

ALIBABA CLOUD

阿里云

数据管理

数据工厂

文档版本：20210811

 阿里云

法律声明

阿里云提醒您在使用或阅读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.数据编排	05
1.1. 数据编排概述	05
1.2. 创建数据流	06
1.3. 发布数据流	08
1.4. 数据流运维	09
2.数据类目	11

1.数据编排

1.1. 数据编排概述

数据管理DMS数据编排是一种低代码开发数据加工任务的工具，您可以组合各类任务节点，形成数据流，通过周期调度运行达到数据加工、数据同步的目的。

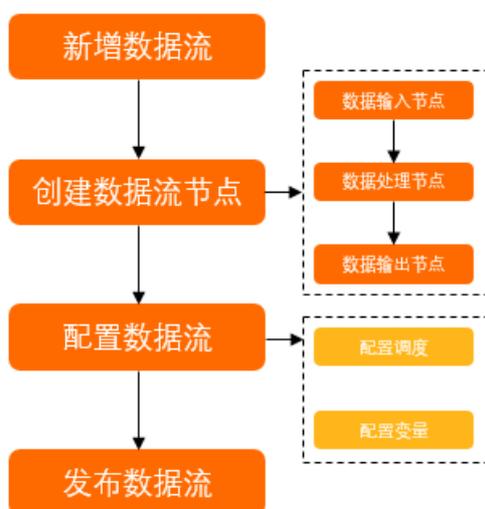
应用场景

DMS数据编排提供了批量处理数据的功能，可应用于以下场景：

② 说明 数据编排功能持续开放中。如需试用该功能或在使用过程中遇到问题，请加钉钉群沟通。钉钉群号：31826394。

- 通过低代码可视化编辑的方式快速搭建离线数仓，支持即席查询、多维分析、数据挖掘、离线计算等数仓应用场景。
- 解决企业复杂大数据批处理难题，支持企业精细化运营、数据营销、智能推荐等大数据业务场景。
- 数据编排功能底层基于Spark开发，可有效提升Hadoop平台运行速度。

创建数据流流程



1. 登录DMS数据管理服务。
2. 在顶部导航栏，选择传输与加工(DTS)。
3. 在左侧导航栏，选择数据编排。
4. 单击新增数据流。
5. 在新增数据流页面，输入数据流名称和描述，单击确认。
6. 在数据流详情页面，创建数据流节点。具体操作，请参见[创建数据流](#)。
7. 单击画布空白区域，配置数据流信息。
 - i. 单击数据流信息页签，在基础属性区域修改数据流名称、责任人、相关人员和切换开启消息通知开关状态。
 - ii. 如果需要调度数据流，在调度配置区域，打开开启调度开关，配置调度。具体配置，请参见[配置调度](#)。

- iii. 单击高级配置页签，设置变量。具体设置，请参见配置变量。
- 8. 发布数据流。具体操作，请参见发布数据流。

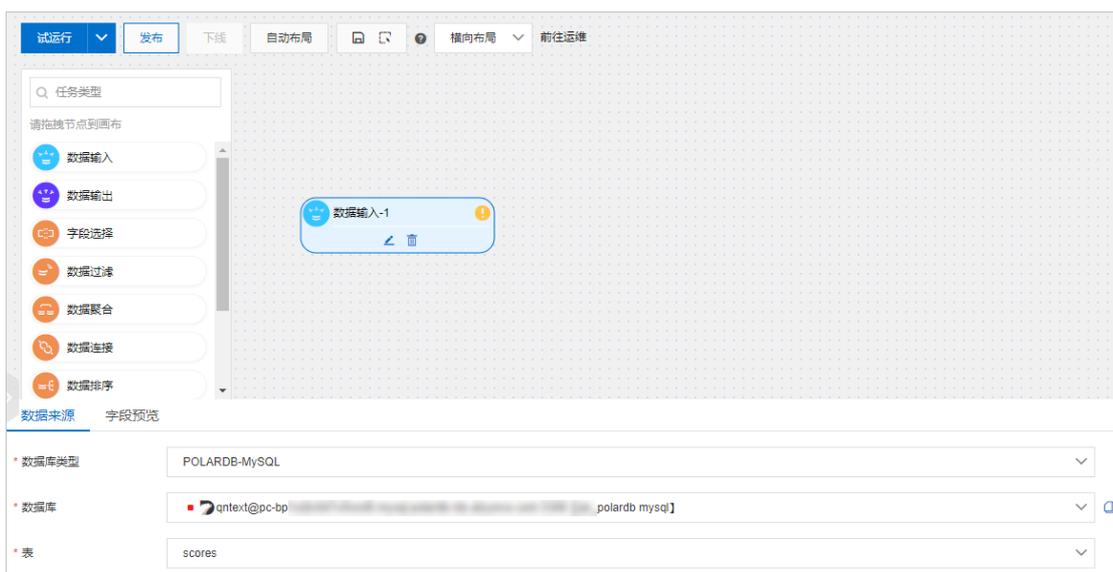
1.2. 创建数据流

在数据管理DMS数据编排中，可以组合各类任务节点，形成数据流，通过周期调度运行达到数据加工、数据同步的目的。

1. 登录DMS数据管理服务。
2. 在顶部导航栏，选择传输与加工(DTS)。
3. 在左侧导航栏，选择数据编排。
4. 单击数据流名称，进入数据流详情页面。
5. 配置数据输入节点。

? **说明** 数据流的第一个节点必须是数据输入，为数据流提供数据源。

- i. 在左侧任务类型列表中，拖拽数据输入节点到画布空白区域。
- ii. 选中数据输入节点，在页面下方的数据来源页签中配置数据源。



配置项	操作
数据库类型	选择源数据库的类型。
数据库	<ol style="list-style-type: none"> a. 输入并选择数据流读取数据的源数据库。 b. 在弹出的登录实例对话框中，输入数据库账号和数据库密码。
表	选择数据流读取数据的源表。

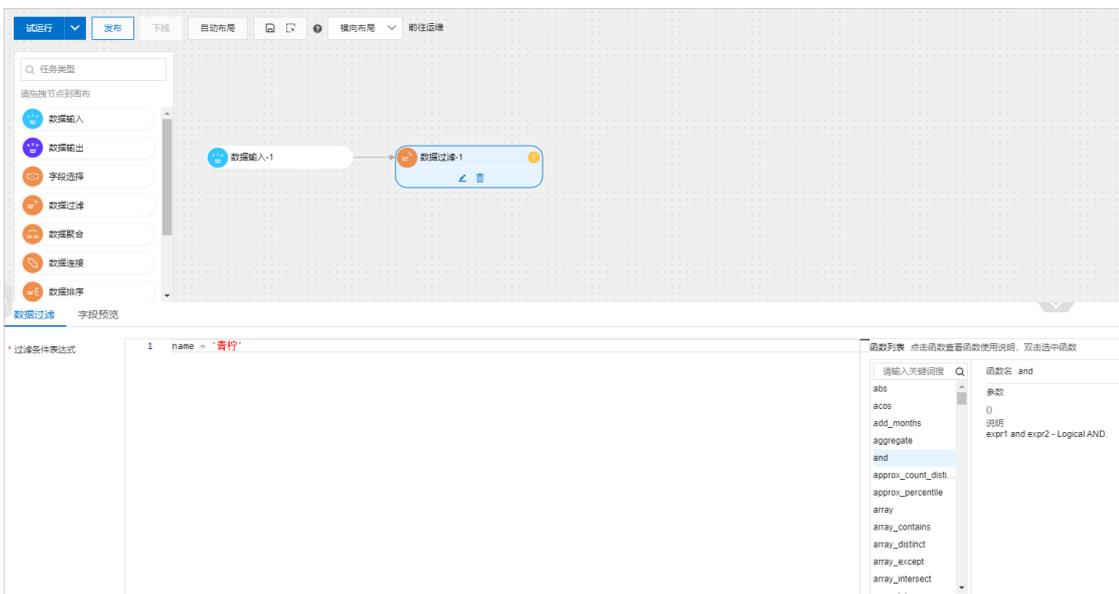
6. 以过滤源表数据为例，配置数据处理节点。

说明 除数据输入和数据输出外，其他任务类型均可作为数据处理节点。

- i. 在左侧任务类型列表中，拖拽数据过滤节点到画布的空白区域。
- ii. 将鼠标放在数据输入节点上，单击节点左侧出现的空心圆点并拉出连接线，连接至数据过滤节点上。



- iii. 选中数据过滤节点，在页面下方的数据过滤页签中配置数据过滤条件表达式。



说明 您可以直接在输入框中输入过滤条件，也可以双击右侧的函数辅助编写。

关于其他数据处理节点的配置，请参见配置转化节点。

7. 配置数据输出节点。

说明 数据流的最后一个节点必须是数据输出，提供处理后的数据要写入的目标表。

- i. 在左侧任务类型列表中，拖拽数据输出节点到画布的空白区域。

ii. 选中数据输出节点，在页面下方的数据输出页签中配置数据输出表。

配置项	描述
数据库	选择数据流输出数据的目标数据库。
表名	选择或输入数据流输出数据的目标表。 <div style="border: 1px solid #add8e6; padding: 5px; background-color: #e0f0ff;"> ? 说明 如果表不存在，在执行时会自动创建表。 </div>
是否先删除表	<ul style="list-style-type: none"> ▪ 开关关闭：不删除目标表。 ▪ 开关打开：删除目标表，在执行数据流时自动创建表。
是否先清空表	<ul style="list-style-type: none"> ▪ 开关关闭：不清空目标表里的数据。 ▪ 开关打开：清空目标表里的数据。
写入方式	两种写入方式可选： <ul style="list-style-type: none"> ▪ INSERT：将数据追加到表的末尾。 ▪ INSERT OVERWRITE：覆盖之前的数据。

iii. 将鼠标放在数据过滤节点上，单击节点左侧出现的空心圆点并拉出连接线，连接至数据输出节点上。

此时，节点右侧的⚠️自动消失，表示数据流的节点上下游依赖完整。

1.3. 发布数据流

完成数据流的配置或变更后，您需要通过发布操作将最新的数据流发布，有效避免部分变更未经确认被直接发布。

1. 登录DMS数据管理服务。
2. 在顶部导航栏，选择传输与加工(DTS)。
3. 在左侧导航栏，选择数据编排。
4. 单击数据流名称，进入数据流详情页面。
5. (可选) 试运行数据流。
 - i. 单击画布左上方的试运行按钮。
 - ii. 在提示对话框，单击确认。

iii. 单击画布的空白区域，单击页面下方的执行日志页签，查看数据流试运行是否成功。

- 如果执行日志的最后一行中出现 `status SUCCEEDED`，表明数据流试运行成功。
- 如果执行日志的最后一行中出现 `status FAILED`，表明数据流试运行失败。

② 说明

- 如果您想提前终止试运行，可以单击画布左上角的停止运行。
- 如果试运行失败，在执行日志中查看执行失败的节点和原因，修改节点的配置后重新尝试。

6. 发布数据流。

- 单击画布左上角的发布。
- 在发布对话框中输入备注信息，并单击确认，发布数据流。

② 说明 如果您不再需要调度该数据流，可以下线数据流：

- 单击画布左上角的下线。
- 在确认对话框中单击确认，下线数据流。

7. 确认发布状态。

- 单击画布右上方的前往运维。
- 在界面右侧，查看数据流是否发布的状态。
 - 已发布：该数据流处于已发布状态。
 - 未发布：该数据流处于未发布状态。

相关文档

[数据编排概述](#)

[数据流运维](#)

1.4. 数据流运维

您可以在数据流运维中心查看数据流的发布状态、发布详情、运行记录等。

1. 登录[DMS数据管理服务](#)。
2. 在顶部导航栏，选择**传输与加工(DTS)**。
3. 在左侧导航栏，选择**数据编排**。
4. 单击数据流名称，进入数据流详情页面。
5. 单击画布右上方前往运维。
6. 在数据流运维页面进行查看或操作。
 - 在数据流运维页面上方，查看数据流的**创建时间**、**修改时间**、**调度配置情况**、**是否发布状态**等基本信息。
 - 单击**运行记录**页签，选择**定时触发**或**手动触发**，查看数据流运行记录。

② 说明

- 定时触发：通过调度或指定时间的方式运行数据流。
- 手动触发：通过手动单击试运行的方式运行数据流。

- 单击状态列前的十，查看数据流运行日志。
- 在操作列中，单击执行历史，查看数据流的操作时间、操作人员和操作内容。
- 在操作列中，选择更多，对不同执行状态的数据流进行终止、重跑、暂停和置成功的操作。

② 说明

- 对于执行成功的数据流，可以进行重跑操作。
- 对于执行失败的数据流，可以将该数据流运行记录的状态置为成功。
- 对于执行中的数据流，可以终止或暂停数据流运行。

- 单击发布列表页签，查看数据流的版本ID，发布人，发布时间、版本详情和DAG图。
您还可以选中任意2个版本ID，单击版本对比，查看版本的对比。

2.数据类目

随着业务的发展，表的数量不断增多，数据管理DMS推出数据类目功能，帮助您对表进行分类，便于管理人员、开发人员及运维人员能更好地管理或使用数据表。

前提条件

- 数据库类型为关系型数据库或数据仓库，更多信息，请参见[支持的数据库类型](#)。
- 您的用户角色为管理员或DBA，关于用户角色信息，请参见[系统角色](#)。

管理数据类目

1. 登录[数据管理DMS 5.0](#)。

 **说明** 如果您需要切换到旧版数据管理DMS，单击页面右下角 ，进入数据管理DMS平台。具体操作，请参见[数据管理DMS 5.0切换至旧版](#)。

2. 在顶部菜单栏中，选择[数据资产](#) > [数据类目](#)。

 **说明** 如果您使用的是旧版数据管理DMS，在顶部菜单栏中，选择[全部功能](#) > [数据工厂](#) > [数据类目](#)。

3. 单击[新增类目](#)。
4. 在[类目名称](#)文本框中，输入类目名称，并单击[确认](#)。
左侧类目树将显示新增的类目。

 **说明** 系统默认创建未归属类目，您可以在该类目中找到所有未归属的表。

5. 在目标类目右侧，单击  图标。

- 新增子类目

您可以选择  > [新增子类目](#)，新增子类目。

最多能创建4层类目目录，若当前目录下已关联表，则不允许新增子类目。

- 编辑类目名称

您可以选择  > [编辑](#)，修改类目名称。

- 删除类目

您可以选择  > [删除](#)，删除类目目录。

若该目录存在子目录或已关联表，则不允许删除。

配置表的数据类目

每张表仅能配置一个类目，若配置了新类目，原先类目会自动解除关联。

1. 登录[数据管理DMS 5.0](#)。

 **说明** 如果您需要切换到旧版数据管理DMS，单击页面右下角 ，进入数据管理DMS平台。具体操作，请参见[数据管理DMS 5.0切换至旧版](#)。

2. 在顶部菜单栏中，选择数据资产 > 数据类目。

 **说明** 如果您使用的是旧版数据管理DMS，在顶部菜单栏中，选择全部功能 > 数据工厂 > 数据类目。

3. 单击选中目标数据类目。

 **说明** 您可以单击数据类目左侧的  按钮，展开子类目。

4. 在目标类目中，单击快速添加。

 **说明** 您也可通过其他方式进行配置：

- 在目标表的表详情页面，单击类目右侧的  按钮进行配置，更多信息，请参见[表详情](#)。
- 在未归属类目中，找到目标表，单击右侧的关联类目，进行配置。

5. 在弹窗中，搜索并勾选目标表，并单击确认。

 **说明** 支持批量选择多张表。

您已配置完成表的数据类目。

取消表的数据类目

您可以在类目目录下，找到目标表，单击右侧的移出类目，取消表的类目关联。