

ALIBABA CLOUD

阿里云

数据传输
数据集成

文档版本：20200820

 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。
<code>Courier</code> 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。
<i>斜体</i>	表示参数、变量。	<code>bae log list --instanceid</code> <i>Instance_ID</i>
<code>[]</code> 或者 <code>[a b]</code>	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
<code>{}</code> 或者 <code>{a b}</code>	表示必选项，至多选择一个。	<code>switch {active stand}</code>

目录

1.配置RDS MySQL间的数据集成任务	05
2.数据集成调度策略说明	11
3.管理数据集成任务	13
4.修改调度策略	14
5.配置监控报警	16
6.查看历史任务	17

1.配置RDS MySQL间的数据集成任务

本文以配置RDS MySQL间的数据集成任务为例，介绍数据集成任务的配置流程和注意事项。

 **说明** 支持全量数据迁移的数据源均支持数据集成功能，详情请参见[支持的数据库和迁移类型](#)。

前提条件

已创建源和目标实例，请参见[创建RDS MySQL实例](#)。

背景信息


DTS的数据集成功能可根据调度策略，定期地将源库中的结构和存量数据迁移至目标库中，通过该功能可以帮助您构建更加灵活的数据仓库。更多详情，请参见[数据集成调度策略说明](#)。

注意事项

如果源数据库没有主键或唯一约束，且所有字段没有唯一性，可能会导致目标数据库中出现重复数据。

功能限制

- DTS支持结构迁移的对象为表、视图、触发器、存储过程、存储函数。

 **说明** 在结构迁移时，DTS会将视图、存储过程或函数中的 `DEFINER` 转换为 `INVOKER`。

- 不支持迁移源库的user信息。迁移完成后，如果您需要调用目标库的视图、存储过程或函数，则需对调用者授予读写权限。

费用说明

迁移类型	链路配置费用	公网流量费用
结构迁移和全量数据迁移	不收费。	通过公网将数据迁移出阿里云时将收费，详情请参见 产品定价 。

操作步骤

本文以华北1（青岛）地域的RDS MySQL实例为例，介绍配置流程。

1. 登录[数据传输控制台](#)。
2. 在左侧导航栏，单击数据集成。
3. 在数据集成任务列表的顶部，选择迁移的目标实例所属地域，本案例选择为华北1（青岛）地域。
4. 单击页面右上角的创建数据集成任务。
5. 配置数据集成任务的源库和目标库连接信息。

1.源库及目标库
2.迁移类型及列表
3.高级配置
4.预检查

* 任务名称:

源库信息

* 实例类型: DTS支持链路类型

* 实例地区:

* RDS实例ID: 其他阿里云账号下的RDS实例

* 数据库账号:

* 数据库密码: 测试连接 ✔ 测试通过

* 连接方式: 非加密连接 SSL安全连接

目标库信息

* 实例类型:

* 实例地区:

* RDS实例ID:

* 数据库账号:

* 数据库密码: 测试连接 ✔ 测试通过

* 连接方式: 非加密连接 SSL安全连接

取消
授权白名单并进入下一步

类别	配置	说明
无	任务名称	DTS会自动生成一个任务名称，建议配置具有业务意义的名称（无唯一性要求），便于后续识别。
源库信息	实例类型	选择RDS实例。
	实例地区	选择源RDS实例所属的地域，本案例选择为华北1（青岛）地域。
	数据库账号	填入源RDS实例的数据库账号，该数据库账号需具备读写权限。 <div style="background-color: #e1f5fe; padding: 5px; margin-top: 5px;"> ? 说明 关于数据库账号的创建和授权方法，请参见创建RDS MySQL数据库账号。 </div>
	数据库密码	填入该数据库账号的密码。 <div style="background-color: #e1f5fe; padding: 5px; margin-top: 5px;"> ? 说明 源库信息填写完毕后，您可以单击数据库密码后的测试连接来验证填入的信息是否正确。如果填写正确则提示测试通过；如果提示测试失败，单击测试失败后的诊断，根据提示调整填写的源库信息。 </div>
	连接方式	根据需求选择非加密连接或SSL安全连接。如果设置为SSL安全连接，您需要提前开启RDS实例的SSL加密功能，详情请参见 设置SSL加密 。
	实例类型	选择RDS实例。

类别	配置	说明
目标库信息	实例地区	选择目标RDS实例所属的地域，本案例选择为华北1（青岛）地域。
	数据库账号	填入目标RDS实例的数据库账号，该数据库账号需具备读写权限。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 5px;"> ? 说明 关于数据库账号的创建和授权方法，请参见创建RDS MySQL数据库账号。 </div>
	数据库密码	填入该数据库账号的密码。 <div style="border: 1px solid #ccc; background-color: #e6f2ff; padding: 5px; margin-top: 5px;"> ? 说明 目标库信息填写完毕后，您可以单击数据库密码后的测试连接来验证填入的信息是否正确。如果填写正确则提示测试通过；如果提示测试失败，单击测试失败后的诊断，根据提示调整填写的目标库信息。 </div>
	连接方式	根据需求选择非加密连接或SSL安全连接。如果设置为SSL安全连接，您需要提前开启RDS实例的SSL加密功能，详情请参见 设置SSL加密 。

6. 配置完成后，单击页面右下角的授权白名单并进入下一步。

7. 配置周期性调度策略并选择迁移对象。

i. 配置周期性调度策略。

1.源库及目标库
2.迁移类型及列表
3.高级配置
4.预检查

* 迁移类型: 结构迁移 全量数据迁移

全量数据迁移过程中，如果源库有数据更新，这部分更新数据不保证被迁移到目标实例。为保证迁移数据的一致性，建议使用增量数据迁移功能或数据同步功能，进行实时的数据同步。当目标实例MySQL的CPU低于2核，建议关闭慢日志。 [查看详情](#)

* 调度策略: 多副本模式，创建新的全量副本到目标端 [查看各调度策略的限制](#)

消耗资源大，需要注意库表名变化

* 调度生效日期: 2020年7月10日 - 2020年7月20日 ?

* 调度周期: 天

* 指定具体时间: 10日 11日 12日 可多选

任务启动时间: 10 : 00

* 任务失败策略: 单次任务失败后，不再执行后续的调度任务 [查看计划调度时间](#)


- 2020年7月10日 10:00:00
- 2020年7月11日 10:00:00
- 2020年7月12日 10:00:00

配置	说明
迁移类型	选中结构迁移和全量数据迁移，DTS将在首次迁移时，迁移源库待迁移对象的结构和存量数据。

配置	说明
调度策略	<ul style="list-style-type: none"> 多副本模式 重新同步模式 基于时间戳的增量模式 <p> 说明 关于调度策略的说明、限制及使用场景请参见调度策略说明。</p>
调度生效日期	<p>单击  图标，然后选择调度生效的起止日期，最大范围为365天。</p>
调度周期	<p>选择调度周期为天或周，然后根据提示设置具体的调度时间点。</p> <p> 说明</p> <ul style="list-style-type: none"> DTS在执行全量数据迁移时将占用源库和目标库一定的读写资源，可能会导致数据库的负载上升，因此建议您充分评估源库和目标库的性能，同时将调度时间点设定在业务低峰期。 设置完成后，可单击任务失败策略右侧的查看计划调度时间来查看设置的调度时间是否符合预期。
任务失败策略	<p>根据业务需求，选择：</p> <ul style="list-style-type: none"> 单次任务失败后，不再执行后续的调度任务 单次任务失败后，忽略失败，并执行下一次调度任务

ii. 选择迁移对象。




配置	说明
迁移对象	<p>在迁移对象框中单击待迁移的对象（库、表或列），然后单击  图标将其移动至已选择对象框。</p> <p>说明</p> <ul style="list-style-type: none">如果调度策略选择为基于时间戳的增量模式，迁移对象不能选择为整库（需为表或列），否则无法在下一步指定时间戳字段。默认情况下，迁移完成后，迁移对象的名称保持不变。如果您需要变更迁移对象在目标实例中的名称，需要使用对象名映射功能，详情请参见库表列映射。如果使用了对象名映射功能，可能会导致依赖这个对象的其他对象迁移失败。

iii. 单击页面右下角的下一步，指定时间戳字段。



8. 单击页面右下角的预检查并启动。

 注意


- 在迁移任务正式启动之前，会先进行预检查。只有通过预检查，DTS才能迁移数据。
- 如果预检查失败，单击具体检查项后的  图标，查看失败详情。根据提示修复后，重新进行预检查。

9. 预检查通过后，单击下一步。

10. 在弹出的购买配置确认对话框，选择链路规格并选中数据传输（按量付费）服务条款。

11. 单击立即购买并启动，数据集成任务将在最近的调度周期（即下次调度时间）正式开始。

<input type="checkbox"/>	ID/名称: [redacted] / 数据集成	状态: 预检通过	停止数据集成任务
2020年7月10日 15:08:39 创建		数据集成任务状态: 运行中	调度生效日期: 2020年7月10日~2020年7月20日 下次调度时间: 2020年7月10日 15:20:00
结构迁移 0%		全量迁移 0%	

 **说明** 该任务将在调度生效时间的截止时间自动结束。如果需要在调度生效的时间内停止该任务，请单击停止数据集成任务。

执行结果

本次任务调度完成后，数据集成任务的状态变更为已完成，DTS将等待下次调度时间并再次执行全量数据迁移。

<input type="checkbox"/>	ID/名称: [redacted] / 数据集成	状态: 已完成	停止数据集成任务
2020年7月10日 15:20:00 创建		数据集成任务状态: 运行中	调度生效日期: 2020年7月10日~2020年7月20日 下次调度时间: 2020年7月11日 15:20:00
结构迁移 100%		全量迁移 100%(已迁移49751行)	

2.数据集成调度策略说明

DTS的数据集成功能可根据调度策略的配置，定期地将源库中的结构和存量数据迁移至目标库中。通过该功能可以帮助您构建更加灵活的数据仓库（例如构建T+1的周期性数仓），本文介绍数据集成的策略、限制及适用场景。

支持的数据库

支持全量数据迁移的数据源均支持数据集成功能，详情请参见[支持的数据库和迁移类型](#)。

费用说明

迁移类型	链路配置费用	公网流量费用
结构迁移和全量数据迁移	不收费。	通过公网将数据迁移出阿里云时将收费，详情请参见 产品定价 。



配置方法

您需要在配置数据集成任务的设置迁移类型及列表环节中，为数据集成任务配置调度策略。

 说明 相关配置案例，请参见[配置RDS MySQL间的数据集成任务](#)。

调度策略说明

调度策略	说明	适用场景
多副本模式	<p>DTS根据调度周期，自动地将源库中待迁移对象的结构和存量数据迁移至目标库中。</p> <p> 说明</p> <ul style="list-style-type: none"> DTS每次调度时，会自动地在目标实例中创建一个带有任务调度时间后缀（格式为 <code>_yyyymmdd_HH_mm_ss</code>）的数据库，用于接收迁移的数据。例如待迁移的数据库为 <code>dtstestdata</code>，任务调度时间为2020年02月10日的11:00:00，那么DTS在该时间执行迁移时，在目标库中创建的库名即为 <code>dtstestdata_20200210_11_00_00</code>。 目标库需具备充足的存储空间用于存放每个周期DTS迁移过来的数据。 	适用于某些测试场景，例如对多个副本历史数据执行测试操作。

调度策略	说明	适用场景
重新同步模式	<p>在第一个调度时间点，DTS自动地将待迁移对象的结构和存量数据迁移至目标库中；在后续的调度时间点，DTS会重新将待迁移对象的结构和存量数据迁移至目标库中。</p> <div style="background-color: #fff9c4; padding: 10px; border: 1px solid #ccc;"> <p> 警告</p> <ul style="list-style-type: none"> 在每次的调度时间点到来之前，您需要手动删除上个调度时间点迁移到目标库中的结构和数据（即删除迁移过来的库和表），否则将会导致数据迁移失败。 如果DTS对该数据源不支持结构迁移，您还需要对照待迁移对象的结构在目标库中创建对应的结构。 </div>	<p>在目标库中仅保持一份最新的存量数据用于测试或数据分析。</p>
基于时间戳的增量模式	<p>DTS会根据时间戳字段，将上次调度后产生的新数据迁移至目标库。该策略存在如下限制：</p> <ul style="list-style-type: none"> 当源和目标均为MySQL数据库时才支持该策略。 源表须具备时间戳字段，且需在配置数据迁移任务时指定每个表的时间戳字段。 源表不能存在外键约束，否则可能导致数据迁移失败（可能在目标库中先更新有外键约束的子表，后更新附表，导致子表插入新数据报错）。 无法迁移被物理删除的数据（例如通过DELETE、TRUNCATE命令删除）。 <div style="background-color: #e1f5fe; padding: 10px; border: 1px solid #ccc;"> <p> 说明</p> <ul style="list-style-type: none"> 如果存在无主键的表，DTS会将整条记录作为主键进行全字段匹配，然后删除旧数据并插入新数据。 新增的数据在写入目标库时，如果遇到主键冲突，DTS将执行REPLACE INTO来插入新数据。 </div>	<p>适用于轻量级的数据迁移场景（无需每次迁移所有的存量数据），例如构建T+1的周期性数仓。</p>

3. 管理数据集成任务



完成数据集成任务的配置后，您可以停止、恢复或删除该数据集成任务。

前提条件

已创建数据集成任务，相关案例请参见[配置RDS MySQL间的数据集成任务](#)。

背景信息

数据集成任务支持的管理操作及对应的影响请参见下表。

操作	影响
停止数据集成任务	<p>数据集成任务将转变为已暂停状态，暂停调度该任务。</p> <p> 说明 处于该状态的任务，可执行恢复数据集成任务，让任务继续运行。</p>
恢复数据集成任务	<p>数据集成任务将转变为运行中或准备启动状态，继续调度该任务。当到达调度时间点时，DTS将正常生成一个数据迁移任务。</p>
删除数据集成任务	<p>数据集成任务将转变为已删除状态，不再调度该任务。</p> <p> 警告 处于该状态的任务，不能再执行恢复数据集成任务，请谨慎执行该操作。</p>

操作步骤

1. 登录[数据传输控制台](#)。
2. 在左侧导航栏，单击数据集成。
3. 在数据集成任务列表的顶部，选择目标数据集成任务所属地域。
4. 找到目标数据集成任务，单击对应的实例ID。
5. 找到目标数据集成任务，单击对应的停止数据集成任务。
6. 在弹出的对话框中，单击确认。
7. 对于处于已暂停状态的数据集成任务，您可以恢复或删除该任务。



- 恢复任务：单击恢复数据集成任务，然后在弹出的对话框中单击确认。
- 删除任务：单击删除数据集成任务，然后在弹出的对话框中单击确认。

4.修改调度策略

对于已经配置完成的数据集成任务，如果由于业务需求变更需要重新调整调度策略，您可以参照本文的方法修改调度策略。

前提条件

已创建数据集成任务，相关案例请参见[配置RDS MySQL间的数据集成任务](#)。

操作步骤

1. 登录[数据传输控制台](#)。
2. 在左侧导航栏，单击[数据集成](#)。
3. 在数据集成任务列表的顶部，选择目标数据集成任务所属地域。
4. 找到目标数据集成任务，单击对应的实例ID。
5. 在页面右上角，单击[修改调度策略](#)。
6. 在弹出的对话框中查看当前的调度策略信息，然后根据业务需求进行调整。

数据集成 ✕

* 调度策略: 多副本模式，创建新的全量副本到目标端 ▼

消耗资源大，需要注意库表名变化

* 调度生效日期: 2020年7月10日 - 2020年7月20日 ☰ ?

* 调度周期: 天 ▼

* 指定具体时间: 10日 11日 12日 13日 14日 可多选

任务启动时间: 15 ↑ ↓ : 20 ↑ ↓

* 任务失败策略: 单次任务失败后，不再执行后续的调度任务 ▼ 查看计划调度时间

- 2020年7月10日 15:20:00
- 2020年7月11日 15:20:00
- 2020年7月12日 15:20:00
- 2020年7月13日 15:20:00
- 2020年7月14日 15:20:00

确认
关闭

配置	说明
调度策略	暂不支持修改，关于调度策略的说明、限制及使用场景请参见 调度策略说明 。
调度生效日期	单击☰图标，然后选择调度生效的起止日期，最大范围为365天。

配置	说明
调度周期	<p>选择调度周期为天或周，然后根据提示设置具体的调度时间点。</p> <p>说明</p> <ul style="list-style-type: none">○ DTS在执行全量数据迁移时将占用源库和目标库一定的读写资源，可能会导致数据库的负载上升，因此建议您充分评估源库和目标库的性能，同时将调度时间点设定在业务低峰期。○ 设置完成后，可单击任务失败策略右侧的查看计划调度时间来查看设置的调度时间是否符合预期。
任务失败策略	<p>根据业务需求，选择：</p> <ul style="list-style-type: none">○ 单次任务失败后，不再执行后续的调度任务○ 单次任务失败后，忽略失败，并执行下一次调度任务

7. 单击确认。

5. 配置监控报警

数据集成任务提供了监控报警功能，您可以通过控制台对重要的监控指标设置报警规则，让您第一时间了解该任务的运行状态。

前提条件

已创建数据集成任务，相关案例请参见[配置RDS MySQL间的数据集成任务](#)。

背景信息

DTS目前支持的监控项为延迟和迁移状态：

- **延迟**：监控增量数据迁移延迟信息，当增量数据迁移的延迟时间超过设定值时（单位为秒），触发报警规则。
- **迁移状态**：监控任务的状态，当任务的状态为异常时，触发报警规则。

 **说明** 由于数据集成任务不涉及增量数据迁移，仅需配置迁移状态监控项。

操作步骤

1. 登录[数据传输控制台](#)。
2. 在左侧导航栏，单击数据集成。
3. 在数据集成任务列表的顶部，选择目标数据集成任务所属地域。
4. 找到目标数据集成任务，单击对应的实例ID。
5. 在左侧导航栏，单击监控报警。
6. 在监控报警页面，对迁移状态监控项进行配置。
 - i. 单击迁移状态监控项操作的修改。



修改监控告警

*监控项：迁移状态


*报警规则：迁移状态 = 异常

*报警联系人手机号码：

[+添加联系人](#)

确定 关闭

- ii. 设置报警联系人手机号码。

 **说明** 如需填写多个报警联系人手机号码，单击添加联系人并填写即可，最多支持5个。

- iii. 单击确定。

6. 查看历史任务

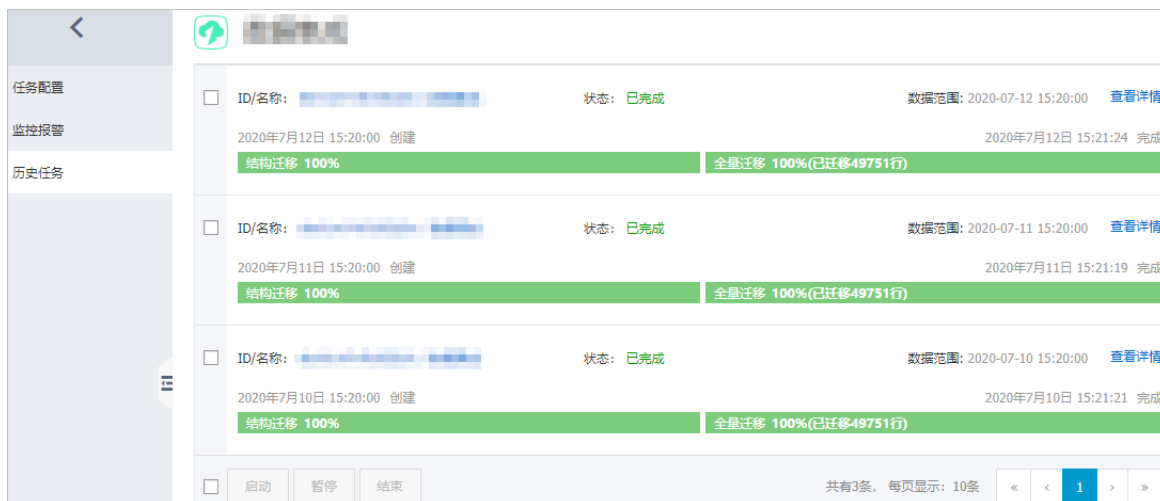
数据集成任务会在每个调度时间点生成一个数据迁移任务，通过DTS控制台您可以查看所有已经运行完成的数据迁移任务，了解该任务的执行情况。

前提条件

已创建数据集成任务，相关案例请参见[配置RDS MySQL间的数据集成任务](#)。

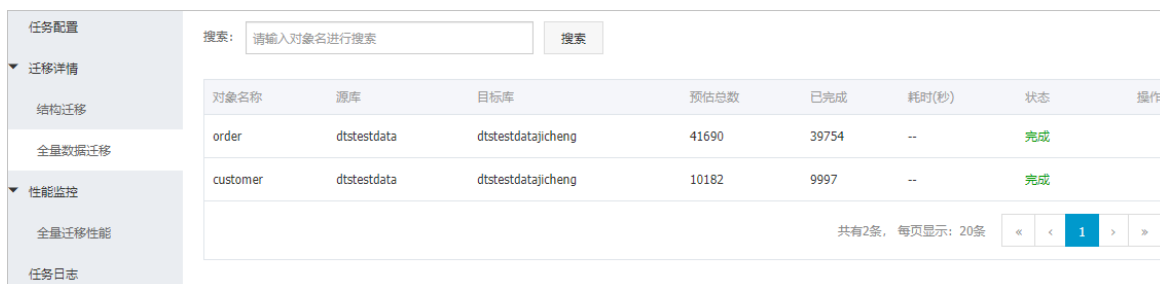
操作步骤

1. 登录[数据传输控制台](#)。
2. 在左侧导航栏，单击[数据集成](#)。
3. 在数据集成任务列表的顶部，选择目标数据集成任务所属地域。
4. 找到目标数据集成任务，单击对应的实例ID。
5. 在左侧导航栏，单击[历史任务](#)，可查看到该数据集成任务已经调度过的所有历史任务信息。



说明 在本页面，您可以查看到各历史任务的运行状态、数据范围等信息。

6. 单击目标历史任务对应的[查看详情](#)，页面将跳转至该任务的全量数据迁移页面。



② 说明 每个历史任务均由一个数据迁移任务构成，您可以像管理数据迁移任务一样管理该历史任务，例如：

- [查看数据迁移进度](#)
- [查看全量迁移链路状态和性能](#)
- [查看任务日志](#)