

迁移中心 MgC

# 用户指南

文档版本 04  
发布日期 2024-02-29



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

<b>1 权限管理</b>	<b>1</b>
1.1 创建用户并授权使用 MgC	1
1.2 MgC 自定义策略	5
1.3 委托权限说明	6
<b>2 配置管理</b>	<b>14</b>
2.1 项目管理	14
2.2 凭证管理	16
<b>3 上云调研</b>	<b>18</b>
3.1 TCO 对比	18
3.2 驱动力调研	21
3.3 准备度调研	21
<b>4 应用发现</b>	<b>23</b>
4.1 创建应用	23
4.2 资源发现与采集（复杂项目）	24
4.2.1 创建公网采集任务	24
4.2.2 创建内网采集任务	32
4.2.3 导入应用关联文件	33
4.2.4 导入工具采集结果	34
4.2.5 导入阿里云主机资源清单	34
4.2.6 查看应用关联分析结果	36
4.3 资源发现与采集（简单项目）	37
4.3.1 创建公网采集任务	37
4.3.2 创建内网采集任务	44
4.4 手动添加资源	46
4.5 采集主机性能	47
4.6 关联应用	48
<b>5 大数据血缘</b>	<b>50</b>
5.1 部署迁移工具	50
5.2 采集元数据	50
5.2.1 创建元数据连接	50
5.2.2 创建并运行采集任务	54
5.2.3 查看元数据	56

5.3 血缘采集.....	57
5.3.1 简介.....	57
5.3.2 导出作业.....	58
5.3.3 创建血缘采集任务.....	59
5.3.4 查看血缘关系.....	60
<b>6 迁移方案设计.....</b>	<b>62</b>
6.1 绑定目的端.....	62
6.2 规格评估.....	63
6.3 购买资源.....	66
6.4 迁移方案配置.....	68
<b>7 迁移集群.....</b>	<b>71</b>
7.1 创建集群.....	71
7.2 集群管理.....	73
7.3 计费说明.....	74
7.4 集群状态说明.....	74
<b>8 迁移工作流.....</b>	<b>76</b>
8.1 创建主机迁移工作流.....	76
8.2 创建可用区迁移工作流.....	78
8.3 创建存储迁移工作流.....	80
8.4 创建对象存储批量迁移工作流.....	90
8.5 添加阶段/步骤.....	93
<b>9 大数据作业迁移.....</b>	<b>95</b>
9.1 复杂度分析.....	95
9.1.1 基础配置.....	95
9.1.2 配置增量感知（可选）.....	96
9.1.3 导入作业包.....	99
9.1.4 迁移配置.....	100
9.2 迁移实施.....	100
9.2.1 自动迁移.....	100
9.2.2 手动迁移.....	101
9.2.3 增量迁移.....	102
<b>10 业务验证.....</b>	<b>103</b>
10.1 大数据运维前风险检查（MRR）.....	103
<b>11 大数据校验.....</b>	<b>105</b>
11.1 简介.....	105
11.2 准备工作.....	106
11.3 创建元数据连接.....	107
11.4 创建表组并添加数据表.....	108
11.5 创建连接.....	109
11.5.1 创建执行机连接.....	109

---

11.5.2 创建大数据云服务连接.....	110
11.5.3 创建大数据组件连接.....	112
11.6 创建统计任务并执行校验.....	113
11.7 查看并导出校验结果.....	122
<b>12 修订记录.....</b>	<b>124</b>

# 1 权限管理

## 1.1 创建用户并授权使用 MgC

如果您需要对您所拥有的MgC进行精细的权限管理，您可以使用[统一身份认证服务](#)（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的华为云账号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用MgC资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- 将MgC资源委托给更专业、高效的其他华为云账号或者云服务，这些账号或者云服务可以根据权限进行代运维。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用MgC服务的其它功能。

本章节为您介绍对用户授权的方法，操作流程如[示例流程](#)所示。

### 前提条件

给用户组授权之前，请您了解用户组可以添加的MgC系统策略，并结合实际需求进行选择。

MgC支持的系统策略及策略间的对比，请参见：[MgC系统策略](#)。若您需要对除MgC之外的其它服务授权，IAM支持服务的所有策略请参见[权限策略](#)。

## 示例流程

图 1-1 给用户授权 MgC 权限流程



### 1. 创建用户组并授权

- **系统策略**: 在IAM控制台创建用户组，并根据[MgC系统策略说明](#)以及实际的权限要求，为用户组授权MgC系统策略，授权范围方案选择“所有资源”。

表 1-1 MgC 系统策略说明

策略名称	描述	策略类别	策略内容
MgC FullAccess	迁移中心管理员权限，拥有操作MgC的所有权限。	系统策略	<a href="#">MgC FullAccess策略内容</a>
MgC ReadOnlyAccess	迁移中心只读权限，仅能查看MgC资源，无法进行操作。	系统策略	<a href="#">MgC ReadOnlyAccess策略内容</a>
MgC DiscoveryAccess	迁移中心资源发现操作权限，拥有操作资源发现功能的权限和只读权限。	系统策略	<a href="#">MgC DiscoveryAccess策略内容</a>
MgC AssessAccess	迁移中心评估操作权限，拥有操作评估功能、资源发现功能的权限和只读权限。	系统策略	<a href="#">MgC AssessAccess策略内容</a>
MgC MigrateAccess	迁移中心迁移操作权限，拥有操作迁移功能、评估功能、资源发现功能的权限和只读权限。	系统策略	<a href="#">MgC MigrateAccess策略内容</a>



策略名称	描述	策略类别	策略内容
MgC AppDiscovery Access	迁移中心应用发现操作权限，拥有操作应用发现功能、资源发现功能的权限和只读权限。	系统策略	<a href="#">MgC AppDiscoveryAccess策略内容</a>
MgC MrrAccess	迁移中心业务验证操作权限，拥有业务验证功能的权限和只读权限。	系统策略	<a href="#">MgC MrrAccess策略内容</a>

- **自定义策略**：如果IAM用户只需要拥有迁移中心 MgC部分操作权限，则使用自定义策略，参见[MgC自定义策略](#)。
2. **创建用户并加入用户组**  
在IAM控制台创建用户，并将其加入[1.创建用户组并授权](#)中创建的用户组。
  3. **用户登录并验证权限**  
新创建的用户登录控制台，切换至授权区域，验证权限：
    - 在“**服务列表**”中选择迁移中心 MgC，进入MgC主界面，根据您授予的权限可以进行对应操作，表示授予的权限已生效。
    - 在“**服务列表**”中选择除迁移中心 MgC外的任一服务，若提示权限不足，表示授予的权限已生效。

## MgC FullAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*",
        "iam:agencies:listAgencies",
        "iam:agencies:createAgency",
        "iam:permissions:grantRoleToAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

## MgC ReadOnlyAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*"
      ]
    }
  ]
}
```

## MgC DiscoveryAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
```

```
{
  "Effect": "Allow",
  "Action": [
    "mgc:*:query*",
    "mgc:*:discovery"
  ]
}
```

## MgC AssessAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:assess",
        "iam:agencies:listAgencies",
        "iam:agencies:createAgency",
        "iam:permissions:grantRoleToAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

## MgC MigrateAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:assess",
        "mgc:*:migrate",
        "iam:agencies:listAgencies",
        "iam:agencies:createAgency",
        "iam:permissions:grantRoleToAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

## MgC AppDiscoveryAccess 策略内容

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:appdiscovery"
      ]
    }
  ]
}
```

## MgC MrrAccess 策略内容

```
{
  "Version": "1.1",
```

```
"Statement": [
  {
    "Action": [
      "mgc:*:query*",
      "mgc:mrr:query",
      "mgc:mrr:update",
      "mgc:mrr:export",
      "mgc:mrr:import",
      "mgc:mrr:upgrade",
      "mgc:mrr:delete",
      "mgc:mrr:check"
    ],
    "Effect": "Allow"
  }
]
```

## 1.2 MgC 自定义策略

如果系统预置的MgC权限，不满足您的授权要求，可以创建自定义策略。

目前华为云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。
- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见：[创建自定义策略](#)。本章为您介绍常用的MgC自定义策略样例。

### MgC 自定义策略样例

- 平台采集、资源管理

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery"
      ]
    }
  ]
}
```

- TCO评估、主机评估

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:assess",
        "iam:agencies:listAgencies",
        "iam:agencies:createAgency",
        "iam:permissions:grantRoleToAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

- ```

    ]
  }
}

```
- 应用关联分析
 

```

{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:appdiscovery"
      ]
    }
  ]
}

```
  - 迁移 workflows
 

```

{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:assess",
        "mgc:*:migrate",
        "iam:agencies:listAgencies",
        "iam:agencies:createAgency",
        "iam:permissions:grantRoleToAgency"
      ]
    }
  ]
}

```

## 1.3 委托权限说明

### 概述

当您启用特定功能时，需要您确认并同意授予必要的权限给MgC，以便我们能够为您提供完整的服务。本节为您介绍需要进行权限委托的场景以及自定义策略详情。

委托授权策略说明：

- 创建自定义策略：如果您尚未设定任何自定义的权限委托策略，系统会自动为您创建一个标准的自定义策略。
- 更新自定义策略：对于已经存在的自定义策略，系统将进行完整性检查。若发现策略内容不全面或缺少必要元素，系统将自动进行补充。

## 跨可用区迁移

| 场景     | 委托对象 | 自定义策略名称                        | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------|------|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 跨可用区迁移 | MgC  | MgC<br>AzMigrationAgencyPolicy | ecs:cloudServers:showServer ( 查询云服务器详情 )<br>ecs:flavors:get ( 查询云服务器规格 )<br>ecs:cloudServerFlavors:get ( 查询云服务器规格详情和扩展信息列表 )<br>ecs:cloudServerQuotas:get ( 查询租户配额 )<br>ecs:servers:list ( 查询云服务器列表 )<br>ecs:cloudServers:list ( 查询云服务器详情列表 )<br>ecs:servers:stop ( 关闭云服务器 )<br>ecs:cloudServers:listServerInterfaces ( 查询云服务器网卡信息 )<br>ecs:cloudServers:createServers ( 创建云服务器 )<br>ecs:cloudServers:listServerBlockDevices ( 查询弹性云服务器磁盘信息 )<br>ecs:cloudServerNics:update ( 云服务器网卡配置私有IP )<br>vpc:publicIps:create ( 创建弹性公网IP )<br>vpc:publicIps:update ( 更新弹性公网IP )<br>vpc:subnets:get ( 查询子网列表或详情 )<br>vpc:networks:get ( 查询网络列表或详情 )<br>vpc:publicIps:list ( 查询弹性公网IP )<br>vpc:publicIps:get ( 查询弹性公网IP详情 )<br>evs:types:get ( 查询云硬盘类型 )<br>evs:volumes:list ( 查询云硬盘列表 )<br>cbr:vaults:get ( 查询指定存储 )<br>cbr:vaults:list ( 查询存储库列表 )<br>cbr:vaults:create ( 创建存储库 )<br>cbr:vaults:addResources ( 添加资源 )<br>cbr:vaults:backup ( 执行备份 )<br>cbr:backups:list ( 查询备份列表 )<br>cbr:tasks:list ( 查询任务列表 )<br>cbr:tasks:get ( 查询单个任务 ) |

| 场景 | 委托对象 | 自定义策略名称 | 细粒度最小使用权限                                                                                                                                                                                                                                  |
|----|------|---------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |      |         | cbr:backups:delete (删除备份)<br>cbr:backups:get (查询指定备份)<br>cbr:vaults:delete (删除存储库)<br>ims:wholeImages:create (制作整机镜像)<br>ims:images:list (查询镜像列表)<br>ims:images:delete (删除镜像)<br>ims:images:get (查询镜像详情)<br>ims:serverImages:create (制作镜像) |

## 新建 TCO 对比

| 场景      | 委托对象 | 自定义策略名称                | 细粒度最小使用权限                                                                                                                  |
|---------|------|------------------------|----------------------------------------------------------------------------------------------------------------------------|
| 新建TCO对比 | MgC  | MgC<br>TcoAgencyPolicy | ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)<br>evs:types:get (查询云硬盘类型)<br>ims:*.get* (ims查询权限)<br>ims:*.list* (ims查询权限) |

## 规格评估

| 场景   | 委托对象 | 自定义策略名称                         | 细粒度最小使用权限                                                                                             |
|------|------|---------------------------------|-------------------------------------------------------------------------------------------------------|
| 规格评估 | MgC  | MgC<br>ServerAssessAgencyPolicy | ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)<br>ims:images:list (查询镜像列表)<br>evs:types:get (查询云硬盘类型) |

## 主机绑定目的端

| 场景      | 委托对象 | 自定义策略名称                                 | 细粒度最小使用权限                                                                                                                    |
|---------|------|-----------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| 主机绑定目的端 | MgC  | MgC<br>ServerBindTarget<br>AgencyPolicy | ecs:cloudServers:showServer ( 查询云服务器详情 )<br>evs:volumes:list ( 查询云硬盘列表 )<br>ecs:cloudServerFlavors:get ( 查询云服务器规格详情和扩展信息列表 ) |

## 主机迁移 workflow

| 场景            | 委托对象 | 自定义策略名称                            | 细粒度最小使用权限                                                                                                                                                                                                   |
|---------------|------|------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 主机迁移 workflow | MgC  | MgC<br>ServerMigrationAgencyPolicy | ecs:cloudServers:showServer ( 查询云服务器详情 )<br>ecs:cloudServers:createServers ( 创建云服务器 )<br>sms:server:migrationServer ( 迁移源端服务器 )<br>sms:server:queryServer ( 查看源端服务器 )<br>ecs:cloudServers:list ( 查看云服务器列表 ) |

## 大数据检查项 ( MRR )

| 场景             | 委托对象 | 自定义策略名称                | 细粒度最小使用权限                                                         |
|----------------|------|------------------------|-------------------------------------------------------------------|
| 大数据检查项 ( MRR ) | MgC  | MgC<br>MrrAgencyPolicy | mrs:cluster:get* ( mrs集群查询权限 )<br>mrs:cluster:list* ( mrs集群查询权限 ) |

## 购买资源

| 场景   | 委托对象 | 自定义策略名称                     | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                      |
|------|------|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 购买资源 | MgC  | MgC<br>PurchaseAgencyPolicy | eps:resources:add (企业项目资源迁入)<br>ecs:cloudServers:createServers (创建云服务器)<br>evs:volumes:list (查询云硬盘列表)<br>ecs:cloudServerFlavors:get (查询云服务器规格详情和扩展信息列表)<br>ecs:cloudServers:list (查询云服务器详情列表)<br>ecs:cloudServers:createServers (创建云服务器)<br>vpc:publicIps:update (更新弹性公网IP)<br>vpc:publicIps:create (创建弹性公网IP) |

## 配置主机购买模板

| 场景       | 委托对象 | 自定义策略名称                             | 细粒度最小使用权限                                                                                                                                          |
|----------|------|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| 配置主机购买模板 | MgC  | MgC<br>PurchaseTemplateAgencyPolicy | iam:projects:listProjects (查询租户项目)<br>eps:enterpriseProjects:list (查看企业项目列表)<br>vpc:subnets:get (查询子网列表或详情)<br>vpc:securityGroups:get (查询安全组列表或详情) |



## 存储迁移 workflow/创建迁移集群

| 场景                   | 委托对象 | 自定义策略名称                         | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------|------|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 存储迁移 workflow/创建迁移集群 | OMS  | OMS<br>ObsMigrationAgencyPolicy | ecs:cloudServers:createServers (创建云服务器)<br>ecs:cloudServers:listServerInterfaces (查询云服务器网卡信息)<br>ecs:cloudServers:showServer (查询云服务器详情)<br>ecs:cloudServers:deleteServers (删除云服务器)<br>ecs:cloudServers:list (查询云服务器详情列表)<br>nat:natGateways:create (创建NAT网关)<br>nat:natGateways:get (查询NAT网关详情)<br>nat:natGateways:delete (删除NAT网关)<br>nat:snatRules:create (创建SNAT规则)<br>nat:snatRules:get (查询SNAT规则详情)<br>nat:dnatRules:list (查询DNAT规则列表)<br>nat:snatRules:list (查询SNAT规则列表)<br>nat:snatRules:delete (删除SNAT规则)<br>nat:natGateways:list (查询NAT网关列表)<br>vpc:securityGroups:create (创建安全组)<br>vpc:securityGroups:delete (删除安全组)<br>vpc:securityGroups:get (查询安全组列表或详情)<br>vpc:securityGroupRules:create (创建安全组规则)<br>vpc:securityGroupRules:get (查询安全组规则列表或详情)<br>vpc:securityGroupRules:delete (删除安全组规则)<br>vpc:securityGroups:get (查询安全组列表或详情) |

| 场景 | 委托对象 | 自定义策略名称 | 细粒度最小使用权限                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|----|------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |      |         | vpcep:epservices:create (创建终端节点服务)<br>vpcep:epservices:get (查询终端节点服务详情)<br>vpcep:permissions:list (查询终端节点服务的白名单列表)<br>vpcep:connections:list (查询连接终端节点服务的连接列表)<br>vpcep:epservices:list (查询终端节点服务列表)<br>vpcep:epservices:delete (删除终端节点服务)<br>vpcep:endpoints:create (创建终端节点)<br>vpcep:endpoints:list (查询终端节点列表)<br>vpcep:endpoints:get (查询终端节点详情)<br>vpcep:endpoints:delete (删除终端节点)<br>vpcep:connections:update (接受或拒绝终端节点的连接)<br>vpcep:permissions:update (批量添加或移除终端节点服务的白名单)<br>lts:topics:get (查询指定日志主题)<br>lts:topics:create (创建日志主题)<br>lts:topics:list (查询日志主题列表)<br>lts:topics:delete (删除指定日志主题)<br>lts:*:* (所有主机组相关的操作)<br>lts:groups:create (创建日志组)<br>lts:groups:get (查询指定日志组)<br>lts:groups:delete (删除指定日志组)<br>aom:*:* (aom的所有权限)<br>apm:icmgr:* (apm采集组件的所有权限) |

| 场景 | 委托对象 | 自定义策略名称                         | 细粒度最小使用权限                               |
|----|------|---------------------------------|-----------------------------------------|
|    | ECS  | ECS<br>ObsMigrationAgencyPolicy | apm:icmgr:* ( apm采集组件的所有权限 )            |
|    | MgC  | -                               | OMS Administrator ( 系统角色, OMS的admin角色 ) |

## 血缘采集任务

| 场景     | 委托对象 | 自定义策略名称                              | 细粒度最小使用权限                                                                                                    |
|--------|------|--------------------------------------|--------------------------------------------------------------------------------------------------------------|
| 血缘采集任务 | MgC  | MgC<br>LineageCollectionAgencyPolicy | obs:object:GetObject ( 获取对象内容、获取对象元数据 )<br>obs:object:PutObject ( PUT上传、POST上传、复制对象、追加写对象、初始化上传段任务、上传段、合并段 ) |

# 2 配置管理

## 2.1 项目管理

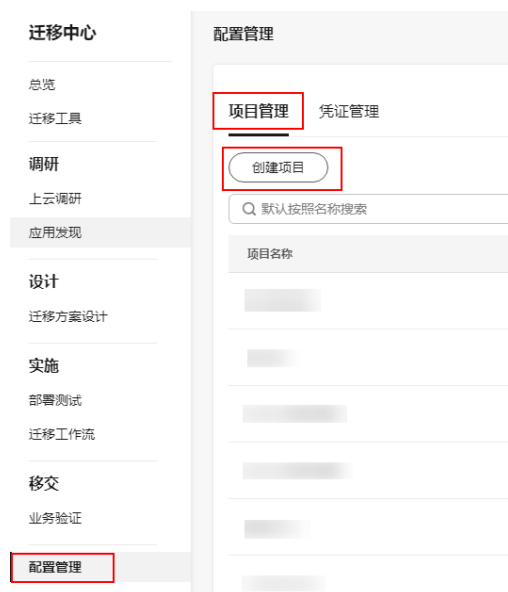
### 操作场景

通过创建独立的项目，可以将不同的操作场景进行区分和隔离，从而实现更好的组织管理和资源分配。例如：可为阿里云迁移至华为云创建一个新的项目，里面将包含阿里云平台的主机、数据库、组件等信息。

### 创建项目

**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“**配置管理**”，进入**配置管理**页面。选择“**项目管理**”页签，单击“**创建项目**”。



**步骤3** 在弹出的窗口中，自定义**项目名称**，并根据实际需求，选择迁移类型，单击“**确认**”。创建项目成功，可在项目管理列表查看。

- 简单应用迁移：适用于应用数量少、应用与资源间依赖关系简单明确以及华为云跨可用区迁移等场景。
- 复杂应用迁移：适用于应用数量多、应用与资源间依赖关系复杂等场景。

×

### 设置项目名称

请输入项目名称

### 选择本次迁移类型

将根据您的选择生成相关流程任务

|                                                                                                                                                                                                         |                                                                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <h4>简单应用迁移</h4> <p>使用迁移中心精简功能，进行快捷资源上云</p> <ul style="list-style-type: none"><li>· 应用数量少或者应用与资源间关系明确</li><li>· 快速进行多种资源迁移上云</li><li>· 特殊场景（如：华为云跨可用区迁移）</li></ul> <p>简单架构 快速上云 跨可用区迁移</p> <p>普通用户迁移</p> | <h4>复杂应用迁移</h4> <p>使用全部功能进行最全面的迁移流程覆盖</p> <ul style="list-style-type: none"><li>· 应用数量多，应用间关系复杂</li><li>· 资源数量多，资源与应用间依赖关系复杂</li></ul> <p>资源数量多 云上结构复杂 分批迁移</p> <p>适合专业服务团队</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

确认取消

----结束

## 修改项目

**步骤1** 单击操作列的“修改”按钮，弹出“修改项目”窗口。

**步骤2** 可以修改项目名称、输入项目描述、设置项目的开始时间和结束时间，单击“确认”，完成修改。

×

### 修改项目

\* 项目名称

项目描述   
0/255

开始时间

结束时间

确认取消

----结束

## 归档项目

归档后的项目将在**当前项目**下拉列表中隐藏，无法选择。单击操作列的“**归档**”按钮，可将项目归档。如想重新选择归档后的项目，单击“**取消归档**”即可。



### 说明

当前项目不允许归档。

## 2.2 凭证管理

凭证是资源采集的唯一认证方式，本节介绍如何添加用于发现云上资源（公网发现）所需要的凭证。内网发现、手动添加所需的凭证添加方式请参考[在Edge添加凭证](#)。

### 认证方式

当前仅支持添加公有云资源类型的凭证。支持的认证方式参见[表2-1](#)。

表 2-1 公有云认证方式

| 资源类型 | 认证方式  | 说明                                                                                   |
|------|-------|--------------------------------------------------------------------------------------|
| 公有云  | AK/SK | 适用于添加 <b>华为云、阿里云、AWS、腾讯云</b> 等云平台凭证。                                                 |
|      | 配置文件  | 适用于添加 <b>谷歌云</b> 平台凭证。配置文件需要上传 <b>谷歌云服务账号</b> 凭据，该凭证必须为 <b>json</b> 格式文件，且大小不能超过4KB。 |
|      | ID/密钥 | 适用于添加 <b>Azure</b> 平台凭证。添加凭证所需的信息获取方法请参见 <a href="#">如何获取添加Azure凭证所需的信息</a> 。        |

### 注意

云上凭证的存储有效期为7天，超过7天后，凭证会过期，需要重新输入。

## 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“**配置管理**”，进入**配置管理**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** 选择“**凭证管理**”页签，单击“**添加凭证**”，弹出“**添加云平台凭证**”窗口。



**步骤4** 按照窗口提示，选择资源类型后，选择认证方式并输入/上传相应凭证，凭证名称为用户自定义，存储位置选择“云上”，单击“确认”。

添加完成后，在凭证管理列表可以看到添加的凭证。

#### 添加云平台凭证

|        |         |
|--------|---------|
| * 使用类型 | 平台      |
| * 类型   | 公有云     |
| * 凭证名称 | 请输入凭证名称 |
| * 认证方式 | AK/SK   |
| * AK   | 请输入AK   |
| * SK   | .....   |
| * 存储位置 | 云上      |

取消

确认

---结束

# 3 上云调研

## 3.1 TCO 对比

通过TCO对比功能，可以自动完成源端云厂商消费账单分析，根据源端产品规格及用量自动匹配华为云上的产品及用量，并对比源端云厂商使用费用和华为云使用费用，为您云间迁移提供参考。

目前支持TCO对比的源端云厂商包括：**AWS、阿里云**。

### 新建 TCO 对比

**步骤1** 登录**迁移中心**管理控制台。

**步骤2** 单击左侧导航栏“**调研>上云调研**”，进入“**上云调研**”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**上云成本分析**框中的“**新建TCO任务**”按钮，右侧弹出“**新建TCO对比**”窗口。

#### 上云调研

迁移前进行应用上云迁移信息调研，搜集源端业务、资源、服务等各方面的需求，为上云迁移设计提供必要的信息

#### 调研工具

##### 上云成本分析

通过TCO对比功能，可以自动完成源端云厂商消费账单分析以及和华为云的成本分析和比较，从而为您云间迁移提供参考。

**新建TCO任务**

**步骤4** 根据**表3-1**，配置参数。



×

### 新建TCO对比

**任务**

\* 任务名称

---

**源端**

\* 云厂商

\* 区域

\* 调研方式   ?

\* 凭证

\* 对比时间段  ?

\* 对比产品

---

**目的端**

\* 区域

表 3-1 参数说明

| 窗口区域 | 参数   | 说明                                                                                                                                                                                                                                                                                                                           |
|------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 任务   | 任务名称 | 用户自定义。                                                                                                                                                                                                                                                                                                                       |
| 源端   | 云厂商  | 选择源端云厂商。<br><b>说明</b><br>当前源端云厂商仅支持AWS和阿里云。                                                                                                                                                                                                                                                                                  |
|      | 区域   | 选择源端所在区域。                                                                                                                                                                                                                                                                                                                    |
|      | 调研方式 | <ul style="list-style-type: none"> <li>“AK/SK”方式，需要选择“凭证”、“对比时间段”和“对比产品”，如未提前<a href="#">添加凭证</a>，请单击“新建凭证”，填写相应参数，校验并保存凭证。<br/><b>说明</b> <ul style="list-style-type: none"> <li>新建凭证中的AK/SK输入源端账号的AK/SK。</li> <li>源端账号拥有只读权限即可。</li> <li>可能产生接口调用相关费用。</li> </ul> </li> <li>“表格上传”方式，通过填写TCO任务模板的方式上传源端信息进行评估分析。</li> </ul> |
| 目的端  | 区域   | 选择目的端所在区域。                                                                                                                                                                                                                                                                                                                   |

**步骤5** 配置完成后，单击“确认”按钮，提交TCO对比任务。

提交成功后可以在TCO任务列表中看到提交的任务，单击**任务名称**，可以查看任务进度，等待任务运行完成后即可**下载账单**、**查看配置**、**产品对比**，进行后续详细分析。

----结束

## 下载账单

系统会从产品计费项的维度分析您指定时间段内的账单，解析出每款产品的规格、月均使用量以及月均消耗，您可以在TCO对比任务完成后下载您的账单，在本地进一步分析。

**步骤1** 在TCO任务列表，单击**任务名称**。

**步骤2** 单击操作列的“**下载账单**”按钮，下载Excel版账单。



----结束

## 查看配置

系统会根据解析出来的产品规格及用量自动匹配华为云上的产品及用量，在匹配策略上系统坚持在不低于您当前配置的前提下优先考虑成本的原则，给出华为云上的产品规格及用量配置。

您可以通过**查看配置**进入产品映射明细列表，在这个列表中，您可以添加、修改、删除产品和规格映射，调整程序预定义的资源映射逻辑，从而获取您真实需要的产品规格列表。

**步骤1** 在TCO任务列表，单击**任务名称**。

**步骤2** 单击操作列的“**查看配置**”按钮，进入“**产品映射**”页面。



**步骤3** 单击映射类型前的  按钮，可以查看映射详情。

----结束

## 产品对比

在使用成本方面，系统会对指定时间内，源端云厂商使用费用和华为云使用费用进行比较，您可以在**产品对比**页面的**对比结果**中查看。

我们还会给出源端云厂商和华为云同等规格配置的产品月均成本比较，您可以在**产品对比**页面的**产品价格对比**中查看。并且可以通过和您的销售经理沟通，针对每种类别产品分别调整价格计算比例（优惠折扣）进一步优化华为云的使用成本。

**步骤1** 在TCO任务列表，单击**任务名称**。

**步骤2** 单击操作列的“**产品对比**”按钮，进入“**产品对比**”页面。



**步骤3** 查看**成本对比结果**和**产品价格对比**，根据实际情况调整价格计算比例（**优惠折扣**）优化华为云的使用成本。

单击页面右上角“**导出结果**”按钮，可以将产品对比结果导出PDF格式到本地。

----结束

## 3.2 驱动力调研

### 操作场景

对上云迁移驱动力的系统评估，有助于帮助您针对性地制定上云策略和迁移方案。

### 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“**调研>上云调研**”，进入“**上云调研**”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**驱动力调研**框中的“**开始调研**”按钮，进入“**驱动力评估**”页面。



**步骤4** 根据实际情况，依次填写/勾选相关信息，单击“**提交**”。

系统会自动生成并下载**上云驱动力评估报告**，您可以通过报告，查看华为云对您上云迁移的建议。

----结束

## 3.3 准备度调研

### 操作场景

对上云迁移准备度的系统评估，以**评估报告**的形式从业务目标、上云策略、IT流程和安全合规等维度详细评估，有助于您制定总体迁移规划和项目计划。

## 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“**调研>上云调研**”，进入“**上云调研**”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**准备度调研框**中的“**开始调研**”按钮，进入“**准备度评估**”页面。



- 步骤4** 根据实际情况，依次填写/勾选相关信息，单击“**提交**”。

系统会自动生成并下载上云**准备度评估**报告，您可以通过报告，查看华为云对您上云迁移的建议。

----**结束**

# 4 应用发现

## 4.1 创建应用

您可以通过创建应用，将采集到的资源添加到应用中，对资源进行功能分组，便于后续进行规格评估和创建迁移 workflow。

### 首次创建

**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>应用发现](#)”，进入应用发现页面。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** 首次进入页面时，在操作流程中，单击手动创建应用后面的“[去创建](#)”，弹出新建应用窗口。

如果已创建过发现任务，则单击应用框中的“[新建应用](#)”按钮，弹出新建应用窗口。

**步骤4** 输入自定义的[应用名称](#)和[描述](#)；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“[创建应用](#)”按钮，应用创建成功，进入[关联资源](#)页面。

- 如果已经进行过资源采集，并且需要将采集到的资源关联到创建的应用，请勾选需要关联到应用的资源，单击“[关联到应用](#)”按钮，完成应用创建和资源关联。
- 如果还未进行资源采集或暂不需要关联资源，单击“[暂不关联资源](#)”按钮，完成应用创建。后续在[资源列表](#)中也可以进行关联操作。

----[结束](#)

### 新增应用

**步骤1** 在应用发现页面，单击应用框中的“[新建应用](#)”按钮，弹出创建应用窗口。

**步骤2** 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](#)选择目的端所在区域，单击“[确定](#)”按钮，新增应用成功，可以在应用列表中查看。

----[结束](#)

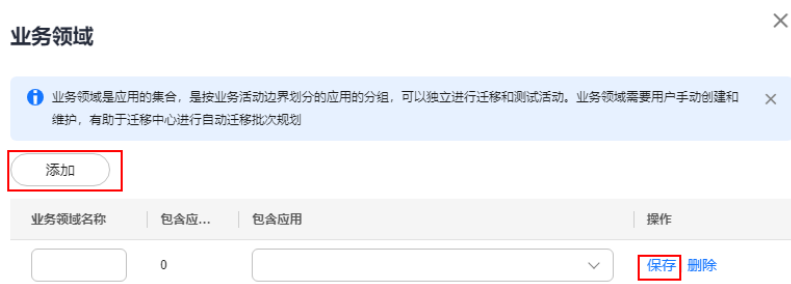
## 管理应用

您可以根据不同的业务领域将应用进行分类管理。

**步骤1** 在应用发现页面，单击应用框中业务领域的“查看”按钮，右侧弹出业务领域窗口。



**步骤2** 单击“添加”按钮，输入业务领域名称，选择该业务包含的应用，单击“保存”。



----结束

## 修改应用

仅简单项目支持。

**步骤1** 在应用列表，单击应用操作列的“修改”，弹出修改应用窗口。

**步骤2** 根据实际情况，修改应用参数后，单击“确定”，完成应用修改。

----结束

## 4.2 资源发现与采集（复杂项目）

### 4.2.1 创建公网采集任务

通过公网在线发现并采集源端云平台资源，获取源端资源详细信息，并整理源端资源与应用间的关联关系。目前支持发现的源端云平台为：阿里云、华为云、AWS、腾讯云、谷歌云、Azure。

- 阿里云和华为云支持采集的资源类型包括：主机、容器、中间件、数据库、大数据、网络和存储。
- AWS和腾讯云支持采集的资源类型包括：主机、数据库、存储。
- 谷歌云和Azure支持采集的资源类型包括：主机、容器、中间件、数据库、存储和网络。

## 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>应用发现](#)”，进入应用发现页面。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** 首次创建任务，在云平台发现区域，单击“[公网发现](#)”按钮，进入创建公网发现任务页面。

非首次创建任务，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。

**步骤4** 根据[表4-1](#)，配置公网发现任务参数。

**表 4-1** 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   | 选择源端云平台，目前支持的云平台有 <a href="#">阿里云</a> 、 <a href="#">华为云</a> 、 <a href="#">AWS</a> 、 <a href="#">腾讯云</a> 、 <a href="#">谷歌云</a> 、 <a href="#">Azure</a> 。                                                                                                                                                                                                                                                                                                                                                                    | 是    |
|      | 凭证   | 选择源端对应的凭证，如未提前 <a href="#">添加凭证</a> ，请单击“ <a href="#">新建凭证</a> ”，填写相应参数，校验并保存凭证。<br><br><ul style="list-style-type: none"> <li><a href="#">阿里云</a>、<a href="#">华为云</a>、<a href="#">AWS</a>、<a href="#">腾讯云</a>认证方式请选择“AK/SK”，并输入源端账号的AK/SK。</li> <li><a href="#">谷歌云</a>认证方式请选择“<a href="#">配置文件</a>”，并上传<a href="#">谷歌云服务账号凭据</a>，该凭证必须为json格式文件，且大小不能超过4KB。</li> <li><a href="#">Azure</a>认证方式请选择“<a href="#">ID/密钥</a>”。添加凭证所需的信息获取方法请参见<a href="#">如何获取添加Azure凭证所需的信息</a>。</li> </ul> <p><b>说明</b><br/>存储位置选择“云上”。</p> | 是    |
|      | 区域   | 选择要发现的源端云平台区域，支持多选。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 是    |

**步骤5** 开启云平台采集，在[资源类型](#)下拉列表，勾选需要进行采集的资源类型（采集项）。当前支持采集的云平台和资源类型参见[表4-2](#)。

表 4-2 支持采集的资源类型

| 云平台 | 资源类型                                                                                               | 子类型                                                                                                                                                                                                                                 |
|-----|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 阿里云 | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 大数据</li> <li>• 数据库</li> </ul> | -                                                                                                                                                                                                                                   |
|     | 中间件                                                                                                | <ul style="list-style-type: none"> <li>• Redis</li> <li>• Kafka</li> </ul>                                                                                                                                                          |
|     | 存储                                                                                                 | <ul style="list-style-type: none"> <li>• 对象存储</li> <li>• 文件存储</li> </ul>                                                                                                                                                            |
|     | 网络                                                                                                 | <ul style="list-style-type: none"> <li>• 云连接</li> <li>• 负载均衡（ALB、CLB）</li> <li>• 专线</li> <li>• 公网域名</li> <li>• 内网域名</li> <li>• 弹性公网IP</li> <li>• 公网NAT网关</li> <li>• 路由表</li> <li>• 安全组</li> <li>• 虚拟私有云</li> <li>• VPN网关</li> </ul> |
| 华为云 | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 大数据</li> <li>• 数据库</li> </ul> | -                                                                                                                                                                                                                                   |
|     | 中间件                                                                                                | <ul style="list-style-type: none"> <li>• Redis</li> <li>• Kafka</li> </ul>                                                                                                                                                          |
|     | 存储                                                                                                 | <ul style="list-style-type: none"> <li>• 对象存储</li> <li>• 文件存储</li> </ul>                                                                                                                                                            |



| 云平台   | 资源类型                                                                                | 子类型                                                                                                                                                                               |
|-------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | 网络                                                                                  | <ul style="list-style-type: none"> <li>● 负载均衡（ELB）</li> <li>● 公网域名</li> <li>● 内网域名</li> <li>● 弹性公网IP</li> <li>● 公网NAT网关</li> <li>● 路由表</li> <li>● 安全组</li> <li>● 虚拟私有云</li> </ul> |
| AWS   | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 数据库</li> </ul>               | -                                                                                                                                                                                 |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
| 腾讯云   | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 数据库</li> </ul>               | -                                                                                                                                                                                 |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
| 谷歌云   | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 容器</li> <li>● 数据库</li> </ul> | -                                                                                                                                                                                 |
|       | 中间件                                                                                 | Redis                                                                                                                                                                             |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
|       | 网络                                                                                  | <ul style="list-style-type: none"> <li>● 弹性公网IP</li> <li>● 路由表</li> <li>● 安全组</li> <li>● 虚拟私有云</li> </ul>                                                                         |
| Azure | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 容器</li> <li>● 数据库</li> </ul> | -                                                                                                                                                                                 |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
|       | 中间件                                                                                 | <ul style="list-style-type: none"> <li>● Redis</li> <li>● Kafka</li> </ul>                                                                                                        |

| 云平台 | 资源类型 | 子类型                                                                                                                                   |
|-----|------|---------------------------------------------------------------------------------------------------------------------------------------|
|     | 网络   | <ul style="list-style-type: none"> <li>弹性公网IP</li> <li>路由表</li> <li>安全组</li> <li>公网NAT网关</li> <li>虚拟私有云</li> <li>负载均衡（ELB）</li> </ul> |

**步骤6**（可选）选择需要采集的源端云平台服务（仅支持阿里云、腾讯云），获取资源与服务/微服务间的调用关系、关联关系以及依赖关系。

### 须知

选择采集的服务时，需要确保所提供的凭证拥有该服务的相关权限。

采集各服务的功能见下表。

| 服务         | 功能                                      |
|------------|-----------------------------------------|
| 资源管理 RM    | 通过采集该服务可以发现更多应用与资源间的关联线索，防止应用与资源间关系链断裂。 |
| 域名解析服务 DNS | 通过采集该服务可以从接入层分析应用间的关联关系。                |
| 应用防火墙 WAF  | 通过采集该服务可以获得接入层的流量方向。                    |
| 负载均衡 SLB   | 通过采集该服务可以获得接入层的流量方向。                    |
| 微服务引擎 MSE  | 通过采集该服务可以获得应用间的关联关系，微服务间的调用关系。          |

**步骤7** 单击“确认”，公网发现任务创建完成，系统自动开始资源发现。

- 您可以在应用发现页面，[查看发现的源端资源列表和资源详情](#)。
- 您可以在任务列表页面，查看任务状态和[查看任务详情](#)。
  - 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。



- 当任务状态为“成功”时，您可以前往应用发现页面，[查看应用关联分析结果](#)，包含：技术架构图、业务架构图、应用列表、查看资源线索、微服务信息等。

**步骤8** 发现任务状态为成功后，根据不同资源类型，进行后续操作。

- 主机资源需要完成[迁移预检查](#)或[深度采集](#)后，才可以进行[迁移方案设计](#)。
- 容器资源需要完成[深度采集](#)后，才可以进行[迁移方案设计](#)。
- 中间件、数据库、大数据、存储等资源可直接进行[迁移方案设计](#)。
- 对象存储资源可以进行[深度采集](#)和[迁移方案配置](#)。

----结束

## 迁移预检查

请按照以下步骤对主机资源进行迁移预检查（源端迁移准备度），确定源端主机是否满足迁移条件。

1. 在源端内网环境中[安装Edge工具](#)并完成注册。
2. 在应用发现页面，选择资源列表页签，单击主机资源的资源数量，进入资源列表页面的主机页签。
3. 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。
4. 单击源端迁移准备度列的“配置”，弹出配置窗口。
5. 根据[表4-3](#)，配置参数。

表 4-3 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。                                                                                                     |
| IP   | 选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入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。                                |
| 端口   | 主机开放端口。 <ul style="list-style-type: none"><li>• Windows主机默认为5985端口，无法修改。</li><li>• Linux主机默认22端口，可根据实际情况修改。</li></ul> |
| 凭证   | 选择主机凭证，如未提前在Edge上 <a href="#">添加源端主机凭证</a> ，请前往Edge控制台添加，并同步至迁移中心。                                                    |

6. 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开始源端迁移准备度检查，当源端迁移准备度列状态会变为已就绪时，代表该主机已满足迁移条件，可以进行下一阶段[迁移方案设计](#)。

## 主机/容器深度采集

请按照以下步骤对主机资源和容器资源进行深度采集。

1. 在源端内网环境中**安装Edge工具**并与迁移中心连接。
2. 在**应用发现**页面，选择**资源列表**页签，单击**主机/容器资源**的**资源数量**，进入**资源列表**页面的**主机/容器**页签。
3. 在需要进行深度采集的**主机/容器采集设备**列，单击“**绑定**”，弹出**绑定采集设备**窗口。  
如果需要批量绑定，可以同时勾选多个资源，单击页面右上角的“**绑定采集设备**”，弹出**绑定采集设备**窗口。
4. 选择源端安装的Edge工具，以及网络接入方式（如果采集的资源与安装Edge的主机处于同一VPC内，可以选择私网接入，否则需要选择公网接入），单击“**确定**”。绑定成功后，资源的采集设备列状态会变为已绑定。
5. 采集设备绑定成功后，还需要为资源绑定凭证，单击资源凭证列的“**绑定**”，弹出**绑定凭证**窗口。
6. 选择资源凭证，如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。
7. 单击“**确定**”，系统会自动检查凭证绑定状态。当采集状态为**已就绪**时，单击采集状态列的“**采集**”进行深度采集。首次深度采集完成后，可以单击采集状态列的“**重新采集**”按钮，进行多次深度采集。采集完成后，可以进行下一阶段**迁移方案设计**。

## 对象存储深度采集

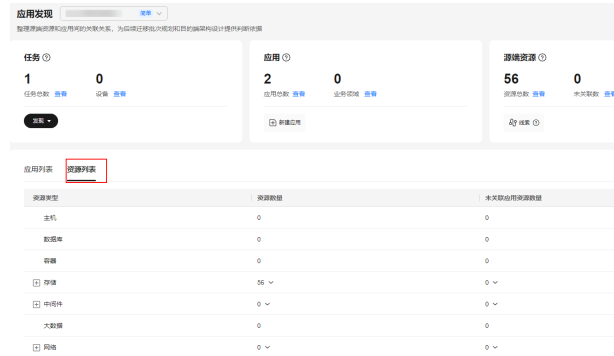
通过深度采集获取对象存储资源的详细信息，以提高迁移集群规格评估的准确性。请按照以下步骤进行**对象存储资源**深度采集。

1. 在源端内网环境中**安装Edge工具**并与迁移中心连接。
2. 在**应用发现**页面，选择**资源列表**页签，单击**存储资源**的**资源数量**，进入**资源列表**页面的**存储**页签。
3. 在对象存储资源列表中，单击**深度采集配置**列的“**配置**”，弹出**配置**窗口。
4. 选择源端安装的Edge工具和相应的资源凭证，单击“**确认**”完成配置。如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。
5. 配置完成后，单击操作列的“**添加前缀**”，弹出**添加前缀**窗口。
6. 输入指定的资源目录路径进行采集，不输入则默认采集全桶资源，单击“**确定**”保存前缀设置。
7. 完成深度采集配置和添加前缀后，单击操作列的“**深度采集**”，系统开始进行对象存储资源的深度采集。支持进行多次深度采集。当采集状态为**采集完成**后，单击资源名称，可以查看采集到的信息。

## 查看发现的源端资源列表和资源详情

发现任务仅能获取源端资源的基本信息，更多详细信息需要通过深度采集获取。

1. 在**应用发现**页面，选择“**资源列表**”页签，可以查看发现的各类型资源数量和未关联应用的资源数量。



2. 单击资源类型或资源数量，可以查看对应的资源类型列表。
3. 单击资源操作列的“查看资源”，可以查看该资源的基本信息。

## 查看任务详情


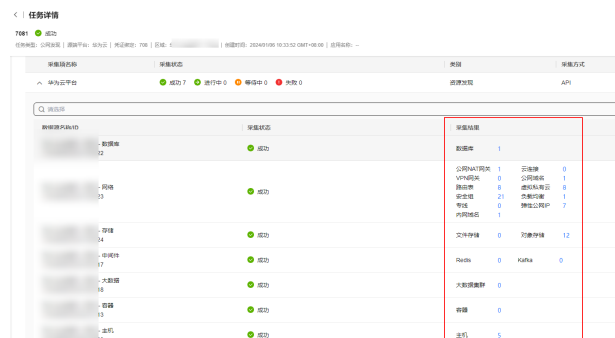
- 步骤1 在应用发现页面，单击任务总数后的“查看”，进入任务列表页面。
- 步骤2 单击采集任务操作列的“查看”，进入任务详情页面，可以查看到该任务包含的采集项。
- 步骤3 单击采集项前的  按钮，可以查看该采集项包含的采集数据源以及各资源类型的采集结果。

图 4-1 采集结果



- 步骤4 在采集结果列，单击资源类型后的数量，可以跳转到对应的资源类型列表页面。

----结束

## 重新采集

只有所有采集源状态为**成功**且采集方式为API的采集项和采集状态为**成功**的数据源，才可以使用重新采集功能；采集失败的数据源和包含采集失败数据源的采集项无法使用重新采集功能，需要删除失败的数据源后通过“**新增数据源**”进行重新采集。

- 步骤1 在任务列表页面，单击采集任务操作列的“查看”，进入任务详情页面。
- 步骤2 选择需要重新采集的**数据源**，单击操作列的“重新采集”，单击“确定”按钮，开始重新采集。

选择需要重新采集的**采集项**，单击操作列的“更多>重新采集”，单击“确定”按钮，开始重新采集。

----结束

## 查看应用关联分析结果

资源采集完成，迁移中心会整理源端资源和应用间的关联关系，详情请[查看应用关联分析结果](#)。

## 导出应用关联信息

在应用发现页面，单击页面右上角的“导出管理”，可以导出接入层调用链、数据库依赖表、中间件依赖表、应用关联表，用于下线分析。

## 4.2.2 创建内网采集任务

适用于暂不支持公网采集的云平台主机，在源端内网环境中部署Edge工具，通过网段扫描或VMware扫描的方式，发现和采集源端主机资源。

### 使用须知

- 目前支持采集的VMware版本包括：VMware 5.0~7.0
- VMware扫描和网段扫描两种方法都采用私有IP结合Edge ID的方式来标识内网中的主机资源。这意味着，一旦完成一次资源采集，如果被采集的主机私有IP发生变更，系统将在下一次采集时将其识别为一个新的主机资源，从而导致主机资源数量增加。为了避免这种情况，建议在迁移过程完成之前，不要更改源端主机的私有IP地址。

### 前提条件

- 已在源端内网环境[安装Edge工具](#)，并与迁移中心连接。
- 已在Edge上[添加源端主机凭证](#)。

### 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>应用发现”，进入应用发现页面。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** 首次创建任务，在云边协同区域，单击“内网发现”按钮，进入创建内网发现任务页面。

非首次创建任务，单击任务框中的“发现>内网发现”，进入创建内网发现任务页面。

**步骤4** 根据[表4-4](#)，配置内网发现任务参数。

表 4-4 内网发现任务参数说明

| 参数   | 说明                                     |
|------|----------------------------------------|
| 任务名称 | 自定义发现任务名称。                             |
| 任务说明 | 输入发现任务说明。                              |
| 设备   | 选择在源端内网环境中安装的 <a href="#">Edge工具</a> 。 |

**步骤5** 根据实际情况，选择使用**网段扫描**或**VMware扫描**发现主机资源。

- 选择**网段扫描**，根据**表4-5**，配置网段扫描参数。

**表 4-5** 网段扫描参数说明

| 参数      | 说明                                                                                                                |
|---------|-------------------------------------------------------------------------------------------------------------------|
| 协议      | 选择通信协议 <b>TCP</b> 或 <b>ICMP</b> 。                                                                                 |
| 网段      | 输入IP地址段，需在以下范围内：<br>- 10.0.0.0 ~10.255.255.255<br>- 172.16.0.0 ~172.31.255.255<br>- 192.168.0.0 ~ 192.168.255.255 |
| Linux   | 填写Linux系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Linux类型主机，可将端口设为0。                                                          |
| Windows | 填写Windows系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Windows类型主机，可将端口设为0。                                                      |

- 选择**VMware扫描**，在**连接地址**填写**vCenter**的IP地址，并选择**vCenter**的凭证。如未提前**添加凭证**，请前往迁移工具Edge添加私有云凭证。凭证中的用户名、密码为**vCenter**的登录用户名和密码。

**步骤6** 单击“**确认**”，内网发现任务创建完成，系统开始主机资源发现。

在**应用发现**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。

**步骤7** 内网发现任务状态为成功后，即可进行主机深度采集。（通过内网发现的主机，已绑定了采集设备，仅需为主机绑定凭证即可进行深度采集。）

1. 在**应用发现**页面，选择**资源列表**页签，单击**主机资源**的**资源数量**，进入**资源列表**页面的**主机**页签。
2. 单击主机凭证列的“**绑定**”，弹出**绑定凭证**窗口。
3. 选择主机凭证，如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。
4. 单击“**确定**”，系统会自动检查凭证绑定状态。当采集状态为**已就绪**时，单击采集状态列的“**采集**”进行深度采集。首次深度采集完成后，可以单击采集状态列的“**重新采集**”按钮，进行多次深度采集。

----结束

### 4.2.3 导入应用关联文件

将您源端配置管理数据库中的应用关联信息，填入导入模板表格中，再将表格导入迁移中心，分析出应用间的依赖及关联关系。

#### 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>应用发现**”，进入**应用发现**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** 首次创建任务，在**工具采集区域**，单击“**导入>应用关联模板**”；非首次创建任务，单击**任务框**中的“**导入>应用关联模板**”，弹出**导入应用关联模板窗口**，单击“**下载模板**”，将导入模板下载到本地。

**步骤4** 打开导入模板，填写模板内的参数，其中标黄的参数为必填。模板参数填写完成并保存后，单击导入窗口的“**添加文件**”，将填写后的模板文件上传至迁移中心。

**步骤5** 上传成功后，单击“**确认**”按钮，系统自动开始采集。

您可以在**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。处理完失败原因后，需要删除该采集任务，重新导入文件进行采集。
- 当任务状态为“**成功**”时，您可以前往**应用发现**页面，[查看应用关联分析结果](#)。

---结束

## 4.2.4 导入工具采集结果

通过线下采集器，采集本地资源信息，并将采集结果上传到迁移中心进行应用关联分析。

### 前提条件

已[创建工具采集任务](#)，并获取采集结果。

### 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>应用发现**”，进入**应用发现**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** 首次创建任务，在**工具采集区域**，单击“**导入>工具采集结果**”按钮；非首次创建任务，单击**任务框**中的“**导入>工具采集结果**”，弹出**导入工具采集结果窗口**。

**步骤4** 单击导入窗口的“**添加文件**”，将工具采集结果的json文件上传至迁移中心。

**步骤5** 上传成功后，单击“**确认**”按钮，系统自动开始采集。

您可以在**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。处理完失败原因后，需要删除该采集任务，重新导入文件进行采集。
- 当任务状态为“**成功**”时，您可以前往**应用发现**页面，[查看应用关联分析结果](#)。

---结束

## 4.2.5 导入阿里云主机资源清单

通过导入阿里云ECS的实例和云盘资源清单（CSV格式），系统将自动生成主机资源信息。利用这些信息，您还可以使用迁移中心的规格评估功能，为您推荐与阿里云资源相匹配的华为云资源规格。



## 导出阿里云 ECS 实例和云盘

**步骤1** 登录阿里云ECS控制台。

**步骤2** 在概览页面，单击“导出数据>实例”，导出所有实例资源清单。



**步骤3** 单击“导出数据>云盘”，导出所有云盘资源清单。

**步骤4** 打开实例资源清单和云盘资源清单，根据实际需求，调整清单内的资源信息。请确保清单中包含了以下必要信息：

| 清单类型 | 必选字段                   |
|------|------------------------|
| 实例清单 | 实例ID、操作系统、实例规格、CPU、内存  |
| 云盘清单 | 实例Id、云盘Id、容量（GiB）、云盘属性 |

----结束

## 导入阿里云主机资源清单

**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>应用发现”，进入应用发现页面。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** 首次创建任务，在工具采集区域，单击“导入>阿里云主机资源清单”按钮；非首次创建任务，单击任务框中的“导入>阿里云主机资源清单”，弹出导入阿里云主机资源清单窗口。

### 须知

仅支持上传UTF-8编码且为中文表头的CSV文件，单个文件大小不能超过15MB。

**步骤4** 单击实例资源清单后的“添加文件”，将[导出的实例资源清单](#)上传。

**步骤5** 实例资源清单上传后，单击云盘资源清单后的“添加文件”，将[导出的云盘资源清单](#)上传。

**步骤6** 实例资源清单与云盘资源清单上传成功后，在资源列表中选择主机，进入主机资源列表，可以查看导入的阿里云主机资源列表。

**步骤7** （可选）将导入的阿里云主机资源[关联到应用](#)，并使用[规格评估](#)功能，获取与阿里云资源相匹配的华为云资源规格。

----结束

## 4.2.6 查看应用关联分析结果

迁移中心会整理源端资源和应用间的关联关系，并且绘制技术架构图和应用依赖图，为后续迁移计划的制定和目的端架构设计提供帮助。

### 前提条件

需要完成以下任意一个任务：

- 已完成[公网采集](#)。
- 已完成[内网采集](#)。
- 已[导入应用关联文件](#)。
- 已[导入工具采集结果](#)。

### 查看技术架构图

**步骤1** 在应用发现页面的应用列表页签，单击右上角的“技术架构图”，可以查看采集到的所有服务、微服务、资源的矩阵图。

**步骤2** 选择服务，单击鼠标右键，可以查看该服务的技术架构图、服务依赖图和包含的微服务列表。

----结束

### 查看业务架构图

**步骤1** 在应用发现页面的应用列表页签，单击右上角的“业务架构图”，可以查看采集到的所有服务矩阵图。

**步骤2** 选择服务，单击鼠标右键，可以查看该服务的技术架构图、服务依赖图和包含的微服务列表。

----结束

### 查看资源列表

**步骤1** 在应用发现页面，选择“资源列表”页签，可以查看发现各类型资源数量和未关联应用的资源数量。

**步骤2** 单击资源类型或资源数量，可以查看对应的资源类型列表。单击资源操作列的“查看资源”，可以查看该资源的基本信息。

----结束

### 查看应用列表

在应用发现页面的应用列表页签，可以查看采集到的所有应用。通过应用列表可以查看应用名称、应用类型、所属业务领域、包含微服务数量以及依赖应用和被依赖应用数量等信息。

- 单击微服务列的数量，可以查看该应用所包含的微服务列表。
- 单击依赖应用列的数量，可以查看该应用所依赖的服务列表。
- 单击被依赖应用列的数量，可以查看依赖该应用的应用列表。

- 单击操作列的“**技术架构图**”，可以查看该应用包含的微服务、资源所属层级以及微服务与资源之间的关联关系。单击微服务/资源节点，可以查看节点详细信息。
- 单击操作列的“**应用依赖图**”，可以查看该应用与依赖服务、被依赖服务之间的关联关系。单击服务节点，可以查看服务的详细信息和依赖服务、被依赖服务列表。

## 查看资源线索

**步骤1** 在应用发现页面，单击源端资源发现框中的“**线索**”按钮，进入**线索**页面。

**步骤2** 选择**资源线索**页签，可以查看采集到的所有资源列表以及资源的详细信息。

- 单击操作列的“**服务依赖图**”，可以查看该资源与依赖服务、被依赖服务之间的关联关系。
- 单击操作列的“**删除**”，可以删除该资源。

----结束

## 查看微服务信息

**步骤1** 在应用发现页面，单击源端资源框中的“**线索**”按钮，进入**线索**页面。

**步骤2** 选择**微服务列表**页签，可以查看采集到的所有微服务列表。

- 单击操作列的“**技术架构图**”，可以查看该微服务所属层级以及与资源之间的关联关系。单击微服务/资源节点，可以查看节点详细信息。
- 单击操作列的“**更多>服务依赖图**”，可以查看该微服务与依赖服务、被依赖服务之间的关联关系。单击微服务节点，可以查看微服务的详细信息和依赖服务、被依赖服务列表。
- 单击操作列的“**更多>删除**”，可以删除该微服务。

----结束

## 导出应用关联信息

在应用发现页面，单击应用框中的“**导出**”，可以导出接入层调用链、数据库依赖表、中间件依赖表、应用关联表，用于下线分析。

# 4.3 资源发现与采集（简单项目）

## 4.3.1 创建公网采集任务

通过公网在线发现并采集源端云平台资源，获取源端资源详细信息。目前支持的源端云平台为：**阿里云、华为云、AWS、腾讯云、谷歌云、Azure**。

- **阿里云和华为云**支持采集的资源类型包括：**主机、容器、中间件、数据库、大数据、网络和存储**。
- **AWS**支持采集的资源类型包括：**主机、数据库、存储**。
- **谷歌云和Azure**支持采集的资源类型包括：**主机、容器、中间件、数据库、存储和网络**。

## 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>应用发现](#)”，选择应用发现页签。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** 首次创建任务，在云平台发现区域，单击“[公网发现](#)”按钮，进入[创建公网发现任务](#)页面。

非首次创建任务，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。

**步骤4** 根据[表4-6](#)，配置公网发现任务参数。

**表 4-6** 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   | 选择源端云平台，目前支持的云平台有 <a href="#">阿里云</a> 、 <a href="#">华为云</a> 、 <a href="#">AWS</a> 、 <a href="#">腾讯云</a> 、 <a href="#">谷歌云</a> 、 <a href="#">Azure</a> 。                                                                                                                                                                                                                                                                                                                                                                    | 是    |
|      | 凭证   | 选择源端对应的凭证，如未提前 <a href="#">添加凭证</a> ，请单击“ <a href="#">新建凭证</a> ”，填写相应参数，校验并保存凭证。<br><br><ul style="list-style-type: none"> <li><a href="#">阿里云</a>、<a href="#">华为云</a>、<a href="#">AWS</a>、<a href="#">腾讯云</a>认证方式请选择“AK/SK”，并输入源端账号的AK/SK。</li> <li><a href="#">谷歌云</a>认证方式请选择“<a href="#">配置文件</a>”，并上传<a href="#">谷歌云服务账号凭据</a>，该凭证必须为json格式文件，且大小不能超过4KB。</li> <li><a href="#">Azure</a>认证方式请选择“<a href="#">ID/密钥</a>”。添加凭证所需的信息获取方法请参见<a href="#">如何获取添加Azure凭证所需的信息</a>。</li> </ul> <p><b>说明</b><br/>存储位置选择“云上”。</p> | 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[资源类型](#)下拉列表，勾选需要进行采集的资源类型（采集项）。当前支持采集的云平台和资源类型参见[表4-7](#)。

表 4-7 支持采集的资源类型

| 云平台 | 资源类型                                                                                               | 子类型                                                                                                                                                                                                                                 |
|-----|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 阿里云 | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 大数据</li> <li>• 数据库</li> </ul> | -                                                                                                                                                                                                                                   |
|     | 中间件                                                                                                | <ul style="list-style-type: none"> <li>• Redis</li> <li>• Kafka</li> </ul>                                                                                                                                                          |
|     | 存储                                                                                                 | <ul style="list-style-type: none"> <li>• 对象存储</li> <li>• 文件存储</li> </ul>                                                                                                                                                            |
|     | 网络                                                                                                 | <ul style="list-style-type: none"> <li>• 云连接</li> <li>• 负载均衡（ALB、CLB）</li> <li>• 专线</li> <li>• 公网域名</li> <li>• 内网域名</li> <li>• 弹性公网IP</li> <li>• 公网NAT网关</li> <li>• 路由表</li> <li>• 安全组</li> <li>• 虚拟私有云</li> <li>• VPN网关</li> </ul> |
| 华为云 | <ul style="list-style-type: none"> <li>• 主机</li> <li>• 容器</li> <li>• 大数据</li> <li>• 数据库</li> </ul> | -                                                                                                                                                                                                                                   |
|     | 中间件                                                                                                | <ul style="list-style-type: none"> <li>• Redis</li> <li>• Kafka</li> </ul>                                                                                                                                                          |
|     | 存储                                                                                                 | <ul style="list-style-type: none"> <li>• 对象存储</li> <li>• 文件存储</li> </ul>                                                                                                                                                            |

| 云平台   | 资源类型                                                                                | 子类型                                                                                                                                                                               |
|-------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|       | 网络                                                                                  | <ul style="list-style-type: none"> <li>● 负载均衡（ELB）</li> <li>● 公网域名</li> <li>● 内网域名</li> <li>● 弹性公网IP</li> <li>● 公网NAT网关</li> <li>● 路由表</li> <li>● 安全组</li> <li>● 虚拟私有云</li> </ul> |
| AWS   | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 数据库</li> </ul>               | -                                                                                                                                                                                 |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
| 腾讯云   | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 数据库</li> </ul>               | -                                                                                                                                                                                 |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
| 谷歌云   | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 容器</li> <li>● 数据库</li> </ul> | -                                                                                                                                                                                 |
|       | 中间件                                                                                 | Redis                                                                                                                                                                             |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
|       | 网络                                                                                  | <ul style="list-style-type: none"> <li>● 弹性公网IP</li> <li>● 路由表</li> <li>● 安全组</li> <li>● 虚拟私有云</li> </ul>                                                                         |
| Azure | <ul style="list-style-type: none"> <li>● 主机</li> <li>● 容器</li> <li>● 数据库</li> </ul> | -                                                                                                                                                                                 |
|       | 存储                                                                                  | <ul style="list-style-type: none"> <li>● 对象存储</li> <li>● 文件存储</li> </ul>                                                                                                          |
|       | 中间件                                                                                 | <ul style="list-style-type: none"> <li>● Redis</li> <li>● Kafka</li> </ul>                                                                                                        |

| 云平台 | 资源类型 | 子类型                                                                                                                                   |
|-----|------|---------------------------------------------------------------------------------------------------------------------------------------|
|     | 网络   | <ul style="list-style-type: none"> <li>弹性公网IP</li> <li>路由表</li> <li>安全组</li> <li>公网NAT网关</li> <li>虚拟私有云</li> <li>负载均衡（ELB）</li> </ul> |

**步骤6** （可选）将采集的资源关联到应用。

- 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](#)；根据实际需求，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](#)”，公网发现任务创建完成，系统自动开始资源发现。

- 您可以在[应用发现](#)页面，[查看发现的源端资源列表和资源详情](#)。
- 您可以在任务列表页面，查看发现任务状态和[查看任务详情](#)。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。



**步骤8** 发现任务状态为成功后，根据不同资源类型，进行后续操作。

- 主机资源需要完成[迁移预检查](#)或[深度采集](#)后，才可以进行[迁移方案设计](#)。
- 容器资源需要完成[深度采集](#)后，才可以进行[迁移方案设计](#)。
- 中间件、数据库、大数据、存储等资源可直接进行[迁移方案设计](#)。
- 对象存储资源可以进行[深度采集](#)和[迁移方案配置](#)。

---结束

## 迁移预检查

请按照以下步骤对主机资源进行迁移预检查（源端迁移准备度），确定源端主机是否满足迁移条件。

- 在源端内网环境中[安装Edge工具](#)并完成注册。
- 在[应用发现](#)页面，选择[资源列表](#)页签，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。

4. 单击源端迁移准备度列的“配置”，弹出配置窗口。
5. 根据表4-8，配置参数。

表 4-8 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。                                                                                                     |
| IP   | 选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入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。                                |
| 端口   | 主机开放端口。 <ul style="list-style-type: none"><li>• Windows主机默认为5985端口，无法修改。</li><li>• Linux主机默认22端口，可根据实际情况修改。</li></ul> |
| 凭证   | 选择主机凭证，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。                                                                      |

6. 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开始源端迁移准备度检查，当源端迁移准备度列状态会变为已就绪时，代表该主机已满足迁移条件，可以进行下一阶段迁移方案设计。

## 主机/容器深度采集

请按照以下步骤对主机资源和容器资源进行深度采集。

1. 在源端内网环境中安装Edge工具并与迁移中心连接。
2. 在应用发现页面，选择资源列表页签，单击主机/容器资源的资源数量，进入资源列表页面的主机/容器页签。
3. 在需要进行深度采集的主机/容器采集设备列，单击“绑定”，弹出绑定采集设备窗口。

如果需要批量绑定，可以同时勾选多个资源，单击页面右上角的“绑定采集设备”，弹出绑定采集设备窗口。
4. 选择源端安装的Edge工具，以及网络接入方式（如果采集的资源与安装Edge的主机处于同一VPC内，可以选择私网接入，否则需要选择公网接入），单击“确定”。绑定成功后，资源的采集设备列状态会变为已绑定。
5. 采集设备绑定成功后，还需要为资源绑定凭证，单击资源凭证列的“绑定”，弹出绑定凭证窗口。
6. 选择资源凭证，如未提前在Edge上添加源端资源凭证，请前往Edge控制台添加，并同步至迁移中心。
7. 单击“确定”，系统会自动检查凭证绑定状态。当采集状态为已就绪时，单击采集状态列的“采集”进行深度采集。首次深度采集完成后，可以单击采集状态列的“重新采集”按钮，进行多次深度采集。采集完成后，可以进行下一阶段迁移方案设计。



## 对象存储深度采集

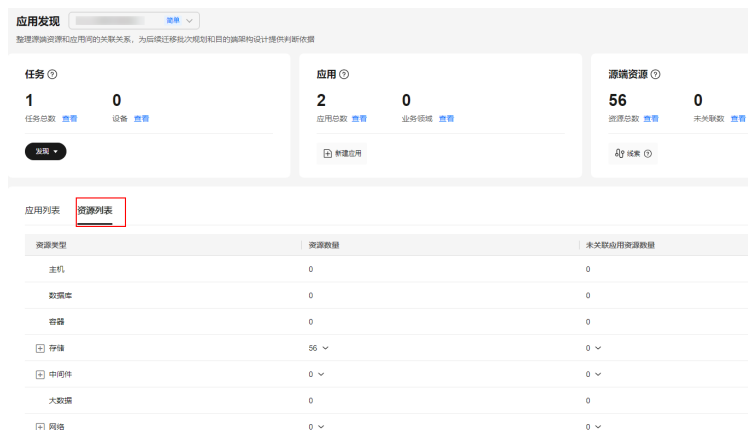
通过深度采集获取对象存储资源的详细信息，以提高迁移集群规格评估的准确性。请按照以下步骤进行**对象存储资源**深度采集。

1. 在源端内网环境中**安装Edge工具**并与迁移中心连接。
2. 在**应用发现**页面，选择**资源列表**页签，单击**存储资源**的**资源数量**，进入**资源列表**页面的**存储**页签。
3. 在对象存储资源列表中，单击**深度采集配置**列的“**配置**”，弹出**配置**窗口。
4. 选择源端安装的Edge工具和相应的资源凭证，单击“**确认**”完成配置。如未提前在Edge上**添加源端资源凭证**，请前往Edge控制台添加，并同步至迁移中心。
5. 配置完成后，单击操作列的“**添加前缀**”，弹出**添加前缀**窗口。
6. 输入指定的资源目录路径进行采集，不输入则默认采集全桶资源，单击“**确定**”保存前缀设置。
7. 完成深度采集配置和添加前缀后，单击操作列的“**深度采集**”，系统开始进行对象存储资源的深度采集。支持进行多次深度采集。当采集状态为**采集完成后**，单击资源名称，可以查看采集到的信息。

## 查看发现的源端资源列表和资源详情

发现任务仅能获取源端资源的基本信息，更多详细信息需要通过深度采集获取。

1. 在**应用发现**页面，选择“**资源列表**”页签，可以查看发现的各类型资源数量和未关联应用的资源数量。



2. 单击资源类型或资源数量，可以查看对应的资源类型列表。
3. 单击资源操作列的“**查看资源**”，可以查看该资源的基本信息。

## 查看任务详情


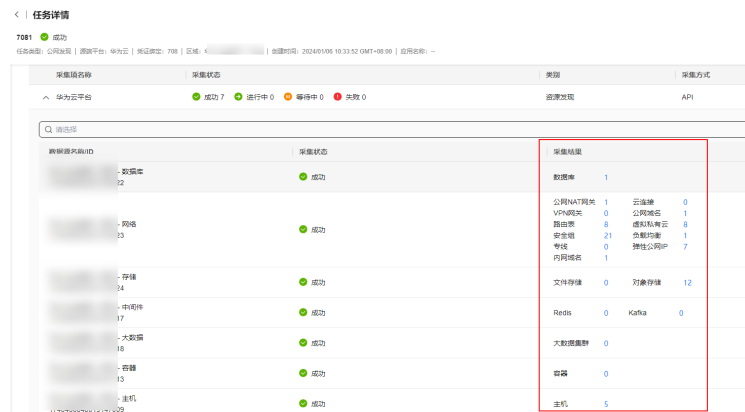
- 步骤1** 在**应用发现**页面，单击**任务总数**后的“**查看**”，进入**任务列表**页面。
- 步骤2** 单击采集任务操作列的“**查看**”，进入**任务详情**页面，可以查看到该任务包含的采集项。
- 步骤3** 单击采集项前的  按钮，可以查看该采集项包含的采集数据源以及各资源类型的采集结果。

图 4-2 采集结果



**步骤4** 在采集结果列，单击资源类型后的数量，可以跳转到对应的资源类型列表页面。

----结束

## 重新采集

只有所有采集源状态为**成功**且采集方式为API的采集项和采集状态为**成功**的数据源，才可以使用重新采集功能；采集失败的数据源和包含采集失败数据源的采集项无法使用重新采集功能，需要删除失败的数据源后通过“**新增数据源**”进行重新采集。

**步骤1** 在任务列表页面，单击采集任务操作列的“**查看**”，进入**任务详情**页面。

**步骤2** 选择需要重新采集的**数据源**，单击操作列的“**重新采集**”，单击“**确定**”按钮，开始重新采集。

选择需要重新采集的**采集项**，单击操作列的“**更多>重新采集**”，单击“**确定**”按钮，开始重新采集。

----结束

## 4.3.2 创建内网采集任务

适用于暂不支持公网采集的云平台主机，在源端内网环境中部署Edge工具，通过**网段扫描**或**VMware扫描**的方式，发现和采集源端主机资源。

## 使用须知

- 目前支持采集的VMware版本包括：**VMware 5.0~7.0**
- VMware扫描和网段扫描两种方法都采用私有IP结合Edge ID的方式来标识内网中的主机资源。这意味着，一旦完成一次资源采集，如果被采集的主机私有IP发生变更，系统将在下一次采集时将其识别为一个新的主机资源，从而导致主机资源数量增加。为了避免这种情况，建议在迁移过程完成之前，不要更改源端主机的私有IP地址。

## 前提条件

- 已在源端内网环境**安装Edge工具**，并与迁移中心连接。
- 已在Edge上**添加源端主机凭证**。

## 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>应用发现**”，进入**应用发现**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** 首次创建任务，在**云边协同**区域，单击“**内网发现**”按钮，进入**创建内网发现任务**页面。

非首次创建任务，单击任务框中的“**发现>内网发现**”，进入**创建内网发现任务**页面。

**步骤4** 根据**表4-9**，配置内网发现任务参数。

**表 4-9** 内网发现任务参数说明

| 参数   | 说明                            |
|------|-------------------------------|
| 任务名称 | 自定义发现任务名称。                    |
| 任务说明 | 输入发现任务说明。                     |
| 设备   | 选择在源端内网环境中安装的 <b>Edge工具</b> 。 |

**步骤5** 根据实际情况，选择使用**网段扫描**或**VMware扫描**发现主机资源。

- 选择**网段扫描**，根据**表4-10**，配置网段扫描参数。

**表 4-10** 网段扫描参数说明

| 参数      | 说明                                                                                                                                                                           |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 协议      | 选择通信协议 <b>TCP</b> 或 <b>ICMP</b> 。                                                                                                                                            |
| 网段      | 输入IP地址段，需在以下范围内： <ul style="list-style-type: none"> <li>- 10.0.0.0 ~10.255.255.255</li> <li>- 172.16.0.0 ~172.31.255.255</li> <li>- 192.168.0.0 ~ 192.168.255.255</li> </ul> |
| Linux   | 填写Linux系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Linux类型主机，可将端口设为0。                                                                                                                     |
| Windows | 填写Windows系统类型主机需要扫描的端口。仅TCP协议需要填写，若不需要扫描Windows类型主机，可将端口设为0。                                                                                                                 |

- 选择**VMware扫描**，在**连接地址**填写**vCenter**的IP地址，并选择**vCenter**的凭证。如未提前**添加凭证**，请前往迁移工具Edge添加私有云凭证。凭证中的用户名、密码为**vCenter**的登录用户名和密码。

**步骤6** 单击“**确认**”，内网发现任务创建完成，系统开始主机资源发现。

在**应用发现**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。

**步骤7** 内网发现任务状态为成功后，即可进行主机深度采集。（通过内网发现的主机，已绑定了采集设备，仅需为主机绑定凭证即可进行深度采集。）

1. 在应用发现页面，选择资源列表页签，单击主机资源的资源数量，进入资源列表页面的主机页签。
2. 单击主机凭证列的“绑定”，弹出绑定凭证窗口。
3. 选择主机凭证，如未提前在Edge上添加源端资源凭证，请前往Edge控制台添加，并同步至迁移中心。
4. 单击“确定”，系统会自动检查凭证绑定状态。当采集状态为已就绪时，单击采集状态列的“采集”进行深度采集。首次深度采集完成后，可以单击采集状态列的“重新采集”按钮，进行多次深度采集。

---结束

## 4.4 手动添加资源

目前仅支持手动添加主机。通过在源端环境中安装Edge工具进行资源采集。

### 前提条件

- 已在源端内网环境安装Edge工具，并与迁移中心连接。
- 已在Edge上添加源端主机凭证。

### 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>应用发现”，进入应用发现页面。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资源类型或资源数量，进入资源列表页面的主机页签。

**步骤4** 单击“添加资源”按钮，弹出添加主机窗口。

**步骤5** 根据表4-11，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 4-11 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。<br>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"><li>• Windows主机默认为5985端口，无法修改。</li><li>• Linux主机默认22端口，可根据实际情况修改。</li></ul> |

| 参数 | 说明                                                                 |
|----|--------------------------------------------------------------------|
| 凭证 | 选择主机凭证，如未提前在Edge上 <a href="#">添加源端主机凭证</a> ，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

## 4.5 采集主机性能

迁移中心支持对主机资源进行性能数据采集，获取源端主机的CPU、内存负载、磁盘IOPS，内网出入流量，网络出入包速率，网络连接数等性能信息。在后续进行目的端规格评估时，可以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。

### 使用须知

- 开始采集后，默认每5分钟进行一次数据采集。
- 建议采集时长≥1小时，若采集时长不足，将无法获取7天和30天最大值以及平均值数据。
- 采集周期默认为7天，7天后会自动停止采集；采集周期内可以进行停止采集和重启采集操作；重启采集后，采集周期重新计算。
- 采集到的性能数据将保留180天，超期后会自动删除。
- 迁移工具Edge建议安装在独立主机上，避免与业务主机共用，以确保采集的稳定性和效率。
- Linux主机：迁移工具Edge最多同时采集1000台Linux主机。当同时采集1000台Linux主机时，安装Edge的主机规格最低要求为8U16G，且可使用内存不小于8G。
- Windows主机：迁移工具Edge最多同时采集500台Windows主机，当同时采集500台Windows主机时，安装Edge的主机规格最低要求为16U32G，且可使用内存不小于8G。

### 前提条件

- 已通过[公网采集/内网采集/手动添加](#)等方式完成主机资源采集。
- 已在源端内网环境[安装Edge工具](#)，并与迁移中心连接。
- 已在迁移工具 Edge上[添加源端主机凭证](#)。

### 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>应用发现](#)”，进入应用发现页面。在当前项目下拉列表中选择[迁移项目](#)。

**步骤3** 选择[资源列表](#)页签，单击主机资源的资源数量，进入[资源列表](#)页面的主机页签。

**步骤4** 为需要进行性能采集的主机配置[迁移工具](#)和[凭证](#)。如果性能采集列的“[开始](#)”按钮未置灰，代表已完成配置，可以跳过此步执行[步骤7](#)。

将鼠标放置性能采集列“开始”按钮处，在弹出的提示框中，单击“去配置”，弹出配置窗口。

**步骤5** 根据表4-12，配置参数。

表 4-12 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。                                                                                                 |
| IP   | 选择主机接入IP，可以为公网IP，也可以是私有IP；在需要使用代理连接主机的情况下，也可以自行输入接入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。                            |
| 端口   | 主机开放端口。 <ul style="list-style-type: none"><li>Windows主机默认为5985端口，无法修改。</li><li>Linux主机默认22端口，可根据实际情况修改。</li></ul> |
| 凭证   | 选择主机凭证，如未提前在Edge上 <a href="#">添加源端主机凭证</a> ，请前往Edge控制台添加，并同步至迁移中心。                                                |

**步骤6** 配置完成后，单击“确认”按钮，系统会校验配置信息的正确性。

**步骤7** 配置信息校验无误后，单击性能采集列的“开始”，开始主机性能采集，此时状态为“采集中”。

开始性能采集后，您还可以进行如下操作：

- **查看采集数据**
  - a. 单击主机名称，进入主机详情页面，在性能信息区域，可以查看采集到的主机性能数据。其中，7天和30天最大值为每小时汇总的定时任务，采集时长需超过1小时才能获取数据。
  - b. 单击操作列的“查看”，可以查看采集周期内每项数据的详细数值和曲线图。
- **停止采集**

单击“停止”，可以暂停性能数据采集。

----结束

## 4.6 关联应用

通过应用将资源进行分组管理，便于后续进行规格评估和创建迁移工作流。

**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>应用发现”，进入应用发现页面。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** 在资源列表页签，选择资源类型，进入对应的资源列表页面。

- 步骤4** 勾选需要关联到同一应用的资源，单击页面左上角的“**关联到应用**”，弹出**关联到应用**窗口。
- 步骤5** 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**创建应用**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**选择目的端所在区域，单击“**确定**”按钮，应用创建成功。
- 步骤6** 选择应用后，单击“**确定**”。在各资源的**已关联应用**列，查看资源所关联应用。
- 结束**

# 5 大数据血缘

## 5.1 部署迁移工具

使用迁移中心的大数据血缘功能前，需要在可以访问源端元数据来源的环境中部署迁移工具Edge。

### 操作步骤

**步骤1** 安装Edge的方法请参考[下载并安装Edge](#)。

**步骤2** 安装成功后，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** Edge与迁移中心连接成功后，添加大数据相关凭证，方法请参考[添加资源凭证](#)。

----结束

## 5.2 采集元数据

### 5.2.1 创建元数据连接

创建元数据来源类型的连接，获取源端大数据数据库表清单。目前支持的连接类型包括：

- **Hive Metastore**连接
- **Doris**连接
- **HBase**连接
- **ClickHouse**连接
- **Glue**连接
- **MaxCompute**连接



### 须知

- 在创建Hive Metastore连接前，需要先获取连接Hive Metastore的IP地址和端口号。这些信息可以通过Hive集群的管理员或在Hive的配置文件中获取。
- 在创建MaxCompute连接前，需要先获取MaxCompute服务所在地域的Endpoint信息。
- 在创建Doris连接前，需要先获取Doris集群部署的服务器IP地址和端口。这些信息可以通过Doris集群管理员或在Doris集群配置文件中获取。
- 在创建HBase连接前，需要先获取HBase集群中至少一个ZooKeeper节点的IP地址和连接端口。这些信息可以通过HBase集群管理员或查看服务器的网络配置来获取。

## 前提条件

已在源端元数据来源的环境中部署迁移工具Edge。

## 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“[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[迁移项目](#)。

**步骤3** 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4** 选择连接类型，单击“[下一步](#)”，进入[详细配置](#)页面。

**步骤5** 根据所选的连接类型，查看对应的参数配置说明，配置连接参数。

- [Hive Metastore连接参数配置说明](#)
- [Doris连接参数配置说明](#)
- [HBase连接参数配置说明](#)
- [ClickHouse组件参数配置说明](#)
- [Glue连接参数配置说明](#)
- [MaxCompute连接参数配置说明](#)

表 5-1 Hive Metastore 连接参数配置说明

| 参数   | 配置说明                                           |
|------|------------------------------------------------|
| 所属源  | 仅需在源端创建元数据连接。                                  |
| 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 | 选择在源端环境中 <a href="#">部署的迁移工具Edge</a> 。         |

| 参数                   | 配置说明                                                                                                                                                                                       |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 启用安全连接               | 选择是否启用安全连接。 <ul style="list-style-type: none"> <li>MetaStore在非安全集权下，可以不启用安全连接。</li> <li>MetaStore在安全集群下，需要启用安全连接并选择凭证。凭证的获取以及添加方法请参考<a href="#">添加资源凭证</a>中的大数据-Hive Metastore。</li> </ul> |
| 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。                                                                                                                                                                  |

表 5-2 Doris 连接参数配置说明

| 参数      | 配置说明                                                                  |
|---------|-----------------------------------------------------------------------|
| 所属源     | 仅需在源端创建元数据连接。                                                         |
| 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    | 选择在源端环境中 <a href="#">部署的迁移工具Edge</a> 。                                |
| Doris凭证 | 选择在Edge添加的Doris凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-Doris”。 |
| 数据库地址   | 填写连接到Doris集群部署的服务器IP地址。                                               |
| 数据库端口   | 填写连接到Doris集群部署的服务器的端口。默认为3306。                                        |
| 数据库名    | 填写Doris数据库名称。                                                         |

表 5-3 HBase 连接参数配置说明

| 参数   | 配置说明                                                                                                                                          |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 所属源  | 仅需在源端创建元数据连接。                                                                                                                                 |
| 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 | 选择在源端环境中 <a href="#">部署的迁移工具Edge</a> 。<br><b>注意</b><br>需要在安装Edge的主机上配置/etc/hosts后，HBase元数据连接才能连接成功。host文件的获取方法请参见 <a href="#">准备开发和运行环境</a> 。 |

| 参数              | 配置说明                                                                  |
|-----------------|-----------------------------------------------------------------------|
| HBase凭证         | 选择在Edge添加的HBase凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-HBase”。 |
| 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。                  |
| Zookeeppeer连接端口 | 填写连接到Zookeeppeer的端口。                                                  |
| HBase 版本        | 选择HBase的版本。                                                           |

表 5-4 ClickHouse 连接参数配置说明

| 参数                   | 配置说明                                                                       |
|----------------------|----------------------------------------------------------------------------|
| 所属源                  | 仅需在源端创建元数据连接。                                                              |
| 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                 | 选择在源端环境中 <a href="#">部署的迁移工具Edge</a> 。                                     |
| ClickHouse凭证（可选）     | 选择在Edge添加的ClickHouse凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-ClickHouse。 |
| 安全集群                 | 选择是否部署在安全集群下。                                                              |
| ClickHouse Server IP | 填写连接到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ClickHouse 服务器通信的 HTTP 端口。默认为8123。                                 |
| HTTP SSL/TLS端口       | 安全集群下，与ClickHouse 服务器通信的HTTPS 端口。                                          |

表 5-5 Glue 连接参数配置说明

| 参数     | 配置说明                                                           |
|--------|----------------------------------------------------------------|
| 所属源    | 仅需在源端创建元数据连接。                                                  |
| 连接名称   | 默认创建名称为“Glue-4位随机数（字母和数字）”，也可以自定义名称。                           |
| 迁移工具   | 选择在源端环境中 <a href="#">部署的迁移工具Edge</a> 。                         |
| Glue凭证 | 选择在Edge添加的Glue凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-Glue。 |
| 区域     | 选择数据所在区域。                                                      |

| 参数         | 配置说明               |
|------------|--------------------|
| Catalog ID | 输入区域对应的Catalog ID。 |

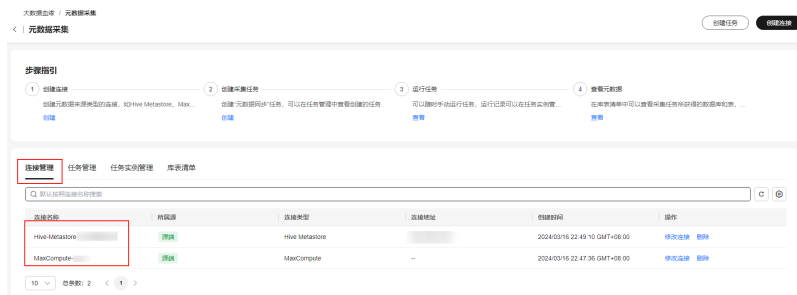
表 5-6 MaxCompute 连接参数配置说明

| 参数       | 配置说明                                                                         |
|----------|------------------------------------------------------------------------------|
| 所属源      | 仅需在源端创建元数据连接。                                                                |
| 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     | 选择在源端环境中部署的迁移工具Edge。                                                         |
| 阿里云凭证    | 选择源端阿里云账号的凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-MaxCompute”。        |
| 项目名称     | 选择需要采集的MaxCompute项目。                                                         |
| Endpoint | 输入MaxCompute服务所在地域的Endpoint信息。各地域Endpoint信息请参考 <a href="#">Endpoint对照表</a> 。 |

**步骤6** 配置完成后，单击“**测试连接**”按钮，系统会校验配置信息并与MetaStore/MaxCompute建立连接，测试通过代表连接成功。

**步骤7** 测试连接通过后，单击“**确认**”按钮，创建元数据连接完成。

**步骤8** 单击“**元数据连接**”，进入元数据采集页面。在**连接管理**页签，可以查看已创建的元数据连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。



----结束

## 5.2.2 创建并运行采集任务

利用元数据连接，创建元数据同步任务，将源端数据库的库、表、字段信息同步至迁移中心。

您还可以通过手动添加、导入模板等方式添加元数据库表信息，方法请参见[查看元数据](#)。

## 前提条件

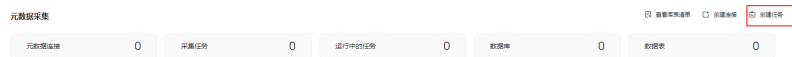
已创建元数据连接。

## 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“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迁移项目。

**步骤3** 在元数据采集区域，单击“创建任务”按钮，右侧弹出创建任务-元数据同步窗口。



**步骤4** 根据表5-7，配置元数据同步任务参数。

表 5-7 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 | 根据实际需求，选择已创建的元数据连接类型。                                                                                 |
| 数据库   | 填写需要进行元数据同步的数据库名称。<br><b>须知</b><br>元数据连接为MaxCompute类型时，该参数为必填参数，不能为空。其他元数据连接类型，该参数为可选参数，不填则代表同步全部数据库。 |
| 并发线程数 | 同时向大数据集群提交的SQL请求最大数量。默认为3，1≤取值范围≤10。                                                                  |

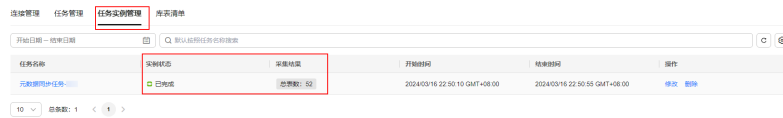
**步骤5** 参数配置完成后，单击“确认”按钮，创建元数据同步任务完成。

**步骤6** 单击“采集任务”，进入元数据采集页面。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“更多>修改”，可以修改任务参数配置。



**步骤7** 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8** 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9** 元数据采集完成后，在元数据列表单击血缘列的“去采集”，可以[创建血缘采集任务](#)。



----结束

### 5.2.3 查看元数据

您可以通过创建采集任务、手动添加、导入模板等方式获取元数据库表清单。

#### 通过采集任务获取元数据

创建采集任务的方法请参考[创建并运行采集任务](#)。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。

#### 手动添加数据表

- 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“[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[迁移项目](#)。
- 步骤3** 在元数据采集区域，单击“[查看库表清单](#)”，进入元数据采集页面的库表清单页签。
- 步骤4** 单击“[新增数据表](#)”按钮，弹出[新增数据表](#)窗口。
- 步骤5** 根据[表5-8](#)，配置参数。

表 5-8 新增数据表参数说明

| 参数        | 说明                                      |
|-----------|-----------------------------------------|
| 元数据连接     | 根据实际需求，选择 <a href="#">已创建的元数据连接</a> 类型。 |
| 数据库名      | 用户自定义。                                  |
| 数据表名      | 用户自定义。                                  |
| 分区信息（可选）  | 用户自定义。                                  |
| 是否分区表（可选） | 根据实际需求，选择该表是否为分区表。                      |
| 源表路径(可选)  | 输入数据表在数据库中的存放路径。                        |
| 业务负责人(可选) | 可以指定数据表所属业务的责任人。                        |
| 分析责任人（可选） | 可以指定数据表的分析责任人。                          |

| 参数     | 说明                            |
|--------|-------------------------------|
| 跨源表    | 根据实际需求设置。设为跨源表后不再显示该表一致性校验结果。 |
| 说明（可选） | 您可以输入用于数据表的说明。                |

**步骤6** 参数配置完成后，单击“**确认**”按钮，完成数据表新增。在库表清单列表中可以查看新增的数据表。

----结束

## 批量导入数据表

**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“**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**迁移项目**。

**步骤3** 在**元数据采集**区域，单击“**查看库表清单**”，进入**元数据采集**页面的**库表清单**页签。

**步骤4** 单击“**导入数据表**”按钮，弹出**导入数据表**窗口。

**步骤5** 根据实际需求，选择**已创建的元数据连接**类型。

**步骤6** 单击“**模板下载**”，将导入模板下载到本地。打开导入模板，将需要添加的数据表对应参数填写到模板内。

### 须知

- 每个模板最多支持导入10,000张表。
- 待导入的表必须属于同一个元数据来源。

**步骤7** 模板参数填写完成并保存后，单击导入窗口的“**添加文件**”按钮，将填写后的模板文件上传。

**步骤8** 上传成功后，单击“**确认**”按钮，系统开始导入，并且会在**任务实例管理**页签中的任务实例列表中，新增一条导入数据表的任务实例记录。您可以通过任务实例查看导入状态。

----结束

## 5.3 血缘采集

### 5.3.1 简介

通过血缘采集，可以建立起数据元素之间的关系和依赖关系，从而形成一个清晰的数据血缘图。

目前支持血缘采集的文件类型以及采集须知，参见下表。

| 文件类型  | 采集须知                                     |
|-------|------------------------------------------|
| SQL作业 | 需要提前导出SQL作业，并上传到华为云OBS桶，再将桶的读写权限委托给迁移中心。 |
| 血缘模板  | 需要下载血缘模板，并填写相应数据。                        |

## 5.3.2 导出作业

本节介绍导出华为云DataArts作业和阿里云DataWorks作业的方法。

### 导出华为云 DataArts 作业


**步骤1** 登录DataArts Studio控制台。

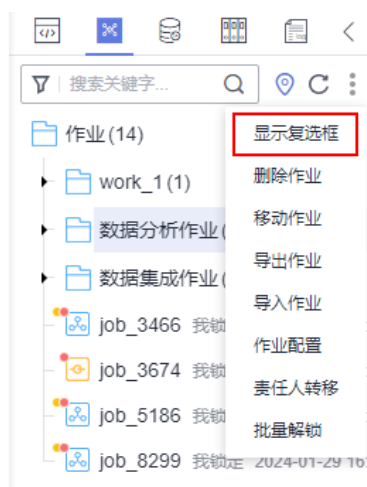
**步骤2** 选择实例，单击“进入控制台”，选择对应工作空间。

已开通实例 [如何进行准备工作?](#) [快速入门](#) [免费赋能课](#)



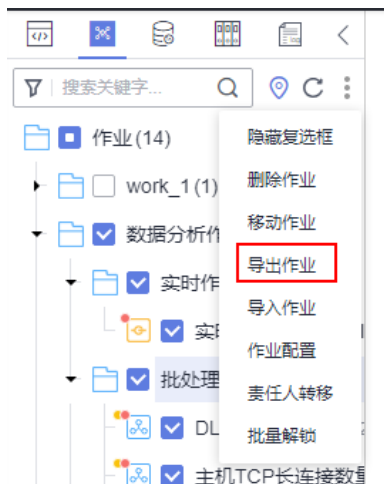
**步骤3** 单击“数据开发”，进入数据开发页面。

**步骤4** 单击作业目录中的 ，选择“显示复选框”。



**步骤5** 勾选需要导出的作业，单击  > 导出作业，弹出的导出作业窗口。





**步骤6** 导出范围选择“导出到OBS”，选择作业保存目录。

**步骤7** 选择需要导出的作业范围和状态，单击“确定”，导出完成后，前往步骤6.选择的OBS目录获取导出结果。

----结束

## 导出阿里云 DataWorks 作业

**步骤1** 登录阿里云DataWorks控制台，单击左侧导航栏的“数据建模与开发 > 数据开发”，在下拉框中选择对应工作空间后单击“进入数据开发”。

**步骤2** 单击左上方的图标，选择全部产品 > 更多 > 迁移助手，进入迁移助手首页。

**步骤3** 参考[创建和查看DataWorks导出任务](#)，创建全量导出任务。

----结束

## 5.3.3 创建血缘采集任务

### 前提条件

已完成[元数据采集](#)。

### 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“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[迁移项目](#)。

**步骤3** 在血缘采集区域，单击“创建任务”，弹出创建血缘采集任务窗口。

**步骤4** 选择文件类型。

- 选择SQL作业。
  - a. 导出需要进行血缘采集的SQL作业。
  - b. 将导出的SQL作业存放到华为云账号下的OBS桶。

- c. 在文件下拉列表中，选择存放在OBS桶的作业文件。
- **选择血缘模板。**
  - a. 单击“**模板下载**”，将血缘模板下载到本地。
  - b. 打开血缘模板，填写模板内的参数。其中，如下参数值为必填字段。
    - 目标数据库(TargetDataset)
    - 目标表(TargetTable)
    - 目标连接名称(TargetConnectionName)
    - 目标组件类型(TargetComponentType)
    - 上游数据库(SourceDataset)
    - 上游表(SourceTable)
    - 上游连接名称(SourceConnectionName)
    - 上游组件类型(SourceComponentType)
    - 作业ID(JobId)

#### 须知

- 模板中**目标组件类型**和**上游组件类型**支持的类型包括**HiveSQL**、**MaxCompute**。
  - 模板中单元格不允许存在公式，否则会导致解析失败。
- c. 模板参数填写完成并保存后，单击窗口的“**添加文件**”，将填写后的模板文件上传至迁移中心。

#### 注意

文件大小不能超过100 MB。

**步骤5** 单击“**确认**”，创建血缘采集任务完成。系统自动开始进行血缘采集。

**步骤6** 单击“**查看任务**”，进入**血缘采集**页面，可以在任务列表中查看采集任务状态。

**步骤7** 当任务状态为“**已完成**”时，单击页面右上角的“**查看血缘**”，可以查看**血缘关系图**。

----结束

## 5.3.4 查看血缘关系

### 前提条件

已**创建血缘采集任务**，且状态为“**已完成**”。

## 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“**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**迁移项目**。
- 步骤3** 在**血缘采集**区域，单击“**查看血缘图**”，进入**血缘图**页面。
- 步骤4** 通过**数据库名**、**表名**筛选需要查看的血缘图。如果进行了大数据校验，还可以通过校验批次进行筛选。

----结束

# 6 迁移方案设计

## 6.1 绑定目的端

仅支持主机类型资源绑定目的端。如果您已经在华为云创建了目的端主机，可以将采集的源端主机与目的端主机进行绑定，绑定后，迁移 workflows 将使用所绑定的目的端作为迁移的目的端。

绑定目的端的主机，在创建迁移 workflows 前，无需进行[评估推荐](#)。

### 前提条件

已完成[资源采集](#)。

### 注意事项

绑定的目的端主机有以下要求：

- 选择绑定的目的端必须关机，如果需要绑定的目的端处于开机状态，请关机后再绑定。
- 迁移 workflows 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 workflows 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。

### 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。
- 步骤3** 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4** 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5** 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。支持目的端磁盘扩容，即可以根据源端主机磁盘的实际使用量，选择小于源端磁盘容量和少于源端磁盘数量的目的端主机。

### 须知

- 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。

**步骤6** 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。并且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。



----结束

## 解除目的端绑定

**步骤1** 在资源配置明细的主机页签，单击已绑定目的端的主机资源操作列的“解除目的端绑定”，弹出解除已有目的端窗口。

**步骤2** 解绑操作无法恢复，确认解除绑定后，单击“是”，完成目的端解绑。

----结束

## 6.2 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。

当前支持规格评估的资源类型为：**主机、数据库、对象存储、容器、中间件**。

### 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 workflow。

## 前提条件

已完成**资源采集**，并将采集的资源关联到**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的**资源总数**、资源的**目的端规格配置情况**以及当前项目下所创建的**应用列表**。

**步骤3** 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4** 在**选择应用**的下拉列表中，选择需要进行资源规格评估的应用。

**步骤5** 在**选择资源区域**，勾选应用内，需要评估的资源。

**步骤6** 根据**表6-1**，配置评估策略。

**表 6-1** 评估策略参数说明

| 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     | <ul style="list-style-type: none"> <li>● 匹配源端规格<br/>根据源端规格推荐最合适的华为云对应资源规格，适合云上源端规格推荐。</li> <li>● 匹配业务场景<br/>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。</li> <li>● 跨可用区迁移<br/>仅适用于华为云弹性云服务可用区下线迁移，该策略只会为应用中的<b>主机</b>资源推荐规格。选择该策略同时需要选择<b>目的可用区</b>。</li> </ul>                                                                                                                                                                                                                                                               |
| 偏好设置     | <ul style="list-style-type: none"> <li>● 性能优先<br/>推荐的目的端规格要求性能最佳。</li> <li>● 价格优先<br/>推荐的目的端规格在满足要求的前提下尽量节约成本。</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                       |
| 高级选项（可选） | <p>主机资源，可以进行以下偏好设置，推荐目的端时优先推荐满足偏好设置要求的主机。</p> <ul style="list-style-type: none"> <li>● 虚拟机规格偏好<br/>选择期望的目的端主机规格类型，支持多选。</li> <li>● 系统盘规格偏好<br/>选择期望的目的端系统盘规格类型。</li> <li>● 数据盘规格偏好<br/>选择期望的目的端数据盘规格类型。</li> <li>● 虚拟机规格调整<br/>选择主机性能规格的推荐策略。 <ul style="list-style-type: none"> <li>- 选择“<b>匹配源端</b>”，会推荐与源端主机性能相同或最接近的CPU核数、内存大小。</li> <li>- 选择“<b>基于性能</b>”，需要先完成<b>源端主机性能采集</b>，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。</li> </ul> </li> </ul> <p><b>须知</b><br/>采集的性能数据越多，推荐的规格越准确，建议主机性能采集时长不少于7天。</p> |
|          | 容器资源可以设置“ <b>集群类型</b> ”、“ <b>集群版本</b> ”、“ <b>容器网络模型</b> ”等规格配置，推荐满足规格配置的集群资源。                                                                                                                                                                                                                                                                                                                                                                                                                                   |

**步骤7** 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8** （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

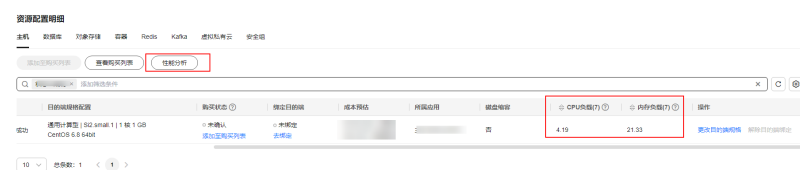
在**迁移方案设计**页面的**应用列表**，单击操作列的“**查看资源配置**”，进入目的端配置表页面。

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。

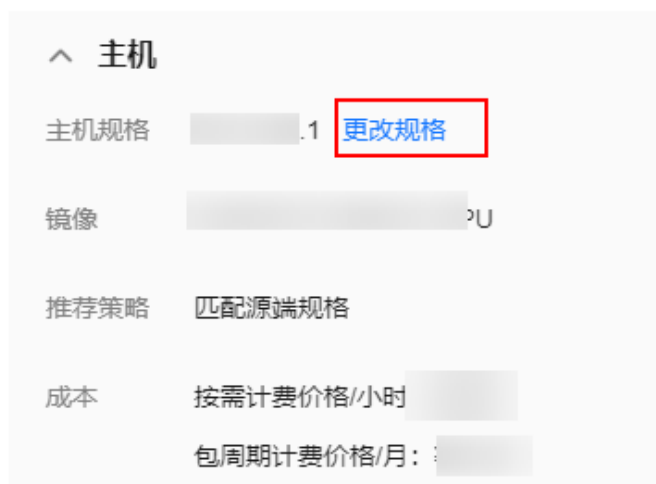


## 更改目的端规格

**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，进行缩容，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缩容，在主机资源配置明细列表中的“**磁盘缩容**”列，“是”代表已缩容；“否”代表未缩容。

### 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主机磁盘缩容，需确保缩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，即使配置缩容，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。

| ^ 磁盘 (1) |              |                                      |         |         |
|----------|--------------|--------------------------------------|---------|---------|
| 是否缩容 否   |              |                                      |         |         |
| 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<br><a href="#">更改规格</a> | ¥12.00  | ¥0.02   |

----结束

## 6.3 购买资源

目的端规格配置完成后，无需切换服务，可以直接在迁移中心批量购买目的端资源。当前仅支持购买主机资源。主机迁移 workflow 支持自动购买主机资源，通过“购买列表”批量购买资源适用于部分主机需要重新部署的场景。通过购买列表成功购买的主机，将自动绑定为迁移目的端，可以使用主机 workflow 继续迁移。



## 前提条件

已对源端资源完成[规格评估](#)。

## 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。
- 步骤3 单击目的端规格配置框中的“目的端配置表”按钮，进入目的端配置表页面。
- 步骤4 在资源配置明细区域的主机页签，单击已完成规格评估主机购买状态列的“添加至购买列表”，购买状态变为“待购买”。
- 步骤5 添加完成后，在页面上方的购买列表框中，单击“查看列表”按钮，进入购买列表页面。
- 步骤6 在操作流程区域，单击“查看我的模板”，右侧弹出“我的模板”窗口，单击“创建模板”按钮，进入创建配置模板窗口。
- 步骤7 在基础信息区域，选择模板类型（当前仅支持主机），自定义模板名称；在配置信息区域，根据[表6-2](#)，配置模板参数。

表 6-2 模板参数配置说明

| 参数    | 配置说明                                                                                                                                  |
|-------|---------------------------------------------------------------------------------------------------------------------------------------|
| 区域    | 选择目的端区域。                                                                                                                              |
| 项目    | 选择目的端区域项目。                                                                                                                            |
| 可用区   | 默认随机分配，也可手动选择。                                                                                                                        |
| 虚拟私有云 | 选择目的端区域创建的虚拟私有云（VPC）。                                                                                                                 |
| 子网    | 选择虚拟私有云所包含的子网，子网网段必须在所选VPC网段范围内。                                                                                                      |
| 安全组   | 所选安全组需要满足： <ul style="list-style-type: none"> <li>• Windows系统开放8899端口、8900端口和22端口。</li> <li>• Linux系统开放22端口。</li> </ul>               |
| 企业项目  | 将购买的主机加入商用生产企业项目。默认项目为default。<br>企业项目管理提供了一种按企业项目管理云资源的方式，帮助您实现以企业项目为基本单元的资源及人员的统一管理。关于创建和管理企业项目的详情，请参见《 <a href="#">企业管理用户指南</a> 》。 |

- 步骤8 参数配置完成后，单击“保存”按钮，在模板列表可以查看创建的模板。
- 步骤9 模板创建完成后，返回购买列表，单击主机购买模板列的“绑定模板”，弹出“绑定购买模板”窗口，在下拉列表中选择创建的模板，单击“确认”。
- 步骤10 为主机绑定购买模板后，单击操作列的“去购买”，当购买状态为“购买成功”时，可以看到购买资源的ID。

----结束

## 6.4 迁移方案配置

迁移中心为多种资源的不同迁移方案配置模板，可以面向迁移 workflow 进行一站式配置，并按应用显示配置进度。

目前提供的迁移方案模板包括：

**对象存储批量迁移：**适用于源端有多个对象存储桶资源需要迁移的场景。如果只有单个对象存储桶资源需要迁移，请直接[创建单个对象存储迁移 workflow](#)。

### 前提条件

- 已通过应用发现的公网采集，完成对象存储资源的发现与采集。
- （可选）已对待迁移的对象存储资源进行深度采集。

### 注意事项

- 创建方案时，待迁移桶列表中的资源所属的平台必须和所选的源端平台一致，且一个方案对应一个目的端区域。
- 一个源端桶资源可以关联多个方案。
- 在一个方案下，一个桶资源的目的端前缀只能有一个，但是源端前缀可以有多个。

### 操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。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。
- 步骤2** 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6-3](#)，配置相关参数。

表 6-3 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   | 选择源端资源所在的云平台。 |
| 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“确认”。您可以通过单击操作列的“修改”，调整桶的迁移方式，迁移方式说明参见[表6-4](#)。

 **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。

表 6-4 迁移方式说明

| 迁移方式 | 说明                                     |
|------|----------------------------------------|
| 全量迁移 | 迁移源端桶/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。列表文件需要存放在目的端桶中。 |
| 前缀迁移 | 通过对象名称或者名称前缀过滤迁移，只迁移与对象名称或名称前缀一致的对象。   |

**步骤7** 在待迁移桶列表中，单击操作列的“修改”，弹出修改迁移设置窗口。

**步骤8** 选择源端凭证，输入对象个数和容量（可选），单击“确认”，完成桶的迁移设置。

- 如果桶的迁移方式为“列表迁移”，需要在列表路径中输入列表文件的存放路径并以“/”结尾。
- 如果桶的迁移方式为“前缀迁移”，在“前缀路径”添加对象名称或者名称前缀，单击“确定”保存前缀。

 **说明**

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，则需要在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

**步骤9** 所有待迁移桶的迁移设置完成后，单击“下一步”，进行目的端配置。

**步骤10** 单击操作列的“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，直接绑定已有集群，也可以通过评估，获取集群规格推荐后，去创建集群。集群的创建方法请参考[创建集群](#)。

**步骤12** 单击“下一步”，进入集群配置，绑定已有迁移集群，本方案下所有资源将通过此集群迁移。

**步骤13** 单击“立即绑定”，右侧弹出[集群选择](#)窗口。

**步骤14** 在集群列表中，选择需要使用的集群，单击“确认”。如果需要对集群进行修改，请参考[集群管理](#)。

**步骤15** 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

# 7 迁移集群

## 7.1 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](#)。

### 操作步骤

- 步骤1** 登录[迁移中心](#)管理控制台。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** 单击页面右上角的“[创建集群](#)”按钮，进入[创建集群](#)页面。
- 步骤3** 根据[表7-1](#)，配置参数。

表 7-1 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。               | -              |
|      | 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    |

| 区域   | 参数        | 配置说明                                                                                                                                                                                                                                                                 | 限制说明                                                                                                                                                                                                                            |
|------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | 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"> <li>规格选定后，后续无法更改。</li> <li>设置的节点数量需要同时满足如下要求：                             <ul style="list-style-type: none"> <li>迁移节点数+列举节点数+1 ≤ 100</li> <li>迁移节点数+列举节点数+1 ≤ 子网可用IP数</li> </ul> </li> </ul> |
|      | 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"> <li>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）。</li> <li>专线：使用专线迁移，输入允许转发或直接请求到专线的IP地址，如：Nginx、网关等。</li> </ul>                                                                                                | -                                                                                                                                                                                                                               |
| -    | 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"> <li>不勾选，则不进行流程控制。</li> <li>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。</li> </ul> <p><b>须知</b><br/>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"> <li>最多可添加五条流量控制规则。</li> <li>时区时间以控制台所选Region的当地时间为准。</li> </ul>                                                                                                                              |

| 区域 | 参数      | 配置说明                                                                                                        | 限制说明 |
|----|---------|-------------------------------------------------------------------------------------------------------------|------|
|    | 启用LTS服务 | <ul style="list-style-type: none"><li>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。</li><li>不启用：不记录存储迁移过程中的日志。</li></ul> | -    |

**步骤4** 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5** 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

## 7.2 集群管理

在集群创建完成后，您可以根据实际需求随时调整集群内的流量控制规则和节点信息。

### 前提条件

已完成[集群创建](#)。

### 添加流量控制规则

**步骤1** 登录[迁移中心](#)管理控制台。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** 在集群列表中，单击操作列的“**管理**”，右侧弹出集群详细信息页面。

**步骤3** 在**流量控制**区域，单击“**添加**”，下方新增一条规则记录。

**步骤4** 在“**开始时间**”、“**结束时间**”和“**限速**”文本框中输入数值。

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。

#### 须知

- 时区时间以控制台所选Region的当地时间为准。
- 最多可以创建5条记录。

**步骤5** 单击“**确认**”，完成流量控制规则的添加。

----结束

## 新增节点

新增迁移节点和列举节点的操作步骤相同，以下步骤以新增迁移节点为例进行介绍。

- 步骤1** 登录[迁移中心](#)管理控制台。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** 在集群列表中，单击操作列的“**管理**”，右侧弹出集群详细信息页面。
- 步骤3** 在节点信息区域，单击“**新增节点>迁移节点**”，弹出**添加节点**页面。
- 步骤4** 输入需要增加的节点数量，节点数量需要同时满足如下要求：
  - 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
- 步骤5** 单击“**确认**”，节点新增完成，系统开始为迁移节点安装迁移插件。添加完成后，可以在节点列表中看到新添加的节点信息。

----结束

## 7.3 计费说明

在创建迁移集群时，会产生以下费用：

- 创建Master节点、迁移节点和列举节点会产生ECS服务费用。具体计费详情请参见[ECS按需计费说明](#)或[ECS价格计算器](#)。
- 如果您使用公网迁移时，会产生NAT网关服务费用。计费详情请参见[NAT网关计费说明](#)或[NAT价格计算器](#)。
- 启用LTS服务时，会产生LTS服务费用。具体计费详情请参见[LTS计费说明](#)或[LTS价格计算器](#)。

## 7.4 集群状态说明

集群所包含的状态以及状态说明参见[表7-2](#)。

表 7-2 集群状态说明

| 状态   | 说明                                                    |
|------|-------------------------------------------------------|
| 创建中  | 正在创建集群资源                                              |
| 创建失败 | 创建集群资源失败                                              |
| 连接中  | 等待master节点上线，并开始创建和安装其他节点                             |
| 健康   | 所有节点均处于在线状态                                           |
| 亚健康  | 部分迁移节点和列举节点在线，并且至少存在一个迁移节点和一个列举节点在线                   |
| 不可用  | 所有迁移节点或列举节点不在线                                        |
| 离线   | master节点离线，可能是网络中断、ECS删除等原因，请排查Vpcep Service、ECS等资源情况 |



| 状态   | 说明              |
|------|-----------------|
| 升级中  | 升级集群版本中         |
| 升级失败 | 升级集群版本失败        |
| 废弃中  | 30天内无任务运行，废弃集群中 |
| 废弃失败 | 废弃集群失败          |
| 删除中  | 删除集群中           |
| 删除失败 | 删除集群失败          |
| 已废弃  | 集群已废弃（vpcep删除）  |
| 等待创建 | 等待创建集群资源        |
| 安装中  | 安装master节点中     |
| 安装失败 | 安装master节点失败    |
| 等待安装 | 等待安装master节点    |
| 等待升级 | 等待升级集群版本        |
| 等待删除 | 等待删除集群          |

# 8 迁移 workflow

## 8.1 创建主机迁移 workflow

本节介绍如何使用主机迁移标准模板创建主机迁移 workflow。

### 前提条件

- 已完成[主机资源采集](#)。
- 已将待迁移的主机资源[关联应用](#)，并对应用内的主机进行[规格评估](#)。规格评估成功后，通过[购买资源](#)，批量购买目的端。如果待迁移应用中的主机已[绑定目的端](#)，则无需进行规格评估。

### 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[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[主机迁移标准模板](#)，单击“[查看步骤详情](#)”，可以查看主机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；步骤类型为“[手动](#)”，表示该步需要用户手动操作。单击右下角的“[开始配置](#)”按钮，进行 workflow 配置。

**步骤5** 根据[表8-1](#)，配置 workflow 参数。

表 8-1 主机迁移 workflow 配置参数说明

| 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 | 选择待迁移资源所关联的应用。 |

| 区域     | 参数      | 说明                                                                                                                                                                                                                                                                                  |
|--------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 迁移网络配置 | 网络类型    | 选择公网迁移，要求目的端服务器配置有“ <b>弹性公网IP (EIP)</b> ”，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。 <ul style="list-style-type: none"> <li>如果源端无法访问公网，需要输入源端代理服务器的私网IP和代理软件指定的代理端口。</li> <li>如果源端代理服务器无法访问公网，需要将SMS-Agent安装包放置到源端能访问下载或者借助代理能访问下载的地址。SMS-Agent安装包可以在SMS控制台下载。</li> </ul>                                               |
| 目的端配置  | 区域      | 目的端所在区域，即规格评估时所购买资源的区域。                                                                                                                                                                                                                                                             |
|        | 项目      | 目的端所在区域的项目。                                                                                                                                                                                                                                                                         |
|        | 虚拟私有云   | <ul style="list-style-type: none"> <li>若源端IP是192.168.X.X，则推荐创建的VPC网段是192.168.0.0/16，同时创建一个子网，网段也是192.168.0.0/16。</li> <li>若源端IP是172.16.X.X，则推荐创建的VPC网段是172.16.0.0/12，同时创建一个子网，网段也是172.16.0.0/12。</li> <li>若源端IP是10.X.X.X，则推荐创建的VPC网段是10.0.0.0/8，同时创建一个子网，网段也是10.0.0.0/8。</li> </ul> |
|        | 子网      | 子网网段与虚拟私有云网段相同。                                                                                                                                                                                                                                                                     |
|        | 安全组     | <ul style="list-style-type: none"> <li>Windows系统开放8899端口、8900端口和22端口。</li> <li>Linux系统当前仅支持文件级迁移，需要开放22端口。</li> </ul> <b>注意</b> <ul style="list-style-type: none"> <li>以上端口，建议只对源端服务器开发。</li> <li>防火墙开放端口与操作系统开放端口保持一致。</li> </ul>                                                  |
| 高级设置   | 迁移后是否开机 | <ul style="list-style-type: none"> <li>选择“否”，迁移完成后目的端服务器自动关机。</li> <li>选择“是”，迁移完成后目的端服务器保持开机状态。</li> </ul>                                                                                                                                                                          |
|        | 网络限速    | <ul style="list-style-type: none"> <li>选择“否”，代表不限流。</li> <li>选择“是”，根据要迁移的源端带宽大小及业务要求，设置带宽大小。</li> </ul>                                                                                                                                                                             |

| 区域 | 参数        | 说明                                                                                                                                                        |
|----|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | 自动安装rsync | <ul style="list-style-type: none"><li>选择“否”，不在源端安装rsync组件。</li><li>选择“是”，在源端自动安装rsync组件。</li></ul> <b>注意</b><br>Linux迁移依赖rsync组件，若源端未安装rsync组件，可能会导致迁移失败。 |
|    | 企业项目      | 选择需要迁移到的企业项目。默认为“default”。                                                                                                                                |

**步骤6** 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7** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动 workflow**窗口，此时 workflow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 workflow。
- 如果您需要在标准 workflow中**添加阶段/步骤**，则单击“**取消**”，此时迁移 workflow状态为“**等待中**”，并未开始迁移。确定 workflow步骤后，单击“**运行**”按钮，开始迁移。

**步骤8** 在迁移 workflow详情页面，可以查看该主机迁移 workflow的**配置信息**和**资源迁移进度**。

- 将鼠标放置主机迁移进度条，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，将鼠标放置进度条，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。

----**结束**

## 8.2 创建可用区迁移 workflow

本节介绍如何使用可用区迁移标准模板创建可用区迁移 workflow。

### 前提条件

- 已完成**主机资源采集**。
- 已将待迁移的主机资源**关联应用**，并已对应用内的主机进行**规格评估**。如果待迁移应用中的主机已**绑定目的端**，则无需进行评估推荐。

### 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**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**可用区迁移标准模板**，单击“**查看步骤详情**”，可以查看可用区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；步骤类型为“**手动**”，表示该步需要用户手动操作。单击“**开始配置**”按钮，进行 workflow 配置。
- 步骤5** 根据**表8-2**参数说明，配置 workflow 参数。

**表 8-2** 可用区 workflow 配置参数说明

| 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        | 选择待迁移资源所关联的应用。                                                                                                                                           |
| 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。当前仅支持华为云 <b>华南-广州</b> 区域。                                                                                                  |
|             | 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<br>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"> <li>选择“是”，生成目的端后，目的端保持关机状态。</li> <li>选择“否”，生成目的端后，目的端自动开机。</li> </ul>                                                 |
|             | 增量备份前源端是否关机 | <ul style="list-style-type: none"> <li>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。</li> <li>选择“否”，workflow 在执行增量备份时，依旧保持开机状态。</li> </ul> |
| 高级设置        | 自动删除镜像      | 选择开启时，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6** 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7** 确认配置信息无误后，单击“**确认创建工作流**”，弹出**启动 workflow**窗口，此时 workflow 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 workflow。
  - 如果您需要在标准 workflow 中**添加阶段/步骤**，则单击“**取消**”，此时迁移 workflow 状态为“**等待中**”，并未开始迁移。确定 workflow 步骤后，单击“**运行**”按钮，开始迁移。
- 步骤8** 在迁移 workflow 详情页面，可以查看该可用区 workflow 的**配置信息**和**资源迁移进度**。
- 将鼠标放置主机迁移进度条，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，将鼠标放置进度条，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。

----结束

## 8.3 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。

当前支持以下区域（Region）：

- 华北-北京四
- 华南-广州
- 华东-上海一
- 华南-广州-友好用户环境

### 前提条件

- 已提前创建**目的端对象桶**或**弹性文件系统**。
- 已提前**创建迁移集群**。

### 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**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“**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。单击“**开始配置**”按钮，进行 workflow 配置。
- 步骤5** 根据**表8-3**，配置基础信息参数。

表 8-3 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。 |

| 参数 | 说明                                                            |
|----|---------------------------------------------------------------|
| 描述 | 用户自定义。                                                        |
| 集群 | 选择用于迁移的集群，集群中需要包含迁移节点和执行节点。如果没有可用集群，可以 <a href="#">新建集群</a> 。 |

**步骤6** 根据表8-4、表8-5，配置源端、目的端信息。

表 8-4 源端信息参数说明

| 参数       | 说明                                                                                                                                                                                                                                                                                                                                          | 备注                |
|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| 数据站点类型   | 目前支持迁移的源端存储类型包括： <ul style="list-style-type: none"> <li>● 华为云 OBS</li> <li>● 阿里云 OSS</li> <li>● 百度云 BOS</li> <li>● 腾讯云 COS</li> <li>● 金山云 KS3</li> <li>● 七牛云 KODO</li> <li>● 优刻得 US3</li> <li>● 亚马逊 S3</li> <li>● 微软 Blob</li> <li>● NAS_SMB</li> <li>● NAS_NFS_V3_MOUNT</li> <li>● NAS_NFS_V3_PROTOCOL</li> <li>● HTTP/HTTPS数据源</li> </ul> | -                 |
| AK       | 输入源端云账号的访问密钥。                                                                                                                                                                                                                                                                                                                               | 数据站点类型选择云平台时填写。   |
| 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。                                                                                                                                                                                                                                                                                                                 |                   |
| 桶类型      | 根据桶的实际类型选择。桶类型可以通过 <a href="#">查看桶的基本信息</a> 获取。                                                                                                                                                                                                                                                                                             | 数据站点类型为华为云OBS时选择。 |
| appid    | 填写腾讯云APPID。<br><b>说明</b><br>您可以在腾讯云控制台账号信息页面查看APPID。                                                                                                                                                                                                                                                                                        | 数据站点类型为腾讯云COS时填写  |

| 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"> <li>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”。</li> <li>txt文件行数不超过100000行。</li> <li>单个列表文件大小不能超过300 MB。</li> <li>列表文件存放目录下的列表文件个数不能超过10000个。</li> <li>列表文件必须是UTF-8无BOM格式编码格式。</li> <li>列表文件中每行长度不要超过65535，否则会导致迁移失败。</li> <li>列表文件的元数据中不能设置“ContentEncoding”，否则会导致迁移失败。</li> <li>列表文件中每行使用制表符（键盘上Tab键）\t分割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。</li> <li>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。</li> </ul> | <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</p>                                                        |
| 路径       | <p>输入待迁移文件路径，格式为：/文件夹。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <p>NAS_SMB/<br/>NAS_NFS_V3_MOUNT/<br/>NAS_NFS_V3_PROTOCOL<br/>时填写。</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</p>                                                        |
| 密码       | <p>输入用户名对应的密码。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <p>NAS_SMB时填写。</p>                                                     |



| 参数         | 说明                                                             | 备注 |
|------------|----------------------------------------------------------------|----|
| Windows上的域 | 输入对应的域名。<br><b>说明</b><br>只需输入.com之前的内容，例如：域名为test.com，则输入test。 |    |

表 8-5 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 <b>华为云OBS</b> 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。<br><b>说明</b><br>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 <a href="#">设置“保存到指定前缀”参数</a> 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 <b>NAS_SMB/NAS_NFS_V3_MOUNT</b> 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 <b>NAS_SMB</b> 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。<br><b>说明</b><br>只需输入.com之前的内容，例如：域名为test.com，则输入test。                                                   |                                               |

**步骤7** 根据表8-6，进行迁移配置。

表 8-6 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。                                                                                                                                                               |
|        | 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。                                                                                                                                                        |
|        | 前缀迁移 | 仅支持云平台数据类型。<br>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。<br><b>说明</b><br>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                    |
| 并发子任务数 | -    | 用户自定义。配置的数目不能超过可用迁移执行节点数 x20。例如：可用的执行节点数为2，则最大子任务数配置数目需要≤40。                                                                                                                   |
| 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。                                                                                                                                                |
|        | 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 | 对于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。                                                                                             |
| 迁移元数据  | -    | 选择是否迁移元数据。<br><ul style="list-style-type: none"> <li>勾选，则迁移源端对象元数据。</li> <li>不勾选，仅迁移ContentType元数据。</li> </ul>                                                                 |
| 清理集群   | -    | 选择 workflow 是否创建清理集群步骤。<br><ul style="list-style-type: none"> <li>勾选，迁移 workflow 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。</li> <li>不勾选，迁移 workflow 不创建清理集群步骤。</li> </ul> |

步骤8 （可选）根据表8-7，配置高级选项。

表 8-7 高级选项参数说明

| 参数        | 说明                                                                                                                                                                                                                                                                                                                                   | 备注                          |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| KMS加密     | <ul style="list-style-type: none"> <li>不勾选，迁移后对象状态与源端保持一致。</li> <li>勾选，所有对象在迁移到OBS桶后都将被加密存储。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>使用KMS服务端加密可能会降低迁移速率10%左右。</li> <li>目的端桶所在区域支持KMS加密时，此参数才可用。</li> </ul>                                                                                               | 仅目的端选择 <b>华为云OBS</b> 时支持配置。 |
| 解冻归档数据    | <p>归档类型的对象存储要实现迁移，必须预先解冻。当源端云服务商支持自动解冻归档类型数据时，可以勾选此选项。目前支持自动解冻归档类型数据的云厂商有：华为云、阿里云、金山云、腾讯云。</p> <ul style="list-style-type: none"> <li>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。</li> <li>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。</li> </ul> <p><b>说明</b></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 <a href="#">过滤策略说明</a> 。                                                                                                                                                                                                                                                                              |                             |
| 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"> <li>不勾选，不接收SMN消息。</li> <li>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。</li> </ul>                                                                                                                                                       |                             |

| 参数   | 说明                                                                                                                                                                                                                                                                                                                                                                                           | 备注 |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 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"> <li>不勾选，不进行流量控制。</li> <li>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。<br/>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。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>限速的数值范围：0~1048576 MB/s。</li> <li>最多可以创建5条记录。</li> <li>时区时间以目的端Region的当地时间为准。</li> </ul> | -  |
| 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"> <li>不勾选，手动启动/暂停迁移任务。</li> <li>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。<br/>例如： <ul style="list-style-type: none"> <li>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。</li> <li>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。</li> </ul> </li> </ul>                                      | -  |

**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 workflow**窗口，此时 workflow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 workflow。
- 如果您需要在标准 workflow中**添加阶段/步骤**，则单击“**取消**”，此时迁移 workflow状态为“**等待中**”，并未开始迁移。确定 workflow步骤后，单击“**运行**”按钮，开始迁移。

**步骤11** 在 workflow详情页面，可以查看该存储迁移 workflow的**配置信息**和**资源迁移进度**。还可以进行如下操作：

- 将鼠标放置资源迁移进度条，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，您需要将鼠标放置在进度条上，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                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 修改并发子任务数           | <ol style="list-style-type: none"> <li>1. 在<b>进度详情</b>区域, 单击期望并发子任务数下方的“<b>修改</b>”, 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x20。例如, 可用的迁移节点数为2, 则配置数目需要≤40。</li> <li>2. 单击“<b>确认</b>”, 完成并发子任务数修改。</li> </ol>                                                                                                                                                                                                                                   |
| 添加迁移限速规则           | <ol style="list-style-type: none"> <li>1. 在<b>迁移速度</b>区域, 单击“<b>添加</b>”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。<br/><b>须知</b> <ul style="list-style-type: none"> <li>- 限速取值范围: 1 MB~1024 GB。</li> <li>- 创建多条限速规则时, 限速时段不能重叠。<br/>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。</li> <li>- 开始时间不能晚于等于结束时间。<br/>例如: 不能设置时间段为23:00~01:00。</li> </ul> </li> <li>2. 设置完成后, 单击“<b>保存</b>”, 完成添加。</li> </ol> |
| 获取迁移失败/跳过/成功资源列表文件 | 在 <b>文件统计</b> 区域, 可以查看存放迁移 <b>失败/跳过/成功</b> 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。                                                                                                                                                                                                                                                                                                                                                       |
| 查看迁移流量统计           | 在 <b>流程统计</b> 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。                                                                                                                                                                                                                                                                                                                                                                                    |

---结束

## 新建集群

**步骤1** 在集群的下拉列表中, 单击“**新建集群**”。

**步骤2** 根据**表8-8**, 配置集群参数。

**表 8-8** 集群参数说明

| 区域   | 参数    | 说明                           |
|------|-------|------------------------------|
| 基础配置 | 集群名称  | 用户自定义。                       |
|      | 虚拟私有云 | 在下拉列表中选择虚拟私有云。               |
|      | 子网    | 子网的可用IP数需要大于等于迁移节点数+列举节点数+1。 |
| 迁移节点 | ECS规格 | 推荐规格8U16G。                   |

| 区域   | 参数      | 说明                                                                                                                      |
|------|---------|-------------------------------------------------------------------------------------------------------------------------|
|      | 数量      | 设置的节点数量需要同时满足如下要求： <ul style="list-style-type: none"> <li>迁移节点数+列举节点数+1 ≤100</li> <li>迁移节点数+列举节点数+1 ≤子网可用IP数</li> </ul> |
| 列举节点 | ECS规格   | 推荐规格8U16G。                                                                                                              |
|      | 数量      | 设置的节点数量需要同时满足如下要求： <ul style="list-style-type: none"> <li>迁移节点数+列举节点数+1 ≤100</li> <li>迁移节点数+列举节点数+1 ≤子网可用IP数</li> </ul> |
| 网络模式 | 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）。                                                            |
|      | 专线      | 使用专线迁移，输入允许转发或直接请求到专线的IP地址，如：Nginx、网关等。                                                                                 |
|      | 启用LTS服务 | <ul style="list-style-type: none"> <li>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。</li> <li>不启用：不记录存储迁移过程中的日志。</li> </ul>          |

----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8-9](#)。

表 8-9 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"> <li>精确匹配，填写文件绝对路径全路径精确匹配，特殊字符用反斜杠 (\) 转义</li> <li>模糊匹配                             <ul style="list-style-type: none"> <li>*匹配任何字符0个或者多个，但不匹配斜杠 (/)</li> <li>使用**匹配任何字符0个或者多个，包括斜杠 (/)</li> <li>?匹配除斜杠 (/) 之外的任何字符，只能匹配一个</li> <li>{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。</li> <li>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</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>除{和}以外，不能出现连续的规则符号，例如：***、*?、**?、?*、?*、*{、}*}*、*}?、?*{、{*}、{ }、{*、}*}、*。</li> <li>{和}之间的模式，只能使用*做通配符。</li> <li>{和}之间的模式，不能有{}嵌套。</li> <li>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。</li> <li>需要配置多个匹配规则时，用英文分号隔开。</li> </ul> |
| 白名单 | <ul style="list-style-type: none"> <li>白名单为空时，迁移所有文件。</li> <li>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。</li> </ul> |                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                            |
| 时间段 | <p>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。</p> <p>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。</p>                         |                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                            |

## 8.4 创建对象存储批量迁移 workflow

本节介绍如何使用对象存储批量迁移模板，批量创建对象存储数据迁移任务，实现对对象存储数据的高效迁移。

### 前提条件

已完成[迁移方案配置](#)。

### 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[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[对象存储批量迁移模板](#)，单击“[查看步骤详情](#)”，可以查看该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。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[创建的对象存储迁移方案](#)。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 workflow 生效，不会修改迁移方案中配置的集群。
- 步骤8** 根据[表8-10](#)，进行迁移配置。

表 8-10 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执行节点数 x20。例如：可用的执行节点数为2，则最大子任务数配置数目需要≤40。 |
| 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。                              |
|        | 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 | 对于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。                                                                                             |
| 迁移元数据 | -    | 选择是否迁移元数据。<br><ul style="list-style-type: none"> <li>勾选，则迁移源端对象元数据。</li> <li>不勾选，仅迁移ContentType元数据。</li> </ul>                                                                 |
| 清理集群  | -    | 选择 workflow 是否创建清理集群步骤。<br><ul style="list-style-type: none"> <li>勾选，迁移 workflow 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。</li> <li>不勾选，迁移 workflow 不创建清理集群步骤。</li> </ul> |

步骤9 （可选）根据表8-11，配置高级选项。

表 8-11 高级选项参数说明

| 参数     | 说明                                                                                                                                                                                                                                                                                                                                   |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| KMS加密  | <ul style="list-style-type: none"> <li>不勾选，迁移后对象状态与源端保持一致。</li> <li>勾选，所有对象在迁移到OBS桶后都将被加密存储。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>使用KMS服务端加密可能会降低迁移速率10%左右。</li> <li>目的端桶所在区域支持KMS加密时，此参数才可用。</li> </ul>                                                                                               |
| 解冻归档数据 | <p>归档类型的对象存储要实现迁移，必须预先解冻。当源端云服务商支持自动解冻归档类型数据时，可以勾选此选项。目前支持自动解冻归档类型数据的云厂商有：华为云、阿里云、金山云、腾讯云。</p> <ul style="list-style-type: none"> <li>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。</li> <li>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。</li> </ul> <p><b>说明</b></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 <a href="#">过滤策略说明</a> 。                                                                                                                                                                                                                                                                              |

| 参数        | 说明                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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"> <li>不勾选，不接收SMN消息。</li> <li>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。</li> </ul>                                                                                                                                                                                                                  |
| 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"> <li>不勾选，不进行流量控制。</li> <li>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。</li> </ul> <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><b>说明</b></p> <ul style="list-style-type: none"> <li>限速的数值范围：0~1048576 MB/s。</li> <li>最多可以创建5条记录。</li> <li>时区时间以目的端Region的当地时间为准。</li> </ul> |

**步骤10** 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。

**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：

- 将鼠标放置资源迁移进度条，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，您需要将鼠标放置在进度条上，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

## 8.5 添加阶段/步骤

您可以在 workflow 标准模板的迁移阶段/步骤，添加卡点，对迁移阶段/步骤进行控制。

### ⚠ 注意

- 只有迁移阶段/步骤状态为**等待中/暂停**，才可以选择该阶段/步骤进行添加。迁移阶段/步骤处于**运行中/完成**状态时，无法添加。
- 如果上一阶段/步骤处于**运行中/暂停/完成**状态，该阶段/步骤只能选择**添加后置阶段/步骤**。
- **可重复**类型的步骤，无法选择**添加后置步骤**。

### 添加阶段

- 步骤1** 在**迁移 workflow 详情**页面，将鼠标放置迁移阶段名称上，在弹出的窗口中，选择**添加前置阶段/添加后置阶段**，弹出**添加阶段**窗口。
- 步骤2** 为迁移阶段自定义**名称**和**说明**，单击“**添加步骤**”，选择步骤类型，为步骤自定义名称和说明，单击“**确认**”按钮。支持添加多个步骤。
- 步骤3** 迁移阶段中的步骤添加完成后，单击“**确认**”按钮，完成迁移阶段的添加。

### 须知

只有自行添加的阶段，可以进行修改、删除操作，模板中的迁移阶段无法进行修改和删除。

----结束

### 添加步骤

- 步骤1** 在**迁移 workflow 详情**页面，将鼠标放置步骤名称上，选择**添加前置步骤/添加后置步骤**，弹出**添加步骤**窗口。
- 步骤2** 根据**表8-12**，选择步骤类型，为步骤自定义名称和说明，单击“**确认**”按钮，添加步骤完成。

表 8-12 步骤类型说明

| 步骤类型 | 说明              |
|------|-----------------|
| 卡点   | 迁移到该步骤时，需要手动确认。 |

**步骤3** 返回迁移阶段，查看添加的步骤。

---

**须知**

只有自行添加的步骤，可以进行修改、删除操作，模板中的步骤无法进行修改和删除。

---

---结束

# 9 大数据作业迁移

## 9.1 复杂度分析

### 9.1.1 基础配置

在进行大数据作业迁移复杂度分析前，需要提供迁移所需的OBS桶和源端基本信息。

#### 准备工作

- 准备一个OBS桶作为大数据作业迁移的数据存储和处理中心，并且华为云账号需要有该桶的读写权限。建议创建一个空的OBS桶，创建方法请参考[创建桶](#)。
- 在凭证管理添加源端云平台凭证，方法请参见[添加凭证](#)。

#### 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](#)>[大数据作业迁移](#)”，进入[大数据作业迁移](#)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[复杂度分析](#)页签，单击“[配置](#)”，弹出[基础配置](#)窗口。

**步骤4** 根据[表9-1](#)，进行参数配置。

表 9-1 基础配置说明

| 参数          | 配置说明                                                        |
|-------------|-------------------------------------------------------------|
| 桶名          | 选择提前准备的用于大数据作业迁移的OBS桶。一旦作业迁移开始，更换桶可能会导致数据不一致和迁移失败，因此需要谨慎选择。 |
| 源端调度平台      | 选择源端作业平台。                                                   |
| 阿里云凭证       | 选择在凭证管理中添加的源端阿里云账号凭证。                                       |
| DataWorks区域 | 选择需要进行分析的大数据作业所在区域。                                         |

**步骤5** 单击“保存”按钮，完成基础信息配置。

----结束

## 9.1.2 配置增量感知（可选）

增量感知功能适用于作业频繁更新的迁移场景，能够跟踪和识别作业的变化，从而实现仅对发生变化的部分进行迁移。

### 使用须知

- 当前仅阿里云DataWorks支持增量感知配置功能，且要求阿里云账号必须开通openevent企业版。
- 输出文件存放路径：{obs\_bucket\_name}/BigData/output/job\_migrate/increment\_event 下的result和logs文件夹。

### 前提条件

已完成[基础配置](#)。

### 第一步：开通阿里云 EventBridge 并创建自定义事件总线

**步骤1** 登录阿里云事件总线EventBridge控制台，在顶部菜单栏，选择区域。区域要与[基础配置](#)所选的DataWorks区域一致。

**步骤2** 在左侧导航栏，选择“事件总线”，进入事件总线页面。

**步骤3** 在事件总线页面的自定义事件总线区域，单击“快速创建”按钮，右侧弹出创建自定义事件总线窗口。

**步骤4** 在总线配置向导，输入自定义事件总线名称（必填）和描述，单击“下一步”。

**步骤5** 事件源和规则无需配置，均选择“跳过”。

**步骤6** 在弹出的直接创建自定义事件总线对话框中单击“确定”，暂不在新创建的自定义事件总线上创建事件规则（包含目标）。

**步骤7** 自定义总线创建完成后，在事件总线页面，找到创建的自定义总线，单击操作列的“详情”。进入总线概览页面，查看自定义总线的基础信息和接入点信息。

----结束

### 第二步：在阿里云 DataWorks 添加事件分发通道

**步骤1** 登录阿里云DataWorks控制台，在左侧导航栏选择“开放平台”，进入开发者后台页面。

**步骤2** 在页面左上角选择区域。区域要与[基础配置](#)所选的DataWorks区域一致。

**步骤3** 单击左侧导航栏的OpenEvent，进入OpenEvent页面。

**步骤4** 单击“添加事件分发通道”，弹出添加事件分发通道对话框，配置要分发事件的工作空间和指定分发到EventBridge中自定义总线。

### 须知

可以给一条事件总线同时配置多个工作空间，同时监听多个工作空间的增量变化。

**步骤5** 单击“**确定**”，完成添加事件分发通道。

**步骤6** 在事件分发通道列表，单击操作列的“**启用**”，启用该事件分发通道后，该工作空间的事件消息将会推送至EventBridge中对应的事件总线。

----结束

## 第三步：在华为云配置函数 workflow

在配置函数 workflow 前，需要先联系MgC大数据迁移专家，获取增量感知jar包。

**步骤1** 联系MgC大数据迁移专家，获取DataWorks增量感知jar包，并将获取的jar包上传至为大数据迁移所准备的OBS桶。

**步骤2** 登录华为云[函数 workflow 控制台](#)，在左侧导航栏选择“**函数>函数列表**”，进入**函数**页面。

**步骤3** 单击页面右上角的“**创建函数**”，进入**创建函数**页面。

**步骤4** 选择创建空白函数，根据**表9-2**，设置基本信息。

表 9-2 基本信息参数配置说明

| 参数   | 配置说明                |
|------|---------------------|
| 函数类型 | 选择 <b>事件函数</b> 。    |
| 区域   | 根据实际需求选择。           |
| 项目   |                     |
| 函数名称 | 用户自定义。              |
| 企业项目 | 根据实际需求选择。           |
| 运行时  | 选择 <b>Java 11</b> 。 |

**步骤5** 基本信息配置完成后，单击“**创建函数**”，完成创建。

**步骤6** 在函数列表，单击创建的函数名称，进入函数详情页面。

**步骤7** 上传jar包。

1. 在**代码**页签，单击“**上传自>OBS地址**”，弹出从OBS上传窗口。
2. 输入存放jar包的OBS桶地址，单击“**确定**”，将jar包上传。

**步骤8** 设置函数执行入口。

1. 在**设置**页签，左侧导航栏选择**常规设置**，**函数执行入口**参数输入：**com.huawei.functiongraph.IncrementalAware.incrementalAwar**，其余参数根据实际需求设置。
2. 单击“**保存**”，完成函数执行入口等常规设置。

**步骤9** 创建时间触发器。

1. 在左侧导航栏选择**触发器**，单击“**创建触发器**”，右侧弹出**创建触发器**窗口。
2. 触发器类型选择“**定时触发器 (TIMER)**”，自定义定时器名称，根据实际需求设置触发规则。

 **说明**

定时触发器功能是为了定时进行增量感知，如果您需要立即触发一次增量感知，可以单击函数 workflow 界面的“**测试**”按钮。

3. 单击“**确定**”，完成触发器创建。

**步骤10** 配置环境变量。

1. 在左侧导航栏选择**环境变量**，请严格按照**表9-3**，配置环境变量。

 **注意**

- 输入华为云账号和阿里云账号的AK、SK需要长期有效，不能使用临时AK、SK。
- 阿里云AK、SK需要有查询EventBridge的权限，华为云AK、SK需要有读写OBS桶的权限。

**表 9-3** 环境变量参数配置

| 参数                    | 参数设置                                                                                                                                                                                                                           |
|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ali_ak                | 阿里云账号的AK。                                                                                                                                                                                                                      |
| ali_sk                | 阿里云账号的SK。                                                                                                                                                                                                                      |
| ali_region_id         | <a href="#">创建自定义事件总线</a> 所选区域的regionId。                                                                                                                                                                                       |
| event_bridge_bus_name | 创建的 <a href="#">阿里云EventBridge自定义事件总线</a> 名称。                                                                                                                                                                                  |
| hw_ak                 | 华为云账号的AK。                                                                                                                                                                                                                      |
| hw_sk                 | 华为云账号的SK。                                                                                                                                                                                                                      |
| obs_bucket_name       | 为大数据迁移所准备的华为云OBS桶名。                                                                                                                                                                                                            |
| obs_endpoint          | 华为云OBS桶所在区域的域名。                                                                                                                                                                                                                |
| time_interval         | <p>查询的时间间隔，建议设置为60分钟，取值范围1~1439内的正整数，单位为分钟。</p> <p><b>注意</b></p> <ul style="list-style-type: none"> <li>- time_interval参数值需要和时间触发器间隔保持一致。</li> <li>- 如果需要改小查询的时间间隔，请确定源端目前没有进行中的增量作业，然后单击“测试”进行一次立即增量补齐后再修改，防止数据遗漏。</li> </ul> |

2. 单击“**保存**”，保存环境变量配置。

----**结束**



### 9.1.3 导入作业包

通过将作业包导入迁移中心，可以对作业包进行全面的分析，从而为迁移计划提供详细的信息和数据支持。支持一次导入多个作业包，一次性评估多个工作空间内的作业信息。

#### 前提条件

- 已完成[基础配置](#)。
- 已从源端导出作业包，并放入准备好的OBS桶的BigData/input/job\_migrate 目录下。

#### 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>大数据作业迁移](#)”，进入[大数据作业迁移](#)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[复杂度分析](#)页签，单击[工作空间列表](#)下方的“[导入作业包](#)”，弹出[导入作业包](#)窗口。


**步骤4** 在OBS桶的BigData/input/job\_migrate 目录下，选择需要进行分析的作业包。支持同时选择多个作业包。

如未提前将作业包放入OBS桶的BigData/input/job\_migrate 目录下，单击“[打开OBS文件夹](#)”，进入job\_migrate 目录将待分析的作业包上传后再重新选择。

**步骤5** 单击“[保存](#)”，导入作业包任务创建完成。系统开始对作业包进行复杂度分析。

**步骤6** 单击“[查看任务](#)”，在分析任务列表中，可以查看任务状态。单击任务名称，可以查看任务详情。

**步骤7** 当任务状态为“[已完成](#)”时，返回[工作空间列表](#)，该列表展示了从作业包中分析出来的源端工作空间，以及每个工作空间包含的作业总数和具体的作业类型。在工作空间列表，您可以进行如下操作：

- 对需要进行迁移的工作空间进行[迁移配置](#)。
- 修改别名，单击**别名**列的  图标，修改自定义别名。
- 选择工作空间是否迁移，在**需要迁移**列，可以设置该空间是否在后续迁移任务中进行迁移。
- 查看工作空间包含的作业类型，单击**作业类型**列的数量，右侧弹出作业类型列表，其中，带有**自动**标识的类型支持自动迁移。

---

#### 须知

阿里云源端PYODPS2 类型的作业，在采集分析后页面上显示的作业类型为PYODPS，属于正常现象。

---

----结束

## 9.1.4 迁移配置

在进行大数据作业迁移前，需要对源端待迁移的工作空间进行一系列的配置工作以确保迁移过程的顺利进行和数据的完整性。

### 前提条件

已[导入作业包](#)且状态为已完成。

### 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>大数据作业迁移](#)”，进入[大数据作业迁移](#)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[复杂度分析](#)页签的[工作空间列表](#)，单击[迁移配置](#)列的“[配置](#)”，右侧弹出[迁移配置](#)窗口。
- 步骤4** 在[基础信息](#)区域，选择“目的端区域”、“DataArts实例”、“DataArts工作空间”以及“云数据迁移 CDM 集群”等参数。
- 步骤5** 在[作业类型](#)区域，根据源端空间中所包含的作业类型，选择待处理的作业类型，并配置相应参数。
- 步骤6** 单击“[保存](#)”，完成迁移配置。迁移配置完成后，即可前往“[迁移实施](#)”页面进行迁移。

----结束

## 9.2 迁移实施

### 9.2.1 自动迁移

在大数据作业迁移的过程中，部分作业类型可以实现自动迁移，同时在迁移后，这些作业的类型会根据华为云的架构和能力进行相应的转化。以下是MgC支持自动迁移的作业类型以及迁移后转化的作业类型：

表 9-4 支持自动迁移的作业类型以及迁移后转化的类型

| DataWorks作业类型 | 转化后类型                                                                               |
|---------------|-------------------------------------------------------------------------------------|
| ODPS_SQL      | <ul style="list-style-type: none"><li>MRS Spark SQL</li><li>DLI Spark SQL</li></ul> |
| EMR_HIVE      | MRS Hive SQL                                                                        |
| EMR_PRESTO    | MRS Presto SQL                                                                      |
| EMR_SPARK_SQL | MRS Spark SQL                                                                       |
| DIDE_SHELL    | Shell脚本                                                                             |
| EMR_SHELL     | Shell脚本                                                                             |

| DataWorks作业类型        | 转化后类型    |
|----------------------|----------|
| AnalyticDB for MySQL | DWS SQL  |
| PYODPS               | Python脚本 |
| PYODPS3              | Python脚本 |
| DI                   | -        |

## 前提条件

已完成[迁移配置](#)。

## 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>大数据作业迁移](#)”，进入大数据作业迁移页面。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** 选择[迁移实施](#)页签，在工作列表区域的“待处理作业”列表，单击“迁移作业”按钮，系统将自动识别并迁移列表中符合自动迁移条件的作业。



- 对于不需要迁移的作业，可以单击操作列的“[设为已处理](#)”，将作业从待处理作业列表中移除，并加入到已处理列表中，系统不会迁移已处理列表中的作业。
- 不支持自动迁移的作业，需要您[手动迁移](#)。
- 对于[迁移失败](#)的作业，将鼠标放在作业列表中显示的“[迁移失败](#)”状态上，可以查看失败原因。处理完失败原因后，单击迁移状态列的“[重新迁移](#)”，系统会立即执行失败作业的再次迁移。

----结束

## 9.2.2 手动迁移

对于不支持自动迁移的作业类型，需要您手动执行迁移。

## 前提条件

已完成[迁移配置](#)。

## 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>大数据作业迁移](#)”，进入大数据作业迁移页面。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** 选择**迁移实施**页签，在**工作列表**区域的“**待处理作业**”列表，单击迁移状态列的“**迁移**”，手动执行该作业的迁移任务。

对于**迁移失败**的作业，将鼠标放在作业列表中显示的“**迁移失败**”状态上，可以查看失败原因。处理完失败原因后，单击迁移状态列的“**重新迁移**”，系统会立即执行失败作业的再次迁移

----结束

## 9.2.3 增量迁移

首次全量迁移完成后，如果源端数据发生变化，可以通过增量迁移将源端数据发生变化的作业重新迁移。

### 前提条件

已通过自动或手动迁移方式，完成首次迁移。

### 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>大数据作业迁移**”，进入**大数据作业迁移**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** 选择**迁移实施**页签，在**工作列表**区域的“**待处理作业**”列表，单击“**获取增量作业**”按钮，系统将自动识别上次迁移后所有发生变化的作业，这些作业将更新至“**待处理作业**”列表中。

单击操作列的“**对比**”，可以查看增量版本与基线版本间的差异。基线版本即上一次迁移版本，每成功迁移一次，都将产生新的基线版本。

**步骤4** 获取增量作业后，再次通过**自动迁移**或**手动迁移**方式，将发生变化的作业迁移。

----结束

# 10 业务验证

## 10.1 大数据运维前风险检查（MRR）

为了保障大数据产品性能和数据的可靠性，迁移中心提供了由运维实践总结而来的大数据检查项列表，包含[手动检查项](#)和[自动检查项](#)。您可以根据检查项，完成大数据转运维前的风险评估。

- 手动检查项当前支持：MapReduce服务(MRS)、数据仓库服务 GaussDB(DWS)、云搜索服务(CSS)、数据治理中心(DataArts Studio)、数据湖探索(DLI)。
- 自动检查项当前支持：MapReduce服务(MRS)。

### 手动检查项

**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“[业务验证](#)”，进入[业务验证](#)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[手动检查项](#)区域，单击“[初始化](#)”按钮，进入[大数据风险检查手动检查项](#)页面。页面上方会实时统计已完成检查的检查项数量、不同风险的检查结果数量以及不同风险等级的检查项数量。

您可以在检查项列表中，查看每项所属云服务、说明、风险等级、风险影响等信息。

**步骤4** 在检查项列表，单击操作列的“[查看详情](#)”，弹出检查项详情，根据提供的[检查方式](#)、[应对措施](#)以及[判断标准](#)，对该检查项进行风险评估。

**步骤5** 单击检查项操作列的“[编辑](#)”，弹出[编辑](#)窗口，根据风险评估结果选择检查结果，并输入检查结果详情和备注，单击“[确认](#)”。此时，检查结果列会显示风险评估结果。检查完成后，单击页面右上角的“[导出](#)”按钮，可以将检查结果以Excel文件格式下载到本地。

----[结束](#)

### 自动检查项

**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“**业务验证**”，进入**业务验证**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**手动检查项**区域，单击“**初始化**”按钮，进入**大数据风险检查手动检查项**页面。

**步骤4** 返回**业务验证**页面，在**自动检查项**区域，单击“**检查详情**”，进入**大数据风险检查自动检查项**页面。

在页面的**检查项**页签，可以看到支持的风险检查项列表，当前仅支持对MRS集群进行风险检查。在检查项列表，单击操作列的“**查看详情**”，弹出检查项详情，可以查看该检查项的**应对措施**、**检查方式**详情以及**判断标准**等信息。

**步骤5** 单击“**导入待检查对象**”按钮，弹出**导入待检测对象**窗口。

**步骤6** 分别选择待检查对象所在**区域**、**项目**以及所属**集群**，单击“**导入**”。导入成功后，选择“**检查对象**”页签，可以查看导入的集群列表。

**步骤7** 在集群列表，单击操作列的“**检查**”，开始对该集群进行风险项检查。如果有多个集群，可以同时勾选后，单击“**批量检查**”，对所有集群进行风险项检查。

单击操作列的“**查看详情**”，弹出集群详情，可以查看该集群的风险项检查结果。

----**结束**

# 11 大数据校验

## 11.1 简介

迁移中心 MgC为多种大数据计算/存储引擎，如Hive、Hbase、Doris和MaxCompute等，提供了一致性校验方法。通过一致性校验，您可以更加放心地进行大数据迁移操作，确保数据的准确性和可靠性，从而避免了数据丢失或不一致的风险。

### 校验方式说明

- 全量校验：全量对比所有数据，适用与需要对比历史数据的场景。
- 日级增量校验：依靠元数据中的创建或更新时间，获取数据变化（增量）分区，再进行校验。支持单日或连续多日增量数据的统计，是最常用的校验方式。
- 小时级增量校验：依靠元数据中的创建或更新时间，获取数据变化（增量）分区，再进行校验。在24小时内自动进行多次校验，可以持续关注数据一致性变化情况。第二日0点自动停止任务。
- 指定日期校验：仅适用于分区为年月日格式的分表，支持周期性指定一个或多个连续日期分区进行校验。分区格式不为年月日的表不会进行校验。

### 支持的源端、目的端组件

| 源端组件                                                                                                                                              | 目的端组件                                                                                                                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>● Hive</li><li>● Hbase</li><li>● Doris</li><li>● MaxCompute</li><li>● ClickHouse</li><li>● Athena</li></ul> | <ul style="list-style-type: none"><li>● Hive</li><li>● DLI</li><li>● MRS(Doris)</li><li>● MRS(Hbase)</li><li>● MRS(ClickHouse)</li></ul> |

## 各组件支持的校验方式

| 组件         | 支持的校验方式                                                                                                    |
|------------|------------------------------------------------------------------------------------------------------------|
| Hive       | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 日级增量校验</li><li>● 小时级增量校验</li><li>● 指定日期校验</li></ul> |
| MaxCompute |                                                                                                            |
| DLI        |                                                                                                            |
| Doris      | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 日级增量校验</li><li>● 小时级增量校验</li></ul>                  |
| Hbase      | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 日级增量校验</li></ul>                                    |
| ClickHouse | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 指定日期校验</li></ul>                                    |
| Athena     | <ul style="list-style-type: none"><li>● 全量校验</li><li>● 日级增量校验</li><li>● 指定日期校验</li></ul>                   |

## 11.2 准备工作

在使用大数据校验功能前，请务必先完成以下必要的准备工作。

### 创建项目

通过创建独立的项目，可以将不同的操作场景进行区分和隔离，从而实现更好的组织管理和资源分配。创建方法请参考[项目管理](#)。

### 准备迁移工具 Edge

在启用大数据校验功能前，需要先在源端环境中安装用于配合大数据校验的迁移工具 Edge。大数据校验场景仅支持Linux版本的Edge，安装方法请参考[下载并安装Linux版本Edge](#)。安装成功后，将Edge与迁移中心进行连接，连接方法请参考[连接迁移中心](#)。

### 选择并启用迁移工具

- 步骤1** 迁移工具Edge与迁移中心连接成功后，登录[迁移中心](#)管理控制台。
- 步骤2** 在左侧导航栏选择“[实施>大数据校验](#)”，进入[大数据校验](#)页面。在页面左上角的**当前项目**下拉列表中选择[创建的项目](#)。
- 步骤3** 单击“[选择并启用迁移工具](#)”按钮，弹出[选择迁移工具](#)窗口。
- 步骤4** 在下拉列表中，选择与迁移中心连接成功的Edge。



- 如迁移工具状态为“**在线**”、“**已启用**”，代表该工具已被用于对数相关操作，单击“**确定**”按钮，继续使用该工具进行后续操作。
- 如迁移工具状态为“**在线**”、“**未启用**”，代表该工具尚未被用于对数，单击“**启用**”按钮，启用该工具进行后续操作。
- 如迁移工具状态为“**离线**”、“**已删除**”，代表该工具已被删除，无法进行对数操作；如果该工具在删除前进行过对数操作，可以单击“**查看**”按钮，查看已有数据。
- 如迁移工具状态为“**离线**”、“**未启用**”，代表该工具无法进行对数相关操作，单击“**查看**”按钮，界面上的所有操作按钮会被禁用。

---

 **注意**

在进行大数据校验前，请确保迁移工具Edge处于在线并启用状态。

---

----结束

## 准备执行机

一致性校验需要将源端和目的端的数据进行对比。执行机作为任务调度和下发单元，负责获取这些数据。由于执行机需要从大数据组件中获取数据，因此需要在执行机安装大数据组件的命令行客户端（MRS/CDH）。通过客户端执行机与大数据组件即可进行交互，从而获取必要的数据库。

同时，为了确保客户端能够成功连接到集群并执行任务，需要确保安装客户端的主机 hosts 文件中包含了集群所有节点的主机名和 IP 地址，以及需要集群的安全组入方向规则将所有端口对客户端节点放开，具体操作请参考[添加安全组规则](#)。

安装 MRS 客户端的方法请参考[安装客户端（3.x 及之后版本）](#)。建议安装客户端的主机规格不小于 16U32G。

## 11.3 创建元数据连接

创建元数据来源类型的连接，获取源端大数据库表清单。目前支持的连接类型包括：

- **Hive Metastore**连接
- **Doris**连接
- **HBase**连接
- **ClickHouse**连接
- **Glue**连接
- **MaxCompute**连接

### 前提条件

已完成[准备工作](#)。

### 操作步骤

创建元数据连接的详细介绍请参考[创建元数据连接](#)。

**注意**

大数据校验场景仅支持Linux版本的Edge，安装方法请参考[下载并安装Linux版本Edge](#)。

## 11.4 创建表组并添加数据表

将源端数据表加入表组中，为后续对数作业做准备。

**须知**

- 每次最多支持导入10,000张表。
- 待导入的表必须属于同一个元数据来源。
- 导入数据表时，系统不会校验数据表加入的分组数量和对数标准。为保持对数的准确性，建议同一数据表加入的分组不超过3个，并避免加入对数标准相同的分组。

### 前提条件

- 已完成[准备工作](#)。
- 已[创建元数据连接](#)。

### 创建表组

**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>大数据校验](#)”，进入[大数据校验](#)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4** 单击“[新增表组](#)”，弹出[新增表组](#)窗口。

**步骤5** 根据[表11-1](#)，配置参数。

表 11-1 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  | 根据实际需求，选择 <a href="#">创建的元数据连接</a> 。<br><b>注意</b><br>表组中的所有表必须来源于同一类型的元数据来源。           |
| 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单击“ <a href="#">查看更多</a> ”进行查看。 |
| 说明（可选） | 您可以输入用于区分表组的说明。                                                                        |

**步骤6** 参数配置完成后，单击“确认”按钮，创建表组完成。在表组列表，可以查看创建的表组信息。

表组创建完成后，即可[导入数据表并加入表组](#)。

----结束

## 导入数据表并加入表组

**步骤1** 在表管理页面，选择库表清单页签。

**步骤2** 单击“导入数据表”按钮，弹出导入数据表窗口。

**步骤3** 选择元数据连接和要加入的表组。

**步骤4** 单击“模板下载”，将导入模板下载到本地。打开导入模板，将需要导入的数据表对应参数填写到模板内。

### 须知

- 每个模板最多支持导入10,000张表。
- 同一表组内的表必须属于同一个元数据来源。
- 模板中单元格不允许存在公式且必须是文本格式，否则会导致解析失败。

**步骤5** 模板参数填写完成并保存后，单击导入窗口的“添加文件”按钮，将填写后的模板文件上传。

**步骤6** 上传成功后，单击“确认”按钮，系统开始导入。在库表清单列表中可以查看导入的数据表。

----结束

## 11.5 创建连接

### 11.5.1 创建执行机连接

对Hive数据进行一致性校验时，需要分别与源端和目的端的执行机建立连接，用于执行数据统计，统计结果将用于数据校验。

### ⚠ 注意

- 需要确保源端和目的端执行机都安装了大数据组件的命令行客户端。
- 源端和目的端执行机均需要创建连接。

## 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>大数据校验”，进入大数据校验页面。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** 在功能入口区域，单击“**连接管理**”，进入**连接管理**页面。
- 步骤4** 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。
- 步骤5** 根据实际情况选择执行机类型，单击“**下一步**”，根据**表11-2**，配置执行机参数。

**表 11-2 执行机参数配置说明**

| 参数      | 配置说明                                                  |
|---------|-------------------------------------------------------|
| 所属源     | 根据实际情况选择。<br><b>须知</b><br>CDH执行机只支持源端，MRS执行机支持源端和目的端。 |
| 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   | 选择执行机的登录凭证。凭证的添加方法请参考 <b>添加资源凭证</b> 中的大数据-执行机。        |
| 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/CDH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。             |

- 步骤6** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与执行机建立连接，测试通过代表与执行机连接成功。
- 步骤7** 测试连接通过后，单击“**确认**”按钮，创建执行机连接完成。
- 步骤8** 在**连接管理**页面，可以查看创建的执行机连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

## 11.5.2 创建大数据云服务连接

对大数据云服务数据进行一致性校验时，需要分别与源端和目的端云服务建立连接，用于数据查询和计算。

目前支持创建连接的大数据云服务包括：

- **数据湖探索（DLI）**
- **Athena**

### 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>大数据校验**”，进入**大数据校验**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** 在功能入口区域，单击“**连接管理**”，进入**连接管理**页面。
- 步骤4** 单击页面右上角的“**新增连接**”，右侧弹出**新增连接**窗口。
- 步骤5** 根据实际情况选择大数据云服务，单击“**下一步**”，进入**详细配置**页面。
- 步骤6** 根据对应的参数配置说明，配置云服务参数。
- [DLI服务参数配置说明](#)
  - [Athena服务参数配置说明](#)

**表 11-3** DLI 服务参数配置说明

| 参数    | 配置说明                                                     |
|-------|----------------------------------------------------------|
| 所属源   | 数据湖探索（DLI）仅支持目的端。                                        |
| 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DLI云服务的凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-DLI。 |
| 区域/项目 | 选择需要校验的数据所在区域/项目。                                        |
| 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。                           |

**表 11-4** Athena 服务参数配置说明

| 参数           | 配置说明                                                           |
|--------------|----------------------------------------------------------------|
| 所属源          | Athena仅支持源端。                                                   |
| 连接名称         | 默认创建名称为“Athena-4位随机数（字母和数字）”，也可以自定义名称。                         |
| Athena凭证     | 选择Athena云服务的凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-Athena。 |
| 区域           | 选择需要校验的数据所在区域。                                                 |
| Catalog Name | 输入需要校验的数据表所在的目录名称。                                             |
| Work Group   | 输入用于管理数据表的工作组。                                                 |
| S3 桶地址       | 输入数据存储桶的路径。                                                    |

- 步骤7** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与云服务建立连接，测试通过代表与云服务连接成功。
- 步骤8** 测试连接通过后，单击“**确认**”按钮，创建云服务连接完成。
- 步骤9** 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

### 11.5.3 创建大数据组件连接

对大数据组件数据进行一致性校验时，需要与大数据组件建立连接，用于数据查询。

目前支持创建连接的大数据组件包括：

- Doris
- HBase
- ClickHouse

#### 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>大数据校验](#)”，进入[大数据校验](#)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** 在功能入口区域，单击“[连接管理](#)”，进入[连接管理](#)页面。

**步骤4** 单击页面右上角的“[新增连接](#)”，右侧弹出[新增连接](#)窗口。

**步骤5** 根据实际情况选择大数据组件，单击“[下一步](#)”，进入[详细配置](#)页面。

**步骤6** 根据对应的参数配置说明，配置执行机参数。

- [Doris组件参数配置说明](#)
- [HBase组件参数配置说明](#)
- [ClickHouse组件参数配置说明](#)

表 11-5 Doris 组件参数配置说明

| 参数      | 配置说明                                                                  |
|---------|-----------------------------------------------------------------------|
| 所属源     | Doris组件仅支持目的端。                                                        |
| 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 | 选择在Edge添加的Doris凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-Doris”。 |
| 数据库地址   | 填写连接到Doris集群部署的服务器IP地址。                                               |
| 数据库端口   | 填写连接到Doris集群部署的服务器的端口。默认为3306。                                        |
| 数据库名    | 填写Doris数据库名称。                                                         |

表 11-6 HBase 组件参数配置说明

| 参数  | 配置说明           |
|-----|----------------|
| 所属源 | HBase组件仅支持目的端。 |

| 参数               | 配置说明                                                                  |
|------------------|-----------------------------------------------------------------------|
| 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          | 选择在Edge添加的HBase凭证。凭证的添加方式请参考 <a href="#">添加资源凭证</a> ，类型选择“大数据-HBase”。 |
| 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。                  |
| Zookeepereer连接端口 | 填写连接到Zookeepereer的端口。默认为2181。                                         |
| HBase 版本         | 选择HBase的版本。                                                           |

表 11-7 ClickHouse 组件参数配置说明

| 参数                   | 配置说明                                                                       |
|----------------------|----------------------------------------------------------------------------|
| 所属源                  | ClickHouse组件仅支持目的端。                                                        |
| 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）     | 选择在Edge添加的ClickHouse凭证。凭证的添加方法请参考 <a href="#">添加资源凭证</a> 中的大数据-ClickHouse。 |
| 安全集群                 | 选择是否部署在安全集群下。                                                              |
| ClickHouse Server IP | 填写连接到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ClickHouse 服务器通信的 HTTP 端口。默认为8123。                                 |
| HTTP SSL/TLS端口       | 安全集群下，与ClickHouse 服务器通信的HTTPS 端口。                                          |

**步骤7** 参数配置完成后，单击“**测试连接**”，系统会校验配置信息并与组件建立连接，测试通过代表与组件连接成功。

**步骤8** 测试连接通过后，单击“**确认**”按钮，创建大数据组件连接完成。

**步骤9** 在**连接管理**页面，可以查看创建的连接列表以及基本信息。单击操作列的“**修改连接**”，可以修改连接参数配置。

----结束

## 11.6 创建统计任务并执行校验

利用创建的源端连接和目的端连接，分别创建源端和目的端统计任务，获取源端和目的端的数据信息，以便于进行数据校验。

支持的大数据组件和统计方式说明请参见[简介](#)。

**注意**

创建的源端统计任务和目的端统计任务的统计方式必须保持一致。

## 前提条件

- 已[创建表组并添加数据表](#)。
- 已根据组件类型，[创建对应的连接](#)。

## 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>大数据校验](#)”，进入[大数据校验](#)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** 在功能入口区域，单击“[任务管理](#)”，进入[任务管理](#)页面。
- 步骤4** 单击页面右上角的“[新增任务](#)”，右侧弹出[新增任务](#)窗口。
- 步骤5** 根据实际需求，选择大数据组件类型和统计方式，单击“[下一步](#)”，进入[详细配置](#)页面。
- 步骤6** 根据对应的参数配置说明，配置所选的统计方式参数。

**说明**

各组件之间配置参数略有区别，请以实际页面为准。

- [全量对数作业参数配置说明](#)
- [日级增量对数参数配置说明](#)
- [小时级增量对数参数配置说明](#)
- [指定日期校验参数配置说明](#)

表 11-8 全量校验参数配置说明

| 区域   | 参数    | 配置说明                                                                                                                   |
|------|-------|------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名   | 默认创建名称为“组件类型全量校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                               |
| 任务配置 | 表组    | 选择需要校验的表所在表组。                                                                                                          |
|      | 执行机连接 | Hive组件配置参数。 <ul style="list-style-type: none"> <li>• 创建源端统计任务时，选择源端执行机连接。</li> <li>• 创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |
|      | DLI连接 | 数据湖探索（DLI）组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                            |



| 区域                 | 参数           | 配置说明                                                                                                                                                                                                                                                                                   |
|--------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | Doris连接      | Doris组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的Doris连接</a> 。                                                                                                                                                                                                                               |
|                    | HBase连接      | HBase组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的HBase连接</a> 。                                                                                                                                                                                                                               |
|                    | ClickHouse连接 | ClickHouse组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的ClickHouse连接</a> 。                                                                                                                                                                                                                     |
|                    | Athena连接     | Athena组件配置参数，仅支持创建源端端统计任务。选择 <a href="#">创建的Athena连接</a> 。                                                                                                                                                                                                                             |
|                    | 执行时间         | 设置作业自动执行时间。当作业激活后，每天都在设置的时间自动执行。                                                                                                                                                                                                                                                       |
|                    | 高级选项         | <ul style="list-style-type: none"> <li>● <b>单客户端并发度</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。</li> <li>● <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。</li> <li>● <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~7200。</li> </ul> |
| 数据筛选               | 非分区表处理策略     | 选择对于非分区表的校验处理策略。 <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul>                                   |
|                    | 高级选项         | <b>单SQL字段上限</b> ：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500                                                                                                                                                                          |
| 执行脚本<br>(Hive配置参数) | 安全配置         | 如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的.keytab文件上传至执行机。 <ul style="list-style-type: none"> <li>● <b>keytab文件地址</b>：输入.keytab文件在执行机的存放地址。</li> <li>● <b>keytab文件用户名</b>：输入.keytab文件对应的用户名。</li> </ul>                                                         |

| 区域                               | 参数   | 配置说明                                                                                                                                                                              |
|----------------------------------|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                  | 执行命令 | <p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> <li>● <b>Beeline</b>: 用于与 Hive交互的命令行工具。</li> <li>● <b>Spark-sql</b>: 用于执行 SQL 查询和分析数据。</li> </ul> |
| MaxCompute参数<br>(MaxCompute配置参数) | -    | 根据实际需求，添加MaxCompute参数。参数详情请参考 <a href="#">MaxCompute文档</a> 。                                                                                                                      |
| 统计配置<br>(HBase配置参数)              | 运行模式 | <p>支持以下两种运行模式：</p> <p>Yarn: 适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。</p> <p>Local: 适用于小规模数据集或开发测试环境，可以快速进行调试和验证。</p>                                                                    |
|                                  | 参数   | 根据所选的运行模式和实际需求，添加命令行参数。                                                                                                                                                           |

表 11-9 日级增量校验参数配置说明

| 区域   | 参数       | 配置说明                                                                                                                          |
|------|----------|-------------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名      | 默认创建名称为“日级增量校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                        |
| 任务配置 | 表组       | 选择需要校验的表所在表组。                                                                                                                 |
|      | 执行机连接    | <p>Hive组件配置参数。</p> <ul style="list-style-type: none"> <li>● 创建源端统计任务时，选择源端执行机连接。</li> <li>● 创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |
|      | DLI连接    | 数据湖探索（DLI）组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                                   |
|      | Doris连接  | Doris组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的Doris连接</a> 。                                                                      |
|      | HBase连接  | HBase组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的HBase连接</a> 。                                                                      |
|      | Athena连接 | Athena组件配置参数，仅支持创建源端统计任务。选择 <a href="#">创建的Athena连接</a> 。                                                                     |
|      | 执行时间     | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。                                                                                             |

| 区域   | 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | 高级选项     | <ul style="list-style-type: none"> <li>● <b>单客户端并发度</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。</li> <li>● <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。</li> <li>● <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~7200。</li> </ul>                                                                                                                                                                                                                                                                              |
| 数据筛选 | 统计范围     | <p>选择需要对哪些时间段内的增量数据进行校验。默认统计24小时增量数据。T代表执行时间，T-n代表执行时间前的n x 24小时。</p> <p>勾选<b>统计连续日期</b>，可以选择多个统计时间。</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|      | 非分区表处理策略 | <p>选择对于非分区表的对数处理策略。</p> <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul>                                                                                                                                                                                                                                                                                                         |
|      | 高级选项     | <ul style="list-style-type: none"> <li>● <b>分区筛选方式</b>：通过“更新时间”和“创建时间”，判断分区是否属于校验范围内。 <ul style="list-style-type: none"> <li>- <b>更新时间</b>：指分区内数据最后一次被修改或更新的时间戳。如果您关心的是分区内数据的最新状态或最近一段时间内的数据变更，可以选择“更新时间”作为筛选条件。</li> <li>- <b>创建时间</b>：指分区被首次创建时的时间戳。如果您关心的是分区从创建开始到某个时间点内的数据，可以选择“创建时间”作为筛选条件。</li> </ul> </li> <li>● <b>分区上限</b>：用于限制查询或处理的分区数量，避免不必要的全表扫描或处理大量不需要的数据。默认为3，取值范围：1~50。<br/>例如，设置分区上限为3，系统只处理降序排序后分区列表中的前3个分区。</li> <li>● <b>单SQL字段上限</b>：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</li> </ul> |

| 区域                               | 参数   | 配置说明                                                                                                                                                                                                                                                                     |
|----------------------------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 执行脚本<br>(Hive配置参数)               | 安全配置 | <p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的<code>.keytab</code>文件上传至执行机。</p> <ul style="list-style-type: none"> <li><b>keytab文件地址</b>：输入<code>.keytab</code>文件在执行机的存放地址。</li> <li><b>keytab文件用户名</b>：输入<code>.keytab</code>文件对应的用户名。</li> </ul> |
|                                  | 执行命令 | <p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> <li><b>Beeline</b>：用于与Hive交互的命令行工具。</li> <li><b>Spark-sql</b>：用于执行SQL查询和分析数据。</li> </ul>                                                                                                 |
| MaxCompute参数<br>(MaxCompute配置参数) | -    | 根据实际需求，添加MaxCompute参数。参数详情请参考 <a href="#">MaxCompute文档</a> 。                                                                                                                                                                                                             |
| 统计配置<br>(HBase配置参数)              | 运行模式 | <p>支持以下两种运行模式：</p> <p>Yarn：适用于大型分布式环境，可以充分利用集群资源，提高任务的并行度和效率。</p> <p>Local：适用于小规模数据集或开发测试环境，可以快速进行调试和验证。</p>                                                                                                                                                             |
|                                  | 参数   | 根据所选的运行模式和实际需求，添加命令行参数。                                                                                                                                                                                                                                                  |

表 11-10 小时级增量校验参数配置说明

| 区域   | 参数      | 配置说明                                                                                                                      |
|------|---------|---------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名     | 默认创建名称为“小时级增量校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                   |
| 任务配置 | 表组      | 选择需要校验的表所在表组。                                                                                                             |
|      | 执行机连接   | <p>Hive组件配置参数。</p> <ul style="list-style-type: none"> <li>创建源端统计任务时，选择源端执行机连接。</li> <li>创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |
|      | DLI连接   | 数据湖探索（DLI）组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                               |
|      | Doris连接 | Doris组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的Doris连接</a> 。                                                                  |

| 区域   | 参数       | 配置说明                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      | 开始时间     | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|      | 高级选项     | <ul style="list-style-type: none"> <li>● <b>单客户端并发度</b>：进行数据统计时，单个执行机的任务并发数。默认为3，取值范围：1~10。</li> <li>● <b>单文件SQL条数</b>：每个需要校验的表会产生一条SQL用以获取数据摘要，每批提交给集群处理的数据文件中所包含的SQL条数。默认为10，推荐取值范围：1~50。</li> <li>● <b>超时时间</b>：每个SQL语句在大数据集群上执行时的最大允许时长，单位为秒（s）。默认为600，推荐取值范围：600~7200。</li> </ul>                                                                                                                                                                                                                                                                                            |
| 数据筛选 | 执行周期     | 设置任务的执行频率和持续时间。                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|      | 非分区表处理策略 | 选择对于非分区表的校验处理策略。 <ul style="list-style-type: none"> <li>● <b>全部校验</b>：全量处理所有非分区表，不做筛选。</li> <li>● <b>全部忽略</b>：忽略所有非分区表，不做校验。</li> <li>● <b>按更新时间过滤</b>：只处理更新时间在校验范围内的非分区表。<b>注意</b>：非分区表非insert写入业务，元数据更新时间存在偏差，按更新时间过滤分区时可能遗漏分区，慎重选择该选项！</li> </ul>                                                                                                                                                                                                                                                                                                                              |
|      | 高级选项     | <ul style="list-style-type: none"> <li>● <b>分区筛选方式</b>：通过“更新时间”和“创建时间”，判断分区是否属于校验范围内。               <ul style="list-style-type: none"> <li>- <b>更新时间</b>：指分区内数据最后一次被修改或更新的时间戳。如果您关心的是分区内数据的最新状态或最近一段时间内的数据变更，可以选择“更新时间”作为筛选条件。</li> <li>- <b>创建时间</b>：指分区被首次创建时的时间戳。如果您关心的是分区从创建开始到某个时间点内的数据，可以选择“创建时间”作为筛选条件。</li> </ul> </li> <li>● <b>分区上限</b>：用于限制查询或处理的分区数量，避免不必要的全表扫描或处理大量不需要的数据。默认为3，取值范围：1~50。<br/>例如，设置分区上限为3，系统只处理降序排序后分区列表中的前3个分区。</li> <li>● <b>单SQL字段上限</b>：每条SQL查询的表的字段上限。单SQL查询的字段数越多，执行时间越长，如果设置的字段数上限过少，则可能形成更多SQL，影响获取校验结果的效率。默认是0，即无上限，取值范围100~500。</li> </ul> |

| 区域                                 | 参数   | 配置说明                                                                                                                                                                                                                                                                     |
|------------------------------------|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 执行脚本<br>(Hive配置参数)                 | 安全配置 | <p>如果大数据集群启用了安全配置（例如Kerberos安全认证），请勾选此选项，并配置安全认证命令。配置前，需要您手动将包含认证密钥的<code>.keytab</code>文件上传至执行机。</p> <ul style="list-style-type: none"> <li><b>keytab文件地址</b>：输入<code>.keytab</code>文件在执行机的存放地址。</li> <li><b>keytab文件用户名</b>：输入<code>.keytab</code>文件对应的用户名。</li> </ul> |
|                                    | 执行命令 | <p>可以通过设置Beeline或Spark-sql命令行参数，对数据进行处理。</p> <ul style="list-style-type: none"> <li><b>Beeline</b>：用于与Hive交互的命令行工具。</li> <li><b>Spark-sql</b>：用于执行SQL查询和分析数据。</li> </ul>                                                                                                 |
| MaxCompute参数<br>(MaxCompute组件配置参数) | -    | <p>根据实际需求，添加MaxCompute参数。参数详情请参考<a href="#">MaxCompute文档</a>。</p>                                                                                                                                                                                                        |

表 11-11 指定日期校验参数配置说明

| 区域   | 参数           | 配置说明                                                                                                                      |
|------|--------------|---------------------------------------------------------------------------------------------------------------------------|
| 基础信息 | 任务名          | 默认创建名称为“指定日期校验-4位随机数（字母和数字）”，也可以自定义名称。                                                                                    |
| 任务配置 | 表组           | 选择需要校验的表所在表组。                                                                                                             |
|      | 执行机连接        | <p>Hive组件配置参数。</p> <ul style="list-style-type: none"> <li>创建源端统计任务时，选择源端执行机连接。</li> <li>创建目的端统计任务时，选择目的端执行机连接。</li> </ul> |
|      | DLI连接        | 数据湖探索（DLI）组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的DLI连接</a> 。                                                               |
|      | ClickHouse连接 | ClickHouse组件配置参数，仅支持创建目的端统计任务。选择 <a href="#">创建的ClickHouse连接</a> 。                                                        |
|      | Athena连接     | Athena组件配置参数，仅支持创建源端统计任务。选择 <a href="#">创建的Athena连接</a> 。                                                                 |
|      | 执行时间         | 设置作业自动执行时间。当作业激活后，每天都会在设置的时间自动执行。                                                                                         |

| 区域                               | 参数       | 配置说明                                                                                                                                                                                                                                                                                                |
|----------------------------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                  | 高级选项     | <ul style="list-style-type: none"> <li>● <b>单客户端并发度</b>: 进行数据统计时, 单个执行机的任务并发数。默认为3, 取值范围: 1~10。</li> <li>● <b>单文件SQL条数</b>: 每个需要校验的表会产生一条SQL用以获取数据摘要, 每批提交给集群处理的数据文件中所包含的SQL条数。默认为10, 推荐取值范围: 1~50。</li> <li>● <b>超时时间</b>: 每个SQL语句在大数据集群上执行时的最大允许时长, 单位为秒(s)。默认为600, 推荐取值范围: 600~ 7200。</li> </ul> |
| 数据筛选                             | 指定日期     | 选择需要对哪些时间段内的增量数据进行校验。默认统计24小时增量数据。T代表执行时间, T-n代表执行时间前的n x 24小时。<br>勾选 <b>统计连续日期</b> , 可以选择多个统计时间。                                                                                                                                                                                                   |
|                                  | 非分区表处理策略 | 选择对于非分区表的对数处理策略。 <ul style="list-style-type: none"> <li>● <b>全部校验</b>: 全量处理所有非分区表, 不做筛选。</li> <li>● <b>全部忽略</b>: 忽略所有非分区表, 不做校验。</li> <li>● <b>按更新时间过滤</b>: 只处理更新时间在校验范围内的非分区表。<b>注意</b>: 非分区表非insert写入业务, 元数据更新时间存在偏差, 按更新时间过滤分区时可能遗漏分区, 慎重选择该选项!</li> </ul>                                       |
|                                  | 高级选项     | <b>单SQL字段上限</b> : 每条SQL查询的表的字段上限。单SQL查询的字段数越多, 执行时间越长, 如果设置的字段数上限过少, 则可能形成更多SQL, 影响获取校验结果的效率。默认是0, 即无上限, 取值范围100~500。                                                                                                                                                                               |
| 执行脚本<br>(Hive配置参数)               | 安全配置     | 如果大数据集群启用了安全配置(例如Kerberos安全认证), 请勾选此选项, 并配置安全认证命令。配置前, 需要您手动将包含认证密钥的 <b>.keytab</b> 文件上传至执行机。 <ul style="list-style-type: none"> <li>● <b>keytab文件地址</b>: 输入<b>.keytab</b>文件在执行机的存放地址。</li> <li>● <b>keytab文件用户名</b>: 输入<b>.keytab</b>文件对应的用户名。</li> </ul>                                          |
|                                  | 执行命令     | 可以通过设置Beeline或Spark-sql命令行参数, 对数据进行处理。 <ul style="list-style-type: none"> <li>● <b>Beeline</b>: 用于与Hive交互的命令行工具。</li> <li>● <b>Spark-sql</b>: 用于执行SQL查询和分析数据。</li> </ul>                                                                                                                            |
| MaxCompute参数<br>(MaxCompute配置参数) | -        | 根据实际需求, 添加MaxCompute参数。参数详情请参考 <a href="#">MaxCompute文档</a> 。                                                                                                                                                                                                                                       |

- 步骤7** 配置完成后，单击“**保存**”按钮，完成统计任务的创建。创建成功后，系统会自动将参数配置同步到迁移工具Edge。您可以在**任务列表**查看创建的统计任务及其配置同步状态。
- 步骤8** 配置同步完成后，您可以选择以下两种方式执行统计任务：
- **自动定时执行**：按照设置的执行时间，自动、定时执行作业。
    - a. 在任务列表，单击统计任务**调度状态**列的“**激活**”按钮，弹出**激活**提示窗。
    - b. 单击“**确认**”按钮，激活自动化执行作业。
  - **手动执行**：手动立即触发一次作业的执行。
    - a. 在任务列表，单击统计任务**操作**列的“**执行作业**”按钮，弹出**执行作业**提示窗。
    - b. 单击“**确认**”按钮，立即执行任务。
- 步骤9** 执行任务后，单击操作列的“**查看实例**”按钮，进入**实例列表**页面。在**实例列表**页面，可以进行如下操作：
- **查看实例状态、进度统计数据、作业执行开始时间和结束时间**等信息。
  - **上传日志**：如果您需要查看作业的执行日志，可以使用**上传日志**功能，将日志文件上传到您的对象存储（OBS）桶中，以便随时查看和分析。使用上传日志功能前，需要您先在迁移工具Edge上配置日志桶，配置方法请参见**配置日志桶**。
  - **取消/终止实例**：如果您需要终止自动定时执行作业或取消正在运行中的作业，可以单击操作列的“**取消/终止实例**”按钮，将执行中的作业取消/终止。
  - **补数**：如果存在结果获取失败的表，可以单击**进度统计**列的“**补数**”按钮，对获取对数摘要失败的表进行重新获取。

----结束

## 11.7 查看并导出校验结果

当源端与目的端的统计任务都执行完成后，在**校验结果**页面可以查看并导出任务的执行结果。校验结果将按照任务开始执行的日期进行分批，如果单日多次执行校验任务，系统只会保留最后一次的校验结果，之前的执行结果会被后续的执行所覆盖。

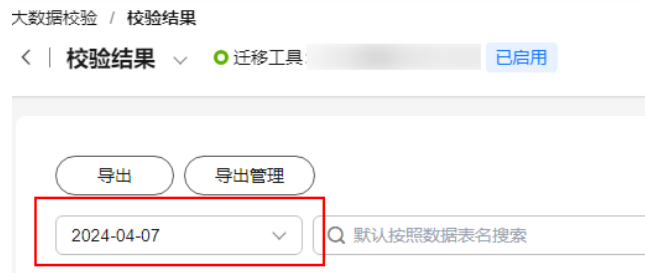
### 前提条件

已分别**创建源端和目的端统计任务并执行校验**。

### 查看校验结果

- 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>大数据校验**”，进入**大数据校验**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** 在功能入口区域，单击“**校验结果**”，进入**校验结果**页面。
- 步骤4** 通过执行日期，筛选出要查看的执行批次。





**步骤5** 在校验结果列表中，可以进行如下操作：

- 查看校验详情：单击分区达标详情列的“查看”按钮，展开该数据表的作业详情。单击操作列的“查看详情”，进入校验详情页面，查看该数据表的校验详细信息。
- 查看执行历史及状态：单击操作列的“查看历史”，可以查看该表的所有执行批次和达标状态。

----结束

## 导出校验结果

**步骤1** 在校验结果页面，单击“导出”按钮，弹出导出校验结果窗口。

**步骤2** 选择需要导出的批次、表组名以及类型，单击“确认”按钮，系统开始导出校验结果。

**步骤3** 导出完成后，单击“导出管理”按钮，右侧弹出导出管理窗口。

**步骤4** 单击操作列的“下载”，可以将对应表的校验结果下载到本地。

----结束

# 12 修订记录

| 发布日期       | 修订记录                                                                                                                         |
|------------|------------------------------------------------------------------------------------------------------------------------------|
| 2024-4-17  | 新增 <b>迁移方案配置</b><br>新增 <b>创建对象存储批量迁移 workflow</b>                                                                            |
| 2024-04-15 | 新增 <b>大数据作业迁移</b> 。                                                                                                          |
| 2024-03-21 | <ul style="list-style-type: none"><li>● 新增<b>大数据血缘</b>。</li><li>● 新增<b>委托权限说明</b>。</li><li>● 新增<b>导入阿里云主机资源清单</b>。</li></ul> |
| 2024-03-11 | <ul style="list-style-type: none"><li>● 公网采集源端新增 Azure。</li><li>● 新增<b>导出作业</b>。</li></ul>                                   |
| 2024-02-29 | 新增 <b>大数据校验</b> 。                                                                                                            |
| 2024-02-20 | 新增 <b>迁移集群</b> 。                                                                                                             |
| 2024-01-31 | <ul style="list-style-type: none"><li>● 新增<b>采集主机性能</b>。</li><li>● <b>规格评估</b>新增基于性能推荐主机规格。</li></ul>                        |
| 2023-12-30 | 新增 <b>购买资源</b> 。                                                                                                             |
| 2023-11-30 | 新增 <b>业务验证</b> 。                                                                                                             |
| 2023-10-30 | 第一次发布。                                                                                                                       |