数据传输服务DTS

快速入门

为了无法计算的价值 | 阿里云

快速入门

购买流程

数据传输目前只支持通过官网购买，本小节简单介绍各种功能的具体购买流程。

购买数据迁移作业

1. 登录数据传输控制台。
2. 进入数据迁移页面，点击右上角”创建数据迁移”，开始迁移任务配置。
3. 完成任务配置后，选择迁移任务规格。
4. 单击 **立即购买并启动**，完成迁移实例购买。

购买数据订阅通道

1. 登录阿里云官网数据传输售卖页面。
2. 选择付费类型，支持包年包月和按量付费。
3. 选择源实例地域及购买链路数量。 其中，源实例地域为需要订阅的RDS实例所在的地域，一旦购买后，就不能更换地域。
4. 单击**立刻购买**和**在线支付订单**，完成链路费用支付。
5. 完成订阅通道购买。

购买数据同步作业

1. 登录阿里云官网数据传输售卖页面。
2. 选择付费类型，支持包年包月和按量付费。
3. 选择源地域、目标地域、链路规格及购买链路数量。 其中，源地域为需要同步的源RDS实例所在的地域。目的地域为需要同步的目标RDS实例所在的地域。链路规格的定义参考数据同步规格定义。
4. 单击**立刻购买**和**在线支付订单**，完成链路费用支付。
5. 完成数据同步链路购买。

购买双向同步实例

1. 登录阿里云官网数据传输售卖页面。
2. 选择付费类型，支持包年包月和按量付费。

3. 选择源实例、源地域、目标实例及目标地域，目前只有RDS For MySQL->RDS For MySQL的同步支持双向同步，所以源实例和目标实例都选择RDS For MySQL。其中，源地域为需要同步的源RDS实例所在的地域。目的地域为需要同步的目标RDS实例所在的地域。
4. 选择同步拓扑，同步拓扑选择双向同步
5. 选择链路规格及购买链路数量。 链路规格的定义参考数据同步规格定义。
6. 单击**立刻购买**和**在线支付订单**，完成链路费用支付。
7. 完成双向同步实例的购买。

目前跨国际同步链路默认不开放，如果要购买跨境同步链路，可以提交工单或直接联系您的客户经理，后端审核通过后可放开购买入口。

创建数据迁移任务

数据传输服务 DTS 提供的数据迁移功能，简单易用，只需 3 个步骤，即可完成整个迁移任务的配置。

本小节以 MySQL->RDS for MySQL 为例，详细介绍数据迁移任务的配置流程，其他存储引擎的配置流程类似。

前提条件

创建 RDS for MySQL 实例上的数据库

在数据迁移过程中，如果待迁移的数据库在目标 RDS for MySQL 实例中不存在，那么 DTS 自动会创建。但是对于如下两种情况，用户需要在配置迁移任务之前，手动创建数据库。

i. 数据库名称不符合：**RDS 定义规范** (由小写字母、数字、下划线、中划线组成，字母开头，字母或数字结尾，最长 64 个字符)

待迁移数据库，在源 MySQL 实例，跟目标 RDS for MySQL 实例中名称不同

对于这两种情况，用户需要在配置迁移任务之前，先在 RDS 控制台完成数据库创建。具体参考 RDS 数据库创建流程。

创建迁移账号

创建迁移任务，需要提供源实例及目标 RDS for MySQL 实例的迁移账号。各种存储引擎具体需要的数据库权限，可以详细参考 数据传输服务 DTS 产品使用手册 的 用户指南-数据迁移 栏目中，针对不同数据库实例类型的具体数据迁移配置方式的介绍。

如果您的源 MySQL 实例迁移账号尚未创建，那么您可以参考 Grant 语法说明，创建满足权限要求的迁移账号。

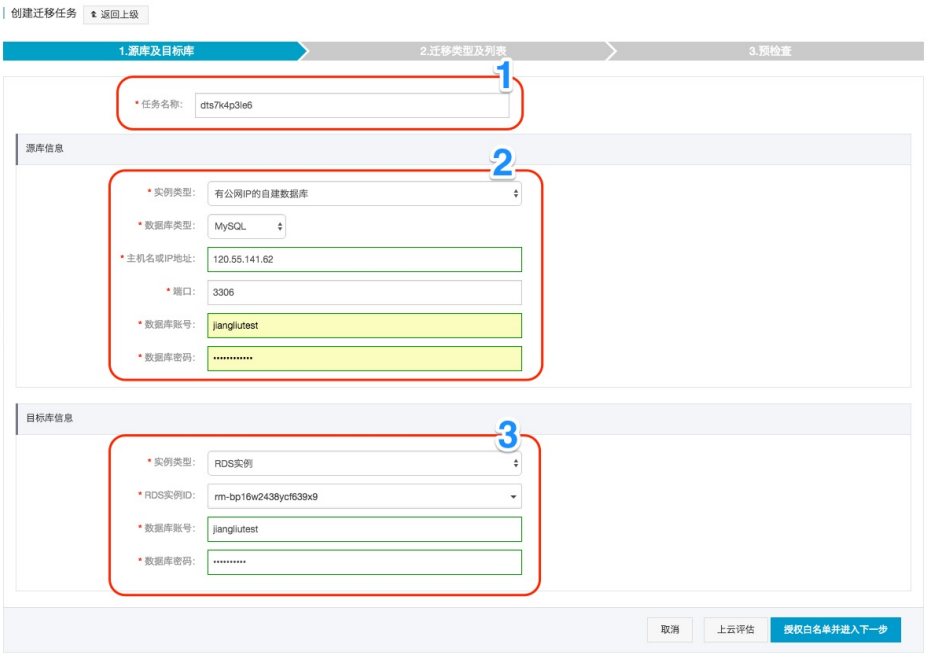
如果您的目标 RDS for MySQL 实例迁移账号尚未创建，那么您可以参考 RDS 账号创建流程，创建待迁移数据库有读写权限的迁移账号。

操作步骤

当数据库、迁移账号都创建完成后，就可以开始配置迁移任务了。下面详细介绍下具体的配置步骤。

1. 进入数据传输 DTS 控制台，点击页面右上角 **创建迁移任务** 开始任务配置。

   配置迁移实例连接信息。



   在这个步骤中，主要配置迁移任务名称，迁移源实例及目标实例连接信息。其中：

   1) 任务名称

   默认情况下，DTS 为每个任务自动生成一个任务名称，您可以修改这个名称，为任务配置一个具备业务意义的名称，便于后续的任务识别。

   2) 源实例连接信息

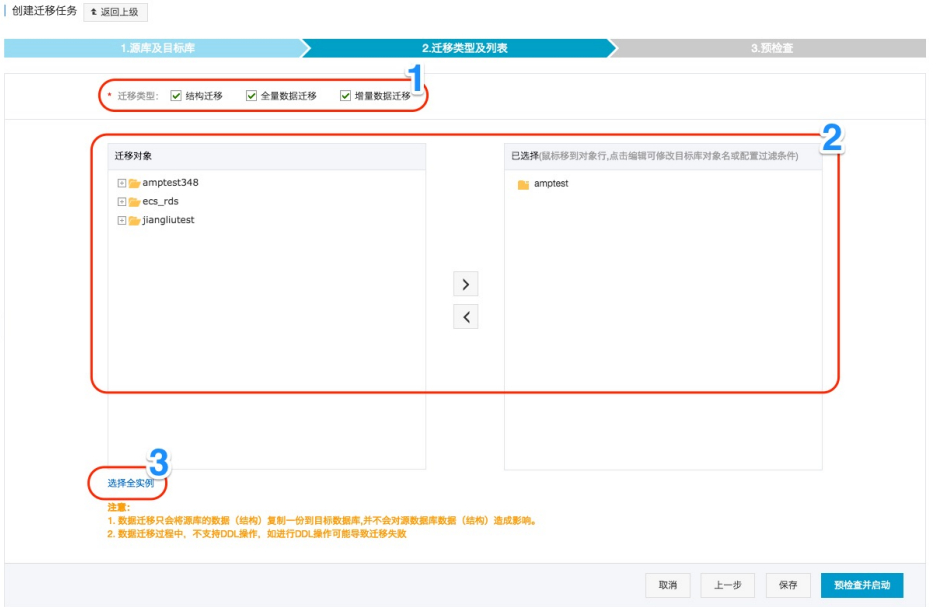
   因为本文档以 MySQL->RDS For MySQL 迁移为例，所以：

   实例类型：**有公网 IP 的自建数据库**
   数据库类型：**MySQL**
   主机名或 IP 地址：**MySQL 实例连接串**
   端口：**MySQL 实例监听端口**
   账号：**MySQL 实例访问账号**
   密码：**上面指定的 MySQL 访问账号对应的密码**

3) 目标 RDS for MySQL 实例连接信息

实例类型：**RDS 实例**
RDS 实例 ID: **选择需要迁移目标 RDS 实例 ID**
数据库账号：**RDS 实例的访问账号**
数据库密码：**上面指定的数据库账号对应的数据库密码**

迁移对象及迁移类型配置。



在这个步骤中，需要配置迁移类型及迁移对象。其中：
1)迁移类型
迁移类型包括结构迁移、全量数据迁移及增量数据迁移。
如果要做全量迁移，那么选择结构迁移 + 全量数据迁移。
如果要做不停机迁移，那么选择结构迁移 + 全量数据迁移 + 增量数据迁移。

2) 迁移对象

迁移对象，需要选择您要迁移的对象。迁移对象选择的粒度可以为：库、表、列三个粒度。
默认情况下，迁移对象迁移到目标库后，对象名跟源库一致。如果您迁移的对象在源实例跟目标实例上名称不同，那么需要使用 DTS 提供的对象名映射功能，详细使用方式可以参考 库表列映射。

预检查。

在迁移任务正式启动之前，会先进行前置预检查，只有预检查通过后，才能成功启动迁移。具体的预检查内容详见 预检查 一节。

如果预检查失败，那么可以点击具体检查项后的按钮，查看具体的失败详情，并根据失败原因修复后，重新进行预检查。

当点击查看后，可以查看到具体的预检查失败原因及修复方法：

当修复完成后，可以在任务列表中，选择这个任务，然后重新启动预检查。

启动迁移任务。

当预检查通过后，我们可以启动迁移任务，任务启动后，可以到任务列表中查看具体的迁移状态及进度。

以上为以 MySQL->RDS For MySQL 迁移为例，数据迁移任务完整的创建流程。其他的实例类型或者存储引擎的配置流程类似。详细的配置流程可以参考 数据传输服务 DTS 产品使用手册 的 用户指南-数据迁移 栏目中，针对不同数据库实例类型的具体数据迁移配置方式的介绍。

文件导入任务配置

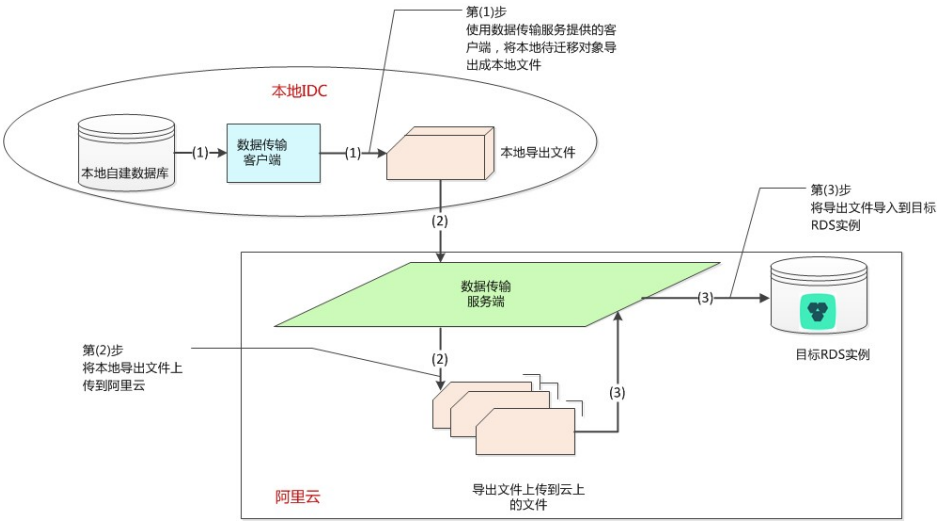
数据传输服务支持通过离线文件导入的方式进行数据库迁移。这种迁移方式主要是为解决封闭网络环境下的数据库的上云迁移。当数据传输服务器无法连通源数据库时，那么数据传输服务提供的数据迁移方式，无法从云端连接本地IDC机房的数据库，进行数据上云迁移。

本小节简单介绍利用文件导入功能实现封闭网络环境下的数据库的上云迁移。

支持数据源

目前文件导入方式仅支持SQLServer的上云迁移。

迁移步骤简介



使用文件导入的实现的迁移方案，整个迁移过程分为三个步骤：

数据导出。

这个步骤是使用数据传输服务提供的客户端将源数据库中需要迁移的数据导出成文件。这个文件的数据存储格式为数据传输DTS定义的中间格式。

为了保证迁移数据的完整性，当使用客户端进行数据导出时，源库最好停写,防止迁移过程中源实例新增写入数据未被迁移到目标库，导致迁移两边数据源数据不一致。

导出文件上传。

这个步骤是，使用数据传输服务提供的客户端，将本地导出的文件上传到阿里云。

数据导入。

这个步骤是将第二步中上传上来的文件，导入到目标RDS实例中。

迁移方法

下面简单介绍使用文件导入功能进行上云迁移的详细步骤。

导出本地数据

下载客户端。

如果您已经下载过DTS客户端软件，那么这个步骤可以忽略。
如果您尚未下载过DTS客户端软件，那么首先需要下载数据传输DTS提供的客户端软件，后续将使用这个客户端软件，进行本地待迁移数据的导出，及导出文件的上传操作。
客户端提供Window64,Linux 两个版本，需要根据本地服务器运行环境选择不同的版本。
客户端的下载入口，在文件导入界面的新手指南，或者是下方的提示链接中，请根据您本地运行服务的操作系统，选择下载不同的版本。

安装客户端。

Window版本

解压客户端。

使用解压软件，将客户端解压到本地。

启动客户端。

进入解压目录，双击**start.bat** 启动客户端。

连接客户端，进入客户端控制台。

使用客户端部署服务器IP:8081 启动客户端WEB控制台。其中8081为WEB服务的监听端口。

Linux版本

解压客户端。

在本地的linux服务器上将客户端解压到本地目录。

启动客户端

进入解压目录，执行如下命令，启动客户端。

```
sh  start.sh
```

启动完成后，使用如下命令确认下，客户端是否启动成功。

```
ps -ef|grep portal
```

如果输出结果如下，那么说明客户端启动成功。

```
[root@dm00 jiangliutest]# ps -ef|grep portal
root      4159     1 50 15:15 pts/0    00:00:36 java -Dmeta.dir=/jiangliutest -Dfile.encoding=UTF-8 -Xms1g -Xmx1g -server -jar dts-portal.war
```

连接客户端。

客户端为WEB应用程序，所以需要找一台部署了浏览器，且能够连通客户端的服务器。在浏览器地址中，输入IP:8081(其中IP为部署客户端的服务器的IP地址，8081为客户端的监听端口), 进入客户端WEB控制台。

导出数据到本地文件。

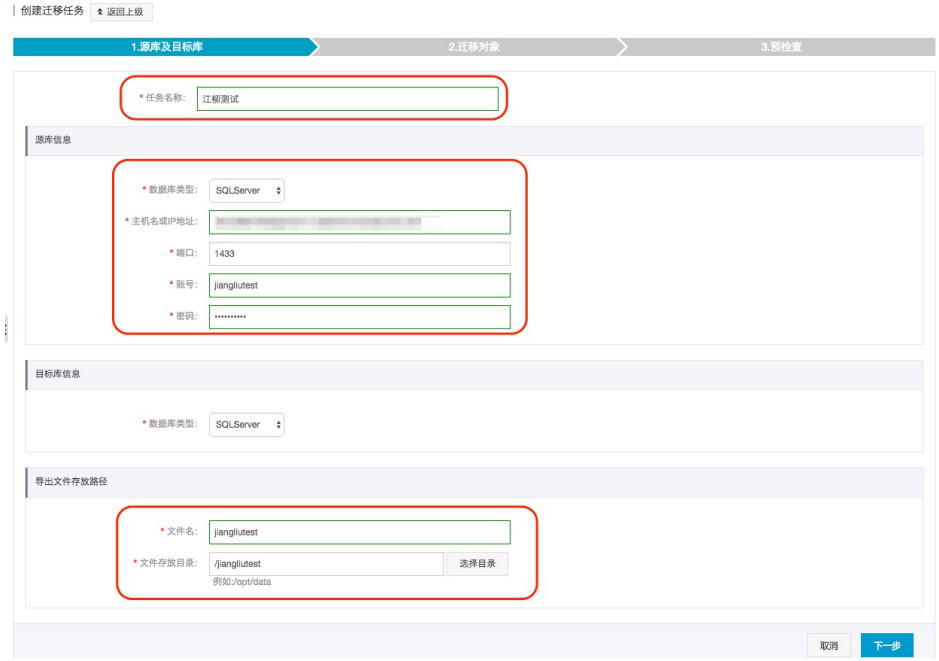
i. 进入客户端控制台，点击右上角的 **数据导出** 按钮，开始数据导出的后续配置。

数据库连接信息及导出文件路径配置。

这一步主要配置 任务名，导出数据库的连接方式以及导出文件的存储路径。其中：

i. 数据导出任务名称：任务名称为导出任务的别名，没有唯一性要求。默认情况下，DTS为每个导出任务自动生成一个名称，可以修改这个名称，为其配置一个具备业务意义的名称，便于后续的任务识别。
ii. 待导出数据库的连接方式。目前只能支持SQLServer，需要填写SQLServer数据库的主机名或IP地址， 监听端口，账号密码。

导出文件的名称及存放路径。



选择导出对象。

这个步骤主要选择导出对象。 目前导出对象可以支持库、表、列三种粒度。

默认情况下，导出对象在目标实例中的对象名跟源实例相同。如果希望对象在目标实例中的名称跟源实例不同，那么需要使用DTS提供的对象名映射功能。具体使用方式参考 **库表列映射**。

预检查。

在迁移任务正式启动之前，会先进行前置预检查，只有预检查通过后，才能成功启动数据导出。

如果预检查失败，那么可以点击具体检查项后的按钮，查看具体的失败详情，并根据失败原因修复后，重新进行预检查。

当预检查通过后，即可启动导出任务。启动完成后，可以到任务列表中查看导出的具体进度。

上传导出文件

当数据导出完成后，即可以将导出文件上传到阿里云。需要使用数据传输客户端进行导出文件上传。具体的上传步骤如下：

上传文件。

当数据导出完成后，在任务列表页，选择需要上传文件的任务，点击右侧的 **上传文件**，开始文件上传的任务配置。

登录阿里云账号。

如果未登录阿里云官网，那么需要进行阿里云账号登录。此时登录的阿里云账号为要数据要导入的目标RDS实例所对应的阿里云账号。

上传信息配置。

当登录阿里云账号后，需要配置任务名称、上传区域及上传文件选择等。其中：

任务名称

这个名称为上传文件任务的别名，没有唯一性要求。建议选择有业务意义的名称，方便后续任务识别

上传区域选择

上传区域是指，导出文件上传到阿里云的哪个地区（包含杭州、北京、青岛、深圳、上海、香港、美西、亚太新加坡），建议选择目标RDS实例所在的区域，可以加快数据导入速度。

上传文件选择。

选择需要上传的导出文件的绝对路径。

完成上面所有上传信息的配置后，即可启动文件上传任务。上传进度可以到客户端任务列表查看。

数据导入

当文件上传完成后，即可登录阿里云数据传输DTS控制台，将上传文件导入到目标RDS实例。

导出文件选择。

进入数据传输DTS控制台的数据迁移页面，点击右上角的 **文件导入**，开始文件导入的配置。

在文件导入的第一步，需要选择需要导入的文件。在**迁移对象**框中，选择刚才上传的文件，将其添加到右边的 **已选择对象** 框中。

选择导入RDS实例。

这个步骤主要配置，需要导入的目标RDS实例的信息，包括RDS实例ID，RDS实例访问账号及密码。这里配置的连接账号必须对待导入对象有读写权限。

预检查。

当导入文件和目标RDS实例配置完成后，即进行导入前的预检查。只有预检查通过后，才能启动数据导入。当数据导入启动后，可以到任务列表查看导入任务的详细进度。

创建实时同步作业

数据传输DTS提供的数据实时同步功能，简单易用。只需3个步骤，即可完成整个同步链路的配置。本小节介绍如何使用数据传输服务快速创建两个实例之间的实时同步作业，实现RDS增量数据的实时同步。

同步限制

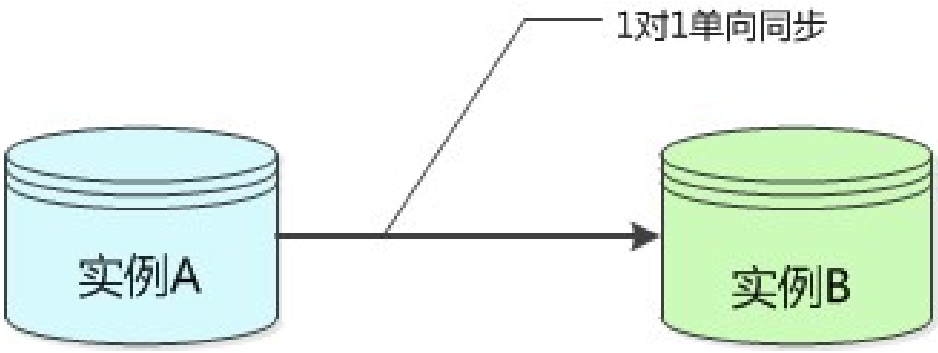
数据源

- 目前实时同步支持MySQL->MySQL/Maxcompute/AnalyticDB/Datahub、DRDS->DRDS/AnalyticDB/Datahub
- MySQL支持RDS MySQL、ECS实例上的自建MySQL、通过专线接入的本地自建MySQL及通过VPN网关接入的本地自建MySQL

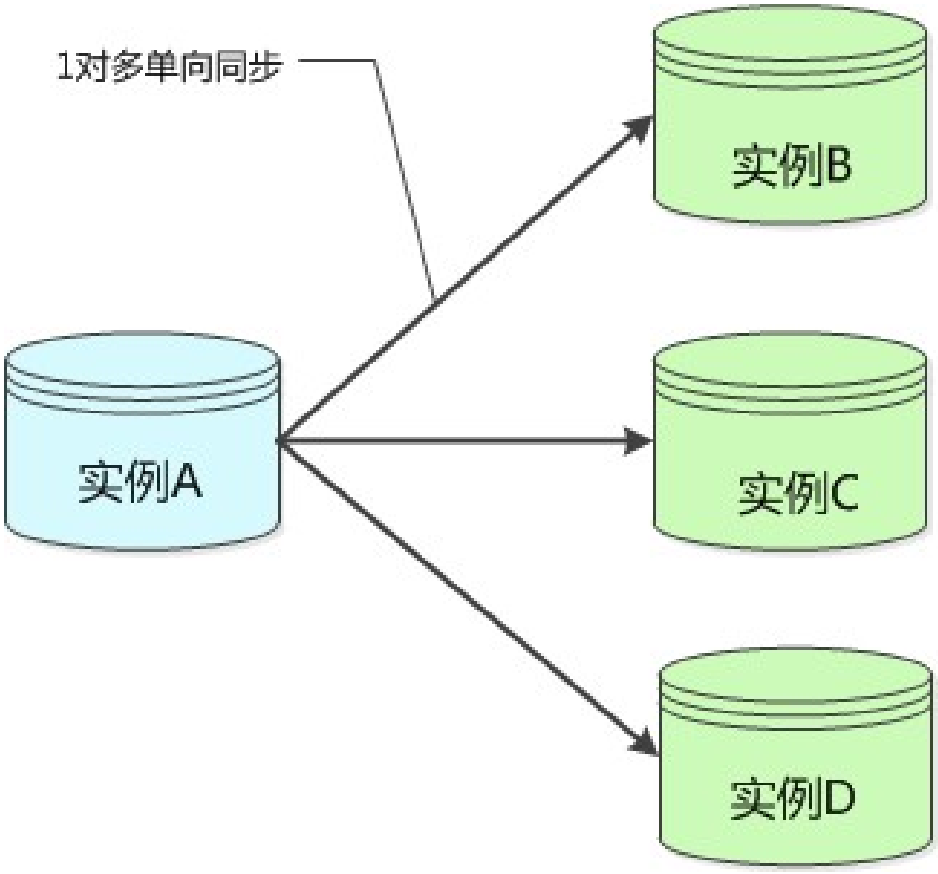
同步架构

目前数据传输服务提供的实时同步功能支持的同步架构有限，其仅能支持如下三种架构：

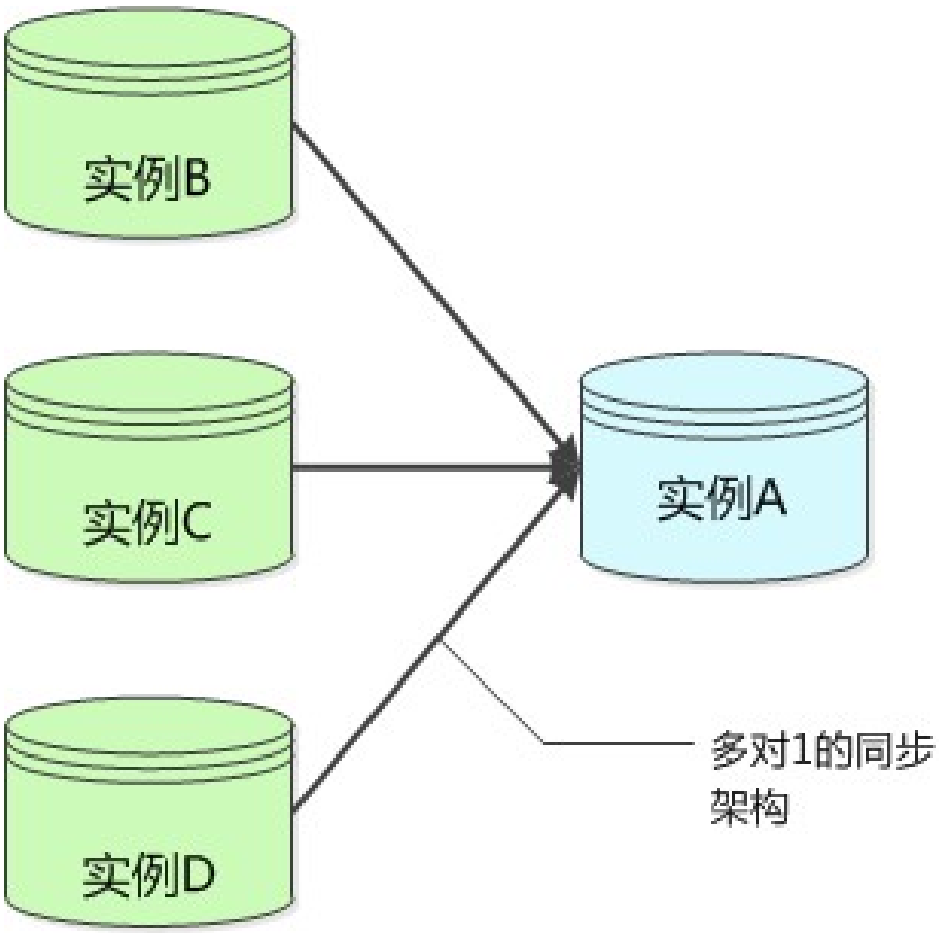
A->B 即两个实例之间的单向同步。且要求B中同步的对象必须为只读，否则可能导致同步链路异常。



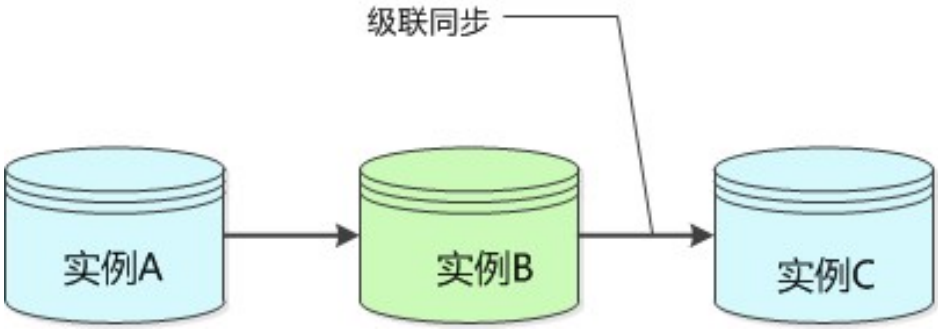
A->B/C/D 即1对多的分发式同步架构,这个架构对目标RDS实例个数没有限制，但是要求目标实例中的同步对象必须为只读，否则可能导致同步链路异常。



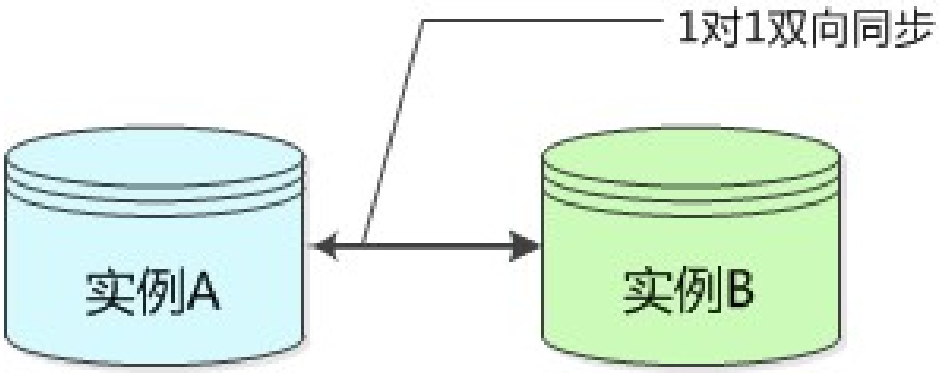
B/C/D->A 即多对1的数据汇总架构。对于多对1的同步架构，要求每个同步链路的同步对象不相同，保证同步完整性。



A->B->C 即级联架构。



A->B->A 即实例A和实例B之间的双向同步架构。



如果用户配置同步链路过程中，配置了这些不支持的同步架构，那么预检查中的**复杂拓扑**检查项会检查失败。

不兼容触发器

如果同步对象为整个库且这个库中包含了会更新同步表内容的触发器，那么可能导致同步数据不一致。

触发器内容为在insert一条数据到a之后，在b中插入一条数据。这种情况下，在同步过程中，如果源实例有a上的insert操作，就会导致b表在源实例跟目标实例数据不一致。

为了解决这个问题，只能将目标实例中的对应触发器删除掉。b表的数据由源实例同步过去。具体解决方案详见最佳实践中的，**触发器存在情况下如何配置同步链路**。

前提条件

在配置同步作业前，要确保同步作业的源及目标RDS实例都已经存在。如果不存在，那么请先购买RDS实例购买RDS实例。

操作步骤

下面我们详细介绍下创建任意两个RDS实例之间的同步链路的具体步骤。

购买同步链路

i. 进入数据传输服务控制台，进入数据同步页面。
ii. 点击控制台右上角“创建同步作业” 开始作业配置。

在链路配置之前需要购买一个同步链路。同步链路目前支持包年包月及按量付费两种付费模式，可以根据需要选择不同的付费模式。

在购买页面需要配置的参数包括：

- 源实例支持MySQL及DRDS,此处选择MySQL。在配置任务时，实例类型选择RDS实例

源地域

源地域为同步链路源RDS实例所在地域。

目标实例DTS支持MySQL、AnalyticDB、Datahub、Elasticsearch及MaxCompute，此处选择MySQL。

目标地域

目标地域为同步链路目标RDS实例所在地域。

实例规格

实例规格影响了链路的同步性能，实例规格跟性能之间的对应关系详见 数据同步规格说明 。

数量

数量为一次性购买的同步链路的数量，如果购买的是按量付费实例，一次最多购买99条链路。

当购买完同步实例，返回数据传输控制台，点击新购链路右侧的“配置同步作业” 开始链路配置。

同步链路连接信息配置

在这一步主要配置：

- 同步作业名称同步作业名称没有唯一性要求，主要为了更方便识别具体的作业，建议选择一个有业务意义的作业名称，方便后续的链路查找及管理。
- 源实例信息此处，实例类型选择：RDS实例; 实例ID: 配置需要同步的源RDS实例的实例ID。

  目标实例信息此处，实例类型选择：RDS实例；实例ID: 配置需要同步的源RDS实例的实例ID。

  当这些内容配置完成后，可以点击“授权白名单并进入下一步”。

授权RDS实例白名单

这个步骤，主要是将数据传输服务器IP添加到同步RDS实例的白名单中。避免因为RDS设置了白名单，数据传输服务器连接不上RDS导致同步作业创建失败。

为了保证同步作业的稳定性，在同步过程中，请勿将这些服务器IP从RDS实例的白名单中删除。

当白名单授权后，点击下一步，进入同步账号创建。

创建目标库上的同步账号

这个步骤主要是在目标RDS实例上创建一个同步账号，账号名字为:dtssyncwriter，在同步过程中，不能删除这个账号，否则会导致同步链路中断。

选择同步对象

当创建完目标RDS实例的同步账号后，即进入同步对象的选择步骤。实时同步的同步对象的选择粒度可以支持到表级别,即用户可以选择同步某些库或是同步某几张表。

如果选择的同步对象为整个库，那么这个库中所有对象的结构变更操作（例如create table,drop view 等），都会同步到目标库。

如果选择的某张表，那么只有这个表的drop/alter/truncate/rename table,create/drop index 的操作会同步到目标库。

需要注意的是rename table操作可能导致同步数据不一致。例如同步对象只包含表A，不包含表B，如果同步过程中源实例执行了**rename A to B**的操作，那么改名后的B表的操作不会被同步到目标库。为了解决这个问题，可以选择同步表A、B对应的整个数据库。

当选择完同步对象后，即进入同步初始化选择。

同步初始化配置

同步初始化配置，初始化是同步链路启动的第一步，它会将源实例中已经存在同步对象的结构及数据在目标实例中初始化，作为后续增量同步数据的基线数据。

同步初始化类型细分为：结构初始化，全量数据初始化。默认情况下，需要选择结构初始化及全量初始化。

预检查

当上面所有选项配置完成后，即进入启动之前的预检查。

当同步作业配置完成后，数据传输服务会进行限制预检查，当预检查通过后，可以点击 **启动** 按钮,启动同步作业。

当同步作业启动之后，即进入同步作业列表。此时刚启动的作业处于**同步初始化**状态。初始化的时间长度依赖于源实例中同步对象的数据量大小。当初始化完成后同步链路即进入**同步中**的状态，此时源跟目标实例的同步链路才真正建立完成。

创建数据订阅通道

数据订阅DTS提供的实时数据订阅功能，可以实时订阅RDS增量数据。轻松帮助业务解决**轻量级缓存更新、业务异步解耦、含ETL逻辑的数据实时同步**等场景。

如果需要使用数据订阅功能，实时订阅消费RDS增量数据，需要进行两个步骤的操作：

1. 在DTS控制台，创建订阅RDS实例的订阅通道。
2. 使用DTS提供的SDK，访问订阅通道，实时订阅并消费增量数据。

本小节主要简单介绍在DTS控制台创建订阅通道的流程。订阅通道配置流程简单，只要3个步骤即可完成整个配置过程。关于订阅通道的管理及SDK使用等参考DTS的产品使用手册。

数据订阅限制

- 只支持RDS For MySQL实例
- MySQL5.6 binlog的binlog_row_image格式必须为full
- MySQL存储引擎只支持myisam,innodb
- MySQL字符集只支持latin1、gbk、utf8、utf8mb4、binary
- 订阅通道保存最近24小时的数据，过时的数据将会从订阅通道中删除。

订阅通道配置步骤

下面详细介绍下创建一条订阅通道的配置步骤。

购买订阅通道。

登录数据传输DTS控制台，选择左侧的 **数据订阅**，进入数据订阅页面。然后，点击页面右上角的 **创建数据订阅** 进入DTS购买页面。

订阅通道支持包年包月及按量付费两种付费模式，可以根据需要选择不同的付费模式。

在购买页面需要配置的参数包括：

源实例地域

源实例地域为需要订阅的RDS实例所在的地域。订购后不支持更换地域，所以选择时，请谨慎。

数量

数量为一次性购买的订阅通道的数量，如果购买的是按量付费实例，一次最多购买99条

当购买完同步实例，返回数据传输控制台，点击新购链路右侧的“配置同步作业” 开始链路配置。

订阅RDS实例ID配置。

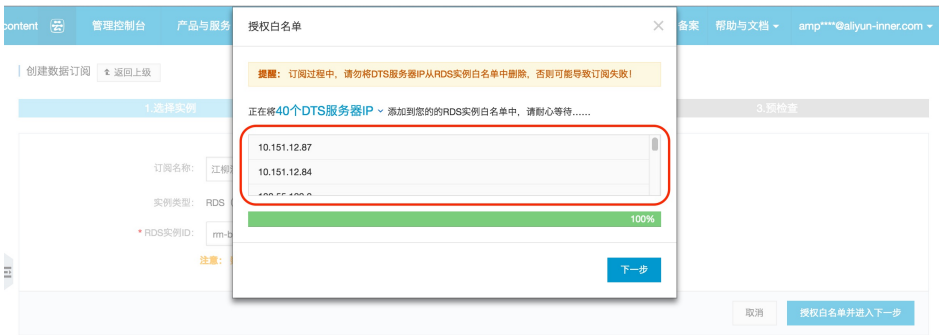
这个主要配置订阅通道名称及需要订阅的RDS实例ID。其中：

- 订阅名称：订阅名称为订阅通道别名，没有唯一性要求。默认情况下，DTS为每个订阅通道自动生成一个名称，可以修改这个名称，为其配置一个具备业务意义的名称，便于后续的通道识别。

  RDS实例ID: 需要订阅增量数据对应的RDS实例的实例ID。

  当这些内容配置完成后，可以点击页面右下角的 **授权白名单并进入下一步**。

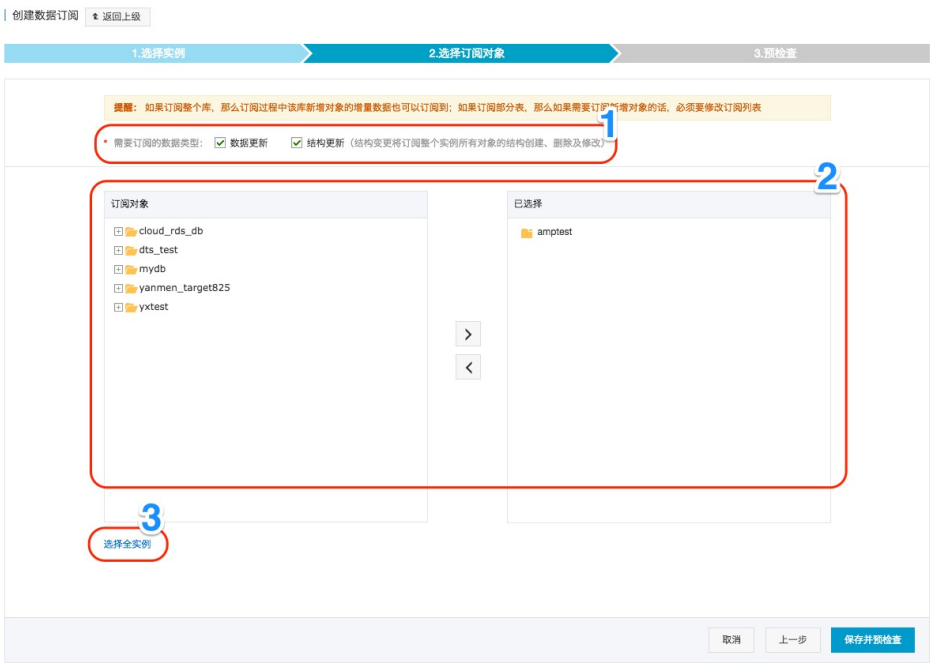
授权RDS实例白名单。



这个步骤，主要是将DTS服务器IP添加到订阅RDS实例的白名单中。避免因RDS配置了白名单，DTS服务器不能连接RDS实例，从而导致订阅通道创建失败。

为了保证订阅通道的稳定性，请勿将DTS服务器IP从RDS实例的白名单中删除。

选择订阅对象。



在这个步骤中，用户需要配置订阅的数据类型及订阅对象。其中:

订阅数据类型：DTS将订阅数据类型细分为**数据变更、结构变更**。

数据变更是指DML(包含insert/delete/update)。结构变更是指DDL(例如：create/drop/alter table)。

如果订阅了结构变更，那么DTS会将整个RDS实例的所有结构变更拉取出来，用户如果只需要部分DDL的话，需要在SDK消费的时候进行过滤。

订阅对象：DTS的订阅对象粒度细分为库、表。即用户可以选择订阅某些库或者是订阅某几张表。

当完成上面的配置后，即进入订阅通道启动之前的前置预检查阶段。

预检查。

当订阅通道配置完成后，DTS会进行前置预检查，当预检查通过后，可以点击右下角的 **开始订阅** ,启动订阅通道。

启动订阅。

当订阅通道刚创建完，会进入初始化阶段，这个阶段一般会持续1分钟左右。初始化完成后，即可在线查看订阅数据，或者使用SDK订阅增量数据。

至此，完成数据订阅通道的配置。