
 - a. 单击来源表显示框区域,在下拉列表中,选择来源表。

⑦ 说明 建议选择生产环境的表。以防任务调度运维时,该对象在生产环境不存在导致操作失败,或者导致自动解析依赖关系失败。

- b. 单击关联逻辑字段显示框区域,在下拉列表中,选择关联逻辑字段。
- c. 选择新增字段。

您也可以在搜索框中,输入字段名称的关键字,快速搜索字段。

- d. 单击字段后•图标。
- e. 在新建字段区域,填写字段的字段英文名和字段名称。

对已添加的字段,您可以单击操作列下的。图标、 图标或 图标,查看字段的计算逻辑、编辑 字段或删除字段。

- 单击保存并校验。
- 如果来源表选中SQL自定义字段,则执行以下操作,添加属性:
 - a. 在代码的编写区域,编写SQL代码。

您可以单击参考示例,查看SQL代码示例。

b. 单击**规范性校验**,校验SQL代码的语法及权限。

如果规范性校验失败,您可以单击格式化,系统会自动调整代码格式。

- c. 单击保存并校验。
- 6. 保存、提交和发布维度逻辑表。

i. 单击页面右上方的画图标,保存维度逻辑表。

ii. 单击页面右上方的 《图标, 提交维度逻辑表。

- iii. 在提交备注页面,填写备注信息。
- iv. 单击确定并提交,完成维度的提交。
- v. 发布维度逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

添加子维度

- 在维度逻辑表页面,查询需要添加子维度的维度逻辑表。
 您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。
- 2. 鼠标悬停至需要添加子维度的维度逻辑表后,再将鼠标悬停至 图标,单击编辑。
- 3. (可选) 解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上方的△图标,解锁维度逻辑表。

当@图标变更为@图标,解锁成功。

- 4. 在该维度逻辑表的主表页面,单击添加子维度。
- 5. 在继承维度页面, 配置参数, 详情请参见继承维度。
- 6. 保存、提交和发布维度逻辑表。
 - i. 单击页面右上方的 图标, 保存维度逻辑表。
 - ii. 单击页面右上方的 🛛 图标, 提交维度逻辑表。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成维度的提交。
 - v. 发布维度逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

2.4.2. 编辑主表

本文为您介绍如何编辑主表添加属性、添加关联维度、定义计算逻辑等操作。

前提条件

完成维度的创建, 详情请参见新建维度。

添加属性

- 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>>。
- 3. 进入维度逻辑表页面:
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。

ii. 单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在左侧导航栏中, 单击 图标。

- 4. 在维度逻辑表左侧功能区域,单击某个维度逻辑表,进入该维度逻辑表的详情页面。
- 5. (可选)解锁维度逻辑表:
 - 如果维度逻辑表已被您自己锁定,则不需要解锁。
 - 如果维度逻辑表已被他人锁定,则需要单击页面右上方的♀图标,解锁维度逻辑表。
- 6. 在维度逻辑表的详情页面,单击页面上方的主表编辑。
- 7. 在主表编辑页面,将鼠标悬停在新建字段上后,单击属性。
- 8. 根据新建属性对话框提示,添加属性,详情请参见添加关联维度、属性和子维度。
- 9. 保存、提交和发布维度逻辑表:
 - i. 完成属性添加后,单击编辑维度逻辑表页面上方的图图标,保存添加的属性。
 - ii. 单击编辑维度逻辑表页面上方的 图标, 提交维度逻辑表。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成维度逻辑表的提交。
 - v. (可选)发布维度逻辑表:
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布维度逻辑表。

添加关联维度

- 1. 在维度逻辑表的详情页面,单击页面上方的主表编辑。
- 2. 在主表编辑页面,将鼠标悬停在新建字段上后,单击关联维度。
- 3. 根据新建关联维度对话框提示,添加维度,详情请参见添加关联维度、属性和子维度。
- 4. 保存、提交和发布维度逻辑表:
 - i. 完成关联维度添加后, 单击编辑维度逻辑表页面上方的图图标, 保存添加的关联维度。
 - ii. 单击编辑维度逻辑表页面上方的 图标, 提交维度逻辑表。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成维度逻辑表的提交。
 - v. (可选)发布维度逻辑表:
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布维度逻辑表。

定义公用计算逻辑

1. 在维度逻辑表的详情页面,单击页面上方的主表编辑。

- 2. 在主表编辑页面, 单击新建字段后的定义公用计算逻辑。
- 3. 根据定义公用计算逻辑对话框提示,单击新建公用计算逻辑。

定义公用计算逻辑 ①		×
Q 请输入搜索关键字	公用计算逻辑	
十 新建公用计算逻辑	定义:逻辑表内可公共使用的SQL片段,即可沉淀并定义为"公用" 临时表或者中间表效果。	计算逻辑",以保证相同的数据处理逻辑得到快捷、方便的复用,达到
	Eg:编辑逻辑表A时,字段order_id,order_type都需要过滤无效常 1)定义逻辑表A的"公用计算逻辑"。将过滤无效商品的SOL片段员 1 s_item 是为了实现过滤无效商品以得到所需数据。而定 2 select item_id, 3 substr(item_name, 1, 64) as item_name 4 from 2)编辑字段时引用"公用计算逻辑"。字段 order_id, order_type的	新品,故而: E义为公用计算逻辑 [s_item] 定义的公用计算逻辑 计算逻辑中引用 [s_item]
暂无任何公用计算逻辑 请新建以便未来引用	1 order_id string 2 select order_id 3 from onedata.s_order a 4 join s_item b 5 on a.item_id-b.item_id 6 where ds='\${bizdate}' 道用场景: 逻辑表内编辑多个字段计算逻辑时,不同字段间会有机段间会存在常见且公共使用的 SQL片段,如定义相同的from及wh 注意事项: 如果逻辑表的"公用计算逻辑"与逻辑表所属项目的薄距离、已存在字 新建保存'公用计算逻辑时请关注是否有项目内重名表、已存在字	1 order_type string 2 select order_item_name 3 from onedata.s_order a 4 join s_item b 5 on a.item_id b.item_id 6 where ds='\${bizdate}' 8回的数据处理场景,如特定数据过滤、数据集合并,所以表内跨字 ere 条件、引用相同的参表 join 或 union后构建的子查询。 表集室名,字段引用的对象会被优先识别为逻辑表的"公用计算逻辑", r段SQL是否会误用"公用计算逻辑",如字段只是引用物理表,请务必
✓ 全选		
		<mark>完成</mark> 点击"完成",仅"已保存"内容会生效

- 4. 根据对话框提示,在文本输入框中输入计算逻辑名称后,单击确认。
- 5. 根据对话框提示,编写计算逻辑的SQL语句:
 - i. 单击参考示例。
 - ii. 根据参考示例,编辑SQL语句。
 - iii. 编写完成后,单击**规范性校验**,检验SQL语句是否符合规范

如果不符合规范,则单击格式化,系统自动调整您编写的SQL语句,使得其符合规范。

iv. 规范校验通过后,单击保存。



同时,您可以删除已有的计算逻辑,具体操作步骤如下:

选中需要删除的计算逻辑,或单击对话框下方的全选。

- 单击批量删除。
- 在删除公用计算逻辑对话框中,单击确认。
- 在二次弹出的删除公用计算逻辑对话框中, 单击OK。
- 6. 单击对话框下方的完成,完成计算逻辑的定义。
- 7. 保存、提交和发布维度逻辑表:
 - i. 完成计算逻辑定义后,单击编辑维度逻辑表页面上方的圆图标,保存定义的计算逻辑。
 - ii. 单击编辑维度逻辑表页面上方的 图标, 提交维度度逻辑表。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成维度逻辑表的提交。
 - v. (可选)发布维度逻辑表:
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布维度逻辑表。

查看计算逻辑

- 1. 在维度逻辑表的详情页面,单击页面上方的主表编辑。
- 在主表编辑页面,单击按类型分组页签,查看主键、关联维度、属性和系统字段。单击按来源分组页签,查看主键字段、系统字段和来源逻辑表。
- 在按类型分组或按来源分组页签,单击需要查看计算逻辑的字段所在行计算逻辑(ALL)的 图标, 查看字段的计算逻辑。

您可以通过以下任何一种方式,快速筛选字段:

- 单击计算逻辑 (ALL) 后的 图标, 选择字段的计算逻辑状态。
- 在搜索框中输入字段名称的关键字。

编辑字段

⑦ 说明 系统仅支持修改关联维度的字段。

您可以编辑主键字段,单击主键字段所在行操作列下 图标,进入**维度详情**页面修改参数,详情请参见编辑 维度。

您可以编辑关联维度字段,操作步骤如下:

- 1. 在维度逻辑表的详情页面,单击页面上方的主表编辑。
- 2. 在主表编辑页面,单击按类型分组或按来源分组,筛选字段。
- 3. 在按类型分组或按来源分组页签,单击需要编辑字段所在行操作列下的 图标。
- 4. 在编辑字段对话框中,根据提示修改字段英文名、字段名称、描述和数据类型。

编辑字段				×
Q 请输入搜索关键词	* 字段英文名	order_id		
ds	* 字段名称	uf		
order_id	描述	请输入字段描述		
				0/128
	数据类型	DOUBLE V		
	更新人/时间 创建人/时间	angen ingenaanse n.e. n.een		
				保存
			取消	确定

5. 单击确定,完成字段的修改。

6. 在确认关闭对话框中,单击确定。

7. 保存、提交和发布维度逻辑表:

i. 完成字段编辑后,单击编辑维度逻辑表页面上方的图图标,保存定义的计算逻辑。

- ii. 单击编辑维度逻辑表页面上方的 图标, 提交维度逻辑表。
- iii. 在提交备注页面,填写备注信息。
- iv. 单击确定并提交,完成维度逻辑表的提交。

- v. (可选)发布维度逻辑表:
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布维度逻辑表。

删除字段

⑦ 说明 系统仅支持删除关联维度的字段。

- 1. 在维度逻辑表的详情页面,单击页面上方的主表编辑。
- 2. 在主表编辑页面, 单击按类型分组或按来源分组, 筛选字段。
- 3. 在按类型分组或按来源分组页签,单击需要删除字段所在行操作列下的 同图标。
- 4. 在确认对话框中,单击确定,完成字段的删除。

2.4.3. 编辑模型关系

本文为您介绍如何编辑关联维度的模型关系。

前提条件

- 完成维度的提交,详情请参见新建维度。
- 维度逻辑表已添加关联维度,详情请参见添加关联维度、属性和子维度。

背景信息

- 系统仅支持编辑关联维度字段的维度模型关系。
- 编辑模型关系的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目为您的数据开发空间,则不需要 选择项目。

iii. 在数据开发页面,单击规范建模。

如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。

- iv. 在规范建模页签, 单击左侧导航栏 4 维度逻辑表图标。
- 在**维度逻辑表**页面,查询需要编辑维度模型关系的维度逻辑表。
 您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。
- 5. 鼠标悬停至需要编辑维度模型关系的维度逻辑表后,再将鼠标悬停至 图标,单击编辑。

- 6. (可选)解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上方的△图标,解锁维度逻辑表。

当 ▲ 图标变更为 ▲ 图标, 解锁成功。

- 7. 在该维度逻辑表的主表页面,单击主表编辑。
- 8. 编辑模型关系。
 - i. 进入编辑模型关系对话框。

您可以通过以下两种方式,进入编辑模型关系对话框:

- 类型分组:
 - a. 在按类型分组页签, 查询需要编辑模型关系的字段。

您也可以在输入框中, 输入字段名称关键字, 快速搜索需要编辑模型关系的字段。

? 说明 系统最多展示6个搜索结果。

- b. 在关联维度或属性区域,单击模型关系列下的编辑模型关系。
- 来源分组:
 - a. 在按来源分组页签,查询需要编辑模型关系的字段。
 您也可以在输入框中,输入字段名称关键字,快速搜索需要编辑模型关系的字段。

⑦ 说明 系统最多展示6个搜索结果。

b. 在来源区域, 单击模型关系列下的编辑模型关系。

ii. 在编辑模型对话框,编辑参数,详情请参见添加关联维度。

iii. 单击确定。

- 9. 保存、提交和发布维度逻辑表。
 - i. 单击页面右上方的 图标, 保存维度逻辑表。
 - ii. 单击页面右上方的 《图标, 提交维度逻辑表。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交,完成维度的提交。
 - v. 发布维度逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

2.4.4. 定义公用计算逻辑

本文为您介绍如何定义维度逻辑表的公用计算逻辑。

前提条件

完成维度的创建, 详情请参见新建维度。

背景信息

- 逻辑表的不同字段间会有相同的数据处理场景,例如数据过滤、数据集合并等。Dataphin支持定义公用计算逻辑,例如定义相同的FROM和WHERE条件。编辑逻辑表内多个字段有相同计算逻辑时,可以引用公用计算逻辑,提高编辑字段计算逻辑的效率。
- 如果逻辑表的公用计算逻辑与逻辑表所属项目的物理表重名,则字段引用的对象会被优先识别为逻辑表的 公用计算逻辑。定义公用计算逻辑时,请您关注是否与逻辑表所属项目的物理表重名。
- 定义公用计算逻辑的权限限制,详情请参见数据开发权限列表。
- 公用计算逻辑仅限维度逻辑表内的字段引用。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
 - iii. 在数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
 - iv. 在规范建模页签, 单击左侧导航栏 盘 建度逻辑表图标。
- 在**维度逻辑表**页面,查询需要定义公用计算逻辑的维度逻辑表。
 您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。
- 5. 鼠标悬停至需要定义公用计算逻辑的维度逻辑表后,再将鼠标悬停至■图标,单击编辑。
- 6. (可选)解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上方的△图标,解锁维度逻辑表。

当@图标变更为@图标,解锁成功。

- 7. 创建公用计算逻辑。
 - i. 在该维度逻辑表的主表页面, 单击**主表编辑**。
 - ii. 单击新建字段后的定义公用计算逻辑。
 - iii. 在定义公用计算逻辑对话框,单击新建公用计算逻辑。
 - iv. 在文本输入框中,输入计算逻辑名称后,单击确认。计算逻辑名称由字母、数字或下划线(_)组合组成。
 - v. 定义公用计算逻辑。
 - a. 在页面右侧代码输入框中,编写计算逻辑的SQL语句。
 您也可以单击参考示例,查看示例代码。

b. 单击**规范性校验**,检验SQL语句是否符合规范。

如果不符合规范,您可以单击格式化,系统自动调整SQL代码格式。

c. 规范校验通过后,单击保存。

对已添加的计算逻辑,您可以执行如下操作。

操作	描述
编辑	编辑公用计算逻辑: a. 在定义公用计算逻辑页面,查询需要编辑的公用计算逻辑。 您也可以在输入框中,输入公用计算逻辑的名称关键字,快速搜索公用 计算逻辑。 b. 单击需要编辑的公用计算逻辑。 c. 在页面右侧,编辑SQL代码。 d. 单击规范性校验。 如果不符合规范,您可以单击格式化,系统自动调整SQL代码语法格 式。 e. 单击保存。
删除	 删除公用计算逻辑: a. 在定义公用计算逻辑页面,查询需要删除的公用计算逻辑。 您也可以在输入框中,输入公用计算逻辑的名称关键字,快速搜索公用计算逻辑。 b. 选中需要删除的公用计算逻辑。 您也可以单击页面下方的全选。 c. 单击批量删除。 d. 在删除公用计算逻辑对话框中,单击确认。 e. 在删除公用计算逻辑对话框中,单击OK。
查看影响字段	查看引用了某个公用计算逻辑的字段: a. 在定义公用计算逻辑页面,查询需要查看引用信息的公用计算逻辑。 您也可以在输入框中,输入公用计算逻辑的名称关键字,快速搜索公用 计算逻辑。 b. 单击需要查看引用信息的公用计算逻辑。 c. 在页面右侧区域,单击影响字段。 d. 在下拉列表中,查看引用了该公用计算逻辑的字段。 您也可以在输入框中,输入字段的名称关键字,快速搜索字段。

操作	描述			
	查看最近一次提交的公用计算逻辑代码: a. 在定义公用计算逻辑页面,查询需要查看代码的公用计算逻辑。 您也可以在输入框中,输入公用计算逻辑的名称关键字,快速搜索公用			
查看已发布内容	计算逻辑。 b. 单击需要查看代码的公用计算逻辑。 c. 在页面右侧区域,单击 已发布内容 。 d. 在下拉列表中,查看最近一次提交的公用计算逻辑代码。			

vi. 单击完成。

8. 保存、提交和发布维度逻辑表。

i. 单击页面右上方的 图标, 保存维度逻辑表。

- ii. 单击页面右上方的 《图标, 提交维度逻辑表。
- iii. 在提交备注页面,填写备注信息。
- iv. 单击确定并提交,完成维度的提交。
- v. 发布维度逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

2.4.5. 编辑字段

本文为您介绍如何编辑维度逻辑表的主键字段、属性字段和关联维度字段。

前提条件

- 完成维度的创建,详情请参见新建维度。
- 完成属性字段或关联维度字段的添加,详情请参见添加关联维度、属性和子维度。

背景信息

- 维度逻辑表必须是由普通维度生成。
- 编辑维度逻辑表字段的权限限制,详情请参见数据开发权限列表。

编辑主键字段

通过编辑主键字段的方式,编辑维度逻辑表对应维度的基本信息和逻辑信息。

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。

- ii. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要选择项目。
- iii. 在数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
- iv. 在规范建模页签, 单击左侧导航栏 2 维度逻辑表图标。
- 在**维度逻辑表**页面,查询需要编辑主键字段的维度逻辑表。
 您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。
- 5. 鼠标悬停至需要编辑主键字段的维度逻辑表后,再将鼠标悬停至图标,单击编辑。
- 6. (可选)解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上方的△图标,解锁维度逻辑表。

当@图标变更为@图标,解锁成功。

- 7. 编辑主键字段。
 - i. 在该维度逻辑表的主表页面, 单击**主表编辑**。
 - ii. 在主键字段区域, 鼠标悬停至操作列下的 团图标。
 - iii. 在提示框中, 单击**维度详情**。
 - iv. 在编辑维度,编辑参数,详情请参见编辑维度。
- 8. 保存、提交和发布维度逻辑表。
 - i. 单击页面右上方的 图标, 保存维度逻辑表。
 - ii. 单击页面右上方的 《图标, 提交维度逻辑表。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交,完成维度的提交。
 - v. 发布维度逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

编辑属性字段或关联维度字段

- 在**维度逻辑表**页面,查询需要编辑属性字段或关联维度字段的维度逻辑表。
 您可以在搜索框中,输入维度逻辑表名称关键字,快速搜索维度逻辑表。
- 2. 鼠标悬停至需要编辑属性字段或关联维度字段的维度逻辑表后,再将鼠标悬停至图标,单击编辑。
- 3. (可选) 解锁维度逻辑表。
 - 如果维度已被您自己锁定,则无需解锁维度逻辑表。
 - 如果维度已被他人锁定,则需要单击编辑维度页面右上方的△图标,解锁维度逻辑表。

当@图标变更为@图标,解锁成功。

4. 在该维度逻辑表的主表页面,单击主表编辑。

在该维度逻辑表的主表页面,您也可以将鼠标悬停至需要编辑的字段后,单击 图标,进入**编辑字段**对 话框。

5. 进入编辑字段对话框。

您可以通过以下方式,进入编辑字段对话框:

- 类型分组:
 - a. 在按类型分组页签, 查询需要编辑的字段。

您也可以在输入框中, 输入字段名称关键字, 快速搜索需要编辑的字段。

⑦ 说明 系统最多展示6个搜索结果。

b. 单击关联维度或属性区域操作列下的 图标。

- 来源分组:
 - a. 在**按来源分组**页签,查询需要编辑的字段。

您也可以在输入框中,输入字段名称关键字,快速搜索需要编辑的字段。

? 说明 系统最多展示6个搜索结果。

b. 单击来源区域操作列下的 图标。

6. 在编辑字段对话框中,您可以编辑所有关联字段和属性字段的字段英文名、字段名称、描述或数据类型。

参数	描述
	字段英文名 由字母、数字或下划线(_)组合组成。
字段英文名	⑦ 说明 字段英文名不能超过64个字符。
	字段名称 由字母、数字、下划线(_)或短划线(-)组合组成。
字段名称	⑦ 说明 字段名称不能超过64个字符。
描述	字段的简单描述。
数据类型	字段的数据类型。

- 7. 单击保存,完成某个字段的编辑。
- 8. 单击确定。
- 9. 保存、提交和发布维度逻辑表。

- i. 单击页面右上方的 图标, 保存维度逻辑表。
- ii. 单击页面右上方的 《图标, 提交维度逻辑表。
- iii. 在提交备注对话框,填写备注信息。
- iv. 单击确定并提交,完成维度的提交。
- v. 发布维度逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的维度逻辑表,即可进入生产环境。

2.4.6. 物理化配置

物理化配置用于配置维度逻辑表的生命周期、分区字段及设置自定义参数。本文为您介绍如何配置维度逻辑 表的物理化参数。

前提条件

完成维度的创建,详情请参见新建维度。

操作步骤

- 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>>。
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击图标。
- 4. 在**维度逻辑表**左侧功能区域,单击某个维度逻辑表,进入该维度逻辑表的详情页面。
- 5. (可选)解锁维度逻辑表。
 - 如果维度逻辑表已被您自己锁定,则不需要解锁。
 - 如果维度逻辑表已被他人锁定,则需要单击页面右上角的∩图标,解锁维度逻辑表。
- 6. 单击页面上方的物理化配置。
- 7. 根据物理化配置对话框提示, 配置参数。

物理化配置	
物理化配置	
* 存储生命周期	7 快速选择 7 14 30 365 天
* 选择分区字段	lt ds
无需物理化字段	
自定义参数设置	
自定义参数: ①	示例: set oneservice.optimize.skewjoin = true/false; set hive.exec.dynamic.partition.mode=nonstrict; (动态分区开启/关闭) set hive.vectorized.execution.enabled = false; (矢量查询开启/关闭) set hive.vectorized.execution.reduce.enabled = false; 0/2048 取消 确定
参数	描述
存储生命周期	选择存储生命周期。系统支持的存储生命周期包括7、14、30和365天。
选择分区字段	选择表的分区。
自定义参数	配置自定义参数,详情请参见 <mark>配置自定义任务参数</mark> 。

- 8. 单击确定,完成维度逻辑表的物理化配置。
- 9. 保存、提交和发布维度逻辑表。
 - i. 完成配置自定义参数后, 单击编辑维度逻辑表页面上方的 图标, 保存物理化配置。
 - ii. 单击编辑维度逻辑表页面上方的 / 图标。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成维度逻辑表的提交。
 - v. (可选)发布维度逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布维度逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布维度逻辑表。

2.4.7. 调度配置

调度配置用于配置维度逻辑表任务节点,在生产环境调度的优先级、调度周期及调度依赖。Dataphin通过各 个节点的调度依赖的配置结果,有序的运行业务流程中各个节点,保障业务数据有效、适时地产出。本文为 您介绍如何配置逻辑表任务节点的调度优先级、调度参数、调度周期及调度依赖。

步骤一:配置逻辑表任务调度的优先级及参数

优先级定义了同一时间同一批待调度任务的优先级。

数据开发·规范建模

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏中的研发。
- 4. 在数据开发页面,按照下图引导,进入**维度逻辑表**页面。



⑦ 说明 上图中①为示例,您根据需要选择目标项目。

- 5. 在**维度逻辑表**列表页面,单击目标维度逻辑表。
- 6. 如果维度逻辑表的负责人不是您自己,则单击目标维度逻辑表页面右上角的。图标,解锁维度逻辑表。

🗼 dim	_ct_lvl1	×									: \$
表信息	调度配置	物理化配置	主表编辑▲	历史信息	去发布>					E	2 🦘 🎯
	-	×+				_					
•		±72			0 &	: (<mark>●</mark> dir 1级	antait		≈ :	
)	- 1	- 199	dataphi	UMT HIM	~ ^		👂 cc1 💼	Aato	Barris .		
		👂 ct1		18.000				*			

- 7. 在目标维度逻辑表页面,单击顶部菜单栏的调度配置。
- 8. 在调度配置页面的基本信息区域,配置逻辑表任务调度的优先级和参数。

调度配置							
基本信息							
逻辑表名称	84.4	84.3					
任务类型	MAX_COMP	UTE_SQL					
负责人	inspire/18	endoproce (1982) Real					
描述	维度逻辑表:	维度逻辑表: 物理化的调度配置					
优先级	中等优先级 🗸 ①				√ ()		
参数配置	请输入配置	请输入配置的参数,支持自定义参数,以;区隔,如keyA=123; keyB=\${yyyymmdd-7};					
in an	节点参数配置	说明					
参数		描述					
优先级		系统默认 持的优先	人选择为 中等优 无级包括最低优劣	先级 。您可以根据 先级、低优先级、	居业务需求修改任 中等优先级、高仿	务调度的优先级。 尤先级和最高优先约	系统支 级。
参数配置		参数配置 则系统调 关于更多	置用于定义逻辑》 周度任务时业务 3的参数配置信,	表任务调度的参数 日期为20210610 息,请参见 <mark>参数</mark> 配	₹。例如,定义了\$ 。]置。	{bizdate}=20210	610,

步骤二:配置调度周期

根据业务场景选择本周期节点的运行,是否需要依赖上一周期本节点或其他节点的运行结果。

- 1. 请参见步骤一: 配置逻辑表任务调度的优先级及参数, 进入调度配置页面。
- 2. 在调度配置页面, 配置调度周期区域的参数。

如果选择了需要依赖上周期本节点的运行结果,则需要配置如下参数。

 調度周期 	「競 () 依赖			
自定义 1 🗸	请输入节点ID 请输入节点Id或输出名	▲ 请选择依赖字段 请输入字段关键字	2	^
		+	3	
编号	描述			

编号	描述
0	选择节点类型。系统支持选择 自定义和本逻辑表内节点 。适用场景说明如下: • 本周期节点是否运行取决于上一周期本节点是否正常产出数据,则需要选择 本逻辑表内节点 。只有上一周期本节点或自定义节点运行成功,才会启动运行本节点。 • 维度逻辑表没有用到某个节点的产出表,但业务上需要依赖该节点的上一周期是 否正常产出数据,则需要选择依赖自定义节点。
2	 如果节点类型选择为自定义,则执行以下操作,选择依赖字段: a.单击节点ID显示框区域。 b.在节点ID列表中,选择节点ID。 您也可以在下拉列表中,选择节点ID。 您也可以在下拉列表中,选择节点ID的关键字,快速搜索节点。 c.单击依赖字段显示框区域,在下拉列表中,选中依赖字段。 您可以在下拉列表中的输入框中,输入字段名称关键字,快速查找字段。如 果所有字段需要依赖,则选中全选。 d.单击确定。 如果节点类型选择为本逻辑表内节点,则执行以下操作,选择依赖字段: a.单击依赖字段显示框区域,在下拉列表中,选中依赖字段。 您可以在下拉列表中的输入框中,输入字段名称关键字,快速查找字段。如 果所有字段需要依赖,则选中全选。 b.单击确定。
3	您可以单击 <mark>+</mark> 图标,添加多个依赖节点。

步骤三:配置调度依赖

上游依赖定义了业务流程之间的调度依赖关系。正确的调度依赖关系可以保障业务数据有效的产出。正确的 依赖关系配置后,无需关心当前节点依赖的数据表是否正常产出数据,是否影响该节点的运行,系统会自动 为您解析任务节点间的关联关系。

- 依赖关系是两个或多个节点之间的调度依赖关系,控制节点运行顺序。同时上游节点的状态将影响其他下 游节点的运行状态。
- 配置了依赖关系的节点调度规则为:首先,上游节点运行完成后,才能调度下游节点。其次,根据节点设定的调度时间判断是否执行调度。
- 您在设定的调度时间之前提交的调度配置,会在设定的调度时间之后生效。而在设定的调度时间之后配置 的依赖关系,只能间隔一天再生成实例。
 - 1. 请参见步骤一: 配置逻辑表任务调度的优先级及参数, 进入调度配置页面。
 - 2. 在调度配置页面, 配置调度依赖区域的参数。

调度依赖默认开启自动解析,系统自动解析该维度逻辑表的物化代码中依赖关系。同时,您也可以手动 添加上游依赖的物理节点和逻辑表节点。

自动解析 (●) 是 ①	▶ 解析输入输出
上游依赖 物理节点 逻辑	● 解析问题提示 ● ● 解析问题提示 ●
父节点输出名称 †	
v2_7_t	2_7 n_13 V2_7_beijin II
逻辑表节点 (本节点) : ld_v2_7	
分区	描述
自动解析	 系统根据配置的来源主表的类型进行自动解析: 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,即可将依赖的逻辑表节点对应的字段添加为上游 依赖。
	为您展现该维度逻辑表所依赖的物理节点和逻辑表节点。您也可以手动添加上游依赖的物理节点,适用场景说明如下: 需要将非来源主表的任务节点,添加为维度逻辑表的上游依赖节点。 节点输出名称的格式不是 项目名.表名 。
上游依赖	 手动添加上游依赖的物理节点,操作说明如下: i. 单击新增上游依赖。 ii. 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。 ⑦ 说明 节点的输出名称是全局唯一的,且不区分大小写。 iii. 单击确定新增。
逻辑表节点(本节点)	维度逻辑表节点的输出名称由系统自动生成。

3. 单击确定,完成任务的调度配置。

4. 保存、提交和发布维度逻辑表。

i. 单击维度逻辑表页面上方的 图标, 保存维度逻辑表。

ii. 单击维度逻辑表页面上方的 / 图标, 提交维度逻辑表。

- iii. 在提交备注页面,填写备注信息。
- iv. 单击确定并提交,完成维度逻辑表的提交。
- v. 如果项目空间的模式为Dev-Prod,则需要发布维度逻辑表至生产环境。如何发布,请参见管理发布 任务。

发布至生产环境的维度逻辑表任务,即可参与生产环境的调度运维。更多信息,请参见<mark>逻辑表任</mark> 务。

2.4.8. 参数配置

参数配置用于定义代码中的参数并赋值。数据研发中, 节点代码根据调度周期动态赋值(例如日期时间), 替换变量后再进行计算。

参数类型说明

调度参数包括系统参数和自定义参数:

• 系统参数可以在代码中直接使用\${bizdate}, \${Node ID}, \${Task ID}、\${业务板块名}和\${项目名}。

参数	描述
\${bizdate}	定义实例计算时对应的业务日期,默认格式为yyyymmdd,即当前日期的前一天。
\${Node ID}	任务节点ID。
\${Task ID}	任务节点生成实例ID。
\${业务板块名}	生产环境的业务板块名或开发环境的业务板块名。 例如,如果系统有Dev-Prod模式生成的生产业务板块为LD_demo,开发业务板块 为LD_demo_dev,则默认生产环境执行代码时\${LD_demo}更新为LD_demo,开发环 境执行代码时\${LD_demo}更新为LD_demo_dev
\${项目名}	生产环境的项目名或开发环境的项目名。 例如,如果系统有Dev-Prod模式生成的生产项目为test_ast,开发业务板块 为test_ast_dev,则默认生产环境执行代码时\${test_ast}更新为test_ast,开发环境 执行代码时\${test_ast}更新为test_ast_dev。

• 自定义参数需要在参数配置区域赋值, 自定义变量名及赋值。在代码中调用变量名即可。

调度参数生效说明

任务配置的参数只有在生产环境,才会参与系统的调度,根据调度周期动态赋值(例如日期时间),替换变 量后再进行计算:

- 如果是Basic开发模式,则需要提交任务节点至生产环境。
- 如果是Dev-Prod开发模式,则需要提交并发布任务节点至生产环境。

开发环境和即席查询模块支持参数识别与设置,但与调度系统规则不同。运行任务节点时,系统不会根据自 定义参数和系统参数的参数值进行计算,而是根据当前界面下,需要用户确认最新输入的参数值。例 如, \${bizdate}执行时,系统会识别并默认填入执行日期为默认值(2019年2月28日查询执行时,执行代码 的\${bizdate}替换为20190228)。

Dataphin调度系统节点配置参数的原则

- 节点代码中引用参数的方式为\${nodeParam}。
- 节点参数配置的格式为 KeyA=ValueA;KeyB=ValueB。不同参数配置之间用英文分号(;)分隔。

定义节点的Key不区分大小写,例如KeyA=ValueA和keya=ValueA定义的参数是一致的。代码中引用配置 参数时,不会区分大小写。例如\${KeyA}、\${keya}或\${keyA}都是引用自定义参数A。

● 节点配置支持参数的引用,引用格式为KeyA=\$KeyB; KeyB=ValueB,使用引用符号\$,则此时相当 于KeyA=ValueB; KeyB=ValueB。

Datapin系统中Voldemor会进行循环引用的校验,因此不支持循环定义。例如KeyA=\$KeyB; KeyB=\$KeyA或bizdate=\$bizdate。

• 支持自定义系统参数,则需要在参数配置区域赋值,例如bizdate=ABC。

如果用户配置的参数和默认参数有重复冲突,则用户配置的参数赋值会覆盖默认参数。例如,配置 了bizdate=ABC,则节点代码中引用bizdate时,该bizdate的值会被替换为ABC,而不是具体日期参数值 yyyymmdd。

Dataphin调度系统(Voldemort)时间参数的配置

- 日调度包括业务日期和预计执行时间两种调度方式。基于业务日期的参数赋值为\${yyyyMMdd};基于预计执行时间的参数赋值为\$[yyyyMMdd]。
- 小时、分钟调度支持基于小时和分钟偏移,实现小时调度、分钟调度,暂不支持秒。HH表示小时,mi(或mm)表示分钟,ss表示秒,小时的偏移:delta/24,分钟的偏移:delta/24/60。例如,在指定执行时间前一小时的表达式为\$[HH-1/24],在业务日期前1分钟为\${HHmmss-1/24/60}。

注意 如果配置的时间参数只精确到日,则小写mm系统识别为月份。当月和分钟同时存在时,为了避免歧义,月份大写为MM,分钟小写为mm,否则系统会统一识别为分钟,从而导致月份参数出错。

当前日期	基于业务日期(T- 1)的配置	实际值	基于预计执行时间 (T)的配置	实际值
2018-01-10	\${yyyymmdd}	20180109	\$[yyyymmdd]	20180110
2018-01-10	\${yyyyMMdd}	20180109	\$[yyyyMMdd]	20180110
2018-01-10	\${yyyymmdd -1}	20180108	\$[yyyymmdd -1]	20180109
2018-01-10	\${yyyymmdd -7}	20180102	\$[yyyymmdd -7]	20180103
2018-01-10	\${yyyy-mm-dd}	2018-01-09	\$[yyyy-mm-dd]	2018-01-10
2018-01-10	\${yyyy-mm}	2018-01	\$[yyyy-mm]	2018-01
2018-01-10	\${yyyy-mm -1}	2017-12	\$[yyyy-mm -1]	2017-12
2018-01-10	\${yyyy}	2018	\$[уууу]	2018
2018-01-10	\${yyyy+1}	2019	\$[yyyy + 1]	2019
2018-01-10	\${yyyy/MM/dd HH:mm:ss}	2018/01/09 00:00:00	\$[yyyy/MM/dd HH:mm:ss]	2018/01/10 00:00:00

当前日期	基于业务日期(T- 1)的配置	实际值	基于预计执行时间 (T)的配置	实际值
2018-01-10	\${yyyy/MM/dd HH:mm:ss-1/24}	2018/01/08 23:00:00	\$[HH:mm:ss- 5/24/60]	23:55:00

2.4.9. 查看历史信息

本文为您介绍如何查看维度逻辑表的历史版本信息。

前提条件

完成维度的创建, 详情参见新建维度。

操作步骤

- 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>>。
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模页面,单击左侧导航栏的图图标。

- 4. 在**维度逻辑表**页面,将鼠标悬停在需要查看历史信息的维度逻辑表后的图标后,单击编辑。
- 5. 在维度逻辑表编辑页面,单击页面上方的历史信息。
- 6. 在历史信息对话框,查看版本信息。

厅	过 定信息						×
版	本信息						
	版本号	提交人	提交时间	备注	操作	生产状态	
	2	443410238-	2019-12-27 16:21:52	1	详情	已发布生产	
	1	delugión (2001)	2019-12-27 16:19:47	1	详情	已发布生产	
					Ę	又消 确定	È

- 7. 在历史信息对话框,单击需要查看版本所在行操作列下的详情。
- 8. 在维度逻辑表某个版本的基本信息页面,查看调度信息、物理化信息和代码等信息。

2.4.10. 下线和删除维度逻辑表

本文为您介绍如何下线和删除不同状态的维度逻辑表。

前提条件

完成维度的创建,详情请参见新建维度。

背景信息

维度逻辑表的状态的说明:

- 没有提交的维度逻辑表,状态为草稿。
- 提交成功的维度逻辑表, 状态为已提交。
- 编辑已提交状态的维度逻辑表,并成功保存,状态更改为开发中。
- 已提交状态的维度逻辑表,成功下线后,状态更改为草稿。

下线和删除维度逻辑表的限制:

- 仅开发中和已提交状态的维度逻辑表,支持下线操作。
- 仅草稿状态的维度逻辑表,支持删除操作。

下线维度逻辑表

- 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>>。
- 3. 进入维度逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的√图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模页面,单击左侧导航栏的图图标。

- 4. (可选) 解锁维度逻辑表。
 - 如果维度逻辑表已被您自己锁定,则不需要解锁。
 - 如果维度逻辑表已被他人锁定,则需要执行以下操作,解锁维度逻辑表:
 - a. 在**维度逻辑表**页面,单击需要下线维度逻辑表。
 - b. 在该维度逻辑表的主表页面,单击页面右上方的 ·· 图标,解锁维度逻辑表。
- 5. 将鼠标悬停在左侧功能区域中该维度逻辑表的 图标后,单击下线。
- 6. 在提示页面,填写备注信息。
- 7. 单击确定并提交,完成维度逻辑表的下线。

删除维度逻辑表

- 1. (可选)解锁维度逻辑表。
 - 如果维度逻辑表已被您自己锁定,则不需要解锁。
 - 如果维度逻辑表已被他人锁定,则需要执行以下操作,解锁维度逻辑表:
 - a. 在**维度逻辑表**页面,单击需要下线维度逻辑表。
 - b. 在该维度逻辑表的主表页面,单击页面右上方的 ·· 图标,解锁维度逻辑表。
- 2. 将鼠标悬停在左侧功能区域中该维度逻辑表的图标后,单击删除。
- 3. 在提示页面,填写备注信息。
- 4. 单击确定并提交,完成维度逻辑表的删除。

2.5. 逻辑表-事实逻辑表

2.5.1. 创建并配置事实逻辑表

业务活动(过程)添加多个属性后生成事实逻辑表。事实逻辑表包含主键、度量和事实属性字段。主键用于 定义事实逻辑表从来源表中的取数逻辑;度量是数值类型的字段,表示某个方面的大小、多少或者程度。本 文为您介绍如何创建及配置事实逻辑表。

前提条件

在开始执行操作前,请确认您已满足以下条件:

- 已完成业务过程的创建。具体操作请参见新建业务过程。
- 已完成数据域的创建。具体操作请参见新建数据域。

操作流程

1. 步骤一: 创建事实逻辑表

定义事实逻辑表的基本信息及主键。

2. 步骤二: 添加度量

为事实逻辑表添加可以度量的字段。

3. 步骤三: 添加事实属性

为事实逻辑表添加业务实体的描述信息。

4. 步骤四: 添加关联维度

为事实逻辑表的事实属性中的外键字段关联维度。

步骤一: 创建事实逻辑表

- 1. 登录Dataphin控制台。
- 2. 按照下图操作指引,进入新建事实逻辑表配置向导页面。

■ Dataphin · 研发	开发⇒ 1 发布 运维	权限				Q、 全局接索	1	1 🗟 🖄 🦨	९ 🧕
DEV 1 v29_UI_Automatic_project / 2 DE v29_UI_Automatic v Prod Prod Dev Dev Dev Bitraisie	欢迎使用Dataphin!				(我的文件
★ 特支変組表 ○ 回 5 → ○ 回 ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○	● 规范建义 》	Hat ⊡ ⊙	公 時3 11日 11日	1 42 ⊗	■ 原子語 ○	标 0	wawa wawa ⊡ ⊙		属性重者
* « •	◎ 款票建模 >>			福 表 0					
2	North Park		800	运续调度 》 运行及维护逻辑表 时间正确有序地生	11数据处理广生的数据任务 "数据。	以确保所有任我都有投资			
			SQL	计算任务 王	VIRTUAL			同步任务 DATAX 同步	•
	□ 任务研发 ≫				SHELL			Pipeline 同步	•
			MR		PYTHON				
	66 X8	引用 »>		Had	2005EEXX	 (○) (□)			

- 3. 在新建事实逻辑表配置向导页面,完成以下操作。
 - i. 在**基本信息**页面, 配置参数。

下图中业务板块和所属项目为您上步图中②所选择的内容,且不支持修改。

新建事实逻	新建事实逻辑表								
		1 基2	本信息		2 ±1	建定义			
业务板块	v29_UI_Automatic_p	project	空间类型	应用层					
所属项目	v29_UI_Automatic_p	project	* 数据域	ui_auto_test(ui_auto 🗸				
* 业务过程	自动化专用(ui_test	t_fct) ∨							
* 明细类型	事务型	~	英文名	fct_ui_test_fct_	最多输入 64 个等	字符	_di		
* 名称	wendang_test								
描述	最多输入 128 个字							0// 20	
								0/128 //	
* 本狙士丰	Data distill chiving	n one id2							
* ALALIA	 ✓ 采用变量替换项[g_onc_nz 目空间, 替换后预	遗:\${Data_d	listill}.chiying_one					
	缺少数据源? 前往同	步或脚本加工							
						取消		下一步	
		144.772							
参数		描还							
数据域		选择业务过	は程所属的	的数据域。					
业务过程 选择需要创建事实逻辑表的业务过程。									

参数	描述
明细类型	明细类型用于区分事实逻辑表的表名。明细类型包括 事务型 和周期快照型,适用场景说明如下: 事务型 :适用于跟踪和度量业务发生过程中事件的场景。例如订单业务发生过程中的支付订单事件。 周期快照型 :适用于周期快照型针对全量的完整业务活动。例如,订单全流程。
英文名	根据明细类型不同,英文名的格式也不同。详细说明如下: 事务型事实逻辑表的英文名格式: fct_业务过程英文名_自定义名称_di 周期快照型事实逻辑表的英文名格式: fct_业务过程英文名_自定义名称_d f 其中,自定义名称的命名规则为: 包含字母、数字或下划线(_)。 不能超过64个字符。
名称	自定义名称的命名规则为: 包含汉字、字母、数字或下划线(_)或短划线(-)。 不能超过64个字符。
描述	填写对事实逻辑表的简单描述。
来源主表	选择事实逻辑表业务数据的来源主表。后续事实逻辑表的度量及事实属性字段将 从该来源主表或该来源主表关联的来源表中选择。 建议您选择生产环境的数据表(非Dev项目产生的数据),避免后续事实逻辑表 在生产环境中运维调度时,因来源主表不在生产环境而无法正常调度该事实逻辑 表,或导致事实逻辑表自动解析依赖关系失败。 如果您选择了生产数据,则后续定义主键的计算逻辑、筛选条件及通过自定义 SQL方式为事实逻辑表引入度量和事实属性字段时,需要明确生产数据的业务板 块或项目。

ii. 配置完成后, 单击下一步。

iii. 在**主键定义**页面, 配置参数。

事实逻辑表的来源数据支持从多个来源表中选取。例如业订单系统,订单的字段可能非常庞大,如 果放在一个表每次操作这个表的性能就会很差,所以您可能将这个表拆成多个表。后续您可以将多 个表的业务数据引入至事实逻辑表中,进行数据分析。主键即用于定义事实逻辑表的取数逻辑,详 细说明如下: 业务数据全部来自来源主表,则是否设定主键选择为否。筛选条件为来源主表数据同步至事实 逻辑表的条件,您可以根据业务情况进行配置。如果不配置筛选条件,则来源主表的记录数与事 实逻辑表的记录数保持一致。

新建事实逻辑表	X
→ 基本信息 2 主键定义	
*来源主表 \${Data_distill}.behavior_gameinfor_01	
是否设定主键 🔵 是 💿 否	
筛选条件	
▷ 格式化 Ø 参考示例 □ 规范性校验 ①	
1	
取消上一步	提交

■ 业务数据除了来自来源主表外,还有其他数据表,则是否设定主键选择为是,则需要定义主键 名称及主键的取数逻辑。

新建事实逻辑表	E					×
		✓ 基本信息 …	(2 主键定义		
* 来源主表	\${2_8_x_test_basic_	_0416}.t_car				
是否设定主键	● 是 ○ 否					
* 字段英文名	请定义字段英文名,	如item_id	* 字段名称	请定义字段名称	, 如商品_id	
* 字段数据类型	STRING	V				
* 字段计算逻辑						
🗋 格式化	🛛 规范性校验	⊘ 参考示例 ①				
1						
						σ×
				取消	上一步	提交

参数	描述						
字段英文名	填写主键字段的英文名称。例如item_id。						
字段名称	填写主键字段的名称。例如商品_id。						
字段数据类型	依据来源主表中映射字段的数据类型选择主键字段的数据类型。例如来源主表 中的字段order_id作为后续的映射,数据类型为STRING,则此处数据类型选 择为STRING。						
	定义来源主表中的映射字段与事实逻辑表主键的映射关系。定义语句示例如下						
字段计算逻辑	select order_id 事实主键字段。 from onedata.s_order 来源主表。 where ds='\${bizdate}' and is_pay='Y' 事实主键字段逻辑。						
	例如来源主表中的order_id这一列数据,在逻辑表重命名为item_id。						

Ⅳ. 完成配置后,单击提交。

步骤二:添加度量

度量是事实逻辑表中可以进行累加的字段。

1. 在步骤一: 创建事实逻辑表创建的事实逻辑表详情页面, 单击添加度量。

≡	D	ataphiı	1 · 研发	开	发≓	发布	运维	权限			
DEV I	v29_UI_Au v29_UI_/	tomatic_project Automat	Prod Dev	✓ fct_u	ui_test_fct	×					
规范	苞建模	数据处理	即席查询	表信息	调度配置	物理化配置	主表编辑▲	历史信息	去发布〉		
*	事实逻辑	諘	Ċ 🖻								
~							1	主表			
	🖻 fct_ui	_test_fct_01_di						■ fct_ui_te 明细类型	st_fct_test		Q ☆ :
	fct_ui	_test_fct_test_di						+ 添加	关联维度		
≈ :								+ 添加	度量		
~								+ 添加	事实属性		
_								ds		业务日期分区,	▲ 格式yy
~										≈	
R				•							

2. 在新建度量对话框, 配置度量字段。

根据事实逻辑表是否已定义主键,选择度量字段的来源表。如果事实逻辑表没有定义主键,则度量字段 仅支持从来源主表中选取;如果事实逻辑表中定义了主键,则度量字段支持从来源主表及其他数据表中 选取。配置度量字段的说明如下:

从来源表或来源主表中已有的字段中选择:下图以无主键事实逻辑表为例,按照下图操作指引,选择并配置字段。

新建度量					×
1 来源表 ● 引入字段 ○ SQL自定义字段					
来源主表 \${Data_distill}.al_01 ①	新建字段			()按引入字段	原始英文名排列
选择新增字段(注:无主键事实逻辑表不支持从不同物理表选择字段)	字段英文名	字段名称	数据类型	字段类型	
Q、请输入搜索关键字 2	3 来源主表: Data_distill.al	_01 4			
已引入的字段 1	> ds		STRING V	度量 🗸	品 区 亩
ds	新增 原始英文名: ds	原始名称:			
□ 未引入的字段 6 ✓					
□ id id					
films_name films_name					
box_office box_office					
price price	全选))			5
				取消	保存并校验

选择度量字段并添加至新建字段区域后,您可以根据业务情况编辑字段的名称及类型。 • SQL自定义字段:按照下图操作指引,为来源表新建字段。

新建度量	×
未源表 ○ 引入字段 ● SQL目定义字段	
□ 格式化	查看公用计算逻辑 ∨
1	
	取消

您可以根据业务需求通过SQL自定义字段作为事实逻辑表的度量字段:

■ 无主键事实逻辑表SQL自定义字段示例。

```
select order_type --度量字段
from onedata.s_order --来源主表
where ds='${bizdate}' and is pay='Y' --度量字段
```

■ 有主键事实逻辑表SQL自定义字段示例。

```
select id as k --as k标识主键关联字段
    ,pay_amt --度量字段
from onedata.s_order_ext --字段逻辑
where type=0 and pt='${bizdate}'
```

步骤三:添加事实属性

事实属性是事实逻辑表中描述业务实体的字段,不具有累加性。

1. 在步骤一: 创建事实逻辑表创建的事实逻辑表详情页面, 单击添加事实属性。

≡	D	ataphiı	1 · 研发	开	友≓	发布	运维	权限			
DEV I	v29_UI_Au v29_UI_/	tomatic_project Automat…⇔∨	Prod	🧟 fct_u	ii_test_fct	×					
±0,3	 * 79 ##	粉层水理	即度查次	表信息	调度配置	物理化配置	主表编辑▲	历史信息	去发布>		
75/651	EV#194	数组内企理	即作旦间								
*	事实逻辑	諘	Ċ₽								
~							1	主表			
	🖻 fct_ui	_test_fct_01_di						● fct_ui_te 明细类型	st_fct_test		Q ≈ :
	🖻 fct_ui	_test_fct_test_di						+ 沃加	羊群维度		
								1000	1/14/14432		
~								+ 添加	度量		
								+ 添加	事实属性		
								ds		业务日期分区,	格式yy
~											
R				•							
-											

2. 在新建事实属性对话框, 配置事实属性字段。

根据事实逻辑表是否已定义主键,选择事实属性字段的来源表。如果事实逻辑表没有定义主键,则事实 属性字段仅支持从来源主表中选取;如果事实逻辑表中定义了主键,则事实属性字段支持从来源主表及 其他数据表中选取。配置事实属性字段的说明如下:

从来源表或来源主表中已有的字段中选择:下图以有主键事实逻辑表为例,按照下图操作指引,选择并配置字段。

新建事实属性							Х
未源表 ③ 引入字段 ○ SQL自定义字段 2							
来源表 2_8_x_test_basic_0416.t_car V ①		新建字段				⑥按引入字目	设原始英文名排列
✓ 系统默认会将项目空间普换为变量形式使用普换后预 宽 % 2_8_x_test_basic_0416).t_car		字段英文名	字段名称	数据类型		字段类型	掘作
关联逻辑字段 benz V ①	1	来源物理表: 2_8_x_test_basic_0)416.t_car / 关联逻辑字段: benz				5
选择新增字段(注:可以从不同物理表选择字段,同一字段可以添加多次) Q 速给入搜索关键字	1	benz 原始英文名: benz	奔驰 原始名称: 奔驰	STRING		事实属性 ∨	品 区 前
v injent/datat/set 4 Ind Bigg	>	Dmw 原始英文名: bmw	宝马 原始名称:宝马	STRING		事实属性 🗸	8 C İİ
volkswagen 大众	新瑁 (1)	ds 原始英文名: ds	原始名称:	STRING	\sim	事实属性 ∨	品区 前
□ mazda □ 马自达							
□ honda 本田							
de cada		全选					6
						取消	保存并校验

选择事实属性字段并添加至新建字段区域后,您可以根据业务情况编辑字段的名称及类型。 • SQL自定义字段:按照下图操作指引,为来源表新建字段。

新建事实属性	×
朱源表 ○ 引入字段 ● SQL目定义字段	
□ 格式化	查看公用计算逻辑 ∨
1	
	0 X
	取消保存并校验

您可以根据业务需求通过SQL自定义字段作为事实逻辑表的事实属性字段:

■ 无主键事实逻辑表SQL自定义字段示例。

```
select order_type --事实属性字段
from onedata.s_order --来源主表
where ds='${bizdate}' and is pay='Y' --事实属性字段
```

■ 有主键事实逻辑表SQL自定义字段示例。

```
select id as k ---"as k"标识主键关联字段
,pay_amt --事实属性字段
from onedata.s_order_ext --字段逻辑
where type=0 and pt='${bizdate}'
```

步骤四:添加关联维度

事实逻辑表的事实属性中的外键字段关联维度。

1. 在步骤一: 创建事实逻辑表创建的事实逻辑表详情页面, 单击添加关联维度。

≡	Da	taphir	1・研发	开	发≠	发布	运维	权限			
DEV I	v29_UI_Autor v29_UI_Au	natic_project tomat…⇔∨	Prod	🖂 fct_	ui_test_fct	×					
物结		数据外理	印度音询	表信息	调度配置	物理化配置	主表编辑▲	历史信息	去发布〉		
Max	bietie 古会)要は言言										
*		x Sarà:	0 🖻								
~			_					±æ	est_fct_test		0 & :
▲	fct_ui_te	st_fct_01_di						明细类	型		~~ .
۲	⊌ tct_ui_te	ist_tct_test_di						+ 添加	口关联维度		
								+ 添加	渡量		
~								+ 添加	叩事实属性		
								ds		业务日期分区,	▲ 格式yy
~										*	
R				•							

2. 在新建关联维度对话框,配置参数。

新建关联维度					х
dim_uitest 💿	关联维度				
	 * 关联维度 ui_auto_test / c 维度类型 普通维度 	lim_uitest uitest 主键及主键计算逻辑 id 🔒			
	编辑关联逻辑) *	联关系		
	* 关联逻辑 必 這项	鉄字段 🗸	= 关联维度主键字段	dim_uitest.id	
	 * 缺联策略 保留原始缺联数 ////////////////////////////////////	- H		∨ ()	
	編編雑度用色 ① *角色英文名 dim_ uitest				
	*角色名称 uitest				
	* 默认值 -110	🕞 恢复到推荐值			
					取消 确定
区域	参数	描述			

区域	参数	描述
关联维度	关联维度	选择您已创建的维度。 选择了维度后,页面为您展示 维度类型 ,同时您可以按照下图操 作指引,查看维度逻辑表的计算逻辑。
编辑关联逻辑	关联逻辑	选择事实逻辑表中的字段,并配置事实逻辑表字段与需要关联维度 逻辑表中的主键的关系。
	缺联策略	缺联策略用于定义来源主表(左表)中存在的字段,但在维度逻辑 表(右表)中不存在的字段,即无法关联的字段的计算逻辑。您可 以选择 保留原始缺联数据和缺联数据使用默认值代替 : • 保留原始缺联数据: 创建派生指标时,保留左表原始数据。 • 缺联数据使用默认值代替:维表中的字段没有与事实逻辑表关 联上,则使用默认值-110。
编辑维度角色	角色英文名、角 色名称	 维度角色具有维度别名作用。多次引用同一维度时不可重名定义, 起到角色扮演效果。您需要定义角色英文名和角色名称: 角色英文名的前缀默认为 dim , 自定义部分的命名规则 为: 包含字母、数字或下划线(_)。 不能超过64个字符。 角色名称的命名规则: 包含汉字、数字、字母、下划线(_)或短划线(-)。 不能超过64个字符。
默认值设定	默认值	如果事实逻辑表的外键字段和维度表关联不上,则系统将 默认值 设定为-110。

您也可以单击对话框左侧的**添加关联维度**,添加多个关联维度。

3. 配置完成后,单击**确定**。

后续步骤

完成事实逻辑表的创建及配置后,您需要为事实逻辑表配置调度参数、生命周期及分区字段等参数。具体操 作请参见<mark>调度配置和物理化配置</mark>。

2.5.2. 物理化配置

物理化配置用于配置事实逻辑表的生命周期、分区字段及设置自定义参数。本文为您介绍如何物理化配置事 实逻辑表。

前提条件

完成事实逻辑表的创建,详情请参见创建并配置事实逻辑表。

操作步骤

- 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>>。
- 3. 进入事实逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击 ፼图标。
- 在事实逻辑表页面的左侧事实逻辑表列表,将鼠标悬停在需要物理化配置的事实逻辑表后的图图标,单击编辑。
- 5. (可选)解锁事实逻辑表。

- 如果事实逻辑表已被您自己锁定,则不需要解锁。
- 如果事实逻辑表已被他人锁定,则需要单击页面右上方的△图标,解锁事实逻辑表。
- 6. 在该事实逻辑表的编辑页面,单击页面上方的物理化配置。
- 7. 根据物理化配置对话框提示, 配置参数。

初理化配直		
物理化配置		
* 存储生命周期	7 快速选择 7 14 30 365 天	
* 选择分区字段	11 ds	
无需物理化字段		
自定义参数设置		
自定义参数: ①	示例: set oneservice.optimize.skewjoin = true/false; set hive.exec.dynamic.partition.mode=nonstrict; (动态分区开启/关闭) set hive.vectorized.execution.enabled = false; (矢量查询开启/关闭) set hive.vectorized.execution.reduce.enabled = false;	0/2048 //
	取消	确定

参数	描述
存储生命周期	选择存储生命周期。系统支持快捷选择的存储生命周期包括7、14、30和365天。
选择分区字段	选择表的分区。
自定义参数	配置自定义参数,详情请参见 <mark>配置自定义任务参数</mark> 。

8. 单击确定,完成事实逻辑表的物理化配置。

- 9. 保存、提交和发布事实逻辑表。
 - i. 单击编辑事实逻辑表页面上方的 图标, 保存物理化配置。
 - ii. 单击编辑事实逻辑表页面上方的 图标, 提交物理化配置。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成事实逻辑表的提交。
 - v. (可选)发布事实逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布事实逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布事实逻辑表。

2.5.3. 修改事实逻辑表

事实逻辑表用于描述业务过程的详细信息。本文为您介绍如何修改事实逻辑表。

前提条件

完成事实逻辑表的创建,详情情参见创建并配置事实逻辑表。

操作步骤

- 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>>。
- 3. 进入事实逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击≥图标。
- 4. 在事实逻辑表页面,将鼠标悬停在需要修改事实逻辑表后的图图标后,单击编辑。
- 5. (可选) 解锁事实逻辑表。
 - 如果事实逻辑表已被您自己锁定,则不需要解锁。
 - 如果事实逻辑表已被他人锁定,则需要单击页面右上方的∩图标,解锁事实逻辑表。
- 在事实逻辑表编辑页面,您可以修改主表、物理化配置和调度配置,详情请参见创建并配置事实逻辑 表、调度配置和物理化配置。
- 7. 保存、提交和发布事实逻辑表。
 - i. 单击编辑事实逻辑表页面上方的 图标, 保存修改。
 - ii. 单击编辑事实逻辑表页面上方的 图标, 提交事实逻辑表。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交,完成事实逻辑表的提交。
 - v. (可选)发布事实逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布事实逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布事实逻辑表。

2.5.4. 调度配置

调度配置是配置节点在调度过程中的调度规则及依赖关系。本文为您介绍如何配置事实逻辑表的调度参数。

前提条件

完成事实逻辑表的创建,详情请参见创建并配置事实逻辑表。

背景信息

调度配置的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的规范建模。
 - 单击快速进入研发项目区域的Dev或Basic(非Data_distill)项目,选择开发数据的项目空间。

⑦ 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏中的研发。
- 4. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的项目空间 (非Data_distill项目)。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不需要 选择项目空间。

5. 进入调度配置页面。

i. 在数据开发页面, 单击规范建模。

ii. 在规范建模的左侧导航栏,单击**□事实逻辑表**图标。

iii. 在事实逻辑表页面的左侧功能区,单击需要配置调度参数的事实逻辑表。

- iv. 解锁事实逻辑表。
 - 如果事实逻辑表已被您自己锁定,则不需要解锁。
 - 如果事实逻辑表已被他人锁定,则需要单击页面右上角的 @ 图标,解锁事实逻辑表。

v. 在该事实逻辑表的页面,单击页面上方的调度配置。

6. 在调度配置页面,配置调度参数。

i. 配置基本信息区域参数。

系统自动生成的逻辑表名称、任务类型和描述,不支持修改。

基本信息		
逻辑表名称	Cons. (1997), (1	
任务类型	MAX_COMPUTE_SQL 回 查看生产任务	
描述	5.899 (company) (CONTRACTOR)	
优先级	高优先级 🗸 🗸	
参数配置	bizdatetime=\$[yyyyMMdd HHmmss - 1/24];today=\$[yyyyMMdd - 1/24];hh=\$[HH - 1/24]	
	方点参数配置说明 	
参数	描述	

参数	描述
优先级	您可以选择节点调度任务执行的优先级,优先级包括: 最低优先级。 低优先级。 中等优先级。 高优先级。 最高优先级。
参数配置	默认为事实逻辑表的数据时效表达式。同时您还可以对代码中所用参数的具体赋 值。节点参数配置是对代码中所用参数的具体赋值。您可以单击 节点参数配置 说明 ,查看Dataphin调度系统节点配置的规则及支持配置的时间参数。

ii. 配置调度周期区域参数。调度配置为逻辑表的数据时效,且不支持修改。

调度周期								
调度配置	小时	\sim	💿 时间段	🔵 时间点				
开始 00:00			间隔 1	\sim	小时	结束	23:59	
cron 表达式 000-23/1**?								
依赖上周期	依赖上周期 💿 不依赖 🗌 依赖							

■ 如果**依赖上周期**选择为**不依赖**,则事实逻辑表参与调度时不依赖上周期。

- 如果**依赖上周期**选择为**依赖**,则执行以下操作,添加依赖节点。
 - a. 单击+图标。
 - b. 根据页面提示, 配置参数。

 ・ 调度周期 ○ 不依赖 ● 依赖 						
	请输入节点ID	へ 请选择依赖字段	2	^		
	请输入节点id或输出名	请输入字段关键字				
		+	3			

编号	描述			
0	选择节点类型。系统支持选择 自定义 和 本逻辑表内节点 。			
2	 如果节点类型选择为自定义,则执行以下操作,选择依赖字段: a. 单击节点D显示框区域。 b. 在节点D列表中,选择节点D。 您也可以在下拉列表的输入框中,输入节点D的关键字,快速搜索 节点。 c. 单击依赖字段显示框区域,在下拉列表中,选中依赖字段。 您可以在下拉列表中的输入框中,输入字段名称关键字,快速查找 字段。如果所有字段需要依赖,则选中全选。 d. 单击确定。 如果节点类型选择为本逻辑表内节点,则执行以下操作,选择依赖字段: a. 单击依赖字段显示框区域,在下拉列表中,选中依赖字段。 您可以在下拉列表中的输入框中,输入字段名称关键字,快速查找 字段。如果所有字段需要依赖,则选中全选。 b. 单击确定。 			
3	您可以单击 <mark>+</mark> 图标,添加多个依赖节点。			

iii. 配置调度依赖区域参数。

上游依赖定义了业务流程之间的调度依赖关系。正确的调度依赖关系可以保障业务数据有效的产出。正确的依赖关系配置后,无需关心当前节点依赖的数据表是否正常产出数据,是否影响该节点的运行,系统会自动为您解析任务节点间的关联关系。

- 依赖关系是两个或多个节点之间的调度依赖关系,控制节点运行顺序。同时上游节点的状态将影响其他下游节点的运行状态。
- 配置了依赖关系的节点调度规则为:首先,上游节点运行完成后,才能调度下游节点。其次,根据节点设定的调度时间判断是否执行调度。

 您在设定的调度时间之前提交的调度配置,会在设定的调度时间之后生效。而在设定的调度时间 之后配置的依赖关系,只能间隔一天再生成实例。

调度依赖默认开启自动解析,系统自动解析该事实逻辑表的物化代码中依赖关系。同时,您也可以 手动添加上游依赖的物理节点和逻辑表节点。

Ⅰ 调度依赖 自动解析 ● 是①	戶 解析输入輸出			
上游依赖 物理节点	逻辑表节点 ①		② 解析问题提示	○ + 新增上游依赖
父节点输出名称	节点名	节点ID (实例ID)	关联源表	负责人 口 一 操作
v2_7_t	v2_7	n_13	v2_7_beijin 🎞	Ť.
逻辑表节点 (本节点) : ld_	_v2_7_	● ● ● 运维	工作台 >	

分区	描述
自动解析	系统会解析代码中的物理节点和逻辑表节点,并查找到与该表名相同的输出名称。输出名称所在的节点作为当前节点的上游依赖。 单击 解析输入输出 ,系统自动解析上游依赖。如果您重复单击 解析输入输出 , 则刷新 父节点输出名称 。
上游依赖	为您展现该逻辑表内的物理节点和逻辑表节点所依赖的父节点。您可以通过以下 两种方式新增上游依赖:
逻辑表节点(本节 点)	事实逻辑表节点的输出名称由系统自动生成。

7. 单击**确定**,完成事实逻辑表的调度配置。

8. 保存、提交和发布事实逻辑表。

i. 单击编辑事实逻辑表页面上方的 图标, 保存调度参数的配置。

ii. 单击编辑事实逻辑表页面上方的 图标, 提交实时逻辑表。

- iii. 在提交备注页面,填写备注信息。
- iv. 单击确定并提交,完成事实逻辑表的提交。
- v. (可选)发布事实逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布事实逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的事实逻辑表,即可进入生产环境。

2.5.5. 查看历史信息

本文为您介绍如何查看事实逻辑表的历史版本信息。

前提条件

完成事实逻辑表的创建,详情参见创建并配置事实逻辑表。

操作步骤

- 1. 登录Dataphin控制台。
- 2. 在Dataphin控制台页面,选择工作区Region后,单击进入Dataphin>>。
- 3. 进入事实逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击≥图标。
- 4. 在事实逻辑表页面,将鼠标悬停在需要查看历史信息的事实逻辑表后的■图标后,单击编辑。
- 5. 在事实逻辑表编辑页面,单击页面上方的历史信息。
- 6. 在历史信息对话框,查看版本信息。

历史信息					×
版本信息					
版本号	提交人	提交时间	备注	操作	生产状态
2	ampro258 -	2019-12-27 16:21:52	1	详情	已发布生产
1	645phin7282	2019-12-27 16:19:47	1	详情	已发布生产
					取消 确定
					-96 H3 96 AL

- 7. 在历史信息对话框,单击需要查看版本所在行操作列下的详情。
- 8. 在事实逻辑表某个版本的基本信息页面,查看调度信息、物理化信息和代码等信息。

基本信息	版本:2 提交人: 🚛	ophin Roddbillud	提交时间: 2019-12-27 16:21:52
▶ 2013年1月1日 - 2013年2月1日 - 20			查看定义详情 🛛
代码查看	C 整段copy	 调度信息 	
1 字段		准保主文分,	
		A2-494-6X-521177-1	
3 OS SIRUMG 亚穷口助疗区,恰近yyyymmdd		描述:	and the second second second second
5		优先级:	中等优先级
6 jie STRING jie		来教配票。	
7 select r_town from te		P POLICE	
° r name v1 string ∻		依赖上周期:	N
10 select id as k, r_nam		上游依赖:	
		当前节点输出:	
12 r_nign_vi buoste 5			
14		 物埋化信息 	
15 r_age_v1 bigint 年(存储生命周期:	36000
16 select id as k, r_age		公区字段。	eh
is rwork_v2 string I		10 L 0 1001	
19 select id as k, r_wor		无需物理化:	未配置
20			
21 r_nouse_vz bigint 22 select id as k, r hou			
23			
24 r_car_v2 string 年,			
25 select id as k, r_car			

2.5.6. 下线和删除事实逻辑表

本文为您介绍如何下线和删除不同状态的事实逻辑表。

前提条件

完成事实逻辑表的创建,详情请参见创建并配置事实逻辑表。

背景信息

事实逻辑表的状态的说明:

- 保存成功的事实逻辑表,状态为草稿。
- 提交成功的事实逻辑表,状态为**已提交**。
- 编辑已提交状态的事实逻辑表,并成功保存,状态更改为开发中。
- 已提交状态的事实逻辑表,成功下线后,状态更改为草稿。

下线和删除事实逻辑表的限制:

- 仅开发中和已提交状态的事实逻辑表,支持下线操作。
- 仅草稿状态的事实逻辑表, 支持删除操作。

下线事实逻辑表

- 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>>。
- 3. 进入事实逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模的左侧导航栏,单击≥图标。

- 4. (可选) 解锁事实逻辑表。
 - 如果事实逻辑表已被您自己锁定,则无需解锁。
 - 如果事实逻辑表已被他人锁定,则执行如下操作,解锁事实逻辑表:
 - a. 在事实逻辑表页面,单击下线的事实逻辑表。
 - b. 在该事实逻辑表的主表页面,单击页面右上角的 ··· 图标。
- 5. 将鼠标悬停在该事实逻辑表的 图标后,单击下线。
- 6. 在提示页面,填写备注信息。
- 7. 单击确定并提交,完成事实逻辑表的下线。

删除事实逻辑表

- 1. (可选) 解锁事实逻辑表。
 - 如果事实逻辑表已被您自己锁定,则无需解锁。
 - 如果事实逻辑表已被他人锁定,则执行如下操作,解锁事实逻辑表:
 - a. 在事实逻辑表页面, 单击需要删除的事实逻辑表。
 - b. 在该事实逻辑表的主表页面,单击页面右上角的 ·· 图标。
- 2. 将鼠标悬停在该事实逻辑表的 图标后,单击删除。
- 3. 在提示页面,填写备注信息。
- 4. 单击确定并提交,完成事实逻辑表的删除。

2.6. 规范定义-原子指标

2.6.1. 新建原子指标

原子指标是对指标统计口径、具体算法的抽象。Dataphin创新性地提出了设计即开发的理念,指标定义同时 也明确了设计统计口径(即计算逻辑),提升了研发效率,并保证了统计结果的一致性。

背景信息

Dataphin根据计算逻辑的差异性,支持创建原子指标和衍生原子指标。详细说明如下:

- 原子指标是原生原子指标。例如,支付金额。
- 衍生原子指标是基于已提交原子指标,设定计算逻辑而构成的指标。例如,一个已提交的原子指标为支付 金额,另一个已提交的原子指标为支付人数,您可以通过新建衍生原子指标的方式统计人均支付金额。

新建原子指标的权限限制,详情请参见数据开发权限列表。

新建原子指标

- 1. 登录Dataphin控制台。
- 2.
- 3. 进入原子指标页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,按照下图操作指引进入新建原子指标页面。



4. 在新建原子指标页面,选择数据域和来源表。

选择来源信息				
业务板块 未遵表 fct	新国项目 te 项目空间关型 应用量	政强制线 test_dataphin ∨		
原子指标列表 全局 模型相关 全局原子指标列表(2)		<u>し</u> + 新建原子指标	来源表模型	
原子指标	来源洋情	状态	最近更新人	操作
product_type 类型	1,01,0,000,000	⊖ 華稿	100000 (0000)	🗹 🔺 🖂 🗄
order_amt order_amt	1.00.00.000.00	已提交	anger (aging	K & % :
参数	描述			

参数	描述
	选择数据域。
数据域	⑦ 说明 划分数据域,详情请参见划分数据域。
来源表	选择来源表。为保证建模标准规范,来源表仅支持维度逻辑表或事实逻辑表。

5. 在新建原子指标页面,单击新建原子指标。

选择来源信息			
业务板块 所服项目 来源表 tct	te 资用空间关型 应用量 数据域	test_dataphin \vee	
原子指标列表 全局 模型相关 全局原子指标列表(2)		し - 新建原子指示 - 新建原子指示	模型
原子描标	来源详情	状态 最近更新人	操作
product_type 类型	10,000,00,0000000000	€ 童稿	☑ ▲ ≅ :
order_amt order_amt	1.18.0.000	☑ 已提交	☑ ▲ ☆ :

6. 在新建原子指标对话框, 配置参数。

新建原子指标								х
* 主要来源字段	请选择主来源字段				~ 0			
* 英文名	请输入英文名			* 名称	请输入名称			
描述								0/128
数据类型	BIGINT V							
* 计算逻辑 (9						是否累加:	● 是 ○ 否
🖻 格式化	化 🖉 参考示例	🛛 规范性校验	(i)					
1								Ū X
						取消	保存	提交
参数		描述						
主要来源字	段	基于所选来	源表,说	赴择事实	表或维度表	中的字段。		

参数	描述
	基于所选主要来源字段,英文名自动填充为字段名,您也可以修改英文名。 英文 名 由字母、数字或下划线(_)组合组成。
英文名	⑦ 说明 英文名 不能超过64个字符。
	名称由汉字、字母、数字、下划线(_)或短划线(-)组合组成。
名称	⑦ 说明 名称不能超过64个字符。
描述	填写原子指标的简单描述。
数据类型	数据类型包 括STRING、BIGINT、DOUBLE、TIMESTAMP、DATETIME和DECIMAL。
是否累加	基于不同大小统计周期的数据,选择是否需要累加计算。例如,最近7天支付金额, 是否需要基于最近1天支付金额进行累加计算。
计算逻辑	基于来源逻辑表模型,定义所需指标的计算逻辑: i. 编写SQL代码。 例如count(distinct order_id),其中order_id为来源逻辑表模型中的字段。 您也可以单击 主键计算逻辑 后的参考示例,可以查看SQL编写示例。 ii. 单击 主键计算逻辑 后面的 规范性校验 ,可以校验您编写的SQL是否符合规范。 如果 规范性校验 校验失败,您可以单击 格式化 ,系统会自动为您调整SQL语句的格式。

- 7. 提交和发布原子指标。
 - i. 单击提交。
 - ii. 在提交备注对话框,填写备注信息。
 - iii. 单击确定并提交。
 - ⅳ. (可选)发布原子指标。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的原子指标,详情请参见管理发布任务。
 - 如果您的开发模式是Basic,则提交成功的原子指标,即可进入生产环境。

新建衍生原子指标

1. 在原子指标页面,单击原子指标后的图图标,选择新建衍生原子指标。

您也可以通过以下方式,进入新建衍生原子指标:

- 单击项目名称后的 图标,选择规范定义 > 衍生原子指标。
- 在页面右侧的数据开发总览区域,单击原子指标下的 图标后,选择新建衍生原子指标。
- 在原子指标页面左下方, 单击原子指标对象列表。在对象列表页的原子指标页签, 单击新建原子

指标后,选择衍生原子指标。

2. 在新建衍生原子指标对话框, 配置参数。

新建衍生原子指标	×
业务板块 * 英文名 请输入英文名	所属项目 项目空间类型 应用层 * 名称 请输入名称
描述	0/128
数据类型 BIGINT V	
● 格式化 ◎ 参考示例	回 规范性校验 ① 引用原子指标
1	
	取消 保存 提交
参数	描述
	英文名 由字母、数字、下划线(_)组合组成。
英文名	⑦ 说明 英文名不能超过64个字符。
	名称 名称由汉字、字母、数字、下划线(_)或短划线(-)组合组成。
名称	⑦ 说明 名称不能超过64个字符。
描述	对衍生指标的简单描述。
数据类型	数据类型包 括STRING、BIGINT、DOUBLE、TIMESTAMP、DATETIME和DECIMAL。
是否累加	基于不同大小统计周期的数据,选择是否需要累加计算。例如,最近7天支付金额, 是否需要基于最近1天支付金额进行累加计算。

参数	描述
计算逻辑	 基于已提交原子指标,组合设定所需指标的统计计算方式: • 在SQL语句输入框中输入衍生指标的计算公式,其中计算公式中引用的原子指标 均为已提交的原子指标。 • a.单击引用原子指标后的 ▼图标后,单击数据域及来源表显示框,在下拉列 表中选择数据域及来源表。 b.鼠标悬停至筛选出来的原子指标后,单击添加,将原子指标添加在右侧的 SQL语句输入框中。 或在 图标后输入原子指标关键字进行快速搜索,单击搜索出来的原子指标后 的添加,将原子指标添加在右侧的SQL语句输入框中。 例如,计算人均下单金额 crt_amt/usr_cnt ,其中crt_amt为下单金额对应的 已提交原子指标,usr_cnt为用户数对应的已提交原子指标。

- 3. 提交衍生原子指标。
 - i. 单击提交。
 - ii. 在提交备注对话框,填写备注信息。
 - ⅲ. 单击确定并提交。
- 4. 发布衍生原子指标。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的衍生原子指标,详情请参见管理发布任务。
 - 如果您的开发模式是Basic,则提交成功的衍生原子指标,即可进入生产环境。

2.6.2. 修改原子指标

原子指标是对指标统计口径和具体算法的抽象,同时定义了统计口径(计算逻辑)。本文为您介绍如何修改 原子指标。

前提条件

完成原子指标的创建,详情请参见新建原子指标。

背景信息

- 原子指标的统计口径变更后,会影响依赖该原子指标生成的派生指标。因此在修改原子指标前,需要您充分考虑对派生指标的影响。
- 修改原子指标权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 2.
- 3.
- 4. 您可以通过以下任何一种方式,进入编辑原子指标页面。
 - 选择需要修改的原子指标,将鼠标悬浮在 图标,单击编辑。



• 单击需要修改的原子指标后,再单击该原子指标属性框中的编辑。

属性查看 > 原子指标属性						
ds 草稿 dd						
□编辑 ◎) 查看同源					
业 冬 标 中 ・	A 100 M 100 M 100					
所属项目:	A100, 20					
项目空间类型:	应用层					
数据域:	test_dataphin					
英文名:	ds					
名称:	dd					
计算逻辑:	a ^g a					
描述:						
数据类型:	bigint					
数据叠加:	是					
统计周期标识:	字段: province, 格式类型: yyyy-MM-dd					

○ a. 单击页面左下角的原子指标对象列表,进入对象列表页的原子指标页面。

b. 单击需要修改原子指标所在行的操作列下的 図图标。

- 5. 根据编辑原子指标页面提示,修改参数,详情请参见新建原子指标。
- 6. 完成参数修改后,单击提交。
- 7. (可选)发布原子指标。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的原子指标,详情请参见管理发布任务。

• 如果您的开发模式是Basic,则不需要发布已提交的原子指标(提交后的原子指标就可以参与调度)。

2.6.3. 查看同源原子指标

Dataphin支持基于一个来源逻辑表生成多个原子指标。本文为您介绍如何查看同源原子指标。

前提条件

完成原子指标的创建, 详情请参见新建原子指标。

背景信息

您可以通过查看同源原子指标的方式,查看与某个原子指标同源的其他原子指标。

⑦ 说明 衍生原子指标不存在同源原子指标。

操作步骤

- 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>>。
- 3. 进入原子指标页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的√图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在左侧导航栏中, 单击, 风图标。
- 4. 在原子指标页面,单击需要查看同源的原子指标后,再单击该原子指标属性框中的查看同源。
- 5. 在同源原子指标页面,查看与该原子指标同源的其他原子指标。

▲ 同源原子指标 ×				: 8
选择来源信息				
业务板块 Vi 未週表 lct_buss_lab2_df	新雕项目	项目空间类型 应用层	数据域 公共地 🗸	
原子指标列表 全局 模型相关 全局原子指标列表(4)		し - 新建原子指标	来源表模型 X	
原子指标	来源洋情	状态	最近更新人	攝作
age23 ceshiceg	1,00,00,00,00,000	⊖ 草稿	NAME AND A DESCRIPTION	C 🛦 🖄 :
age ceshi	Conductor Strengther	⊖ 草稿	NAME AND ADDRESS	ZA 🕾 :

2.6.4. 克隆原子指标

Dataphin支持基于一个原子指标克隆同源原子指标。本文为您介绍如何克隆原子指标。

前提条件

完成原子指标的创建, 详情请参见新建原子指标。

操作步骤

- 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>>。
- 3. 进入原子指标页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的√图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在左侧导航栏中, 单击<u>风</u>图标。

- 在原子指标页面,单击需要克隆的原子指标后,将鼠标悬停在该原子指标属性框中的图标,后,单击克
 隆。
- 5. 根据**克隆原子指标**对话框提示,配置参数,详情请参见新建原子指标。

克隆原子指标											×
* 主要来源字段	fct_b	10,000			~	0					
* 英文名	age23			* 名称	ceshiceg						
描述											0/128
数据类型	BIGINT V	统计周期标识	字段	ds		~	格式类型	yyyy-MM-dd	~	0	
* 计算逻辑(0								是否累加: 🤇	是	() 否
🔓 格式化	化 🖉 参考示例	1 同 规范性校验									
1 fc	t_bus										N N
						[取消	4	呆存	提	交

- 6. 单击**提交**,完成参数的配置。
- 7. 在提交备注页面,填写备注信息。
- 8. 单击确定并提交,提交原子指标。
- 9. (可选)发布原子指标。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的原子指标,详情请参见管理发布任务。
 - 如果您的开发模式是Basic,则不需要发布已提交的原子指标(提交后的原子指标就可以参与调度)。

2.6.5. 查看相关派生指标

本文为您介绍查看原子指标的相关派生指标。

前提条件

完成原子指标的创建和提交,详情请参见新建原子指标。

背景信息

- 因为系统仅支持已提交状态的原子指标构建派生指标,所以仅限已提交状态的原子指标支持查看相关派 生指标。
- 查看相关派生指标的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的规范建模。
 - 单击快速进入研发项目区域的Dev或Basic(非Data_distill)项目,选择开发数据的项目空间。

? 说明

- 如果您通过**快速开始相关工作**进入数据开发模块,则跳过步骤3。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏中的研发。
- 4. 在数据开发页面,单击项目名称后的 ☑图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目空间为您的数据开发空间,则不需要 选择项目空间。

- 右数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模。
- 6. 在规范建模页签,单击左侧导航栏的八原子指标图标。
- 在原子指标页面,单击需要查看相关派生指标的原子指标。在原子指标属性框,鼠标悬停至 图标
 后,单击相关派生指标。

您也可以通过以下方式,查看相关派生指标:

- i. 在原子指标页面, 单击页面下方的原子指标对象列表。
- ii. 鼠标悬停至需要查看相关派生指标的原子指标 图标后,单击相关派生指标。
- 8. 在对象列表页页面,查看该原子指标的相关派生指标。

2.6.6. 创建派生指标

Dat aphin支持基于某个原子指标创建派生指标。本文为您介绍如何基于原子指标创建派生指标。

前提条件

完成原子指标的创建,详情请参见新建原子指标。

操作步骤

- 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>>。
- 3. 进入原子指标页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在左侧导航栏中, 单击<u>风</u>图标。

- 在原子指标页面,单击需要创建派生指标的原子指标后,将鼠标悬停在该原子指标属性框中的 图标
 后,单击创建派生指标。
- 5. 根据**新建派生指标**页面提示,配置参数,详情请参见新建派生指标。
- 6. 完成参数配置后,提交派生指标。
 - i. 完成参数配置后, 单击**预生成派生指标**。
 - ii. 完成参数配置后,单击**提交**。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 完成备注信息填写后, 单击确定并提交。
- 7. (可选)发布原子指标。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的原子指标,详情请参见管理发布任务。
 - 如果您的开发模式是Basic,则不需要发布已提交的原子指标(提交后的原子指标就可以参与调度)。

2.6.7. 下线和删除原子指标

本文为您介绍如何下线、下线并删除和删除不同状态的原子指标。

前提条件

完成原子指标的创建,详情请参见新建原子指标。

背景信息

- 原子指标的状态的说明:
 - 保存成功的原子指标,状态为**草稿**。
 - 提交成功的原子指标,状态为已提交。
 - 编辑已提交状态的原子指标,并成功保存,状态更改为开发中。
 - **已提交**状态的原子指标,成功下线后,状态更改为草稿。
- 下线原子指标的限制:

- 仅开发中和已提交状态的原子指标, 支持下线操作。
- 仅支持超级管理员、项目管理员和开发者角色的用户下线原子指标。
- 当且仅当原子指标无依赖引用时(派生指标引用及其他原子指标引用),才支持下线操作。
- 删除原子指标的限制:
 - Q草稿状态的原子指标,支持删除操作。
 - 仅支持超级管理员、项目管理员和开发者角色的用户删除原子指标。
- 下线并删除原子指标的限制:
 - 仅开发中和已提交状态的原子指标,支持下线并删除操作。
 - 仅支持超级管理员、项目管理员和开发者角色的用户下线并删除原子指标。
 - 当且仅当原子指标无依赖引用时(派生指标引用及其他原子指标引用),才支持下线并删除操作。

删除原子指标

- 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>>。
- 3. 进入原子指标页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在左侧导航栏中, 单击<u>风</u>图标。

- 4. 您可以通过以下任何一种方式,进入删除原子指标的提示对话框。
 - 选择需要删除的原子指标,将鼠标悬停在__图标,单击删除。
 - 单击需要下线的原子指标后,将鼠标悬停在该原子指标属性框中 图标后,单击删除。

属性查看 > 原子指标属性					
age231 草稿 ceshiceg1					
☑ 编辑 ◎ 查看同源 : 1					
□ 克隆					
业务板块 Init_开发					
所属项目 ect_开发					
的建派生指标 项目空间类型					
数据域 🕁 下线					
英文名 也 删除 2					
名称: ceshiceg1					
计算逻辑: 品					
描述:					
数据类型: bigint					
数据叠加: 是					
统计周期标识: 字段: ds,格式类型: yyyy-MM-dd					

- a. 单击页面左下角的原子指标对象列表,进入对象列表页的原子指标页面。
 b. 将鼠标悬停在需要删除原子指标所在行的操作列下的 图标,单击删除。
- 5. 在提示对话框,填写备注信息。
- 6. 单击确定并提交,完成原子指标的删除。

下线原子指标

在原子指标页面,您可以通过以下任何一种方式,进入下线原子指标的提示对话框。
 选择需要下线的原子指标,将鼠标悬停在 图标,单击下线。



○ 单击需要下线的原子指标后,将鼠标悬停在该原子指标属性框中□图标后,单击下线。



○ a. 单击页面左下角的原子指标对象列表,进入对象列表页的原子指标页面。

b. 将鼠标悬停在需要下线原子指标所在行的操作列下的图图标,单击下线。

- 2. 在提示对话框,填写备注信息。
- 3. 单击确定并提交,完成原子指标的下线。

下线并删除原子指标

- 1. 在原子指标页面,您可以通过以下任何一种方式,进入下线并删除原子指标的提示对话框。
 - 选择需要下线并删除的原子指标,将鼠标悬停在 图标,单击下线并删除。
 - 单击需要下线的原子指标后,将鼠标悬停在该原子指标属性框中 图标后,单击**下线并删除**。

屢性查看 > 原子指标属性
y_b 已提交 y_b
☑ 编辑 ④ 查看同源 : 1
草稿 已发; 心 克隆
<>> 相关派生指标
业务板块 的建派生指标 Init_开发
所属项目 🕁 下线 ect_开发
项目空间类型 上下线并删除 2
数据域: 公 共域
英文名: y_b
名称: y_b
计算逻辑: 『
描述: y_b22
数据类型: bigint
数据叠加: 是

- a. 单击页面左下角的原子指标对象列表,进入对象列表页中的原子指标页面。
 - b. 将鼠标悬停在需要下线并删除原子指标所在行的操作列下的 图标,单击下线并删除,完成原子 指标的下线并删除。
- 2. 在提示对话框,填写备注信息。
- 3. 单击确定并提交,完成原子指标的下线并删除。

2.7. 规范定义-业务限定

2.7.1. 新建业务限定

业务限定用于圈定统计的业务范围。本文为您介绍如何新建业务限定。

前提条件

完成数据域的创建,详情请参见新建数据域。

背景信息

- 系统支持基于维度逻辑表或事实逻辑表构建业务限定。
- 业务限定用于标准化定义派生指标的条件限制,而原子指标用于标准化定义派生指标的计算逻辑。
- 为保证构建统一、标准和规范的派生指标,业务限定在业务板块内只能归属于一个来源逻辑表。

操作步骤

- 1. 登录Dataphin控制台。
- 2. 在Dataphin控制台页面,选择工作区后,单击进入Dataphin>>。
- 3. 进入业务限定页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。
 - iii. 在数据开发页面,单击规范建模。
 - iv. 在**规范建模**的左侧导航栏,单击委图标。
- 4. 在业务限定页面,您可以通过以下任何一种方式,进入新建业务限定页签。
 - 在**业务限定**页面,单击项目名称后的<mark>●</mark>图标,选择**规范定义 > 业务限定**。



- 在业务限定页面,单击业务限定后的■图标。
- 在数据开发的总览页面,单击业务限定下的 图标。

≡	Dataphin · 研发	开发≠ 发布	运维 权限		Q 全局搜索		늅 🕹 @ 라 💽
Den V.	na_r_beging_pic_ (b ⊕ ×) Prod Dev	欢迎使用Dataphin!					
规范	#模 ① 数据处理 即席查询						
*	业务过程 ひと		A 1		*		
	Q、输入关键字	◆规范定义 ≫				-	V
*	▼ ▲ 公共域 (1)		維度	业务过程 1 ③	原子指标 ● ◎	业务限定 [] ◎	派生描标 ● ◎
	~ buss		自动创建		Λ (\uparrow	
\$			Ý				·
~		(
		♥ 以油建模 ≫	維度逻辑表	事实逻辑表			汇总逻辑表
							+ 9

 ・ 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页面。单击新建业 务限定。

≡	Dataphin · 研发	开发≠	发布 运维	权限		Q、 全局撤出	¢	峃	🗢 ଉ	ଣ୍ଡ	
1977 192	Lealing pr	Ⅲ 对象列表页 维度 业务过程 质	× 仔指标 业务限定	派生描标						: 8	3
规范	建模 の 数据处理 即席查询	业务限定列表						2 + 新建业务	限定 (/ =	5
*	业务限定 じ 田 Q 输入关键字	Q 请输入关键字	所屬项目:	va_/_pering_project_							
	▼ La 公共城 (1)	所雇项目	C.T. Seepeg_projec			~					
٠		业务限定		数据域	开发状态	最近更新人	最近更新时间	所雇项目		操作	
*		X_age A		公共城	🔮 已提交	dataphic 7x807/1x1	2028-04-09 21:45:22	<pre>x0_7_teljing_proje x0_7_teljing_proje</pre>	ı A	25 i	
2											
	业务限定对参列表 1 >										

5. 在新建业务限定页签,选择数据域和来源表。

				: 8
选择来源信息				
业务板块 v2_T_memory_Moment_开展 来源表 fct_buss_tab1_di	所愿项目 va_f_Jeepe_project_dev	项目空间类型 应用层	設掘城 公 共城 >	
 业务限定所在来源 全局 模型相关 全局业务限定列表(0) 		し + 新建业务限定	来源表模型	
	똷	时还没有任何业务限定文件哦 + 新建业务限定		

- 6. 在新建业务限定页签,单击新建业务限定。
- 7. 在新建业务限定页面,配置参数。

新建业务限定	Ē							×
* 主要来源字段	请选择主动	来源字段			~ 0			
* 英文名	请输入英	文名		* 名称	请输入名称			
描述								0/128
* 计算逻辑	■ 格式化	⊘ 参考示例	🗊 规范性校验	0				
								(III)
						取消	保存	提交

参数	描述					
主要来源字段	选择业务限定的主要来源字段。					
英文名	填写业务限定的英文名。英文名由字母、数字或下划线(_)组合组成。					
名称	填写业务限定的名称。名称由汉字、字母、数字或下划线(_)或短划线(-)。					
描述	填写对业务限定的简单描述。					
	填写业务限定的计算逻辑。 i. 单击 计算逻辑 后面的 参考示例 ,可以查看SQL编写示例。					
计算逻辑	<pre>select province from dataphin_test where ds='\${bizdate}';</pre>					
	ii. SQL编写完成后,单击计算逻辑后的规范性校验,可以校验您编写的SQL是否 符合规范。					

8. 提交业务限定。

- i. 完成参数配置后, 单击**提交**。
- ii. 在提交备注页面,填写备注信息。
- iii. 单击确定并提交,完成业务限定的提交。
- 9. (可选)发布业务限定。
 - 如果您的开发模式是Dev-Prod模式,则需要发布新建的业务限定,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则不需要发布新建的业务限定。

2.7.2. 克隆业务限定

本文为您介绍如何克隆已有的业务限定。

前提条件

完成业务限定的创建, 详情请参见新建业务限定。

背景信息

操作步骤

- 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>>。
- 3. 进入业务限定页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模的左侧导航栏,单击≤图标。

- 4. 在业务限定页面,您可以通过以下任何一种方式,进入新建业务限定页面。
 - 在业务限定页的面左侧业务限定列表,选择需要克隆业务限定所在的数据域。单击某个业务限定后, 再将鼠标悬停在页面右侧弹出的该业务限定属性框中的 图标,单击克隆。
 - 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页签。将鼠标悬停

 在需要克隆业务限定所在行的操作列下的 图标,单击克隆。
- 5. 在**克隆业务限定**对话框,配置参数,详情请参见新建业务过程。
- 6. 提交业务限定。
 - i. 完成参数配置后, 单击提交。
 - ii. 在提交备注页面,填写备注信息。
 - iii. 单击确定并提交,完成业务限定的提交。
- 7. (可选)发布业务限定。
 - 如果您的开发模式是Dev-Prod模式,则需要发布业务限定,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则不需要发布业务限定。

2.7.3. 查看同源业务限定

Dataphin支持基于一个来源逻辑表生成多个业务限定。本文为您介绍如何查看同源业务限定。

前提条件

完成业务限定的创建,详情请参见新建业务限定。

操作步骤

- 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>>。
- 3. 进入业务限定页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。

ii. 在数据开发页面,单击项目名称后的☑图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模的左侧导航栏,单击≤图标。

- 4. 在业务限定页面,您可以通过以下任何一种方式,进入同源业务限定页面。
 - 在**业务限定**页面,选择需要查看同源业务限定所在的数据域。单击某个业务限定后,再单击该业务限 定属性框中的**查看同源**。
 - 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页面。将鼠标悬停
 在需要查看同源业务限定所在行的操作列下的▲图标。
- 5. 在同源业务限定页面,查看与该业务限定同源的其他业务限定。

2.7.4. 查看相关派生指标

本文为您介绍查看业务限定的相关派生指标。

前提条件

完成业务限定的创建和提交,详情请参见新建业务限定。

背景信息

- 因为系统仅支持已提交和开发中状态的业务限定构建派生指标,所以仅限已提交和开发中状态的业务限 定查看相关派生指标。
- 查看相关派生指标的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的规范建模。
 - 单击快速进入研发项目区域的Dev或Basic(非Data_distill)项目,选择开发数据的项目空间。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。
- 4. 在数据开发页面,单击项目名称后的型图标后,单击Dev或Basic页签,选择数据开发的项目空间 (非Data_distill项目)。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不需要 选择项目空间。

5. 在数据开发页面,单击规范建模。

如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模。

- 6. 在规范建模页签,单击左侧导航栏 when when a constraint of the constraint of th
- 在业务限定页面,单击需要查看相关派生指标的业务限定。在业务限定属性框,鼠标悬停至图标后,单击相关派生指标。

您也可以通过以下方式,查看相关派生指标:

- i. 在**业务限定**页面, 单击页面下方的**业务限定对象列表**。
- ii. 单击需要查看相关派生指标的原子指标后≤≤图标。
- 8. 在派生指标列表页面,查看基于该业务限定生成的派生指标。

2.7.5. 修改业务限定

业务限定用于圈定统计的业务范围。本文为您介绍如何修改业务限定。

前提条件

完成业务限定的创建,详情请参见新建业务限定。

背景信息

业务限定的计算逻辑修改成功后,相关的派生指标的统计分析逻辑随之更新。

操作步骤

- 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>>。
- 3. 进入业务限定页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模的左侧导航栏,单击≤图标。

- 4. 在**业务限定**页面,您可以通过以下任何一种方式,进入修改业务限定页签。
 - 在业务限定页面的左侧业务限定列表,选择需要修改业务限定所在的数据域。单击某个业务限定后, 再单击页面右侧弹出的该业务限定属性框中的编辑。
 - 在业务限定页面的左侧业务限定列表,选择需要修改业务限定所在的数据域。将鼠标悬停在某个业务
 限定后的图标,单击编辑。
 - 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页面。单击需要修 改业务限定所在行的操作列下的应图标。
- 5. 在编辑业务限定对话框,修改配置参数,详情请参见新建业务限定。

编辑业务限定								B	提交内容 🗸 🗙 🗙
* 主要来源字段	fct_bi				~	0			
* 英文名	x_age			* 名称	x_age				
描述	x_age								5/128
* 计算逻辑	◎ 格式化 1 fct_but	⊘ 参考示例 uss_tab2_df.d	同 规范性校验 im_gd.age>18 、	① cheng					
							取消	保存	提交

6. 提交业务限定。

- i. 完成参数配置后, 单击**提交**。
- ii. 在提交备注页面,填写备注信息。
- iii. 单击确定并提交,完成业务限定的提交。
- 7. (可选)发布业务限定。
 - 如果您的开发模式是Dev-Prod模式,则需要发布业务限定,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则不需要发布业务限定。

2.7.6. 下线和删除业务限定

本文为您介绍如何下线、下线并删除和删除不同状态的业务限定。

前提条件

完成业务限定的创建, 详情请参见新建业务限定。

背景信息

- 业务限定的状态的说明:
 - 保存成功的业务限定,状态为草稿。
 - 提交成功的业务限定, 状态为**已提交**。
 - 编辑已提交状态的业务限定,并成功保存,状态更改为开发中。
 - **已提交**状态的业务限定,成功下线后,状态更改为草稿。
- 下线业务限定的限制:
 - 仅开发中和已提交状态的业务限定,支持下线操作。
 - 仅限无相关派生指标的业务限定支持下线操作。
- 删除业务限定的限制: 仅草稿状态的业务限定, 支持删除操作。
- 下线并删除业务限定的限制:
 - 仅开发中和已提交状态的业务限定,支持下线并删除操作。
 - 仅限无相关派生指标的业务限定支持下线并删除操作。

删除业务限定

- 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>>。
- 3. 进入业务限定页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。
 - 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。
 - iii. 在数据开发页面,单击规范建模。
 - iv. 在规范建模的左侧导航栏,单击≤图标。
- 4. 在**业务限定**页面,您可以通过以下任何一种方式,进入删除业务限定的提示页面。
 - 在业务限定页面的左侧业务限定列表,选择需要删除业务限定所在的数据域。将鼠标悬停在某个业务
 限定■图标后,单击删除。
 - 在业务限定页面的左侧业务限定列表,选择需要删除业务限定所在的数据域。单击某个业务限定后, 将鼠标悬停在页面右侧弹出的该业务限定属性框中的□图标后,单击删除。
 - 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页面。单击需要删除业务限定所在行的操作列下的□图标后,单击删除。
- 5. 在提示页面,填写备注信息。
- 6. 单击提交并提交,完成业务限定的删除。

下线业务限定

- 1. 在业务限定页面,您可以通过以下任何一种方式,进入下线业务限定的提示页面。
 - 在业务限定页面的左侧业务限定列表,选择需要下线业务限定所在的数据域。将鼠标悬停在某个业务
 限定■图标后,单击下线。
 - 在业务限定页面的左侧业务限定列表,选择需要下线业务限定所在的数据域。单击某个业务限定后, 将鼠标悬停在页面右侧弹出的该业务限定属性框中的□图标后,单击下线。
 - 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页面。单击需要下
 线业务限定所在行的操作列下的 图标后,单击下线。
- 2. 在提示页面,填写备注信息。
- 3. 单击提交并提交,完成业务限定的下线。

下线并删除业务限定

- 1. 在业务限定页面,您可以通过以下任何一种方式,进入下线并删除业务限定的提示页面。
 - 在**业务限定**页面的左侧业务限定列表,选择需要下线并删除业务限定所在的数据域。将鼠标悬停在某 个业务限定■图标后,单击下线并删除。
 - 在业务限定页面左下方,单击业务限定对象列表,进入对象列表页的业务限定页面。单击需要下 线并删除业务限定所在行的操作列下的 图标后,单击下线并删除。

- 在业务限定页面的左侧业务限定列表,选择需要下线并删除业务限定所在的数据域。单击某个业务限定后,将鼠标悬停在页面右侧弹出的该业务限定属性框中的图标后,单击下线并删除。
- 2. 在提示页面,填写备注信息。
- 3. 单击提交并提交,完成业务限定的下线并删除。

2.8. 规范定义-派生指标

2.8.1. 新建派生指标

派生指标用于圈定原子指标统计业务的范围。本文为您介绍如何新建派生指标。

前提条件

完成数据域的创建,详情请参见新建数据域。

背景信息

Dataphin支持构建标准、规范及无二义性的派生指标。派生指标的说明如下:

- 派生指标由原子指标、业务限定、统计周期和统计粒度组成。
- 系统支持基于维度逻辑表或事实逻辑表构建派生指标。
- 派生指标在业务板块内只能归属于一个来源逻辑表。
- 因为原子指标和派生指标可能来自不同数据域的逻辑表,所以派生指标可能归属于多个数据域。
- 新建派生指标的权限限制, 详情请参见数据开发权限列表。

操作步骤

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

您可以通过以下方式,快速进入工作区:

- 选择工作区地域后, 单击**快速开始相关工作**区域中相应的功能模块。
- 选择工作区地域后,单击**快速进入研发项目**后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3的i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3的ii。

3. 进入派生指标页面。

- i. 在Dataphin首页, 单击顶部菜单栏中的研发。
- ii. 在数据开发页面,单击项目名称后的▼图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要再次选择项目。
- iii. 在数据开发页面,单击规范建模。

iv. 在规范建模的左侧导航栏,单击**云派生指标**图标。

如果进入数据**开发**页面后,系统默认进入**规范建模**页签,则无需再次单击**规范建模**页签。 4. 您可以通过以下方式,进入**新建派生指标**页签。

- 单击左侧导航栏项目名称后的 图标,选择规范定义 > 派生指标。
- 在派生指标页面,单击派生指标后的■图标。
- 在数据开发的总览页面,单击派生指标下的 图标。
- • 在派生指标页面左下方,单击派生指标对象列表。在对象列表页的派生指标页面,单击新建派生指标。
- 5. 构建派生指标。
 - i. 在新建派生指标页签, 选择原子指标。
 - ii. 选择原子指标后,单击下一步。
 - iii. 在创建派生指标区域, 配置参数。

参数	描述
统计粒度	选择派生指标的统计粒度。您可以单击 新建统计粒度 ,增加派生指标的统计粒 度,但系统仅支持添加3个统计粒度。
统计周期	统计周期用于定义派生指标来源数据的时间跨度。例如,选择了最近7天,则表示统计最近7天内的事实行为数据。 您可以单击 新建统计周期 ,增加派生指标的统计周期,但系统仅支持添加3个统 计周期。
业务限定	选择派生指标的业务限定,系统仅支持选择与所选原子指标具有相同数据时效的 业务限定。您可以单击 新建业务限定 ,增加派生指标的业务限定,但系统仅支 持添加3个业务限定。

iv. 完成参数配置后, 单击**预生成派生指标**。

- 6. (可选)在编辑派生指标页面,您可以修改派生指标英文名和派生指标名称。
- 7. 保存和提交派生指标。
 - i. 单击**保存**,保存派生指标。
 - ii. 单击提交, 提交派生指标。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交,完成派生指标的提交。
- 8. (可选)发布派生指标至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布新建的派生指标,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的派生指标,即可进入生产环境。

2.8.2. 修改派生指标

派生指标用于圈定原子指标统计业务的范围。本文为您介绍如何修改派生指标。

前提条件

完成派生指标的创建,详情请参见新建派生指标。

操作步骤

- 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>>。
- 3. 进入**派生指标**页签。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的**▼**图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在**规范建模**的左侧导航栏,单击**骤**图标。

- 4. 在派生指标页签, 您可以通过以下任何一种方式, 进入编辑派生指标页对话框。
 - 在派生指标页签的左侧派生指标列表,选择需要修改派生指标所在的数据域。单击某个派生指标后, 再单击页面右侧弹出的该派生指标属性框中的编辑。
 - 在派生指标页签的左侧派生指标列表,选择需要修改派生指标所在的数据域。将鼠标悬停在某个派生指标后的图图标,单击编辑。
 - 在派生指标页签左下方,单击派生指标对象列表,进入对象列表页的派生指标页签。单击需要修 改派生指标所在行的操作列下的应图标。
- 5. 在编辑派生指标对话框,修改配置参数,详情请参见新建派生指标。

编辑派生指标				×
派生指标英文名	car_1d_ age			
* 派生指标名称	最近1天_test02_test_01			
描述	最近1天_test02_test_01			19/128
		取消	保存	提交

6. (可选)保存派生指标。

如果派生指标不需要提交,则单击保存,保存修改后的派生指标。

7. 提交派生指标。

i. 单击提交, 提交派生指标。

- ii. 在提交备注页面,填写备注信息。
- iii. 单击确定并提交,完成派生指标的提交。
- 8. (可选)发布派生指标。
 - 如果您的开发模式是Dev-Prod模式,则需要发布派生指标,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则不需要发布派生指标。

2.8.3. 下线和删除派生指标

本文为您介绍如何下线、下线并删除和删除不同状态的派生指标。

前提条件

完成派生指标的创建,详情请参见新建派生指标。

背景信息

- 派生指标的状态的说明:
 - 保存成功的派生指标,状态为草稿。
 - 提交成功的派生指标,状态为**已提交**。
 - 编辑已提交状态的派生指标,并成功保存,状态更改为开发中。
 - 已提交状态的派生指标,成功下线后,状态更改为草稿。
- 下线、删除和下线并删除派生指标的限制说明:
 - 仅开发中和已提交状态的派生指标,支持下线操作。
 - · 仅草稿状态的派生指标,支持删除操作。
 - 仅开发中和已提交状态的派生指标,支持下线并删除操作。

删除派生指标

- 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>>。
- 3. 进入派生指标页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击< 图标。
- 4. 在派生指标页面,您可以通过以下任何一种方式,进入删除派生指标的提示页面。
 - 在派生指标页面的左侧派生指标列表,选择需要删除派生指标所在的数据域。将鼠标悬停在某个派生指标
 图标后,单击删除。
 - 在派生指标页面左下方,单击派生指标对象列表,进入对象列表页的派生指标页面。单击需要删 除派生指标所在行的操作列下的 圖 图标。

- 5. 在提示页面,填写备注信息。
- 6. 单击提交并提交,完成派生指标的删除。

下线派生指标

- 1. 在派生指标页面,您可以通过以下任何一种方式,进入下线派生指标的提示页面。
 - 在派生指标页面的左侧派生指标列表,选择需要下线派生指标所在的数据域。将鼠标悬停在某个派生指标
 图标后,单击下线。
 - 在派生指标页面的左侧派生指标列表,选择需要下线派生指标所在的数据域。单击某个派生指标后, 将鼠标悬停在页面右侧弹出的该派生指标属性框中的 图标后,单击下线。
 - 在派生指标页面左下方,单击派生指标对象列表,进入对象列表页的派生指标页面。单击需要下 线派生指标所在行的操作列下的圆标。
- 2. 在提示页面,填写备注信息。
- 3. 单击提交并提交,完成派生指标的下线。

下线并删除派生指标

- 1. 在派生指标页面,您可以通过以下任何一种方式,进入下线并删除派生指标的提示页面。
 - 在派生指标页面的左侧派生指标列表,选择需要下线并删除派生指标所在的数据域。将鼠标悬停在某个派生指标图标后,单击下线并删除。
 - 在派生指标页面左下方,单击派生指标对象列表,进入对象列表页的派生指标页面。单击需要下 线并删除派生指标所在行的操作列下的画图标。
 - 在派生指标页面的左侧派生指标列表,选择需要下线并删除派生指标所在的数据域。单击某个派生指标后,将鼠标悬停在页面右侧弹出的该派生指标属性框中的□图标后,单击下线并删除。
- 2. 在提示页面,填写备注信息。
- 3. 单击提交并提交,完成派生指标的下线并删除。

2.8.4. 查看相关汇总逻辑表

Dataphin支持查看某个派生指标所关联的汇总逻辑表。本文为您介绍查看派生指标的相关汇总逻辑表。

前提条件

完成派生指标的创建,详情请参见新建派生指标。

背景信息

因为系统只能汇聚**已提交**状态的派生指标至汇总逻辑表中,所以仅限**已提交**状态的派生指标支持查看相关 汇总逻辑表。

操作步骤

- 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>>。
- 3. 进入派生指标页面。
 - i 在Datanhin首而 单击顶部莖单栏由的研发

ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

iii. 在数据开发页面,单击规范建模。

iv. 在规范建模的左侧导航栏,单击<Ⅰ图标。

- 4. 在派生指标页面,您可以通过以下任何一种方式,进入汇总逻辑表页面。
 - 在派生指标页面左下方,单击派生指标对象列表,进入对象列表页的派生指标页面。将鼠标悬停

 在需要查看相关汇总逻辑表的派生指标所在行的操作列下的图标后,单击查看汇总表。
 - 在派生指标页面的左侧派生指标列表,选择需要查看相关汇总逻辑表的派生指标所在的数据域。单击 某个派生指标后,单击页面右侧弹出的该派生指标属性框中的查看相关。
- 5. 在汇总逻辑表页面,查看该派生指标的相关汇总逻辑表。

2.9. 逻辑表-汇总逻辑表

2.9.1. 新建汇总逻辑表

汇总逻辑表用于统计以统计粒度(维度)为主题的所有数据。本文为您介绍如何新建汇总逻辑表。

前提条件

- 完成维度的创建,详情请参见新建维度。
- 完成派生指标的创建,详情请参见新建派生指标。

背景信息

汇总逻辑表的说明如下:

- 汇总逻辑表由一个特定的统计粒度(维度)和该统计粒度相关的统计指标(原子指标)组成。
- 汇总逻辑表的统计指标(原子指标)具有相同的统计粒度(维度)。
- 如果派生指标和汇总逻辑表的统计粒度一致,则系统会自动加入派生指标至汇总逻辑表。
 汇总逻辑表的统计指标包括以下两种来源:
- 基于系统自动汇聚的派生指标,构建汇总逻辑表的统计指标。
- 通过创建自定义指标的方式,构建汇总逻辑表的统计指标。

新建汇总逻辑表的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后, 单击**快速开始相关工作**区域中相应的功能模块。
 - 选择工作区地域后,单击快速进入研发项目后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3的i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3的ii。

3. 进入汇总逻辑表页面。

- i. 在Dataphin首页, 单击顶部菜单栏中的研发。
- ii. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。
 如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目为您的数据开发空间,则不需要再次选择项目。
- iii. 在数据开发页面,单击规范建模。如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
- iv. 在规范建模的左侧导航栏,单击回汇总逻辑表图标。
- 4. 您可以通过以下方式,进入新建汇总逻辑表页面。
 - 单击左侧导航栏项目名称后的 图标,选择逻辑表 > 汇总逻辑表。
 - 在汇总逻辑表页面,单击汇总逻辑表后的■图标。
 - 在数据开发的总览页面,单击汇总逻辑表下的 图标。
- 5. 在新建汇总逻辑表配置向导页面,完成以下操作。
 - i. 在主键信息页面, 选择统计粒度。

统计粒度为已提交或已发布的维度。

- ii. 选择粒度后, 单击进入下一步。
- 6. 提交汇总逻辑表。
 - i. 完成新建汇总逻辑表配置向导页面的参数配置后, 单击提交。
 - ii. 在提交备注对话框,填写备注信息。
 - iii. 单击确定并提交,完成汇总逻辑表的提交。
- 7. (可选)发布汇总逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布新建的汇总逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的汇总逻辑表,即可进入生产环境。

2.9.2. 编辑汇总逻辑表

本文为您介绍如何编辑汇总逻辑表。

前提条件

完成创建汇总逻辑表,详情请参见新建汇总逻辑表。

背景信息

您可以通过编辑汇总逻辑表的方式,执行以下操作:
- 添加派生指标至汇总逻辑表。
- 添加自定义指标至汇总逻辑表。
- 修改汇总逻辑表中的派生指标。
- 查看汇总逻辑表中派生指标和自定义指标的详情。
- 归类汇总逻辑表中的派生指标和自定义指标。
- 删除汇总逻辑表中的自定义指标。

编辑汇总逻辑表的权限限制,详情请参见数据开发权限列表。

添加派生指标

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的规范建模。
 - 单击快速进入研发项目区域的Dev或Basic(非Data_distill)项目,选择开发数据的项目空间。

? 说明

- 如果您通过**快速开始相关工作**进入数据开发模块,则跳过步骤3。
- 如果您通过**快速进入研发项目**进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏中的研发。
- 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的项目空间 (非Data distill项目)。

如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目空间为您的数据开发空间,则不需要 选择项目空间。

- 5. 进入汇总逻辑表页面。
 - i. 在数据开发页面, 单击规范建模。

如果您进入数据开发页面后,默认进入规范建模页签,则不需要再次单击规范建模页签。

- ii. 在规范建模的左侧导航栏,单击了汇总逻辑表图标。
- 6. 鼠标悬停至需要添加派生指标的汇总逻辑表后,再将鼠标悬停至该汇总逻辑表的图图标后,单击编辑。
- 7. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则单击页面右上方的 △图标。

当高图标变更为高图标后,则解锁成功。

- 8. 添加派生指标。
 - i. 在该汇总逻辑表的编辑页面, 单击派生指标页签。

如果您进入该汇总逻辑表页面后,默认进入派生指标页签,则不需要再次单击派生指标页签。

ii. 在派生指标页签, 单击新建派生指标。

- iii. 在新建派生指标页面, 配置参数, 详情请参见新建派生指标。
- 9. 提交和发布汇总逻辑表。
 - i. 派生指标添加完成后,单击编辑汇总逻辑表页面上方的 《图标,提交汇总逻辑表。
 - ii. 在提交备注页面,填写备注信息。
 - iii. 单击确定并提交,完成汇总逻辑表的提交。
 - iv. (可选)发布汇总逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布汇总逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的汇总逻辑表,即可进入生产环境。

修改派生指标

- 1. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:
 - a. 在**汇总逻辑表**页面,单击需要修改派生指标的汇总逻辑表。
 - b. 单击该汇总逻辑表编辑页面右上方的 高图标。

当高图标变更为高图标后,则解锁成功。

- 2. 在派生指标页签, 单击需要修改的派生指标操作列下的编辑。
- 3. 在编辑派生指标对话框,修改参数。

编辑派生指标						×
派生指标英文名	order_count_1d_	cesghi				
* 派生指标名称	最近1天_order_cd	unt				
描述	最近1天_order_cd	unt				16/128
			取消	保	存	提交
参数	描述					
	派生指杨	英文名由字f	母、数字或下划]线(_)组合	组成。	
派生指标英文名	? i	,明 派生指	旨标英文名 不俞	指超过54个字	符。	

参数	描述			
	派生指标名称由汉字、数字、字母及下划线(_)或短划线(-)。			
派生指标名称	⑦ 说明 派生指标英文名不能超过64个字符。			
描述	添加或修改派生指标的描述。			

- 4. 单击提交。
- 5. 在提交备注页面,填写备注信息。
- 6. 单击确定并提交,完成派生指标的编辑。
- 7. 提交汇总逻辑表。
 - i. 单击汇总逻辑表页面上方的 / 图标, 提交汇总逻辑表。
 - ii. 在提交备注页面,填写备注信息。
 - iii. 单击确定并提交,完成汇总逻辑表的提交。
- 8. (可选)发布汇总逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布汇总逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的汇总逻辑表,即可进入生产环境。

查看派生指标详情

- 1. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:
 a. 在汇总逻辑表页面,单击需要查看派生指标详情的汇总逻辑表。
 - b. 单击该汇总逻辑表编辑页面右上方的△图标。

当高图标变更为高图标后,则解锁成功。

- 2. 在派生指标页签,单击需要查看的派生指标操作列下的详情。
- 3. 在派生指标属性框中,查看派生指标的所属业务板块、项目及原子指标等信息。

归类派生指标

- 1. (可选) 解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:
 - a. 在**汇总逻辑表**页面,单击需要归类派生指标的汇总逻辑表。
 - b. 单击该汇总逻辑表编辑页面右上方的叠图标。

当 图标变更为 图标后,则解锁成功。

2. 在派生指标页签, 单击需要归类的派生指标操作列下的分类。

3. 在指标归类对话框,您可以创建分类和归类指标。

指标归类			×
最近1天_order_count order_count_1d		1 归类至 Q 搜索或创建分类	+创建分类 2 创建分类
已归类至:(0)		2 暂时没有任何分组,请先创建: 请输入分组名称,快速创建	创建
		取消	确定
编号	描述		
٠	 · 归类指标: a.选中分类。 您也可以在搜 b.单击确定,另 · 创建分类: a.单击创建分 b.在对话框的输 	覺索框中,输入已有分类名称快速搜索分类。 完成指标归类。 €。 前入框中,输入分类名称后,单击 确定 。	
2	如果您还没有分类, ! i. 在输入框中, 输 ii. 单击 创建 。 ⑦ 说明 如果该	则执行以下操作新建分类: 入分类名称。 该汇总逻辑表已有分类,则系统不展示②区域。	

添加自定义指标

您可以通过添加自定义指标的方式,将本项目中您负责的物理表中的字段,添加至汇总逻辑表中进行分析计算。

- 1. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:

a. 在汇总逻辑表页面,单击需要添加自定义指标的汇总逻辑表。

b. 单击该汇总逻辑表编辑页面右上方的 @ 图标。

当 图标变更为 图标后,则解锁成功。

2. 在汇总逻辑表编辑页面, 单击自定义指标。

如果您进入该汇总逻辑表页面后,默认进入自定义指标页签,则不需要再次单击自定义指标页签。

- 3. 在自定义指标页签, 单击新建自定义指标。
- 4. 在新建自定义指标配置向导页面,完成以下操作。
 - i. 在主键信息页面, 选择来源物理表、维度主键及关联的逻辑。
 - ii. 单击关联并下一步。
 - iii. 在汇总逻辑表描述信息页面,选中左侧区域的字段后,单击添加。 您可以在搜索框中,输入字段英文名的关键字,快速搜索字段。
 - iv. 在右侧功能区域的新增列表页签,填写字段英文名和字段名称。 您也可以单击操作列下的删除,删除某个字段。
 - v. 单击确定,完成自定义指标的添加。
- 5. 提交汇总逻辑表。
 - i. 自定义指标添加完成后,单击编辑汇总逻辑表页面上方的 图标,提交汇总逻辑表。
 - ii. 在提交备注页面,填写备注信息。
 - iii. 单击确定并提交,完成汇总逻辑表的提交。
- 6. (可选)发布汇总逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布汇总逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的汇总逻辑表,即可进入生产环境。

查看自定义指标详情

- 1. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:

a. 在汇总逻辑表页面, 单击需要查看自定义指标详情的汇总逻辑表。

b. 单击该汇总逻辑表编辑页面右上方的册图标。

当高图标变更为高图标后,则解锁成功。

2. 在自定义指标页签, 单击需要查看自定义指标的操作列下的详情。

您可以通过以下方式,快速搜索需要查看的自定义指标:

- 在搜索框中, 输入自定义指标名称关键字。
- 选中某个物理表名或单击──图标后,选中某个汇总逻辑表指标分类。

3. 在该自定义指标的属性框中,查看表结构和表信息。

删除自定义指标

- 1. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:
 a. 在汇总逻辑表页面,单击需要删除自定义指标的汇总逻辑表。
 b. 单击该汇总逻辑表编辑页面右上方的高图标。

当高图标变更为高图标后,则解锁成功。

- 在自定义指标页签,单击需要删除的自定义指标操作列下的删除。
 您可以通过以下方式,快速查询需要删除的自定义指标:
 - 在搜索框中, 输入自定义指标名称关键字。
 - 选中某个物理表名或单击 ≥图标后,选中某个汇总逻辑表指标分类。
- 3. 在确认对话框中, 单击确定, 完成指标的删除。
- 4. 提交汇总逻辑表。
 - i. 单击编辑汇总逻辑表页面上方的 / 图标, 提交汇总逻辑表。
 - ii. 在提交备注页面,填写备注信息。
 - iii. 单击确定并提交,完成汇总逻辑表的提交。
- 5. (可选)发布汇总逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布汇总逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的汇总逻辑表,即可进入生产环境。

归类自定义指标

- 1. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则执行以下操作,解锁汇总逻辑表:
 a. 在汇总逻辑表页面,单击需要归类自定义指标的汇总逻辑表。
 - b. 单击该汇总逻辑表编辑页面右上方的△图标。

当@图标变更为@图标后,则解锁成功。

- 在自定义指标页签,单击需要归类的自定义指标操作列下的分类。
 您可以通过以下方式,快速查询需要归类的自定义指标:
 - 在搜索框中, 输入自定义指标名称关键字。
 - 选中某个物理表名或单击 ⊠图标后,选中某个汇总逻辑表指标分类。
- 3. 在指标归类对话框,您可以创建分类和归类指标。

指标归类			Х
最近1天_order_count order_count_1d		 1 /ul>	+创建分类 12 创建分类
已归类至:(0)		2 暂时没有任何分组,请先创建 请输入分组名称,快速创建	创建
		取消	确定
编号	描述		
0	 归类指标: a. 查询分类。 在搜索框中, b. 选中分类。 c. 单击确定, 完 创建分类: a. 在输入框中, b. 单击创建分类; 	输入已有分类名称,快速搜索分类。 6成指标归类。 输入分类名称。 餐。	
2	如果您还没有分类, 『 i. 在输入框中, 输 ii. 单击 创建 。 ⑦ 说明 如果该	则需要新建分类: 入分类名称。 《汇总逻辑表已有分类,则系统不展示②区域。	

2.9.3. 物理化配置汇总逻辑表

物理化配置用于配置汇总逻辑表的自定义参数。本文为您介绍如何对汇总逻辑表进行物理化配置。

前提条件

完成汇总逻辑表的创建,详情请参见新建汇总逻辑表。

背景信息

汇总逻辑表默认存储生命周期为36000天,暂不支持自定义修改。

操作步骤

- 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>>。
- 3. 进入汇总逻辑表页面。
 - i. 在Dataphin首页,单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的▼图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击回图标。
- 在 汇总逻辑表页面的左侧汇总逻辑表列表,将鼠标悬停在需要物理化配置的汇总逻辑表后的图图标,单击编辑。
- 5. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定,则需要单击页面右上方的叠图标,解锁汇总逻辑表。
- 6. 在该汇总逻辑表的编辑页面,单击页面上方的物理化配置。
- 7. 根据物理化配置对话框提示, 配置自定义参数, 详情请参见配置自定义任务参数。
- 8. 提交汇总逻辑表。
 - i. 完成配置自定义参数后,单击编辑汇总逻辑表页面上方的 / 图标。
 - ii. 在提交备注页面,填写备注信息。
 - iii. 单击确定并提交,完成汇总逻辑表的提交。
- 9. (可选)发布汇总逻辑表。
 - 如果您的开发模式是Dev-Prod模式,则需要发布汇总逻辑表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布汇总逻辑表。

2.9.4. 查看汇总逻辑表信息

汇总逻辑表用于统计以统计粒度(维度)为主题的所有数据。本文为您介绍如何查看汇总逻辑表信息。

前提条件

完成汇总逻辑表的创建,详情请参见新建汇总逻辑表。

背景信息

查看汇总逻辑表信息的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后,单击快速开始相关工作区域中相应的功能模块。
 - 选择工作区地域后,单击**快速进入研发项目**后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3的i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3的ii。
- 3. 进入汇总逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Dat a_dist ill)项目,且项目为您的数据开发空间,则不需要 再次选择项目。

- iii. 在数据开发页面,单击规范建模。
 如果进入数据开发页面后,系统默认进入规范建模页签,则无需再次单击规范建模页签。
- iv. 在规范建模的左侧导航栏,单击回汇总逻辑表图标。
- 在 汇总逻辑表页面的左侧汇总逻辑表列表,鼠标悬停至需要查看详情的汇总逻辑表 图标后,单击编辑。
- 5. (可选)解锁汇总逻辑表。
 - 如果汇总逻辑表已被您自己锁定,则不需要解锁。
 - 如果汇总逻辑表已被他人锁定且您需要修改汇总逻辑表信息,则单击页面右上方的叠图标。

当高图标变更为高图标后,则解锁成功。

6. 在该汇总逻辑表的编辑页面,单击页面上方的表详情。

在汇总逻辑表信息页面,您可以执行如下操作:

- 查看汇总逻辑表的**主键信息**。
- 查看汇总逻辑表描述信息或修改描述。
- 7. 单击确定。
- 8. (可选)提交和发布汇总逻辑表。

如果您修改了**汇总逻辑表描述信息**中的**描述、名称**或**英文名**信息,则需要提交或发布汇总逻辑表至生 产环境。提交和发布汇总逻辑表的操作步骤如下:

- i. 单击页面右上方的 √ 图标。
- ii. 在提交备注页面,填写备注信息。
- iii. 单击确定并提交,完成汇总逻辑表的提交。
- iv. (可选)发布汇总逻辑表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布汇总逻辑表,详情请参见发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的汇总逻辑表,即可进入生产环境。

2.9.5. 删除汇总逻辑表

汇总逻辑表用于统计以统计粒度(维度)为主题的所有数据。本文为您介绍如何删除汇总逻辑表。

前提条件

- 完成汇总逻辑表的创建,详情请参见新建汇总逻辑表。
- 完成汇总逻辑表依赖的派生指标的删除。如何删除派生指标,请参见删除派生指标。

背景信息

系统仅支持删除没有派生指标和自定义指标的汇总逻辑表。

操作步骤

- 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>>。
- 3. 进入汇总逻辑表页面。
 - i. 在Dataphin首页, 单击顶部菜单栏中的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标,选择Dev或Basic(非Data_distill)项目。

如果您当前访问的是Dev或Basic(非Data_distill)项目,且项目空间为您的数据开发空间,则不 需要选择项目空间。

- iii. 在数据开发页面,单击规范建模。
- iv. 在规范建模的左侧导航栏,单击 图标。
- 4. 将鼠标悬停在需要删除汇总逻辑表后的图标,单击删除。

注意 如果汇总逻辑表存在依赖的派生指标,则无法删除汇总逻辑表。

- 5. 在提示页面,填写备注信息。
- 6. 单击确定并提交,完成汇总逻辑表的删除。

2.10. 规范建模-建模引擎

Dataphin逻辑表开发、提交和调度运行,会自行根据系统内置规则对逻辑表模型进行校验、物理化,提供查询使用。标准规范、可靠稳定及高效的逻辑表研发过程依赖于智能建模引擎服务。

逻辑表查询SQL

Dat aphin支持逻辑表雪花模型的查询,即支持基于[逻辑模型.关联维度.关联维度.….属性](如订单.买家.会员 类型.类型值)查询所需数据,提高查询类SQL的编码效率。



- 事实逻辑表/维度逻辑表查询
 - SQL格式说明如下。

```
select Column,
```

LTable.Column, //逻辑表名.字段名

```
LTable.Role(relation - dimension).....Dimension_Column, //逻辑表名.角色名(关联维
```

度字段名).维度属性字段名

```
from LD.LTable //业务板块名.逻辑表名
```

where ...

```
select Column,
```

```
LTable_Alias.Column,
LTable_Alias.Role(relation - dimension).....Dimension_Column
from LD.LTable Alias
where ...
```

。 SQL格式举例如下。

- 汇总逻辑表查询
 - 。 SQL格式说明如下。

```
select Column,
LTable.Column,
LTable.Dimension(granularity).....Dimension_Column,
from LD.LTable
where ...
```

select Column, Alias.Column, Alias.Dimension (granularity).....Dimension_Column from LD.LTable Alias where ...

。 SQL格式举例如下。

```
select customer id,
       c.dim customer.address line1,
       c.dim_customer.dim_tax_rate.tx_name,
       crt trd comm amt 30d trd valid,
       acct cnt td actv account,
       watch sec cnt td actv watches,
       avg crt trd comm 7d,
       crt trd acct cnt 7d,
       crt trd cnt 1d,
       avg crt trd comm 1d,
       crt trd comm amt 1d trd valid,
       crt trd acct cnt 1d,
       watch sec cnt td inac watches,
       watch sec cnt td,
       acct cnt td,
       crt trd cnt 7d
from LD DEMO.DWS CUSTOMER c
where ds='${bizdate}'
limit 100;
```

? 说明

- 如果维度逻辑表两个角色引用的字段有重名, 需要给两个字段命别名, 以防字段名冲突。
- 仅主维度的维度逻辑表可以被查询(子维度的维度逻辑表不可查询),包含被作为关联维度(取角色名)的情况。
- 子维度的维度逻辑表包含的字段, 合并在主维度的维度逻辑表字段list中。
- 组成汇总逻辑表统计粒度的维度,可以通过类似关联维度方式引用,汇总逻辑表 select *可获 取[维度_主键]数据。为防全表扫描过慢,建议使用 select 汇总表.维度表.主键获取主键, select 汇总表.维度表.属性获取属性信息。

计算逻辑重复性校验

Dataphin核心价值之一定义无二义性,且命名、计算逻辑无重复。提交规范定义、逻辑表对象时,Dataphin 会校验对象的英文名、中文名、计算逻辑完全没有重复。如果计算逻辑存在重复,则进行系统提示,避免误 创建同名不同义、或者同义不同名的对象。

新建或编辑之后,在您提交并发布时间周期、维度、维度逻辑表字段、事实逻辑表、事实逻辑表字段、原子 指标、业务限定、汇总逻辑表自定义指标时,会基于AST抽象语法树结构进行解析,在命名空间内进行计算 逻辑(或表达式)重复性校。如果校验有重复,则提示对象计算逻辑重复。

模型多版本动态刷新

动态刷新可以提高逻辑表提交及物理化的效率,增加逻辑表变更的灵活性,也可以降低物理表变更及历史数 据变更迁移的计算引擎源资源消耗。在SQL任务调度时,系统会基于当前最新的逻辑表物理化情况自动识别 和路由对应的物理表,从而获取查询所需的数据。

逻辑表的版本信息,可以在逻辑表编辑页面上方的菜单栏中,单击版本信息查看。

3.编码研发

3.1. 数据开发概述

数据开发包括编码研发和规范建模。编码研发用于构建计算任务,例如创建SQL代码任务, Shell任务, Python 任务, MR任务和Spark任务;规范建模用于构建逻辑化的数据模型。

功能介绍

 规范建模:基于已明确的业务需求,从业务顶层开始设计,规范定义数据的标准及计算口径。同时,根据 业务需求完成数据模型的设计。

规范建模为您提供可视化、标准化、规范化和自动化定义数据,而非传统手动写SQL代码的方式,这样就 避免了统计指标计算口径不一致的情况。可视化构建数据仓库的模型,逻辑表提交后即生成对应智能黑盒 调度任务,自动化生产数据。

⑦ 说明 仅绑定数据板块的项目支持规范建模功能。

- 数据处理:通过编写代码的方式构建复杂的数据模型、构建同步任务和构建不同类型的代码任务。
- 即席查询:即主题式查询,面向业务主题的数据查询,屏蔽了物理模型中技术特性带来的影响,基于逻辑 模型从业务视角出发对外提供查询服务。

Dataphin支持Basic和Dev-Prod的数据开发模式:

- Basic项目绑定了Basic数据板块,系统支持规范建模功能。Basic项目绑定了Prod数据板块,则系统不 支持规范建模功能,仅支持数据处理和即席查询功能。
- Dev项目支持规范建模、数据处理和即席查询功能。Prod项目支持规范建模和数据处理功能。

数据开发界面介绍

Dataphin	研2 Prod Dev	v324_regression_test_project	集成 开发 发布 运维	5 9 全局搜索	麿 酋 ぱ �、 (🧕
	_ Ů ∄ ⊡				
我的工作流 3		欢迎使用Dataphin!			
最近打开			6		
规范建模		◎ 別U220単位 //	•	51-163# //	
📄 维度逻辑表					MaxCompute任务
🔜 事实逻辑表					POL D
<u>人</u> 原子指标		指示 + ◎	に想逻辑表 ◎		 交
★ 业务限定				发布管理	MAX_COMPU + SPARK_JAR +
× ##15		•		独立的环境运行提交后的任务和数 据,以更好地验证数据可用性,同	SHELL + PYTHON +
1410				时方便进行生产数据的发档管理。	VIRTUAL +
🗹 汇总逻辑表		*			FLINK(F)@
数据处理	∞ '		JL/45. EB star	没布	
🚬 计算任务		÷ ©	± 0	\downarrow	SQL For D
同步任务	0220112				
表管理				200	DATASTREAM +
计算模版					数据表
				运用 網路 运行及维护逻辑责和教授外理产生	元表 + 镜像表 +
📄 即席查询		事实逻辑表	维度逻辑表	的数据任务,以确保所有任务都有 按照时间正确有序地生产数据。	
亩 回收站)_pg	● ◎	• •		政据集成 反 体

② 说明 如果您的Dataphin实例为基础研发版,则不支持规范建模。

功能区编号	描述
0	当前打开的项目名称,以及项目的环境类型(Prod或Dev)。您可以单击项目名称后的 <mark>▼</mark> 图标,在下拉列表中,依次选择项目分类(Prod、Dev或Basic)及项目名称,切换项 目。同时,您也可以在搜索框中输入项目名称关键词,快速搜索项目,提高了搜索的效 率。
2	环境类型标识按钮,单击 图标,切换Dev和Prod项目。Prod项目表示生产环境的项目,Dev项目表示开发环境项目。 例如,当前您访问的是Dev项目(编号②的环境类型标识为 PPB 图标),单击 图标后,切 换至Prod项目(编号②的环境类型标识为 PPB 图标)。 ⑦ 说明 Basic项目没有环境类型标识按钮。
3	快速新建规范建模、数据处理、回收站、即席查询任务的入口。同时为您记录最近打开的 任务,便于您快速查询。
4	刷新任务列表、新建文件和新建文件夹的按钮。
\$	在全局搜索框中输入物理表、逻辑表、规范定义、函数或资源等名称的关键字,搜索目标 对象。您可以单击 研发、萃取和资产 页签,快速搜索目标对象。 同时,您也可以单击 图标,在代码搜索框中,输入代码的关键字快速搜索代码。 ⑦ 说明 • 系统仅支持搜索已提交、开发中和已发布状态的代码 • 系统仅支持搜索版本升级后(2020年7月14日)提交过的代码。
6	展示了规范建模的快速入口。
0	展示了统一运维的流程。
8	展示了数据处理支持的编码任务,并提供了创建任务的入口。

实时引擎适配功能说明

功能特性	Blink	Ververica Flink	开源Flink
适配离线引擎	MaxCompute	MaxCompute	Hadoop(CDH5、 CDH6、CDP、 FusionInsight)TDH
自定义数据源	支持	待支持	待支持
元表	支持	支持	支持

功能特性	Blink	Ververica Flink	开源Flink
镜像表	支持	支持	支持
任务调试	支持	支持	不支持
测试表	支持	支持	支持
Flink SQL任务	支持	支持	支持
Flink SQL任务粗粒度配置	支持	支持	支持
Flink SQL任务细粒度配置	支持	支持	不支持
Flink SQL任务智能资源配 置	支持	支持	不支持
SQL建议(实时模式)	支持	不支持	不支持
DataStream任务	支持	支持	支持
DataStream任务粗粒度 配置	支持	支持	支持
DataStream任务细粒度 配置	支持	支持	不支持
DataStream任务智能资 源配置	支持	支持	不支持
实时代码模板	支持	待支持	待支持
资源	支持	支持	支持
自定义函数	支持	支持	支持
监控告警	支持	不支持,需在FlinkVVP平 台配置	支持
监控告警-业务延时过高	支持	不支持	不支持
监控告警-TPS超过范围	支持	不支持	支持
监控告警-失败率超过配置	支持	不支持	支持
监控告警-数据滞留超过配 置	支持	不支持	不支持
监控告警-Checkpoint失 败配置	支持	不支持	支持
监控告警-反压持续时间超 过配置	不支持	不支持	支持

数据开发·<mark>编码研发</mark>

智能数据构建与管理 Dataphin

功能特性	Blink	Ververica Flink	开源Flink
资产-元表血缘	支持	支持	支持
资产-镜像表血缘	支持	支持	支持
元表权限管理	支持	支持	支持
镜像表权限管理	支持	支持	支持
自定义实时函数权限管理	支持	支持	支持
元表数据质量	支持	支持	支持
跨租户发布	不支持	支持	待支持

实时数据源支持情况

数据源类型	数据源名称	Ververica Flink	Blink	开源Flink
大数据存储	MaxCompute	支持	支持	不支持
关系型数据库	PolarDB-X(原 DRDS)	支持	支持	不支持
关系型数据库	AnalyticDB for MySQL 3.0	支持	支持	不支持
消息队列	Log Service	支持	支持	不支持
NoSQL	Aliyun HBase	支持	支持	不支持
大数据存储	DataHub	支持	支持	不支持
消息队列	Kafka	支持	支持	支持
NoSQL	Tablestore	支持	支持	不支持
消息队列	RocketMQ	支持	支持	不支持
NoSQL	Cloud HBase	不支持	不支持	不支持
消息队列	Swift	不支持	不支持	不支持
NoSQL	Redis	支持	支持	不支持
消息队列	Random	不支持	不支持	不支持
NoSQL	TSDB	不支持	不支持	不支持
大数据存储	Hologres	支持	支持	不支持

智能数据构建与管理 Dat aphin

数据源类型	数据源名称	Ververica Flink	Blink	开源Flink
关系型数据库	HybridDB for MySQL	不支持	不支持	不支持
NoSQL	Lindorm	不支持	不支持	不支持
消息队列	MetaQ	不支持	不支持	不支持
NoSQL	Tair	不支持	不支持	不支持
关系型数据库	TDDL	不支持	不支持	不支持
消息队列	TimeTunnel	不支持	不支持	不支持
大数据存储	Elastic Search	支持	支持	支持
关系型数据库	AnalyticDB for MySQL 2.0	不支持	支持	不支持
实时自定义源	-	不支持	支持	不支持
关系型数据库	Postgres CDC	支持	不支持	支持
关系型数据库	MySQL CDC	支持	不支持	支持
关系型数据库	SQL Server CDC	不支持	不支持	不支持
关系型数据库	RDS for MySQL (DRDS)	支持	不支持	不支持
关系型数据库	DRC	不支持	不支持	不支持
关系型数据库	light.DRC	不支持	不支持	不支持
消息队列	AntQ	不支持	不支持	不支持
大数据存储	Explorer	不支持	不支持	不支持
NoSQL	AntHBase	不支持	不支持	不支持
NoSQL	ZDAL HBase	不支持	不支持	不支持
NoSQL	GBase	不支持	不支持	不支持
关系型数据库	MySQL PB Proxy	不支持	不支持	不支持
NoSQL	TBase	不支持	不支持	不支持
NoSQL	Apache HBase	不支持	不支持	支持
大数据存储	Apache Hive	不支持	不支持	不支持
大数据存储	Apache HDFS	不支持	不支持	不支持

Dataphin SQL语法解析器关键字

Dataphin中语法解析器的保留字和关键字与MaxCompute的保留字和关键词一致,详细内容请参见保留字与 关键字。

3.2. 新建资源

资源管理用于存储、管理代码研发过程中所需的文件,如JAR、JSON、Python等资源文件。本文为您介绍如 何新建资源。

背景信息

根据业务场景的需求,您可以将计算任务研发过程中所需的文件,上传至Dataphin平台,在代码研发过程中 引用该文件。

管理资源的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后,单击快速开始相关工作区域中的数据研发。
 - 选择工作区地域后,单击快速进入研发项目后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3的i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3的i和ii。

3. 进入资源管理页面。

- i. 在Dataphin首页, 单击研发。
- ii. 在数据开发页面,单击项目名称后的☑图标,单击Dev或Basic页签,选择数据开发的项目空间。 如果您当前访问的是Dev或Basic项目,且为数据开发的项目空间,则不需要再次选择项目空间。
- iii. 在数据开发页面,单击数据处理。 如果进入数据开发页面,系统默认进入数据处理页签,则无需再次单击数据处理。
- iv. 在左侧导航栏, 单击**⊡资源管理**图标。
- 4. 在资源管理页面,您可以通过以下方式,进入新建资源页面。
 - 单击资源管理后的 图标。
 - 单击项目名称后的 · 图标,选择数据处理 > 资源。
 - 在开发首页, 单击资源管理下的 图标。
- 5. 在新建资源对话框中, 配置参数。

新建资源			×
* 类型			
* 名称	请输入64字符内名称		
* 描述	请输入128字符内描述		
			0/128
上传文件	土 请点击选择文件上传		
* 计算类型	请选择计算类型		~
选择目录	资源管理		
		取消	提交

参数	描述
类型	选择上传文件的 类型 。系统默认支持上传file、jar、Python类型文件及其他 (others)。
名称	填写资源的 名称 。资源的 名称 在项目内不允许重复。 名称 由字母、数字、下划线 (_)或英文句号(.)组合组成。
描述	填写对资源简单的描述。
上传文件	根据已选的 类型 ,选择上 传文件 。
计算类型	选择资源的计算类型。
选择目录	选择已创建的文件夹作为资源的目录。

- 6. 单击提交,完成资源的提交。
- 7. 在提交备注对话框,填写备注信息。
- 8. 单击确定并提交。
- 9. 发布资源至生产环境。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的资源,详情请参见管理发布任务。
 - 如果您的开发模式是Basic,则提交成功的资源,即可进入生产环境。

3.3. 新建用户自定义函数

离线计算函数用于管理离线计算任务代码开发过程中用到的SQL函数,包括计算引擎源常用且系统默认支持的函数和用户自定义函数,前者不支持编辑操作。本文为您介绍如何新建用户自定义函数。

前提条件

已完成资源的创建。具体操作,请参见新建资源。

背景信息

> 文档版本: 20220708

Dataphin系统根据函数类型定义了函数的目录,帮助您更好地管理函数。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后, 单击**快速开始相关工作**区域中的数据研发。
 - 选择工作区地域后,单击**快速进入研发项目**后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过**快速开始相关工作**进入数据开发模块,则跳过步骤3.i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3.ii。

3. 进入函数管理页面。

- i. 在Dataphin首页, 单击顶部菜单栏的研发。
- ii. 在数据开发页面,单击项目名称后的
 图标,选择数据开发的项目空间(Dev或Basic项目)。
 如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。
- iii. 在数据开发页面,单击数据处理。
- iv. 在数据处理页签, 单击左侧导航栏的 2图标。
- 4. 在函数管理页面,进入新建函数对话框。

系统支持新建的函数类型,根据Dataphin系统的计算引擎不同而不同:

- 如果Dataphin系统的计算引擎为MaxCompute,则系统支持选择MAXC函数或FLINK函数。
- 如果Dataphin系统的计算引擎为ADB for PostgreSQL,则系统支持选择ADB函数或FLINK函数。
 您可以通过以下任何一种方式,进入新建函数对话框:
- 单击函数管理后的■图标,选择新建函数类型。
- 单击项目名称后的 · 图标,选择数据处理 > 函数后选择函数类型。
- 在数据开发首页,单击ADB函数、MAXC函数或FLINK函数下的 图标。
- 5. 在新建函数对话框, 配置参数。

新建函数		×
* 名称	请输入64字内的名称,支持英文,数字,下划线,首字母仅支持英文	ζ
选择资源	请选择资源	
* 类名	请输入类名,如test_udf.UDFGETSrcId	
* 类型		\sim
* 命令格式	请输入命令格式,如bigint weekday (datetime date)	
* 使用文档	请蝓入500字内描述,着重说明用途、参数介绍、使用用例 简单一段使用用例SQL,如:	4
选择目录	MAXC函数-用户定义函数	~
	取消提交	

参数	描述
名称	填写名称。名称由字母,数字,下划线(_)组合组成,且首字符仅限字母。
选择资源	选择资源文件。下拉框中提供的是匹配当前项目的资源名称。 ⑦ 说明 • 当前,仅支持选择PYTHON、JAR格式的文件进行定义函数。 • 选择多个资源时,仅限选择同一种资源类型。 • 如果您还没有资源,则需要创建资源。具体操作,请参见新建资源。
类名	填写UDF函数类名。不同类型资源的类名配置如下: • 当资源类型为JAR时,配置的类名格式为 Java包名称.实际类名 。 例如, com.dataphin.examples.udf 为Java包的名称,TestExample为实际类名,则类名参数配置 为 com.dataphin.examples.udf.TestExample 。 • 当资源类型为Python时,配置的类名格式为 Python资源名称.实际类名 。 例如, Dataphin_Test 为Python资源名称,UdfTest为实际类名,则类名参数配置为Dataphin_Test.UdfTest。

参数	描述					
类型	选择类型,下拉框中可供选择的类型包含窗口、统计、数值、字符串、时间、ip地 址相关函数、URL、编译码、业务、其他。					
命令格式	填写命令格式。命令格式即函数引用格式,例如, bigintweekday (datetime date) 。					
使用文档	填写函数使用描述,例如。 select get_week_date("20170810",0,2),获取8月10日,本周周二日期。 from cndata.dual					
选择目录	系统默认为当前函数类型的目录。如果您需要修改,则系统仅支持修改该函数类型 目录下的子目录。 例如,您新建的是MAXC函数,则系统自动为您选中MAXC函数为目录。如果您需要 修改目录,则系统仅支持选择MAXC函数目录下的子目录。					

6. 提交用户自定义函数。

- i. 完成参数配置后, 单击**提交**。
- ii. 在提交备注页面,填写备注信息。
- iii. 单击**确定并提交**,完成用户自定义函数的创建,即将自定义函数注册至MaxCompute或ADB for PostgreSQL。

⑦ 说明 如果自定义函数引用的资源有更新,则需重新提交自定义函数,这样注册至 MaxCompute或ADB for PostgreSQL的自定义函数才会更新。

您可以通过**即席查询**(具体操作,请参见查询并下载数据)编写SQL代码(SQL代码中引用该函数)验证 函数是否符合预期效果。SQL查询语句示例如下。

```
select
get_week_date("20170810",0,2),--获取8月10日本周周二日期。
from cndata.dual
```

- 7. (可选)发布用户自定义函数至生产环境。
 - 如果您的开发模式是Dev-Prod,则需要发布已提交的用户自定义函数。具体操作,请参见管理发布任务。
 - 如果您的开发模式是Basic,则提交成功的用户自定义函数,即可进入生产环境。

3.4. 新建MAXC任务

3.4.1. 新建SQL任务

本文为您介绍如何基于Dataphin构建SQL类型的离线计算任务。

背景信息

SQL离线计算任务适用于对已有数据进行加工处理,生成符合业务场景需求的数据的场景。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的数据研发。
 - 单击快速进入研发项目区域的Dev或Basic项目,选择开发数据的项目空间。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3.i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3.ii。
- 3. 进入计算任务页面。
 - i. 在Dataphin首页, 单击顶部菜单栏的研发。
 - ii. 在数据开发页面,单击项目名称后的
 図标后,单击Dev或Basic页签,选择数据开发的项目空间
 (Dev或Basic项目)。
 如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。
 - iii. (可选)在数据开发页面,单击数据处理页签。如果进入数据开发页面后,系统默认进入数据处理页签,则无需再次单击数据处理页签。
 - iv. 在数据处理页签, 单击左侧导航栏 一计算任务图标。
- 4. 在计算任务页面,进入新建SQL任务的新建文件对话框。
 - 单击计算任务后的 图标,选择MAXC任务 > MAX_COMPUTE_SQL。
 - 单击项目名称后的 图标,选择数据处理 > 计算任务 > MAX_COMPUTE_SQL。
 - 在开发首页,单击MAXC_SQL后的⊡图标。
- 5. 编写并运行代码。

i. 在**新建文件**对话框,配置参数。

新建文件					×
* 名称	SQL				
* 调度类型	● 周期性节点	() 手动节点			
描述	SQL				1,
选择目录	代码管理				~
			取消	确定	
参数		描述			
名称		填写离线计算任约	务的名称,例	如SQL。	
调度类型		选择任务的调度参 ■ 周期性节点, ■ 手动节点,需	类型。 调度类 自动参与系线 要手动触发信	₽型包括: 充的周期性调度 £务的运行。	o
描述		填写对任务的简单	单描述。		

选择离线计算任务的目录。

ii. 单击确定。

选择目录

iii. 在代码编写页面,编写离线计算任务的SQL代码。

	<u>881</u> -	
	代码编	写 调度配置 历史信息 去发布 >
	1	SELECT * from LD_Dualdws_s3 where ds='20190801'
		select * from LD_Dual dws_e where ds='20190801'
		select * from LD_Dual dim_e where ds='20190801'
		select * from LD_Dual dws_e where ds='20190801'
		已提交

		跨项目

D.		逻辑表
4		DML-Dev_Prod
		select * from LD_hm_0 t_hm_0203_01_aa2_di;
2		<pre>select * from \${LD_hr .fct_hm_0203_01_aa2_di;</pre>
		select * from LD_hm_0v.fct_hm_0203_01_aa2_di;
		DML-Basic
		select * from ld_hm_tlumo;
		select * from \${ld_hndim_lumo;
	19	物理表
	20	DDL
	21	insert into table test1yf_2 values('1');汪释3
		create table (id int,str string);
		insert into tablel values('1','2');
	24	DML-DevProd
	25	select answer
	26	from sers;

iv. 检查代码。

代码编写完成后,单击**预编译**,系统帮助您检查编写的SQL代码的语法。

如果预编译失败,您可以单击页面上方的**格式化**,系统自动帮助您调整语法格式。调整完格式,您 可以单击**刷新**,刷新页面上的代码。

- v. 单击页面右上方的执行,运行代码。
- 6. (可选)配置调度参数。
 - 如果离线计算任务的调度类型为**周期性节点**,则需要配置调度参数,详情请参见调度配置。
 - 如果离线计算任务的调度类型为手动节点,需要手动触发任务的调度。
- 7. 在代码编写页面,保存并提交SQL任务。
 - i. 单击页面右上方的 图标, 保存代码。
 - ii. 单击页面右上方的 图标, 提交代码。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交。
- 8. (可选)发布SQL任务。
 - 如果您的开发模式是Dev-Prod模式,则需要发布SQL任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功后的SQL任务,即可参与生产环境的调度。

3.4.2. 新建MapReduce任务

本文为您介绍如何基于Dataphin构建MapReduce类型的离线计算任务。

前提条件

完成JAR资源包的上传,详情请参见新建资源。

背景信息

MapReduce任务中引用到的JAR资源包需提前创建,因此需要您在资源管理中上传JAR资源包,然后在 MapReduce代码任务中引用。

操作步骤

- 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>>。

3.

- 4. 在计算任务页面,进入新建MapReduce任务文件对话框。
 - 单击计算任务后的■图标,选择MAXC任务 > MAX_COMPUTE_MR。
 - 单击项目名称后的 图标,选择数据处理 > 计算任务 > MAX_COMPUTE_MR。
 - 在开发首页, 单击MAXC_MR后的 ··图标。
- 5. 编写并运行代码。

i. 在新建文件对话框, 配置参数。

新建文件				×
. 52	115			
* 冶称	MK			
* 调度类型	● 周期性节点	○ 手动节点	Į	
描述	MR			
选择目录	代码管理			~
			取消	确定

参数	描述
名称	填写离线计算任务的名称,例如MR。
调度类型	选择任务的调度类型。调度类型包括: 周期性节点,自动参与系统的周期性调度。 手动节点,需要手动触发任务的运行。
描述	填写对任务的简单描述。
选择目录	选择离线计算任务的目录。

ii. 单击确定。

iii. 在**代码编写**页面,编写MapReduce离线计算任务的代码,详情请参见MapReduce。

🞆 MR		•				
代码编写	调度配置	历史信息				
1 @re 2 add 3 jar 4	source_refer jar mr_odps -resources	rence{"mr_d .jar as mo momo.jar	odps.jan"} omo.jar -f; -classpath	mr_odps.jar	hive.WordCountOdps	wc_in wc_out;

代码示例如下:

```
@resource_reference{"mr_odps.jar"}
add jar mr_odps.jar as momo.jar -f;
jar -resources momo.jar -classpath mr_odps.jar hive.WordCountOdps wc_in wc_out;
```

iv. 检查代码。

代码编写完成后,单击**预编译**,系统帮助您检查编写的SQL代码的语法。

如果预编译失败,您可以单击页面上方的**格式化**,系统自动帮助您调整语法格式。调整完格式,您 可以单击**刷新**,刷新页面上的代码。

v. 代码编写完成后,单击页面右上方的执行,运行代码。

- 6. (可选)配置调度参数。
 - 如果离线计算任务的调度类型为**周期性节点**,则需要配置调度参数,详情请参见调度配置。
 - 如果离线计算任务的调度类型为手动节点,需要手动触发任务的调度。
- 7. 在代码编写页面,保存并提交MapReduce任务。

i. 单击页面右上方的 图标, 保存代码。

ii. 单击页面右上方的 图标, 提交代码。

- iii. 在提交备注页面,填写备注信息。
- ⅳ. 单击确定并提交。
- 8. (可选)发布MapReduce任务。
 - 如果您的开发模式是Dev-Prod模式,则需要发布MapReduce任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功后的MapReduce任务,即可参与生产环境的调度。

3.4.3. 新建Spark任务

本文为您介绍如何基于Dataphin构建Spark类型的任务。

前提条件

完成上传JAR或PYTHON文件,详情请参见新建资源。

背景信息

Spark任务中引用到的JAR和PYTHON文件需要提前创建,因此您在资源管理中先上传JAR和PYTHON文件,然后在Spark任务中引用。

操作步骤

- 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>>。

3.

- 4. 在计算任务页面,进入新建文件对话框。
 - 单击计算任务后的 函标,选择MAXC任务 > SPARK_JAR_ON_MAX_COMPUTE。
 - 单击项目名称后的 图标,选择数据处理 > 计算任务 > SPARK_JAR_ON_MAX_COMPUTE。
 - 在开发首页,单击MAXC_SPARK_JAR后的 · 图标。

5. 编写并运行代码。

i. 在新建文件对话框, 配置参数。

新建文件				×
* 名称	Spark_python			
* 调度类型	● 周期性节点	○ 手动节点	Į	
描述	Spark			1
选择目录	代码管理			~
			取消	确定

参数	描述
名称	填写离线计算任务的名称,例如Spark_python。
调度类型	选择任务的调度类型。调度类型包括: 周期性节点,自动参与系统的周期性调度。 手动节点,需要手动触发任务的运行。
描述	填写对任务的简单描述。
选择目录	选择离线计算任务的目录。

ii. 单击确定。

iii. 在代码编写页面,编写Spark离线计算任务的代码。

代码示例如下。

```
@resource_reference{"spark.py"}
spark-submit
--deploy-mode cluster
--conf spark.hadoop.odps.task.major.version=cupid_v2
--conf spark.hadoop.odps.end.point=http://service.cn.maxcompute.aliyun.com/api
--conf spark.hadoop.odps.runtime.end.point=http://service.cn.maxcompute.aliyun-inc.
com/api
--master yarn
spark.py
```

其中, resource reference{} 为引用JAR或PYTHON文件资源包,其它语句均为固定语法。

iv. 检查代码。

代码编写完成后,单击预编译,系统帮助您检查编写的SQL代码的语法。

如果预编译失败,您可以单击页面上方的**格式化**,系统自动帮助您调整语法格式。调整完格式,您 可以单击**刷新**,刷新页面上的代码。

v. 单击页面右上方的执行,运行代码。

6. (可选)配置调度参数。

- 如果离线计算任务的调度类型为周期性节点,则需要配置调度参数,详情请参见调度配置。
- 如果离线计算任务的调度类型为手动节点,需要手动触发任务的调度。
- 7. 在代码编写页面,保存并提交Spark任务。
 - i. 单击页面右上方的

 回图标,保存代码。
 - ii. 单击页面右上方的 图标, 提交代码。
 - iii. 在提交备注页面,填写备注信息。
 - iv. 单击确定并提交。
- 8. (可选)发布Spark任务。
 - 如果您的开发模式是Dev-Prod模式,则需要发布Spark任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功后的Spark任务,即可参与生产环境的调度。

3.5. 开发流批一体任务

3.5.1. 新建镜像表

镜像表帮助您统一管理Flink计算任务中的实时表和离线表。本文为您介绍如何生成镜像表。

背景信息

如果Flink任务的生效模式为实时和离线,则新建计算任务开发即有实时表又有离线表。Dataphin支持新建镜 像表,镜像表映射了实时表和离线表。您在开发计算任务时只需要引用镜像表,系统会为您自动生成实时实 例和离线实例,提高了计算任务的研发效率。

步骤一:新建镜像表

- 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>>。

您可以通过以下方式,快速进入工作区:

- 选择工作区地域后, 单击快速开始相关工作区域中的数据研发。
- 选择工作区地域后,单击快速进入研发项目后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3.i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3.ii。

3. 进入表管理页面。

- i. 在Dataphin首页, 单击顶部菜单栏的研发。
- ii. 在数据开发页面,单击项目名称后的
 图标,选择数据开发的项目空间(Dev或Basic项目)。
 如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。
- iii. 在数据开发页面,单击数据处理。

iv. 在数据处理页签, 单击左侧导航栏 表管理图标。

- 4. 在表管理页面,您可以通过以下方式,进入新建镜像表对话框。
 - 单击表管理后的 图标后,选择镜像表。
 - 单击项目名称后的 图标,选择数据处理>表>镜像表。
- 5. 在新建镜像表对话框, 配置镜像表名称、选择目录和描述。
- 6. 单击确定。
- 7. 在主表开发页面,添加字段。

系统支持SQL导入字段、批量导入字段和单行导入字段的方式,新建主表字段:

- 。 SQL语句导入字段。
 - a. 单击SQL导入字段。
 - b. 在SQL导入对话框,单击参考示例,查看代码示例,编写导入字段的代码,代码示例如下。

```
create table dwi_pub_hbd_cate_mtr (
  rowkey VARCHAR comment 'rowkey',
  stat_date VARCHAR comment 'stat_date',
  keymin VARCHAR comment 'keymin',
  PRIMARY KEY(rowkey)
)
```

- c. 单击确定。
- 批量导入字段。
 - a. 单击批量导入。
 - b. 在**批量导入**对话框, 查看代码示例, 编写批量导入字段的代码, 代码示例如下。

```
rowkey||VARCHAR||rowkey
rowkey1||VARCHAR||rowkey1
rowkey2||VARCHAR||rowkey2
stat_date||VARCHAR||stat_date
keymin||VARCHAR||keymin
```

- c. 单击确定。
- 单行添加字段。
 - a. 单击单行添加。
 - b. 在单行添加对话框,填写字段名和描述,选择字段类型。
 - c. 单击确定。

如果您已添加了字段,还需继续添加字段,则单击主表显示框右上方的**回**图标后,选择添加字段的方式。

步骤二:关联实时表和离线表

1. 在主表开发页面, 单击选择关联表。



2. 在新建关联关系对话框,填写关联的实时表和离线表。

新建关联关系	系							×
● 关联实时表	Ē			• 关联离线表				
* 选择项目	PJ_			*选择项目	PJ_	. N		
* 选择实时表	adb_a	1 物理表	~	* 选择离线表	c	ch DATAHU	в	\sim
字段名	字段类型	字段描述		字段名	字段类型	字段描述		
stat_tz	STRING	统计日期时区	1999 - 1999 -	id	STRING	ххх		1
stat_time	STRING	统计时间		name	STRING	XXX		
stat_date	STRING	统计日期		action	STRING	X0000X		ି
stat_hour	STRING	统计小时		result	STRING	2015(NXXX ⁸²¹⁴⁾		
						取消	确定	

关联表	参数	描述	
关联实时表	选择项目	选择关联实时表的项目。	
	选择实时表	选择实时表。	
关联离线表	选择项目	选择关联离线表的项目。	
	选择离线表	选择离线表。	

- 3. 单击**确定**。
- 4. 关联字段。

您可以通过以下方式,关联字段:

○ 手动关联

选中实时表或离线表中需要关联的字段后,单击主表中需要关联的字段。关联后的字段使用连线连接。

实时表				主表		
		Q ()		🚾 tata11112		Q 🖻
01 stat_tz	STRING	统计日期时区		01 rowkey	VARCHAR	rowkey
02 stat_time	STRING	统计时间		02 stat_date	VARCHAR	stat_date
03 stat_date	STRING	统计日期	7	03 keymin	VARCHAR	keymin
04 stat_hour	STRING	统计小时	Ľ	04 vagaga1	VARCHAR	-
05 stat_min	STRING	统计分钟		100000000000000000000000000000000000000	2389999889999	28816819 m

您也可以单击连线上的口后,在删除字段关联对话框,单击确定,删除已关联的字段。

。 自动关联

单击页面上方的**自动关联**,系统会自动关联镜像表与实时表和离线表同名的字段,并生成连线。 如果已有关联字段,则自动关联会导致原有连线会被覆盖。

• 批量关联

a. 单击页面上方的批量关联。

b. 在**批量关联**对话框,单击参考示例。参考示例编写关联字段的代码。

镜像表字段,实时表字段,离线表字段 镜像表字段,实时表字段,离线表字段 镜像表字段,实时表字段,离线表字段

c. 单击确定。

步骤三:预览镜像关系

在镜像表开发页面,页面右侧为您展示**镜像关系预览**面板。在**镜像关系预览**面板,查看实时表、离线表和 镜像表之间的映射关系。

如果您已关闭了镜像关系预览面板,则单击页面右上方的。图标,打开镜像关系预览面板。

步骤四:提交或发布镜像表

- 1. 在镜像表开发页面,单击页面上方的圆图标,保存镜像表。
- 2. 单击页面上方的 刻图标。
- 3. 在提交备注对话框,填写备注信息。
- 4. 单击确定并提交。
- 5. (可选)发布镜像表至生产环境。

根据当前项目的开发模式,选择是否需要镜像表:

- 如果项目为Basic模式,则无需发布镜像表,提交成功的镜像表即可进入生产环境。
- 如果项目为Dev-Prod模式,则需要发布镜像表至生产环境。

3.5.2. 新建元表

元表帮助您统一配置与管理计算任务开发过程中用到的输入表、输出表、维表。本文为您介绍如何新建元 表。

背景信息

- 元表只有在解锁的状态下,才支持编辑、移动、克隆和删除等操作。
 - 如果元表已被您自己锁定,则无需解锁元表。
 - 如果元表已被他人锁定,则执行以下操作,解锁元表:
 - a. 在元表页面, 单击需要编辑、移动、克隆或删除等操作的元表。
 - b. 在元表的详情页面,单击右上方的 @ 图标。

当 图标变更为 图标时,该实时元表解锁成功。

• 管理元表的权限限制,详情参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后,单击快速开始相关工作区域中的数据研发。
 - 选择工作区地域后,单击快速进入研发项目后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3.i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3.ii。
- 3. 进入表管理页面。
 - i. 在Dataphin首页, 单击顶部菜单栏的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标,选择数据开发的项目空间(Dev或Basic项目)。
 如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。
 - iii. 在数据开发页面,单击数据处理。
 - iv. 在数据处理页签, 单击左侧导航栏 表管理图标。
- 4. 在表管理页面,您可以通过以下方式,进入新建元表对话框。
 - 单击表管理后的■图标后,选择元表。
 - 单击项目名称后的 图标,选择数据处理>表>元表。
 - 在数据开发首页,单击元表下的→图标。
- 5. 在新建元表对话框, 配置元表名称、数据源、选择目录和描述。

不同数据源的配置参数不同。请根据对话框提示完成参数的配置。

- 6. 单击确定。
- 7. (可选)添加字段。
 - 通过SQL导入的方式添加字段:
 - a. 在实时元表页面,鼠标悬停至添加字段上,选择SQL导入。
 - b. 在SQL导入对话框中,编写SQL代码,代码示例如下。

```
create table dwi pub hbd cate mtr (
   rowkey VARCHAR comment 'rowkey',
   stat date VARCHAR comment 'stat date',
   keymin VARCHAR comment 'keymin',
   PRIMARY KEY(rowkey)
 )
  with (
     type='alihbase',
     diamondGroup='null',
     zkQuorum='hbasemaster74000.sg94.tbsite.net,hbasemaster74001.sg94.tbsite.net
, hbasemaster74002.sg94.tbsite.net, hbase74000.sg94.tbsite.net, hbase74001.sg94.tbsi
te.net',
     diamondKey='null',
     zkNodeParent='/group-sg94-lzd-mix',
      columnFamily='info',
      tableName='dwi pub hbd cate mtr 001',
      stringWriteMod='true'
 );
```

c. 单击确定, 完成字段的添加。

- 通过批量导入的方式添加字段:
 - a. 在实时元表页面,鼠标悬停至**添加字段**上,选择**批量导入**。
 - b. 批量导入对话框中,编写代码,代码示例如下。

```
ID, INT, 描述, false, true, name, INT, 描述, false, true
```

- c. 单击确定。
- 通过单行添加的方式添加字段:
 - a. 在实时元表页面,鼠标悬停至**添加字段**上,选择单行添加。
b. 在单行添加对话框, 配置参数。

单行添加				×
* 字段名	请输入字段名			
* 字段类型	请输入字段类型			~
* 是否主键) 是 💿 否			
* 是否Header①	○ 是 🧿 否			
描述	请输入描述			
				0/128
			取消	确定

c. 单击**确定**,完成字段的添加。

同时,您可以对已添加的字段,执行如下操作。

操作	描述
编辑	执行如下操作,编辑字段: i. 在元表页面,单击操作列下的 图标。 ii. 根据编辑页面的提示,修改参数。 iii. 完成参数修改后,单击 完成 。
删除	执行如下操作,删除字段: i. 在元表页面,单击操作列下的画图标。 ii. 在提示对话框中,单击确定并提交,完成字段的删除。
排序	执行如下操作,重新排序字段: i. 在元表页面,单击 排序 。 ii. 根据页面提示,单击 完成 ,完成字段的排序。
搜索	在搜索框中输入字段的名称或字段名称关键字进行搜索字段。
刷新	单击 • 图标, 重新解析元表中的字段。

8. 保存并提交元表。

i. 在元表的开发页面, 单击右上角的画图标, 保存元表。

ii. 在元表的开发页面,单击右上角的

iii. 在提交备注对话框,填写备注信息。

iv. 单击确定并提交。

- 9. (可选)发布元表至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布元表,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的元表,即可进入生产环境。

对已添加的元表,您可以执行如下操作。

操作	描述
查看元表版本信息	 a. 在表管理页面,单击需要查看的元表。 b. 在元表详情页面,单击版本信息,可以查看历史版本提交情况: 单击操作列下的 图标,查看版本对比。 同时,单击回滚到对比版本,回滚到历史版本。 单击操作列下的 图标,查看代码。
查看元表信息	a. 在 表管理 页面,单击需要查看的元表。 b. 在元表详情页面,单击 元表信息 ,可以查看元表的基本信息和引用信息。
修改元表	 您可以通过修改元表的方式,修改元表的基本信息和字段: a. 在表管理页面,将鼠标悬停在需要修改的元表后的 图标,单击编辑。 b. 修改元表的基本信息: a. 在实时元表详情页面,单击页面上方的元表信息。 b. 在元表信息对话框,修改参数。 如果元表是基于系统提供的实时数据源类型构建的,则您可以修改数据源、来源topic、描述及参数KV配置。 如果元表是基于自定义实时数据源构建的,则您可以修改数据源、描述及参数KV配置。 c. 单击确定。 c. 单击确定。 c. 在实时元表详情页面,您可以添加、修改或删除元表的字段。 d. 提交或发布元表。
移动元表	a. 在 表管理 页面,将鼠标悬停在需要移动的元表后的 <mark>。</mark> 图标,单击 移动 。 b. 在移动文件对话框,选择目录。 c. 单击 确定 ,完成元表的移动。
克隆元表	a. 在 表管理 页面,将鼠标悬停在需要克隆的元表后的 <mark>。</mark> 图标,单击 克隆 。 b. 在 克隆元表 对话框,配置参数。 c. 单击 确定 。

操作	描述
	⑦ 说明 系统仅支持删除没有被计算任务引用的元表。
删除元表	a. 在 表管理 页面,将鼠标悬停在需删除的元表后的 <mark>,</mark> 图标,单击 删除 。
	b.在 提示 对话框,填写备注信息。 c.单击 确定并提交 ,完成元表的删除。

3.5.3. 新建自定义源

您可以根据业务场景的需求,自定义实时数据源。本文为您介绍如何新建自定义源。

前提条件

上传JAR文件,详情请参见新建资源。

背景信息

自定义源即自定义实时数据源。系统支持通过JAR文件,自定义实时数据源。当实时数据源提交后,即可基于该实时数据源新建实时计算元表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后,单击快速开始相关工作区域中的数据研发。
 - 选择工作区地域后,单击**快速进入研发项目**后的Dev或Basic页签,选择相应的项目。

⑦ 说明

- 如果您通过**快速开始相关工作**进入数据开发模块,则跳过步骤3。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3和4。
- 3. 进入自定义源页面。
 - i. 在Dataphin首页,单击顶部菜单栏的研发。
 - ii. 在数据开发页面,单击项目名称后的
 図标后,单击Dev或Basic页签,选择数据开发的项目空间
 (Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则无需选择项目空间。

- iii. 在数据开发页面,单击数据处理页签。如果在数据开发页面,默认进入数据处理页签,则无需再次单击数据处理页签。
- iv. 在数据处理页签, 单击左侧导航栏的图自定义源图标。
- 4. 在自定义源页面,进入新建实时自定义源对话框。

- 单击自定义源后的 图标。
- 单击项目名称后的 图标,选择数据处理 > 自定义源。
- 在数据开发首页, 单击自定义源下的□图标。
- 5. 在新建自定义源对话框中, 配置参数。

新建实时自动	E义源 ×
* 名称	请输入实时自定义源名称,新建后名称不可更改
∗选择JAR包	请选择资源
* ClassType	请选择ClassType V
* ClassPath	请输入classPath
* 选择目录	实时自定义源 🗸
描述	请输入描述, 128字符以内 0(128
	0/20 //
	取消 提交

参数	描述
名称	自定义数据源的名称。系统仅支持英文字符、数字、下划线(_)组合组成。命名的限制如下: • 系统不支持名称的首字母是数字。 • 系统不支持名称与已提交的自定义源名称重复。
选择JAR包	自定义数据源的JAR包。系统仅支持选择已提交的JAR包。
ClassType	JAR包的ClassType。系统支持选择class和tableFactoryClass。
ClassPath	ClassPath的路径。
选择目录	选择自定义数据源的目录。
描述	实时数据源的描述。

6. 单击**提交**,提交新建的自定义源。

7. 根据提交备注对话框提示,填写备注信息。

8. 单击确定并提交,完成自定义源的提交。

- 9. (可选)发布自定义源至生产环境。
 - 如果您的开发模式是Dev-Prod,则需要发布自定义源,详情请参见管理发布任务。系统仅支持删除没 有被已提交的实时计算元表引用的自定义源。
 - 如果您的开发模式是Basic,则提交成功的自定义源,即可进入生产环境。
 对已添加的自定义源,您可以执行如下操作。

操作	描述
查看自定义源详情	 Dataphin支持查看自定义源的基本信息、引用信息和版本信息: a. 在自定义源页面,单击左侧功能区域中需要查看信息的自定义源。 b. 在自定义源属性框中,查看当前版本的基本信息、引用信息和版本信息: 版本信息区域,执行以下操作: 鼠标悬停至 ◎ 图标后,查看对应版本的详细信息。 单击某个版本操作列下的 ◎ 图标,回滚到该版本。在提示对话框,单击确定,自定义实时数据源的版本就可以回滚到该版本。 引用信息区域,单击引用的实时计算元表,查看实时计算元表详情。 ② 说明 当基于自定义源构建的实时计算元表提交后,在该自定义源的属性框中才可以看到引用信息。
修改自定义源	 a. 在自定义源页面,将鼠标悬停在需要修改的自定义源后的 图标,单击编辑。 b. 在编辑自定义源对话框,修改参数。 ⑦ 说明 系统不支持编辑自定义源的名称。 c. 提交或发布自定义源。
移动自定义源	 a. 在自定义源页面,将鼠标悬停在需要移动的自定义源后的图图标,单击移动。 b. 在移动文件对话框,选择目录。 c. 单击确定,完成自定义源的移动。
删除自定义源	 ⑦ 说明 系统仅支持删除没有被已提交的实时计算元表引用的自定义源。 a. 在自定义源页面,将鼠标悬停在需删除的自定义源后的 图标,单击删除。 b. 在提示对话框,填写备注信息。 c. 单击确定并提交,完成自定义源的删除。

3.5.4. 新建代码模板

代码模板为数据研发过程中实时计算任务的开发提供模板,提高研发效率。本文为您介绍如何新建代码模板。

背景信息

- 新建Flink_Template_SQL类型的实时计算任务中引用代码模板,提高代码的开发效率。
- 代码模板只有在解锁的状态下,才支持编辑、移动、克隆和删除等操作。
 - 如果代码模板已被您自己锁定,则无需解锁代码模板。
 - 如果代码模板已被他人锁定,则执行以下操作,解锁代码模板:
 - a. 在代码模板页面, 单击需要编辑、移动、克隆或删除等操作的代码模板。
 - b. 如果代码模板已被他人锁定,则单击代码模板代码页面右上方的圆图标。

当68 图标变更为 88 图标时,代码模板解锁成功。

• 管理代码模板的权限限制, 详情参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入工作区:
 - 选择工作区地域后,单击快速开始相关工作区域中的数据研发。
 - 选择工作区地域后,单击快速进入研发项目后的Dev或Basic页签,选择相应的项目。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3.i。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3.ii。
- 3. 进入代码模板页面。
 - i. 在Dataphin首页, 单击顶部菜单栏的研发。
 - ii. 在数据开发页面,单击项目名称后的☑图标,选择数据开发的项目空间(Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

- iii. 在数据开发页面,单击数据处理。
- iv. 在左侧导航栏, 单击**同代码模板**图标。
- 4. 在代码模板页面,您可以通过以下任何一种方式,进入新建代码模板页面。
 - 单击代码模板后的■图标。
 - 单击项目名称后的 图标,选择数据处理 > 代码模板。
 - 在开发首页, 单击代码模板下的 ⊡图标。

5. 在新建代码模板对话框,填写模板名称、选择目录和描述。

新建实时计算	算模板	×
* 模版名称	请输入模板名称, 新建后不可修改	
* 选择目录	test	~
描述	请輸入描述,128字符以内	
		0/128
* `_` Tip: 新建相	模板输入代码后需要手动解析配置任务参数	
	取消	定

- 6. 单击**确定**。
- 7. 代码模板的代码开发。
 - i. 在模板代码开发页面,编写代码。
 - ii. 单击页面上方的模板信息, 配置代码中的参数值。
 - iii. 根据页面提示,填写代码模板中用到参数的参数说明和默认值。
 - iv. 填写完成后, 单击确定, 保存模板信息。

您可以在新建Flink_Template_SQL类型的实时计算任务中引用该模板,同时也可以在实时计算任务中对模板中的参数进行修改。

8. 保存并提交代码模板。

- i. 完成代码开发后, 单击页面右上方的 图标, 保存代码模板。
- ii. 单击页面右上方的 图标, 提交代码模板。
- iii. 在提交备注对话框,填写备注信息。
- iv. 单击确定并提交。
- 9. (可选)发布代码模板至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布代码模板,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交代码模板后,即可进入生产环境。
 对已添加的代码模板,您可以执行如下操作。

操作

描述

操作	描述
查看版本信息	 a. 在代码模板页面,单击需要查看的代码模板。 b. 在代码模板的代码页面,单击页面上方的版本信息: 单击某个版本操作列下的 图标,查看版本对比。 在版本对比对话框,您可以单击资源配置对比和模板参数对比页签,查 看版本的对比信息。同时您还可以执行以下操作,回滚版本: a. 在代码对比和模板参数对比页面,单击页面下方的回滚到对比版 本。 b. 在提示对话框,单击确定,回滚至该版本。 单击某个版本操作列下的 图标,查看代码的基本信息。
查看模板信息	 a. 在代码模板页面,单击需要查看模板的代码模板。 b. 在代码模板的代码页面,单击页面上方的模板信息。在模板信息页面,查看模板的基本信息和参数配置。 在基本信息区域,修改描述。 在参数配置区域,修改参数的参数说明和默认值。 c. 单击确定,完成代码模板的查看。

操作	描述
查看引用信息	只有代码模板被引用后,您才可以查看该代码模板的引用信息。 a. 在代码模板页面,单击左侧功能区域中需要查看引用信息的代码模板。 b. 在代码模板的代码页面,单击页面上方的引用信息。在引用信息页面,查看 代码模板被引用的信息。 同时,您还可以批量启用最新版本和查看不同版本的代码: • 批量启用最新版本: a. 选中多个最新版本后,单击 批量启用最新版本 前的函图标。 b. 在提示对话框中,单击启用。 • 查看代码: a. 单击引用此模板任务后的画图标,查看代码。 b. 单击查看代码页面上方的⊠图标,关闭查看代码页面。
修改代码模板	a. 在 代码模板 页面,鼠标悬停至需要修改的代码模板 <mark>。</mark> 图标,单击 编辑 。 b. 修改代码模板的代码和模板信息。 c. 提交或发布代码模板。
移动代码模板	 a. 在代码模板页面,将鼠标悬停在需要移动的代码模板后的图图标,单击移动。 b. 在移动文件对话框,选择目录。 c. 单击确定,完成代码模板的移动。
克隆代码模板	 a. 在代码模板页面,将鼠标悬停在需要移动的代码模板后的图图标,单击克隆。 b. 在克隆代码模板对话框,配置参数。 c. 单击确定。

操作	描述
删除代码模板	 ⑦ 说明 系统仅支持删除没有被实时计算任务引用的代码模板。 a. 在代码模板页面,将鼠标悬停在需删除的代码模板后的图图标,单击删除。
	b.在 提示 对话框,填写备注信息。 c.单击 确定并提交 ,完成代码模板的删除。

3.5.5. 创建Flink SQL任务

Flink SQL是为了简化计算模型、降低您使用Flink门槛而设计的一套符合标准SQL语义的开发语言。本文为您介绍如何基于Dat aphin研发Flink SQL类型的计算任务。

背景信息

项目绑定了Flink计算源,即可新建Flink SQL任务。Flink SQL任务支持处理离线与实时计算数据。

新建的Flink SQL任务默认开启实时模式,即只处理实时计算任务。您可以手动开启离线模式,处理离线计算任务。同时,Flink SQL任务支持处理离线和实时任务。关于Flink SQL更多信息,请参见Flink SQL参考。

注意事项

当前, Flink任务的计算引擎仅支持Blink。

步骤一:新建Flink SQL任务

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏研发。
- 4. 按照下图指引,进入新建Flink_SQL对话框。



5. 在新建Flink_SQL对话框,配置参数后,单击确定。

新建 Flink_\$	SQL	×
* 夕称	datanhin test	
. EU.		
* 远择日录	实时计异仕务	~
* 资源队列	root.blink_just	_fortest1
* 引擎版本	blink-3.2.2	~
		取消 确定
参数		说明
名称		名称的命名规则如下: • 只能包含小写英文字母、数字、下划线(_)。 • 名称的长度范围为3~62个字符。 • 项目内的名称不支持重复。 • 名称仅支持以英文字母开头。
选择目录		计算任务所属的目录。
资源队列		选择该项目所绑定实时计算源中的资源队列。
引擎版本		选择当前资源队列所支持的版本。

步骤二:开发Flink SQL任务的代码

- 在Flink SQL任务代码页面,编写任务的代码。
 您可以单击页面右上方的格式化,系统自动调整SQL代码格式。
- 2. 单击页面右上方的预编译,校验代码任务的语法及权限问题。
- 3. 单击页面右上方的调试,代码任务采样数据并进行本地调试,保障代码任务的正确性。

i. 在调试配置对话框的选择采样模式页签中,选择调试的模式后,单击下一步。

调试配置							×
		1 选择采样模	武	2 采柏	间试数据		
		当前仅支持单种模式	、调试,选择模式质	后请采样对应模式表数	救据进行调试		
SN SUP ^{ar B}					R		
S ^{al} cuper ^a		空时横式调试		ja ja	容线横式调试		
	采样对应的实时 在FLINK S	物理表数据,完成数 tream模式下进行本均	据采样后将 也调试	采样对应的离线物 在FLINK Ba	理表数据,完成数据系 tch模式下进行本地调	彩样后将 试	
Suber Al	iminter Super	daninter SuperAc	ininian Super				
					取消	下 ⁻	步

ii. 在**采样调试数据**页签中,为元数据进行采样调试。

您可以通过自动抽样和上传数据的方式,为元数据进行采样调试。适用场景说明如下:

 自动抽样:自动抽样到的数据是随机的,所以适用于对采集到的数据没有限制的场景。选择自动 抽样后,需要配置抽样条数。

↓ 注意 如果元数据表中没有数据,则自动抽样将采集不到数据。

上传本地数据:自动抽样时元数据采集不到或数据抽样的逻辑比较严格,例如从100万条数据中抽取其中1条数据,这样采集效率就很低,可以选择手动上传本地数据。

上传本地数据前需要先下载样例,根据下载的样例编辑需要上传的数据,单击**上传**后,数据自动 填充至**元数据采样**区域。

⊘ 选择采样模式			2 采样调试数据	2
数据表列表 所有数据表都需要已采样,才可进行调试 ————————————————————————————————————	元数据采样	<u>.</u>	样例下载 > 土	とた 住 で し は の 抽样
源表 v2_8_0_stream_dev.moyi_data X 未采样	54 	id	name	action

■ 手动输入:适用于采集的数据比较少,或者需要修改已采集到的数据的场景。

iii. 完成所有数据表的元数据采样后,单击页面下方的确定。

iv. 在Result页面, 查看调试数据、中间结果和调试结果。

调试	数据 中间	间结果	调试结果 													
<	tmp_multiple	e_days	tmp_multi	ple_days												
	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A
1	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
2	null	null	null	null	null	nuli	null	null	null	null	null	null	null	null	null	null
3	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
4	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
5	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null

步骤三: 配置Flink SQL任务

1. 在任务配置页面,确认在实时模式页签。

在实时模式页签,为您展示任务运行所使用的资源队列、引擎版本、资源配置类型及自动调优。

• 资源配置		
* 资源队列	root.dt-streamcompu	
	切换后提交生效的任务执行可能有差异	
* 引擎版本	blink-3.4.4 V	
* 资源配置类型	○ 系统推荐配置	
antistan" ar	 自定义配置 	
资源配置记录	暂未保存资源记录	🎴 保存当前配置为新记录
* 自动调优		
STREETS SLOWCH	CARDON CONTRACTOR CONTRACTOR	
	最大CU数 1 / 总CU: 2147483647	
	期望最大内存 1 MB	

资源队列和引擎版本默认为创建Flink SQL任务时选择的配置。同时,您也可以修改资源队列和引擎版本。

i. 选中自定义配置为资源配置类型,并单击去配置。

ii. 在资源配置页面,为您展示一张拓扑图,图中每个方框代表了一个计算任务,都可以进行独立配置。每个Group代表着Group内部的节点可以存放在一台机器进行计算,可以有效避免数据的网络传播,提升性能。图中当前的资源配置就是默认为您展示系统推荐的资源配置。



iii. 单击需要配置资源的Group右上角的①后,在自定义配置Group执行参数对话框配置参数后,单 击确定。

		自定义配置 Group 执	行参数		2 ×
		* core :	0.25		
GROUP-1	GRC	* heap_memory :	1260		
TT4TableSource-speed_boat_dev_dwd_a e_trd_pay_mix-Stream->SourceConversio n(table:[builtin, default,_DataStreamTable	Gro ct: (* parallel :	4		
_0, source. [1141ableSource-speed_boat _dev_dwd_ae_trd_pay_mix]], fields:(f0))>> MiniBatchAssigner(intervalMs: 5000)>>c	HASH er_i buy	* direct_memory :	0		
CORE 0.25 HEAP MEMORY 1260.00 PARALLEL 4	CC HE PA	* native_memory :	20		
				取消	确定
参数	描述				
core	通常,core配置成 值不能超过1,0.2	ർ0.25。即一个CPU可以 25即可,0.125是最低	以支持4个线 下限了,同B	程同时运算,因 时支持8个线程。	国此core的取 ,

参数	描述
heap_memory	单位为MB, heap_memory是堆内存,供Java应用程序使用的内存。 heap_memory及其内部各组成的大小可以通过JVM的一系列命令行参数来控 制,在一般的blink程序中,都会需要一定的heap_memory开销,,例如申请一 定的heap_memory作为程序的缓存等,因此您可以按程序的规模来设置其大 小。
parallel	表示同时并发的线程数量,用户可以选择合适的数量来运行自己的任务,并不是 越大越好,越大代表你资源申请的越多,反而对性能有抑制。 通常,一个简单计算节点每秒可以处理2000~4000条之间的数据。 ① 注意 如果源头是tt,tt的queue大小决定了parallel的上限,不能超过这个数字,否则程序将报错。
direct_memory	单位为MB, direct_memory并不是虚拟机运行时数据区的一部分,也不是Java 虚拟机规范中定义的内存区域。但是这部分内存也被频繁的使用,而且也可能导 致OutOfMemoryError异常出现。如果您的程序有使用igraph或者swift,可以 适当配置其大小,如16-32MB。 在Java NIO (New Input/Output)类,引入了一种基于通道(Channel)与缓 冲区(Buffer)的I/O方式,direct_memory可以使用Native函数库直接分配堆 外内存,然后同一个存储在Java堆中的DirectByteBuffer对象作为这块内存的引 用进行操作。这样能在一场场景中显著提高性能,因为避免了在Java堆和Native 堆中来回复制数据。
native_memory	单位为MB, native_memory没有相应的参数来控制大小,其大小依赖于操作系 统进程的最大值(对于32位系统就是3~4G,各种系统的实现并不一样),以及 生成的Java字节码大小、创建的线程数量、维持Java对象的状态信息大小(用于 GC)以及一些第三方的包。 native memory存放下面4种信息: 管理Java heap的状态数据(用于GC)。 JNI调用,也就是Native Stack。 JIT(即使编译器)编译时使用native memory,并且JIT的输入(Java字节 码)和输出(可执行代码)也都是保存在native memory。 NIO direct buffer。

iv. 按照下图指引,进入自定义配置Operator执行参数对话框并配置参数。完成参数配置后单击确 定。

GROUP-11	UP-11	自定义配置Operator-11执行参数 ×
GlobalGroupAggregate(groupBy: (stat_mi n), select:(stat_min, COUNT_RETRACT(d istinct\$0 count\$0) AS pay, and by: cnt m	OPERATOR-11 2	
tr_037, COUNT_RETRACT(distinct\$1 cou nt\$1) AS pay mord ->Calc(select ((CAS	GlobalGroupAggregate(groupBy: (stat min), select:(stat_min, COUNT_RETRA	* core: 0.25
TCAST('b2e7:md5\\u0004AE:biz\\u0004	CT(distinct\$0 count\$0) AS pay_ord_byr cnt mtr 037. COUNT RETRACT(di	and the second sec
CORE 0.25	CORE	* heap_memory: 102
HEAP MEMORY 510.00 PARALLEL 4	HEAP MEMORY 102.00 NATIVE MEMORY 614.00 PARALLEL 4	* parallel : 4
		* direct_memory : 0
		* native_memory: 614
		* state_size : 0
		* chain_strategy: ALWAYS NEVER HEAD
		取消 确定

自定义配置Operator执行参数对话框中core、heap_memory、parallel、direct_memory和 native_memory参数解释请参见上一步骤,下表仅对state_size、chain_strategy参数进行解释。

描述
保持默认0即可。
 chain_strategy用于定义多个节点的链接策略,Dataphin支持的链接策略包括: Always:默认参数,即和其他节点均部署在一台机器上,没有特殊要求。 Never:节点不会和其他节点放在一台机器上,即需要独立部署。 Head:可以接受和其他节点放在一台机器上,但是只能作为Group的头节点。
↓ 注意 目前, Head和Never极少出现, 默认Always即可。

v. 配置完成后, 单击页面右上方的保存。

↓ 注意 如果您想继续使用系统推荐的资源配置,则单击页面右上方的重置为系统初始。

2. (可选)保存资源配置记录。

↓ 注意 仅实时模式支持保存资源配置记录。

如果资源配置类型为自定义配置,则单击资源信息记录后的保存当前配置为新纪录。在保存资源记录对话框中,输入资源记录名称后,单击确定。

对已有的资源信息记录,您可以执行以下操作。

操作	描述
查看版本信息	单击某个记录 操作 列下的 图标,查看版本信息。

操作	描述
启用记录	i. 单击某个记录的操作列下的 问图标。 ii. 在提示对话框中,单击确定。
删除记录	i. 单击某个记录操作列下的 _面 图标。 ii. 在提示对话框中,单击确定

3. (可选)打开自动调优开关, 配置最大CU数和期望最大内存, 开启自动调优。

↓ 注意 仅实时模式支持自动调优功能。

4. (可选)配置实时任务的时间参数。

实时任务的时间参数为stat_date,用于方式实时计算任务的运行时间的偏移,例如,您需要计算当天 某个指标聚合值,为了防止时间偏移,则您需要设置state_date大于当天零点,过滤掉偏移的时间点。

为了规避在**任务参数**处经常漏掉配置stat_date,您只需要在实时任务配置的属性配置中新增 stat_date的kv配置,其中Value是一个基于业务时间的表达式,同时您也可以配置多个时间参数,使用 半角分号(;)分割。例如,stat_date=\${yyyyMMdd-1},则任务运行过程中的开始执行时间为 \${yyyyMMdd-1}。

• 属性配置	
时间参数	stat_date=\${yyyyMMdd-1}

5. 配置Flink SQL任务的依赖关系。

在任务配置面板的依赖关系区域,配置依赖关系。

赖关系 自动	的解析			
游依赖				手动添加上游
輸出名称	节点名	节点ID (实例ID)	负责人	5 操作
virtı	virtua			
_11	_110(and the second se	- SoperPort	Ē
325	3255	_		
virtı	virtua			
_11	_110(-	Ē
325568	3255			
前节点				手动添加输出
輸出名称	节点名	节点ID (实例ID)	负责人	5.操作
mo	mc		10000	÷
ql_	ql_			ш <
mo	mc		Concerning State	
al	al			

参数	描述
	当节点的任务类型为SQL时,您可以单击 自动解析 ,系统会解析代码中的表,并查 找到与该表名相同的输出名称。输出名称所在的节点作为当前节点的上游依赖。 如果代码中引用项目变量或不指定项目,则系统默认解析为生产项目名,以保证生 成调度的稳定性。例如,开发项目名称为 onedata_dev :
	• 如果代码里指定 select * from s_order ,则调度解析依赖为 onedata. s_order 。
自动解析	 如果代码里指定 select * from \${onedata}.s_order ,则调度解析依赖 为 onedata.s_order 。
	• 如果代码里指定 select * from onedata.s_order , 则调度解析依赖为 onedata.s_order 。
	• 如果代码里指定 select * from onedata_dev.s_order , 则调度解析依 赖为 onedata_dev.s_order 。

参数	描述				
上游依赖	 通过执行如下操作,添加该节点任务调度时依赖的上游节点: i. 单击手动添加上游。 ii. 在新建上游依赖对话框中,您可以通过以下两种方式搜索依赖节点: 输入所依赖节点的输出名称的关键字进行搜索节点。 输入virtual搜索虚拟节点(每个租户或企业在初始化时都会有一个根节点)。 ⑦ 说明 节点的输出名称是全局唯一的,且不区分大小写。 iii. 单击确定新增。 同时您还可以单击操作列下的 图标,删除已添加的依赖节点。 				
当前节点	 通过执行如下操作,设置当前节点的输出名称,根据需要您可以设置多个输出名称,供其他节点依赖使用: i.单击手动添加输出。 ii.在新增当前节点输出对话框中,填写输出名称。输出名称的命名规则请尽量统一,一般命名规则为 生成项目名.表名 且不区分大小写,以标识本节点产出的表,同时其他节点更好地选择调度依赖关系。 例如,开发项目名称为 onedata_dev ,建议将输出名称设置为 onedata .s_order 。如果您将输出名称设置为 onedata_dev.s_order ,则仅限代码 select * from onedata_dev.s_order 能解析出上游依赖节点。 ii.单击确定新增。 同时您还可以对当前节点已添加的输出名称执行如下操作: 单击操作列下的 图标,删除已添加的输出名称。 如果该节点已提交或已发布,且被任务所依赖(任务已提交),则单击操作列下的 图标,查看下游节点。 				

6. (可选) 配置Flink SQL任务的调度参数

i. 在任务配置面板的离线模式页签的调度配置区域,配置任务的调度参数。

• 调度配置		
时间属性	● 正常调度 ○ 空跑调度	
暂停调度	• 否 是	SuperAdmin
调度周期	日 ∨ 00:00	
cron 表达式	0 0 0 * * ?	SuperAdmin(Su SuperAdmin(Su
最近的生成时间	2020-10-27 00:00:00 2020-10-28 00 2020-10-31 00:00:00	:00:00 2020-10-29 00:00:00 2020-10-30 00:00:00
依赖上周期	• 否	
	○是	
优先级	中等优先级	
参数配置	请输入配置的参数,支持自定义参数	, 以; 区隔, 如keyA=123; keyB=\${yyyymmdd-7};
	节点参数配置说明	

参数	描述
时间属性	选择时间属性。时间属性包括: 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。
暂停调度	暂停调度 选择是后,即可暂停该任务的调度,会按照下面的调度周期时间配置 调度,但是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。

参数	描述					
调度周期	 调度周期可选择日、周、月、小时和分钟: 日调度,即调度任务每天自动运行一次。新建周期任务时,系统默认的时间 周期为每天0点运行一次。您可以根据需要,单击 ◎ 图标,指定运行的时间 点。 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。您可以 根据需要,单击 ◎ 图标,指定运行的时间点。 如果您没有指定日期,为保证下游实例正常运行,系统会生成实例后直接设 置为运行成功,而不会真正执行任何逻辑,也不会占用资源。 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。您可 以根据需要,单击 ◎ 图标,指定运行的时间点。 如果在没有被指定的日期时,为保证下游实例正常运行,系统会每天生成实 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。 小时调度,即每天指定的时间段内,调度任务按间隔时间数的时间间隔运行 一次。或选择指定的时间点,调度系统会自动为任务生成实例并运行。您可 以根据业务需求选中时间段或时间点: 如果您选中了时间段,您可以单击开始或结束后的 ◎ 图标,指定运行的 开始和结束时间。同时您可以单击间隔后的 ◎ 图标,在下拉列表中选择间 隔时间。 如果您选中了时间点,单击下拉列表框,在下拉列表中选择时间点。 例如,每天00:00~23:59的时间段内,每隔1小时会自动调度一次,因此调度 系统会自动为任务生成实例并运行。 分钟调度,即每天指定的时间段内,调度任务按间隔时间数的时间间隔运行 一次。您可以单击开始或结束后的 ◎ 图标,指定运行的开始和结束时间。同 时您可以单击间隔后的 ◎ 图标,在下拉列表中选择间隔时间。 					
依赖上周期	根据业务场景选择本周期节点的运行,是否需要依赖上一周期本节点或其他节点 的运行结果。 选择节点类型。系统支持选择 自定义和当前 。适用场景说明如下: 本周期节点是否运行取决于上一周期本节点是否正常产出数据,则需要选 择 当前 。只有上一周期本节点运行成功,才会启动运行本节点。 代码任务没有用到某个节点的产出表,但业务上需要依赖该节点的上一周期 是否正常产出数据,则需要选择依赖 自定义 节点。					

参数	描述
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。优先级包括: 最低优先级。 低优先级。 中等优先级。 高优先级。 最高优先级。
参数配置	您可以对代码中所用参数的具体赋值。单击 节点参数配置说明 ,查看Dataphin 调度系统的配置规则及支持配置的时间参数。

- ii. 单击确定。
- 7. (可选)配置Flink SQL任务的参数。

在任务配置面板,任务参数配置区域,通常需要配置批任务的bizdate和流批任务的stat_date参数。

•	任务参	数
	01	# excatly-once语义
	02	blink.checkpoint.mode=EXACTLY_ONCE
	03	# checkpoint间隔时间,单位毫秒
	04	blink.checkpoint.interval.ms=180000
	05	blink.checkpoint.timeout.ms=600000
	06	# 使用niagara作为statebackend,以及设定state数据生命周期,单位毫秒
	07	state.backend.type=niagara
	08	state.backend.niagara.ttl.ms=129600000
	09	# 开启5秒的microbatch
	10	#blink.microBatch.allowLatencyMs=5000
	11	#blink.miniBatch.allowLatencyMs=5000
	12	#blink.miniBatch.size=20000
	13	# local 优化,默认已经开启
	14	#blink.localAgg.enabled=true
	15	# 开启partial优化,解决count distinct热点
0	流任务	: 定义格式为 stat_date>='xxxx' 。
0	批任务	: 定义格式为 stat_date='xxxx' 或 set project.table.batch.`part
	te}*'	· •

步骤四: 提交Flink SQL任务

- 1. 在Flink SQL任务代码开发页面,单击页面右上方的圆图标,保存Flink SQL任务。
- 2. 在Flink SQL任务代码开发页面,单击页面右上方的 图标,提交Flink SQL任务。
- 3. 在提交备注对话框,填写备注信息。

- 4. 单击确定并提交,完成Flink SQL任务的提交。
- 5. (可选)发布Flink SQL任务至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布实时计算任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的实时计算任务,即可进入生产环境。

后续步骤

在运维中心查看并运维Flink SQL任务,保证任务的正常运行。具体操作,请参见<mark>实时任务或实时实例</mark>。

3.5.6. 创建Flink Template SQL任务

3.5.6.1. 步骤一: 创建并调试Flink Template SQL计算任务

Flink Template SQL计算任务是基于已有的代码模板开发的任务。当您需要开发一批代码逻辑及结构相同的 计算任务时,可以先创建代码模板,再创建Flink Template SQL计算任务,大大提高了代码研发的效率。本 文为您介绍如何创建Flink_Template_SQL类型的计算任务。

前提条件

在您开始执行操作前,请确认已完成代码模板的创建及提交。具体操作,请参见新建代码模板。

使用限制

Flink Template SQL计算任务的代码,不支持编辑。如果您需要编辑Flink Template SQL任务的代码,则需要进入实时计算模板页面进行修改。

模板参数	任务属性	任务配置	代码结构	版本信息	SQL建议				🖒 刷新
· ў·模板信 如需编辑	E务代码, ^{重任务代码,词}	只读模式不可	丁编辑 的实时计算模板	ī 5 5 去实时计:	算模板 义	模板信息 模板名称: dfasdf 人: t	模板版本: 1 🗸	模板负责	

新建Flink Template SQL任务

- 1. 登录Dataphin控制台。
- 2. 在Dataphin首页,单击顶部菜单栏研发。
- 3. 在数据开发页面,按照下图指引,进入新建Flink_Template_SQL页面。

DEV		<mark>> 1 Prod</mark> Dev	Ð	৵৸	······································
焎	远建模 数 据	数理 2 🛛 🛛	席查询	~~	他使用Dataphini
«»		C A	∃ 🗄	5	
0		æ 7 adatop ^{nwy} (5	MAXC任务	· > 	◆ 规范定义 》
Supe	 ▶ 0-0719 (1) ▶ 0-0728 (2) 		通用脚本		FLINK_SQL
	▶ 00-nj-test070	8 (3)		ľ	FLINK_DATASTREAM

4. 在新建Flink_Template_SQL对话框,配置参数后,单击确定。

新建 Flink_Template_SQL					
* 名称	请输入任务名称,新建后名称不可修改				
* 选择目录	实时计算任务	~			
* 资源队列	请选择资源队列	~			
* 引擎版本	请选择引擎版本	~			
* 引用模板	请选择引用模版	~			
模板版本	~				
	取消	确定			

参数	描述
名称	名称的命名规则如下: • 只能包含小写英文字母、数字、下划线(_)。 • 名称的长度范围为3~62个字符。 • 项目内的名称不支持重复。 • 名称仅支持以英文字母开头。
选择目录	实时计算任务所属的目录。
资源队列	选择Flink Template SQL任务运行的资源队列,该资源队列为当前项目所绑定的实 时计算源中的资源队列。
引擎版本	资源队列的版本。
引用模板	选择需要引用的模板。

预编译及调试Flink Template SQL任务

如果需要修改模板输入的参数默认值,则进入模板参数页面进行修改。



1. 单击页面右上方的预编译, 校验代码任务的语法及权限问题。

2. 单击页面右上方的调试,代码任务采样数据并进行本地调试,保障代码任务的正确性。

- i. 在调试配置对话框的选择采样模式页签中,选择调试的模式后,单击下一步。 调试配置 X 1 选择采样模式 ------ 2 采样调试数据 当前仅支持单种模式调试,选择模式后请采样对应模式表数据进行调试 实时模式调试 离线模式调试 采样对应的实时物理表数据,完成数据采样后将 采样对应的离线物理表数据,完成数据采样后将 在FLINK Stream模式下进行本地调试 在FLINK Batch模式下进行本地调试 取消

ii. 在**采样调试数据**页签中,为元数据进行采样调试。

您可以通过自动抽样和上传数据的方式,为元数据进行采样调试。适用场景说明如下:

■ 自动抽样: 自动抽样到的数据是随机的, 所以适用于对采集到的数据没有限制的场景。选择自动 抽样后,需要配置抽样条数。

↓ 注意 如果元数据表中没有数据,则自动抽样将采集不到数据。

■ 上传本地数据: 自动抽样时元数据采集不到或数据抽样的逻辑比较严格, 例如从100万条数据中 抽取其中1条数据,这样采集效率就很低,可以选择手动上传本地数据。

上传本地数据前需要先下载样例,根据下载的样例编辑需要上传的数据,单击上传后,数据自动 填充至**元数据采样**区域。

✓ 选择采样模式			2 采样调试数据	
数据表列表 所有数据表都需要已采样,才可进行调试 ————————————————————————————————————	元数据采样	*	样例下载 ∨ ▲	上传 ① 自动抽样
源表 v2_8_0_stream_dev.moyi_data X 未采样	5- 	id	name	action

■ 手动输入:适用于采集的数据比较少,或者需要修改已采集到的数据的场景。

iii. 完成所有数据表的元数据采样后,单击页面下方的确定。

iv. 在Result页面, 查看调试数据、中间结果和调试结果。

调试	数据 中间	间结果	调试结果 													
<	tmp_multipl	e_days	tmp_multi	ple_days												
	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A
1	null	null	null	null	null	null	null	null	null	null	nuli	null	null	null	null	null
2	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
з	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
4	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
5	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null

完成Flink Template SQL计算任务默认为开启实时模式,您可以手动开启离线模式。

3.5.6.2. 步骤二: 配置并提交Flink Template SQL任务

完成Flink Template SQL任务的创建及调试后,需要为Flink Template SQL任务配置调度资源、时间参数、依赖关系及任务参数。本文为您介绍如何配置并提交Flink Template SQL任务。

配置说明

配置操作	适用场景	使用限制
配置Flink Template SQL任务 的资源	Dataphin默认为新建的Flink Template SQL任务使用系统推荐配置。 当系统推荐的资源不符合用户业务场景的性能要求时,您可以自定义 任务调度的资源。	仅实时模式支持配 置资源。
配置实时任务的时 间参数	为了规避在任务参数处经常漏掉配置stat_date,您只需要在实时任 务配置的属性配置中新增stat_date的kv配置。	仅实时模式支持配 置时间参数。
配置依赖关系	配置任务运行时的上下游依赖关系。	实时模式和离线模 式均支持配置依赖 关系。
调度配置	为离线周期任务配置调度参数。	仅离线模式支持配 置调度参数。
配置任务参数	配置Flink Template SQL任务运行的参数,例 如 log4j.rootLogger=INFO,file 。	实时模式和离线模 式均支持配置任务 参数。
提交或发布Flink Template SQL任务 至生产环境	提交或发布Flink Template SQL任务至生产环境。	不涉及

前提条件

在执行操作前,请确认您已完成Flink Template SQL任务的创建及调试。具体操作,请参见步骤一:创建并调试Flink Template SQL计算任务

配置Flink Template SQL任务的资源

↓ 注意 仅实时模式支持配置资源。

- 1. 在Flink Template SQL任务代码开发页面,单击页面上方的任务配置。
- 2. 在任务配置页面,确认在实时模式页签。

在实时模式页签,为您展示任务运行所使用的资源队列、引擎版本、资源配置类型及自动调优。

• 资源配置		
* 资源队列	root.dt-streamcompu	
	切换后提交生效的任务执行可能有差异。清谨慎操作	
* 引擎版本	blink-3.4.4	
* 资源配置类型	 系统推荐配置 自定义配置 ③ 去配置》 	
资源配置记录	暂未保存资源记录	💾 保存当前配置为新记录
* 自动调优		
	最大CU数 1 / 总CU: 2147483647	
	期望最大内存 1 MB	

资源队列和**引擎版本**默认为创建Flink Template SQL任务时选择的配置。同时,您也可以修改资源队列 和引擎版本。

- 3. 选中自定义配置为资源配置类型,并单击去配置。
- 4. 在资源配置页面,为您展示一张拓扑图,图中每个方框代表了一个计算任务,都可以进行独立配置。每 个Group代表着Group内部的节点可以存放在一台机器进行计算,可以有效避免数据的网络传播,提升 性能。图中当前的资源配置就是默认为您展示系统推荐的资源配置。



i. 单击需要配置资源的Group右上角的①后,在自定义配置Group执行参数对话框配置参数后,单 击确定。

		自定义配置 Group 执	行参数		- 2 ×
	1	USERSON CONTRACTOR	0.05		101 1010
0		* core :	0.25		
GROUP-1	GRC	* heap_memory :	1260		
TT4TableSource-speed_boat_dev_dwd_a e_trd_pay_mix-Stream->SourceConversio n(table:[builtin, default,_DataStreamTable	Gro ct: (set	* parallel :	4		
_0, source: [1141ableSource-speed_boat _dev_dwd_ae_trd_pay_mix]], fields:(f0))-> MiniBatchAssigner(intervalMs: 5000)->c	HASH er_i buy	* direct_memory :	0		A second second
CORE 0.25 HEAP MEMORY 1260.00 PARALLEL 4	CC HE PA	* native_memory :	20		
				取消	确定
参数	描述				
core	通常,core配置成 值不能超过1,0.2	戊0.25。即一个CPU可나 25即可,0.125是最低	以支持4个线 下限了,同时	程同时运算,因 时支持8个线程。	引此core的取

参数	描述
heap_memory	单位为MB, heap_memory是堆内存,供Java应用程序使用的内存。 heap_memory及其内部各组成的大小可以通过JVM的一系列命令行参数来控 制,在一般的blink程序中,都会需要一定的heap_memory开销,,例如申请一 定的heap_memory作为程序的缓存等,因此您可以按程序的规模来设置其大 小。
parallel	表示同时并发的线程数量,用户可以选择合适的数量来运行自己的任务,并不是 越大越好,越大代表你资源申请的越多,反而对性能有抑制。 通常,一个简单计算节点每秒可以处理2000~4000条之间的数据。
direct_memory	单位为MB, direct_memory并不是虚拟机运行时数据区的一部分,也不是Java 虚拟机规范中定义的内存区域。但是这部分内存也被频繁的使用,而且也可能导 致OutOfMemoryError异常出现。如果您的程序有使用igraph或者swift,可以 适当配置其大小,如16-32MB。 在Java NIO (New Input/Output)类,引入了一种基于通道(Channel)与缓 冲区(Buffer)的I/O方式,direct_memory可以使用Native函数库直接分配堆 外内存,然后同一个存储在Java堆中的DirectByteBuffer对象作为这块内存的引 用进行操作。这样能在一场场景中显著提高性能,因为避免了在Java堆和Native 堆中来回复制数据。
native_memory	单位为MB, native_memory没有相应的参数来控制大小,其大小依赖于操作系 统进程的最大值(对于32位系统就是3~4G,各种系统的实现并不一样),以及 生成的Java字节码大小、创建的线程数量、维持Java对象的状态信息大小(用于 GC)以及一些第三方的包。 native memory存放下面4种信息: 管理Java heap的状态数据(用于GC)。 JNI调用,也就是Native Stack。 JIT(即使编译器)编译时使用native memory,并且JIT的输入(Java字节 码)和输出(可执行代码)也都是保存在native memory。 NIO direct buffer。

ii. 按照下图指引,进入自定义配置Operator执行参数对话框并配置参数。完成参数配置后单击确定。

GROUP-11 CROUP-11		自定义配置Operator-11执行参数 ×
GlobalGroupAggregate(groupBy: (stat_mi n), select:(stat_min, COUNT_RETRACT(d istinctS0 countS0) AS pay, ord, byr, ont m	OPERATOR-11	
tr_037, COUNT_RETRACT(distinct\$1 cou nt\$1) AS pay_mord>Calc(select: ((CAS	GlobalGroupAggregate(groupBy: (stat min), select:(stat_min, COUNT_RETRA	* core : 0.25
	cnt mtr 037. COUNT RETRACT(di	* heap_memory: 102
HEAP MEMORY 510.00 PARALLEL 4	CORE 0.25 HEAP MEMORY 102.00 NATIVE MEMORY 614.00 PARALLEL 4	* parallel: 4
		* direct_memory : 0
		* native_memory: 614
		* state_size : 0
		* chain_strategy: ALWAYS NEVER HEAD
		取消 确定

自定义配置Operator执行参数对话框中core、heap_memory、parallel、direct_memory和 native_memory参数解释请参见上一步骤,下表仅对state_size、chain_strategy参数进行解释。

参数	描述
state_size	保持默认0即可。
chain_strategy	 chain_strategy用于定义多个节点的链接策略,Dataphin支持的链接策略包括: Always:默认参数,即和其他节点均部署在一台机器上,没有特殊要求。 Never:节点不会和其他节点放在一台机器上,即需要独立部署。 Head:可以接受和其他节点放在一台机器上,但是只能作为Group的头节点。
	↓ 注意 目前, Head和Never极少出现, 默认Always即可。

iii. 配置完成后,单击页面右上方的保存。

↓ 注意 如果您想继续使用系统推荐的资源配置,则单击页面右上方的重置为系统初始。

5. (可选)保存资源配置记录。

↓ 注意 仅实时模式支持保存资源配置记录。

如果资源配置类型为自定义配置,则单击资源信息记录后的保存当前配置为新纪录。在保存资源记录对话框中,输入资源记录名称后,单击确定。

对已有的资源信息记录,您可以执行以下操作。

操作	描述
查看版本信息	单击某个记录 操作 列下的 图标,查看版本信息。

操作	描述
启用记录	i. 单击某个记录的操作列下的 _€ 图标。 ii. 在提示对话框中,单击确定。
删除记录	i. 单击某个记录操作列下的 _面 图标。 ii. 在提示对话框中,单击确定

6. (可选)打开自动调优开关, 配置最大CU数和期望最大内存, 开启自动调优。

↓ 注意 仅实时模式支持自动调优功能。

配置实时任务的时间参数

任务配置 实时模式	离线模式	⑤ 当前生效模式: 实时	×
2.117月0日天主	 		
资源配置记录	暂未保存资源记录	💾 保存当前配置为新记录	Ē
* 自动调优	最大CU数 1 1 1 成CU:200 期望最大内存 1 MB		
ann samaise Sa	er AdminiSu Super AdminiSu Super AdminiSu	SuperAdminiSu SuperAdminiSu	an san
▲属性配置 时间参数	请输入配置的参数,支持自定义参数,以; 区隔, 续	@keyA=123; keyB=\${yyyymmdd-7};	

实时任务的时间参数为stat_date,用于方式实时计算任务的运行时间的偏移,例如,您需要计算当天某个指标聚合值,为了防止时间偏移,则您需要设置state_date大于当天零点,过滤掉偏移的时间点。

为了规避在**任务参数**处经常漏掉配置stat_date,您只需要在实时任务配置的属性配置中新增stat_date的kv 配置,其中Value是一个基于业务时间的表达式,同时您也可以配置多个时间参数,使用半角分号(;)分 割。例如,stat_date=\${yyyyMMdd-1},则任务运行过程中的开始执行时间为\${yyyyMMdd-1}。

属性配置
 时间参数 stat_date=\${yyyyMMdd-1}

配置依赖关系

↓ 注意 实时模式和离线模式均支持配置依赖关系。

配置 实时模	式 离线模式		 当前生 	效模式: 实时
依赖关系 自 上游依赖	动解析		antaburror a	手动沃加下的
輸出名称	节点名	节点ID (实例ID) 负责/	L L	操作
		A Station		
		暂无数据		
当前节点			00000000	手动添加输出
输出名称	节点名	节点ID (实例ID) 负责人	L .	操作
dws_	dws_	n_' 70· 诺为((58937)	市 く

参数	描述
上游依赖	 您可以通过自动解析和手动添加两种方式,为Flink Template SQL任务节点添加上游依赖的节点: 单击自动解析,Dataphin会解析Flink Template SQL任务代码中的表,并查找到与该表名相同的输出名称。输出名称所在的节点作为当前节点的上游依赖。如果代码中引用项目变量或不指定项目,则系统默认解析为生产项目名,以保证生成调度的稳定性。例如,开发项目名称为 onedata_dev : 如果代码里指定 select * from s_order ,则调度解析依赖为 onedata.s_order 。 如果代码里指定 select * from \${onedata}.s_order ,则调度解析依赖为 onedata.s_order 。 如果代码里指定 select * from onedata.s_order ,则调度解析依赖为 onedata.s_order 。 如果代码里指定 select * from onedata_dev.s_order ,则调度解析依赖为 onedata.s_order 。 如果代码里指定 select * from onedata_dev.s_order ,则调度解析依赖为 onedata.gev.s_order 。 如果需要添加其他节点作为当前标签的上游节点,则需要手动添加上游依赖的物理节点。 ① 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 。

参数	描述		
	通过执行如下操作,设置当前节点的输出名称,根据需要您可以设置多个输出名称,供 其他节点依赖使用:		
	1. 单击手动添加输出。		
	 在新增当前节点输出对话框中,填写输出名称。输出名称的命名规则请尽量统一,一般命名规则为 生成项目名.表名 且不区分大小写,以标识本节点产出的表,同时其他节点更好地选择调度依赖关系。 		
当前节点	例如,开发项目名称为 onedata_dev ,建议将输出名称设置为 onedata.s_ order 。如果您将输出名称设置为 onedata_dev.s_order ,则仅限代码 s elect * from onedata_dev.s_order 能解析出上游依赖节点。		
	3. 单击确定新增。		
	同时您还可以对当前节点已添加的输出名称执行如下操作:		
	 单击操作列下的 图标,删除已添加的输出名称。 		
	● 如果该节点已提交或已发布,且被任务所依赖(任务已提交),则单击操作列下的 Ⅰ		
	图标,查看下游节点。		

调度配置

↓ 注意 仅支持为离线模式配置调度参数。

任务配置 实时	模式 离线模式			1 当前:	生效模式: 实时	
* 离线模式 🔵	and and webs					
• 调度配置						
时间属性	 正常调度 否 是 	空跑调度				
调度周期	Η ∨	00:00				
cron 表达式	000**?					
最近的生成时间	2021-07-31 00:00:0 2021-08-04 00:00:0	0 2021-08-01 00: 0	:00:00 2021-08	-02 00:00:00 202	1-08-03 00:00:00	
依赖上周期	 ● 否 ● 是 					
优先级	中等优先级		- YER_1			
时间参数	请输入配置的参数。	, 支持自定义参数,	以; 区隔, 如	keyA=123; keyB=\$	\${yyyymmdd-7};	
	节点参数配置说明					

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。 			
依赖上周期	根据业务场景选择本周期节点的运行,是否需要依赖上一周期本节点或其他节点的运行 结果。 选择节点类型。系统支持选择自定义和当前。适用场景说明如下: •本周期节点是否运行取决于上一周期本节点是否正常产出数据,则需要选择当前。只 有上一周期本节点运行成功,才会启动运行本节点。 •代码任务没有用到某个节点的产出表,但业务上需要依赖该节点的上一周期是否正常 产出数据,则需要选择依赖自定义节点。			
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。			
时间参数	您可以对代码中所用参数的具体赋值。单击 节点参数配置说明 ,查看Dataphin调度系 统的配置规则及支持配置的时间参数。			

配置任务参数

在任务参数配置区域,配置Flink Template SQL任务的参数。

任	务团	配置	实时模式	离线模式		1 当前	生效模式:	实时
				NON THE REPORT OF				
	 任务参数 				批	任务参数推荐	中档	\sim
		01	# for medium	i-level jobs				
		02	#blink.operate	or.name.max.length=-1				
		03	# sql 部分					
		04	# 优化参数					
		05	sql.optimizer.	agg.phase.enforcer=TWO_PF	HASE			
		06	# 维表使用流-	-样的 lookup join, batch odp	s cacheall 时等效 [:]	于 mapjoin 效果	e.	
		07	sql.optimizer.	use.lookup-join.in-batch=tru	e			
		08	sql.optimizer.	lookup-join-to-hash-join.ena	bled=true			
		09	# 3.7.0 开始支持, 一路拆多路再 union all 操作优化成 expand 算子, 减少 output 边					
		10	sql.optimizer.	union-project-to-expand.ena	bled=true			
		11	sql.optimizer.	non-deterministic-calc-merg	e.enabled=true			
		12	sql.optimizer.	partitioned-lookup-join.buck	et=64			
		13	# 算子资源配置	<u>-</u>				
		14	# hash-agg 単	9并发管理内存大小,默认32				
		15	sql.resource.h	ash-agg.table.memory.mb=2	256			
		16	sql.resource.h	ash-agg.table-prefer-memo	y-mb=256			
		17	# sort 单并发管	管理内存大小,默认32				
		18	sql.resource.s	ort.buffer.memory.mb=64				
		19	# external-bu	ffer单并发管理内存大小,默认	10			
		20	sql.resource.e	external-buffer.memory.mb=0	54			
		21	sql.resource.h	ash-join.table.memory.mb=2	256			
		22	sql.resource.s	ource.never-chaining-strateg	y.enabled=true			
		23	# 粗粒度作业资	资源配置				

提交或发布Flink Template SQL任务至生产环境

1. 按照下图指引,提交Flink Template SQL任务至生产环境。



2. 如果项目空间的开发模式为Dev-Prod,则需要发布Flink Template SQL任务至生产环境。

3.5.7. 创建Flink DataStream计算任务

Flink DataStream计算任务是基于Java环境研发的计算任务,用于进行数据的实时处理。本文为您介绍如何 基于JAR资源文件,构建Flink DataStream实时计算任务。

前提条件

在您开始执行操作前,请确认已上传已经完成开发的Datstream作业的JAR包至Dataphin平台,详情请参见新 建资源。

使用限制

Dataphin支持超级管理员、项目管理员和开发者创建Flink DataStream计算任务。

步骤一:新建Flink DataStream任务

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏研发。
- 4. 按照下图指引,进入新建 Flink DataStream对话框。



5. 在新建Flink_DataStream对话框,配置参数后,单击确定。
| 新建 Flink_DataStream × | | | | |
|-----------------------|-------------------|--------|--|--|
| | | | | |
| * 名称 | 请输入任务名称,新建后名称不可修改 | | | |
| * 选择目录 | 实时计算任务 | \sim | | |
| * 选择资源 | 请选择该任务依赖的资源包 | | | |
| * 类名 | 请输入该任务将使用资源的完整类名 | | | |
| * 资源队列 | 请选择资源队列 | \sim | | |
| * 引擎版本 | 请选择引擎版本 | \sim | | |
| | | | | |
| | 取消 确定 | | | |
| 公 | 47.44 | | | |

参数	描述
名称	 名称的命名规则如下: 只能包含小写英文字母、数字、下划线(_)。 名称的长度范围为3~62个字符。 项目内的名称不支持重复。 名称仅支持以英文字母开头。
选择目录	实时计算任务所属的目录。
选择资源	该实时任务依赖的资源包。
类名	使用资源的完整类名。
资源队列	该项目所绑定的实时计算源中的资源队列。
引擎版本	当前资源队列所支持的版本。

步骤二:开发Flink DataStream任务的代码

- 在Flink DataStream任务代码页面,编写任务的代码。
 您可以单击页面右上方的格式化,系统自动调整SQL代码格式。
- 2. 单击页面右上方的预编译, 校验代码任务的语法及权限问题。
- 3. 单击页面右上方的调试,代码任务采样数据并进行本地调试,保障代码任务的正确性。

i. 在调试配置对话框的选择采样模式页签中,选择调试的模式后,单击下一步。

调试配置							×
		1 选择采样模	武	2 采柏	间试数据		
		当前仅支持单种模式	、调试,选择模式质	后请采样对应模式表数	救据进行调试		
SN SUP ^{ar B}					R		
S ^{al} cuper ^a		空时横式调试		ja ja	容线横式调试		
	采样对应的实时 在FLINK S	物理表数据,完成数 tream模式下进行本均	据采样后将 也调试	采样对应的离线物 在FLINK Ba	理表数据,完成数据系 tch模式下进行本地调	彩样后将 试	
Suber Al	iminter Super	daninter SuperAc	ininian Super				
					取消	下 ⁻	步

ii. 在**采样调试数据**页签中,为元数据进行采样调试。

您可以通过自动抽样和上传数据的方式,为元数据进行采样调试。适用场景说明如下:

 自动抽样:自动抽样到的数据是随机的,所以适用于对采集到的数据没有限制的场景。选择自动 抽样后,需要配置抽样条数。

↓ 注意 如果元数据表中没有数据,则自动抽样将采集不到数据。

上传本地数据:自动抽样时元数据采集不到或数据抽样的逻辑比较严格,例如从100万条数据中抽取其中1条数据,这样采集效率就很低,可以选择手动上传本地数据。

上传本地数据前需要先下载样例,根据下载的样例编辑需要上传的数据,单击**上传**后,数据自动 填充至**元数据采样**区域。

⊘ 选择采样模式			2 采样调试数据	2
数据表列表 所有数据表都需要已采样,才可进行调试 ————————————————————————————————————	元数据采样	<u>.</u>	样例下载 > 土	とた 住 で し は の 抽样
源表 v2_8_0_stream_dev.moyi_data X 未采样	54 	id	name	action

■ 手动输入:适用于采集的数据比较少,或者需要修改已采集到的数据的场景。

iii. 完成所有数据表的元数据采样后,单击页面下方的确定。

iv. 在Result页面,查看调试数据、中间结果和调试结果。

调订	:数据 中	间结果	调试结果 													
<	tmp_multip	le_days	tmp_mult	iple_days												
	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A	字段A
1	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
2	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
3	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
4	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null
5	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null	null

步骤三: 配置Flink DataStream任务的资源

- 1. 在Flink DataStream任务代码开发页面,单击页面上方的任务配置。
- 2. 在任务配置页面,确认在实时模式页签。

在实时模式页签,为您展示任务运行所使用的资源队列、引擎版本、资源配置类型及自动调优。

• 资源配置			
* 资源队列	root.dt-streamcompu		
1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 -	切换后提交生效的任务执行可能有差异墙谨慎操作		
* 引擎版本	blink-3.4.4 V		
* 资源配置类型	● 系统推荐配置		
(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	● 自定义配置		
资源配置记录	暂未保存资源记录	💾 保	存当前配置为新记录
* 自动调优			
	最大CU数 1 / 总CU: 2147483647		
	期望最大内存 1 MB		

资源队列和**引擎版本**默认为创建Flink Dat aSt ream任务时选择的配置。同时,您也可以修改资源队列和 引擎版本。

- 3. 选中自定义配置为资源配置类型,并单击去配置。
- 4. 在资源配置页面,为您展示一张拓扑图,图中每个方框代表了一个计算任务,都可以进行独立配置。每 个Group代表着Group内部的节点可以存放在一台机器进行计算,可以有效避免数据的网络传播,提升 性能。图中当前的资源配置就是默认为您展示系统推荐的资源配置。



i. 单击需要配置资源的Group右上角的①后,在自定义配置Group执行参数对话框配置参数后,单 击确定。

		自定义配置 Group 执	行参数		- 2×
11		* core :	0.25		
GROUP.1	GRO	* heap_memory :	1260		
TT4TableSource-speed_boat_dev_dwd_a e_trd_pay_mix-Stream->SourceConversio n(table:/builtin, default, DataStreamTable	Gro ct: (set	* parallel :	4		
_0, source: [TT4TableSource-speed_boat _dev_dwd_ae_trd_pay_mix]], fields:(f0))-> MiniBatchAssigner(intervalMs: 5000)->c	HASH HASH er_i buy	* direct_memory :	0		
CORE 0.25 HEAP MEMORY 1260.00	CC	* native_memory :	20		
PARALLEL 4				取当	确定
				HI XI	
参数	描述				
core	通常,core配置成 值不能超过1,0.2	ർ0.25。即一个CPU可以 25即可,0.125是最低	以支持4个线 下限了,同E	程同时运算,因 时支持8个线程。	此core的取

参数	描述
heap_memory	单位为MB, heap_memory是堆内存,供Java应用程序使用的内存。 heap_memory及其内部各组成的大小可以通过JVM的一系列命令行参数来控 制,在一般的blink程序中,都会需要一定的heap_memory开销,,例如申请一 定的heap_memory作为程序的缓存等,因此您可以按程序的规模来设置其大 小。
parallel	表示同时并发的线程数量,用户可以选择合适的数量来运行自己的任务,并不是 越大越好,越大代表你资源申请的越多,反而对性能有抑制。 通常,一个简单计算节点每秒可以处理2000~4000条之间的数据。
direct_memory	单位为MB, direct_memory并不是虚拟机运行时数据区的一部分,也不是Java 虚拟机规范中定义的内存区域。但是这部分内存也被频繁的使用,而且也可能导 致OutOfMemoryError异常出现。如果您的程序有使用igraph或者swift,可以 适当配置其大小,如16-32MB。 在Java NIO (New Input/Output)类,引入了一种基于通道(Channel)与缓 冲区(Buffer)的I/O方式,direct_memory可以使用Native函数库直接分配堆 外内存,然后同一个存储在Java堆中的DirectByteBuffer对象作为这块内存的引 用进行操作。这样能在一场场景中显著提高性能,因为避免了在Java堆和Native 堆中来回复制数据。
native_memory	单位为MB, native_memory没有相应的参数来控制大小,其大小依赖于操作系 统进程的最大值(对于32位系统就是3~4G,各种系统的实现并不一样),以及 生成的Java字节码大小、创建的线程数量、维持Java对象的状态信息大小(用于 GC)以及一些第三方的包。 native memory存放下面4种信息: 管理Java heap的状态数据(用于GC)。 JNI调用,也就是Native Stack。 JIT(即使编译器)编译时使用native memory,并且JIT的输入(Java字节 码)和输出(可执行代码)也都是保存在native memory。 NIO direct buffer。

ii. 按照下图指引,进入自定义配置Operator执行参数对话框并配置参数。完成参数配置后单击确定。

GROUP-11 CROUP-	1	自定义配置Operator-11执行参数 ×
GlobalGroupAggregate(groupBy: (stat_mi n), select:(stat_min, COUNT_RETRACT(d istinctSo.countS0) AS nav. ord. bur. ont.m.	OPERATOR-11	
tr_037, COUNT_RETRACT(distinct\$1 cou nt\$1) AS pay mord ->Calc(select) (ICAS	GlobalGroupAggregate(groupBy: (stat min), select:(stat_min, COUNT_RETRA	* core : 0.25
TCAST('b2e7:md5\\u0004AE:biz\\u0004	CT(distinct\$0 count\$0) AS pay_ord_byr cnt mtr 037, COUNT RETRACT(di	
CORE 0.25	CORE	* heap_memory: 102
HEAP MEMORY 510.00 PARALLEL 4	HEAP MEMORY 102.00 NATIVE MEMORY 614.00 PARALLEL 4	* parallel: 4
		* direct_memory : 0
		* native_memory: 614
		* state_size : 0
		* chain_strategy: ALWAYS NEVER HEAD
		取消 确定

自定义配置Operator执行参数对话框中core、heap_memory、parallel、direct_memory和 native_memory参数解释请参见上一步骤,下表仅对state_size、chain_strategy参数进行解释。

参数	描述
state_size	保持默认0即可。
chain_strategy	 chain_strategy用于定义多个节点的链接策略,Dataphin支持的链接策略包括: Always:默认参数,即和其他节点均部署在一台机器上,没有特殊要求。 Never:节点不会和其他节点放在一台机器上,即需要独立部署。 Head:可以接受和其他节点放在一台机器上,但是只能作为Group的头节点。
	↓ 注意 目前, Head和Never极少出现, 默认Always即可。

iii. 配置完成后,单击页面右上方的保存。

↓ 注意 如果您想继续使用系统推荐的资源配置,则单击页面右上方的重置为系统初始。

5. (可选)保存资源配置记录。

↓ 注意 仅实时模式支持保存资源配置记录。

如果资源配置类型为自定义配置,则单击资源信息记录后的保存当前配置为新纪录。在保存资源记录对话框中,输入资源记录名称后,单击确定。

对已有的资源信息记录,您可以执行以下操作。

操作	描述
查看版本信息	单击某个记录 操作 列下的 图标,查看版本信息。

操作	描述
启用记录	i. 单击某个记录的操作列下的 _€ 图标。 ii. 在提示对话框中,单击确定。
删除记录	i. 单击某个记录操作列下的 _画 图标。 ii. 在提示对话框中,单击确定

6. (可选)打开自动调优开关, 配置最大CU数和期望最大内存, 开启自动调优。

↓ 注意 仅实时模式支持自动调优功能。

步骤四: 配置实时任务的时间参数

任务配置 实时模式	离线模式	● 当前生效模式: 实时	×
· www.mexe			
资源配置记录	暂未保存资源记录	💾 保存当前配置为新记录	
* 自动调优	最大CU数 1 人 は CU 200		
• 属性配置 时间参数	请输入配置的参数,支持自定义参数,以; 区隔, 如	ΩkeyA=123; keyB=\${yyyymmdd-7};	

实时任务的时间参数为stat_date,用于方式实时计算任务的运行时间的偏移,例如,您需要计算当天某个指标聚合值,为了防止时间偏移,则您需要设置state_date大于当天零点,过滤掉偏移的时间点。

为了规避在**任务参数**处经常漏掉配置stat_date,您只需要在实时任务配置的属性配置中新增stat_date的kv 配置,其中Value是一个基于业务时间的表达式,同时您也可以配置多个时间参数,使用半角分号(;)分 割。例如,stat_date=\${yyyyMMdd-1},则任务运行过程中的开始执行时间为\${yyyyMMdd-1}。

• 属性配置 时间参数 stat_date=\${yyyyMMdd-1}

步骤五: 配置依赖关系

- 1. 在Flink DataStream任务代码开发页面,单击页面上方的任务配置。
- 2. 在任务配置面板的依赖关系区域, 配置依赖关系。

THANK .				手动添加上游
俞出名称 2000	节点名	节点ID (实例ID)	负责人	公操作
virtu	virtua			
.11	_110(and the second se	- Saperha	Ō
25	3255			
irtı	virtua			
.11	_1100	and the second se	-	茴
25568	3255			
俞出名称	节点名	节点ID (实例ID)	负责人	S.操作 ^{。din}
10	mc		-	Super
I_	ql_			ā 🖓
	me		is service the	1.51
	ql_		1.	i 🖓

参数	描述
上游依赖	 您可以通过自动解析和手动添加两种方式,为Flink DataStream任务节点添加上游依赖的节点: 单击自动解析,Dataphin会解析Flink DataStream任务代码中的表,并查找到与该表名相同的输出名称。输出名称所在的节点作为当前节点的上游依赖。如果代码中引用项目变量或不指定项目,则系统默认解析为生产项目名,以保证生成调度的稳定性。例如,开发项目名称为 onedata_dev : 如果代码里指定 select * from s_order ,则调度解析依赖为 onedat a.s_order 。 如果代码里指定 select * from \${onedata.s_order ,则调度解析依赖为 onedata.s_order 。 如果代码里指定 select * from onedata.s_order ,则调度解析依赖 为 onedata.s_order 。 如果代码里指定 select * from onedata.s_order ,则调度解析依赖 为 onedata.gev.s_order ,则调度解析 依赖为 onedata.dev.s_order ,则调度解析 依赖为 onedata_dev.s_order 。 如果需要添加其他节点作为当前标签的上游节点,则需要手动添加上游依赖的物 理节点。 如果需要添加其他节点作为当前标签的上游节点,则需要手动添加上游依赖的物 建节点。 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	 通过执行如下操作,设置当前节点的输出名称,根据需要您可以设置多个输出名称,供其他节点依赖使用: i.单击手动添加输出。 ii.在新增当前节点输出对话框中,填写输出名称。输出名称的命名规则请尽量统一,一般命名规则为 生成项目名.表名 且不区分大小写,以标识本节点产出的表,同时其他节点更好地选择调度依赖关系。 例如,开发项目名称为 onedata_dev ,建议将输出名称设置为 onedata .s_order 。如果您将输出名称设置为 onedata_dev.s_order ,则仅限代码 select * from onedata_dev.s_order 能解析出上游依赖节点。 ii.单击确定新增。 同时您还可以对当前节点已添加的输出名称执行如下操作: 单击操作列下的意图标,删除已添加的输出名称。 如果该节点已提交或发布。且被任务所依赖(任务已提交),则单击操作列下的 图标,查看下游节点。

(可选)

步骤六:配置任务参数

Flink DataStream任务代码中的通用参数,您可以通过任务参数进行批量配置。

- 1. 在Flink DataStream任务代码开发页面,单击页面上方的任务配置。
- 2. 在任务配置面板,任务参数配置区域,配置参数。

步骤七:提交或发布Flink DataStream任务

1. 按照下图指引,提交Flink DataStream任务至生产环境。



2. 如果项目空间的开发模式为Dev-Prod,则需要发布Flink DataStream任务至生产环境。如何发布Flink DataStream任务至生产环境,请参见发布任务。

后续步骤

在运维中心查看并运维Flink DataStream任务,保证任务的正常运行。具体操作,请参见<mark>实时任务或实时实</mark> 例。

3.6. 新建ADB任务

3.6.1. 新建ADB for PostgreSQL任务

本文为您介绍如何构建ADB for PostgreSQL类型的计算任务。

背景信息

新建ADB for PostgreSQL任务的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的数据研发。
 - 单击快速进入研发项目区域的Dev或Basic项目,选择开发数据的项目空间。

? 说明

- 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。

 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的项目空间 (Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则无需选择项目空间。

- 5. 进入计算任务页面。
 - i. 在数据开发页面,单击数据处理页签。
 如果进入数据开发页面后,系统默认进入数据处理页签,则无需再次单击数据处理。
 - ii. 在**数据处理**页签,单击左侧导航栏**砅计算任务**图标。
- 6. 在计算任务页面,进入新建ADB for PostgreSQL任务的新建文件的对话框。
 - 单击计算任务后的 图标,选择ADB任务 > ADB_FOR_PG。
 - 单击项目名称后的 图标,选择数据处理 > 计算任务 > ADB_FOR_PG。
 - 在开发首页, 单击ADB_FOR_PG后的 ⊡图标。
- 7. 编写并运行代码。
 - i. 在新建文件对话框, 配置参数。

新建文件		×
* 名称	adb	
* 调度类型	● 周期性节点 ○ 手动节点	
描述	请输入任务描述	1.
选择目录	请选择目录	\sim
	取消	角定

参数	描述
名称	填写离线计算任务的名称,例如adb。
调度类型	选择离线计算任务调度的类型: 周期性节点:周期性任务可以参与生产环境的调度。 手动节点:手动任务需要触发其运行。
描述	填写对任务的简单描述。
选择目录	选择离线计算任务的目录。

ii. 单击确定。

iii. 在代码编写页面,编写SQL代码。

iv. 检查代码。

代码编写完成后,单击预编译,系统帮助您检查编写的SQL代码的语法。

如果预编译失败,您可以单击页面上方的**格式化**,系统自动帮助您调整语法格式。调整完格式,您 可以单击**刷新**,刷新页面上的代码。

- v. 单击页面右上方的执行,运行代码。
- 8. (可选)配置调度参数。
 - 如果离线计算任务的调度类型为周期性节点,则需要配置调度参数,详情请参见调度配置。
 - 如果离线计算任务的调度类型为手动节点,则需要手动触发离线计算任务的调度。
- 9. 在代码编写页面,保存、提交和发布ADB for PostgreSQL任务。
 - i. 单击右上方的 图标, 保存代码。
 - ii. 单击右上方的 图标, 提交代码。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交。
 - v. 发布ADB for PostgreSQL任务至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布ADB for PostgreSQL任务,详情请参见管理发布 任务。
 - 如果您的开发模式是Basic模式,则提交成功的离线计算任务,即可参与生产环境的调度。

3.6.2. ADB for PostgreSQL支持的语法

本文为您介绍Dataphin的计算引擎为AnalyticDB for PostgreSQL时,系统支持的SQL命令及语法。

- CREATE DATABASE
- CREATE INDEX
- CREATE ROLE
- CREATE SEQUENCE
- CREATE USER
- CREATE VIEW
- CREATE SCHEMA
- CREATE TABLE
- CREATE TABLE AS
- SELECT
- ALTER DATABASE
- ALTER EXTERNAL TABLE
- ALTER INDEX
- ALT ER SCHEMA
- ALTER TABLE
- DROP DATABASE
- DROP INDEX
- DROP SCHEMA

- DROP ROLE
- DROP SEQUENCE
- DROP USER
- DROP VIEW
- DROP TABLE
- **BEGIN**
- CHECKPOINT
- CLOSE
- CLUSTER
- COMMENT
- COMMIT
- DEALLOCATE
- DECLARE
- DELETE
- END
- EXPLAIN
- FET CH
- LOAD
- LOCK
- MOVE
- PREPARE
- REASSIGN OWNED
- **REINDEX**
- RELEASE SAVEPOINT
- RESET
- ROLLBACK
- ROLLBACK TO SAVEPOINT
- SAVEPOINT
- SELECT INTO
- SET
- SET ROLE
- SET SESSION AUTHORIZATION
- SET TRANSACTION
- SHOW
- START TRANSACTION
- TRUNCATE
- UPDATE
- VACUUM
- VALUES

CREATE DATABASE

创建数据库。

```
CREATE DATABASE name [ [WITH] [OWNER [=] dbowner]

[TEMPLATE [=] template]

[ENCODING [=] encoding]

[CONNECTION LIMIT [=] connlimit ] ]
```

CREATE INDEX

定义索引。

```
CREATE [UNIQUE] INDEX name ON table
[USING btree|bitmap|gist]
( {column | (expression)} [opclass] [, ...] )
[ WITH ( FILLFACTOR = value ) ]
[TABLESPACE tablespace]
[WHERE predicate]
```

更多信息请参见CREATE INDEX。

CREATE ROLE

```
定义数据库角色(用户或组)。
```

CREATE ROLE name [[WITH] option [...]]

更多信息请参见CREATE ROLE。

CREATE SEQUENCE

定义序列生成器。

```
CREATE [TEMPORARY | TEMP] SEQUENCE name
[INCREMENT [BY] value]
[MINVALUE minvalue | NO MINVALUE]
[MAXVALUE maxvalue | NO MAXVALUE]
[START [ WITH ] start]
[CACHE cache]
[[NO] CYCLE]
[OWNED BY { table.column | NONE }]
```

更多信息请参见CREATE SEQUENCE。

CREATE USER

定义默认具有LOGIN权限的数据库角色。

CREATE USER name [[WITH] option [...]]

更多信息请参见CREATE USER。

CREATE VIEW

定义视图。

```
CREATE [OR REPLACE] [TEMP | TEMPORARY] VIEW name
[ ( column_name [, ...] ) ]
AS query
```

更多信息请参见CREATE VIEW。

CREATE SCHEMA

定义SCHEMA。

```
CREATE SCHEMA schema_name [AUTHORIZATION username]
[schema_element [ ... ]]
CREATE SCHEMA AUTHORIZATION rolename [schema_element [ ... ]]
```

更多信息请参见CREATE SCHEMA。

CREATE TABLE

定义表。

```
CREATE [[GLOBAL | LOCAL] {TEMPORARY | TEMP}] TABLE table_name (
[ { column_name data_type [ DEFAULT default_expr ]
   [column constraint [ ... ]
[ ENCODING ( storage directive [,...] ) ]
1
   | table constraint
   | LIKE other_table [{INCLUDING | EXCLUDING}
                     {DEFAULTS | CONSTRAINTS}] ... }
   [, ... ] ]
   )
   [ INHERITS ( parent_table [, ... ] ) ]
   [ WITH ( storage parameter=value [, ... ] )
   [ ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP} ]
   [ DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY ]
   [ PARTITION BY partition_type (column)
       [ SUBPARTITION BY partition type (column) ]
          [ SUBPARTITION TEMPLATE ( template spec ) ]
       [...]
    ( partition spec )
        | [ SUBPARTITION BY partition type (column) ]
          [...]
    ( partition_spec
     [ ( subpartition spec
          [(...)]
        ) ]
    )
```

更多信息请参见CREATE TABLE。

CREATE TABLE AS

从查询的结果中定义新的表。

```
CREATE [ [GLOBAL | LOCAL] {TEMPORARY | TEMP} ] TABLE table_name
[(column_name [, ...])]
[ WITH ( storage_parameter=value [, ... ] ) ]
[ON COMMIT {PRESERVE ROWS | DELETE ROWS | DROP}]
[TABLESPACE tablespace]
AS query
[DISTRIBUTED BY (column, [ ... ] ) | DISTRIBUTED RANDOMLY]
```

更多信息请参见CREATE TABLE AS。

SELECT

从表或视图中检索行。

```
[ WITH with_query [, ...] ]
SELECT [ALL | DISTINCT [ON (expression [, ...])]]
* | expression [[AS] output_name] [, ...]
[FROM from_item [, ...]]
[WHERE condition]
[GROUP BY grouping_element [, ...]]
[HAVING condition [, ...]]
[WINDOW window_name AS (window_specification)]
[{UNION window_name AS (window_specification)]
[{UNION | INTERSECT | EXCEPT} [ALL] select]
[ORDER BY expression [ASC | DESC | USING operator] [NULLS {FIRST | LAST}] [, ...]]
[LIMIT {count | ALL}]
[OFFSET start]
[FOR {UPDATE | SHARE} [OF table_name [, ...]] [NOWAIT] [...]]
```

更多信息请参见SELECT。

ALTER DATABASE

修改数据库属性。

ALTER DATABASE name [WITH CONNECTION LIMIT connlimit] ALTER DATABASE name SET parameter { TO | = } { value | DEFAULT } ALTER DATABASE name RESET parameter ALTER DATABASE name RENAME TO newname ALTER DATABASE name OWNER TO new owner

更多信息请参见ALTER DATABASE。

ALTER EXTERNAL TABLE

改变外部表的定义。

ALTER EXTERNAL TABLE name RENAME [COLUMN] column TO new_column ALTER EXTERNAL TABLE name RENAME TO new_name ALTER EXTERNAL TABLE name SET SCHEMA new_schema ALTER EXTERNAL TABLE name action [, ...]

更多信息请参见ALTER EXTERNAL TABLE。

ALTER INDEX

改变索引的定义。

```
ALTER INDEX name RENAME TO new_name
ALTER INDEX name SET TABLESPACE tablespace_name
ALTER INDEX name SET ( FILLFACTOR = value )
ALTER INDEX name RESET ( FILLFACTOR )
```

更多信息请参见ALTER INDEX。

ALTER SCHEMA

改变模式的定义。

ALTER SCHEMA name RENAME TO newname ALTER SCHEMA name OWNER TO newowner

更多信息请参见ALTER SCHEMA。

ALTER TABLE

改变的表的定义。

```
ALTER TABLE [ONLY] name RENAME [COLUMN] column TO new_column
ALTER TABLE name RENAME TO new_name
ALTER TABLE name SET SCHEMA new_schema
ALTER TABLE [ONLY] name SET
DISTRIBUTED BY (column, [ ... ] )
| DISTRIBUTED RANDOMLY
| WITH (REORGANIZE=true|false)
ALTER TABLE [ONLY] name action [, ... ]
ALTER TABLE [ONLY] name action [, ... ]
ALTER TABLE name
[ ALTER PARTITION { partition_name | FOR (RANK(number))
| FOR (value) } partition_action [...] ]
partition_action
```

更多信息请参见ALTER TABLE。

DROP DATABASE

删除数据库。

DROP DATABASE [IF EXISTS] name

更多信息请参见DROP DATABASE。

DROP INDEX

删除索引。

DROP INDEX [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

更多信息请参见DROP INDEX。

DROP SCHEMA

删除SCHEMA。

DROP SCHEMA [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

更多信息请参见DROP SCHEMA。

DROP ROLE

删除数据库角色。

```
DROP ROLE [IF EXISTS] name [, ...]
```

更多信息请参见DROP ROLE。

DROP SEQUENCE

删除序列。

DROP SEQUENCE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

更多信息请参见DROP SEQUENCE。

DROP USER

删除数据库用户。

DROP USER [IF EXISTS] name [, ...]

更多信息请参见DROP USER。

DROP VIEW

删除视图。

DROP VIEW [IF EXISTS] name [, ...] [CASCADE | RESTRICT]

更多信息请参见DROP VIEW。

DROP TABLE

删除表。

```
DROP TABLE [IF EXISTS] name [, ...] [CASCADE | RESTRICT]
```

更多信息请参见DROP TABLE。

BEGIN

启动事务块。

```
BEGIN [WORK | TRANSACTION] [transaction_mode]
    [READ ONLY | READ WRITE]
```

更多信息请参见BEGIN。

CHECKPOINT

强制事务记录检查点。

CHECKPOINT

更多信息请参见CHECKPOINT。

CLOSE

关闭游标。

CLOSE cursor_name

更多信息请参见CLOSE。

CLUSTER

根据索引对磁盘上的堆存储表,重新排序。

```
CLUSTER indexname ON tablename
CLUSTER tablename
CLUSTER
```

更多信息请参见CLUSTER。

COMMENT

定义或修改对象的注释。

智能数据构建与管理 Dat aphin

COMMENT ON { TABLE object name | COLUMN table name.column name | AGGREGATE agg_name (agg_type [, ...]) | CAST (sourcetype AS targettype) | CONSTRAINT constraint name ON table name | CONVERSION object name | DATABASE object name | DOMAIN object name | FILESPACE object name | FUNCTION func name ([[argmode] [argname] argtype [, ...]]) | INDEX object name | LARGE OBJECT large object oid | OPERATOR op (leftoperand_type, rightoperand_type) | OPERATOR CLASS object name USING index method | [PROCEDURAL] LANGUAGE object_name | RESOURCE QUEUE object name | ROLE object name | RULE rule name ON table name | SCHEMA object name | SEQUENCE object name | TABLESPACE object_name | TRIGGER trigger name ON table name | TYPE object name | VIEW object name } IS 'text'

更多信息请参见COMMENT。

COMMIT

提交当前事务。

COMMIT [WORK | TRANSACTION]

更多信息请参见COMMIT。

DEALLOCATE

取消分配预编译的语句。

DEALLOCATE [PREPARE] name

更多信息请参见DEALLOCATE。

DECLARE

定义游标。

```
DECLARE name [BINARY] [INSENSITIVE] [NO SCROLL] CURSOR
[{WITH | WITHOUT} HOLD]
FOR query [FOR READ ONLY]
```

更多信息请参见DECLARE。

DELETE

从表中删除行。

```
DELETE FROM [ONLY] table [[AS] alias]
[USING usinglist]
[WHERE condition | WHERE CURRENT OF cursor_name ]
```

更多信息请参见DELETE。

END

结束当前事务。

END [WORK | TRANSACTION]

更多信息请参见END。

EXPLAIN

展示语句的查询计划。

EXPLAIN [ANALYZE] [VERBOSE] statement

更多信息请参见EXPLAIN。

FETCH

通过游标获取查询结果的行。

FETCH [forward_direction { FROM | IN }] cursorname

更多信息请参见FETCH。

LOAD

加载或重新加载共享库文件。

LOAD 'filename'

更多信息请参见 LOAD。

LOCK

锁住一张表。

LOCK [TABLE] name [, ...] [IN lockmode MODE] [NOWAIT]

更多信息请参见LOCK。

MOVE

放置游标。

MOVE [forward_direction {FROM | IN}] cursorname

更多信息请参见MOVE。

PREPARE

准备一个执行的语句。

PREPARE name [(datatype [, \dots])] AS statement

更多信息请参见PREPARE。

REASSIGN OWNED

修改数据库角色。

REASSIGN OWNED BY old role [, ...] TO new role

更多信息请参见REASSIGN OWNED。

REINDEX

重新构建索引。

REINDEX {INDEX | TABLE | DATABASE | SYSTEM} name

更多信息请参见REINDEX。

RELEASE SAVEPOINT

销毁定义过的SAVEPOINT。

RELEASE [SAVEPOINT] savepoint_name

更多信息请参见RELEASE SAVEPOINT。

RESET

恢复系统配置参数的值为默认值。

```
RESET configuration_parameter RESET ALL
```

更多信息请参见RESET。

ROLLBACK

中止当前事务。

ROLLBACK [WORK | TRANSACTION]

更多信息请参见ROLLBACK。

ROLLBACK TO SAVEPOINT

将当前事务回滚到某个SAVEPOINT。

ROLLBACK [WORK | TRANSACTION] TO [SAVEPOINT] savepoint_name

更多信息请参见ROLLBACK TO SAVEPOINT。

SAVEPOINT

在当前事务定义一个新的SAVEPOINT。

SAVEPOINT savepoint_name

更多信息请参见SAVEPOINT。

SELECT INTO

从查询结果中定义一个新的表。

```
[ WITH with_query [, ...] ]
SELECT [ALL | DISTINCT [ON ( expression [, ...] )]]
 * | expression [AS output_name] [, ...]
INTO [TEMPORARY | TEMP] [TABLE] new_table
[FROM from_item [, ...]]
[WHERE condition]
[GROUP BY expression [, ...]]
[HAVING condition [, ...]]
[{UNION | INTERSECT | EXCEPT} [ALL] select]
[ORDER BY expression [ASC | DESC | USING operator] [NULLS {FIRST | LAST}] [, ...]]
[LIMIT {count | ALL}]
[OFFSET start]
[FOR {UPDATE | SHARE} [OF table_name [, ...]] [NOWAIT]
[...]]
```

更多信息请参见SELECT INTO。

SET

改变数据库配置参数的值。

```
SET [SESSION | LOCAL] configuration_parameter {TO | =} value |
    'value' | DEFAULT}
SET [SESSION | LOCAL] TIME ZONE {timezone | LOCAL | DEFAULT}
```

更多信息请参见SET。

SET ROLE

设置当前会话当前角色的标识符。

```
SET [SESSION | LOCAL] ROLE rolename
SET [SESSION | LOCAL] ROLE NONE
RESET ROLE
```

更多信息请参见SET ROLE。

SET SESSION AUTHORIZATION

设置会话角色标识符和当前会话当前角色的标识符。

SET [SESSION | LOCAL] SESSION AUTHORIZATION rolename SET [SESSION | LOCAL] SESSION AUTHORIZATION DEFAULT RESET SESSION AUTHORIZATION

更多信息请参见SET SESSION AUTHORIZATION。

SET TRANSACTION

设置当前事务的特征。

SET TRANSACTION [transaction_mode] [READ ONLY | READ WRITE] SET SESSION CHARACTERISTICS AS TRANSACTION transaction_mode [READ ONLY | READ WRITE]

更多信息请参见SET TRANSACTION。

SHOW

显示当前系统配置参数的值。

```
SHOW configuration_parameter SHOW ALL
```

更多信息请参见SHOW。

START TRANSACTION

开始事务块。

```
START TRANSACTION [SERIALIZABLE | READ COMMITTED | READ UNCOMMITTED]
[READ WRITE | READ ONLY]
```

更多信息请参见START TRANSACTION。

TRUNCATE

清空表的所有行。

```
TRUNCATE [TABLE] name [, ...] [CASCADE | RESTRICT]
```

更多信息请参见TRUNCATE。

UPDATE

更新表的行。

```
UPDATE [ONLY] table [[AS] alias]
SET {column = {expression | DEFAULT} |
(column [, ...]) = ({expression | DEFAULT} [, ...]) } [, ...]
[FROM fromlist]
[WHERE condition | WHERE CURRENT OF cursor name ]
```

更多信息请参见UPDATE。

VACUUM

垃圾收集和选择性分析数据库。

VACUUM [FULL] [FREEZE] [VERBOSE] [table] VACUUM [FULL] [FREEZE] [VERBOSE] ANALYZE [table [(column [, ...])]]

更多信息请参见VACUUM。

VALUES

计算一组行。

```
VALUES ( expression [, ...] ) [, ...]
[ORDER BY sort_expression [ASC | DESC | USING operator] [, ...]]
[LIMIT {count | ALL}] [OFFSET start]
```

更多信息请参见VALUES。

3.7. 新建通用任务

3.7.1. 创建、配置并提交离线计算模版

Dataphin支持创建Shell和Python任务的离线计算模版。后续在开发Shell和Python任务时,可以引用该离线 计算模版以提高任务的研发效率。本文为您介绍如何创建、配置并提交离线计算模版。

背景信息

当多个Shell或Python任务的代码逻辑相似,只有部分配置项或输入参数不同时,您可以将代码写入到离线 计算模版中,并将不同的配置项和输入参数定义为模版变量参数。后续任务引用了该离线计算模版后,可以 给离线计算模版中的模版变量参数赋值,以实现代码公共逻辑的单点维护和快速复用,提升了任务代码研发 的效率。

通常每个任务有独立的运行资源,大批量任务同时运行时会导致运行资源占用率高,影响任务运行的并发性。Dataphin支持引用了同一个离线计算模板的多个任务可以共享运行资源,以保障其他任务正常运行。您需要为离线计算模板开启共享运行资源开关,后续引用了该模板的多个任务可以共享运行在同一个资源上, 其他任务就可以使用Dataphin剩余的资源。如何开启离线计算模板的共享运行资源开关,请参见步骤三:配置离线计算模版。

使用限制

- 仅支持创建Shell和Python任务的离线计算模版。
- 支持超级管理员、项目管理员和开发者角色的用户创建、配置并提交离线计算模版。
 如何获取项目管理员和开发者角色,请参见添加项目成员。
- 仅支持超级管理员角色的用户开启离线计算模板的共享运行资源开关。

操作流程

1. 步骤一: 创建离线计算模版

创建离线计算模版任务。

2. 步骤二:开发离线计算模版的代码

开发离线计算模版的代码。

3. 步骤三: 配置离线计算模版

配置离线计算模版的模版变量参数及运行资源。

4. 步骤四:保存、调试及提交离线计算模版

保存、调试和提交离线计算模版后,任务才可以引用该离线计算模版。

步骤一: 创建离线计算模版

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏的**研发**。 默认进入数据**开发**页面。
- 4. 在数据开发页面,按照下图操作指引,进入新建离线计算模版。

	开发≓	发布	运维	权限						Q、全局撤索		1 🕅	🖞 🕹	
DEV hadcopilintelist, D v2_9_0_auto_08	欢迎使用D	ataphin!											Ĵ	我的文件
	86 86 5	▶规范控义》		<u>建度</u> 1 ③		<u>业</u> 得过程 ■ ◎		} ₩:	▲ 子描标 〕 ③	2	29項定 ■ ◎	itte الله الله	 生道标 ○ 	易性查看
Ξ]3 - Σ		● 数据建模 ≫) 22448 2 ⊗	
		智能计算			2	运转调查 进行及推 时间正确	≫ 户逻辑表和 制序地生产	欺握处理产生的数据付 数据。	海,以确保所有任	時都神技探				
						计算任	5					同步任务		
		■ 任务研发 ≫			Sul			SHELL				Pipeline 🗐 🖽		
					MR			PYTHON						
			((文)#	31月 ≫			Hado	↓ op@2.#X ⊙		です。 注意語語 1 ②				R

5. 在新建离线计算模版对话框, 配置参数。

新建离线计算模版						
* 模版名称	请输入模板名称, 新建后不可修改					
*节点类型	请选择节点类型	~				
* 选择目录	离线计算模板	×				
描述	请输入描述, 128字符以内					
		0/128				
ў 提示: 养	所建模板输入代码后需要手动解析配置任务参数					
	取消	确定				

参数	描述
模版名称	 命名规则如下: 交持任意字符。 全局唯一。 不能超过64个字符。
节点类型	当前版本仅支持Python和Shell任务。根据您后续开发的任务类型进行选择,例如后 续您需要基于离线计算模版开发Shell任务,则节点类型选择为Shell。
选择目录	<image/>
描述	填写对离线计算模版简单的描述。

6. 单击确定,完成离线计算模版。

完成离线计算模版创建后,默认进入离线计算模版的代码开发页面。开发代码请参见步骤二:开发离线 计算模版的代码。

步骤二:开发离线计算模版的代码

在离线计算模版的代码开发页面,进行开发相应语言(Python或Shell)的代码。

≡	Dataphin · 研发	开发;异 发布 运维	权限	Q、 全局撤去	1 國 🖞 🗸	<u>-</u> ८	
	// hadoop@itetetetete/ ↓ v2_9_0_auto_08♀∨ Prod Prod Prod Prod Prod Prod Prod Prod Prod Prod	wendang_lest1 X 极质信息学 版本信息 引用信息		■通訊の際業品を	ich: 😡 #Eilin 🔒	:8 8 √ ©	我的文件
* 	televice televice	i ros edec).					- 属性重都
	I) dimode_1116					Х*Ф	

代码说明如下:

- 支持的语法与Python、Shell的语法一致。Python语法请参见Python官网。
- 支持定义模版变量参数,格式为 @@{模板变量参数名称} ,模版变量参数名称只能包含字母、数字和下划 线(_)且只能以字母开头。例如 @@{variable} 。

Shell语言的代码示例如下。

```
#!/bin/bash
# 本示例用于检测某个FTP上的文件是否已经上传完成。
# ftpserver作为模板变量参数,后续可以为其配置默认值,这样每个引用该模板的任务共用同一个FTP服务器,也可
以在任务上指定其他服务器。
ftp ip=00{ftpserver}
# username、password作为模板变量参数,后可以为参数开启加密,防止用户名密码泄露。
username=00{username}
password=@@{password}
# 检测FTP文件是否上传完成。
check ftp file() {
filename=$1
ftp -niv <<EOF
open $ftp ip # $ftp ip是Shell局部变量。
user $username $password
cd /file
ls $filename
bye
EOF
}
cnt=0
while (( $cnt <= 300 ))
do
 if check ftp file ${filename} | grep ${filename} # ${filename}为任务参数。
 then
  echo "文件已上传完成"
   exit O
 else
   echo "文件未上传完成"
  cnt=$(( $cnt + 1 ))
  sleep(60s)
 fi
done
echo "超过300分钟, 文件仍未上传完成, 异常退出"
exit 1
```

完成离线计算模版的代码开发后,需要配置该离线计算模版的模版变量参数及运行资源。具体操作请参见<mark>步</mark>骤三:配置离线计算模版。

步骤三: 配置离线计算模版

- 1. 在离线计算模版的代码开发页面,单击页面上方的模版信息。
- 2. 在模版信息面板中, 配置参数。

模版	信息		Х
基本條	息		
档	誠名称	wendang_test1	
	负责人	Recommendation (Second Second	
	描述	请输入对模版的描述内容	0/128
运行費	記置		
* 共	淳运行资	遊源① 🔵 开启 💿 关闭	
参数配	記置		刷新参数
参	数配置		
	参数	效名称: abc	
	* 参数	放说明 : 请输入对该参数的说明信息	
	剽	认值 : 请输入对该参数的默认值,可在任务中修改	
	参数加	1密①: 〇 开启 💿 关闭	
		取消	确定

模版信息面板中的基本信息区域的模版名称和负责人的参数不支持修改;运行配置用于开启离线计算 模板的共享运行资源开关;参数配置用于配置代码中的模版变量参数的说明、设定默认值及是否加密, 如果您修改了代码中的模版变量参数,您可以单击刷新参数后,在模版信息面板查看最新模版变量参数。

区域	参数	描述
基本信息	描述	填写离线计算模版的简单描述。
运行配置	共享运行资源	开启共享运行资源后,引用该模版的任务可以共享运行资源。该操 作仅限超级管理员执行,如果您需要开启,请联系超级管理员。
	参数说明	填写代码中的参数说明,便于开发者理解。
	默认值	为代码中的参数赋值。后续您可以在引用了该模版的任务中修改该 参数值,运行任务后即可生效。
参数配置	<u>置</u> 参数加密	开启参数加密后,参数默认值将以密文形式存储,以保护敏感数据。后续在引用该模版的任务中也是无法查看明文的参数默认值, 任务运行时Dataphin会自动解密参数默认值。 关闭参数加密后,已配置的参数默认值将自动清空。

3. 单击确定,完成离线计算模版的配置。
 完成离线计算模版配置后,需要保存、调试和提交模版代码。具体操作请参见步骤四:保存、调试及提

交离线计算模版。

步骤四:保存、调试及提交离线计算模版

1. 在离线计算模版的代码开发页面,按照下图操作指引,保存并调试离线计算模版。

模版(🖳 調試 🔿 刷新 😡 規范	iin 🔒 🖪 ⊄ 🦁
1				
2				
З				
4				
5	ftp_ip-@@{ftpserver}			
6				
7	# username、password作为模板变量参数,后可以为	参数开启加密,防止用户名密码泄露。		
8	username=@@{username}	Y		
9	password-@@{password}			
10				
11	# 检测FTP文件是否上传元成。			
12	check_ttp_tile() {	* ftpserver: 请输入默认值		
1.5	filename=pi			
14	TUP -NIV KKEUF	* password: 请插入款认值		
16	usen Susenname Snassword	 38から10521月 		
17	cd /file	• usemane. Intel ANAL		
18	ls \$filename	* flonamo: 海给入时注意		
19	hve	· Incidente · Ing (E) / (A/A/12)		
20	EOF			
21	}			
22		取消 确定 4		
23	cnt=0			

在参数填写对话框,您需要为代码中的模版变量参数赋值。本文的代码示例中的模版变量参数为abc和 bizdate,因此需要为其赋值后,才能调试开发的代码是否可以正常运行。

2. 调试成功后,按照下图操作指引,提交离线计算模版。

模版信息 ◇ 版本信息 引用信息	🖳 调试 🖒 刷新 싎 格式化 😪 规范说明 🔒 🛛 🗖 🛛
1 echo "@@{abc}" 2 echo "\${bizdate}"	*提交备注 2
	请输入此次提交的相关备注说明,如变更信息 0/128
	填写备注以说明此次提交的变更内容 当前项目,提交成功后,请前往待发布列表,以发布对象至生产环境。 3 取消 确定并提交

提交成功后,您可以按照下图操作指引,查看离线计算模版的版本信息。

开发≠	发布	运维	权限	Q 全局捜索			ti di	Ļ	2	
Sh wend	ang_test1	×							÷	串
模版信息	、 版本信息			版本信息						X
1 ec	ho "@@{abc}" ho "\${bizdat	e}"				2				
		-,		版本号	提交人	提交时间	备注	生产	操作	
				v1	35.pe24d=10(2021-12-22 19:59:58	ok	⊘ 2		

同时在版本信息面板您也可以进行不同版本的代码对比及查看代码:

• 按照下图操作指引,进行版本对比。版本对比包括代码对比和模版变量参数对比。

版信息∳ 版本信 1 echo "@@{ab o	息 引用信息	2	版本信息					×
2 echo "\${bi	版本对比		×		揭交时间	备注	生产	副作
	代码对比 模版参数对比			nin(2021-12-22 19:59:58	ok	0e [08
	编辑中版本	对比版本号: 1						1
	1 echo "@@{abc}" 2 echo "\${bizdate}"	1 echo "@@{abc}" 2 echo "\${bizdate}"						
			- 5					
		复制对比版本代码	(K) 关闭	3				

○ 按照下图操作指引,查看代码。

他が信号や 版本信号 31時信号 1 echo " <mark>論(cho</mark>)"	版本信息				×
2 echo "\$(bizdate)"	版本号	提交人	提交时间	备注	生产 操作
	v1	lage Advit (2021-12-22 19:59:58	ok	se m <mark>e</mark>

后续步骤

基于已创建的离线计算模版创建对应的任务,具体操作请参见新建Shell任务或创建Python计算任务。后续 您也可以在离线计算模版的代码开发页面,查看离线计算模版被任务引用的情况。

極振信目令 版本信見 到用信見 <mark>1</mark> 1 echo "発信abc)"	引用信息	2	3 🖂
2 echo "\${bizdate}"	引用此構板的任务	开发者	操作
	gsg	Second Second Second Second	⊟

3.7.2. 新建Shell任务

本文为您介绍如何基于Dat aphin构建Shell类型的离线计算任务。

前提条件

完成上传JSON资源包,详情请参见新建资源。

背景信息

Shell任务中引用到的JSON资源包需提前创建,因此需要您在资源管理中上传JSON资源包,然后在Shell任务中引用。本文以MySQL数据库为例为您介绍,请您下载JSON文件。

新建Shell任务的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的数据研发。

- 单击快速进入研发项目区域的Dev或Basic项目,选择开发数据的项目空间。
 - ? 说明
 - 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3。
 - 如果您通过**快速进入研发项目**进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。
- 4. 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的项目空间 (Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

- 5. 进入计算任务页面。
 - i. 在数据**开发**页面,单击**数据处理**页签。 如果进入数据**开发**页面后,系统默认进入**数据处理**页签,则无需再次单击**数据处理**页签。
 - ⅲ. 在**数据处理**页签,单击左侧导航栏<mark>砅</mark>计算任务图标。
- 6. 在计算任务页面, 进入新建文件对话框。
 - 单击计算任务后的 图标,选择通用 脚本 > SHELL。
 - 单击项目名称后的 图标,选择数据处理 > 计算任务 > SHELL。
 - 在开发首页, 单击SHELL后的 · 图标。
- 7. 编写并运行代码。

i. 在新建文件对话框, 配置参数。

新建文件						×
*	请输入仕务名称					
* 调度类型	● 周期性节点	○ 手动节点				
描述	请输入任务描述					
				h		
引用模版①	请选择目录			\sim	□查看计算模版 >>	
选择目录	计算任务			\sim		
				取消	确定	
			SUPE			

参数	描述
名称	填写离线计算任务的名称,例如DataX。
调度类型	选择任务的调度类型。调度类型包括: 周期性节点:自动参与系统的周期性调度。 手动节点:需要手动触发任务的运行。
描述	填写对任务的简单描述。
引用模板	如果开发的Shell任务与已有的代码模板相似,您可以引用离线代码模板,提升 代码研发的效率。
选择目录	选择离线计算任务的目录。

ii. 单击确定。

iii. 在代码编写页面,编写Shell离线计算任务的代码。

代码如下。

```
@resource_reference{"datax.json"}
python $DATAX_HOME/bin/datax.py datax.json #Dataphin系统已内置DataX的安装目录为DATAX_H
OME/bin/datax.pyo
```

其中, resource_reference{}用于调用已上传的datax.json资源文件。

iv. 检查代码。

代码编写完成后,单击**预编译**,系统帮助您检查编写的SQL代码的语法。

如果预编译失败,您可以单击页面上方的**格式化**,系统自动帮助您调整语法格式。调整完格式,您 可以单击**刷新**,刷新页面上的代码。

- v. 单击页面右上角的执行,运行代码。
- 8. 配置调度参数。

- 如果离线计算任务的调度类型为周期性节点,则需要配置调度参数,详情请参见调度配置。
 - 如果离线计算任务的调度类型为手动节点,需要手动触发任务的调度。
- 9. 在代码编写页面,保存、提交和发布Shell任务。
 - i. 单击页面右上方的 图标, 保存代码。
 - ii. 单击页面右上方的 图标, 提交代码。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交。
 - v. 发布Shell任务至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布Shell任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功的Shell任务,即可参与生产环境的调度。

3.7.3. 新建Virtual任务

虚拟节点是租户配置调度参数时可选的依赖节点。本文为您介绍如何基于Dat aphin构建Virt ual类型的虚拟节点。

背景信息

在开始构建数据模型前,您可以创建基于该数据模型的虚拟节点,以便于该数据模型中其他任务配置调度参数时依赖该节点。

新建Virtual任务的权限限制,详情请参见数据开发权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的数据研发。
 - 单击快速进入研发项目区域的Dev或Basic项目,选择开发数据的项目空间。

? 说明

- 如果您通过**快速开始相关工作**进入数据开发模块,则跳过步骤3。
- 如果您通过**快速进入研发项目**进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。
- 在数据开发页面,单击项目名称后的☑图标后,单击Dev或Basic页签,选择数据开发的项目空间
 (Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

- 5. 进入**计算任务**页面。
 - i. 在数据开发页面, 单击数据处理页签。

如果进入数据开发页面后,系统默认进入数据处理页签,则无需再次单击数据处理页签。

- ii. 在数据处理页签,单击左侧导航栏 🐼 计算任务图标。
- 6. 在计算任务页面, 进入新建Virtual任务的新建文件对话框。
 - 单击计算任务后的 图标,选择通用脚本 > VIRTUAL。
 - 单击项目名称后的 图标,选择数据处理 > 计算任务 > VIRTUAL。
 - 在开发首页,单击VIRT UAL后的 ·· 图标。
- 7. 编写并运行代码。
 - i. 在新建文件对话框, 配置参数。

新建文件				2	×
* 名称	Virtual				
* 调度类型	● 周期性节点	○ 手动节点			
描述	Virtual				//
选择目录	代码管理			~	
		[取消	确定	

参数	描述
名称	填写离线计算任务的名称,例如Virtual。
调度类型	选择任务的调度类型: 周期性节点:自动参与系统的周期性调度。 手动节点:需要手动触发任务的运行。
描述	填写对任务的简单描述。
选择目录	选择离线计算任务的目录。

- ii. 单击确定。
- iii. 检查代码。

代码编写完成后,单击**预编译**,系统帮助您检查编写的SQL代码的语法。

如果预编译失败,您可以单击页面上方的**格式化**,系统自动帮助您调整语法格式。调整完格式,您 可以单击**刷新**,刷新页面上的代码。

- iv. 在代码编写页面, 编写SQL代码。
- 8. 配置调度参数。
 - 如果离线计算任务的调度类型为**周期性节点**,则需要配置调度参数,详情请参见<mark>调度配置</mark>。
- 如果离线计算任务的调度类型为**手动节点**,需要手动触发任务的调度。
- 9. 在代码编写页面,保存、提交和发布Virtual任务。
 - i. 单击页面右上方的 图标, 保存代码。
 - ii. 单击页面右上方的 图标, 提交代码。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交。
 - v. 发布Virtual任务至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布Virtual任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则提交成功后的Virtual任务,即可参与生产环境的调度。

3.7.4. 创建Python计算任务

为了满足丰富的Python应用场景,Dataphin支持创建可以使用Python语法的Python计算任务。本文为您介 绍如何基于Dataphin构建Python计算任务和定义Python UDF。

使用限制

Python 3.7无法向下兼容python 2.7,无法直接升级历史的python 2任务。

2.9.3版本后, Dataphin默认支持研发Python 3.7计算任务。系统仅支持草稿状态的Python任务修改版本。

注意事项

Python UDF在SQL任务中执行时,需要指定特定执行版本的兼容。系统默认执行Python 2.7版本,如果需要执行Python 3.7,则执行以下命令,在任务级别开启版本Python 3.7。

```
set odps.isolation.session.enable=true;
set odps.sql.python.version=cp37;
```

背景信息

Python 3.7更能满足更多样化的大数据处理的诉求,例如 list.clear() 。Python 2.7和Python 3.7的更多 信息,请参见Python。

前提条件

如果您需要定义Python UDF,则需要上传Python文件,详情请参见新建资源。

操作步骤

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。
- 4. 按照下图指引,进入新建文件对话框。

≡	Dataphin 研发	.1211 () • ×	集成 开发发 发布 运维	Q 全局搜索	e le té 🧟 💿
	H\$1465 O 🖸 🖸				8
	TIQ #XXX88# 4 MAX_COMF	PUTE_SQL			
		PUTE_MR		▲ 統一還進 >>	■ 数据处理 ≫
	> III SPARK_JAP	R_ON_MAX_COMPUTE			MaxCompute任务
	SHELL				
					sul .
▲		1885A + ©	() ()		提 交
۲	1 47 (F)			反布管理	MAX_COMPUT E SPARK_JAR_OE
~	· Di stant (r.t.			all Diff 利益(1) 建分面的任务和政策。 以更好地验证数据可用性,同时方便进 行生命地理处则全部理	SHELL 🖻 PYTHON 🖻
				11	VIRTUAL 🗈
	li și jeni și e	*			ELINIK CLER
	L GUING		~	发布	1 CHARLESS
	3	原子描标 → ◎	业务限定	4	SQL
0	 In Solid 				
B	 In the part of the second secon			269	DATASTREAM 📄
	 In traffi 				数据表
	 In subscription 			活在構成	元表 : 镜像表 :
-	 In signal, 3 (3) 	本 亦沒語:	他們可以是	运17次年产业建筑中1853年124日 王1385 提任务,实践中有任务都有按照时间 工作方,实践中方型	
	 (a) (b) (b) 	± ⊙	+ ©		数据集成
•	> 10 (1440)709 (4)				Pipeline 由 整座迁移 由
	 In second pill 				
۵	* 10 million (1999) (2)	■ 研发工具箱 ≫			
ò		函数 🕒 💿	澄源	Ē	

您也可以通过以下方式进入**新建文件**对话框:

○ 单击项目名称后的 图标,选择数据处理 > 计算任务 > Python。

	Dataphin · 研发	开发≠	发布 运维 权限					
DEV	DEV 2_7测试							
	□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □							
规》	范建模 数据处理 即席查规范定)	۷ >						
<	计算任务	2 3						
	▼ Q、输入关键字 数据处理	型 > 计算任务 >	MAX_COMPUTE_SQL					
U	▶ 00-nj_test1215 (5)	同时任务	MAX_COMPUTE_MR					
datap	▶ 🔲 0512 (1)	表〉	SPARK_JAR_ON_MAX_COMPUTE					
	▶ a_test (7)	自定义源	SHELL					
2	▶ chen (5)	代码模板	PYTHON					
	▶ i junze_test (1)	函数 >	VIRTUAL					
	▶ 📄 moyi_0716moyi_0716 (1)	资源	FLINK_SQL					
₽ō	▶ 📄 moyiAutoTest (35)		FLINK TEMPLATE SOL					
_	▶ MoyiAutoTest7 (18)							
~	▶ nj_test (1)		FLINK_DATASTREAM					

○ 在开发首页,单击任务研发区域的Python后的⊡图标。

	Dataphin · 研发	开发≠ 发布	运维 权限	Q、全局搜索	日間の
DEV	2_738id sh_stream_dem				
规	克建模 数据处理 即席查询			计算任务	
(>	计算任务 🖒 🛆 🖻 🖃		MAXC_SQL +	VIRTUAL +	FLINK_SQL +
ວ	▼ Q、输入关键字	□ 任务研发 》	MAXC_SPARK_JAR 🗄	SHELL +	FLINK_TEMPLATE_SQL
	▶ 00-nj_test1215 (5)		MAXC_MR +	PYTHON 🕒	FLINK_DATASTREAM

5. 编写并运行代码。

i. 在新建文件对话框, 配置参数。

新建 PYTH	ON 任务	X
* 任务名称	请输入任务名称	
* 调度类型	● 周期任务 ○ 手动任务	
引用模版③	请选择目录	∨ 「「「查看计算模版 >>
选择目录	代码管理	×
描述	请输入任务描述	
		"
		取消 确定

参数	描述
任务名称	填写代码任务的名称,例如Python。
调度类型	选择任务的调度类型。调度类型包括: 周期性节点,自动参与系统的周期性调度。 手动节点,需要手动触发任务的运行。
引用模板	如果开发的Shell任务与已有的代码模板相似,您可以引用离线代码模板,提升 代码研发的效率。
选择目录	选择代码任务的目录。
描述	填写对任务的简单描述。

- ii. 单击确定。
- iii. 在Python任务代码编写页面,选择需要创建的Python任务版本后,编写代码。

🛛 🕅	刀标 🗙 📴 aa 🛛 🗙	÷
🕑 运行	新 🗋 保存任务 🗘 提交 🕢 去返布 🔍 去运業 💿 🔒 😁 Python 3.7 \vee	
1	#!/usr/bin/env python I	
2	# -*- coding: utf-8 -*-	
3		
4	# Author:	
5	# Created Time: 2022-04-12 17:50:17	
6		
7	#	
8		
•		

定义Python UDF时,需要在代码中添加 import Python文件名 语句,导入上传的Python文件。

开发Python 3.7计算任务时,需要根据业务场景安装如下资源包。以下安装包已经预制至系统内, 您在开发代码过程中,只需要在代码开始部分添加 import 资源包名 语句即可,例如 import configparser 。

资源包	版本	适用场景
configparser	>=3.5.0	读取配置文件
DateTime	无	数据处理
jumpssh	无	使用跳板机连接服务器
mysql-connector-python	>=8.0.11	连接并操作MySQL
numpy	无	算法基础处理
pandas	无	算法基础处理
psycopg2	>=2.7.4	连接操作PostgreSQL
pyhs2	>=0.6.0	连接并操作HDFS
pyodps	>=0.7.16	ODPS操作,适用ODPS
requests	>=2.4.0	算法基础处理
scikit-learn	无	算法基础处理
scipy	无	算法基础处理
setuptools	>=3.0	Pthon基础功能库
yarn-api-client	>=0.2.3	Yarn-api客户端
Matplotlib	无	算法基础处理

需要选择代码示例如下。

```
#!/usr/bin/Python
# -*- coding: latin-1 -*-
import time
import datetime
import base64
import hashlib
import json
import sys
import csv
from odps import ODPS
def main():
    print "Hello World"
```

建议您在Python文件前两行对编码进行主动注释 ,防止执行代码时使用系统编码,导致执行结果 报错。

iv. 单击页面右上角的执行,运行代码。

- 6. (可选)配置调度参数。
 - 如果离线计算任务的调度类型为周期性节点,则需要配置调度参数,详情请参见调度配置。
 - 如果离线计算任务的调度类型为手动节点,需要手动触发任务的调度。
- 7. 按照下图指引,保存并提交Python任务。

[1] [1]	^{1标} <mark>1</mark> 2 ×	sh aa		>	<			
🕑 运行	行 🗋 保存任务 🔂 提交			0		Θ	Python 3.7	
1	<pre>#!/usr/bin/env python</pre>							
2	# -*- coding: utf-8 -	* _						
3	****	*********	******	#####	#			
4	# Author: Superhand.	baytan Adarbay	datapılar)					
5	# Created Time: 2022-	04-12 17:50):17					
6	****			#####	#			
7	#				=			
8								

- 8. 如果项目空间的模式为Dev-Prod,则发布Python任务至生产环境。具体操作,请参见管理发布任务。
- 9. 查看调度任务。具体操作,请参见脚本任务或手动任务。

3.7.5. Dataphin PyODPS

Dataphin PyODPS是基于MaxCompute PyODPS开发的Dataphin Python版本的SDK。在Dataphin中使用 Python时,您可以使用Dataphin PyODPS安全地访问Dataphin项目内存储在MaxCompute的数据。

背景信息

Dataphin PyODPS依赖于MaxCompute的 PyODPS。

前提条件

在使用Dataphin PyODPS前,请确认您已将Dataphin计算引擎设置为MaxCompute。具体操作请参见设置 Dataphin实例的计算引擎。

使用限制

仅计算引擎为MaxCompute的Dataphin支持Dataphin PyODPS。

获取待操作数据对象的权限

- 申请逻辑表权限,详情请参见管理逻辑表权限。
- 申请物理表权限,详情请参见管理物理表权限。
- 上传资源到Python任务所在的项目,并提交或发布至生产环境。具体操作,请参见新建资源。
- 需要在Python任务所在的项目内创建函数,并提交或发布至生产环境。具体操作,请参见Python UDF或Java UDF最佳实践。

使用说明

Dataphin PyODPS安装包已内置在Dataphin中,您可以在Dataphin中直接调用。详细说明如下:

- 当前, Dataphin支持的PyODPS包括项目空间、表、SQL、资源和函数。MaxCompute的其他PyODPS暂时 不支持直接在Dataphin上调用。
- 在Dataphin上调用PyODPS时,与直接在MaxCompute上调用,命令执行上基本一致。详细说明如下表所示。

PyODPS	是否一致	备注
SQL	不一致	在Dataphin中使用PyODPS访问项目内的表时,需要在表名称前加 项目或业务板块的信息,详细说明如下: • 访问物理表的SQL语句格式为 \${Dataphin项目英文名}.物理 表名 。 例如,Dataphin中有项目ProjectA,则Dataphin自动生成项目变 量\${ProjectA}。ProjectA内有一张物理表TableA,则您可以使 用 \${ProjectA}.TableA 来引用这张表。动态变 量\${ProjectA}在开发环境执行时默认被替换为Project A_dev,在 生产环境执行时默认被替换为ProjectA。当然您也可以直接指定 开发项目 (ProjectA_dev) 或生产项目 (ProjectA)。 • 访问逻辑表的SQL语句格式为 \${Dataphin业务板块英文名}. 逻辑表名 。 例如,Dataphin中有业务板块LD_BizUnitA,则Dataphin自动生 成业务板块变量\${LD_BizUnit A}。业务板块LD_BizUnitA内有一张 逻辑表DimTableA。您可以使用 \${LD_BizUnitA}.DimTable A 来引用这张表,该变量在开发环境执行时默认被替换 为LD_BizUnitA_dev,在生产环境执行时默认被替换 为LD_BizUnitA。当然您也可以直接指定开发项目 (LD_BizUnitA_dev)或生产项目(LD_BizUnitA)。
表		不支持跨Dataphin项目访问存储在MaxCompute的数据对象,也就
项目空间		定文持功回当則Pytnon仕务所在的Dataphin项目内存储在 MaxCompute的数据对象。
资源	一致	例如,当前Python任务所在的项目英文名为ProjectB,如果您执行 odps.get_project('Test_Project') ,则会报以下错误。
函数		Exception: Dataphin odps client auth check failed, not allow to operate project Test_Project

操作过程中有无权限的报错时,您需要获取对应数据对象的权限。具体操作,请参见获取待操作数据对象的 权限。

使用示例

```
from dataphin import odps
odps.execute sql("select * from ${[业务板块]}.[逻辑表名]")
```

3.8. 创建并配置机器学习任务

机器学习任务用于调用您在机器学习平台中构建的任务,并按照节点配置进行调度生产。本文为您介绍如何 创建并配置机器任务。

功能上线时间

- 华东2(上海)、华北2(北京): 2022年03月03日
- 华南1(深圳)、华东1(杭州): 2022年03月08日

前提条件

在开始执行操作前,请确认您已满足以下条件:

- 基于机器学习的PAI-Designer完成工作流的创建,具体操作请参见创建及管理工作流。
- 项目的机器学习开关已打开,具体操作请参见创建Basic项目、创建Prod和Dev项目。

信息设置: \	v29_UI_Automatic_project(v29	9_UI_Autom…	已选: Dev-Pro	d 模式 🛈			×
名称设置					命名空间		
* 公用名称	v29_UI_Automatic_project				业务板块	LD_v29_UI_Automatic_project	
* 公用英文名	v29_UI_Automatic_project				* 空间类型	应用层	
项目设置项目	目的英文名及名称的命名,将基于公用名	, 按规则自动生成					
* 计算源类型	MAXC 🔄 FLINK 🗌 HOLO 🛈		* 1	18] 开〇(PAI		
	③ 请至少选择一个计算源,生产/开发计	算源设置同步					

使用限制

- 当前, Dataphin仅支持导入基于MaxCompute计算资源创建的PAI-Designer工作流。
- 仅计算引擎为MaxCompute的Dataphin支持创建机器学习任务。

步骤一: 创建机器学习任务文件

- 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>>。
- 在Dataphin首页,单击顶部菜单栏的研发。
 默认进入数据开发页面。
- 4. 在数据开发页面,按照下图操作指引,进入新建文件对话框。

≡	Dataphin · 研发	开发 ◎ 发布 运维	权限	الترجيبي المحافظ	Q、 全局提案	I 🗟	뇹 🕫 오 📀
	pai_smoke_test >> Prod	欢迎使用Dataphin!					
2 <>>					٨		TAK
0	→ ■ 目录 (1) 通用脚本	◆ 秋范定义 » 3 ^{>} 5	(8) (8) (8) (8) (8) (8) (8) (8) (8)	10.00	▲ 原子描标 日 ②	业务限定 □ ◎	派生潮标 □ ◎
8	mangoong tel \$2343 pal_smoke_test01 pai_smoke_test02	PALDESIGNER)		
_	sql_test	◎ 設振建構 ≫	後度 建度 変編表 ⊗	◆ 帯兵芝雄表 ③ ③			二.思逻辑表
20	esq_test_03	4 智能计算		运推调度 >>			
Σ Φ	VER CLEUTER	and another data	and the second second	运行从理户逻辑表100%。 时间正确有序地生产数据。	也想 [,] "主的說過在後,以機能所有在後都有投始 (——		20 20
	·		NAVO 001	计算任务	FUNK OOL		
		□ 任务研发 ≫	MAXC_SPARK_JAR	SHELL D	FLINK_TEMPLATE_SQL	E S Pipe	ine ≣∌ ⊡
danas	an ann an		3 Blig-young	dataphin			
		▲▲文件引用 >>	MAXCEXX		法務結理		元夜 元夜 〇

5. 在新建文件对话框中, 配置参数。

新建文件				×
* 名称	请输入任务名称			
* 调度类型	● 周期性节点	○ 手动节点		
描述	请输入任务描述			1
边拴日求	计异任务			\sim
			取消	确定

参数	描述
名称	 命名规则如下: 不支持竖线()、\/、半角冒号(:)、半角问号(?)、尖括号(<>)、星号(*)半角单引号(")。 全局唯一。 不能超过64个字符。
调度类型	选择机器学习任务后续调度的类型,支持选择 周期性节点和手动节点 ,详细说明如下: • 周期性节点:即创建的机器学习任务为周期任务,后续会自动参与Dataphin的调度,更多关于周期任务的描述,请参见通用功能说明。 • 手动节点:即创建的机器靴子任务为手动任务,后续调度需要您手动触发,更多 关于手动任务的描述,请参见手动任务。
描述	填写对机器学习任务的简单描述。

参数	描述							
选择目录	默认选择为计算任 标文件夹为计算模 ■ Dataphin・研建 ででででです。 ででででです。 ででででです。 ででででです。 ででででです。 ででででです。 ででででです。 でででです。 でででです。 でででです。 でででです。 でのでのでのでのでのでのでのでのでのでのでのでのでのでのでのでのでのでのでの	务。同时 版的目录 知daphin!	t您也可以れ ≹。 ○ ○ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■	在计算任会 Burg Burg Burg Burg Burg Burg Burg Burg		示文件实所	后,选择1	亥目
		□ 任务研究 >	MAXC_SQL MAXC_SPARK_JAR	・ ・ ・ いRTUAL ・ SHELL	FUNK_SQL FUNK_TEMPLATE_SQL		RIPES DATAX RIP Pipelne RIP	
		662048(III) >		PYTHON FUNKEER	PLINC DATASTREAM	Riskans Samana Samana	1	

6. 单击确定,完成机器学习任务文件的创建。

步骤二:开发并配置机器学习任务

1. 在上述已创建的机器学习任务开发页面,按照下图操作指引,导入机器学习中已开发好的工作流。

≡	Dataphin · 研发	开发 。 发布 运维 权限	। 🗈 🕆 🕹 🖉 🖉
	pai smoke test 💎 🎧 Prod		: 8
_		⑥ 导入工作法 10 以新工作法 10 去纳维工作法 目 调度配置 目 历史信息 ◎ 运行 ④ 去没有 ④ 去没有	£ 2 4 ⊚
规	電磁機 数据处理 印度查询	a the safety with a star with with with with	deneghter deneghter deneghter
in	计算任务 〇品田田	导入工作流	
	▼ Q 編入关键字	「工作法 Dataphin_Pai_Designer > 🥑	
`	▶ 圖目录 (1) 🗈 🗄		
datas	💭 🗖 gsgag 👘 🕄		
	mangxing		
	🧰 pai_smoke_test01		and the second
	pai_smoke_test02		
	sql_test		
90	📰 sql_test_02		and a second second
_	sql_test_03		
~	🔲 test_test_fuli		and and and
2	國 课业设计问题	\$P\$100000000000000000000000000000000000	
		El sueda ATT Level - MEN data A	
		导入工作流	

导入后,为您展示在机器学习中开发的工作流。

≡	Dataphin · 研发	开发≒	发布	运维 权	限	Q 全府	局搜索		ë 🗘	@ L\$	
DEV	0217 test dev	PAI_Designer	•							: 8	***
	Dev		○ 刷新工作流	🛛 去编辑工作流	モ 🗉 调度配置	🗉 历史信息 🛛 🕒					影的文
规范望	建模 ① 数据处理 即席查询	1000 and						مى ^{مەربىي} تەتبىرىي			中 中
43	计算任务 🗘 🗳 🖃 🖃										50
2											南御
J	▶ ms_test (1)					Pai	读数据表				10
Bates	rjz_test (0)					/					
B	🕨 🔤 zhaoshun (6)										
	fuli_test11				-			Pai 拆分			
	🚾 fuli_test_11	4			Pai 相关系数矩阵	F					
	PAI_Designer 🛛 🗄										
₽ō	🔝 sql_test										
-	🗃 test					Pai 线	性回归				
Σ	S test_shell										
		1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 - 1997 -									
								Pai 预测			
								_		\$	
										Ð	
								Pai 回归模型评估		Q	
								_		×	

同时,您也可以单击页面上方的**去编辑工作流**,进入机器学习中生成该工作流的页面进行编辑,如何 编辑请参见<mark>创建及管理工作流</mark>。

≡	Dataphi	n ・研发		开发≒	发布	运维	权限			、 全局搜索				Ļ				
DEV	0217 test dev	Prod	•	PAI_Designer	•												ŝ	
		Dev		🖆 导入工作流	○ 刷新工作流	☑ 去编辑	軍工作流	🗉 调度配置	🗉 历史信息	⑤ 运行	💮 去发布	의, 去运维			6	9 1	0	的女仆
规范	建模 ① 数据处理		查询															#
4 3	计算任务		• 🖬															河
С										Pai 读数据表								世世
	► ms_test (1)																	
3050	i rjz_test (0)								/		- Ale							
100	Image: Participation between the second s		1.42															
*	🧰 fuli_test11											Pai 拆分						
_	<pre>fuli_test_11</pre>			4				1 相天系数矩阵	•									
=	PAI_Designer		Ø:															
ΨU	🛃 sql_test									4								
-	iest								Pa	ai 线性回归								
Σ	si test_shell		1.12															
												Pai 预测						
			1.12														•	
																	€	
												Pai 回归模型	平估				Q	
																	x	

更新了工作流后,您可以单击页面上方的刷新工作流,就可以刷新机器学习任务中的工作流。

2. 如果机器学习任务为周期任务,则按照下图操作指引,配置调度参数。

≡	Dataphin · 研发	开发 发布 运维 权限 Q:	全地統定 (島) 谷 🗘 💽
	pal smoke test • V	3 9999 •	: 8:
-	, page and a dev	□ 日日、日本市場、「日本市場」「日本市場」「日本市場」「日本市場」」「日本市場」」「日本市場」「日本市場」「日本市場」」「日本市場」「日本市場」「日本市場」」「日本市場」「日本市場」「日本市場」」「日本市場」「日本市場」「日本市場」「日本市場」」「日本市場」「日本市場」「日本市場」「日本市場」」「日本市場」「日本市場」」「日本市場」「日本市場」「日本市場」「日本市場」」「日本市場」「日本市場」「日本市場」「日本市場」「日本市場」「日本市場」「日本市場」「日本市場」」「日本市場」」「日本市場」「日本市」」「日本」」「日本	×
112	范建模 数据处理 即席查询		2
43	计算任务 ① 品 回 团	• 基本信息	
2		and a sea of the sea o	
Ŭ	▶ 🛅 目录 (1)	R 百点4 發光, 清視	交場度 目 查想PROD市点
datal	0 9090	Pai 表版型-1	MER .
	mangxing	Unical TA_OLUG	nun -
&	pai_smoke_test01	Pal 模型-1 负张人 dataphin	
_	pai_smoke_test02		个字符内的描述文案
-	📕 sql_test	Pal 写数据表-2	
40	sql_test_02	・視覚配置	
-	sqLtest_03	・ 「 「 「 「 」 「 」 「 」 「 」 「 」 「 」 「 」 「 」	★ 空節調度
Σ	test_test_tuli		
		現成所明 日	✓ 00:00 [©]
		cron 表达式 000**?	
100		報告的生産時間 2022-00-1 2022-00-2	7 00:00:00 2022-02-18 00:00:00 2022-02-19 00:00:00 2022-02-20 00:00:00 1 00:00:00
		林範上期期 ④ 否	
		<u>ه</u> ()	
datal		优先级 中毒优先	a and
		时间参数 语绘入后	置的参数,支持自定义参数,以;区隔,如keyA=123;keyB=\$(yyyymmdd-7);
		节点争取配	型 洗明

在调度配置面板,您也可以查看该机器学习任务的基本信息,同时支持修改负责人及描述。

i. 配置**调度配置**区域参数。

调度配置				ahin	nin	motion	>
					- Carlos V	Garao	0.000
● 调度配置							
时间属性	● 正常调度	【 🗌 空跑调度					
暂停调度	• 否	是					600000
调度周期	Ε ∨	00:00					
cron 表达式	000**?						datasph
最近的生成时间	2022-02-17 2022-02-21	00:00:00 2022-02 00:00:00	-18 00:00:00 2022-02	-19 00:00:00	2022-02-20 00):00:00	
依赖上周期	• 否						Astaph
) 是						
优先级	中等优先级	ł	Y.,				
时间参数	请输入配置	的参数, 支持自定	义参数, 以; 区隔, 如k	eyA=123; ke	eyB=\${yyyymmo	ld-7};	dated b
	节点参数配置	说明					
 依赖关系 上游依赖 	自动解析 物理节点	逻辑表节点 ①				手动液	动上游
輸出名称	Ŧ	节点名	节点ID (实例ID)	负责人	datapana.		操作
						取消	确定
参数		描述					
时间属性		选择时间属性 正常调度 正常调度	主。时间属性 包括: 按照调度周期的日 :按照调度周期的日 送返回成功,没有真	寸间配置调, 寸间配置调, 正的执行信	度,并正常拔 度,但都是空 E 务。	い行,通常任务 2.跑执行,即一社	默认选中 周度到该
暂停调度		选择 是 后,即 一旦调度到该 行,但后面还]可暂停该任务的调 [任务会直接返回失 [会继续使用的场景	度,会按照 败,不会拔 。	踩下面的调度/ 1行。通常用 ⁻	周期时间配置调 于某个任务暂时]度,但是 不用执

参数	描述
调度周期	 用于定义周期任务生成对应的周期实例的调度周期,支持选择日、周、月、小时和分钟,详细说明如下: 日调度,即调度实例每天自动运行一次。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,即调度实例每周的特定几天,在特定时间点自动运行一次。 月调度,即调度实例在每月的特定几天,在特定时间点自动运行一次。 小时调度,即每天指定的时间段内或时间点,调度实例按时间间隔数及开始结束时间或按每天的时间点运行一次。 分钟调度,即每天指定的时间段内,调度实例按时间间隔数及开始结束时间运行一次。
依赖上周期	用于定义周期任务是否需要依赖本节点或其他节点上个调度周期的运行结果。依 赖节点支持选择当前和自定义。选择了自定义后,您需要通过节点名称、输出 名称或节点ID,查找需要依赖的节点。 依赖上周期 音
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。
时间参数	时间参数是对代码中所用时间参数的具体赋值。您可以单击 节点参数配置说 明,查看Dataphin调度系统支持配置时间参数的类型及生效说明。

ii. 配置**依赖关系**区域参数。

调	度配置						
	时间参数 请输入商	配置的参数, 支持自治	定义参数,以;[⊠隔,如keyA=	=123; keyB=\${yyyymi	mdd-7};	
	节点参数理	配置说明					
•	依赖关系 自动解析						
200	上游依赖 物理节点	逻辑表节点 ①				手动液	励上游
	輸出名称	节点名) DI点芇	(实例ID)	负责人		操作
9000							
			群 甲s ^{Ti} unga hag	无数据 			data f
	当前节点					手动添	泇輸出
900	輸出名称	节点名	节点ID((实例ID)	负责人		操作
1000			- 	无数据			
	(Jator	Galeri	0.850		(Laso)	取消	确定
4811	参数	描述					

参数	描述
上游依赖	上游依赖不能为空。您可以通过自动解析和手动添加的方式为机器学习任务添加 上游依赖,详细说明如下: • 自动解析 Dataphin根据机器学习任务的工作流中的读数据表自动解析到上游依赖。首 先Dataphin自动获取工作流组件中的表及字段信息,假设获取到的表名为 Table;其次Dataphin自动获取对应的MaxCompute项目绑定的Dataphin项 目名,假设为Project;最后将获取到的节点 Project.Table 添加至上游 依赖区域。如果Dataphin获取不到对应的项目或对应的节点,则无法自动解 析出上游依赖任务。 • 手动添加 按照下图操作指引,手动添加上游依赖节点。 • 你额关系 自动解析 • ************************************
当前节点	定义当前节点的输出名称,根据需要您可以设置多个输出名称,供其他节点依赖 使用。 按照下图操作指引,添加当前节点的输出名称。

步骤三:运行并提交机器学习任务

1. 在上述已创建的机器学习任务开发页面,按照下图操作指引,运行机器学习任务。

≡	Dataphin · 研发								۵ 🧕
	noi emetre test			1					
	par_sinoke_test V V O Dev			田 历史信息					
规范		all ^{an} el ^{ante} el	alle andre andre	a	en en en en	and a second	and the second		
45									
~									
<u>۲</u>									
					6.00° 6.00°				
				参数填与	2				
				* bizdate 20220101	遺充準備				
		100 A. 100 A. 100 A.			1				
_		4		退探境充空值后,将参数值清空持	并重庆,且不阻塞点击确定按钮	据表-1			
5									
_					-9K 709 - 50 725				
					Pal 耳题3	· 国現-1			

查看运行日志。

⑥ 导入工作流 〇 刷新工作流	☑ 去编辑工作流	🗄 调度配置	🗏 历史信息	⑤运行	🕑 去发布	心, 去运维						8	8 1 0
	hin balantinin	distance for	(atop)	in do	rangitin	dataqinin dataqinin	dataphie	dataphin	datasphin	distantion.	data phin	and a stranger	dataphin
						real Breacharter 1							
						Pal 写数据表-1							
运行日志			1000										₽ ¥
et con.alibaba.dataphin.com et con.alibaba.dataphin.com et con.alibaba.dataphin.com et con.alibaba.dataphin.com et con.alibaba.dataphin.com et con.alibaba.dataphin.com et con.alibaba.dt.valdemort et gissu.util.concurrent.fut et jssu.util.concurrent.fut et jssu.util.concurrent.fut	mon, pai.api.Pai(limon, pai.Pai(limon, pai.Pai(limon, pai.api.Pai(limon, pai.api.Pai(limon, pai.api.Pai(limon, pai.api.Pai(limon, pai.Pai(limon, pai.api.Pai(limon, p	ent.executeRetr ent.executeRet	y(PaiClient.ja y(PaiClient.ja y(PaiClient.ja y(PaiClient.ja aiClient.java: iDesignerOpera iDesignerOpera or.call(BaseOp or.call(BaseOp tors.java:511)	va:318) va:318) va:318) va:296) va:296) va:337) 196) tor.submitP tor.submitP tor.execute tor.execute erator.java erator.java	aiJob(PaiDesi (PaiDesigner0 (PaiDesigner0 ::334) ::115)	zor@perator.java:399) perator.java:232) perator.java:203)			aan	~			

2. 运行成功后,按照下图操作指引,提交机器学习任务至生产环境。

pai_	smoke_test02							
图 导入。	し作為 〇風的	新土作流 凶	去编辑上作流	曰 历史信息	⑤ 运行 (4) 去发布	4 去运堆		
								* 提交留注 2
								请输入此次提交的相关备注说明,如变更信息
						daterin daterin		
						Pai 读数据表-1		0/128
								杨司条注则道明此分理去的恋菌内网
								当前项目,提交成功后,请前往侍发布列表,以发布对象至生产环境。
								取消 确定并提交
						Pai 写数据表-1		

3. 如果您使用的是Dev-Prod的项目,提交成功后需要发布机器学习任务至生产环境,具体操作请参见发布 任务。

后续步骤

后续您需要在运维中心运维管理机器学习任务及生成的实例,详情请参见运维中心概述。

3.9. 调度配置

调度配置用于配置节点在调度过程中的调度规则及依赖关系。本文为您介绍如何配置计算任务的调度参数。

前提条件

完成创建计算任务:

- 如果计算引擎为MaxCompute,详情请参见新建MAXC任务。
- 如果计算引擎为ADB for PostgreSQL,详情请参见新建ADB任务。

背景信息

- 系统仅支持周期性节点的代码任务进行调度配置。
- 依赖关系是两个或多个节点之间的语义连接关系,同时上游节点的状态将影响其他下游节点的运行状态。
- 配置了依赖关系的节点调度规则为:首先,上游节点运行完成后,才能调度下游节点。其次,根据节点设定的调度时间判断是否执行调度。
- 您在设定的调度时间之前提交的调度配置,会在设定的调度时间之后生效。而在设定的调度时间之后配置 的依赖关系,只能间隔一天再生成实例。
- 调度配置的权限限制,详情请参见数据引入权限列表。

操作步骤

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的数据研发。
 - 单击快速进入研发项目区域的Dev或Basic项目,选择开发数据的项目空间。
 - ? 说明
 - 如果您通过快速开始相关工作进入数据开发模块,则跳过步骤3。
 - 如果您通过**快速进入研发项目**进入数据开发模块,则跳过步骤3和4。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。
- (可选)在数据开发页面,单击项目名称后的☑图标,单击Dev或Basic页签,选择数据开发的项目空间(Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

- 5. 进入离线计算任务的调度配置页面。
 - i. 在数据开发页面,单击数据处理。
 如果进入数据开发页面后,系统默认进入数据处理页签,则无需再次单击数据处理页签。
 - ii. 在左侧导航栏中, 单击 🐼 计算任务图标。
 - iii. 在计算任务页面,查询需要配置调度参数的计算任务。您可以在搜索框中,输入计算任务名称关键字,快速搜索计算任务。
 - iv. 单击需要配置调度参数的计算任务。

- v. (可选)解锁计算任务。
 - 如果需要配置调度参数的计算任务已被您自己锁定,则无需解锁。
 - 如果需要配置调度参数的计算任务已被他人锁定,则单击**代码编写**页面右上方的**圆**图标。

当38标变更为588标时,解锁成功。

vi. 在该离线计算任务的代码编写页面,单击调度配置。

- 6. 在调度配置页面, 配置调度参数。
 - i. 配置基本信息区域参数。

系统自动生成的名称、节点ID、节点类型不支持修改,系统支持修改负责人和描述。

ii. 配置调度配置区域参数。

参数	描述
时间属性	选择 时间属性。时间属性 包括: 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。
暂停调度	暂停调度 选择是后,即可暂停该任务的调度,会按照下面的调度周期时间配置 调度,但是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。

参数	描述
调度周期	 調度周期可选择日、周、月、小时和分钟: 日调度,即调度任务每天自动运行一次。新建周期任务时,系统默认的时间周期为每天0点运行一次。您可以根据需要,单击 图标,指定运行的时间点。 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。您可以根据需要,单击 图标,指定运行的时间点。 如果您没有指定日期,为保证下游实例正常运行,系统会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。 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。您可以根据需要,单击 图标,指定运行的时间点。 如果在没有被指定的日期时,为保证下游实例正常运行,系统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。 小时调度,即每天指定的时间段内,调度任务按间隔时间数的时间间隔运行一次。或选择指定的时间点,调度系统会自动为任务生成实例并运行。您可以根据业务需求选中时间段或时间点: 如果您选中了时间段,您可以单击开始或结束后的 图标,指定运行的开始和结束时间。同时您可以单击间隔后的 图标,在下拉列表中选择时间点。 如果您选中了时间点,单击下拉列表框,在下拉列表中选择时间点。 分钟调度,即每天指定的时间段内,调度任务按间隔时间数的时间间隔运行一次。您可以单击开始或结束后的 图标,指定运行的开始和结束时间。同时您可以单击间隔后的 图标,指定运行的开始和结束时间。同时您可以单击间隔后的 图标,在下拉列表中选择间隔时间。
依赖上周期	根据业务场景分析是否需要选中 依赖上周期 。如果您选择是,则需要选择依赖 上个调度周期的节点: 如果您选择了 当前 ,则表示依赖当前节点的上个调度周期。 如果您选择了自定义,则需在输入框中输入依赖上周期节点的名称关键字进 行搜索节点。
优先级	您可以选择节点调度任务执行的优先级,优先级包括: 最低优先级。 低优先级。 中等优先级。 高优先级。 最高优先级。

参数	描述
参数配置	节点参数配置是对代码中所用参数的具体赋值。您可以单击 节点参数配置说 明,查看Dataphin调度系统支持配置节从而支持节点调度时,参数值以变量替 换执行点的时间参数及配置原则,进行配置参数。

iii. 配置依赖关系区域参数。

参数	描述							
	当节点的任务类型为SQL时,您可以单击 开始解析 ,系统会解析代码中的物理表 和逻辑表,并查找到与该表名相同的输出名称。输出名称所在的节点作为当前节 点的上游依赖。							
	如果代码中引用项目变量或不指定项目,则系统默认解析为生产项目名,以保证 生成调度的稳定性。例如,开发项目名称为 onedata_dev :							
白动解析	■ 如果代码里指定 select * from s_order ,则调度解析依赖为 oneda ta.s_order 。							
	■ 如果代码里指定 select * from \${onedata}.s_order ,则调度解析 依赖为 onedata.s_order 。							
	■ 如果代码里指定 select * from onedata.s_order , 则调度解析依赖 为 onedata.s_order 。							
	■ 如果代码里指定 select * from onedata_dev.s_order ,则调度解析 依赖为 onedata_dev.s_order 。							
	上游依赖包括物理节点和逻辑表节点。逻辑表节点上游依赖仅支持系统解析。如 果业务上没有上游依赖,您需要添加根节点(root节点)为上游依赖,便于统一 管理。							
	执行如下操作,添加该代码任务调度时依赖的上游物理节点:							
	a. 单击 手动添加上游 。							
	b. 在新建上游依赖对话框中,您可以通过以下两种方式搜索依赖节点:							
上游依赖	■ 输入所依赖节点的输出名称的关键字进行搜索节点。							
	■ 输入root搜索虚拟节点(每个租户或企业在初始化时都会有一个根节 点)。							
	⑦ 说明 节点的输出名称是全局唯一的,且不区分大小写。							
	c. 单击确定新增。							
	同时您还可以单击操作列下的 。 图标,删除已添加的依赖节点。							

参数	描述
	通过执行如下操作,设置当前节点的输出名称,根据需要您可以设置多个输出名称,供其他节点依赖使用: a. 单击 手动添加输出 。
	b. 在新增当前节点输出对话框中,填写输出名称。输出名称的命名规则请尽量统一,一般命名规则为生成项目名.表名且不区分大小写,以标识本节点产出的表,同时其他节点更好地选择调度依赖关系。
当前节点	例如,开发项目名称为 onedata_dev ,建议将输出名称设置为 oned ata.s_order 。如果您将输出名称设置 为 onedata_dev.s_order ,则仅限代码 select * from onedata _dev.s_order 能解析出上游依赖节点。
	c. 单击 确定新增 。
	同时您还可以对当前节点已添加的输出名称执行如下操作:
	单击操作列下的 图标, 删除已添加的输出名称。
	如果该节点已提交且被任务所依赖(任务已提交),则单击操作列下的 ③图
	标,查看下游节点。

7. 单击确定,完成代码任务的调度配置。

8. 保存、提交和发布离线计算任务。

i. 在离线计算任务的配置页面, 单击 图标, 保存离线计算任务。

- ii. 在离线计算任务的配置页面,单击√图标,提交离线计算任务。
- iii. 在提交备注对话框,填写备注信息。
- iv. 单击确定并提交。
- v. (可选)发布离线计算任务至生产环境。
 - 如果您的开发模式是Dev-Prod模式,则需要发布离线计算任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则不需要发布离线计算任务,提交后的离线计算任务就可以参与 生产环境的调度。

3.10. 管理计算任务

本文为您介绍如何编辑、移动、重命名和删除计算任务等操作。

前提条件

完成计算任务的创建,详情请参见MAXC任务、ADB任务、Flink任务或通用任务。

背景信息

- 计算任务只有在解锁的状态下,才支持编辑、移动、重命名和删除等操作。
 - 如果计算任务已被您自己锁定,则无需解锁。

- 如果计算任务已被他人锁定,则需要解锁任务:
 - a. 在计算任务页面, 鼠标悬停至 图标, 单击编辑。
 - b. 在代码开发页面,单击右上方的 图标。

当圖图标变更为圖图标时,解锁成功。

- 计算任务的状态说明:
 - 保存成功的计算任务, 状态为**草稿**。
 - 提交成功的计算任务, 状态为已提交。
 - 编辑已提交状态的计算任务,并成功保存,状态更改为开发中。
 - 已提交状态的计算任务,成功下线后,状态更改为草稿。
- 下线、删除和下线并删除计算任务的限制:
 - 仅开发中和已提交状态的计算任务支持下线操作。
 - Q草稿状态的计算任务支持删除操作。
 - 仅开发中和已提交状态的计算任务支持下线并删除操作。
- 新建的Flink_SQL、Flink_Template_SQL类型的计算任务默认开启实时模式,您可以手动开启离线模式。 Flink_SQL、Flink_Template_SQL类型的计算任务支持处理实时和离线数据。
- 下文中涉及的概念说明。

概念	描述
Flink任务	Flink_SQL、Flink_Template_SQL和Flink_DataStream类型的计算任务。
MAXC任务或ADB任务	SQL、MapReduce、Spark、Shell、Python、Virtual和ADB for PostgreSQL类型的 计算任务。
计算任务	Flink_SQL、Flink_Template_SQL、Flink_DataStream、SQL、MapReduce、 Spark、Shell、Python、Virtual和ADB for PostgreSQL类型的计算任务。

编辑MAXC任务或ADB任务

- 1. 登录Dataphin控制台。
- 在Dataphin控制台页面,选择工作区地域后,单击进入Dataphin>>。
 您可以通过以下方式,快速进入数据开发模块:
 - 单击快速开始相关工作区域的数据研发。
 - 单击快速进入研发项目区域的Dev或Basic项目,选择开发数据的项目空间。

? 说明

- 如果您通过**快速开始相关工作**进入数据开发模块,则跳过步骤3。
- 如果您通过快速进入研发项目进入数据开发模块,则跳过步骤3和4。

3.

4. 鼠标悬停至需要编辑的计算任务后,再将鼠标悬停至图标,单击编辑。

- 5. 在代码开发页面,编辑计算任务的代码和调度参数,详情请参见新建SQL任务、新建MapReduce任务、新建 Spark任务、新建Shell任务、创建Python计算任务、新建Virtual任务和新建ADB for PostgreSQL任务。
- 6. 保存、提交和发布计算任务。
 - i. 单击页面右上角的 图标, 保存计算任务。
 - ii. 单击页面右上角的 图标, 提交计算任务。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交。
 - v. (可选)发布计算任务。
 - 如果您的开发模式是Dev-Prod模式,则需要发布计算任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布计算任务。

编辑Flink任务

- 1. 在代码开发页面,您可以修改计算任务的代码、资源配置、依赖关系、任务参数、调度参数、通用配置的参数、或**模板参数**(Flink_Template_SQL类型的计算任务),详情请参见创建Flink SQL任务、步骤一: 创建并调试Flink Template SQL计算任务和创建Flink DataStream计算任务。
- 2. 保存、提交和发布Flink任务。
 - i. 单击页面右上角的 图标, 保存计算任务。
 - ii. 单击页面右上角的 图标, 提交计算任务。
 - iii. 在提交备注对话框,填写备注信息。
 - iv. 单击确定并提交。
 - v. (可选)发布计算任务。
 - 如果您的开发模式是Dev-Prod模式,则需要发布计算任务,详情请参见管理发布任务。
 - 如果您的开发模式是Basic模式,则无需发布计算任务。

修改Flink任务的通用配置

- 1. 在代码开发页面, 单击页面上方的通用配置。
- 2. 在通用配置面板,修改责任人、资源队列和引擎版本。

注意 资源队列切换后,提交并执行任务,可能导致运行结果有差异,请谨慎操作。

3. 单击确定。

修改模板参数

⑦ 说明 仅FLINK_TEMPLATE_SQL类型的计算任务支持修改模板参数。

- 1. 在代码开发页面, 单击页面上方的模板参数。
- 2. 在模板参数面板,修改参数值。
- 3. 单击确定。

查看Flink任务的代码结构

- 1. 在代码开发页面,单击页面上方的代码结构。
- 2. 在代码结构面板,查看源表(SOURCE)、维度逻辑表(DIM)表、输出表(SINK)、中间表 (SELECT)、过滤(FILTER)、分组统计(GROUP)、关联(JOIN)、合并(UNION)、取前N位数 (TOPN)、快照(SNAPSHOT)、复杂事件处理(CEP)、自定义函数(UDF)和其他(Other)等信 息。
- 3. 单击 图标,关闭面板。

修改Flink任务的生效模式

- 1. 解锁任务。
 - 如果计算任务已被您自己锁定,则无需解锁。
 - 如果计算任务已被他人锁定,则需要解锁代码。
 - a. 在计算任务页面,鼠标悬停至图标,单击编辑。
 - b. 在代码开发页面,单击右上方的圖图标。

当 图标变更为 图标时,解锁成功。

- 2. 在代码开发页面, 单击页面上方的定向配置。
- 3. 在定向配置面板, 单击实时模式或离线模式页签。
- 4. 在实时模式或离线模式页签,单击对应页签的 💿 或 💽 图标,打开或关闭对应模式。

⑦ 说明 如果当前生效模式仅为一种时,则系统不支持关闭该模式。

查看Flink任务的版本信息

- 1. 在计算任务页面, 单击左侧功能区域中需要查看版本信息的计算任务。
- 2. 在代码开发页面,单击页面上方的版本信息。
- 3. 在版本信息面板,您可以执行以下操作。
 - 单击某个版本所在行操作列下的□□图标,查看版本对比。

在版本对比对话框,您可以单击资源配置对比、任务参数对比或代码对比页签,查看版本的对比信息。同时您还可以执行以下操作,回滚版本:

- a. 在代码对比、资源配置对比或任务参数对比页面,单击页面下方的回滚到对比版本。
- b. 在提示对话框中, 单击确定。
- 单击某个版本所在行操作列下的 图标, 查看代码的基本信息。

查看MAXC任务或ADB任务的历史信息

- 1. 在计算任务页面, 单击需要查看历史信息的计算任务。
- 在该任务的代码编写页面,单击页面上方的历史信息,查看版本信息。
 在历史信息面板,您还可以执行以下操作:
 - 单击操作列下的详情。在基本信息页面,查看版本的详细信息。

- 单击操作列下的代码对比。在版本对比页面,比较历史版本和当前版本的代码,您也可以单击回滚
 到对比版本,回滚历史版本。
- 单击操作列下的查看代码, 查看代码内容。
 - ⑦ 说明 如果您仅提交了一个版本,则只能查看当前版本的信息。

移动计算任务

- 1. 鼠标悬停至需要移动的计算任务后,再将鼠标悬停至 图标,单击移动。
- 2. 在移动文件对话框,选择目录。
- 3. 单击确定,完成移动计算任务。

重命名计算任务

- 1. 鼠标悬停至需要克隆的计算任务后,再将鼠标悬停至图标,单击重命名。
- 2. 在文本框中, 输入计算任务的名称。
- 3. 按下Enter键或单击非输入框区域,完成重命名计算任务。

克隆计算任务

- 1. 鼠标悬停至需要重命名的计算任务后,再将鼠标悬停至图标,单击克隆。
- 2. 在克隆计算任务对话框,填写新任务名称,并选择目录。
- 3. 单击确定。
- 4. 开发、提交或发布代码任务,详情请参见MAXC任务、ADB任务、Flink任务或通用任务。

筛选任务

在数据处理页面,单击 图标,快速筛选我创建的数据处理对象。

快捷去运维

在计算任务代码开发页面,单击页面上方的去运维,快速进入运维中心,运维该任务。

快捷去发布

在计算任务代码开发页面,单击页面上方的去发布,快速进入发布中心,发布该任务。

下线计算任务

- 1. 鼠标悬停至需要下线的计算任务后,再将鼠标悬停至 图标,单击下线。
- 2. 在提示对话框,填写备注信息。
- 3. 单击确定并提交,完成下线计算任务。

下线并删除计算任务

1. 鼠标悬停至需要下线并删除的计算任务后,再将鼠标悬停至图标,单击下线并删除。

2. 在提示对话框,填写备注信息。

3. 单击确定并提交,完成下线并删除计算任务。

删除计算任务

- 1. 鼠标悬停至需要删除的计算任务后,再将鼠标悬停至图标,单击删除。
- 2. 在提示对话框,填写备注信息。
- 3. 单击确定并提交,完成删除计算任务。

4.即席查询

4.1. 概述

即席查询提供数据查询功能。系统会根据您所设置的计算类型自动识别切换SQL类型。

因为Dat aphin独有的逻辑表模型以业务板块为命名空间,所以引用逻辑表的SQL语法有一定的限制,详细内容如下:

- 系统支持引用逻辑表的方式为业务板块.逻辑表。
- 系统支持查询逻辑表中字段的方式为[业务板块.逻辑表.关联维度角色.....关联维度角色.维度逻辑表字段]。
 如果作为Select或者Where等设置的内容,则仅需要填写[逻辑表.关联维度角色.....关联维度角色.维度逻辑表字段],就可以实现关联维度对象的字段查找使用。

⑦ 说明 虚拟维度的主键字段不支持用于即席查询,虚拟维度仅用于构建统计粒度。

4.2. 查询并下载数据

您可以根据当前的业务情况自定义并执行查询语句及下载查询的数据。例如,当完成计算任务开发后,您可 以通过即席查询验证计算任务是否符合预期。本文为您介绍如何查询并下载数据。

前提条件

在您开始执行操作前,请确认您已开启项目下载数据的权限。具体操作请参见管理项目空间的权限和计算 源。

使用限制

Dataphin默认最多返回10000条查询结果的记录。您可以在查询语句后增加 limit *返回条数*,自定义返回 记录的条数,例如 select * from ht5 out limit 20; ,即返回20条记录。

注意事项

查询数据过程中,如果任务执行超过24小时仍然没有结束,则Dataphin会自动终止任务进程,避免占用大量运行资源,影响其他任务的运行。

查询数据

查询汇总逻辑表数据时为避免大量资源的浪费,在开始查询汇总逻辑表数据前,您可以先关闭使用 select * 语法查询汇总逻辑表数据的权限,具体操作请参见设置即席查询的引擎。

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏的**研发**。 默认进入数据**开发**页面。
- 4. 在数据开发页面,按照下图操作指引,进入新建文件对话框。

数据开发·<mark>即席查询</mark>

■ Dataphin · 研发	开发≓ 发布	运维 权限	Q、 全局搜索	日間 道 🛵	a 🕗
DEV 時間第55 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	欢迎使用Dataphin!				我的文件
則病查询 ○ 品 □ □ Q、输入关键字 4 , model action (1) Instruction (0)	✔ 规范定义 》	推度 注 ◎	业%过程 业%过程 □ ◎ + ◎	wsggz wsggz wsggz withh withh w withh w withh w withh w withh withh withh withh withh withh withh withh withh wih	調査
	◆ 数据建模 >>	自动创建 》 推度逻辑表 ②	事实逻辑表 ②		
 fuxioo-test gsgagafaf lost lost_desensitze lost_desensitze 	智能计算	- 20	這個項度》 這行乃進於度但裏和說服於理产生的数据任务, 时间正确有序地生产数据。	以确保所有任务都有按照	() 反因

5. 在新建文件对话框,配置参数。

新建文件				×
* 名称	请输入任务名称			
描述	请输入任务描述			
选择目录	临时代码			\sim
		取消	确定	

参数	描述
名称	填写即席查询名称。命名规则如下: • 除竖线()、半角冒号(:)、半角问号(?)、尖括号(<>)、星号(*)、半 角引号(")外,其余字符均支持。 • 不能超过64个字符。
描述	填写对即席查询的简单描述。

参数	描述							
	默认选择为 临时代 标文件夹为即席查试	码 。同时您也 旬的目录。	2可以在 即	席查询页	面创建目	标文件夹质	后,选择该	(目)
		开发 ≕ 发布	运维 权限	Q 全	局指定		奋 🕹 🕄	9
	Dev / Masses Prod Prod ① dqe_demo_dev マーレ Prod ■ 現約課題 約2005211 明政会論時	欢迎使用Dataphin!						我的文件
		✔ 短范定义 ≫			<u>Å</u>		STATING.	夷性查看
选择目录	> iii guzhi_demo (1) iii ms_test1 (0)			0	€⊚	₽ ©	• •	
	・ ■ 元広(1) ・ ■ 元広(1) ・ ■ 元広(1) ・ ■ 元広(1) ・ ■ 元広(1) ・ □ 305, kell ・ □ c ・ □ c ・ □ consequenté	◆ 数数建模 ≫	他度逻辑表 ②	- 事实逻辑表 于 ⊗			〕 汇总逻辑表 〒 ⊗	L
	 Acceptorymetric Rociano test grogagatat test 			运输 调度 》 运行及维护逻辑表和数据注意产生的数据任务,以确保所有任 时间正确有传统主产数据。		是所有任务都有按照	58 (64 50)	
	test_dev_prod		-					

- 6. 单击确定,完成即席查询文件的创建。
- 7. 编写查询数据的代码,并单击页面右上方的执行。

≡	Dataphin 研发	Brod v32_odp	s_1211		集成	开发	发布	运维			奋 🧔	
	即席查询		🧱 eee	×								: 8
			🕑 运行 🛄 预编	降 🗎 保存任务 🚔 格式体	≤ ⊚ ⊜							
			1 DESCRIBE	adm_ms_test01								助
	 In Sec. 		2 3 select *	from adm_ms_test01 whe	re ds=20211221							手
	• 10 TABLE		4 5 incent T	NTO adm ms tost01 PART	TTON(dr=20211221)		22 ivinor	ing: 180 25 ident	top: 111 town 1 33	AAAA (second AAA (second))		m
	 		6	and ada_as_cestor rate.	11011(03-202112221	WALDED (,	ing ,100,20, 0000	(), <u>,</u> ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	·····, ·····, ····)		. .
	• (a) (600 (6)		7 create to	able if not exists test	1201(- 1
◬	💼 aliyan _ taat		9 customer	r_id STRING,								
۲			10 total_p 11 partition	rice DOUBLE) ned by (ds STRING);								
	i Mari		12									
	•		13 INSERT II	NTO TESTIZOI PARTITION	ds=20211221) VAL	ies ("www	/nn , 2222	,12.12)				
	 and, but 		< 15 select *	from test1201 where de	-20211221;							

查询数据的代码说明:

○ 查询逻辑表数据的代码说明。

```
SELECT * FROM 业务板块公用英文名.逻辑表名 WHERE ds='${bizdate}';
```

■ *业务板块公用英文名*: 需要查询逻辑表所在业务板块的公用英文名。

您可以在业务板块页面,查看业务板块公用英文名。

≡	Datap	hin·规划				
业务	/冬坂中	业务板块 (81)				
w ź	3.共定义	BASIC LD v297 zhtest basic	1	板块参数 ①		
项目		 更新信息 描述: 122 	Ē	时间分区 ① 中文名 时间分区	英文名 ds	Ø
□ IJ 源	阿管理	C PROD LD_v297_zhtest	2000 1999	数据类型 STRING 描述 时间分区	默认值 NULL	
3 数	如据源	DEV LD_v297_zhtest_dev	ž	业务日期 ①		Ø
स	算源	① 更新信息 描述:		参致 dizdate	参致值 yyyyMMdd	

■ *逻辑表名*:需要查询的逻辑表名。

■ *ds='\${bizdate}'*: 定义查询的分区日期, \${bizdate} 为Dataphin内置日期参数, 默认格式为 yyyymmdd, 即当前日期的前一天。Dataphin也支持输入具体的日期, 例如 ds=`20210223'。

代码示例。

SELECT * FROM LD_newretail.dws_xnwd_03 WHERE ds='\${bizdate}';

↓ 注意 如果关闭了使用 select ★ 语法查询汇总逻辑表的权限,则 select ★ 语句运行会报错,需要指定具体字段。

• 查询物理表数据的代码说明。

SELECT * FROM 项目公用英文名.物理表名 WHERE ds='\${bizdate}';

■ *项目公用英文名*:需要查询物理表所在项目的公用英文名。

您可以在项目的信息设置页面,查看项目的公用英文名。

信息设置: v297_test_project(v297_demo) 已选: Dev-Prod 模式 ①	X
名称设置	命名空间
* 公用名称 v297_demo	业务板块 LD_297_test_project ∨
* 公用英文名 v297_test_project	* 空间类型 公共层 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
项目设置 项目的英文名及名称的命名,将基于公用名,按规则自动生成	
项目1: DEV 项目2: PI	ROD
 ★ 计算源 请至少选择一个计算源,生产开发计算源设置同步 ★ 计算源 请 	青至少选择一个计算源,生产/开发计算源设置同步
计算源类型 🔽 MAXC 🗌 FLINK 🗌 HOLO 🛈 计算源类	差型 🔽 MAXC 🗌 FLINK 🗌 HOLO ③
* MAXC v297_test_project_dev v * MAX	XC v297_test_project ∨
	取消 确定

- 物理表名: 需要查询的物理表名。
- *ds='\${bizdate}'*: 定义查询的分区日期, \${bizdate} 为Dataphin内置日期参数, 默认格式为 yyyymmdd, 即当前日期的前一天。Dataphin也支持输入具体的日期, 例如 ds=`20210223'。

代码示例。

SELECT * FROM beijing ast.ads perform taok all goods day WHERE ds='\${bizdate}';

查询语句执行成功后,您可以在Result页签查看数据。

代码编写							🕑 执行	■ 预编译	🔒 格式(Ł 🔂	9
1 select * from	de_addresse_	where ds=\${b	izdate};							11150/11100 1111/10	
2											
3											
4											
5										心	
6											*
										_	
Console Result											≽
		_							: ~		
A											
1 address_id 🗑	address 🕎	address2 🕎	district 🕎	city_id 🛛	postal_code 🛛	phone 🕎	last_update 🛛	ds 🕎			

下载数据

查询语句执行成功后,按照下图操作指引,下载查询到的数据。

1 select * from	
	· · · · · · · · · · · · · · · · · · ·
6	— *
Console Result	<mark>2</mark> ×
result 0 X	
A B C D E F G H I	
1 address_id 🗹 address 🔍 address2 🔍 district 🔍 city_id 🔍 postal_code 🔍 phone 🔍 last_update 🔍 ds 💟	

4.3. 管理即席查询

本文为您介绍如何进行编辑、移动、重命名、删除即席查询任务,帮助您更好的管理即席查询任务。

编辑即席查询任务

- 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>>。
- 3. 在Dataphin首页,单击顶部菜单栏的研发。
- 4. 在数据开发页面,单击项目名称后的**∑**图标,选择数据开发的项目空间(Dev或Basic项目)。

如果您当前访问的是Dev或Basic项目,且项目空间为您的数据开发空间,则不需要选择项目空间。

- 5. 在数据开发页面,单击即席查询。
- 6. 在即席查询页面,将鼠标悬停在需修改的即席查询任务的

÷

图标后,单击**编辑**。

7. (可选)在该即席查询文件的代码编写页面,单击页面右上方的高图标,偷锁。

当您需要编辑他人创建的文件或您的文件被他人锁定时,只有偷锁成功后,您才可以编辑即席查询文件。

 如果您需要编辑他人创建的文件或您的文件被他人锁定时,则需要偷锁。偷锁成功后,您才可以编辑 即席查询文件。

- 如果即席查询文件已被您锁定,则不需要偷锁。
- 8. 编写并运行即席查询任务。
 - i. 在该即席查询文件的**代码编写**页面,编写代码任务。

编写代码时,引用表的说明:

- 使用跨项目物理表时,名称需以项目英文名为前缀。例如 select * from cloudtest_dev.tabl e4 ,其中table4为物理表。
- 逻辑表查询时,名称需以业务板块的英文名为前缀。例如 select province from ld_practice .dim province ,其中dim_province为逻辑表。
- 查询开发环境数据时,需生产环境名称后加上_dev,系统会自动根据生产环境中的业务板块及 项目生成对应的变量。例如,业务板块LD_Trade,则系统自动生成业务板块变量\${LD_Trade}。 该变量在开发环境执行时默认被替换为LD_Trade_dev,在生产环境执行时默认被替换 为LD_Trade。您也可以执行时设置固定的值,提高代码在不同环境执行的灵活性。
- ii. 完成编写后,单击**预编译**,检验SQL语句是否复合规范。

您也可以单击格式化,系统自动调整您编写的SQL语句,使得其符合规范。

- iii. 规范校验通过后,单击执行,查看运行SQL结果是否成功。
- 9. SQL运行成功后,单击页面右上方的 图标,即可保存即席查询任务。

删除即席查询任务

⑦ 说明 仅支持删除您已锁定的即席查询任务。

1. 在即席查询页面,将鼠标悬停在需删除的即席查询任务的

:

图标后,单击**删除**。

- 2. 在提示对话框中,填写备注信息。
- 3. 单击确定并提交,完成即席查询任务的删除。

移动即席查询任务

⑦ 说明 仅支持移动您已锁定的即席查询任务。

- 1. 在即席查询页面,将鼠标悬停在需移动的即席查询任务的
 - Ξ

图标后*,*单击**移动**。

- 2. 在移动文件对话框中,选择选择目录。
- 3. 单击确定,完成即席查询任务的移动。

重命名即席查询任务

⑦ 说明 仅支持重命名您已锁定的即席查询任务。

1. 在即席查询页面,将鼠标悬停在需重命名的即席查询任务的

3

图标后*,*单击重**命名**。

2. 根据页面提示,在文本输入框中输入即席查询任务的名称。

3. 按下Enter键,完成即席查询任务的重命名。

5.SQL使用说明 5.1. MaxCompute SQL与函数支持

5.1.1. 常用命令

本文为您介绍Dataphin支持的MaxCompute SQL常用开发命令。

변入项目空间(use)支持查看项目空间属性(setproject)不支持设置项目空间属性(setproject)不支持2空石2空石2空石増加RAM账号体系(list accountproviders)支持週除RAM账号体系(remove)支持Set是作Setset5Show flagsshow flags不支持ボロビアOjectProtection)ホ支持※加可信项目空间(add trustedproject)不支持管看可信项目空间(list trustedproject)不支持管看可信项目空间(list trustedproject)不支持管有项目空间的安全配置(show)不支持	类别	常用命令	是否支持
호급项目空间属性 (setproject)지支持···································		进入项目空间(use)	支持
ගුබହබඹ ගුබହබඹ ແ ගුබාඔහිබබ සේගුබම සේගා ප්ත්ය ප්ත්ය ප්ත්ය ප්ත්ය 		查看项目空间属性(setproject)	不支持
项目空间操作출看项目空间账号体系 (list accountproviders)支持増加RAM账号体系 (add)支持潮除RAM账号体系 (remove)支持set操作setsetshow flagsshow flags不支持자支持ボロビー水市の目空间数据保护 (ProjectProtection)不支持添加可信项目空间 (add trustedproject)不支持養除可信项目空间 (list trustedproject)不支持宣看可信项目空间 (list trustedproject)不支持宣看项目空间 (list trustedproject)不支持		设置项目空间属性(setproject)	不支持
별加RAM账号体系 (add)支持 ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●	项目空间操作	查看项目空间账号体系(list accountproviders)	支持
 Set操作 SetSet支持SetSet支持Show flagsshow flags不支持 		增加RAM账号体系(add)	支持
set操作setset支持show flagsshow flags不支持자료		删除RAM账号体系(remove)	支持
Settart show flagsshow flags 不支持	cot语作	setset	支持
开启项目空间数据保护 (ProjectProtection)不支持添加可信项目空间 (add trustedproject)不支持移除可信项目空间 (remove trustedproject)不支持查看可信项目空间 (list 	Set珠叶	show flagsshow flags	不支持
安全操作添加可信项目空间 (add trustedproject)不支持移除可信项目空间 (remove trustedproject)不支持查看可信项目空间 (list trustedprojects)不支持查看项目空间的安全配置 (show SecurityConfiguration)不支持		开启项目空间数据保护 (ProjectProtection)	不支持
安全操作移除可信项目空间 (remove trustedproject)不支持查看可信项目空间 (list trustedprojects)不支持查看项目空间的安全配置 (show SecurityConfiguration)不支持		添加可信项目空间(add trustedproject)	不支持
查看可信项目空间(list trustedprojects) 不支持 查看项目空间的安全配置(show SecurityConfiguration) 不支持	安全操作	移除可信项目空间(remove trustedproject)	不支持
查看项目空间的安全配置(show SecurityConfiguration) 不支持		查看可信项目空间(list trustedprojects)	不支持
		查看项目空间的安全配置(show SecurityConfiguration)	不支持
添加用户 (add user) 支持		添加用户(add user)	支持
删除用户(remove user) 支持		删除用户(remove user)	支持
查看用户列表(list users) 支持		查看用户列表(list users)	支持
创建角色(create role) 不支持		创建角色(create role)	不支持
用户和角色操作 查看角色列表(list roles) 支持	用户和角色操作	查看角色列表(list roles)	支持

类别	常用命令	是否支持
	为用户授予角色(grant)	支持
	删除角色中的用户(revoke)	支持
	删除角色(drop role)	不支持
^运 扣 品 <i>作</i>	为角色或用户授权(grant)	支持
[文代][末]]F	撤销角色或用户授权(revoke)	支持
	创建表(create table)	支持
	修改表的所有人(alter table)	支持
主品作	删除表(drop table)	支持
衣床TF	查看表或视图信息(desc)	支持
	查看分区信息(desc partition)	支持
	列出空间下的表(show tables)	支持
	添加分区(alter table add)	支持
	删除分区(alter table drop)	支持
分区和列操作	添加列或注释(alter table add columns)	支持
	修改列名及注释(alter table change column)	支持
	查看实例信息(show)	不支持
	查看实例状态(status)	不支持
亦 /周语/在	查看正在执行的实例信息(top instance)	不支持
大四球日	停止实例(kill)	不支持
	获取实例对应的作业信息(desc)	不支持
	获取实例对应的作业运行日志信息 (wait)	不支持
	添加资源(add)	不支持
	查看资源列表(list)	支持
	为资源创建别名(alias)	不支持
资源操作		

类别	常用命令	是否支持
	下载资源(get)	不支持
	删除资源(drop)	不支持
	注册函数(create function)	不支持
云 北 坛	注销函数(drop function)	不支持
函数抹TF	查看函数(desc function)	不支持
	查看函数列表(list functions)	不支持
Tuppoll提作	上传数据(Tunnel upload)	不支持
T unnet珠TF	下载数据(Tunnel download)	不支持
时区配置操作	时区配置操作	不支持
其它操作	计量预估(cost sql)	不支持

5.1.2. SQL使用限制

本文为您介绍基于Dataphin使用MaxCompute SQL的限制项。

限制项	最大值/限制条件	分类	说明
表名长度	128字节	长度限制	表名、列名中不能有特殊字符,以字母 开头,且只能用英文小写字母(a- z)、英文大写字母(A-Z)、数字和 下划线(_)。
注释长度	1024字节	长度限制	长度不超过1024字节的有效字符串。
表的列定义	1200个	数量限制	单表的列定义个数最多为1200个。
单表分区数	60000个	数量限制	单表的分区个数最多为60000个。
表的分区层级	6级	数量限制	在表中创建的分区层次不能超过6级。
屏显	10000行	数量限制	SELECT语句屏显最多输出10000行。
INSERT 目标个数	256个	数量限制	MULTI-INSERT 场景,目标表的数 量限制为256个。
UNION ALL	256个	数量限制	UNION ALL 场景,最多允许合并 256个表。
MAPJOIN	128个	数量限制	MAPJOIN 场景,最多允许连接128 个小表。
数据开发·SQL使用说明

限制项	最大值/限制条件	分类	说明
MAPJOIN 内存限 制	512 MB	数量限制	MAPJOIN 场景,所有小表的内存不能超过512 MB。
ptinsubq	1000行	数量限制	子查询中存在分区列时,子查询的返回 结果不能超过1000行。
SQL语句长度	2 MB	长度限制	SQL语句的最大长度为2 MB。包括您使 用SDK调用SQL的场景。
WHERE 子句条件个 数	256个	数量限制	WHERE 子句中的条件个数最大为 256个。
列记录长度	8 MB	数量限制	表中单个单元的最大长度为8 MB。
IN的参数个数	1024	数量限制	IN的最大参数限制,例如 in (1,2,3,1024) 。如 果 in() 的参数过多,会影响编译 性能。1024为建议值,不是限制值。
jobconf.json	1 MB	长度限制	jobconf.json 的大小为1 MB。当 表包含的分区数量较多时,大小可能超 过 jobconf.json ,超过1 MB。
视图	不可写	操作限制	视图不支持写入,不支 持 INSERT 操作。
列的数据类型	不可修改	操作限制	不允许修改列的数据类型及列位置。
Java UDF函数	不允许 为 ABSTRACT 或 者 STATIC 。	操作限制	Java UDF函数不能 为 ABSTRACT 或 STATIC 。
最多查询分区个数	10000个	数量限制	最多查询分区个数不能超过10000个。
SQL执行计划长度	1 MB	长度限制	MaxCompute SQL生成的执行计划不 能超过1 MB, 否则会触发 FAILED: ODPS-0010000:System internal error - The Size of Plan is too large 报错。

⑦ 说明 上述MaxCompute SQL限制项均无法修改。

5.1.3. 数据类型

当前,Dataphin使用MaxCompute 2.0作为计算引擎。本文为您介绍Dataphin支持的MaxCompute 2.0的数 据类型。

适用场景

适用于在2020年04月之前无存量数据的MaxCompute项目,且该项目依赖的产品组件支持2.0数据类型版本。

基础数据类型

类型	常量示例	描述
TINYINT	1Y、-127Y	8位有符号整型。 取值范围:-128~127。
SMALLINT	327675、-1005	16位有符号整型。 取值范围:-32768~32767。
INT	1000、-15645787	32位有符号整型。 取值范围:-2 ³¹ ~2 ³¹ -1。
BIGINT	1000000000L、-1L	64位有符号整型。 取值范围:-2 ⁶³ +1~2 ⁶³ -1。
BINARY	unhex('FA34E10293CB42848573A4E399 37F479')	二进制数据类型,目前长度限制为8 MB。
FLOAT	cast(3.14159261E+7 as float)	32位二进制浮点型。
DOUBLE	3.14159261E+7	64位二进制浮点型。

类型	常量示例	描述
DECIMAL(precision,scale)	3.5BD、 99999999999999999998D	 10进制精确数字类型。 precision:表示最多可以表示多少位的数字。取值范围: 1 <= precision <= 38 。 scale:表示小数部分的位数。取值范围: 0 <= scale <= 18 。 如果不指定以上两个参数,则默认为 decimal(38,18) 。 ⑦ 说明 同一个表中不能同时存在新老Decimal数据类型。 执行 setproject odps.sql.hive.compatible=true;命令开启Hive兼容模式时, Decimal(precision, scale) 类型在数据导入(Tunnel Upload)和SQL运算过程中,如果数据的小数位数超过scale的大小,则会进行四舍五入的处理。如果整数部分超出限制,则会报错。
VARCHAR(n)	无	变长字符类型,n为长度。 取值范围:1~65535。
CHAR(n)	无	固定长度字符类型,n为长度。最大取值 255。长度不足则会填充空格,但空格不 参与比较。
STRING	"abc"、'bcd'、"alibaba"、'inc'	字符串类型,目前长度限制为8 MB。

类型	常量示例	描述
DATE	DATE'2017-11-11'	日期类型,格式为 yyyy-mm-dd 。 取值范围: 0000-01-01~9999-12-31。
DAT ET IME	DAT ET IME'2017-11-11 00:00:00'	日期时间类型。 取值范围: 0000-01-01 00:00:00.000~9999-12-31 23:59:59.999, 精确到毫秒。
T IMEST AMP	TIMESTAMP'2017-11-11 00:00:00.123456789'	时间戳类型。 取值范围: 0000-01-01 00:00:00.00000000~99999-12-31 23:59:59.999999999, 精确到纳秒。 ⑦ 说明 TIMESTAMP类型本身与 时区无关, 在任何时 区, TIMESTAMP类型存储的都是从 Epoch (UTC 1970-01-01 00:00:00) 开始的一个偏移量。您 可以通过内建函数对TIMESTAMP类 型数据执行与时区相关的计算, 如 cast(as string) 将TIMESTAMP类型数据 按照当前时区值转换成STRING类 型。
BOOLEAN	True, False	BOOLEAN类型。 取值范围:True、False。

数据类型说明如下:

- 上述的各种数据类型均可以为NULL。
- SQL中的INT关键字是32位整型。

```
--将a转换为32位整型。
cast(a as INT)
```

- 整型常量的语义会默认为INT类型。例如 SELECT 1 + a; 中的整型常量1会被作为INT类型处理。如果常量过长,超过了INT的值域而又没有超过BIGINT的值域,则会作为BIGINT类型处理;如果超过了BIGINT的值域,则会被作为DOUBLE类型处理。
- 隐式转换
 - 部分隐式类型转换会被禁用。例如,STRING->BIGINT、STRING->DATETIME、DOUBLE->BIGINT、 DECIMAL->DOUBLE、DECIMAL->BIGINT有精度损失或者报错的风险。禁用类型可以通过CAST函数强制 进行数据类型转换。
 - VARCHAR类型常量可以通过隐式转换为STRING常量。
- 表、函数以及UDF

- 。参数涉及2.0数据类型的内置函数,在2.0数据类型版本下可以正常使用。
- UDF包含的数据类型都会按照2.0数据类型进行解析重载。
- 分区列支持STRING、VARCHAR、CHAR、TINYINT、SMALLINT、INT、BIGINT数据类型。
- STRING常量支持连接,例如, abc和xyz会解析为abcxyz。
- 给DECIMAL字段插入常量时,常量的写法需要与常量定义中的格式保持一致。例如下面示例代码中的 3.5 BD 。

insert into test_tb(a) values (3.5BD)

DATETIME查询显示的时间值不包含毫秒。Tunnel命令通过 -dfp 来指定时间格式,可以指定到毫秒显示,例如 tunnel upload -dfp 'yyyy-MM-dd HH:mm:ss.SSS'。关于Tunnel命令的更多信息,请参见Tunnel命令。

复杂数据类型

类型	定义方法	构造方法
ARRAY	array<int></int>array<struct<a:int, b:string="">></struct<a:int,>	 array(1, 2, 3) array(array(1, 2), array(3, 4))
МАР	map<string, string=""></string,>map<smallint, array<string="">></smallint,>	 map("k1", "v1", "k2", "v2") map(1S, array('a', 'b'), 2S, a rray('x', 'y'))
STRUCT	 struct<x:int, y:int=""></x:int,> struct<field1:bigint, field2:a<br="">rray<int>, field3:map<int, int="">></int,></int></field1:bigint,> 	<pre>named_struct(`x', 1, `y', 2) named_struct(`field1', 100L, ` field2', array(1, 2), `field3', map(1, 100, 2, 200))</pre>

⑦ 说明 MaxCompute的复杂数据类型可以被嵌套使用,相关的内建函数说明请参见ARRAY、MAP或STRUCT。

5.1.4. 运算符

运算符用于执行程序代码运算。本文为您介绍Dataphin支持的MaxCompute的运算符,包括关系运算符、算术运算符、位运算符和逻辑运算符。

关系运算符

运算符	说明
A=B	 A或B为NULL,返回NULL。 A等于B,返回TRUE,否则返回FALSE。

运算符	说明
A<>B	 A或B为NULL,返回NULL。 A不等于B,返回TRUE,否则返回FALSE。
A <b< td=""><td> A或B为NULL,返回NULL。 A小于B,返回TRUE,否则返回FALSE。 </td></b<>	 A或B为NULL,返回NULL。 A小于B,返回TRUE,否则返回FALSE。
A<=B	 A或B为NULL,返回NULL。 A小于等于B,返回TRUE,否则返回FALSE。
A>B	 A或B为NULL,返回NULL。 A大于B,返回TRUE,否则返回FALSE。
A>=B	 A或B为NULL,返回NULL。 A大于等于B,返回TRUE,否则返FALSE。
A IS NULL	A为NULL,返回TRUE,否则返回FALSE。
A IS NOT NULL	A不为NULL,返回TRUE,否则返回FALSE。
A LIKE B	如果A或B为NULL,返回NULL。A为字符串,B为要匹配的模式,如果匹配,返回TRUE,否则 返回FALSE。 • 百分号(%)匹配任意多个字符。 • 下划线(_)匹配单个字符。 • 要匹配百分号(%)或下划线(_)要用转义符表示为 '%' 或 '_' 。 'aaa' like 'a_'= TRUE 'aaa' like 'a&' = TRUE 'aaa' like 'a&' = TRUE 'aaa' like 'a&' = TRUE 'aab' like 'a\\%b'= TRUE 'axb' like 'a\\%b'= FALSE
A RLIKE B	A是字符串,B是字符串常量或者正则表达式。如果匹配成功,返回TRUE,否则返回FALSE。 如果B为空串会报错。如果A或B为NULL,返回NULL。
A IN B	 B为一个集合,如果A在B中,返回TRUE,否则返回FALSE。 如果A为NULL,返回NULL。 如果B仅包含一个NULL元素,即AIN(NULL),则返回NULL。 B必须是常数集合且至少包含一个项,且所有项的数据类型需要一致。 ⑦ 说明 如果B集合中包含NULL元素同时也包含其它元素,则将NULL视为与B集合中其它元素的数据类型一致。

运算符	说明
BET WEEN AND	表达式为 A [NOT] BETWEEN B AND C 。 • A、B或C为NULL,则为空。 • A大于等于B且小于等于C,返回TRUE,否则返回FALSE。
IS [NOT] DISTINCT FROM	表达式为 A IS [NOT] DISTINCT FROM B 。详情请参见IS DISTINCT FROM和IS NOT DISTINCT FROM。

算术运算符

运算符	说明
A+B	A或B为NULL,返回NULL,否则返回A+B的结果。
A-B	A或B为NULL,返回NULL,否则返回A-B的结果。
A*B	A或B为NULL,返回NULL,否则返回A×B的结果。
	A或B为NULL,返回NULL,否则返回A÷B的结果。
A/B	⑦ 说明 如果A和B为BIGINT类型,返回结果为DOUBLE类型。
A%B	A或B为NULL,返回NULL,否则返回A÷B并取余数的结果。
+A	仍然返回A。
-A	如果A为NULL,返回NULL,否则返回-A。
A DIV B	A或B为NULL,返回NULL,否则返回A DIV B的结果。

常见用法如下。

SELECT age+10, age-10, age%10, -age, age*age, age/10, age div 10 FROM user;

? 说明

- 只有STRING、BIGINT或DOUBLE类型的参数才能参与算术运算,日期和布尔类型不允许参与运算。
- STRING类型在参与运算前会隐式转换为DOUBLE类型。
- BIGINT和DOUBLE类型共同参与计算时,系统会将BIGINT类型隐式转换为DOUBLE类型再进行计算,返回结果为DOUBLE类型。
- A和B都是BIGINT类型,执行A/B运算,返回结果为DOUBLE类型。执行上述其他运算,返回BIGINT 类型。

位运算符

运算符	说明
A&B	返回A和B进行按位与运算的结果。例如1&2返回0,1&3返回1,NULL和任何值按位与运算都为 NULL。A和B必须为BIGINT类型。
A B	返回A和B进行按位或运算的结果。例如1 2返回3,1 3返回3,NULL和任何值按位或运算都为NULL。 A和B必须为BIGINT类型。
A B	字符串连接操作符。例如 a b c 相当于 CONCAT(a, b, c) 。

⑦ 说明 位运算符不支持隐式转换,只允许BIGINT类型参与运算。

逻辑运算符

运算符	说明
	TRUE and TRUE=TRUE
	TRUE and FALSE=FALSE
	FALSE and TRUE=FALSE
	FALSE and FALSE=FALSE
A and B	FALSE and NULL=FALSE
	NULL and FALSE=FALSE
	TRUE and NULL=NULL
	NULL and TRUE=NULL
	NULL and NULL=NULL
	TRUE or TRUE=TRUE
	TRUE or FALSE=TRUE
	FALSE or TRUE=TRUE
	FALSE or FALSE=FALSE
	FALSE or NULL=NULL
A or B	NULL or FALSE=NULL
	TRUE or NULL=TRUE
	NULL or TRUE=TRUE
	NULL or NULL=NULL

运算符	说明
	A是NULL,返回NULL。
NOT A	A是TRUE,返回FALSE。
	A是FALSE,返回TRUE。

⑦ 说明 逻辑运算符只允许BOOLEAN类型参与运算,不支持隐式类型转换。

5.1.5. DDL语句

MaxCompute中的DDL语句包括对表、分区、列、生命周期和视图的操作。本文为您介绍Dataphin支持的MaxCompute的DDL语句。

表操作

功能	当前是否支持
创建表	是 ⑦ 说明 对于创建的外部表, Dataphin支持使用 msck repair补全外部表分区数据语法格式。具体操 作,请参见创建OSS外部表。
修改表的注释	是
修改表的修改时间	是
修改表的聚簇属性	支持增加聚簇属性不支持删除聚簇属性
重命名表	是
删除表	是
查看表或视图信息	是
查看分区信息	是
列出项目空间下的表和视图	是
列出所有分区	是
修改表的所有人	否
清空非分区表里的数据	否

功能	当前是否支持
查看建表语句	否

分区和列操作

功能	是否支持
添加分区	支持添加单个分区不支持添加多个分区
删除分区	 支持删除单个分区 不支持删除多个分区 不支持根据过滤条件删除分区
添加列或注释	是
修改列名	是
修改列注释	是
修改分区的更新时间	是
修改分区值	是
合并分区	否

生命周期操作

功能	是否支持
修改表的生命周期	是
禁止或恢复生命周期	是

视图操作

功能	是否支持
创建或更新视图	是
重命名视图	是
删除视图	是

5.1.6. DML操作

MaxCompute中的DML语句包括插入、更新和删除数据操作。本文为您介绍Dataphin支持的MaxCompute的DML操作语句。

功能	是否支持
插入或覆写数据(INSERT INTO INSERT OVERWRITE)	 支持 insert {into overwrite} table [partition (<pt_spec>)] <select_st atement=""> from <from_statement>;</from_statement></select_st></pt_spec> 不支持指定列插入数据: insert into table <t able_name=""> [partition (<pt_spec>)] (<col_name> ,<col_name>) <select_statement> from <from_statement>;</from_statement></select_statement></col_name></col_name></pt_spec></t> 不支持插入数据排序: insert {into overwrite } table <table_name> [partition (<pt_spec>)] [(<col_name> [,<col_name>)] <sele ct_statement=""> from <from statement=""> [zord er by <zcol_name> [, <zcol_name>]];</zcol_name></zcol_name></from></sele></col_name></col_name></pt_spec></table_name>
插入或覆写动态分区数据(DYNAMIC PART IT ION)	是
多路输出(MULT I INSERT)	是
VALUES	 支持 insert … values 不支持 values table
LOAD	否
COMMON TABLE EXPRESSION (CTE)	 支持 with table as (select) insert in to tab select from table 支持 with table as (select) select . from table
更新或删除数据(UPDATE DELETE)	支持 delete from <table_name> [where <where_condition>]</where_condition></table_name>

5.1.7. DQL操作

MaxCompute中的DQL语句包括SELECT、JOIN、WHERE子句等。本文为您介绍Dataphin支持的MaxCompute的DQL操作语句。

功能	是否支持
命令格式	是
列表达式(select_expr)	 支持 * 支持 distinct 不支持正则表达式
WHERE子句(where_condition)	Dataphin 2.9.2及之后版本支持UDF分区裁剪

智能数据构建与管理 Dat aphin

功能	是否支持
当SQL语句设置了group by属性,即 set odps.sql.groupby.position.alias=true; , gro up by中的整型常量会被当做select的列序号处理,详情 请参见GROUP BY分组查询 (col_list)	是
基础子查询	 支持 select <select_expr> from (<select_s tatement="">) <sq_alias_name>;</sq_alias_name></select_s></select_expr> 不支持 select (<select_statement>) from < table_name>;</select_statement>
IN SUBQUERY	是
NOT EXISTS SUBQUERY	是
SCALAR SUBQUERY	是
交集、并集和补集	 支持 intersect 不支持 intersect all 不支持 intersect distinct
并集	 支持 union all 支持 union [distinct]
补集	 支持 except 支持 minus 不支持 except distinct 不支持 minus distinct 不支持 except all 不支持 minus distinct
JOIN	是
SEMI JOIN(半连接)	是
MAPJOIN HINT	是
Lateral View	是
GROUPING SET S	是
COMMON TABLE EXPRESSION (CTE)	是
CLONE T ABLE	否
SELECT TRANSFORM	否

功能	是否支持
EXPLAIN	否
执行语序	否

5.1.8. SQL增强操作

本文为您介绍Dataphin支持的SQL增强操作。

功能	是否支持
EXPLAIN	是

5.1.9. 函数操作

本文为您介绍Dataphin支持的函数操作。

功能	是否支持
函数操作	是
MAX_PT	是 支持使用Dataphin项目名称进行检索数据。例 如,Dataphin的项目英文名称为biz_cdm,该项目下有 分区表biz_order、分区键为ds。则您可以执行以下语句 查询数据。
	<pre>select * from biz_cdm.biz_order where ds = max_pt('biz_cdm.biz_order');</pre>