ALIBABA CLOUD

Alibaba Cloud

DataWorks

快速入门

文档版本：20220324

阿里云

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。

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 `cd /d C:/window` 命令，进入Windows系统文件夹。
斜体	表示参数、变量。	`bae log list --instanceid` *Instance_ID*
[] 或者 [a|b]	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 {active|stand}`

目录

1.入门概述

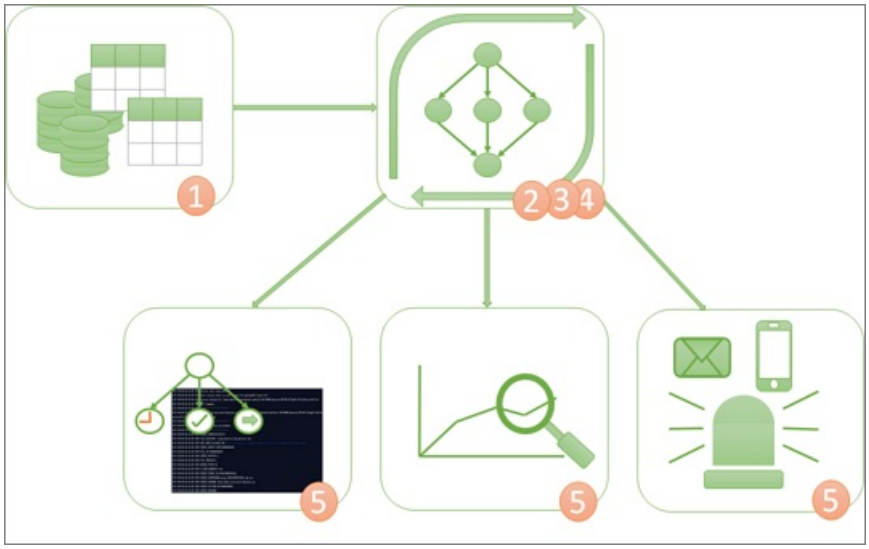
本模块将指引您快速完成一个完整的数据开发和运维操作。

> **说明**
>
> - 如果您是第一次使用DataWorks，请确认已经根据准备工作模块的操作，准备好账号和工作空间角色等内容后，登录DataWorks控制台，单击相应工作空间后的**进入数据开发**，即可进行数据开发操作。
> - 本模块的操作在标准模式的工作空间下进行。如果您使用的是简单模式的工作空间，操作步骤同标准模式。但在提交任务时，不会区分开发环境和生产环境。

通常，通过DataWorks的工作空间实现数据开发和运维包含以下操作：

1. 建表并上传数据
2. 创建业务流程
3. 创建同步任务
4. 设置周期和依赖
5. 运行及排错
6. 使用临时查询快速查询SQL（可选）

下图为数据开发和运维的基本流程。



2.建表并上传数据

本文以创建表bank_data和result_table为例，为您介绍如何通过DataWorks创建表并上传数据。

前提条件

您在**工作空间配置**页面添加**MaxCompute**计算引擎实例后，当前页面才会显示**MaxCompute**目录。详情请参见配置工作空间。

背景信息

表bank_data用于存储业务数据，表result_table用于存储数据分析后产生的结果。

创建表bank_data

1. 进入**数据开发**页面。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
2. 在**数据开发**页面，鼠标悬停至图标，单击**MaxCompute > 表**。

您也可以打开相应的业务流程，右键单击**MaxCompute**，选择**新建 > 表**。

3. 在**新建表**对话框中，输入**表名**为bank_data，单击**提交**。

> 注意
> - 表名不能超过100个字符，且必须以字母开头，不能包含中文或特殊字符。
> - 如果绑定多个实例，则需要选择MaxCompute引擎实例。

4. 在表的编辑页面，单击**DDL模式**。
5. 在**DDL模式**对话框中，输入如下建表语句，单击**生成表结构**。

```
CREATE TABLE IF NOT EXISTS bank_data
(
 age             BIGINT COMMENT '年龄',
 job             STRING COMMENT '工作类型',
 marital         STRING COMMENT '婚否',
 education       STRING COMMENT '教育程度',
 default         STRING COMMENT '是否有信用卡',
 housing         STRING COMMENT '房贷',
 loan            STRING COMMENT '贷款',
 contact         STRING COMMENT '联系途径',
 month           STRING COMMENT '月份',
 day_of_week     STRING COMMENT '星期几',
 duration        STRING COMMENT '持续时间',
 campaign        BIGINT COMMENT '本次活动联系的次数',
 pdays           DOUBLE COMMENT '与上一次联系的时间间隔',
 previous        DOUBLE COMMENT '之前与客户联系的次数',
 poutcome        STRING COMMENT '之前市场活动的结果',
 emp_var_rate    DOUBLE COMMENT '就业变化速率',
 cons_price_idx  DOUBLE COMMENT '消费者物价指数',
 cons_conf_idx   DOUBLE COMMENT '消费者信心指数',
 euribor3m       DOUBLE COMMENT '欧元存款利率',
 nr_employed     DOUBLE COMMENT '职工人数',
 y               BIGINT COMMENT '是否有定期存款'
);
```

创建表的更多SQL语法请参见创建表。

6. 在**确认操作**对话框中，单击**确认**。
7. 生成表结构后，在**基本属性**模块输入表的**中文名**，并分别单击**提交到开发环境**和**提交到生产环境**。

> 说明 本示例以标准模式的工作空间为例。如果您使用的是简单模式的工作空间，仅单击**提交到生产环境**即可。

8. 在左侧导航栏，单击**表管理**。
9. 在**表管理**页面，双击打开相应的表名，查看表信息。

创建表result_table

1. 在**数据开发**页面，鼠标悬停至图标，单击**MaxCompute > 表**。

   您也可以打开相应的业务流程，右键单击**MaxCompute**，选择**新建 > 表**。
2. 在**新建表**对话框中，输入**表名**为result_table，单击**提交**。
3. 在**DDL模式**对话框中，输入如下建表语句，单击**生成表结构**。

```
CREATE TABLE IF NOT EXISTS result_table
(
 education    STRING COMMENT '教育程度',
 num          BIGINT COMMENT '人数'
);
```

4. 在**确认操作**对话框中，单击**确认**。
5. 生成表结构后，在**基本属性**区域输入表的**中文名**，并分别单击**提交到开发环境**和**提交到生产环境**。
6. 在左侧导航栏，单击**表管理**。
7. 在**表管理**页面，双击打开相应的表名，查看表信息。

本地数据上传至bank_data

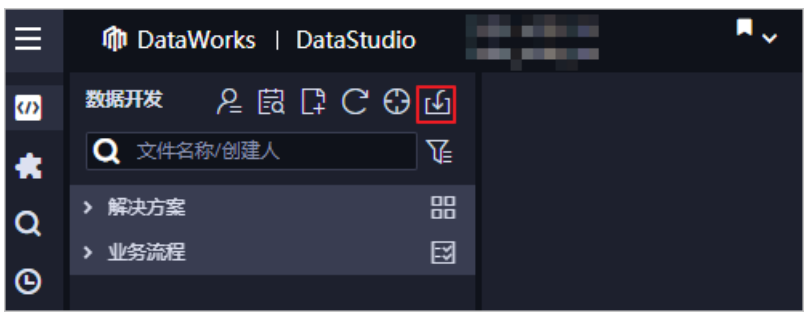
DataWorks支持以下操作：

- 上传本地的文本文件至工作空间的表中。
- 通过数据集成模块，从多个不同的数据源导入业务数据至工作空间。

> **说明** 本地文本文件上传的限制如下：
> - 文件类型：仅支持.txt、.csv和.log类型的文件。
> - 文件大小：不能超过30MB。
>
> 如果您需要上传超过30MB的文件，则可以使用如下方式：
> - 将数据文件上传至OSS，使用MaxCompute外部表映射的方式获取OSS中相应的文件数据。上传数据至OSS，详情请参见上传文件，MaxCompute外部表映射，详情请参见外部表。
> - 将数据文件上传至OSS，使用数据集成功能将OSS的数据同步至MaxCompute表。上传数据至OSS，详情请参见上传文件，同步OSS数据至MaxCompute表，详情请参见通过向导模式配置任务。
> - 使用**数据分析 > 数据上传**功能。
>
> - 操作对象：支持分区表导入和非分区表导入，但不支持分区值为中文、and（&）、星号（*）等特殊字符。

以导入本地文件banking.txt至DataWorks为例，操作如下：

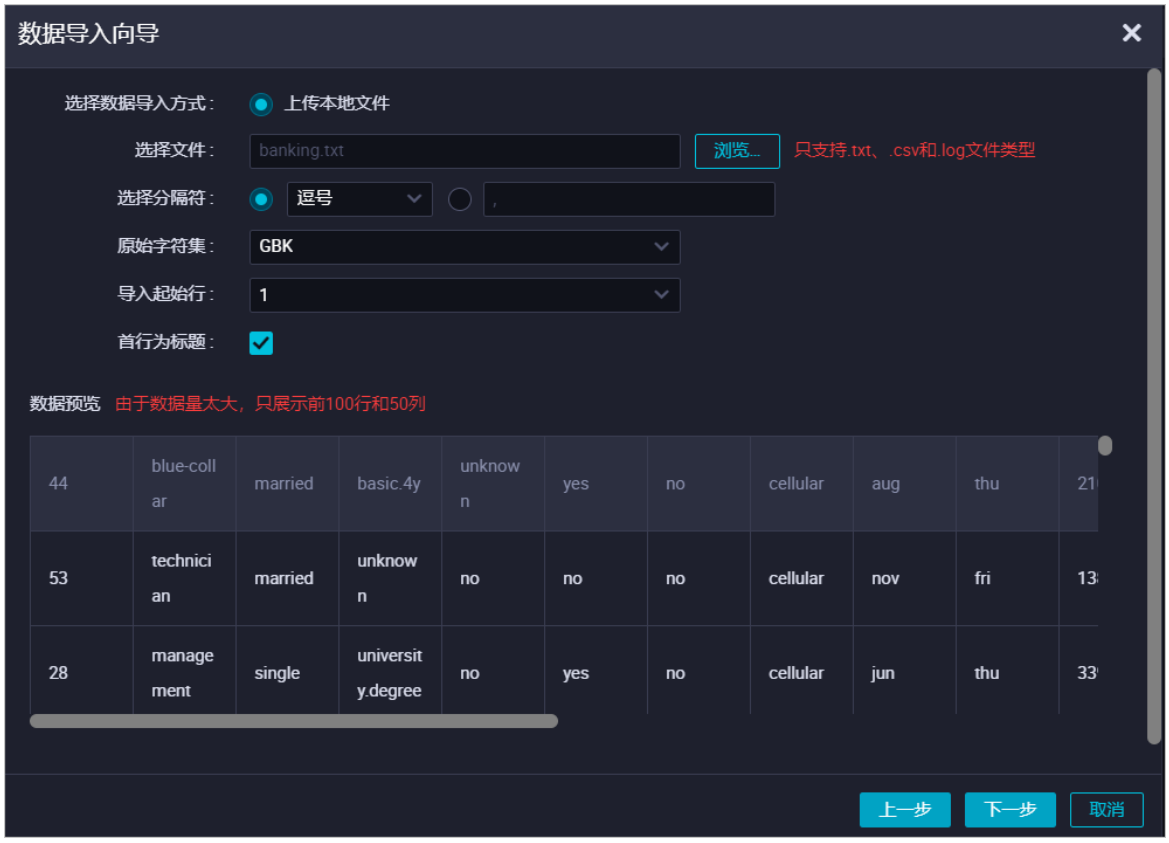
1. 在**数据开发**页面，单击图标。



2. 在**数据导入向导**对话框中，至少输入3个字母来搜索需要导入数据的bank_data表，单击**下一步**。

> **说明** 如果您创建表后无法在此处搜索到该表，您可以先在数据地图进行手工同步表操作后，再在此处尝试搜索该表，手工同步详情可参考文档：手工同步表。

3. **选择数据导入方式**为**上传本地数据**，单击**选择文件**后的**浏览...**。选择本地数据文件，配置导入信息。



参数	描述
选择数据导入方式	默认**上传本地文件**。
选择文件	单击**浏览...**，选择本地需要上传的文件。
选择分隔符	包括**逗号**、**Tab**、**分号**、**空格**、**|**、**#**和**&**等分隔符，此处选择**逗号**。
原始字符集	包括**GBK**、**UTF-8**、**CP936**和**ISO-8859**，此处选择**GBK**。
导入起始行	选择导入的起始行，此处选择**1**。
首行为标题	根据自身需求，设置首行是否为标题。本示例无需选中**首行为标题**。
数据预览	您可以在此处进行数据预览。 **说明** 如果数据量过大，仅展示前100行和前50列的数据。

4. 单击**下一步**。
5. 选择目标表字段与源字段的匹配方式，本示例选择**按位置匹配**。

6. 单击**导入数据**。

后续步骤

现在，您已经学习了如何创建表并上传数据，您可以继续下一个教程。在该教程中，您将学习如何通过创建、配置和提交业务流程，对工作空间的数据进行深入分析和计算。详情请参见创建业务流程。

3.创建业务流程

本文为您介绍如何创建业务流程，在业务流程中创建节点并配置依赖关系。完成创建后，您可以利用数据开发功能，对工作空间的数据进行深入分析和计算。

前提条件

开始本操作前，请确保您已经在工作空间中准备好业务数据表bank_data和其中的数据，以及结果表result_table。详情请参见建表并上传数据。

背景信息

DataWorks的数据开发功能支持在业务流程中，通过可视化拖拽来完成节点间的依赖设置。您可以通过操作业务流程的方式，实现对数据的处理和相互依赖。一个工作空间下支持创建多个业务流程，详情请参见管理业务流程。

创建业务流程

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 在**数据开发**页面，鼠标悬停至图标，单击**业务流程**。
5. 在**新建业务流程**对话框中，输入**业务名称**和**描述**。

   > 注意 业务名称必须是大小写字母、中文、数字、下划线（_）以及小数点（.），且不能超过128个字符。

6. 单击**新建**。

创建节点并配置依赖关系

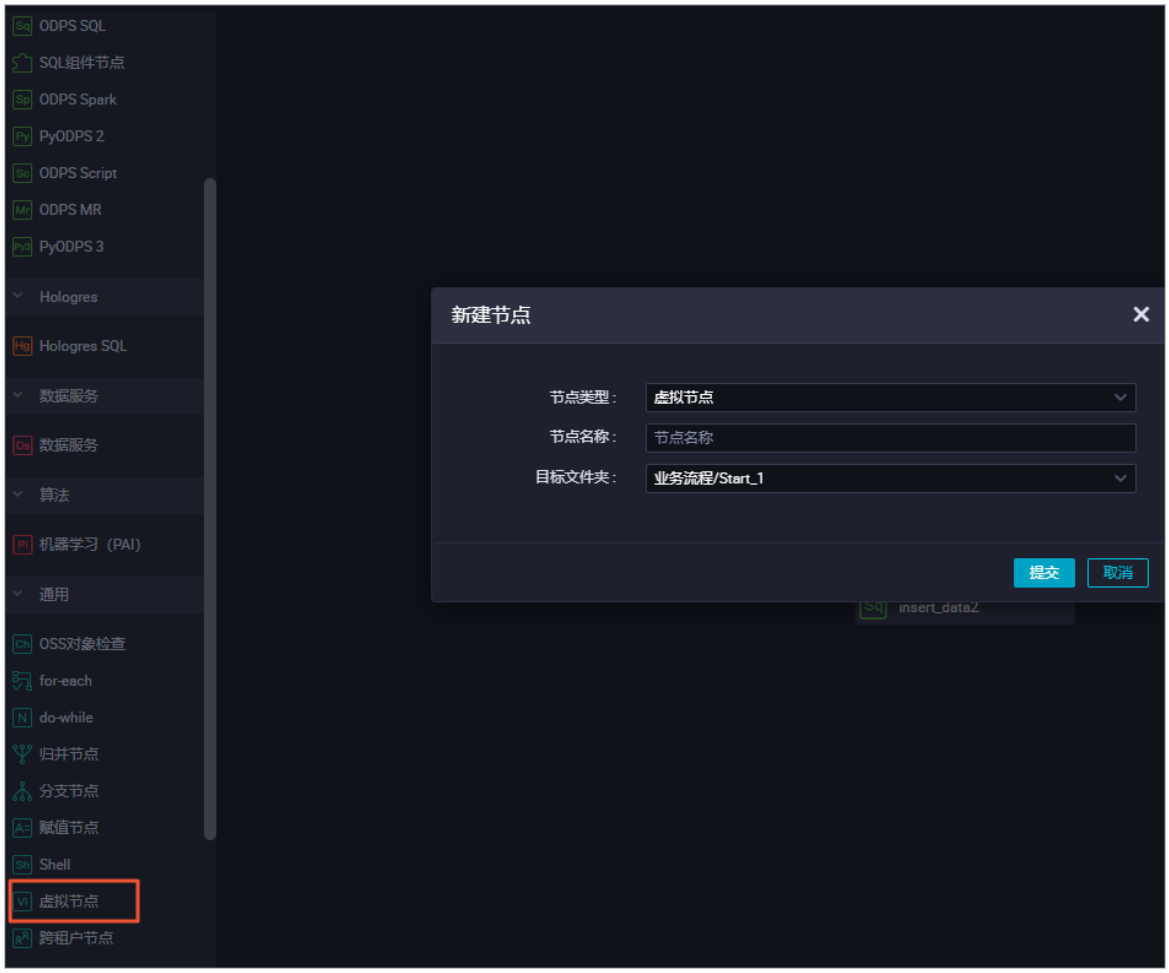
在业务流程中创建一个虚拟节点（start）和ODPS SQL节点（insert_data），并配置依赖关系为insert_data依赖于start。

> 注意
> - 虚拟节点属于控制类型节点，在业务流程运行过程中，不会对数据产生任何影响，仅用于实现对下游节点的运维控制。
> - 虚拟节点在被其它节点依赖的情况下，如果被运维人员手动设置为运行失败，则下游未运行的节点将因此无法被触发运行。在运维过程中，可以防止上游的错误数据进一步扩展。
> - 业务流程中，虚拟节点的上游节点通常会被设置为工作空间根节点。工作空间根节点的格式为`工作空间名称_root`。
> - DataWorks会为节点自动添加一个节点名的输出，结构为工作空间名称.节点名称。如果一个工作空间下有两个同名的节点，请修改其中一个节点的节点输出。

建议您在设计业务流程时，默认创建一个虚拟节点作为业务流程的根节点，来控制整个业务流程。设计业务流程的操作如下：

1. 双击业务流程名称进入开发面板，单击**通用>虚拟节点**。

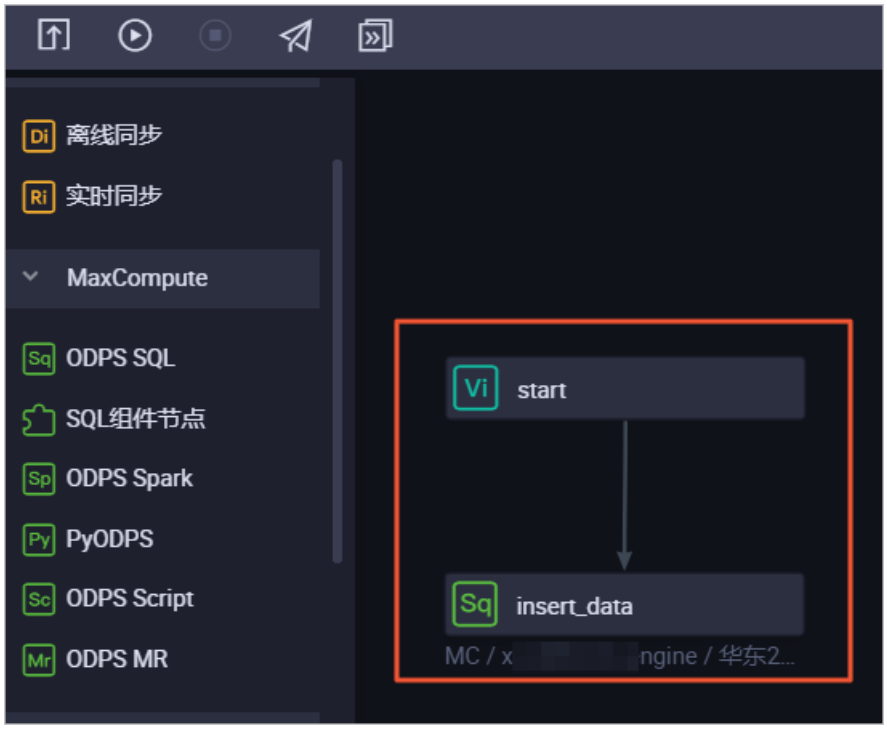
您也可以用鼠标拖拽**虚拟节点**至右侧的开发面板。



2. 在**新建节点**对话框中，输入**节点名称**为start，单击**提交**。

   > **注意** 节点名称必须是大小写字母、中文、数字、下划线（_）以及小数点（.），且不能超过128个字符。

3. 以同样的操作新建**ODPS_SQL**节点，命名为insert_data。
4. 通过拖拽连线，设置start节点为insert_data节点的上游节点。

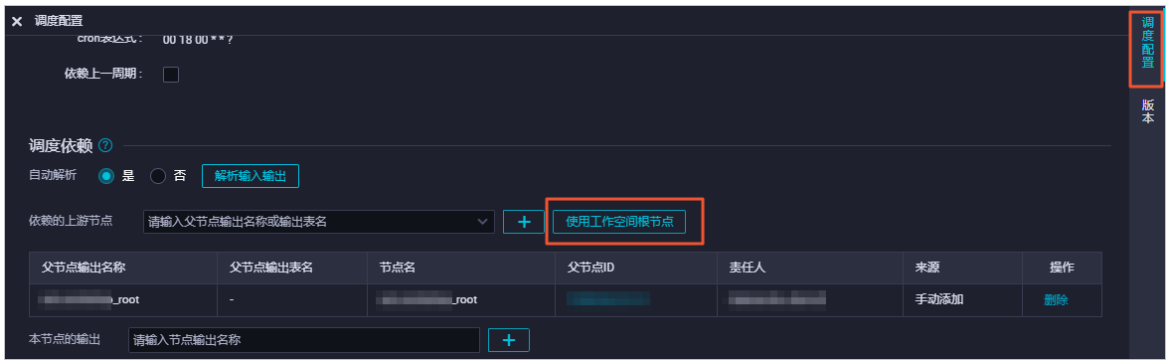


配置虚拟节点的上游依赖

在业务流程中，虚拟节点通常作为整个业务流程的控制器，是业务流程中所有节点的上游节点。

通常**使用工作空间根节点**作为虚拟节点**依赖的上游节点**：

1. 双击虚拟节点名称，进入节点的编辑页面。
2. 单击节点编辑页面右侧的**调度配置**。
3. 在**调度依赖**区域，单击**使用工作空间根节点**，设置虚拟节点的上游节点为工作空间根节点。



4. 保存并提交节点。

> **注意** 您需要设置节点的**重跑属性**和**依赖的上游节点**，才可以提交节点。

   i. 单击工具栏中的图标，保存节点。

   ii. 单击工具栏中的图标。

   iii. 在**提交新版本**对话框中，输入**变更描述**。

   iv. 单击**确认**。

编辑和运行ODPS SQL节点

本节将在ODPS_SQL节点insert_data中，通过SQL代码，查询不同学历的单身人士贷款买房的数量并保存结果，以便后续节点继续分析或展现。

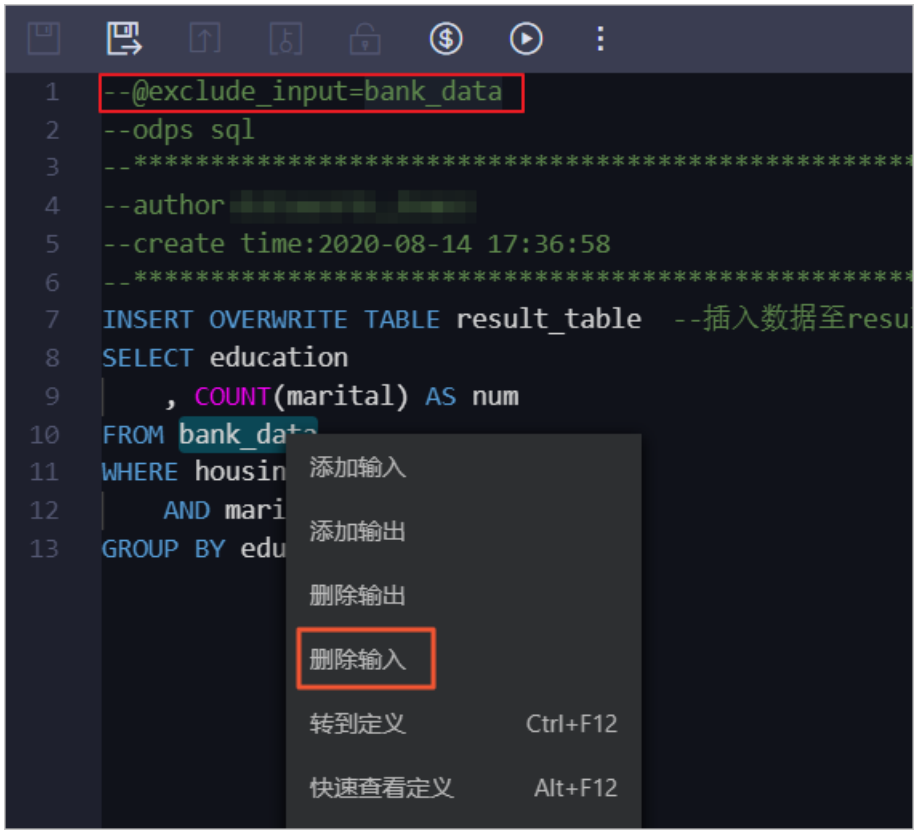
1. 打开ODPS SQL节点的编辑页面，输入下述代码。

   具体语法说明请参见SQL概述。

```
INSERT OVERWRITE TABLE result_table  --插入数据至result_table中。
SELECT education
    , COUNT(marital) AS num
FROM bank_data
WHERE housing = 'yes'
    AND marital = 'single'
GROUP BY education;
```

2. 右键单击代码中的bank_data，选中**删除输入**。

   建表并上传数据中创建的bank_data表为非周期性调度产出的表，您可以在select非周期性调度产出表的节点代码编辑页，右键相应的表名，进行删除输入的操作。您也可以在代码的最上方添加一条规则的注释，操作完成后自动解析将不会解析该依赖。



> **说明** 由于DataWorks的调度依赖主要保障的是调度节点定时更新的表数据，通过节点调度依赖保障下游取数没有问题，所以不是DataWorks平台上调度更新的表，平台无法监控。当存在非周期性调度生产数据的表，有节点select这类表数据时，您需要手动删除通过select自动生成的依赖的上游节点配置。

3. 单击工具栏中的图标，防止代码丢失。

4. 单击图标。

   运行结束后，即可在页面下方查看运行日志和结果。

提交业务流程

1. 运行并调试ODPS_SQL节点insert_data后，返回业务流程页面。
2. 单击图标。
3. 在**提交**对话框中，选择需要提交的节点，输入**备注**，并选中**忽略输入输出不一致的告警**。
4. 单击**提交**。

   业务流程提交后，即可在**业务流程**下的节点列表查看节点提交状态。如果节点名称左侧存在图标，表示该节点已提交；如果不存在图标，表示该节点未提交。

后续步骤

现在，您已经学习了如何创建和提交业务流程，您可以继续下一个教程。在该教程中，您将学习如何通过创建同步任务，将数据回流至不同类型的数据源中。详情请参见创建数据同步任务。

4.创建同步任务

本文为您介绍如何通过创建同步任务，导出MaxCompute中的数据至MySQL数据源中。

前提条件

- 您需要首先通过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。

  > **说明** 如果是通过自定义资源组调度RDS的数据同步任务，必须把自定义资源组的机器IP也加入RDS的白名单中。

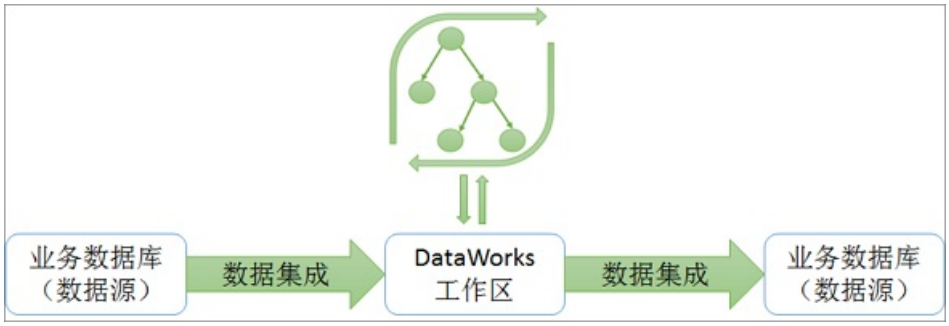
- 如果您使用的是RDS MySQL数据库，请在RDS MySQL数据库中创建表odps_result，建表语句如下所示。

```
CREATE TABLE `ODPS_RESULT` (
`education`  varchar(255) NULL ,
`num`  int(10) NULL
);
```

  建表完成后，您可以执行 `desc odps_result;` 语句，查看表详情。

背景信息

在DataWorks中，通常通过数据集成功能，定期导入系统中产生的业务数据至工作区。SQL任务进行计算后，再定期导出计算结果至您指定的数据源中，以便进一步展示或运行使用。



目前数据集成功能支持从RDS、MySQL、SQL Server、PostgreSQL、MaxCompute、OCS、DRDS、OSS、Oracle、FTP、DM、HDFS和MongoDB等数据源中，导入数据至工作空间或从工作空间导出数据。详细的数据源类型列表请参见支持的数据源与读写插件。

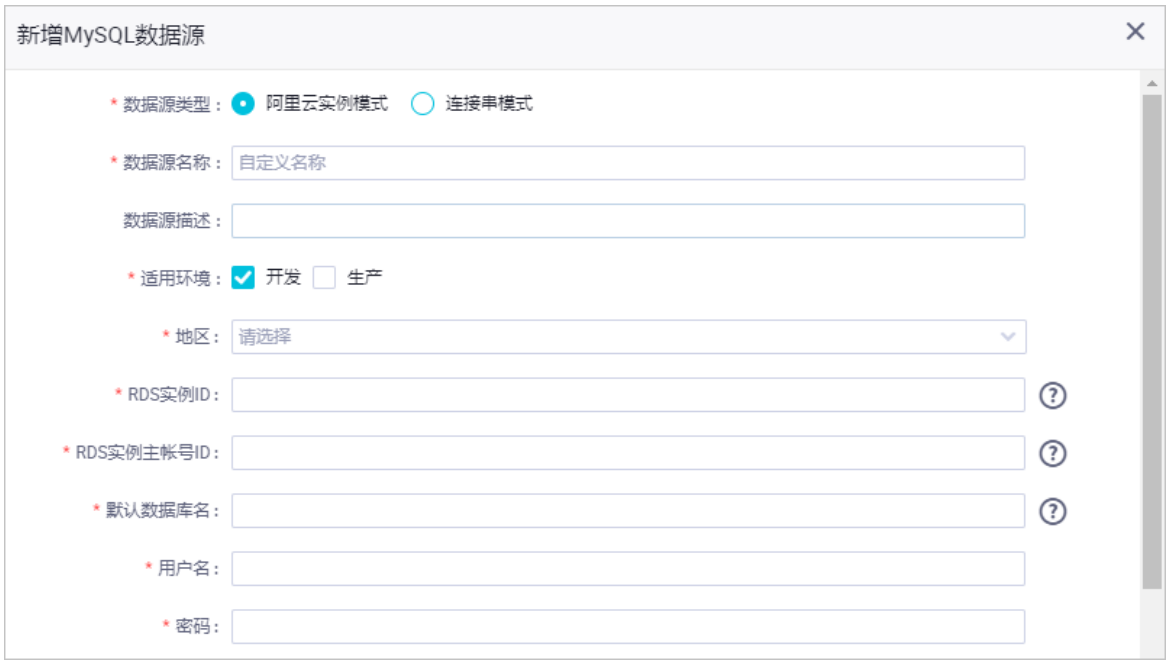
新增数据源

> **说明** 仅项目管理员角色可以新建数据源，其它角色的成员仅支持查看数据源。

1. 进入**数据源管理**页面。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据集成**。
   iv. 在左侧导航栏，单击**数据源 > 数据源列表**，进入**工作空间管理 > 数据源管理**页面。
2. 在**数据源管理**页面，单击右上角的**新增数据源**。

3. 在**新增数据源**对话框中，选择数据源类型为**MySQL**。
4. 在**新增MySQL数据源**对话框，配置各项参数。

   此处以创建**阿里云实例模式**类型为例。



参数	描述
数据源类型	当前选择的数据源类型为**阿里云实例模式**。
数据源名称	数据源名称必须以字母、数字、下划线（_）组合，且不能以数字和下划线（_）开头。
数据源描述	对数据源进行简单描述，不得超过80个字符。
适用环境	可以选择**开发**或**生产**环境。 **说明** 仅标准模式工作空间会显示该配置。
地区	选择相应的地域。
RDS实例ID	您可以进入RDS控制台，查看RDS实例ID。
RDS实例主账号ID	实例购买者登录DataWorks控制台，鼠标悬停至右上角的用户头像，查看账号ID。

参数	描述
默认数据库名	此处配置的是该数据源对应的默认数据库名称。后续配置同步任务的说明如下： ○ 配置整库同步（包含实时和离线）或同步解决方案任务时，您可以选择相应RDS实例下所有具有权限的数据库。 ○ 配置离线同步任务，当您选择使用多个数据库时，则每个数据库均需要配置一个数据源。
用户名	登录数据库的用户名称。
密码	登录数据库的密码。密码中避免使用@符号。

5. 测试资源组连通性。

   在**数据集成**和**任务调度**页签下，分别单击相应资源组后的测试连通性，连通状态为可连通时，表示连通成功。

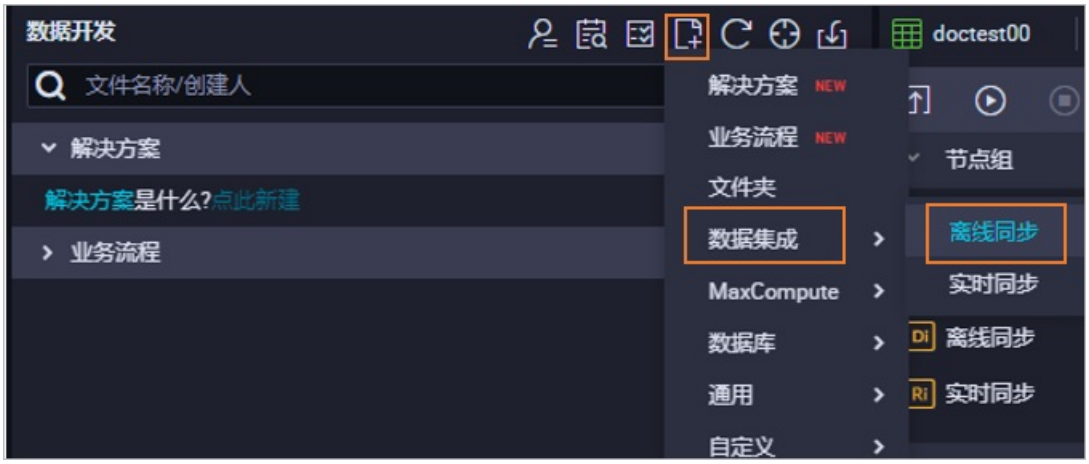
   > **说明**
   > - 数据同步时，一个任务只能使用一种资源组。
   > - 您需要测试每种资源组的连通性，以保证同步任务使用的数据集成资源组能够与数据源连通，否则将无法正常执行数据同步任务。
   > - 如果您需要同时测试多种资源组，请选中相应资源组后，单击批量测试连通性。详情请参见选择网络连通方案。

6. 测试连通性通过后，单击**完成**。

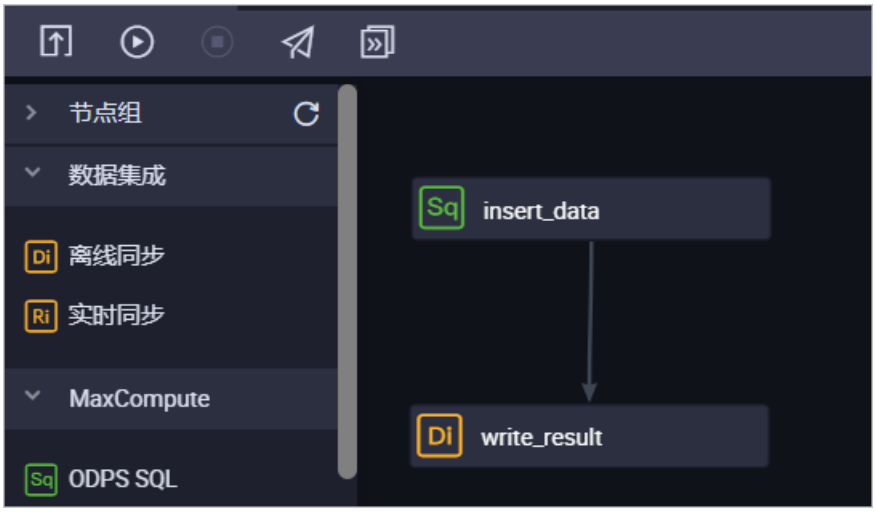
新建并配置同步节点

本节将新建一个同步节点write_result并进行配置，目的是把表result_table中的数据写入至自己的MySQL数据库中。具体操作如下：

1. 切换至**数据开发**面板，新建一个离线同步节点write_result。



2. 在**业务流程**页面，设置write_result节点的上游节点为insert_data节点。



3. 在**离线同步节点**页面，选择**数据源**（**ODPS > odps_first**）、**表**（result_table）为数据来源。
4. 选择您新建的MySQL数据源中的表（odps_result）为数据去向。

5. 选择字段的映射关系，左侧的源头表字段和右侧的目标表字段为一一对应关系。

6. 在**通道控制**区域，配置作业速率上限和脏数据检查规则。

   配置完成上述操作后，请进行通道控制。

参数	描述
任务期望最大并发数	数据同步任务内，可以从源并行读取或并行写入数据存储端的最大线程数。向导模式通过界面化配置并发数，指定任务所使用的并行度。
同步速率	设置同步速率可以保护读取端数据库，以避免抽取速度过大，给源库造成太大的压力。同步速率建议限流，结合源库的配置，请合理配置抽取速率。
错误记录数	错误记录数，表示脏数据的最大容忍条数。
分布式处理能力	数据同步时，可以将任务切片分散到多台执行节点上并发执行，提高同步速率。该模式下，配置较大任务并发数会增加数据存储访问压力，如需使用该功能，请提前评估数据存储的访问负载。该功能仅支持在独享数据集成资源组配置。

7. 预览保存。

   完成上述配置后，上下滚动鼠标即可查看任务配置。确认无误后，单击工具栏中的图标。

提交数据同步任务

同步任务保存后，返回业务流程。单击工具栏中的图标，提交同步任务至调度系统中。调度系统会根据配置的属性，从第二天开始自动定时执行。

后续步骤

现在，您已经学习了如何创建同步任务，将数据导出至不同类型的数据源中，您可以继续下一个教程。在该教程中，您将学习如何设置同步任务的调度属性和依赖关系。详情请参见设置周期和依赖。

5.配置调度和依赖属性

本文以配置任务write_result的调度周期为周调度为例，为您介绍如何设置DataWorks的调度属性和依赖属性。

前提条件

请确保您已创建任务write_result，详情请参见创建同步任务。

背景信息

DataWorks具有强大的调度能力，支持根据时间、依赖关系的节点触发机制。DataWorks可以为您保障每日千万级别的任务，根据DAG关系准确、准时运行，并且支持分钟、小时、天、周和月多种调度周期配置，详情请参见时间属性配置说明。

配置调度属性

1. 进入**数据开发**页面。
   i. 登录DataWorks控制台。
   ii. 在左侧导航栏，单击**工作空间列表**。
   iii. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
2. 在相应的业务流程下，双击打开离线同步节点write_result的编辑页面。
3. 单击编辑页面右侧导航栏的**调度配置**。

> 说明 手动业务流程中创建的节点需要手动触发，无法通过调度执行。

4. 在**时间属性**区域，配置节点的调度时间属性。

参数	描述
实例生成方式	包括**T+1次日生成**和**发布后即时生成**。详情请参见实例生成方式：发布后即时生成实例。

参数	描述
调度类型	○ **正常调度**：按照调度周期配置的定时时间启动调度，正常执行任务（即会真实跑数据）。 ○ **暂停调度**：按照调度周期配置的定时时间启动调度，但节点状态被置为暂停（即不会真实跑数据）。 ○ **空跑调度**：按照调度周期配置的定时时间启动调度，但该节点为空跑状态（即不会真实跑数据）。
调度周期	节点的运行周期（年、月、周、日、小时和分钟），此处示例设置每周一、周二的 `00:00` 点启动调度。
cron表达式	根据您配置的调度时间默认显示，不可以更改。
超时定义	当任务运行时长超过超时时间，任务将自动终止运行。 ○ 超时时间对周期实例、补数据实例、测试实例均生效。 ○ 超时时间默认值为3~7天，系统根据实际负载情况动态调整默认的任务超时时间，范围为3～7天不等。 **说明** ○ 如果您使用独享调度资源组运行任务，**超时定义**最大值可设置为72小时；如果您使用公共调度资源组运行任务，**超时定义**最大值可设置为168小时。 ○ 由于任务执行时间过长而导致任务超时终止，仍会收取该任务产生的流量、计算等费用。
重跑属性	包括**运行成功或失败后皆可重跑**、**运行成功后不可重跑**，**运行失败后可以重跑**和**运行成功或失败后皆不可重跑**。
出错自动重跑	如果**重跑属性**设置为**运行成功或失败后皆可重跑**和**运行成功后不可重跑**，**运行失败后可以重跑**时，会显示该属性，可以配置任务出错自动重跑。如果设置为**运行成功或失败后皆不可重跑**，则不会显示该属性，即任务出错不会自动重跑。
重跑次数	当勾选**出错自动重跑**后，您需要配置**重跑次数**。
重跑间隔	当勾选**出错自动重跑**后，您需要配置**重跑间隔**。默认每次重跑的间隔为30分钟，最小支持设置为1分钟，最大支持设置为30分钟。
生效日期	节点的有效日期，请根据自身需求进行设置。

更多时间属性介绍，详情请参见时间属性配置说明。

配置依赖属性

配置离线同步节点的调度属性后，继续配置离线同步节点的依赖属性。

依赖属性中可以配置节点的上游依赖，表示即使当前节点的实例已经到定时时间，也必须等待上游节点的实例运行完毕，才会触发运行。

例如，当前节点的实例将在上游insert_data节点的实例运行完毕后，才会触发执行。

> **说明** 从业务维度看，节点依赖关系设置就是下游节点等待上游节点产出表数据后，下游节点再对该表数据进行下一步操作，比如对上游产出的表数据进行进一步清洗，或者将上游清洗的结果表数据回流至其他数据库，但这些都需要等待上游节点执行成功（上游节点产出表数据）后才可以进行的操作，节点依赖关系的设置，保障的就是下游节点执行时，依赖的上游数据已经产出。关于调度依赖的逻辑说明详情，您可以参考文档：同周期调度依赖逻辑说明。

在调度系统中，每一个工作空间中默认会创建一个工作空间名称_root节点作为根节点。如果本节点没有上游节点，可以直接依赖根节点。

提交并发布节点

1. 在write_result节点的编辑页面，单击工具栏中的图标。
2. 提交节点。

   > **注意** 您需要设置节点的**重跑属性**和**依赖的上游节点**，才可以提交节点。

   i. 单击工具栏中的图标。

   ii. 在**提交新版本**对话框中，输入**备注**。

   iii. 单击**确认**。

   如果您使用的是标准模式的工作空间，提交成功后，开发环境便有了该节点，但如果您需要将节点发布至生产环境进行周期性调度，请单击右上角的**任务发布**，在此界面将节点发布至生产环境。具体操作请参见发布任务。

   节点只有提交至调度系统中，才会从第二天开始，自动根据调度属性配置的周期，在各时间点生成实例，并定时运行。

   > **说明** 如果是23:30以后提交的节点，则调度系统从第3天开始，才会自动周期生成实例并定时运行。

后续步骤

现在，您已经学习了如何设置离线同步节点的调度属性和依赖关系，您可以继续下一个教程。在该教程中，您将学习如何对提交的节点进行周期运维，并查看日志排错。详情请参见运行及排错。

6.运行及排错

本文为您介绍如何实现节点的运行、运维，并通过查看日志进行排错。

在设置周期和依赖的操作中，您配置了每周二凌晨2点执行离线同步节点。提交节点后，需要到第2天才能看到调度系统自动执行的结果。DataWorks为您提供测试运行、补数据和周期运行三种触发方式，帮助您确认实例运行的定时时间、相互依赖关系、数据结果产出是否符合预期。

- 测试运行：手动触发方式。如果您仅需要确认单个节点的定时情况和运行，建议您使用测试运行。
- 补数据运行：手动触发方式。如果您需要确认多个节点的定时情况和相互依赖关系，或者需要从某个根节点开始重新执行数据分析计算，建议您使用补数据运行。
- 周期运行：系统自动触发方式。提交成功的节点，调度系统在第二天0点起会自动触发当天不同时间点的运行实例，并在定时时间达到时检查各实例的上游实例是否运行成功。如果定时时间已到并且上游实例全部运行成功，则当前实例会自动触发运行，无需人工干预。

> **说明** 手动触发和自动调度的调度系统与周期生成实例的规则一致：
> - 无论周期选择小时、分钟、日、月或周，节点在每一个日期都会生成对应的实例。
> - 仅在指定日期的对应实例，会定时运行并生成运行日志。
> - 非指定日期的对应实例不会实际运行，而是在满足运行条件时，将状态直接转换为成功，因此不会有运行日志生成。

测试运行

1. 单击当前页面左上角的图标，选择**全部产品 > 运维中心（工作流）**，进入**运维中心**页面。
2. 在左侧导航栏，单击**周期任务运维 > 周期任务**。
3. 单击相应节点列表后的**测试**。
4. 在**冒烟测试**对话框中，输入**冒烟测试名称**，并**选择业务日期**，单击**确定**。
5. 自动跳转至**测试实例**页面，单击相应的实例，即可在右侧查看实例DAG图。

   右键单击实例，您可以查看该实例的依赖关系和详细信息，并进行终止运行、重跑等具体操作。

   > **说明**
   > - 测试运行是手动触发节点，只要到定时的时间，立即运行，自动忽略实例的上游依赖关系。
   > - 根据前文所述的实例生成规则，配置为每周二凌晨2点运行的节点write_result，测试运行时选择的业务日期是周一（业务日期=运行日期-1），实例会在2点真正运行。如果不是周一，则实例在2点转换为成功状态，且没有日志生成。

补数据运行

如果需要确认多个节点的定时情况和相互依赖关系，或者需要从某个根节点开始重新执行数据分析计算，您可以进行补数据操作。

1. 在**运维中心**页面，单击左侧导航栏中的**周期任务运维 > 周期任务**。
2. 单击相应节点列表后的**补数据 > 当前节点**。
3. 配置**补数据**对话框中的参数，单击**确定**。

参数	描述
补数据名称	输入补数据名称。
选择业务日期	选择补数据的业务日期，业务日期为 `运行日期-1` 。
当前任务	默认为当前节点，不可以更改。
是否并行	可以选择**不并行**或指定允许几组任务同时运行。

4. 自动跳转至**补数据实例**页面，单击相应的实例，即可看到实例DAG图。

   右键单击实例，可以查看该实例的依赖关系和详细信息，并进行终止运行、重跑等具体操作。

   > 说明
   > - 补数据任务的实例依赖前一天，例如补2017-09-15到2017-09-18时间段内的任务，如果15号的实例运行失败了，则16号的实例也不会运行。
   > - 根据前文所述的实例生成规则，配置为每周二凌晨2点运行的节点write_result，补数据运行时选择的业务日期是周一（业务日期=运行日期-1），实例会在2点真正运行。如果不是周一，则实例在2点转换为成功状态，且没有日志生成。

周期自动运行

周期自动运行，由系统根据所有节点的调度配置自动触发，所以页面没有操作入口。您可以通过以下两种方式查看实例信息和运行日志：

- 在**运维中心**页面，单击左侧导航栏中的**周期任务运维 > 周期实例**，选择业务日期或运行日期等参数，搜索write_result节点对应的实例后，右键查看实例信息和运行日志。
- 选择**周期实例**页面中相应的节点实例并单击，即可看到实例DAG图。

  右键单击实例，可以查看该实例的依赖关系和详细信息并进行终止运行、重跑等具体操作。

  > 说明
  > - 如果上游节点未运行，下游节点也不会运行。
  > - 如果节点的实例初始状态为未运行，当定时时间到达时，调度系统会检查该实例的全部上游实例是否运行成功。
  > - 只有上游实例全部运行成功，且定时时间到达的实例，才会被触发运行。
  > - 处于未运行状态的实例，请确认上游实例已经全部成功且已到定时时间。

7.使用临时查询快速查询SQL（可选）

如果您已经创建了DataWorks工作空间（MaxCompute项目），可以直接使用DataWorks临时查询功能，快速书写SQL语句操作MaxCompute。

进入临时查询

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击**工作空间列表**。
3. 选择工作空间所在地域后，单击相应工作空间后的**进入数据开发**。
4. 在左侧导航栏，单击**临时查询**。
5. 在**临时查询**面板，右键单击**临时查询**，选择**新建节点 > ODPS SQL**。
6. 在**新建节点**对话框中，输入**节点名称**，并选择**目标文件夹**。

   > **说明** 节点名称的长度不能超过128个字符。

7. 单击**提交**。

运行SQL

现在，您可以在新建的临时查询节点中运行MaxCompute支持的SQL语句，详情请参见SQL概述。

以运行一个DDL语句新建表为例，输入建表语句，单击▶即可。

```
create table if not exists sale_detail
(
shop_name     string,
customer_id   string,
total_price   double
)
partitioned by (sale_date string,region string);
-- 创建一张分区表sale_detail
```

您可以查看本次运行的费用预估，单击**运行**。

您可以在下方的日志窗口，查看运行情况和最终结果。如果本次运行成功，结果为OK。

您可以使用同样的方法执行查询语句。