

ALIBABA CLOUD

阿里云

智能数据构建与管理 Dataphin 快速入门

文档版本：20220527

 阿里云

法律声明

阿里云提醒您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。

1. 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。
2. 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。
3. 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。	 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。	 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 <code>cd /d C:/window</code> 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] 或者 [a b]	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
{ } 或者 {a b}	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1. Dataphin使用流程	05
2. 步骤一：准备工作	06
3. 步骤二：规划数仓	07
4. 步骤三：引入数据	13
5. 步骤四：规范定义	18
6. 步骤五：规范建模	19
7. 步骤六：开发环境补数据	28
8. 步骤七：验证数据	29
9. 步骤八：发布任务	31
10. 步骤九：生产环境补数据	32
11. 步骤十：查看运行结果	33

1.Dataphin使用流程

本教程将指引您快速完成一次完整的数仓建设。

基于Dataphin构建数仓的基本流程如下。

主流程	说明	操作指导
准备工作	开通和配置Dataphin、设置Dataphin计算引擎、创建MaxCompute项目和准备数据源。	步骤一：准备工作
规划数仓	规划本入门教程的数仓，包括创建业务板块、计算源、数据源、项目及项目中的成员。	步骤二：规划数仓
数据集成	同步源数据源中的数据至目标数据源。	步骤三：引入数据
规范定义	基于Dataphin支持的数据研发模块，分析数据并明确统计指标。包括维度、业务过程、原子指标、业务限定和派生指标。	步骤四：规范定义
规范建模	基于规范定义部分中数据的圈定，利用Dataphin中的研发模块进行数据的研发。	步骤五：规范建模
开发环境补数据	在开发环境中补数据，包括维度逻辑表、事实逻辑表、汇总逻辑表和管道任务。验证数据模型是否正确及符合预期。	步骤六：开发环境补数据
验证数据	您可以通过即席查询来验证数据，验证通过的数据即可发布至生产环境。	步骤七：验证数据
发布任务	发布开发环境的数据至生产环境，参与生产环境的运维调度。	步骤八：发布任务
生产环境补数据	发布至生产环境的数据次日才会参与调度，本教程为了让您快速熟悉智能数据建设与治理的流程，采用补数据的方式，模拟生产环境调度任务生成实例。	步骤九：生产环境补数据
查看运行结果	查看生产环境中数据运行结果是否符合预期，即汇总逻辑表是否能够统计到各省份办公用品和技术产品最近1天的销售总额。	步骤十：查看运行结果

2. 步骤一：准备工作

本文为您介绍本入门教程的准备工作。

创建MaxCompute（ODPS）项目

主流程	描述	操作指导
准备阿里云账号	创建阿里云账号。	准备阿里云账号
配置AccessKey	授权主账号的AccessKey至Dataphin。	同步RAM用户账号至Dataphin
创建RAM用户	创建主账号的RAM用户。	准备RAM用户
开通MaxCompute	购买MaxCompute。	开通MaxCompute和DataWorks
设置计算引擎	设置Dataphin的计算引擎为MaxCompute。	设置Dataphin实例的计算引擎
创建工作空间	创建Dataphin的计算源： <ul style="list-style-type: none"> 创建开发环境的计算源，配置如下参数，其余参数保持默认： <ul style="list-style-type: none"> 工作空间名称配置为test_dev_odps。 选择计算引擎服务选择为MaxCompute。 实例显示名称配置为test。 创建生产环境的计算源，配置如下参数，其余参数保持默认： <ul style="list-style-type: none"> 工作空间名称配置为test_prod_odps。 选择计算引擎服务选择为MaxCompute。 实例显示名称配置为test1。 	创建MaxCompute项目

开通和配置Dataphin

主流程	描述	操作指导
开通Dataphin	购买Dataphin。	开通Dataphin
配置AccessKey	授权主账号的AccessKey至Dataphin。	同步RAM用户账号至Dataphin
设置计算引擎	设置Dataphin的计算引擎为MaxCompute。	设置Dataphin实例的计算引擎

准备数据源

本入门教程以统计某年办公用品、技术产品两类产品各省份每天的销售总额为例。

- 通过RDS创建MySQL实例，获取RDS实例ID，并在RDS控制台添加白名单，详情请参见[创建RDS MySQL实例](#)。本教程中RDS数据源名称以dataphin为例。

说明

- 地域选择华东2（上海）。
- 网络类型选择专有网络。

- 下载本教程中使用的数据表格[company_sales_record_copy](#)。
- 上传数据表格company_sales_record_copy至新建的RDS数据源，详情请参见[将Excel的数据导入数据库](#)。

注意 如果您的RDS实例中已经有数据表，您也可以不使用company_sales_record_copy数据表。通过执行如下命令查询建表的DDL语句。

```
show create table <table name>;
```

后续您可以依据该表的DDL语句，创建同步数据的目标表。如何创建同步数据的目标表，请参见[步骤三：引入数据](#)。

3. 步骤二：规划数仓

本文为您介绍如何规划本入门教程的数仓。

前提条件

在开始执行操作前，请确认您已满足以下条件：

- 完成Dataphin的开通及配置，详情请参见[步骤一：准备工作](#)。
- 完成Dataphin数据源的构建，详情请参见[步骤一：准备工作](#)。
- 完成MaxCompute项目的创建，详情请参见[步骤一：准备工作](#)。

创建业务板块和数据域

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[规划](#)。
4. 按照下图指引，进入新建业务板块对话框。



5. 创建业务板块。
 - i. 在新建业务板块对话框，选择Dev-Prod模式后，单击下一步。
 - ii. 在新建业务板块对话框中，配置参数。

参数	描述
公用英文名	输入为beginner_test。
公用名称	输入为入门教程。
描述	输入简单的描述。例如，入门教程。
图标	选择 图标。
业务板块管理员	您可以选择业务板块管理员，也可以不选择。

- iii. 单击确定，完成业务板块（LD_beginner_test和LD_beginner_test_dev）的创建。

6. 创建交易域。
 - i. 在LD_beginner_test业务板块页面，单击页面右侧的[新建数据域](#)。
 - ii. 在新建数据域对话框中，配置参数。

参数	描述
数据域名称	输入test_dataphin。
数据域英文名	输入test_dataphin。
英文缩写	输入test。
描述	输入简单的描述，例如交易。

- iii. 单击确定。

创建计算源

1. 在规划页面，单击左侧导航栏中的[计算源](#)。

2. 创建Dev项目计算源。

- i. 在计算源页面中，鼠标悬停至页面右上方的新增计算源。
- ii. 在新建计算源对话框中，配置参数。

新建计算源
✕

* 计算源类型 MaxCompute ▼

* 计算源名称 test_dev_odps

计算源描述 Dev项目的计算源 9/128

* Endpoint http://service.cn-shanghai.maxcompute.aliyun-inc.com/api

* Project Name test_dev_odps

* Access ID tD6D [REDACTED]

* Access Key

测试连接 取消 提交

参数	描述
计算源类型	默认为MaxCompute，不支持修改。
计算源名称	输入test_dev_odps。
计算源描述	输入Dev项目的计算源。
Endpoint	默认为Dataphin实例计算引擎的Endpoint，不支持修改。
Project Name	输入test_dev_odps（Dataworks创建的MaxCompute（ODPS）项目名称）。
Access ID	访问密钥中的AccessKey ID，您可以通过 用户信息管理 页面获取。
Access Key	访问密钥中的AccessKey Secret，您可以通过 用户信息管理 页面获取。

- iii. 单击测试连接。
- iv. 测试连接成功后，单击提交，完成Dev项目计算源的创建。

3. 创建Prod项目计算源。

- i. 在计算源页面中，鼠标悬停至页面右上方的新增计算源，单击离线计算源。

ii. 在离线计算源对话框中，配置参数。

新建计算源
✕

* 计算源类型 ▾

* 计算源名称

计算源描述 10/128

* Endpoint

* Project Name

* Access ID

* Access Key

测试连接
取消
提交

参数	描述
计算源类型	默认为 MaxCompute ，不支持修改。
计算源名称	输入test_prod_odps。
计算源描述	输入Prod项目的计算源。
Endpoint	默认为Dataphin实例计算引擎的Endpoint，不支持修改。
Project Name	输入test_prod_odps（Dataworks创建的MaxCompute（ODPS）项目名称）。
Access ID	访问密钥中的AccessKey ID，您可以通过 用户信息管理 页面获取。
Access Key	访问密钥中的AccessKey Secret，您可以通过 用户信息管理 页面获取。

iii. 单击**测试连接**。

iv. 测试连接成功后，单击**提交**，完成Prod项目计算源的创建。

创建项目

1. 在规划页面，单击左侧导航栏中的**项目管理**。
2. 在**新建项目**对话框中，选择**Dev-Prod**模式后，单击**下一步**。
3. 在**新建项目**对话框中，配置如下参数，其他参数保持默认。

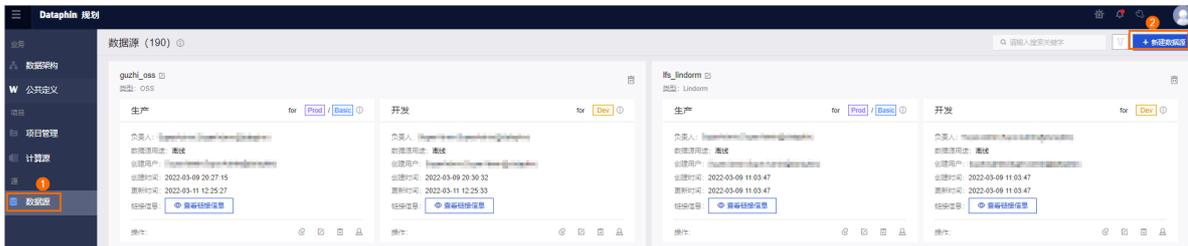
参数	描述
公用名称	输入入门教程。
公用英文名	输入beginner_test。
业务板块	选择LD_beginner_test。
空间类型	选择应用层。

参数	描述
项目1: Dev	选择离线计算源为test_dev_odps。
项目2: Prod	选择离线计算源为test_prod_odps。
描述	输入简单的描述，例如入门教程。
沙箱白名单	添加沙箱白名单： <ol style="list-style-type: none"> i. 单击新建。 ii. 在访问地址输入框中，输入RDS ID.mysql.rds.aliyuncs.com。在访问地址输入框中，输入3306。 iii. 单击图标。

4. 单击**确定**，完成项目（beginner_test和beginner_test_dev）的创建。

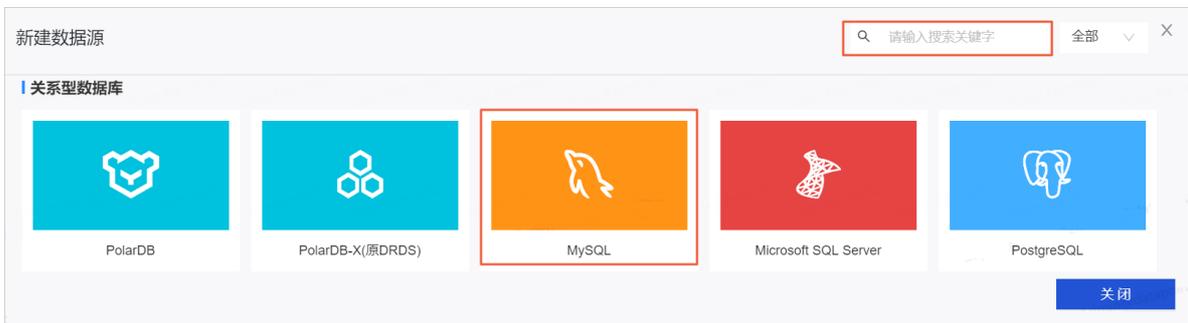
创建源数据源

1. 按照下图指引，进入**新建数据源**页面。



2. 在**新建数据源**对话框的**关系型数据库**区域，选择**MySQL**。

如果您最近使用过MySQL，也可以在**最近使用**区域选择MySQL。同时，您也可以在快速搜索框中，输入MySQL的关键词，快速筛选。



3. 在**新建MySQL数据源**对话框，配置数据源信息。

返回新建数据源 > 新建MySQL数据源
✕

* 数据源名称

* 版本

数据源描述 0/128

* 数据源配置 "生产+开发"数据源 "生产"数据源

生产数据源

* JDBC URL

* 用户名

* 密码

* 类型 公网数据库 阿里云数据库 ECS(VPC)自建数据库

* 访问方式 VPC网络 公网IP

* 区域

* VPC ID

* RDS ID

提示：当前使用的数据库为VPC网络环境下阿里云数据库，连接请指定授权IP白名单：100.104.238.64/26

开发数据源

* JDBC URL

* 用户名

* 密码

* 类型 公网数据库 阿里云数据库 ECS(VPC)自建数据库

* 访问方式 VPC网络 公网IP

* 区域

* VPC ID

* RDS ID

提示：当前使用的数据库为VPC网络环境下阿里云数据库，连接请指定授权IP白名单：100.104.238.64/26

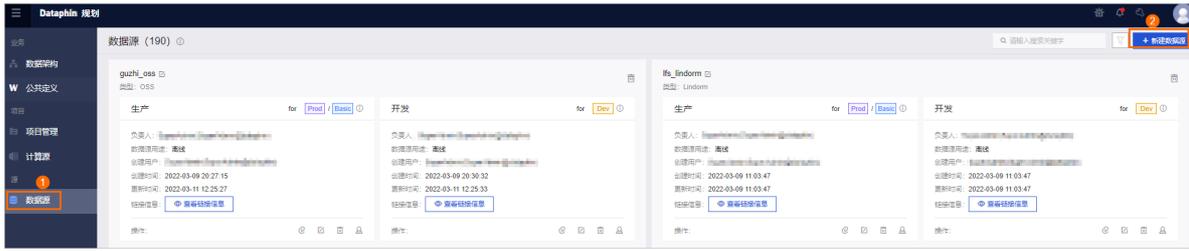
测试连接 取消 确定

参数	描述
数据源类型	选择MySQL。
数据源名称	输入为dataphin。
版本	选择为MySQL 5.x。
数据源描述	对数据源简单的描述。
数据源配置	选择生产+开发数据源。
JDBC URL	输入格式为jdbc:mysql://RDS ID.mysql.rds.aliyuncs.com:3306/dataphin。将RDS ID更换为您创建的RDS实例ID。
用户名	输入为dataphin。
密码	输入您新建RDS时创建的用户名为dataphin的密码。
类型	选择阿里云数据库。
访问方式	选择VPC网络。
区域	选择华东2（上海）。
VPC ID	输入您新建实例RDS时选择的专有云网络的VPC ID。
RDS ID	输入您创建的RDS实例ID。

4. 单击测试连接。
5. 连通性测试通过后，单击确定。

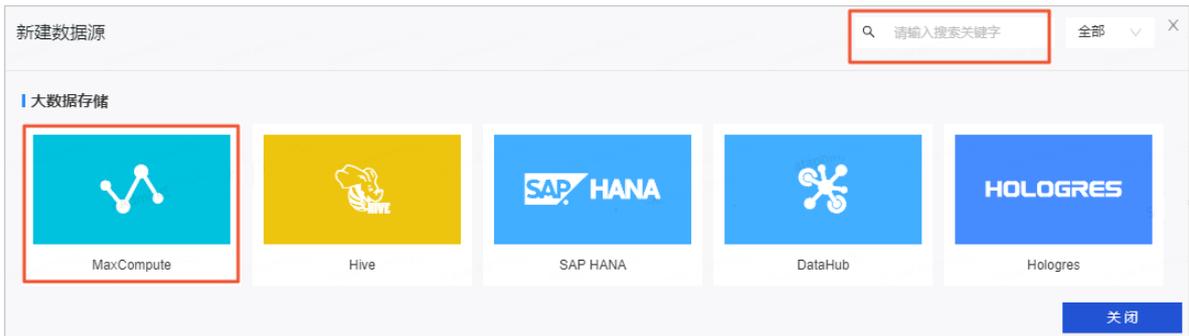
创建目标数据源

1. 按照下图指引，进入新建数据源页面。



2. 在新建数据源对话框的大数据存储区域，选择MaxCompute。

如果您最近使用过MaxCompute，也可以在最近使用区域选择MaxCompute。同时，您也可以在最搜索框中，输入MaxCompute的关键词，快速筛选。



3. 在新建MaxCompute数据源对话框，配置数据源信息。

参数	描述
数据源类型	选择MaxCompute。
数据源名称	输入为目标数据源。
数据源描述	输入目标数据源。
数据源配置	选择生产+开发数据源。
Endpoint	输入HTTP协议的链接地址service.cn.maxcompute.aliyun.com/api。
Project Name	开发环境输入为test_dev_odps，生产环境输入为test_prod_odps。
Access ID	访问密钥中的AccessKey ID，您可以通过 用户信息管理 页面获取。
Access Key	访问密钥中的AccessKey Secret，您可以通过 用户信息管理 页面获取。

4. 单击测试连接。

5. 连通性测试通过后，单击确定。

4. 步骤三：引入数据

本文为您介绍如何将数据源的数据集成您的工作空间。

前提条件

- 完成数据源的创建，详情请参见[创建源数据源](#)。
- 完成目标数据源的创建，详情请参见[创建目标数据源](#)。

背景信息

- 如果您是在2020年4月份之后购买的Dataphin，则数据同步能力已升级为数据集成。数据集成为您提供简单高效、安全可靠的数据同步平台。
- 如果您是在2020年4月份之前购买的Dataphin，则数据源的数据通过数据同步的方式，同步到您的工作空间，详情请参见[数据同步](#)。

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[数据集成](#)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，单击Dev页签，选择beginner_test_dev为数据开发的项目空间。
 - iii. 在[数据开发](#)页面，鼠标悬停至顶部菜单栏中的[开发](#)上，单击[集成](#)。
4. 创建管道开发脚本。
 - i. 在[数据集成](#)页面，鼠标悬停在图标，单击[离线单条管道](#)。
 - ii. 在[创建管道开发脚本](#)对话框，配置参数。

参数	描述
管道名称	输入test。
调度类型	调度类型选择周期性节点。
描述	填写对离线单条管道的简单描述。
选择目录	默认目录为离线管道。

- iii. 单击[确定](#)。
5. 开发离线管道脚本。
 - i. 在[离线单条管道开发](#)页面，单击右上方的[组件库](#)。
 - ii. 单击输入前的图标后，选中MySQL组件并拖动至左侧的管道画布中。

iii. 鼠标悬停至组件框内并右键单击，选择属性配置。在MySQL输入配置对话框，配置参数。

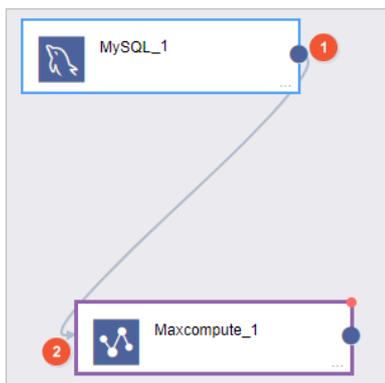


参数	描述
步骤名称	保持默认。
数据源	选择数据源（dataphin）。
来源表量	选择单表。
表	选择来源表company_sales_record_copy。
切分键	无需添加切分键。
输入过滤	无需添加过滤条件。
输出字段	查看输出字段。

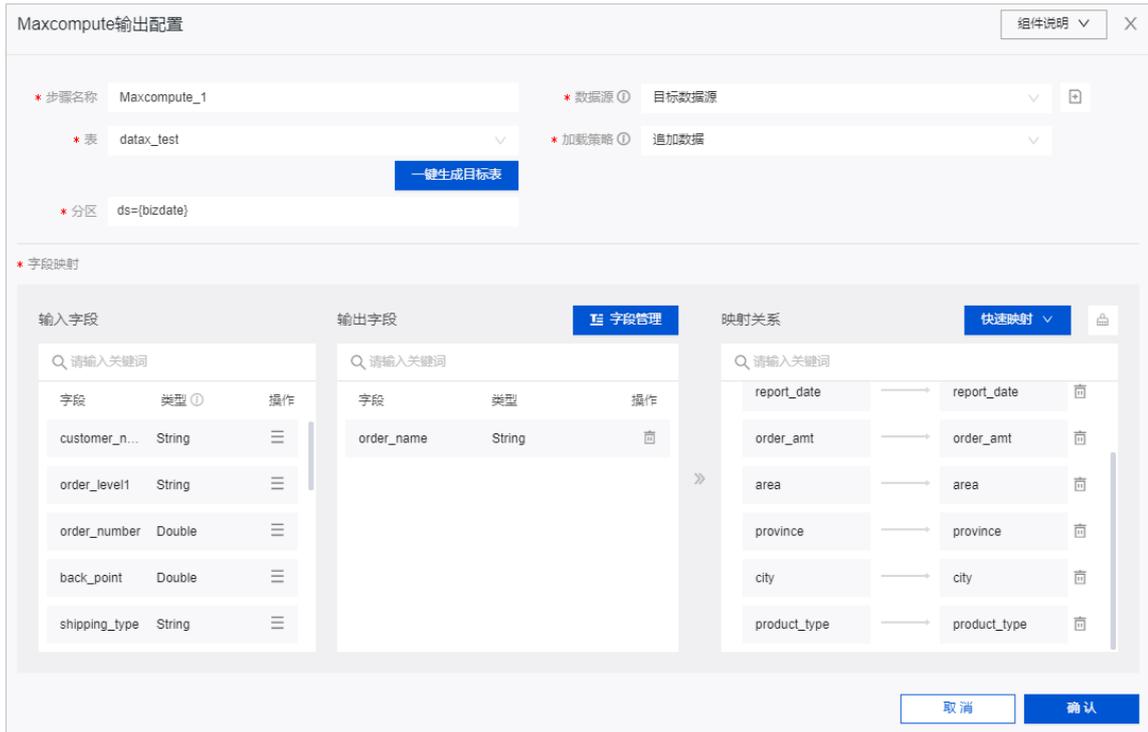
iv. 单击确认，完成输入组件的属性配置。

v. 单击输出前的 > 图标后，选中MaxCompute组件并拖动至左侧的管道画布中。

vi. 单击下图中输入组件（MySQL）①处后拖动并指向输出组件（MaxCompute）的②处，形成有向连线。



vii. 鼠标悬停至输出组件框内并右键单击，选择属性配置。在MaxCompute输出配置对话框，配置参数。



参数	说明
步骤名称	保持默认。
数据源	选择目标数据源。
表	<p>创建目标表：</p> <ol style="list-style-type: none"> 单击一键生成目标表。 在代码输入框中，输入建表语句。 <pre>CREATE TABLE IF NOT EXISTS datax_test (order_id bigint comment '订单号', `area` string comment '区域', province string comment '省份', city string comment '城市', product_type string comment '类型', order_name string comment '客户名称', report_date datetime comment '日期', order_amt double comment '销售额') PARTITIONED BY (`ds` STRING);</pre> <ol style="list-style-type: none"> 单击新建。
加载策略	选择 追加数据 。
分区	输入ds=\${bizdate}。
输入字段	根据上游的输入，为您展示输入字段。
输出字段	为您展示输出字段。
快速映射	<p>映射关系选择为同名映射的操作步骤：</p> <ol style="list-style-type: none"> 单击快速映射后的图标。 选择同名映射。 在提醒对话框中，单击确定。

viii. 单击**确认**，完成输出组件的属性配置。

6. 单击管道开发脚本页面左上方的预览。在对话框中，**bizdate**填写为20200810，单击确定。

order_id	report_date	customer_name	order_level1	order_number	order_amt	back_point	shipping_type	profit_amt	shipping_cos
13729	2013-01-01...		其它	9	872.48	0.08	空运	-342.91	35
28774	2013-01-01...		高级	33	180.36	0.1	火车	-111.8	4.69
37537	2013-01-02...		低级	43	4083.19	0.07	大卡	-1049.85	45
37537	2013-01-02...		低级	32	4902.38	0.05	火车	1438.49	7.07
37537	2013-01-02...		低级	4	1239.06	0	大卡	-193.08	48.8

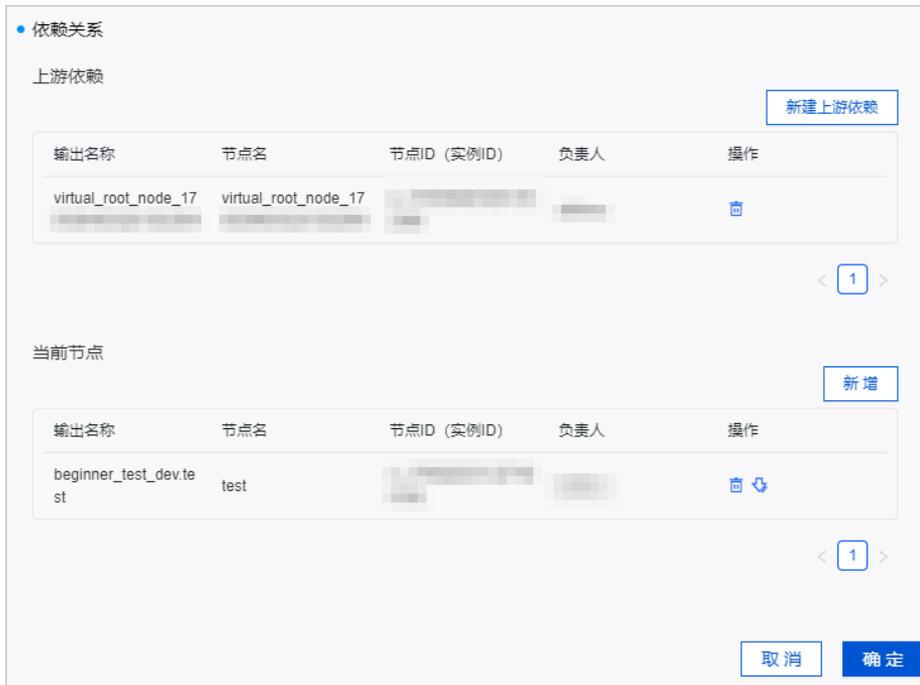
7. 单击管道开发脚本页面左上方的执行。在对话框中，**bizdate**填写为20200810，单击确定。集成源数据源的数据至您的目标数据源。

运行日志		步骤度量	预览结果			
步骤名称	线程号	输入 (条记录)	输出	读取	写入	时间
MySQL_1	0	17136	0	0	17136	2020-08-10 18:14:27
Maxcompute_1	0	0	17136	17136	0	2020-08-10 18:14:27

8. 配置管道脚本的调度参数。

- i. 在管道脚本页面，单击调度配置。

ii. 在调度配置页面，配置依赖关系区域参数，其他参数保持默认。



参数	描述
上游依赖	添加上游依赖： <ol style="list-style-type: none"> 单击新建上游依赖。 在新建上游依赖对话框，输入virtual_root_node，搜索并选择虚拟节点。 单击确定新增。
当前节点	添加当前节点的输出节点： <ol style="list-style-type: none"> 单击新增。 在新增当前节点输出对话框，输入beginner_test_dev.test。 单击确定新增。

iii. 单击确定。

9. 保存、提交和发布离线单条管道脚本。

- 单击页面右上方的图标，保存管道脚本。
- 单击页面右上方的图标，提交管道脚本。
- 在提交备注对话框，输入备注信息。
- 单击确定并提交。
- (可选) 发布离线单条管道脚本至生产环境。
 - 如果您的开发模式是Dev-Prod，则需要发布已提交的离线单条管道脚本，详情请参见[管理发布任务](#)。
 - 如果您的开发模式是Basic，则提交成功的离线单条管道脚本，即可参与生产环境的调度。

5.步骤四：规范定义

规范定义是指以维度建模作为理论基础，构建总线矩阵，划分并定义数据域、业务过程、维度、原子指标、业务限定、时间周期和派生指标。本文为您介绍在本教程中如何进行规范定义。

实例数据说明

本示例以统计办公用品、技术产品两类产品各省份每天的销售总额为例，在company_sales_record_copy表中，选取如下字段进行分析。

字段	描述
order_id	表示订单号。
area	表示区域。
province	表示省份。
city	表示城市。
product_type	表示产品类型。
customer_name	表示客户名称。
report_date	表示订单日期。
order_amt	表示销售额。

基本概念

- 维度

维度是度量的基础，用来反映业务的一类属性，这类属性的集合构成一个维度，也可以称为一个实体对象。在划分数据域、构建总线矩阵时，需要结合对业务过程的分析定义维度。本教程中，以创建某年办公用品、技术产品两类产品各省份每天的销售总额指标为目标，故以province为维度进行数据建模。

- 业务过程

业务过程是指企业的业务活动事件，如下单、支付、退款都是业务过程。通常业务过程是企业活动中的事件，因此业务过程是一个不可拆分的行为事件。本教程中业务过程即下单购买，使用order_pay标识。

- 原子指标

基于某一业务事件行为下的度量，是业务定义中不可再拆分的指标，具有明确业务含义的名词。原子指标=业务过程（动作）+度量，如支付（事件）金额（度量）。本教程中以order_amt为度量并对其进行汇总创建原子指标，则sum（order_amt）是原子指标。

- 业务限定

业务限定为统计的业务范围的圈定。为保障所有统计指标统一、标准、规范地构建，业务限定在业务板块内唯一，并唯一归属于一个来源逻辑表，计算逻辑也以该来源逻辑表模型的字段为基础进行定义。本教程中以 product_type='技术产品' 和 product_type='办公用品' 为业务限定。

- 派生指标

派生指标=原子指标+业务限定+统计周期+维度（统计粒度）。派生指标即为对原子指标进行业务统计范围的圈定。本教程中派生指标为各省份办公用品、技术产品最近一天的销售总额，省份（province）为维度，下单购买产品为业务过程，销售总额为原子指标即sum（order_amt），产品类型限定为技术产品和办公用品为业务限定，统计周期为最近一天。

6.步骤五：规范建模

本文为您介绍如何基于本教程完成数据建模。

前提条件

- 完成数据源及项目空间的准备，详情请参见[步骤一：准备工作](#)。
- 完成数仓规划，详情请参见[步骤二：规划数仓](#)。
- 完成数据集成，详情请参见[步骤三：引入数据](#)。
- 完成规范定义，详情请参见[步骤四：规范定义](#)。

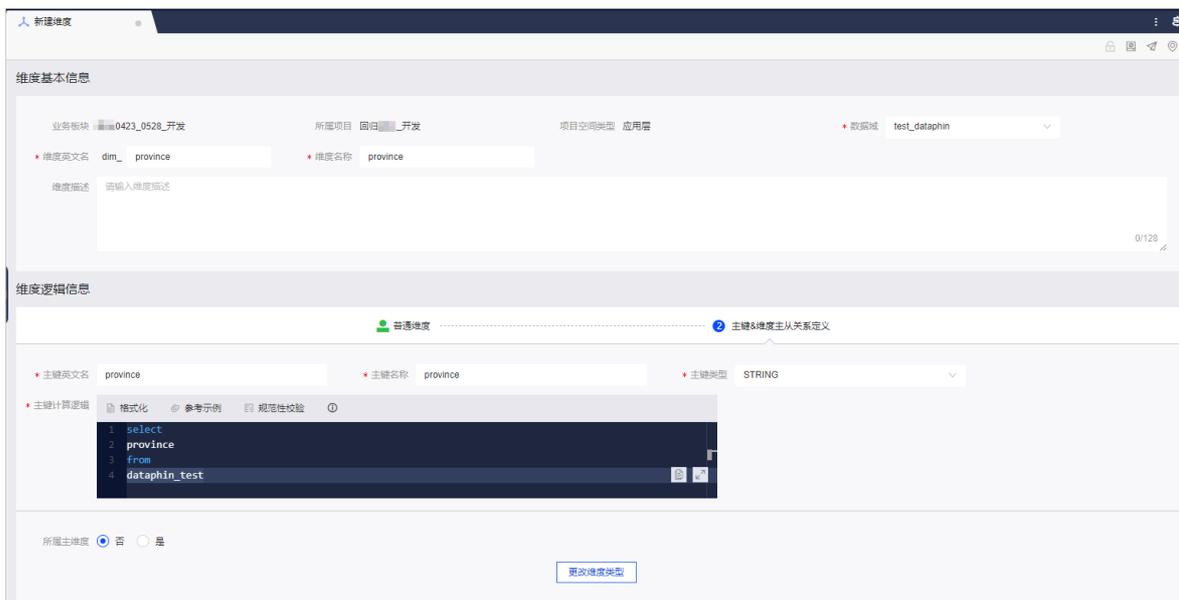
新建维度

完成维度创建后，系统自动生成维度逻辑表dim_province。

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页签，选择LD_beginner_test_dev为数据开发的项目空间。
 - 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。

 - iv. 在规范建模页签，单击左侧导航栏的维度图标。
4. 在维度页面，单击图标。
5. 在新建维度页面，配置参数。



区域	参数	描述
维度基本信息	数据域	选择test_dataphin。
	维度英文名	输入province。
	维度名称	输入province。
	维度描述	输入对维度的简单描述。
	主键英文名	输入province。
	主键名称	输入province。

区域	参数	描述
维度逻辑信息	主键类型	选择STRING。
	主键计算逻辑	<p>支持通过编写SQL来定义维度的主键计算逻辑：</p> <ul style="list-style-type: none"> i. 单击主键计算逻辑后面的参考示例，可以查看SQL编写示例，输入内容如下。 <pre>select province from datax_test</pre> <ul style="list-style-type: none"> ii. SQL编写完成后，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。

6. 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. 在提交备注对话框，输入备注信息。
 - iv. 单击确定并提交。
7. 查看并提交维度逻辑表。
 - i. 在规范建模页签，单击左侧导航栏中的维度逻辑表图标。
 - ii. 在维度逻辑表页面，单击dim_province，查看系统自动生成的维度逻辑表。
 - iii. 单击页面右上方的图标，保存维度逻辑表。
 - iv. 单击页面右上方的图标，提交维度逻辑表。
 - v. 在提交备注对话框，输入备注信息。
 - vi. 单击确定并提交。

新建业务过程和事实逻辑表

1. 在规范建模页签，单击左侧导航栏中的业务过程图标。
2. 在业务过程页面，单击图标。
3. 在新建业务过程对话框，配置如下参数。

新建业务过程 ×

业务板块 入门教程... 所属项目 入门教程_开发 项目空间类型 应用层

* 数据域 test_datap... ▾

* 英文名 order_pay

* 名称 order_pay

描述

0/128

参数	描述
数据域	选择test_dataphin。

参数	描述
英文名	输入order_pay。
名称	输入order_pay。
描述	输入对业务过程的简单描述。

4. 单击提交。
5. 在提交备注页面，输入备注信息。
6. 单击确定并提交，完成业务过程的创建。
7. 创建事实逻辑表。
 - i. 在业务过程页面，单击数据域（test_dataphin）下的order_pay。
 - ii. 在业务过程属性框，鼠标悬停至  图标后，单击创建逻辑表。
 - iii. 在新建事实逻辑表对话框，配置如下参数。

新建事实逻辑表 ×

1 基本信息
2 主键定义

业务板块 入门教程_开发 空间类型 应用层

所属项目 入门教程_开发 * 数据域 test_dataphin(test_d... ▾

* 业务过程 order_pay(order_pay) ▾

* 明细类型 事务型 ▾ 英文名 fct_order_pay_ 最多输入 64 个字符 _di

* 名称 shishibiao

描述 最多输入 128 个字符 0/128

* 来源主表 beginner_test.datax_test ①

采用变量替换项目空间，替换后预览-\${beginner_test}.datax_test
缺少数据源？前往同步或脚本加工

取消
下一步

参数	描述
数据域	选择test_dataphin。
业务过程	选择order_pay (order_pay) 。
明细类型	选择事务型。
英文名	无需配置。
名称	输入shishibiao。
描述	输入事实逻辑表的简单描述。
来源主表	选择beginner_test.datax_test。

- iv. 完成参数配置后，单击下一步。
 - v. 在主键定义页面，是否设定主键选择为否。
 - vi. 单击提交，完成事实逻辑表（fct_order_pay_di）的创建。
8. 配置事实逻辑表。

- i. 在fct_order_pay_di逻辑事实表页面，单击添加度量。
- ii. 在新建度量对话框，配置参数。

新建度量

来源表 引入字段 SQL自定义字段

来源主表: \${beginner_test}.datax_test

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)

字段及描述

order_id 订单号	+
area 区域	+
province 省份	+
city 城市	+

新建字段

字段英文名	字段名称	数据类型	操作
area	区域	STRING	🔍 ✎ 🗑
province	省份	STRING	🔍 ✎ 🗑
city	城市	STRING	🔍 ✎ 🗑
order_id	订单号	BIGINT	🔍 ✎ 🗑
product_type	类型	STRING	🔍 ✎ 🗑

来源主表: beginner_test.datax_test

原始英文名: area 原始名称: 区域

原始英文名: province 原始名称: 省份

原始英文名: city 原始名称: 城市

原始英文名: order_id 原始名称: 订单号

原始英文名: product_type 原始名称: 类型

参数	描述
来源表	选择引入字段，来源主表系统默认选择为\${beginner_test}.datax_test。
新建字段	单击area、province、city、order_id、product_type和order_amt后的+图标，添加字段。

- iii. 完成参数配置后，单击保存并校验，完成度量的添加。
- iv. 单击添加关联维度。

- v. 在新建关联维度对话框，配置关联维度信息和编辑关联逻辑参数，其他参数保持默认。

- 关联维度信息选择test_dataphin数据域下的dim_province维度逻辑表。
- 编辑关联逻辑选择度量下的province。

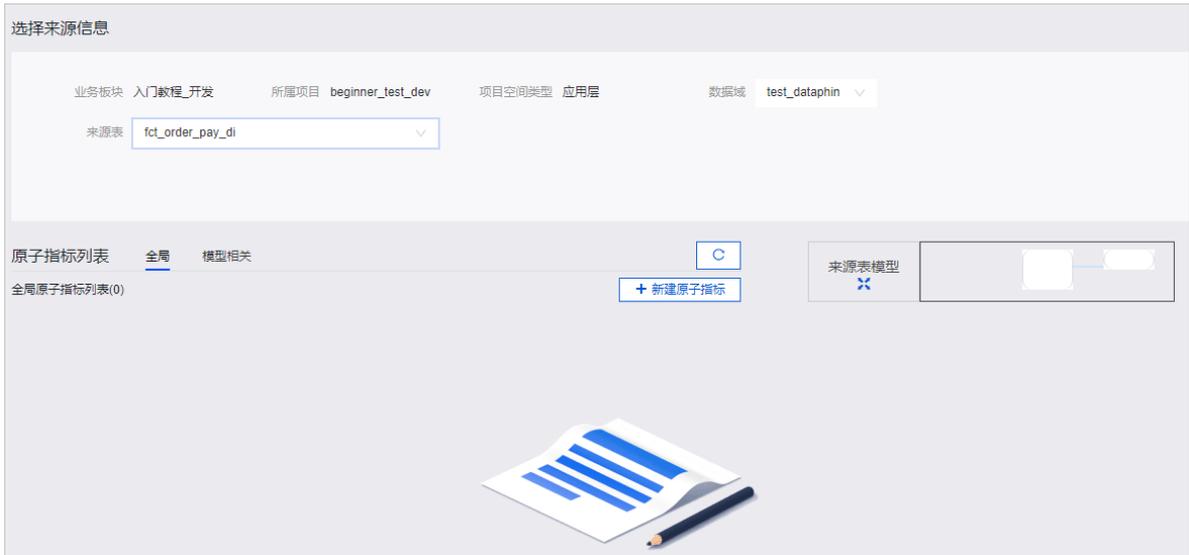
- vi. 完成参数配置后，单击确定。

9. 保存和提交事实逻辑表。

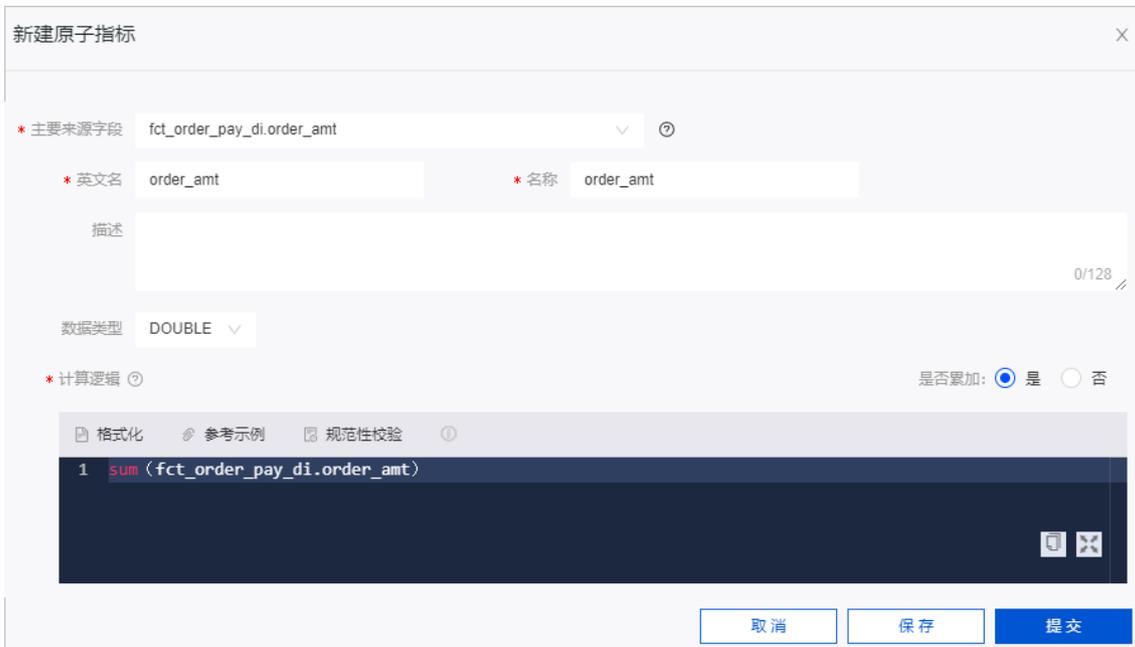
- i. 单击页面右上方的图标，保存事实逻辑表。
- ii. 单击页面右上方的图标，提交事实逻辑表。
- iii. 在提交备注对话框，输入备注信息。
- iv. 单击确定并提交。

新建原子指标

1. 在规范建模页签，单击左侧导航栏中的原子指标图标。
2. 在原子指标页面，单击图标后，选择新建原子指标。
3. 在新建原子指标页面，选择来源表为fct_order_pay_di后，单击新建原子指标。



4. 在新建原子指标对话框，配置参数。



参数	描述
主要来源字段	选择fct_order_pay_di.order_amt。
英文名	输入order_amt。
名称	输入order_amt。
数据类型	选择DOUBLE。
计算逻辑	输入sum (fct_order_pay_di.order_amt)。

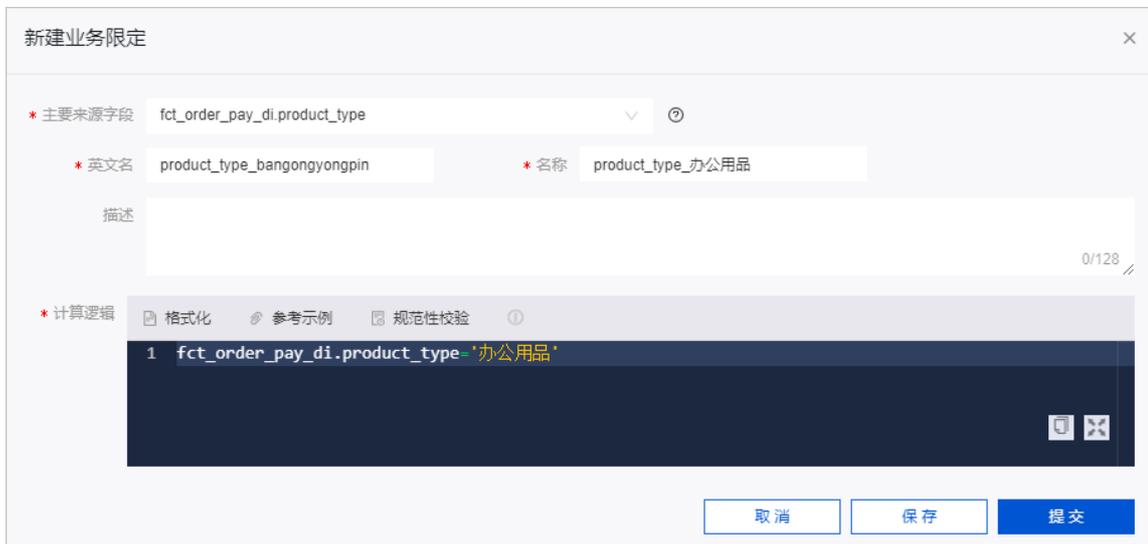
5. 单击规范性校验，可以校验您编写的SQL是否符合规范。
6. 校验通过后，单击提交。
7. 在提交备注页面，输入备注信息。
8. 单击确定并提交，完成原子指标的创建。

新建业务限定

- 1. 在规范建模页签，单击左侧导航栏中的业务限定图标。
- 2. 在业务限定页面，单击图标。
- 3. 在新建业务限定页面，来源表选择为fct_order_pay_di后，单击新建业务限定。



- 4. 新建业务限定（产品类型办公用品）。
 - i. 在新建业务限定对话框，配置参数。



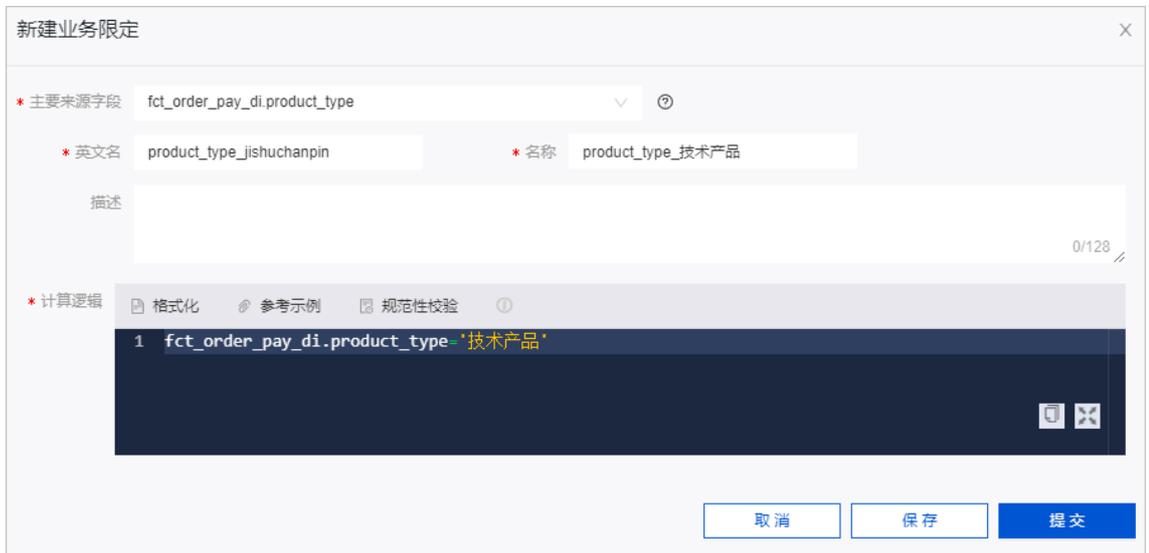
参数	描述
主要来源字段	选择fct_order_pay_di.product_type。
英文名	输入product_type_bangongyongpin。
名称	输入product_type_办公用品。
计算逻辑	输入计算逻辑： a. 输入fct_order_pay_di.product_type='办公用品'。 b. SQL编写完成后，单击规范性校验，可以校验您编写的SQL是否符合规范。

- ii. 单击提交。
 - iii. 在提交备注页面，输入备注信息。
 - iv. 单击确定并提交，完成业务限定的创建。
- 5. 新建业务限定（产品类型技术产品）。

i. 单击新建业务限定。



ii. 在新建业务限定对话框，配置参数。



参数	描述
主要来源字段	选择fct_order_pay_di.product_type。
英文名	输入product_type_jishuchanpin。
名称	输入product_type_技术产品。
计算逻辑	输入计算逻辑： a. 输入fct_order_pay_di.product_type='技术产品'。 b. SQL编写完成后，单击规范性校验，可以校验您编写的SQL是否符合规范。

新建派生指标

完成派生指标创建后，系统自动生成汇总逻辑表dws_province。

1. 在规范建模页面，单击左侧导航栏中的派生指标图标。
2. 在新建派生指标页面，单击回图标。
3. 在新建派生指标页面，选择原子指标为test_dataphin/order_amt。
4. 单击下一步。
5. 在创建派生指标页面，配置参数。

原子指标: order_amt order_amt 来源逻辑表: fct_order_pay_di

1 确定统计粒度

* 统计粒度:

[+ 新建统计粒度](#) 1/3

2 确定统计周期

* 统计周期:

[+ 新建统计周期](#) 1/3

3 确定业务限定

业务限定:

[+ 新建业务限定](#) 2/3

[返回上一步](#) [预生成派生指标](#)

参数	描述
统计粒度	选择dim_province province下的fct_order_pay_di.dim_province。
统计周期	选择最近1天。
业务限定	选择product_type_办公用品后，单击新建业务限定，选择product_type_技术产品。

- 6. 单击预生成派生指标。
- 7. 提交派生指标。
 - i. 在编辑派生指标页面，确认输入信息后，单击提交。
 - ii. 在提交备注页面，输入备注信息。
 - iii. 单击确定并提交，生成派生指标的创建。

7. 步骤六：开发环境补数据

本文为您介绍如何为开发环境中的维度逻辑表、事实逻辑表、汇总逻辑表和管道任务补数据。

前提条件

- 完成数据建模，详情请参见[步骤五：规范建模](#)。
- 完成数据集成，详情请参见[步骤三：引入数据](#)。

背景信息

您需要为数据表dim_province、fct_order_pay_di、dws_province和管道任务test补数据，验证数据模型是否符合预期。数据表的补数据操作相同，下文以dim_province为例介绍。

逻辑表任务补数据

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. 进入生产环境的逻辑表运维页面。
 - i. 在Dataphin首页，单击顶部菜单栏的研发。
 - ii. （可选）在数据开发页面，单击项目名称后的图标，单击Dev页签。选择beginner_test_dev项目。
 - iii. 在数据开发页面，单击顶部菜单运维。
 - iv. 在运维页面，单击逻辑表运维页签，系统默认进入逻辑表任务页面。
4. 在逻辑表任务页面，打开dim_province文件夹，单击维度逻辑任务节点。
5. 在右侧DAG图中，右键单击维度逻辑任务节点，选择补数据。
6. 在补数据对话框，业务日期输入2020-08-10~2020-08-10，选择下游选择否。
在补数据对话框，系统会自动生成实例名称，您也可以手动修改实例名称。
7. 单击确定。
8. 查看补数据实例的运行结果。
 - i. 单击全局运维页签。
 - ii. 在全局运维页签，单击左侧导航栏的图标。
 - iii. 在离线补数据实例页面，打开维度逻辑表补数据实例文件夹dim_province后，再单击维度逻辑表补数据实例节点。
 - iv. 在右侧的DAG图中，右键单击维度逻辑表dim_province节点，选择查看运行日志。
 - v. 在运行日志页面，查看运行日志。

管道任务补数据

1. 在全局运维页签，单击左侧导航栏中的图标。
2. 在离线周期任务页面，单击您开发的管道任务test。
3. 在右侧的DAG图中，右键单击管道任务test节点，选择补数据。
4. 在补数据对话框，将业务日期输入2020-08-10~2020-08-10，选择下游选择否。
在补数据对话框，系统会自动生成实例名称，您也可以手动修改实例名称。
5. 单击确定。
6. 查看补数据实例的运行结果。
 - i. 在全局运维页签，单击左侧导航栏的图标。
 - ii. 在离线补数据实例页面，打开管道任务补数据实例文件夹test后，再单击管道任务补数据实例节点。
 - iii. 在右侧的DAG图中，右键单击管道任务test节点，选择查看运行日志。
 - iv. 在运行日志页面，查看运行日志。

8.步骤七：验证数据

本文为您介绍如何通过即席查询来验证数据。

前提条件

完成维度逻辑表、汇总逻辑表、事实逻辑表和管道任务补数据操作，详情请参见[步骤六：开发环境补数据](#)。

背景信息

验证开发环境中开发的数据模型是否符合预期，即派生指标是否能够统计到各省份办公用品、技术产品最近1天的销售总额。

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. 进入即席查询页面。
 - i. 在Dataphin首页，单击顶部菜单栏的研发。
 - ii. （可选）在数据开发页面，单击项目名称后的图标，单击Dev页签。选择beginner_test_dev项目。
 - iii. 在数据开发页面，单击即席查询页签。
4. 在即席查询页签，单击图标。
5. 在新建文件对话框，配置参数。

新建文件 ✕

* 名称

描述

选择目录

参数	描述
名称	填写为dim_province。
描述	填写简单描述。
选择目录	保持默认。

6. 单击确定。
7. 在代码编写页面，编写代码并验证数据是否符合预期。
 - i. 编写以下命令后，单击页面上方的执行，验证数据是否同步成功。

```
select * from beginner_test_dev.datax_test where ds>0;
```

order_id	area	province	city	product_type	order_name	report_date	order_amt	ds	»
13729	华南	广东省	广州市	办公用品	IN	2013-01-01 00:00:00	872.48	20200810	
28774	华东	安徽省	淮北市	办公用品	IN	2013-01-01 00:00:00	180.36	20200810	
37537	东北	辽宁省	辽阳市	家具产品	IN	2013-01-02 00:00:00	4083.19	20200810	
37537	东北	辽宁省	辽阳市	办公用品	IN	2013-01-02 00:00:00	4902.38	20200810	
37537	华南	广东省	梅州市	家具产品	IN	2013-01-02 00:00:00	1239.06	20200810	
44069	华南	广东省	清远市	办公用品	IN	2013-01-02 00:00:00	137.63	20200810	

- ii. 编写以下命令后，单击页面上方的执行，验证维度逻辑表数据。

```
select * from LD_beginner_test_dev.dim_province where ds>0;
```

<	ds	province
1	20200811	广东省
2	20200811	安徽省
3	20200811	辽宁省
4	20200811	辽宁省
5	20200811	广东省
6	20200811	广东省

- iii. 编写以下命令后，单击页面上方的执行，验证事实逻辑表数据。

```
select * from LD_beginner_test_dev.fct_order_pay_di where ds >0;
```

order_amt	ds	province	order_id	area	product_type	city
872.48	20200811	广东省	13729	华南	办公用品	广州市
180.36	20200811	安徽省	28774	华东	办公用品	淮北市
4083.19	20200811	辽宁省	37537	东北	家具产品	辽阳市
4902.38	20200811	辽宁省	37537	东北	办公用品	辽阳市
1239.06	20200811	广东省	37537	华南	家具产品	梅州市
137.63	20200811	广东省	44069	华南	办公用品	清远市

- iv. 编写以下命令后，单击页面上方的执行，验证汇总逻辑表数据。

```
select * from LD_beginner_test_dev.dws_province where ds>0;
```

<	order_amt_1d_product_type_bangongyongpin	province	ds	order_amt_1d_product_type_jishuchanpin
1	77089.14000000003	上海市	20200810	125966.561
2	32175.46	云南省	20200810	12841.545000000002
3	424030.23999999999	内蒙古省	20200810	839224.48999999995
4	362623.840000000026	北京市	20200810	512812.57200000016
5	192066.54000000004	吉林省	20200810	491833.00499999995
6	32521.799999999996	四川省	20200810	113547.01000000002

9. 步骤八：发布任务

本文为您介绍如何将开发环境中开发的任务发布至生产环境。发布至生产环境的任务会自动参与生产环境的调度。

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. 进入发布页面。
 - 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，单击Dev页签，选择beginner_test_dev项目。
 - iii. 在数据开发页面，单击顶部菜单栏的**发布**。
4. 在发布页面，单击规范建模页签。
5. 发布维度和维度逻辑表。
 - i. 在规范建模页签，选中dim_province（维度逻辑表）和dim_province（维度）。
 - ii. 单击页面下方的发布。
 - iii. 在发布对话框，单击确定。
6. 发布派生指标、原子指标、业务过程、业务限定和事实逻辑表。
 - i. 在规范建模页签，选中product_type_bangongyongpin、product_type_jishuchanpin、order_amt、fct_order_pay_di和order_pay。
 - ii. 单击页面下方的发布。
 - iii. 在发布对话框，单击确定。
7. 发布汇总逻辑表。
 - i. 在规范建模页签，选中order_amt_1d_product_type_bangongyongpin、order_amt_1d_product_type_jishuchanpin和dws_province。
 - ii. 单击页面下方的发布。
 - iii. 在发布对话框，单击确定。
8. 查看发布结果。
 - i. 单击左侧导航栏的**发布记录列表**。
 - ii. 单击规范建模页签。
 - iii. 在规范建模页签，查看发布状态。

10. 步骤九：生产环境补数据

本文为您介绍如何为生产环境中的维度逻辑表、事实逻辑表和汇总逻辑表补数据。

前提条件

任务发布至生产环境，详情请参见[步骤五：发布任务](#)。

背景信息

发布至生产环境的数据次日才会参与调度，本教程为了让您快速熟悉智能数据构建与管理的流程，采用补数据的方式，模拟生产环境调度任务生成实例。

您需要为数据表dim_province、fct_order_pay_di、dws_province和管道任务test补数据。数据表的补数据操作相同，下文以dim_province为例介绍。

逻辑表任务补数据

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. 进入生产环境的逻辑表运维页面。
 - 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，单击Prod页签，选择beginner_test项目。
 - iii. 在数据开发页面，单击顶部菜单运维。
 - iv. 在运维页面，单击逻辑表运维页签。系统默认进入逻辑表任务页面。
4. 在逻辑表任务页面，打开维度逻辑表任务（dim_province）文件夹后，单击维度逻辑表（dim_province）节点。
5. 在右侧的DAG图中，右键单击维度逻辑任务节点，选择补数据。
6. 在补数据对话框，本教程中的业务日期输入2020-08-10~2020-08-10，选择下游选择否。
在补数据对话框，系统会自动生成实例名称，您也可以手动修改实例名称。
7. 单击确定。
8. 查看补数据实例的运行结果。
 - i. 单击全局运维页签。
 - ii. 在全局运维页签，单击左侧导航栏的图标。
 - iii. 在离线补数据实例页面，打开维度逻辑表补数据实例文件夹dim_province后，再单击维度逻辑表补数据实例节点。
 - iv. 在右侧的DAG图中，右键单击维度逻辑表dim_province节点，选择查看运行日志。
 - v. 在运行日志页面，查看运行日志。

管道任务补数据

1. 在全局运维页签，单击左侧导航栏中的图标。
2. 在离线周期任务页面，单击您开发的管道任务test。
3. 在右侧的DAG图中，右键单击管道任务test节点，选择补数据。
4. 在补数据对话框，将业务日期输入2020-08-10~2020-08-10，选择下游选择否。
在补数据对话框，系统会自动生成实例名称，您也可以手动修改实例名称。
5. 单击确定。
6. 查看补数据实例的运行结果。
 - i. 在全局运维页签，单击左侧导航栏的图标。
 - ii. 在离线补数据实例页面，打开管道任务补数据实例文件夹test后，再单击管道任务补数据实例节点。
 - iii. 在右侧的DAG图中，右键单击管道任务test节点，选择查看运行日志。
 - iv. 在运行日志页面，查看运行日志。

11.步骤十：查看运行结果

本文为您介绍查看生产环境数据的运维调度结果。

背景信息

查看生产环境中数据运行结果是否符合预期，即汇总逻辑表是否能够统计到各省份办公用品和技术产品最近1天的销售总额。

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**即席查询**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，单击Dev页签。选择beginner_test_dev项目。
 - iii. 在数据开发页面，单击**即席查询**页签。
4. 在**即席查询**页签，单击图标。
5. 在**新建文件**对话框，配置参数。

参数	描述
名称	输入prod_dws_province。
描述	输入简单描述。
选择目录	保持默认。

6. 单击**确定**。
7. 在**代码编写**页面，编写命令后，单击页面上方的**执行**，验证数据运行结果。

```
select * from LD_beginner_test.dws_province where ds>0;
```

<<	order_amt_1d_product_type_bangongyongpin	province	ds	order_amt_1d_product_type_jishuchanpin
1	77089.14000000003	上海市	20200810	125966.561
2	32175.46	云南省	20200810	12841.545000000002
3	424030.23999999999	内蒙古省	20200810	839224.48999999995
4	362623.84000000026	北京市	20200810	512812.57200000016
5	192066.54000000004	吉林省	20200810	491833.00499999995
6	32521.799999999996	四川省	20200810	113547.01000000002