解决方案实践

# 快速搭建 DataX 数据库迁移环境

文档版本1.0.0发布日期2024-10-12





#### 版权所有 © 华为技术有限公司 2024。 保留一切权利。

非经本公司书面许可,任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部,并不得以任何形式传播。

#### 商标声明

#### 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束,本文档中描述的全部或部分产品、服务或 特性可能不在您的购买或使用范围之内。除非合同另有约定,华为公司对本文档内容不做任何明示或暗示的声 明或保证。

由于产品版本升级或其他原因,本文档内容会不定期进行更新。除非另有约定,本文档仅作为使用指导,本文 档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 安全声明

## 漏洞处理流程

华为公司对产品漏洞管理的规定以"漏洞处理流程"为准,该流程的详细内容请参见如下网址: https://www.huawei.com/cn/psirt/vul-response-process 如企业客户须获取漏洞信息,请参见如下网址: https://securitybulletin.huawei.com/enterprise/cn/security-advisory

目录

1 方案概述	1
2 资源和成本规划	
3 实施步骤	5
3.1 准备工作	5
3.2 快速部署	8
3.3 开始使用	
3.4 快速卸载	24
4 附录	
5 修订记录	27



## 应用场景

该方案帮助您基于**DataX**快速实现异构数据源之间的高效同步迁移上云。它是一款开源、高效、易用的异构数据源离线同步工具,致力于实现包括关系型数据库(MySQL、Oracle等)、HDFS、Hive、ODPS、HBase、FTP等各种异构数据源之间稳定高效的数据同步功能。

## 方案架构

该解决方案基于华为云部署DataX数据库迁移环境,帮助您在华为云上快速实现异构数据源之间的高效同步迁移。解决方案架构图如下:



#### **图 1-1** 方案架构图

部署该方案中需要使用的资源:

- 创建一台Flexus<mark>云服务器X实例</mark>,部署数据库迁移的DataX同步环境。
- 创建一个弹性公网IP EIP,并绑定到云服务器,用于提供访问公网和被公网访问能力。

• 创建安全组,通过配置安全组规则,为云服务器提供安全防护。

## 方案优势

- 并行执行
   采用了分布式架构和多线程技术,能够并行地从多个数据源抽取数据,并同时将数据加载到目标数据源中,大大提高了数据同步的效率。
- 易配置管理
   DataX 提供了直观的 JSON 配置文件格式,用户只需编写简单的配置文件,指定数据源、目标数据源、数据转换规则等信息,即可启动数据同步任务。
  - 数据转换处理灵活 在数据同步过程中,DataX 提供了丰富的数据转换功能。可以进行字段映射、数 据清洗、格式转换、数据过滤等操作,确保数据在不同数据源之间的一致性和准 确性。

# 约束与限制

- 该解决方案部署前,需注册华为账号并开通华为云,完成实名认证,且账号不能 处于欠费或冻结状态。如果计费模式选择"包年包月",请确保账户余额充足以 便一键部署资源的时候可以自动支付;或者在一键部署的过程进入费用中心,找 到"待支付订单"并手动完成支付。
- 如果选用IAM委托权限部署资源,请确保使用的华为云账号有IAM的足够权限,具体请参考创建rf\_admin\_trust委托;如果使用华为主账号或admin用户组下的IAM子账户可不选委托,将采用当前登录用户的权限进行部署。



该解决方案主要部署如下资源,以下费用仅供参考,具体请参考华为云官网<mark>价格详</mark> 情,实际收费以账单为准。

华为云服务	配置示例	每月预估花费
Flexus云服务器X实例	● 区域:华北-北京四	248.00元
	• 计费模式:包年包月	
	<ul> <li>规格: Flexus云服务器X实例   性能 模式(关闭)   x1.2u.4g   2vCPUs   4GiB</li> </ul>	
	• 镜像: CentOS 7.6 64bit	
	● 系统盘: GPSSD   100GB	
	● 数据盘: GPSSD   100GB	
	• 购买量: 1	
弹性公网IP EIP	● 按需计费: 0.82元/GB	0.82 元/GB
	● 区域:华北-北京四	
	<ul> <li>计费模式:按需计费</li> </ul>	
	● 线路: 动态BGP	
	<ul> <li>公网带宽: 按流量计费</li> </ul>	
	• 购买数量:1	
合计		248.00元 + EIP流 量费用

表 2-2 资源和成本规格-云服务器部署(排	安需计费)
------------------------	-------

华为云服务	配置示例	每月预估花费
Flexus云服务器X实例	• 按需计费: 0.42元/小时	301.68元
	● 区域:华北-北京四	
	<ul> <li>计费模式:按需计费</li> </ul>	
	<ul> <li>规格: Flexus云服务器X实例   性能 模式(关闭)   x1.2u.4g   2vCPUs   4GiB</li> </ul>	
	● 镜像:CentOS 7.6 64bit	
	● 系统盘: GPSSD   100GB	
	● 数据盘: GPSSD   100GB	
	• 购买时长: 720小时	
	● 购买量:1	
弹性公网IP EIP	● 按需计费: 0.82元/GB	0.82 元/GB
	● 区域:华北-北京四	
	<ul> <li>计费模式:按需计费</li> </ul>	
	● 线路: 动态BGP	
	<ul> <li>公网带宽: 按流量计费</li> </ul>	
	● 购买数量:1	
合计		301.68元 + EIP流 量费用



- 3.1 准备工作3.2 快速部署
- 3.3 开始使用
- 3.4 快速卸载

# 3.1 准备工作

当您首次使用华为云时注册的账号,则无需执行该准备工作,如果您使用的是IAM用 户账户,请确认您是否在admin用户组中,如果您不在admin组中,则需要为您的账号 授予相关权限,并完成以下准备工作。

# 创建 rf\_admin\_trust 委托 (可选)

**步骤1** 进入华为云官网,打开<mark>控制台管理</mark>界面,鼠标移动至个人账号处,打开"统一身份认证"菜单。

图 3-1 控制台管理界面



图 3-2 统一身份认证菜单

简体中文	ETRETRETRET KURTUR
基本信息	已实名认证
安全设置	
我的凭证	
统一身份认	、证
切换角色	
标签管理	
操作日志	
	退出

**步骤2** 进入"委托"菜单,搜索"rf\_admin\_trust"委托。

**图 3-3** 委托列表

委托 ①					(	创建委托
部除您还可以创建45个委托						
金部英型  ★ば名称ID 合	্ rf_admin_trust রান্যাক ≙	★任时长 <u>△</u>	Alteria A	#22 合	122.4%	X
rf_admin_trust	云服务 RFS	永久 一	2024/10/17 15:07:18 GMT+08:00		授权 修改 删除	

- 如果委托存在,则不用执行接下来的创建委托的步骤
- 如果委托不存在时执行接下来的步骤创建委托
- **步骤3** 单击步骤2界面中的"创建委托"按钮,在委托名称中输入"rf\_admin\_trust",委托 类型选择"云服务",输入"RFS",单击"完成"。



步骤5 在搜索框中输入"Tenant Administrator"并勾选搜索结果,单击"下一步"。

上一歩
 横定

#### **图 3-6** 选择策略

1 选择策	(2) RER-MICEN (3) Rol				
载托"rf_adn	in_trust"将拥有所造策略				新建策略
118	510 AR#ESHQEESHQR	全部类型	✓ 所有云服务	<ul> <li>         、         (         根除接条 ~ )         (         Tenant Administrator         </li> </ul>	×IQ
	80	类型			
	Tenant Administrator	7.48.0			

步骤6选择"所有资源",并单击"确定"完成配置。

#### 图 3-7 设置最小授权范围

授权	
✓ 选择预整 — 2 设置最小技程规范据 3 元化	
● 根据当前忽所态择的策略,系统推荐以下授权范围方案。更像于它最小化提权,可进行选择,了解如何根据您的应用场量选择合适的提权范围方案	×
选择授权范围方案	
(例)所有资源 提权后, MM用户可以供提权限使用账号中所有资源,包括企业项目、区域项目和全局服务资源, 展开其他方案	

步骤7 "委托"列表中出现 "rf\_admin\_trust" 委托则创建成功。

#### **图 3-8** 委托列表

委托 ①						创建委托
部除 您还可以创建45个委托						
全部类型 >	Q rf_admin_trust					×
要托名称/ID      令	委托对象 令	委托时长 令	创建时间 🔶	描述 令	操作	
rf_admin_trust	云服务 RFS	永久 一	2024/10/17 15:07:18 GMT+08	:00	授权 修改 删除	

----结束

# 3.2 快速部署

本章节主要帮助用户快速部署"快速搭建DataX数据库迁移环境"解决方案。

#### **表 3-1** 参数说明

参数名称	类型	是否可选	参数解释	默认值
subnet_id	strin g	必填	子网ID。如使用已有虚拟私有云 VPC,请填写已有VPC子网ID;如使 用新建VPC须填写new,新建VPC名 称同云服务器名称,子网名称: {ecs_name}-subnet。获取已有VPC 子网ID。	new
security_g roup_nam e	strin g	必填	安全组名称。该模板使用新建安全 组。取值范围:1-64个字符,支持数 字、字母、中文、_(下划线)、-(中 划线)、.(点)。	creating-a- database- migration- environment- with-datax
ecs_name	strin g	必填	云服务器名称。不允许重名。取值范 围:1-56个字符组成,支持字母、数 字、中文、下划线(_ )、中划线 (- )、英文句号(. )。	creating-a- database- migration- environment- with-datax
ecs_flavor	strin g	必填	云服务器实例规格,支持弹性云服务器 ECS及华为云Flexus 云服务器X实例。Flexus 云服务器X实例规格ID命名规则为x1.?u.?g,例如2vCPUs4GiB规格ID为x1.2u.4g,具体华为云Flexus 云服务器X实例规格请参考控制台。具体规格请参考官网弹性云服务器规格清单。	x1.2u.4g
system_di sk_size	num ber	必 填	云服务器系统盘大小,磁盘类型默认 为通用型SSD。以GB为单位,取值范 围为40~1,024,不支持缩盘。	100
data_disk_ size	num ber	必 填	云服务器数据盘大小,磁盘类型默认 为通用型SSD。以GB为单位,取值范 围为10-32,768。	100
ecs_passw ord	strin g	必填	云服务器初始密码。取值范围:长度 为8-26位,密码至少必须包含大写字 母、小写字母、数字和特殊字符(! @\$%^=+[{}]:,./?)中的三种,密码 不能包含用户名或用户名的逆序。管 理员帐户为root。	空
charging_ mode	strin g	必 填	计费模式,默认自动扣费,取值为 prePaid(包年包月)或postPaid (按需计费)	postPaid

参数名称	类型	是否可选	参数解释	默认值
charging_ unit	strin g	必填	包年或包月,有效值为"year"或 "month"。当charging_mode(计 费模式)为prePaid时,此选项为必 填项。	month
charging_ period	num ber	必填	包年包月时长,当charging_unit取 值为"year",取值范围为1-3;取 值为"month",取值范围为1-9。 当charging_mode(计费模式)为 prePaid时,此选项为必填项。	1

# **步骤1** 登录**华为云解决方案实践**,选择"快速搭建DataX数据库迁移环境",单击"一键部署",跳转至解决方案创建资源栈界面。

#### 图 3-9 解决方案实施库



步骤2 在选择模板界面中,单击"下一步"。

#### **图 3-10** 选择模板

< 立即创建资	2029년	
<b>1</b> 259885 —	② 配置参款 ———— ③ 光源性论量 ———— ④ 配置喇叭	
* 创建方式	已和國家 在可能化傳播器的證	
* 模板來源	我的爆集         URL         上級爆集           地分配用板板板上手で取り出版法、機能や心の原本を感じたりだけ、(1*菜本3ml21.pxn)、         (1*菜本3ml21.pxn)、         (1*菜本3ml21.pxn)、	
★ 模板 URL	Mgs:/Ibconsentation-samples.abs:cn-sem14-mptu samm至少需要考虑要考虑是不会,文明不可能是114m。	
	● 资源编标器员不会在管理管理之外的结果使用定上物的数据,我们不会对您的编版进行加密,对于参数中的编译数据,推荐使用KMS或DEVHT管进行加密。目前控制台中支持自动使用KMS如图的编版参数。	
		G G G
		<b>T</b> -4

**步骤3**在配置参数界面中,参考表 表1 参数说明 完成自定义参数填写,单击"下一步"。

#### **图 3-11 配置参数**

立即创建资源栈	reifer 3 estivate 4 reifer		
★ 资源线名称 creatin 资源线名	g a-dalabase migration environment with-datax 第公讲文或高文开始、支持中文、大小写高文、数字、下刻纸、中和纸、120个字符之内、 <b>数用线名称不能要名。</b>		
相述 快速開	國DataXIII 集團任務环境		
	18255 A		
<b>记置参数</b> 请输入关键字搜索参数名	称 Q M 機械振動物の分類激励率 ①		
受数名称	-	类型	織送
subnet_id	new	string	子何D, 如使用已有些料料有云 VPC,谦填弓已有VPC子何D; 如使用射理VPC原填写new,射理VPC高标号无限务器名称,子宛名称:{ecs_name}-subn
security_group_name	creating-a-database-migration-environment-with-datax	string	安金给名称,读微胶很用新建安金结,取描范围:1-64个字符,觉得数字、字母、中文、_(下划)统),+(中划)统),-(作),数认:creating-a-database-ml
ecs_name	creating-a-database-migration-environment-with-datax	string	云振身器名称,不允许重名,取道范围:1-59个字符组成,支持字母、数字、中文、下划线(_)、中划线(-)、英文号号(.)。默认:creating-d-databa…
ecs_flavor	x1.2u.4g	string	云陽易難奕彻現態,支持神性云陽易難 ECS及除为云Flexus 云陽易離X寀例,Flexus 云服易離X寀例規图ID命名規則为x1.?u.?g,例如2vCPUs4GB規图D
system_disk_size	100	number	云磁各轉系統最大小,磁曲與型款以为通用型SS0,以GB为单位,取道范围为40-1,024,不支持编曲。款以为100。
* data_disk_size	100	number	云振身轉動攝盘大小,磁盘與國歌以为適用型SSD。以GB为单位,取道范围为10-32,768。數以为100,
			上一步 下

**步骤4** (可选,如果使用华为主账号或admin用户组下的IAM子账户可不选委托)在资源设置 界面中,在权限委托下拉框中选择"rf\_admin\_trust"委托,单击"下一步"。

#### **图 3-12** 资源栈设置

<   立即创建资源线	
⑦ 选择模板 ──── ()	S KEBAR — () KEMAL
LAMIQ現費IE	AND DECEMBER OF A CONTRACTOR AND A CONTR
	● 用于不是型的现象にあれる本本本は重要用于他の加速分析器。 6月間を10月で有一般といわり目前と見たけ、可知の進き数据の集集等を影響電源集集を影響電源集集を影響電源集集を計算電源
医液淀量	HARRANJAN, MACHARLENSKANDL-TACHERNA, ARTICERRANJANAC,
图8/程/4	Delinedende, deligibilitzation,
	54

步骤5 在配置确认界面中,单击"创建执行计划"。

**图 3-13** 配置确认

模拟 ———— (一) 配置参数	x 🕗 XIIINA		
在漂繞排服务使用免费,但近深	\$内包含的资源可能需要付费。目前留不支持在当前页查询价格,您可通过 创	建执行计划 (免费) 后获取	级给霸用.
擇欄板			
<b>那线名称</b>	creating-a-database-migration-environment-with-datax		组织发 使温度强度DataX数据库证标环境
電参数 2			
教名称	an a	类型	描述
ubnet_id	new	string	子句D、如使用已有或拟私有云 VPC、清晰有已有VPC子问D:如使用新疆VPC须频写new、新疆VPC名称母云服装器名称,子母名称:(ecs_name)-subnet、政策已有VPC子
ecurity_group_name	creating-a-database-migration-environment-with-datax	string	安全協名称,法構版使用新聞安全編,取贏百團;1-64个字符,支持数字、字母、中文、_(下划线)、-(中划线)、-(点),默认: creating-a-database-migration-environment-
cs_name	creating-a-database-migration-environment-with-datax	string	云服务器名称,不允许重名,取重范围:1-56个字符组成,支持字母、数字、中文、下划线(_)、中划线()、英文句号(),默认:creating-a-database-migration-environ
ics_flavor	×1.2u.4g	string	云服房器实例烦恼,支持神性云服存器 ECS现华为云Fixeus 云服房器X实例,Fixeus 云服房器X实例烦恼/D命名规则为r1.7u.7g,例如2xCPUs40B烦恼/D为r1.2u.4g,具体华
	100	number	云展势震奏纯盘大小,磁盘角型默认为通用型SSD,以GB为单位,取值范围为40~1,024,不变持缩盘。默认为100。
ystem_disk_size		number	云服务器数据曲大小、磁曲线型默认为道用型SSD,以GB为单位、取蛋范面为10-32,768,默认为100,
ystem_disk_size	100		
ystem_disk_size lata_disk_size cs_password	100	string	云服身醫初始密码。取遺征國:长度为0-28位,密码至少必须如食大喝字母。小喝字母。数字和時務字符(l@s%/=+{{}}::-??)中的三种,密码不能包含用户名或用户名的进序。
ystem_disk_size	100 	string	王振政権の決定時、記道王王、北京大学校2、王明王ン公院放出大学学校、小学学校、数学校5時件学行(1894~-101.7)中的三件、王明不能放出時から武司からの出す、 は第三年、記した王王、 2015年2月、18月7日の今日(18月1日)、18月2日の中国(18月1日)、第11日の中国人

步骤6 在弹出的创建执行计划框中,自定义填写执行计划名称,单击"确定"。

 $\times$ 

#### **图 3-14** 创建执行计划

创建执行计划							
1) 通过执行计划,可以	预览您的资源变更信息。						
	[						
★ 执行计划名称	executionPlan_20241010_1018_4b39						
描述	请输入对执行计划的描述						
		0/255 4					
	确定	取消					

#### 步骤7 单击"部署",并且在弹出的执行计划确认框中单击"执行"。

#### **图 3-15** 执行计划

<   creating-a-database-migration-environment-with-datax								影钟 更新模板或参数	С	
基本信用	送源	输出 導	种	欄板	执行计划					
1	8								请输入关键字	Q
	执行计划名称ID			状态	翡用預估 ⑦	9009360	描述	操作		
	executionPlan_20241010_1018_4b39 c13a8c57-1c44-4e80-9d25-222f8b10d20c			创建成功,特部署	查審測用時間	2024/10/10 10:19:20 GMT+08:00	-	部署 部除		

**图 3-16**执行计划确认

<b>执行计划</b>		
执行计划名称	状态	创建时间
executionPlan_20241010_101	创建成功,	2024/10/10 10:19:20 GMT+08
executionPlan_20241010_101 ① 确定执行后,资源栈会按照 费要求,可能会产生费用。	创建成功,… 。 该计划更新,并且:	2024/10/10 10:19:20 GMT+08. 会开通模板内的资源,根据资源付

- **步骤8** (可选)如果计费模式选择"包年包月",在余额不充足的情况下(所需总费用请参考资源和成本规格-云服务器部署(包年包月))请及时登录费用中心,手动完成待支付订单的费用支付。
- **步骤9** 待"事件"中出现"Apply required resource success",表示该解决方案已经部署完成。大约等待10-15分钟,DataX数据库同步环境即可搭建完成。

#### 图 3-17 部署完成

<   creating-a-dat	abase-migratic	影除 更新博技或参数	С			
基本信息 资源	輸出 制格	模板 执行计划				
					资源名称 > 请给入关键字	Q
事件时间 😜		血体菜園 Δ	即件描述	资源名称 类型	关联资源ID	
2024/10/10 10:22:29	GMT+08:00	日本	Apply required resource success.	-	-	
2024/10/10 10:22:26	GMT+08:00	生成完成	husweicloud_compute_instance.ecs: Creation complete after 1m15s [id=c3962686.7e0e.4463.8089. af52c2aab40a]	ecs ECS	c3962686-7e0e-4463-8089-af52c2aab40a	

----结束

# 3.3 开始使用

## 验证 DataX 迁移服务正常运行

🛄 说明

- 有关DataX中间件说明请参考DataX开源地址。
- 该解决方案使用的资源版本: Java版本: JDK-22。
- 请参考添加安全组规则,配置对应数据库服务器IP地址白名单,以便能正常使用服务。
- 注意: 该服务分为新建VPC和已有VPC两个版本,保证DataX服务器和目标数据库服务器在一个子网下,实现私网的连接与访问,提高迁移数据的安全性。
- **步骤1** 登录Flexus云服务器X实例控制台,选择创建的云服务器,单击远程登录,进入Linux 云服务器。

 $\times$ 

#### **图 3-18** 登录云服务器

lexus云服务器X实例 ⑦										Ē-€	eisti 🕻	购买FI	exusX
● 減差您参加Flexus云服务器X实例使用体验调研,您宝贵的意见和建议是	被心持续提升?	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	源动力,感谢您的	参与!									×
我的X实例: 华南-广州 (10) 半北-北京四 (7) 日 亚太-新加坡 (3)	华北-乌兰寨	病一 (1)	华东·上海一 (1)										Q
开机 关机 重局 重置素码 <b>更多 ~</b>	) (77)	)											
<ul> <li>、 默认按照名称搜索、过速</li> </ul>											0	00	۲
_ 名称/ID ⊖	监控	安全	状态 ⊖	可用区 ⊖	規格:模像 ⊖	操作系统类型 🖯	IP地址 ⊖	计赛模式 ⊖	企业项目 ⊖	标签 ⊖	操作		
creating-a-database-migration-environment-with-datax	e	0	运行中	可用区1	2vCPUs   4GiB CentOS 7.6 64bit	Linux	192.168.1	按需计赛 2024/10/10 1	default	-	远程登录	更多。	

#### 图 3-19 登录 Flexus 云服务器

登录Linux云服务器								
<b>请确保安全组已放通CloudShell连接实例使用的端口(默认使用22端口)</b> 优势:操作更流畅,命令支持复制粘贴,支持浏览输出历史和多终端分区布局。 <mark>了解更多</mark> CloudShell登录								
其他方式								
<ol> <li>使用控制台提供的VNC方式登录</li> <li>立即登录</li> <li>2、使用Putty、Xshell等工具登录Linux云服务器, 了解更多</li> </ol>								
a) 下载安装登录工具,例如Putty。 b) 输入弹性公网IP。								

#### 步骤2 在Linux云服务器中输入账号和密码后回车。

**步骤3** 输入命令: cd /opt/datax进入该目录,执行python3 bin/datax.py job/job.json查看回显,如回显为下图红框所示,则表示datax服务成功搭建。

图 3-20 执行 datax 运行命令



#### **图 3-21** 查看执行回显

			"encodina": "UTF-8"							
1										
}										
2024-10-17	14-25-22 055 [m	ain] TNEO	ParfTraca - ParfTraca	traceId-job -1 jeEna	bla-falea					
2024-10-17	14:25:32.955 [ma	ain] INFO	JobContainer - DataX	iobContainer starts jo	b.					
2024-10-17	14:25:32.956 [ma	ain] INFO	JobContainer - Set jo	0 = bId						
2024-10-17	14:25:32.984 [jo	bb-0] INFO	JobContainer - jobCor	itainer starts to do p	repare					
2024-10-17	14:25:32.984 []C	DD-0 INFO	JobContainer - Datax	Reader.Job [streamrea	der] do prepare work					
2024-10-17	14:25:32.984 [jo	bb-0] INFO	JobContainer - jobCon	tainer starts to do s	plit					
2024-10-17	14:25:32.985 [jo	bb-0] INFO	JobContainer - Job se	et Channel-Number to 1	channels.					
2024-10-17	924-10-17 14:25:32.985 [job-0] INFO JobContainer - DataX Reader Job [streamreader] splits to [1] tasks.									
2024-10-17	14:25:32.965 []C	ob-0] INFO	JobContainer - Jatax	tainer starts to do s	chedule	ISKS.				
2024-10-17	14:25:33.025 [je	bb-0] INFO	JobContainer - Sched	ller starts [1] taskGr	oups.					
2024-10-17	14:25:33.032 [jo	bb-0] INFO	JobContainer - Runnir	ng by standalone Mode.						
2024-10-17	14:25:33.037 [ta	askGroup-0]	INFO TaskGroupConta	ner - taskGroupld=[0]	start [1] channels i	or [1] tasks.				
2024-10-17	14:25:33.039 Tta	askGroup-01	INFO Channel - Chan	hel set record speed l	imit to -1. No tos ac	tivated.				
2024-10-17	14:25:33.049 [ta	askGroup-0]	INFO TaskGroupConta	iner - taskGroup[0] ta	skId[0] attemptCount	1] is started				
2024-10-17	14:25:33.149 [ta	askGroup-0]	INFO TaskGroupConta	iner - taskGroup[0] ta	skId[0] is successed,	used[102]ms				
2024-10-17	14:25:33.150 Lta	askGroup-0]	StandAloneJobContain	ner - taskGroup[0] co	mpleted it's tasks. 100000 records 2000	1888 hytes   Speed 2	53 91KB/s 18888 FRCOT	de/e   Error A recorde	A bytes I All Task W	
itWriterTim	∎e 0.052s   All	Task WaitP	ReaderTime 0.062s   Pe	centage 100.00%	100000 100000, 2000	ooo bytes - opeed r			, o bytes   - xee rusk w	
2024-10-17	14:25:43.048 [jo	bb-0] INFO	AbstractScheduler - S	Scheduler accomplished	all tasks.					
2024-10-17	14:25:43.048 []0	bb-0] INFO	JobContainer - DataX	Writer.Job [streamwri	ter] do post work.					
2024-10-17	14:25:43.048 []0 14:25:43.048 []0	ob-0] INFO	JobContainer - DataX	iohId [0] completed s	derj do post work. uccessfully.					
2024-10-17	14:25:43.049 [jo	bb-0] INFO	HookInvoker - No hoo	invoked, because bas	e dir not exists or i	s a file: /opt/data	ax/hook			
2024-10-17	14:25:43.051 [jo	ob-0] INFO	JobContainer -							
Lt	otal cpu intoj =	*>			L minDeltaCou					
	-1.00%		-1.00%	u	mtnuettacpu   -1.00%					
C+	total oc infal -									
	NAME		totalGCCount	maxDeltaGCCount	minDeltaGCCount	totalGCTime	maxDeltaGCTime	I minDeltaGCTime		
	G1 Young 0	Generation				0.000s	0.000s	0.000s		
	G1 Old Ger	neration	10	0	0	0.000s	0.0005	0.000s		
	G1 Concurr	rent GC	1.0	1.6	10	0.000s	0.0005	0.000s		
2024-10-17	14:25:43.051 [jo	ob-0] INFO	JobContainer - PerfT	race not enable!						
2024-10-17	14:25:43.052 [jo	bb-0] INFO	StandAloneJobContaine	erCommunicator - Total	100000 records, 2600	1000 bytes   Speed 2	53.91KB/s, 10000 recor	ds/s   Error 0 records	, 0 bytes   All Task W	
itWriterTum	te 0.052s   All	Task Waith	teaderTime 0.062s   Per	centage 100.00%						
任务启动时期	51	: 202	24-10-17 14:25:32							
任务结束时刻	<u>я</u>	: 202	4-10-17 14:25:43							
任务总计和	র্ষ		105							
任务于均加重			253.91KB/S							
<b>读出记录总</b>	<u>ک</u>		100000							
读写失败总数	ά –		θ							

----结束

## 简介 DataX 迁移配置模板文件生成

#### 🗀 说明

- DataX的数据同步工作,需要使用json文件来保存配置信息。配置writer、reader等信息,我 们可以使用命令来生成一份配置的json模板,在这个模板上进行修改,生成最终的json文件。
- 注意:目标数据库writer配置中的jdbcUrl字段值为字符串。

#### **步骤1** 运行如下命令生成模板文件,将其中的{reader}替换成自己想要的reader组件名字,将 其中的{writer}替换成自己想要的writer组件名字。

python3 bin/datax.py -r {reader} -w {writer}

#### 支持的reader:

cassandrareader drdsreader hbase094xreader hbase20xsqlreader loghubreader oceanbasev10reader oraclereader otsstreamreader sqlserverreader tdenginereader clickhousereader ftpreader hbase11xreader hdfsreader mongodbreader odpsreader ossreader postgresqlreader starrocksreader tsdbreader datahubreader gdbreader hbase11xsqlreader kingbaseesreader mysqlreader opentsdbreader otsreader rdbmsreader streamreader txtfilereader

#### 支持的writer:

adbpgwriter databendwriter elasticsearchwriter hbase11xsqlwriter hologresjdbcwriter mongodbwriter ocswriter osswriter selectdbwriter tdenginewriter adswriter datahubwriter ftpwriter hbase11xwriter kingbaseeswriter mysqlwriter odpswriter otswriter sqlserverwriter tsdbwriter cassandrawriter doriswriter gdbwriter hbase20xsqlwriter kuduwriter neo4jwriter oraclewriter postgresqlwriter starrockswriter txtfilewriter clickhousewriter drdswriter hbase094xwriter hdfswriter loghubwriter oceanbasev10writer oscarwriter rdbmswriter streamwriter

#### 以MySQL迁移SQL Server为例,运行如下命令生成模板,红色字典区域为json数据模 板:

python3 bin/datax.py -r mysqlreader -w sqlserverwriter

#### 图 3-22 生成模板



步骤2 模板字段简介

- 1. job
  - 配置文件的主要对象,包含整个数据迁移任务的配置信息。
- 2. setting
  - 位于job对象下,用于设置数据迁移任务的全局参数,如并发通道数
     (channel),该参数决定了DataX在执行数据迁移时使用的并行线程数。
- 3. content
  - 位于job对象下,是一个数组,包含多个数据迁移任务的具体配置。每个任务都包含reader和writer两个部分。
- 4. reader

- 位于content数组的每个对象中,用于配置数据源读取器。
  - name:读取器的名称,如mysqlreader表示MySQL数据源读取器。
    - parameter:读取器的参数配置,包含数据源连接信息、要读取的字段 等。
      - username:数据源的用户名。
      - password:数据源的密码。
      - column:要读取的字段列表。
      - connection:数据源连接信息,包括JDBC URL和要读取的表名等。
- 5. writer

- 位于content数组的每个对象中,用于配置目标数据库写入器。
  - name:写入器的名称,如sqlserverwriter表示SQL Server目标数据库写入器。
  - parameter:写入器的参数配置,包含目标数据库连接信息、要写入的 字段等。
    - username: 目标数据库的用户名。
    - password:目标数据库的密码。
    - column:要写入的字段列表,应与数据源读取器中的字段列表对应。
    - postSql:写入数据到目的表后,会执行这里的标准语句。
    - preSql:写入数据到目的表前,会先执行这里的标准语句,例如: 清空表数据。
    - connection:目标数据库连接信息,包括JDBC URL和要写入的表名等。

**图 3-23** 模板示例:





----结束

# 数据库 MySQL 迁移 MySQL 操作验证

#### 须知

- 源端数据库服务器和目标数据库服务器,分别需对DataX服务器开放对应数据库 3306访问端口,请参考修改安全组规则,配置IP地址白名单,以便DataX服务可以 连接源端和目标端数据库。
- 目标端数据库如果和DataX数据库在同一子网下,模板中目标端数据库可配置私网。
- 注意:DataX迁移数据库时,reader、writer插件要匹配数据库的版本,DataX服务 对应数据库的reader插件路径在DataX工作目录下的plugin/reader/{reader}/libs 下,writer插件路径在DataX工作目录下的plugin/writer/{writer}/libs下。
- 本次以MySQL-8.0.39版本示例测试数据库迁移同步。
- 步骤1 迁移前的源端和目标端的数据库、表信息展示。

图 3-24 源端数据库、表数据

	_							
mysql> use mysql_origin; Database changed								
<pre>mysql&gt; select * from mysql_origin_table;</pre>								
id   name   score								
1   Michael   60.00     2   David   80.00     3   Emma   67.00     4   Amelia   93.00     5   James   83.00								
++ 5 rows in set (0.00 sec)								
mysql>								

图 3-25 目标端数据库、表数据

```
mysql> use mysql_target;
Database changed
mysql> select * from mysql_target_table_2;
Empty set (0.00 sec)
mysql>
```

**步骤2** 迁移Json模板配置(读取源端数据库表中的id, name部分字段信息,迁移到目的端数 据库)

图 3-26 Json 迁移模板



**步骤3** 输入命令: cd /opt/datax进入该目录,执行python3 bin/datax.py mysql2mysql.json,查看服务迁移结果。

### **图 3-27** 查看执行结果

id,name	9.422 []OD-0] INFO	UriginalContPretreat	mentutii - table:[mysq	<pre>L_target_table_2] all</pre>	columns:[			
J. 2024-10-12 10:10:2 INSERT INTO %s (id	9.456 [job-0] INFO	OriginalConfPretreat	mentUtil - Write data					
<ol> <li>which idbcUrl l</li> </ol>	ike:[idbc:mysal://	1):3306/mvs	dl target?vearIsDateTv	pe=false&zeroDateTime	Behavior=convertToNul	l&rewriteBatchedStater	ments=true&tinvInt1isBit=false]	i i i i i i i i i i i i i i i i i i i
2024-10-12 10:10:2	9.457 [job-0] INFO	JobContainer - jobCon	ntainer starts to do p	repare				i de la companya de l
2024-10-12 10:10:2	9.457 [job-0] INFO	JobContainer - DataX	Reader.Job [mysqlread	er] do prepare work .				i de la companya de l
2024-10-12 10:10:2	9.457 [job-0] INFO	JobContainer - DataX	Writer.Job [mysqlwrit	riter.Job [mysqlwriter] do prepare work .				
2024-10-12 10:10:2	9.457 [job-0] INFO	JobContainer - jobCon	ntainer starts to do s	plit				i de la companya de l
2024-10-12 10:10:2	9.458 [job-0] INFO	JobContainer - Job se	et Channel-Number to 1	channels.				i de la companya de l
2024-10-12 10:10:2	9.460 []00-0] INFO	JobContainer - Datax	Weiter Job [mysqlread	erj splits to [1] tas	KS.			i de la companya de l
2024-10-12 10:10:2	9 485 [job-0] INFO	JobContainer - jobCon	ntainer starts to do s	chedule	~>·			i de la companya de l
2024-10-12 10:10:2	9.487 [job-0] INFO	JobContainer - Sched	uler starts [1] taskGr	oups.				i de la companya de l
2024-10-12 10:10:2	9.490 [iob-0] INFO	JobContainer - Runnig	ng by standalone Mode.					i de la companya de l
2024-10-12 10:10:2	9.494 [taskGroup-0	] INFO TaskGroupConta	iner - taskGroupId=[0]	start [1] channels f	or [1] tasks.			i de la companya de l
2024-10-12 10:10:2	9.496 [taskGroup-0	] INFO Channel - Chann	nel set byte_speed_lim	it to -1, No bps acti	vated.			i de la companya de l
2024-10-12 10:10:2	9.496 [taskGroup-0	] INFO Channel - Chann	nel set record_speed_l	imit to -1, No tps ac	tivated.			i de la companya de l
2024-10-12 10:10:2	9.520 [taskGroup-0	] INFO TaskGroupConta	iner - taskGroup[0] ta	skId[0] attemptCount[	1] is started			i de la companya de l
2024-10-12 10:10:2	9.522 [0-0-0-reade	r] INFO CommonRdbmsRea	ader\$Task - Begin to r	ead record by Sql: [s	elect id name from my	sql_origin_table	101 101 101 101 10	i de la companya de l
] ]dbcUrt:[]dbc:my	SQL://	:3306/mysql_origin/yea	risbaterype=raise&zero	paterimeBenavior=conv	ertioNuclatinyIntiisB	and actinin toble	edstatements=true].	i de la companya de l
1_idbcUr1+[idbc+my	a.399 [0-0-0-reade	-3306/mysel origin2yeau	rTeDateTune-false&zero	DateTimeRebayior-conv	ectToNull&tiovIntlieR	it_falee&rewriteBatch	adStatements-true]	i de la companya de l
2024-10-12 10:10:2	9.921 [taskGroup-0	1 INFO TaskGroupConta	iner - taskGroup[0] ta	skId[0] is successed.	used[415]ms			i de la companya de l
2024-10-12 10:10:2	9.921 [taskGroup-0	INFO TaskGroupConta	iner - taskGroup[0] co	mpleted it's tasks.				i de la companya de l
2024-10-12 10:10:3	9.516 [job-0] INFO	StandAloneJobContain	erCommunicator - Total	5 records, 32 bytes	Speed 3B/s, 0 recor	ds/s   Error 0 records	s, 0 bytes   All Task WaitWrite	erTime 0.000s   All Task Wa
itReaderTime 0.025	s   Percentage 100	.00%						1
2024-10-12 10:10:3	9.517 [job-0] INFO	AbstractScheduler - S	Scheduler accomplished	all tasks.				i de la companya de l
2024-10-12 10:10:3	9.517 [job-0] INFO	JobContainer - DataX	Writer.Job [mysqlwrit	er] do post work.				i de la companya de l
2024-10-12 10:10:3	9.517 []0D-0] INFO	JobContainer - DataX	Reader.Job [mysqlread	erj do post work.				i de la companya de l
2024-10-12 10:10:3	9.517 [job-0] INFO	HookTowakar No book	jobid [0] completed s	a dir not aviete or i	s a file: (ent/datax/	hook		i de la companya de l
2024-10-12 10:10:3	9 528 [job-0] INFO	JobContaiper -	k divoked, because bas		s a rece. Jope/dataxy	HOOK		i de la companya de l
ftotal cp	u infol =>							i de la companya de l
av	erageCou	maxDeltaC	pu	minDeltaCpu				1
- 1	.00%	-1.00%		-1.00%				i i i i i i i i i i i i i i i i i i i
								i i i i i i i i i i i i i i i i i i i
								i i i i i i i i i i i i i i i i i i i
[total gc	unfo] =>						a construction of the second	i i i i i i i i i i i i i i i i i i i
N	AME 1 Verne Concertion	totalGCCount	maxDeltaGCCount	minueltaccount	totalGCI1me	maxDeltaGCTime	MinDeltaGlime	i i i i i i i i i i i i i i i i i i i
9	1 Old Generation				0.0095	0.0095	0.0095	i i i i i i i i i i i i i i i i i i i
G	1 Concurrent 6C	2	2	2	0.0003	0.0003	0.0000	1
								i i i i i i i i i i i i i i i i i i i
2024-10-12 10:10:3	9.520 [job-0] INFO	JobContainer - PerfT	race not enable!					i i i i i i i i i i i i i i i i i i i
2024-10-12 10:10:3	9.520 [job-0] INFO	StandAloneJobContain	erCommunicator - Total	5 records, 32 bytes	Speed 3B/s, 0 recor	ds/s   Error 0 records	s, 0 bytes   All Task WaitWrit	erTime 0.000s   All Task Wa
itReaderTime 0.025	s   Percentage 100	.00%						
2024-10-12 10:10:3	9.521 [job-0] INFO	JobContainer -						
法发展密时到	: 20	24-10-12 10:10:28						
住在台上拆时	20	24-10-12 10:10:39						i i i i i i i i i i i i i i i i i i i
住名平均语言		38/s						
记录写入读度		Orec/s						
读出记录总数		5						
读写失败总数								

图 3-28 查看目的端数据库写入情况



----结束

# 数据库 MySQL 迁移 Oracle 操作验证

#### 须知

- 源端数据库服务器需对DataX服务器开放对应数据库3306访问端口,目标端数据库 服务器需对DataX服务器开放对应数据库1521访问端口,请参考修改安全组规则, 配置IP地址白名单,以便DataX服务可以连接源端和目标端数据库。
- 目标端数据库如果和DataX数据库在同一子网下,模板中目标端数据库可配置私网。
- 注意:DataX迁移数据库时,reader、writer插件要匹配数据库的版本,DataX服务 对应数据库的reader插件路径在DataX工作目录下的plugin/reader/{reader}}/libs 下,writer插件路径在DataX工作目录下的plugin/writer/{writer}/libs下。
- 本次以MySQL-8.0.39, Oracle-19c版本示例测试异构数据库迁移同步。

步骤1 迁移前的源端和目标端的数据库、表信息展示。

<b>图</b> 3-29 源端数据库、表数据
mysql> use mysql_origin; Database changed mysql> select * from mysql_origin_table; +++
id   name   score   ++
1   Michael   60.00     2   David   80.00     3   Emma   67.00     4   Amelia   93.00     5   James   83.00
5 rows in set (0.00 sec) mysql> 📕

图 3-30 目标端数据库、表数据



**步骤2** 迁移Json模板配置(读取源端数据库表中的id, name部分字段信息,迁移到目的端 Oracle数据库 ) 图 3-31 Json 迁移模板



**步骤3** 输入命令: cd /opt/datax进入该目录,执行python3 bin/datax.py mysql2oracle.json,查看服务迁移结果。



**图 3-32** 查看执行结果

#### 图 3-33 查看目的端数据库写入情况

SQL> conn oracle_target/123456; Connected. SQL> select * from oracle_target_table;				
no rows selected				
SQL> select * from oracle_target_table;				
ID				
NAME				
1 Michael				
2 David				
3 Emma				
ID NAME				
4 Amelia				
5 James				

----结束

可视化编辑器

C

# 3.4 快速卸载

须知 • 如选择"新建VPC"一键部署该方案,目标数据库服务器和DataX服务器在同一VPC 下,须对目标数据库服务器<mark>切换其他VPC</mark>后,即可按照下述步骤进行卸载。 步骤1 登录资源编排服务 RFS,进入"资源栈",选择创建的资源栈名称,单击"删除"。 图 3-34 一键卸载 资源编排服务 🔤 资源栈 ③ | 资源线名称 v | creating-a-database-migrati nvironment-with-datax × | Q 🛞 C 资源栈 更新时间 ⊖ 操作 资源线集 横板车 2024/10/10 10:22:29 GMT+08:00 創珍 更新

步骤2 在弹出的删除资源栈确认框中,输入"Delete",单击"确定",即可卸载解决方案。

 $\times$ 

图 3-35 删除资源栈确认

|--|

您确定要删除该资源栈及资源栈内资源吗?	删除后不能恢复,	请谨慎操作

资源栈名称		状态	创建时间		
creating-a-database-m	igration-en	部署成功	2024/10/10	) 10:19:19 GMT	+08:00
资源列表 (8)					
云产品名称	物理资源名称	'ID		资源状态	
弹性云服务器	creating-a-dat	abase-migration	environmen	生成完成	<b>^</b>
虚拟私有云	creating-a-dat	abase-migration	environmen	生成完成	
虚拟私有云	3c8288e9-868	31-4c63-a3d1-cb	d23116d5e8	生成完成	
虚拟私有云	ca72f63b-0fff-	4a9f-a52e-6d3a	8ae232e3	生成完成	
虚拟私有云	78c98248-a90	)e-4fb5-a3c3-05	74eb355118	生成完成	
虚拟私有云	creating-a-dat	abase-migration	-environmen	生成完成	
		1			*
如您确认要删除资源栈或其资源,请输入Delete以确认删除					
Delete					
				确定	取消

----结束



# 名词解释

基本概念、云服务简介、专有名词解释

- 华为云Flexus云服务器X实例: Flexus云服务器X实例是新一代面向中小企业和开 发者打造的柔性算力云服务器。Flexus云服务器X实例功能接近ECS, 同时还具备 独有特点,例如Flexus云服务器X实例具有更灵活的vCPU内存配比、支持热变配 不中断业务变更规格、支持性能模式等。
- 弹性公网IP EIP:提供独立的公网IP资源,包括公网IP地址与公网出口带宽服务。 可以与弹性云服务器、裸金属服务器、虚拟IP、弹性负载均衡、NAT网关等资源 灵活地绑定及解绑。



#### **表 5-1** 修订记录

发布日期	修订记录
2024-10-17	第一次正式发布。