

ALIBABA CLOUD

阿里云

DataWorks

常见问题

文档版本：20220712

 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 <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.DataWorks交流钉钉群	06
2.数据集成任务及数据集成资源组常见问题	07
2.1. 问题汇总	07
2.2. 资源组操作及网络连通	09
2.3. 实时同步常见问题	13
2.4. 离线同步常见问题	16
3.调度配置	28
3.1. 调度参数	28
3.2. 依赖关系	34
4.运维中心	42
4.1. 问题汇总	42
4.2. 等待资源	44
4.3. 任务未运行	48
4.4. 任务实例状态异常	54
4.5. 为什么会空跑	57
4.6. 任务冻结与解冻	59
4.7. 补数据	62
4.8. 任务运行成功没有数据	66
4.9. 任务运行失败	67
4.10. 下线节点	70
4.11. 孤立节点	72
4.12. 其他问题	75
5.数据开发与运行	77
6.账号与权限	91
6.1. 用户权限控制	91
6.2. 操作权限控制	93

6.3. 操作审计	95
7.资源组	97
7.1. 独享资源组	97
7.2. 自定义资源组	103
8.数据源	114
8.1. 数据源权限管理	114
8.2. 添加数据源典型问题场景	115
9.数据治理	119
10.智能监控	121
11.数据服务	126
12.安全中心	128
13.App Studio	130

1.DataWorks交流钉钉群

DataWorks作为阿里云大数据平台操作系统，对接各种大数据计算引擎，以all in one box的方式提供专业高效、安全可靠的全域智能大数据平台，高效率完成数据全链路研发流程，建设企业数据治理体系，同时提供优质高效的交流服务，本文为您介绍DataWorks的交流钉钉群。

说明 从2009年飞天大数据平台写下第一行代码开始，DataWorks历经10年发展，形成一套成熟的产品功能体系，满足企业数据中台搭建需求。在阿里巴巴内部，每天有数万数据、算法开发工程师正在使用DataWorks。

扫描下方二维码或电脑端[点击这里](#)加入DataWorks钉钉交流群，扫码加入后，您会同时加入组织总群“飞天大数据平台”与部门群“DataWorks 交流群|主群”。

阿里云 | DataWorks

产品购买与解决方案咨询请联系群主~

DataWorks钉钉部门总群开通

加入此群，您可以收获...

- 产品最新活动
- 产品有限内测通道
- 产品最新直播
- 好友邀请无人数上限
- 产品用户交流
- 产品团队支持
- 大虎山专家支持
- 后续各类外部群逐渐减少支持

**使用钉钉
扫码加入**

DataWorks钉钉部门总群开通

加入DataWorks产品交流钉钉群，您可以获得专属产品技术支持，技术直播活动，产品新功能内测等福利，与阿里云大数据专家线上交流！

2.数据集成任务及数据集成资源组常见问题

2.1. 问题汇总

写入/读取MySQL报错：App

资源组操作及网络连通问题

- 数据同步网络连通问题
 - 在做离线同步时，我们需要先了解哪些DataWorks及其网络能力？
 - 同步ECS自建的数据库的数据时，如何保障数据库与DataWorks的网络连通？
 - 数据库和DataWorks不在同一个地域（Region）下，进行数据同步时如何保障数据库与DataWorks的网络连通？
 - 数据库和DataWorks不在同一个阿里云账号下，进行数据同步时如何保障数据库与DataWorks的网络连通？
 - 访问VPC环境下的数据库测试连通性失败，该如何处理？
- 资源组配置操作问题
 - 已经购买独享资源组，但在数据源测试连通性时或任务执行时找不到该资源组怎么办？
 - 如何通过日志查看任务执行在什么资源组类型上？
 - 如何排查自定义调度资源组等待gateway？

实时同步常见问题

- 实时同步任务配置须知
 - 实时同步任务支持哪些数据源？
 - 实时同步任务为什么不建议使用公网？
 - 实时同步字段格式问题
 - 实时同步数据时，如何处理TRUNCATE？
 - 如何提高实时同步的速度和性能？
 - 实时同步是否支持在界面运行？
- 实时同步Kafka数据

Kafka实时同步报错：Startup mode for the consumer set to timestampOffset, but no begin timestamp was specified.
- 实时同步MySQL数据
 - MySQL实时同步报错：Cannot replicate because the master purged required binary logs.
 - MySQL实时同步报错：MySQLBinlogReaderException
 - MySQL实时同步报错：show master status' has an error!
 - MySQL实时同步报错：parse.exception.PositionNotFoundException: can't find start position forxxx
 - 实时同步MySQL数据源的数据时，一开始读到数据，一段时间后无法读到数据，怎么处理？
- 实时同步Hologres数据

实时同步Hologres报错：permission denied for database xxx

离线同步常见问题

- 离线同步任务运维常见问题
 - 为什么数据源测试连通性成功，但是离线同步任务执行失败？
 - 如何切换数据集成任务执行资源组？
 - 脏数据如何排查和定位？
- 非具体插件报错原因与解决方案
 - 如何处理编码格式设置/乱码问题导致的脏数据报错？
 - 数据同步时报错：[TASK_MAX_SLOT_EXCEED]:Unable to find a gateway that meets resource requirements. 20 slots are requested, but the maximum is 16 slots.
 - 任务存在SSRF攻击Task have SSRF attacks如何处理？
 - 离线同步报错：OutOfMemoryError: Java heap space
 - 离线同步任务执行偶尔成功偶尔失败
 - 离线同步常见问题
 - 离线同步报错Duplicate entry 'xxx' for key 'uk_uk_op'如何处理？
 - 离线同步报错plugin xx does not specify column如何处理？
- 具体插件报错原因及解决方案
 - 添加MongDB数据源时，使用root用户时报错
 - MongDB使用的authDB库为admin库，如何同步其他业务库数据？
 - 读取MongDB时，如何在query参数中使用timestamp实现增量同步？
 - 读取OSS数据报错：AccessDenied The bucket you access does not belong to you.
 - 读取OSS文件是否有文件数限制？
 - 写入Redis使用hash模式存储数据时，报错如下：Code:[RedisWriter-04], Description:[Dirty data]. - source column number is in valid!
 - 写入/读取MySQL报错：Application was streaming results when the connection failed. Consider raising value of 'net_write_timeout/net_read_timeout、' on the server.
 - 读取MySQL数据库报错The last packet successfully received from the server was 902,138 milliseconds ago
 - 读取PostgreSQL数据报错，该如何处理？
 - 报错Communications link failure，该如何处理？
 - 读取MaxCompute（ODPS）表数据报错The download session is expired.
 - 写入MaxCompute(ODPS) 报错block失败：Error writing request body to server
 - 写入DataHub时，一次性写入数据超限导致写入失败如何处理？
- 离线同步场景及解决方案
 - 离线同步常见问题
 - 离线同步任务如何自定义表名？
 - 配置离线同步节点时，无法查看全部的表，该如何处理？
 - 读取MaxCompute（ODPS）表数据时，添加一行注意事项
 - 读取MaxCompute（ODPS）表数据时，如何同步分区字段？
 - 读取MaxComputet（ODPS）表数据时，如何同步多个分区数据？
 - 如何处理表列名有关键字导致同步任务失败的情况？

- 读取Loghub同步某字段有数据但是同步过来为空
- 读取Loghub同步少数据
- 读取Loghub字段映射时读到的字段不符合预期
- 读取kafka配置了endTime来指定所要同步的数据的截止范围，但是在目的数据源中发现了超过这个时间的数据。
- 写入OSS出现随机字符串如何去除？
- MySQL分库分表如何将分表同步到一张MaxCompute中

2.2. 资源组操作及网络连通

- 数据同步网络连通问题
 - 在做离线同步时，我们需要先了解哪些DataWorks及其网络能力？
 - 同步ECS自建的数据库的数据时，如何保障数据库与DataWorks的网络连通？
 - 数据库和DataWorks不在同一个地域（Region）下，进行数据同步时如何保障数据库与DataWorks的网络连通？
 - 数据库和DataWorks不在同一个阿里云账号下，进行数据同步时如何保障数据库与DataWorks的网络连通？
 - 访问VPC环境下的数据库测试连通性失败，该如何处理？
- 资源组配置操作问题
 - 已经购买独享资源组，但在数据源测试连通性时或任务执行时找不到该资源组怎么办？
 - 如何通过日志查看任务执行在什么资源组类型上？
 - 如何排查自定义调度资源组等待gateway？

在做离线同步时，我们需要先了解哪些DataWorks及其网络能力？

做数据集成同步前，您需要非常了解以下信息：

- 需要同步的数据库所在的VPC，Vswitch，地域（Region）等信息和DataWorks所在地域（Region）情况。
- 需要同步的数据库和DataWorks是否存在跨账号，跨region等特殊情况。

如果在任务同步时遇到问题，您可以先参考[支持的数据源与读写插件](#)，进行自助排查。

如果您在数据源测试连通性时遇到问题，您可以先根据[配置资源组与网络连通](#)文档，选择对应的解决方案进行排查。

在使用独享数据集成资源组时，需按照以下流程操作：购买资源组、绑定数据库网络连通的专有网络、评估是否添加路由、数据库白名单配置、独享资源组绑定所属工作空间。详情请参见[新增和使用独享数据集成资源组](#)。

同步ECS自建的数据库的数据时，如何保障数据库与DataWorks的网络连通？

如果您需要使用独享数据集成资源组通过内网同步ECS自建的数据库，您需要为独享数据集成资源组做相关的网络配置，详情可参见[新增和使用独享数据集成资源组](#)，连通方案核心点为：

- 独享数据集成资源组绑定ECS所在VPC，会自动添加一个路由指向VPC网段（该路由不建议人为删除哦，可能会导致访问其他数据库失败，导致任务报错。）
- 在数据库白名单列表中添加独享数据集成资源组绑定的交换机网段。详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。

数据库和DataWorks不在同一个地域（Region）下，进行数据同步时如何保障数据库与DataWorks的网络连通？

操作前，您可以先根据[配置资源组与网络连通](#)文档，选择对应的网络连通解决方案，连通方案核心点为：

- 如果您使用公网同步跨地域（Region）的数据库，您需要将独享资源组本身的EIP添加到数据库白名单列表中，详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。

 说明 公网会产生公网流量费，详情可参见文档[公网流量计费说明](#)。

- 如果您需要通过内网同步跨账号的数据库，您需要做以下操作：
 - i. 两个阿里云账号间的网络需要先使用网络打通工具相关云产品打通，如VPN、高速通道等。
 - ii. 独享数据集成资源组绑定已经和另一个地域（Region）网络打通的VPC。
 - iii. 绑定VPC后，再自定义路由选择本地IDC，添加目标数据库IP。
 - iv. 在数据库添加独享资源组绑定的交换机网段（注意是绑定的交换机网段）详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。

数据库和DataWorks不在同一个阿里云账号下，进行数据同步时如何保障数据库与DataWorks的网络连通？

操作前，您可以先根据[配置资源组与网络连通](#)文档，选择对应的解决方案进行排查。

- 如果您使用公网同步跨账号的数据库，您需要将独享资源组本身的EIP添加到数据库白名单列表中，详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。

 说明 公网会产生公网流量费，详情可参见文档[公网流量计费说明](#)。

- 如果您需要通过内网同步跨账号的数据库，您需要做以下操作：
 - i. 两个阿里云账号间的网络需要先使用网络打通工具相关云产品打通，如VPN、高速通道等。
 - ii. 独享数据集成资源组绑定已经和另一个阿里云账号下网络打通的VPC。
 - iii. 绑定VPC后，再自定义路由选择“本地IDC”，添加目标数据库IP。
 - iv. 在数据库添加独享资源组绑定的交换机网段（注意是绑定的交换机网段）详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。

访问VPC环境下的数据库测试连通性失败，该如何处理？

- 如果您使用VPC连接地址添加的数据源：
 - i. 请确认您已经为独享数据集成资源组绑定好数据库所在的VPC。
 - ii. 已经将独享数据集成资源组绑定的交换机网段添加至数据库白名单。详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。
- 如果您使用公网地址添加的数据源，使用独享数据集成资源组测试连通性失败时，请确认您是否已经将独享数据集成资源组本身的EIP添加至数据库白名单。详情请参见：[使用独享数据集成资源组执行任务需要在数据库添加的IP白名单](#)。

 说明 公网会产生公网流量非，详情可参考文档：[公网流量计费说明](#)。

已经购买独享资源组，但在数据源测试连通性时或任务执行时找不到该资源组怎么办？

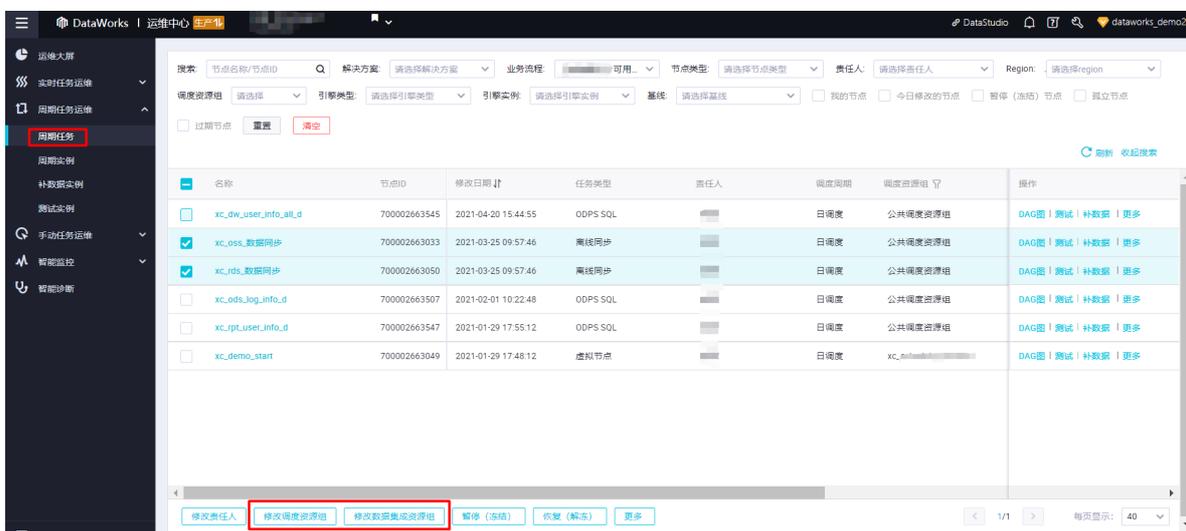
请确认独享资源组已经绑定DataWorks工作空间。详情请参见：[新增和使用独享数据集成资源组](#)。

如何通过日志查看任务执行在什么资源组类型上？

- 任务运行在默认的资源组上，日志中会出现如下信息：`running in Pipeline[basecommon_group_XXXXX]`。
- 任务运行在数据集成自定义资源组上，日志中会出现如下信息：`running in Pipeline[basecommon_XXXXXX]`。
- 任务运行在独享数据集成资源上，日志中会出现如下信息：`running in Pipeline[basecommon_S_resource_group_XXX]`。

如何切换任务在调度系统中执行所用的资源组

- 在生产运维中心切换任务执行使用的调度资源组和数据集成任务执行资源组：



- 在数据开发界面走发布流程修改生产环境任务执行使用的资源组。

说明 以下方式修改资源组请记得发布，标准模式下提交操作生效范围是开发环境，发布后操作才会在生产环境周期任务自动调度时才会生效，发布后您可以在生产运维中心周期任务检查资源组修改是否生效。

1. 修改任务调度使用的资源组：



2. 修改数据集成任务执行资源组：



如何排查自定义调度资源组等待gateway?

登录DataWorks控制台，在左侧导航栏，单击资源组列表，进入自定义资源组页签。单击任务使用的调度资源后的服务器管理，查看服务器的状态是否停止，或是否被其它任务占用。

如果以上排查无法解决问题，请执行下述命令重启服务。

```
su - admin /home/admin/alisatasknode/target/alisatasknode/bin/serverctl restart
```

2.3. 实时同步常见问题

● 实时同步任务配置须知

- 实时同步任务支持哪些数据源？
- 为什么实时同步任务延迟较大？
- 实时同步任务为什么不建议使用公网？
- 实时同步字段格式问题
- 实时同步数据时，如何处理TRUNCATE？
- 如何提高实时同步的速度和性能？
- 实时同步是否支持在界面运行？

● 实时同步Kafka数据

Kafka实时同步报错：Startup mode for the consumer set to timestampOffset, but no begin timestamp was specified.

● 实时同步MySQL数据

- MySQL实时同步报错：Cannot replicate because the master purged required binary logs.
- MySQL实时同步报错：MySQLBinlogReaderException
- MySQL实时同步报错：show master status' has an error!
- MySQL实时同步报错：parse.exception.PositionNotFoundException: can't find start position forxxx
- 实时同步MySQL数据源的数据时，一开始读到数据，一段时间后无法读到数据，怎么处理？

● 实时同步Hologres数据

实时同步Hologres报错：permission denied for database xxx

实时同步任务支持哪些数据源？

实时同步支持的数据源请参考文档：[实时同步支持的数据源](#)。

实时同步任务为什么不建议使用公网？

实时同步任务使用公网时，会存在以下风险：

- 网络可能不稳定，丢包等时常发生，影响同步性能。
- 安全性不高。

实时同步字段格式问题

数据集成实时同步在同步MySQL、Oracle、Loghub和PolarDB类型的数据至DataHub或Kafka时，会在同步的目标端添加5个附加列，以进行元数据管理、排序去重等操作。详情请参见[实时同步字段格式](#)。

为什么实时同步任务延迟较大？

实时同步任务的延迟比较大时，可能原因为：

- 一般源端变更数据量少或过多会导致延迟。
- 网络问题会导致延迟，实时同步任务不建议使用公网。
- 起始位点比较早，追平数据要一定时间。

Kafka实时同步报错： `Startup mode for the consumer set to timestampOffset, but no begin timestamp was specified.`

请重置启动位点。

MySQL实时同步报错： `Cannot replicate because the master purged required binary logs.`

MySQL实时同步报错：`Cannot replicate because the master purged required binary logs. Replicate the missing transactions from elsewhere, or provision a new slave from backup.` 时，可能是因为在MySQL未找到消费位点的binlog记录，请检查您MySQL的binlog保留时间，同步任务启动时请将该位点将配置在这个时间范围内。

说明 如果您订阅不到binlog，可以尝试重置位点到当前时间。

MySQL实时同步报错： `MySQLBinlogReaderException`

MySQL实时同步报错：`MySQLBinlogReaderException: The database you are currently syncing is the standby database, but the current value of log_slave_updates is OFF, you need to enable the binlog log update of the standby database first.` 时，可能是因为备库没有开启binlog，如果您要同步备库，需要做备库级联开启binlog，可以找DBA寻求帮助。

开启binlog的操作详情可参见[步骤3：开启MySQL的Binlog](#)。

MySQL实时同步报错： `show master status' has an error!`

MySQL实时同步报错：`show master status' has an error!`，报错详情为 `Caused by: java.io.IOException: message=Access denied; you need (at least one of) the SUPER, REPLICATION CLIENT privilege(s) for this operation, with command: show master status` 时，可能是因为数据源没有开启对应数据库的权限。

数据源配置账号需要拥有数据库的SELECT、REPLICATION SLAVE、REPLICATION CLIENT权限。给数据源添加数据库对应权限的操作详情可参见[步骤2: 创建账号并配置权限](#)。

MySQL实时同步报错： `parse.exception.PositionNotFoundException: can't find start position forxxx`

同步未找到位点，请重置位点。

实时同步MySQL数据源的数据时，一开始读到数据，一段时间后无法读到数据，怎么处理？

1. 可在数据库执行以下命令，查看当前这个数据库实例正在写入的binlog文件。

```
show master status
```

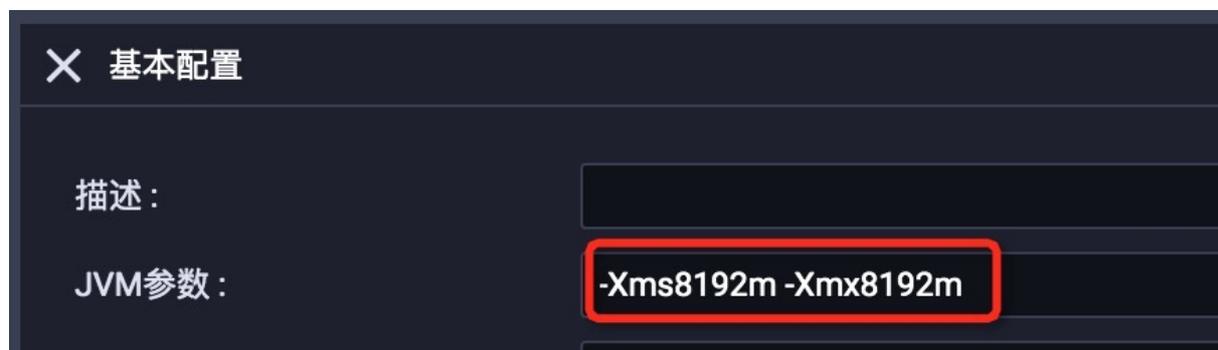
2. 对比日志中读到的binlog文件，在日志中搜 `journalName=MySQL-bin.000001,position=50`，确认是否有数据写入数据库。
3. 如果有数据在写入，但是binlog却没有往前推进，请联系DBA处理。

实时同步数据时，如何处理TRUNCATE？

实时同步支持TRUNCATE的，在增量合并的时候会生效。如果选择忽略TRUNCATE，可能会导致进行实时数据同步时出现多的数据。

如何提高实时同步的速度和性能？

如果同步写入速度较慢，可以适当增加写入端并发数，调整JVM参数，JVM参数与同步库数量无关，和变更频率有关。在当前资源组机器允许情况下，内存给的越大，Full GC频率会越小，实时同步性能相应也会越高。



实时同步Hologres报错： `permission denied for database xxx`

实时同步Hologres数据时，需要在Hologres给当前操作用户Hologres实例的admin权限（需要有创建schema的权限），操作详情可参见[Hologres权限模型概述](#)。

实时同步是否支持在界面运行？

实时同步不支持在DataWorks的界面上直接运行，您需要在配置好实时同步任务后，提交并发布实时同步节点后，进入生产环境运行该节点。详情请参见[创建、编辑、提交和运维实时同步节点](#)。

2.4. 离线同步常见问题

本文为您介绍离线同步的相关问题。

- 离线同步任务运维常见问题
 - [为什么数据源测试连通性成功，但是离线同步任务执行失败？](#)
 - [如何切换数据集成任务执行资源组？](#)
 - [脏数据如何排查和定位？](#)
 - [如何排查离线同步任务运行时间长的问题？](#)
- 非插件报错原因与解决方案
 - [如何处理编码格式设置/乱码问题导致的脏数据报错？](#)
 - [数据同步时报错：\[TASK_MAX_SLOT_EXCEED\]:Unable to find a gateway that meets resource requirements. 20 slots are requested, but the maximum is 16 slots.](#)
 - [任务存在SSRF攻击Task have SSRF attacks如何处理？](#)
 - [离线同步报错：OutOfMemoryError: Java heap space](#)
 - [离线同步任务执行偶尔成功偶尔失败](#)
 - [离线同步报错Duplicate entry 'xxx' for key 'uk_uk_op'如何处理？](#)
 - [离线同步报错plugin xx does not specify column如何处理？](#)
 - [如何处理表列名有关键字导致同步任务失败的情况？](#)
- 具体插件报错原因及解决方案
 - [添加MongoDB数据源时，使用root用户时报错](#)
 - [MongoDB使用的authDB库为admin库，如何同步其他业务库数据？](#)
 - [读取MongoDB时，如何在query参数中使用timestamp实现增量同步？](#)
 - [读取OSS数据报错：AccessDenied The bucket you access does not belong to you.](#)
 - [读取OSS文件是否有文件数限制？](#)
 - [写入Redis使用hash模式存储数据时，报错如下：Code:\[RedisWriter-04\], Description:\[Dirty data\]. - source column number is in valid!](#)
 - [写入/读取MySQL报错：Application was streaming results when the connection failed. Consider raising value of 'net_write_timeout/net_read_timeout、' on the server.](#)
 - [读取MySQL数据库报错The last packet successfully received from the server was 902,138 milliseconds ago](#)
 - [读取PostgreSQL数据报错，该如何处理？](#)
 - [报错Communications link failure，该如何处理？](#)
 - [读取MaxCompute（ODPS）表数据报错The download session is expired.](#)
 - [写入MaxCompute\(ODPS\) 报错block失败：Error writing request body to server](#)
 - [写入DataHub时，一次性写入数据超限导致写入失败如何处理？](#)
 - [RestAPI Writer报错：通过path:\[\] 找到的JSON串格式不是数组类型](#)
- 离线同步场景及解决方案

- 离线同步任务如何自定义表名？
- 配置离线同步节点时，无法查看全部的表，该如何处理？
- 读取MaxCompute（ODPS）表数据时，添加一行注意事项
- 读取MaxCompute（ODPS）表数据时，如何同步分区字段？
- 读取MaxCompute（ODPS）表数据时，如何同步多个分区数据？
- 如何处理表列名有关键字导致同步任务失败的情况？
- 读取Loghub同步某字段有数据但是同步过来为空
- 读取Loghub同步少数据
- 读取Loghub字段映射时读到的字段不符合预期
- 读取kafka配置了endTime来指定所要同步的数据的截止范围，但是在目的数据源中发现了超过这个时间的数据。
- 写入OSS出现随机字符串如何去除？
- MySQL分库分表如何将分表同步到一张MaxCompute中

为什么数据源测试连通性成功，但是离线同步任务执行失败？

- 如果您之前测试连通性通过，请重新测试连通性，确认资源组与数据库当前是可联通状态（确认数据库端无变更）。
- 请检查测试连通性成功的资源组和任务执行所用的资源组是不是同一个。

查看任务运行资源：

- 任务运行在默认的资源组上，日志中会出现如下信息：`running in Pipeline[basecommon_group_XXXXXX]`。
 - 任务运行在数据集成自定义资源组上，日志中会出现如下信息：`running in Pipeline[basecommon_XXXXXX]`。
 - 任务运行在独享数据集成资源上，日志中会出现如下信息：`running in Pipeline[basecommon_S_res_group_XXX]`。
- 如果凌晨任务调度时偶尔失败，但重跑又能成功，请检查数据库在报错时间点的负载情况。

如何切换数据集成任务执行资源组？

您可以在DataStudio修改数据集成任务调试所用的资源组，同时，也支持您在运维中心修改任务调度时所使用的数据集成任务执行资源组。配置详情请参见：[新增和使用独享数据集成资源组](#)。

脏数据如何排查和定位？

脏数据定义：单条数据写入目标数据源过程中发生了异常，则此条数据为脏数据。因此只要是写入失败的数据均被归类于脏数据。

脏数据影响：脏数据将不会成功写入目的端。您可以控制是否允许脏数据产生，并且支持控制脏数据条数，数据集成默认允许脏数据产生，您可以在同步任务配置时指定脏数据产生条数。详情可参考：[通过向导模式配置离线同步任务](#)。

- 任务设置允许脏数据：当脏数据产生时，任务继续执行，但脏数据将会被舍弃，不写入目的端。
- 任务控制允许产生的脏数据条数：
 - 设置脏数据允许条数为0条，则当脏数据产生时，任务将失败退出。
 - 设置脏数据允许条数为x条，则当脏数据产生条数超过x条时，任务将失败退出，当脏数据产生条数小于x条时，任务将继续运行，但脏数据将会被舍弃，不写入目的端。

脏数据实时场景分析：

● 场景一：

- 报错现象：

```
{ "message": "写入 ODPS 目的表时遇到了脏数据：第[3]个字段的数据出现错误，请检查该数据并作出修改 或者您可以增大阈值，忽略这条记录.", "record": [{"byteSize":0,"index":0,"type":"DATE"}, {"byteSize":0,"index":1,"type":"DATE"}, {"byteSize":1,"index":2,"rawData":0,"type":"LONG"}, {"byteSize":0,"index":3,"type":"STRING"}, {"byteSize":1,"index":4,"rawData":0,"type":"LONG"}, {"byteSize":0,"index":5,"type":"STRING"}, {"byteSize":0,"index":6,"type":"STRING"} ] }
```
- 如何处理：该日志中可以看出脏数据的字段，第三个字段异常。
 - 脏数据是writer端报的，要检查下writer端的建表语句。odps端该表字段指定的字段大小小于MySQL端该字段数据大小。
 - 数据同步原则：来源端数据源的数据要能写入目的端数据源（来源端和目的端类型需要匹配，字段定义的大小需要匹配），即源端数据类型需要与写端数据类型匹配，源端是VARCHAR类型的数据不可写到INT类型的目标列中；目标端的数据类型定义的大小需要可以接收源端映射字段实际数据大小，源端是long、varchar、double等类型的数据，目的端均可用string、text等大范围类型接纳。
 - 脏数据报错不清晰时，需要复制出打印出的脏数据的一整条，观察其中的数据，和目的端数据类型比较，看哪一条或哪一些不合规范。

比如：

```
{ "byteSize":28,"index":25,"rawData":"ohOM71vdGKqXOqtmtriUs5QqJsf4", "type":"STRING" }
```

byteSize：字节数；index:25，第26个字段；rawData：具体值（即value）；type：数据类型。

● 场景二：

- 报错现象：datax读取MySQL中有空值时报脏数据。
- 如何处理：请检查源端存在空值的字段将写入目标端什么类型的字段中，如果类型不一致将会报错。例如，string类型的null写入目标端int类型字段时会报错。

如何排查离线同步任务运行时间长的问題？

可能原因1：执行时间过长

- 前置语句或后置语句（例如：preSql和postSql）在数据库执行耗时过长，导致任务运行慢。
- 未合理配置切分键，导致任务运行慢。

离线同步会根据切分键（splitPk）来对数据进行分片，数据同步根据此配置启动并发任务进行数据同步，提高数据同步的效能。（具体插件是否需要配置切分键，请参考具体插件文档）。

解决方案1：

- 若配置前置或后置语句，建议使用添加了索引的字段进行数据过滤。
- 支持设置切分键的场景下，建议合理配置切分键。以MySQL Reader插件切分键配置为例：
 - 推荐splitPk用户使用表主键，因为表主键通常情况下比较均匀，因此切分出来的分片也不容易出现数据热点。
 - 目前splitPk仅支持整型数据切分，不支持字符串、浮点和日期等其他类型。如果指定其他非支持类型，将使用单通道进行同步。
 - 如果不填写splitPk，包括不提供splitPk或者splitPk值为空，数据同步视作使用单通道同步该表数据。

可能原因2：等待数据集成任务执行资源

解决方案2：若日志出现长时间WAIT状态，说明当前任务运行所使用的独享数据集成资源组剩余可运行的并发数不足以运行当前任务。具体原因及解决方案详情请参见：[为什么数据集成任务一直显示wait？](#)。

说明 由于离线同步任务通过调度资源组下发到数据集成执行资源组上执行，所以一个离线同步任务将同时耗费一个调度资源，若离线同步任务长时间运行未释放资源，除了阻塞其他离线任务运行外，可能还将阻塞其他类型的调度任务运行。

如何处理编码格式设置/乱码问题导致的脏数据报错？

● 报错现象：

如果数据中包括表情符，在同步过程中可能会报错脏数据：`[13350975-0-0-writer] ERROR StdoutPluginCollector - 脏数据 {"exception":"Incorrect string value: '\\xF0\\x9F\\x98\\x82\\xE8\\xA2...' for column 'introduction' at row 1","record":[{"byteSize":8,"index":0,"rawData":9642,"type":"LONG"},],"type":"writer"}。`

● 可能原因：

- 数据库相关编码未设置为utf8mb4，导致同步表情符报错。
- 源端的数据本身就是乱码。
- 数据库和客户端的编码不一样。
- 浏览器编码不一样，导致预览失败或乱码。

● 解决方案：

针对产生乱码的不同原因，选择相应的解决方法：

- 如果您的原始数据乱码，需首先处理好原始数据，再进行同步任务。
- 数据库和客户端编码格式不一致，需先修改编码格式。
- 浏览器编码和数据库或客户端编码格式不一致，需先统一编码格式，然后进行数据预览。

您可以尝试以下操作：

- i. JDBC格式添加的数据源修改utf8mb4：`jdbc:mysql://xxx.x.x.x:3306/database?com.mysql.jdbc.faultInjection.serverCharsetIndex=45`。
- ii. 实例ID形式添加数据源：在数据库名后拼接，格式为 `database?com.mysql.jdbc.faultInjection.serverCharsetIndex=45`。
- iii. 修改数据库相关的编码格式为utf8mb4。例如，在RDS控制台修改RDS的数据库编码格式。

说明 设置RDS数据源编码格式命令：`set names utf8mb4`。查看RDS数据库编码格式命令：`show variables like 'char%'`。

数据同步时报错： `[TASK_MAX_SLOT_EXCEED]:Unable to find a gateway that meets resource requirements. 20 slots are requested, but the maximum is 16 slots.`

● 可能原因：

设置并发数太大导致没有足够的资源。

● 解决方案：

减小离线同步任务的并发数设置。

- 如果您使用向导模式配置离线同步任务，需要调小配置通道控制中的任务期望最大并发数，详情可参见[通过向导模式配置离线同步任务](#)。

- 如果您使用脚本模式配置离线同步任务，需要调小配置通道控制中的concurrent参数，详情可参见[通过脚本模式配置离线同步任务](#)。

任务存在SSRF攻击 Task have SSRF attacks 如何处理？

目前基于数据库内网的数据源不能使用公共数据集成资源组运行任务，请更换使用独享数据集成资源组运行任务，独享数据集成资源组的使用详情可文档：[新增和使用独享数据集成资源组](#)。

离线同步报错： OutOfMemoryError: Java heap space

出现上述报错后，您需要：

1. 如果您使用独享数据集成资源组的话，可以调整jvm参数。
2. 如果插件配置支持的参数中有batchsize或者maxfilesize，可以调小对应的值。
您可以查看各插件是否支持上述参数，进入[支持的数据源与读写插件](#)文档后，点击对应插件查看参数详情。
3. 调小并发数。
 - 如果您使用向导模式配置离线同步任务，需要调小配置通道控制中的任务期望最大并发数，详情可参见[通过向导模式配置离线同步任务](#)。
 - 如果您使用脚本模式配置离线同步任务，需要调小配置通道控制中的concurrent参数，详情可参见[通过脚本模式配置离线同步任务](#)。
4. 如果是文件同步，例如同步OSS文件等，请减少读取的文件数。

离线同步任务执行偶尔成功偶尔失败

出现离线同步任务偶尔执行失败的情况，可能是因为白名单配置不全导致的，您需要检查数据库白名单是否配置完全。

- 使用独享数据集成资源组时：
 - 如果您此前已经将独享数据集成资源组的弹性网卡IP添加到了数据源的白名单中，后续如果资源组有扩容，您需要及时更新数据源的白名单，将扩容后的资源组的弹性网卡IP添加到数据源白名单中。
 - 为了避免出现资源组扩容后需要再次添加白名单的情况，建议您直接添加独享数据集成资源组绑定的交换机网段作为数据库白名单。详情请参见[添加白名单](#)。
- 使用公共数据集成资源组时：

请确保公共数据集成资源组对应地域的所有IP地址段都已经添加到数据源的白名单配置中。详情请参考文档[使用公共数据集成（调试）资源组执行任务时需要在数据库添加的IP白名单](#)。

如果白名单无异常，请确认数据库负载是否过高导致连接中断。

添加MongoDB数据源时，使用root用户时报错

使用root用户添加MongoDB数据源时报错，是因为添加MongoDB数据源时，必须使用用户需要同步的表所在的数据库创建的用户名，不可以使用root用户。

例如需要导入name表，name表在test库，则此处数据库名称为test，需要使用test数据库中创建的用户名。

MongDB使用的authDB库为admin库，如何同步其他业务库数据？

请在数据源配置中库名填写业务库名，确保username权限正常，若数据源连通性报错auth failed可先忽略。需要通过脚本模式配置任务，在同步任务JSON配置parameter配置项中添加"authdb":"admin"参数。

读取MongoDB时，如何在query参数中使用timestamp实现增量同步？

可以使用赋值节点先对date类型时间处理成timestamp，将该值作为MongoDB数据同步的入参，详情请参考文档：[MongoDB时间戳类型字段如何实现增量同步？](#)

读取OSS数据报错： AccessDenied The bucket you access does not belong to you.

数据源配置的ak用户没有该bucket的权限，请为OSS数据源配置的ak用户授予该bucket相应权限。

读取OSS文件是否有文件数限制？

OSSreader插件读取OSS执行时间超过RDS侧允许的最大文件数datax本身没有限制，主要看任务配置的jvm参数，但是不建议object参数配置*，容易导致内存溢出。

写入Redis使用hash模式存储数据时，报错如下： Code: [RedisWriter-04], Description: [Dirty data]. - source column number is in valid!

● 产生原因：

Redis在使用hash模式存储时，hash的attribute和value要成对出现。例如：`odpsReader: "column": ["id", "name", "age", "address",]`，目标端在配置RedisWriter: "keyIndexes": [0, 1]，在Redis中，id和name将作为key，age将作为attribute，address将作为value存到Redis中的hash类型的Redis中，如果odps源端只配置了两列，则不可以使用hash模式去存Redis缓存，会报该异常。

● 解决方案：

假如只想用两列，那就配置Redis的tring模式去存信息，如果必须要用hash存，源端至少配置三列。

写入/读取MySQL报错： Application was streaming results when the connection failed. Consider raising value of 'net_write_timeout/net_read_timeout、' on the server.

● 报错原因：

- net_read_timeout：datax将RDS MySQL根据splitpk对MySQL的数据进行拆分成数条等量的取数sql（select取数语句），执行时某条sql执行时间超过RDS侧允许的最大运行时间。
- net_write_timeout：等待将一个block发送给客户端的超时时间过小导致。

● 解决方案：

在数据源URL的连接上增加该参数，net_write_timeout/net_write_timeout设置稍微大一些，或者在RDS控制台调整该参数。

● 改进建议：

如果任务可以重跑建议设置任务自动重跑。

参数名称	当前值	默认值	是否可配置	取值范围	备注
local_infile	OFF	ON	否	[ON OFF]	
loose_opt_rds_last_error_gtid	ON	ON	否	[ON OFF]	
innodb_ft_max_token_size	84	84	是	[10-84]	
innodb_undo_log_truncate	OFF	OFF	否	[ON OFF]	
loose_performance_schema_users_size	10000	10000	是	[-1-1048576]	
innodb_max_undo_log_size	1073741824	1073741824	否	[10485760-18446744073709551615]	
slave_parallel_type	LOGICAL_CLOCK	LOGICAL_CLOCK	否	[DATABASE LOGICAL_CLOCK]	
loose_rds_check_core_file_enabled	ON	ON	否	[ON OFF]	
loose_thread_pool_enabled	ON	OFF	否	[ON OFF]	
innodb_adaptive_hash_index	OFF	ON	否	[ON OFF]	
innodb_sync_spin_loops	30	30	否	[0-4294967295]	
net_write_timeout	60	60	否	[-1-18446744073709551615]	
flush_time	0	0	否	[0-31536000]	
lower_case_table_names	1	1	是	[0 1]	
sha256_password_proxy_users	OFF	OFF	否	[ON OFF]	
character_set_filesystem	binary	binary	否	[utf8 latin1 gbk binary]	
innodb_flush_sync	ON	ON	否	[ON OFF]	
tls_version	TLSv1,TLSv1.1,TLSv1.2	TLSv1,TLSv1.1,TLSv1.2	是	[TLSv1,TLSv1.1,TLSv1.2 TLSv1,TLSv1...	
key_cache_division_limit	100	100	否	[1-100]	

例如：`jdbc:mysql://192.168.1.1:3306/lizi?useUnicode=true&characterEncoding=UTF8&net_write_timeout=72000`

读取MySQL数据库报错 `The last packet successfully received from the server was 902,138 milliseconds ago`

CPU使用正常，内存使用较高可能导致连接被断开。

如果您确认可以自动重跑的话，建议您设置任务出错自动重跑，详情参考文档：[时间属性配置说明](#)

读取PostgreSQL数据报错，该如何处理？

- 问题场景：离线同步工具同步PostgreSQL数据时，报错如下：`org.postgresql.util.PSQLException: FATAL: terminating connection due to conflict with recovery`
- 可能原因：出现该情况的原因是从数据库中拉取数据的时间较长，请增加`max_standby_archive_delay`和`max_standby_streaming_delay`的参数值，详情请参见[Standby Server Events](#)

报错 `Communications link failure`，该如何处理？

- 读取报错
 - 问题现象：

读取数据时，报错如下：`Communications link failure The last packet successfully received from the server was 7,200,100 milliseconds ago. The last packet sent successfully to the server was 7,200,100 milliseconds ago. - com.mysql.jdbc.exceptions.jdbc4.CommunicationsException: Communications link failure`
 - 可能原因：

数据库执行SQL查询速度慢，导致MySQL读超时。
 - 解决方法：
 - 排查是否设置了`where`过滤条件，以确保筛选字段已添加索引。
 - 排查源数据表的数据是否太多。如果数据太多，建议拆分为多个任务。
 - 查询日志找到执行阻塞的SQL，咨询数据库管理员解决。
- 写入报错

o 问题现象：

写入数据时，报错如下：
`Caused by: java.util.concurrent.ExecutionException: ERR-CODE: [TDDL-4614][ERR_EXECUTE_ON_MYSQL] Error occurs when execute on GROUP 'xxx' ATOM 'dockerxxxxx_xxxx_trace_shard_xxxx': Communications link failure The last packet successfully received from the server was 12,672 milliseconds ago. The last packet sent successfully to the server was 12,013 milliseconds ago. More...`

o 可能原因：

慢查询导致SocketTimeout。TDDL默认连接数据的SocketTimeout是12秒，如果一个SQL在MySQL端执行超过12秒仍未返回，会报4614的错误。当数据量较大或服务端繁忙时，偶尔会出现该错误。

o 解决方法：

- 建议数据库稳定后，重新运行同步任务。
- 联系数据库管理员调整该超时时间。

离线同步报错 Duplicate entry 'xxx' for key 'uk_uk_op' 如何处理？

- 报错现象：`Error updating database. Cause: com.mysql.jdbc.exceptions.jdbc4.MySQLIntegrityConstraintViolationException: Duplicate entry 'cfc68cd0048101467588e97e83ffd7a8-0' for key 'uk_uk_op'。`
- 可能原因：数据集成同步任务不允许同一时间运行相同节点的不同实例（即相同JSON配置的同步任务不能同一时间运行多个，例如5分钟为周期的同步任务，由于上游延迟，在0点5分时调起了原本0点0分的实例和0点5分的实例，这样会导致其中一个实例无法调起，或者任务实例在运行时又进行了补数据、重跑等操作）。
- 解决方案：错开实例运行时间，如果为小时任务建议设置[自依赖](#)。

离线同步报错 plugin xx does not specify column 如何处理？

出现上述报错，有可能是因为同步任务的字段映射配置不正确，或者是插件没有正确配置column。

1. 检查是否有做字段映射。
2. 检查插件是否有正确配置column。

读取MaxCompute (ODPS) 表数据报错 The download session is expired.

- 报错现象：
`Code:DATAX_R_ODPS_005:读取ODPS数据失败, Solution:[请联系ODPS管理员]. RequestId=202012091137444331f60b08cda1d9, ErrorCode=StatusConflict, ErrorMessage=The download session is expired.`
- 可能原因：
离线同步读取MaxCompute数据时，使用的是MaxCompute的tunnel命令来进行上传下载数据。Tunnel的Session在服务端的生命周期为24小时，所以离线同步任务如果执行超过24小时任务会失败退出，关于tunnel介绍详情请参见[使用说明](#)。
- 解决方案：
您可以适当调大离线同步任务并发数，合理规划同步的数据量确保任务在24小时内同步完成。

写入MaxCompute(ODPS) 报错block失败： Error writing request body to server

- 报错现象：

```
Code:[OdpsWriter-09], Description:[写入数据到 ODPS 目的表失败.] - ODPS 目的表写 block:0 失败, uploadId=[202012081517026537dc0b0160354b]. 请联系 ODPS 管理员处理. - java.io.IOException: Error writing request body to server.
```

- 可能原因：
 - 可能原因1：数据类型异常，即来源数据不符合odps数据类型规范。比如4.2223这个数据值写入odps的decimal (18,10) 数据类型的场景。
 - 可能原因2：odps块/通信异常等
- 解决方案：

请转换数据类型，使用符合数据类型规范的数据。如果数据类型排查后仍有报错，可提交工单咨询排查。

写入DataHub时，一次性写入数据超限导致写入失败如何处理？

- 报错现象：


```
ERROR JobContainer - Exception when job runcom.alibaba.datax.common.exception.DataXException: Code:[DatahubWriter-04], Description:[写数据失败.] - com.aliyun.datahub.exception.DatahubServiceException: Record count 12498 exceed max limit 10000 (Status Code: 413; Error Code: TooLargePayload; Request ID: 20201201004200a945df0bf8e11a42)
```
- 可能原因：

报错原因为datax一次性提交到DataHub的数据量超过DataHub的限制。影响提交到DataHub数据量的配置参数主要为：

 - **maxCommitSize**：表示datax累积的buffer数据，待积累的数据大小达到maxCommitSize 大小（单位MB）时，批量提交到目的端。默认是1,048,576，即1MB数据。
 - **batchSize**：DataX-On-Flume积累的Buffer数据条数，待积累的数据条数达到batchSize大小（单位条数）时，批量提交至目的端。
- 解决方案：

调小maxCommitSize和batchSize的参数值。

离线同步任务如何自定义表名？

如果您表名较规律，例如orders_20170310、orders_20170311和orders_20170312，按天区分表名称，且表结构一致。您可以结合调度参数通过脚本模式配置离线同步任务自定义表名，实现每天凌晨自动从源数据库读取前一天表数据。

例如，今天是2017年3月15日，自动从源数据库中读取orders_20170314的表的数据导入，以此类推。



在脚本模式中，修改来源表的表名为变量，例如orders_`\${tablename}`。由于按天区分表名称，且需要每天读取前一天的数据，所以在任务的参数配置中，为变量赋值为tablename=\${yyyymmdd}。

说明

更多调度参数的使用问题请参考文档[调度参数概述](#)

配置离线同步节点时，无法查看全部的表，该如何处理？

您在配置离线同步节点时，选择来源区域仅默认显示所选数据源中的前25张表。如果数量过大，您可以输入表名进行搜索或使用脚本模式开发。

读取MaxCompute（ODPS）表数据时，添加一行注意事项

1. 可以输入常量，输入的值需要使用英文单引号包括，例如 'abc'、'123'等。
2. 可以配合调度参数使用，如 '\${bizdate}'等，调度参数具体使用请参考文档[调度参数概述](#)。
3. 可以输入您要同步的分区列，如分区列有pt等。
4. 如果您输入的值无法解析，则类型显示为 '自定义'。
5. 不支持配置odps函数。
6. 如果您手工编辑添加的列显示自定义（例如添加了MaxCompute分区列，LogHub数据预览未预览到的列），不会影响实际任务执行。

读取MaxCompute（ODPS）表数据时，如何同步分区字段？

请在字段映射处选择添加一行，输入分区列名，如分区列有pt等；

读取MaxComput（ODPS）表数据时，如何同步多个分区数据？

读取数据所在的分区信息。

- ODPS的分区配置支持linux shell通配符，*表示0个或多个字符，?表示任意一个字符。
- 默认情况下，读取的分区必须存在，如果分区不存在则运行的任务会报错。如果您希望当分区不存在时任务仍然执行成功，则可以切换至脚本模式执行任务，并在ODPS的Parameter中添加 `"successOnNoPartition": true` 配置。

例如，分区表test包含pt=1,ds=hangzhou、pt=1,ds=shanghai、pt=2,ds=hangzhou、pt=2,ds=beijing四个分区，则读取不同分区数据的配置如下：

- 如果您需要读取pt=1,ds=hangzhou分区的数据，则分区信息的配置为 `"partition": "pt=1,ds=hangzhou"`。
- 如果您需要读取pt=1中所有分区的数据，则分区信息的配置为 `"partition": "pt=1,ds=*"` 。
- 如果您需要读取整个test表所有分区的数据，则分区信息的配置为 `"partition": "pt=*,ds=*"` 。

此外，您还可以根据实际需求设置分区数据的获取条件（以下操作需要转脚本模式配置任务）：

- 如果您需要指定最大分区，则可以添加 `/*query*/ ds=(select MAX(ds) from DataXODPSReaderPPR)` 配置信息。
- 如果需要按条件过滤，则可以添加相关条件 `/*query*/ pt+表达式` 配置。例如 `/*query*/ pt>=20170101 and pt<20170110` 表示获取pt分区中，20170101日期之后（包含20170101日期），至20170110日期之前（不包含20170110日期）的所有数据。

说明

`/*query*/` 表示将其后填写的内容识别为一个where条件。

如何处理表列名有关键字导致同步任务失败的情况？

- 报错原因：column中含有保留字段，或者column配置中含有数字开头的字段。
- 解决方法：数据集成同步任务转脚本模式配置，对column配置中的特殊字段添加进行转义。脚本模式配

置任务请参考：[通过脚本模式配置离线同步任务](#)。

- MySQL的转义符为 ``关键字``。
- Oracle和PostgreSQL的转义符为 `"关键字"`。
- SQLServer的转义符为 `[关键字]`。

MySQL为场景示例：

```

{
  "stepType": "mysql",
  "parameter": {
    "envType": 0,
    "datasource": "wpw_test_mysql",
    "column": [
      "id",
      "`order`",
      "`add`"
    ],
    "connection": [
      {
        "datasource": "wpw_test_mysql",
        "table": [
          "abc"
        ]
      }
    ]
  }
}

```

- 以MySQL数据源为例：

- 执行下述语句，新建一张表aliyun。 `create table aliyun (`table` int ,msg varchar(10));`
- 执行下述语句，创建视图，为table列取别名。 `create view v_aliyun as select `table` as col1 ,msg as col2 from aliyun;`

说明

- table是MySQL的关键字，在数据同步时，拼接出来的代码会报错。因此需要通过创建视图，为table列起别名。
- 不建议使用关键字作为表的列名。

- 通过执行上述语句，为有关键字的列取列名。在配置同步任务时，选择v_aliyun视图代替aliyun表即可。

读取Loghub同步某字段有数据但是同步过来为空

此插件字段大小写敏感，请检查loghub reader的column配置。

读取Loghub同步少数据

数据集成取的是数据进入loghub的时间，请在Loghub控制台检查数据元数据字段receive_time是否在任务配置的时间区间内。

读取Loghub字段映射时读到的字段不符合预期

如果出现此现象请在界面手动编辑column。

读取kafka配置了endTime来指定所要同步的数据的截止范围，但是在目的数据源中发现了超过这个时间的数据。

kafkareader在读数据的时候，是以batch的形式进行的，在读到的一批数据中，如果有超过endTime的，则停止同步，但是超出endTime的这部分数据也会被写入到目的端数据源。

- 您可以通过skipExceedRecord配置项来指定是否同步超出的数据。详细使用方法见[Kafka Reader](#)。【不建议设置为不同步，会有造成丢数据】
- 您可以通过配置kafka的max.poll.records配置项来指定一次拉取数据的量，结合并发数，来控制可能超出的数据量。超出的数据量 < max.poll.records * 并发数

写入OSS出现随机字符串如何去掉？

OSS Writer写入的文件名，OSS使用文件名模拟目录的实现。OSS对于Object的名称有以下限制：使用"object": "datax"，写入的Object以datax开头，后缀添加随机字符串。文件数由实际切分的任务数有关。

如果您不需要后缀随机UUID，建议您配置"writeSingleObject": "true"，详情请参见[OSS Writer](#)文档的writeSingleObject说明。

详情请参考文档：[\[OSS writer\] \(OSS Writer\)](#)

MySQL分库分表如何将分表同步到一张MaxCompute中

您可以参考MySQLreader插件文档进行任务配置：[MySQL Reader](#)

RestAPI Writer报错：通过path:[] 找到的JSON串格式不是数组类型

RestAPI Writer提供两种写入模式，当同步多条数据的时候需要将dataMode配置为multiData，详情请参见[RestAPI Writer](#)，同时需要在RestAPI Writer脚本中新增参数dataPath:"data.list"。

参数	描述	是否必须	默认值
url	RESTful接口地址。	是	无
dataMode	RESTful请求传递的JSON数据的格式。 <ul style="list-style-type: none">• oneData：一次请求只传递一条数据，有几条数据就进行几次请求。• multiData：一次请求传递一批数据，根据reader端切分的task数确定请求次数。	是	无

 **注意** 配置Column时不需要加“data.list”前缀。

3. 调度配置

3.1. 调度参数

本文为您介绍调度参数的相关问题。

- 调度参数典型场景
 - 调度参数往前取一个小时，如何处理跨天参数替换的问题
 - 如何处理表的分区格式年月日时分秒间带空格情况：pt=yyyy-mm-dd hh24:mi:ss
 - 节点定时时间为每天20:00，节点内参数取值为定时时间（\$cyctime或者\$[yyyymmddhh24miss]），由于上游节点执行异常导致当前节点第二天0:00才调起，这个时候参数取值是20:00还是0:00?
 - 如何配置ODPS Spark节点中的时间参数?
 - 如何对不便于进行二次处理调度参数返回值的节点进行取值?
 - 直接使用MaxCompute日期函数取值和通过调度参数取值有什么区别?
- 调度参数测试运行常见问题
 - 如何在数据开发测试调度参数?
 - FAILED: ODPS-0130161:[1,84] Parse exception - invalid token '\$'
 - params format error, please check your params (key=values)
- 运行按钮的赋值逻辑差异

运行，高级运行和开发环境冒烟测试赋值逻辑有什么区别
- 调度参数运维与查看
 - 如何查看生产环境调度参数替换情况?
 - 如何通过日志查看当前实例执行时调度参数替换情况?
 - 如何定义调度参数的时区?
 - 节点中调度参数已配置且提交发布，但返回值始终无变化。

调度参数往前取一个小时，如何处理跨天参数替换的问题

- 问题描述：表分区格式day=yyyymmdd,hour=hh24，且任何执行的时候小时都取上个小时数据，\$[yyyymmdd] \$[hh24-1/24]是可以满足，但是0点的时候会变成当天的23点，怎么办?
- 解决办法：您可以修改参数的计算公式，修改datetime为\$[yyyymmdd-1/24]，hour的计算公式仍然是\$[hh24-1/24]。

配置：

- 代码里 `day=datetime, hour={hour},`
- 节点配置参数赋值 `datetime=[yyyymmdd-1/24],hour=[hh24-1/24]`

实时场景：

- 如果一个实例的定时时间是2021-7-21 00:00:00，减1小时便是昨天，则\$[yyyymmdd-1/24]的值是20210720、\$[hh24-1/24]的值是23。
- 如果一个实例的定时时间为2021-7-21 01:00:00的实例，减1小时还是今天，则\$[yyyymmdd-1/24]的值是20210721、\$[hh24-1/24]的值是00。

如何处理表的分区格式年月日时分秒带空格情况：pt=yyyy-mm-dd hh24:mi:ss

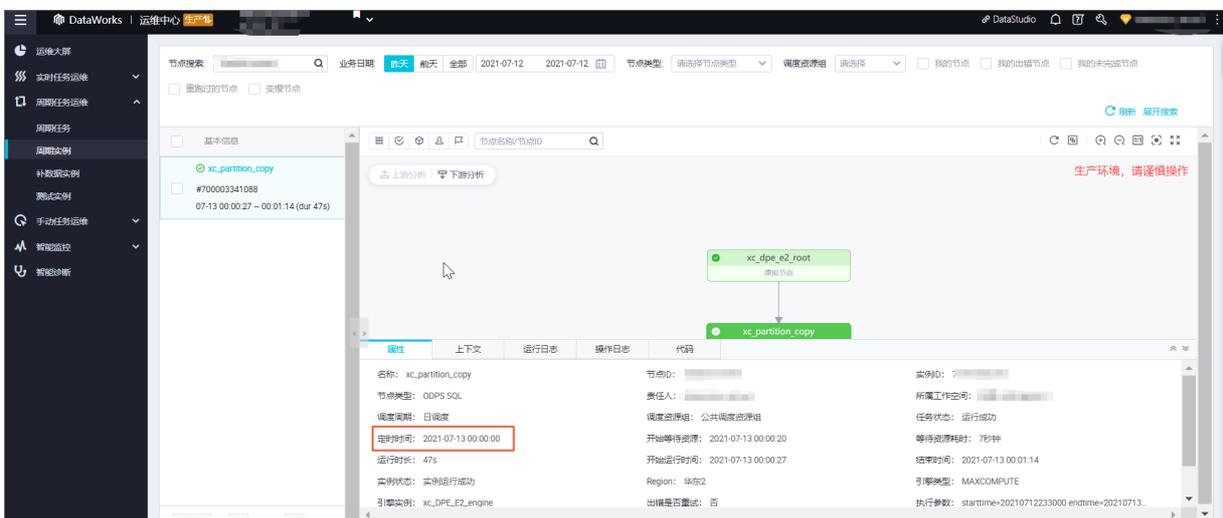
注意 调度参数赋值中不支持带空格。

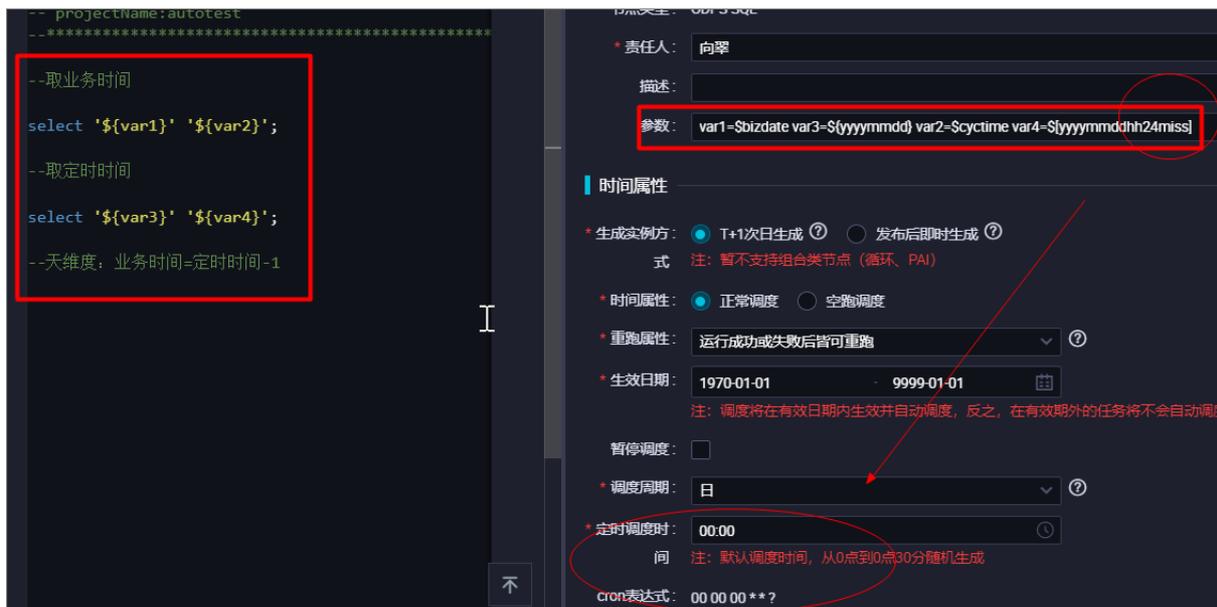
您可以使用两个自定义变量分别赋值年月日和时分秒，参数dateime=\${yyyy-mm-dd}、hour=\${hh24:mi:ss}，分别获取日期和时间后，在代码中用空格拼接为pt=\${dateime} \${hour}。



节点定时时间为每天20:00，节点内参数取值为定时时间（\$cyctime或者\${yyyymmddhh24miss}），由于上游节点执行异常导致当前节点第二天0:00才调起，这个时候参数取值是20:00还是0:00？

实例的定时时间可以理解为预期状态下实例运行时间，定时时间与实例实际运行时间是可能存在差异的（任务实际运行时，可能会因为等待资源等原因导致实际运行时间与定时时间不一致，实例定时时间是绝对的，不会随着实例的运行时间改变而改变），调度参数这里取的是实例的定时时间，节点调度参数的替换值在实例生成时就已经替换，不会随着实例实际运行时间的改变而改变，所以您无需顾虑调度参数取值是否会变动。

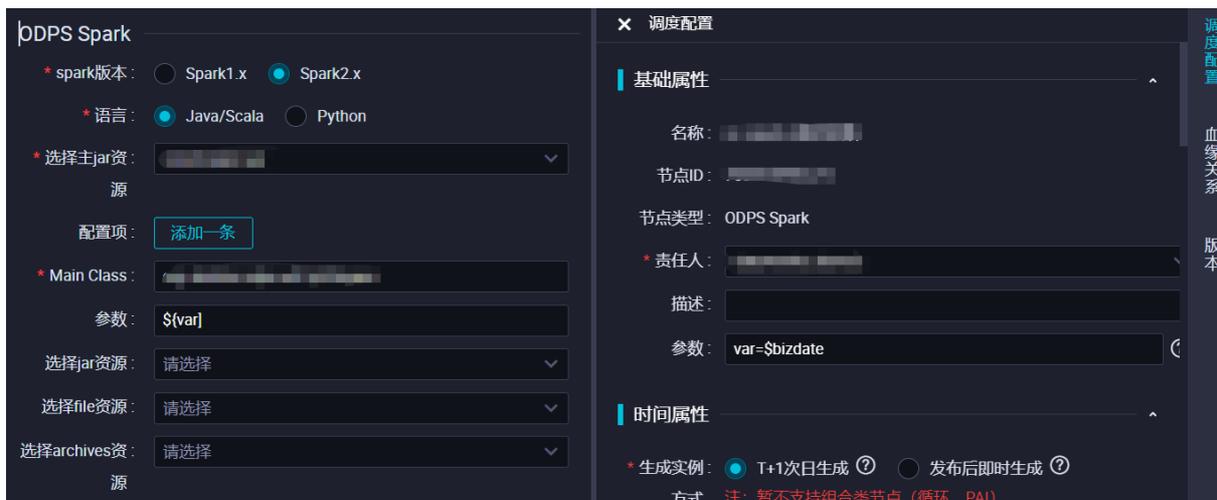




如何配置ODPS Spark节点中的时间参数？

您在创建ODPS Spark节点时，需要在编辑页面配置节点使用的参数，系统会顺序执行。

配置代码中变量后，需要单击编辑页面右侧的调度配置，在参数面板为上述变量赋值，调度参数使用详情请参见本文内容。



如何对不便于进行二次处理调度参数返回值的节点进行取值？

部分节点（例如，离线同步节点）配置了调度参数后，调度参数的返回值不能直接使用，需要进行转换等二次处理。该类节点，您可以配置赋值节点为该节点的上游，使用赋值节点先引用调度参数并对参数返回结果进行转换处理，再通过节点上下文功能，将最终的参数结果传递给下游的目标节点。配置赋值节点，详情请参见[赋值节点](#)，配置节点上下文，详情请参见[配置节点上下文](#)

直接使用MaxCompute日期函数取值和通过调度参数取值有什么区别？

- 直接使用MaxCompute日期函数取值，其返回结果是执行计算任务的机器所显示的时间。当在不同时间执行同一个实例时，返回时间会改变。
- 通过调度参数取值，其返回结果是基于定时时间进行计算的。无论何时执行同一个实例，返回时间均不会改变。

如何在数据开发测试调度参数？

由于调度参数只有走调度系统才能被自动替换，所以如果您需要在数据开发测试调度参数在调度系统的替换情况，您可以选择开发环境冒烟测试。

说明 由于数据集成任务没有开发环境测试功能，所以如果您需要测试调度参数替换情况，建议您建一个sql节点使用开发环境冒烟测试先测试调度参数，测试符合预期后，再在数据集成节点中使用。

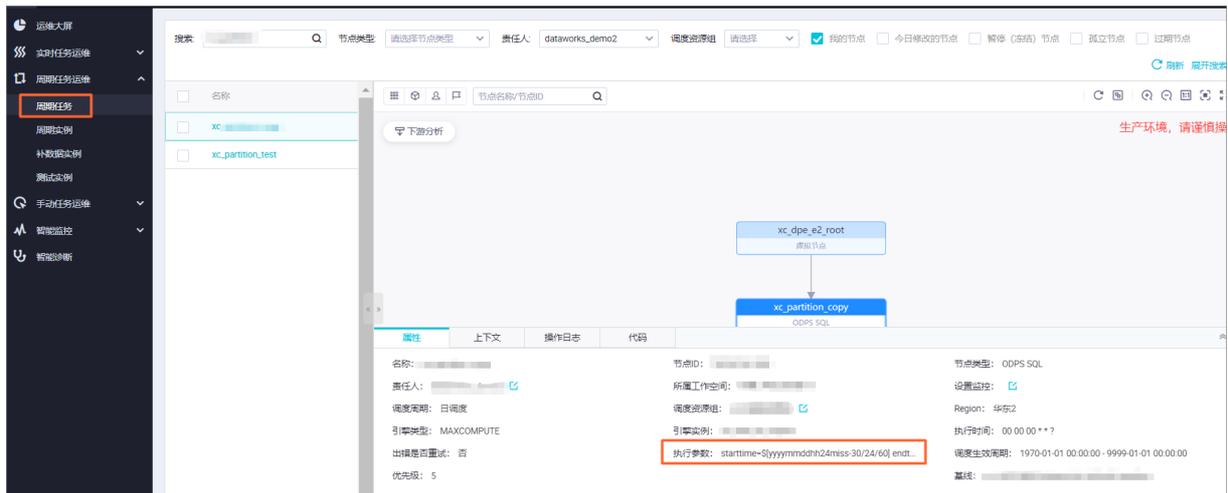
FAILED: ODP5-0130161:[1,84] Parse exception - invalid token '\$'

报错原因：调度参数未赋值或赋值出现异常。

解决办法：

1. 检查是否有给调度参数赋值。
2. 检查配置的调度参数是否符合要求（详情参考文档：[调度参数概述](#)）

注意 修改参数配置后记得提交发布，发布后请在生产运维中心周期任务的属性面板查看参数是否是最新的配置。



params format error, please check your params (key=values)

1. 检查是否为变量赋值
2. 检查调度参数那是不是多加了空
3. 节点名中点和中文不能一块使用

time①=②\${yyyyymmdd③hh24:mi:ss]time1=\${yyyyymmdd]④time2=\${hh24:mi:ss]（①②③④为空格所在位置）。

调度参数的等号（=）两端不可以加空格，即示例中的①②处。

调度参数赋值中不支持空格，即示例中的③处。

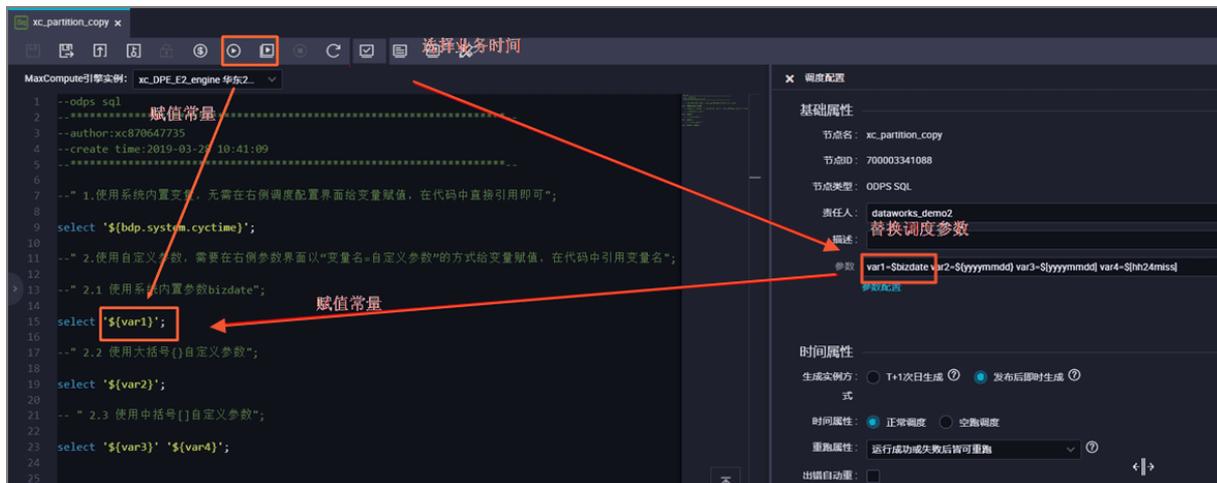
两个调度参数间用一个空格分隔，即示例中的④处。

运行，高级运行和开发环境冒烟测试赋值逻辑有什么区别

- 运行：第一次运行后，您需要手动为该变量赋值常量，第一次赋值会被记录。如果对代码进行修改，该变量依旧使用第一次运行时的赋值。
- 高级运行：高级运行时，您需要手动为当前代码中的变量进行赋值常量，如果修改了代码中的变量，请使

用高级运行给变量重新赋值常量。

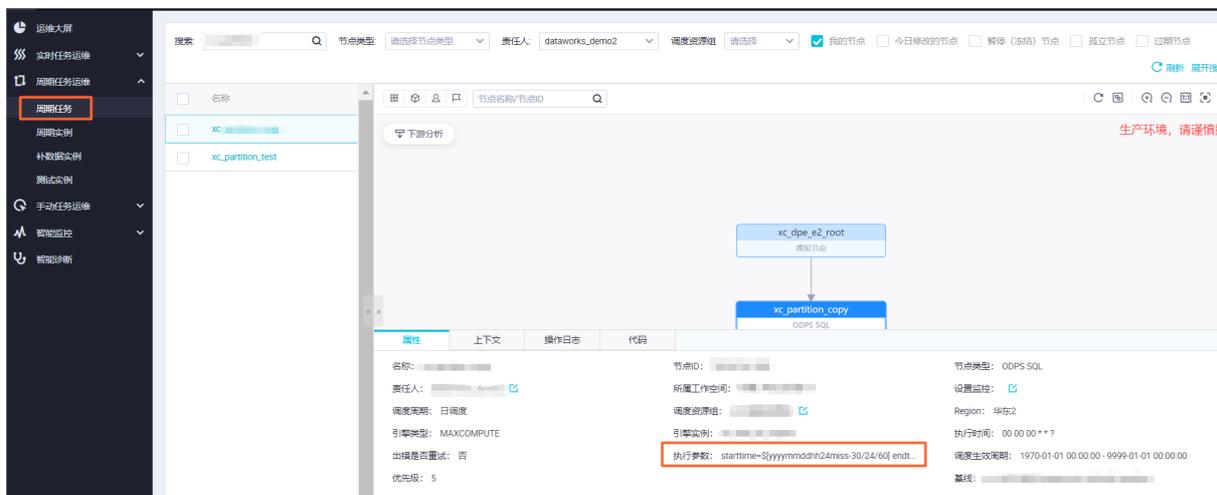
- 开发环境执行冒烟测试：通过手动选择业务日期模拟节点自动调度的场景，获取当前业务日期下调度参数的替换值。



说明 如果要修改任务执行资源组，请使用“高级运行”。

如何查看生产环境调度参数替换情况？

如果您在数据开发修改了调度参数并且提交发布，请在生产运维中心周期任务属性面板检查调度参数是否符合预期，如果不符合预期，请检查任务发布，发布包界面，是否发布生成功。



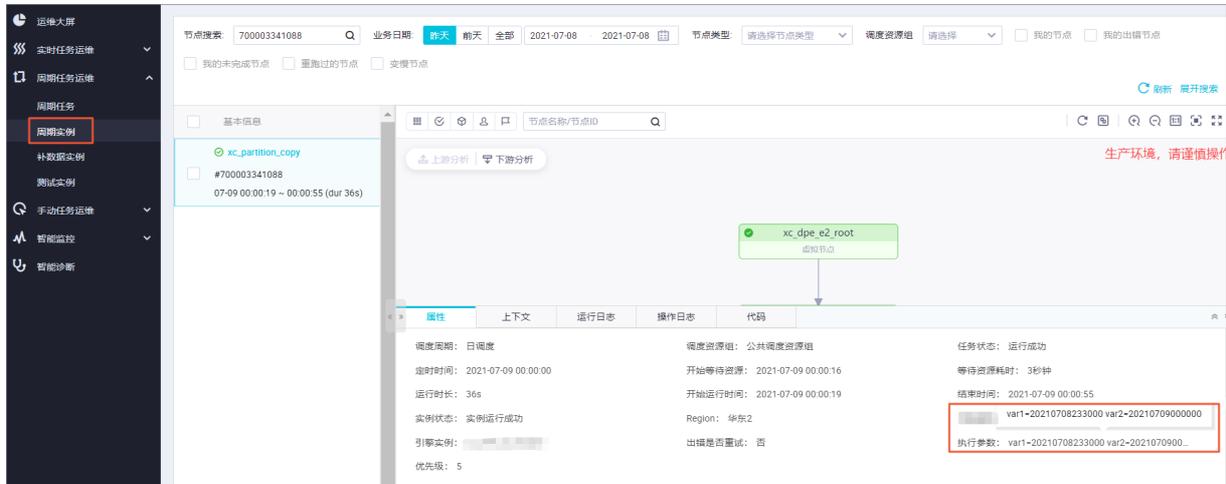
在周期实例属性面板查看单个实例调度参数替换情况。

注意 单个实例不管是否已经运行过，实例的执行参数会实时更新为生产周期任务最新的参数配置。

实时场景：

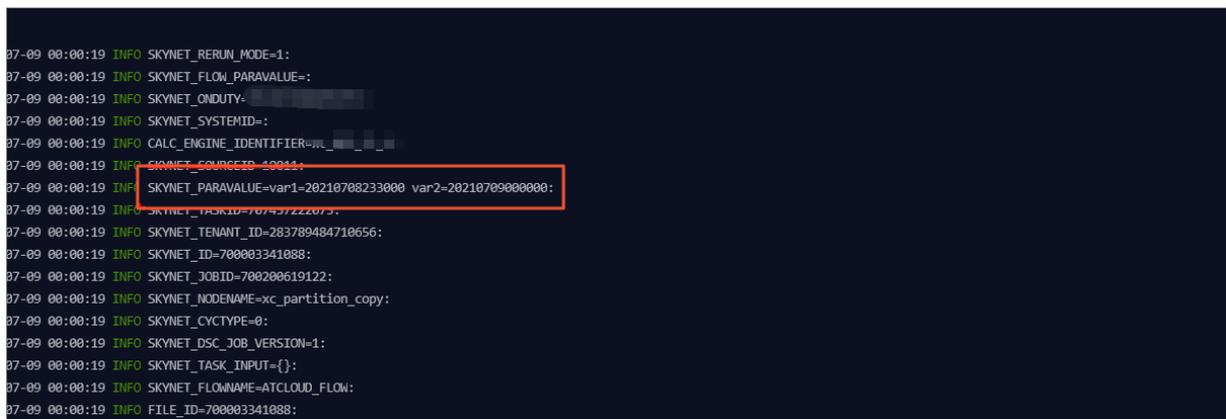
- 节点A当天周期实例A1调度参数time1取值为\$bizdate，当天实例已执行成功，代码中参数取值即为bizdate业务时间（具体值）。
- 在当天某一时间点调度参数time1取值由\$bizdate改为\$cyctime，那么当天实例执行参数此时会显示为cyctime定时时间（具体值）。

- 当前实例重跑，会用最新的参数配置：即time1=\$cyctime
- 如果要看调度参数变更前的实例使用的调度参数替换情况，请在日志中查看，详情参考常见问题：如何在日志中查看调度参数替换情况[如何通过日志查看当前实例执行时调度参数替换情况？](#)。



如何通过日志查看当前实例执行时调度参数替换情况？

在代码中找到 `SKYNET_PARAVALUE` 字样。



如何定义调度参数的时区？

夏令时支持现转实例引擎和参数分区计算，在进入和退出夏令时当天能正确运行。以美国西部时间为例，说明如下：

- 进入夏令时会产生23个实例，当天03:00往前偏移10分钟的结果是01:50。进入夏令时当天定时时间2点的任务会空跑。
- 退出夏令时会产生24个实例，当天03:00往前偏移10分钟的结果是02:50。

进入夏令时当天的日、周和月调度节点，如果设置的运行时间正好是被跳过的时间段，则当天生成的实例运行时间会是当天的0点整。

节点中调度参数已配置且提交发布，但返回值始终无变化。

您需要排查调度参数是否被同名业务流程参数覆盖。排查操作，详情请参见[使用流程参数](#)。当排查到存在同名业务流程参数时，如果不需要使用该业务流程参数，则可以将其删除，如果需要保留该业务流程参数，则需要修改节点调度参数的名称。

3.2. 依赖关系

本文为您介绍依赖关系的相关问题。

- 调度依赖配置您需要先了解以下几个方面
 - 什么是调度依赖
 - 为什么需要设置调度依赖
 - DataWorks上如何设置调度依赖？
 - 哪些场景下不支持设置调度依赖？
 - 不需要依赖的表如何删除？
- 调度依赖配置常见问题
 - 为什么自动解析的父节点名会报依赖的父节点输出（表）不存在？
 - 为什么有些节点的本节点的输出中，下游节点名称、下游节点ID有这些数据，有些为空并且不能手动编辑。
 - 不需要依赖的依赖关系如何删除？
 - 下游任务依赖上游任务的运行规则是什么？
 - 节点的输出名称的作用是什么？
 - 一个节点可以有多个输出名称吗？
 - 多个节点往同一张表写数据，自动解析报错节点输出名相同且提交报错，可以有相同的输出名称吗？
 - 使用自动解析依赖关系时，如何不解析到中间表？
 - 业务流程最上游节点的父节点配置应该如何配置？
- 节点删除、变更常见问题
 - 为什么在节点A搜索上游节点输出名时，搜索到了节点B不存在的输出名？
 - 为什么节点删除报错当前节点存在子节点，下线失败，但是在调度配置界面看已经没有该节点的依赖？
- 跨周期依赖配置实时场景
 - 为什么运维中心有些依赖线是虚线？
 - 小时任务设置自依赖对自身及下游节点有什么影响？
 - 日调度任务依赖小时调度任务运行的不同场景，依赖关系如何配置？
 - 天任务依赖小时任务时，直接设置依赖关系，天任务几点会运行？
 - 天任务依赖小时任务时，如何实现天任务依赖小时任务当天具体某个实例，而非依赖所有的小时实例？
 - 天任务依赖小时任务时，如何实现天任务依赖小时任务昨天所有的小时实例。而非当天所有小时实例？
 - 什么情况下配置依赖上一周期的依赖项为本节点？
 - 下游节点同时依赖多个任务运行时，依赖关系如何配置？
 - 上游A是小时任务，下游B是日任务，任务每天在A任务全部执行完成后要汇总执行一次，小时任务运行到第二天，天任务是否会运行，参数是否会影响
 - 节点A每天每小时整点执行一次，节点B每天运行一次。如何设置节点A每天第一次运行成功后，节点B便开始执行？
- 调度依赖配置实时场景
 - 有A、B、C三个任务，如何实现每个小时执行一次A->B->C（A执行完了B再执行，B执行完了C再执行）的任务流程？
 - 如何设置跨业务流程，同region跨项目依赖？

- 其他常见问题
 - 已配置失败重跑的任务在运行失败后未重跑，并报错Task Run Timed Out, Killed by System!!!

什么是调度依赖

调度依赖就是节点间的上下游依赖关系，在DataWorks中，上游任务节点运行完成且运行成功，下游任务节点才会开始运行。

② 说明

挂上依赖关系后，当前节点执行的必要条件之一是：当前节点依赖的父节点需要执行成功。

为什么需要设置调度依赖

配置调度依赖后，可以保障调度任务在运行时能取到正确的数据（当前节点依赖的上游节点成功运行后，DataWorks通过节点运行的状态识别到上游表的最新数据已产生，下游节点再去取数）。避免下游节点取数据时，上游表数据还未正常产出，导致下游节点取数出现问题。

DataWorks上如何设置调度依赖？

在DataWorks上，上游节点的输出作为下游节点的输入，形成节点依赖关系。

② 说明

- sql任务会以如下方式自动配置节点输入输出：
 - select一张表会将自动解析作为本节点依赖的上游。
 - insert/create一张表，该表会自动解析作为本节点的输出。
- 数据集成任务需要手动将产出表以projectname.tablename方式作为节点输出，下游对同步产出表进行清洗时，自动解析才能解析到。
- 因为需要唯一的输出定位到唯一的节点形成节点依赖关系，所以节点输出（projectname.tablename）需要唯一。

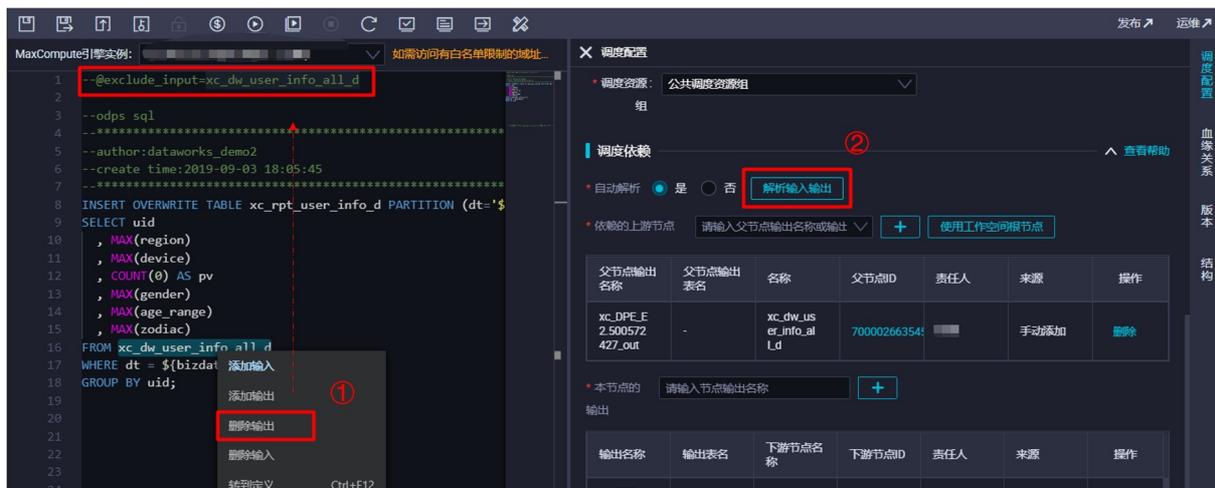
哪些场景下不支持设置调度依赖？

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。非周期性调度生产数据的表包括：

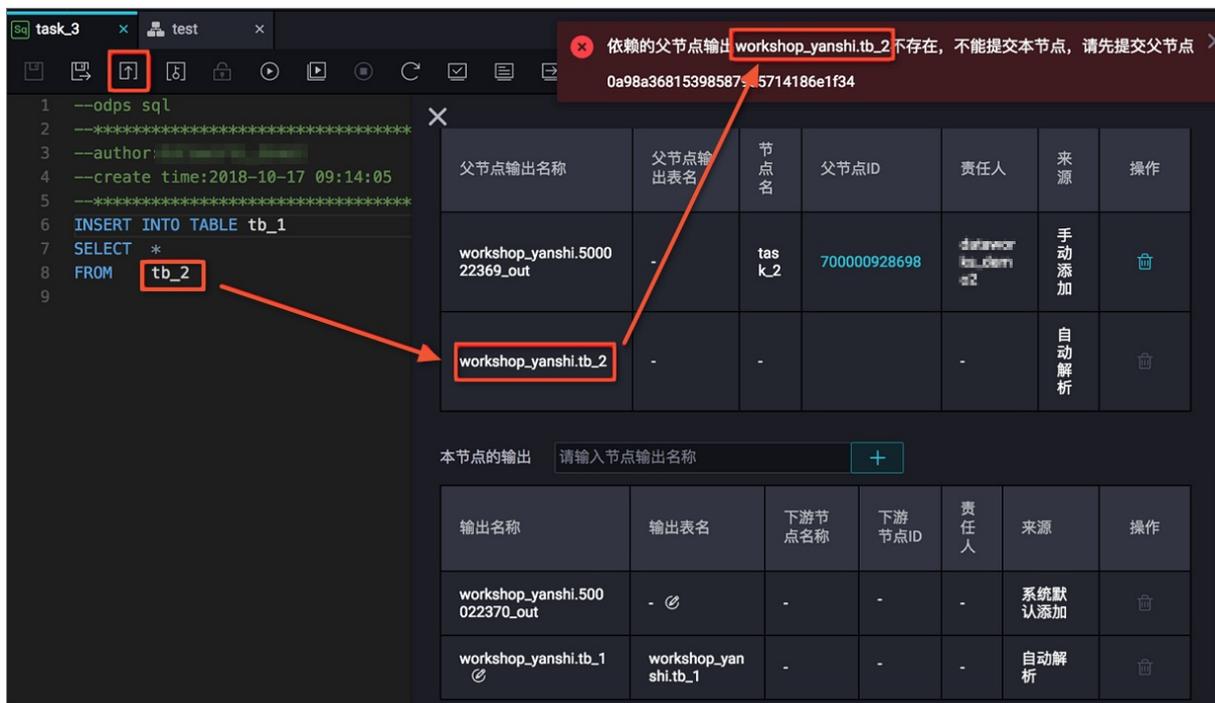
- 从本地上传到DataWorks的表
- 维表
- 非DataWorks调度产出的表
- 手动任务产出的表

不需要依赖的表如何删除？

代码中找到表右键删除输入再重新选择自动解析



为什么自动解析的父节点名会报依赖的父节点输出（表）不存在？



这里报错并不是说表不存在，只是没有通过表找到产出表数据的节点挂上节点依赖。

出现上述情况有以下两种原因：

- 上游节点未提交，提交后可以再次尝试。
- 上游节点已经提交，但上游节点的输出名不是workshop_yanshi.tb_2。

说明

- 如果tb_2为同步任务产出的表，请在产出tb_2表数据的同步任务节点输出那手动以 `projectname.tablename` 的方式添加为节点输出，详情请参考文档[同周期调度依赖逻辑说明](#)。
- 如果tb_2为非调度节点每天更新的表，请在代码中找到表右键删除输入，再重新自动解析。

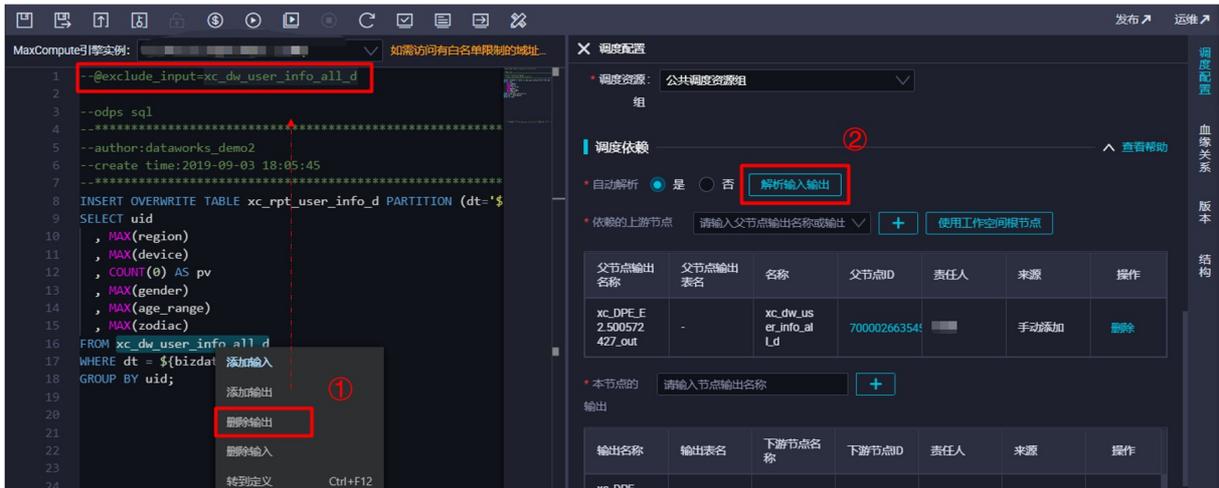
非调度节点每天更新的表请参考常见问题：[哪些场景下不支持设置调度依赖？](#)

为什么有些节点的本节点的输出中，下游节点名称、下游节点ID有这些数据，有些为空并且不能手动编辑。

节点依赖关系是下游节点去挂上游节点依赖。如果当前节点下游无子节点，则无内容。待本节点下游配置子节点后，便会自动解析出内容。

不需要依赖的依赖关系如何删除？

代码中找到表右键删除输入再重新选择自动解析



下游任务依赖上游任务的运行规则是什么？

在DataWorks调度系统中，节点设置上下游依赖关系是为了保障数据有效产出和获取。所以是否设置依赖关系可以从数据是否存在强相关性去判断。详情可以参考文档：[同周期调度依赖逻辑说明](#)

节点的输出名称的作用是什么？

节点的输出名称用于建立节点间的依赖关系。假设节点A的输出名称是ABC，而节点B将ABC作为它的输入，则节点A与节点B之间便建立了上下游关系。

一个节点可以有多个输出名称吗？

可以。节点输出相当于当前节点的定义，下游节点如果需要依赖当前节点，可以通过引用本节点的任何一个输出名称作为下游节点的父节点输出名称，即可与当前节点建立依赖关系。

多个节点往同一张表写数据，自动解析报错节点输出名相同且提交报错，可以有相同的输出名称吗？

不可以。节点输出名和节点、表一样，需要在租户级别唯一。这样自动解析可以根据唯一的输出找到唯一的节点挂上节点依赖关系，所以如果实际场景有多个节点产出数据至同一张表，那么需要确定下游节点自动解析这张表时应该挂哪个节点依赖（哪个节点最晚写入这张表数据，确保下游节点取数无问题即可），并且修改其他节点的输出，确保输出唯一。

使用自动解析依赖关系时，如何不解析到中间表？

在SQL代码中选中中间表名并右键单击删除输入或删除输出，再次执行自动解析输入输出即可。

业务流程最上游节点的父节点配置应该如何配置？

如果是业务流程的业务开始节点，您可以选择在业务流程内添加一个虚拟节点作为业务流程的开始节点。虚拟节点的上游可以设置为工作空间根节点。

关于虚拟节点的使用可以参考文档 [\(虚拟节点\)](#)

为什么在节点A搜索上游节点输出名时，搜索到了节点B不存在的输出名？

因为依赖关系解析是基于已经提交和发布的节点信息来进行搜索，如果节点B提交成功后，您又删除了节点B的输出名称且未提交至调度系统，则在节点A上仍然能搜到节点B已删除的输出名。

为什么节点删除报错当前节点存在子节点，下线失败，但是在调度配置界面看已经没有该节点的依赖？

当前节点在开发环境和生产环境都没有节点依赖当前节点时，当前节点才能被下线。

您可以分别进入开发环境运维中心和生产环境运维中心进行确认。

为什么运维中心有些依赖线是虚线？

虚线是设置了跨周期依赖，跨周期依赖介绍详情请参考文档：[场景2：依赖上一周期的结果时，如何配置调度依赖](#)

小时任务设置自依赖对自身及下游节点有什么影响？

- 对当前节点执行影响：上游节点上一周期实例运行成功后下一周期实例才会运行。
实时场景：假设小时任务0点开始每小时调度一次，1点的实例需要等待0点的实例执行成功，1点的实例才会执行。
- 对下游节点执行影响：假设下游节点为天调度任务，下游天节点依赖由直接依赖多个小时实例变为直接依赖上游某一个小时实例，小时实例由于存在自依赖，所以下游天任务实际上为间接依赖了上游多个小时实例。

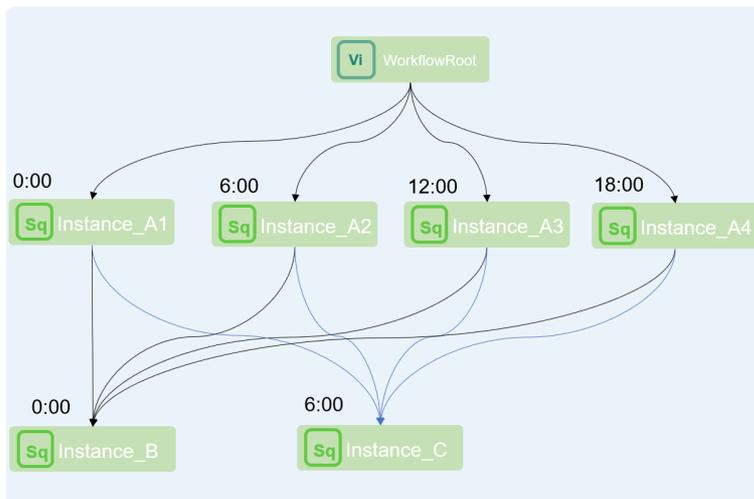
日调度任务依赖小时调度任务运行的不同场景，依赖关系如何配置？

- 场景一：日调度任务依赖小时调度任务当天的所有小时实例。

天调度任务直接依赖小时调度任务时，会依赖小时调度任务当天所有实例。

天任务直接依赖小时任务，小时任务不设置自依赖

小时调度节点A配置依赖上一周期本节点前



定时时间0:00的天调度节点B和定时时间6:00的天调度节点C依赖每6小时调度一次的小时节点A。

- A节点每日在0:00-23:59时间段内,每8小时调度一次
- 节点A每天生成4个实例,
 - 第一个实例定时时间0:00,
 - 第二个实例定时时间6:00,
 - 第三个实例定时时间12:00,
 - 第四个实例定时时间18:00

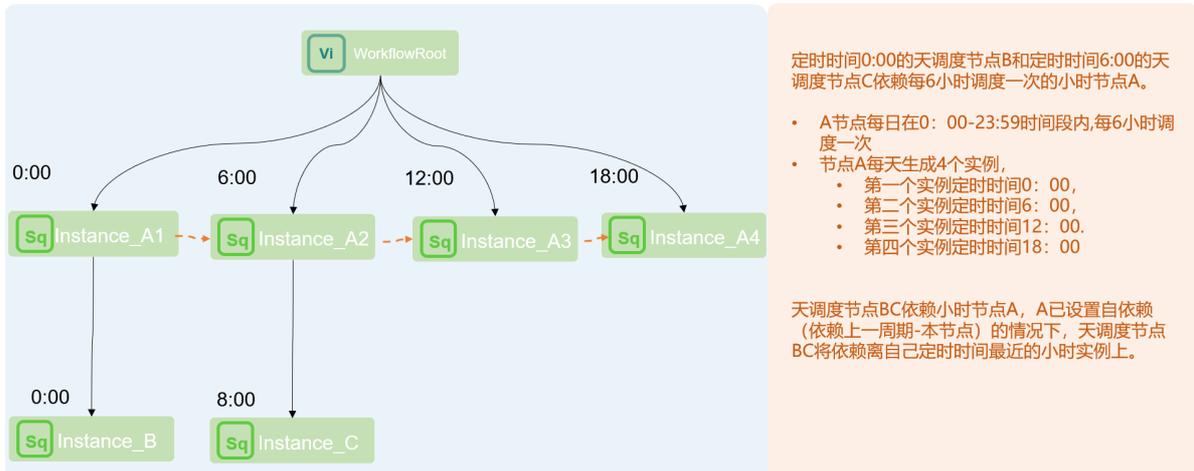
天调度节点BC直接依赖小时节点A (A不设置自依赖的情况下), 天调度节点BC将依赖小时节点A当天所有的小时实例。

- 场景二：日调度任务依赖小时调度任务当天具体某个实例。

- 小时调度任务配置：小时调度任务设置自依赖，即小时任务调度配置界面依赖的上一周期处选择本节点。
- 日调度任务配置：天调度任务直接依赖小时任务。即天调度任务的输入（依赖的上游节点）配置小时任务。

天任务依赖小时任务，小时任务设置自依赖

小时调度节点A配置依赖上一周期本节点后



- 场景三：日调度任务依赖小时调度任务昨天所有的小时实例。
 - 日调度任务调度配置跨周期依赖小时调度，即日调度任务调度配置界面选择依赖上一周期，选择自定义，填小时任务节点id。
 - 日调度任务调度配置去掉小时任务同周期依赖，即同周期依赖（依赖的上游节点）处，去掉小时任务同周期依赖。

说明 如果日调度任务对小时调度任务做过跨周期依赖，那需要检查是否同周期依赖已经去掉。否则日调度任务会同时依赖小时调度任务当天所有实例和昨天所有实例。

天任务依赖小时任务时，直接设置依赖关系，天任务几点会运行？

原理说明：当天任务直接依赖小时任务时：天任务会依赖小时任务当天所有实例，即等到小时任务当天最后一个实例执行成功，天任务才会执行。

实时场景：

- 假设小时任务0点开始，每小时调度一次。那么天任务需要等待24个小时任务全部运行完成才能开始运行天任务。
- 运维中心依赖关系查看：天任务右键查看父节点会显示依赖小时任务当天所有实例，即依赖24个小时实例。（依赖线：实线）

天任务依赖小时任务时，如何实现天任务依赖小时任务当天具体某个实例，而非依赖所有的小时实例？

原理说明：当天任务依赖小时任务当天具体某个小时实例时，小时任务只需要设置自依赖，天任务定时时间与具体某个小时实例定时时间保持一致即可。

实时场景：当天任务需要依赖小时任务当天定时时间12点的实例时

- 依赖关系设置：

- 上游小时任务配置：小时任务需要设置自依赖，即在调度配置，在时间属性区域，选中依赖上一周期 > 本节点。
- 下游天任务配置：天任务设置定时时间为12点。
- 运维中心依赖关系查看：
 - 天实例右键查看父节点会显示依赖小时任务当天定时时间为12:00的实例。（依赖线：实线）
 - 小时实例右键查看父节点会显示上游父节点为上一个小时实例，定时时间12:00的实例依赖定时时间11:00的实例（依赖线：虚线，小时任务设置了跨周期依赖，依赖项：本节点）

天任务依赖小时任务时，如何实现天任务依赖小时任务昨天所有的小时实例。而非当天所有小时实例？

原理说明：天任务依赖小时任务昨天所有小时实例，天任务需要跨周期依赖小时任务。

实时场景：当天任务需要依赖小时任务当天所有实例。

- 依赖关系设置：
 - 下游天任务配置：跨周期依赖小时任务，即在调度配置，时间属性区域，选择依赖上一周期 > 小时任务节点id。
 - 上游小时任务配置：无需配置
- 运维中心依赖关系查看：

下游天实例右键查看父节点会显示依赖小时任务昨天所有小时实例。（依赖线：虚线，天任务设置了跨周期依赖小时任务）

什么情况下配置依赖上一周期的依赖项为本节点？

业务场景：如果当前节点本周期任务运行需要依赖上一周期当前节点产出的节点数据时，当前节点可以设置自依赖。即当前节点上一周期实例执行完成后，当前节点下一周期的实例才去执行，避免当前实例取数时，上一周期实例还在执行（数据还未产出）。

- 本节点依赖自己上一周期产出的数据，如果您需要确认上一周期产出数据的时间，请在任务的调度配置 > 时间属性区域，配置依赖上一周期 > 本节点。
- 小时任务依赖天任务，当上游天任务运行完成时，已到下游小时任务的多个周期定时时间，导致小时任务多周期并发调度。此时您可以在任务的调度配置 > 时间属性区域，配置依赖上一周期 > 本节点。

下游节点同时依赖多个任务运行时，依赖关系如何配置？

如果下游节点同时配置上了依赖多个任务，是否需要依赖所有节点需要您进行业务判断，如果表数据有强相关依赖的，建议将所有节点都设置上依赖关系。节点是否需要设置依赖关系，请参考[为什么需要设置调度依赖](#)

例如，下游节点C同时依赖日调度任务B及当天的小时调度任务A，小时调度任务输出结果为表A，日调度任务输出结果为表B，下游节点C需要依赖表A及表B运行。

假设下游节点C中涉及查询表A及表B的相关数据，当您仅配置依赖的上游节点为小时调度任务A，而未配置依赖的上游节点为日调度任务B，则可能在上游节点的日调度任务B未执行完成时，下游节点C便开始执行节点任务，导致获取日调度任务的输出表B的数据失败，任务运行报错。因此，该示例中，您需要将日调度任务及小时调度任务均配置为下游节点C依赖的上游节点。

如果下游节点对上游表没有强依赖关系，即下游节点取任何时候上游表数据都没有问题（即使上游节点最新表数据也未产出），这种情况下，可以不设置节点依赖关系。

上游A是小时任务，下游B是日任务，任务每天在A任务全部执行完成后要汇总执行一次，小时任务运行到第二天，天任务是否会运行，参数是否会影响

日任务B直接依赖小时任务A当天的实例，日任务B会汇总当天的小时任务。如果小时任务完成最后一个实例的运行，已经到了第二天，则下游的日任务仍会执行。只是运行时间不同，调度参数替换不受影响。

节点A每天每小时整点执行一次，节点B每天运行一次。如何设置节点A每天第一次运行成功后，节点B便开始执行？

配置节点A时，需要勾选依赖上一周期，并选择本节点。节点B的定时时间配置为0点，则每天自动调度实例中，节点B的实例仅依赖节点A在0点生成的实例，即节点A的第一个实例。

有A、B、C三个任务，如何实现每小时执行一次A->B->C（A执行完了B再执行，B执行完了C再执行）的任务流程？

1. 依赖关系设置：将A、B、C的依赖关系设置为A的输出为B的输入，B的输出为C的输入。
2. 调度频率设置：由于调度配置为节点维度的配置，所以如果3个节点都需要小时调度，需要A、B、C的调度周期均设置为小时才可。

如何设置跨业务流程，同region跨项目依赖？

原理说明：上游节点的输出作为下游节点的输入形成节点依赖关系。

在需要设置依赖关系的节点输入处添加需要依赖的节点（跨项目，跨业务流程）的节点输出即可。

已配置失败重跑的任务在运行失败后未重跑，并报错Task Run Timed Out, Killed by System!!!

- 报错内容：当目标任务调度配置的时间属性中，重跑属性配置为运行成功或失败后皆可重跑或运行成功后不可重跑，运行失败后可以重跑，实际执行任务时，该任务在运行失败后未重跑，并产生 Task Run Timed Out, Killed by System!!! 报错。
- 可能原因：该任务调度配置的时间属性中，配置了超时时间。当任务的运行时长超过超时时间，任务自动终止运行。由于超时导致运行失败的任务不会触发任务重跑。
- 解决方案：当任务运行出现超时失败时，失败重跑机制不生效，您需要手动重启相关任务。

4. 运维中心

4.1. 问题汇总

补数据

- 补数据功能说明
- 为什么小时分钟任务补数据选择了并行但实际不生效？
- 为什么补数据选择业务时间后不运行，实例显示黄色，实例状态显示等待时间？
- 为什么补数据报错调起的节点运行时间不在所选业务时间范围内？
- 为什么有节点选择了补数据，但是没有生成补数据实例？

等待资源

- 为什么会出现等待资源？
- 为什么任务一直在等待gateway调度资源？
- 为什么数据集成任务一直显示wait？

为什么会空跑

- 什么是空跑？
- 为什么会空跑？
 - i. 场景一：周月实例空跑周期
 - ii. 场景二：实时生成的过期实例
 - iii. 场景三：设置成功实例
 - iv. 场景四：实例属性为空跑
 - v. 场景五：临时工作流中未被选择的实例

日调度任务空跑排查方法

任务未运行

- 任务运行条件
- 任务到定时时间，为什么还没运行？

任务实例状态异常

- 运维中心周期任务没有找到该任务
- 有周期任务，但是周期实例一个都没有
- 有周期任务，有其他周期实例，但是这个任务没有产生周期实例
- 有周期任务，有周期实例但未运行
- 任务运行必须满足的条件有哪些？

任务冻结与解冻

- 冻结周期任务和周期实例的影响
- 为什么做了冻结操作，任务还是执行？

任务运行成功没有数据

- 场景一：节点状态成功，有执行日志
- 场景二：节点状态成功，无执行日志

任务运行失败

- 重跑任务失败
 - 已配置失败重跑的任务在运行失败后未重跑，并报错Task Run Timed Out, Killed by System!!!
 - 任务运行失败
- MaxCompute类型节点
 - ODPS-0420095: Access Denied - Authorization Failed [4093], You have NO privilege to do the restricted operation on {acs:odps:*:projects/xxxx}. Access Mode is AllDenied.
 - ODPS-0420061: Invalid parameter in HTTP request - Fetched data is larger than the rendering limitation. Please try to reduce your limit size or column number
 - 多线程执行DataWorks节点任务时数据乱序，该如何处理？
- AnalyticDB for MySQL类型节点

使用公共资源组运行AnalyticDB for MySQL任务失败，该如何处理？
- 数据库类节点

MySQL节点报错：暂不支持MySQL节点驱动
- 调度配置常见问题
 - 任务运行失败
 - 任务运行失败
- 通用类型节点
 - For-each, do-while、pai等节点如何在运维中心查看日志
 - 分支节点运行报错：error in your condition run fail
 - FTP Check节点任务，运行报错：None Ftp connection info!!
 - FTP Check节点任务，运行报错：Connect Failed
 - FTP Check节点任务，运行报错：The current time has exceeded the end-check time point!
 - FTP Check节点任务，运行报错：File not Exists or exceeded the end-check time point!
- 资源组

报错：no available machine resources under the task resource group，该如何处理？

下线节点

- 如何下线节点
- 节点存在子节点，下线失败如何处理？
- 如何确认生产任务是否下线成功？
- 如何找回下线节点

孤立节点

- 什么是孤立节点

- [如何处理孤立节点？](#)

其他问题

- 资源组
 - [日志和实例保留多长时间？](#)
 - [为什么不能在调度资源组上进行大数据计算？](#)

- 其他

[如何设置任务优先级？](#)

4.2. 等待资源

本文为您介绍等待资源的相关问题。

- [为什么会出现等待资源？](#)
- [为什么任务一直在等待gateway调度资源？](#)
- [为什么数据集成任务一直显示wait？](#)

为什么会出现等待资源？

- 问题现象

调度将任务下发到对应执行引擎上，此过程可能会出现以下等待资源问题：

- 等待调度资源组。
- 数据集成任务等待数据集成资源。
- 计算任务等待引擎计算资源。

- 产生原因

资源组下可用资源是有限的，如果有任务长时间执行未释放资源必定会阻塞其他任务执行，关于任务下发机制，详情请参见：[DataWorks资源组概述](#)。

为什么任务一直在等待gateway调度资源？

- 问题现象

任务日志中显示等待gateway调度资源。

- 产生原因

当前资源组里的任务并发数达到上限。

- 解决措施

需要等待运行中的任务执行完成释放资源，或者对独享调度资源组进行扩容。您可以通过以下方式查看资源信息。

- 使用公共调度资源组，您可以在[运维中心](#)的[运维大屏](#)页面，查看公共资源组使用率。

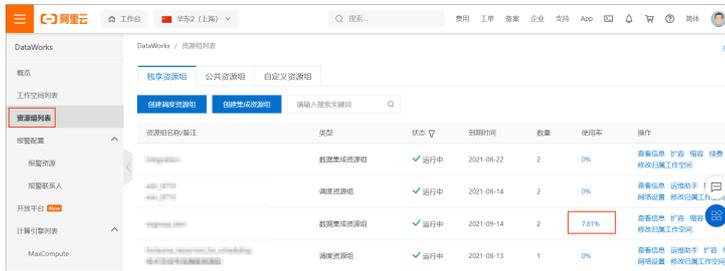
 **说明** 每天0点~9点是DataWorks的任务高峰期，公共调度资源紧张，可能出现任务等待资源的情况。您可以错峰运行调度任务或者在[DataWorks控制台](#)添加独享调度资源组缓解资源紧张情况。

- 使用独享调度资源组，您可以通过DataWorks控制台或者[运维中心](#)查看该资源组当前执行的任务列表及

资源水位。

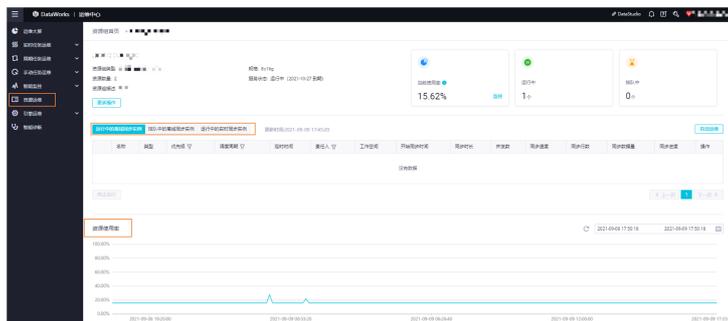
- 通过DataWorks控制台查看资源使用情况。

在DataWorks控制台的左侧导航栏，单击资源组列表，在资源组列表页面查看资源使用率，单击使用率查看资源组信息、资源使用率等。

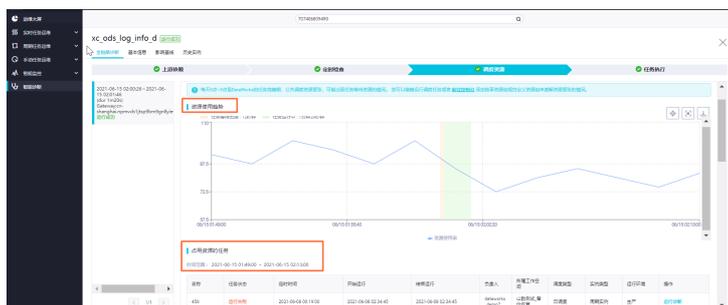


- 通过运维中心查看资源使用情况。

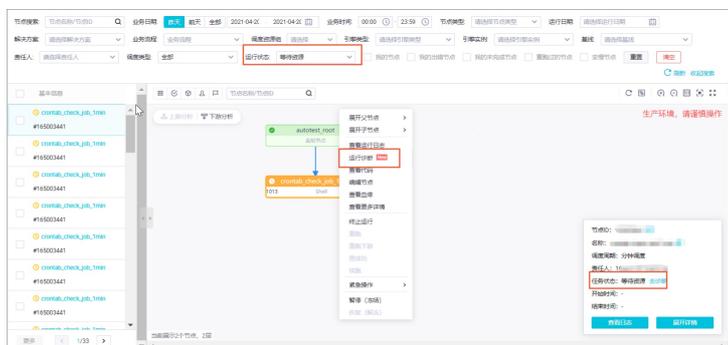
在运维中心页面的左侧导航栏，单击资源运维，查看资源的使用率等信息。



智能诊断可以查看当前任务执行、排队情况，以及当前任务等待资源时，哪些任务占用资源。



在等待资源的实例节点，右键选择运行诊断 > 调度资源，查看当前任务运行状态为等待资源时，仍在运行的任务。



说明 不同规格的独享调度资源组支持同时跑的最大实例数不同，详情请参考[独享调度资源组计费说明：包年包月](#)。

为什么数据集成任务一直显示wait?

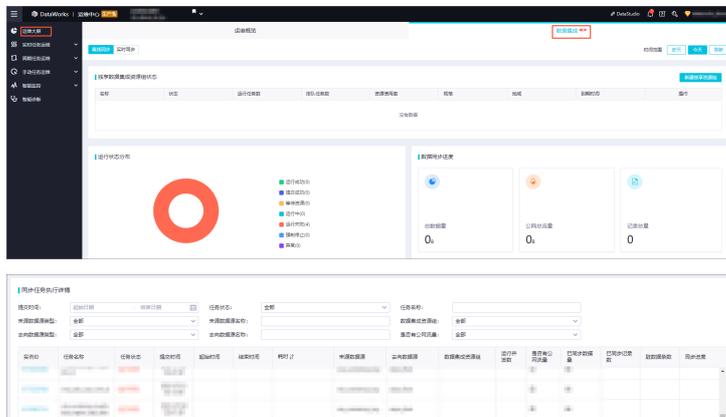
- 问题现象

数据集成任务日志中显示wait。
- 产生原因

当前数据集成资源组上执行的任务超过了机器支持的并发上限，任务在等待数据集成执行资源。
- 解决措施
 - 检查日志（任务配置）中concurrent参数值，如果并发任务设置过大，需根据机器支持的最大并发数与当前同时消耗的并发数配置任务并发。

- 如果单个任务并发设置超过了机器支持的最大并发数，请将这些并发任务全部停掉，修改任务并发数后再去执行。

在运维中心页面左侧导航栏，单击运维大屏，在运维大屏页面的数据集成页签，查看同步任务执行详情。

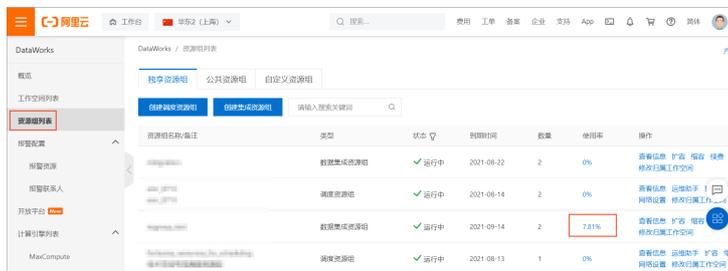


说明

- 任务并发配置需要结合**机器支持的最大并发数**，以及需要同时运行的任务数及并发和来分配单个任务并发数配置。
- 一个数据集成任务将占用一个调度资源，若任务长时间未运行成功，可能将阻塞其他任务运行。

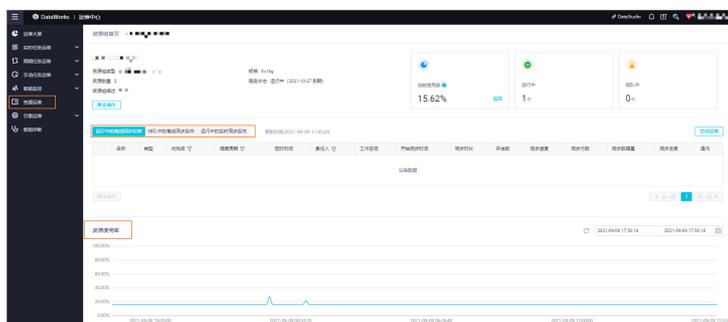
- 等待运行中的任务执行完成释放资源，或者对独享数据集成资源组进行扩容。您可以通过以下方式查看资源信息。
 - 通过Dat aWorks控制台查看资源使用情况。

在Dat aWorks控制台的左侧导航栏，单击资源组列表，在资源组列表页面查看资源使用率，单击使用率查看资源组信息、资源使用率等。



- 通过运维中心查看资源使用情况。

在运维中心页面的左侧导航栏，单击资源运维，查看资源的使用率和运行中的任务等信息。



说明 不同规格的独享数据集成资源组支持同时跑的最大实例数不同，详情请参考：[独享数据集成资源组计费说明：包年包月](#)。

4.3. 任务未运行

本文为您介绍任务未运行的相关问题。

- **任务运行条件**
- **任务到定时时间，为什么还没运行？**

任务运行条件

运维中心通过不同状态标识区分状态类型。

正常调度任务运行需要多个条件：调度资源、定时时间、上游节点运行状态、调度属性为正常调度（未被冻结）。

序号	状态类型	状态标识	运行流程图
----	------	------	-------

序号	状态类型	状态标识	运行流程图
1	运行成功状态		
2	未运行状态		
3	运行失败状态		
4	正在运行状态		
5	等待状态		
6	暂停/冻结状态		

- 紫色（冻结）：说明该实例被冻结了，任务将不会运行，并且阻塞下游，可以单击展开详情>操作日志，在操作记录中查看相关记录。
- 黄色（等待）：等待任务定时时间，单击右下角展开详情，在属性中查看相关信息。

说明

- 出现等待资源状态，说明当前项目下正在运行的任务达到上限，可以右键实例在运行诊断的调度资源部分查看任务等待资源时，哪些任务正在占用资源，详情请参见[等待资源](#)。
- 出现等待时间状态，说明当前实例定时时间还未到。

- 灰色（未运行）：需右键灰色实例，查看上游父节点的状态。建议使用[运行诊断](#)和[上游分析](#)。
- 蓝色（运行中）：说明实例正在运行，但任务若长时间未运行成功，详情请参见[等待资源](#)。

说明 上游没有出现以上任何一种状态，整个业务流程均为灰色，说明上游依赖关系变动导致业务流程都被孤立。详情请参见[孤立节点](#)。

任务到定时时间，为什么还没运行？

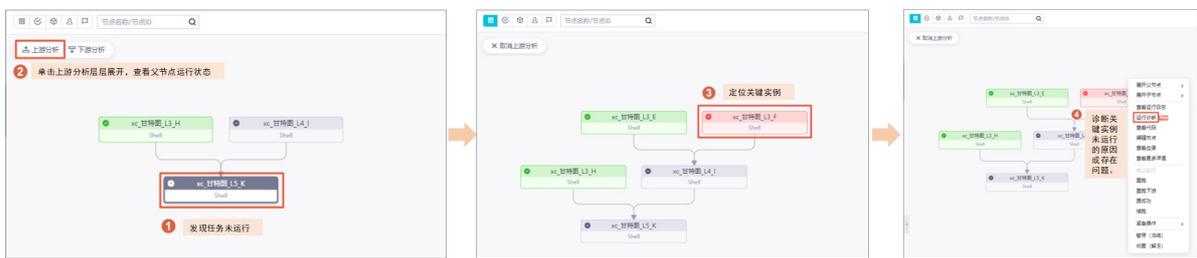
● 问题现象

在周期实例运行时，实例定时时间和实例开始运行时间有时并不完全一致，或任务到了定时时间还没运行。



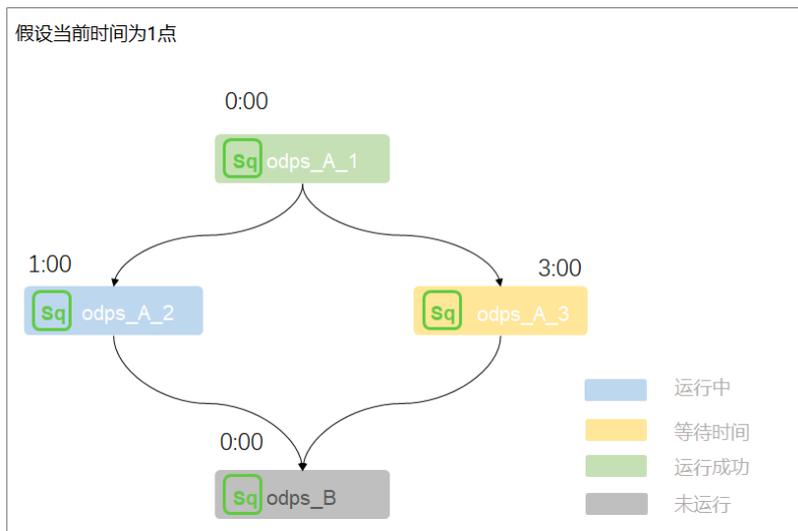
● 问题排查

您可以先通过上游分析功能定位阻塞当前任务运行的关键实例，再通过智能诊断分析当前任务阻塞原因，快速定位问题。



DataWorks上任务未运行通常有4种情况，通过以下示例进行分析说明。以下图示为三层依赖，实际场景可能不止三层，但逻辑一样。

- 场景一：该节点所有依赖的父节点还未运行成功（未运行节点状态：灰色）。



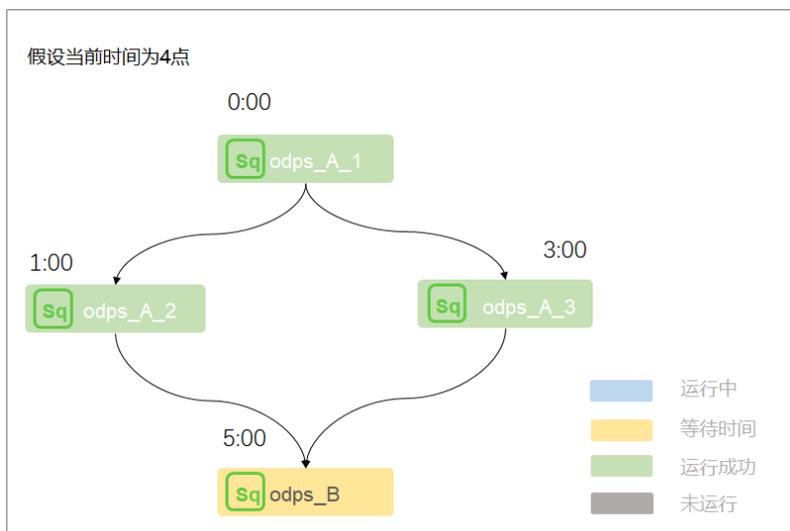
场景：假设当前时间为1点，A2任务定时时间为1点，A3定时时间为3点，B定时时间为0点，B任务依赖A2、A3节点。

分析：下游B节点定时时间为0点，此时已经到了任务定时时间，但由于上游父节点定时时间未到，所以下游节点B需要等到上游父节点A2，A3定时时间到了并且任务执行成功后，B任务才会执行。如果上游A2、A3任务有一个失败，那下游B任务将不会调度起来。所以B任务当前的实例状态为未运行。

结论：当该节点所有的依赖的父节点都已经运行完成时，当前节点才会运行。

- ④ **说明** 若上游一直处于运行中的状态，您可以参考以下解决方案。
- 非离线同步任务一直处于运行中的状态，您可以工单咨询对应引擎同学查看具体原因。
 - 离线同步任务一直处于运行中的状态，可能处于长时间等待任务执行资源，或实际运行过程中部分逻辑处理较慢，详情请参见[如何排查离线同步任务运行时间长的问题](#)。

- 任务定时时间未到（等待时间节点状态：黄色）。

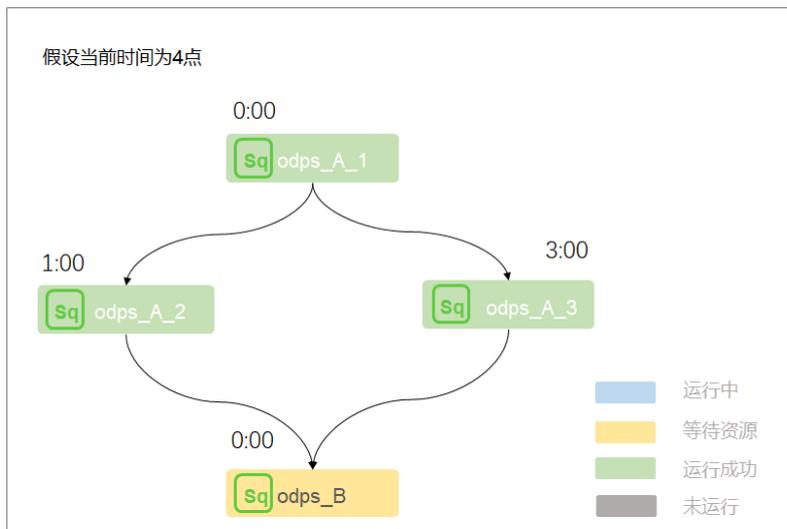


场景：假设当前时间为4点，A2任务定时时间为1点，A3定时时间为3点，B定时时间为5点，B任务依赖A2、A3节点。

分析：下游B节点定时时间为5点，父节点A2、A3都已经执行成功，但由于B任务还未到定时时间，所以B任务当前的实例状态为等待时间。

结论：任务定时时间已到，当前节点才会运行。

- 项目下没有足够的调度资源可让任务运行（等待资源节点状态：黄色）。



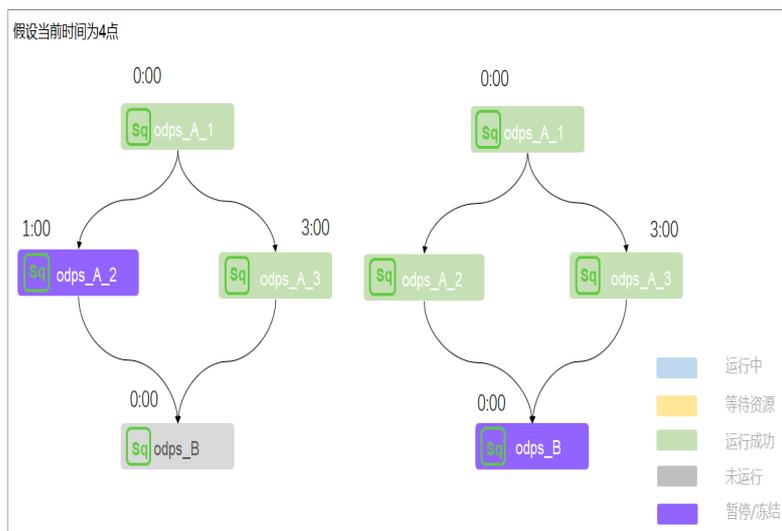
场景：假设当前时间为4点，A2任务定时时间为1点，A3定时时间为3点，B定时时间为0点，B任务依赖A2、A3节点。

分析：下游B节点定时时间为0点，父节点A2、A3都已经执行成功，但由于当前B节点使用的调度资源组下资源不够，导致任务没有调度资源，所以B任务当前的实例状态为等待资源。

结论：工作空间调度资源充足时，任务才会运行。（等待资源状态时，日志中显示当前项目下任务并发达到上限，正在等待gateway资源）

④ 说明 如果任务运行在独享调度资源组，您可以在DataWorks控制台查看独享调度资源组，通过资源组使用率查看资源组下正在执行的任务及资源组水位，或者右键实例使用智能诊断功能，查看当前任务在等待资源时，哪些任务正在占用资源，详情请参见，等待资源。

- 任务被冻结（任务被冻结节点状态：紫色）。



场景：假设当前时间为4点，A2任务定时时间为1点，节点状态为暂停调度，A3定时时间为3点，B定时时间为0点，B任务依赖A2、A3节点。

分析：通过第一条可知，当前节点执行需要父节点全都执行成功，由于左侧图中A2节点是冻结状态，所以下游B节点不会执行。右侧图，父节点都执行成功，但由于冻结节点不会调度，所以右侧图B节点不会执行。

结论：父节点被冻结或当前节点被冻结，任务将不会运行。

4.4. 任务实例状态异常

本文为您介绍实例异常排查的相关问题。

- 运维中心周期任务没有找到该任务
- 有周期任务，但是周期实例一个都没有
- 有周期任务，有其他周期实例，但是这个任务没有产生周期实例
- 有周期任务，有周期实例但未运行
- 任务运行必须满足的条件有哪些？

运维中心周期任务没有找到该任务

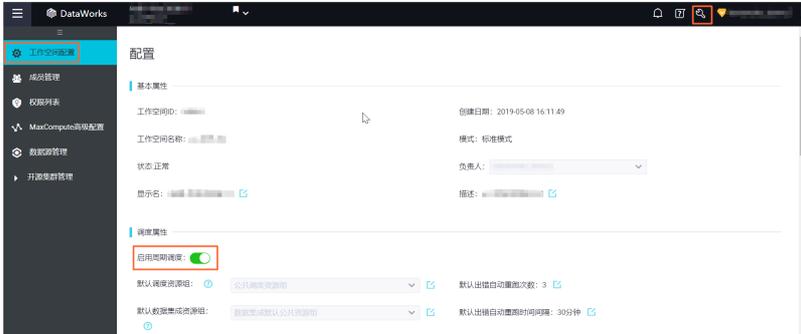
- 问题现象
运维中心周期任务没有找到目标任务。
- 产生原因
节点尚未发布到生产环境。
- 解决措施
请检查下是否发布失败。

有周期任务，但是周期实例一个都没有

- 问题现象
有周期任务，但是周期实例一个都没有。

- 产生原因
工作空间可能没有开启调度。

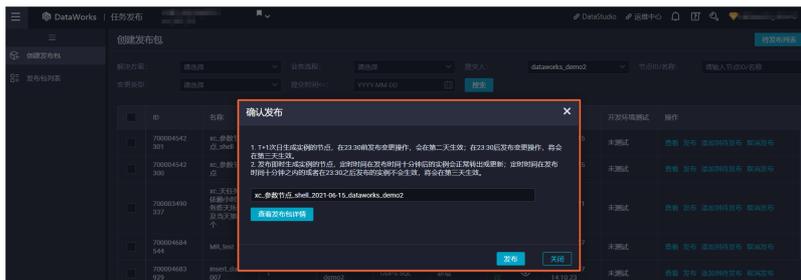
- 解决措施
按照下图所示，在工作空间配置页面，检查工作空间是否开启调度。



说明 如有需要请协调主账号开启工作空间调度。

有周期任务，有其他周期实例，但是这个任务没有产生周期实例

- 问题现象
有周期任务，有其他周期实例，但是这个任务没有产生周期实例。
- 产生原因
 - 任务在23:30之后发布。
 - 目标节点没有挂在其他父节点下，成了孤立节点。
 - 目标节点及上游节点不在有效期内。
- 解决措施
 - 检查任务是否在23:30之后发布。



- 右键单击周期任务DAG图，检查该节点是否没有挂在其他父节点下，成了孤立节点，收到孤立节点报警请第一时间处理。

说明 节点依赖关系变动，导致节点没有上游依赖。请重新提交上下游设置节点依赖关系。

- 检查本节点及上游节点是否在有效期内。过期任务不产生周期实例，请在数据开发调度配置界面修改生效时间。



有周期任务，有周期实例但未运行

- 问题现象

有周期任务，有周期实例但未运行。

- 产生原因

父节点未执行成功，导致下游未执行。

- 解决措施

右键灰色实例DAG 展开多层父节点，检查任务是否有正在运行、运行中、运行失败、等待、冻结状态。

? 说明 更多任务未运行的排查与处理请参见[任务未运行](#)。

父节点状态

- 实例状态：紫色（冻结）

说明该实例被冻结了，任务将不会运行，并且阻塞下游，可以在实例详情，操作记录中查看相关记录。
- 实例状态：黄色（等待）

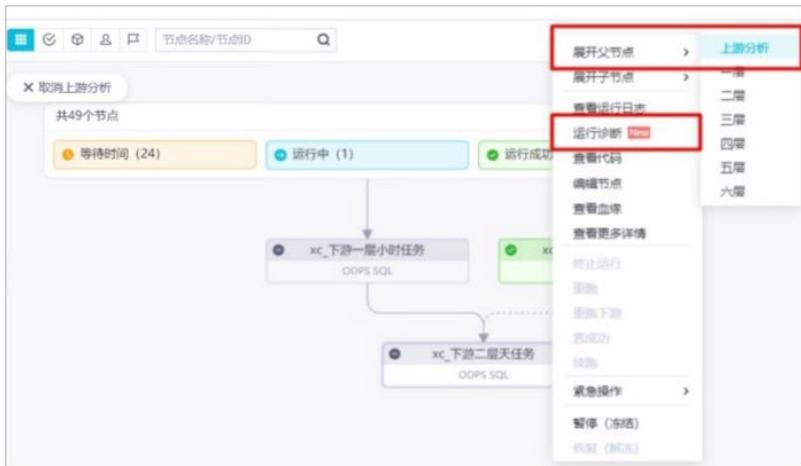
等待任务定时时间，点开右下角查看实例详情，看任务定时时间。

等待调度资源（日志中会显示），若出现此类等待说明当前项目下正在运行的任务达到上限，可以右键实例在运行诊断的调度资源部分查看任务等待资源时，哪些任务正在占用资源。
- 实例状态：灰色（待运行）

请继续右键灰色实例，查看上游父节点，看其上游状态。

如果都是灰色，请根据第二条检查是否没有挂在其他父节点下，孤立节点不运行。

建议您使用智能诊断和上游分析。



说明

- 正常调度任务运行需要有多个条件：调度资源、定时时间、上游节点运行状态、正常调度（未被冻结）。
- 上游没有出现以上任何一种状态，整个业务流程均为灰色，说明上游依赖关系变动导致业务流程都被孤立。

任务运行必须满足的条件有哪些？

任务运行必须满足以下条件：

- 任务定时时间已到（等待时间的节点状态：黄色）。
- 该节点所有依赖的父节点都已经运行完成（运行成功节点状态：绿色，运维中心DAG图展开父节点）。
- 项目下有足够的资源可运行任务（等待资源节点状态：黄色，日志中显示当前项目下任务并发达到上限，正在等待gateway资源）。
- 任务正常调度（任务被冻结节点状态：紫色）。

4.5. 为什么会空跑

本文为您介绍空跑的相关问题。

- 什么是空跑？
- 为什么会空跑？
 - 场景一：周月实例空跑周期
 - 场景二：实时生成的过期实例
 - 场景三：设置成功实例
 - 场景四：实例属性为空跑
 - 场景五：临时工作流中未被选择的实例
- 日调度任务空跑排查方法

什么是空跑？

空跑的实例任务状态显示成功（正常调度），运行日志为空，无执行时长，不会实际跑数据。

为什么会空跑？

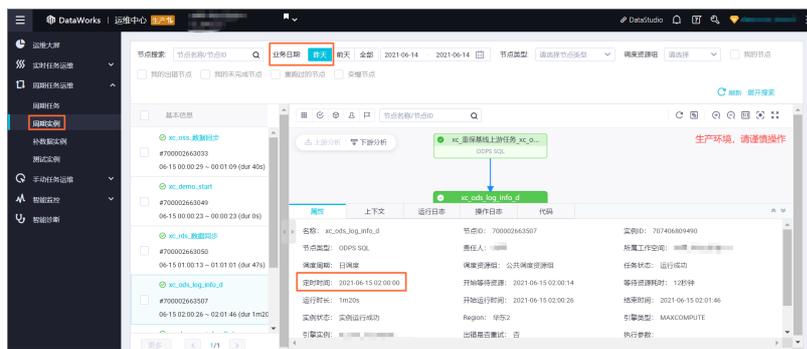
场景一：周月实例空跑周期

任务定时时间为定时每周几或者“每月定时几号”跑的任务，任务只有在指定调度时间才会真实跑数据，在非指定时间内，非调度时间内其他时间会生成空跑实例，不会实际执行任务，该实例状态为周月实例空跑周期。



如果周期实例需要真实跑数据，任务定时时间需要设置为发布时间点十分钟后的未来时间。或者通过补数据补业务时间昨天数据，明天周期实例会正常执行。

周月调度任务补数据空跑解决方法：您可以按下图观察一下当天周期实例的定时时间和业务时间，如果任务为月调度任务定时每月1号调度，那补数据需要选择业务时间为月末，周任务同理。



场景二：实时生成的过期实例

当使用发布后及时生成实例方式生成周期实例时，只有任务定时时间在任务发布时间点十分钟后的实例会真实跑数据，发布时间点十分钟前的实例会生成一个空跑实例，不会真实跑数据，该实例状态为实时生成的过期实例。



说明 具体详情请参考[实施转实例](#)。

场景三：设置成功实例

当实例状态为失败时，若触发置成功，则该实例状态会成功，不会真实跑数据，且不阻塞下游任务执行，该实例状态为设置成功实例。



场景四：实例属性为空跑

检查数据开发调度配置界面该节点时间属性是否设置为空跑调度，该实例状态为实例属性为空跑。



场景五：临时工作流中未被选择的实例

当补数据时，假设C节点依赖B节点，B节点依赖A节点，如果补数据选择补AC节点，则该B节点实例状态为临时工作流中未被选择的实例。



日调度任务空跑排查方法

如果是日调度任务请检查该节点的调度配置界面是不是将节点配置为空跑。

? 说明 T+1：业务时间为今天的数据明天才会去跑。

4.6. 任务冻结与解冻

本文介绍冻结解冻的相关问题。

- 冻结周期任务和周期实例的影响
- 冻结周期任务对补数据、测试操作产生的实例影响
- 冻结的实例解冻后如何重新运行？
- 为什么做了冻结操作，任务还是执行？
- 如何查看哪些用户对任务做过哪些操作？

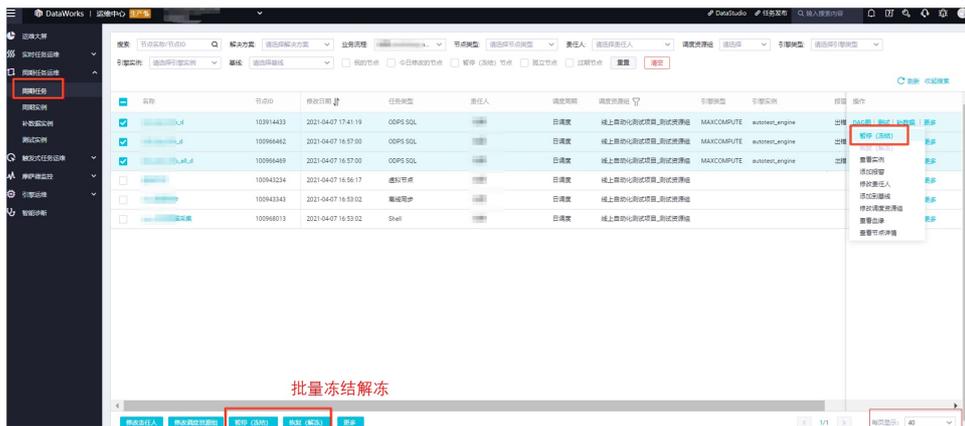
冻结周期任务和周期实例的影响

● 冻结与解冻周期任务

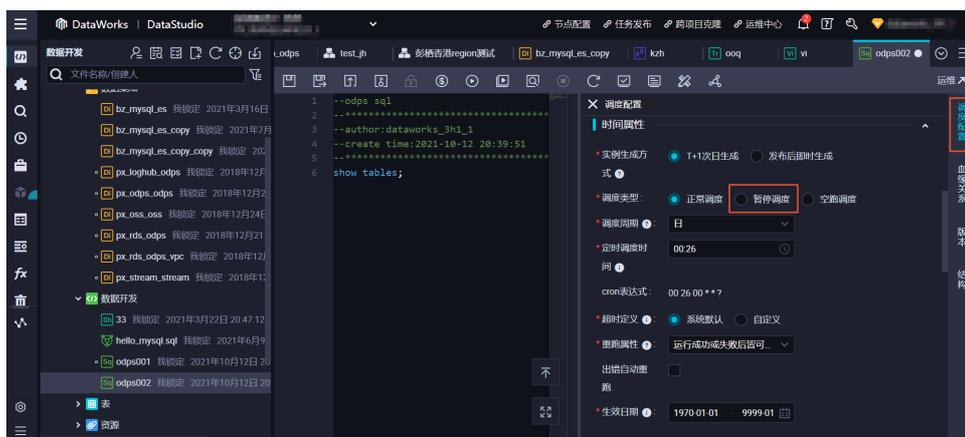
系统每晚会根据周期任务生成第二天要运行的周期实例，冻结周期任务会导致之后生成的周期实例处于冻结状态，并导致依赖该周期实例的下游实例无法运行。

❓ 说明 冻结周期任务当天，已经生成的周期实例会正常调度。冻结操作第二天，周期实例会冻结，并且会阻塞下游所有节点执行。

○ 在运维中心解冻周期任务。



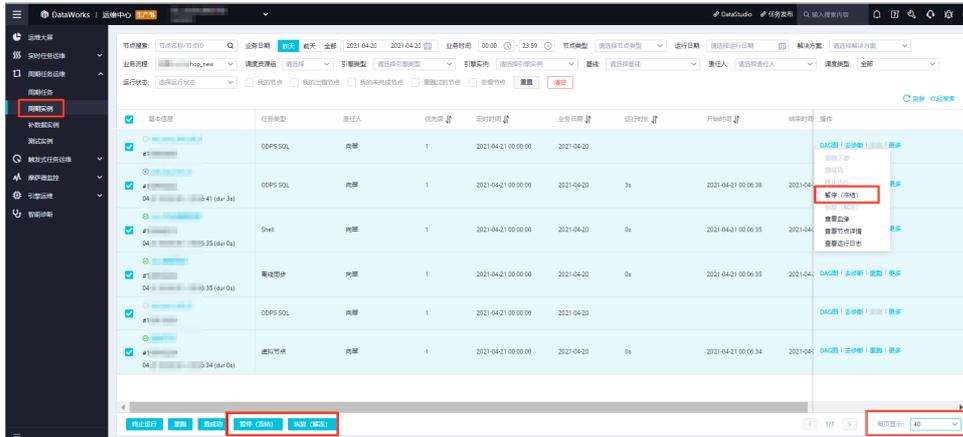
○ 在数据开发解冻周期任务。



❓ 说明 在数据开发（DataStudio）调度配置的调度类型里选择暂停调度或者正常调度，需重新提交发布，才会冻结或解冻生产调度的周期任务。

● 冻结与解冻周期实例

冻结周期实例只作用于该实例，如果只解冻周期实例，周期任务仍然是冻结状态，那么第二天的周期实例仍然是冻结状态。



注意 冻结解冻均人为操作。如有疑问，可查看相应操作记录。

冻结周期任务对补数据、测试操作产生的实例影响

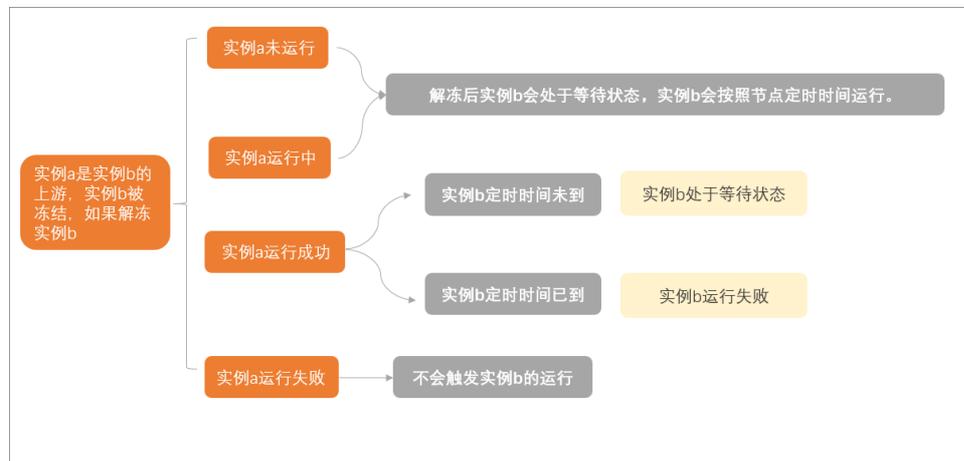
补数据与测试操作都是根据当前的周期任务生成的实例快照，如果周期任务是冻结状态，那么生成的补数据实例和测试实例也会是冻结状态。

冻结的实例解冻后如何重新运行？

冻结的实例解冻后，会根据调度参数配置的定时时间及上游依赖的父节点运行状态确定是否触发当前实例的运行。

假设实例a是实例b的上游，实例b被冻结，如果解冻实例b：

- 场景一：实例a未运行，那么解冻后实例b会处于等待状态，实例b会按照节点定时时间运行。
- 场景二：实例a运行中，那么解冻后实例b会处于等待状态，实例b会按照节点定时时间运行。
- 场景三：实例a运行成功。
 - 如果实例b的定时时间未到，那么实例b会处于等待状态。
 - 如果实例b定时时间已到，那么实例b会运行失败，如果您需要运行实例b，请单击实例b操作列的重跑，重新运行当前实例。实例运行成功后，会根据当前实例配置的依赖关系触发下游实例的正常调度。
- 场景四：实例a运行失败则不会触发实例b的运行。实例a运行失败请参见任务未运行处理。



为什么做了冻结操作，任务还是执行？

- 问题现象

做了冻结操作，任务还是执行。

- 产生原因及解决措施

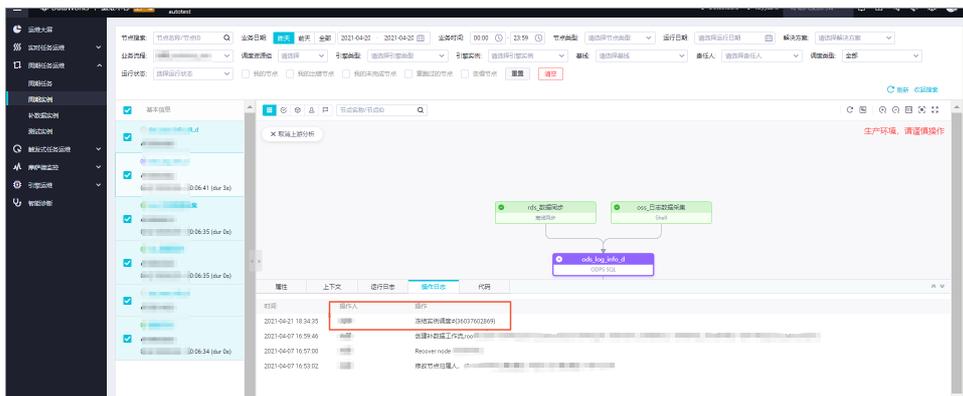
平台每晚会根据周期任务，生成第二天要跑的周期实例并自动调度。您可以手动针对当前周期任务触发补数据和测试操作，并根据当前实例快照信息生成补数据实例与测试实例，根据以上说明的周期任务与周期实例，测试实例和补数据实例的关系，您需要先查看以下几点。

- 请确认您冻结的任务是周期任务。
- 周期任务的冻结操作对已经生成的实例无效。
 - 对冻结操作当天的周期实例不生效（周期实例为前一天晚上批量生成）。
 - 对冻结周期任务操作前触发的补数据实例与测试实例不生效。

如何查看哪些用户对任务做过哪些操作？

- 基于实例或任务操作日志查看：

您可以在运维中心的周期任务或者周期实例面板查看操作日志。



- 基于节点版本对比查看：

若通过发布流程发布的节点变更操作，在实例或任务操作日志无详细记录，您在通过DataStudio界面，通过节点版本与生产环境版本对比，获取版本变更的详细情况。详情请参见：[版本](#)。

- 基于运维中心操作历史查看：

运维与管理员可在操作历史界面查看空间下用户在生产运维中心页面进行的相关操作。详情请参见：[操作历史](#)。

4.7. 补数据

本文为您介绍补数据的相关问题。

- [补数据功能说明](#)
- [为什么小时分钟任务补数据选择了并行但实际不生效？](#)
- [为什么补数据选择业务时间后不运行，实例显示黄色，实例状态显示等待时间？](#)
- [为什么补数据选择业务时间昨天和今天，也会出现等待时间的情况？](#)
- [为什么补数据选择补0点~1点生成了多个实例？](#)
- [为什么进行大量补数据操作时，会出现等待资源情况（实例显示黄色，实例状态为等待资源）？](#)

- 为什么补数据报错调起的节点运行时间不在所选业务时间范围内？
- 为什么有节点选择了补数据，但是没有生成补数据实例？

补数据功能说明

补数据支持补历史一段时间区间的数据或者需要补未来一段时间的数据时，可以选择补数据功能。节点使用的调度参数会根据补数据选择的业务时间自动替换为对应的值。将MySQL增量数据写入MaxCompute对应的分区中的示例如下。



为什么小时分钟任务补数据选择了并行但实际不生效？

● 问题现象

小时分钟任务补数据选择了并行但实际不生效。

● 产生原因

补数据并行控制是，补一段以天为维度的业务日期区间的数据以及几天的实例是否同时执行。此功能不控制小时、分钟任务当天的所有实例是否并发执行，当天小时分钟任务的实例是否并发执行与您小时分钟任务是否设置了自依赖有关。自依赖使用说明，请参见[场景2：依赖上一周期的结果时，如何配置调度依赖](#)。

● 解决措施

- 选择不并行，一个补数据实例下的多个业务日期串行执行，即上一个业务日期的补数据实例执行完，下一个业务日期的补数据实例才会执行。
- 选择并行，您可以设置同时使用2组、3组、4组或5组等多个补数据实例进行补数据，即多个补数据实例下有多个业务日期并行执行。

实时场景：假设小时、分钟节点选择补一个星期数据。

- 如果小时、分钟节点设置了自依赖，那么小时、分钟节点每天的实例会一个一个执行。
- 如果小时节点任务没有设置自依赖，那么小时、分钟节点每天的实例一块执行。

为什么补数据选择业务时间后不运行，实例显示黄色，实例状态显示等待时间？

● 问题现象

补数据选择业务时间后不运行，实例显示黄色，实例状态显示等待时间。

● 产生原因

如果您补数据选择的业务日期实例，其定时时间为未来时间时，会出现等待时间的情况。

● 解决措施

您可以选择立即执行未来时间的数据。



说明

- 如果您补数据选择的业务时间（实例定时时间）为未来时间，当不勾选此选框时，实例会出现等待时间的情况（实例显示黄色，实例状态为等待时间）。
- 如果您补数据选的业务时间（实例定时时间）为未来时间，当勾选此选框时，实例会马上执行。

为什么补数据选择业务时间昨天和今天，也会出现等待时间的情况？

● 问题现象

补数据选择业务时间昨天和今天，会出现等待时间的情况。

● 产生原因

从平台维度来看，业务时间昨天的数据，今天调度运行。补数据补对应业务时间的数据相当于重新执行了一次对应业务时间的周期实例。

如果您需要过滤今天的周期实例，您需要在周期实例面板选择过滤业务时间昨天的实例，实例的定时时间为今天。



为什么补数据选择补0点~1点生成了多个实例？

● 问题现象

补数据选择补0点~1点生成了多个实例。

● 产生原因

补数据生成多少个实例与您节点配置的定时时间有关。

- 假设小时任务设置0: 00开始至23: 59分每小时调度一次。如果选择补数据区间为00: 00~1: 00, 那么将产生两个实例, 实例定时时间分别为0: 00, 1: 00。
- 假设分钟任务设置0: 00开始至23: 59分每30分钟调度一次。如果选择补数据区间为00: 00~1: 00, 那么将产生三个实例, 实例定时时间分别为0: 00, 00: 30, 1: 00。

为什么进行大量补数据操作时, 会出现等待资源情况 (实例显示黄色, 实例状态为等待资源)?

● 问题现象

进行大量补数据操作时, 出现等待资源情况 (实例显示黄色, 实例状态为等待资源)。

● 产生原因

资源组可执行任务最大并发数有上限, 如果当前任务执行并发数超过资源组所支持的最大并发数时, 将会出现等待资源的情况。

🔗 说明 等待资源排查, 详情请参见[等待资源](#)。

为什么补数据报错调起的节点运行时间不在所选业务时间范围内?

● 问题现象

补数据报错调起的节点运行时间不在所选业务时间范围内。

● 产生原因

小时分钟任务需要选择具体的调度时间区间才可以正常生成补数据实例并执行。



为什么有节点选择了补数据, 但是没有生成补数据实例?

● 问题现象

某节点选择了补数据, 但是没有生成补数据实例。

● 产生原因

不在生效日期范围内的节点不生成实例，请检查需要补数据的节点生效时间是否在有效期内。



4.8. 任务运行成功没有数据

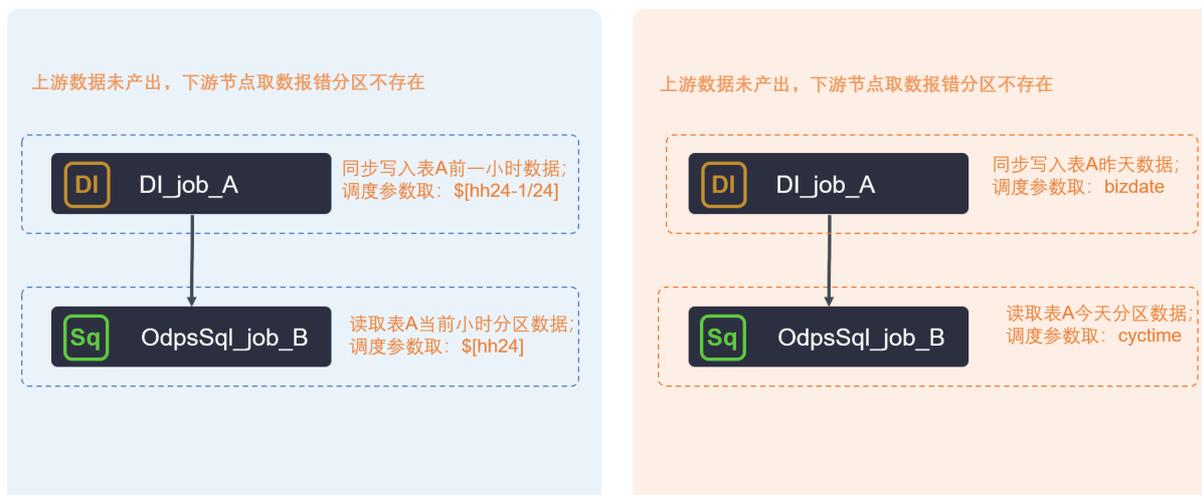
本文汇总介绍任务执行成功没有数据的相关问题。

- 场景一：节点状态成功，有执行日志
- 场景二：节点状态成功，无执行日志

场景一：节点状态成功，有执行日志

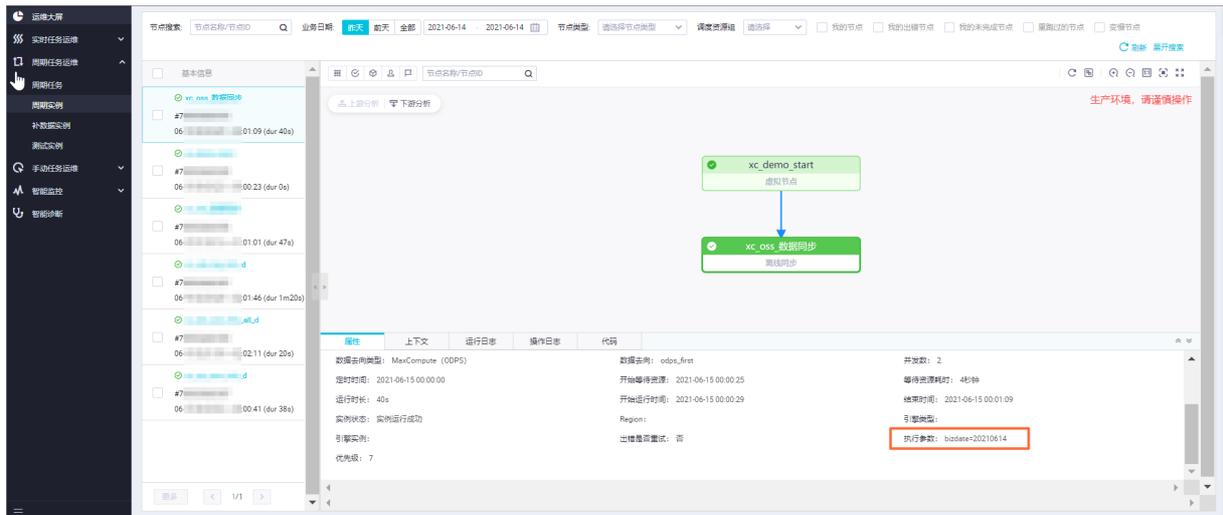
节点运行完成说明节点里面的逻辑完成了，但是周期实例自动调度时，节点没有取到上游数据或者报错不存在在该表分区。手动重跑实例可以正常取到数据。

1. 没有依赖产出该上游表数据的节点任务。
2. 有依赖产出表的节点，但是依赖的节点产生的表分区不是想要的，那就说明依赖的周期不对，可以在周期实例参数配置及日志详情里面查看一天的实例上下游参数替换情况。



请重新设置节点依赖关系

查看上下游节点参数配置



场景二：节点状态成功，无执行日志

节点空跑，请检查属性界面节点状态。

4.9. 任务运行失败

本文汇总介绍任务运行失败的相关问题。

- 重跑任务失败
 - 已配置失败重跑的任务在运行失败后未重跑，并报错Task Run Timed Out, Killed by System!!!
- MaxCompute类型节点
 - ODPS-0420095: Access Denied - Authorization Failed [4093], You have NO privilege to do the restricted operation on {acs:odps:*:projects/xxxx}. Access Mode is AllDenied.
 - ODPS-0420061: Invalid parameter in HTTP request - Fetched data is larger than the rendering limitation. Please try to reduce your limit size or column number
 - 多线程执行DataWorks节点任务时数据乱序，该如何处理？
- AnalyticDB for MySQL类型节点
 - 使用公共资源组运行AnalyticDB for MySQL任务失败，该如何处理？
- 通用类型节点
 - For-each, do-while、pai等节点如何在运维中心查看日志
 - 分支节点运行报错：error in your condition run fail
 - FTP Check节点任务，运行报错：None Ftp connection info!!
 - FTP Check节点任务，运行报错：Connect Failed
 - FTP Check节点任务，运行报错：The current time has exceeded the end-check time point!
 - FTP Check节点任务，运行报错：File not Exists or exceeded the end-check time point!
- 资源组
 - 报错：no available machine resources under the task resource group，该如何处理？

已配置失败重跑的任务在运行失败后未重跑，并报错 `Task Run Timed Out, Killed by System!!!`

• 报错内容：

当目标任务调度配置的时间属性中，重跑属性配置为运行成功或失败后皆可重跑或运行成功后不可重跑，运行失败后可以重跑，实际执行任务时，该任务在运行失败后未重跑，并产生 `Task Run Timed Out, Killed by System!!!` 报错。

• 可能原因：

该任务调度配置的时间属性中，配置了超时时间。当任务的运行时长超过超时时间，任务自动终止运行。由于超时导致运行失败的任务不会触发任务重跑。

• 解决方案：

当任务运行出现超时失败时，失败重跑机制不生效，您需要手动重启相关任务。

ODPS-0420095: Access Denied - Authorization Failed [4093], You have NO privilege to do the restricted operation on {acs:odps:*:projects/xxxx}. Access Mode is AllDenied.

1. 先确认odps引擎有没有欠费。
2. 小扳手 > MaxCompute高级配置里面，下面的权限有没有打开。



ODPS-0420061: Invalid parameter in HTTP request - Fetched data is larger than the rendering limitation. Please try to reduce your limit size or column number

查询语句加一下limit，如果要查看更多可以将数据导出，超过10000条请使用tunnel导出。

多线程执行DataWorks节点任务时数据乱序，该如何处理？

MaxCompute表是无序读取的，如果您未进行自定义设置，查询结果也是无序的。

MaxCompute同步数据是无序存储的。如果您对数据的顺序有要求，请在数据同步成功后，对数据进行排序。例如，在SQL语句中，需要添加order by xx limit n对数据进行排序。

使用公共资源组运行AnalyticDB for MySQL任务失败，该如何处理？

您需要开通独享调度资源组并连通VPC网络，再运行AnalyticDB for MySQL任务。详情请参见[数据源测试连通性](#)

MySQL节点报错：暂不支持MySQL节点驱动

- 问题描述添加MySQL数据源时，选择了非连接串模式创建的数据源，导致运行任务时失败，报错信息为 `sql execute failed! 暂不支持的jdbc驱动`。
- 问题原因出现上述报错通常都是选择了非连接串模式创建的MySQL数据源导致。
- 解决方案重新选择使用连接串模式创建的数据源。您可以参考[配置MySQL数据源](#)进入数据源管理页面，单击目标数据源操作列的编辑，在数据源编辑页面查看创建数据源时所使用的模式。

For-each, do-while、pai等节点如何在运维中心查看日志

找到实例右键查看内部节点

分支节点运行报错：`error in your condition run fail`

- 分支条件需要符合python语法的条件语句。
- 如果上游赋值节点的输出是字符串，分支节点引用变量时记得给变量加引号。

FTP Check节点任务，运行报错：`None Ftp connection info!!`

- 报错内容：FTP Check节点检测FTP数据源中Done文件是否存在时，任务运行失败，错误提示为 `None Ftp connection info!!`。
- 可能原因：FTP数据源配置有误，FTP Check节点未获取到FTP数据源的信息。
- 解决方案：您可以进入数据源管理页面，检查所使用的数据源配置是否正确。进入页面的操作，详情请参见[数据源管理](#)。如果没有可用的数据源，则需要新建FTP数据源，详情请参见[配置FTP数据源](#)。

FTP Check节点任务，运行报错：`Connect Failed`

- 报错内容：FTP Check节点检测FTP数据源中Done文件是否存在时，任务运行失败，错误提示为 `Connect Failed`。
- 可能原因：FTP数据源连接FTP服务端失败。
- 解决方案：您可以使用telnet IP 端口命令，检查FTP服务端运行是否正常。IP和端口为FTP数据源的IP和端口信息，您可以登录[DataWorks控制台](#)，进入数据源管理页面查看。进入数据源管理页面的操作，详情请参见[数据源管理](#)。

FTP Check节点任务，运行报错：`The current time has exceeded the end-check time point!`

- 报错内容：FTP Check节点检测FTP数据源中Done文件是否存在时，任务运行失败，错误提示为 `The current time has exceeded the end-check time point!`。
- 可能原因：表示当前检测时间已经超过检查FTP数据源中Done文件是否存在的Check停止时间，任务直接失败。
- 解决方案：您可以根据业务需求，为FTP Check节点任务重新配置一个合理的Check停止时间，详情请参见[配置检测策略](#)。

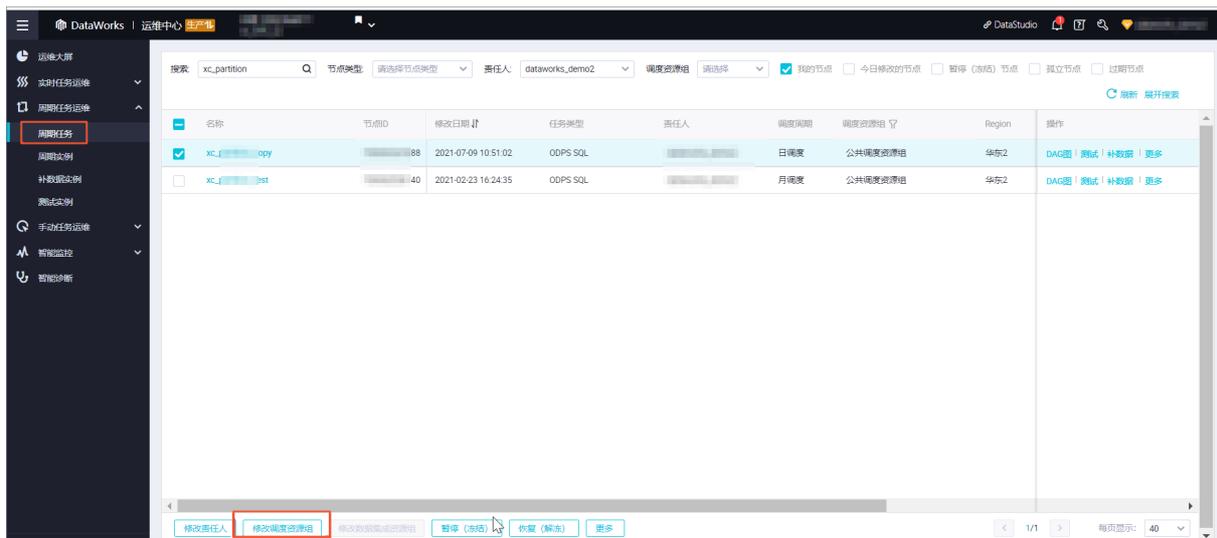
FTP Check节点任务，运行报错：`File not Exists or exceeded the end-check time point!`

- 报错内容：FTP Check节点检测FTP数据源中Done文件是否存在时，任务运行失败，错误提示为 `File not Exists or exceeded the end-check time point!`。

- 可能原因：当前检测任务在Check停止时间前最终未检测到FTP数据源中存在Done文件，或当前检测任务的检测时间已经超过配置的Check停止时间。
- 解决方案：该报错为FTP Check节点检测FTP数据源中Done文件不存在的预期报错，收到该报错时，DataWorks不会触发启动FTP Check节点的下游任务。

报错： no available machine resources under the task resource group ，该如何处理？

- 问题现象运维中心报错如下。 no available machine resources under the task resource group
- 解决方法：在运维中心页面的左侧导航栏，单击周期任务运维 > 周期任务修改任务执行使用的调度资源组。



4.10. 下线节点

本文汇总介绍下线节点的相关问题。

- 如何下线节点
- 节点存在子节点，下线失败如何处理？
- 如何确认生产任务是否下线成功？
- 如何找回下线节点

如何下线节点

数据开发删除开发节点后，会在任务发布界面有条待发布的下线操作记录，将该下线操作发布生产并且发布成功，生产任务就会下线。



发布操作受角色权限和流程控制约束，并非一定成功，请确保发布状态成功。如果发布失败，请在发布包界面查看发布包状态。



说明 点击右上角工作空间配置成员管理查看与运维权限同学。

节点存在子节点，下线失败如何处理？

当前节点在开发环境和生产环境都没有节点依赖当前节点时，当前节点才能被下线。

说明 下线节点有可能导致其他依赖该节点的任务运行异常，请与各依赖该节点的任务负责人做好沟通。

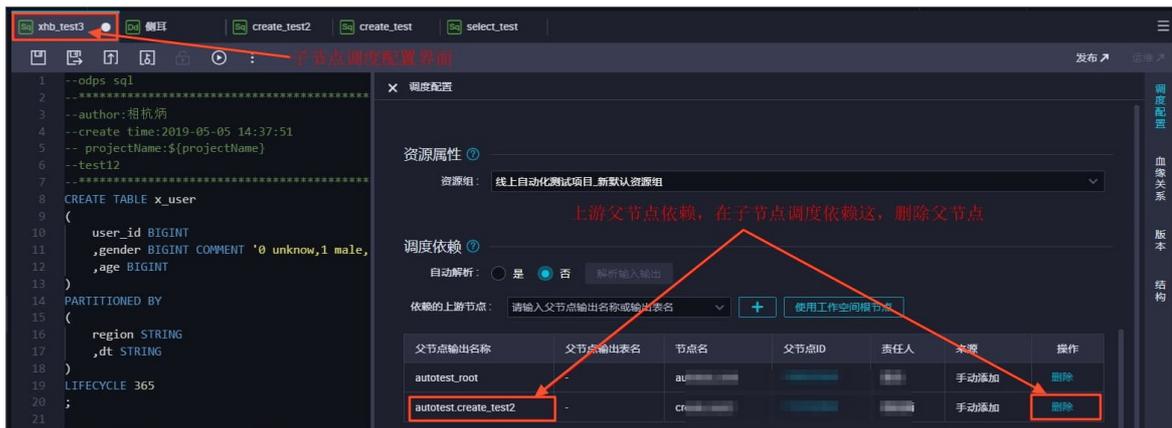
操作引导：

1. 如果子节点是您自己的节点。

可以到数据开发界面打开此子节点的调度配置界面，去掉该要下线的节点父节点依赖关系，然后保存 > 提交（去掉开发环境该下线节点依赖）> 发布（去掉生产环境该下线节点依赖）

2. 如果不是您自己的节点。

可以分别到开发运维中心和生产运维中心搜索这个子节点的信息，查看该节点的负责人并与之沟通，请他修改该节点的上游依赖（修改操作同上），不要让该节点依赖您需要下线的节点，并提交（去掉开发环境该下线节点依赖）> 发布（去掉生产环境该下线节点依赖），之后您可以删除要下线的这个节点。



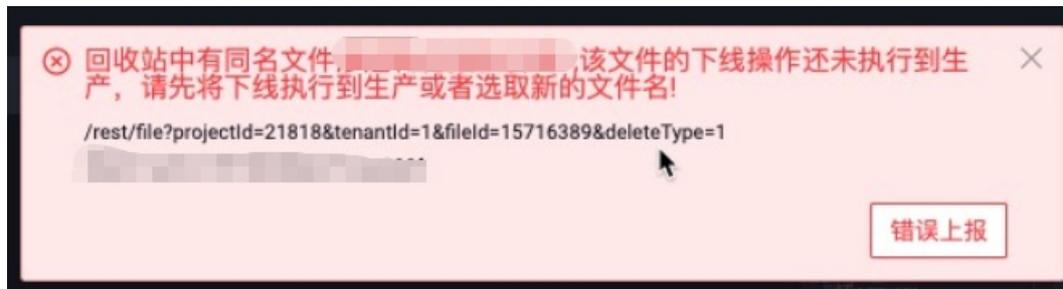
如何确认生产任务是否下线成功？

执行下线操作后请在开发运维中心和生产运维中心“周期任务”里根据节点ID再次查询该节点查看节点是否还存在。

警告 下线节点属高危操作，需谨慎！！

如何找回下线节点

节点删除后会放在回收站，如果需要，可以到回收站去还原代码。



说明 还原操作需要节点开发和生产都已经下线完成，如果遇到如上报错，请参考本文如何下线节点操作，将下线操作发布生产后再还原。

4.11. 孤立节点

本文为您介绍孤立节点的相关问题。

- 什么是孤立节点
- 如何处理孤立节点?

什么是孤立节点

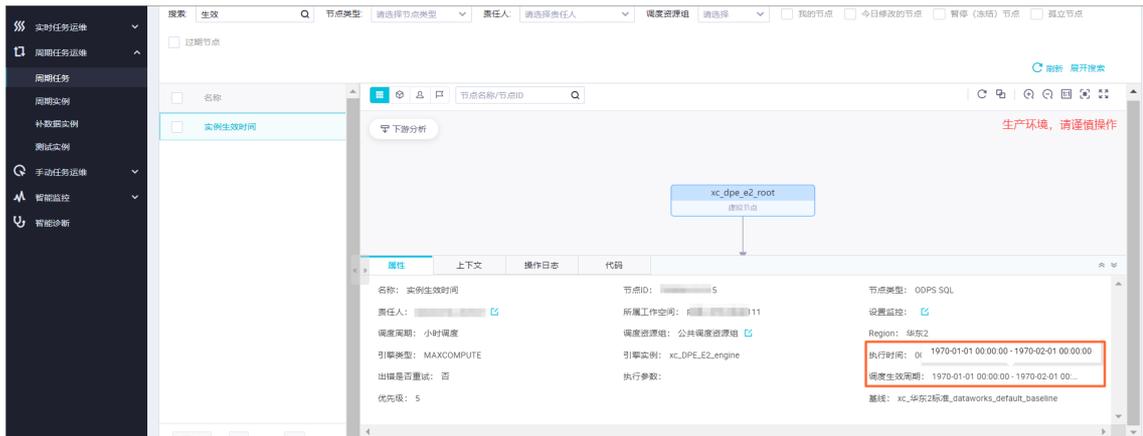
孤立节点指节点在周期任务或周期实例处展开父节点没有依赖任何父节点，这类型节点在自动调度时不会被调度调起，所以孤立节点不会自动调度运行，如果该孤立节点下游依赖较多，可能会造成较严重的后果。孤立节点产生后会自动告警，如果收到孤立节点告警请及时处理。



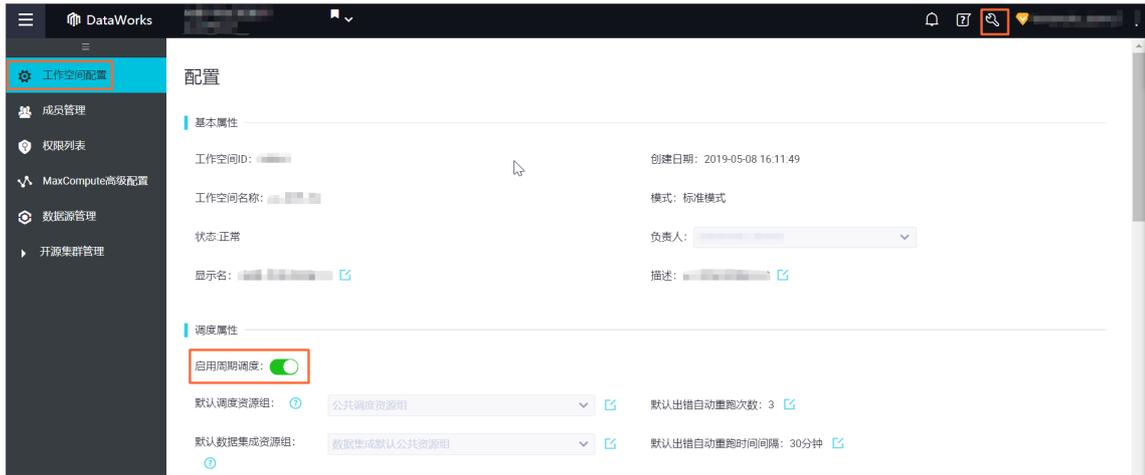
如何处理孤立节点?

1. 孤立节点产生的可能原因及解决方案:

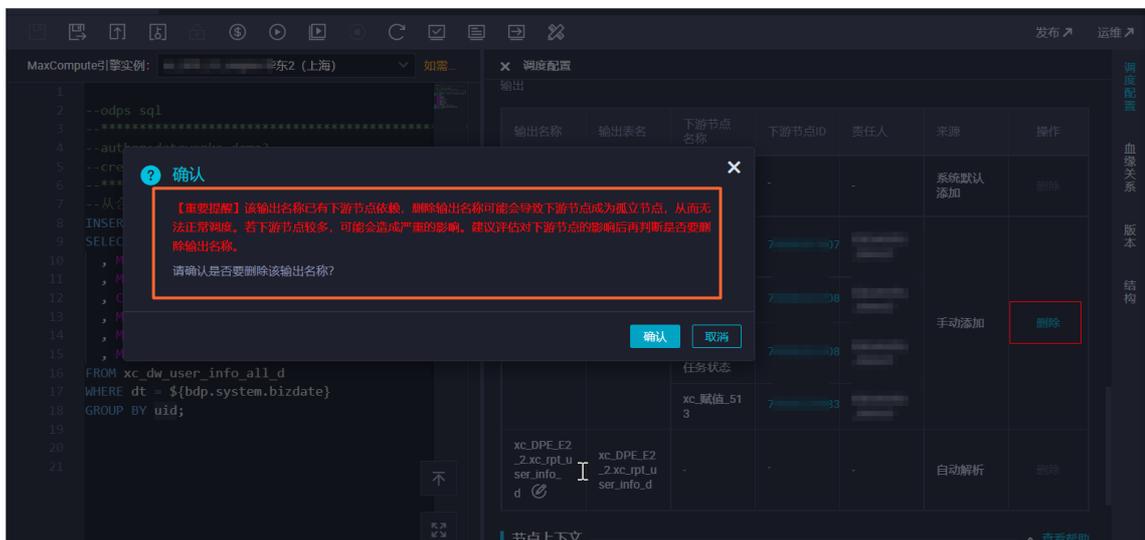
- 上游节点不存在，在生产运维中心周期任务或周期实例中搜不到，请重新为当前孤立节点设置上游依赖，上游节点发布后请在生产运维中心周期任务确认任务已经成功发布上线。
- 上游节点不在调度生效周期内，请在数据开发调度配置界面重新设置节点生效时间。



- 跨项目依赖的上游节点所在项目未开启调度。请找到对应项目owner开启项目调度或者去除该跨项目节点依赖。



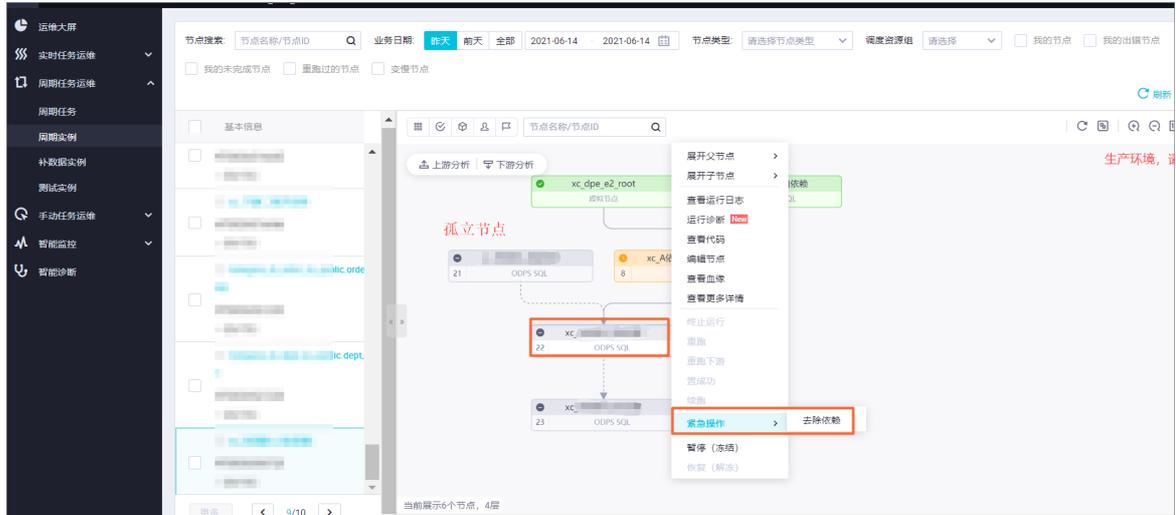
- o 上游父节点的输出名称被修改或删除。



警告 强制变更节点输出可能会导致下游所有任务不调度，请业务内部协调谨慎操作。若下线节点请先移除待下线节点的下游节点依赖后再去操作。

- 2. 下游依赖了孤立节点临时可供参考的解决方案：

孤立节点将导致下游节点无法正常运行。在紧急情况下，如果您确认对孤立节点的依赖不会影响当前节点的数据产出，您可以取消对孤立节点的依赖，让当前节点运行起来。



3. 修改孤立节点告警接收人

孤立节点告警是默认告警规则，默认告警给节点责任人，主账号可以修改是否开启该规则。点开详情可以修改告警方式及添加除节点责任人外的其他告警接受人。



4.12. 其他问题

本文汇总介绍运维中心其他常见问题。

- 任务与实例
 - 周期任务与周期实例、补数据实例、测试实例是什么关系？
- 资源组
 - 日志和实例保留多长时间？
 - 为什么不能在调度资源组上进行大数据计算？
- 其他
 - 如何设置任务优先级？
 - 如何查看任务优先级？

周期任务与周期实例、补数据实例、测试实例是什么关系？

DataWorks每晚将根据周期任务，批量生成第二天自动调度的周期实例，周期实例自动生成并且自动调度，您可以针对周期任务手动触发生成补数据实例和测试实例并执行。

 **说明** 新建节点当天，您可以选择发布后及时生成实例，即任务发布运维中心后，您可以在周期实例中看到该实例。

日志和实例保留多长时间？

公共调度资源组实例保留一个月（30天）、日志保留一周（7天）。

独享调度资源组任务实例、日志保留一个月（30天）。

 **说明**

- 完成的实例日志大于3M时，每天会定时进行清理。
- 实例生成后不支持手动删除，如果任务不需要再运行，可以选择冻结周期任务或实例。冻结周期任务和冻结实例的操作影响不同，详情请参见：[任务冻结与解冻](#)。

为什么不能在调度资源组上进行大数据计算？

调度资源组主要用于调度任务，资源有限，并不适合用来完成计算任务。MaxCompute具有海量数据处理能力，推荐您通过MaxCompute进行大数据计算。

如何设置任务优先级？

您可以通过设置节点所在基线的优先级来实现节点优先级调整。基线优先级高（取值[1,3,5,7,8]，数值越大，优先级越高），DataWorks上任务调度时将优先获得调度资源。

通过调整基线优先级来调整周期任务每日实例的优先级。

1. 进入[运维中心](#)页面。
2. 在左侧导航栏，单击[智能监控 > 基线管理](#)。
3. 单击相应基线后的[编辑](#)。
4. 在[修改基线](#)对话框中，设置优先级。
5. 单击完成。

如何查看任务优先级？

您可以在[运维中心 > 实例属性](#)可查看任务优先级。

5. 数据开发与运行

本文为您介绍数据开发的相关问题。

- 资源
 - 如何在pyodps调用第三方包？
 - 如何在节点内使用资源？
 - 如何下载上传到DataWorks的资源？
 - 超过30M的资源如何上传？
 - 通过odpscmd上传的资源如何在DataWorks上使用？
 - 在DataWorks如何上传自己本地编写的jar并执行？
 - 如何实现在DataWorks上使用MaxCompute表资源？
- pyodps
 - Python资源是否支持调用另一个Python资源？
 - PyODPS是否支持调用自定义函数来使用第三方包？
 - PyODPS 3中调用Pickle文件报错：_pickle.UnpicklingError: invalid load key, '\xef.
 - 如何删除MaxCompute资源？
- 节点与业务流程
 - 节点删除后如何还原？
 - 节点被删除，实例是否受影响？
 - 如何查看节点版本？
 - 如何查看业务节点提交状态？
 - 修改后的节点任务提交发布至生产环境后，是否会覆盖之前生产环境的错误节点？
 - 如何将工作空间下的节点代码导出？
 - 当业务流程中包含多个节点时，是否可以批量配置节点的调度信息？
 - 如何克隆业务流程？
- 表
 - 如何可视化新建表？
 - 如何对生产表添加字段？
 - 如何删除表？
 - 如何将本地数据上传到MaxCompute表中？
 - 为什么EMR通过DataWorks可视化建表失败报错：call emr exception？
 - 如何在开发环境访问生产环境数据？
 - 如何控制表数据查询以后是否可以下载？
 - 如何下载超过1万条的表数据？
 - 为什么EMR通过DataWorks可视化建表失败报错：call emr exception？
- 运行日志及保留时长
 - 数据开发界面历史执行日志如何获取？
 - 数据开发运行历史保留多长时间？

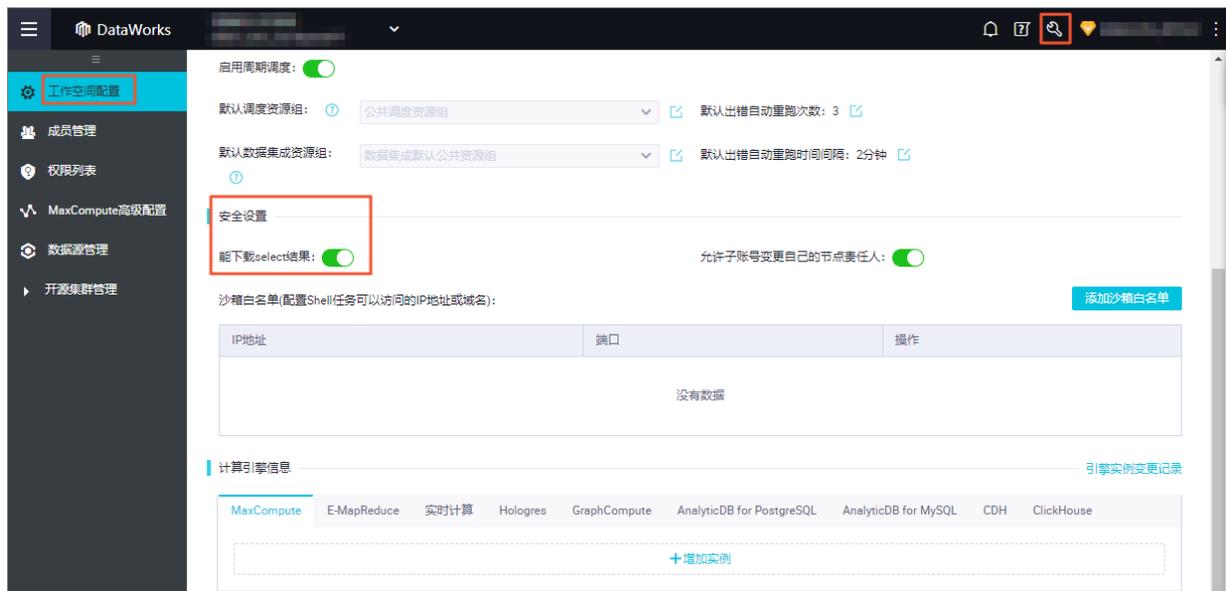
- 批量操作
 - 如何对节点、资源、函数等进行相关批量修改相关属性的操作？
 - 如何在数据开发页面批量修改业务流程下节点使用的调度资源组？
- BI对接
 - 使用Power BI连接MaxCompute报错，该如何处理？
- API调用
 - Openapi调用报错access is forbidden. Please first activate DataWorks Enterprise Edition or Flagship Edition.
 - 如何获取pythonsdk调用案例？
- 其他
 - 任务执行没有instanceid，如何关闭odps加速模式？

如何在pyodps调用第三方包？

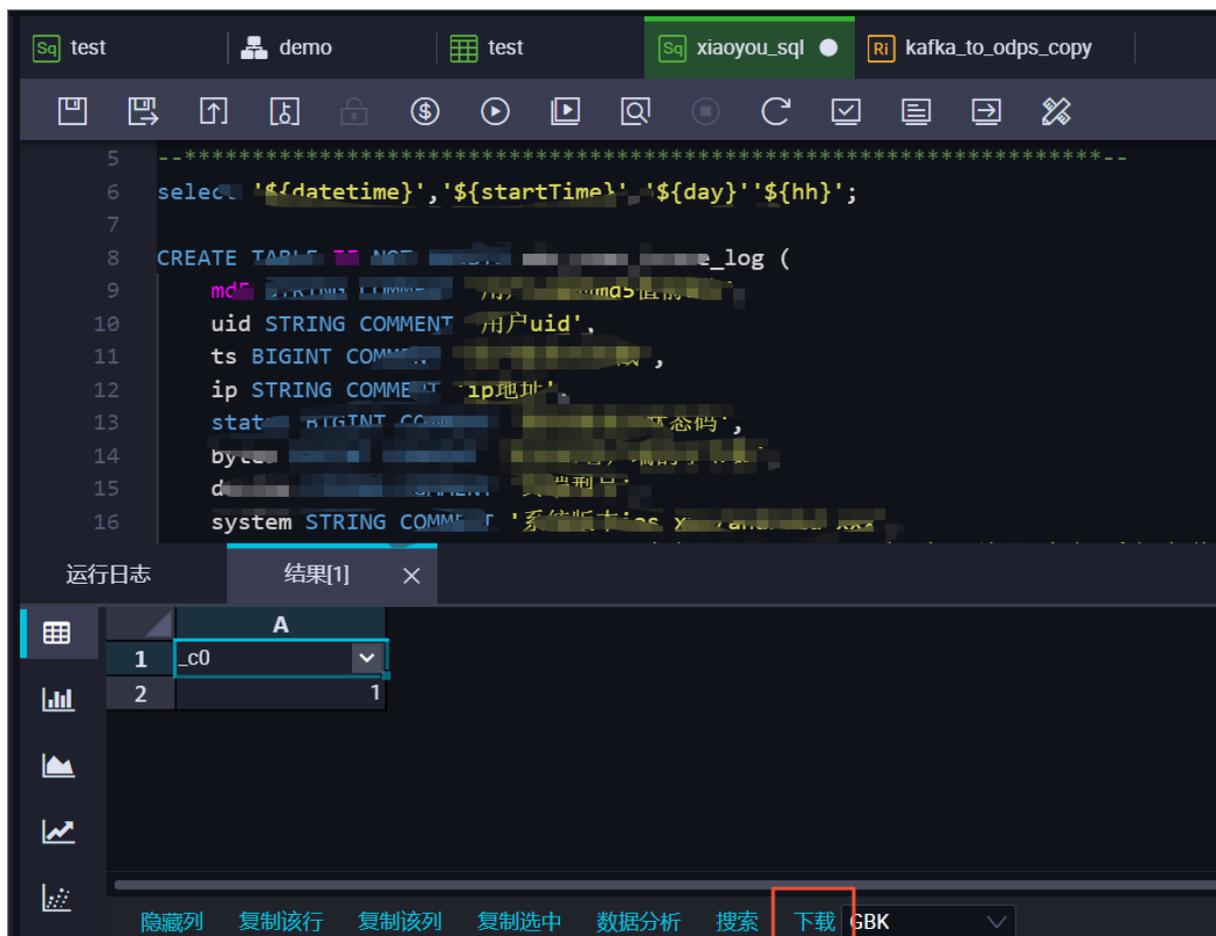
需要使用独享调度资源组进行操作，详情请参考文档：[在PyODPS节点中调用第三方包](#)

如何控制表数据查询以后是否可以下载？

在DataWorks下载数据的前提是要将下载功能打开，如果没有下载入口，说明工作空间未开启下载，如果有需要可以联系主账号或者工作空间管理员在工作空间管理处开启。



查询数据后在查询结果右下角会有下载入口，如下图所示：



DataWork界面下载由于引擎限制只能下载10000条数据。

如何下载超过1万条的表数据？

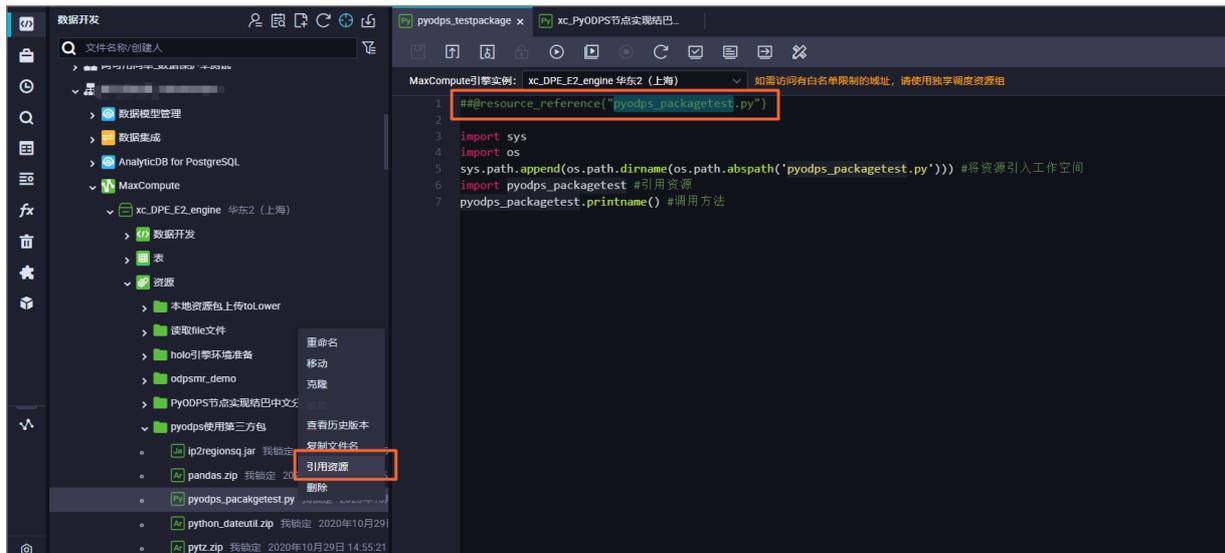
需要通过MaxCompute tunnel操作：[SQLTask配合Tunnel实现大量数据导出示例](#)

为什么EMR通过DataWorks可视化建表失败报错：call emr exception?

- 可能原因：EMR所在的ECS集群的安全组缺失必要安全组配置。绑定EMR计算引擎使用时，您需要添加以下安全组策略，否则可能会出现上述建表失败提示。
- 如何处理：查看EMR所在的ECS集群的安全组配置情况，补充上述安全组策略。

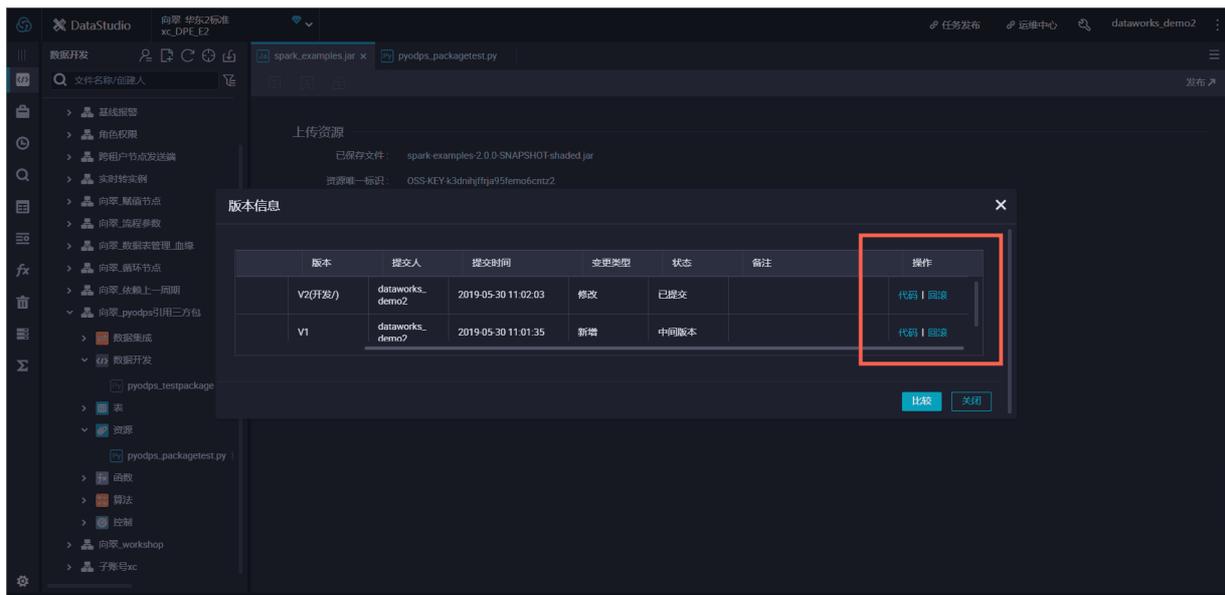
如何在节点内使用资源？

右键目标资源节点，选择引用资源。



如何下载上传到DataWorks的资源?

右键目标资源节点，选择查看历史版本。



超过30M的资源如何上传?

超过30M的资源需要通过[tunnel] (使用Tunnel命令上传下载数据) 上传，上传后再使用MaxCompute资源功能添加到DataWorks进行后续的资源使用操作。详情请参考常见问题：[通过odpscmd上传的资源如何在DataWorks上使用?](#)

通过odpscmd上传的资源如何在DataWorks上使用?

odpscmd上传的资源在DataWorks上使用请在MaxCompute资源处将资源添加到数据开发。



在DataWorks如何上传自己本地编写的jar并执行？

您需要在数据开发界面将该jar作为资源上传到DataWorks，节点使用该jar资源时，您需要先右键目标资源节点 > 引用资源（会在节点最上方自动多出一条注释），然后直接使用资源名执行。



例如：在Shell节点内 `##@resource_reference{"test.jar"} java -jar test.jar`

如何实现在DataWorks上使用MaxCompute表资源？

目前DataWorks不支持直接使用图形界面上上传MaxCompute表资源。如果您想在DataWorks上使用MaxCompute表资源，操作如下：

1. 在DataWorks数据开发页面上上传一个File类型资源，资源名称和表名称一致，本示例为userlog3.txt。

说明

请勿选中上传为ODPS资源。

2. 完成上传后，通过MaxCompute CLI客户端添加MaxCompute表资源，示例如下：`add table userlog3 -f;`
3. 完成添加后，在DataWorks上使用MaxCompute表资源时，直接使用上传的资源即可。

Python资源是否支持调用另一个Python资源？

两个资源在同一个工作空间的前提下，Python资源支持调用另一个Python资源。

PyODPS是否支持调用自定义函数来使用第三方包？

如果您不想通过DataFrame的map方法使用test函数，PyODPS同样支持直接调用自定义函数来导入第三方包，详情请参见[PyODPS使用第三方包](#)

PyODPS 3中调用Pickle文件报错：`_pickle.UnpicklingError: invalid load key, '\xef.`

如果您的代码中存在特殊字符，请将代码打成zip再上传，然后在代码中进行解压使用。

如何删除MaxCompute资源？

资源创建完成后，如果需要删除资源，简单模式下直接右键单击所需删除的资源即可，标准模式下需要先在开发环境中删除资源，再在生产环境中删除资源，以下以删除生产环境的资源为例，为您演示删除资源的操作步骤。

说明

如果您的DataWorks为标准模式，则开发环境和生产环境分离，在DataWorks的DataStudio删除资源时，只删除了开发环境的资源，生产环境的资源需要将删除操作发布到生产环境，生产环境的资源才会被删除下线。

1. 在开发环境中删除资源。在对应业务流程下，选择MaxCompute > 资源，右键待删除的资源后单击删除，在弹出的页面单击确认。



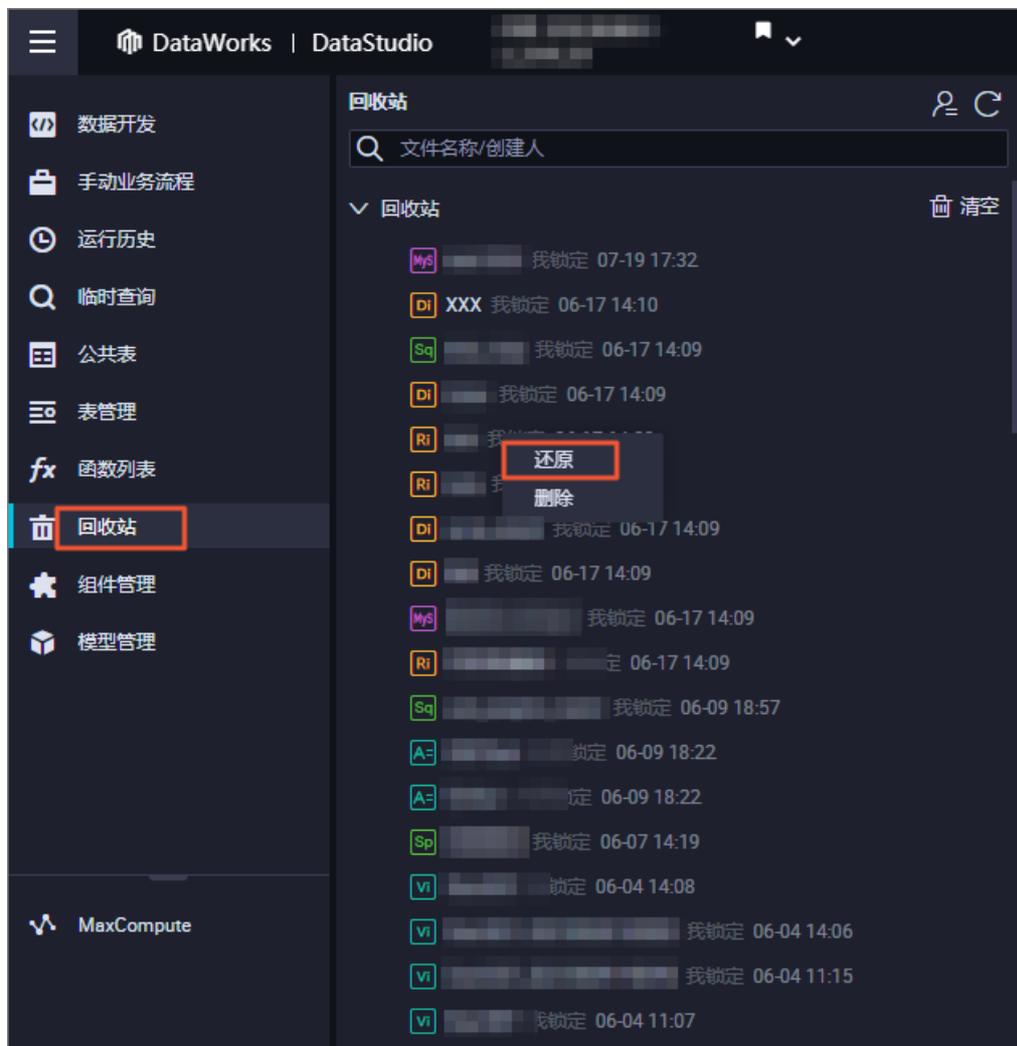
2. 在生产环境中删除资源。删除操作会在任务发布界面生成一条待发布的记录，发布后才能生产任务才会下线。您可以在数据开发DataStudio页面的右上角单击任务发布，过滤变更类型为下线，找到上述步骤下线资源的变更发布包后，单击操作列的发布，在弹出的页面中单击发布。



完成发布后，生产环境的资源才会被删除。

节点删除后如何还原？

节点删除后可以进入回收站进行还原。



如何查看节点版本？

打开单个节点配置界面可以查看节点的版本

注意
已提交才会生成版本。

	文件ID	版本	提交人	提交时间	操作
<input checked="" type="checkbox"/>	500572428	V12(开发)		2021-03-23 18:38:39	代码 回滚
<input checked="" type="checkbox"/>	500572428	V11(生产)		2021-01-29 17:54:40	代码 回滚
<input type="checkbox"/>	500572428	V10		2021-01-29 17:46:42	代码 回滚
<input type="checkbox"/>	500572428	V9		2021-01-29 17:06:42	代码 回滚
<input type="checkbox"/>	500572428	V8		2020-10-14 15:00:31	代码 回滚
<input type="checkbox"/>	500572428	V7		2020-09-23 10:43:23	代码 回滚
<input type="checkbox"/>	500572428	V6		2020-05-19 17:30:15	代码 回滚
<input type="checkbox"/>	500572428	V5		2020-05-19 11:12:42	代码 回滚
<input type="checkbox"/>	500572428	V4		2020-01-09 11:11:06	代码 回滚
<input type="checkbox"/>	500572428	V3		2019-10-09 15:20:42	代码 回滚
<input type="checkbox"/>	500572428	V2		2019-09-03 20:54:01	代码 回滚
<input type="checkbox"/>	500572428	V1		2019-09-03 19:52:35	代码 回滚

如何克隆业务流程？

您可以使用节点组功能：[节点组](#)

如何将工作空间下的节点代码导出？

您可以使用迁移助手功能：[概述](#)

如何查看业务节点提交状态？

如果您希望查看业务节点提交状态，可以选择[数据开发](#) > [业务流程](#)，展开对应的业务流程列表，即可查看当前业务流程下所有节点状态。如果节点名称左侧存在图标，表示该节点已提交；如果节点名称左侧不存在图标，表示该节点未提交。

当业务流程中包含多个节点时，是否可以批量配置节点的调度信息？

DataWorks不支持对业务流程配置调度信息，对于业务流程内的节点，您无法批量配置调度信息，只能逐个配置节点的调度信息。例如，一个业务流程中包含20个节点，您只能为每个节点逐个配置调度信息。

节点被删除，实例是否受影响？

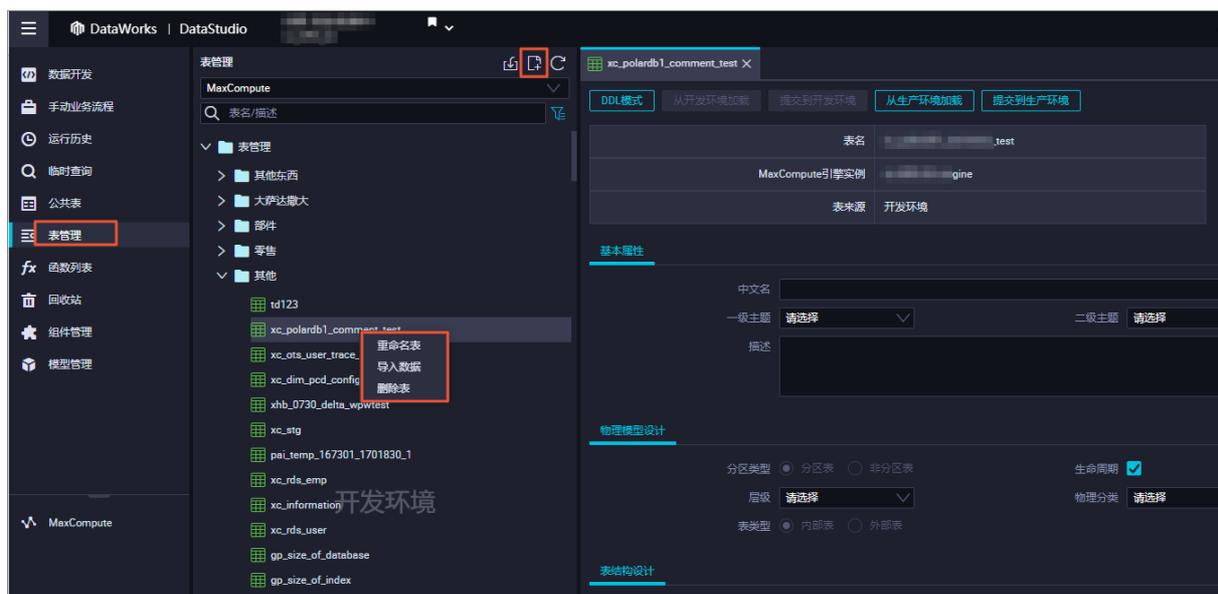
调度系统每天会根据时间属性为该任务生成对应的一个或多个实例，当一个任务运行一段时间后被删除，则实例不会被删除。当实例在删除任务后被触发运行时，会由于找不到需要运行的代码而失败。

修改后的节点任务提交发布至生产环境后，是否会覆盖之前生产环境的错误节点？

不会覆盖之前的节点，未运行的实例会用最新代码运行，不会删除已生成的节点实例。如果调度参数有变化，需要重新生成实例去运行。

如何可视化新建表？

可以在数据开发，表管理和业务流程内的表容器中建表。



如何对生产表添加字段？

主账号可以在表管理页面，对生产表添加字段后，提交至生产环境。

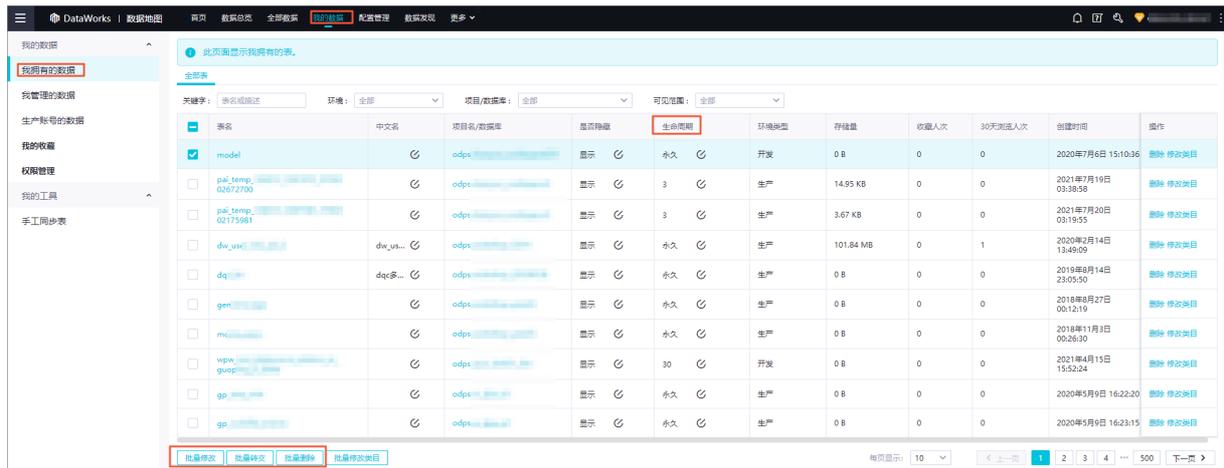
子账号需要拥有运维或项目管理员角色，才可以在表管理页面，对生产表添加字段后，提交至生产环境。

如何删除表？

删除开发表：可以在数据开发界面删除。

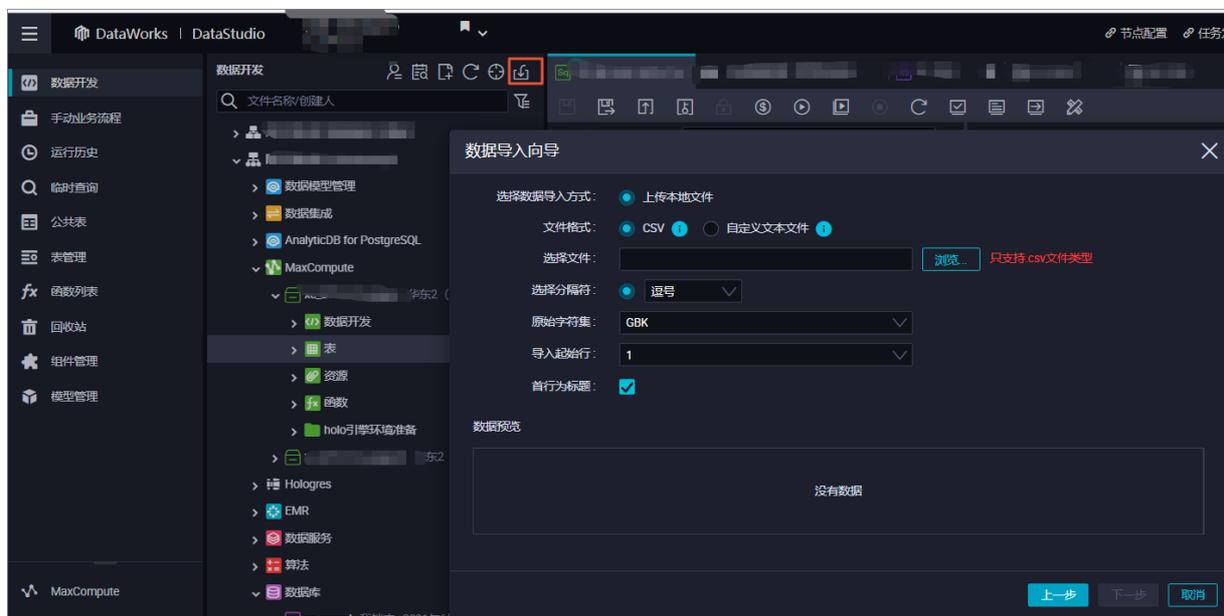
删除生产表：

- 生产表请前往数据地图我的数据中删除表。
- 您可以创建一个ODPS SQL节点，输入并执行DROP语句。创建ODPS SQL节点的详情请参见[创建ODPS SQL节点](#)，删除表的语法格式请参见[表操作](#)。



如何将本地数据上传到MaxCompute表中？

您可以进入数据开发找到导入表功能导入本地数据。



在EMR计算引擎中建表失败，界面提示：`call emr exception?`

● 可能原因：

EMR所在的ECS集群的安全组缺失必要安全组配置。绑定EMR计算引擎使用时，您需要添加以下安全组策略，否则可能会出现上述建表失败提示。

- 授权策略：允许
- 协议类型：自定义TCP
- 端口范围：8898/8898
- 授权对象：100.104.0.0/16

● 解决方法：

查看EMR所在的ECS集群的安全组配置情况，补充上述安全组策略。

如何在开发环境访问生产环境数据？

标准模式下，如果在数据开发界面需要查询生产环境的数据，请使用项目名.表名进行访问。

如果是简单模式升级到标准模式，在数据开发界面需要查询生产环境数据，则需要先申请生产者角色权限，再使用项目名.表名进行访问。申请权限具体请参考[申请表权限](#)

数据开发界面历史执行日志如何获取？

您可以进入数据开发界面左侧边栏找到运行历史模块，查看历史执行日志。

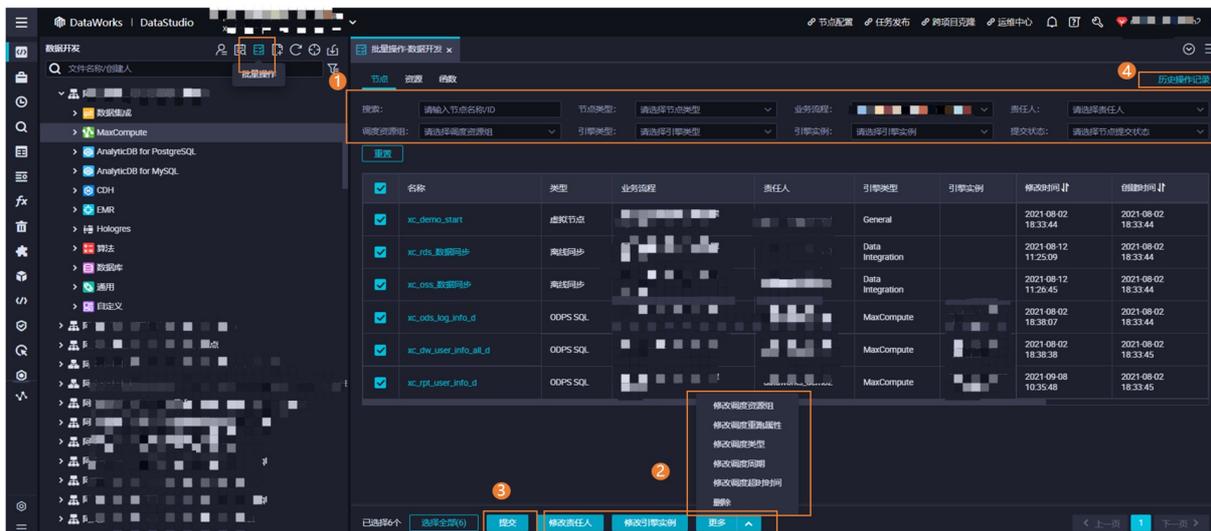
数据开发运行历史保留多长时间？

数据开发界面的运行历史默认保留3天。

? 说明 生产运维中心日志与实例保留时长请参考：[日志和实例保留多长时间？](#)

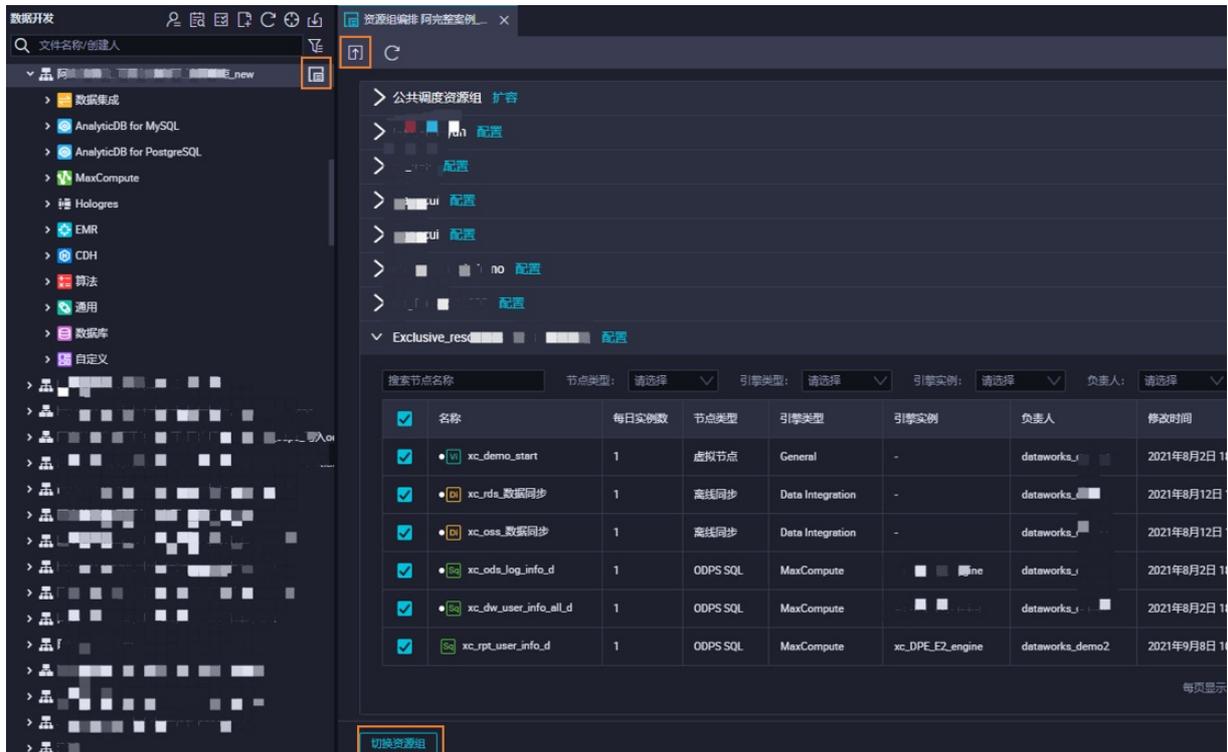
如何对节点、资源、函数等进行相关批量修改相关属性的操作？

您可以在数据开发界面左侧工具栏中进入批量操作界面。支持针对节点、资源、函数进行批量操作，批量修改完成后，您可以批量提交，并在任务发布界面批量发布，让节点变更操作生产环境生效。



如何在数据开发页面批量修改业务流程下节点使用的调度资源组？

您可以在数据开发页面指定的业务流程名处，进入资源组编排，您可以在此页面批量修改该业务流程内的节点指定的调度资源组，修改完成后，您可以选择提交按钮进行批量提交，并在任务发布界面批量发布，让节点使用的调度资源组变更操作在生产环境生效。



使用Power BI连接MaxCompute报错，该如何处理？

目前MaxCompute不支持连接Power BI，建议您使用交互式分析（Hologres）进行连接，详情请参见[访问域名](#)

Openapi调用报错 access is forbidden. Please first activate DataWorks Enterprise Edition or Flagship Edition.

openapi需要DataWorks企业版和旗舰版。详情请参见：[DataWorks OpenAPI概述](#)

如何获取pythonsdk调用案例？

您可以在对应的API页面点击调试查看python sdk的demo。

任务执行没有instanceid，如何关闭odps加速模式？

获取instanceid需要关掉加速模式。

说明 DataWorks只支持下载1W条数据，超过1W的数据需要通过tunnel下载，下载时需要instanceid、

```

...
提交任务成功
任务执行中...
任务执行中...
任务执行中...
恭喜您！您的任务被选中以 MaxCompute 查询加速模式运行
MaxCompute 查询加速模式运行sql成功
SQL: SELECT a.real_name
...

```

在运行任务的时候在odps sql节点编辑器里加上 `set odps.mcqa.disable=true;`（需要和其他select语句一块执行）。

6. 账号与权限

6.1. 用户权限控制

本文汇总介绍用户权限控制的相关问题。

- [为什么子账号进入DataWorks控制台后看不到具体项目？](#)
- [如何将子账号加入到具体工作空间下？](#)
- [如何授权子账号创建DataWorks工作空间的权限？](#)
- [主账号如何给子账号添加Aliyundataworksfullaccess权限？](#)
- [如何为子账号授权购买相关服务的权限？](#)
- [如何为子账号授权资源组绑定与解绑工作空间的权限？](#)
- [怎么创建MaxCompute仅有查询权限的自定义角色？](#)
- [如何设置子账号手机邮箱？](#)
- [人员离职注意事项](#)

为什么子账号进入DataWorks控制台后看不到具体项目？

DataWorks控制台只展示当前账号已经被加入的工作空间列表，子账号需要被加入到DataWorks工作空间后才可以看到在控制台工作空间列表中看到具体工作空间。

如何将子账号加入到具体工作空间下？

工作空间管理员可以在工作空间管理处将子账号加入到具体工作空间下。详情请参见：[添加工作空间成员和角色](#)。

如何授权子账号创建DataWorks工作空间的权限？

需要主账号在[阿里云访问控制](#)为子账号添加Aliyundataworksfullaccess权限。

主账号如何给子账号添加Aliyundataworksfullaccess权限？

在阿里云访问控制添加：



如何为子账号授权购买相关服务的权限？

需要主账号在[阿里云访问控制](#)为子账号添加AliyunBSSOrderAccess权限。

如何为子账号授权资源组绑定与解绑工作空间的权限？

需要子账号拥有该工作空间的管理员角色。

怎么创建MaxCompute仅有查询权限的自定义角色？

可以参考文档：[查询权限](#)

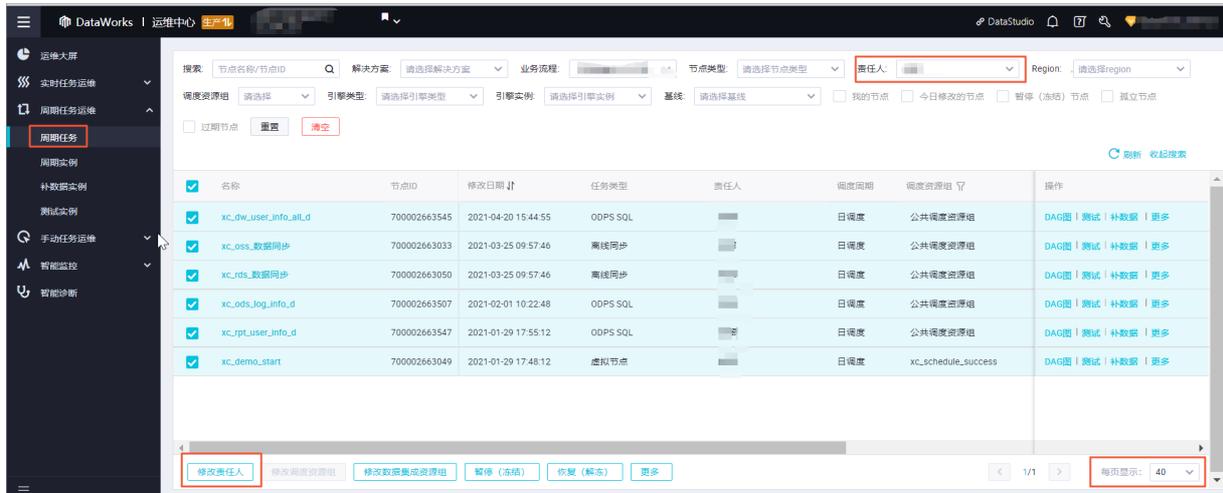
如何设置子账号手机邮箱？

如果子账号需要接收报警，主账号需要为子账号在Dat aWorks控制台的报警配置中为子账号设置个人邮箱手机号，详情请参见：[查看和设置报警联系人](#)

人员离职注意事项

人员离职前，请在Dat aWorks将相关任务转交后再去访问控制将该RAM用户移除，如果直接在RAM访问控制移除用户可能会导致Dat aWorks侧出现一些报错。

批量转交名下节点



说明 请及时更新告警规则，检查值班表配置。



6.2. 操作权限控制

本文介绍操作权限控制的相关问题。

如何控制工作空间下各用户可操作哪些功能权限？

您可以根据业务场景，为子账号授予不同的预设角色，来控制其是否有某个功能点权限，同时，您也可以为子账号授予不同的自定义DataWorks空间角色，来控制其是否有某个模块读写权限。关于预设角色拥有的对应功能权限列表请参考文档：[附录：预设角色权限列表（空间级）](#)。关于自定义DataWorks角色您可以参考文档：[角色及成员管理：空间级](#)。

如何控制工作空间下各用户是否有对应引擎资源的操作权限？

当授权用户工作空间角色后，该用户是否拥有引擎权限与引擎类别和引擎配置有关。

- 使用MaxCompute引擎时的权限问题逻辑说明：
 - DataWorks预设角色与MaxCompute引擎开发项目的Role存在权限映射关系（DataWorks预设角色默认拥有MaxCompute开发项目映射的role所有的引擎层面的权限）。

- DataWorks预设角色与MaxCompute引擎生产项目的Role没无权限映射关系（DataWorks预设角色无法直接操作生产引擎项目上的资源）。

说明 上述说明在DataWorks上的具体体现为，管理员角色和开发角色的成员，有DataWorks上大部分功能权限，有DataWorks开发环境(MaxCompute引擎开发项目)所有权限，但默认没有生产环境(MaxCompute生产项目)的操作权限。如果子账号需要在开发环境访问生产表，需要在数据地图**权限申请**申请该表的操作权限。

任务执行环境	实时场景
数据开发执行任务（开发环境执行任务）	<ul style="list-style-type: none"> ■ 场景一：<code>select col1 from tablename</code> 用个人账号访问开发环境下该表。即用个人账号访问：<code>projectname_dev.tablename</code>。 ■ 场景二：<code>select col1 from projectname.tablename</code> 用个人账号访问生产环境下该表。即用个人账号访问：<code>projectname.tablename</code>。 <p>说明 阿里云RAM用户（非调度引擎指定账号）默认无权限访问生产环境，生产表权限需要在数据地图申请后才可使用。</p>
生产运维中心执行任务（生产环境执行任务）	<p>场景：<code>select col1 from tablename</code> 用调度引擎指定账号访问生产环境下该表。即调度引擎指定账号访问：<code>projectname.tablename</code>。</p>

- 使用E-MapReduce引擎时的权限逻辑说明与权限控制：

- 逻辑说明：当前工作空间使用的是E-MapReduce引擎时，预设角色与引擎无直接关系。

模式	执行环境	执行账号	原理
快捷模式	数据开发执行任务（开发环境执行任务）	统一使用Hadoop账号执行	
	生产运维中心执行任务（生产环境执行任务）		
安全模式	数据开发执行任务（开发环境执行任务）	使用在引擎配置时候指定的开发环境指定的账号访问所有引擎资源。	通过为DataWorks工作空间下的成员配置LDAP权限映射，实现控制每个子账号在DataWoks操作时的EMR底层权限控制的目的。在DataWorks上使用阿里云主账号或RAM用户在下发代码的同时，EMR集群内会匹配对应的同名用户来运行任务。
	生产运维中心执行任务（生产环境执行任务）	使用在引擎配置时候指定的生产环境指定的账号访问所有引擎资源。	

说明
DataWorks成员与EMR用户权限映射关系详情可参考：[准备工作：绑定EMR引擎](#)

- 权限控制：管理者可以使用EMR集群内的Ranger组件对每个用户进行权限管控，最终实现不同阿里云主账号、任务责任人或RAM用户在DataWorks上运行EMR任务时，拥有对应不同的数据权限。

● 使用其他引擎时的权限问题

当前工作空间使用的其他引擎时，在数据开发界面执行任务访问引擎资源时是否有权限与您在引擎配置中配置账号有关。

如何控制只能从内网环境访问DataWorks控制台？

如果您希望您所在企业的用户只能通过企业内网访问DataWorks控制台，请进入RAM控制台，仅放通您所在企业的内网地址映射出的公网地址。

关于RAM控制台的详细操作，请参见[设置RAM用户安全策略](#)

6.3. 操作审计

本文汇总介绍操作审计的相关问题。

- [如何获取界面一些操作审计日志，比如界面下载数据操作？](#)
- [重要数据如何通过事前设置来获得事后溯源能力？](#)
- [MaxCompute表权限的权限审计](#)
- [表数据，节点删除如何恢复？](#)
- [如何进行节点版本对比与版本回滚？](#)

如何获取界面一些操作审计日志，比如界面下载数据操作？

DataWorks已集成至操作审计（ActionTrail）中，您可以在ActionTrail中查看及检索阿里云账号最近90天的DataWorks行为事件日志。后续可以通过ActionTrail将事件日志投递至日志服务LogStore或指定的OSS Bucket中，实现对事件的监控和告警，满足及时审计、问题回溯分析等需求。详情请参考：[通过操作审计查询行为事件日志](#)。

重要数据如何通过事前设置来获得事后溯源能力？

如果您文件比较重要防止文件泄露，您可以通过数据保护伞功能的脱敏配置里面的数据水印功能溯源来实现。详情请参考文档：[数据保护伞数据脱敏管理：数据脱敏管理](#)

MaxCompute表权限的权限审计

您可以进入安全中心，在数据访问控制，权限审计处，查看表都有谁有具体的什么权限，以及申请的权限有效期，并支持回收表权限。[权限审计](#)

表数据，节点删除如何恢复？

- 节点代码恢复：数据开发回收站可以还原近期删除的节点，但注意节点还原后节点ID是新生成的。使用详情可参考文档：[回收站](#)
- MaxCompute表数据恢复：提供数据备份与恢复功能，系统会自动备份数据的历史版本（例如被删除或修改前的数据）并保留一定时间，相关资源可参考：[备份与恢复](#)

如何进行节点版本对比与版本回滚？

您可以在数据开发界面找到该节点，在节点右侧配置界面的版本处进行版本对比与回滚操作。详情可参考文档：[版本](#)。

7. 资源组

7.1. 独享资源组

本文为您介绍DataWorks独享资源组的相关问题。

- 独享资源组使用场景
 - [独享调度资源组使用场景](#)
 - [独享数据集成资源组使用场景](#)
- 独享资源组网络配置
 - [独享资源组访问VPC环境下数据的前提条件是什么？](#)
 - [如何查看数据源的网络环境？](#)
 - [添加独享资源组白名单](#)
- 独享资源组商业化行为
 - [如何对资源组进行续费？](#)
 - [如何对资源组进行扩容、缩容和规格变更操作？](#)
- 独享资源组使用细则
 - [如何切换任务执行时使用的调度资源组？](#)
 - [如何切换数据集成任务执行时使用的数据集成资源组？](#)
 - [如何将资源组与工作空间进行绑定和解绑？](#)

独享调度资源组使用场景

如果非数据集成任务需要访问VPC环境下的数据库、有白名单访问控制的数据库时，需要使用独享调度资源组，独享调度资源组的使用详情可参考文档：[新增和使用独享调度资源组](#)

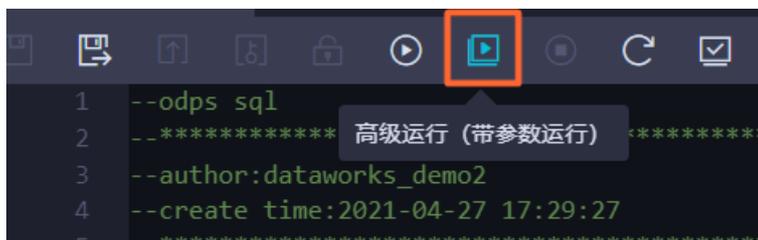
独享数据集成资源组使用场景

如果您需要同步VPC环境下的数据库，您需要使用独享数据集成资源组，独享数据及后才能资源组的使用详情可参考文档：[新增和使用独享数据集成资源组](#)

如何切换任务执行时使用的调度资源组？

1. 切换在数据开发测试运行时任务使用的资源组：

如果您需要切换测试运行时的调度资源组，请选择高级运行。



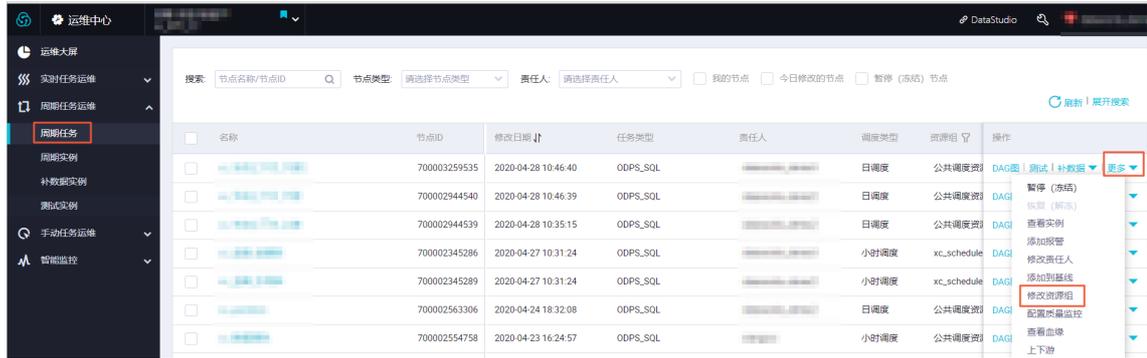
2. 切换生产调度运行时任务使用的调度资源组：
 - 在数据开发界面通过发布流程切换生产任务执行时使用的调度资源组。

在数据开发界面打开节点调度配置界面，在资源属性区域选择任务调度运行时需要使用的资源组。并提交发布。详情请参见配置资源属性。



- 在生产运维中心修改任务执行时使用的调度资源组。

左上角的☰图标，选择全部产品 > 任务运维 > 运维中心（工作流），进入生产运维中心页面,选择周期任务。



说明 虚拟节点不占用资源，所以无需也不支持修改调度资源组。

如何切换数据集成任务执行时使用的数据集成资源组？

1. 切换数据集成任务开发环境测试运行时使用数据集成任务执行资源组：

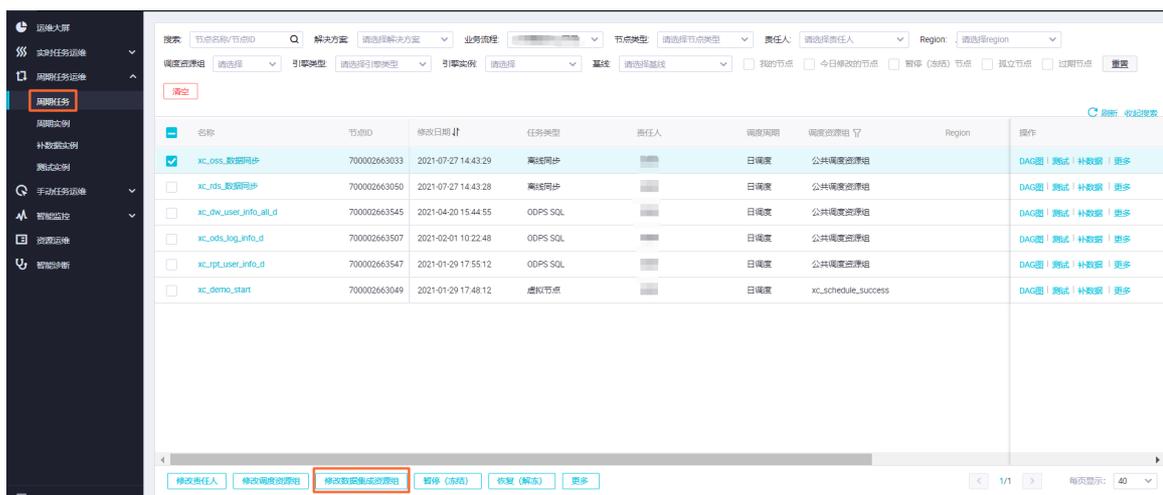
在数据开发界面打开离线同步任务，选择右侧的数据集成资源组配置。



说明 如果生产环境执行也需要使用该资源组，请修改资源组后进行节点提交发布操作。

2. 切换生产调度运行时使用数据集成任务执行资源组：

左上角的☰图标，选择全部产品 > 任务运维 > 运维中心（工作流），进入生产运维中心页面,选择周期任务。



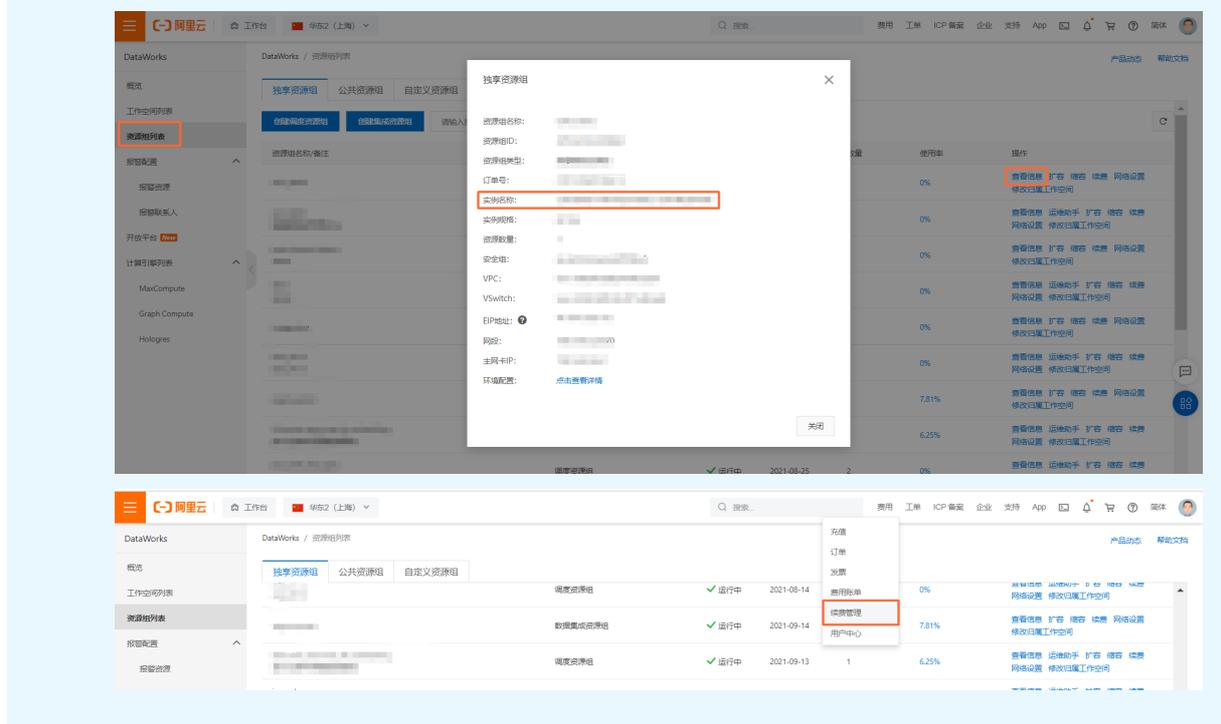
如何对资源组进行续费？

在Dat aWorks控制台组员组列表，选择具体资源组进行续费。



说明

- 如果子账号需要进行相关操作，请参考文档[如何为子账号授权购买相关服务的权限？](#)
- 如果您需要修改资源组续费方式（手动续费，自动续费），您可以根据资源组的实例名称在阿里云续费管理页面进行处理。



如何对资源组进行扩容、缩容和规格变更操作？

您可以在Dat aWorks控制台资源列表中对指定资源组进行扩容，缩容和规格变更操作。



说明

- 如果子账号需要进行相关操作，请参考文档[如何为子账号授权购买相关服务的权限？](#)
- 扩容、缩容操为增加或减少资源组下不同规格的资源组数量。

独享资源组访问VPC环境下数据的前提条件是什么？

您在Dat aWorks使用独享资源组访问数据库，需要确保以下内容：

- 购买独享资源时，需要选择数据源所在的可用区，并为独享资源组绑定专有网络。
- 独享资源绑定的专有网络和交换机，需要和数据源保持一致。
- 如果数据库有白名单限制，您需要添加独享资源组EIP地址、独享资源组绑定的专有网络（或交换机）网段至数据库的白名单和安全组访问许可内。详情请参见[添加白名单](#)和[场景示例：ECS自建数据库的安全组配置](#)。

如何查看数据源的网络环境？

- 查看RDS数据源的网络环境，方式如下：
 - i. 登录[RDS控制台](#)。
 - ii. 在实例列表页面，单击相应的实例名称。
 - iii. 在左侧导航栏，单击数据库连接。
 - iv. 单击切换交换机。



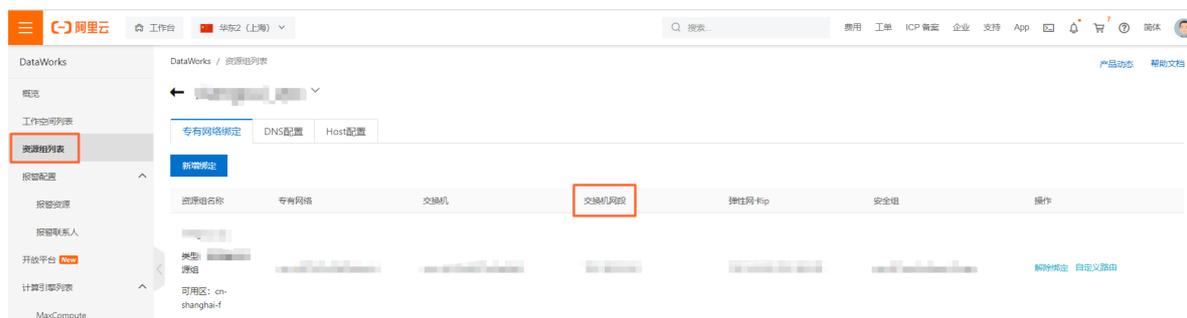
- v. 在弹出的切换交换机对话框中查看数据源所在的交换机。
- 查看ECS自建数据源的网络环境，方式如下：
 - i. 登录[ECS控制台](#)。
 - ii. 在实例页面，单击相应的实例名称。
 - iii. 在实例详情页面，查看该实例的配置信息。



添加独享资源组白名单

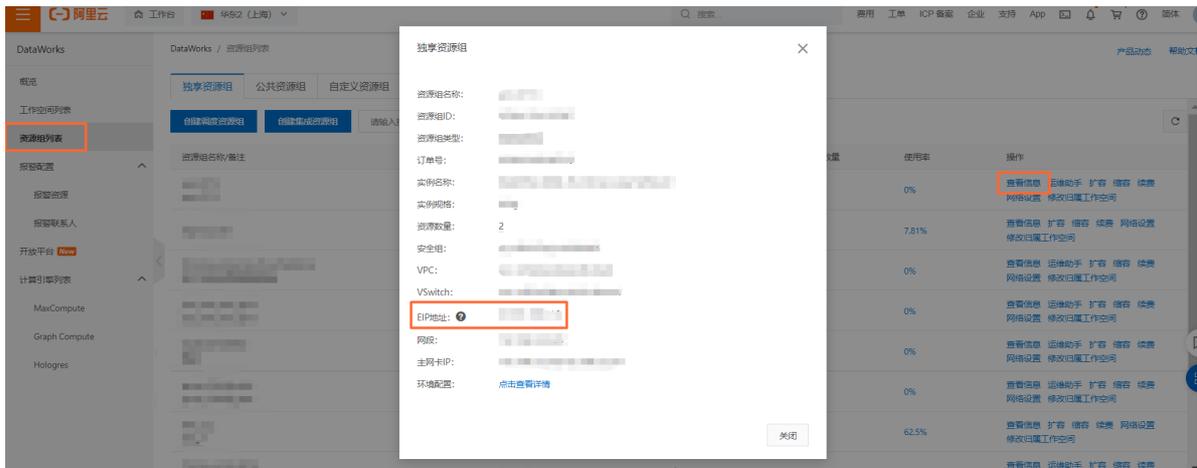
- 如果您使用独享数据集成资源组走VPC内网同步数据，请在数据库白名单列表中添加独享数据集成资源组绑定的交换机网段。获取独享绑定的交换机网段信息如下：

在Dat aWorks控制台独享资源组页签下，单击目标独享数据集成资源组后的网络设置，查看交换机网段并将其添加至数据库白名单列表中。



- 如果您使用独享数据集成资源组走公网同步数据，请在数据库白名单列表中添加独享数据集成资源组本身的EIP地址。获取独享数据集成资源组EIP地址如下：

在DataWorks控制台独享资源组页签下，单击数据集资源组后的查看信息，复制对话框中的EIP地址至数据库白名单列表中。



说明 如果您之后对独享数据集资源组进行了扩容操作，请检查此处待添加的EIP是否有变动，为了避免任务执行出错，请在扩容后第一时间更新数据库添加的白名单。

如何将资源组与工作空间进行绑定和解绑？

请在DataWorks控制台资源列表进行操作。



说明 需要有工作空间管理员权限才可以进行此操作。

7.2. 自定义资源组

本文将为您介绍自定义资源组的相关问题。

- 如何安装监控？
- 如何处理添加自定义调度资源服务器失败的情况？
- 如何处理服务器突然终止服务且重启失败的情况？
- 如何处理自定义调度资源下的任务长期等待资源的情况？
- 如何暂停和初始化Agent服务？
- 重启ECS后，如何使Agent自行恢复服务？
- 如何处理无法启动自定义资源组的情况？
- 自定义资源组有哪些应用场景？
- 自定义资源组有哪些使用限制？
- 安装自定义资源组后，需要查看哪些配置？
- 如何监控Agent进程的运行状态？

- DataWorks有哪些调度分类？
- 如何使用自定义资源组？
- 如何处理添加ECS资源组时报错的情况？
- 如何处理添加自定义资源组后状态不可用的情况？
- 如何处理ECS状态正常，但Shell任务执行失败的情况？
- 如何处理查看DataWorks运行日志，找不到对应文件的情况？
- 如何处理自定义资源组下任务执行OOM，无法分配内存给作业线程的情况？
- 如何处理数据集成自定义资源组DataX组件log4j问题？

如何安装监控？

如果使用自定义调度资源组出现问题，需要首先确认Agent是否安装监控。操作如下所示：

1. 登录机器，切换至root账号下。
2. 执行下述命令。

```
wget https://di-agent-image-cn-shanghai.oss-cn-shanghai.aliyuncs.com/install_monitor.sh  
-O install_monitor.sh
```

3. 执行下述脚本。

```
sh install_monitor.sh
```

如何处理添加自定义调度资源服务器失败的情况？

出现添加自定义调度资源服务器失败，服务状态一直显示为停止，可能有以下原因：

- 界面注册的机器名称和实际的机器名称不一致。

确认方法如下：

- 经典网络下（仅华东2支持），需要确保填写的主机名称和IP是在ECS上执行命令 `hostname` 和 `hostname -i` 返回的名称和IP。

 **说明** 请注意是否修改过主机名称。如果修改过主机名称，请进入 `/etc/hosts` 查看是否绑定对应的host，确认绑定后，使用绑定的hosts到界面注册使用。

- 专有网络下，请确保填写的是在ECS上执行命令 `dmidecode | grep UUID` 返回的值。

 **说明**

- 如果没有安装dmidecode，请首先进行安装。
- 不同版本的dmidecode返回的UUID区分大小写。
- 机器名称需要区分大小写。

如果确认是该原因导致的问题，请根据下述方法解决：

- i. 删除原来注册的服务器。
- ii. 填写正确的主机名称（UUID）和IP，重新注册服务器。

- 初始化命令使用错误。

确认方法如下：

- i. 登录ECS执行下述命令。

```
cat /home/admin/alisatasknode/target/alisatasknode/conf/config.properties | grep driver
```

- ii. 登录DataWorks控制台。
- iii. 在左侧导航栏，单击资源组列表。
- iv. 单击自定义资源组页签。
- v. 单击相应资源组后的服务器初始化。
- vi. 对照ECS执行的命令结果和服务器初始化窗口显示的username是否一致。

如果确认是该原因导致的问题，请根据页面正确的初始化语句，重新进行初始化即可解决。

说明

- 在界面注册好服务器后，您可以在资源组列表 > 自定义资源组页面进行服务器初始化操作。
不同资源组的初始化命令不同，且不可互相使用。
- 请务必完全复制服务器初始化对话框中的命令，并按照顺序执行。
- 专有网络的机器需要执行经典网络的初始化语句。

- 时间与北京时间相差5分钟以上。

确认方法如下：

- i. 登录ECS。
- ii. 执行命令 `date`，查看返回的时间是否与北京时间相差5分钟以上。

如果确认是该原因导致的问题，请在非业务影响的情况下，同步ECS时间至北京时间时区。

- 目录权限点有误。

确认方法如下：

- i. 登录ECS，执行命令 `ps -ef | grep zoo | grep -v cdp`。
- ii. 确认返回一个进程的是否是admin账号。

如果是admin账号，请继续确认 `/home/admin/alisatasknode` 的目录及子目录下是否都是admin权限。

如果有root权限，请根据下述操作进行修复：

- i. 切换至root账号下，执行 `chown admin:admin /home/admin -R`。
- ii. 切换至admin账号下，执行命令 `/home/admin/alisatasknode/target/alisatasknode/bin/serverctl restart` 重启Agent。

- 执行 `install.sh` 时报错。

确认方法如下：

- i. 执行命令 `install.sh`。
- ii. 查看 `/home/admin/alisatasknode/logs` 下是否有日志文件生成。如果没有，则说明没有安装成功。

如果确认是该原因导致的问题，解决方法如下：

- i. 确认当前ECS的OS是否是Cent OS5、Cent OS6或Cent OS7版本。如果不是，请切换为该类版本并重新初始化。
- ii. 执行命令 `执行/opt/taobao/java/bin/java -V`，确认JDK版本是否返回1.8版本。
- iii. 执行命令 `ls -al /opt/taobao`，确认该目录权限是否为admin。如果不是，请执行 `chown admin:admin /opt/taobao -R`，并重新执行初始化语句。

如何处理服务器突然终止服务且重启失败的情况？

服务器使用了一段时间，出现突然终止服务，重启无效果或重启失败的情况，可能有以下原因：

- 不同账户启动过Agent，导致目录权限不一致。

确认方法如下：

- i. 登录ECS，切换至root账号。
- ii. 执行命令 `ps -ef | grep zoo | grep -v cdp`。

如果返回两个进程，则说明是该原因导致，解决方法如下：

- i. 登录ECS，对上述 `ps` 出的两个进程执行 `kill -9`。
- ii. 执行命令 `chown admin:admin /home/admin/ -R`。
- iii. 切换至admin账号。
- iv. 执行命令 `/home/admin/alihatasknode/target/alihatasknode/bin/serverctl restart`，重启Agent。

- 句柄数被过多占用。

确认方法如下：

- 登录ECS，执行命令 `grep "temporarily unavailable" /home/admin/alihatasknode/logs/alihatasknode.log`。如果有结果返回，则说明有问题。
- 重启Agent失败，报错为 `Caused by: java.io.IOException: error=11, Resource temporarily unavailable`。

如果确认是该原因导致的问题，请根据下述方法解决：

- i. 切换至root账号下，执行命令 `ps -ef | grep zoo | grep -v cdp`。
- ii. 对上述 `ps` 出的进程全部执行 `kill -9` 命令。
- iii. 执行命令 `chown admin:admin /home/admin/ -R`。
- iv. 切换至admin账号。
- v. 执行命令 `/home/admin/alihatasknode/target/alihatasknode/bin/serverctl restart`，重启Agent。

- 专有网络的ECS的UUID发生变化。

- i. 登录ECS，执行命令 `dmidecode | grep UUID`，确认返回的UUID是否全部为小写字母（假设原来是大写）。
- ii. 将返回的结果和资源组列表 > 自定义资源组页面服务器管理对话框中的结果进行对比。

如果确认是该原因导致的问题，请在自定义资源组页面删除原来大写UUID的服务器，重新添加小写UUID的服务器。

说明 如果出现服务器无法删除，报错 `remove node failed, exception: [3006:ERROR_GATEWAY_EXIST_TASKS]:gateway tasks not empty` 的情况，请记录服务器所在区域，并复制报错内容，[提交工单](#)进行咨询。

如何处理自定义调度资源下的任务长期等待资源的情况？

出现使用自定义调度资源运行的任务，长时间等待资源未能运行的情况，可能有以下原因：

- 服务器终止服务。

确认方法如下：

- i. 登录DataWorks控制台。
- ii. 在左侧导航栏，单击资源组列表。
- iii. 单击自定义资源组页签。
- iv. 单击相应服务器后的服务器管理，查看当前服务器的状态是否终止。

如果确认服务器终止服务，请登录ECS，启动Agent。

- 服务器暂停服务。

○ 确认方法如下：

- a. 登录ECS。
- b. 查看 `/home/admin/alisaasknode/logs/alisaasknode_status.log` 日志。

日志会实时显示服务器的状态。如果显示BUSY或HANGUP，说明当前服务器存在大任务进程占用资源的情况。

○ 解决方法如下：

- a. 执行命令 `ps -ef | grep taskexec`，查看执行任务的进程。
- b. 通过日志排查占用资源的任务。

如果发现任务异常，请进入DataWorks页面终止任务的运行。等待两分钟后，服务器即可自动恢复服务。

- 服务器Agent异常。

确认方法如下：

- 执行命令 `df -h`，查看磁盘是否存在100%的使用率。
- 查看CPU、内存等指标是否过高。

如果确认是该原因导致的问题，请在机器本身的异常处理完成后，重启Agent。

如何暂停和初始化Agent服务？

如果您需要暂停Agent服务，请根据以下两种情况进行选择：

- 如果是在资源组列表 > 自定义资源组页面添加的Agent，您可以单击相应服务器后的服务器管理，在管理服务器对话框中单击暂停即可。
- 如果是在数据集成 > 资源组页面添加的Agent，则不支持暂停Agent服务，请[提交工单](#)进行咨询。

如果您需要初始化Agent服务，操作如下：

1. 切换至root下，执行 `ps -ef | grep zoo | grep alisa` 命令。

2. 执行 `kill -9` 命令，结束 `ps` 获得的进程的运行。
3. 删除 `/home/admin/alisatasknode` 目录。
4. 在新的空目录下执行 `install.sh` 命令。

 说明 请确保下载的为相应区域的install.sh文件。

重启ECS后，如何使Agent自行恢复服务？

重启ECS后，您可以通过下述操作，使Agent自行恢复服务：

1. 登录机器，切换至root账号。
2. 执行命令：

```
wget https://di-agent-image-cn-shanghai.oss-cn-shanghai.aliyuncs.com/install_monitor.sh
-O install_monitor.sh
```

3. 执行脚本：`sh install_monitor.sh`。

如何处理无法启动自定义资源组的情况？

出现使用经典网络的方式注册专有网络的机器后，无法启动自定义资源组的情况，您可以执行命令 `tail -f /home/admin/alisatasknode/logs/alisatasknode.log`，查看运行日志确定问题原因。

以专有网络UUID的方式注册的服务器名称，可能是初始化语句有误，需要进行修改。

解决方法如下：

1. 登录DataWorks控制台。
2. 在左侧导航栏，单击资源组列表。
3. 单击自定义资源组页签。
4. 单击相应服务器后的服务器初始化。
5. 根据服务器初始化对话框中提示进行操作。

 说明 其中step3命令中的 `enable_uuid=false`，需要修改为 `enable_uuid=true`。

自定义资源组有哪些应用场景？

- 保证运行资源：由于集群共享公共资源组，会存在水位变高导致任务长时间等待的情况。如果您对任务有较高的资源使用需求，可以使用自定义资源组来自建任务运行集群。
- 连通网络：由于公共资源组无法连通VPC环境下的数据库，您可以使用自定义资源组进行网络连通。
- 用于调度资源组：调度槽位资源紧张的情况下，您可以使用自定义资源组。
- 提升并发能力：公共资源组的运行槽位有限，您可以通过自定义资源组扩大槽位资源，允许更多的并发任务同时调度运行。

自定义资源组有哪些使用限制？

- 一台ECS只能注册到一个自定义资源组下，一个自定义资源组可以添加多个ECS。
- 经典网络和专有网络注册的区别为：经典网络是主机名称，专有网络是UUID。
- 一个自定义资源中只允许存在一种网络类型。
- 不支持运行手动任务实例。

- ECS需要具备公网访问能力，ECS可以配置公网IP、EIP、NAT网关SNAT。

安装自定义资源组后，需要查看哪些配置？

通过DataWorks界面引导，完成自定义资源组的安装后，请登录ECS查看Agent插件的以下信息：

- 默认安装路径：`/home/admin/`，默认路径下通常会有以下目录信息：
 - `alisatasknode`：Agent有关配置和命令所在目录。
 - `datax`和`datax-on-flume`：数据同步插件库和配置所在目录。

```
[root@iZ...Z ~]# cd /home/admin/
[root@iZ...Z admin]# ls
alisatasknode  datax3  datax-on-flume
```

- Agent有关命令：目前支持对Agent进程进行stop、start和restart等命令操作。

```
/home/admin/alisatasknode/target/alisatasknode/bin/serverctl start/stop/restart
```

- 运行日志：Agent运行日志有以下两个存放路径：
 - `/home/admin/alisatasknode/taskinfo/`：存放Shell脚本运行的日志信息，和DataWorks节点运行日志页面中查看的结果一致。
 - `/home/admin/alisatasknode/logs/`：`alisatasknode.log`日志文件中存放的是Agent插件的运行信息，例如接收到的任务运行操作、Agent心跳状态等。
 - `/home/admin/datax3/log/`：存放数据集成任务的详细运行日志，如果任务执行失败，请查看该部分日志查找原因。

如何监控Agent进程的运行状态？

您可以通过下述方法监控Agent进程的运行状态，在监控到Agent进程退出后，可以及时进行恢复：

1. root用户登录ECS机器。
2. 执行命令

```
wget https://di-agent-image-cn-shanghai.oss-cn-shanghai.aliyuncs.com/install_monitor.sh
-O install_monitor.sh
```

3. 执行命令 `sh install_monitor.sh`，监控日志默认存放在`/home/admin/alisatasknode/monitor/monitor.log`中。

DataWorks有哪些调度分类？

自定义资源组在DataWorks调度体系中使用，当前DataWorks调度体系分为一级调度资源和二级运行资源：

- 一级调度资源：进入运维中心页面，单击左侧导航栏中的周期任务运维 > 周期实例。右键单击相应实例的DAG图，选择查看更多详情，即可在属性面板查看一级调度资源。
- 二级运行资源：进入数据集成 > 自定义资源组页面，即可查看二级任务运行资源。

如何使用自定义资源组？

- 配置一级调度资源

登录DataWorks控制台，单击左侧导航栏中的资源列表，进入自定义资源组页面进行新建。

 说明 在该页面创建的调度资源适用于Shell任务，配置的是一级调度资源。

- 配置二级运行资源
 - i. 登录DataWorks控制台，进入数据集成 > 自定义资源组页面，单击右上角的新增自定义资源组进行新建，详情请参见[新增和使用自定义数据集成资源组](#)。

 说明 该页面添加的资源组仅用于数据同步任务，配置的是二级运行资源。

- ii. 新建完成后，能且只能在数据开发页面配置任务时，在右侧导航栏的数据集成资源组配置中进行选择。

如何处理添加ECS资源组时报错的情况？

添加ECS资源组时，报错 `gateway already exists` 的解决方法如下：

1. 报错提示对应ECS已经在gateway中注册过，因为一个ECS只能添加到一个自定义资源组下，所以需要您在资源列表和数据集成 > 自定义资源组页面，查看是否存在同名或同UUID的自定义资源组。
2. 如果工作空间中没有发现对应的自定义资源组，则需要收集request ID信息进行咨询。

如何处理添加自定义资源组后状态不可用的情况？

- 查看 `alisatasknode.log` 日志，确认是否有心跳上报302的情况。如果上报心跳302，请排查下述问题：
 - 查看UUID是否一致。对比自定义资源组页面的UUID信息和ECS上执行命令 `dmidecode|grep UUID` 的结果是否一致。

 说明 此处的UUID区分大小写。

- 如果UUID不一致，需要填写正确的UUID并重新安装Agent。

 说明 `dmidecode`命令在3.0.5版本及之前版本会是大写的方式显示UUID，如果升级到3.1.2或以上版本，则会小写显示UUID，此时会导致心跳异常，需要重新安装Agent进行恢复。

- 确认 `config.properties`中配置的用户名及密码是否和自定义资源组添加界面一致，如果不一致，请参见Agent插件安装界面给出的命令重新安装Agent。
 - 如果UUID和密码信息正确，请查看 `config.properties`中的字段 `node.uuid.enable`。针对VPC类型的ECS，该字段的值需要为true。如果VPC类型下 `node.uuid.enable`的值为false，则修改为true，重启Agent进程即可。
- 查看 `alisatasknode.log` 日志，确认是否存在 `connection timeout` 的信息。如果存在，则可以进行如下处理：
 - i. 查看ECS是否有公网能力，例如公网IP、EIP、NAT网关SNAT IP，请执行 `ping www.taobao.com` 确认是否可以连通。
 - ii. 如果ECS有公网能力，请查看ECS的安全组配置中内网出方向或公网出方向是否有访问限制。如果有访问限制，需要对gateway的IP和端口放行。

如何处理ECS状态正常，但Shell任务执行失败的情况？

- 查看 `alisatasknode.log` 日志中具体的报错信息，可以根据T3_0699121848关键字进行搜索。
- 登录ECS，切换至admin用户下，执行 `python -V` 命令查看Python版本是否为2.7或2.6版本。

Agent当前支持的是Python2.7或2.6版本，通常Python版本不对会导致replace user hive conf error的错误。

```
2019-01-17 19:18:01,790 INFO [pool-2-thread-1] [FileAgent.java:264] [] - 正式文件生成成功。正式文件路径: /home/admin/alisatasknode/taskinfo//20190117/phoenix/19/17/57/k0613hp0osh2cr
2019-01-17 19:18:01,792 INFO [pool-2-thread-1] [CodeFormatUtils.java:352] [1] - [T3_069911838]开始替换代码中的参数变量
2019-01-17 19:18:01,792 INFO [pool-2-thread-1] [CodeFormatUtils.java:362] [1] - [T3_069911838]替换代码中的参数变量完毕
2019-01-17 19:18:01,792 ERROR [pool-2-thread-1] [SetUserConf.java:25] [] - Replace user hive conf error.Params from envmap is not correct. SET_USER_CONF:null IDE_HIVE_USER_REGION:null
2019-01-17 19:18:01,802 INFO [pool-2-thread-1] [NodeLocalLog.java:53] [1] - AlisaNode主进程写日志结束，关闭文件。文件路径:/home/admin/alisatasknode/taskinfo//20190117/phoenix/19/17/57
2019-01-17 19:18:30,117 INFO [pool-11-thread-1] [AlisaNodeWrapperListener.java:138] [1] - 任务[T3_069911838]The task process was abnormal exit, system set task rerun!!!
```

如何处理查看DataWorks运行日志，找不到对应文件的情况？

```
2019-02-01 15:06:28 INFO ALISA_TASK_ID=golf9rw44bohm1jtz7u//
2019-02-01 15:06:28 INFO ALISA_TASK_EXEC_TARGET=golf9rw44bohm1jtz7u//
2019-02-01 15:06:28 INFO ALISA_TASK_PRIORITY=1:
2019-02-01 15:06:28 INFO --- Invoking Shell command line now ---
2019-02-01 15:06:28 INFO =====
/home/admin/alisatasknode/taskinfo//20190201/phoenix/15/06/23/7/golf9rw44bohm1jtz7u//
那个文件或目录
2019-02-01 15:06:29 INFO =====
2019-02-01 15:06:29 INFO Exit code of the Shell command 27
```

请登录ECS并切换至admin用户，执行命令 `sh -x脚本名`，确认是否可以正常执行，根据报错信息进行调试。

如何处理自定义资源组下任务执行OOM，无法分配内存给作业线程的情况？

自定义资源组下任务执行OOM，下图为运行报错日志，提示无法分配内存给作业线程。

```
2019-03-26 15:12:41.063 [job-63992276] INFO JobContainer - Running by local Mode.
2019-03-26 15:12:41.073 [taskGroup-0] INFO TaskGroupContainer - taskGroupId=[0] start [1] channels for [1] tasks.
2019-03-26 15:12:41.080 [taskGroup-0] INFO Channel - Channel set byte_speed_limit to -1, No bps activated.
2019-03-26 15:12:41.080 [taskGroup-0] INFO Channel - Channel set record_speed_limit to -1, No tps activated.
Exception in thread "taskGroup-0" com.alibaba.datax.common.exception.DataXException: Code:[Framework-02], Description:[DataX引擎运行过程出错，具体原因请
at java.lang.Thread.start0(Native Method)
at java.lang.Thread.start(Thread.java:714)
at com.alibaba.datax.core.taskgroup.TaskGroupContainer$TaskExecutor.doStart(TaskGroupContainer.java:452)
at com.alibaba.datax.core.taskgroup.TaskGroupContainer.start(TaskGroupContainer.java:244)
at com.alibaba.datax.core.taskgroup.runner.TaskGroupContainerRunner.run(TaskGroupContainerRunner.java:24)
at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1142)
at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:617)
at java.lang.Thread.run(Thread.java:745)
java.lang.OutOfMemoryError: unable to create new native thread
at java.lang.Thread.start0(Native Method)
at java.lang.Thread.start(Thread.java:714)
```

自定义资源组创建时设置的内存数决定了资源组可以提供的槽位能力。资源组内的系统进程和Agent进程会占用一部分内存，不能将ECS实例所有内存都用于槽位资源，会导致高并发下某些作业OOM。

建议您调小自定义资源组下的内存数设置，可以考虑预留2 GB的空间给系统和Agent进程使用。如果还有其它进程，则需要预留更多的内存。

如何处理数据集成自定义资源组DataX组件log4j问题？

当您使用DataWorks数据集成自定义资源组时，如果使用DataX相关组件涉及log4j问题，可采用以下步骤进行修复：

1. 下载log4j-core:
 - i. 下载 *apache log4j*，详细地址请参见[apache log4j下载地址](#)。
 - ii. 将 *apache log4j* 文件解压 (`tar zxvf apache-log4j-2.17.1-bin.tar.gz`)，在目录中找到 `log4j-core-2.17.1.jar`。
2. 修复log4j问题:
 - i. 上传jar包至临时目录。
将下载的新版 `log4j-core-2.17.1.jar` 上传至ECS的临时目录。例如， `/tmp/`。
 - ii. 进入工具包安装目录。
在数据集成自定义资源组机器上，使用 `cd /home/admin/datax3/` 语句，进入 `datax3` 工具包的安装目录 `/home/admin/datax3/`。

iii. 确认待替换jar包。

使用 `find . -name "*log4j-core*" -exec ls {} \;` 语句查询 `log4j-core` 的jar包列表。

待替换的jar包列表如下图所示。

```
[root@iZbp1ds61ad4fgi23j3ys0Z ~]# cd /home/admin/datax3/
[root@iZbp1ds61ad4fgi23j3ys0Z datax3]# find . -name "*log4j-core*" -exec ls {} \;
./plugin/reader/otsreader/libs/log4j-core-2.0.2.jar
./plugin/reader/hivereader/libs/log4j-core-2.6.2.jar
./plugin/reader/hdfsreader/libs/log4j-core-2.6.2.jar
./plugin/reader/otsstreamreader/libs/log4j-core-2.0.2.jar
./plugin/reader/ossreader/libs/log4j-core-2.6.2.jar
./plugin/writer/otswriter/libs/log4j-core-2.0.2.jar
./plugin/writer/hivewriter/libs/log4j-core-2.6.2.jar
[root@iZbp1ds61ad4fgi23j3ys0Z datax3]#
```

iv. 备份原 `log4j-core jar` 包。

使用 `find . -name "*log4j-core*" -exec mv {} /tmp \;` 语句，备份原 `log4j-core jar` 包至 `/tmp` 目录。

v. 替换新 `log4j-core jar` 包。

将新版 `/tmp/log4j-core-2.17.1.jar` 文件复制至被移除文件（即待替换的jar包列表）所在目录。示例语句如下。

```
cp /tmp/log4j-core-2.17.1.jar ./plugin/writer/hivewriter/libs/
cp /tmp/log4j-core-2.17.1.jar ./plugin/writer/otswriter/libs/
cp /tmp/log4j-core-2.17.1.jar ./plugin/reader/otsstreamreader/libs/
cp /tmp/log4j-core-2.17.1.jar ./plugin/reader/otsreader/libs/
cp /tmp/log4j-core-2.17.1.jar ./plugin/reader/ossreader/libs/
cp /tmp/log4j-core-2.17.1.jar ./plugin/reader/hivereader/libs/
cp /tmp/log4j-core-2.17.1.jar ./plugin/reader/hdfsreader/libs/
```

vi. 检查替换结果。

使用 `find . -name "*log4j-core*" -exec ls {} \;` 语句，检查新版 `log4j-core jar` 包的替换情况。

```
[admin@ /home/admin/datax3]
$find . -name "*log4j-core*" -exec ls {} \;
./plugin/writer/hivewriter/libs/log4j-core-2.17.1.jar
./plugin/writer/otswriter/libs/log4j-core-2.17.1.jar
./plugin/reader/otsstreamreader/libs/log4j-core-2.17.1.jar
./plugin/reader/otsreader/libs/log4j-core-2.17.1.jar
./plugin/reader/ossreader/libs/log4j-core-2.17.1.jar
./plugin/reader/hivereader/libs/log4j-core-2.17.1.jar
./plugin/reader/hdfsreader/libs/log4j-core-2.17.1.jar

[admin@ /home/admin/datax3]
$
```

② 说明 完成替换后，请您留意调度任务的执行情况，如有问题请[提交工单](#)及时反馈。

如果您需要回滚旧版jar包，请将备份的jar文件恢复到原有lib目录，并移除新版 `log4j-core-2.17.1.jar` 即可。命令语句如下。

```
rm -rf ./plugin/reader/otsreader/libs/log4j-core-2.17.1.jar
rm -rf ./plugin/reader/hivereader/libs/log4j-core-2.17.1.jar
rm -rf ./plugin/reader/hdfsreader/libs/log4j-core-2.17.1.jar
rm -rf ./plugin/reader/otsstreamreader/libs/log4j-core-2.17.1.jar
rm -rf ./plugin/reader/ossreader/libs/log4j-core-2.17.1.jar
rm -rf ./plugin/writer/otswriter/libs/log4j-core-2.17.1.jar
rm -rf ./plugin/writer/hivewriter/libs/log4j-core-2.17.1.jar
cp /tmp/log4j-core-2.0.2.jar ./plugin/reader/otsreader/libs/
cp /tmp/log4j-core-2.6.2.jar ./plugin/reader/hivereader/libs/
cp /tmp/log4j-core-2.6.2.jar ./plugin/reader/hdfsreader/libs/
cp /tmp/log4j-core-2.0.2.jar ./plugin/reader/otsstreamreader/libs/
cp /tmp/log4j-core-2.6.2.jar ./plugin/reader/ossreader/libs/
cp /tmp/log4j-core-2.0.2.jar ./plugin/writer/otswriter/libs/
cp /tmp/log4j-core-2.6.2.jar ./plugin/writer/hivewriter/libs/
```

8. 数据源

8.1. 数据源权限管理

本文为您汇总数据源权限管理相关的问题。

- 哪些角色可以管理数据源的权限？
- 是否支持再次分享已被分享的数据源？
- 数据源的创建者是否有权控制该数据源？
- 什么是私有模式的数据源？
- 哪些用户能够收回数据源的分享权限？
- 如何检查数据源分享关系的移除、关联情况？
- 分享数据源后，其资源组的连通状态是否会被同步分享？
- 分享数据源后，引用过该数据源的任务关联信息是否会被同步分享？

哪些角色可以管理数据源的权限？

租户所有者、租户管理员、空间管理员和项目所有者等角色，可以对数据源进行权限管理。

是否支持再次分享已被分享的数据源？

DataWorks不支持二次权限分享。

数据源的创建者是否有权控制该数据源？

数据源权限不受Owner粒度权限管理的控制，所以创建者无权控制该数据源。

租户数据源的Operator机制完善，Owner信息不完善。Owner可能被移除，降为开发角色等低权限角色。为保障数据源的安全，仅允许具备相应权限的角色控制数据源。

什么是私有模式的数据源？

如果一个数据源仅被分享给自己，该数据源为私有模式，仅自己可以查看和使用该数据源。

其他人即使具备数据源的控制权限，也无法查看该数据源。

哪些用户能够收回数据源的分享权限？

- 具备数据源权限管理权限的用户，能够在数据源权限管理的入口，调整数据源的分享情况。
- 如果以可编辑模式分享的权限，被分享人可以执行取消分享操作。
- 如果以只读模式分享的权限，被分享人不可以执行取消分享操作。

如何检查数据源分享关系的移除、关联情况？

- 一个数据源分享给目标用户后，源数据源可以控制分享关系，随时解除分享，且无需检查已分享数据源的任务依赖情况。
- 一个数据源以可编辑模式分享给目标用户后，如果目标用户执行取消分享操作，会检查该分享数据源的任务依赖情况。如果有任务依赖，您需要先删除任务再删除数据源。

分享数据源后，其资源组的连通状态是否会被同步分享？

不会。因为被分享的数据源是一个新数据源，被分享者具备的资源组并非和原始资源组完全一致，需要单独测试连通性。

分享数据源后，引用过该数据源的任务关联信息是否会被同步分享？

不会。因为被分享的数据源是一个新数据源，与源数据源关联的任务无任何关系。

8.2. 添加数据源典型问题场景

DataWorks添加数据源的典型问题包括连通性问题、参数问题和权限问题。

连通性问题

② 说明

- 如果您使用的是RDS数据源，建议您首先为RDS添加白名单，详情请参见[添加白名单](#)。
- 如果您使用的是ECS上自建数据库，建议您首先为ECS添加安全组，详情请参见[场景示例：ECS自建数据库的安全组配置](#)。

连通性问题主要体现为测试连通性失败：

- 问题现象：添加MySQL数据源时，网络类型选择为经典网络，单击**测试连通性**时失败，报错信息为 `测试连接失败，测试数据源联通性失败，连接数据库失败，数据库连接串...异常消息：Communications link failure. The last packet sent successfully to the server was 0 milliseconds ago.The driver has not received any packets from the server`。

解决方法：出现上述报错通常都是网络连通性问题导致。建议检查您的网络是否可达、防火墙是否对该IP或端口有相关限制，以及安全组是否已配置对IP或端口放行。

- 问题现象：添加阿里云MongoDB数据源，测试数据源连通性失败，报错如下。

```
error message: Timed out after 5000 ms while waiting for a server that matches ReadPreferenceServerSelector{readPreference=primary}. Client view of cluster state is {type=UNKNOWN, servers=[..] error with code: PROJECT_DATASOURCE_CONN_ERROR
```

解决方法：处理此类问题时，首先需要确定您的DataWorks工作空间所处地域。使用阿里云MongoDB，需要确定网络类型是否为VPC。VPC环境下MongoDB不支持数据连通性测试（使用方案一可以避免该问题）。

VPC环境下阿里云MongoDB数据同步有以下两种方案：

- 方案一：通过公网进行数据同步
 - a. 数据源配置时，数据源类型选择**连接串模式**。
 - b. VPC环境下，您的MongoDB需要开通公网访问。
 - c. 在MongoDB上放行相关白名单IP，详情请参见[添加白名单](#)。
 - d. 进行数据连通性测试。
- 方案二：配置自定义资源组，从内网进行数据同步
 - a. 准备一台和MongoDB同区域、同网络的ECS作为调度资源，详情请参见[新增和使用自定义数据集成资源组](#)。
 - b. 将这台ECS的IP加入MongoDB的白名单或者安全组。
 - c. 数据源测试连通时直接确定保存（不支持测试连通性）。
 - d. 修改资源组为自定义调度资源，测试运行。

 说明 请务必添加相应的白名单。

- 问题现象：添加自建MongoDB数据源，测试数据源连通性失败。

解决方法：

- 数据源配置时，数据源类型选择**连接串模式**。
- 如果是VPC环境下ECS上自建的MongoDB，需要开通公网访问。
- 确认网络和端口之间是否能够连通，检查ECS的防火墙和安全组设置。
- 确认自建的数据源涉及的安全访问限制、权限的限制和远程登录的情况。
- 确认访问地址 `host:port`、数据库名和用户名是否填写正确。

 说明

添加MongoDB数据源时，使用的用户名必须是用户需要同步的这张表所在的数据库创建的用户名，不能用root。

例如需要导入name表，name表在test库，则此处数据库名称填写为test。

用户名为指定数据库中创建的用户名，不要使用root。例如之前指定的是test库，则用户名需使用test数据库中创建的账户。

- 问题现象：VPC环境下添加Redis数据源，测试数据源连通性失败，报错如下。



解决方法：Redis添加数据源时如果没有公网IP，需要保证数据源和DataWorks工作空间地域一致，通过新增调度资源完成数据源的打通。

- 问题现象：添加MongoDB数据源，已经配置白名单，测试数据源连通性仍然失败，报错如下。

```
error message: Timed out after 5000 ms while waiting for a server that matches ReadPreferenceServerSelector{readPreference=primary}
```

解决方法：VPC网络的MongoDB数据源和DataWorks的默认资源组在内网上是不通的，所以无法直接进行同步任务，需要通过公网或者自定义资源组的方式进行连通。

- 问题现象：Docker中安装的MySQL如何添加到数据源？

解决方法：Docker中安装的MySQL直接用服务器的公网IP组成的JDBC地址无法连接，连通性测试无法通过。您需要将MySQL的端口映射到宿主机上，使用映射出的端口链接。

- 问题现象：配置Redis数据源失败，测试数据源连通性失败报错如下。

```
error message: java.net.SocketTimeoutException: connect timed out
```

解决方法：目前DataWorks不支持Redis通过内网添加数据源，建议您为Redis数据源开通公网访问能力。数据源配置时，选择**连接串模式**，通过公网连接。

- 问题现象：新增阿里云RDS数据源时，测试连通性不通。

解决方法：

- i. 当RDS数据源测试连通性不通时，需要到自己的RDS上添加数据同步机器IP白名单，详情请参见[添加白名单](#)。

 **说明** 如果使用自定义资源组调度RDS的数据同步任务，必须把自定义资源组的机器IP也加到RDS的白名单中。

- ii. 确保添加的信息正确：RDS实例ID和RDS实例主账号ID、用户名、密码数据库名必须确保正确。

- 问题现象：新增自建ECS中的MySQL数据源时，数据源测试连通性不通。

解决方法：

- i. 确认网络和端口之间是否能连通，检查ECS的防火墙以及安全组设置。
- ii. 确认自建的数据库涉及的安全访问限制、权限的限制和远程登录的情况。
- iii. 确认添加的用户名、密码、JDBC URL中的IP地址和端口的信息是否正确。
- iv. 在VPC的环境下购买的ECS，只能用脚本模式运行任务，在添加数据源时测试连通性不能成功。购买ECS后，您可以添加自定义资源，将同步任务下发到相应的资源组运行。

参数问题

- 问题现象：添加MySQL类型数据源时，单击测试连通性，报错如下。

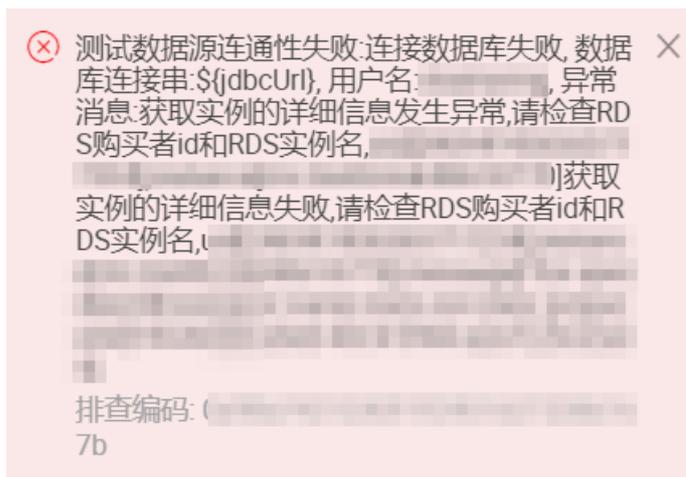
测试连接失败，测试数据源连通性失败，连接数据库失败...异常消息：No suitable driver found for...

解决方法：出现上述情况可能是JDBC URL格式填写错误导致，填写JDBC URL时，请不要在URL中添加空格或任何特殊字符。正确格式为：`jdbc:mysql://ServerIP:Port/Database`。

- 问题现象：使用用户名root添加MongoDB数据源时报错。

解决方法：添加MongoDB数据源时，使用的用户名必须是用户需要同步的表所在的数据库创建的用户名，不能用root。例如需要导入name表，name表在test库，则此处数据库名称填写为test。用户名为指定数据库中创建的用户名，不要使用root。例如之前指定的是test库，则用户名需使用test数据库中创建的账户。

- 问题现象：添加RDS数据源失败，数据库无法连接，报错如下。



解决方法：需要检查填写的UID是否为子账号的UID，此处要填写RDS所属主账号的UID，才可以成功添加数据源。

- 问题现象：添加ODPS默认数据源时，报测试连通性失败。

解决方法：ODPS默认数据源无需添加，默认为odps_fisrt。

- 问题现象：DataWorks的数据源支持HybridDB for PostgreSQL吗？
解决方法：支持，添加时选择关系型数据库PostgreSQL即可。
- 问题现象：没有外网地址的DRDS实例，配置数据源时，是否支持将实例的内网地址映射为自定义的域名？
解决方法：需要严格按照格式来，目前不支持域名映射的方式。
- 问题现象：添加RDS数据源时，为什么白名单已添加，依然报错提示 `user not exist ip white list reference`。
解决方法：出现这种情况通常是由于用户名输入错误，请检查输入的用户名是否正确。

权限问题

- 问题现象：添加AnalyticDB for MySQL数据源时，测试数据连通性报错如下。

```
连接数据库失败，数据库连接串：${jdbcUrl}，用户名：XXXXXX，异常消息：You don't have privilege for connecting database 'dw', userId=RAM$XXX, schemaId=XX
```


解决方法：您需要检查在数据源中填写的子账号是否有AnalyticDB for MySQL的访问权限。
分析型数据库用户基于阿里云账号进行认证，用户建立的数据库属于该用户，且可以授权给其他用户访问其数据库下的表，所以连接的用户需要在AnalyticDB for MySQL上进行授权。
- 问题现象：子账号无权限查看数据源，无法创建数据源，提示您没有权限进行此操作。
解决方法：仅项目管理员权限的RAM子账号才可以增删改数据源。

9.数据治理

- 数据地图
 - 为什么数据地图数据总览页存储量和存储趋势图相差较大?
 - 数据地图血缘展示延迟问题
 - 数据地图新建表搜不到
 - 当前表业务逻辑变更如何通知下游?

• 数据保护伞

数据保护伞为什么有时候查询脱敏有时候查询不脱敏

为什么数据地图数据总览页存储量和存储趋势图相差较大?

存储量是实时的数据，存储趋势图（折线图）是某一时刻的快照，如果您现在的计算量很少（产生的临时表比较少）的话，两个确实会有一些差距。

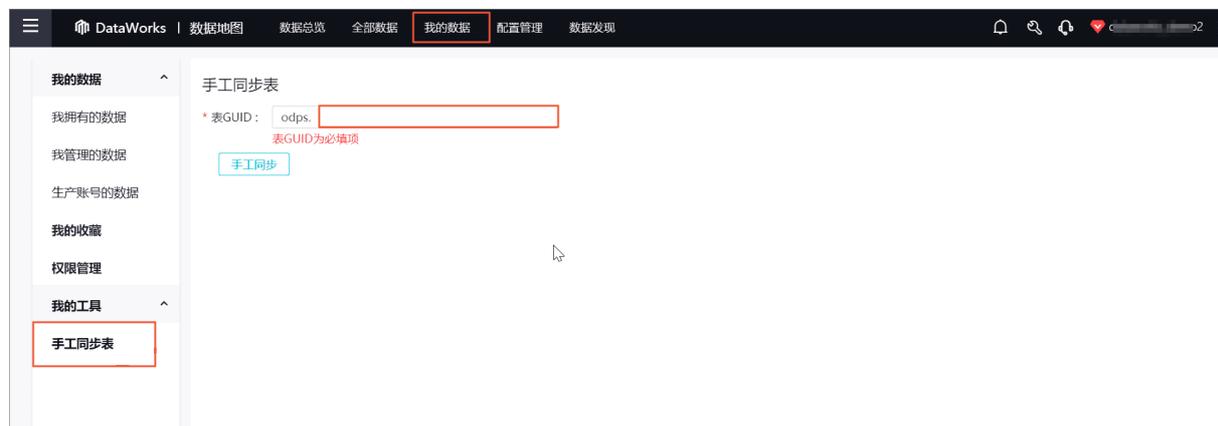
数据地图血缘展示延迟问题

血缘有最少一天的延迟

数据地图新建表搜不到

若数据地图表和表目前状态不一致，请使用数据地图手工同步功能手动同步该表最新状态。

- 表管理数据地图表结构未更新
- 新建表在数据地图找不到
- 数据集成搜不到新建的表



当前表业务逻辑变更如何通知下游?

您可以通过数据地图表详情页的血缘信息 > 影响分析中批量以邮件方式告知下游相关责任人。



数据保护伞为什么有时候查询脱敏有时候查询不脱敏

80%数据符合规则才会脱敏

10. 智能监控

本文为您汇总智能监控的相关问题。

- [在运维中心设置报警后收不到，该如何排查？](#)
- [不需要接收任务的报警，该如何处理？](#)
- [为什么在基线实例页面会显示基线状态为空基线？](#)
- [为什么开启的基线破线未报警？](#)
- [变慢的任务是否可以不报警？](#)
- [为什么未收到出错任务的报警？](#)
- [如何处理夜间收到报警的情况？](#)

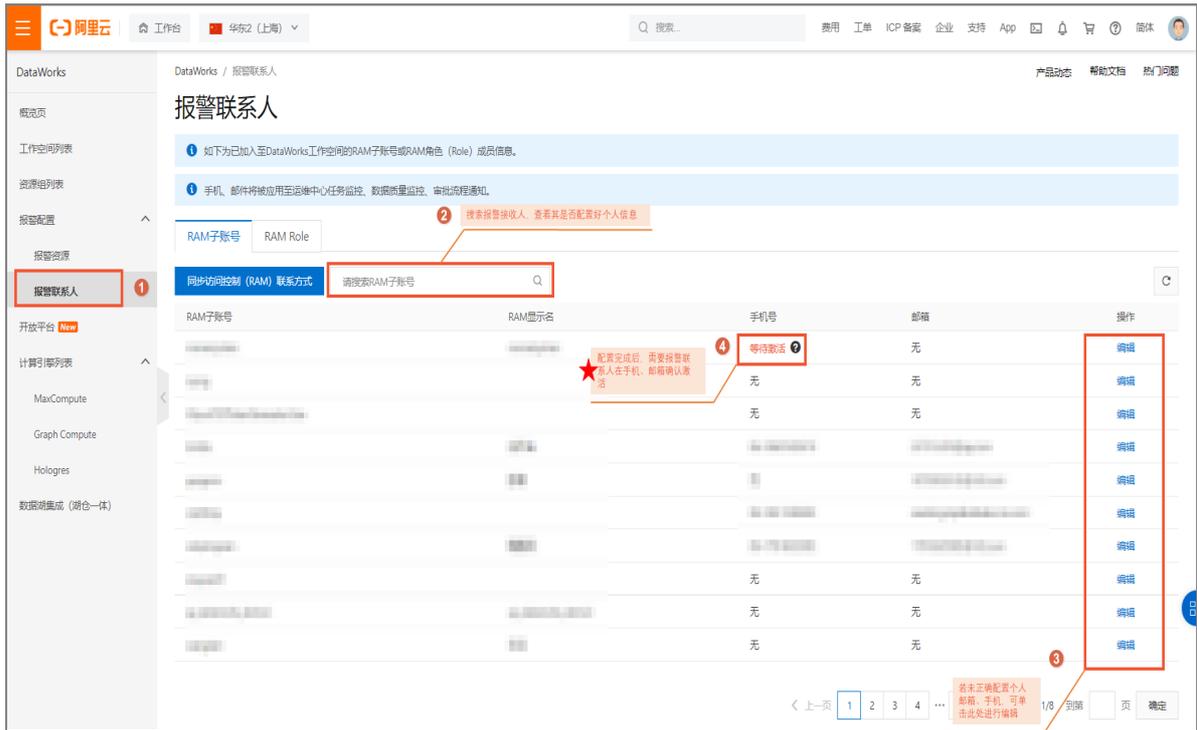
在运维中心设置报警后收不到，该如何排查？

此场景下，我们需要先确认是否有真实触发报警，在真实触发报警但未收到的情况下，我们根据不同的报警类别，将其再细分为两个场景：[配置电话、短信和邮箱报警但收不到，该如何排查](#)，和[配置钉钉群报警后但收不到报警信息，该如何排查](#)，以下为具体排查思路：

- **确认是否有成功触发报警？**
 - 如果是调度任务报警，您可以在[周期实例](#)面板确认实例运行状态，是否可以正常触发报警，关于各个规则触发条件可参考文档[以下文档](#)：
自定义规则报警触发条件可参考：[自定义规则](#)，基线报警触发条件可参考文档：[智能监控](#)。
 - 如果是实时同步任务报警，您可以在[运维中心](#)界面左侧菜单栏[实时任务运维](#)中选择[实时同步任务](#)，查看实时同步任务运行详情。
- **有真实触发报警，但配置的电话、短信和邮箱未收到报警，该如何排查？**

此场景下您需要先了解，如何在DataWorks配置报警联系人的手机、邮箱。

[电话、短信和邮箱报警](#)，将通过DataWorks控制台[报警配置](#)，[报警联系人](#)页面所设置的个人信息来报警。主要配置步骤如下图所示：



如果报警配置后未收到，您可以在报警联系人处确认以下几点：

- 已经在此处配置好个人信息。
- 新增或变更个人信息后，已经在手机、邮箱激活个人手机、邮箱。

说明

- 阿里云主账号及有AliyunDataWorksFullAccess权限的子账号，可以为其他RAM子账号设置个人信息。配置详情可参考文档：[查看和设置报警联系人](#)
- 如果未正确设置报警接收人的手机或邮箱信息，系统会向基本接收管理中产品的欠费、停服、即将释放等信息的消息接收人发送告警信息，导致报警接收人未收到告警短信或邮件。

- 有真实触发报警，但配置钉钉群接收报警未收到报警，该如何排查？

请确认以下几点：

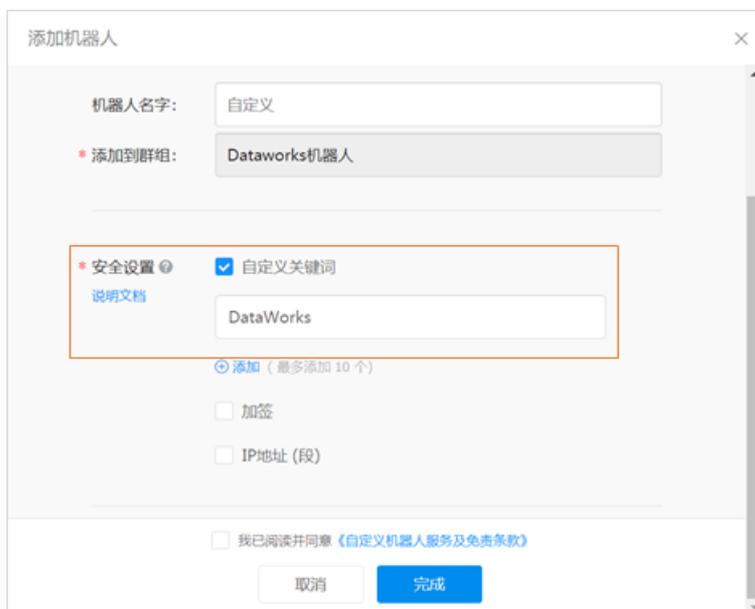
○ 1. 规则配置页面输入的钉钉机器人地址是否正确？

- 调度任务监控（基线监控与自定义规则监控）配置的钉钉机器人WebHook地址是否正确，是否多了空格。
- 实时同步报警配置需要配置的是钉钉群机器人TOKEN信息。



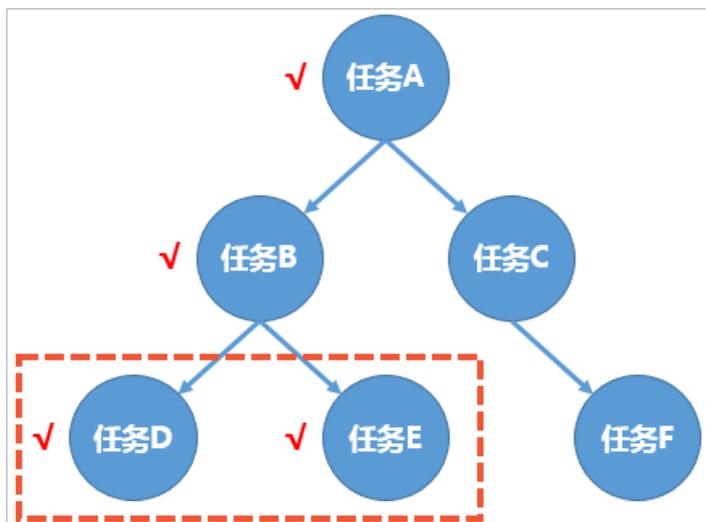
○ 2. 钉钉群机器人配置是否正确？

为保障钉钉群可以正常接收到报警信息，钉钉群机器人配置中、安全设置仅支持添加自定义关键词，并且自定义关键词需要包含DataWorks（注意区分大小写），关于钉钉机器人配置详情可参考文档：[配置钉钉机器人报警](#)。



不需要接收任务的报警，该如何处理？

一条基线纳入监控后，该基线上的任务及基线上任务的上游的所有任务都会被监控。如果基线上的任务或节点的上游节点影响到基线上受保障的任务的数据产出，则智能监控会生成一个事件，事件默认报警给任务责任人。详情请参见[智能监控](#)。



如上图所示，假设整个DataWorks只有6个任务节点，任务D和任务E属于受保障基线上的节点。任务D和任务E及它们所有的上游节点，均会被纳入监控范围。即上图中的任务A、任务B、任务D和任务E出现异常（出错或变慢），也会被智能监控察觉，而任务C和任务F不受监控。

- 如果任务D和任务E本身不需要接收报警，请联系基线责任人，从基线上移除任务D和任务E。
- 任务A和任务B是受保障的任务D和任务E的上游任务，虽然不在基线上，但会影响基线数据的产出。一旦任务A和任务B出错或变慢，便会产生事件报警，默认报警给任务责任人。

如果任务A和任务B无需接收报警，请联系任务责任人删除任务D和任务E对它们的依赖。

为什么在基线实例页面会显示基线状态为空基线？

以下几个将会导致空基线产生：

- 场景一：节点只能存在一条基线上。当任务从一条基线添加到另一条基线，实际上是节点在基线上的迁移，当一条基线上一个任务都没有的时候（原基线上的任务全都被移走），这条开启状态的基线会成为空基线。
- 场景二：基线新建当天会显示空基线，基线开启第二天正常产生基线实例。
- 场景三：小时基线对应的实例周期填写错误。

说明 假设小时任务定时每天6点，18点执行，那么此任务周期数为2，在小时基线配置时，任务第一个周期为6点，第二个周期为18点。

为什么开启的基线破线未报警？

基线开关开启的基线监控是针对任务的。如果所有的任务都正常，即使破线也不会报警，因为所有的任务都运行正常，无法判断出哪个任务出错。

任务都正常但基线仍破线的原因，通常有以下原因：

- 设置的基线时间不合理。
- 任务的依赖有问题，即使基线破线也不报警。

变慢的任务是否可以不报警？

任务变慢报警一定要满足以下两个条件：

- 任务处于重要的基线上游。
- 任务和往常比较，确实存在变慢的情况。

您可以在事件管理页面查看下游基线信息，并和下游监控任务的基线方确认任务变慢的影响程度。

- 如果任务变慢的影响不大，可以选择忽略。
- 如果确认要为下游方负责，请维护好任务。

为什么未收到出错任务的报警？

并不是所有任务出错后都会报警，任务需要满足下述条件之一，才会在出错后进行报警：

- 处于某条开启状态的基线的上游。关于基线报警的说明请参考文档：[基线管理](#)。
- 设置了相关的自定义提醒规则。关于自定义报警规则设置，详情请参考文档：[自定义规则](#)。

如何处理夜间收到报警的情况？

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. 单击左上角的图标，选择全部产品 > 任务运维 > 运维中心。
3. 在左侧导航栏，单击智能监控 > 事件管理。
4. 在事件管理页面关闭报警。您可以通过以下两种方式关闭夜间的报警：
 - o 处理报警事件，在处理期间会暂停事件报警：
 - a. 单击相应事件后的处理。
 - b. 在处理事件对话框中，设置预计处理时间。
 - c. 单击确定。

 说明 事件的处理操作记录会被记录，并且在处理期间暂停报警。

- o 忽略报警事件，永久关闭事件报警：
 - a. 单击相应事件后的忽略。
 - b. 在忽略事件对话框中，单击确定。

 说明 事件的忽略操作记录会被记录，忽略后会永久关闭事件报警。

11. 数据服务

本文为您汇总数据服务的相关问题。

- [为什么数据服务无法连接ECS自建的数据库？](#)
- [是否必须开通API网关？](#)
- [如何配置数据源？](#)
- [数据服务是否支持MaxCompute（ODPS）数据源？](#)
- [向导模式生成API与脚本模式有何功能差异？](#)
- [数据服务中的API分组的作用是什么，与API网关中的分组有什么关联？](#)
- [如何合理设置API分组？](#)
- [最多可以创建多少个API分组？](#)
- [什么情况下要开启API返回结果分页功能？](#)
- [生成API是否支持POST请求？](#)
- [生成API是否支持HTTPS协议？](#)

为什么数据服务无法连接ECS自建的数据库？

数据服务不支持ECS自建的数据库，配置内网地址时，会出现网络不通的情况。

是否必须开通API网关？

API网关提供了API托管服务，如果您的API计划对外开放调用，则必须先开通API网关服务。

如何配置数据源？

数据源需要在工作空间管理 > 数据源管理页面进行配置。配置完成后，数据服务会自动读取数据源信息。详情请参见[配置数据源](#)。

数据服务是否支持MaxCompute（ODPS）数据源？

数据服务不支持直接连接MaxCompute（ODPS）类型的数据源，您需要通过Hologres类型的数据源间接连接和加速查询MaxCompute数据源。Hologres数据源的详情请参见[配置Hologres数据源](#)。

Hologres为您提供更高效和稳定的实时数仓服务。如果您在旧版本中通过Lightning数据源连通MaxCompute数据源，建议您将其迁移至Hologres。

向导模式生成API与脚本模式有何功能差异？

脚本模式的功能更为强大，详情请参见[通过脚本模式生成API](#)。

数据服务中的API分组的作用是什么，与API网关中的分组有什么关联？

API分组是一个特定功能或场景的API集合，是数据服务中API的最小组织单元，对应于API网关中的分组概念。数据服务中的API发布至API网关时，系统会在API网关中自动创建一个同名的分组。

如何合理设置API分组？

通常将解决同一个问题或者相似功能的API放在一个分组当中。例如，您可以把根据城市名称和根据经纬度查询天气的两个API，放在一个名为查询天气的API分组中。

最多可以创建多少个API分组？

目前一个云账号下最多可以创建100个API分组。

什么情况下要开启API返回结果分页功能？

默认情况下，API最多只会返回2000条查询结果。因此当API返回结果有可能超过2000条时，请开启返回结果分页功能。当API无请求参数时，通常返回结果会比较多，系统会强制开启返回结果分页。

生成API是否支持POST请求？

生成API支持GET和POST请求。

生成API是否支持HTTPS协议？

生成API支持HTTP和HTTPS协议。

12.安全中心

本文汇总了DataWorks安全中心的相关问题。

- [通过安全中心可以申请什么权限？](#)
- [数据管理和安全中心是什么关系？](#)
- [为什么在申请时，有时可以选择字段，有时不可以选择？](#)
- [提交申请后，需要谁进行审批？](#)
- [为什么提交了一个申请，在我的申请中看到两个申请单？](#)
- [为什么有的字段只申请1个月权限，审批完成后查看变为永久？](#)
- [为什么有的表和字段没有申请权限，但能看到有权限？](#)
- [为什么并没有审批某个待我审批中的申请单，却没有了？](#)
- [查询某个工作空间和环境，提示MaxCompute项目异常，无法进行操作，该如何处理？](#)
- [为什么交还或回收某个字段权限却无效？](#)
- [为什么主账号不能申请权限？](#)
- [是否可以在安全中心页面查看以前在数据管理页面的申请或审批记录？](#)
- [是否可以通过安全中心的申请记录来回收权限？](#)
- [之前在数据管理提交的权限申请，仍未通过审批，是否需要重新申请？](#)
- [如何设置字段的LabelSecurity？](#)

通过安全中心可以申请什么权限？

您可以通过安全中心页面申请DataWorks工作空间内的表权限，包括开发环境和生产环境。

数据管理和安全中心是什么关系？

安全中心是数据管理中与权限和安全相关功能的升级替代产品。此前已经在数据管理模块中申请的权限和通过 `odpscmd grant` 命令授权的权限，仍然可以在安全中心 > 我的权限中显示。

如果您需要通过可视化的方式进行新的权限申请、审批操作，请进入安全中心进行操作，数据管理模块后续将不再支持权限的申请和审批。

为什么在申请时，有时可以选择字段，有时不可以选择？

如果该工作空间开启了LabelSecurity，即可在申请时选择字段，未开启则仅支持整表申请。

提交申请后，需要谁进行审批？

提交的申请需要项目管理员或表Owner进行审批，其中任何一个审批通过或拒绝，则审批完成。

为什么提交了一个申请，在我的申请中看到两个申请单？

因为您的申请单中包含的数据表的表Owner不同，安全中心会按照表Owner对于申请单自动进行拆分。

为什么有的字段只申请1个月权限，审批完成后查看变为永久？

说明字段的安全等级为0或者小于等于您账号的安全等级。

为什么有的表和字段没有申请权限，但能看到有权限？

出现该情况有以下两种可能：

- 除安全中心外，管理员还可通过控制台命令行给您授权。
- 如果您是通过安全中心进行了申请，则说明字段的安全等级为0或者小于等于您账号的安全等级。

为什么并没有审批某个待我审批中的申请单，却没有了？

因为其它项目管理员或表Owner在您之前完成了审批，因此该申请单已成为完结状态，将不再显示于您的待我审批中。

查询某个工作空间和环境，提示MaxCompute项目异常，无法进行操作，该如何处理？

请发送提示框及框内的错误编码给项目管理员，由其进行问题的排查和解决。

为什么交还或回收某个字段权限却无效？

只能交还或回收字段的安全等级大于账号安全等级的字段，对于安全等级为0或者小于等于账号安全等级的字段，无法进行字段权限的交还或回收。

为什么主账号不能申请权限？

主账号默认具有所有权限，无需单独申请权限，因此对主账号无需的操作。例如，申请权限等进行隐藏，不会影响到主账号的正常使用。

是否可以在安全中心页面查看以前在数据管理页面的申请或审批记录？

目前安全中心和数据管理中的申请或审批记录没有进行关联，如果需要查看在数据管理进行申请或审批的历史记录，请跳转至数据管理页面进行查看。

是否可以通过安全中心的申请记录来回收权限？

目前安全中心并非唯一的授权渠道，为了最大程度地支持权限回收，权限审计中列出的是所有用户的ACL权限，不区分授权渠道。您可以基于目前的权限现状进行回收，无需通过申请记录来操作。

之前在数据管理提交的权限申请，仍未通过审批，是否需要重新申请？

安全中心和数据管理中的申请审批记录未进行关联，您需要重新申请。

如何设置字段的LabelSecurity？

您需要进入数据地图设置字段的LabelSecurity。

13.App Studio

本文将为您汇总App Studio的相关问题。

- [Main函数中打断点后启动会出现什么问题？](#)
- [启动时间过长该如何处理？](#)
- [是否可以在Watch中添加变量？](#)

Main函数中打断点后启动会出现什么问题？

```
public class Main {  
    public static void main(String[] args) {  
        SpringApplication.run(Main.class, args) ; // 在该行打断点后，开始启动调试。  
    }  
}
```

当停止在该断点时，单击Step Over，线程可能无法恢复。

问题被解决之前不要进行该操作，目前DataWorks暂不支持多线程编程调试。

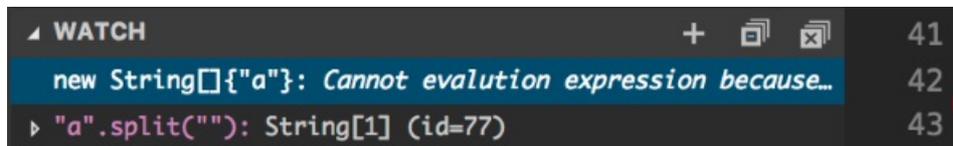
启动时间过长该如何处理？

如果在代码函数定义的地方增加断点，再单击启动调试，会导致程序启动时间较长。

该情况属于正常现象，请尽量使用行断点，而不要使用方法断点。

是否可以在Watch中添加变量？

无法在Watch中添加某些变量。



由上图可见，第一个会抛异常，第二个会失败。