

迁移中心 MgC

最佳实践

文档版本 12
发布日期 2024-12-02



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



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

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 主机迁移权限配置	1
2 主机迁移相关最佳实践	7
2.1 主机迁移网络要求和配置	7
2.2 线下主机迁移上云	8
2.3 阿里云主机迁移至华为云	13
2.3.1 概述	13
2.3.2 准备工作	14
2.3.3 步骤一：下载并安装 Edge	15
2.3.4 步骤二：采集阿里云 ECS 主机资源	15
2.3.5 步骤三：评估迁移准备度	17
2.3.6 步骤四：目的端规格评估	19
2.3.7 步骤五：创建主机迁移 workflow	22
2.4 主机一站式跨可用区迁移	22
2.5 华为云主机跨可用区迁移	24
2.6 将主机迁移至 Flexus 云服务器 X 实例（原云耀云服务器 X 实例）	27
2.7 迁移后主机私有 IP 保持不变	30
2.8 批量修改与还原源端 Linux 主机的 Host 配置	34
2.8.1 概述	34
2.8.2 准备工作	35
2.8.3 配置脚本	36
2.8.3.1 配置 update_hosts_linux.sh 脚本	36
2.8.3.2 配置 rollback_hosts_linux.sh 脚本	41
2.9 批量修改与还原源端 Windows 主机的 Host 配置	44
2.9.1 概述	45
2.9.2 准备工作	46
2.9.3 脚本示例	47
2.9.3.1 配置 update_hosts_win.ps1 脚本	47
2.9.3.2 配置 rollback_hosts_win.ps1 脚本	52
2.9.4 常见问题	56
2.9.4.1 如何启用 PowerShell 远程管理模块	56
2.9.4.2 如何启用 WinRM 服务	56
2.9.4.3 执行脚本后报错：连接到远程服务器失败，且登录凭证信息正确	56
3 存储迁移相关最佳实践	58

3.1 迁移其他云平台存储数据至华为云.....	58
3.2 OBS 多前缀批量迁移.....	74
3.3 归档存储（冷存储）数据迁移.....	85
3.4 SFS1.0 迁移至 SFS 3.0.....	92
3.4.1 使用须知.....	92
3.4.2 准备工作.....	93
3.4.3 创建集群.....	93
3.4.4（可选）网络打通.....	95
3.4.5 创建存储迁移 workflow.....	97
3.4.6（可选）清理集群.....	101
3.5 NAS 到 NAS 迁移和割接方案.....	102
3.6 文件存储批量迁移.....	104
3.7 通过 HTTP 迁移 MinIO 存储数据到华为云 OBS.....	105
3.8 通过 HTTP 迁移 Ceph 存储数据到华为云 OBS.....	112
4 目的端主机磁盘扩容.....	120
5 调整目的端磁盘分区.....	124
6 采集 Azure 容器 AKS 资源.....	127
7 采集谷歌云 GCP 容器 GKE 资源.....	129
8 采集 AWS 容器资源.....	131
9 大数据迁移后数据校验.....	135
9.1 MaxCompute 迁移至 DLI 数据校验.....	135
9.2 MRS ClickHouse 迁移至 MRS ClickHouse 数据校验.....	141
9.3 阿里云 EMR ClickHouse 迁移至 MRS ClickHouse 数据校验.....	148
9.4 阿里云 ClickHouse 迁移至 MRS ClickHouse 数据校验.....	154
9.5 阿里云 ClickHouse 迁移至 CloudTable ClickHouse 数据校验.....	161
9.6 MRS Doris 迁移至 MRS Doris 数据校验.....	166
9.7 MRS Hive/CDH/EMR 迁移至 MRS Hive 数据校验.....	172
9.8 MaxCompute 迁移至 MRS Hive 数据校验.....	178
9.9 MRS HBase 迁移至 MRS HBase 数据校验.....	184
9.10 DeltaLake（有元数据）迁移至 MRS DeltaLake 数据校验.....	190
9.11 DeltaLake（无元数据）迁移至 MRS DeltaLake 数据校验.....	196
10 不使用 EIP 场景下进行大数据数据迁移.....	202

1 主机迁移权限配置

概述

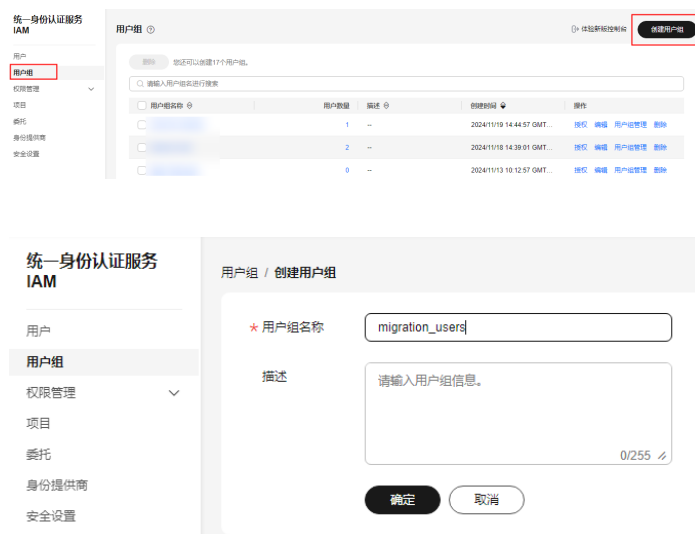
1. 创建一个新的用户组migration_users，赋予MgC和SMS所需要的最小权限，后续使用的AK/SK权限以此为准。
2. 用户通过属于admin组的用户，创建属于migration_users组的、仅支持编程访问的用户mgc-user（同时勾选访问密钥，在创建用户后下载访问密钥），该用户无法通过华为云console密码登录，仅支持编程访问。
3. 后续提供创建mgc-user用户时下载的AK/SK给Edge，该AK/SK负责Edge注册到MgC以及迁移过程中的接口调用鉴权。

步骤一：创建用户组

步骤1 登录华为云IAM管理控制台。

步骤2 在统一身份认证服务，左侧导航栏，选择“用户组”，单击右上方的“创建用户组”。

图 1-1 创建用户组

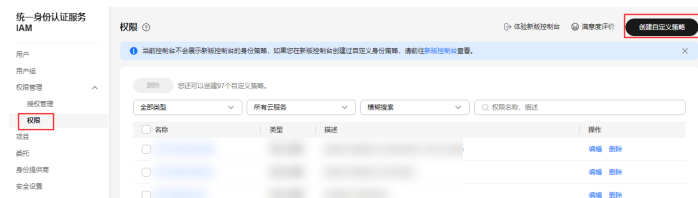


----结束

步骤二：创建权限策略

步骤1 在统一身份认证服务，左侧导航栏，选择“权限管理>权限”，单击右上方的“创建自定义策略”。

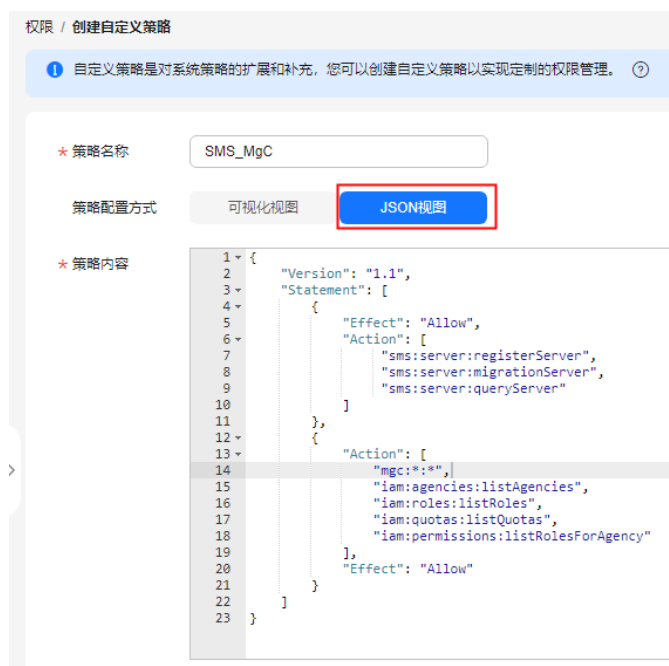
图 1-2 创建自定义策略



步骤2 创建SMS-全局云服务策略。输入“策略名称”，策略配置方式选择“JSON视图”，将以下内容复制粘贴到策略内容中。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "sms:server:registerServer",
        "sms:server:migrationServer",
        "sms:server:queryServer"
      ]
    },
    {
      "Action": [
        "mgc:*:*",
        "iam:agencies:listAgencies",
        "iam:roles:listRoles",
        "iam:quotas:listQuotas",
        "iam:permissions:listRolesForAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

图 1-3 创建 SMS-全局云服务策略



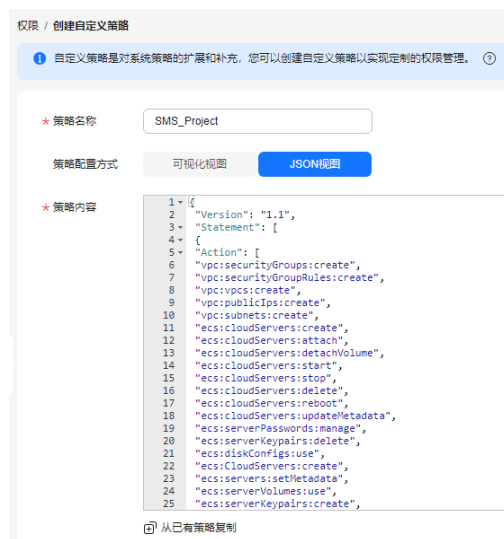
步骤3 创建SMS-项目级云服务策略。输入自定义“策略名称”，策略配置方式选择“JSON视图”，将以下内容复制粘贴到策略内容中。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "vpc:securityGroups:create",
        "vpc:securityGroupRules:create",
        "vpc:vpcs:create",
        "vpc:publicIps:create",
        "vpc:subnets:create",
        "ecs:cloudServers:create",
        "ecs:cloudServers:attach",
        "ecs:cloudServers:detachVolume",
        "ecs:cloudServers:start",
        "ecs:cloudServers:stop",
        "ecs:cloudServers:delete",
        "ecs:cloudServers:reboot",
        "ecs:cloudServers:updateMetadata",
        "ecs:serverPasswords:manage",
        "ecs:serverKeypairs:delete",
        "ecs:diskConfigs:use",
        "ecs:CloudServers:create",
        "ecs:servers:setMetadata",
        "ecs:serverVolumes:use",
        "ecs:serverKeypairs:create",
        "ecs:serverInterfaces:use",
        "ecs:serverGroups:manage",
        "ecs:securityGroups:use",
        "ecs:servers:unlock",
        "ecs:servers:rebuild",
        "ecs:servers:lock",
        "ecs:servers:reboot",
        "evs:volumes:use",
        "evs:volumes:create",
        "evs:volumes:update",
        "evs:volumes:delete",
        "evs:volumes:attach",
        "evs:volumes:detach",
```



```
"evs:snapshots:create",  
"evs:snapshots:delete",  
"evs:snapshots:rollback",  
"ecs:*:get*",  
"ecs:*:list*",  
"evs:*:get*",  
"evs:*:list*",  
"vpc:*:list*",  
"vpc:*:get*",  
"ims:*:get*",  
"ims:*:list*",  
],  
"Effect": "Allow"  
}  
]  
}
```

图 1-4 SMS 项目级资料



----结束

步骤三：授权

步骤1 在统一身份认证服务，左侧导航栏，选择“用户组”。

步骤2 在用户组列表中，单击**步骤一**创建的用户组操作列的“授权”。

图 1-5 为用户组授权



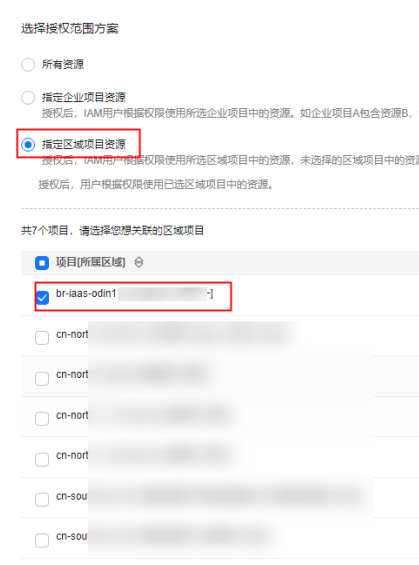
步骤3 搜索并勾选**步骤二**创建的两个自定义策略，单击“下一步”。

图 1-6 选择自定义策略



步骤4 授权范围方案选择“指定区域项目资源”，选择指定区域项目，用户可以根据权限使用该区域项目中的资源。

图 1-7 选择指定区域项目



步骤5 单击“确定”，完成用户组授权。

----结束

步骤四：创建用户

步骤1 在统一身份认证服务，左侧导航栏，选择“用户”，单击右上方的“创建用户”。

图 1-8 创建用户



步骤2 输入自定义用户名，去勾选“管理控制台访问”，单击“下一步”。

图 1-9 配置基本信息



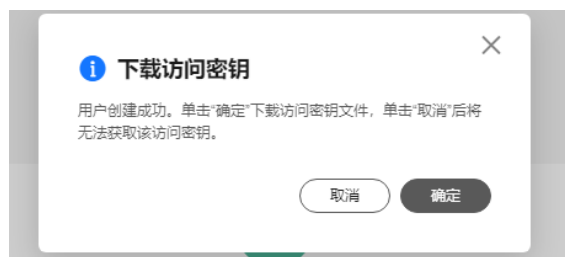
步骤3 勾选步骤一创建的用户组，单击“创建用户”。

图 1-10 选择用户组



步骤4 创建用户成功，弹出的下载访问密钥窗口，单击“确定”，将AK/SK下载到本地。

图 1-11 下载访问密钥



----结束

2 主机迁移相关最佳实践

2.1 主机迁移网络要求和配置

背景说明

主机迁移过程中主要涉及的流量分为两部分：控制流和数据流。

- **控制流**：控制流是指源端主机与云服务管理面之间的通信。为确保迁移过程的顺利进行，需要检查源端主机出方向是否有限制，并确保源端主机能够访问迁移过程中依赖的云服务，包括SMS、IAM、ECS、EVS、VPC、IMS等服务的域名。
- **数据流**：数据流涉及从源端到目的端的数据传输。为确保数据流的畅通，需要检查目的端服务器的安全组配置，确保目的端安全组对源端IP开放了指定的迁移端口。如果源端主机无法直接访问公网或与目的端直接通信，需要配置代理服务器。在这种情况下，需要额外考虑以下因素：
 - 代理服务器的配置，确保它能够正确转发源端主机的流量。
 - 代理服务器的安全组设置，确保对源端IP开放了代理端口。

MgC 通用主机迁移网络场景

MgC当前不支持完全无公网的迁移场景。

1. 在与源端互通的内网环境中安装迁移工具Edge。安装方法和要求请参考[安装Windows版本的Edge](#)。
2. 参考MgC的[线下主机迁移上云](#)最佳实践，完成步骤二至步骤五的操作。

注意

需要确保目的端安全组对源端服务器开放指定迁移端口。安全组的配置方法请参考[如何配置目的端服务器安全组规则?](#)

3. **创建主机迁移 workflow**。
 - 如果源端可以访问公网，在迁移网络配置选择公网迁移。

图 2-1 公网迁移



- 如果源端无法访问公网，您需要准备一台可以访问公网的代理服务器并安装代理软件，方法可参考[配置代理服务器访问华为云](#)中的步骤1。对该代理服务器的网络要求如下：
 - 该代理机器可以在源端内网，也可以在云上，只需和源端待迁移服务器内网互通即可。
 - 需要确保代理服务器入方向代理端口对源端IP开放。

配置完代理服务器后，在迁移网络配置选择私网迁移，输入代理服务器的私网IP和代理软件指定的代理端口。

图 2-2 私网迁移



源端无法访问公网，源端和目的端不能直接互通

该场景MgC暂不支持，可以使用SMS服务进行迁移。通过一台中转代理服务器同时转发管理面和数据面流量。详细介绍请参考[SMS常见网络场景下的网络要求和配置方法](#)。

2.2 线下主机迁移上云

适用场景

本文介绍如何使用迁移中心MgC将线下主机迁移到华为云。

准备工作

- 建议在源端内网环境中准备一台用于安装Edge的Windows主机，并确保该Windows主机满足以下要求：
 - 可以[连接外网](#)（公网），并检查是否能够访问MgC和IoTDA服务的域名，具体需要检查的域名请查看[域名列表](#)。
 - PowerShell 版本在**3.0**及以上。
 - 推荐规格不小于**4U8G**。
 - 安全组的出方向规则中，需要包含8883端口。
 - 关闭该主机上的杀毒、防护类软件，该类软件会拦截Edge的执行，导致迁移 workflow 执行失败。

注意

不建议将Edge安装在源端待迁移的主机上，主要原因包括：

- **资源消耗**：Edge在采集和迁移运行过程中会消耗CPU和内存资源。在迁移任务量大时，这可能会对源端的业务运行造成影响。
- **端口占用**：Edge会占用源端主机上的一些端口，也可能对源端业务产生影响。

- 源端主机需要对安装Edge的Windows主机开放指定端口：
 - Windows开放5985端口。
 - Linux开放22端口。
- 如果源端是Windows主机，需要开启WinRM远程访问，并与安装Edge的主机建立连接，方法请参考[Windows服务器WinRM服务配置与故障排查](#)。
- 使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，获取方法请参见[准备工作](#)。
- 在MgC控制台[创建项目](#)。

使用须知

在创建主机迁移 workflow 前，请仔细阅读并了解以下使用须知内容。

项目	须知
源端下载带宽	用于源端服务器下载迁移Agent。 <ul style="list-style-type: none"> ● 推荐带宽：每台机器不低于30 Mbits/s。 ● 共用带宽建议：若多台机器共用，平均不低于50 Mbit/s。
迁移带宽	<ul style="list-style-type: none"> ● 重要性：影响数据传输速度和迁移时间。 ● 推荐值与预估时长：详情请参见迁移需要多长时间?
CPU和内存要求	<ul style="list-style-type: none"> ● 内存：至少预留520 MB。 ● CPU：Linux系统预留不少于0.3核，Windows系统不低于1核。
系统兼容性列表	支持迁移的操作系统列表，请参见 兼容性列表 。
主机迁移重要声明	了解迁移过程中的重要信息和免责声明，详情请参见 主机迁移服务重要声明有哪些?
约束与限制	确保迁移过程符合服务条款和限制条件，详情请参见 主机迁移约束与限制 。
计费说明	了解迁移过程中可能产生的费用，详情请参见 计费说明 。
权限配置	使用IAM进行精细的权限管理，详情请参见 权限配置 。
网络配置及迁移端口	确保源端和目的端的网络连接和端口开放满足迁移要求，详情请参见 网络配置及端口开放 。

步骤一：下载并安装 Edge

参考[下载并安装Edge](#)，安装Edge工具并与迁移中心连接。

步骤二：添加主机

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 在[资源列表](#)单击主机或主机的资源数量，进入[资源列表](#)页面的[主机](#)页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

步骤4 单击列表上方的“[添加资源](#)”按钮，弹出添加主机窗口。

图 2-3 添加主机



步骤5 根据[表2-1](#)，填写相应参数后，单击“[确认](#)”，系统会自动检查凭证绑定状态和进行信息采集。

表 2-1 添加主机参数说明

参数	说明
名称	用户自定义
采集设备	选择源端注册的Edge设备。
类型	主机实际系统类型
IP	填写主机IP。 如果采集的主机与安装Edge的主机处于同一VPC内，可以填写私有IP，否则需要填写公网IP。
端口	主机开放端口。 <ul style="list-style-type: none"> Windows主机默认为5985端口，无法修改。 Linux主机默认22端口，可根据实际情况修改。

参数	说明
凭证	选择源端主机凭证，如未提前在Edge上 添加源端主机凭证 ，请前往Edge控制台添加，并同步至迁移中心。

步骤6 添加成功后，在**资源列表**页面的**主机**页签，可以查看添加的主机资源。

----结束

步骤三：关联应用

将待迁移主机关联应用，在后续的规格评估和创建迁移 workflow 时，选择该应用，即可对应用内的所有主机进行迁移评估推荐和创建迁移任务。

步骤1 在**资源列表**页面的**主机**页签，勾选需要关联到同一应用的主机，单击页面左上角的“**关联到应用**”，弹出**关联到应用**窗口。

步骤2 在下拉列表中选择需要关联的应用。未提前[创建应用](#)，单击“**新建应用**”，输入自定义的**应用名称**和**描述**；根据实际需求，选择**业务场景**和**使用场景**；目的端区域选择目的端所在区域，单击“**确定**”按钮，应用创建成功。

步骤3 选择应用后，单击“**确定**”。在主机的**已关联应用**列，查看资源所关联应用。

----结束

(可选) 步骤四：绑定已有目的端

如果您已经在华为云创建了目的端主机，可以将源端主机与目的端主机进行绑定。绑定后，迁移 workflow 将使用所绑定的目的端作为迁移的目的端。绑定已有目的端的主机，无需进行评估推荐，请直接[创建主机迁移 workflow](#)。

如果您不进行目的端绑定，请执行**步骤五：规格评估**操作。

注意

绑定的目的端主机有以下要求：

- 所绑定的目的端主机能够被格式化。迁移 workflow 将使用所绑定的目的端作为迁移的目的端，迁移过程中该目的端主机会被格式化，然后将源端数据迁移到该主机。
- 如果后续主机迁移 workflow 是公网迁移，所绑定目的端主机需要能访问公网。
- 选择绑定的目的端主机所在区域需要与源端主机[关联的应用](#)处于同一区域。

步骤1 单击左侧导航栏的“**设计>迁移方案设计**”，进入**迁移方案设计**页面。

步骤2 单击目的端规格配置框中的“**目的端配置表**”，进入**目的端配置表**页面。

步骤3 在**资源配置明细**的**主机**页签，单击主机资源**绑定目的端**列的“**去绑定**”，弹出**绑定目的端**窗口。

步骤4 选择源端主机[关联的应用](#)所在区域以及区域下的项目，在所选项目下，选择要绑定的目的端主机，单击“**确认**”。

目的端绑定成功后，该主机资源绑定目的端列的状态会变为“已绑定”，单击“详情”可以查看绑定的目的端规格。

---结束

步骤五：规格评估

评估推荐是根据源端资源规格、性能和应用场景以及特定的成本、可用性、性能、安全合规等需求，为您推荐最合适的华为云对应资源规格，通过规格评估确定迁移目的端。

📖 说明

已绑定目的端的主机资源，无需进行规格推荐，可直接创建迁移工作流。

- 步骤1** 在迁移方案设计页面，单击目的端规格配置框中的“规格评估”按钮，进入评估单应用页面。
- 步骤2** 在选择应用的下拉列表中，选择源端主机[关联的应用](#)。
- 步骤3** 在选择资源区域，勾选应用内，需要评估的主机。
- 步骤4** 根据[表2-2](#)，配置评估策略。

表 2-2 评估策略参数说明

参数	子参数	说明
购买区域	-	选择在华为云购买资源的区域。推荐您选择靠近目标用户的区域，可以减少网络时延，提高访问速度。
推荐策略	匹配源端规格	根据源端规格推荐最合适的华为云对应资源规格。详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。
	匹配业务场景	以源端业务场景为基础，结合华为云最佳实践进行资源规格推荐。详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。
优先级设置	性能优先	推荐的目的端规格要求性能最佳。
	价格优先	推荐的目的端规格在满足要求的前提下尽量节约成本。
偏好设置	虚拟机类型偏好（可选）	选择期望的目的端主机类型，支持多选。
	虚拟机规格族偏好（可选）	根据选择的主机类型，罗列出包含的规格族类型，选择期望的目的端主机规格族，支持多选。 须知 勾选“仅显示适合专属主机规格族”，则只能选择在购买区域所选Region下专属主机所包含的规格，并且虚拟机类型偏好将无法选择。
	系统盘规格偏好（可选）	选择期望的目的端系统盘规格类型。

参数	子参数	说明
	数据盘规格偏好 (可选)	选择期望的目的端数据盘规格类型。
	虚拟机规格调整	选择主机性能规格的推荐策略。 详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。

步骤5 配置完成后，单击“确认”，创建评估成功。

步骤6 在迁移方案设计页面的应用列表，单击应用操作列的“查看资源配置”，进入目的端配置表页面。

在资源配置明细区域，可以查看应用内主机资源的评估推荐结果，其中目的端规格配置是根据源端规格配置和评估策略推荐的华为云对应规格配置，成本预估是华为云推荐规格所需费用；单击主机资源操作列的“更多>更改目的端规格”，可以修改目的端主机规格、磁盘存储类型和资源总用量。

----结束

步骤六：创建主机迁移 workflow

完成已上[步骤一~步骤五](#)的操作后，即可[创建主机迁移 workflow](#)，将线下主机迁移上云。

2.3 阿里云主机迁移至华为云

2.3.1 概述

本最佳实践主要介绍从阿里云ECS迁移至华为云ECS的详细步骤和注意事项。

本指导包含以下关键步骤：

- 准备工作：** 确保具备必要的账号以及账号满足权限要求。
- 安装Edge：** 下载并安装MgC服务所需的云边协同组件Edge。
- 资源采集：** 采集源端阿里云ECS主机的信息。
- 迁移准备：** 对源端阿里云ECS主机进行配置检查、网络测试和迁移分组。
- 规格评估：** 评估并配置华为云目的端主机的规格。
- workflow 创建：** 创建并配置迁移 workflow 并启动迁移。

使用须知

在创建主机迁移 workflow 前，请仔细阅读并了解以下使用须知内容。

项目	须知
源端下载带宽	用于源端服务器下载迁移Agent。 <ul style="list-style-type: none"> 推荐带宽：每台机器不低于30 Mbits/s。 共用带宽建议：若多台机器共用，平均不低于50 Mbit/s。
迁移带宽	<ul style="list-style-type: none"> 重要性：影响数据传输速度和迁移时间。 推荐值与预估时长：详情请参见迁移需要多长时间?
CPU和内存要求	<ul style="list-style-type: none"> 内存：至少预留520 MB。 CPU：Linux系统预留不少于0.3核，Windows系统不低于1核。
系统兼容性列表	支持迁移的操作系统列表，请参见 兼容性列表 。
主机迁移重要声明	了解迁移过程中的重要信息和免责声明，详情请参见 主机迁移服务重要声明有哪些?
约束与限制	确保迁移过程符合服务条款和限制条件，详情请参见 主机迁移约束与限制 。
计费说明	了解迁移过程中可能产生的费用，详情请参见 计费说明 。
权限配置	使用IAM进行精细的权限管理，详情请参见 权限配置 。
网络配置及迁移端口	确保源端和目的端的网络连接和端口开放满足迁移要求，详情请参见 网络配置及端口开放 。

2.3.2 准备工作

为了确保迁移过程的顺利进行，需要完成以下必要的准备工作。

华为账号准备

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。

阿里云访问密钥（AK/SK）准备

检查阿里云迁移账号是否拥有AK/SK以及只读访问云服务器ECS的权限（**AliyunECSReadOnlyAccess**），如果没有，请参考如下步骤生成AK/SK并添加权限。

1. 登录阿里云RAM控制台。
2. 在左侧导航栏，选择**身份管理 > 用户**。
3. 在**用户**页面，单击**用户名称**，进入用户详情页面。
4. 在**用户 AccessKey**区域，单击“**创建 AccessKey**”，生成AccessKey ID和AccessKey Secret。
5. 添加权限。
在**用户**页面，单击RAM用户**操作**列的**添加权限**，授予RAM用户**只读访问云服务器ECS的权限（AliyunECSReadOnlyAccess）**。



创建迁移项目

在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。

2.3.3 步骤一：下载并安装 Edge

迁移工具Edge是一款专为配合MgC设计的软件，用于源端资源发现与信息采集，同时还提供了 workflow 模块，用于支持云端迁移 workflow 任务的执行。

操作步骤

步骤1 建议在源端内网环境中准备一台Windows主机，用于安装迁移工具Edge，对该主机的要求以及Edge的安装方法请参见[安装Windows版本的Edge](#)。

注意

不建议将Edge安装在源端待迁移的主机上，主要原因包括：

- **资源消耗**：Edge在采集和迁移运行过程中会消耗CPU和内存资源。在迁移任务量大时，这可能会对源端的业务运行造成影响。
- **端口占用**：Edge会占用源端主机上的一些端口，也可能对源端业务产生影响。

步骤2 安装完成并注册用户后，将Edge与迁移中心进行连接，方法请参考[连接迁移中心](#)。

步骤3 连接成功后，需要在Edge上添加源端待迁移主机凭证，凭证的添加方法请参考[添加资源凭证](#)。资源类型请根据实际情况选择“**Windows主机**”或“**Linux主机**”。

----结束

2.3.4 步骤二：采集阿里云 ECS 主机资源

前提条件

- 已完成[准备工作](#)。
- 已在[源端环境中安装Edge](#)，并与迁移中心连接。

操作步骤

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的当前项目下拉列表中选择在完成[准备工作](#)时所创建的迁移项目。

步骤3 单击任务框中的“[发现>公网发现](#)”，进入[创建公网发现任务](#)页面。

图 2-4 公网发现



步骤4 根据[公网发现任务参数说明](#)，配置公网发现任务参数。

表 2-3 公网发现任务参数说明

区域	参数	说明	是否必选
基础配置	任务名称	自定义发现任务名称。	是
	任务说明	输入发现任务说明。	否
任务配置	源端	选择阿里云。	是
	凭证	选择阿里云对应的凭证，如未提前 添加凭证 ，请单击“ 新建凭证 ”，填写相应参数，校验并保存凭证。 须知 新建凭证中的认证方式请选择“AK/SK”，并输入源端阿里云账号的AK/SK。该账号的AK/SK需要拥有“只读访问云服务器ECS的权限（AliyunECSReadOnlyAccess）”。	是
	区域	选择源端主机所在区域，支持多选。	是
资源发现	云平台采集	开启云平台采集，并在 资源类型 下拉列表，勾选“ 主机 ”资源。	是
	关联到应用（可选）	应用是资源分组的概念，将资源关联到应用，后续的评估推荐和迁移 workflows 都将以应用作为整体进行（可以自由添加和删除资源到应用中）。 <ul style="list-style-type: none"> 已提前创建应用，在“应用”的下拉列表中选择要关联的应用。 未提前创建应用，单击“新建应用”，弹出新建应用窗口，输入自定义的应用名称和描述；根据实际需求，选择业务场景和使用场景；区域选择目的端区域，单击“确定”按钮，应用创建成功。 	否

步骤5 单击“确认”，采集阿里云ECS主机任务创建完成，系统自动开始资源采集。

- 在资源采集页面，单击任务总数后的“查看”，进入任务列表页面。可以查看任务状态和任务详情。如果任务状态为“失败”，单击操作列的“查看”，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。
- 在资源采集页面的“资源列表”，单击主机资源类型或数量，进入资源列表页面的主机页签，可以查看采集的主机资源列表和资源详情。

----结束

2.3.5 步骤三：评估迁移准备度

该操作主要用于校验源端待迁移主机的基本配置、网络环境和迁移分组。

前提条件

- 已完成[阿里云ECS主机资源采集](#)。
- 已在Edge[添加源端主机凭证](#)。

操作步骤

步骤1 在资源采集页面的资源列表，单击主机资源类型或数量，进入资源列表页面的主机页签。

资源列表

资源类型	资源数量
主机	138
容器	0
中间件	2
数据库	1
大数据	0
网络	496
存储	116

步骤2 在列表上方选择“目标场景>主机迁移”，此时源端迁移准备度列会出现“配置”按钮，状态为“未就绪”。

< | 资源列表

名称(注册表实例ID)	IP地址	规格(资源类型)	平台	采集设备	凭证	迁移准备度	迁移时间	已采集	源端迁移准备度	操作
...	未就绪	未就绪	配置

步骤3 将鼠标移动到源端迁移准备度列的“未就绪”状态上，可以看到迁移准备度需要完成的相关配置，包括：绑定采集设备、绑定凭证、迁移预检测（自动触发）和关联应用。单击“采集设备”或“凭证”后的“配置”，弹出配置窗口。

< | 资源列表

名称(注册表实例ID)	IP地址	规格(资源类型)	平台	采集设备	凭证	迁移准备度	迁移时间	已采集	源端迁移准备度	操作
...	未就绪	未就绪	配置

目标场景：主机迁移

绑定采集设备、绑定凭证、迁移预检测、关联应用

采集设备	配置
凭证	配置
迁移预检测	配置
关联应用	配置

步骤4 根据表2-4，配置参数。

表 2-4 迁移准备度参数配置说明

参数	配置说明
类型	根据实际的主机操作系统类型选择。
采集设备	选择在源端环境安装的Edge工具。当资源较多时，可以同时勾选多个资源，并通过列表上方的“绑定采集设备”进行批量绑定。
IP	选择主机接入IP，可以为公网IP，也可以是私有IP；迁移预检查通过后，该IP会作为后续主机迁移工作流的资源接入IP。迁移时Edge会通过接入IP访问源端主机。
端口	源端主机对Edge安装主机开放端口。 <ul style="list-style-type: none"> Windows主机默认为5985端口，无法修改。 Linux主机默认22端口，可根据实际情况修改。
凭证	选择在Edge添加的源端主机凭证，如未提前在Edge上添加源端主机凭证，请前往Edge控制台添加，并同步至迁移中心。

步骤5 配置完成后，单击“确认”按钮，系统会校验配置信息的正确性并自动执行“迁移预检查”，包括：校验Edge设备与源端的连通性，校验凭证的有效性，并自动对主机规格基础信息进行补充性采集。检查时长大概需要10秒钟。

步骤6 由于迁移工作流是以应用为基本单位，迁移应用下的所有资源，因此需要将待迁移资源关联到应用，方法请参见[关联到应用](#)。当所有配置和迁移预检测校验通过后，源端迁移准备度状态会变为已就绪，此时代表该主机已满足迁移条件。



注意

为降低迁移风险，建议单个应用下的主机资源不超过30台；如果待迁移主机数量较多，您可以将主机分配到多个应用中。

----结束

2.3.6 步骤四：目的端规格评估

规格评估是根据源端资源规格、性能和应用场景以及特定的成本、可用性、性能、安全合规等需求，为您推荐最合适的华为云对应资源规格，通过规格评估确定迁移目的端。

前提条件

已完成[迁移准备度评估](#)，并将资源关联到应用。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 单击左侧导航栏的“设计>迁移方案设计”，进入迁移方案设计页面。在迁移方案设计页面，可以查看当前项目下所创建的应用列表。

步骤3 在应用列表中，单击应用操作列的“规格评估”，进入评估单应用页面。



步骤4 在选择资源区域，勾选应用内，需要评估的主机资源。

步骤5 根据[表2-5](#)，配置评估策略。

表 2-5 评估策略参数说明

参数	子参数	说明
购买区域	-	选择在华为云购买资源的区域。推荐您选择靠近目标用户的区域，可以减少网络时延，提高访问速度。
推荐策略	匹配源端规格	根据源端规格推荐最合适的华为云对应资源规格。详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。
	匹配业务场景	以源端业务场景为基础，结合华为云最佳实践进行资源规格推荐。详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。
优先级设置	性能优先	推荐的目的端规格要求性能最佳。
	价格优先	推荐的目的端规格在满足要求的前提下尽量节约成本。
偏好设置	虚拟机类型偏好 (可选)	选择期望的目的端主机类型，支持多选。

参数	子参数	说明
	虚拟机规格族偏好 (可选)	根据选择的主机类型，罗列出包含的规格族类型，选择期望的目的端主机规格族，支持多选。 须知 勾选“仅显示适合专属主机规格族”，则只能选择在购买区域所选Region下专属主机所包含的规格，并且虚拟机类型偏好将无法选择。
	系统盘规格偏好 (可选)	选择期望的目的端系统盘规格类型。
	数据盘规格偏好 (可选)	选择期望的目的端数据盘规格类型。
	虚拟机规格调整	选择主机性能规格的推荐策略。 详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。

步骤6 配置完成后，单击“**创建评估**”，主机规格评估创建成功。规格评估完成后，可以[查看评估推荐结果](#)和[主机性能数据](#)，其中目的端规格配置即为该资源的迁移目的端。

步骤7 (可选) 您还可以进行如下操作：

- **更改目的端规格**：根据实际需求，可以调整目的端主机规格和磁盘规格，并支持磁盘扩容。
- **绑定目的端**：根据评估结果，如果您在华为云上已有满足要求的主机，可以选择并绑定该主机作为目的端主机。

----结束

查看评估推荐结果

在**迁移方案设计**页面的**应用列表**，单击操作列的“**查看资源配置**”，进入目的端配置表页面。

在**资源配置明细**区域，可以查看应用内各类型资源的评估推荐结果，其中**目的端规格配置**是根据源端规格配置和评估策略推荐的华为云对应规格配置，**成本预估**是华为云推荐规格所需费用。



查看主机性能数据

在**资源配置明细**的主机列表中，可以查看每台主机最后采集日期7天/30天内的**CPU负载**和**内存负载**平均值。单击“**性能分析**”，可以查看所有主机的性能分布统计。



更改目的端规格

- 步骤1** 在资源配置明细区域，单击主机资源操作列的“更改目的端规格”，右侧弹出资源详情窗口。
- 步骤2** 单击主机规格后的“更改规格”，可以修改目的端主机规格和镜像。

规格配置



- 步骤3** 单击磁盘对应的目的端规格处的“更改规格”，可以修改目的端磁盘的存储类型和资源总用量（磁盘容量）。Linux主机的磁盘容量可以根据源端主机磁盘的实际使用量，进行扩容，即小于源端磁盘容量。更改磁盘规格后，系统会自动判断是否进行了磁盘扩容，在主机资源配置明细列表中的“磁盘扩容”列，“是”代表已扩容；“否”代表未扩容。

须知

- 系统盘容量取值范围为：40 G~1024 G
- 数据盘容量取值范围为：10 G~32768 G
- 当前仅支持Linux主机磁盘扩容，需确保扩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置扩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

资源类型	源端规格	目的端规格	包周期价格/月	按需价格/小时
系统盘	SATA 40 GB	普通IO 40 GB 更改规格	¥12.00	¥0.02

----结束

2.3.7 步骤五：创建主机迁移 workflow

当完成所有[准备工作](#)和[步骤一~步骤四](#)的操作后，即可前往[迁移 workflow 界面](#)，[创建主机迁移 workflow](#)。



2.4 主机一站式跨可用区迁移

适用场景

本文介绍如何使用迁移中心MgC将华为云上的云主机从可用区A快速搬迁到可用区B。适用于30台以下单批次、跨可用区主机迁移，无需手动采集、分组、评估，一站式解决创建工作流前所有步骤。

准备工作

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，获取方法请参见[准备工作](#)。

操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“**总览**”，进入**总览**页面。
- 步骤3** 在**新手入门**选择**跨可用区迁移**，在**单批次小数量主机迁移用户一站式跨可用区迁移**区域，单击“**开始创建**”按钮。



- 步骤4** 在弹出一站式跨可用区迁移窗口，自定义应用名称，选择要迁移到的目的可用区。

步骤5 单击“开始执行”，系统会自动采集当前账号下所选区域主机信息，并自动创建应用将主机加入到该应用内，开始迁移评估。

步骤6 评估执行完成后，单击“关闭”按钮，进行 workflow 配置。

步骤7 根据表2-6参数说明，配置 workflow 参数。

表 2-6 参数说明

区域	参数	说明
workflow 详情	名称	用户自定义
	描述	用户自定义
应用程序	应用名称	选择步骤4中自定义的应用名称。
迁移设置	区域	选择待迁移可用区所在Region，与创建的应用配置保持一致。
	目的可用区	选择要迁移到目的端可用区，与创建的应用配置保持一致。
	网络设置	当前只支持“保留源端配置”。
	目的虚拟机创建策略	立即创建 workflow 立即开始进行源端备份，源端各自独立备份完成后，立即开始生成镜像和目的端主机。
	目的端是否关机	<ul style="list-style-type: none"> 选择“是”，生成目的端后，目的端保持关机状态。 选择“否”，生成目的端后，目的端自动开机。
	增量备份前源端是否关机	<ul style="list-style-type: none"> 选择“是”，workflow 在执行增量备份前，会关闭源端主机，这样可以最大限度保证新建目的端数据与源端数据的一致性。 选择“否”，workflow 在执行增量备份时，依旧保持开机状态。
	创建系统盘镜像	<ul style="list-style-type: none"> 选择“是”，将基于源端系统盘创建系统盘镜像，便于后续目的端重装系统。 选择“否”，不创建系统盘镜像。
高级设置	自动删除镜像	选择开启，业务割接后自动删除迁移过程中产生的备份、快照及镜像文件。
	保持主网卡IP不变	选择开启时，将目的端的主网卡私有IP切换为源端主网卡私有IP，并绑定源端主网卡绑定的EIP，此时源端主网卡私有IP将设置为随机IP，并解绑EIP。以上操作不支持自动回退，需要您手动回退。

步骤8 workflow 配置完成后，单击“下一步：确认并创建”按钮，确认配置信息无误后，单击“确认创建 workflow”。迁移 workflow 创建完成，可在 workflow 列表查看。

注意

此时迁移 workflow 状态为“等待中”，并未开始迁移。

步骤9 单击 workflow 名称，进入 workflow 详情页面，**workflow 步骤**为迁移模板的标准步骤，您也可以在工作流步骤中自行**添加阶段/步骤**。

步骤10 确定 workflow 步骤后，单击“运行”按钮，开始迁移。

- 在**workflow 步骤**页签可以监控迁移进展。如果迁移步骤中包含**手动操作**，需要您进行手动确认后，才能进行后续步骤。
- 在**主机**页签，可以查看每台主机的迁移状态。

----结束

2.5 华为云主机跨可用区迁移

适用场景

本文介绍如何使用迁移中心MgC将华为云可用区的主机搬迁到其他可用区，适用于大批量主机跨可用区迁移。30台以下的小规模、单批次迁移请使用[主机一站式跨可用区迁移](#)。

准备工作

- 使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，获取方法请参见[准备工作](#)。
- 在MgC控制台[创建项目](#)。

步骤一：采集可用区主机资源

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“调研>资源采集”，进入资源采集页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击任务框中的“发现>公网发现”，进入创建公网发现任务页面。

图 2-5 公网发现



步骤4 根据表2-7，配置公网发现任务参数。

表 2-7 公网发现任务参数说明

区域	参数	说明	是否必选
基础配置	任务名称	自定义发现任务名称。	是
	任务说明	输入发现任务说明。	否
任务配置	源端	选择华为云。	是
	凭证	选择源端对应的凭证，如未提前 添加凭证 ，请单击“ 新建凭证 ”，填写相应参数，校验并保存凭证。 说明 新建凭证中的AK/SK输入源端账号的AK/SK。	是
	区域	选择源端主机所在区域，支持多选。	是

步骤5 开启云平台采集，在资源类型下拉列表中勾选**主机**。

步骤6 （可选）将采集的主机资源关联到应用。

- 已提前[创建应用](#)，在“应用”的下拉列表中选择要关联的应用。
- 未提前创建应用，单击“[新建应用](#)”，弹出[新建应用](#)窗口，输入自定义的**应用名称**和**描述**；**业务场景**选择“[跨可用区迁移](#)”；**使用场景**根据实际情况选择；选择目的端所在区域和可用区，单击“[确定](#)”按钮，应用创建成功。

步骤7 单击“[确认](#)”，公网发现任务创建完成，系统开始自动发现**步骤4**任务配置所选区域内的主机资源。

- 您可以在[资源采集](#)页面的“[资源列表](#)”，单击主机资源的**资源数量**，进入[资源列表](#)页面的**主机**页签。
- 在[资源采集](#)页面，单击**任务总数**后的“[查看](#)”，进入[任务列表](#)页面，查看任务状态。如果任务状态为“[失败](#)”，单击操作列的“[查看](#)”，可以查看采集失败的数据源，将鼠标放到数据源的采集状态上，可以查看失败原因。

----**结束**

步骤二：关联应用

如果在采集可用区主机资源的**步骤6**已经为主机关联了应用，则可以跳过该步骤，进行下一步[规格评估](#)。

步骤1 在[资源列表](#)页面的**主机**页签，勾选需要关联到同一应用的主机，单击页面左上角的“[资源管理](#)>[关联应用](#)”，弹出[关联应用](#)窗口。

步骤2 在下拉列表中选择需要关联的应用。未提前[创建应用](#)，单击“[新建应用](#)”，弹出[新建应用](#)窗口，输入自定义的**应用名称**和**描述**；**业务场景**选择“[跨可用区迁移](#)”；**使用场景**根据实际情况选择；选择目的端所在区域和可用区，单击“[确定](#)”按钮，应用创建成功。

步骤3 选择应用后，单击“确定”。在主机的**已关联应用**列，查看资源所关联应用。

----结束

步骤三：规格评估

步骤1 在**迁移方案设计**页面，单击**目的端规格配置框**中的“规格评估”按钮，进入**评估单应用**页面。

步骤2 在**选择应用**的下拉列表中，选择源端主机**关联的应用**。

步骤3 在**选择资源区域**，勾选应用内，需要评估的主机。

步骤4 根据**表2-8**，配置评估策略。

表 2-8 评估策略参数说明

参数	说明
购买区域	选择目的端所在区域。
推荐策略	选择 跨可用区迁移 ，同时选择 目的可用区 。
优先级设置	<ul style="list-style-type: none">性能优先 推荐的目的端规格要求性能最佳。价格优先 推荐的目的端规格在满足要求的前提下尽量节约成本。
偏好设置	可以设置“虚拟机类型”、“虚拟机规格族”、“系统盘规格”、“数据盘规格”，推荐目的端时优先推荐满足以上要求的主机。

步骤5 配置完成后，单击“确认”，创建评估成功。

步骤6 在**迁移方案设计**页面的**应用列表**，单击应用操作列的“查看资源配置”，进入**目的端配置表**页面。

在**资源配置明细**区域，可以查看应用内主机资源的评估推荐结果，其中**目的端规格配置**是根据源端规格配置和评估策略推荐的华为云对应规格配置，**成本预估**是华为云推荐规格所需费用；单击主机资源操作列的“更多>更改目的端规格”，可以修改目的端主机规格、磁盘存储类型和资源总用量。

----结束

步骤四：创建可用区迁移 workflow

完成已上**步骤一~步骤三**的操作后，即可**创建跨可用区迁移 workflow**。

2.6 将主机迁移至 Flexus 云服务器 X 实例（原云耀云服务器 X 实例）

概述

相对于其他弹性云服务器产品，[Flexus云服务器 X实例](#)提供了用户可自定义的精细化产品规格，用来最大化精确匹配用户业务的实际资源需求，以降低用户上云成本，避免资源浪费。为了降低用户选择或自定义X实例规格的复杂度，在迁移场景下，可借助MgC的源端主机性能采集功能，获取源端主机CPU和内存的实际使用情况，再基于性能的评估策略帮助您有效评估并精准调整资源规格配置，实现性能提升与成本控制的双重目标。

适用范围

主机来源：适用于阿里云、华为云、AWS、腾讯云、Azure等云厂商的主机，以及自主添加的线下主机。

使用MgC服务进行相关操作，步骤请参考下表。

操作	步骤
Flexus云服务器 X实例评估+手动购买	准备工作+步骤1~7+ 自行购买 。
Flexus云服务器 X实例评估+手动购买+使用MgC workflows迁移上云	准备工作+步骤1~8+ 自行购买 ，并在步骤7绑定自行购买的X实例主机。
Flexus云服务器 X实例评估+自动购买+使用MgC workflows迁移上云	准备工作+步骤1~8。

准备工作

- **华为账号准备：**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。
- **创建迁移项目：**
在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。

操作步骤

步骤1 下载并安装迁移工具Edge。

在源端内网环境中准备一台Windows主机，用于安装迁移工具Edge，如果您没有空闲的云主机，可以购买一台按需收费的主机，规格推荐4U8G。对该主机的要求以及Edge的安装方法请参见[安装Windows版本的Edge](#)。

步骤2 连接迁移中心。

Edge安装完成并注册用户后，将Edge与迁移中心进行连接，方法请参考[连接迁移中心](#)。

步骤3 添加资源凭证。

Edge与迁移中心连接成功后，需要在Edge上添加源端待迁移主机凭证，凭证的添加方法请参考[添加资源凭证](#)。资源类型请根据实际情况选择“**Windows主机**”或“**Linux主机**”。

步骤4 采集主机资源。

MgC提供了三种主机采集方式，您可以根据实际场景和您源端主机所在的云平台，选择适合的采集方式。

图 2-6 采集方式



- 源端主机位于第三方云平台：源端为**阿里云、华为云、AWS、腾讯云、谷歌云、Azure**等云平台主机，推荐使用**公网采集**，也可以使用各场景通用的**手动添加**方式。
- 源端主机位于线下IDC：推荐使用**内网采集**，通过**网段扫描**或**VMware扫描**，采集源端主机资源，也可以使用各场景通用的**手动添加**方式。

步骤5 采集主机性能。

根据采集到的源端主机实际性能负载推荐目的端主机CPU核数、内存大小等性能规格。更多详情以及采集方法请参见[采集主机性能](#)。

须知

- 如果是每日固定业务，推荐性能采集的时长不少于1天。
- 如果是周期性波动业务，推荐性能采集的时长覆盖整个波动周期，以便后续更准确的评估源端主机性能。

步骤6 关联应用。

由于后续的规格评估和创建迁移 workflows 都是以应用为单位，因此需要将待迁移主机关联到应用，方法请参见[关联到应用](#)。

步骤7 规格评估。通过配置评估策略，推荐出X实例规格的主机。

1. 单击MgC控制台左侧导航栏的“**设计>迁移方案设计**”，进入**迁移方案设计**页面。在迁移方案设计页面，可以查看当前项目下所创建的应用列表。
2. 在应用列表中，单击应用操作列的“**规格评估**”，进入**评估单应用**页面。



3. 在**选择资源区域**，勾选应用内，需要评估的资源。
4. 根据**表2-9**，配置评估策略，即可优先推荐出X实例。



表 2-9 评估策略参数配置说明

参数	子参数	说明
购买区域	-	选择在华为云购买资源的区域。推荐您选择靠近目标用户的区域，可以减少网络时延，提高访问速度。
推荐策略	匹配源端规格	根据源端规格推荐最合适的华为云对应资源规格。
优先级设置	价格优先	推荐的目端规格在满足要求的前提下尽量节约成本。
偏好设置（必选） 针对迁移到 Flexus 云服务器X实例的场景，建议选择“按性能负责调整”的规格调整。	虚拟机类型偏好（可选）	选择“通用计算型”。详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。
	虚拟机规格族偏好（可选）	选择期望的目的端主机规格族。
	系统盘规格偏好（可选）	选择期望的目的端系统盘规格类型。
	数据盘规格偏好（可选）	选择期望的目的端数据盘规格类型。
	虚拟机规格调整	选择“按性能负责调整”，需要先完成 源端主机性能采集 ，获取源端主机性能负载数据，再通过设置数据处理规则，推荐满足条件的CPU核数、内存大小。采集的性能数据越多，推荐的规格越准确。 详细的推荐逻辑说明参见 主机评估策略推荐逻辑说明 。

- 配置完成后，单击“**创建评估**”，主机规格评估创建成功。规格评估完成后，可以[查看评估推荐结果](#)和[查看主机性能数据](#)，其中目的端规格配置即为该资源的迁移目的端。
- （可选）您还可以进行如下操作：
 - 更改目的端规格**：根据实际需求，可以调整目的端主机规格和磁盘规格，并支持磁盘扩容。
 - 绑定目的端**：根据评估结果，如果您在华为云上已有满足要求的主机，可以选择并绑定该主机作为目的端主机。

步骤8 创建主机迁移工作流。

当完成以上所有步骤后，即可前往**迁移 workflow 界面**，[创建主机迁移工作流](#)。



----结束

2.7 迁移后主机私有 IP 保持不变

迁移中心 MgC 的主机迁移工作流，提供了保持 IP 不变功能，可以实现主机迁移后保持私有 IP 不变，减少因 IP 变动而需要对业务代码进行的修改工作。

免责声明

- 业务可用性风险**

本功能旨在帮助用户在迁移后保持主机的私有 IP 地址与源端一致，但不保证切换 IP 后的业务可用性。需要用户自行评估并承担由此带来的风险。
- 回退说明**

主机迁移工作流提供自动化的 IP 不变功能，但不提供自动化回退功能。如果用户在使用该功能后遇到问题需要回退，仅支持**手动回退**。
- IP 冲突风险提示**

源端与目的端在私有 IP 相同的情况下，可能会引发 IP 冲突，进而导致业务不可用。
- 未知风险提示**

由于迁移无法对业务进行深入的嗅探和分析，因此使用 IP 不变功能后可能出现其他未被预见的问题。
- 用户责任**

用户应当在迁移前进行充分的测试和准备，并在迁移后自行排查和解决可能出现的问题。我们建议在正式迁移前，用户先在测试环境中进行模拟迁移，以评估潜在风险并制定相应的应对策略。

约束与限制

- **目的端主机关机**
在执行IP不变的迁移操作前，目的端主机必须处于关机状态。如果目的端未关机，IP不变功能将尝试自动关机以确保迁移的顺利进行。
- **指定子网**
需要指定最终切换的VPC中的子网，并且该子网的网段必须与源端的网段保持一致。
- **网卡绑定限制**
目的端主机应只绑定一个网卡，不绑定任何额外的扩展网卡。
- **IP版本支持**
仅支持IPv4地址的切换。

准备工作

- **华为账号准备：**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。
- **创建迁移项目：**
在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。

操作步骤

步骤1 下载并安装迁移工具Edge。

在源端内网环境中准备一台Windows主机，用于安装迁移工具Edge。对该主机的要求以及Edge的安装方法请参见[安装Windows版本的Edge](#)。

步骤2 连接迁移中心。

Edge安装完成并注册用户后，将Edge与迁移中心进行连接，方法请参考[连接迁移中心](#)。

步骤3 添加资源凭证。

Edge与迁移中心连接成功后，需要在Edge上添加源端待迁移主机凭证，凭证的添加方法请参考[添加资源凭证](#)。资源类型请根据实际情况选择“Windows主机”或“Linux主机”。

步骤4 采集主机资源。

MgC提供了三种主机采集方式，您可以根据实际场景和您源端主机所在的云平台，选择适合的采集方式。

图 2-7 采集方式



- 源端主机位于第三方云平台：源端为阿里云、华为云、AWS、腾讯云、谷歌云、Azure等云平台主机，推荐使用公网采集，也可以使用各场景通用的手动添加方式。
- 源端主机位于线下IDC：推荐使用内网采集，通过网段扫描或VMware扫描，采集源端主机资源，也可以使用各场景通用的手动添加方式。

步骤5 关联应用。

由于后续的规格评估和创建迁移 workflows 均是以应用为单位，因此需要将待迁移主机关联到应用，方法请参考[关联到应用](#)。

步骤6 规格评估。

根据源端主机规格、性能和应用场景等，为您推荐最合适的华为云对应主机规格，通过规格评估确定迁移目的端。方法请参考[规格评估](#)。

主机规格评估完成后，可以根据实际需求，选择是否进行如下操作：

- **更改目的端规格**：调整目的端主机规格和磁盘规格，支持磁盘缩容。
- **绑定目的端**：根据评估结果，如果您在华为云上已有满足要求的主机，可以选择并绑定该主机作为目的端主机。

步骤7 创建主机迁移 workflow。当完成以上所有步骤后，即可前往迁移 workflow 界面，创建主机迁移 workflow。

注意

- 建议单个 workflow 中包含的主机数量不超过100台。
- 同一用户同时进行迁移的主机数量上限为1000台，超出此数量的主机会在 workflow 第一步暂停并处于等待状态，直至有资源释放或主机完成迁移，等待中的主机将按照创建时间顺序依次自动执行 workflow 后续步骤。
- 如果您是首次使用MgC的主机迁移 workflow，需要您同意华为云账号进行相关的委托授权。委托权限所包含的细粒度权限请前往[委托权限说明](#)查看。

1. 在左侧导航栏选择“**实施>迁移 workflow**”，进入迁移 workflow 页面。在页面左上角的**当前项目**下拉列表中选择在准备工作中创建的**迁移项目**。
2. 选择**主机迁移标准模板**，单击“**创建主机 workflow**”，进入配置 workflow 页面。



3. 在 workflow 详情区域，自定义 workflow 名称和描述。
4. 在**应用程序**区域，选择**步骤5**待迁移主机关联的应用。

须知

应用内的主机，需要完成[规格评估](#)或[绑定目的端](#)。

5. 在**迁移网络配置**区域，选择网络类型。
 - 选择**公网迁移**，要求目的端服务器配置有“**弹性公网IP (EIP)**”，迁移时会使用目的端EIP。

- 选择**私网**迁移，需要提前配置专线、VPN、对等连接或同VPC子网，确保源端能够通过私网访问目的端。
6. 在**目的端配置区域**，选择与源端网络打通的VPC和子网作为中转VPC。

须知

配置的VPC和子网，对已绑定目的端的主机不生效。如果所选应用中的主机均已绑定目的端，则无需进行目的端配置。

7. 在**高级设置区域**，**保持IP不变**参数选择“是”，仔细阅读高危操作提示并了解使用该功能的**免责声明**后，单击“确认”，启动IP不变功能。

说明

本文只介绍**保持IP不变**参数的配置方法，其余参数根据实际需求配置。

高级设置

迁移后是否开机 是 否

网络限速 是 否
网络限速可以避免迁移过程中对现有网络资源造成过大压力

自动安装rsync 是 否
rsync是linux主机迁移必要的应用程序，检测到源端linux服务器缺少rsync时会自动安装

保持IP不变 是 否
支持目的端私网IP与源端保持一致，目的端未处于关机状态时，会自动进行关机处理

保持IP不变-虚拟私有云 不选择时, 将自动创建

保持IP不变-子网 不选择时, 将自动创建

企业项目 请选择

仅对新创建的目的端主机生效，对已绑定目的端的主机不生效，仅支持已启用的商用生产项目

8. 在**保持IP不变-虚拟私有云**的下拉列表中，选择包含源端私有IP网段的VPC；在**保持IP不变-子网**的下拉列表中，选择包含源端私有IP的子网。选择后，系统会对所选的子网与源端私有IP进行匹配，只有匹配成功，才能继续完成工作流的创建。
9. 工作流配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。
10. 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。
- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
 - 如果您需要在标准工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。
11. 在迁移工作流详情页面，可以查看该主机迁移工作流的**配置信息**和**资源迁移进度**。当资源的迁移进度完成“启动迁移Agent”这一步后，会自动在SMS控制台的迁移服务器列表创建一条迁移任务，关于MgC和SMS主机资源之间的映射关系请查看[MgC主机工作流资源与SMS迁移服务器列表之间的映射关系](#)。
- 将鼠标移动到主机迁移进度条，在弹窗中可以查看该主机更多迁移信息。
 - 当迁移步骤（进度条）到达需要**手动确认**的步骤时，将鼠标移动到进度条，在弹窗中单击**步骤状态**后的“**确认**”，才会继续进行后续迁移步骤。
 - 当迁移步骤（进度条）到达“**调整磁盘分区**”步骤时，系统会识别目的端主机是否进行了磁盘扩容。

- 是，则需要前往[SMS控制台](#)，对该主机进行“调整磁盘分区”操作，方法请参考[设置目的端](#)中的[是否调整分区](#)参数配置说明。调整完成后，单击[步骤状态](#)后的“确认”继续进行后续迁移步骤。
- 否，则跳过该步骤。
- 迁移步骤中的“增量同步”为可重复步骤，在进行业务验证前，可以重复执行。
- 当迁移步骤（进度条）到达“业务割接”时，代表迁移完成。请前往目的端进行业务验证，验证无误后，手动将业务切换到目的端。切换完成后单击“确认”系统将自动进行后续“源端专项清理”和“迁移任务清理”操作。

----结束

手动回退

如果您在迁移时，使用了“保持IP不变”功能后遇到问题需要回退，可以选择手动切换VPC的方式进行回退。具体操作请参考[切换虚拟私有云](#)。

2.8 批量修改与还原源端 Linux 主机的 Host 配置

2.8.1 概述

背景介绍

当您迁移Linux主机时，需要确保源端主机能够识别并解析相关云服务的接口域名。这通常涉及到在每台源端主机上编辑hosts文件。如果迁移的主机数量众多，手动编辑将非常耗时。为了简化这一过程，我们提供了一个批量修改的脚本示例。您可以利用这个脚本，快速地将域名解析IP批量推送到所有源端主机的hosts文件中。

脚本说明

脚本是基于Shell语言开发，仅支持在Linux系统上运行，实现对源端待迁移Linux主机的/etc/hosts文件的批量更新和批量恢复。

脚本执行限制：每次执行脚本时最多支持同时修改100台主机的hosts配置，以避免脚本执行时间过长。

脚本功能与使用场景详细说明参见下表。

脚本名	功能描述	使用场景
update_hosts_linux.sh	批量更新源端Linux主机的/etc/hosts文件。脚本会自动记录操作日志，提示异常信息，并生成操作总结。	在迁移Linux主机之前，执行此脚本以更新源端主机的hosts配置。

脚本名	功能描述	使用场景
rollback_hosts_linux.sh	批量恢复源端Linux主机的/etc/hosts文件。脚本会自动记录操作日志，提示异常信息，并生成操作总结。	在完成Linux主机迁移后，执行此脚本以恢复源端主机的hosts配置到迁移前的状态。

2.8.2 准备工作

CSV 文件准备

创建一个包含源端主机信息的CSV文件，确保文件能被访问，并且格式符合要求。格式如下：

```
username,ip,port,password
```

- **username**：该字段表示登录源端主机时所使用的用户名。为了确保脚本具有足够的权限执行修改操作，通常应使用具有管理员权限的用户名，如root。
- **ip**：该字段表示源端主机的内网IP地址。
- **port**：该字段表示SSH服务监听的端口号。默认情况下，Linux系统使用端口22。如果源端主机的SSH服务运行在自定义端口上，请在此处指定正确的端口号。
- **password**：该字段表示登录源端主机时所需的密码。脚本将使用此密码通过SSH协议自动登录到源端主机。

注意

- CSV文件的第一行是标题行，不会被脚本解析
- CSV文件的每一行对应一台主机信息，并且每一列信息之间用英文逗号分隔。
- 确保格式正确和信息准确，不要出现多余的空格、逗号或无效的IP地址等。

示例：

```
username,ip,port,password
root,192.168.1.10,xx,examplePass123
root,192.168.1.11,xx,examplePass456
```

Hosts 文件准备

创建一个文本文件，包含需要追加到源端主机/etc/hosts文件中的内容，并确保该文件能被访问。文件内写入相关云服务的接口域名映射，并以**#Migration-proxy-start**开始，以**#Migration-proxy-end**结束。相关云服务的接口域名映射需要根据具体的环境而定，请联系对应局点的环境接口人获取。

示例：

```
#Migration-proxy-start
xxx.xxx.xxx.xxx iam.xxx.com
xxx.xxx.xxx.xxx ecs.xxx.com
xxx.xxx.xxx.xxx evs.xxx.com
xxx.xxx.xxx.xxx ims.xxx.com
xxx.xxx.xxx.xxx obs.xxx.com
xxx.xxx.xxx.xxx eps.xxx.com
```



```
xxx.xxx.xxx.xxx vpc.xxx.com  
#Migration-proxy-end
```

日志目录准备

- **配置日志目录：**脚本会默认使用 `/var/log/update_hosts` 作为日志存储目录。目录不存在的话，脚本会自动创建，如果需要更改日志存储路径，请修改脚本中的 `LOG_DIR` 参数的值。
- **检查日志目录权限：**确保当前用户对日志目录有写入权限。如果权限不足，需要调整目录权限或选择其他路径。

检查源端主机连通性

- **网络连接检查：**确保执行脚本的主机能够通过网络访问所有源端主机。
- **SSH端口开放检查：**确保源端主机的SSH端口（默认是22）已开放，允许从执行脚本的主机通过SSH协议进行连接。

SSH 访问权限设置

- **SSH服务配置：**确保所有源端主机上的SSH服务已启用并运行正常。大多数Linux服务器发行版默认启用SSH服务。
- **SSH密码认证：**确保所有源端主机的SSH服务已配置为接受密码认证，以便脚本可以使用CSV文件中提供的密码进行自动登录。

脚本执行机环境配置

确保执行脚本的Linux主机上安装了以下必要的工具和命令：

- **SSH工具：**用于建立与远程主机的安全连接。
- **sed工具：**用于编辑和修改/etc/hosts文件内容。
- **setsid命令：**用于在SSH连接中避免交互需求（通常用于避免密码输入时的交互）。
- **mktemp命令：**用于创建临时文件或目录。

执行权限和用户权限配置

- **用户权限配置：**确保执行脚本的用户具有对日志目录、CSV文件和hosts文件的读写权限。
- **脚本执行权限配置：**确保 `update_hosts_linux.sh` 和 `rollback_hosts_linux.sh` 脚本文件具有执行权限。使用 `chmod +x update_hosts_linux.sh` 和 `chmod +x rollback_hosts_linux.sh` 命令赋予脚本执行权限。

2.8.3 配置脚本

2.8.3.1 配置 update_hosts_linux.sh 脚本

根据提供的脚本示例，结合实际情况，修改示例中的相关配置。

前提条件

已完成[准备工作](#)。

操作步骤

步骤1 在执行脚本的主机上创建一个名为“**update_hosts_linux.sh**”的文件，然后将以下脚本示例的内容复制到文件中。如果您已经通过SSH连接到源端Linux主机，可以直接使用vim创建和编辑脚本文件，步骤如下：

1. 在Vim编辑器中按 i 进入插入模式。
2. 复制并粘贴脚本代码，完成后按 Esc。
3. 输入 :wq 保存并退出。

```
#!/bin/bash

# Configuration
# Log directory path: Used to store run logs, error logs, and summary logs.
# If the directory doesn't exist, the script will create it automatically.
LOG_DIR="/var/log/update_hosts"

# Run log file path: Records detailed information about the script's execution.
RUN_LOG="$LOG_DIR/run.log"

# Error log file path: Records any errors that occur during the script's execution.
ERROR_LOG="$LOG_DIR/error.log"

# Summary log file path: Records a summary of the script's execution, including the number of successful
and failed servers.
SUMMARY_LOG="$LOG_DIR/summary.log"

# CSV file path: Contains information about the target hosts (must be manually created and configured).
CSV_FILE="target_servers.csv"

# Hosts content file path: Contains the content to be appended to each target host's /etc/hosts file (must
be manually created and configured).
HOSTS_FILE="hosts_content.txt"
DEFAULT_PORT=22
SSH_TIMEOUT=10

# Initialize log directory and files
initialize_logs() {
    mkdir -p "$LOG_DIR"
    echo "===== " >> "$RUN_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new update execution" >> "$RUN_LOG"
    echo "===== " >> "$RUN_LOG"

    echo "===== " >> "$ERROR_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new update execution" >> "$ERROR_LOG"
    echo "===== " >> "$ERROR_LOG"

    echo "===== " >> "$SUMMARY_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new update execution" >> "$SUMMARY_LOG"
    echo "===== " >> "$SUMMARY_LOG"
}

# Log info function
log_info() {
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG"
}

# Log error function
log_error() {
    echo "[ERROR] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG" "$ERROR_LOG"
}

# Read server information from CSV file
read_servers_from_csv() {
    local csv_file="$1"
    local servers=()
    local header_skipped=false
```

```

if [ ! -f "$csv_file" ]; then
    log_error "CSV file '$csv_file' not found."
    exit 1
fi

# Ensure file ends with a newline character
sed -i -e '$a\ ' "$csv_file"

while IFS=, read -r username ip port password; do
    # Skip header row
    if [ "$header_skipped" = false ]; then
        header_skipped=true
        continue
    fi
    # Skip empty and invalid rows
    if [[ -z "$username" || -z "$ip" ]]; then
        continue
    fi
    port=${port:-$DEFAULT_PORT} # Use default port 22
    # Ensure port is numeric
    if ! [[ "$port" =~ ^[0-9]+$ ]]; then
        log_error "Invalid port '$port' for $username@$ip. Skipping this server."
        continue
    fi
    servers+=("$username@$ip:$port:$password")
done < "$csv_file"

echo "${servers[@]}"
}

# Read hosts content from TXT file
read_hosts_content_from_txt() {
    local txt_file="$1"
    if [ -f "$txt_file" ]; then
        cat "$txt_file"
    else
        log_error "Hosts content file '$txt_file' not found."
        exit 1
    fi
}

# Initialize log files
initialize_logs

# Read server information from CSV file
servers=$(read_servers_from_csv "$CSV_FILE")

# Read hosts content from TXT file
hosts_content=$(read_hosts_content_from_txt "$HOSTS_FILE")

# Counters for success and failure
success_count=0
failure_count=0
failed_servers=()

# Iterate over each server and push hosts content
for server in "${servers[@]}; do
    # Extract user, IP, port, and password information
    IFS=: read -r user_host port pass <<< "$server"
    IFS=@ read -r user ip <<< "$user_host"

    log_info "Starting update for $user@$ip:$port"

    # Create temporary script and SSH_ASKPASS script
    tmp_script=$(mktemp)
    askpass_script=$(mktemp)

    cat <<EOF > "$tmp_script"

```

```
#!/bin/bash
# Backup hosts file
if [ ! -f /etc/hosts.bak ]; then
    cp /etc/hosts /etc/hosts.bak
fi

# Remove old Migration-proxy section
sed -i '/#Migration-proxy-start/,/#Migration-proxy-end/d' /etc/hosts

# Append new Migration-proxy section
echo "$hosts_content" >> /etc/hosts
EOF

cat <<EOF > "$askpass_script"
#!/bin/bash
echo "$pass"
EOF

chmod +x "$tmp_script" "$askpass_script"

# Set SSH_ASKPASS environment variable and use ssh to connect to the target machine and execute the
temporary script
export SSH_ASKPASS="$askpass_script"
export DISPLAY=:0

ssh_output=$(mktemp)
setsid ssh -o BatchMode=no -o ConnectTimeout=$SSH_TIMEOUT -o StrictHostKeyChecking=no -p "$port"
"$user@$ip" "bash -s" < "$tmp_script" 2> "$ssh_output"
ssh_status=$?

if [ $ssh_status -eq 0 ]; then
    log_info "Updated hosts on $ip:$port successfully"
    ((success_count++))
else
    ssh_error=$(cat "$ssh_output")
    case $ssh_status in
        1)
            log_error "General error occurred while updating hosts on $ip:$port: $ssh_error"
            ;;
        2)
            log_error "Misuse of shell builtins while updating hosts on $ip:$port: $ssh_error"
            ;;
        255)
            if [[ "$ssh_error" == *"Permission denied"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Permission denied (password may be incorrect or
username is wrong)"
            elif [[ "$ssh_error" == *"Connection refused"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Connection refused (port may be incorrect or SSH
service not running on target)"
            elif [[ "$ssh_error" == *"No route to host"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: No route to host (network unreachable)"
            elif [[ "$ssh_error" == *"Host key verification failed"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Host key verification failed"
            elif [[ "$ssh_error" == *"Connection timed out"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Connection timed out"
            else
                log_error "SSH login failed for $user@$ip:$port: $ssh_error"
            fi
            ;;
        *)
            log_error "An unknown error occurred while updating hosts on $ip:$port: $ssh_error"
            ;;
    esac
    failed_servers+="$user@$ip:$port"
    ((failure_count++))
fi

# Remove temporary scripts and SSH output file
rm -f "$tmp_script" "$askpass_script" "$ssh_output"
```

```
done

# Calculate failure and success percentages
total_count=${#servers[@]}
failure_percentage=$(echo "scale=2; ($failure_count / $total_count) * 100" | bc)
success_percentage=$(echo "scale=2; ($success_count / $total_count) * 100" | bc)

# Output summary result and log to file
summary_content=$(cat <<EOF
=====
[SUMMARY] $(date '+%Y-%m-%d %H:%M:%S') - Execution Update Summary
=====
Total number of servers: $total_count
Number of successful updates: $success_count
Number of failed updates: $failure_count
Success rate: $success_percentage%
Failure rate: $failure_percentage%
-----
EOF
)

if [ $failure_count -gt 0 ]; then
    summary_content+="Failed servers:\n"
    for server in "${failed_servers[@]}; do
        summary_content+=" - $server\n"
    done
fi
summary_content+="-----"

# Output summary result to log file and terminal
echo -e "$summary_content" | tee -a "$SUMMARY_LOG"

log_info "Script execution completed. Check $SUMMARY_LOG for summary."
```

步骤2 根据实际情况，修改脚本中的以下配置项参数：

- **LOG_DIR="/var/log/update_hosts"**
 - **描述：**日志目录路径，用于存放运行日志、错误日志和总结日志的文件夹路径。
 - **参数默认值：** /var/log/update_hosts
 - **修改建议：**修改为当前用户具有写入权限的目录路径。
 - **修改示例：**LOG_DIR="/home/username/update_hosts_logs"
- **CSV_FILE="target_servers.csv"**
 - **描述：**CSV文件路径，包含源端主机信息。
 - **参数默认值：** target_servers.csv
 - **修改建议：**使用绝对路径，或确保相对路径是正确的。
 - **修改示例：**CSV_FILE="/home/username/configs/servers.csv"
- **HOSTS_FILE="hosts_content.txt"**
 - **描述：**Hosts文件存放路径，包含要追加到源端主机 /etc/hosts 文件中的内容。
 - **参数默认值：** hosts_content.txt
 - **修改建议：**使用绝对路径，或确保相对路径是正确的。
 - **修改示例：**HOSTS_FILE="/home/username/configs/hosts_content.txt"

步骤3 配置项参数修改完成并保存后，在终端窗口（如果有图形化界面可以按Ctrl+Alt+T打开终端）执行脚本：

```
./update_hosts_linux.sh
```

脚本会在终端窗口中输出日志信息，并在执行完毕后生成一个执行结果报告，可以在 LOG_DIR 指定目录中的 summary.log 文件中查看。

----结束

2.8.3.2 配置 rollback_hosts_linux.sh 脚本

根据提供的脚本示例，结合实际情况，修改示例中的相关配置。

前提条件

已完成[准备工作](#)。

操作步骤

步骤1 在执行脚本的主机上创建一个名为“rollback_hosts_linux.sh”的文件，然后将以下脚本示例的内容复制到文件中。如果您已经通过SSH连接到源端Linux主机，可以直接使用vim创建和编辑脚本文件，步骤如下：

1. 在Vim编辑器中按 i 进入插入模式。
2. 复制并粘贴脚本代码，完成后按 Esc。
3. 输入 :wq 保存并退出。

```
#!/bin/bash

# Configuration

# Log directory path: Used to store run logs, error logs, and summary logs.
# If the directory doesn't exist, the script will create it automatically.
LOG_DIR="/var/log/update_hosts"

# Run log file path: Records detailed information about the script's execution.
RUN_LOG="$LOG_DIR/run.log"

# Error log file path: Records any errors that occur during the script's execution.
ERROR_LOG="$LOG_DIR/error.log"

# Summary log file path: Records a summary of the script's execution, including the number of successful
and failed servers.
SUMMARY_LOG="$LOG_DIR/summary.log"

# CSV file path: Contains information about the target hosts (must be manually created and configured).
CSV_FILE="target_servers.csv"
DEFAULT_PORT=22
SSH_TIMEOUT=10

# Initialize log directory and files
initialize_logs() {
    mkdir -p "$LOG_DIR"
    echo "===== " >> "$RUN_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new rollback execution" >> "$RUN_LOG"
    echo "===== " >> "$RUN_LOG"

    echo "===== " >> "$ERROR_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new rollback execution" >> "$ERROR_LOG"
    echo "===== " >> "$ERROR_LOG"

    echo "===== " >> "$SUMMARY_LOG"
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - Starting new rollback execution" >>
"$SUMMARY_LOG"
    echo "===== " >> "$SUMMARY_LOG"
}
```

```

# Log info function
log_info() {
    echo "[INFO] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG"
}

# Log error function
log_error() {
    echo "[ERROR] $(date '+%Y-%m-%d %H:%M:%S') - $1" | tee -a "$RUN_LOG" "$ERROR_LOG"
}

# Read server information from CSV file
read_servers_from_csv() {
    local csv_file="$1"
    local servers=()
    local header_skipped=false

    if [ ! -f "$csv_file" ]; then
        log_error "CSV file '$csv_file' not found."
        exit 1
    fi

    # Ensure file ends with a newline character
    sed -i -e '$a\ ' "$csv_file"

    while IFS=, read -r username ip port password; do
        # Skip header row
        if [ "$header_skipped" = false ]; then
            header_skipped=true
            continue
        fi
        # Skip empty and invalid rows
        if [[ -z "$username" || -z "$ip" ]]; then
            continue
        fi
        port=${port:-$DEFAULT_PORT} # Use default port 22
        # Ensure port is numeric
        if ! [[ "$port" =~ ^[0-9]+$ ]]; then
            log_error "Invalid port '$port' for $username@$ip. Skipping this server."
            continue
        fi
        servers+=("$username@$ip:$port:$password")
    done < "$csv_file"

    echo "${servers[@]}"
}

# Initialize log files
initialize_logs

# Read server information from CSV file
servers=$(read_servers_from_csv "$CSV_FILE")

# Counters for success and failure
success_count=0
failure_count=0
failed_servers=()

# Iterate over each server and execute rollback
for server in "${servers[@]}; do
    # Extract user, IP, port, and password information
    IFS=: read -r user_host port pass <<< "$server"
    IFS=@ read -r user ip <<< "$user_host"

    log_info "Starting rollback for $user@$ip:$port"

    # Create temporary script and SSH_ASKPASS script
    tmp_script=$(mktemp)
    askpass_script=$(mktemp)

```

```

cat <<EOF > "$tmp_script"
#!/bin/bash
# Backup hosts file
if [ ! -f /etc/hosts.bak ]; then
    cp /etc/hosts /etc/hosts.bak
fi

# Remove old Migration-proxy section
sed -i '/#Migration-proxy-start/,/#Migration-proxy-end/d' /etc/hosts
EOF

cat <<EOF > "$askpass_script"
#!/bin/bash
echo "$pass"
EOF

chmod +x "$tmp_script" "$askpass_script"

# Set SSH_ASKPASS environment variable and use ssh to connect to the target machine and execute the
temporary script
export SSH_ASKPASS="$askpass_script"
export DISPLAY=:0

ssh_output=$(mktemp)
setsid ssh -o BatchMode=no -o ConnectTimeout=$SSH_TIMEOUT -o StrictHostKeyChecking=no -p "$port"
"$user@$ip" 'bash -s' < "$tmp_script" 2> "$ssh_output"
ssh_status=$?

if [ $ssh_status -eq 0 ]; then
    log_info "Rolled back hosts on $ip:$port successfully"
    ((success_count++))
else
    ssh_error=$(cat "$ssh_output")
    case $ssh_status in
        1)
            log_error "General error occurred while rolling back hosts on $ip:$port: $ssh_error"
            ;;
        2)
            log_error "Misuse of shell builtins while rolling back hosts on $ip:$port: $ssh_error"
            ;;
        255)
            if [[ "$ssh_error" == *"Permission denied"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Permission denied (password may be incorrect or
username is wrong)"
            elif [[ "$ssh_error" == *"Connection refused"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Connection refused (port may be incorrect or SSH
service not running on target)"
            elif [[ "$ssh_error" == *"No route to host"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: No route to host (network unreachable)"
            elif [[ "$ssh_error" == *"Host key verification failed"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Host key verification failed"
            elif [[ "$ssh_error" == *"Connection timed out"* ]]; then
                log_error "SSH login failed for $user@$ip:$port: Connection timed out"
            else
                log_error "SSH login failed for $user@$ip:$port: $ssh_error"
            fi
            ;;
        *)
            log_error "An unknown error occurred while rolling back hosts on $ip:$port: $ssh_error"
            ;;
    esac
    failed_servers+="$user@$ip:$port"
    ((failure_count++))
fi

# Remove temporary scripts and SSH output file
rm -f "$tmp_script" "$askpass_script" "$ssh_output"
done

```



```
# Calculate failure and success percentages
total_count=${#servers[@]}
failure_percentage=$(echo "scale=2; ($failure_count / $total_count) * 100" | bc)
success_percentage=$(echo "scale=2; ($success_count / $total_count) * 100" | bc)

# Output summary result and log to file
summary_content=$(cat <<EOF
=====
[SUMMARY] $(date '+%Y-%m-%d %H:%M:%S') - Execution Rollback Summary
=====
Total number of servers: $total_count
Number of successful rollbacks: $success_count
Number of failed rollbacks: $failure_count
Success rate: $success_percentage%
Failure rate: $failure_percentage%
-----
EOF
)

if [ $failure_count -gt 0 ]; then
    summary_content+="Failed servers:\n"
    for server in "${failed_servers[@]"; do
        summary_content+=" - $server\n"
    done
fi
summary_content+="======"

# Output summary result to log file and terminal
echo -e "$summary_content" | tee -a "$SUMMARY_LOG"

log_info "Script execution completed. Check $SUMMARY_LOG for summary."
```

步骤2 根据实际情况，修改脚本中的以下配置项参数：

- **LOG_DIR="/var/log/rollback_hosts"**
 - **描述：** 日志目录路径。
 - **参数默认值：** /var/log/rollback_hosts
 - **修改建议：** 修改为当前用户具有写入权限的目录路径。
 - **修改示例：** LOG_DIR="/home/username/rollback_hosts_logs"
- **CSV_FILE="target_servers.csv"**
 - **描述：** CSV 文件路径，包含源端主机信息。
 - **参数默认值：** target_servers.csv
 - **修改建议：** 使用绝对路径，或确保相对路径是正确的。
 - **修改示例：** CSV_FILE="/home/username/configs/servers.csv"

步骤3 配置项参数修改完成并保存后，在终端窗口（如果有图形化界面可以按Ctrl+Alt+T打开终端）执行脚本：

```
./rollback_hosts_linux.sh
```

脚本会在终端窗口中输出日志信息，并在执行完毕后生成一个执行结果报告，可以在 LOG_DIR 指定目录中的 summary.log 文件中查看。

----结束

2.9 批量修改与还原源端 Windows 主机的 Host 配置

2.9.1 概述

背景介绍

当您迁移Windows主机时，需要确保源端主机能够解析相关云服务的接口域名。这通常涉及到在每台源端主机上编辑hosts文件。如果迁移的主机数量众多，手动编辑将非常耗时。为了简化这一过程，我们提供了一个批量修改的脚本示例。您可以利用这个脚本，快速地将域名解析IP批量追加写入到所有源端主机的hosts文件中。

脚本说明

脚本是基于Powershell语言开发，仅支持在Windows系统上运行，实现对源端待迁移Windows主机的Hosts文件的批量更新和批量恢复。Hosts文件路径C:\Windows\System32\drivers\etc\hosts

脚本执行限制：每次执行脚本时最多支持同时修改100台主机的hosts配置，以避免脚本执行时间过长。

脚本功能与使用场景详细说明参见下表。

脚本名	功能描述	使用场景
update_hosts_win.ps1	批量更新源端Windows主机的hosts文件。脚本会自动记录操作日志，提示异常信息，并生成操作总结。	在迁移Windows主机之前，执行此脚本以更新源端主机的hosts配置。
rollback_hosts_win.ps1	批量恢复源端Windows主机的hosts文件。脚本会自动记录操作日志，提示异常信息，并生成操作总结。	在完成Windows主机迁移后，执行此脚本以恢复源端主机的hosts配置到迁移前的状态。

PowerShell 版本和依赖项要求

PowerShell 远程处理依赖于 Windows 远程管理 (WinRM)，由 Windows Management Framework (WMF) 提供。

在 PowerShell 上运行远程会话，本地计算机和远程计算机必须满足以下要求：

- Windows PowerShell 3.0 或更高版本（推荐 WMF 5.1）。
- Microsoft .NET Framework 4.0 或更高版本。
- Windows 远程管理 3.0 或更高版本。

如果要在 Windows PowerShell 2.0 上运行远程会话，本地计算机和远程计算机必须满足以下要求：

- Windows PowerShell 2.0 或更高版本。
- Microsoft .NET Framework 2.0 或更高版本。
- Windows 远程管理 2.0版本。

在 PowerShell 3.0及更高版本上运行的功能（如会话断开和重连），仅在两台计算机都使用 PowerShell 3.0 或以上版本时可用。

使用以下命令可以检查 PowerShell 版本：
`$PSVersionTable`

2.9.2 准备工作

配置 PowerShell 执行策略

检查执行策略，确保PowerShell允许运行脚本。打开PowerShell，输入以下命令检查当前执行策略：

```
Get-ExecutionPolicy
```

以下是不同的返回结果说明：

- **Restricted**：不允许执行任何脚本。
- **AllSigned**：只能运行由可信发布者签名的脚本。
- **RemoteSigned**：允许运行本地创建的脚本，但远程下载的脚本必须经过签名。
- **Unrestricted**：允许运行所有脚本，但会在运行从互联网下载的脚本时发出警告。
- **Bypass**：不阻止任何脚本的执行，不会显示任何警告或提示。

如果执行策略为**Restricted**或**AllSigned**，请通过以下命令临时更改策略，以允许运行本地创建的脚本和已签名的远程脚本。

```
Set-ExecutionPolicy RemoteSigned -Scope Process
```

该命令仅在当前PowerShell会话中更改执行策略，并在会话结束后恢复为默认策略。

CSV 文件准备

创建一个包含源端主机信息的CSV文件，确保文件能被访问，并且格式符合要求。格式如下：

```
username,ip,port,password
```

- **username**：该字段表示登录源端主机时所使用的用户名。为了确保脚本具有足够的权限执行修改操作，通常应使用具有管理员权限的用户名，如Administrator。
- **ip**：该字段表示源端主机的内网IP地址。
- **port**：该字段表示WinRM服务监听的端口号（默认是5985）。
- **password**：该字段表示登录源端主机时所需的密码。脚本将使用此密码通过WinRM协议自动登录到源端主机。

注意

- CSV文件的第一行是标题行。
- CSV文件的每一行对应一台主机信息，并且每一列信息之间用英文逗号分隔。
- 确保格式正确和信息准确，不要出现多余的空格、逗号或无效的IP地址等。

示例：

```
username,ip,port,password  
Administrator,192.168.1.10,xx,examplePass123  
Administrator,192.168.1.11,xx,examplePass456
```

Hosts 文件准备

创建一个文本文件，包含需要追加到源端主机hosts文件中的内容，并确保该文件能被访问。文件内写入相关云服务的接口域名映射，并以**#Migration-proxy-start**开始，以**#Migration-proxy-end**结束。相关云服务的接口域名映射需要根据具体的环境而定，请联系对应局点的环境接口人获取。

示例：

```
#Migration-proxy-start
xxx.xxx.xxx.xxx iam.xxx.com
xxx.xxx.xxx.xxx ecs.xxx.com
xxx.xxx.xxx.xxx evs.xxx.com
xxx.xxx.xxx.xxx ims.xxx.com
xxx.xxx.xxx.xxx obs.xxx.com
xxx.xxx.xxx.xxx eps.xxx.com
xxx.xxx.xxx.xxx vpc.xxx.com
#Migration-proxy-end
```

日志目录准备

- **配置日志目录：**脚本默认使用 C:\Users\Public\Hosts_Script_Logs 作为日志存储目录。目录不存在的话，脚本会自动创建，如果需要更改日志存储路径，请修改脚本中的 **\$logDir** 变量的值。
- **检查日志目录权限：**确保当前用户对日志目录有写入权限。如果权限不足，需要调整目录权限或选择其他路径。

检查网络连通性

- **网络连接检查：**确认执行脚本的本地计算机能够通过网络访问所有源端Windows主机的IP地址和端口。源端Windows主机需要对执行脚本的主机开放 **5985**端口。
- **防火墙配置：**检查并配置本地计算机和源端主机的防火墙，确保允许通过WinRM（Windows远程管理）连接进行远程PowerShell会话。
- **启用WinRM服务：**确保所有源端Windows主机上的WinRM服务已启用并正常运行。可以在源端主机上使用以下命令启用WinRM：
Enable-PSRemoting -Force

2.9.3 脚本示例

2.9.3.1 配置 update_hosts_win.ps1 脚本

根据提供的脚本示例，结合实际情况，修改示例中的相关配置。

前提条件

已完成[准备工作](#)。

操作步骤

- 步骤1** 在执行脚本的主机上创建一个名为“**update_hosts_lwin.ps1**”的文件，然后将以下脚本示例的内容复制到文件中。

```
# Configuration
# Path to the CSV file with server information. Must exist before running the script.
$csvFile = "C:\Users\Public\target_servers.csv" # Manually configure
# Path to the hosts content file. Must exist before running the script.
```

```

$hostsFile = "C:\Users\Public\hosts_content.txt" # Manually configure

# Directory for storing log files. Will be created if it doesn't exist.
$logDir = "C:\Users\Public\Hosts_Script_Logs" # Automatically created

# Log file for general run information.
$runLog = Join-Path $logDir "run.log" # Automatically created

# Log file for error messages.
$errorLog = Join-Path $logDir "error.log" # Automatically created

# Log file for summary information.
$summaryLog = Join-Path $logDir "summary.log" # Automatically created

# Initialize log directory and files
function Initialize-Logs {
    if (-not (Test-Path $logDir)) {
        New-Item -Path $logDir -ItemType Directory
    }
    Add-Content -Path $runLog -Value "======"
    Add-Content -Path $runLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting new
update execution"
    Add-Content -Path $runLog -Value "======"

    Add-Content -Path $errorLog -Value "======"
    Add-Content -Path $errorLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting
new update execution"
    Add-Content -Path $errorLog -Value "======"

    Add-Content -Path $summaryLog -Value "======"
    Add-Content -Path $summaryLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') -
Starting new update execution"
    Add-Content -Path $summaryLog -Value "======"
}

# Log info function
function Log-Info {
    param (
        [string]$message
    )
    $logMessage = "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Write-Output $logMessage
}

# Log error function
function Log-Error {
    param (
        [string]$message
    )
    $logMessage = "[ERROR] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Add-Content -Path $errorLog -Value $logMessage
    Write-Output $logMessage
}

# Read server information from CSV file
function Read-ServersFromCSV {
    param (
        [string]$csvFile
    )
    if (-not (Test-Path $csvFile)) {
        Log-Error "CSV file '$csvFile' not found."
        exit 1
    }
    return Import-Csv -Path $csvFile
}

```

```

# Read hosts content from TXT file
function Read-HostsContentFromTXT {
    param (
        [string]$hostsFile
    )
    if (-not (Test-Path $hostsFile)) {
        Log-Error "Hosts content file '$hostsFile' not found."
        exit 1
    }
    return Get-Content -Path $hostsFile -Raw
}

# Add to TrustedHosts
function Add-ToTrustedHosts {
    param (
        [string]$ip
    )
    # Check current TrustedHosts list
    $trustedHostsPath = "WSMan:\localhost\Client\TrustedHosts"
    $trustedHosts = (Get-Item $trustedHostsPath).Value

    if ($trustedHosts -eq $null -or $trustedHosts -eq "") {
        # Set the initial trusted host
        Set-Item $trustedHostsPath -Value $ip -Force
        Log-Info "Set initial TrustedHosts value to $ip"
    } elseif ($trustedHosts -notlike "**$ip*") {
        # Add new IP to TrustedHosts if not already present
        $updatedTrustedHosts = if ($trustedHosts -eq "") { $ip } else { "$trustedHosts,$ip" }
        try {
            Set-Item $trustedHostsPath -Value $updatedTrustedHosts -Force
            Log-Info "Added $ip to TrustedHosts"
        } catch {
            Log-Error "Failed to add $ip to TrustedHosts: $_"
        }
    } else {
        Write-Host "TrustedHosts list already contains IP $ip."
    }
}

# Initialize log files
Initialize-Logs

# Verify CSV file
if (-not (Test-Path $csvFile)) {
    Log-Error "CSV file '$csvFile' not found."
    exit 1
}

# Verify hosts file
if (-not (Test-Path $hostsFile)) {
    Log-Error "Hosts content file '$hostsFile' not found."
    exit 1
}

# Read server information from CSV file
$servers = Read-ServersFromCSV -csvFile $csvFile

# Read hosts content from TXT file
$hostsContent = Read-HostsContentFromTXT -hostsFile $hostsFile

# Counters for success and failure
$successCount = 0
$failureCount = 0
$failedServers = @()

# Remote script block
$remoteScriptBlock = {
    param (
        [string]$hostsContent
    )
}

```

```

)
$hostsFilePath = 'C:\Windows\System32\drivers\etc\hosts'

# Read the file content
$content = Get-Content -Path $hostsFilePath

# Initialize flag
$inBlock = $false
$newContent = @()

# Traverse file content
foreach ($line in $content) {
    if ($line -match '#Migration-proxy-start') {
        $inBlock = $true
    }
    if (-not $inBlock) {
        $newContent += $line
    }
    if ($line -match '#Migration-proxy-end') {
        $inBlock = $false
        continue
    }
}

# Remove trailing empty lines
while ($newContent[-1] -eq "") {
    $newContent = $newContent[0..($newContent.Count - 2)]
}

# Write the new content back to the file
$newContent | Set-Content -Path $hostsFilePath

# Append new Migration-proxy section
Add-Content -Path $hostsFilePath -Value $hostsContent

Write-Output 'Successfully updated hosts file on remote server.'
}

# Main script logic
Log-Info "Script execution started."

foreach ($server in $servers) {
    $username = $server.username
    $ip = $server.ip
    $password = $server.password

    if (-not $username -or -not $ip -or -not $password) {
        Log-Error "Invalid server entry: $username, $ip, $password. Skipping."
        continue
    }
}

Log-Info "Starting update for $username@$ip"

$securePassword = ConvertTo-SecureString $password -AsPlainText -Force
$credential = New-Object System.Management.Automation.PSCredential ($username, $securePassword)

Add-ToTrustedHosts -ip $ip

try {
    $session = New-PSSession -ComputerName $ip -Credential $credential -ErrorAction Stop

    Invoke-Command -Session $session -ScriptBlock $remoteScriptBlock -ArgumentList $hostsContent

    Remove-PSSession -Session $session
    Log-Info "Updated hosts on $ip successfully"
    $successCount++
}
catch {
    Log-Error "Failed to update hosts on $ip: $_"
}

```

```

    $failedServers += "$username@${ip}"
    $failureCount++
  }
}

# Calculate failure and success percentages
$totalCount = $servers.Count
if ($totalCount -gt 0) {
    $failurePercentage = [math]::Round(($failureCount / $totalCount) * 100, 2)
    $successPercentage = [math]::Round(($successCount / $totalCount) * 100, 2)
} else {
    $failurePercentage = 0
    $successPercentage = 100
}

# Output summary result and log to file
$summaryContent = @"
=====
[SUMMARY] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Execution Update Summary
=====
Total number of servers: $totalCount
Number of successful updates: $successCount
Number of failed updates: $failureCount
Success rate: $successPercentage%
Failure rate: $failurePercentage%
-----
"@

if ($failedServers.Count -gt 0) {
    $summaryContent += "Failed servers:`n"
    foreach ($server in $failedServers) {
        $summaryContent += " - $server`n"
    }
}
$summaryContent += "=====

# Output summary result to log file and terminal
$summaryContent | Add-Content -Path $summaryLog
Write-Output $summaryContent
Log-Info "Script execution completed. Check $summaryLog for summary."

```

步骤2 根据实际情况，修改脚本中的以下配置项参数：

- **\$logDir = "C:\Users\Public\Hosts_Script_Logs"**
 - **描述：** 日志目录路径，用于存放运行日志、错误日志和总结日志的文件夹路径。
 - **参数默认值：** C:\Users\Public\Hosts_Script_Logs
 - **修改建议：** 修改为当前用户具有写入权限的目录路径。
 - **修改示例：** \$logDir = "C:\Users\username\Documents\Hosts_Script_Logs"
- **\$csvFile = "C:\Users\Public\target_servers.csv"**
 - **描述：** **CSV文件**存放路径，包含源端主机信息。
 - **参数默认值：** C:\Users\Public\target_servers.csv
 - **修改建议：** 使用绝对路径，或确保相对路径是正确的。如果CSV文件路径发生变化，需要更新填写的路径。
 - **修改示例：** \$csvFile = "C:\Users\username\Documents\servers.csv"
- **\$hostsFile = "C:\Users\Public\hosts_content.txt"**
 - **描述：** **Hosts文件**存放路径，包含要追加到源端主机hosts文件中的内容。
 - **参数默认值：** C:\Users\Public\hosts_content.txt
 - **修改建议：** 使用绝对路径，或确保相对路径是正确的。

- **修改示例:** \$hostsFile = "C:\Users\username\Documents\hosts_content.txt"

步骤3 配置项参数修改完成并保存后，以管理员身份打开PowerShell窗口，使用以下命令执行脚本：

```
.\update_hosts_win.ps1
```

脚本会在终端窗口中输出日志信息，并在执行完毕后生成一个执行结果报告，可以在 \$logDir 指定目录中的 summary.log 文件中查看。

---结束

2.9.3.2 配置 rollback_hosts_win.ps1 脚本

根据提供的脚本示例，结合实际情况，修改示例中的相关配置。

前提条件

已完成[准备工作](#)。

操作步骤

步骤1 在执行脚本的主机上创建一个名为“**rollback_hosts_win.ps1**”的文件，然后将以下脚本示例的内容复制到文件中。

```
# Configuration
# Path to the CSV file with server information. Must exist before running the script.
$csvFile = "C:\Users\Public\target_servers.csv" # Manually configure

# Directory for storing log files. Will be created if it doesn't exist.
$logDir = "C:\Users\Public\Hosts_Script_Logs" # Automatically created

# Log file for general run information.
$runLog = Join-Path $logDir "run.log" # Automatically created

# Log file for error messages.
$errorLog = Join-Path $logDir "error.log" # Automatically created

# Log file for summary information.
$summaryLog = Join-Path $logDir "summary.log" # Automatically created

# Initialize log directory and files
function Initialize-Logs {
    if (-not (Test-Path $logDir)) {
        New-Item -Path $logDir -ItemType Directory
    }
    Add-Content -Path $runLog -Value "======"
    Add-Content -Path $runLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting new restore execution"
    Add-Content -Path $runLog -Value "======"

    Add-Content -Path $errorLog -Value "======"
    Add-Content -Path $errorLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting new restore execution"
    Add-Content -Path $errorLog -Value "======"

    Add-Content -Path $summaryLog -Value "======"
    Add-Content -Path $summaryLog -Value "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Starting new restore execution"
    Add-Content -Path $summaryLog -Value "======"
}

# Log info function
```

```

function Log-Info {
    param (
        [string]$message
    )
    $logMessage = "[INFO] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Write-Output $logMessage
}

# Log error function
function Log-Error {
    param (
        [string]$message
    )
    $logMessage = "[ERROR] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - $message"
    Add-Content -Path $runLog -Value $logMessage
    Add-Content -Path $errorLog -Value $logMessage
    Write-Output $logMessage
}

# Read server information from CSV file
function Read-ServersFromCSV {
    param (
        [string]$csvFile
    )
    if (-not (Test-Path $csvFile)) {
        Log-Error "CSV file '$csvFile' not found."
        exit 1
    }
    return Import-Csv -Path $csvFile
}

# Add to TrustedHosts
function Add-ToTrustedHosts {
    param (
        [string]$ip
    )
    # Check current TrustedHosts list
    $trustedHostsPath = "WSMan:\localhost\Client\TrustedHosts"
    $trustedHosts = (Get-Item $trustedHostsPath).Value

    if ($trustedHosts -eq $null -or $trustedHosts -eq "") {
        # Set the initial trusted host
        Set-Item $trustedHostsPath -Value $ip -Force
        Log-Info "Set initial TrustedHosts value to $ip"
    } elseif ($trustedHosts -notlike "**$ip*") {
        # Add new IP to TrustedHosts if not already present
        $updatedTrustedHosts = if ($trustedHosts -eq "") { $ip } else { "$trustedHosts,$ip" }
        try {
            Set-Item $trustedHostsPath -Value $updatedTrustedHosts -Force
            Log-Info "Added $ip to TrustedHosts"
        } catch {
            Log-Error "Failed to add $ip to TrustedHosts: $_"
        }
    } else {
        Write-Host "TrustedHosts list already contains IP $ip."
    }
}

# Initialize log files
Initialize-Logs

# Verify CSV file
if (-not (Test-Path $csvFile)) {
    Log-Error "CSV file '$csvFile' not found."
    exit 1
}

# Read server information from CSV file

```

```

$servers = Read-ServersFromCSV -csvFile $csvFile

# Counters for success and failure
$successCount = 0
$failureCount = 0
$failedServers = @()

# Remote script block
$remoteScriptBlock = {
    param ()

    $hostsFilePath = 'C:\Windows\System32\drivers\etc\hosts'

    # Read the file content
    $content = Get-Content -Path $hostsFilePath

    # Initialize flag
    $inBlock = $false
    $newContent = @()

    # Traverse file content
    foreach ($line in $content) {
        if ($line -match '#Migration-proxy-start') {
            $inBlock = $true
        }
        if (-not $inBlock) {
            $newContent += $line
        }
        if ($line -match '#Migration-proxy-end') {
            $inBlock = $false
            continue
        }
    }

    # Remove trailing empty lines
    while ($newContent[-1] -eq "") {
        $newContent = $newContent[0..($newContent.Count - 2)]
    }

    # Write the new content back to the file
    $newContent | Set-Content -Path $hostsFilePath

    Write-Output 'Successfully restored hosts file on remote server.'
}

# Main script logic
Log-Info "Script execution started."

foreach ($server in $servers) {
    $username = $server.username
    $ip = $server.ip
    $password = $server.password

    if (-not $username -or -not ${ip} -or -not $password) {
        Log-Error "Invalid server entry: $username, ${ip}, $password. Skipping."
        continue
    }

    Log-Info "Starting restore for $username@${ip}"

    $securePassword = ConvertTo-SecureString $password -AsPlainText -Force
    $credential = New-Object System.Management.Automation.PSCredential ($username, $securePassword)

    Add-ToTrustedHosts -ip $ip

    try {
        $session = New-PSSession -ComputerName ${ip} -Credential $credential -ErrorAction Stop

        Invoke-Command -Session $session -ScriptBlock $remoteScriptBlock
    }
}

```

```

Remove-PSSession -Session $session
Log-Info "Restored hosts on ${ip} successfully"
$successCount++
}
catch {
Log-Error "Failed to restore hosts on ${ip}: $_"
$failedServers += "$username@${ip}"
$failureCount++
}
}

# Calculate failure and success percentages
$totalCount = $servers.Count
if ($totalCount -gt 0) {
    $failurePercentage = [math]::Round(($failureCount / $totalCount) * 100, 2)
    $successPercentage = [math]::Round(($successCount / $totalCount) * 100, 2)
} else {
    $failurePercentage = 0
    $successPercentage = 100
}

# Output summary result and log to file
$summaryContent = @"
=====
[SUMMARY] $(Get-Date -Format 'yyyy-MM-dd HH:mm:ss') - Execution Rollback Summary
=====
Total number of servers: $totalCount
Number of successful restores: $successCount
Number of failed restores: $failureCount
Failure rate: $failurePercentage%
Success rate: $successPercentage%
-----
"@

if ($failedServers.Count -gt 0) {
    $summaryContent += "Failed servers:`n"
    foreach ($server in $failedServers) {
        $summaryContent += " - $server`n"
    }
}
$summaryContent += "=====

# Output summary result to log file and terminal
$summaryContent | Add-Content -Path $summaryLog
Write-Output $summaryContent
Log-Info "Script execution completed. Check $summaryLog for summary."

```

步骤2 根据实际情况，修改脚本中的以下配置项参数：

- **\$logDir = "C:\Users\Public\Hosts_Script_Logs"**
 - **描述：** 日志目录路径，用于存放运行日志、错误日志和总结日志的文件夹路径。
 - **参数默认值：** C:\Users\Public\Hosts_Script_Logs
 - **修改建议：** 修改为当前用户具有写入权限的目录路径。
 - **修改示例：** \$logDir = "C:\Users\username\Documents\Hosts_Script_Logs"
- **\$csvFile = "C:\Users\Public\target_servers.csv"**
 - **描述：** **CSV文件**存放路径，包含源端主机信息。
 - **参数默认值：** C:\Users\Public\target_servers.csv
 - **修改建议：** 使用绝对路径，或确保相对路径是正确的。如果CSV文件路径发生变化，需要更新填写的路径。
 - **修改示例：** \$csvFile = "C:\Users\username\Documents\servers.csv"

步骤3 配置项参数修改完成并保存后，以管理员身份打开PowerShell窗口，使用以下命令执行脚本：

```
.\rollback_hosts_win.ps1
```

脚本会在终端窗口中输出日志信息，并在执行完毕后生成一个执行结果报告，可以在 \$logDir 指定目录中的 summary.log 文件中查看。

----结束

2.9.4 常见问题

2.9.4.1 如何启用 PowerShell 远程管理模块

通常Windows Server 2012及更高版本已默认启用PowerShell远程处理。如果更改了设置，可以通过以下方法重新启用PowerShell远程管理。

步骤1 以管理员权限启动PowerShell。

步骤2 执行如下命令，即可启用PowerShell远程管理。

```
Enable-PSRemoting
```

步骤3 验证配置。在PowerShell中输入如下命令：

```
New-PSSession
```

- 如果配置成功，则该命令会在本地计算机上创建会话，并返回一个会话对象。返回示例：

```
Id Name      ComputerName State ConfigurationName
-----
1 Session1 localhost    Opened  Microsoft.PowerShell
```

- 如果配置失败，请参考 PowerShell 帮助文档中的 [about_Remote_Troubleshooting](#) 获取解决方案。

----结束

2.9.4.2 如何启用 WinRM 服务

步骤1 以管理员权限启动PowerShell。

步骤2 执行如下命令，即可自动启动 WinRM 服务并配置其远程访问设置。

```
Enable-PSRemoting -Force
```

----结束

2.9.4.3 执行脚本后报错：连接到远程服务器失败，且登录凭证信息正确

问题现象

执行脚本后报错，错误信息为 “[ERROR] Failed to update hosts on xxx.xxx.x.xx: [xxx.xxx.x.xx] 连接到远程服务器 xxx.xxx.x.xx 失败有关详细信息，请参阅 about_Remote_Troubleshooting 帮助主题”。

问题分析

出现该报错是因为：远程服务器的WinRM服务未启动或者配置不正确。

解决方法

- 步骤1** 验证执行机能否 ping 通远程服务器。如果可以ping通，请执行步骤2。
- 步骤2** 确认远程服务器的 WinRM 服务5985端口是否已对执行机开放。确认开放请执行步骤3。
- 步骤3** 检查远程服务器的登录凭证是否正确。确认凭证正确无误后，请执行后续步骤。
- 步骤4** 登录连接失败的远程服务器，以管理员身份打开 PowerShell。
- 步骤5** 执行如下命令，然后重新执行脚本。

```
Enable-PSRemoting -Force
```

----结束

3 存储迁移相关最佳实践

3.1 迁移其他云平台存储数据至华为云

适用场景

通过独立专享迁移集群以及配置迁移专线，简单、快捷实现对象存储、文件存储一站式上云，提升上云效率。本文介绍如何使用迁移中心MgC将其他云平台的对象存储或文件存储数据迁移到华为云。

当前支持以下区域（Region）：

- 拉美-圣地亚哥
- 拉美-圣保罗
- 土耳其-伊斯坦布尔
- 亚太-曼谷
- 亚太-新加坡
- 亚太-雅加达
- 中东-利雅得
- 华北-北京四
- 华东-上海一

优势

- 专享迁移集群，减少资源竞争，提高迁移效率。
- 支持多种云平台对象存储、文件存储服务，支持自建网络文件服务。支持迁移的云平台 and 存储类型包括：
 - 华为云 OBS
 - 阿里云 OSS
 - 百度云 BOS
 - 腾讯云 COS
 - 金山云 KS3

- 七牛云 KODO
- 优刻得 US3
- 亚马逊 S3
- 微软云 Blob
- NAS_GFS
- NAS_SMB
- NAS_NFS_V3_MOUNT
- NAS_NFS_V3_PROTOCOL
- HTTP/HTTPS数据源

约束与限制

- 关于存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些?](#)
- 源端为阿里云OSS：当源端对象的存储类型是深度冷归档存储时，无法使用MgC提供的解冻归档数据功能直接迁移，需要在源端先手动解冻后再迁移。

逻辑架构图

通过迁移中心进行存储数据迁移的逻辑架构图如图3-1所示。

图 3-1 存储迁移逻辑架构图



准备工作

- 使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，获取方法请参见[准备工作](#)。
- 在MgC控制台[创建项目](#)。
- 确保源端账号和目的端账号拥有迁移需要的权限，详细介绍请参见[如何获取源端桶和目的端桶权限?](#)。
- （可选）涉及专线场景需要提前打通VPC或者保证在目的端ECS上能够ping通源端内网域名和内网桶域名。

步骤一：创建目的端

迁移中心支持将对象存储数据迁移到弹性文件系统，也支持将文件存储数据迁移到 OBS桶。请根据实际需求，在华为云提前[创建目的端对象桶](#)或[创建弹性文件系统](#)。

步骤二：创建集群

迁移集群是专为存储工作流提供协作中的核心组件，通过集群可以创建和管理迁移节点、列举节点，部署和升级迁移插件，确保存储工作流顺利运行。详细说明和创建步骤请参见[创建集群](#)。

如果是专线迁移场景，需要在高级设置中添加域名映射配置，对域名和桶域名同时绑定源端专线内网ip。

图 3-2 高级设置

IP地址	域名	操作
<input type="text"/>	<input type="text"/>	<input type="button" value="确定"/> <input type="button" value="取消"/>

步骤三：创建存储迁移工作流

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“[实施>迁移工作流](#)”，进入迁移工作流页面。在页面左上角的当前项目下拉列表中选择[迁移项目](#)。

步骤3 单击页面右上角的“[创建迁移工作流](#)”按钮，进入创建迁移工作流页面。

步骤4 选择[存储迁移标准模板](#)，单击“[开始配置](#)”按钮，进行工作流配置。单击“[查看步骤详情](#)”，可以查看存储迁移模板所包含的迁移阶段和步骤，以及每个阶段和步骤的详细描述。其中，步骤类型为“[自动](#)”，表示该步由系统自动完成。

步骤5 如[图3-3](#)所示，根据[表3-1](#)，配置基础信息参数。

图 3-3 基础信息

基础信息
请填写名称、描述，并选择区域、集群

* 名称

* 区域

描述 0/255

* 集群 [去新建](#)

表 3-1 基础信息参数说明

参数	说明
名称	用户自定义。
区域	下拉菜单中选择目的端所在区域。需要与集群所在区域一致。
描述	用户自定义。
集群	选择 步骤二创建的迁移集群 。

步骤6 如[图3-4](#)所示，根据[表3-2](#)、[表3-3](#)，配置源端、目的端信息。

图 3-4 源端/目的端信息

源端信息
请填写源端信息，根据不同站点类型配置不同参数

* 数据站点类型

* AK

* 桶

* SK

* 对象存储终端地址

目的端信息
请填写目的端信息，根据不同站点类型配置不同参数

* 数据站点类型

* AK

* 桶

* SK

对象存储终端地址

指定前缀

表 3-2 源端信息参数说明

参数	说明	备注
数据站点类型	<p>目前支持迁移的源端存储类型包括：</p> <ul style="list-style-type: none"> ● 华为云 OBS ● 阿里云 OSS ● 百度云 BOS ● 腾讯云 COS ● 金山云 KS3 ● 七牛云 KODO ● 优刻得 US3 ● 亚马逊 S3 ● 微软 Blob ● NAS_GFS ● NAS_SMB ● NAS_NFS_V3_MOUNT ● NAS_NFS_V3_PROTOCOL ● HTTP/HTTPS数据源 	-
AK、SK	输入源端云账号的访问密钥和私有密钥。请确保输入的AK、SK正确，并拥有源端桶的相应权限。	数据站点类型选择云平台时填写。
桶	输入源端需要迁移的桶名。	
对象存储终端地址	输入源端桶所在区域的终端节点（Endpoint）信息。	
桶类型	根据桶的实际类型选择。桶类型可以通过 查看桶的基本信息 获取。	数据站点类型为华为云OBS时选择。
appid	<p>填写腾讯云APPID。</p> <p>说明 您可以在腾讯云控制台账号信息页面查看APPID。</p>	数据站点类型为腾讯云COS时填写
鉴权类型	<p>鉴权类型的选择支持以下两种：</p> <ul style="list-style-type: none"> ● 连接字符串。 ● 存储账户名称+密钥。 <p>获取以上两种鉴权类型的方式请参考：如何获取微软云的访问凭证。</p>	数据站点类型为微软云BLOB时配置。

参数	说明	备注
列表路径	<p>输入存放列表文件的路径。列表文件存放地址必须与目的端桶处于同一区域。</p> <p>将源端待迁移的文件URL和对象名称写入列表文件（每行只能写一个URL和对象名称），列表文件需要提前创建。</p> <p>列表文件约束与限制：</p> <ul style="list-style-type: none"> 列表文件类型必须为.txt，其他文件类型不做处理，并且该文件元数据中的“ContentType”只能为：“text/plain”。 txt文件行数不超过100000行。 单个列表文件大小不能超过300 MB。 列表文件存放目录下的列表文件个数不能超过10000个。 列表文件必须是UTF-8无BOM格式编码格式。 列表文件中每行长度不要超过65535，否则会导致迁移失败。 列表文件的元数据中不能设置“ContentEncoding”，否则会导致迁移失败。 列表文件中每行使用制表符（键盘上Tab键）\t分隔URL和目的端对象名称，请勿使用空格。格式为：[URL][制表符][目的端对象名称]，其中源端对象名称如果包含中文、特殊字符必须使用URL Encode对URL编码；目的端对象名称如果包含中文、特殊字符也需要使用URL Encode编码。 列表文件中每行不要添加无效空格，否则会将空格作为对象名，导致迁移失败。 	<p>数据站点类型选择HTTP/HTTPS数据源时填写。</p>
文件存储服务地址	<p>输入源端文件系统的挂载地址，格式为：IP地址+:/ xxx，例如：192.1.1.1:/0001</p>	<p>数据站点类型选择NAS_SMB/NAS_NFS_V3_MOUNT/NAS_NFS_V3_PROTOCOL时填写。</p>
路径	<p>输入待迁移文件路径，格式为：/文件夹。</p>	
用户名	<p>输入具有访问源端文件系统所有文件权限的用户名，例如：administrator。</p>	<p>数据站点类型选择NAS_SMB时填写。</p>
密码	<p>输入用户名对应的密码。</p>	

参数	说明	备注
Windows上的域	输入对应的域名。 说明 只需输入.com之前的内容，例如：域名为test.com，则输入test。	

表 3-3 目的端信息参数说明

参数	说明	备注
数据站点类型	根据源端存储类型，选择华为云对应的存储类型。	-
AK	输入目的端华为云账号的访问密钥。	数据站点类型选择 华为云OBS 时填写。
SK	输入目的端华为云账号的私有密钥。	
桶	选择目的端OBS桶。	
对象存储终端地址	输入目的端OBS桶所在区域的终端节点（Endpoint）信息。 说明 可在OBS桶的基本信息查看。	
指定前缀	将对象迁移到指定前缀，例如：源端对象路径为/A/B/C.txt，目的端前缀配置为/D，迁移后目的端对象路径为/D/A/B/C.txt。详细介绍请参考 设置“保存到指定前缀”参数 中的功能介绍。	
文件存储服务地址	输入目的端文件系统的挂载地址，格式为：IP地址+:/ xxx，例如：192.1.1.1:/0001	数据站点类型选择 NAS_SMB/NAS_NFS_V3_MOUNT 时填写。
路径	输入目的端接收文件的路径，格式为：/文件夹。	
用户名	输入具有访问目的端文件系统所有文件权限的用户名，例如：administrator。	数据站点类型选择 NAS_SMB 时填写。
密码	输入用户名对应的密码。	
Windows上的域	输入对应的域名。 说明 只需输入.com之前的内容，例如：域名为test.com，则输入test。	

步骤7 如图3-5所示，根据表3-4，进行迁移配置。

图 3-5 迁移配置

迁移配置
请选择任务类型、并发子任务以及同名覆盖方式。选择是否开启迁移元数据以及清理场景

* 迁移任务类型 全量迁移 列表迁移 前缀迁移

* 列举方式 串行列举 并行列举
源端非PFS桶的默认列举方式

* 并发子任务数 ①

* 同名覆盖方式 ① 条件覆盖 不覆盖 直接覆盖
覆盖大小不相同或源端最后修改时间大于目的端最后修改时间的对象。

迁移元数据
为保证迁移任务正常运行，仍将为您迁移ContentType元数据

清理集群
勾选后将自动生成清理集群步骤，您可以继续配置额外要删除的配置，包括NAT网关、安全组等

表 3-4 迁移配置参数说明

参数	子参数	说明
迁移任务类型	全量迁移	迁移源端桶/文件夹内的全部数据。

参数	子参数	说明
	列表迁移	<p>通过对象列表文件过滤迁移，只迁移对象列表文件中的对象。</p> <p>在“列表路径”输入存放在目的端桶的对象列表文件路径。对象列表文件的约束与限制如下：</p> <ul style="list-style-type: none"> ● 目的端为OBS <ul style="list-style-type: none"> - 对象列表文件大小不能超过30 MB。 - 对象列表文件类型必须为.txt类型，并且元数据ContentType只能为text/plain。 - 对象列表文件必须是UTF-8无BOM格式编码格式。 - 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。 - 对象列表文件行长度不超过16KB，否则迁移失败。 - 对象列表文件元数据中不能设置ContentEncoding，否则迁移失败。 - 对象列表文件行数不超过10000行。 ● 目的端为NAS <ul style="list-style-type: none"> - 对象列表文件大小不能超过30 MB。 - 对象列表文件类型必须为.txt类型。 - 对象列表文件必须是UTF-8无BOM格式编码格式。 - 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。 - 对象列表文件行长度不超过16KB，否则迁移失败。 - 对象列表文件行数不超过10000行。
	前缀迁移	<p>仅支持云平台数据类型。</p> <p>在“前缀路径”输入对象名称或者名称前缀，则只迁移与输入对象名称或名称前缀一致的对象。</p> <p>须知</p> <ul style="list-style-type: none"> ● 输入对象前缀路径时，如果对象文件在源端桶的根目录下，则直接添加前缀；如果对象文件存在目录路径，则需要对对象前缀前添加对象文件所在的目录路径。格式为：文件夹/对象前缀 ● 多个前缀之间用英文逗号分隔。
并发子任务数	-	<p>用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。</p>

参数	子参数	说明
同名覆盖方式	不覆盖	<p>对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。</p> <p>警告</p> <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	<p>对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。</p>
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； <p>说明 对于空文件夹，覆盖策略与文件相同。</p>
迁移元数据	-	<p>选择是否迁移元数据。</p> <ul style="list-style-type: none"> 勾选，则迁移源端对象元数据。 不勾选，仅迁移Content-Type、Content-Encoding元数据。
清理集群	-	<p>选择工作流是否创建清理集群步骤。</p> <ul style="list-style-type: none"> 勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPCEP服务等。 不勾选，迁移工作流不创建清理集群步骤。

步骤8（可选）如图3-6所示，根据表3-5，配置高级选项。

图 3-6 高级选项




表 3-5 高级选项参数说明

参数	说明	备注
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。	-
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。 	仅目的端选择 华为云OBS 时支持配置。
解冻归档数据	<ul style="list-style-type: none"> 不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。 勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。 <p>说明</p> <p>在数据迁移时，系统会对所有数据发起取回操作，源端云服务商收取相应的API调用费用和存储费用。</p>	-
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。	

参数	说明	备注
CDN迁移支持	<p>当默认域名不能满足您的迁移需求，比如您需要一个自定义域名以便使用CDN降低源端出口下载流量的费用，可以勾选此选项。在域名输入框中输入自定义域名，并选择传输协议。HTTP协议安全级别低于HTTPS协议，推荐使用HTTPS协议。</p> <p>源端为阿里云OSS、腾讯云COS时，如果使用了CDN鉴权，还需要选择“鉴权类型”并输入“鉴权密钥”。</p>	
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。 	
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 <p>例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</p> <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。 	
定时迁移	<p>设置迁移任务在指定时间内自动迁移。</p> <ul style="list-style-type: none"> 不勾选，手动启动/暂停迁移任务。 勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。 <p>例如：</p> <ul style="list-style-type: none"> 设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 	

步骤9 workflow配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

- 步骤10** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。
- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
 - 如果您需要在标准工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。
- 步骤11** 在工作流详情页面，可以查看该存储迁移工作流的**配置信息**和**资源迁移进度**。还可以进行如下操作：
- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
 - 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
 - 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
 - 设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
 - 节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。
- 步骤12**（可选）单击资源迁移进度条或进度条弹窗中的“**资源迁移进度详情**”，右侧弹出该资源迁移详情页面，可以查看**任务概览**和**进度详情**等信息。还可以进行如下操作：

操作	步骤
更换集群	<p>只有OMS任务状态（即工作流状态）为“已暂停”时，可以进行“更换集群”操作。</p>  <p>1. 在任务概览区域，单击迁移集群名称后的“更换”。</p> <p>2. 在下拉列表中选择新集群，单击“确认”，更换集群后，系统开始自动运行。</p>
修改定时迁移	<p>1. 在任务概览区域，单击定时迁移后的“修改”。</p> <p>2. 设置开始时间和结束时间，单击“确认”。</p>
修改并发子任务数	<p>1. 在进度详情区域，单击期望并发子任务数下方的“修改”，配置期望并发的子任务数。配置要求：配置的数目不能超过可用迁移节点数x10。例如，可用的迁移节点数为2，则配置数目需要≤20。</p> <p>2. 单击“确认”，完成并发子任务数修改。</p>

操作	步骤
添加迁移限速规则	<p>1. 在迁移速度区域，单击“添加”按钮，可以增加一条限速规则，用于限制该资源在某个时间段内的迁移速度。</p> <p>须知</p> <ul style="list-style-type: none"> - 限速取值范围：1 MB~1024 GB。 - 创建多条限速规则时，限速时段不能重叠。 例如：已有一条规则，限速时段是8:00~12:00，那么7:00~13:00，7:00~8:00，9:00~12:00...有时间重叠的规则，都无法创建成功。 - 开始时间不能晚于等于结束时间。 例如：不能设置时间段为23:00~01:00。 <p>2. 设置完成后，单击“保存”，完成添加。</p>
获取迁移失败/跳过/成功资源列表文件	<p>在文件统计区域，可以查看存放迁移失败/跳过/成功的资源列表文件路径。单击路径，可以进入存放文件的OBS桶，获取对应列表文件。</p>
查看迁移流量统计	<p>在流程统计区域，可以查看该资源近1小时、近6小时、近24小时、整个周期迁移流量使用情况。</p>

----结束

过滤策略说明

过滤策略参数填写通配符规则说明和约束说明参见[表3-6](#)。

表 3-6 过滤策略参数说明

参数	说明	填写规则	约束说明
黑名单	命中黑名单规则，则文件不进行迁移/一致性对比。支持精确匹配和模糊匹配。	<ul style="list-style-type: none"> 精确匹配 全路径精确匹配，填写文件绝对路径，特殊字符用反斜杠 (\) 转义。 	<ul style="list-style-type: none"> 除{和}以外，不能出现连续的规则符号，例如：<code>***</code>、<code>*?、**?、?*</code>、<code>*、?***</code>、<code>*{*、*}*、*}*?、?{*、*}</code>、<code>{*}</code>、<code>{*}</code>、<code>{*、*}</code>、<code>*,*</code>。 {和}之间的模式，只能使用*做通配符。 {和}之间的模式，不能有{}嵌套。 文件绝对路径在黑名单和白名单同时命中，则黑名单生效，即不做迁移。 需要配置多个匹配规则时，用英文分号隔开。
白名单	<ul style="list-style-type: none"> 白名单为空时，迁移所有文件。 白名单不为空时，仅对绝对路径命中白名单规则的文件进行迁移/一致性对比。 	<p>注意 配置黑、白名单注意事项：</p> <ul style="list-style-type: none"> 如果文件存储服务地址是以冒号加斜杠 (:/) 结尾，请确保在黑、白名单中填写的路径是相对于挂载地址的绝对路径。 例如，文件存储服务地址为 "192.1.1.1:/", 挂载地址为 "/mnt/turbo", 则填写 "/mnt/turbo" 下的绝对路径。 挂载信息示例： <pre>[root@oms-cluster-ecs filter_test]# pwd /mnt/sts_turbo/mgc/filter_test [root@oms-cluster-ecs filter_test]# ll drwxr-xr-x 2 root root 0 Aug 16 15:27 test2 -rw-r--r-- 1 root root 5 Aug 16 15:27 test2.log</pre> 如果要使黑名单精确匹配以避免迁移 "test2.log" 文件，则应在黑名单中填写路径：<code>/mgc/filter_test/test2.log</code> 如果文件存储服务地址在冒号加斜杠 (:/) 之后还包含其他路径，配置黑、白名单时，除了填写挂载地址下的绝对路径外，还要加上文件存储服务地址:/之后的部分。 例如，文件存储服务地址为 "192.1.1.1:/mgc-test", 挂载地址为 "/mnt/turbo", 则填写 "mgc-test+/mnt/turbo" 下的绝对路径。 挂载信息示例： <pre>[root@oms-cluster-ecs execution-service]# cd /mnt/turbo/autotest/filter_test/ [root@oms-cluster-ecs filter_test]# ll -rw-r--r-- 1 root root 14 Aug 8 09:22 test1.log drwxr-xr-x 1 root root 4096 Aug 8 09:22 test2 -rw-r--r-- 1 root root 14 Aug 8 09:22 test2.log</pre> 如果要使黑名单精确匹配以避免迁移 "test2" 文件，则应在黑名单中填写路径：<code>/mgc-test/autotest/filter_test/test2</code> <ul style="list-style-type: none"> 模糊匹配 	

参数	说明	填写规则	约束说明
		<ul style="list-style-type: none"> - *匹配任何字符0个或者多个，但不匹配斜杠(/) - 使用**匹配任何字符0个或者多个，包括斜杠(/) - ?匹配除斜杠(/)之外的任何字符，只能匹配一个 - {和}定义元素之间的选择：包含一个以逗号分隔的模式列表，其中任何一种模式都可能匹配，可以包含通配符。 - 特殊字符可以在它们之前使用\进行转义，\后面为通配符时，表示转义，否则表示\本身 	
时间段	配置时间段，根据配置，迁移最后修改时间在配置时间段内的文件或者文件夹。 开始时间和结束时间可以配置为空，为空时表示不限制，时间可以精确到分钟。		

过滤策略场景规则示例以及举例见下表。

须知

例如第一级目录为“test”，那么：

- NAS的路径格式为：“/test”。
- OBS的路径格式为：“test”。

场景	NAS规则示例	OBS规则示例	说明
文件路径以xx结尾(特定后缀)	/xx /**xx	xx/**xx	xx可以是包含*和?的表达式
根路径下第一层xx开始的文件	/xx* • /testssss 匹配 • /test/xx 不匹配	/xx* • testssss 匹配 • test/xx 不匹配	
文件路径以xx开始	/xx** /xx/**	xx** xx/**	
文件名包含xx	**xx*	**xx*	
文件路径包含xx	**xx**	**xx**	

场景	NAS规则示例	OBS规则示例	说明
文件路径以xx开始，以yy结尾	/xx**yy	xx**yy	xx, yy可以是包含*和?的表达式
文件路径以xx, yy结尾(特定后缀)	**{xx,yy}	**{xx,yy}	
文件名包含xx, yy	**{xx,yy}*	**{xx,yy}*	
文件路径包含xx, yy	**{xx,yy}**	**{xx,yy}**	

3.2 OBS 多前缀批量迁移

本文将详细介绍如何使用前缀列表文件（包含多个前缀的列表文件）来筛选并批量迁移OBS桶内的对象。

准备工作

- 使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，获取方法请参见[准备工作](#)。
- 在MgC控制台[创建项目](#)。
- 将源端云平台账号对应的AK/SK添加至MgC，作为采集源端桶的凭证。认证方式选择AK/SK，添加方法请参考[添加资源凭证](#)。
- 确保源端账号和目的端账号拥有迁移需要的权限，详细介绍请参见[如何获取源端桶和目的端桶权限？](#)。
- 在华为云创建用于接收源端对象存储数据的OBS桶（即目的端桶），创建方法请参考[创建桶](#)。如果您已有目的端桶，则无需重新创建。
- 为每个待迁移的源端桶，创建一个对应的前缀列表文件（即一个桶对应一个前缀列表文件），前缀列表文件要求如下：
 - 文件格式为.txt，大小不超过2 MB。
 - 每行只能包含一个前缀，每个前缀的长度不超过1024个字符。
 - 前缀总数（即行数）不超过1000。

使用须知

- **支持的区域**
当前支持在以下区域（Region）进行批量迁移，其它Region请使用线下工具RDA或者咨询其它方案进行迁移。
 - 拉美-圣地亚哥
 - 拉美-圣保罗
 - 土耳其-伊斯坦布尔
 - 亚太-曼谷
 - 亚太-新加坡
 - 亚太-雅加达
 - 中东-利雅得

- 华北-北京四
- 华东-上海一
- **内网迁移**
支持在同一Region内通过内网进行迁移。
- **软连接处理**
MgC不支持源端路径为软链接的迁移。如果源端包含软链接，请进行如下处理：
 - 填写实际的文件路径。
 - 迁移完成后，手动在目的端创建相应的软链接。

步骤一：发现并采集源端桶

步骤1 登录[迁移中心管理控制台](#)。

步骤2 单击左侧导航栏的“[调研>资源采集](#)”，进入[资源采集](#)页面。在页面左上角的当前项目下拉列表中选择在完成[准备工作](#)时所创建的迁移项目。

步骤3 单击任务框中的“[发现>公网发现](#)”，进入[创建公网发现任务](#)页面。

图 3-7 公网发现



步骤4 根据[表3-7](#)，完成[基础配置](#)和[任务配置](#)区域的参数配置。

基础配置
请填写任务名称与任务说明

任务名称

任务说明 (可选)

任务配置
请选择源端平台、凭证及资源所在区域

源端

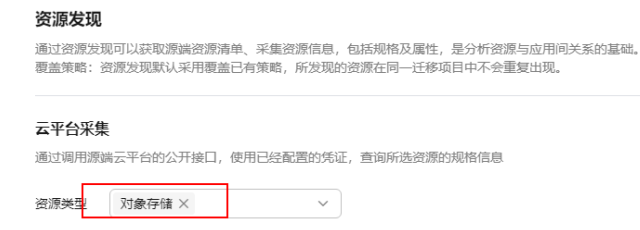
凭证
源端平台访问口令，用于发现源端资源

区域

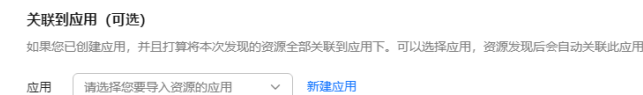
表 3-7 基础配置和任务配置区域参数配置说明

区域	参数	说明	是否必选
基础配置	任务名称	自定义发现任务名称。	是
	任务说明	输入发现任务说明。	否
任务配置	源端	选择源端云平台，此处选择“华为云”。	是
	凭证	选择在完成准备工作时添加的源端账号凭证，如未提前添加凭证，请单击“新建凭证”，认证方式选择“AK/SK”，输入源端账号的AK/SK，校验并保存凭证。	是
	区域	选择要发现的源端云平台区域，支持多选。	是

步骤5 在资源发现区域的资源类型下拉列表，选择“对象存储”。源端为阿里云和腾讯云平台时，需要开启“云平台采集”，才能选择资源类型。



步骤6 将采集的对象存储资源关联到应用中。



- 如果已提前创建应用，在“应用”的下拉列表中选择要关联的应用。
- 未提前创建应用，单击“新建应用”，弹出新建应用窗口，输入自定义的应用名称和描述，选择业务场景和使用场景；区域选择目的端所在区域，单击“确定”按钮，应用创建成功。

步骤7 单击“确认”，创建公网采集对象存储资源任务完成，系统自动开始资源采集。

步骤8 在资源采集页面，单击任务总数后的“查看”，进入任务列表页面。



当任务状态为“成功”时，采集完成。



----结束

步骤二：创建集群

在创建迁移集群时会产生部分其他服务费用，详情请参见[计费说明](#)。

须知

为了保证迁移的稳定性和数据安全，创建的迁移集群不支持用户自行登录。如您有特殊需求必须登录，请联系技术支持人员。

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“部署>迁移集群”，进入迁移集群页面。

步骤3 单击页面右上角的“创建集群”按钮，进入创建集群页面。

如果是首次创建迁移集群，需要您同意华为云账号进行相关的委托授权。委托权限所包含的细粒度权限请前往[委托权限说明](#)查看。



步骤4 根据[表3-8](#)，配置参数。

表 3-8 集群参数配置说明

区域	参数	配置说明	限制说明
基础配置	集群名称	自定义名称。	同一账户下集群名称不可重名。
	区域	选择创建集群的区域。	与目的端区域保持一致。
	集群类型	选择集群使用场景。	当前仅支持存储迁移场景。
节点创建	Master节点ECS规格	用于管理迁移节点和列举节点。	节点数量为1，无法修改。
	迁移节点ECS规格	迁移节点用于执行迁移和对比的动作，推荐规格8U16G。	<ul style="list-style-type: none"> 规格选定后，后续无法更改。 设置的节点数量需要同时满足如下要求： <ul style="list-style-type: none"> - 迁移节点数+列举节点数+1 ≤ 100 - 迁移节点数+列举节点数+1 ≤ 子网可用IP数
	列举节点ECS规格	列举节点用于执行列举动作，推荐规格8U16G。	
网络配置	虚拟私有云	在下拉列表中选择虚拟私有云。	-
	子网	选择与虚拟私有云网段相同且满足可用IP数的子网。	子网的可用IP数≥迁移节点数+列举节点数+1。
	模式选择	<ul style="list-style-type: none"> 公网：使用公网迁移，并选择公网网关。如果没有可用网关，选择“新建网关”，选择网关规格和EIP（EIP最多支持同时勾选20个）。 内网：使用内网迁移，适用于同一区域内的迁移场景。 专线：使用专线迁移，通过专线直接访问源端数据。搭建云专线的方法可参考云专线 DC。 	-

区域	参数	配置说明	限制说明
高级设置	DNS配置（可选）	填写DNS服务器IP地址，用于追加/etc/resolv.conf文件中的nameserver参数值。多个DNS服务器地址使用英文逗号分隔，例如： 192.0.2.1,192.0.2.2	最多可输入3个DNS服务器地址。
	域名映射配置（可选）	添加域名到IP地址的映射，用于修改/etc/hosts文件。	最多可添加500个IP地址映射。
-	流量控制	设置迁移任务在指定时段的最大流量带宽。 <ul style="list-style-type: none"> 不勾选，则不进行流量控制 勾选，进行流量控制。在“开始时间”、“结束时间”和“限速”文本框中输入数值。 须知 例如：设置开始时间08:00、结束时间12:00、限制最大流量20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。	<ul style="list-style-type: none"> 最多可添加五条流量控制规则。 时区时间以控制台所选Region的当地时间为准。
	启用LTS服务	<ul style="list-style-type: none"> 启用：记录存储迁移过程中的日志，便于查看和定位迁移中出现的问题。 不启用：不记录存储迁移过程中的日志。 	-

步骤5 参数配置完成后，单击“确认”开始创建集群。在迁移集群列表中查看集群状态，状态详情参见[集群状态说明](#)。如果集群状态为创建失败，将鼠标移动到“创建失败”上，可以查看失败原因。在解决失败原因后，可以通过操作列的“更多 > 重试”功能，重新创建集群。

---结束

步骤三：迁移方案配置

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“设计>迁移方案配置”，进入迁移方案配置页面。单击页面右上角的“创建迁移方案”按钮，右侧弹出迁移方案模板窗口。



步骤3 单击对象存储批量迁移框中的“开始配置”按钮，进入对象存储批量迁移页面。



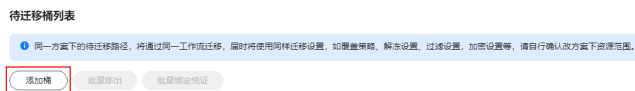
步骤4 在基础配置区域，根据表3-9，配置相关参数。



表 3-9 基础配置参数配置说明

参数	配置说明
方案名称	用户自定义。
描述（可选）	用户自定义。
源端平台	与 步骤一 采集对象存储资源时，所选的源端云平台保持一致。此处选择“华为云”。
目的端桶区域	选择目的端桶所在区域。

步骤5 在待迁移桶列表区域，单击“添加桶”按钮，右侧弹出添加桶窗口。



步骤6 勾选需要迁移的桶，单击操作列的“修改”，调整桶的迁移方式为“前缀迁移”，单击“确认”，添加该桶到待迁移桶列表中。



注意

- 所选的资源必须都来源于基础配置所选的源端平台。
- 最多添加 100 个桶。

步骤7 绑定源端凭证。



- **单个绑定源端凭证：**在待迁移桶列表中，单击每一行操作列的“修改”，在弹出的修改迁移设置窗口中选择源端凭证进行绑定。
- **批量绑定源端凭证：**在列表中勾选多个桶，单击列表上方“批量绑定凭证”按钮，批量绑定源端凭证。

步骤8 导入前缀列表文件。

在待迁移桶列表中，单击每一行操作列的“导入前缀”，弹出导入前缀窗口。将该桶对应的前缀列表文件上传后，单击“确认”按钮。



步骤9 完成所有待迁移桶的凭证绑定和前缀列表导入后，单击“下一步”，进行目的端配置。

步骤10 单击每一行操作列的“修改”，选择目的端凭证和目的端桶，输入目的端前缀（可选），单击“保存”。



步骤11 所有资源的目的端配置完成后，单击“下一步”，进行集群评估。集群评估为可选操作，您可以跳过该步骤，也可以通过评估，获取推荐的集群规格后，对创建的集群进行修改，请参考[集群管理](#)。

步骤12 单击“下一步”，进入集群配置，单击“立即绑定”，右侧弹出[集群选择](#)窗口。

步骤13 在集群列表中，选择**步骤二创建的集群**，单击“确认”，完成集群绑定，绑定后本方案下所有资源将通过此集群迁移。

须知

只有集群状态为**健康**或者**亚健康**时，可以进行选择该集群进行绑定。

步骤14 集群配置完成后，单击“确定”，迁移方案创建完成。在**迁移方案配置**页面的**方案列表**，可以查看已创建的迁移方案。

- 如果需要对迁移方案配置进行调整，可以单击操作列的“设计”进行修改。
- 当**设计进度**为**已完成**时，单击操作列的“创建 workflow”，即可**创建对象存储批量迁移 workflow**。

----结束

步骤四：创建对象存储批量迁移 workflow

须知

- 单个对象大小不能超过： $500 \text{ MB} \times 10000 = 4.76837158203125 \text{ TB}$ ，否则可能会导致迁移失败。
- 在迁移过程中，系统将自动在目的端桶创建一个名为“oms”的临时文件夹。**严禁手动操作此文件夹**，包括但不限于修改、删除或添加内容。对该文件夹进行的任何手动操作都可能导致迁移流程中断或失败。

步骤1 登录**迁移中心管理控制台**。

步骤2 在左侧导航栏选择“实施>迁移 workflow”，进入**迁移 workflow**页面。

步骤3 单击页面右上角的“创建迁移 workflow”按钮，进入**选择模板**页面。



步骤4 选择**对象存储批量迁移模板**，单击“开始配置”按钮，进行 workflow 配置。

步骤5 在**基础信息**区域，输入自定义的 workflow **名称**和**描述**。

步骤6 在**迁移方案**区域，选择**步骤三创建的迁移方案**。选择方案后，可以看到该方案的配置概览。单击“查看详情”，可以查看该方案的更多信息。



步骤7 在**迁移集群**区域，选择迁移使用的集群。默认使用迁移方案中配置的集群，您也可以在下拉列表中重新选择集群，选择后仅对当前工作流生效，不会修改迁移方案中配置的集群。

步骤8 根据**表3-10**，进行迁移配置。

表 3-10 迁移配置参数说明

参数	子参数	说明
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。
同名覆盖方式	不覆盖	对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。 警告 <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； 说明 对于空文件夹，覆盖策略与文件相同。
迁移元数据	-	选择是否迁移元数据。 <ul style="list-style-type: none"> 勾选，则迁移源端对象元数据。 不勾选，仅迁移Content-Type、Content-Encoding元数据。

步骤9 （可选）根据**表3-11**，配置高级选项。

表 3-11 高级选项参数说明

参数	说明
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。

参数	说明
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。
解冻归档数据	<ul style="list-style-type: none"> 不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。 勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。 <p>说明</p> <p>在数据迁移时，系统会对所有数据发起取回操作，源端云服务商收取相应的API调用费用和存储费用。</p>
过滤配置	<p>通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见过滤策略说明。</p>
CDN迁移支持	<p>当默认域名不能满足您的迁移需求，比如您需要一个自定义域名以便使用CDN降低源端出口下载流量的费用，可以勾选此选项。在域名输入框中输入自定义域名，并选择传输协议。HTTP协议安全级别低于HTTPS协议，推荐使用HTTPS协议。</p> <p>源端为阿里云OSS、腾讯云COS时，如果使用了CDN鉴权，还需要选择“鉴权类型”并输入“鉴权密钥”。</p>
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 <p>例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</p> <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。

步骤10 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

- 步骤11** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，此时工作流已创建完成。
- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
 - 如果您需要在工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。
- 步骤12** 在工作流详情页面，可以查看该工作流的**配置信息**和**资源迁移进度**。还可以进行如下操作：
- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
 - 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
 - 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
 - **设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
 - **节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

----结束

步骤五：清理集群（可选）

当您完成数据迁移并且集群不再需要时，可以直接删除闲置的集群以释放资源。方法请参考[删除集群](#)。

3.3 归档存储（冷存储）数据迁移

背景概述

冷存储是一种针对不频繁访问数据的长期备份解决方案，例如，平均一年访问一次的归档存储或平均几年访问一次的深度归档存储。这种存储方式具有高安全性、持久性和成本低等优势。

在迁移冷存储数据时，一个主要的考虑因素是数据访问的延迟。由于数据通常处于“冻结”状态，因此在需要访问时必须先进行数据恢复（即解冻）。解冻方式分为“**标准解冻**”和“**加急解冻**”，区别如下：

- **标准解冻**：适用于对访问时间要求不是特别紧急的情况，解冻耗时大约需要3~5小时。
- **加急解冻**：适用于需要快速访问数据的情况，解冻时间可以在1~5分钟内完成，但可能涉及额外的费用。

迁移中心的存储工作流提供了自动“**解冻归档数据**”功能，采用**加急解冻**的方式完成数据恢复。

支持的区域

当前支持以下区域（Region）：

- 拉美-圣地亚哥
- 拉美-圣保罗
- 土耳其-伊斯坦布尔
- 亚太-曼谷
- 亚太-新加坡
- 亚太-雅加达
- 中东-利雅得
- 华北-北京四
- 华东-上海一

约束与限制

- 关于存储迁移的约束与限制请查看[存储迁移的约束与限制有哪些？](#)
- 源端为阿里云OSS：当源端对象的存储类型是深度冷归档存储时，无法使用MgC提供的解冻归档数据功能直接迁移，需要在源端先手动解冻后再迁移。

准备工作

- **华为账号准备**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。
- **创建迁移项目**
在MgC控制台为本次迁移创建独立的项目，方法请参考[项目管理](#)。
- **创建目的端OBS桶**
在华为云创建一个用于接收源端数据的OBS桶，区域选择您希望迁移到的目的端区域。创建方法请参考[创建桶](#)。
- **权限配置**
确保源端账号和目的端账号拥有迁移需要的权限，详细介绍请参见[如何获取源端桶和目的端桶权限？](#)。
- **创建集群**
通过集群可以创建Master节点、迁移节点和列举节点，确保存储工作流的顺利运行。创建方法请参考[创建集群](#)。

创建迁移工作流

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“[实施>迁移工作流](#)”，进入[迁移工作流](#)页面。在页面左上角的当前项目下拉列表中选择在完成准备工作时所[创建的迁移项目](#)。

步骤3 单击页面右上角的“[创建迁移工作流](#)”按钮，进入[选择模板](#)页面。



步骤4 选择存储迁移标准模板，单击“开始配置”按钮，进行 workflow 配置。



步骤5 根据表3-12，配置基础信息参数。

基础信息
请填写名称、描述，并选择区域、集群

* 名称 * 区域

描述

* 集群 [去新建](#)

表 3-12 基础信息参数配置说明

参数	说明
名称	用户自定义。
区域	在下拉菜单中选择 目的端桶 所在区域。
描述	用户自定义。
集群	选择在完成准备工作时， 创建的集群 作为迁移集群。

步骤6 根据表3-13、表3-14，配置源端、目的端信息。

源端信息
请填写源端信息，根据不同站点类型配置不同参数

* 数据站点类型

* AK * SK

* 桶 * 对象存储终端地址

目的端信息
请填写目的端信息，根据不同站点类型配置不同参数

* 数据站点类型

* AK * SK

* 桶 对象存储终端地址

指定前缀

表 3-13 源端信息参数说明

参数	说明
数据站点类型	根据实际的源端云平台选择。
AK	输入源端云账号的访问密钥。
SK	输入源端云账号的私有密钥。
桶	输入源端需要迁移的桶名。
对象存储终端地址	输入源端桶所在区域的终端节点（Endpoint）信息。 例如：站点类型选择阿里云OSS，源端桶所在区域为华东1（杭州），则输入“oss-cn-hangzhou.aliyuncs.com”。
桶类型	数据站点类型为 华为云OBS 时选择。根据桶的实际类型选择。桶类型可以通过 查看桶的基本信息 获取。
appId	数据站点类型为 腾讯云COS 时填写。填写腾讯云APPID。 说明 您可以在腾讯云控制台账号信息页面查看APPID。

表 3-14 目的端信息参数说明

参数	说明
数据站点类型	选择“华为云OBS”。
AK	输入华为云迁移账号的访问密钥和私有密钥。
SK	
桶	选择 创建的目的端OBS桶 。

参数	说明
对象存储终端地址	输入目的端桶所在区域的终端节点（Endpoint）信息。 例如：目的端桶所在区域为华北-北京四，则输入“obs.cn-north-4.myhuaweicloud.com”。 说明 可在OBS桶的基本信息查看。
指定前缀	可选参数。将对象迁移到指定前缀，例如：源端对象路径为/A/B/C.txt，目的端前缀配置为/D，迁移后目的端对象路径为/D/A/B/C.txt。详细介绍请参考 设置“保存到指定前缀”参数 中的功能介绍。

步骤7 根据表3-15，进行迁移配置。



表 3-15 迁移配置参数说明

参数	子参数	说明
迁移任务类型	列表迁移	通过列表文件过滤迁移，只迁移列表文件中的对象。
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。
同名覆盖方式	不覆盖	对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。 警告 <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。

参数	子参数	说明
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； <p>说明 对于空文件夹，覆盖策略与文件相同。</p>
清理集群	-	<p>选择工作流是否创建清理集群步骤。</p> <ul style="list-style-type: none"> 勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPCEP服务等。 不勾选，迁移工作流不创建清理集群步骤。


步骤8 单击高级选项（可选）前的  按钮，展开高级选项，勾选“解冻归档数据”。其余功能根据实际需求选择并配置，各功能的说明请参见表3-16。



表 3-16 高级选项功能说明

功能	说明
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。

功能	说明
解冻归档数据	<ul style="list-style-type: none"> ● 不勾选：如果遇到归档类型对象，则直接判定该对象迁移失败并跳过，继续迁移其余对象。 ● 勾选：如果遇到归档类型对象，则自动解冻该对象并进行迁移。如果解冻失败，则判定该对象迁移失败并跳过，继续迁移其余对象。 <p>说明 在数据迁移时，系统会对所有数据发起取回操作，源端云服务商会收取相应的API调用费用和存储费用。</p>
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> ● 不勾选，不接收SMN消息。 ● 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> ● 不勾选，不进行流量控制。 ● 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> - 限速的数值范围：0~1048576 MB/s。 - 最多可以创建5条记录。 - 时区时间以目的端Region的当地时间为准。
定时迁移	<p>设置迁移任务在指定时间内迁移。</p> <ul style="list-style-type: none"> ● 不勾选，手动启动/暂停迁移任务。 ● 勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。 <p>例如：</p> <ul style="list-style-type: none"> - 设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 - 设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。

步骤9 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤10 确认配置信息无误后，单击“**确认创建工作流**”，弹出启动工作流窗口，此时工作流已创建完成。

- 如果您希望立即开始迁移，则单击“**确认**”按钮，立即运行工作流。
- 如果您需要在工作流中**添加阶段/步骤**，则单击“**取消**”，此时迁移工作流状态为“**等待中**”，并未开始迁移。确定工作流步骤后，单击“**运行**”按钮，开始迁移。

步骤11 在工作流详情页面，可以查看该工作流的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标移动到资源迁移进度条，在弹窗中可以查看该资源更多迁移信息。
- 当迁移步骤（进度条）到达需要**手动确认**的步骤时，您需要将鼠标移动到进度条上，并在弹出的窗口中单击**步骤状态**后的“**确认**”。只有完成这一操作，系统才会继续进行后续的迁移步骤。
- 在**基础信息**区域，单击集群名称后的“**管理**”，右侧弹出集群详细信息页面。在这个页面上，您可以进行如下操作：
 - **设置流量控制规则**：您可以添加、编辑或删除流量控制规则，以确保集群内的流量按照您的需求进行管理。
 - **节点管理**：您可以根据实际需求，添加/删除迁移节点和列举节点，也可以对现有节点进行插件升级。

----结束

3.4 SFS1.0 迁移至 SFS 3.0

3.4.1 使用须知

- **支持的区域**

当前支持在以下区域（Region）进行SFS迁移，其它Region请使用线下工具RDA或者咨询其它方案进行迁移。

 - 拉美-圣地亚哥
 - 拉美-圣保罗
 - 土耳其-伊斯坦布尔
 - 亚太-曼谷
 - 亚太-新加坡
 - 亚太-雅加达
 - 中东-利雅得
 - 华北-北京四
 - 华东-上海一
- **内网迁移**

支持在同一Region内通过内网进行迁移。
- **软连接处理**

MgC不支持源端路径为软链接的迁移。如果源端包含软链接，请进行如下处理：

 - 填写实际的文件路径。
 - 迁移完成后，手动在目的端创建相应的软链接。

3.4.2 准备工作

华为账号准备

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。

创建迁移项目

在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。

创建弹性文件系统

在目的端创建相应的弹性文件系统，用于迁移后的数据存储和管理，方法请参见[创建弹性文件系统](#)。

3.4.3 创建集群

通过集群，您可以创建和管理迁移节点、列举节点，部署和升级迁移插件，并确保存储 workflows 顺利运行。

创建集群将产生相关服务费用，具体费用请参见[计费说明](#)。

须知

创建的集群节点暂不支持直接登录，如有需要请联系技术支持人员。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。

步骤2 在左侧导航栏选择“**部署>迁移集群**”，进入[迁移集群](#)页面。在页面左上角的当前项目下拉列表中选择在完成[准备工作](#)时所创建的迁移项目。

步骤3 单击页面右上角的“**创建集群**”按钮，进入[创建集群](#)页面。



如果是首次创建迁移集群，需要您同意华为云账号进行相关的委托授权。委托权限所包含的细粒度权限请前往[委托权限说明](#)查看。



步骤4 根据表3-17，配置参数。

表 3-17 集群参数配置说明

区域	参数	配置说明	限制说明
基础配置	集群名称	自定义名称。	同一账户下集群名称不可重名。
	区域	选择创建集群的区域。	与目的端区域保持一致。
	集群类型	选择集群使用场景。	当前仅支持存储迁移场景。
节点创建	Master节点ECS规格	用于管理迁移节点和列举节点，默认创建，无需配置。	规格与迁移节点规格一致
	迁移节点ECS规格	迁移节点用于执行迁移和对比的动作，推荐规格为C系列规格，大小是8U16G。	<ul style="list-style-type: none"> 规格选定后，后续无法更改。 设置的节点数量需要同时满足如下要求： <ul style="list-style-type: none"> 迁移节点数+列举节点数+1 ≤ 100 迁移节点数+列举节点数+1 ≤ 子网可用IP数
	列举节点ECS规格	列举节点用于执行列举动作，推荐规格为C系列规格，大小是8U16G。	
网络配置	虚拟私有云	在下拉列表中选择虚拟私有云。	-
	子网	选择与虚拟私有云网段相同且满足可用IP数的子网。	子网的可用IP数≥迁移节点数+列举节点数+1。
	模式选择	选择内网模式。	同一个Region是内网迁移无需绑定EIP。

区域	参数	配置说明	限制说明
-	流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，则不进行流量控制。 勾选，进行流量控制。在“开始时间”、“结束时间”和“限速”文本框中输入数值。 <p>须知 例如：设置开始时间08:00、结束时间12:00、限制最大流量20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。</p>	<ul style="list-style-type: none"> 最多可添加五条流量控制规则。 时区时间以控制台所选Region的当地时间为准。
	启用LTS服务	<ul style="list-style-type: none"> 启用：记录存储迁移过程中的日志，便于查看和定位迁移中出现的问题。 不启用：不记录存储迁移过程中的日志。 	-

步骤5 参数配置完成后，单击“确认”开始创建集群。在迁移集群列表中查看集群状态，状态详情参见[集群状态说明](#)。

步骤6 迁移集群创建成功后，您可以进行如下操作。

- 当集群状态为**健康**或者**亚健康**时，单击集群操作列的[创建工作流](#)，可以配置存储迁移工作流。
- 在[集群管理](#)界面中，您可以添加/删除节点、流量控制规则等。

----结束

3.4.4（可选）网络打通

当创建的集群选择的VPC与SFS服务使用的VPC不是同一个时，需要进行额外的网络配置，以确保SFS 1.0、SFS 3.0和集群之间的网络连通。

SFS 1.0 网络打通

SFS 1.0通过创建对等连接来打通网络，方法请参考[创建相同账户下的对等连接](#)。

SFS 3.0 网络打通

SFS 3.0通过配置VPC终端节点并添加与迁移集群相同的VPC来实现网络打通。

步骤1 参考[配置VPC终端节点](#)，购买VPC终端节点。其中“虚拟私有云”参数配置选择[创建集群](#)时所选择的VPC。

步骤2 配置VPC终端节点完成后，登录弹性文件服务管理控制台，在左侧导航栏选择“通用文件系统 > 文件系统列表”进入文件系统列表页面。



步骤3 单击文件系统名称，进入该文件系统详情页面。在左侧导航栏选择“权限管理”，单击“添加授权”按钮。



步骤4 根据表3-18，配置授权参数，单击“确定”添加完成。



表 3-18 参数配置说明

参数	配置说明
VPC	选择 创建集群 时所选择的VPC。
读写权限	选择“读写”。
用户权限	在下拉列表中选择“root用户不匿名（no_root_squash）”。
授权地址条目	选择“所有IP地址”。

----结束

3.4.5 创建存储迁移 workflow

本节介绍如何使用存储迁移标准模板创建存储迁移 workflow。

前提条件

- 已完成准备工作。
- 已创建集群。
- (可选) 已打通SFS 1.0、3.0和集群之间的网络

操作步骤

步骤1 登录**迁移中心**管理控制台。

步骤2 在左侧导航栏选择“**实施>迁移 workflow**”，进入**迁移 workflow**页面。在页面左上角的**当前项目**下拉列表中选择在完成**准备工作**时所创建的迁移项目。

步骤3 单击页面右上角的“**创建迁移 workflow**”按钮，进入**选择模板**页面。



步骤4 选择**存储迁移标准模板**，单击“**开始配置**”按钮，进行 workflow 配置。



步骤5 根据**表3-19**，配置基础信息参数。

基础信息
请填写名称、描述，并选择区域、集群

* 名称 * 区域

描述

* 集群 [去新建](#)

表 3-19 基础信息参数说明

参数	说明
名称	用户自定义。
区域	下拉菜单中选择目的端所在区域。
描述	用户自定义。
集群	选择 创建的集群 作为迁移集群。

步骤6 根据表3-20、表3-21，配置源端、目的端信息。

源端信息
请填写源端信息，根据不同站点类型配置不同参数

* 数据站点类型

* 文件存储服务地址 * 路径

目的端信息
请填写目的端信息，根据不同站点类型配置不同参数

* 数据站点类型



* 文件存储服务地址 * 路径

表 3-20 源端信息参数说明

参数	说明	备注
数据站点类型	选择“NAS_NFS_V3_PROTOCOL”	-
文件存储服务地址	输入源端文件系统的挂载地址，格式为：IP地址+:/或IP地址+:/ xxx，例如：192.1.1.1:/或192.1.1.1:/0001	-
路径	输入待迁移文件路径，格式为：/文件夹	当输入的是/时，表示迁移整个弹性文件系统。

表 3-21 目的端信息参数说明

参数	说明	备注
数据站点类型	选择“NAS_NFS_V3_MOUNT”	-

参数	说明	备注
文件存储服务地址	输入目的端文件系统的挂载地址。挂载地址可以在文件系统列表页面，单击 挂载地址 后面的  图标复制获取。 	-
路径	输入目的端接收文件的路径，格式为：/ 文件夹。	当输入的是/时，表示迁移到目的端的根目录。

步骤7 根据表3-22，进行迁移配置。



表 3-22 迁移配置参数说明

参数	子参数	说明
迁移任务类型	全量迁移	迁移在 源端信息 配置的待迁移文件路径下/文件夹内的全部数据。
	列表迁移	通过列表文件过滤迁移，只迁移列表文件中的对象。 在“ 列表路径 ”输入存放在目的端桶的对象列表文件路径。 对象列表文件的约束与限制如下： <ul style="list-style-type: none"> 对象列表文件大小不能超过30 MB。 对象列表文件类型必须为.txt类型。 对象列表文件必须是UTF-8无BOM格式编码格式。 对象列表文件每行只能包含一个对象名称，且对象名称使用URL Encode编码。 对象列表文件行长度不超过16KB，否则迁移失败。 对象列表文件行数不超过10000行。
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。

参数	子参数	说明
同名覆盖方式	不覆盖	<p>对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。</p> <p>警告</p> <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； <p>说明 对于空文件夹，覆盖策略与文件相同。</p>
迁移元数据	-	<p>选择是否迁移元数据。</p> <ul style="list-style-type: none"> 勾选，则迁移源端对象元数据。 不勾选，仅迁移Content-Type、Content-Encoding元数据。
清理集群	-	<p>选择工作流是否创建清理集群步骤。</p> <ul style="list-style-type: none"> 勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPCEP服务等。 不勾选，迁移工作流不创建清理集群步骤。

步骤8 （可选）根据[表3-23](#)，配置高级选项。

表 3-23 高级选项参数说明

参数	说明
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。

参数	说明
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。
定时迁移	<p>设置迁移任务在指定时间内迁移。</p> <ul style="list-style-type: none"> 不勾选，手动启动/暂停迁移任务。 勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。 例如： <ul style="list-style-type: none"> 设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。

步骤9 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤10 确认配置信息无误后，单击“**确认创建工作流**”，弹出启动工作流窗口，单击“**确认**”按钮，立即运行工作流。

步骤11 在工作流列表，单击工作流名称，进入工作流详情页面，可以查看该迁移工作流的**配置信息和资源迁移进度**。

----**结束**

3.4.6 （可选）清理集群

当您完成数据迁移并且集群不再需要时，可以直接删除闲置的集群以释放资源。

步骤1 在左侧导航栏选择“**部署>迁移集群**”，进入**迁移集群**页面。

步骤2 在集群列表，选择需要清理的集群，单击操作列的“更多 > 删除”，弹出删除集群窗口。



步骤3 选择是否同步删除关联资源后，单击“确认”，删除集群。



----结束

3.5 NAS 到 NAS 迁移和割接方案

概述

- **适用场景：**适用于NAS到NAS迁移过程中源端业务持续运行的场景。
- **目标：**通过多次同步方案迁移增量对象数据，减少源端和目的端数据差异，确保数据迁移的准确性和业务连续性。
- **源端支持的类型：**NAS_SMB、NAS_NFS_V3_MOUNT、NAS_NFS_V3_PROTOCOL、NAS_GFS。
- **目的端支持的类型：**NAS_SMB、NAS_NFS_V3_MOUNT。

注意事项

迁移过程中，会调用源端云服务商的接口，所产生的费用都遵循源端云服务商的计费规则并由其向您收取。

操作流程

操作步骤	说明
准备工作	<ul style="list-style-type: none"> 注册华为账号、开通华为云，并为账户充值。 获取账号迁移权限。 获取目的端账号的AK/SK。 在MgC控制台创建迁移项目 在MgC控制台创建迁移集群。 创建目的端弹性文件系统。
步骤一：创建全量迁移工作流	使用MgC存储迁移工作流完成源端数据的全量迁移。
步骤二：创建增量迁移工作流	在全量迁移完成后，再创建增量迁移工作流，进行多次增量数据迁移，以减少数据差异。
步骤三：灰度切流（可选）	在条件允许的情况下，建议进行灰度切流，即部分业务切换到目的端，以验证业务的连续性和数据一致性。
步骤四：业务割接	在多次增量迁移后，确定一个可接受的最后一次增量迁移时间窗，用于最终的业务割接。

准备工作

- 华为账号准备**
 使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。
- 创建迁移项目**
 在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。
- 创建集群**
 为本次迁移创建独立的迁移集群，方法请参见[创建集群](#)。
- 创建弹性文件系统**
 在目的端创建相应的弹性文件系统，用于迁移后的数据存储和管理，方法请参见[创建弹性文件系统](#)。

步骤一：创建全量迁移工作流

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“[实施>迁移工作流](#)”，进入[迁移工作流](#)页面。在页面左上角的当前项目下拉列表中选择在完成[准备工作](#)时所创建的迁移项目。

步骤3 参考[创建存储迁移工作流](#)，创建全量迁移任务。

----结束

步骤二：创建增量迁移工作流

在全量迁移完成后，参考[创建存储迁移工作流](#)，创建一个与[步骤一](#)相同的新的存储迁移工作流以进行增量迁移。需要注意在配置工作流的[迁移配置](#)区域，同名覆盖方式选

择“条件覆盖”，其余参数配置与步骤一创建的工作流保持一致。配置完成后，启动新的存储迁移工作流，开始增量迁移。

图 3-8 选择同名覆盖方式

迁移配置
请选择任务类型、并发子任务以及同名覆盖方式。选择是否开启迁移元数据以及清理场景

* 迁移任务类型 全量迁移 列表迁移 前缀迁移

* 并发子任务数

* 同名覆盖方式 条件覆盖 不要覆盖 直接覆盖 CRC64
覆盖大小不相同或源端最后修改时间大于目的端最后修改时间的对象。

迁移元数据
为保证迁移任务正常运行，仍将为您迁移ContentType元数据

清理集群
勾选后将自动生成清理集群步骤，您可以继续配置额外要删除的配置，包括NAT网关、安全组等

步骤三：灰度切流（可选）

待首次增量迁移完成后，如果有条件，在目的端进行灰度切流验证业务。

步骤四：业务割接

多次增量迁移后，确定一个业务割接时间窗。停止源端业务，完成最后一次增量迁移，并启动目的端写业务。

3.6 文件存储批量迁移

本文介绍如何使用迁移中心（MgC）的迁移方案配置功能和文件存储批量迁移模板，以高效完成文件存储的批量迁移工作。

- **源端支持的类型：**NAS_NFS_V3_MOUNT、NAS_NFS_V3_PROTOCOL、NAS_GFS。
- **目的端支持的类型：**NAS_NFS_V3_MOUNT。

约束与限制

存储迁移的约束与限制同样适用于文件存储批量迁移，详情请查看[存储迁移的约束与限制有哪些？](#)

准备工作

- **华为账号准备**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户。获取方法请参考[准备工作](#)。
- **创建迁移项目**
在MgC控制台为本次迁移创建独立的项目，方法请参见[项目管理](#)。
- **创建弹性文件系统**

在目的端创建相应的弹性文件系统，用于迁移后的数据存储和管理，方法请参见[创建弹性文件系统](#)或[创建高性能文件系统](#)。

步骤一：创建迁移集群

迁移集群是专为存储 workflow 提供协作的核心组件，通过集群可以创建和管理迁移节点、列举节点，部署和升级迁移插件，确保存储 workflow 顺利运行。为本次迁移创建独立的迁移集群，方法请参见[创建集群](#)。

步骤二：创建文件存储迁移方案

将源端需要迁移到同一目的端区域的文件存储资源添加到同一迁移方案中，并为每个资源配置相应的目的端，然后绑定迁移集群。详细操作步骤请参考[配置文件存储批量迁移方案](#)。

步骤三：创建文件存储迁移 workflow

选择迁移方案，配置迁移策略，完成迁移方案内资源的批量迁移。详细操作步骤请参考[创建文件存储批量迁移 workflow](#)。

3.7 通过 HTTP 迁移 MinIO 存储数据到华为云 OBS

本文介绍如何使用迁移中心（MgC）的存储迁移 workflow，实现通过 HTTP 协议将 MinIO 存储的数据迁移至华为云对象存储服务（OBS）的方法。

支持的区域

当前支持以下区域（Region）：

- 拉美-圣地亚哥
- 拉美-圣保罗
- 土耳其-伊斯坦布尔
- 亚太-曼谷
- 亚太-新加坡
- 亚太-雅加达
- 中东-利雅得
- 华北-北京四
- 华东-上海一

注意

单个对象大小不能超过： $500 \text{ MB} \times 10000 = 4.76837158203125 \text{ TB}$ ，否则可能会导致迁移失败。

准备工作

- 华为账号准备

使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。

- **创建迁移项目**

在MgC控制台为本次迁移创建独立的项目，方法请参考[项目管理](#)。

- **创建目的端OBS桶**

在华为云创建一个用于存放URL列表文件和接收源端数据的OBS标准存储桶，区域选择您希望迁移到的目的端区域。创建方法请参考[创建桶](#)。

如果使用IAM用户进行迁移，该IAM用户需要拥有目的端桶的读写权限。授权方法请参考[对单个IAM用户授予桶的读写权限](#)。

- **创建集群**

通过集群可以创建Master节点、迁移节点和列举节点，确保存储工作流的顺利运行。创建方法请参考[创建集群](#)。

步骤一：设置 MinIO 文件共享并生成下载链接

以下步骤中提到的minioLink、bucket-public01、http://100.85.xxx.xx:xxxx等应替换为您实际的MinIO服务别名、存储桶名称和MinIO服务的访问地址。

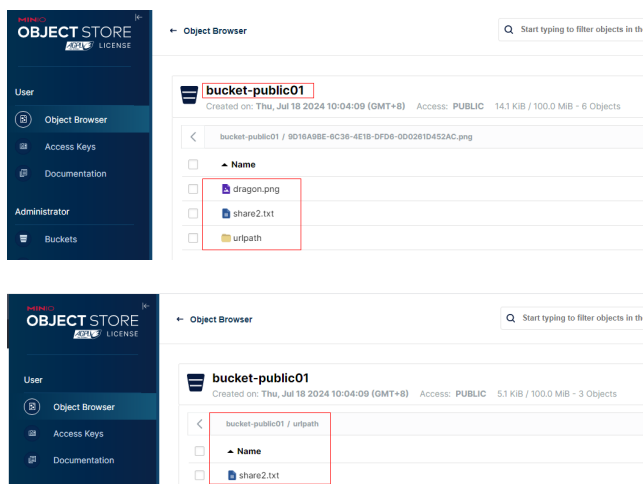
步骤1 在部署MinIO的服务器上，打开命令行界面。

步骤2 使用mc命令设置文件共享下载，命令格式如下：

```
mc share download --recursive --expire=<EXPIRE-IN-HOURS> <MINIO-ALIAS>/<BUCKET-NAME>/
```

- --recursive：递归地为子文件夹中的文件设置共享。
- --expire=<EXPIRE-IN-HOURS>：设置共享链接的过期时间，单位为小时。最长不得超过7天（168小时）。
- <MINIO-ALIAS>：部署MinIO服务时自定义的服务别名。
- <BUCKET-NAME>：存储桶名称。

例如，您的存储桶名为bucket-public01，桶中有2个文件(dragon.png, share2.txt)和1个子文件夹(urlpath)，子文件夹中有1个文件(share2.txt)。



如果您希望生成存储桶中所有文件及子文件夹中文件的共享下载链接，过期时间为12小时，输出的结果保存到test2.txt文件中。您可以执行如下命令：

```
mc share download --recursive --expire=12h minioLink/bucket-public01/ >> test2.txt
```

执行命令后，将生成3个文件的共享下载链接并保存在test2.txt文件中。

```
[root@oms-cluster-ecs-1797559186613161988-d32e ~]# mc share download --recursive --expire=12h minioLink/bucket-public01/
URL: http://100.85.118.88:9000/bucket-public01/dragon.png
Expire: 12 hours 0 minutes 0 seconds
Share: http://100.85.118.88:9000/bucket-public01/dragon.png?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=admin%2F2024072424T090221Z&X-Amz-Expires=43200&X-Amz-SignedHeaders=host&X-Amz-Signature=80df3a76c21afb74ebb9cfd151cd84934c9c7e4alc47c901a10a86

URL: http://100.85.118.88:9000/bucket-public01/share2.txt
Expire: 12 hours 0 minutes 0 seconds
Share: http://100.85.118.88:9000/bucket-public01/share2.txt?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=admin%2F2024072424T090221Z&X-Amz-Expires=43200&X-Amz-SignedHeaders=host&X-Amz-Signature=7c183cedc0b4a8455b7bf2753c86c3938b7b7b5ffe22f020f29b1f

URL: http://100.85.118.88:9000/bucket-public01/urlpath/share2.txt
Expire: 12 hours 0 minutes 0 seconds
Share: http://100.85.118.88:9000/bucket-public01/urlpath/share2.txt?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=admin%2F20240724T090221Z&X-Amz-Expires=43200&X-Amz-SignedHeaders=host&X-Amz-Signature=63c322ca1594fe7abe2f500fd3e42d208a7581921cd728
```

步骤3 打开生成的URL列表文件（在本文示例中为test2.txt），您将看到每个文件的共享下载链接。通过浏览器尝试访问这些链接，以验证生成的URL是否有效。

如果链接无法访问，请检查文件的访问权限是否设置为可读，并重新生成链接。

步骤4 根据MgC的要求，编辑URL列表文件内容，使其包含文件的共享下载链接和文件名，格式为：

```
<SHARED-URL> <FILE-NAME>
```

共享下载链接和文件名之间用制表符隔开，子文件夹的文件名需要包含子文件夹名称。关于URL列表文件的更多要求和限制，请查看[存储迁移的约束与限制有哪些？](#)。

例如：

```
http://100.85.xxx.xx:xxxx/bucket-public01/dragon.png dragon.png
http://100.85.xxx.xx:xxxx/bucket-public01/share2.txt share2.txt
http://100.85.xxx.xx:xxxx/bucket-public01/urlpath/share2.txt urlpath/share2.txt
```

步骤5 按要求编辑完所有URL链接并检查无误后，保存该URL列表文件。

----结束

步骤二：上传 URL 列表文件至 OBS 桶

步骤1 登录[OBS管理控制台](#)，在左侧导航栏选择“桶列表”。

步骤2 在桶列表中，单击[创建的目的端OBS桶](#)，进入“对象”页面。

步骤3 单击“新建文件夹”，在“文件夹名称”中输入自定义名称（例如：minioUrl），单击“确定”。

步骤4 单击上一步创建的文件夹名称，进入该文件夹，单击“上传对象”，系统弹出“上传对象”对话框。

步骤5 通过以下任意方式，将待迁移对象的[URL列表文件](#)（在本文示例中为test2.txt）上传至文件夹。

- 拖拽URL列表文件至“上传对象”区域框内，单击“上传”。
- 单击“上传对象”区域框内的“添加文件”，选择URL列表文件进行添加，单击“上传”。

----结束

步骤三：创建存储迁移 workflow

步骤1 登录[迁移中心管理控制台](#)。

步骤2 在左侧导航栏选择“[实施>迁移 workflow](#)”，进入[迁移 workflow](#)页面。在页面左上角的当前项目下拉列表中选择在完成准备工作时所[创建的迁移项目](#)。

源端信息
请填写源端信息，根据不同站点类型配置不同参数

* 数据站点类型

* 列表路径

目的端信息
请填写目的端信息，根据不同站点类型配置不同参数

* 数据站点类型

* AK * SK

* 桶 对象存储终端地址

指定前缀

表 3-25 源端信息参数说明

参数	说明
数据站点类型	选择“HTTP/HTTPS数据源”。
列表路径	填写在 步骤二存放URL列表文件 的文件夹名称（在本文示例中为minioUrl/），注意要在文件夹名称后面加上“/”。

表 3-26 目的端信息参数说明

参数	说明
数据站点类型	选择“华为云OBS”。
AK	输入华为云迁移账号的访问密钥和私有密钥。该账号需要拥有目的端桶的读写权限。
SK	
桶	选择 创建的目的端OBS桶 。
对象存储终端地址	输入目的端桶所在区域的终端节点（Endpoint）信息。 例如：目的端桶所在区域为华北-北京四，则输入“obs.cn-north-4.myhuaweicloud.com”。 说明 可在OBS桶的基本信息查看。
指定前缀	可选参数。将对象迁移到指定前缀，例如：源端对象路径为/A/B/C.txt，目的端前缀配置为/D，迁移后目的端对象路径为/D/A/B/C.txt。详细介绍请参考 设置“保存到指定前缀”参数 中的功能介绍。

步骤7 根据[表3-27](#)，进行迁移配置。



表 3-27 迁移配置参数说明

参数	子参数	说明
迁移任务类型	列表迁移	通过列表文件过滤迁移，只迁移列表文件中的对象。
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。
同名覆盖方式	不覆盖	对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。 警告 <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； 说明 对于空文件夹，覆盖策略与文件相同。
清理集群	-	选择工作流是否创建清理集群步骤。 <ul style="list-style-type: none"> 勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPC服务等服务。 不勾选，迁移工作流不创建清理集群步骤。

步骤8 （可选）根据表3-28，配置高级选项。

表 3-28 高级选项参数说明

参数	说明
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。
定时迁移	<p>设置迁移任务在指定时间内迁移。</p> <ul style="list-style-type: none"> 不勾选，手动启动/暂停迁移任务。 勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。 例如： <ul style="list-style-type: none"> 设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。

- 步骤9** workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。
- 步骤10** 确认配置信息无误后，单击“**确认创建工作流**”，弹出启动工作流窗口，单击“**确认**”按钮，立即运行工作流。
- 步骤11** 在工作流列表，单击工作流名称，进入工作流详情页面，可以查看该迁移工作流的**配置信息和资源迁移进度**。

---结束

3.8 通过 HTTP 迁移 Ceph 存储数据到华为云 OBS

本文介绍如何使用迁移中心（MgC）的存储迁移工作流，实现通过HTTP协议将Ceph存储的数据迁移至华为云对象存储服务（OBS）的方法。

支持的区域

当前支持以下区域（Region）：

- 拉美-圣地亚哥
- 拉美-圣保罗
- 土耳其-伊斯坦布尔
- 亚太-曼谷
- 亚太-新加坡
- 亚太-雅加达
- 中东-利雅得
- 华北-北京四
- 华东-上海一

注意

单个对象大小不能超过： $500 \text{ MB} \times 10000 = 4.76837158203125 \text{ TB}$ ，否则可能会导致迁移失败。

准备工作

- **华为账号准备**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，并获取账号/IAM用户的访问密钥（AK/SK）。获取方法请参考[准备工作](#)。
- **创建迁移项目**
在MgC控制台为本次迁移创建独立的项目，方法请参考[项目管理](#)。
- **创建目的端OBS桶**
在华为云创建一个用于存放URL列表文件和接收源端数据的OBS标准存储桶，区域选择您希望迁移到的目的端区域。创建方法请参考[创建桶](#)。
如果使用IAM用户进行迁移，该IAM用户需要拥有目的端桶的读写权限。授权方法请参考[对单个IAM用户授予桶的读写权限](#)。

- **创建集群**

通过集群可以创建Master节点、迁移节点和列举节点，确保存储工作流的顺利运行。创建方法请参考[创建集群](#)。

步骤一：生成 Ceph 文件共享下载链接

以下步骤中提到的bucket01、http://100.93.xxx.xx:7480等应替换为您实际的Ceph S3存储桶名称和Ceph RGW服务的Web访问地址和端口。

示例中的以下参数，请替换为实际值：

- <BUCKET-NAME>：存储桶名称。
- < FILE-NAME >：创建的json文件名。
- <DOMAIN>：Ceph RGW（RADOS Gateway）服务的实际域名或IP地址。
- <PORT>：Ceph RGW服务的实际访问端口号。

如果您要迁移的桶数据可以直接通过浏览器访问，请跳过步骤1，直接进行[步骤2](#)操作。

步骤1 查看源端桶是否已经设置了访问策略，命令格式如下：

```
s3cmd info s3://<BUCKET-NAME>
```

- 如果返回结果中Policy参数值为“none”，代表没有设置桶的访问策略。请执行[步骤2](#)。

```
s3://bucket01/ (bucket):
Location: default
Payer: BucketOwner
Expiration Rule: none
Policy: none
CORS: none
ACL: admin: FULL_CONTROL
```

- 如果返回结果中Policy参数值有配置信息，请拷贝策略信息并保存，用于数据迁移完成后恢复到此策略。

```
s3://bucket01/ (bucket):
Location: default
Payer: BucketOwner
Expiration Rule: none
Policy: {
  "Statement": [{
    "Effect": "Allow",
    "Principal": "*",
    "Action": "s3:GetObject",
    "Resource": "*"
  }]
}
CORS: none
ACL: admin: FULL_CONTROL
```

步骤2 在安装s3cmd工具的服务器上，打开文本编辑器，创建一个xxx.json文件（名称自定义），其中包含S3存储桶策略，允许从指定的S3存储桶中获取对象。将以下内容复制并粘贴到编辑器中，替换 <BUCKET-NAME> 为实际的S3存储桶名称，保存创建的json文件并退出编辑器。

```
{
  "Statement": [{
    "Effect": "Allow",
    "Principal": "*",
    "Action": "s3:GetObject",
```

```
    "Resource": "arn:aws:s3::: <BUCKET-NAME>/*"
  }
}
```

更多参数设置请参考[Amazon S3 存储桶策略的示例](#)。

步骤3 使用 s3cmd 命令行工具设置存储桶策略，允许公开访问存储桶中的文件。命令格式如下：

```
s3cmd setpolicy <FILE-NAME>.json s3://<BUCKET-NAME>
```

其中 <FILE-NAME> 替换为 [步骤2](#) 创建的 json 文件名称，<BUCKET-NAME> 替换为实际的 S3 存储桶名称。

步骤4 使用以下命令列出存储桶中的所有文件，并将结果导出到文本文件：

```
s3cmd ls -r s3://<BUCKET-NAME> >> <FILE-URL>.txt
```

其中 <BUCKET-NAME> 替换为实际的 S3 存储桶名称，<FILE-URL> 替换为您希望保存文件列表的本地文件名（例如：s3url.txt）。

步骤5 打开生成的列表文件（在本文示例中为 s3url.txt），您将看到桶内所有文件的共享 s3 地址列表。将“s3://及其之前的所有参数”替换为“http://<DOMAIN>:<PORT>/”，以生成可通过浏览器访问的 URL（即文件共享下载链接）。在浏览器中输入此 URL，如果能够成功访问文件，则说明设置正确。如果访问被拒绝（AccessDenied），则需要重新执行 [步骤1](#)~[步骤4](#) 设置访问策略。

例如：Ceph RGW 服务的实际域名或 IP（DOMAIN）为“100.93.xxx.xx”，端口号（PORT）为“7480”，存储桶名为“bucket01”，桶中有 2 个文件（dragon.png, index.html），生成的共享 s3 地址列表为：

```
2024-07-26 03:09      3987 s3://bucket01/dragon.png
2024-07-26 02:01      1701 s3://bucket01/index.html
```

将“s3://及其之前的所有参数”替换为“http://<DOMAIN>:<PORT>/”，即“日期...s3://”替换为“http://100.93.xxx.xxx:7480”，生成的 URL 列表为：

```
http://100.93.xxx.xxx:7480/bucket01/dragon.png
http://100.93.xxx.xxx:7480/bucket01/index.html
```

步骤6 根据上一步的方法和要求，将列表文件中的所有共享 s3 地址替换为 URL 地址。如果列表文件中的 s3 地址较多，可以使用文本编辑器（如 NotePad++）进行批量替换。

步骤7 根据 MgC 的要求，编辑 URL 列表文件内容，使其包含文件的共享下载链接和文件名，格式为：

```
<SHARED-URL> <FILE-NAME>
```

共享下载链接和文件名之间用制表符（Tab 键）隔开，子文件夹的文件名需要包含子文件夹名称。关于 URL 列表文件的更多要求和限制，请查看[存储迁移的约束与限制有哪些？](#)。

例如：

```
http://100.93.xxx.xxx:7480/bucket01/dragon.png dragon.png
http://100.93.xxx.xxx:7480/bucket01/index.html index.html
```

步骤8 按要求编辑完所有 URL 链接并检查无误后，保存该 URL 列表文件。

----结束

步骤二：上传 URL 列表文件至 OBS 桶

步骤1 登录[OBS管理控制台](#)，在左侧导航栏选择“桶列表”。

步骤2 在桶列表中，单击[创建的目的端 OBS 桶](#)，进入“对象”页面。

- 步骤3** 单击“新建文件夹”，在“文件夹名称”中输入自定义名称（例如：cephUrl），单击“确定”。
 - 步骤4** 单击上一步创建的文件夹名称，进入该文件夹，单击“上传对象”，系统弹出“上传对象”对话框。
 - 步骤5** 通过以下任意方式，将待迁移对象的**URL列表文件**（在本文示例中为s3url.txt）上传至文件夹。
 - 拖拽URL列表文件至“上传对象”区域框内，单击“上传”。
 - 单击“上传对象”区域框内的“添加文件”，选择URL列表文件进行添加，单击“上传”。
- 结束

步骤三：创建存储迁移 workflow

- 步骤1** 登录**迁移中心管理控制台**。
- 步骤2** 在左侧导航栏选择“**实施>迁移 workflow**”，进入**迁移 workflow**页面。在页面左上角的当前项目下拉列表中选择在完成准备工作时所**创建的迁移项目**。
- 步骤3** 单击页面右上角的“**创建迁移 workflow**”按钮，进入**选择模板**页面。



- 步骤4** 选择**存储迁移标准模板**，单击“**开始配置**”按钮，进行 workflow 配置。



- 步骤5** 根据**表3-29**，配置基础信息参数。

基础信息
请填写名称、描述，并选择区域、集群

* 名称 * 区域

描述 0/255

* 集群 [去新建](#)

表 3-29 基础信息参数配置说明

参数	说明
名称	用户自定义。
区域	在下拉菜单中选择 创建的目的端桶 所在区域。
描述	用户自定义。
集群	选择在完成准备工作时， 创建的集群 作为迁移集群。

步骤6 根据表3-30、表3-31，配置源端、目的端信息。

源端信息
请填写源端信息，根据不同站点类型配置不同参数

* 数据站点类型

* 列表路径

目的端信息
请填写目的端信息，根据不同站点类型配置不同参数

* 数据站点类型

* AK * SK

* 桶 对象存储终端地址

指定前缀

表 3-30 源端信息参数说明

参数	说明
数据站点类型	选择“HTTP/HTTPS数据源”。
列表路径	填写在 步骤二存放URL列表文件 的文件夹名称（在本文示例中为cephUrl/），注意要在文件夹名称后面加上“/”。

表 3-31 目的端信息参数说明

参数	说明
数据站点类型	选择“华为云OBS”。

参数	说明
AK	输入华为云迁移账号的访问密钥和私有密钥。该账号需要拥有目的端桶的读写权限。
SK	
桶	选择 创建的目的端OBS桶 。
对象存储终端地址	输入目的端桶所在区域的终端节点（Endpoint）信息。 例如：目的端桶所在区域为 华北-北京四 ，则输入“obs.cn-north-4.myhuaweicloud.com”。 说明 可在OBS桶的基本信息查看。
指定前缀	可选参数。将对象迁移到指定前缀，例如：源端对象路径为/A/B/C.txt，目的端前缀配置为/D，迁移后目的端对象路径为/D/A/B/C.txt。详细介绍请参考 设置“保存到指定前缀”参数 中的功能介绍。

步骤7 根据表3-32，进行迁移配置。



表 3-32 迁移配置参数说明

参数	子参数	说明
迁移任务类型	列表迁移	通过列表文件过滤迁移，只迁移列表文件中的对象。
并发子任务数	-	用户自定义。配置的数目不能超过可用迁移节点数x10。例如：可用的迁移节点数为2，则最大子任务数配置数目需要≤20。
同名覆盖方式	不覆盖	对于源端和目的端存在的同名文件，跳过源端文件，保留目的端文件。 警告 <ul style="list-style-type: none"> 首次迁移时选择“不覆盖”，不会迁移源端中父文件夹的属性到目的端，可能会造成文件夹属性不完整。为规避这一风险，首次迁移时请谨慎使用“不覆盖”策略。 选择“不覆盖”策略进行数据迁移时，如果迁移任务因错误中断后重启或从暂停状态恢复，可能会造成部分数据未完全迁移，而任务显示成功的现象，影响数据的完整性。为规避这一风险，请谨慎使用“不覆盖”策略。

参数	子参数	说明
	直接覆盖	对于源端和目的端存在的同名文件，源端文件直接覆盖目的端文件。
	条件覆盖	<ul style="list-style-type: none"> 对于文件迁移：源端和目的端存在的同名文件，保留最后修改时间较新的文件；如果最后修改时间相同，比较文件大小，大小不同则源端覆盖目的端文件；如果最后修改时间和大小都相同，则跳过； 对于文件夹迁移：当文件迁移完成后再比对文件夹大写、时间和权限，如果文件夹大小不同/时间不相同/权限不相同，则源端覆盖目的端文件夹； <p>说明 对于空文件夹，覆盖策略与文件相同。</p>
清理集群	-	<p>选择工作流是否创建清理集群步骤。</p> <ul style="list-style-type: none"> 勾选，迁移工作流将自动创建清理集群步骤，您还可以继续配置是否删除NAT网关、安全组、VPCEP服务等。 不勾选，迁移工作流不创建清理集群步骤。

步骤8 （可选）根据表3-33，配置高级选项。

表 3-33 高级选项参数说明

参数	说明
目的端存储类型	选择迁移到华为云OBS后，对象的存储类型。存储策略说明请参见 存储类型介绍 。
KMS加密	<ul style="list-style-type: none"> 不勾选，迁移后对象状态与目的端桶级服务端加密状态保持一致。 勾选，所有对象在迁移到OBS桶后都将被加密存储。 <p>说明</p> <ul style="list-style-type: none"> 使用KMS服务端加密可能会降低迁移速率10%左右。 目的端桶所在区域支持KMS加密时，此参数才可用。
过滤配置	通过黑、白名单以及时间段过滤迁移文件。过滤策略详细介绍请参见 过滤策略说明 。
SMN消息发送配置	<p>选择是否使用SMN消息通知功能，该功能可以通知您迁移任务的结果。</p> <ul style="list-style-type: none"> 不勾选，不接收SMN消息。 勾选，接收SMN消息。选择SMN主题，会向该主题的订阅者发送消息；选择消息发送语言；选择发送消息的触发条件，支持同时勾选。

参数	说明
流量控制	<p>设置迁移任务在指定时段的最大流量带宽。</p> <ul style="list-style-type: none"> 不勾选，不进行流量控制。 勾选，进行流量控制。分别设置“开始时间”、“结束时间”和“限速”数值。 例如：设置开始时间08:00、结束时间12:00、限速为20 MB/s，当迁移任务运行到（08:00~12:00）该时间段时，最大迁移速度将限制为20 MB/s。该时间段以外，迁移速度不做限制。 <p>说明</p> <ul style="list-style-type: none"> 限速的数值范围：0~1048576 MB/s。 最多可以创建5条记录。 时区时间以目的端Region的当地时间为准。
定时迁移	<p>设置迁移任务在指定时间内迁移。</p> <ul style="list-style-type: none"> 不勾选，手动启动/暂停迁移任务。 勾选，设置任务在指定时间范围内进行迁移，指定时间以外则暂停迁移。 例如： <ul style="list-style-type: none"> 设置开始时间08:00、结束时间12:00，则迁移任务会在每天早晨08:00至中午12:00该时间段进行迁移。该时间段以外，则迁移任务暂停。 设置开始时间12:00、结束时间08:00，则迁移任务会在每天中午12:00至次日早晨08:00该时间段进行迁移。该时间段以外，则迁移任务暂停。

步骤9 workflows配置完成后，单击“**下一步：确认并创建**”按钮，进入确认并创建页面。

步骤10 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动工作流**窗口，单击“**确认**”按钮，立即运行工作流。

步骤11 在工作流列表，单击工作流名称，进入工作流详情页面，可以查看该迁移工作流的**配置信息和资源迁移进度**。

----结束

步骤四：恢复桶的访问策略

迁移完成后，恢复源端桶的访问策略。

- 如果在**步骤1**中执行命令的返回结果为 Policy: none，请执行如下命令，删除添加的公开访问策略：
s3cmd delpolicy s3://<BUCKET-NAME>
- 如果在**步骤1**中执行命令的返回结果包含访问策略，请按以下步骤恢复。
 - 执行如下命令，删除添加的公开访问策略：
s3cmd delpolicy s3://<BUCKET-NAME>
 - 执行如下命令，恢复访问策略为保存的原有策略。
s3cmd setpolicy <保存的原策略>.json s3://<BUCKET-NAME>

4 目的端主机磁盘扩容

迁移中心MgC支持目的端主机磁盘扩容功能，允许您根据源端主机磁盘的实际使用量，选择磁盘容量更小、数量更少的目的端主机进行迁移。

使用须知

- 系统盘容量取值范围为：40 G~1024 G
- 数据盘容量取值范围为：10 G~32768 G
- 当前仅支持Linux主机磁盘扩容，需确保扩容后容量大于源端实际使用量。
- 跨可用区迁移场景仅支持扩容，即使配置扩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

获取源端主机磁盘信息

在进行目的端主机磁盘扩容前，需要先完成源端主机资源的发现与深度采集，获取源端主机的磁盘信息。

步骤1 请根据实际情况，选择主机发现与采集方法。

- [公网采集](#)
- [内网采集](#)
- [手动添加](#)

步骤2 完成主机资源的发现与深度采集后。在资源列表的主机页签，单击主机名称，进入主机详情页面。



步骤3 在主机详情页面的磁盘信息区域，可以查看到该源端主机的磁盘数量以及实际使用情况。这些数据可以为目的端磁盘调整提供参考。

磁盘类型	磁盘名称	云硬盘类型	磁盘分区格式	大小	已使用空间	利用率
系统盘	dbvvrda	--	MBR	50 GB	1.28 GB	2.56
数据盘	dbvvrdb	--	MBR	20 GB	24 KB	0.01

---结束

通过绑定目的端进行磁盘扩容

- 步骤1** 单击左侧导航栏的“设计>迁移方案设计”，进入迁移方案设计页面。
- 步骤2** 单击目的端规格配置框中的“目的端配置表”，进入目的端配置表页面。
- 步骤3** 在资源配置明细的主机页签，单击主机资源绑定目的端列的“去绑定”，弹出绑定目的端窗口。
- 步骤4** 选择源端主机[关联的应用](#)所在区域以及区域下的项目，在所选项目下，选择要绑定的目的端主机。您可以参考[源端主机磁盘信息](#)并结合实际情况，选择一台磁盘容量小于源端、磁盘数量少于源端的目的端主机。



注意

需要确保所绑定的目的端主机磁盘容量大于源端实际使用量。

例如，源端主机系统盘为50G，数据盘为20G，但实际使用量远低于此，在磁盘容量与数量上均有扩容空间。因此，目的端可选择绑定一台仅包含40G系统盘的主机，即可满足业务需求。

绑定目的端

基础配置

区域: [下拉菜单] ?

项目: [下拉菜单] ?

目的端服务器选择 [更多目的端选择规则](#)

名称: [下拉菜单] | 全局搜索: 请输入名称 [输入框] [清除]

源端: [下拉菜单] | 推荐目的端: 操作系统: UNKNOWN | 系统盘: 50 GB | 数据盘: 20 GB

! 当前仅Linux主机支持磁盘扩容，需确保扩容后容量大于源端实际使用量。跨可用区迁移仅支持扩容，即使配置扩容，系统也会自动创建一个与源端磁盘容量相同的目的端磁盘。

名称	状态	操作系统	磁盘信息	私网IP	公网IP
[选择]	开机	Huawei Clou...	系统盘: 40 GB	[私网IP]	-----
n [选择]	开机	CentOS 7.8 ...	系统盘: 40 G...	[私网IP]	-----

[取消] [确认]

- 步骤5** 单击“确认”。目的端绑定成功后，系统会自动判断所绑定的目的端是否进行了磁盘扩容，在“磁盘扩容”列，“是”代表已扩容；“否”代表未扩容。

源端名称	源端规格名称	状态	目的端规格名称	源端状态	绑定目的端	成本预估	所属应用	磁盘扩容	操作
[源端名称]	--11核1 GB UNKNOWN	运行中	通用计算型(S2) 4核11核1 GB	运行中	[目的端名称]	[成本]	[应用]	是	[操作]

步骤6 进行目的端磁盘扩容后，即可前往[创建主机迁移 workflow](#)。当迁移步骤（进度条）到达“[调整磁盘分区](#)”步骤时，系统会识别目的端主机是否进行了磁盘扩容。

- 是，则会在该步暂停，您需要前往[SMS控制台](#)，对该主机进行“调整磁盘分区”操作，方法请参考[调整目的端磁盘分区](#)。调整完成后，单击步骤状态后的“[确认](#)”继续进行后续迁移步骤。



- 否，则跳过该步骤，继续进行后续迁移步骤。

----结束

通过更改目的端规格进行磁盘扩容

步骤1 参考[规格评估](#)，获取目的端规格配置。

步骤2 主机规格评估成功后，在[资源配置明细](#)区域，单击主机资源操作列的“[更改目的端规格](#)”，右侧弹出[资源详情](#)窗口。



步骤3 单击磁盘对应的目的端规格处的“[更改规格](#)”，弹出[修改](#)窗口。



步骤4 参考[源端主机磁盘信息](#)并结合实际情况，修改目的端磁盘的资源总用量（磁盘容量）。

例如，源端主机系统盘为50G，数据盘为20G，但实际使用量远低于此，因此可以将目的端系统盘调整为40G，数据盘调整为10G。

注意

需要确保目的端主机磁盘容量大于源端实际使用量。

步骤5 修改完成后，单击“确认”。系统会自动判断是否进行了磁盘扩容，“是”代表已扩容；“否”代表未扩容。

资源类型	源端规格	目的端规格	包周期价格/月	按需价格/小时
系统盘	-- 50 GB	高IO 40 GB 更改规格		
数据盘	-- 20 GB	高IO 20 GB 更改规格		

步骤6 进行目的端磁盘扩容后，即可前往[创建主机迁移工作流](#)。当迁移步骤（进度条）到达“调整磁盘分区”步骤时，系统会识别目的端主机是否进行了磁盘扩容。

- 是，则会在该步暂停，您需要前往[SMS控制台](#)，对该主机进行“调整磁盘分区”操作，方法请参考[调整目的端磁盘分区](#)。调整完成后，单击步骤状态后的“确认”继续进行后续迁移步骤。



- 否，则跳过该步骤，继续进行后续迁移步骤。

----结束

5 调整目的端磁盘分区

使用主机工作流进行迁移时，如果系统检测到目的端主机磁盘进行了扩容，则需要前往主机迁移服务(SMS)，对该主机进行“调整磁盘分区”操作。

操作步骤

步骤1 登录SMS控制台。

步骤2 在左侧导航树中，单击“迁移服务器”，进入迁移服务器列表页面。



步骤3 根据MgC主机工作流中的资源名称，在迁移服务器列表找到对应的服务器，单击目的端列的“设置目的端”，进入迁移配置页面。

步骤4 勾选高级配置后的“现在配置”，打开高级配置项。

步骤5 单击“是否调整分区”配置项后面的“调整磁盘分区”按钮，弹出磁盘分区调整窗口。



步骤6 根据 workflow 配置的目的端规格，调整目的端磁盘大小、数量以及分区大小，确保调整后与 workflow 配置的目的端规格一致。

图 5-1 Linux 磁盘分区调整



说明

- Linux支持LVM调整，可以选择物理卷和逻辑卷是否迁移以及调整大小。
- Linux Btrfs文件系统暂时不支持磁盘分区调整。
- Linux系统分区，swap分区是否迁移不可选，默认为“是”，必须进行迁移。
- LVM迁移卷组，可通过卷组配置页，左上方的按钮组，选择全部迁移或暂不迁移。
- LVM中的逻辑卷如果是是否迁移都选择“否”，则卷组不迁移，对应的物理卷是否迁移也会全部自动切换成“否”。

步骤7 配置完成后，单击“**下一步：确认调整**”。确认调整无误后，单击“**确定**”，完成目的端磁盘分区调整。

步骤8 单击右下角的“**下一步：目的端配置**”，进入目的端配置页签。

步骤9 在**已有服务器**中，选择在MgC绑定或者新创建的同名目的端（目的端主机名称可以在迁移 workflow 中单击“**创建目的端主机**”步骤查看），单击“**下一步：确认配置**”，进入**确认配置**页签。

步骤10 确认信息无误后，单击“**保存配置**”按钮，仔细阅读“迁移条件须知和风险提示”后，单击“**确定**”。

步骤11 返回MgC主机迁移 workflows，在“**调整磁盘分区**”步骤单击**步骤状态**后的“**确认**”继续进行后续迁移步骤。

----**结束**

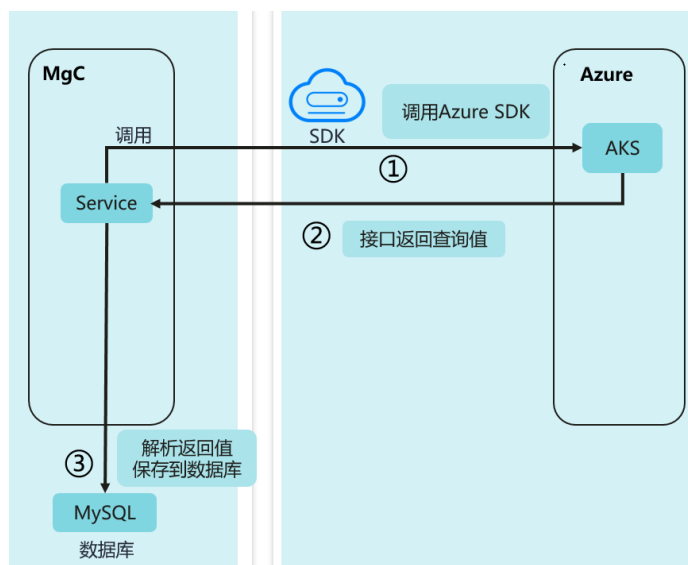
6 采集 Azure 容器 AKS 资源

在进行 Azure Kubernetes Service (AKS) 资源迁移前，通过 MgC 进行资源采集是关键步骤，它为后续迁移操作提供了必要的数据库支持。本节介绍使用 MgC 采集 AKS 资源的基本原理、准备工作、账号权限要求以及具体的操作步骤，旨在帮助用户高效、准确地完成 AKS 资源的数据采集工作。

采集原理

使用 MgC 采集 Azure 容器服务 AKS 资源的原理图，如图 6-1 所示。

图 6-1 采集 AKS 资源原理图



1. **调用SDK**：通过 Azure 提供的 SDK，MgC 调用 AKS API，获取容器、虚拟机等资源信息。
2. **接口返回**：MgC 接收 API 调用的返回值，这些返回值通常包含丰富的资源数据。
3. **解析数据**：MgC 将返回的数据进行解析，提取关键信息，如节点数、虚拟机核数等。然后将这些关键信息保存到数据库中，为后续分析和迁移提供支持。

准备工作

- **华为账号准备**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，注册华为账号以及创建IAM用户的方法请参考[准备工作](#)。
- **创建项目**
在MgC控制台为本次采集创建独立的项目（建议创建简单项目），方法请参考[项目管理](#)。
- **Azure认证信息准备**
提前获取AKS所在应用客户端的密码、购买AKS的订阅ID，对应应用的租户ID、客户端（应用）ID。这些认证信息的获取方法请参见[如何获取添加Azure凭证所需的信息？](#)
- **添加源端凭证**
在MgC控制台添加Azure认证信息作为采集凭证，方法请参考[添加凭证](#)。

权限要求

需要确保添加的Azure凭证所属的应用，在购买AKS资源的资源组和订阅里拥有如下权限：

- Microsoft.ClassicCompute/virtualMachines/read
- Microsoft.Insights/MetricDefinitions/Read
- Microsoft.Management/getEntities/action

以上权限的配置方法请参考[如何配置采集Azure容器资源所需权限？](#)

操作步骤

完成准备工作并确认Azure账号权限满足要求后，即可前往MgC控制台创建采集任务，具体操作步骤请参考[公网发现与采集](#)。

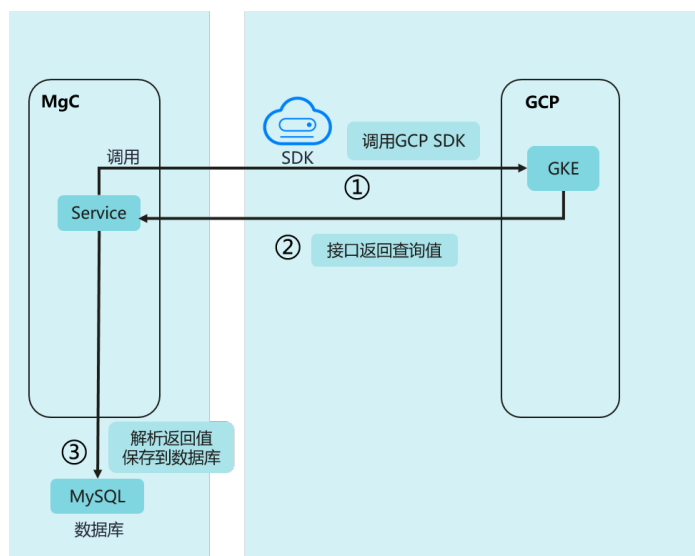
7 采集谷歌云 GCP 容器 GKE 资源

在进行Google Kubernetes Engine (GKE) 资源迁移前，通过MgC进行资源采集是关键步骤，它为后续迁移操作提供了必要的数据库支持。本节介绍使用MgC采集GKE资源的基本原理、准备工作、账号权限要求以及具体的操作步骤，旨在帮助用户高效、准确地完成GKE资源的数据采集工作。

采集原理

使用MgC采集GCP容器服务GKE资源的原理图，如图7-1所示。

图 7-1 采集 GKE 资源原理图



1. **调用SDK**：通过GCP提供的SDK，MgC调用GKE API，获取容器实例，Region，网络，子网等资源信息。
2. **接口返回**：MgC接收API调用的返回值，这些返回值通常包含丰富的资源数据。
3. **解析数据**：MgC将返回的数据进行解析，提取关键信息，如节点数、子网段等。然后将这些关键信息保存到数据库中，为后续分析和迁移提供支持。

准备工作

- **华为账号准备**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户。注册华为账号以及创建IAM用户的方法请参考[准备工作](#)。
- **创建项目**
在MgC控制台为本次采集创建独立的项目（建议创建简单项目），方法请参考[项目管理](#)。
- **GCP账号凭据准备**
提前获取GKE所属GCP账号的凭据（密钥文件）。
- **添加源端凭证**
在MgC控制台添加GCP账号的密钥文件作为采集凭证，方法请参考[添加凭证](#)。认证选择“配置文件”，并上传GCP服务账号的密钥文件，该文件必须为json格式，且大小不能超过4KB。

账号权限要求

在进行Google Kubernetes Engine（GKE）资源采集时，需要确保GCP账号拥有如下权限：

- container.clusters.list
- compute.regions.get
- compute.networks.list
- compute.subnetworks.list

操作步骤

完成准备工作并确认GCP账号权限满足要求后，即可前往MgC控制台创建采集任务，具体操作步骤请参考[公网发现与采集](#)。

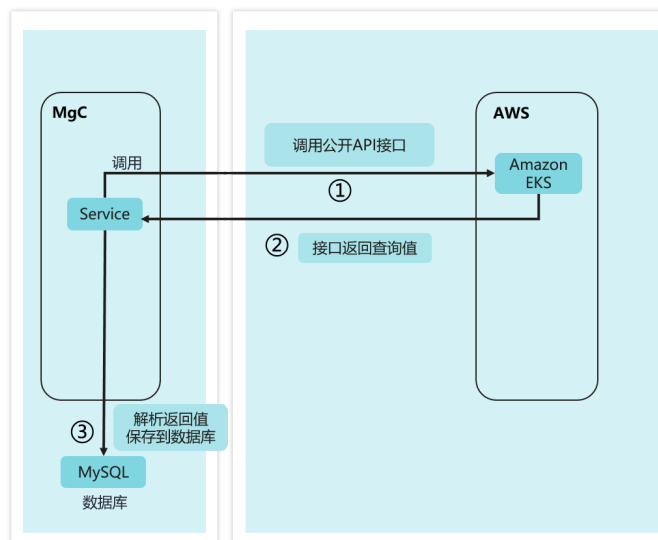
8 采集 AWS 容器资源

本节介绍使用MgC采集AWS 容器资源的基本原理、准备工作、账号权限要求以及具体的操作步骤，旨在帮助用户高效、准确地完成AWS 容器资源的数据采集工作。

公网发现原理

使用MgC通过公网发现AWS容器资源的原理图，如图8-1所示。

图 8-1 采集 AWS 容器资源原理图

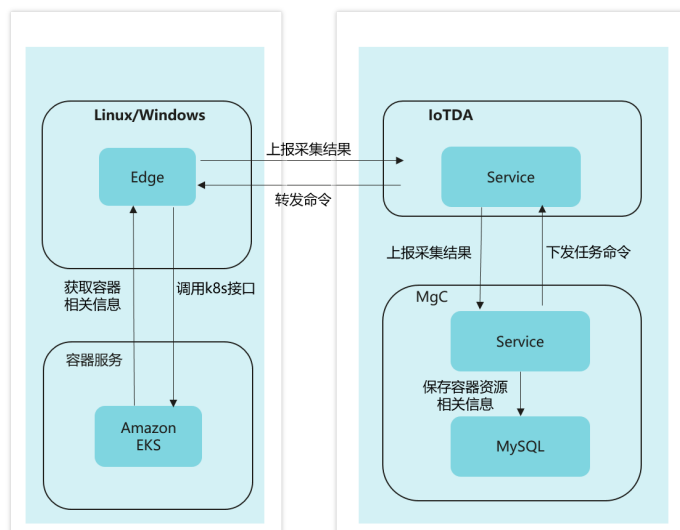


1. **调用SDK**：通过调用Amazon EKS服务提供的API，获取容器、虚拟机等资源信息。
2. **接口返回**：MgC接收API调用的返回值，这些返回值通常包含丰富的资源数据。
3. **解析数据**：MgC将返回的数据进行解析，提取关键信息，如节点数、虚拟机核数等。然后将这些关键信息保存到数据库中，为后续分析和迁移提供支持。

深度采集原理

MgC对AWS 容器资源进行深度采集的原理图，如图8-2所示。

图 8-2 AWS 容器深度采集原理图



对AWS容器资源进行深度采集的过程详细说明如下：

1. **迁移中心下发命令：**迁移中心向Edge发出采集容器资源信息的命令。
2. **Edge通过凭证登录容器集群：**Edge使用用户提供的凭证登录至目标容器集群中。
3. **K8s接口采集信息：**Edge通过Kubernetes（K8s）API接口与容器集群交互，采集容器集群的详细信息，包括容器的详细规格、节点配置、持久卷配置以及网络策略等。
4. **信息上报给迁移中心：**Edge将采集到的信息上报给迁移中心。
5. **迁移中心解析并保存信息：**迁移中心接收到上报的信息后，会进行解析，提取出有用的信息，并将其保存到数据库中。

准备工作

- **华为账号准备**
使用MgC之前，您需要拥有一个华为账号或一个可用于访问MgC的IAM用户，注册华为账号以及创建IAM用户的方法请参考[准备工作](#)。
- **创建项目**
在MgC控制台为本次采集创建独立的项目（建议创建简单项目），方法请参考[项目管理](#)。
- **AWS账号凭证准备**
提前获取待采集资源所属AWS账号的AK/SK，方法请参考[获取AWS访问密钥（Access keys）](#)。
- **添加源端凭证**
在MgC控制台添加AWS账号凭证作为采集凭证，方法请参考[添加凭证](#)。
- **获取AWS容器集群的登录配置文件**
配置文件将作为深度采集的凭证，以允许Edge通过提供的凭证登录至AWS容器集群中，并使用K8s接口采集容器集群的详细信息。

权限要求

在进行Amazon EKS资源采集时，需要确保AWS账号拥有如下权限：

- eks:DescribeCluster
- eks:ListClusters
- ec2:DescribeInstances
- ec2:DescribeSubnets
- cloudwatch:GetMetricStatistics

创建公网发现任务

完成准备工作并确认AWS账号凭证满足权限要求后，即可前往MgC控制台创建公网发现任务，具体操作步骤请参考[公网发现与采集](#)。其中，源端选择**AWS**；凭证选择准备工作中添加的源端凭证；区域根据实际需求选择；资源类型选择**容器**。

图 8-3 创建公网发现任务

任务配置
请选择源端平台、凭证及资源所在区域

源端 阿里云 华为云 AWS 腾讯云 Azure 七牛云 金山云

凭证
源端平台访问口令，用于发现源端资源

区域

资源发现
通过资源发现可以获得源端资源清单、采集资源信息，包括规格及属性，是分析资源与应用间关系的基础。
覆盖策略：资源发现默认采用覆盖已有策略，所发现的资源在同一迁移项目中不会重复出现。

云平台采集
通过调用源端云平台的公开接口，使用已经配置的凭证，查询所选资源的规格信息

资源类型

深度采集

通过[公网发现任务](#)，获取AWS容器资源列表后，即可对列表中的容器资源进行深度采集，以获取容器的详细规格、节点配置、持久卷配置以及网络策略等详细信息。具体操作步骤请参考[容器深度采集](#)。

在Edge添加采集凭证时，类型选择**容器**，配置文件上传AWS容器集群的登录配置文件。

图 8-4 在 Edge 添加深度采集凭证



创建凭证

* 使用类型 容器

* 类型 容器

* 凭证名称 请输入名称

* 认证方式 配置文件

* 选择文件 点击右侧按钮先添加再上传 添加文件 ①

取消 确认

9 大数据迁移后数据校验

9.1 MaxCompute 迁移至 DLI 数据校验

本文主要介绍将阿里云 MaxCompute 数据迁移至华为云数据湖探索 DLI 后，使用迁移中心 MgC 进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具 Edge 并注册用户，方法请参见 [安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端“**大数据-MaxCompute**”凭证和目的端“**大数据-数据湖搜索（DLI）**”凭证，认证方式选择“**AK/SK**”。凭证的添加方法请参考[添加资源凭证](#)。

- DLI侧的AK/SK获取方法请参考[如何获取AK/SK](#)。
- MaxCompute侧的AK/SK获取方法请参考[查看RAM用户的AccessKey信息](#)。

步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在元数据采集区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“**MaxCompute**”，单击“**下一步**”。

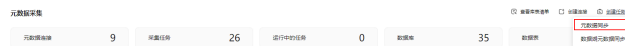


步骤8 根据**MaxCompute连接参数配置说明**，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-1 MaxCompute 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
阿里云凭证	选择 步骤4 在Edge侧添加的 MaxCompute 凭证。
项目名称	输入需要采集的MaxCompute项目名称。项目名称可在MaxCompute侧控制台获取。
Endpoint	输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 Endpoint对照表 。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据**元数据同步任务参数配置说明**，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。



表 9-2 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择 步骤8 创建的元数据连接。
数据库	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。 须知 元数据连接为MaxCompute类型时，该参数为必填参数，不能为空。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



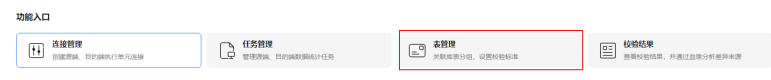
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。

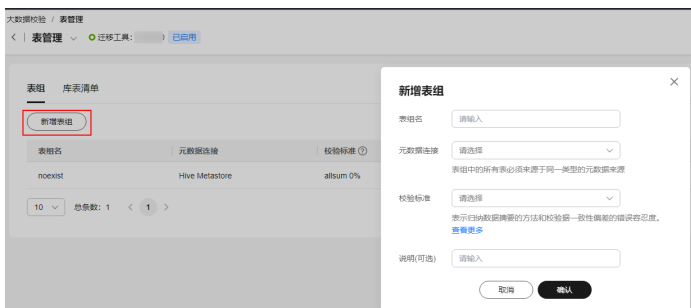


表 9-3 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“**连接管理**”，进入**连接管理**页面。



步骤19 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。连接类型选择“**数据湖探索 (DLI)**”，单击“**下一步**”，进入**详细配置**页面。



步骤20 根据[DLI服务参数配置说明](#)，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 9-4 DLI 服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“DLI-4位随机数（字母和数字）”，也可以自定义名称。
DLI凭证	选择 步骤4 在Edge侧添加的DLI凭证。如果所选凭证与登录迁移中心账号凭证相同，请勾选“ 所选凭证与登录迁移中心凭证相同 ”，系统将根据所选的区域，自动查询账号在该区域下的项目。
区域	输入需要校验的数据（即目的端数据）所在区域的代码。例如：ap-southeast-1。可以参考 终端节点 获取不同区域的区域代码信息。
项目	输入需要校验的数据所在的项目ID。获取方法请参考 获取项目的名称和ID 。

参数	配置说明
队列名称	输入需要进行校验的DLI队列名称。队列类型必须是SQL队列。

步骤21 在MgC控制台，创建**MaxCompute**统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择步骤[步骤16](#)创建的表组。

- 任务类型请选择大数据组件中的“**MaxCompute**”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤22 在MgC控制台，创建**数据湖探索 DLI**统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择步骤[步骤16](#)创建的表组。

- 任务类型请选择大数据组件中的“**数据湖探索 DLI**”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤23 当**MaxCompute**统计任务和**数据湖探索 DLI**统计任务分别执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.2 MRS ClickHouse 迁移至 MRS ClickHouse 数据校验

本文主要介绍MRS ClickHouse版本间数据迁移后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知
在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端MRS ClickHouse凭证和目的端MRS ClickHouse凭证，类型选择“大数据-ClickHouse”，认证方式选择“用户名/密码”。凭证的添加方法请参考[添加资源凭证](#)。

步骤5 在MgC控制台左侧导航栏选择“调研 > 大数据血缘”，进入[大数据血缘](#)页面。在页面左上角的当前项目下拉列表中选择[步骤2](#)创建的项目。



步骤6 在元数据采集区域，单击“创建连接”，右侧弹出创建连接窗口。



步骤7 连接类型选择“ClickHouse”，单击“下一步”。



步骤8 根据ClickHouse组件参数配置说明，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-5 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择步骤3.与MgC建立连接的Edge工具。
ClickHouse凭证（可选）	选择步骤4.在Edge添加的源端 MRS ClickHouse凭证。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到源端ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。
HTTP端口	非安全集群下，与源端ClickHouse 服务器通信的 HTTP 端口。 可以登录源端集群的FusionInsight Manager，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“http_port”参数获取。
HTTP SSL/TLS端口	安全集群下，与源端ClickHouse 服务器通信的HTTPS 端口。 可以登录源端集群的FusionInsight Manager，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“https_port”参数获取。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据[元数据同步任务参数配置说明](#)，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。



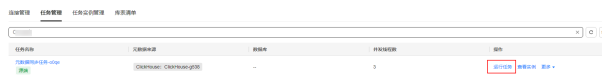
表 9-6 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择 步骤8 .创建的元数据连接。
数据库（可选）	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3， $1 \leq \text{取值范围} \leq 10$ 。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

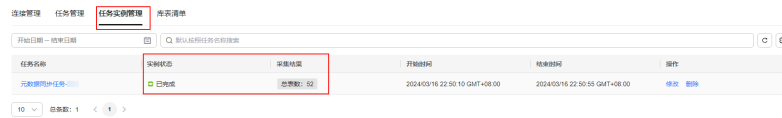
步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



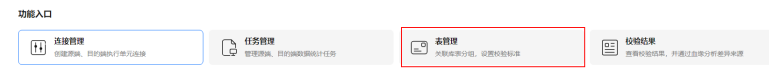
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据**新增表组参数说明**，配置参数后，单击“确认”按钮，创建表组。

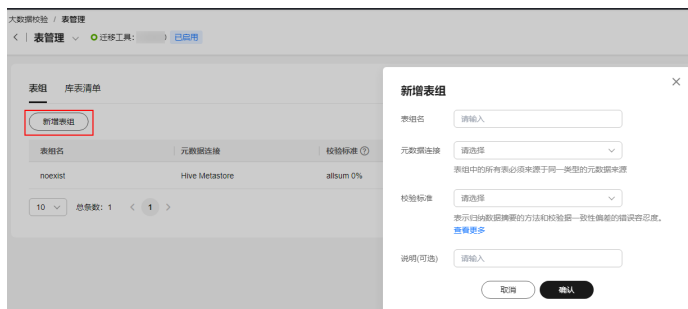


表 9-7 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。

参数	说明
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在**表管理**页面，选择**库表清单**页签，勾选要加入同一表组的数据表，单击列表上方的“**表组管理 > 批量加入组**”，在弹出的**修改表组**窗口中，选择要加入的表组，单击“**确认**”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在**大数据校验**页面的**功能入口**区域，单击“**连接管理**”，进入**连接管理**页面。



步骤19 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。连接类型选择“**ClickHouse**”，单击“**下一步**”，进入**详细配置**页面。

新增连接

1 选择连接类型 ———— 2 详细配置

执行机

执行机是获取Hive数据的任务执行单元，需要预先安装各大数据组件的客户端

MRS执行机

CDH执行机

大数据云服务

大数据相关云服务，提供数据查询、计算等服务

数据湖探索 (DLI)

Athena

大数据组件

提供大数据数据查询能力的组件

Doris

HBase

ClickHouse

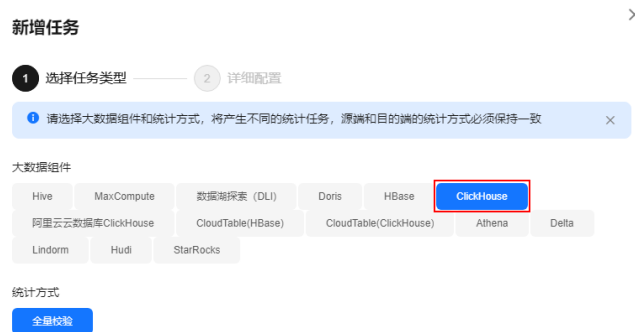
步骤20 根据[ClickHouse组件参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-8 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
ClickHouse凭证（可选）	选择 步骤4 .在Edge添加的目的端MRS ClickHouse凭证。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到MRS ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。
HTTP端口	非安全集群下，与MRS ClickHouse 服务器通信的 HTTP 端口。 可以登录目的端集群的 FusionInsight Manager ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“http_port”参数获取。
HTTP SSL/TLS端口	安全集群下，与MRS ClickHouse 服务器通信的HTTPS 端口。 可以登录目的端集群的 FusionInsight Manager ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“https_port”参数获取。

步骤21 分别创建源端和目的端MRS ClickHouse统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- **任务类型**请选择大数据组件中的“ClickHouse”。



- **统计方式**仅支持“全量校验”。

步骤22 当任务执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.3 阿里云 EMR ClickHouse 迁移至 MRS ClickHouse 数据校验

本文主要介绍将阿里云EMR ClickHouse数据迁移至华为云MRS ClickHouse后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加[阿里云EMR ClickHouse凭证](#)和目的端MRS ClickHouse凭证，类型选择“**大数据-ClickHouse**”，认证方式选择“**用户名/密码**”。凭证的添加方法请参考[添加资源凭证](#)。

阿里云EMR ClickHouse的凭证填写登录EMR ClickHouse Server的用户名和密码。您可以在EMR控制台ClickHouse服务配置页面的server-users页签，查看用户名和密码（users.default.password）。

步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在**元数据采集**区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“**ClickHouse**”，单击“**下一步**”。

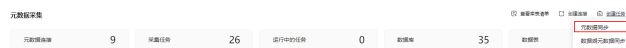


步骤8 根据[ClickHouse组件参数配置说明](#)，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 9-9 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
ClickHouse凭证（可选）	选择 步骤4 在Edge添加的 阿里云EMR ClickHouse凭证 。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到 阿里云EMR ClickHouse 服务器的IP地址。通常是运行ClickHouse的主机IP地址。
HTTP端口	非安全集群下，与阿里云EMR ClickHouse 服务器通信的HTTP 端口。 可以在EMR控制台ClickHouse服务的 配置 页面，单击 server-config 页签，查看“http_port”参数获取。
HTTP SSL/TLS端口	安全集群下，与阿里云EMR ClickHouse 服务器通信的HTTPS 端口。 可以在EMR控制台ClickHouse服务的 配置 页面，单击 server-config 页签，查看“http_port”参数获取。

步骤9 在**元数据采集**区域，单击“**创建任务 > 元数据同步**”按钮，右侧弹出**创建任务-元数据同步**窗口。



步骤10 根据**元数据同步任务参数配置说明**，配置元数据同步任务参数后，单击“**确认**”按钮，创建元数据同步任务。



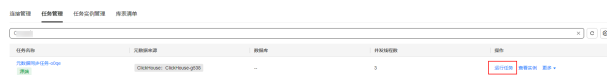
表 9-10 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择 步骤8 创建的元数据连接。
数据库（可选）	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



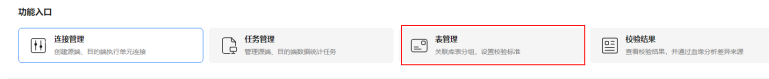
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。

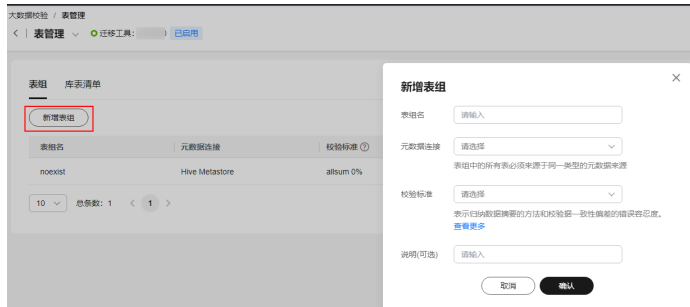


表 9-11 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“**连接管理**”，进入连接管理页面。



步骤19 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。连接类型选择“**ClickHouse**”，单击“**下一步**”，进入**详细配置**页面。

新增连接

1 选择连接类型 ———— 2 详细配置

执行机

执行机是获取Hive数据的任务执行单元，需要预先安装各大数据组件的命令客户端，如MRS客户端、CDH客户端

MRS执行机 CDH执行机

大数据云服务

大数据相关云服务，提供数据查询、计算等服务

数据湖探索 (DLI) Athena CloudTable(ClickHouse) CloudTable(HBase)

大数据组件

提供大数据数据查询能力的组件

Doris HBase **ClickHouse** StarRocks

步骤20 根据[ClickHouse组件参数配置说明](#)，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 9-12 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
ClickHouse凭证（可选）	选择 步骤4 .在Edge添加的目的端 MRS ClickHouse 凭证。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到MRS ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。

参数	配置说明
HTTP端口	非安全集群下，与MRS ClickHouse 服务器通信的 HTTP 端口。 可以登录目的端集群的 FusionInsight Manager ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“http_port”参数获取。
HTTP SSL/TLS端口	安全集群下，与MRS ClickHouse 服务器通信的HTTPS 端口。 可以登录目的端集群的 FusionInsight Manager ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“https_port”参数获取。

步骤21 分别创建源端阿里云EMR ClickHouse和目的端MRS ClickHouse统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- 任务类型请选择大数据组件中的“ClickHouse”。



- 统计方式仅支持“全量校验”。

步骤22 当任务执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.4 阿里云 ClickHouse 迁移至 MRS ClickHouse 数据校验

本文主要介绍将阿里云 ClickHouse数据迁移至华为云MRS ClickHouse后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在[总览](#)页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端阿里云 ClickHouse凭证和目的端MRS ClickHouse凭证，类型选择“大数据-ClickHouse”，认证方式选择“用户名/密码”。凭证的添加方法请参考[添加资源凭证](#)。

步骤5 在MgC控制台左侧导航栏选择“调研 > 大数据血缘”，进入大数据血缘页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在元数据采集区域，单击“创建连接”，右侧弹出创建连接窗口。



步骤7 连接类型选择“阿里云云数据库 ClickHouse”，单击“下一步”。

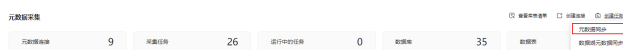


步骤8 根据[ClickHouse组件参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-13 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“阿里云云数据库ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
ClickHouse凭证（可选）	选择 步骤4 在Edge添加的源端阿里云 ClickHouse凭证。
数据库链接	填写连接源端ClickHouse集群的外网地址。可以在 集群信息 页面，查看外网地址。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据[元数据同步任务参数配置说明](#)，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。



表 9-14 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择 步骤8 创建的元数据连接。
数据库（可选）	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



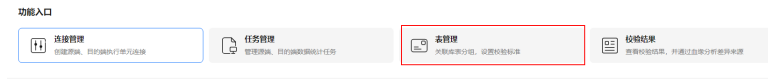
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。

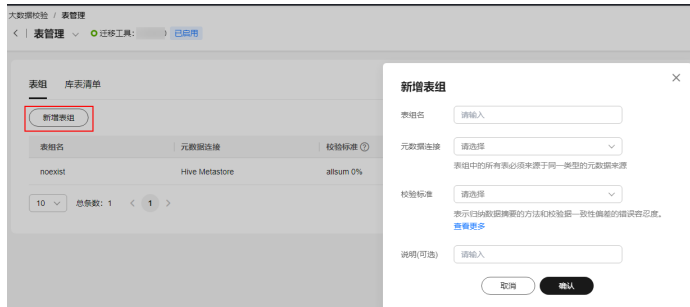


表 9-15 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



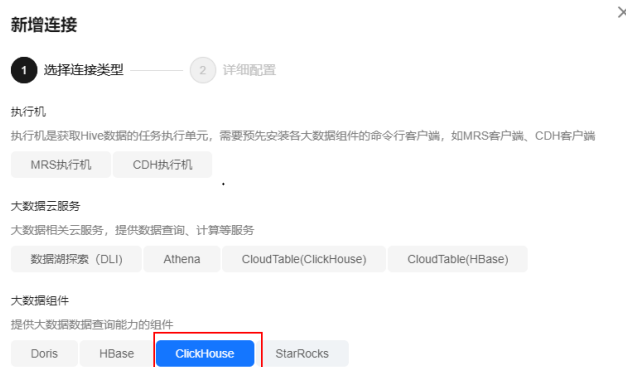
须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“连接管理”，进入连接管理页面。



步骤19 单击页面右上角的“新增连接”，右侧弹出新增连接窗口。连接类型选择“ClickHouse”，单击“下一步”，进入详细配置页面。



步骤20 根据[ClickHouse组件参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

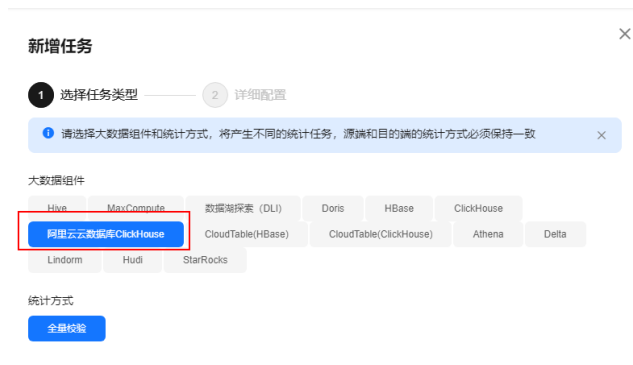
表 9-16 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
ClickHouse凭证（可选）	选择 步骤4 在Edge添加的目的端MRS ClickHouse凭证。
安全集群	选择是否部署在安全集群下。
ClickHouse Server IP	填写连接到MRS ClickHouse服务器的IP地址。通常是运行ClickHouse的主机IP地址。
HTTP端口	非安全集群下，与MRS ClickHouse 服务器通信的 HTTP 端口。 可以登录目的端集群的 FusionInsight Manager ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“http_port”参数获取。

参数	配置说明
HTTP SSL/TLS端口	安全集群下，与MRS ClickHouse 服务器通信的HTTPS 端口。 可以登录目的端集群的 FusionInsight Manager ，选择“集群 > 服务 > ClickHouse > 配置 > 全部配置”，搜索“https_port”参数获取。

步骤21 创建阿里云云数据库 ClickHouse统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- **任务类型**请选择大数据组件中的“阿里云云数据库 ClickHouse”。



- 统计方式仅支持“**全量校验**”。

步骤22 创建MRS ClickHouse统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- **任务类型**请选择大数据组件中的“**ClickHouse**”。



- 统计方式仅支持“**全量校验**”。

步骤23 当阿里云云数据库 ClickHouse统计任务和MRS ClickHouse统计任务执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.5 阿里云 ClickHouse 迁移至 CloudTable ClickHouse 数据校验

本文主要介绍将阿里云 ClickHouse数据迁移至华为云CloudTable ClickHouse后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端阿里云 ClickHouse凭证和目的端 CloudTable ClickHouse凭证，类型选择“大数据-ClickHouse”，认证方式选择“用户名/密码”。凭证的添加方法请参考[添加资源凭证](#)。

步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在元数据采集区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“**阿里云云数据库 ClickHouse**”，单击“**下一步**”。



步骤8 根据**ClickHouse组件参数配置说明**，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 9-17 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“阿里云云数据库ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
ClickHouse凭证（可选）	选择 步骤4 在Edge添加的源端 阿里云 ClickHouse 凭证。
数据库链接	填写连接源端ClickHouse集群的外网地址。可以在 集群信息 页面，查看外网地址。

步骤9 在元数据采集区域，单击“**创建任务 > 元数据同步**”按钮，右侧弹出**创建任务-元数据同步**窗口。



步骤10 根据**元数据同步任务参数配置说明**，配置元数据同步任务参数后，单击“**确认**”按钮，创建元数据同步任务。



表 9-18 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择 步骤8 创建的元数据连接。
数据库（可选）	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3， $1 \leq \text{取值范围} \leq 10$ 。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



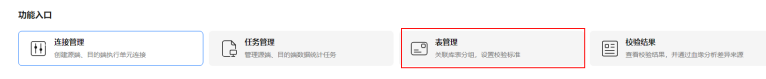
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。

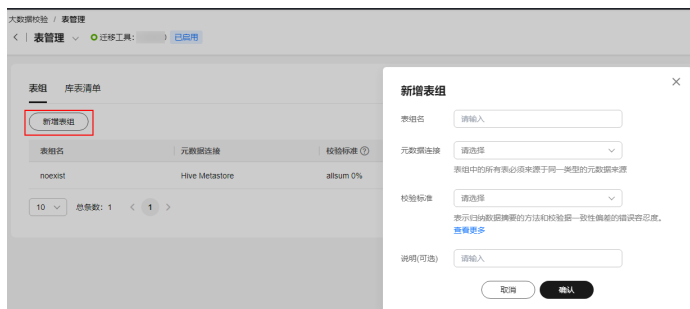


表 9-19 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择步骤8.创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“查看更多”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



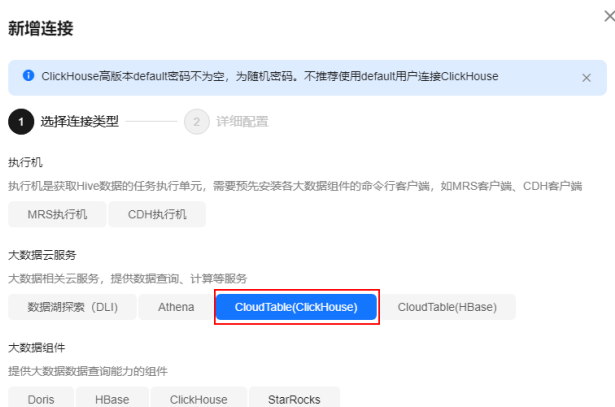
须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“连接管理”，进入连接管理页面。



步骤19 单击页面右上角的“新增连接”，右侧弹出新增连接窗口。连接类型选择“CloudTable(ClickHouse)”，单击“下一步”，进入详细配置页面。



步骤20 根据[ClickHouse组件参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

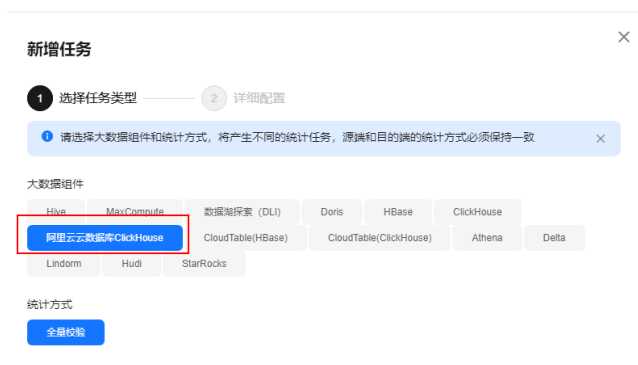
表 9-20 ClickHouse 组件参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“CloudTable-ClickHouse-4位随机数（字母和数字）”，也可以自定义名称。

参数	配置说明
CloudTable (ClickHouse)凭证	选择 步骤4 .在Edge添加的目的端 CloudTable ClickHouse 凭证。
数据库链接	填写连接到CloudTable(ClickHouse)的数据库链接地址。可以在ClickHouse集群的 基本信息 获取，对应参数为“ 集群访问地址 ”。

步骤21 创建**阿里云云数据库 ClickHouse**统计任务并执行，方法请参考**创建统计任务并执行校验**。需要注意任务配置中的表组，选择**步骤16**.创建的表组。

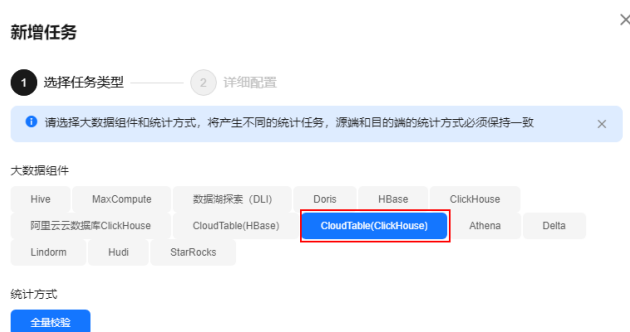
- **任务类型**请选择大数据组件中的“**阿里云云数据库 ClickHouse**”。



- 统计方式仅支持“**全量校验**”。

步骤22 创建**CloudTable(ClickHouse)**统计任务并执行，方法请参考**创建统计任务并执行校验**。需要注意任务配置中的表组，选择**步骤16**.创建的表组。

- **任务类型**请选择大数据组件中的“**CloudTable(ClickHouse)**”。



- 统计方式仅支持“**全量校验**”。

步骤23 当**阿里云云数据库 ClickHouse**统计任务和**CloudTable(ClickHouse)**统计任务执行完成后（实例状态为“已完成”），在**校验结果**页面可以查看并导出任务的执行结果，方法请参考**查看并导出校验结果**。

----结束

9.6 MRS Doris 迁移至 MRS Doris 数据校验

本文主要介绍MRS Doris版本间数据迁移后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知
在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端 MRS Doris凭证和目的端MRS Doris凭证，类型选择“大数据-Doris”，认证方式选择“用户名/密码”。凭证的添加方法请参考[添加资源凭证](#)。

步骤5 在MgC控制台左侧导航栏选择“调研 > 大数据血缘”，进入大数据血缘页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在元数据采集区域，单击“创建连接”，右侧弹出创建连接窗口。



步骤7 连接类型选择“Doris”，单击“下一步”。



步骤8 根据[Doris连接参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-21 Doris 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
Doris凭证	选择 步骤4 在Edge侧添加的源端MRS Doris凭证。
数据库地址	填写连接到源端Doris数据库的Doris FE实例IP地址。 Doris FE实例IP地址可通过登录MRS集群的Manager界面，单击“集群 > 服务 > Doris > 实例”，查看任一FE实例的IP地址。
数据库端口	填写连接到源端Doris数据库的连接端口。 数据库连接端口为Doris FE的查询连接端口，可以通过登录Manager，单击“集群 > 服务 > Doris > 配置”，查询Doris服务的“query_port”参数获取。
数据库名	填写源端Doris数据库名称。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据[元数据同步任务参数配置说明](#)，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。

表 9-22 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择 步骤8 创建的元数据连接。
数据库	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。

步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



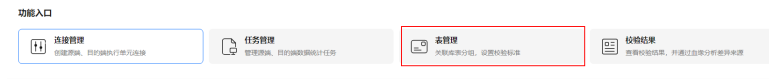
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据[新增表组参数说明](#)，配置参数后，单击“确认”按钮，创建表组。

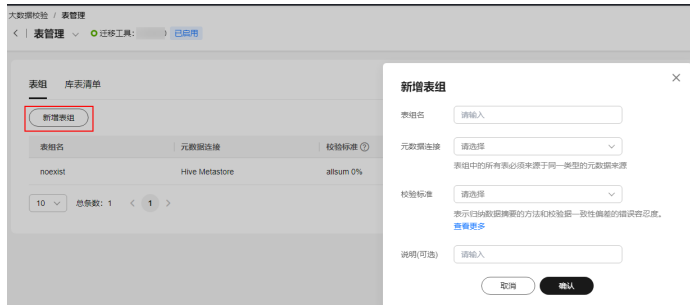


表 9-23 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“连接管理”，进入连接管理页面。



步骤19 单击页面右上角的“新增连接”，右侧弹出新增连接窗口。连接类型选择“Doris”，单击“下一步”，进入详细配置页面。



步骤20 根据表9-24，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-24 Doris 服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“Doris-4位随机数（字母和数字）”，也可以自定义名称。
Doris凭证	选择步骤4.在Edge侧添加的目的端MRS Doris凭证。
数据库地址	填写连接到目的端Doris数据库的Doris FE实例IP地址。 Doris FE实例IP地址可通过登录MRS集群的Manager界面，单击“集群 > 服务 > Doris > 实例”，查看任一FE实例的IP地址。
数据库端口	填写连接到目的端Doris数据库的连接端口。 数据库连接端口为Doris FE的查询连接端口，可以通过登录Manager，单击“集群 > 服务 > Doris > 配置”，查询Doris服务的“query_port”参数获取。
数据库名	填写待校验数据所在的MRS Doris数据库名称。

步骤21 在MgC控制台，创建Doris统计任务并执行，方法请参考创建统计任务并执行校验。需要注意任务配置中的表组，选择步骤步骤16.创建的表组。

- 任务类型请选择大数据组件中的“Doris”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤22 当任务执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.7 MRS Hive/CDH/EMR 迁移至 MRS Hive 数据校验

本文主要介绍MRS Hive版本间数据迁移以及将自建CDH/EMR环境上的数据迁移至华为云MRS Hive之后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“[创建迁移项目](#)”。



步骤2 在弹出的窗口中，自定义[项目名称](#)，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“[确认](#)”。

须知

在左侧导航栏的“[配置管理 > 项目管理](#)”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端“**MRS/CDH执行机**”凭证和目的端“**MRS执行机**”凭证，认证方式选择“**用户名/密码**”。凭证的添加方法请参考[添加资源凭证](#)。

须知

如果源端为**MRS Hive**且在**安全集群**下（开启Kerberos认证），还需要在Edge上添加**源端Hive Metastore**凭证，类型选择“**大数据-Hive Metastore**”，认证方式选择“**用户名/密钥**”。凭证文件需要同时上传core-site.xml、hivemetastore-site.xml、hive-site.xml、krb5.conf、user.keytab五个文件。获取方式参见[大数据-Hive Metastore凭证文件获取方法](#)。

步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在**元数据采集**区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“**Hive Metastore**”，单击“**下一步**”。

创建连接



步骤8 根据[Hive Metastore连接参数配置说明](#)，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-25 Hive Metastore 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Hive-Metastore-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
启用安全连接	选择是否启用安全连接。 <ul style="list-style-type: none"> MetaStore在非安全集群下，可以不启用安全连接。 MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证选择步骤4在Edge侧所添加的源端Hive Metastore凭证。
Hive版本	选择源端Hive的实际版本。
MetaStore节点IP	填写连接MetaStore的IP地址。
MetaStore Thrift服务端口	填写连接MetaStore的端口，默认为9083。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据[元数据同步任务参数配置说明](#)，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。

创建任务-元数据同步

任务名称

元数据同步任务-tyl9

默认创建名称为“任务类型-4位随机数（字母和数字）”

元数据连接

Hive-Metastore-pln8

数据库（可选）

多个数据库名用英文逗号拼接

0/1,000

并发线程数

3

元数据同步任务将采用多线程并发进行，增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）宿主计算机资源

表 9-26 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择步骤8.创建的元数据连接。
数据库（可选）	填写需要进行元数据同步的数据库名称。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

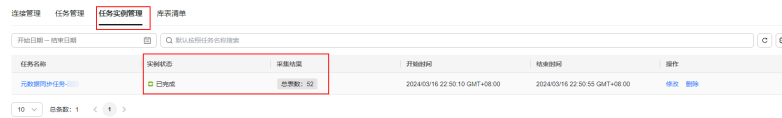
步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



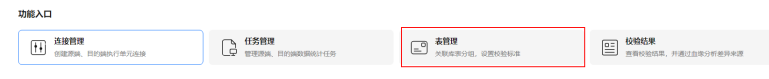
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据**新增表组参数说明**，配置参数后，单击“确认”按钮，创建表组。

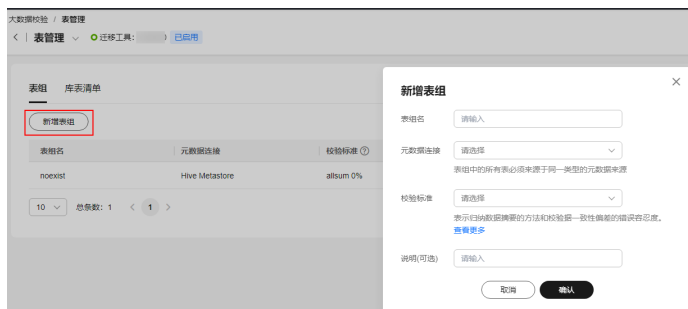


表 9-27 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。

参数	说明
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



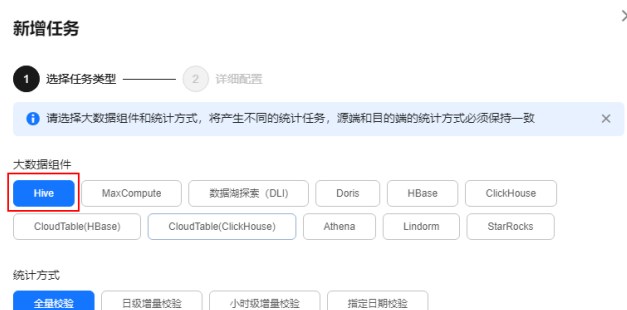
须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 分别创建源端执行机连接和目的端执行机连接，方法请参考[创建执行机连接](#)。执行机凭证分别选择[步骤4](#)在Edge侧添加的源端和目的端执行机凭证。

步骤19 在MgC控制台，分别创建源端和目的端Hive统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- **任务类型**请选择大数据组件中的“Hive”。



- **统计方式**按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤20 当源端与目的端的统计任务都执行完成后（实例状态为“已完成”），在校验结果页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.8 MaxCompute 迁移至 MRS Hive 数据校验

本文主要介绍将阿里云 MaxCompute 数据迁移至华为云 MRS Hive 后，使用迁移中心 MgC 进行数据一致性校验的步骤和方法。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具 Edge 并注册用户，方法请参见 [安装 Linux 版本的 Edge 工具](#)。

操作步骤

步骤1 登录 [迁移中心](#) 管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考 [连接迁移中心](#)，将 Edge 工具与云端 MgC 控制台建立连接。

步骤4 连接成功后，需要在 Edge 上分别添加源端“**大数据-MaxCompute**”凭证、目的端“**大数据-Hive Metastore**”凭证和目的端“**MRS 执行机**”凭证，认证方式分别选择“**AK/SK**”、“**用户名/密钥**”、“**用户名/密码**”。凭证的添加方法请参考 [添加资源凭证](#)。

- **MaxCompute**侧的AK/SK获取方法请参考[查看RAM用户的AccessKey信息](#)。
- **Hive Metastore**凭证文件的获取方法请参考[大数据-Hive Metastore凭证文件获取方法](#)。

步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在元数据采集区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“**MaxCompute**”，单击“**下一步**”。

创建连接

1 选择连接类型 ———— 2 详细配置

大数据组件

提供大数据元数据管理服务的大数据组件

Hive Metastore Doris HBase ClickHouse Glue

大数据云服务

大数据相关云服务，提供数据查询、计算等服务

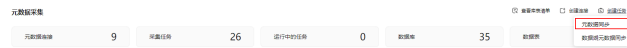
MaxCompute

步骤8 根据[MaxCompute连接参数配置说明](#)，配置连接参数后，单击“**测试连接**”按钮，测试通过代表连接成功。

表 9-28 MaxCompute 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“MaxCompute-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择 步骤3 与MgC建立连接的Edge工具。
阿里云凭证	选择 步骤4 在Edge侧添加的 MaxCompute 凭证。
项目名称	输入需要采集的MaxCompute项目名称。项目名称可在MaxCompute侧控制台获取。
Endpoint	输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 Endpoint对照表 。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据元数据同步任务参数配置说明，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。



表 9-29 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择步骤8.创建的元数据连接。
数据库	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。 须知 元数据连接为MaxCompute类型时，该参数为必填参数，不能为空。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。

步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



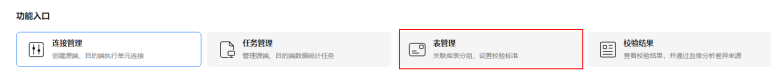
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择步骤3.与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。

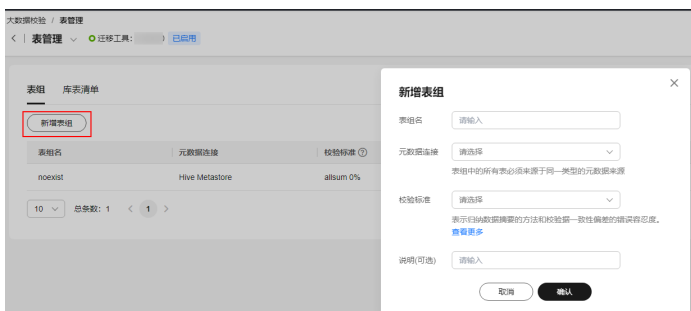


表 9-30 新增表组参数说明

参数	说明
表组名	用户自定义

参数	说明
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“连接管理”，进入连接管理页面。



步骤19 单击页面右上角的“新增连接”，右侧弹出新增连接窗口。连接类型选择“MRS执行机”，单击“下一步”，进入详细配置页面。



步骤20 根据表9-31，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-31 MRS 执行机参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“执行机类型-4位随机数（字母和数字）”，也可以自定义名称。
执行机凭证	选择步骤4.在Edge侧添加的目的端MRS执行机凭证。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口，默认为22端口。
安装目录	填写MRS客户端的安装目录路径。即安装客户端时，“./install.sh”所填写的地址。
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

步骤21 在MgC控制台，创建Hive统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择步骤步骤16.创建的表组；执行机连接选择步骤20.创建的MRS执行机连接。

- **任务类型**请选择大数据组件中的“Hive”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤22 当任务执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

---结束

9.9 MRS HBase 迁移至 MRS HBase 数据校验

本文主要介绍MRS HBase版本间数据迁移后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

准备工作

- 需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。
- 需要将源端和目的端所有节点的主机名和IP地址映射添加到安装Edge主机的/etc/hosts文件中，配置连接才能连接成功。
 - 在安装Edge工具的主机上，打开/etc/hosts文件进行编辑。
 - 在/etc/hosts文件中，为每个源端和目的端节点添加一行，格式如下：
IP地址 主机名
例如，节点的IP地址是192.168.1.1，主机名是source-node-01，则添加：
192.168.1.1 source-node-01
 - 完成所有映射的添加后，保存并关闭/etc/hosts文件。
 - 为了确保配置正确，可以通过ping命令测试主机名解析是否成功。例如：
ping source-node-01

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“[创建迁移项目](#)”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“[确认](#)”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端MRS HBase凭证和目的端MRS HBase凭证，类型选择“**大数据-HBase**”，认证方式选择“**用户名/密钥**”。凭证的添加方法请参考[添加资源凭证](#)。其中用户名输入HBase客户端登录用户名，密钥需要上传以下配置文件：

- 非安全集群（未开启Kerberos认证）需要上传core-site.xml，hdfs-site.xml，yarn-site.xml，mapred-site.xml和hbase-site.xml等5个配置文件。配置文件的获取方法请参考[准备运行环境配置文件](#)。集群是否开启Kerberos认证，可通过[查看集群基本信息](#)来确认。
- 安全集群（开启Kerberos认证）需要上传core-site.xml，hdfs-site.xml，yarn-site.xml，krb5.conf，user.keytab，mapred-site.xml和hbase-site.xml等7个文件。其中krb5.conf，user.keytab两个文件是集群用户的凭证，获取方法如下。其余5个配置文件的获取方法请参考[准备运行环境配置文件](#)。
 - a. [登录FusionInsight Manager界面](#)，选择“系统 > 权限 > 用户”。
 - b. 在用户名中选择developuser，单击操作“更多 > 下载认证凭据”下载认证凭据文件。
 - c. 将下载的凭据文件解压即可得到用户的“user.keytab”文件与“krb5.conf”文件。

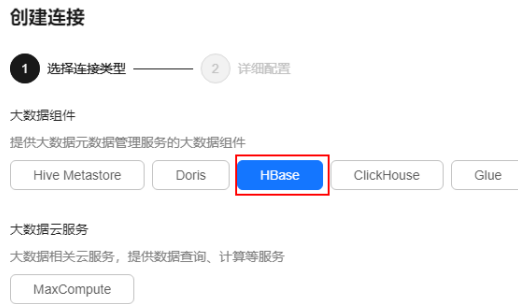
步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在元数据采集区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“HBase”，单击“下一步”。



步骤8 根据表9-32，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-32 HBase 连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择步骤3.与MgC建立连接的Edge工具。
HBase凭证	选择步骤4.在Edge添加的源端MRS HBase凭证。
安全集群	选择是否部署在安全集群下。
Zookeeper节点IP	填写连接到Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。
Zookeeper连接端口	填写连接到Zookeeper的端口。
HBase 版本	选择HBase的版本。

步骤9 在元数据采集区域，单击“创建任务 > 元数据同步”按钮，右侧弹出创建任务-元数据同步窗口。



步骤10 根据元数据同步任务参数配置说明，配置元数据同步任务参数后，单击“确认”按钮，创建元数据同步任务。

表 9-33 元数据同步任务参数配置说明

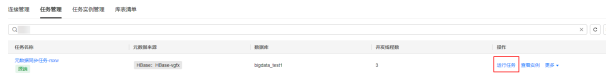
参数	配置说明
任务名称	默认创建名称为“元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择步骤8.创建的元数据连接。

参数	配置说明
数据库（可选）	填写需要进行元数据同步的数据库名称，如果有多个数据库，库名之间用逗号隔开。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3，1≤取值范围≤10。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



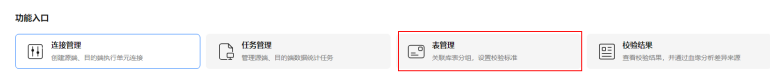
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。



表 9-34 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择步骤8.创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“查看更多”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 在大数据校验页面的功能入口区域，单击“连接管理”，进入连接管理页面。



步骤19 单击页面右上角的“新增连接”，右侧弹出新增连接窗口。连接类型选择“HBase”，单击“下一步”，进入详细配置页面。



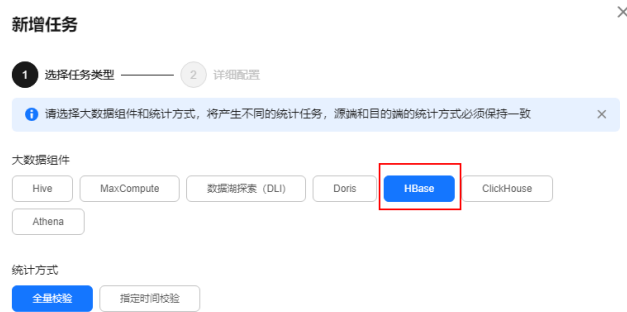
步骤20 根据表9-35，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-35 HBase 服务参数配置说明

参数	配置说明
所属源	选择目的端。
连接名称	默认创建名称为“HBase-4位随机数（字母和数字）”，也可以自定义名称。
HBase凭证	选择步骤4.在Edge侧添加的目的端MRS HBase凭证。
安全集群	选择是否部署在安全集群下。
Zookeeper节点IP	填写连接到Zookeeper的IP地址。通常是部署Zookeeper的服务器公网IP地址或内网IP地址。
Zookeeper连接端口	填写连接到Zookeeper的端口。默认为2181。
HBase 版本	选择HBase的版本。

步骤21 在MgC控制台，创建HBase统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择步骤[步骤16](#).创建的表组。

- **任务类型**请选择大数据组件中的“HBase”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤22 当任务执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.10 DeltaLake（有元数据）迁移至 MRS DeltaLake 数据校验

本文主要介绍将自建DeltaLake环境上的数据迁移至华为云MRS DeltaLake之后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

须知

DeltaLake存在元数据，可通过数据湖元数据同步方式进行数据采集。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。

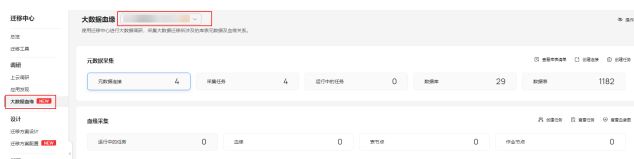


步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端“**Delta Lake执行机**”凭证和目的端“**MRS Delta Lake执行机**”凭证，类型选择“**大数据-执行机**”，认证方式选择“**用户名/密码**”。凭证的添加方法请参考[添加资源凭证](#)。



步骤5 在MgC控制台左侧导航栏选择“**调研 > 大数据血缘**”，进入**大数据血缘**页面。在页面左上角的当前项目下拉列表中选择**步骤2**创建的项目。



步骤6 在**元数据采集**区域，单击“**创建连接**”，右侧弹出**创建连接**窗口。



步骤7 连接类型选择“**Delta Lake（有元数据）**”，单击“**下一步**”。



步骤8 根据表9-36，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-36 Delta Lake（有元数据）连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Delta-Lake-有元数据-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择步骤3.与MgC建立连接的Edge工具。
执行机凭证	选择步骤4.在Edge侧添加的源端Delta Lake执行机凭证。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口，默认为22端口。
Spark客户端目录	填写Spark客户端的bin目录的绝对路径
环境变量地址	请添加环境变量文件的绝对路径，示例：/opt/bigdata/client/bigdata_env，如果该参数不为空，则环境变量文件在命令执行前将被自动“source”
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

步骤9 在元数据采集区域，单击“创建任务 > 数据湖元数据同步”按钮，右侧弹出创建任务-数据湖元数据同步窗口。



步骤10 根据表9-37，配置数据湖元数据同步任务参数后，单击“确认”按钮，创建数据湖元数据同步任务。



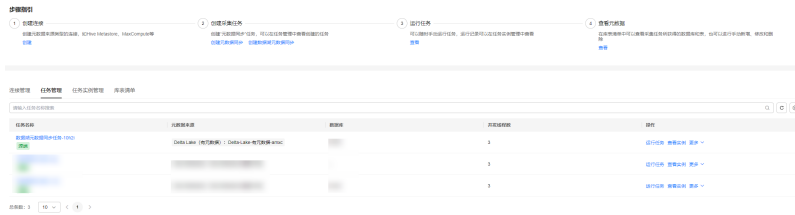
表 9-37 元数据同步任务参数配置说明

参数	配置说明
任务名称	默认创建名称为“数据湖元数据同步任务-4位随机数（字母和数字）”，也可以自定义名称。
元数据连接	选择步骤8.创建的元数据连接。
数据库（可选）	填写需要进行元数据同步的数据库名称。不填则代表同步全部数据库。
并发线程数	设置采集线程并发数。默认为3， $1 \leq \text{取值范围} \leq 10$ 。增加线程并发数，可以提高同步效率，但会占用更多源端连接和迁移工具（Edge）所安装主机的计算资源。
自定义参数	可以通过自定义参数设置同步的特定表或分区、过滤条件等。 <ul style="list-style-type: none"> ● 如果需要同步的是阿里云 EMR环境中的元数据，请添加如下参数： <ul style="list-style-type: none"> - 参数名：conf - 参数值：spark.sql.catalogImplementation=hive ● 如果源端是阿里云 EMR Delta 2.2版本并且是通过Delta 2.3版本依赖访问时，请添加如下参数： <ul style="list-style-type: none"> - 参数名：master - 参数值：local

步骤11 在元数据采集区域，单击“采集任务”，进入元数据采集页面。



步骤12 在任务管理页签，可以查看已创建的元数据同步任务列表以及参数配置信息。单击操作列的“运行任务”，即可触发一次任务执行。每执行一次任务，会生成一条任务实例。



步骤13 单击操作列的“查看实例”，进入任务实例管理页签。可以查看该任务的运行记录，以及每条任务实例的状态和采集结果。当元数据同步任务的实例状态为“已完成”且采集结果有数值时，在库表清单页签可以查看采集到的元数据列表。



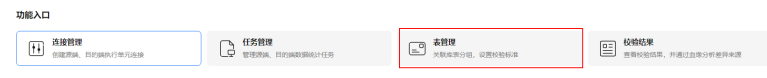
步骤14 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择步骤3.与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤15 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤16 单击“新增表组”，弹出新增表组窗口。根据新增表组参数说明，配置参数后，单击“确认”按钮，创建表组。

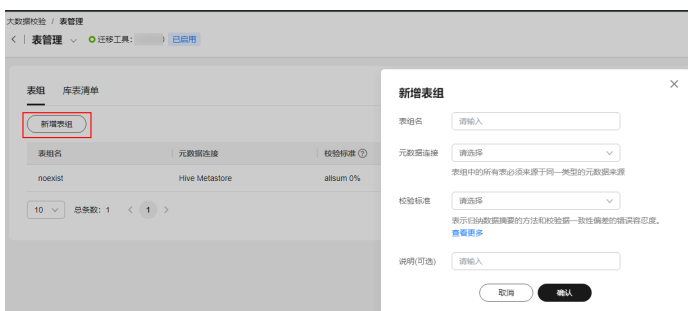


表 9-38 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤17 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤18 分别创建源端执行机连接和目的端执行机连接，方法请参考[创建执行机连接](#)。执行机凭证分别选择[步骤4](#)在Edge侧添加的源端和目的端执行机凭证。

步骤19 在MgC控制台，分别创建源端和目的端Delta统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- **任务类型**请选择大数据组件中的“Delta”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤20 当源端与目的端的统计任务都执行完成后（实例状态为“已完成”），在校验结果页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

9.11 DeltaLake（无元数据）迁移至 MRS DeltaLake 数据校验

本文主要介绍将自建DeltaLake环境上的数据迁移至华为云MRS DeltaLake之后，使用迁移中心 MgC进行数据一致性校验的步骤和方法。

须知

DeltaLake不存在元数据，无法通过数据湖元数据同步方式进行数据采集，需要通过导入数据的方式获取元数据信息。

准备工作

需要先在源端内网环境中安装用于配合数据校验的工具Edge并注册用户，方法请参见[安装Linux版本的Edge工具](#)。

操作步骤

步骤1 登录[迁移中心](#)管理控制台。在总览页面，单击“创建迁移项目”。



步骤2 在弹出的窗口中，自定义项目名称，仔细查看并了解项目类型的适用场景后，选择项目类型，单击“确认”。

须知

在左侧导航栏的“配置管理 > 项目管理”页签可以查看和管理已创建的项目。



步骤3 创建项目完成后，参考[连接迁移中心](#)，将Edge工具与云端MgC控制台建立连接。

步骤4 连接成功后，需要在Edge上分别添加源端“Delta Lake执行机”凭证和目的端“MRS Delta Lake执行机”凭证，类型选择“大数据-执行机”，认证方式选择“用户名/密码”。凭证的添加方法请参考[添加资源凭证](#)。



步骤5 在MgC控制台左侧导航栏选择“调研 > 大数据血缘”，进入大数据血缘页面。在页面左上角的当前项目下拉列表中选择[步骤2](#)创建的项目。



步骤6 在元数据采集区域，单击“创建连接”，右侧弹出创建连接窗口。



步骤7 连接类型选择“Delta Lake（无元数据）”，单击“下一步”。



步骤8 根据表9-39，配置连接参数后，单击“测试连接”按钮，测试通过代表连接成功。

表 9-39 Delta Lake（无元数据）连接参数配置说明

参数	配置说明
所属源	选择源端。
连接名称	默认创建名称为“Delta-Lake-无元数据-4位随机数（字母和数字）”，也可以自定义名称。
迁移工具	选择步骤3.与MgC建立连接的Edge工具。
执行机凭证	选择步骤4.在Edge侧添加的源端Delta Lake执行机凭证。
执行机IP	填写连接执行机的IP地址。
执行机连接端口	填写连接执行机的端口，默认为22端口。
Spark客户端目录	填写Spark客户端的bin目录的绝对路径
环境变量地址	请添加环境变量文件的绝对路径，示例：/opt/bigdata/client/bigdata_env，如果该参数不为空，则环境变量文件在命令执行前将被自动“source”
SQL文件地址	填写用于存放执行作业时生成的SQL文件的文件夹地址。您需要拥有该文件夹的读写权限。 须知 迁移完成后，您需要手动清理该路径下生成的文件夹，以释放存储空间。

步骤9 在元数据采集区域，单击“查看库表清单”或“数据表”，进入库表清单页签。



步骤10 单击“导入数据表”按钮，弹出导入数据表窗口。



步骤11 单击“xlsx模板下载”，将导入模板下载到本地。打开模板填写对应信息后保存。

步骤12 单击导入数据表窗口中的“添加文件”按钮，选择填写后的模板文件进行上传，然后单击“确认”按钮，完成数据表导入。导入完成后，在库表清单页签可以查看导入的数据列表。



步骤13 在左侧导航栏选择“实施>大数据校验”，进入大数据校验页面。首次使用大数据校验功能时，需要先选择并启用迁移工具。单击“选择并启用迁移工具”按钮，弹出选择迁移工具窗口。在下拉列表中，选择**步骤3**与迁移中心连接成功的Edge。



注意

在进行大数据校验时，请确保迁移工具Edge一直处于在线并启用状态。

步骤14 在功能入口区域，单击“表管理”，进入表管理页面的表组页签。



步骤15 单击“新增表组”，弹出新增表组窗口。根据**新增表组参数说明**，配置参数后，单击“确认”按钮，创建表组。

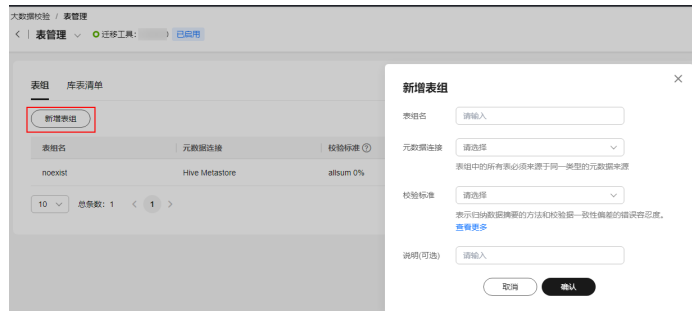


表 9-40 新增表组参数说明

参数	说明
表组名	用户自定义
元数据连接	选择 步骤8 创建的元数据连接。 注意 表组中的所有表必须来源于同一个元数据来源。
校验标准	选择可以接受的数据差异最大值（即错误容忍度）。MgC提供了多种类型的对数标准可供选择，关于对数标准的详细说明，可以单击新增表组弹窗中的“ 查看更多 ”进行查看。
说明（可选）	您可以输入用于区分表组的说明。

步骤16 在表管理页面，选择库表清单页签，勾选要加入同一表组的数据表，单击列表上方的“表组管理 > 批量加入组”，在弹出的修改表组窗口中，选择要加入的表组，单击“确认”。



须知

如果需要新增数据表，可以通过导入数据表的方式添加，方法请参考[导入数据表并加入表组](#)。

步骤17 分别创建源端执行机连接和目的端执行机连接，方法请参考[创建执行机连接](#)。执行机凭证分别选择[步骤4](#)在Edge侧添加的源端和目的端执行机凭证。

步骤18 在MgC控制台，分别创建源端和目的端Delta统计任务并执行，方法请参考[创建统计任务并执行校验](#)。需要注意任务配置中的表组，选择[步骤16](#)创建的表组。

- **任务类型**请选择大数据组件中的“Delta”。



- 统计方式按实际需求选择，各统计方式说明请参见[校验方式说明](#)。

步骤19 当源端与目的端的统计任务都执行完成后（实例状态为“已完成”），在[校验结果](#)页面可以查看并导出任务的执行结果，方法请参考[查看并导出校验结果](#)。

----结束

10 不使用 EIP 场景下进行大数据数据迁移

本节介绍不具备公网场景下Edge通过SNAT/VPN等方式完成大数据迁移和对数的操作方法。本文以阿里云作为源端，Edge安装在阿里云ECS进行举例。

步骤一：配置华为云 VPN

步骤1 登录华为云管理控制台，选择“网络 > 虚拟专用网络 VPN”。

步骤2 配置VPN网关。

1. 在左侧导航树选择“虚拟专用网络 > 企业版-VPN网关”。
2. 单击“创建站点入云VPN网关”，根据界面提示配置参数。
3. 配置完成后，单击“立即购买”。

步骤3 配置对端网关。需要创建主、备两个对端网关。

1. 在左侧导航树选择“虚拟专用网络 > 企业版-对端网关”。
2. 单击“创建对端网关”，根据界面提示配置参数。其，标识选择IP Address，并填写阿里云侧网关的公网IP。
3. 配置完成后，单击“立即创建”。

步骤4 配置VPN连接。创建两条VPN连接，分别连接华为云VPN网关和阿里云对端网关。

1. 在左侧导航树选择“虚拟专用网络 > 企业版-VPN连接”。
2. 单击“创建VPN连接”，选择已创建的VPN网关和对端网关，并填入对端网关的子网地址（确保子网地址不重合）。



策略配置选择“自定义配置”，并确保配置信息与阿里云侧保持一致，以确保连通性。



3. 配置完成后，单击“立即购买”。

----结束

步骤二：配置阿里云 VPN

步骤1 登录阿里云管理控制台，选择“产品与服务 > 网络和CDN > 混合云网络 > VPN网关”。

步骤2 配置VPN网关。

1. 单击“创建VPN网关”，根据界面提示配置参数。
2. 配置完成后，单击“立即购买”。

步骤3 配置用户网关。

1. 在左侧导航树选择“VPN > 用户网关”。
2. 单击“创建用户网关”，根据界面提示配置参数。
3. 配置完成后，单击“确定”。

步骤4 配置VPN连接：

1. 在左侧导航树选择“VPN > IPsec连接”。
2. 单击“创建IPsec连接”，选择步骤2配置的VPN网关，并配置策略参数，确保与华为云侧的策略配置保持一致。
3. 配置完成后，单击“确定”。

步骤5 配置路由信息。在阿里云上增加到华为云VPC子网的路由信息。

1. 在左侧导航树选择“VPN > VPN网关”。
2. 单击VPN网关名称，在“目的路由表”页签下，单击“添加路由条目”，根据界面提示配置参数。

----结束

步骤三：创建阿里云 NAT 网关并配置

参考[创建和管理公网NAT网关实例](#)，创建阿里云NAT网关并配置SNAT和DNAT条目。

步骤1 登录阿里云管理控制台，选择“产品与服务 > 网络和CDN > 混合云网络 > NAT网关”。

步骤2 创建公网NAT网关。

1. 在公网NAT网关页面，单击“创建公网NAT网关”，根据界面提示配置参数。
2. 配置完成后，单击“立即购买”。

步骤3 配置SNAT条目。

1. 在公网NAT网关页面，找到步骤2创建的公网NAT网关实例，在操作列单击“设置SNAT”。
2. 在SNAT管理页签，单击“创建SNAT条目”，根据界面提示配置参数。
3. 配置完成后，单击“确定创建”。

步骤4 配置DNAT条目。

1. 在公网NAT网关页面，找到步骤2创建的公网NAT网关实例，在操作列单击“设置DNAT”。
2. 在DNAT管理页签，单击“创建DNAT条目”，根据界面提示配置参数。
3. 配置完成后，单击“确定创建”。

----结束

步骤四：配置安全组

需要分别在华为云和阿里云进行安全组配置。

步骤1 在华为云控制台配置安全组，允许Edge所在主机的内网IP地址访问。

1. 登录华为云管理控制台。
2. 在服务列表，选择“网络 > 虚拟私有云”，进入虚拟私有云列表页面。
3. 在左侧导航栏，选择“访问控制 > 安全组”，进入安全组列表页面。
4. 在安全组列表中，找到集群所在的安全组，单击操作列的“配置规则”，进入安全组规则配置页面。
5. 在“入方向规则”页签，单击“添加规则”，弹出“添加入方向规则”对话框。
6. 需要添加一条允许通过TCP协议访问27080端口的规则。在源地址填写Edge所在主机的内网IP地址。



7. 入方向规则设置完成后，单击“确定”。

步骤2 在阿里云控制台配置安全组，允许Edge所在主机的公网IP访问。

1. 登录阿里云ECS管理控制台。
2. 在左侧导航栏，选择“网络与安全 > 安全组”。
3. 找到Edge所在主机的安全组，在操作列中，单击“管理规则”。
4. 在入方向页签，单击“快速添加”。授权策略选择“允许”，授权对象为Edge所在主机的公网IP，端口选择全部。
5. 入方向规则设置完成后，单击“确定”。

----结束

步骤五：搭建大数据数据迁移环境

参考大数据数据迁移的[准备工作](#)，完成迁移环境搭建。其中，购买ECS选择在阿里云购买一台ECS实例，要求该ECS实例的私有IP地址能够通过SNAT访问外部网络。Edge安装完成后，使用SNAT的公网IP访问Edge，注册用户并[连接迁移中心](#)。

步骤六：创建大数据迁移任务

完成步骤一到步骤五的操作，并了解大数据校验任务的[使用须知](#)后。按照以下步骤创建大数据数据迁移任务。

步骤1 [创建源端MaxCompute连接](#)。

步骤2 [创建目的端DLI连接](#)。

步骤3 创建[元数据迁移任务](#)或[数据迁移任务](#)并执行。

----结束