

DTS控制台一本通

云运维工程师从入门到精通

作者：手辨



- 阿里云工程师多年运维心得
- 真实案例手把手教你数据迁移
- 最新版DTS数据订阅功能详细说明





云服务技术大学
云产品干货高频分享



云服务技术课堂
和大牛零距离沟通



阿里云开发者“藏经阁”
海量免费电子书下载

目录

第一章	什么是 DTS	6
第二章	DTS 概览页	7
第三章	DTS 数据迁移	8
3.1	地域	8
3.2	文件导入	8
3.3	创建迁移任务页面	9
3.3.1	实例类型	10
3.3.2	实例地区	16
3.3.3	端口	18
3.3.4	数据库类型	18
3.3.5	数据库账号	19
3.3.6	数据库密码	21
3.3.7	连接方式	21
3.3.8	测试连接	22
3.3.9	授权白名单并进入下一步	23
3.3.10	迁移类型	26
3.3.11	预检查并启动	28
3.3.12	立即购买并启动	32
3.4	任务列表	33
3.4.1	ID/ 名称	33
3.4.2	查看原因并修复	43
3.4.3	启动任务	44

3.4.4	查看详情	44
3.4.5	创建类似任务	45
3.4.6	升级	45
3.4.7	监控报警	46
3.4.8	修改密码	46
3.4.9	暂停任务	46
3.4.10	结束任务	49
3.4.11	释放任务	49
第四章 DTS 数据订阅		50
4.1	创建数据订阅页面	50
4.1.1	商品类型	51
4.1.2	功能	51
4.1.3	订阅实例类型	51
4.1.4	源实例地域 - 订阅	52
4.2	任务列表	52
4.2.1	配置订阅通道	52
4.2.2	订阅 ID/ 名称	56
4.2.3	消费时间点	61
4.2.4	数据范围	61
4.2.5	修改订阅对象	61
4.2.6	新增消费组	61
4.2.7	更多	61
4.3	使用 Client 订阅数据	62
4.3.1	下载示例代码	62
4.3.2	导入示例代码	62
4.3.3	配置 Client	64

4.3.4	启动 Client	66
4.3.5	打印消费信息	66

第五章 DTS 数据同步 68

5.1	地域	69
5.2	创建同步作业页面	69
5.2.1	商品类型	69
5.2.2	功能	70
5.2.3	源实例	70
5.2.4	源实例地域	70
5.2.5	目标实例	70
5.2.6	目标实例地域	70
5.2.7	同步拓扑	70
5.2.8	网络类型	72
5.2.9	同步链路规格	72
5.3	任务列表	72
5.3.1	配置同步链路	73
5.3.2	实例 ID/ 作业名称	76
5.3.3	同步概况	78
5.3.4	暂停任务	78
5.3.5	升级	79
5.3.6	更多	79

第六章 操作日志 83

第一章 什么是 DTS

DTS 全称为数据传输服务 DTS，它最常见的用途是把本地数据库的数据迁移、同步到阿里云，或者把阿里云的数据迁移、同步到本地。它更多的是一个数据迁移产品，它最终做的事情就是把数据从源端抽取（复制源端数据，源端数据依然在）然后写入到目标端。另外，它还提供数据订阅的功能（关于数据订阅，后面章节会进行讨论），支持对增量数据进行订阅。简单的可以理解成下图这种架构，如图 1-1。



图 1-1

第二章 DTS 概览页

概览页应该是每一个多地域部署售卖的阿里云产品，都应该提供的功能。DTS 概览页展示了当前账户下，各个地域下，不同 DTS 任务类型的分布情况。如下图 2-1。点击对应的数字，可以跳转到具体的任务列表。

地域	迁移实例	订购实例	同步实例	失败	成功
华东1 (杭州)	56	6	35	17	39
华东2 (上海)	4		2	1	3
华北1 (北京)	45		19	14	31
华北2 (北京)	16	1	5	2	14
华东3 (深圳)	3		3	3	3
华东4 (青岛)					
华南1 (深圳)					
华南2 (海口)					

图 2-1

该页面，需要注意如下 1 点：

- 如果遇到概览页展示存在 DTS 实例，而点击数字进行跳转，或者在实际的迁移页面并没有实例存在，这个问题一般是后台数据问题需要阿里云进行订正。如果您遇到，请反馈阿里云售后核实。

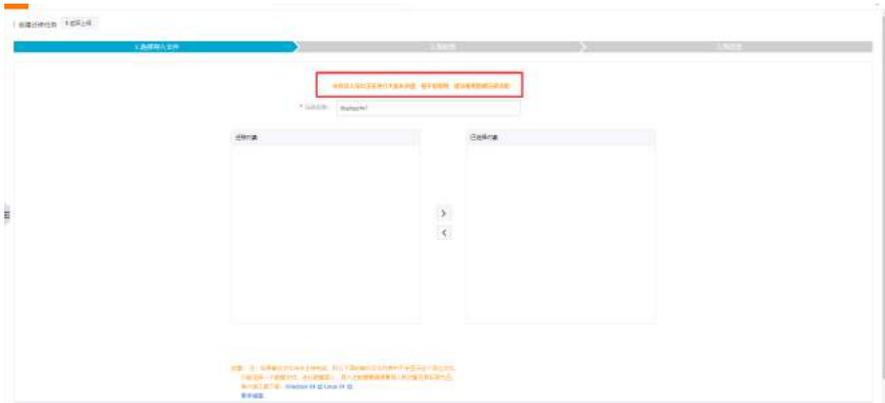


图 3-2

3.3 创建迁移任务页面

图 3-1 标记②的“创建迁移任务”，可以进行数据迁移的任务配置，点击后，会进入 DTS 迁移任务的配置界面，如图 3-3，需要注意：

- DTS 目前的实现是**逻辑迁移**。
- 逻辑迁移的意思是指 DTS 会借助 SELECT 的方式抽取源库已存在的数据，然后再通过 INSERT 的方式写入到目标库。
- 目前还不支持**物理文件**层面的迁移。

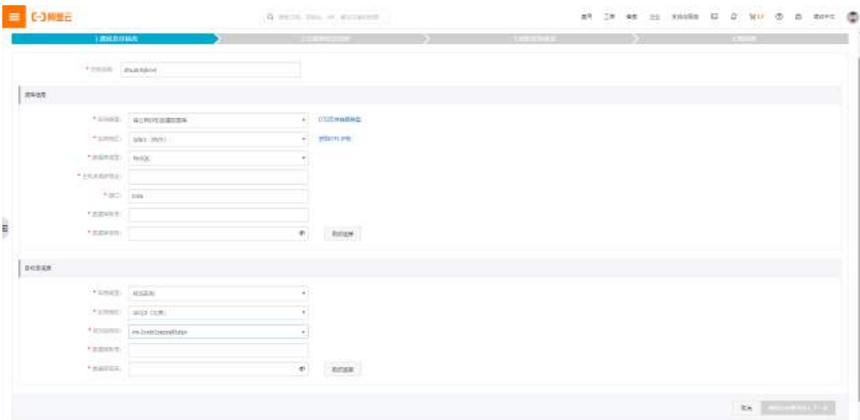


图 3-3

DTS 迁移任务主要分为 2 部分进行配置，即源库信息和目标库信息，源库是指要使用 DTS 迁移的源端的数据库，目标库是指待迁入数据的数据库。比如我们要把 MySQL A 数据库的数据迁移到 MySQL B 数据库，这里，MySQL A 数据库 A 就是源库。目标库就是 MySQL B 数据库。下面对源库以及目标库配置过程的具体项目进行讨论。

3.3.1 实例类型

如图 3-3，顾名思义，就是 DTS 支持迁移的源端数据库接入的类型，并且 DTS 对不同的接入类型的实例，有不同的支持粒度，主要分为如下几种。无论是哪一种，最终实现的核心目的是能让 DTS 的服务器集群连接到源端或者目标端的数据库。

3.3.1.1 有公网 IP 自建的数据库

这是最简单的一种接入类型，就是指您的源端的数据库有公网 IP 地址，这个地址允许其他程序（这里特指 DTS）可以通过您的公网 IP 连接到您的源端数据库，这种实例类型配置比较灵活，但是容易受到公网网络环境的影响，进而影响迁移的速率。当我们选择这个类型时，页面的信息也会按照“有公网 IP 自建的数据库”进行展示，如图 3-4。您需要把源端数据库的公网连接地址填写到“主机名或者 IP 地址”里，这个需要再次强调，这个 IP 地址必须能够让 DTS 服务器访问到，即连通性必须正常。如果连通性有问题，在“测试连接”时会出错。我们会在“测试连接”部分详细讨论。这里有 4 点需要说明：

- 虽然这里特指的是“自建数据库”，但是只要是公网可达的（比如 RDS 的公网地址），都可以使用这个方式进行传输。
- 这里指的 IP 并不是只可以填写 IP，也可以填写域名（比如 RDS 的公网连接字符串）。
- 自建 MySQL 数据库一般都有一个参数 `bind_address`，这个是指 MySQL 接受（监听）来自于哪个 IPV4 或者 IPV6 地址的连接，为了避免出现 DTS 无法连接的情况，建议设置成“*”或者“0.0.0.0”。

- 请一定确保，自建数据库所在主机的防火墙没有拦截 DTS 服务器地址的入方向的访问。服务器地址我们会在“[3.3.2 实例地区](#)”讨论。

The screenshot shows a configuration form for connecting to a self-built database. The fields are as follows:

- * 实例类型:** 有公网IP的自建数据库 (Dropdown menu)
- * 实例地区:** 华东1 (杭州) (Dropdown menu)
- * 数据库类型:** MySQL (Dropdown menu)
- * 主机名或IP地址:** (Empty text input)
- * 端口:** 3306 (Text input)
- * 数据库账号:** (Empty text input)
- * 数据库密码:** (Empty text input with a show/hide icon)

Additional elements include a blue link "DTS支持链路类型" next to the instance type field, a blue link "获取DTS IP段" next to the instance region field, and a grey button labeled "测试连接" (Test Connection) at the bottom right.

图 3-4

3.3.1.2 通过专线 /VPN 网关 / 智能接入网关接入的自建数据库

相比公网的不安全而言。这个方式实现了非“公网”传输。但是这里的非“公网”传输需要借助其他的产品来实现，这些产品是：[高速通道专线](#)、[智能接入网关](#)、[VPN 网关](#)。它们最终实现的是把本地自建数据库和阿里云的链路打通。通过非“公网”的方式进行连接。最终只要连通性正常。就可以进行传输。如下图 3-5。这里有 5 点需要特别说明：

- 虽然它们都可以实现非“公网”方式连接。但是实现方式是不一样的。专线才是真正意义上的内网传输。而智能接入网关和 VPN 网关都是依赖公网实现的。只是实现了非“公网”方式连接（打通了本地与阿里云内网）。
- 选择这个方式，需要填写 VPC ID（“已和源端数据库联通的 VPC”）。请注意，这里不是随便填写一个 VPC ID 就可以了。这个 VPC ID 必须进行了[高速通道专线](#)、[智能接入网关](#)、[VPN 网关](#)的相关配置打通了链路才可以。没有做任何配置的 VPC ID 即使填写上也无法联通（测试连接会出错）。具体配置方式涉及各个产品的较多内容，您如果需要可参考具体产品的帮助文档。

- 虽然这里特指的是“自建数据库”，但是只要是通过这三种方式实现网络可达的（比如 VPC 下的 RDS），都可以使用这个方式进行传输。
- “IP 地址”这部分，只可以填写 IP 地址。如何填写域名或者字符会提示“请输入合法的 ip 地址”。
- 配置好高速通道专线、智能接入网关、VPN 网关之后，还需要配置 DTS 与它们之间路由，配置路由需要参考：https://help.aliyun.com/document_detail/117525.html?spm=a2c4g.11186623.6.590.2d2f6487R0gxRt。

The screenshot shows a configuration form for DTS with the following fields and options:

- * 实例类型:** 通过专线/VPN网关/智能接入网关接入的自建数据库 (Dropdown menu)
- * 实例地区:** 华东1 (杭州) (Dropdown menu)
- * 已和源端数据库联通的VPC:** 请选择专有网络 (Dropdown menu)
- * 数据库类型:** MySQL (Dropdown menu)
- * IP地址:** (Text input field)
- * 端口:** 3306 (Text input field)
- * 数据库账号:** (Text input field)
- * 数据库密码:** (Text input field with a show/hide icon)

Additional links and text on the right side of the form include: "DTS支持链路类型", "操作指南", and "其他阿里云账号下的专有网络". A "测试连接" button is located at the bottom right.

图 3-5

3.3.1.3 无公网 IP:Port 的数据库 (通过数据库网关 DG 接入)

如下图 3-6，选择[这个接入方式](#)，也可以实现非“公网”接入。它的实现方式是要在源端数据库所在的主机上安装一个数据库网关。这个数据库网关与智能接入网关、VPN 网关相似，依赖于公网。但是免去了复杂的配置步骤与较高的成本。这种接入方式的使用量较少。使用时需要注意如下 1 点：

- 这里配置时需要选择“数据库网关 ID”，这里不是选择了 ID 就可以了，还需要对这个网关进行安装和添加数据库。也就是要保证网关“状态”运行正常。

* 实例类型:	<input type="text" value="无公网IP:Port的数据库(通过数据库网关DG接入)"/>	DTS支持链路类型
* 实例地区:	<input type="text" value="华东1 (杭州)"/>	
* 数据库网关ID:	<input type="text" value="请选择数据库网关ID"/>	创建数据库网关
* 数据库类型:	<input type="text" value="MySQL"/>	
* 数据库地址:	<input type="text"/>	
* 数据库账号:	<input type="text"/>	
* 数据库密码:	<input type="text"/>	<input type="button" value="测试连接"/>

图 3-6

3.3.1.4 通过云企业网 CEN 接入的自建数据库

如下图 3-7，[云企业网](#)简称 CEN(Cloud Enterprise Network)，它不是一种具体的接入方式，即它并不是专线、VPN 这种连接能力。它是一个能力提供者或者平台。它提供一种组网的能力，可以实现专线、VPN 等的网络互通与管理。选择这个接入方式后，需要选择“云企业网实例 ID”以及“已和源库互联的 VPC 网络”。这里特指“自建数据库”。但是无论是自建还是云数据库，只要连接打通，都可以使用这种方式。

* 实例类型:	<input type="text" value="通过云企业网CEN接入的自建数据库"/>	DTS支持链路类型
* 实例地区:	<input type="text" value="华东1 (杭州)"/>	
* 云企业网实例ID:	<input type="text" value="请选择云企业网实例ID"/>	
* 已和源库互联的VPC网络:	<input type="text"/>	
* 数据库类型:	<input type="text" value="MySQL"/>	
* IP地址:	<input type="text"/>	
* 端口:	<input type="text" value="3306"/>	
* 数据库账号:	<input type="text"/>	
* 数据库密码:	<input type="text"/>	<input type="button" value="测试连接"/>

图 3-7

3.3.1.5 ECS 上的自建数据库

顾名思义，就是指在阿里云 ECS 服务器上搭建的自建数据库，如下图 3-8，选择这个接入方式非常适合 ECS 服务器上的数据库进行迁移。您只需要选择对应的“ECS 实例 ID”，填写相关数据库信息即可。此处需要注意 2 点：

- ECS 服务器有安全组的限制。正常情况下，当选择完 ECS 的实例 ID，填写完数据库的连接信息，然后点击 DTS 源端数据库的“测试连接”后，DTS 会自动把 DTS 服务器的 IP 端添加到 ECS 的安全组里面。如图 3-9。如果您遇到连通性问题，请先检查这一点是否正常。
- 除了安全组，ECS 上部署的操作系统内还有防火墙，请确保防火墙也做了响应的入方向放行规则。

* 实例类型: ECS上的自建数据库 DTS支持链路类型

* 实例地区: 华东1 (杭州)

* ECS实例ID: 请选择ECS实例

* 数据库类型: MySQL

* 端口: 3306

* 数据库账号:

* 数据库密码:

* 连接方式: 非加密连接 SSL安全连接

测试连接

图 3-8

访问规则

入方向 出方向

手动添加 快速添加 全部编辑

操作策略	优先级	协议类型	端口范围	接收对象	规则	创建时间	操作
允许	1	全部	目标-1/1	源:100.104.12.0/24	DTS源端添加安全组, 请勿删除	2020年4月19日 16:47:09	编辑 删除
允许	1	全部	目标-1/1	源:100.104.41.128/26	DTS源端添加安全组, 请勿删除	2020年4月19日 16:47:09	编辑 删除
允许	1	全部	目标-1/1	源:118.55.129.0/24	DTS源端添加安全组, 请勿删除	2020年4月19日 16:47:09	编辑 删除
允许	1	全部	目标-1/1	源:10.151.112.0/24	DTS源端添加安全组, 请勿删除	2020年4月19日 16:47:09	编辑 删除

图 3-9

3.3.1.6 RDS 实例

这里特指您购买的阿里云的 **RDS 数据库** 实例，如下图 3-10，选择这个接入方式后，只需要填写对应的“RDS 实例 ID”以及账密即可。DTS 后台会自动的通过实例 ID 查询对应 RDS 的连接地址和端口。并且，DTS 还支持“其他阿里云账号下的 RDS 实例”的迁移。您可以在阿里云 B 账号下迁移阿里云 A 账号的 RDS 数据库。跨账号迁移的操作与配置比较复杂，可以参考该云栖文档：<https://yq.aliyun.com/articles/353204?spm=a2c4e.11155435.0.0.6ff363b5zVAhec>。

此处注意如下 1 点：

- RDS 数据库产品有一个安全限制，就是白名单。同“[3.3.1.5 ECS 上的自建数据库](#)”的自动添加安全组行为类似，DTS 也会自动把 DTS 服务器的地址段添加到 RDS 数据库的白名单中。添加完后的白名单信息您无法在云 RDS 白名单页面看到。

* 实例类型: RDS实例 DTS支持链路类型

* 实例地区: 华东1 (杭州)

* RDS实例ID: 请选择RDS实例 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

图 3-10

3.3.1.7 云 MONGODB 实例

这里特指您购买的阿里云的 **MONGODB 数据库** 实例，如下图 3-11，选择这个接入方式后，只需要填写对应的云 MONGODB 的实例 ID、认证数据库以及账密即可。同“[3.3.1.6 RDS 实例](#)”的自动添加白名单行为一样，DTS 也会自动把 DTS 服务器的地址段添加到云 MONGODB 数据库的白名单中。添加完后的白名单信息您无法在云 MONGODB 的白名单页面看到。

* 实例类型: 云数据库MongoDB DTS支持链路类型

* 实例地区: 华东1 (杭州)

* MongoDB实例ID: dds-bp1bd82d5c7c7644

* 数据库名称: 账号验证数据库

* 数据库账号:

* 数据库密码: 测试连接

图 3-11

3.3.1.8 PolarDB

这里特指您购买的阿里云的 [PolarDB 数据库](#)实例，如下图 3-12，选择这个接入方式后，只需要填写对应的云 PolarDB 的实例 ID 以及相关的连接信息即可。同“[3.3.1.6 RDS 实例](#)”的自动添加白名单行为一样，DTS 也会自动把 DTS 服务器的地址段添加到云 PolarDB 数据库的白名单中。添加完后的白名单信息您无法在云 PolarDB 的白名单页面看到。

* 实例类型: PolarDB DTS支持链路类型

* 实例地区: 华东1 (杭州)

* PolarDB实例ID: 请选择PolarDB实例

* 数据库账号:

* 数据库密码: 测试连接

图 3-12

3.3.2 实例地区

图 3-3 中的“实例地区”选择后也就意味着 DTS 会使用这个地区的服务器

(DTS 在很多地区都部署了服务器) 连接源端数据库进行数据的抽取和传输 (简单说就连接数据库后执行 select 查询获取数据)。选择不同的实例类型, 在实例地区选择上有 2 点不同, 主要如下:

- 如果实例类型选择的是“有公网 IP 自建的数据库”, 实例地区这里理论上选择任何一个都可以 (因为公网 IP 任何地方都可达), 建议选择与数据库所在地域物理距离相近的地区。比如自建数据库的机房在北京, 则建议选择华北 2 地区。
- 除“有公网 IP 自建的数据库”外的其他实例类型在选择实例地区时, 则需要按照实例所在的地域进行选择。比如 RDS 实例在华东 1, 则地域必须选择华东 1。这样的话, DTS 的华东 1 服务器集群才可以正常连接华东 1 的 RDS。

重点再说明一下实例类型部分说明的问题, 因为 DTS 的服务器非常多, 为了能够让 DTS 顺利的链接上源端的数据库, 假设您源端数据库做了防火墙 (自建)、安全组 (ECS)、白名单 (RDS) 等安全设置, 您还需要把 DTS 这个地区的所有服务器 (有的客户不想放行所有服务器, 这个暂时无法满足) 的地址进行放行。需要放行的 DTS 服务器的网段点击图 3-3 的“获取 DTS IP 段”查看, 如下图 3-13。



3.3.3 端口

图 3-3 的“端口”部分用来指定需要 DTS 访问的数据库端口，只有某些实例类型才需要填写该项目。指定了 IP 只能说明您的数据库运行在这台主机上，但是这台主机上可能运行了很多的应用程序。端口就用来说明要访问哪个应用程序。请注意您的网络防火墙或者安全组（特指 ECS）配置。对对应端口的访问进行放行。

3.3.4 数据库类型

图 3-3 的“数据库类型”部分用来选择迁移的数据库类型，目前阿里云 DTS 支持的源端数据库类型有：Oracle、MySQL、SQLServer、PostgreSQL、Mongodb、DB2、TIDB。这里需要正确的选择源端数据库的类型，如果您源端是 MySQL 则需要选择 MySQL。您选择什么样的数据库类型意味着 DTS 的应用程序使用哪种数据库驱动进行连接您的源端数据库，所以请务必选择正确。这里有 3 个注意点：

- 源库的数据库类型与目标库的数据库类型建议一致，因为兼容性最好。以下图 3-14 为例，当选择实例类型为 RDS，实例 ID 是一台 MySQL 实例时，目标端数据库类型或者实例类型时会展示出 SQLServer、Postgresql。目前 DTS 除了支持 MySQL->Oracle、Mysql->Postgresql、MySQL->MySQL、MySQL-> PolarDB MySQL、MySQL-> AnalyticDB MySQL、MySQL-> DRDS 外，其他的数据库类型暂不支持，所以如果此时目标选择 SQLserver 实例，迁移会出现异常。其他的实例类型和数据库类型同理。
- 若源端数据库为 Sqlsever 时，当开启增量迁移时，一个 DTS 任务只允许一个数据库进行增量迁移（这取决于其事务日志）。全量迁移支持多个数据库。
- 若源端数据库为 PostgreSQL 时，无论是否开启增量迁移，一个 DTS 任务只允许迁移一个数据库（这取决于其 XLOG）。

The image shows two screenshots of the DTS configuration interface. The top screenshot is for MySQL, with the 'RDS实例ID' dropdown menu open, showing 'i863t2'. A red arrow points to this dropdown with the text 'MySQL' and '其他阿里云账号下的RDS实例'. The bottom screenshot is for SQLserver, with the 'RDS实例ID' dropdown menu open, showing ':46ch'. A red arrow points to this dropdown with the text 'SQLserver'. Both screenshots include fields for '实例类型', '实例地区', '数据库账号', and '数据库密码', along with a '测试连接' button and a '连接方式' section with radio buttons for '非加密连接' (selected) and 'SSL安全连接'.

图 3-14

3.3.5 数据库账号

图 3-3 的“数据库账号”指的是您需要 DTS 使用哪个账号连接您的源端的数据库。这里出现问题最多的是 MySQL 数据库的账号问题以及 MONGODB 数据库的账号问题。主要有如下 4 点：

- MySQL 的数据库账号组成是 user 和 host 两个元素，即：user@host 这种格式。这就意味着，如果 user 相同而 host 不同，这不是相同的账号。比如如下 2 个账户：alitest@%' 与 alitest@'10.0.0.1'。当您使用“alitest”账户在 IP 为 10.0.0.1 的客户端主机访问您的数据库的时候，鉴权时账户用的是 alitest@'10.0.0.1'。当您使用“alitest”账户在 IP 为 10.0.0.2 的客户端主机访问您的数据库的时候，由于“alitest”的 host 只有 % 和 10.0.0.1。而 10.0.0.1 并不符合，所以鉴权时账户用的是 alitest@%'。了解这点非常重要。

- 所以，我非常建议您创建一个独立的账号进行 DTS 的迁移，并且账号的 host 建议是 '%'。因为 DTS 服务器的网段非常多，并且没有规律（请参考图 3-5）。% 可以有效的避免账户连接多问题。
- MySQL 遇到最多的数据库账号连接类问题是这个错误：
Access denied for user 'xxxx'@'xxxx' (using password: YES)
这个错误在连接时产生的原因有 2 个：账户错误或者密码错误。您如果遇到，请参考这里排查：
<https://yq.aliyun.com/articles/158321?spm=a2c4e.11155435.0.0.6ff-363b5cXNuax>
- MONGODB 的账户链接时，还需要填写一个“数据库名称”项，如下图 3-15。这个是指 MONGODB 账户的 [authentication database](#)。authentication database 是指创建 MONGODB 账户的时候所在的数据库。比如一个 MONGODB 数据库为 alitest。我们执行 use alitest 切换到 alitest 数据库下，然后执行 db.createUser() 创建一个 test 账户，密码是 123。当我们使用 test 账户登陆 mongodb 的时候，数据库名称 (authentication database) 就是 alitest。

The screenshot shows a configuration form for DTS. The fields are as follows:

- * 实例类型: 有公网IP的自建数据库 (Dropdown menu)
- * 实例地区: 华东1 (杭州) (Dropdown menu)
- * 数据库类型: MongoDB (Dropdown menu)
- * 主机名或IP地址: (Empty text input)
- * 端口: 27017 (Text input)
- 数据库名称: (Empty text input, highlighted with a red box)
- 数据库账号: (Empty text input)
- 数据库密码: (Empty text input with a show/hide icon)
- * 连接方式: 非加密连接 SSL安全连接(该选项目前仅支持MongoDB Atlas版本)

Additional elements: "DTS支持链路类型" (blue link), "获取DTS IP段" (blue link), and a "测试连接" (Test Connection) button.

图 3-15

3.3.6 数据库密码

图 3-3 的“数据库密码”指的是使用的“[数据库账号](#)”的连接密码。数据库密码遇到的问题较少。请重点关注“[数据库账号](#)”章节的相关问题。

3.3.7 连接方式

当实例类型选择“ECS上的自建数据库”，数据库类型选择“MySQL”时，DTS支持连接方式的选择，即支持SSL加密的连接方式。当勾选“SSL安全连接”时，需要上传SSL的证书等文件，如下图 3-16。其中CA根证书是必传的。另外，DTS目前只支持SSL加密。请注意如下 2 点：

- 如您要使用 SSL 安全连接，请确保您的源端数据库以及目标端数据库做了相应的 SSL 配置，关于如何配置数据库的 SSL 非本书的重点。不在此赘述。
- 并非所有的实例类型以及数据库类型 DTS 都支持 SSL 访问。

* 实例类型: ECS上的自建数据库 DTS支持该类型

* 实例地区: 华东1 (杭州)

* ECS实例ID: 请选择ECS实例

* 数据库类型: MySQL

* 端口: 3306

* 数据库账号:

* 数据库密码:

* 连接方式: 非加密连接 SSL安全连接

* CA根证书:

Client证书:

Client私钥:

图 3-16

3.3.8 测试连接

重点中的重点，图 3-3 的“测试连接”点击后，会使用我们配置的这些连接信息去连接源端或者目标端的数据库，针对 ECS、RDS、云 MONGODB、云 REDIS 等云数据库，还会进行安全组、白名单的检查和添加操作，检查对应的云 ECS 实例和云数据库实例的安全组、白名单里是否有 DTS 的安全组或者白名单，没有则添加。测试分为 ping、telnet、数据库协议（比如 MySQL JDBC Connect）三个层面，测试结果如下图 3-17。需要注意如下 4 点：

- 如果测试连接失败。分为 ping 失败，telnet 失败以及数据库协议（比如 MySQL JDBC Connect）失败。我们只需要关注 telnet 和数据库协议即可。因为 ping 现在的测试结果并不能说明问题，可以忽略它的测试结果。
- telnet 如果成功，说明 DTS 可以正常通过您配置的地址与端口访问到对应的应用程序。如果失败，则说明 DTS 无法通过您配置的地址与端口访问到您的应用程序。此时则需要进行网络防火墙或者安全组的排查或者抓取网络报文排查。
- 数据库协议（比如 MySQL JDBC Connect）如果成功，则说明 DTS 可以通过您配置的地址、端口、数据库账号、数据库密码等信息，成功访问到您的数据库。如果失败，一般会有错误提示，对 MySQL 来说，最常见的是 Access denied for user 'xxxx'@'xxxx' (using password: YES)，这点，我们在“[3.3.5 数据库账号](#)”讨论过。其他的连接问题，可以参考这个云栖文档：<https://yq.aliyun.com/articles/158321?spm=a2c4e.11155435.0.0.6ff-363b5aL5jmA>
- 有时点击测试连接后，会一直转圈，没有任何结果。遇到这个情况。可以忽略测试连接，直接点击“授权白名单并进入下一步”。换句话说，测试连接并不是必须要点击的。如果配置的源库连接有问题，我们点击“授权白名单并进入下一步”跳转到新页面时会出错（因为此时会真正的去源库取数据库的对象信息，如下图 3-18）。如果配置的目标库连接有问题，我们在“预检查”时会出错（因为预检查会检查目标库连接）。所以测试连接一般情况下可点可不点。大多数情况下不需要执着于这一步。

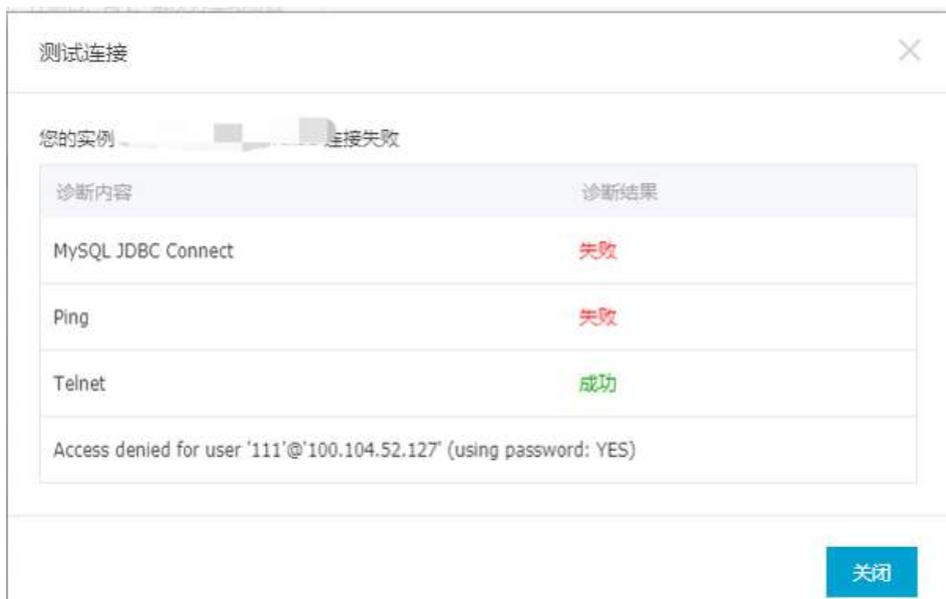


图 3-17

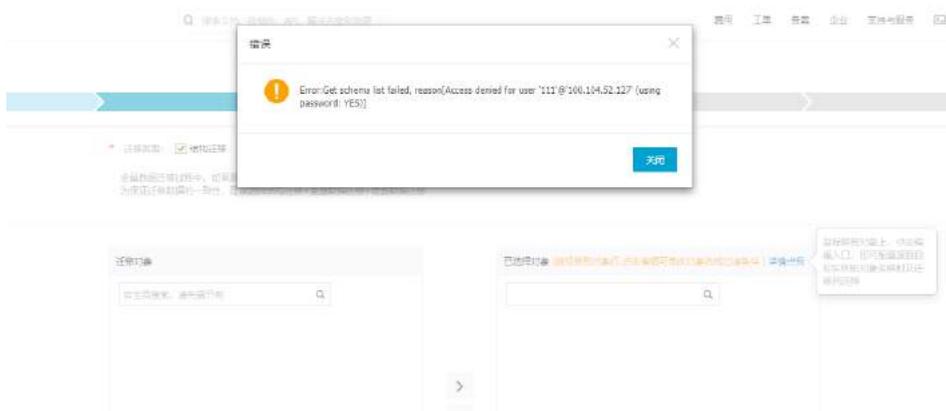


图 3-18

3.3.9 授权白名单并进入下一步

图 3-3 的“授权白名单并进入下一步”点击后，会和测试连接一样，也会进行安全组与白名单的检查和添加，然后会进入数据库迁移对象选择页面，如下图

3-19。这是一个库级别（数据库级别的同步，DTS 还支持表级别和列级别同步）同步的任务截图。这个页面的左侧显示的是源端数据库的对象信息（如图例里的 dtstest、mysqltest 数据库）。右侧显示的是要迁移到目标数据库的对象信息（如图例里的 dtstestdata 数据库），如果您要更改迁移到目标数据库的对象的名字，可以把鼠标放到“dtstestdata”上，然后点击右侧的编辑，更改库名（除了更改库名外，表级别同步还支持修改表名以及列名）。该功能为 DTS 的库表列映射，如下图 3-20。

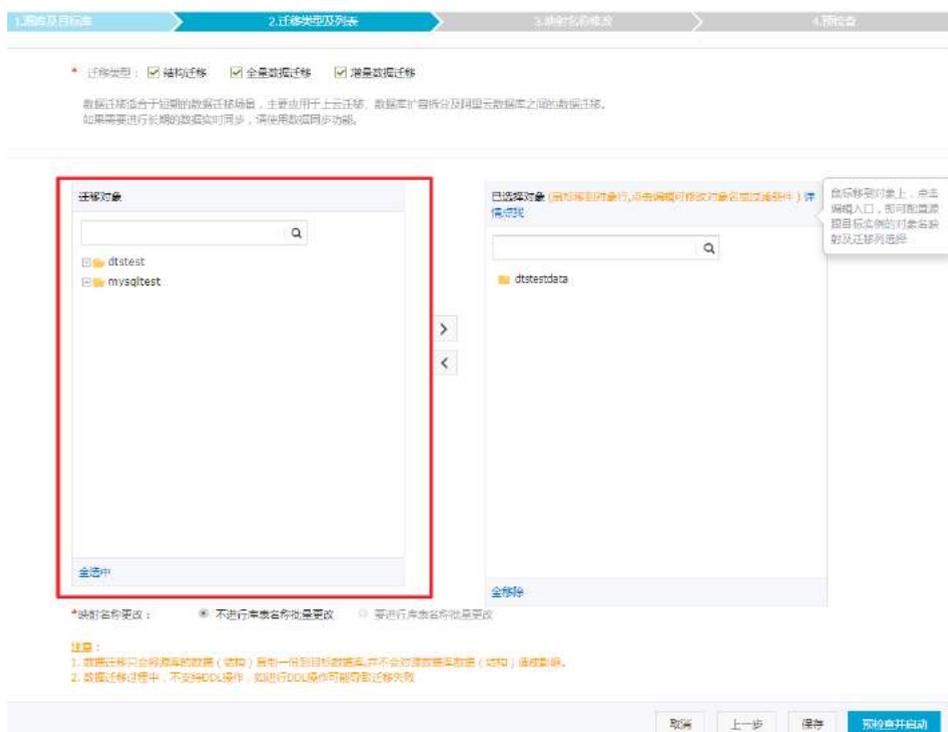


图 3-19



图 3-20(1)



图 3-20(2)

3.3.10 迁移类型

3.3.10.1 结构迁移

图 3-19 里的“结构迁移”是指是否迁移源端数据库的对象结构信息。这些结构包括表结构、索引、视图、存储过程等。DTS 会通过 Select 的方式，获取源库的数据对象的结构。为什么要进行结构迁移？什么时候需要结构迁移？原因分别如下：

- 结构迁移针对的是关系型数据库，非关系型数据库不需要（比如 MongoDB），对关系型数据库来说，数据要存储，必须先建表。所以 DTS 需要先进行结构迁移，创建表结构才能进行后面的数据迁移。
- 当目标数据库里面没有对应的表对象时，才需要选择结构迁移。如果目标数据库里面已经有了对应的数据库的表对象，不需要选择结构迁移（比如目标库已经有表 A，此时选择结构迁移，DTS 会在目标库创建表 A，因为表 A 已经存在，DTS 创建会出错，提示 1050 - Table xxxx already exists）。

结构迁移遇到的最常见的问题是：

- 源端或者目标端的数据库的数据库对象非常非常多，因为 DTS 要以 Select 的方式查询这些信息。数据库对象过多，容易造成查询超时，这类问题比较难解决。如果您遇到这种问题，请反馈阿里云售后。

如果要查看 DTS 在源端或者目标端的进程状态，大多数数据库都可以在源端或者目标端的执行数据库的相关查询命令，以 MySQL 为例，可以执行 `:show processlist` 确认会话信息。

3.3.10.2 全量数据迁移

图 3-19 里的“全量数据迁移”是指是否迁移源端数据库的数据（表的记录）。这是真正的数据的迁移，它迁移的是源端数据库表里已经存在的数据（这些数据可能是历史数据，早已经写入，也可能刚刚写入不久的数据，非未来新增的数据），简单的说，DTS 会通过 Select 获取源端数据库里的表的数据。然后通过 Insert、Replace into、Update、Delete 的方式，写入到目标端。DTS 使用 Select 查询对应表的数据，不是一性全部查询的。一般是对的数据进行分片（常见的是使用主键

进行分割), 然后并行查询各个分片的数据。DTS 写入目标端数据也是并行写入的。这也就导致有 2 个问题:

- DTS 迁移完成后, 目标数据库的数据空间大小比源端数据库的数据空间大小大, 这是因为并行写入产生数据空洞导致。如果您遇到这种问题, 可以在业务允许的期间执行 `Optimize table xxx` 对表的空间进行收缩 (Optimize table 有产生 MDL 阻塞的可能, 执行时请评估执行时间和业务影响)。
- Optimize table 大多数情况下, 都能够成功的对表的空间进行收缩。极少情况下无法对表的空间进行收缩。如果遇到无法收缩的情况, 可以反馈阿里云售后。

如果要查看 DTS 在源端或者目标端的进程状态, 大多数数据库都可以在源端或者目标端的执行数据库的相关查询命令, 以 MySQL 为例, 可以执行 `:show processlist` 确认会话信息。

3.3.10.3 增量数据迁移

图 3-19 里的“增量数据迁移”是指 DTS 通过解析源端数据库的相关数据库日志 (比如 MySQL 的 Binlog、SQLserver 的 Transaction Log、MONGODB 的 oplog 等), 把全量迁移开始之后的增量数据, 实时的同步到目标端。即源端有一条 Insert, DTS 会解析这些日志, 生成一个同样的 Insert 发送到目标端, 实现增量数据迁移, 增量迁移目前不支持触发器的迁移。

很多客户在问, DTS 是如何保证数据完整性的? DTS 通过全量 + 增量的方式保证数据完整性, 但是在如下 2 个情况下, DTS 无法保证数据的完整性:

- 迁移的对象里, 没有主键或者唯一键的数据库对象, 无法保证完整性。因为 DTS 是分批次抽取和并发写入数据, 没有主键时会造成数据重复或者数据缺失的情况。请确保迁移对象有主键或者唯一键。
- 多对一的相同数据库对象的数据迁移, 无法保证数据完整性。多对一是指多个源库通过多个任务, 把源端多个数据库对象的数据迁移到目标库的同一个数据库对象里。

如果要查看 DTS 在源端或者目标端的进程状态, 大多数数据库都可以在源

端或者目标端的执行数据库的相关查询命令，以 MySQL 为例，可以执行 :show processlist 确认会话信息。DTS 的增量会在源端的 MySQL 数据库启动一个 binlog dump 进程(如果您搭建过 MySQL 主从，MySQL 本身的主从也是会启动一个 binlog dump 进程)。

DTS 的增量数据迁移延迟是无法保证的，正常情况下 DTS 的增量迁移是秒级延迟，但是当遇到一些 DDL、大量更新时或者 DTS 规格达到瓶颈等情况时，增量数据迁移延迟会增高。如果您遇到大的延迟(比如超过 1000S)，可与阿里云售后反馈。

3.3.11 预检查并启动

当迁移对象以及迁移类型配置好后，点击“预检查并启动”将会进行下面的检查环节，检查环节顺利通过后，会启动 DTS 任务开始迁移数据。检查的目的是避免一些已知的问题导致 DTS 迁移异常，常见的检查项目如下图 3-21。预检查失败后，无法进行下一步，需要您针对出错的检查项进行处理，当您处理了预检查失败的检查项后。可以点击启动任务(启动方式为图 3-1 中任务列表的“启动任务”按钮。)重新启动预检查。



图 3-21

3.3.11.1 源库连接性检查

检查**预检查的 DTS 服务器**是否可以通过迁移任务页面的相关配置正常连接源端数据库。需注意如下 2 点：

- DTS 的服务器有很多，这些服务器有的部署了预检查程序，有的部署了“测试连接”程序，有的部署了结构迁移的程序，有的部署了全量迁移的程序，有的部署了增量迁移的程序。这也就意味着，任何一个环节的连接正常都不能保证下一个或者其他环节连接也正常（比如数据库等防火墙做了限制）。您可能会遇到预检查正常但是后续的结构迁移连接异常，也可能遇到结构迁移任务正常但是后续的全量迁移连接异常的情况。这些都是有可能的。
- 遇到连接异常的情况，首先要有一个判断，这个异常是网络本身 (TCP/IP) 不通还是应用层面的问题。如果是网络层面的，则需要查防火墙以及链路配置。如果网络没问题只是在应用层出现的异常，则需要查造成这个应用异常的原因。举一个例子，前面提到过的这个错误：Access denied for user 'xxxx'@'xxxx' (using password: YES)。这个错误意味着，网络可达，只是在数据库进行账密认证的时候出错了（应用层）。如果网络不可达，我们甚至无法进行数据库应用层面的这些认证。了解这点非常重要。

3.3.11.2 源库权限检查

要使 DTS 通过您在创建迁移任务页面填写的账户信息获取源端数据库的数据，就需要对这个账户进行一写数据库对象的授权。不同的数据库类型以及迁移类型需要不同的数据库权限。以 MySQL 为例，结构、全量迁移只需要 SELECT 权限即可。而增量迁移则需要 REPLICATION SLAVE、REPLICATION CLIENT、SHOW VIEW 和 SELECT 权限。

3.3.11.3 目的库连接性检查

检查**预检查的 DTS 服务器**是否可以通过迁移任务页面的相关配置正常连接目标端数据库。注意点与“[3.3.11.1 源库连接性检查](#)”相同。

3.3.11.4 目的库权限检查

要使 DTS 把从源端获取到的数据写入目标端数据库，也需要对应的写入权限。同样以 MySQL 为例，结构、全量和增量需要 SELECT 和 INSERT、UPDATE、DELETE、CREATE 等等读写的权限。

3.3.11.5 存储引擎检查

这个一般是针对 MySQL 数据库的迁移，因为 MySQL 有很多数据库引擎。DTS 不支持 FEDERATED、MRG_MyISAM 或 TokuDB 存储引擎的迁移。

3.3.11.6 源库版本检查

检查源库的版本是否符合 DTS 产品的要求，以 MySQL 为例，DTS 只支持如下几个版本：5.1、5.5、5.6、5.7、8.0。其他数据库类型的版本限制可以参考这里：https://help.aliyun.com/document_detail/26618.html?spm=a2c4g.11186623.2.7.3a392a8770k06x#concept-26618-zh。

3.3.11.7 同名对象存在性检查

目的是检查目标数据库中是否存在和待迁移对象同名的对象。当我们选择结构迁移的时候，DTS 会在目标创建对象结构，如果目标数据库已经有同名的对象结构了，就会创建失败。

这点我们在“3.3.10.1 结构迁移”讨论过。

3.3.11.8 数据库可用性检查

检查目标实例是否已经有了对应的数据库，比如我们要把源端 alitest 的数据库迁移到目标端。我们需要先在目标端手动创建 alitest 实例。避免预检查失败，请注意如下 1 点：

- 很多情况下，DTS 会自动的在目标端创建数据库，DTS 自动创建数据库的前提是账户的权限正常。不过也有例外（比如字符集问题，数据库名称不合法等），如果遇到 DTS 无法自动创建数据库的情况，请手动创建后重启预检查。

3.3.11.9 源库 binlog 开启检查

当迁移类型勾选增量迁移时，才会检查该项，在“3.3.10.3 增量数据迁移”我们讨论过，增量迁移是通过一些日志来实现的。这个检查项是为了检查数据库这方面的日志是否开启。如果源端是 MySQL 数据库，则检查 binlog。

3.3.11.10 源库 binlog 模式检查

开启了相关日志后，要检查源数据库的 Binlog 模式是否为 ROW。建议更改 .cnf 配置文件，改完配置文件的参数后，请对源端数据库进行一次重启。

3.3.11.11 源库 binlog_row_image 是否为 FULL

如果您源端是 MySQL5.6 及以上的数据库版本时，开启增量迁移后，会检查 binlog_row_image 参数。建议更改 .cnf 配置文件，改完配置文件的参数后，请对源端数据库进行一次重启。

3.3.11.12 源库 server_id 检查

如果您源端是 MySQL 数据库，则开启增量迁移后，会检查源端数据库的 server id 参数。server_id 值需要是大于 2 的整数。请注意如下 1 点：

- 建议更改 .cnf 配置文件，更改完配置文件参数后，请对源端数据库进行一次重启

3.3.11.13 源库 binlog 存在性检查

这个检查项目主要检查 show binary logs 列出来的的文件是否存储，检查这个的目的是因为很多情况下，我们删除 MySQL 的 binlog 时，是直接从操作系统目录 remove 掉的，并没有通过 MySQL 的 purge 进行删除。这也就导致数据库里还存在着 binlog 文件的记录，而操作系统目录里已经没有了。遇到这种情况出现，您可以通过执行 purge 来清理 MySQL 数据库记录的 binlog 文件。使之与操作系统目录的文件数量和编号对应。

3.3.11.14 MySQL 密码格式检查

检查源库使用数据库密码格式是否为老版。主要检查数据库的参数 old_passwords 是否为 1。如果您的 old_passwords 为 1，建议修改为 0。修改密码格式可能会影响业务的正常运行，请谨慎操作。

3.3.11.15 复杂拓扑存在性检查

DTS 支持的迁移拓扑一般有如下 5 种：一对一、一对多、级联、多对一、双向迁移（非双向同步），其中双向迁移是不支持的拓扑结构，如果后台检测到这种拓扑结构，会出错。如果这个检查项您出现预检查失败，一般可以联系和授权阿里云售后跳过概检查项（预检查检测环节可以手动跳过的前提是双向迁移的数据对象以及数据不冲突，避免造成数据混乱）。

3.3.11.16 补充

如果预检查出现这种错误检查项“unexpected error”，请与阿里云售后反馈确认。

3.3.12 立即购买并启动

预检查通过后，点击下一步，进入购买页面，如下图 3-22。选择规格后，点击“立即购买并启动”正式运行 DTS 任务。这里需要注意如下 3 点：

- 对 DTS 迁移来说，只进行结构迁移和全量迁移，目前是免费的
- 选择增量迁移后，才会收费，暂停同样收费。
- DTS 迁移只支持后付费（按量计费），不支持包年包月。

购买配置确认 ×

链路地域: **华东1 (杭州)**

链路规格: small medium **large**
xlarge 2xlarge

付费类型: 按使用时间付费

配置费用: **1.25** /小时 ?

公网流量费: **¥ 0.00** /GB

注: 增量迁移完成后, 请及时释放实例, 从而停止费用。查看计费详情

《数据传输 (按量付费) 服务条款》

取消 立即购买并启动

图 3-22

3.4 任务列表

图 3-1 标记③的处为所选择地域的任务列表。包括了预检查中、失败、迁移中、暂停、完成等各个状态的任务，具体功能如下。

3.4.1 ID/名称

显示 DTS 任务的 ID 以及名称，点击这个超链接后，进入任务详情页面，如下图 3-23。

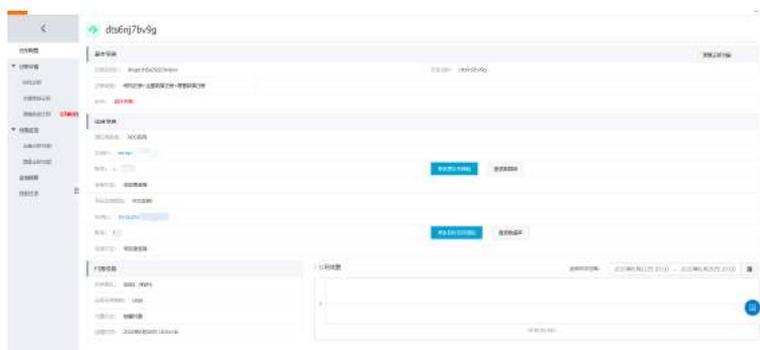


图 3-23

3.4.1.1 任务配置

该页面展示了任务的详细配置信息，包括源端与目标端的配置信息。如果迁移任务使用的源端和目标端的数据库账户的密码更改导致 DTS 任务异常，可以点击“修改原 / 目标实例密码”进行订正。修改密码需要使用旧密码（如果您旧密码忘记无法修改，此时需要重建任务）。如下图 3-24。



图 3-24

3.4.1.2 迁移详情

此功能展示了 DTS 迁移过程中各个环节的迁移状态，详细说明如下。

3.4.1.2.1 结构迁移

如下图 3-25，展示了结构迁移页面迁移的数据库结构的信息（此处只展示了 tables、views、functions、procedures）。如果结构迁移时迁移的数据库对象有异常，会在“状态”列显示。点击“查看创建语法”可以查看 DTS 是以何种 SQL 语句在目标端创建对象结构的。您也可以自行创建目标数据库的表结构信息，不使用 DTS 进行结构迁移（如果您目标数据库已经有对应的表结构了，在“[3.3.10 迁移类型](#)”选择时，无需选择结构迁移）。



图 3-25

结构迁移部分失败分为 2 点：

- DTS 通过 SQL 语句在源端数据库获取数据库对象的相关信息时异常，如下图 3-26 的例子就是 DTS 通过 SELECT 语句在源端执行 SQL 时，出现 Unknown error 的错误。这个图里有 2 个关键的信息，其中“err msg”指的是 SQL 语句执行时的错误信息（此处为 Unknown error）。“sqls”指的是 DTS 在源端执行哪一个 SQL 查询哪一个对象时出现了 Unknow error（此处是 select routine_schema,routine_name... from information_schema.ROUTINES where... 和 select parameter_mode,parameter_name... from information_schema.PARAMETERS WHERE specific_schema=? and ...）。判断这个问题的原因可以拿这两个 SQL 在源端数据库执行复现问题，确认 Unknown error 的原因。

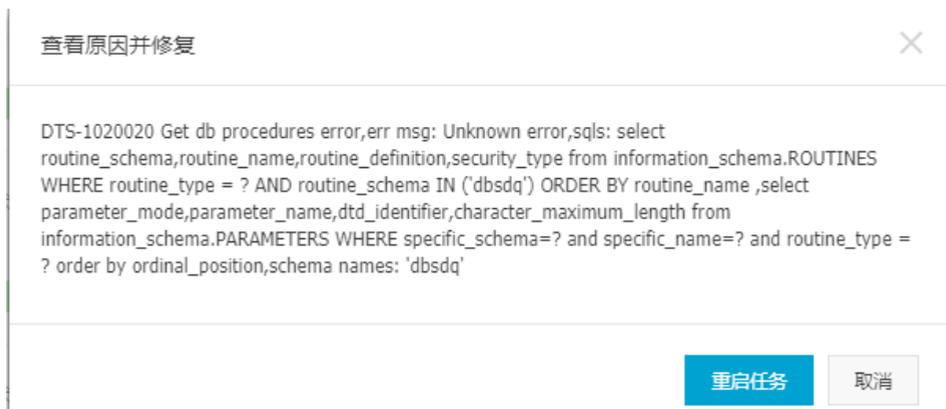


图 3-26

- DTS 通过源端数据库获取到对象的相关信息后，DTS 内部会组合成 DDL 语句在目标端数据库进行执行。当 DTS 把这些 DDL 语句在目标端执行时异常。如下图 3-27 的例子就是 DTS 在目标执行 DDL 时出现了异常。这个错误比较简单，指的是 DTS 在目标端数据库创建表时，提示表已存在。要解决这个问题，需要去目标端查看下，是否有对应的表以及存在。根据错误，找解决方案。

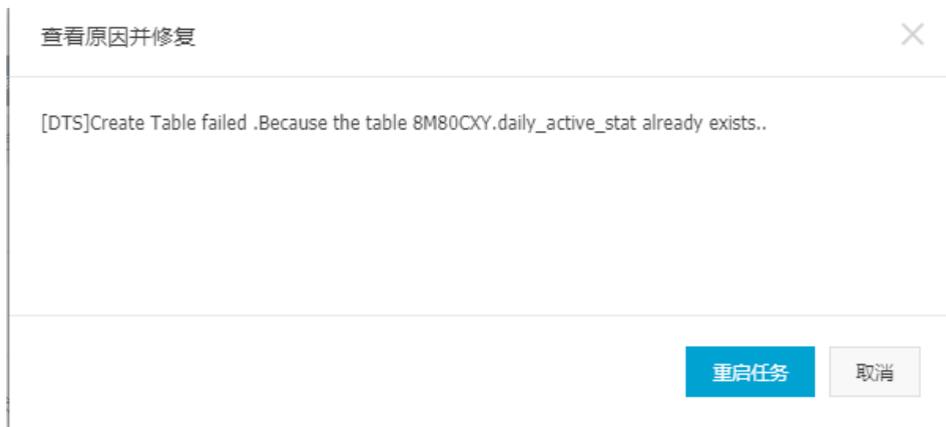


图 3-27

3.4.1.2.2 全量迁移

如下图 3-28，是结构迁移部分顺利运行完成后，进行全量迁移的展示页面。

为什么结构迁移在前？因为只有先创建表，才能往目标实例写数据。全量迁移要从源端抽取数据然后把数据写入目标端，这会导致源端与目标端数据库实例的负载（CPU,IOPS 等）增加，因为全量迁移的是源端数据库表里已经存在的数据（这些数据可能是历史数据，早已经写入，也可能刚刚写入不久的数据，非未来新增的数据）。

任务名称	源实例	目标实例	迁移总量	已迁移	耗时(小时)	状态	操作
testdb2	aliott	dtsott	0	1	--	成功	

图 3-28

与结构迁移相同的是全量迁移出现问题也分为 2 点：

- DTS 通过 SQL 语句（一般是 SELECT）在源端数据库获取数据库表的数据的时候异常。如下图 3-29 就是在源端使用 SELECT 抽取数据时，由于 MAX_ALLOW_PACKET 参数设置过小导致数据查询异常。您可以通过调整参数的大小来解决该问题，如果参数已经调到最大后依然无法避免这种情况，请反馈阿里云售后。

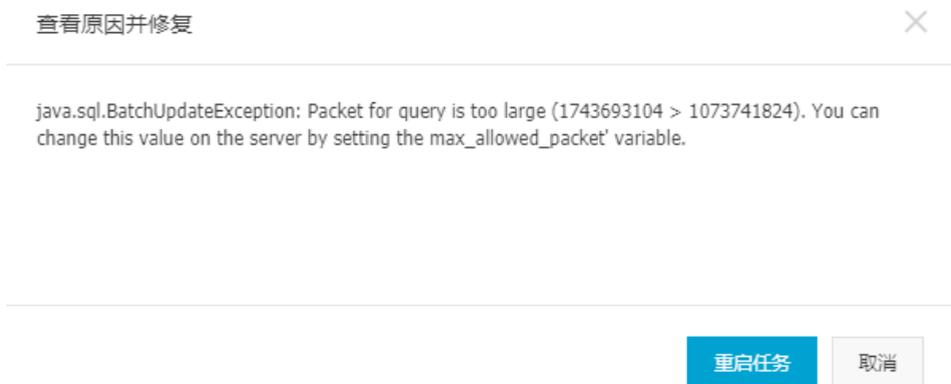


图 3-29

- DTS 通过源端数据库获取到的数据库数据，DTS 内部会组合成 DML 语句（INSERT、UPDATE、DELETE 等）在目标端数据库进行执行。当 DTS 把

这些 DML 语句在目标端执行时异常。如下图 3-30 是指 DTS 在把数据写入目标端数据库时，提示表不存在。遇到该问题，需要在目标端数据库查看数据库的表是否正常可写入以及 DTS 目标端的数据库账户权限是否正常。

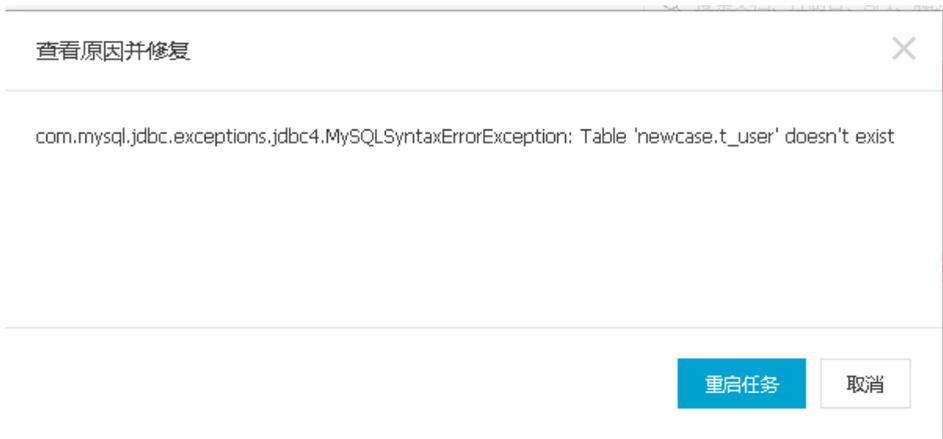


图 3-30

3.4.1.2.3 增量迁移

如下图 3-31，是 DTS 进行增量时的对象展示页面。此处可以看到有一个“待修复”标识，说明在增量环节出现了异常，需要修复。具体的异常和修复方案，我们会在“3.4.2 查看原因并修复”章节详细说明。增量迁移不会导致源端数据库的负载有明细增加。由于要把增量数据写入目标端，目标端数据库实例的负载可能会增加。

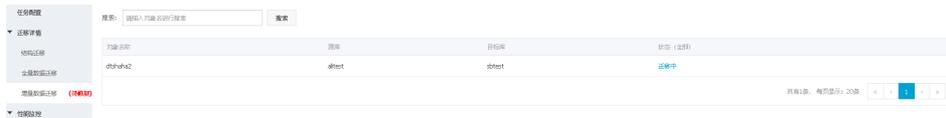


图 3-31

再次强调下，dts 的增量迁移的实现方法，大多和数据库本身的复制功能相关，借助的是源端数据库的一些日志来实现，以 mysql 为例，dts 借助的是 binlog，dts

启动增量的时候，会在源端启动一个 binlog dump 进程，源端实例会推送 binlog 的记录给 dts，dts 解析这些记录，然后在目标实例应用，实现增量同步。

同时，DTS 的增量数据迁移延迟是无法保证的，正常情况下 DTS 的增量迁移是秒级延迟，但是当遇到一些 DDL、大量更新时或者 DTS 规格达到瓶颈等情况时，增量数据迁移延迟会增高。如果您遇到大的延迟（比如超过 1000S），可与阿里云售后反馈。

• 3.4.1.3 性能监控

该功能展示了全量和增量迁移阶段，DTS任务的迁移性能，详细说明如下。

• 3.4.1.3.1 全量迁移性能

前面我们讨论过，DTS 进行全量迁移，实际上就是在源端执行 SELECT 获取数据，然后再把获取的数据写入到目标端的过程。如下图 3-32，“全量迁移链路拓扑”形象的展示了这个过程，并且也展示了全量抽取与全量写入两个环节的 BPS(全量迁移阶段 DTS 全量迁移服务的 IOPS，单位：Bytes/Second) 与 RPS(全量迁移阶段 DTS 全量迁移服务的记录数量，单位：Records/Second) 信息。同时“全量迁移性能”部分也展示了全量迁移的四个关键指标以及“指标含义”：全量迁移流量，原、目标实例 RPS，源库网络延迟、目标库网络延迟，源库 SQL 执行 RT、目标库 SQL 执行 RT。



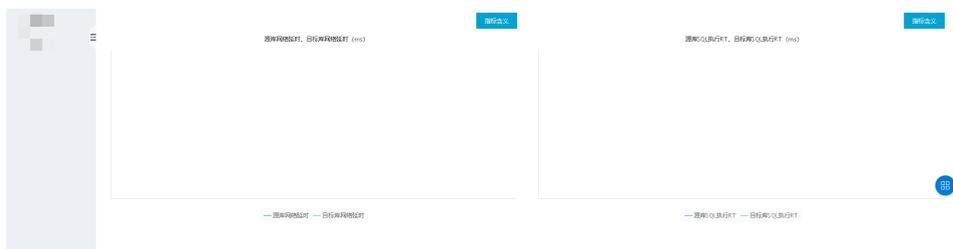


图 3-32

3.4.1.3.2 全量迁移慢的定位与处理

在“3.4.1.3.1 全量迁移性能”章节了解了全量迁移的性能指标，本章节讨论下遇到全量迁移慢如何分析定位与处理。全量迁移慢分为 3 个部分：源端抽取慢、DTS 服务本身对数据处理慢、目标端写入慢。

- 源端抽取慢：要判断这一点，首先要在源端查看 DTS 的会话状态，以 MySQL 为例，可以通过执行 `select * from information_schema.processlist` 与 `show processlist` 获取，查到会话状态后有如下 3 种情况：
 - 没有 DTS 的任何会话，遇到这个情况，说明 DTS 与源端实例的连接可能存在问题。
 - DTS 的会话存在，但是状态全部是 `sleep`，遇到这个情况，说明 DTS 抽取已经完成，或者 DTS 抽取存在异常。
 - DTS 会话存在，部分会话在执行 `query` 获取数据，遇到这个情况，说明 DTS 正在抽取源端数据库的数据，状态正常，您需要关注下源端数据库的负载，避免出现性能瓶颈。
- DTS 服务本身对数据处理慢：因为 DTS 负责把源端数据库数据获取后写入到目标端，DTS 服务本身的链路规格以及内存资源都可能会导致全量迁移慢的情况，比较常见的是 DTS 内存较低导致在迁移大数据对象时频繁 OOM(DTS 的错误表现是 `Java heap space`)，遇到这方面的问题，可以反馈阿里云售后核实。
- 目标写入慢：要判断这一点，同样需要在目标端查看 DTS 的会话状态，以 MySQL 为例，可以通过执行 `select * from information_schema.processlist` 与 `show processlist` 获取，查到会话状态后有如下 3 种情况：

- 没有任何的 DTS 会话在执行 DML 操作，说明 DTS 与目标实例的链接可能存在问题。
- DTS 的会话存在，但是状态全部是 sleep，遇到这个情况，说明 DTS 与目标端数据库连通性正常，但是 DTS 无法在目标端执行 SQL。
- DTS 会话存在，部分会话在执行 DML 写入数据，遇到这个情况，说明 DTS 状态正常。如果 DML 耗时太长，需要关注下阻塞，数据库负载等方面。

3.4.1.3.3 增量迁移性能

DTS 增量迁移分为 3 个阶段，如下图 3-33：

1. DTS 通过数据库的日志，获取源端的新增数据（全量迁移之后的新增数据），这部分工作由 DTS 采集模块实现。
2. DTS 拿到这些数据后，在 DTS 侧缓存起来，以便后面写入目标端数据库。这部分工作由 DTS 缓存模块实现。
3. DTS 获取缓存起来的增量数据，进行处理后把这些数据应用到目标端数据库，这部分工作由 DTS 写入模块实现。

同样，增量迁移时，每个环节也都有性能指标。重点关注的指标是增量迁移的 RPS 和迁移延迟。增量抽取不会对源端的数据库性能产生大的影响。



图 3-33

3.4.1.3.4 增量迁移延迟的定位与处理

通过“[3.4.1.3.3 增量迁移性能](#)”章节说明了解了 DTS 增量迁移阶段的 3 个步骤，当出现增量延迟的时候，也会出在这三个部分：DTS 采集模块获取源端增量日志延迟、DTS 本身服务异常导致数据处理延迟、DTS 写入模块写入目标数据库延迟。最长遇到问题的时采集延迟和写入延迟：

- DTS 采集模块获取源端增量日志延迟：这个主要出现在源端的日志缺失、日志过大或者源端数据库连接异常时出现，以 MySQL 为例，如果手动清理掉了 DTS 还未获取的源端 binlog，此时 DTS 采集模块就会异常。还有如果源端的 binlog 单个文件特别大（比如大事务产生了 100G 的单个 binlog 文件），DTS 也可能存在拉取文件超时的情况，导致延迟。如果要判断增量采集模块是否有异常或者延迟，以 MySQL 为例，DTS 是通过在源端建立一个 binlog dump 的会话实现的 binlog 的抽取，故可以通过执行 `select * from information_schema.processlist` 与 `show processlist` 获取，查到会话状态后有如下 3 中情况：
 - binlog dump 进程不存在，说明 DTS 与源端数据库的连通性可能存在问题。
 - binlog dump 进程存在，且状态为 `Master has sent all binlog to slave; waiting for more updates`，如下图 3-34。如果是这种状态，可以基本说明源端数据库的 binlog 已经全部发到了 DTS 侧。

Sleep		
Daemon	1755752	Waiting on empty queue
Binlog Dump GTID	1755619	Master has sent all binlog to slave; waiting for more updates
Connect	1755619	Waiting for master to send event
Connect	15	Slave has read all relay log; waiting for more updates
Connect	1755619	Waiting for master to send event

图 3-34

- binlog dump 进程存在，但是状态非 `Master has sent all binlog to slave; waiting for more updates`，可以说明源端数据库正在发送 binlog 文件给 DTS，中间可能存在一些问题或者耗时较长。

- DTS 写入模块写入目标数据库延迟：如果 DTS 源端数据库抽取数据非常的及时，那出现延迟很可能是在写入目标数据库的时候写入较慢或者并发不高导致。遇到这类问题，主要排查如下 3 个方面。
 - 目标数据库的 IO、CPU 负载等资源是否正常，如果这些正常，可以排除性能瓶颈。
 - 目标数据库的 SQL 执行状态是否正常，比如是否耗时较长或者遇到了阻塞，以 MySQL 为例，可以通过执行 `select * from information_schema.processlist` 与 `show processlist` 获取会话的执行状态，查看 DML 语句的执行情况与耗时，问题判断和分析与“[3.4.1.3.2 全量迁移慢的定位与处理](#)”章节的“目标写入慢”部分相同。

3.4.1.4 监控报警

此功能用来监控 DTS 迁移任务的延迟与状态情况，如下图 3-35。DTS 迁移只支持监控这两项，您可以设置这两项的报警接收的手机号。

DTS 的增量数据迁移延迟是无法保证的，正常情况下 DTS 的增量迁移是秒级延迟，但是当遇到一些 DDL、大量更新时或者 DTS 规格达到瓶颈等情况时，增量数据迁移延迟会增高。如果您遇到大的延迟（比如超过 1000S），可与阿里云售后反馈。



图 3-35

3.4.2 查看原因并修复

当 DTS 任务出现异常的时候，我们需要查看异常原因以及尝试修复它，从“[3.4.1.2.3 增量迁移](#)”章节可以看到，这个任务增量异常，我们要查看异常的原因，

需要查看任务列表页面，如下图 3-35 和图 3-36。这里的迁移错误是因为 DTS 使用的源端账户缺少 SUPER 或者 REPLICATION CLIENT 权限导致增量异常。处理方法则是把账户的 REPLICATION CLIENT 权限加上然后重启任务即可。



图 3-36

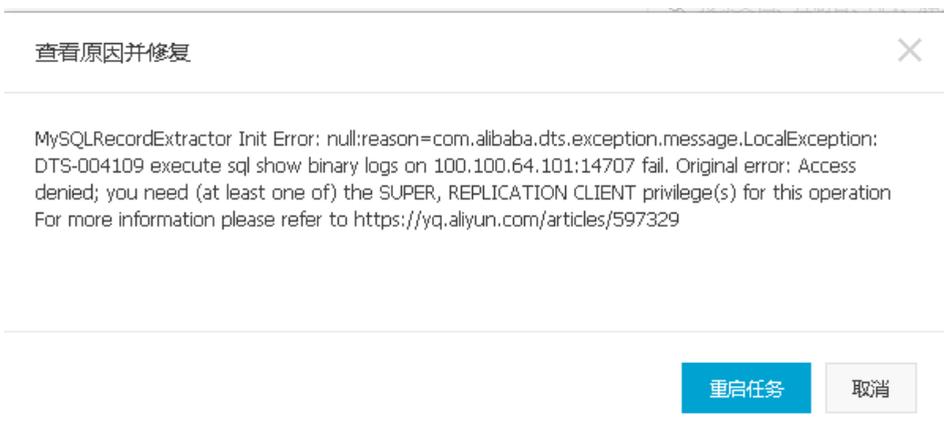


图 3-37

3.4.3 启动任务

当任务暂停，失败时，可以点击启动任务来启动。

3.4.4 查看详情

查看详情功能与“[3.4.1 ID/名称](#)”功能相同。

3.4.5 创建类似任务

该功能用来创建与当前 DTS 任务近似的新任务，点击后，会跳转新的创建页面，如下图 3-37。该功能不推荐使用，如果您需要创建新的任务，请点击“[3.3 创建迁移任务页面](#)”。

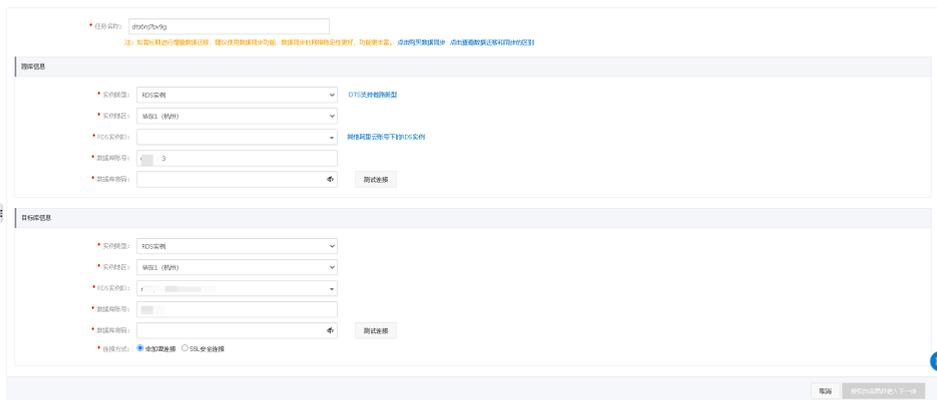


图 3-38

3.4.6 升级

当 DTS 任务规格不足时，比如我们通过“[3.4.1.3 性能监控](#)”章节讨论的性能指标判断出迁移性能的瓶颈在 DTS，此时就需要升级 DTS 的任务规格。需要注意如下 3 点：

- 如果您配置 DTS 任务时，没有选择增量，在任务配置时是不收费的，但是如果点击升级 DTS 任务（无增量的 DTS 任务），则会产生费用，如下图 3-38。
- DTS 的规格现在还只能升级，不能降级。
- 不同的 DTS 数据迁移规格，性能不同，具体可以参考这里的测试结果：https://help.aliyun.com/document_detail/26606.html?spm=a2c4g.11186623.6.559.2ab-2179776BAy8。

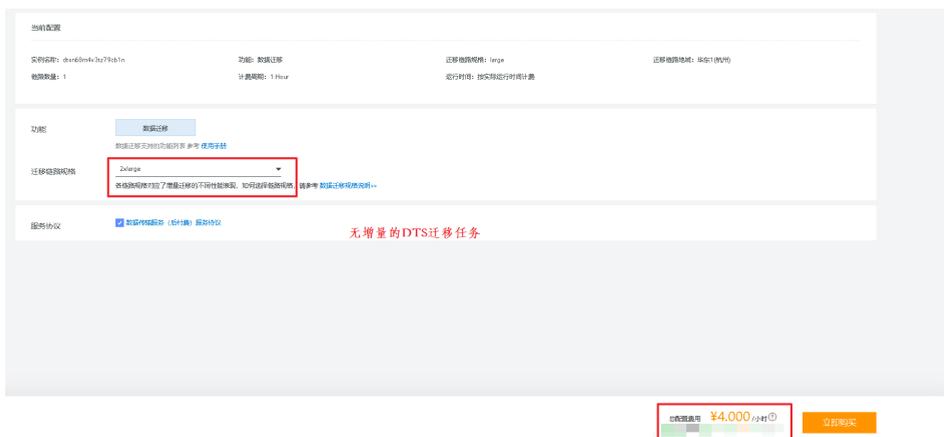


图 3-39

3.4.7 监控报警

该功能与“[3.4.1.4 监控报警](#)”功能相同。

3.4.8 修改密码

该功能与“[3.4.1.1 任务配置](#)”的修改密码功能相同，修改密码只支持密码修改，账户不可修改。并且修改密码需要使用旧密码（如果您旧密码忘记无法修改，此时需要重建任务）。

3.4.9 暂停任务

暂停任务一般可以用于当因为 DTS 的迁移操作导致源端数据库或者目标端数据库负载消耗较大时，此时暂停任务来缓解数据库的压力（点击暂停后，请查看数据库种 DTS 会话的状态，如果 DTS 会话长时间未结束可以考虑手动 KILL 掉未结束的会话）。这是比较特殊的情况下的使用。

大多数情况下不推荐暂停任务（全量任务与增量任务都不推荐暂停），主要原因为如下 2 点：

- **全量任务**：因为全量迁移阶段，DTS 会对源端已存在的数据进行抽取，抽取这些数据无法短时间内完成（数据量大时可能要经历好几个小时），如果我们在全量任务阶段暂停了全量任务，然后又重启，全量数据的迁移会重新开始，比如有 A、B、C、D 四个表要进行迁移，此时 A 表迁移完成，D 表还未开始迁移，正在迁移 B 表（假设已经完成 70%）和 C 表（假设已经完成 30%），如果此时暂停全量迁移，B 表和 C 表的源端数据库 SELECT 请求将会停止，除非您已经放弃进行全量迁移，否则最终还是要重新启动全量任务的，如果再次启动全量任务，DTS 会重新对 B 表和 C 表进行数据的迁移，也就是意味着前面的 B 表（假设已经完成 70%）和 C 表（假设已经完成 30%）已经迁移完成的这些数据并不会续传（DTS 不支持续传）。这也就导致如下 2 个问题：
 - 如果 B 表和 C 表没有主键或者唯一键，数据可能会出现重复甚至是丢失等不一致的情况。
 - 如果 B 表和 C 表有主键或者唯一键，DTS 在重新抽取源库的 B 表和 C 表的数据写入目标端时，可能会出现主键冲突的情况，影响迁移效率。
- **增量迁移**：对增量迁移来说，暂停任务比全量任务安全很多，但是也存在如下 3 点注意事项，**这是重点**：
 - 增量迁移任务暂停时间过长，会导致增量失败。原因在于“3.4.1.3.3 增量迁移性能”章节中我们讨论过，DTS 的增量迁移会有一个 DTS 缓存模块对源端数据库的增量数据进行缓存。这个缓存模块的数据有效期为 7 天（大多数 DTS 任务都是 7 天，少量的 DTS 任务非 7 天）。当我们暂停 DTS 任务超过 7 天时，DTS 增量任务会出现异常。比如我们在 2020 年 6 月 12 日 15:00:00 进行了任务暂停，此时 DTS 采集模块不停止，一直实时的获取源端数据库的增量数据（不会停止），然后这些增量数据由 DTS 缓存模块缓存，缓存的增量 DTS 写入模块暂停，停止写入，写入的数据最后一次时间点为 2020 年 6 月 12 日 15:00:00。随着时间的推移，我们在 2020 年 6 月 20 日 15:00:00 启动任务，由于 2020 年 6 月 12 日 15:00:00~2020 年 6 月 20 日 15:00:00 期间，DTS 增量采集模块和缓存模块一直在工作，此时 DTS 增量采集模块

采集的增量数据是最新的(2020年6月20日15:00:00时间点的),而由于DTS缓存模块的增量数据只保存7天(不是所有DTS任务都是7天,请以具体情况为准),此时DTS缓存模块里的增量的数据范围是2020年6月13日15:00:00~2020年6月20日15:00:00。但是由于2020年6月12日15:00:00停止时,DTS写入模块停止了写入,DTS写入模块的写入时间点依然在2020年6月12日15:00:00,2020年6月20日15:00:00启动后,DTS写入模块继续写入去DTS缓存模块里寻找2020年6月20日15:00:00的数据,开始写入,由于DTS缓存模块里的增量数据已经变成了2020年6月13日15:00:00~2020年6月20日15:00:00,此时DTS写入模块会失败。DTS整个任务也就会失败。遇到这种情况,解决办法有3个:

- 如果源端数据库的相关增量数据的日志还存在(以MySQL为例,源端数据库的binlog还一直存在并未删除),该问题的修复方案可以使DTS的采集模块重新从2020年6月12日15:00:00采集增量数据,DTS缓存模块重新缓存2020年6月12日15:00:00的增量数据,这样DTS写入模块也就可以正常从2020年6月12日15:00:00写入增量数据
- 如果源端数据库的相关增量数据的日志已经不存在被删除,也可以选择放弃2020年6月12日15:00:00~2020年6月13日15:00:00的增量数据,直接让DTS写入模块从2020年6月13日15:00:00开始写入,这会导致数据不一致(2020年6月12日15:00:00~2020年6月13日15:00:00的增量数据未迁移,产生了丢失),该操作非特殊情况不建议。
- 如果源端数据库的相关增量数据的日志已经不存在被删除,且要保证数据一致性。您可以重新进行一次DTS任务(重新进行结构+全量+增量的迁移),保证数据一致性。
- DTS迁移任务暂停后,由于DTS写入模块暂停,再次重新(7天内)启动后会产生延迟(因为要重新写入未写入的增量数据,追赶最新的增量数据),这个延迟理论上是逐渐降低的。

3.4.10 结束任务

使用该功能，有 4 种情况，分别是：

- DTS 未选择增量迁移，全量迁移或者结构迁移完成后，任务自动结束（此时不需手动结束）
- DTS 未选择增量迁移，但是全量迁移或者结构迁移出现异常，该异常无法修复，放弃该 DTS 任务时，可以手动结束该 DTS。
- DTS 选择了增量迁移，增量迁移一直在正常的同步增量数据，您选择在合适的时间切换业务，停止增量迁移时，可以使用结束任务。
- DTS 选择了增量迁移，但是增量任务异常或者无法修复，您放弃该 DTS 任务时，手动结束该 DTS。

无论哪一种情况，结束任务意味着完全放弃这个任务的所有已完成或者未完成的任务环节，结束后无法恢复。

3.4.11 释放任务

“[3.4.10 结束任务](#)”只是把任务状态由“运行中”、“暂停中”、“迁移失败”等改成“已完成”。任务信息依然在任务列表展示，如下图 3-40。要把该任务从任务列表删除，则需要释放它。



图 3-40

第四章 DTS 数据订阅

数据订阅其实就是通过 API 获取 [DTS 缓存模块](#) 的数据，在 DTS 数据迁移章节，DTS 缓存模块的数据只能由 DTS 写入模块获取然后写入目标端数据库。数据订阅把这个 DTS 写入模块进行了开放，我们可以借助 DTS 提供的 DTS 缓存模块的 API 进行增量数据的获取。要使用 DTS 数据订阅来获取 DTS 的缓存模块的增量数据，需要新购 DTS 数据订阅实例，这也就意味着，您拥有 DTS 数据迁移并不能对 DTS 缓存模块进行订阅。购买后，会显示到 DTS 数据订阅的任务列表，如下图 4-1。DTS 数据订阅之前分为新版和旧版，旧版已经在 2020 年 3 月 31 日下线，现在只有新版，本章节将对新版 DTS 数据订阅的功能进行详细的说明。注意：

- DTS 数据订阅只能订阅增量数据，无法订阅全量数据。

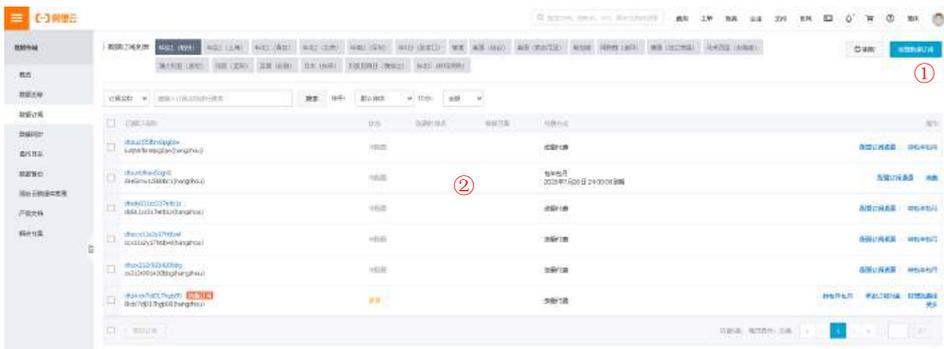


图 4-1

4.1 创建数据订阅页面

要创建数据订阅 DTS 实例，需要点击图 4-1 标记①处的“创建数据订阅”功能。点击后，跳转到创建页面，如下图 4-2。下面对这个页面的选择项做详细的说明。

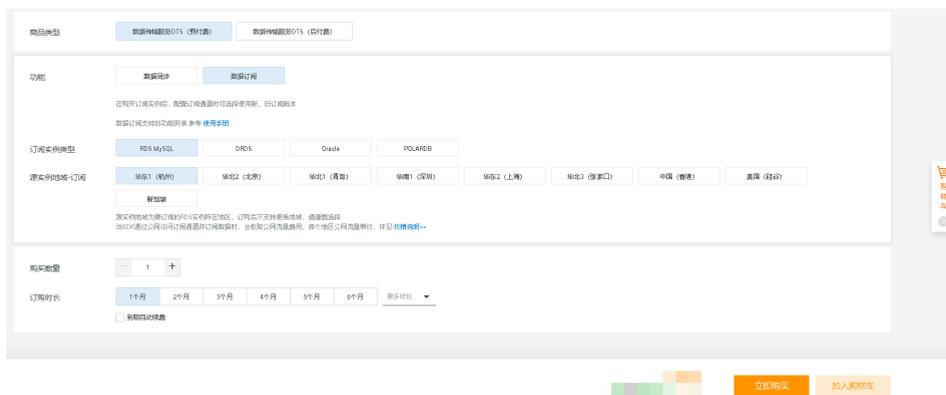


图 4-2

4.1.1 商品类型

商品类型是指购买的数据库 DTS 产品的计费类型，预付费是指包年包月，按年或者月进行计费，费用需要提前支付。后付费是按量付费，即按小时进行计费（按小时计费的意思是只有当这个 DTS 被创建并且被您配置运行后，就会按小时进行计费。只创建出来不配置不会进行计费）

4.1.2 功能

当选择预付费时，这里只能选择数据同步和数据订阅，数据同步我们会在“第五章 DTS 数据同步”章节说明。当选择后付费时，这里还会多一个数据迁移的选项，在这里创建数据迁移时，提示是收费的，**请注意**。这个收费只有在您配置 DTS 迁移选择增量时才会进行收费，不进行增量迁移不会收费。

4.1.3 订阅实例类型

当功能选择数据订阅后，我们需要选择订阅的实例类型，即您需要 DTS 获取哪一种数据库类型的增量数据，目前 DTS 支持的订阅数据库类型为 MySQL、DRDS、Oracle、POLARDB。

4.1.4 源实例地域 – 订阅

要订阅的数据库实例所在的地域，这里一定要选择正确，因为这决定了DTS使用哪里的订阅服务器去连接您的源端数据库。

配置选择完成后，点击支付，等待创建，创建完成后，如下图4-3。新创建的实例，默认并没有新版订阅的标记。



图 4-3

4.2 任务列表

图4-1 标记②的处为所选择地域的任务列表。展示了当前地域下运行中，未配置以及异常等状态的订阅任务，具体功能如下。

4.2.1 配置订阅通道

创建出来未配置的实例，默认是没有“新版订阅”的标识的，只有当配置完后，才会显示“新版订阅的标识”，点击配置订阅通道配置订阅，跳转到新的页面，如下图4-4。各配置项的详细说明如下。

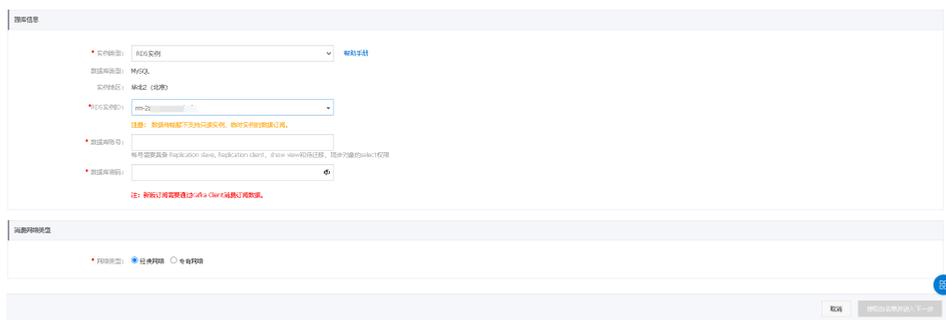


图 4-4

4.2.1.1 实例类型

该处实例类型与“[3.3.1 实例类型](#)”介绍的含义相同。

4.2.1.2 RDS 实例 ID

由于我们在“[4.1.3 订阅实例类型](#)”创建时，选择的是 MySQL，我们在“[4.2.1.1 实例类型](#)”选择的是 RDS 实例，则需要我们填写 RDS 实例 ID，如果我们选择的 RDS 是 5.7 及以上的 RDS 实例，则还需要填写 RDS 的数据库账号和数据库密码。当选择的是 RDS 5.6 实例时则不需要，如下图 4-5。原因在于针对 RDS5.6 实例，DTS 会在 RDS 内部创建专属于 DTS 的迁移账户，而对于 RDS5.7 及以上的，DTS 并不会这样做，非常推荐您使用 RDS5.7 及以上的高权限账户进行配置。



图 4-5 (经典网络)

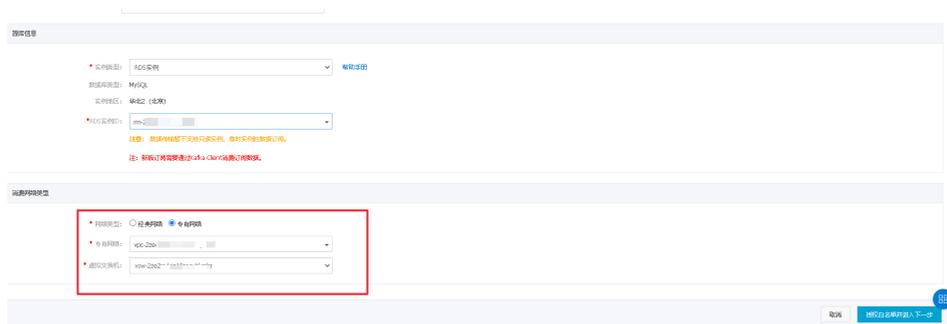


图 4-5 (专有网络)

4.2.1.3 数据库账号 / 数据库密码

由于 DTS 要获取数据库的增量数据，数据库的账号权限一定要具备增量日志的获取权限，具体参考“[3.3.5 数据库账号](#)”和“[3.3.11.2 源库权限检查](#)”。

4.2.1.4 消费网络类型

分为经典网络和专有网络，这里的选择代表的是您后面使用 Client 订阅数据的时候，如何配置。当 DTS 订阅创建完成后，会生成 2 个地址：公网和私网，如“[4.2.2 订阅 ID/名称](#)”章节的图 4-9。如果选择经典网络，生成的私网地址就是一个经典网络的地址（要访问该私网地址需要同地域经典网络的 ECS），如果选择的是专有网络，生成的私网地址就是一个专有网络的地址（要访问该私网地址需要同低于同 VPC 的 ECS）。

4.2.1.5 授权白名单并进入下一步

当配置完后，点击“授权白名单并进入下一步”进入下一页面（功能说明参考“[3.3.9 授权白名单并进入下一步](#)”），如下图 4-6。该页面的订阅数据类型包括正常的数据库更新以及对象结构更新，当把对象从左侧选择到右侧后，不支持数据迁移的库表列映射。需要注意的是如下 3 点：

- 数据更新，只订阅 insert, update, delete 这类操作
- 结构更新订阅的是**整个实例**所有对象的结构创建、删除及修改，您需要使用数据订阅客户端过滤所需的数据。

- 如果选择整个库作为订阅对象，那么该库中新增对象的增量数据也可以被订阅到。如果选择某个表作为订阅对象，那么只有这个表的增量数据可以被订阅到。此时如果需要加入一个新的表作为订阅对象，您需要将其加入至订阅对象中，具体参考“[4.2.5 修改订阅对象](#)”章节。



图 4-6

4.2.1.6 保存并预检查

该功能与“[3.3.11 预检查并启动](#)”含义相同。如下图 4-7。



图 4-7

正常配置后，点击关闭按钮，任务列表如下图 4-8。



图 4-8

4.2.2 订阅 ID/名称

显示 DTS 任务的 ID 以及名称，点击这个超链接后，进入任务详情页面，如下图 4-9。

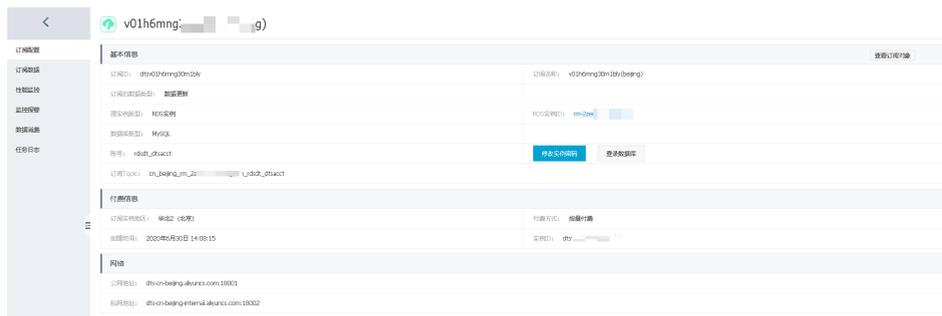


图 4-9

4.2.2.1 订阅配置

该页面是 DTS 数据订阅的配置信息展示页面。其中基本信息页面的账户为 rdsdt_dtsacct，这是由于我选择的 RDS 是 5.6 实例，DTS 使用的内部账号进行的订阅。“订阅 Topic”、“网络”这两个地方非常重要，后面进行“Client 下游消费”时会用到。重点说下“网络”：

- 公网是任何地方都可达的，只要消费 Client 程序部署的这台主机可以访问公网即可。

- 私网的使用有特殊的限制，私网分为经典网络私网和 VPC 网络私网，取决于“[4.2.1.4 消费网络类型](#)”章节如何配置的消费类型。经典网络私网可以让部署在经典网络 ECS 上的消费 Client 进行链接与消费。VPC 网络私网可以让部署在专有网络 ECS (这个 ECS 所在的 VPC 需要与部署 DTS 时选择的 VPC 相同，如图 4-5) 上的消费 Client 进行链接与消费。

4.2.2.2 订阅数据

该页面可以查看已经订阅的数据，这些数据保存在 DTS 缓存模块中，如下图 4-10。这个页面不支持 Oracle 订阅的数据查看。

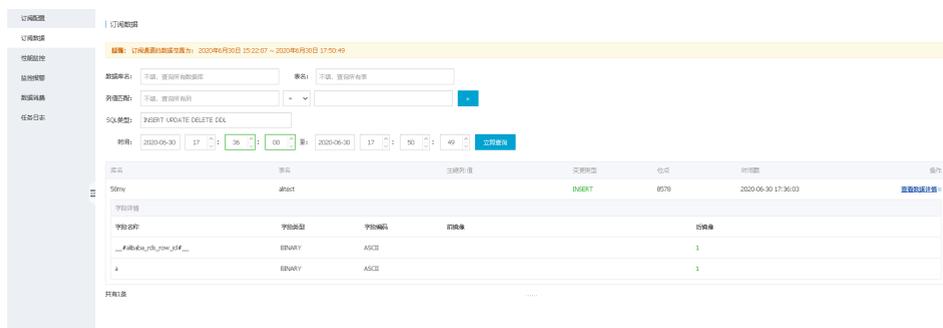


图 4-10

4.2.2.3 性能监控

该页面展示了 DTS 订阅的拓扑图以及监控指标，如下图 4-11。从拓扑图里可以看到，这个拓扑图其实就是我们在“[第四章 DTS 数据订阅](#)”讨论的内容，只是把 DTS 数据迁移的采集模块和缓存模块单独拿出来，DTS 数据迁移的 DTS 增量写入模块变成了“下游消费 Client”。

在“[3.4.1.3.4 增量迁移延迟的定位与处理](#)”章节的定位和排查方法，也可以用到这里，DTS 数据订阅客户遇到问题最多的是订阅消费延迟，也就是下游消费 Client 消费 DTS 缓存的增量数据有延迟，比如 DTS 缓存的增量数据是 2020-06-13 15:00:00，而下游消费 Client 只消费到 2020-06-13 12:00:00 的增量数据，这

就差了3个小时。出现这种情况的原因一般是很多客户的下游消费 Client 里面并不只是单纯的进行数据消费，还会对这些消费的数据进行业务上的逻辑处理，从而导致这样的延迟出现。如遇这种情况，强烈建议将业务逻辑处理与消费逻辑分离，保证下游消费 Client 可以实时的消费到数据，这样也可以避免因消费延迟过大导致增量缓存模块数据过期的问题。



图 4-11

4.2.2.4 监控报警

该功能与“[3.4.1.4 监控报警](#)”章节功能相同。如下图 4-12 和图 4-13。

报警ID	报警名称	报警状态	报警人
	订阅任务	已启用	
	订阅任务 - 报警	已启用	

图 4-12



图 4-13

4.2.2.5 数据消费

要使用 Client 进行订阅的消费 (获取缓存模块的增量数据), 需要先在数据消费页面 (如下图 4-14) 创建消费组。



图 4-14

4.2.2.5.1 新增消费组

如下图 4-15, 新增消费组时需要填写 3 个信息: 消费组名称、账号、密码。这三个信息可以随意填写, 点击确定后, 该页面会显示“消费组 ID/ 名称”、“堆积记录数”、“消息延迟”等信息, 如下图 4-16。请注意如下 2 点:

- 您可以创建多个消费组, 给不同的下游消费 Client 使用, 您也可以把一个消费组给不同的下游消费 Client 使用 ([注意 kafka 中 SUBSCRIBE_MODE_NAME 参数设置](#))。
- 只有当下游 Client 客户端开始进行消费时, 这些堆积记录数, 消息延迟才会显示。



图 4-15

订阅组名称	消费位点	堆积记录数	消息延迟(s)	消息	操作
dtsync7621180qr	atext	-	-	atext	修改策略 删除

图 4-16

4.2.2.5.2 消费位点

如下图 4-17，下游开始消费后，消费位点会进行展示。消费时间点是指“下游消费 Client”对这个 DTS 订阅的缓存模块的增量数据消费的位点，如“[4.2.2.3 性能监控](#)”所述。由于我暂停了下游消费 client，所以此时的消费位点一直在 7 月 1 日 15:21:56。堆积记录数和消息延迟都很大。当重新启动下游消费 Client 时，消费位点会随着 Client 的消费而推进，如下图 4-18。

订阅组名称	消费位点	堆积记录数	消息延迟(s)	消息	操作
dtsync7621180qr	20200701日 15:21:56	161953	149483	atext	修改策略 删除

图 4-17

订阅组名称	消费位点	堆积记录数	消息延迟(s)	消息	操作
dtsync7621180qr	20200701日 08:40:52	0	7	atext	修改策略 删除

图 4-18

4.2.2.5.3 堆积记录数

该消费组，有多少记录数没有被下游消费 Client 消费。

4.2.2.5.4 消息延迟

该消费组的订阅数据延迟情况。

4.2.3 消费时间点

此处，新版订阅的消费时间点一直为空，因为新版的消费时间点可能有多个（多个消费组），具体的消费时间点需要查看“[4.2.2.5.1 新增消费组](#)”页面。

4.2.4 数据范围

数据范围是指 DTS 的缓存模块里的增量数据的范围（默认 7 天），超过这个范围的增量数据已经过期，无法订阅。故当订阅不在数据范围内的增量数据时，下游消费 Client 会出现异常。具体异常我们会在“[4.3 使用 Client 订阅数据](#)”讨论。

4.2.5 修改订阅对象

当我们要新增 DTS 的订阅对象时，可以点击该功能进行添加。

4.2.6 新增消费组

该功能与“[4.2.2.5.1 新增消费组](#)”章节功能相同。

4.2.7 更多

如下图 4-19，更多功能也包含了多个功能项。重点说明如下节点。



图 4-19

4.2.7.1 查看示例代码

要进行数据订阅的数据消费，需要使用下游消费 Client，这个下游消费 Client 可以使用 0.11 版本至 1.1 版本的 Kafka 客户端消费订阅数据，DTS 也提供了 MySQL 数据订阅的消费示例代码，点击可以跳转到 github，获取 MySQL 的消费示例代码。示例代码的配置，我们会在“[4.3 使用 Client 订阅数据](#)”讨论，请注意如下 1 点：

- 示例代码，只是针对 MySQL 数据订阅进行消费的，Oracle 数据订阅的不适用，您还需要编写 Oracle 的示例代码。

4.2.7.2 释放订阅

后付费（按量付费）的 DTS 订阅实例，支持手动释放。预付费（包年包月）的 DTS 订阅实例不支持手动释放。

4.2.7.3 重新配置

如果需要重新配置当前的 DTS 订阅实例，可以使用该功能。

4.3 使用 Client 订阅数据

本章节，我们以 DTS 提供的示例代码为例，讨论如何配置下游消费 Client，进行数据的消费。

4.3.1 下载示例代码

点击“[4.2.7.1 查看示例代码](#)”，示例代码一般在这个地址：https://github.com/LioRoger/subscribe_example?spm=5176.2020520151.6108.d2.3e0a75b02MxgZf。

4.3.2 导入示例代码

解压示例代码后，使用 Eclipse 导入示例代码（也可使用其他工具），该示例代码是基于 Maven 的，所以要提前配置好 Maven。导入步骤如下图 4-20、图 4-21、图 4-22。

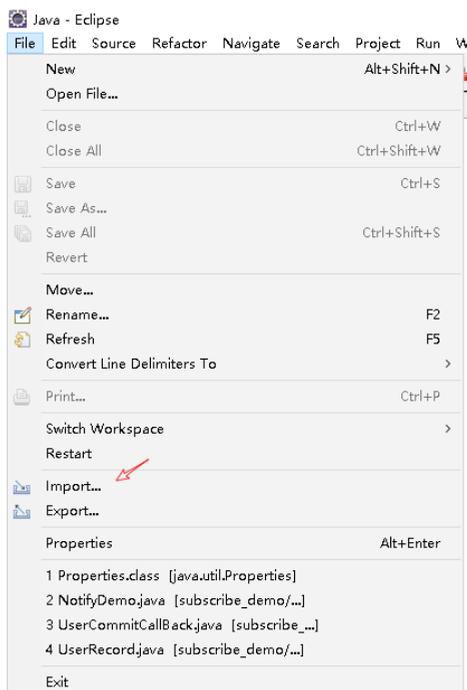


图 4-20

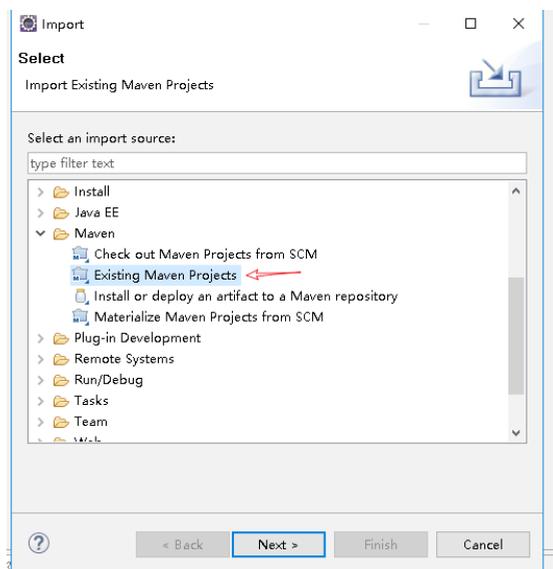


图 4-21

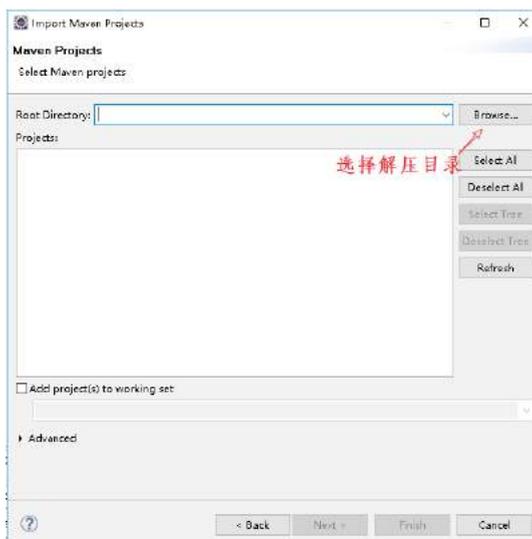


图 4-22

4.3.3 配置 Client

导入完成后，会看到如下图 4-23 的结构图。找到。

/subscribe_demo/src/main/java/NotifyDemo.java 文件。

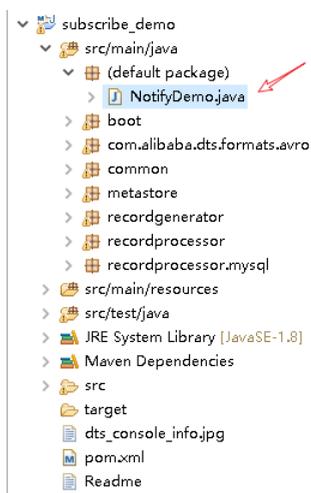


图 4-23

打开该文件，打开后，配置如下图 4-24 部分设置。其中有如下 7 个重点配置项：

USER_NAME：是指在“[4.2.2.5.1 新增消费组](#)”章节创建消费组时填写的账号。

PASSWORD_NAME：是指在“[4.2.2.5.1 新增消费组](#)”章节创建消费组时填写的密码。

SID_NAME：是指在“[4.2.2.5.1 新增消费组](#)”章节创建消费组后，页面显示的“消费组 ID/ 名称”中的 ID。

GROUP_NAME：此处，填写消费组 ID。

KAFKA_TOPIC：是指在“[4.2.2 订阅 ID/ 名称](#)”章节图 4-9 中的“订阅 Topic”。

KAFKA_BROKER_URL_NAME：是指在“[4.2.2 订阅 ID/ 名称](#)”章节图 4-9 中的“网络”。此处使用的是公网，您也可以使用内网，使用内网有条件限制，请参考章节“[4.2.1.4 消费网络类型](#)”与“[4.2.2.1 订阅配置](#)”。

INITIAL_CHECKPOINT_NAME：用来设置消费 Client 启动时的消费的时间点。这里如果设置的时间点不在数据 DTS 订阅的数据范围，会出现异常。

USE_CONFIG_CHECKPOINT_NAME：用于设置当这个消费 Client 启动时，是否用 INITIAL_CHECKPOINT_NAME 的值来重置消费时间点，当第一次 Client 启动时，消费 Client 的消费时间点也会随着消费数据进行推进（推进的时间点在 INITIAL_CHECKPOINT_NAME 之后），此时如何停止这个消费 Client 再次启动，就会有 2 个选择：继续按照之前消费的时间点继续消费或者重新按照 INITIAL_CHECKPOINT_NAME 设置的时间点重新消费。这个参数就是用来控制这两个选项的，true 的话表示每次启动都重置这个订阅 Client 的消费时间点（重置成 INITIAL_CHECKPOINT_NAME），false 的话则继续消费（从上一次消费 Client 停止时的消费时间点继续）。

SUBSCRIBE_MODE_NAME：这个配置项用来指定消费的模式，分为 assign 和 subscribe，这两个设置是 kafka 的设置，不再赘述。

```

public static Properties getConfigs() {
    Properties properties = new Properties();
    // user password and gid for auth
    properties.setProperty(USER_NAME, "alitest");
    properties.setProperty(PASSWORD_NAME, "66666666");
    properties.setProperty(SID_NAME, "dtsyqk78121jDbgz");
    // kafka consumer group general same with sid
    properties.setProperty(GROUP_NAME, "dtsyqk78121jDbgz");
    // topic to consume, partition is 0
    properties.setProperty(KAFKA_TOPIC, "cn_beijing_rm_2zed39edfc72gyibn_rdsdt_dtaacct");
    // kafka broker id1
    properties.setProperty(KAFKA_BROKER_URL_NAME, "dte-cn-beijing.aliyuncs.com:18001");
    // initial checkpoint for first seek (a timestamp to set, eg 1566180200 if you want (Mon Aug 19 10:03:21 CST 2019))
    properties.setProperty(INITIAL_CHECKPOINT_NAME, "1593567000");
    // if force use config checkpoint when start, for checkpoint reset
    properties.setProperty(USE_CONFIG_CHECKPOINT_NAME, "true");
    // use consumer assign or subscribe interface
    // when use subscribe mode, group config is required, kafka consumer group is enabled
    properties.setProperty(SUBSCRIBE_MODE_NAME, "assign");
    return properties;
}

```

图 4-24

4.3.4 启动 Client

配置完成后, 启动 Client, 如下图 4-25。启动后, Console 窗口会打印 log 信息。

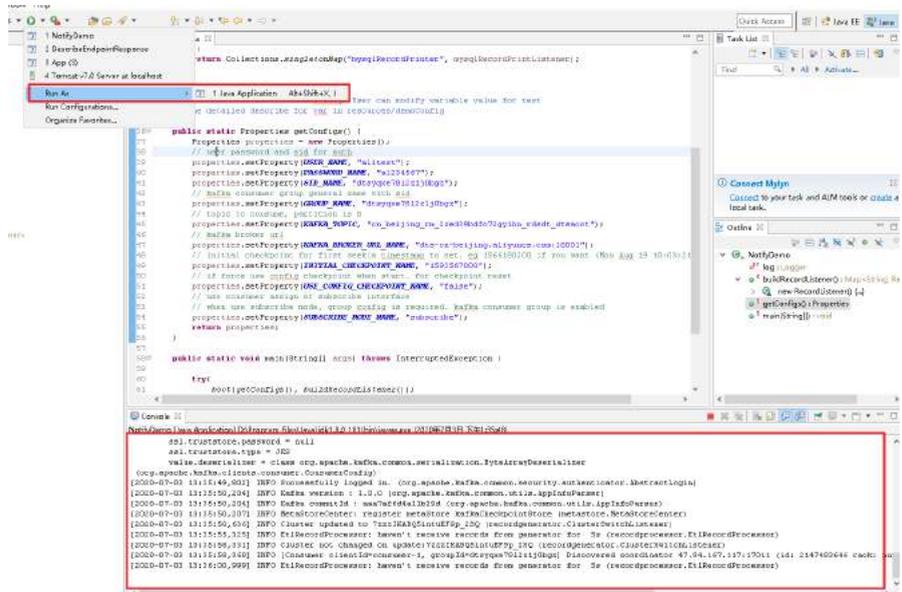


图 4-25

4.3.5 打印消费信息

为了确定订阅数据正常, 可以在 Console 窗口打印出 Client 订阅到的数据信息, 此时需要 Demo 代码中的一个注释, 如下图 4-26, 把 log.info(ret) 的注释去

掉。重新启动。



图 4-26

重新启动后，当源端数据库出现增量数据更新时，Client 会消费这个 SQL 然后在 Console 窗口打印出来，如下图 4-27。可以看到插入记录，指的是在 a 列（表的列名）写入了一个值为 2。

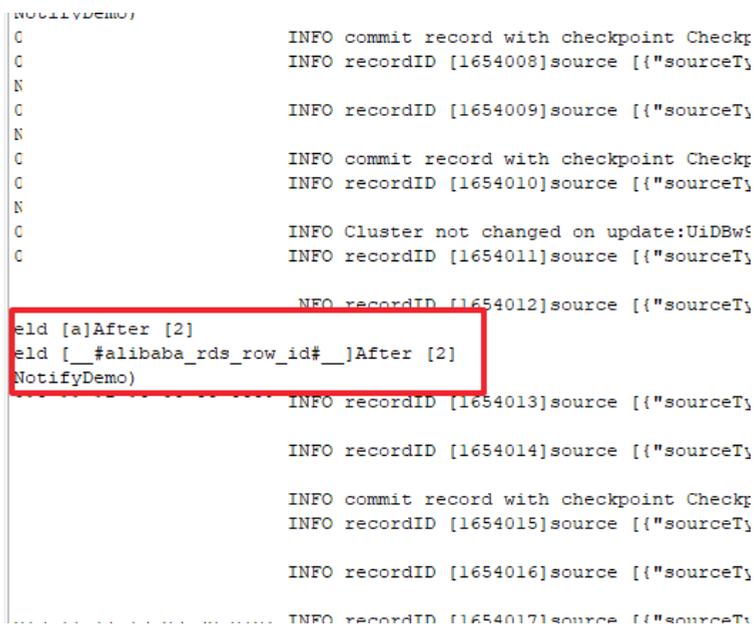


图 4-27

第五章 DTS 数据同步

前面了解了“[第三章 DTS 数据迁移](#)”，本章节详细讨论 DTS 数据同步。数据迁移与数据同步从逻辑结构与实现上来说，是完全一样的，同样分为结构迁移、全量迁移、增量迁移三个部分。只是数据同步的增量迁移是默认开启的（不可关闭）。所以无论是数据迁移还是数据同步，都可以实现实时的数据同步。数据迁移与数据同步在有重叠，那为什么还要有数据同步呢？数据同步比数据迁移也有很多的好处。两者的区别主要有如下 4 点：

- 数据同步需要先购买再使用，支持预付费（包年包月）和后付费（按量付费）2 种，而 DTS 迁移只支持后付费（按量付费）。
- 数据同步支持跨地域的内网同步（如华东 1 的 rds-> 华北 1 的 RDS）。而数据迁移对跨地域的迁移需要一侧有外网地址才可以，不支持内网迁移。
- 数据同步支持重新配置功能，数据迁移不支持，数据迁移如需要重新配置，需要新建任务。
- 数据同步支持在同步过程中增加同步对象与删除同步对象，而数据迁移不支持，数据迁移要新增和删除迁移对象，需要重新配置任务。

如下图 5-1 是 DTS 数据同步的页面。

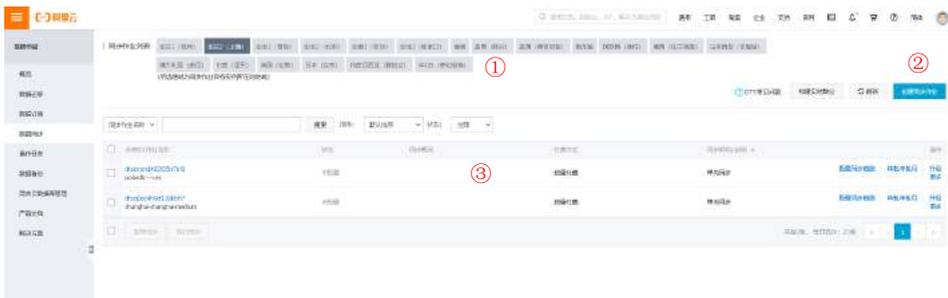


图 5-1

5.1 地域

图 5-1 标记①处为地域信息，此处可以切换各个地域查看不同地域的实例信息。这里选择的地域为“同步作业目标实例所在的地域”，假设您创建了一个华东 1 到华北 1 的数据同步任务，这个任务会展示到华北 1 地域。推荐从概览页进入。

5.2 创建同步作业页面

图 5-1 标记②的“创建同步作业”，点击后，跳转到 DTS 数据同步购买页面，如下图 5-2。下面详细说下各项配置项。

商品类型: 数据同步DTS (按付费) | 数据同步DTS (包付费)

功能: 数据迁移 | **数据同步** | 数据订阅

源实例: MySQL
支持RDS MySQL、阿里云ECS自建MySQL、本地自建MySQL及其他云厂商MySQL

源实例地域: 华东1 (杭州) | 华东2 (上海) | 华北2 (北京) | 华北3 (青岛) | 华北3 (张家口) | 华南1 (深圳) | 美国 (硅谷) | 美国 (弗吉尼亚)
新加坡 | 中国 (香港) | 阿联酋 (迪拜) | 德国 (法兰克福) | 马来西亚 (吉隆坡) | 印度 (孟买) | 印度尼西亚 (雅加达) | 澳大利亚 (悉尼)
日本 (东京) | 华北5 (呼和浩特) | 英国 (伦敦)

源地域为同步数据源实例所在地域，订购后不支持更换地域，请谨慎选择

目标实例: MySQL
支持RDS MySQL、ECS上自建MySQL及通过专线连接的自建MySQL或其他云厂商的RDS

目标实例地域: 华东1 (杭州) | 华东2 (上海) | 华北1 (北京) | 华北2 (北京) | 华北3 (青岛) | 华南1 (深圳) | 华北5 (呼和浩特)

目标地域为同步数据目标实例所在地域，订购后不支持更换地域，请谨慎选择
如需购买其国家或地区数据同步实例，请提交工单申请

同步拓扑: 串行同步 | **双向同步**

支持2个MySQL、2个PolarDB for MySQL、MySQL与PolarDB for MySQL、2个云Redis自建实例之间双向同步

网络类型: **专线**
为网络传输提供使用专线服务模式，目前仅支持专线模式，DTS提供专线，用户无需单独购买高速通道

同步规格规格: medium
规格选择参考: [mysql规格同步性能200 records/s](#)、[small规格同步性能2000 records/s](#)、[medium规格同步性能5000 records/s](#)、large规格限制，请参考 [数据同步规格说明](#)

购买数量: 1

立即购买 ¥1,020/小时

图 5-2

5.2.1 商品类型

该商品类型的含义与“[4.1.1 商品类型](#)”相同。

5.2.2 功能

该功能的含义与“[4.1.2 功能](#)”相同。

5.2.3 源实例

源实例是指 DTS 数据同步的源端数据库类型，DTS 支持 MySQL、DRDS(新名称为 POLARDB-X)、Redis、PostgreSQL、Polardb、SQLserver 的数据库同步。

5.2.4 源实例地域

待迁移的源实例所在的地域。此处的选择规则与“[3.3.2 实例地区](#)”类似。

5.2.5 目标实例

目标实例是指 DTS 数据同步的目标数据库类型，不同的“源实例”对应不同的目标实例，比如源实例为 MySQL 时，DTS 数据同步支持的目标实例有：MySQL、MaxCompute、Datahub、ADS for MySQL、kafka、POLARDB 等，比如源实例为 SQLserver，DTS 数据同步支持的目标实例只有 ADB for PostgreSQL。随着 DTS 的迭代升级，支持的类型也会不断的进行丰富。您配置时请以具体的页面展示为准。

5.2.6 目标实例地域

待迁移的目标实例所在的地域。此处的选择规则与“[3.3.2 实例地区](#)”类似。

5.2.7 同步拓扑

当前 DTS 数据同步支持单向同步和双向同步，单向同步就是 A→B 的同步，双向同步除了 A→B 外，还支持 B→A。双向同步相比单向同步要考虑的问题更多，比如数据冲突处理、rename 等数据库命令、DDL 命令处理等问题。单向与双向同步有如下几点需要注意。

5.2.7.1 单向同步

配置单向同步时，请注意如下 7 点：

- 如果同步对象为单个或多个表（非整库），那么在数据同步时，请勿对源库的同步对象使用 gh-ost 或 pt-online-schema-change 等类似工具执行在线 DDL 变更，否则会导致同步失败。
- 暂不支持中国（香港）可用区 A 的 RDS MySQL 实例配置数据同步。
- 配置数据同步，RDS MySQL 实例必须具备内网地址。
- 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
- 全量初始化过程中，并发 INSERT 会导致目标实例的表碎片，全量初始化完成后，目标实例的表空间比源实例的表空间大。
- 当同步对象为整个库，且库中的触发器（TRIGGER）会更新库内某个表时，可能导致源和目标库的数据不一致。请关闭目标库触发器。
- 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，如果同步过程中源实例将表 A 重命名为表 B，那么表 B 将不会被同步到目标库。为避免该问题，您可以在数据同步配置时，选择同步表 A 和表 B 所在的整个数据库作为同步对象。

5.2.7.2 双向同步

配置双向同步时，请注意如下 7 点：

- 为保障双向同步链路的稳定性，对于同一张表的 DDL 更新只能在其中一个同步方向进行同步。即一旦某个同步方向配置了 DDL 同步，则在反方向上不支持 DDL 同步，只进行 DML 同步。
- 对于数据冲突的检测，DTS 目前支持的冲突检测策略为 TaskFailed（遇到冲突，任务报错退出）、Ignore（遇到冲突，直接使用目标实例中的冲突记录）、Overwrite（遇到冲突，直接覆盖目标实例中的冲突记录）。
- 全量初始化过程中，并发 INSERT 会导致目标集群的表碎片，全量初始化完

成后，目标集群的表空间比源集群的表空间大。

- 如果数据同步的源集群没有主键或唯一约束，且记录的全字段没有唯一性，可能会出现重复数据。
- 如果同步对象为单个或多个表（非整库），那么在数据同步时，请勿对源库的同步对象使用 `gh-ost` 或 `pt-online-schema-change` 等类似工具执行在线 DDL 变更，否则会导致同步失败。
- 当同步对象为整个库，且库中的触发器（TRIGGER）会更新库内某个表时，可能导致源和目标库的数据不一致，请关闭目标库触发器。
- RENAME TABLE 操作可能导致同步数据不一致。例如同步对象只包含某个表，如果同步过程中源实例对该表执行了重命名操作，那么该表的数据将不会同步到目标库。为避免该问题，您可以在数据同步配置时将该表所属的整个数据库作为同步对象。

5.2.8 网络类型

DTS 数据同步通过专线进行内网数据传输，这也保证了 DTS 数据同步可以实现跨地域内网同步。

5.2.9 同步链路规格

DTS 数据同步也有链路规格的说法，DTS 数据同步的链路规格有 4 种，性能由低到高分别是 `micro`、`small`、`medium`、`large`（无限制规格，`large` 规格的线上运行性能依赖于网络环境、源实例和目标实例的性能。）。

性能可以参考这里 https://help.aliyun.com/document_detail/26605.html?spm=a2c4g.11186623.6.560.4f8d5f5fzFwoJ。

5.3 任务列表

图 5-1 标记③的处为所选择地域的任务列表。该页面展示与“[3.4 任务列表](#)”的页面含义相同。当您购买完 DTS 数据同步作业后，此处会显示，比如刚才创建

了一个华东 1-> 华东 2 的数据同步，点击华东 2 地域的任务列表，任务显示如下图 5-3。

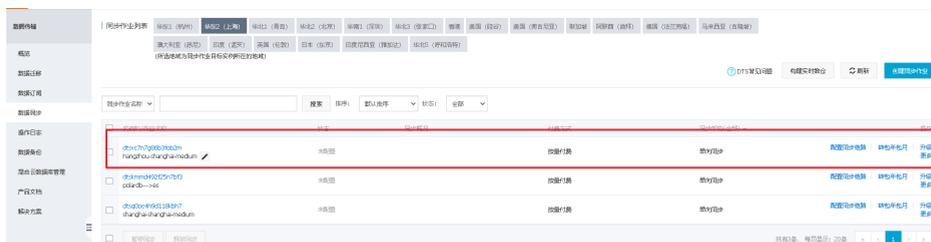


图 5-3

5.3.1 配置同步链路

DTS 数据同步购买完后，需要对其配置，才可以进行启动，点击“配置同步链路”跳转到 DTS 数据同步配置页面。如下图 5-4。由于之前购买 DTS 数据同步时选择的实例类型，地域都已经选定，此处的展示比 DTS 数据迁移也有一些区别。

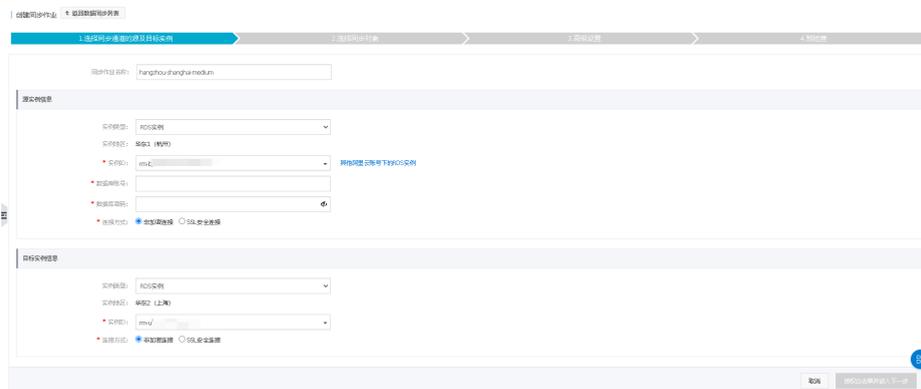


图 5-4

5.3.1.1 实例类型

该实例类型的含义与“[3.3.1 实例类型](#)”含义相同。

5.3.1.2 实例 ID

由于我们实例类型选择的时 RDS 实例，此处要选择一个 RDS 数据库实例，DTS 会自动检索华东 1 地域下您的实例信息，然后展示出来。该配置项与“4.2.1.2RDS 实例 ID”含义相同。

5.3.1.3 数据库账号 / 数据库密码

该配置项与“[4.2.1.3 数据库账号 / 数据库密码](#)”含义相同。

5.3.1.4 连接方式

该配置项与“[3.3.7 连接方式](#)”的含义相同。

5.3.1.5 授权白名单并进入下一步

该功能与“[3.3.9 授权白名单并进入下一步](#)”含义相同。点击后，跳转新的页面，如下图 5-5。该页面的含义与功能与“[3.3.9 授权白名单并进入下一步](#)”相同。

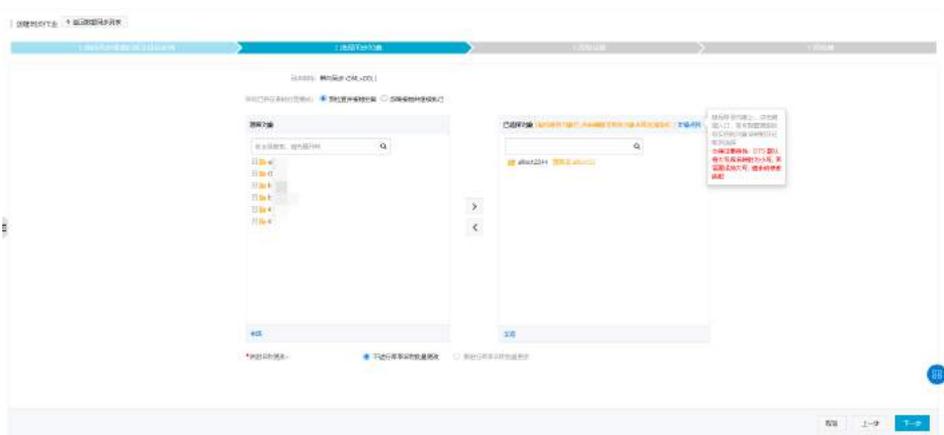


图 5-5

5.3.1.6 迁移类型

点击图 5-5 右下角的下一步，进入迁移类型选择页面。如下图 5-6。从这个页面也可以看到，您只能配置结构初始化（结构迁移）与全量数据初始化（全量数据迁

移), DTS 数据同步默认开启增量数据迁移。



图 5-6

5.3.1.7 预检查并启动

该功能与“[3.3.11 预检查并启动](#)”含义相同。如下图 5-7, 预检查完成且全部成功后, 点击关闭, DTS 数据同步开始进行数据同步, 如下图 5-8。



图 5-7



图 5-8

5.3.2 实例 ID/ 作业名称

该功能与“[3.4.1 ID/名称](#)”、“[4.2.2 订阅 ID/名称](#)”含义相同。点击后，如下图 5-9。

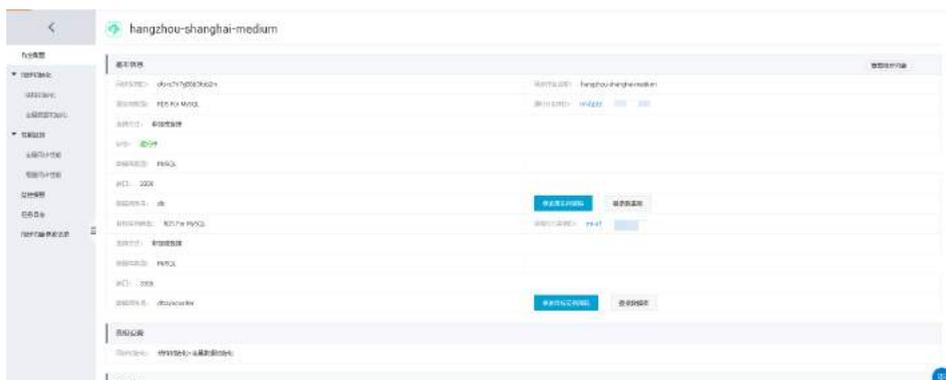


图 5-9

5.3.2.1 作业配置

该页面含义与“[3.4.1.1 任务配置](#)”、“[4.2.2.1 订阅配置](#)”含义相同。

5.3.2.2 同步初始化

该功能与“[3.4.1.2 迁移详情](#)”含义相同。如下图 5-10 和图 5-11。

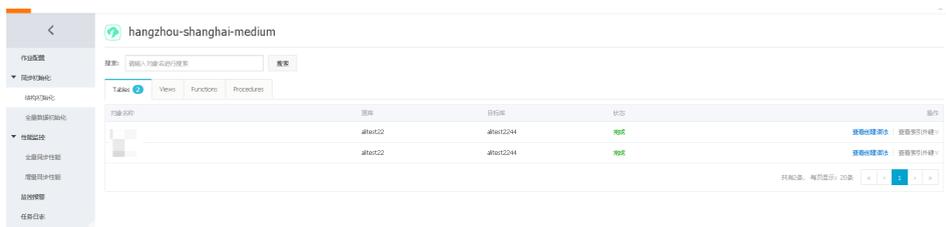


图 5-10



图 5-11

5.3.2.3 性能监控

该功能与“[3.4.1.3 性能监控](#)”含义相同，如下图 5-12 和图 5-13。

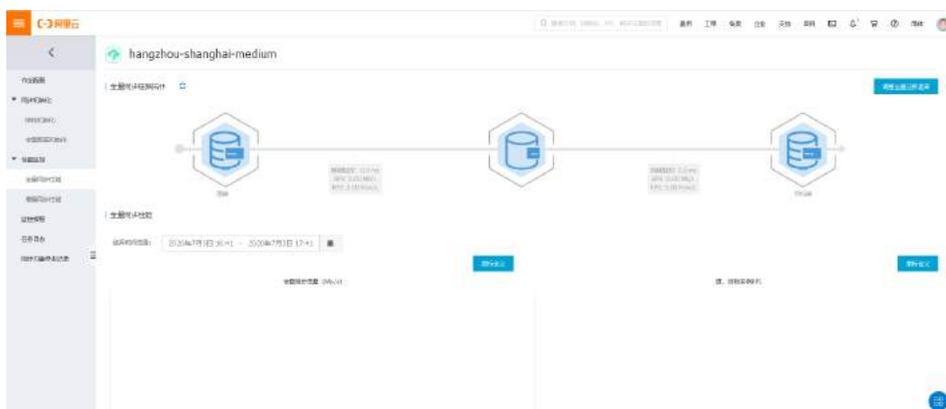


图 5-12



图 5-13

5.3.2.4 监控报警

该页面含义与“[3.4.7 监控报警](#)”相同。如下图 5-14。

报警名	状态	报警规则	报警联系人
同步延迟	已关闭		操作
同步失败	已关闭	同步状态 = 异常	操作 修改
同步速度	已关闭		操作 修改

5.3.3 同步概况

图 5-8 的“同步概况”，指的是这个同步任务的迁移速度以及延迟情况，DTS 数据同步的延迟与迁移速度与“DTS 数据迁移”的逻辑相同，同样是从源端获取数据（全量与增量），然后把数据写入到目标数据库。当出现延迟或者迁移速度不达标时，可以根据“[5.3.2.3 性能监控](#)”说明进行排查。

5.3.4 暂停任务

该功能与“[3.4.9 暂停任务](#)”含义相同。

5.3.5 升级

该功能与“[3.4.6 升级](#)”含义相同。

5.3.6 更多

点击更多，可以对 DTS 数据同步任务进行更多配置，如下图 5-15。重点说下“修改同步对象”与“转为双向同步”这两个功能。



图 5-15

5.3.6.1 修改同步对象

DTS 数据同步支持在任务运行过程中增加或者删除同步对象，点击“修改同步对象”功能，跳转到新的对象配置页面，如下图 5-16，在这个页面，您可以把不需要同步的对象移除，也可以把新增需要同步的对象添加到同步列表，比如此处，我们

把 sbtest 数据库加入同步对象 (新增同步对象)。

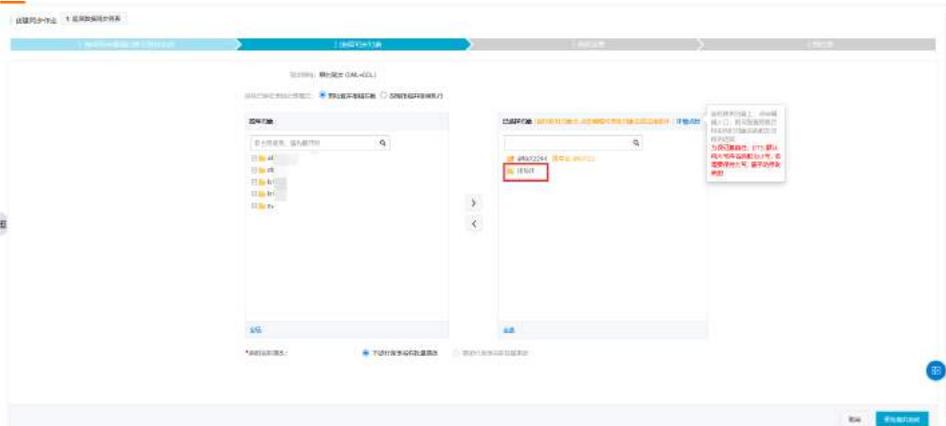


图 5-16

点击预检查并启动后，预检查通过后，DTS 同步状态会变成“新增对象初始化中”，如下图 5-17。这个状态可能会持续很久，DTS 的新增同步对象是通过启动一个子 DTS 同步任务完成的，也就是 DTS 的后台会起到一个当前任务下的子任务按照结构 + 全量 + 增量 (何种迁移类型取决于主任务创建时如何配置) 迁移新增的对象。然后等迁移完成后，合并到主任务中 (合并时是增量迁移合并，也就是说合并的前提是子任务已经正常进入了增量迁移)。合并完成后“新增对象初始化中”会小时。如果您遇到“新增对象初始化中”持续很长时间 (比如 1 天) 可以反馈阿里云售后核实。需要注意，这类新增对象的操作不可取消。



图 5-17

5.3.6.1 转为双向同步

单向数据同步可以直接升级为双向同步，只有任务状态为正常同步中状态时，该选项才可选择（比如当任务状态为新增对象初始化中时，该功能不可用）。点击后，任务列表中该任务会变成如下图 5-18 的展示形式。

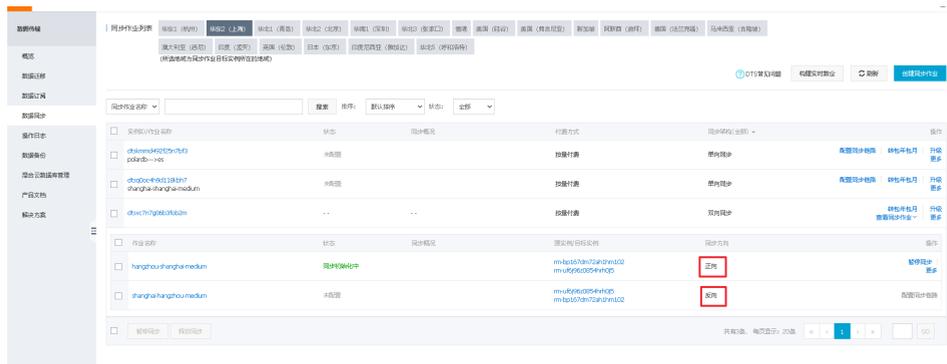


图 5-18

其中，之前配置好的任务变成了正向，为了实现双向同步，我们还需要配置反向同步任务才可以，点击“配置同步链路（等待由灰色变为蓝色）”，跳转配置页面如下图 5-19。

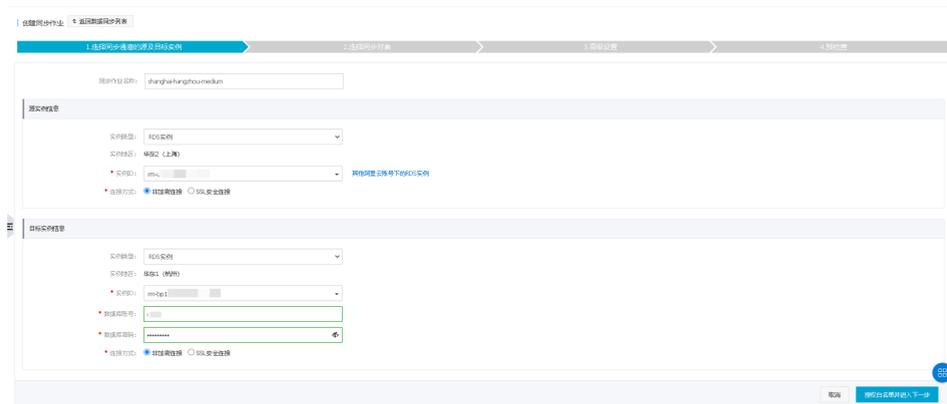


图 5-19

继续点击“授权白名单并进入下一步”，跳转对象选择页面，如下图 5-20，选择 DML 同步类型、冲突修复策略以及同步对象后，点击下一步，选择迁移类型，然后启动预检查。至此，双向同步配置完成。

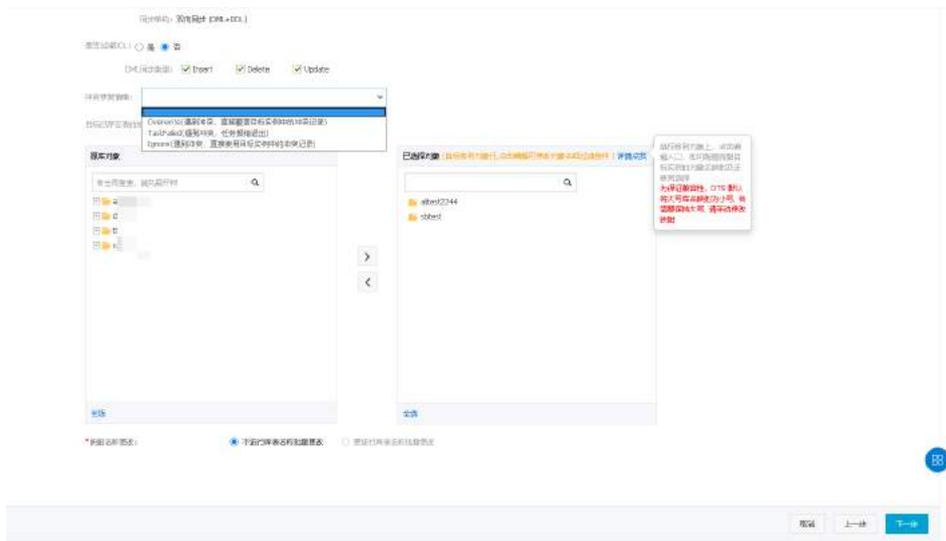
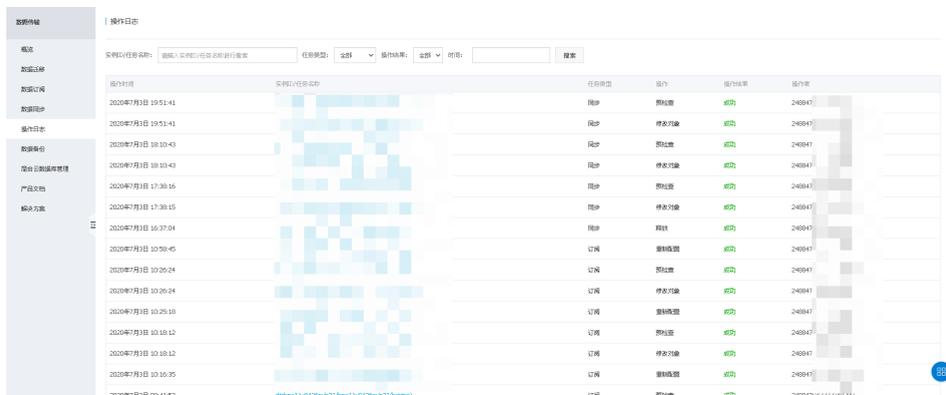


图 5-20

第六章 操作日志

如下图 6-1，记录了阿里云账户对 DTS 的操作记录，需要注意如下 1 点：

- 这里的记录，只记录当前账户的操作记录，这里的“当前账户”是指登陆阿里云控制台的账户，包括主账户和子账户，比如有 2 个子账户分别是 A 与 B。B 账户登陆阿里云 DTS 控制台无法看到 A 账户的操作记录（只能看到 B 账户自己的）。



操作时间	实例ID/任务名称	任务类型	操作	操作结果	操作者
2020年7月31日 19:51:41		同步	修改表	成功	240047
2020年7月31日 19:51:41		同步	修改表	成功	240047
2020年7月31日 19:10:43		同步	修改表	成功	240047
2020年7月31日 18:10:43		同步	修改表	成功	240047
2020年7月31日 17:38:16		同步	修改表	成功	240047
2020年7月31日 17:38:15		同步	修改表	成功	240047
2020年7月31日 16:37:04		同步	修改	成功	240047
2020年7月31日 10:58:45		订阅	重新配置	成功	240047
2020年7月31日 10:20:24		订阅	修改表	成功	240047
2020年7月31日 10:20:24		订阅	修改表	成功	240047
2020年7月31日 10:20:18		订阅	重新配置	成功	240047
2020年7月31日 10:18:12		订阅	修改表	成功	240047
2020年7月31日 10:16:35		订阅	重新配置	成功	240047
2020年7月31日 09:41:42		订阅	修改表	成功	240047

图 6-1



云服务技术大学
云产品干货高频分享



云服务技术课堂
和大牛零距离沟通



阿里云开发者“藏经阁”
海量免费电子书下载