Migration Center

User Guide

Issue 19

Date 2024-11-30

HUAWEI

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process.* For details about this process, visit the following web page:
https://www.huawei.com/en/psirt/vul-response-process
For vulnerability information, enterprise customers can visit the following web page:
https://securitybulletin.huawei.com/enterprise/en/security-advisory

Contents

1 Permissions Management

1.1 Creating a User and Granting MgC Permissions

This section describes how to use **Identity and Access Management (IAM)** for fine-grained permissions control on your Migration Center (MgC) resources. With IAM, you can:

- Create IAM users for employees based on the organizational structure of your enterprise. Each IAM user has their own security credentials for accessing MgC resources.
- Grant only the permissions required for users to perform a specific task.
- Entrust a Huawei Cloud account or cloud service to perform professional and efficient O&M on your MgC resources.

If your Huawei Cloud account does not need individual IAM users, then you may skip over this section.

This section describes the procedure for granting permissions (see **Process Flow**).

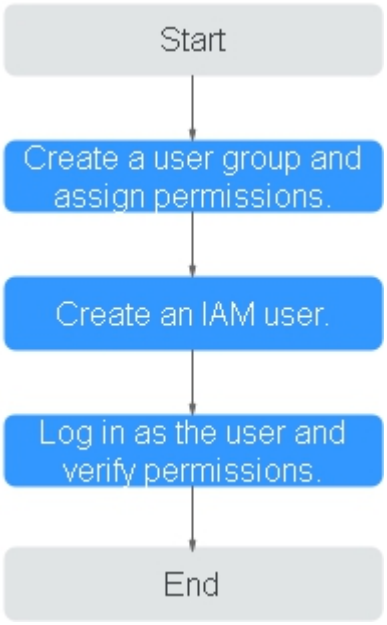
Prerequisites

Before assigning permissions to user groups, you should learn about system-defined policies supported by MgC and select the policies based on service requirements.

For details about the system-defined policies supported by MgC and the comparison between them, see **System-defined Policies**. For the system-defined policies of other services, see **System-defined Permissions**.

Process Flow

Figure 1-1 Granting MgC permissions to users



1. **Create a user group and assign permissions** to it.
   - **System-defined policy**: Create a user group on the IAM console, and assign the MgC system-defined policies to the user group based on the **Description of MgC System-defined Policies** and actual permissions requirements. Select **All Resources** for **Scope**.

     Table 1-1 MgC system-defined policies

Policy Name	Description	Policy Type	Policy Content
MgC FullAccess	Administrator permissions for MgC. Users with these permissions can perform all operations on MgC data.	System-defined policy	**MgC FullAccess Policy Content**
MgC ReadOnlyAccess	Read-only permissions for MgC. Users with permissions can only view MgC data.	System-defined policy	**MgC ReadOnlyAccess Policy Content**

Policy Name	Description	Policy Type	Policy Content
MgC DiscoveryAccess	Permissions for resource discovery of MgC. Users with these permissions can use the resource discovery function of MgC and view MgC data.	System-defined policy	**MgC DiscoveryAccess Policy Content**
MgC AssessAccesss	Permissions for application assessment of MgC. Users with these permissions can use the resource discovery and application assessment functions of MgC and view MgC data.	System-defined policy	**MgC AssessAccesss Policy Content**
MgC MigrateAccesss	Permissions for application migration of MgC. Users with these permissions can use the resource discovery, application assessment, and application migration functions of MgC and view MgC data.	System-defined policy	**MgC MigrateAccess Policy Content**
MgC AppDiscoveryAccess	Permissions for resource discovery of MgC. Users with these permissions can use the resource discovery and data collection functions of MgC and view MgC data.	System-defined policy	**MgC AppDiscoveryAccess Policy Content**
MgC MrrAccess	Permissions for service verification of MgC. Users with these permissions can use the service verification function of MgC and view MgC data.	System-defined policy	**MgC MrrAccess Policy Content**

  - **Custom policy**: If an IAM user only needs specific MgC permissions, create custom policies. For details, see **MgC Custom Policies**.

2. **Create an IAM user and add it to the user group**.

   Create a user on the IAM console and add the user to the group created in **1**.

NOTICE

When the IAM user access certain MgC functions for the first time, the IAM user needs to authorize MgC through agencies. Therefore, the IAM user must have the permissions required to create agencies. There are two methods for the IAM user to obtain the required permissions.

- (Recommended) Ask the administrator to log in to the MgC console and access the corresponding functions to complete the authorization.
- Ask the administrator to grant the corresponding system-defined MgC permissions and the **iam:agencies:createAgency**, **iam:permissions:grantRoleToAgency**, and **iam:roles:updateRole** permissions on the IAM console.

3. **Sign in** and verify permissions.

   Sign in to the IAM console using the created IAM user, and in the authorized region, perform the following operations:

   - Choose **Service List** > **Migration Center**. On the MgC console, perform operations based on the granted permissions. If you can, the granted permissions have taken effect.
   - Choose another service from **Service List**. If a message appears indicating that you have insufficient permissions to access the service, the granted permissions have taken effect.

MgC FullAccess Policy Content

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

MgC ReadOnlyAccess Policy Content

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

MgC DiscoveryAccess Policy Content

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



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

MgC AssessAccess Policy Content

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:assess",
        "iam:agencies:listAgencies",
        "iam:roles:listRoles",
        "iam:quotas:listQuotas",
        "iam:permissions:listRolesForAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

MgC MigrateAccess Policy Content

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "mgc:*:query*",
        "mgc:*:discovery",
        "mgc:*:assess",
        "mgc:*:migrate",
        "iam:agencies:listAgencies",
        "iam:roles:listRoles",
        "iam:quotas:listQuotas",
        "iam:permissions:listRolesForAgency"
      ],
      "Effect": "Allow"
    }
  ]
}
```

MgC AppDiscoveryAccess Policy Content

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


MgC MrrAccess Policy Content

```
{
        "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 Custom Policies

Custom policies can be created to supplement the system-defined policies of MgC.

You can create custom policies in either of the following ways:

- Visual editor: Select cloud services, actions, resources, and request conditions. This does not require knowledge of policy syntax.
- JSON: Create a JSON policy from scratch or based on an existing policy.

For details, see **Creating a Custom Policy**. The following section contains examples of common MgC custom policies.

Example MgC Custom Policies

- Allowing users to perform platform collection and resource management

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

- Allowing users to perform TCO analysis and server assessment

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



```
        }
    ]
}
```

- Allowing users to perform application dependency mapping

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

- Allowing users to use migration workflows

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

1.3 Agency Permissions

Overview

To use some functions of MgC, you must delegate MgC the required permissions so that we can provide you with complete services. This section describes the scenarios where authorization is required and what custom permission policies will be created.

The system may create a new custom policy or update an existing policy during the authorization.

- If there is no available custom policy, the system automatically creates a new one.
- If there is an available custom policy but it does not contain required permissions, the system automatically updates the policy.

Creating a Cross-AZ Migration Workflow

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Creating a cross-AZ migration workflow	MgC	MgC AzMigrationAgencyPolicy	ecs:cloudServers:showServer (Query details about an ECS) ecs:flavors:get (Querying ECS Flavors) ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information) ecs:cloudServerQuotas:get (Querying quotas of a tenant) ecs:servers:list (Querying ECSs) ecs:cloudServers:list (Querying details about ECSs) ecs:servers:stop (Stopping an ECS) ecs:cloudServers:listServerInterfaces (Querying NICs of an ECS) ecs:cloudServers:createServers (Creating an ECS) ecs:cloudServers:listServerBlockDevices (Querying information about the disks attached to an ECS) ecs:cloudServerNics:update (Configuring a private IP address for a NIC of an ECS) ecs:availabilityZones:list (Listing AZs) ecs:servers:start (Starting an ECS) ecs:cloudServers:changeNetworkInterface (Updating attributes of a specified NIC on an ECS) ecs:serverInterfaces:get (Querying ECS NICs) ecs:cloudServers:get (Query details about an ECS) vpc:publicIps:create (Creating an EIP) vpc:publicIps:update (Updating an EIP) vpc:subnets:get (Querying subnets) vpc:networks:get (Querying networks) vpc:publicIps:list (Querying EIPs) vpc:publicIps:list (Querying details about an EIP)

Scenario	Delegated Object	Custom Policy	Minimal Permissions
			vpc:ports:get (Querying ports or querying details about a port) vpc:ports:delete (Deleting a port) vpc:ports:update (Updating a port) vpc:ports:create (Creating a port) evs:types:get (Querying EVS disk types) evs:volumes:list (Querying EVS disks) cbr:vaults:get (Querying a specified vault) cbr:vaults:list (Querying vaults) cbr:vaults:create (Creating a vault) cbr:vaults:addResources (Associating resources) cbr:vaults:backup (Creating a restore point) cbr:backups:list (Querying backups) cbr:tasks:list (Querying tasks) cbr:tasks:get (Querying details about a task) cbr:backups:delete (Deleting a backup) cbr:backups:get (Querying a backup) cbr:vaults:delete (Deleting a vault) ims:wholeImages:create (Creating a full-ECS image) ims:images:list (Querying images) ims:images:delete (Deleting an image) ims:images:get (Querying details about an image) ims:serverImages:create (Creating an image)

Migration Cost Analysis

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Creating a migration cost analysis task	MgC	MgC TcoAgencyPolicy	ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information) evs:types:get (Querying EVS disk types) ims:*:get* (Querying details about an image) ims:*:list* (Querying images)

Getting Target Recommendations

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Getting target recommendations	MgC	MgC ServerAssessAgencyPolicy	ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information) ims:images:list (Querying images) evs:types:get (Querying EVS disk types) deh:dedicatedHosts:get (Obtaining details about a DeH) deh:dedicatedHosts:list (Listing DeHs)

Binding a Source Server to an Existing Target Server

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Binding a source server to an existing target server	MgC	MgC ServerBindTarge-tAgencyPolicy	ecs:cloudServers:showServer (Query details about an ECS) evs:volumes:list (Querying EVS disks) ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information)

Creating a Server Migration Workflow

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Creating a Server Migration Workflow	MgC	MgC ServerMigrationAgencyPolicy	ecs:cloudServers:showServer (Query details about an ECS) ecs:cloudServers:createServers (Creating an ECS) sms:server:migrationServer (Migrating a source server) sms:server:queryServer (Querying source servers) ecs:cloudServers:list (Querying ECSs) ecs:cloudServers:listServerBlockDevices (Querying information about the disks attached to an ECS) ecs:cloudServerQuotas:get (Querying quotas of a tenant) vpc:publicIps:create (Creating an EIP) ecs:cloudServers:get (Query details about an ECS)

Purchasing Resources

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Purchasing resources	MgC	MgC PurchaseAgencyPolicy	eps:resources:add (Adding resources to an enterprise project) ecs:cloudServers:createServers (Creating an ECS) evs:volumes:list (Querying EVS disks) ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information) ecs:cloudServers:list (Querying details about ECSs) ecs:cloudServers:createServers (Creating an ECS) vpc:publicIps:update (Updating an EIP) vpc:publicIps:create (Creating an EIP)

Configuring a Server Purchase Template

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Configuring a server purchase template	MgC	MgC PurchaseTemplateA-gencyPolicy	iam:projects:listProjects (Querying projects) eps:enterpriseProjects:list (Querying enterprise projects) vpc:subnets:get (Querying subnets or querying details about a subnet) vpc:securityGroups:get (Querying security groups or querying details about a security group) vpc:vpcs:get (Querying VPC details)

Creating Migration Plans

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Creating a server migration plan (importing target server configurations from an OBS bucket)	MgC	MgC ImportTargetConfigurationAgencyPolicy	obs:object:GetObject (Obtaining object content and metadata) obs:bucket:ListBucket (Listing objects in a bucket) obs:bucket:ListAllMyBuckets (Listing buckets)
Creating a server migration plan (exporting target server configurations)		MgC ExportTargetConfigurationAgencyPolicy	ims:images:list (Listing images) ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information)
Creating a batch object storage migration plan (configuring target buckets)		MgC ListObsBucketsAgencyPolicy	obs:bucket:ListBucket (Listing objects in a bucket) obs:bucket:ListAllMyBuckets (Listing buckets)

Creating a Migration Cluster

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Creating a migration cluster	OMS	OMS ObsMigrationAgencyPolicy	ecs:cloudServers:createServers (Creating an ECS) ecs:cloudServers:listServerInterfaces (Querying NICs of an ECS) ecs:cloudServers:showServer (Query details about an ECS) ecs:cloudServers:deleteServers (Deleting ECSs) ecs:cloudServers:list (Querying details about ECSs) nat:natGateways:create (Creating a NAT Gateway) nat:natGateways:get (Querying details about a NAT gateway) nat:natGateways:delete (Deleting a NAT gateway) nat:snatRules:create (Creating an SNAT rule) nat:snatRules:get (Querying details about an SNAT rule) nat:dnatRules:list (Querying DNAT rules) nat:dnatRules:list (Querying DNAT rules) nat:snatRules:delete (Deleting an SNAT rule) nat:natGateways:list (Querying NAT gateways) vpc:securityGroups:create (Creating a security group) vpc:securityGroups:delete (Deleting a security group) vpc:securityGroups:get (Querying security groups or querying details about a security group) vpc:securityGroupRules:create (Creating a security group rule) vpc:securityGroupRules:get (Querying security group rules or querying details about a security group rule)

Scenario	Delegated Object	Custom Policy	Minimal Permissions
			vpc:securityGroupRules:delete (Deleting a security group rule) vpcep:epservices:create (Creating a VPC endpoint service) vpcep:epservices:get (Querying details about a VPC endpoint service) vpcep:permissions:list (Querying whitelist records of a VPC endpoint service) vpcep:connections:list (Querying connections of a VPC endpoint service) vpcep:epservices:list (Querying VPC endpoint services) vpcep:epservices:delete (Deleting a VPC endpoint service) vpcep:endpoints:create (Creating a VPC endpoint) vpcep:endpoints:list (Querying VPC endpoints) vpcep:endpoints:get (Querying details about a VPC endpoint) vpcep:endpoints:delete (Deleting a VPC endpoint) vpcep:connections:update (Accepting or rejecting a VPC endpoint) vpcep:permissions:update (Batch adding or deleting whitelist records of a VPC endpoint service) lts:topics:get (Querying a log topic) lts:topics:create (Creating a log topic) lts:topics:list (Querying log topics) lts:topics:delete (Deleting a log topic) lts:*:* (Performing all operations on host groups)

Scenario	Delegated Object	Custom Policy	Minimal Permissions
			lts:groups:create (Creating a log group) lts:groups:get (Querying details about a log group) lts:groups:delete (Deleting a log group) aom:*:* (Full permissions for AOM) apm:icmgr:* (Full permissions for the APM collection component)
	ECS	ECS ObsMigrationAgencyPolicy	apm:icmgr:* (Full permissions for the APM collection component)

Creating a Storage Migration Workflow

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Creating a storage migration workflow	MgC	-	OMS Administrator (system-defined role, the administrator role of the OMS)

Importing RVTools Data

Scenario	Delegated Object	Custom Policy	Minimal Permissions
Importing RVTools data	MgC	MgC OfflineCollectionAgencyPolicy	obs:object:GetObject (Obtaining object content and metadata) obs:bucket:ListBucket (Listing objects in a bucket) obs:bucket:ListAllMyBuckets (Querying buckets)

1.4 IAM User Permissions

With IAM, you can grant fine-grained permissions to IAM users under your account by configuring permission policies.

The following table lists the permissions that are required for IAM users to use different MgC functions. For details about how to configure permission policies, see **Creating a Custom Policy**.

Function	Permission
Configuring a server purchase template	vpc:vpcs:list (Listing VPCs) vpc:subnets:get (Querying subnets or querying details about a subnet) vpc:publicIps:list (Querying EIPs) vpc:securityGroups:get (Querying security groups or querying details about a security group) eps:enterpriseProjects:list (Listing enterprise projects) ecs:availabilityZones:list (Listing AZs)
Creating a server migration workflow	vpc:vpcs:list (Listing VPCs) vpc:vpcs:get (Querying VPC details) vpc:subnets:get (Querying subnets or querying details about a subnet) vpc:publicIps:list (Querying EIPs) vpc:publicIps:get (Querying details about an EIP) vpc:securityGroups:get (Querying security groups or querying details about a security group) eps:enterpriseProjects:list (Listing enterprise projects) eps:enterpriseProjects:get (Querying details about an enterprise project)
Getting server recommendations	ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information) ecs:cloudServers:list (Querying details about ECSs) ecs:cloudServers:showServer (Query details about an ECS) ecs:flavors:get (Querying ECS flavors) ims:images:list (Querying images) ims:images:get (Querying details about an image) evs:volumes:list (Querying EVS disks) evs:types:get (Querying EVS disk types)
Creating a Cross-AZ migration workflow	ecs:availabilityZones:list (Listing AZs)

Function	Permission
Configuring Product Mappings for TCO Analysis	ecs:cloudServerFlavors:get (Querying details about flavors and extended flavor information) ims:images:list (Querying images) evs:types:get (Querying EVS disk types)
Creating a storage migration workflow	Tenant Guest (read-only permissions for all cloud services except IAM) OMS Administrator (full permissions for OMS)
Creating a migration cluster	Tenant Guest (read-only permissions for all cloud services except IAM) OMS Administrator (full permissions for OMS) nat:natGateways:list (Querying NAT gateways) smn:topic:list (Querying a topic)
Creating an Agency	iam:agencies:listAgencies (Querying agencies based on specified conditions) iam:roles:listRoles (Listing permissions) iam:quotas:listQuotas (Listing quotas) iam:permissions:listRolesForAgency (Listing permissions of an agency) iam:agencies:createAgency (Creating an agency) iam:permissions:grantRoleToAgency (Granting specified permissions to an agency) iam:roles:createRole (Creating a custom policy) iam:roles:updateRole (Modifying a custom policy)

2 Settings Management

2.1 Managing Migration Projects

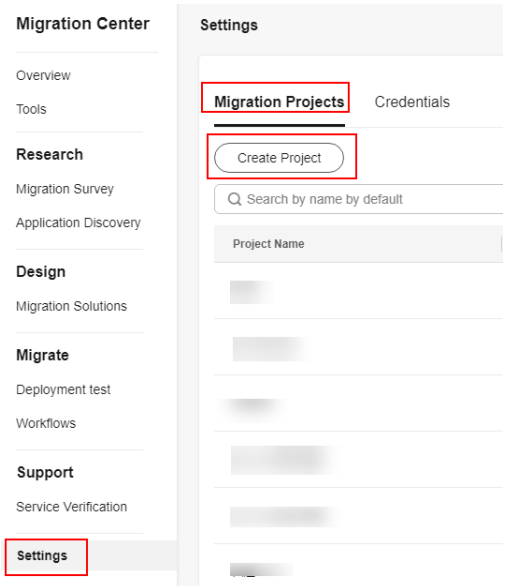
Scenarios

You can create a migration project to isolate and manage resources. For example, you can create a project for migration from Alibaba Cloud to Huawei Cloud. The information about your servers, databases, and components on Alibaba Cloud and the migration progress of these resources will be stored in this project.

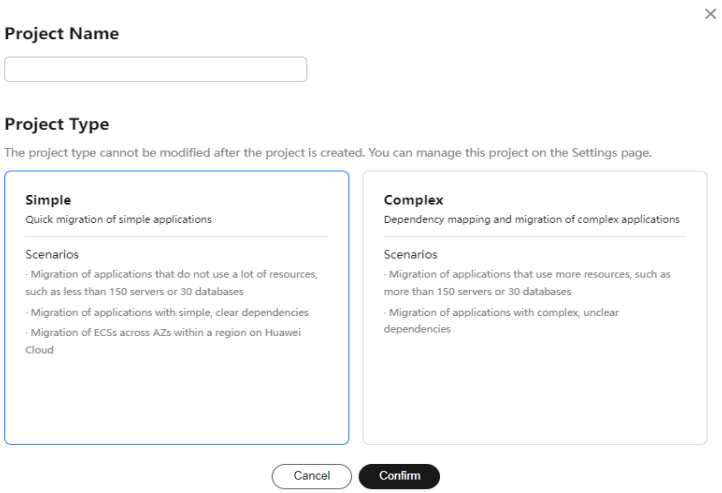
Creating a Migration Project

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Settings**. Under **Migration Projects**, click **Create Project**.



Step 3 In the displayed dialog box, enter a project name, select a project type, and click **Confirm**. The project is created and you can view it in the migration project list.

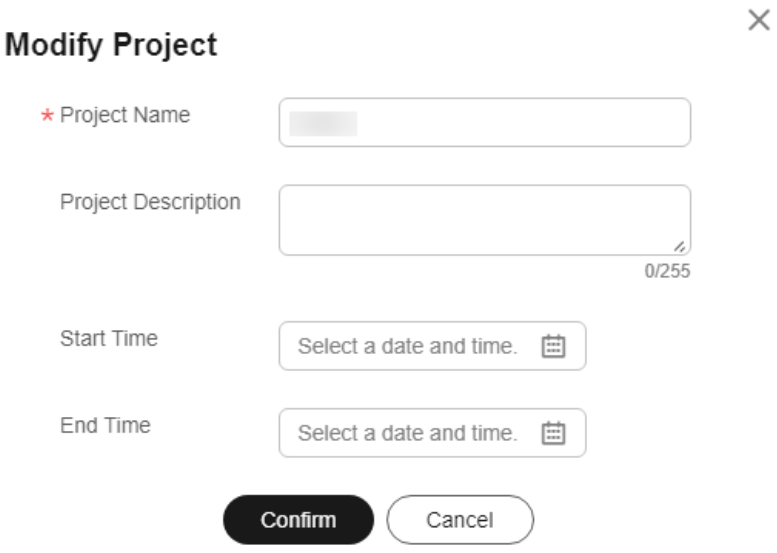


----End

Modifying a Migration Project

Step 1 Locate the project and click **Modify** in the **Operation** column.

Step 2 In the displayed dialog box, modify the project name, project description, start time, and end time, and click **Confirm**.



----End

Archiving a Migration Project

Archived projects will not be displayed in the migration project drop-down list on the top of the page. Locate the project and click **Archive** in the **Operation** column to archive the project. If you want to display an archived project in the drop-down list, you need to unarchive it first.

NOTE

The migration project being currently used cannot be archived.

2.2 Managing Credentials

Credentials are used by MgC to authenticate your identity during resource discovery and data collection. This section describes how to add credentials required by MgC to discovering cloud resources. To learn how to add credentials required by Edge to discover on-premises resources or required for manually adding resources to MgC, see **Adding Resource Credentials**.

Supported Authentication Types

Currently, only credentials of public cloud resources can be added. For details about the supported authentication types, see **Table 2-1**.

Table 2-1 Supported authentication types for public cloud resources

Resource Type	Authentication Type	Description
Public cloud	AK/SK	AK/SK pairs of cloud platforms, such as Huawei Cloud, Alibaba Cloud, AWS, Tencent Cloud, Qiniu Cloud, and Kingsoft Cloud
	Configuration file	Credential configuration files of Google Cloud. A configuration file that contains credentials for Google Cloud service accounts can be uploaded to MgC. The file must be in JSON format and cannot exceed 4 KB.
	ID/Secret	Azure credentials. To learn how to obtain Azure credentials, see **How Do I Obtain Azure Credentials?**

CAUTION

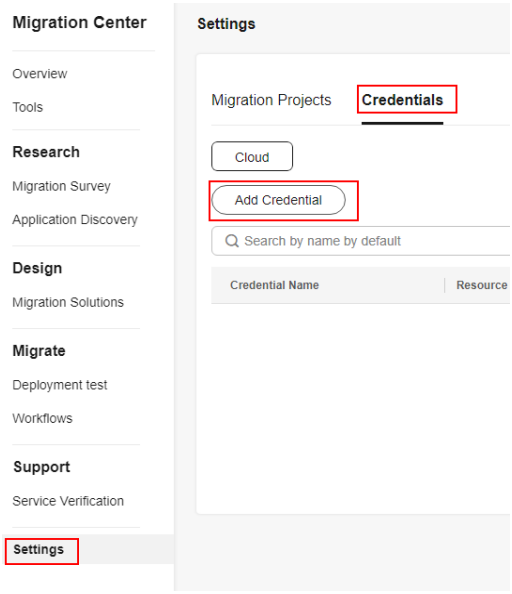
Your cloud credentials are stored in the MgC database and remain valid for seven days. After this period, you need to add the credentials to MgC again if you still want to discover or migrate your cloud resources.

Procedure

Step 1 Sign in to the **MgC** console.

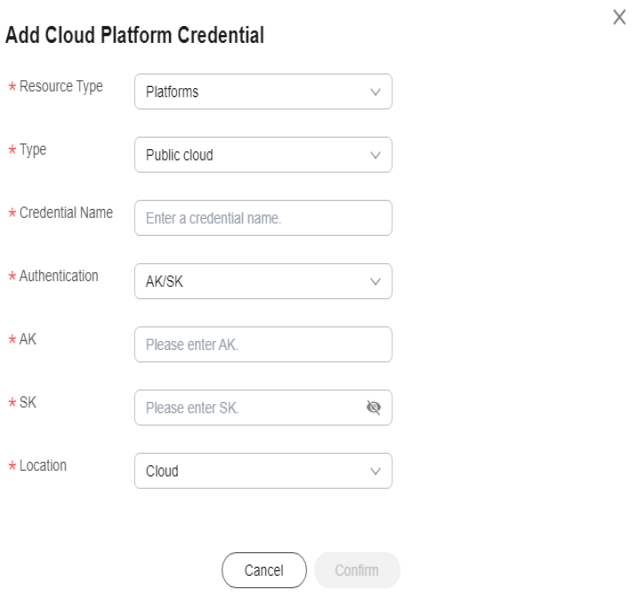
Step 2 In the navigation pane on the left, choose **Settings**. Select a **migration project** in the upper left corner of the page.

Step 3 On the **Credentials** tab, click **Add Credential**.



Step 4 Select a resource type and authentication type as prompted. Specify a credential name, enter or upload your credential, and click **Confirm**.

After the credential is added, you can choose **Credentials** > **Cloud** to view the added credential.



----End

3 Migration Surveys

3.1 TCO Analysis (Credential)

You can share your source cloud credentials, and MgC will analyze your source bills, identify what and how many source resources you used, match the source resources with Huawei Cloud alternatives, and then compare the cost of your source environment with the cost on Huawei Cloud.

The supported source clouds include **AWS** and **Alibaba Cloud**.

NOTICE

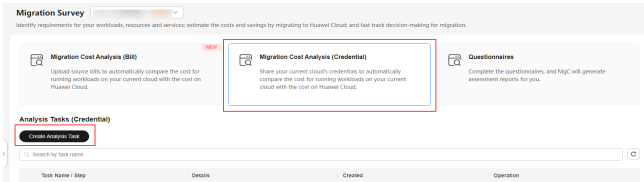
A maximum of 9,999 servers can be included in an analysis task.

Creating an Analysis Task

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Surveys**. Select a **migration project** in the upper left corner of the page.

Step 3 Click the **TCO Analysis (Credential)** card and then click **Create Analysis Task**.



Step 4 Configure the parameters listed in **Table 3-1**.

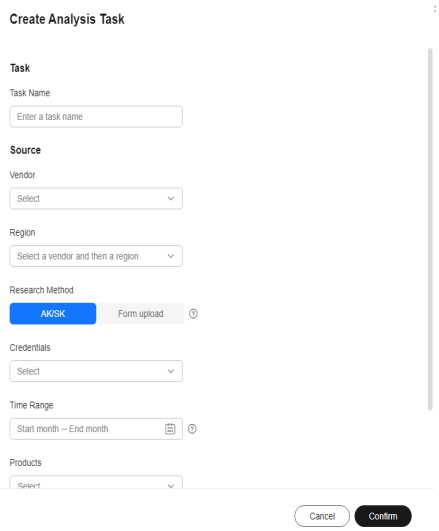


Table 3-1 Parameters required for creating an analysis task

Parameter	Description
Task Name	Enter a task name.
Vendor	Select the source cloud vendor. **NOTE** Currently, AWS and Alibaba Cloud can be chosen.
Region	Select the region where your source resources are located.
Research Method	• If you select **AK/SK**, specify **Credential**, **Time Range**, and **Products**. If you have not add the source cloud credential to MgC, **add it** by choosing **Create** from the credential drop-down list. Configure the parameters for adding a credential and click **Verify and Save**. **NOTE** – Enter the AK/SK pair of the source cloud account when you add the credential to MgC. – The source account must have read-only permissions. – Fees for API requests may be incurred. • If you select **Form upload**, download the template, fill it out, and upload the file. **NOTICE** A maximum of 9,999 servers can be contained in the file.
Target Region	Select the region you want to migrate to.

Step 5 Click **Confirm**.

After the task is submitted, it will be displayed in credential-based analysis task list. You can click the task name to view the task progress. After the task is finished, you can **download bills**, **view product mappings**, and **obtain analysis results** for further analysis.

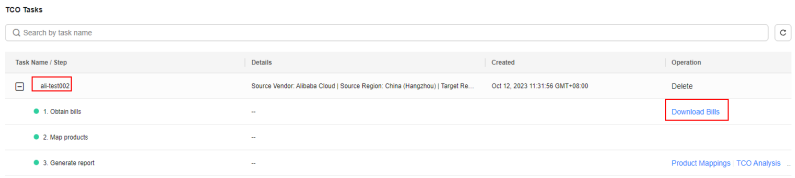
----End

Downloading Bills

MgC analyzes your source bills for the specified period; identifies the billing items; and obtains the specifications, monthly usage, and monthly consumption of each product you used. You can download your bills after the analysis is finished.

Step 1 In the credential-based analysis task list, click the task name.

Step 2 Click **Download Bills** in the **Operation** column. The bills will be downloaded as an Excel file.



----End

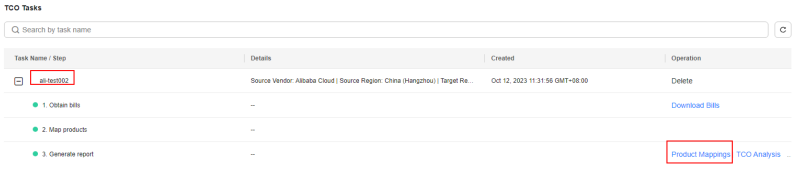
Viewing Product Mappings

MgC automatically recommends Huawei Cloud alternatives based on the specifications and usage of source resources. The recommendations prioritize cost-effectiveness while maintaining the current specifications.

You can view the product mapping details and adjust the mappings as needed.

Step 1 In the credential-based analysis task list, click the task name.

Step 2 Click **Product Mappings** in the **Operation** column.



Step 3 Click on the left of a resource type to view the mapping details.

NOTE

- Mapping refers to the aligning of source products to target products.
- Standard mapping means a source product is mapped to the equivalent Huawei Cloud product.
- Non-standard mapping means a source product is mapped to a Huawei Cloud product that is closest to the source product. You can change the mapping if needed.

----End

Viewing Analysis Results

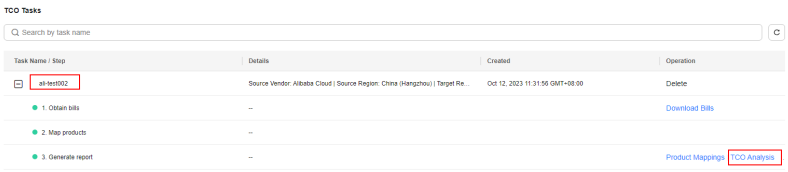
You can view how much you would save if you migrate to Huawei Cloud.

You can also compare the average monthly costs of a certain type of resource between the source cloud vendor and Huawei Cloud. In addition, you can get

discounts for each type of resource from your sales manager to save more money on Huawei Cloud.

Step 1 In the credential-based analysis task list, click the task name.

Step 2 Click **TCO Analysis** in the **Operation** column.



Step 3 View **Analysis Details** and **Price Differences**, and adjust the estimated costs of Huawei Cloud resources based on the discounts you got.

Click **Export Report** in the upper right corner. The analysis report will be exported to a PDF file.

----End

3.2 TCO Analysis (Bill)

You can upload your source bills to MgC, and it will analyze the bills, identify what and how many source resources you used, match the source resources with Huawei Cloud alternatives, and then compare the cost of your source environment with the cost on Huawei Cloud.

Notes and Constraints

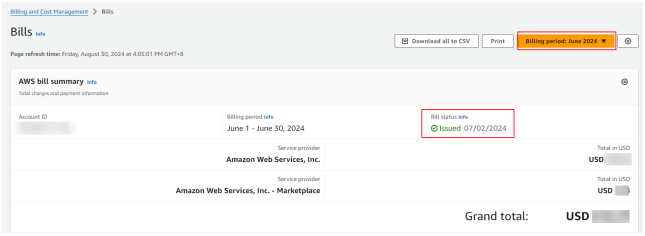
- The supported source cloud is AWS.
- The system can map 17 AWS products such as EC2, RDS, and S3 to Huawei Cloud alternatives. Any products not supported will be excluded from the analysis.
  - Amazon EC2 (Compute)
  - Amazon EC2 (EBS)
  - Amazon EC2 (NAT Gateway)
  - Amazon RDS (MySQL)
  - Amazon RDS (Aurora MySQL)
  - Amazon EKS (Fargate)
  - Amazon ECS (Fargate)
  - Amazon DataTransfer
  - Amazon VPC (VPN)
  - Amazon VPC (VPCEP)
  - Amazon ElastiCache
  - Amazon DocumentDB
  - Amazon ELB
  - Amazon S3
  - Amazon DynamoDB

  - Amazon CloudWatch
  - Amazon ECR
- A maximum of 100 analysis tasks can be created in each MgC migration project, and the task names must be unique.
- The bill file to be uploaded must meet the following requirements:
  - The file must be a PDF file in English or Portuguese. Image PDF files are not supported.
  - The file size cannot exceed 50 MB.
  - The file name cannot contain spaces.

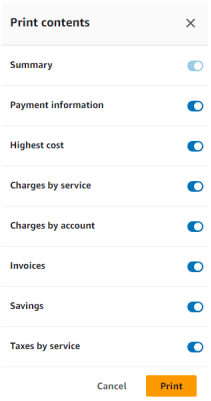
Preparations

- **Create a migration project**.

  On the MgC console, create a migration project. For details, see **Managing Migration Projects**.
- **Download AWS bills.**
  a. Open the **AWS Billing and Cost Management console**.
  b. On the **Bills** page, select the month for **Billing period**. Under the AWS bill summary section, confirm that the **Bill status** appears as **Issued**.



  c. Click **Print** in the upper right corner of the page, select the content to be printed, and click **Print**. **Highest cost** and **Charges by service** are mandatory.



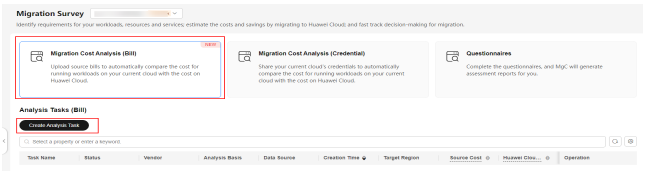
  d. In the displayed dialog box, select **Save as PDF** and click **Save**.

Creating an Analysis Task

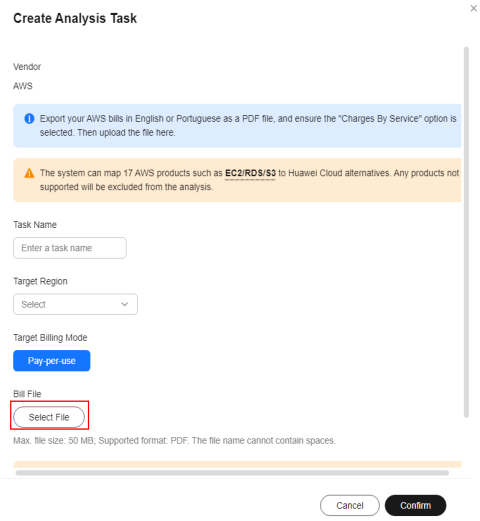
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Surveys**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 Click the **TCO Analysis (Bill)** card and click **Create Analysis Task**. If you are using this function for the first time, you need to delegate MgC certain permission. After you agree to the authorization, the **Create Analysis Task** page is displayed on the right.



Step 4 Enter a task name, select the region you want to migrate to, and click **Select File** to upload the exported bill file to MgC.



Step 5 Click **Confirm**. After the task is submitted, you can view the task status, total source cost, and total Huawei Cloud cost.

When the task status changes to **Completed**, you can perform the following operations:

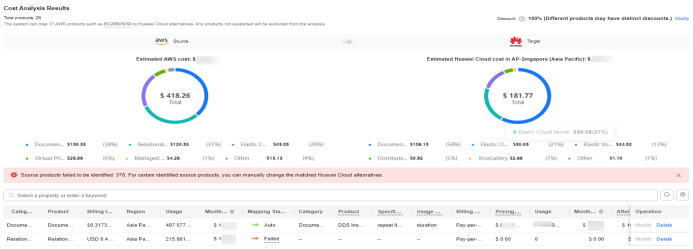
- **View product mappings**. You can view what and how many source products you used, whether there are Huawei Cloud alternatives, and what are the alternatives.
- **Configuring discount assumptions** to adjust the prices of Huawei Cloud alternatives and further optimize the cost on Huawei Cloud.
- **Export cost analysis results**. You can download the analysis results to your local PC for manual adjustment and analysis.

----End

Viewing Product Mappings

After the cost analysis task is complete, you can view the types and quantities of all products identified from the source bills, whether there are Huawei Cloud alternatives, and what are the alternatives.

In the bill-based analysis task list, click the task name in the **Task Name** column. On the task details page, view the product mapping details.



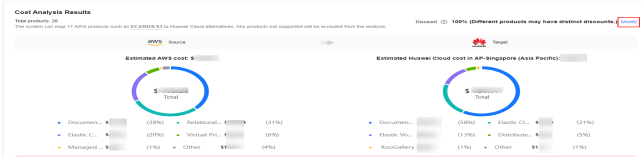
For source products that cannot be automatically mapped to Huawei Cloud alternatives, you can **export the analysis results**. Then manually enter the Huawei Cloud product specifications and usage based on the pricing model provided on the Huawei Cloud official website to estimate the final cost.

Configuring Discount Assumptions

You can configure discount assumptions for individual Huawei Cloud products to save more money. For more detailed information on available discounts, you can contact the Huawei Cloud commercial representative.

Step 1 In the bill-based analysis task list, click the task name in the **Task Name** column.

Step 2 In the upper right corner of the page, click **Modify** next to **Discount**. Then, configure discount assumptions for each product to optimize your savings.



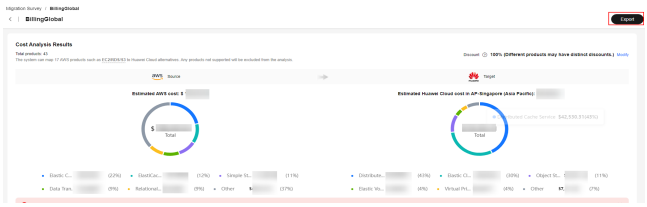
----End

Exporting Cost Analysis Results

After a cost analysis task is complete, you can download the analysis results to your local PC for manual adjustment and analysis.

- **Method 1**: In the bill-based analysis task list, locate the task and click **Export** in the **Operation** column. Then the analysis results will be downloaded as an Excel file.

- **Method 2**: On the task details page, click **Export** in the upper right corner of the page. Then the analysis results will be downloaded as an Excel file.



3.3 Assessing Migration Motivations

Scenarios

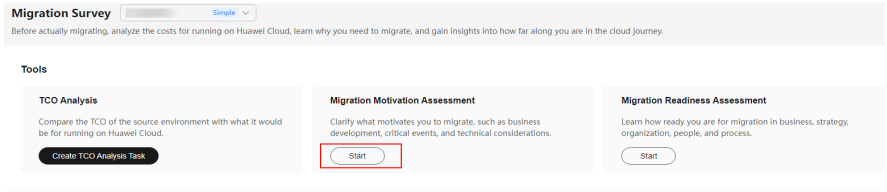
A motivation assessment helps you plan more feasible cloud migration paths and solutions.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Surveys**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Start** in the **Migration Motivation Assessment** pane.



Step 4 Answer all questions and click **Submit**.

Then MgC generates and downloads the assessment report on your motivations for cloud migration. You can get Huawei Cloud suggestions on your migration to the cloud in the report.

----End

3.4 Assessing Migration Readiness

Scenarios

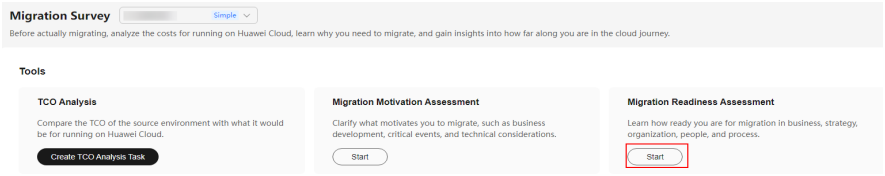
A cloud readiness assessment helps you learn how you are ready for migration in terms of business, strategy, organization, finance, and security. It helps you design a complete migration path and specific project plans.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Surveys**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Start** in the **Migration Readiness Assessment** pane.



Step 4 Answer all questions and click **Submit**.

Then MgC generates and downloads the assessment report on your readiness for cloud migration. You can get Huawei Cloud suggestions on your migration to the cloud in the report.

----End

4 Resource Discovery

4.1 Complex Project

4.1.1 Discovering Resources over the Internet

Before migrating, you need to discover your resources and collect their details. This section describes how to discover resources provisioned on cloud platforms such as Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Google Cloud, Azure, Qiniu Cloud, and Kingsoft Cloud over the Internet. Then you can organize the associations between resources and applications. The resource types supported by MgC depend on the source cloud platform.

- Alibaba Cloud and Huawei Cloud: servers, containers, middleware, databases, networks, and storage
- Tencent Cloud: servers, databases, and storage
- AWS , Google Cloud, and Azure: servers, containers, middleware, databases, storage, and networks
- Qiniu Cloud and Kingsoft Cloud: object storage

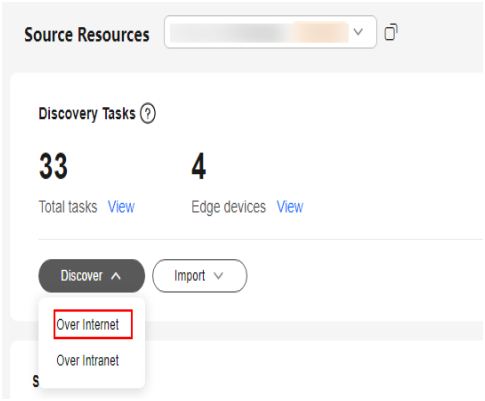
Creating a Discovery Task

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Discover** > **Over Internet**.

Figure 4-1 Discovery over Internet



Step 4 Configure the parameters listed in **Table 4-1**.

Table 4-1 Parameters for creating an Internet-based discovery task

Area	Parameter	Description	Mandatory
Basic Settings	Task Name	Specify a name for the task.	Yes
	Task Description	Describe the task.	No
Task Settings	Source Platform	Select the source cloud platform. Currently, Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Google Cloud, Azure, Qiniu Cloud, and Kingsoft Cloud are supported.	Yes

Area	Parameter	Description	Mandatory
	Credential	Select the credential for accessing the source cloud platform. If no credential is available, choose **Create** to add one. For details, see **Managing Credentials**. • If the source cloud platform is Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Qiniu Cloud, or Kingsoft Cloud, select **AK/SK** for **Authentication** and enter the AK/SK pair of your source cloud account. • If the source cloud platform is Google Cloud, select **Configuration File** for **Authentication** and upload the configuration file that contains your Google Cloud service account credentials. The file must be in JSON format and cannot exceed 4 KB. • If the source cloud platform is Azure, Select **ID/Secret** for **Authentication**. To learn how to obtain Azure credentials, see **How Do I Obtain Azure Credentials?**	Yes
	Region	Select the region where your source environment is located. Multiple regions can be selected.	Yes

Step 5 Enable **Cloud Platform Collection**, choose the resource types (collection items) to be collected from the **Resource Type** drop-down list. For resource types supported for each cloud platform, see **Table 4-2**.

Figure 4-2 Selecting resource types

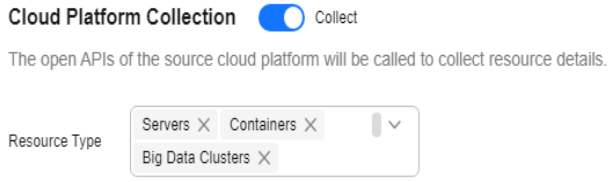


Table 4-2 Supported resource types

Cloud Platform	Resource Type	Subtype
Alibaba Cloud	• Servers • Containers • Big data clusters • Databases	-
	Middleware	• Redis • Kafka
	Storage	• Object storage • File storage
	Networks	• Cloud connections • Load balancers (ALB and CLB) • Private lines • Public domain names • Private domain names • EIPs • Public NAT gateways • Route tables • Security groups • VPCs • VPN gateways
Huawei Cloud	• Servers • Containers • Big data clusters • Databases	-
	Middleware	• Redis • Kafka
	Storage	• Object storage • File storage
	Networks	• Load balancers (ELB) • Public domain names • Private domain names • EIPs • Public NAT gateways • Route tables • Security groups • VPCs

Cloud Platform	Resource Type	Subtype
AWS	• Servers • Containers • Databases	-
	Middleware	• Redis • Kafka
	Networks	• Load balancers (ELB) (Currently, only CLB balancers are supported.) • Public domain names • Private domain names • EIPs • Public NAT gateways • Route tables • Security groups • VPCs
	Storage	• Object storage • File storage
Tencent Cloud	• Servers • Databases	-
	Storage	• Object storage • File storage
Google Cloud	• Servers • Containers • Databases	-
	Middleware	Redis
	Storage	• Object storage • File storage
	Networks	• EIPs • Route tables • Security groups • VPCs
Azure	• Servers • Containers • Databases	-
	Storage	• Object storage • File storage

Cloud Platform	Resource Type	Subtype
	Middleware	• Redis • Kafka
	Networks	• EIPs • Route tables • Security groups • Public NAT gateways • VPCs • Load balancers
Qiniu Cloud	Storage	Object storage
Kingsoft Cloud	Storage	Object storage

Step 6 (Optional) In the **Application Discovery** area, select the collection items to identify invocation chains between resources and applications and map application dependencies.

NOTICE

Before selecting a collection item, ensure that the account of the provided credential has the permissions for accessing the collection item.

The following table lists what information each collection item can provides.

Collection Item	Function
Resource Management	Collecting data from this service helps you find out how different applications and resources are connected, which helps prevent disruptions in their relationships.
DNS	Collecting DNS data helps gain insights into how different applications interact at the access layer.
WAF	Collecting WAF data helps gain insights into application traffic flows.
LB	Collecting LB data helps gain insights into applications traffic flows.
MSE	Collecting MSE data helps analyze how different applications and microservices interact.

Step 7 Click **Confirm**. After this Internet-based discovery task is created, MgC starts discovering details about resources and applications.

- You can **view the list and details of discovered resources** on the **Resource** page.
- You can **view the task status and details** on the task list page.

Step 8 Wait until the task status changes to **Succeeded**. Then perform subsequent operations for different types of resources.

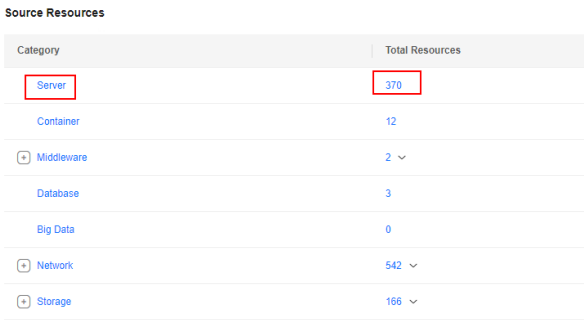
- For discovered servers, **perform a deep collection** or **check their migration readiness** and then design a **migration solution** or **migration plan**.
- For discovered containers, **perform a deep collection**and **design a migration solution**.
- For discovered middleware, databases, and storage resources, **design migration solutions Deep collection** is available for AWS RDS (for MySQL, MariaDB, Aurora, PostgreSQL, SQL Server and Oracle) and AWS DocumentDB databases.
- For discovered object storage resources, perform a **deep collection** and configure **a migration plan**.

----End

Check Migration Readiness for Servers

Perform the following steps to check whether your source servers meet the migration requirements:

1. **Install Edge** in the source intranet environment and connect Edge to MgC.
2. In the **Resources** area, click **Server** in the **Category** column or the number in the **Total Resources** column.



3. Select the servers to migrate together and click **Group as Application** above the list.
   - If you have **created an application**, select the application from the drop-down list and click **OK**.
   - If you have not created an application, click **Create Application** in the displayed dialog box. Then enter an application name and description, select a business scenario, environment, and target region, and click **Create**. Then click **OK**.
4. On the top of the server list, choose **Migration Scenario** > **Server migration**.
5. For each source server, click **Configure** in the **Migration Readiness** column.

6. Configure the parameters listed in **Table 4-3**.

Table 4-3 Parameters for configuring migration readiness

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.
Access IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration.
Port	Enter the port on the source server that allows access from the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC. **NOTICE** The account provided in the credential must have sufficient permissions, so Edge can collect necessary server details. To perform a deep collection for the server to collect as much as details, the credential you provided must meet the following requirements: • Linux: **root** and its password • Windows: **administrator** and its password

7. Click **Confirm**. The system checks whether the source server can be accessed using the information you provided and whether the server can be migrated. If **Ready** shows up in the **Migration Readiness** column, the source server can be migrated. Then you can **design a migration solution** or **design a migration plan** to migrate the server.

Performing a Deep Collection for Servers

Perform the following steps to perform a deep collection for your source servers:

1. Ensure that **Edge** has been installed in the source intranet environment and has been connected to MgC.
2. In the **Resources** area, click **Server** in the **Category** column or the number in the **Total Resources** column.
3. For each server whose details need to be collected, in the **Edge Device** column, click **Configure**. Configure the parameters listed in **Table 4-4**.

Table 4-4 Parameters for configuring a deep collection

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.
Access IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration.
Port	Enter the port on the source server that allows access from the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC. **NOTICE** The account provided in the credential must have sufficient permissions, so Edge can collect necessary server details. To perform a deep collection for the server to collect as much as details, the credential you provided must meet the following requirements: • Linux: **root** and its password • Windows: **administrator** and its password

4. Click **Confirm**. Then the system automatically starts executing a deep collection. When **Collected** shows up in the **Deep Collection** column, the collection is complete. You can **design a migration solution** or **design a migration plan** to migrate your servers.

Performing a Deep Collection for Containers

Perform the following steps to perform a deep collection for your containers:

1. Ensure that **Edge** has been installed in the source intranet environment and has been connected to MgC.
2. In the **Resources** area, click **Container** in the **Category** column or the number in the **Total Resources** column.
3. For each container resource whose details need to be collected, click **Associate** in the **Edge Device** column.

   To associate multiple container resources with one Edge device, select them and choose **Manage Device Association** above the list.
4. Select the Edge device in the source environment and click **OK**. Wait until **Associated** appears in the **Edge Device** column.

5. For each container resource, in the **Credential** column, click **Associate** to configure their credentials.
6. Select the resource credential. If the credential has not been added to MgC, go to the Edge console, **add the credential** to Edge, and synchronize it to MgC.
7. Click **OK**. Then MgC checks whether the resource can be accessed using the associated credential. When **Ready** appears in the **Deep Collection** column, click **Collect** to perform a deep collection. You can click **Collect Again** in the **Deep Collection** column to perform a second deep collection if needed. After the collection is complete, click the resource name to view the collected details.

Performing a Deep Collection for Object Storage Resources

Follow the steps below to perform a deep collection for your object storage resources, so the system can use the collected details to recommend rightsized migration clusters.

1. Ensure that **Edge** has been installed in the source intranet environment and has been registered with MgC.
2. On the **Resource Discovery** page, in the **Resources** area, click **Storage** in the **Category** column or the number in the **Total Resources** column.
3. Locate an object storage resource and click **Configure** in the **Deep Collection** column.
4. Select the Edge device in the source environment and the credential used for accessing the resource, and click **Confirm**. If the credential has not been added to MgC, go to the Edge console and **add the resource credential** to Edge and synchronize it to MgC.

> **NOTICE**
>
> To perform a deep collection for your Azure object storage resources, you need to provide your storage account and access key. For details about how to obtain the credentials, see **How Do I Obtain the Required Credentials Before Using MgC to Perform a Deep Collection for My Azure Object Storage Resources?**

5. Click **Add Prefix** in the **Operation** column.
6. Enter a prefix to filter the objects whose details need to be collected. If this parameter is not specified, all objects in the bucket are collected by default. Click **OK** to save the prefix settings.
7. Click **Deep Collection** in the **Operation** column. The system starts collecting path details. You can perform a deep collection on a resource for multiple times. When **Collection Status** changes to **Completed**, click the resource name to view the collected information.

Performing a Deep Collection for Databases

Deep collection is available for AWS RDS (for MySQL, MariaDB, Aurora, PostgreSQL, SQL Server, and Oracle) and AWS DocumentDB databases. The data

collected includes database version, engine, server character set, average transactions per second (TPS), and queries per second (QPS), and other key performance metrics (KPIs). The collected data depends on the database type.

1. Ensure that **Edge** has been installed in the source intranet environment, can access source databases, and has been registered with MgC.
2. On the **Resource Discovery** page, in the **Resources** area, click **Database** in the **Category** column or the number in the **Total Resources** column.
3. In the database list, filter all discovered AWS databases by applying the **Vendor** filter. Locate a supported database and in the **Edge Device** column, click **Associate**.

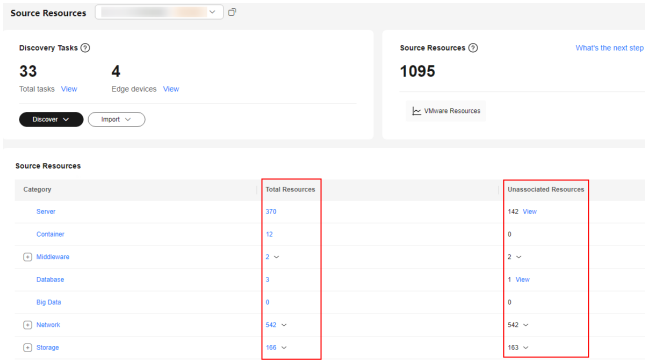
   To associate multiple databases with one Edge device, select them and choose **Manage Device Association** above the list.
4. Select your Edge device. For **Access Setting**, if the selected resource is in the same VPC as the Edge device, select **Private access**. Otherwise, select **Public access**. Then click **OK**. Wait until **Associated** appears in the **Edge Device** column.
5. In the **Credential** column, click **Associate** to associate the credential for accessing the resource.
6. Select the database credential. If the credential has not been added to MgC, go to the Edge console, **add the credential** to Edge, and synchronize it to MgC.
7. Click **OK**. Then MgC checks whether the resource can be accessed using the associated credential. When **Ready** appears in the **Deep Collection** column, click **Collect** to perform a deep collection. You can click **Collect** in the **Deep Collection** column to perform a deep collection again if needed.
8. Wait for the deep collection to complete. Then click the database name to go to the database details page. In the **Database Information** area, you can view the collected details.

Viewing Discovered Source Resources

A discovery task can collect only basic information about source resources. More detailed information needs to be collected through a deep collection.

1. On the **Resource Discovery** page, in the **Resources** area, view the number of discovered resources by category and the number of resources that are not associated with applications.

Figure 4-3 Viewing the resource list

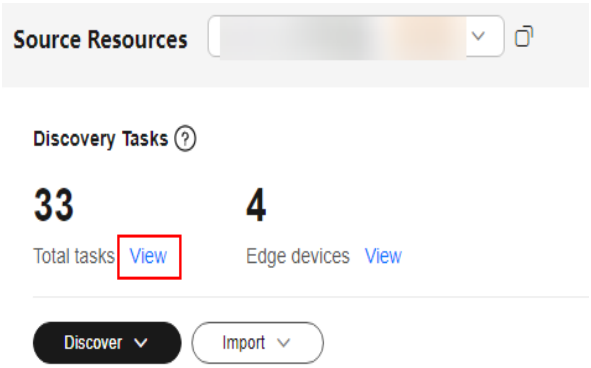


2. Click a resource category or number to view the resource list.
3. Click a resource name to view its details.

Viewing Discovery Tasks

Step 1 On the **Resource Discovery** page, click **View** next to **Total tasks** to go to the task list.

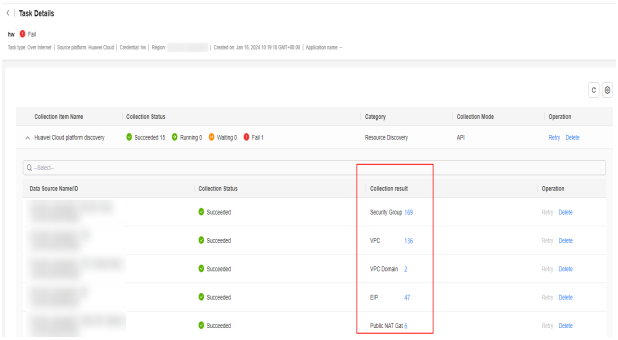
Figure 4-4 Viewing the task list



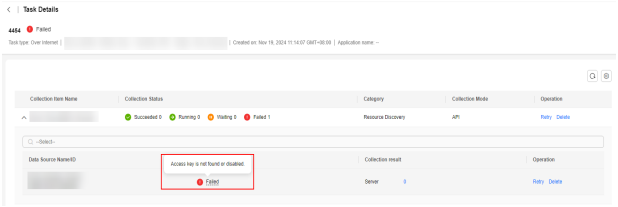
Step 2 Locate a discovery task, and click **View** in the **Operation** column.

Step 3 Click ⌄ before a collection item to view the data sources contained in the collection item and the collection results.

Figure 4-5 Viewing collection results



If a data source fails to be collected, you can move the cursor to the collection status of the data source to view the failure cause.



Step 4 In the **Collection Result** column, you can click the number next to a resource category to view the corresponding resource list.

----End

4.1.2 Discovering Resources over an Intranet

This section describes how to discover servers in on-premises environments. Before getting started, you need to install Edge in the source intranet. Then you can discover servers by network range or VMware host.

Precautions

- Only VMs in VMware vSphere 5.0 to 7.0 can be discovered.
- When the system scans VMware VMs or scans servers on specified network ranges, it uses the servers' private IP addresses and the ID of the used Edge device to identify discovered servers. If a server's private IP address changes after a collection is complete, the server will be identified as a new one during the next collection, and the total number of discovered servers will increase. To avoid this, you are advised not to change private IP addresses of source servers before the migration is complete.

Prerequisites

- You have **installed Edge** in the source intranet environment and have connected the Edge device to MgC.
- You have **added source server credentials** to Edge.

NOTICE

To perform a deep collection for your source servers to collect as much as details, provide server credentials that meet the following requirements:

- Linux: **root** and its password
- Windows: **administrator** and its password

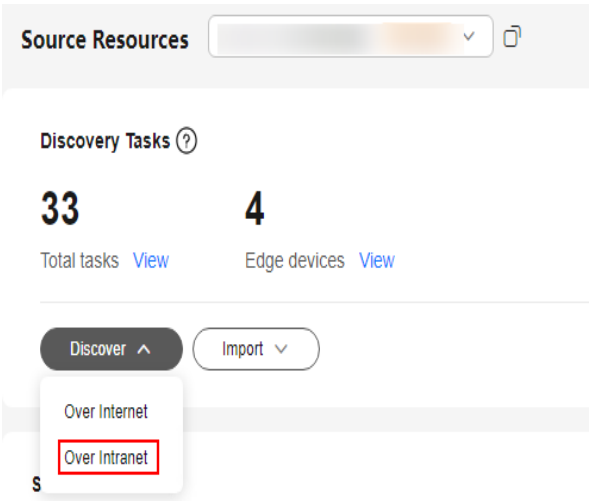
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Discover** > **Over Intranet**.

Figure 4-6 Discovery over an intranet



Step 4 Configure the parameters listed in **Table 4-5**.

Table 4-5 Parameters for creating an intranet-based discovery task

Parameter	Description
Task Name	Enter a task name.
Task Description	Describe the task.
Edge Device	Select the **Edge device** in the source intranet environment.

Step 5 Enable **Scan Network Range** or **Scan VMware VMs** to discover servers as needed.

- If **Scan Network Range** is enabled, configure parameters listed in **Table 4-6**.

  Table 4-6 Parameters for scanning a network range

Parameter	Description
Protocol	Select the communication protocol **TCP** or **ICMP**.
Network Range	Enter an IP address range which must fall within: – 10.0.0.0 – 10.255.255.255 – 172.16.0.0 – 172.31.255.255 – 192.168.0.0 - 192.168.255.255
Linux	Enter the port for scanning Linux servers. This parameter is available only if you choose the TCP protocol. If you need to skip Linux servers during the scan, set this parameter to **0**.

Parameter	Description
Windows	Enter the port for scanning Windows servers. This parameter is available only if you choose the TCP protocol. If you need to skip Windows servers during the scan, set this parameter to **0**.

- If **Scan VMware VMs** is enabled, in the **IP Address** text box, enter the IP address of the vCenter Server where the VMs will be discovered, and select the credential for accessing the vCenter Server. All VMs managed by the vCenter Server will be discovered. If the vCenter Server's credential has not been added to Edge, **add the resource** to Edge. When adding the credential, enter the username and password for logging in to the vCenter Server.

Step 6 Click **Confirm**. The intranet-based discovery task is created, and MgC starts collecting details about source servers.

On the **Source Resources** page, in the **Discovery Tasks** card, click **View** next to **Total tasks**.

Step 7 Wait until the task status changes to **Succeeded**, and perform a deep collection.

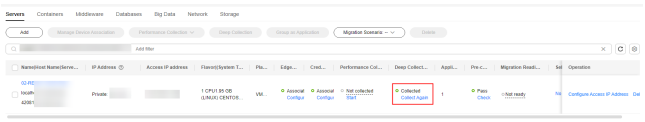
1. On the **Source Resources** page, in the resource list, click **Server** in the **Category** column or the number in the **Total Resources** column.
2. In the **Edge Device** or **Credential** column, click **Configure**.
3. Configure the parameters listed in **Table 4-7**.

Table 4-7 Parameters for configuring a deep collection

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.
Access IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration.
Port	Enter the port on the source server that allows access from the Edge device. – By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. – By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to Edge, go to the Edge console and **add the server credential** to Edge and synchronize it to MgC.

4. Click **Confirm**. Then the system automatically starts executing a deep collection. When **Collected** shows up in the **Deep Collection** column, the collection is complete. You can **design a migration solution** or **design a migration plan** to migrate your servers.

Figure 4-7 Discovery completed



----End

4.1.3 Importing Application Associations

You can import application details obtained from the configuration management database in your source environment to MgC. Once imported, MgC will analyze these details to identify application dependencies and associations.

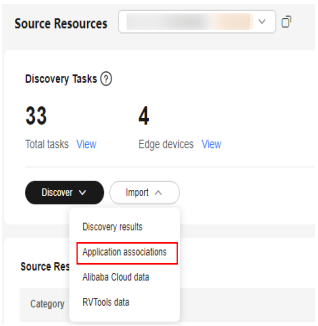
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Import** > **Application associations**. Download the template for importing application associations.

Figure 4-8 Downloading the template for importing application associations



Step 4 Populate the template with your application details and save the file. In the file, fields highlighted in yellow are mandatory. Then click **Select File** to upload the saved file to MgC.

Step 5 Click **Confirm**.

You can view the task in the task list.

- If the task status is **Failed**, click **View** in the **Operation** column to view the data source that failed to be collected. You can move the cursor to the collection status to view the failure cause. After handling the failure cause, you need to delete the failed task and perform a second import.

- If the task status is **Succeeded**, you can go to the **Applications** page to **view the organized application dependencies**.

----End

4.1.4 Importing Discovery Results

You can use offline collectors to discover your resources and applications and collect their details, and import the collected details to MgC for automated application dependency analysis.

Prerequisites

You have **created collection tasks using offline collectors** and have obtained the collection results.

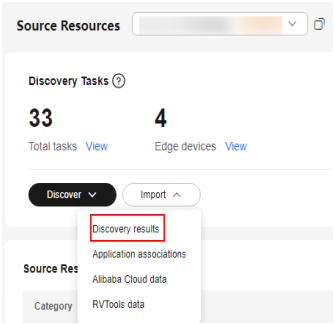
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Import** > **Discovery results**.

Figure 4-9 Importing discovery results



Step 4 In the **Import Discovery Results** dialog box, click **Select File** to upload the JSON file that stores the discovery results to MgC.

Step 5 Click **Confirm**.

You can view the task in the task list.

- If the task status is **Failed**, click **View** in the **Operation** column to view the data source that failed to be collected. You can move the cursor to the collection status to view the failure cause. After handling the failure cause, you need to delete the failed task and perform a second import.
- If the task status is **Succeeded**, you can go to the **Applications** page to **view the organized application dependencies**.

----End

4.1.5 Importing Details of Alibaba Cloud Resources

You can import CSV-formatted resource details from Alibaba Cloud directly into MgC. You first need to export the resource details from Alibaba Cloud. After the

import is complete, you can view the resource details on the MgC console. With the imported details, you can get recommendations for suitable Huawei Cloud resources.

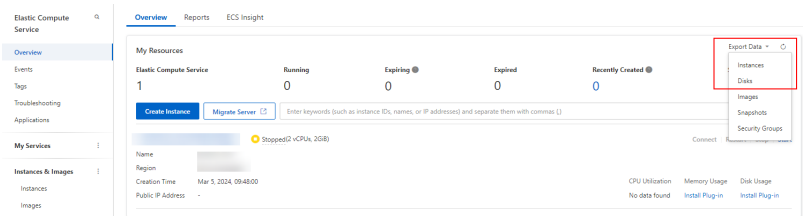
The supported Alibaba Cloud resources are:

- Elastic Compute Service (ECS) instances and cloud disks
- Object Storage Service (OSS) buckets
- ApsaraDB RDS for MySQL instances
- ApsaraDB RDS for GaussDB instances

Exporting Data of Alibaba Cloud ECS Instances and Disks

Step 1 Sign in to the Alibaba Cloud ECS console.

Step 2 On the **Overview** page, choose **Export Data** > **Instances** to export a list of all ECS instances.



Step 3 Choose **Export Data** > **Disks** to export a list of all cloud disks.

Step 4 Open the instance list and the disk list and adjust the resource information if needed. Ensure that the lists contain the necessary information.

List	Mandatory Fields
Instance list	Instance ID, OS, instance specifications, CPU, and memory
Disk list	Instance ID, disk ID, capacity (GiB), and disk attributes

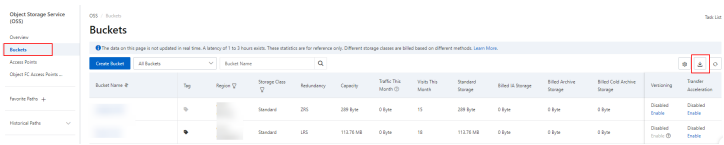
----End

Exporting Data of Alibaba Cloud OSS Buckets

Step 1 Sign in to the OSS console.

Step 2 In the left navigation pane, choose **Buckets**.

Step 3 Click the  icon in the upper right corner of the list. The **Export Bucket List** dialog box is displayed.



Step 4 Select the information to be exported and click **OK**.

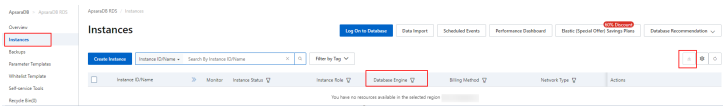
----End

Exporting Data of Alibaba Cloud ApsaraDB RDS for MySQL Instances

Step 1 Sign in to the Alibaba Cloud ApsaraDB RDS for MySQL console.

Step 2 In the navigation pane on the left, choose **Instances**. Select a region in the upper part of the page.

Step 3 Filter all MySQL instances in the **Database Engine** column and click the  icon in the upper right corner of the list.



Step 4 In the displayed dialog box, select the information to be exported. Click **OK**. The system exports the instance information to a CSV file and automatically downloads the file.

Step 5 Modify the exported CSV file so MgC can parse the file content.

1. Use a file editor to open the CSV file and search for **PrivatePort(A network port,)** in the table header.
2. Delete the comma (,) from this field and save the file. The correct format is **PrivatePort(A network port)**.

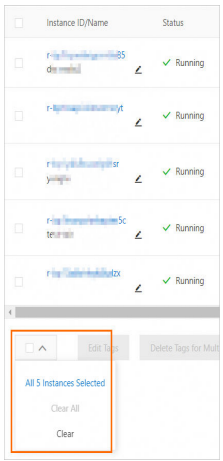
----End

Exporting Data of Alibaba Cloud ApsaraDB RDS for Redis Instances

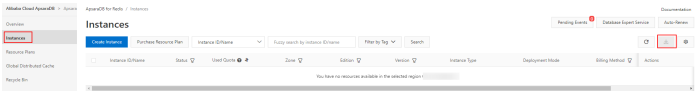
Step 1 Sign in to the Alibaba Cloud ApsaraDB for Redis console.

Step 2 In the navigation pane on the left, choose **Instances**. Select a region in the upper part of the page.

Step 3 In the lower left corner of the page, expand and click **Select All**.



Step 4 Click the  icon in the upper right corner of the list. In the dialog box that is displayed, select the information to be exported.



Step 5 Click **OK**. The system exports the instance information to a CSV file and automatically downloads the file.

Step 6 Modify the exported CSV file so MgC can parse the file content.

1. Use a file editor to open the CSV file.
2. Add a comma (,) at the end of the table heading line and save the file.

   The following shows an example table header before modification:

   ```
   ID,Name,Quantity
   ```

   The following shows the table header after modification:

   ```
   ID,Name,Quantity,
   ```

----End

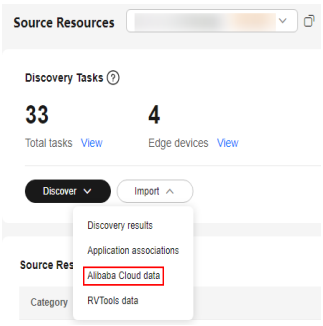
Importing Details of Alibaba Cloud Resources

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Import** > **Alibaba Cloud Data**.

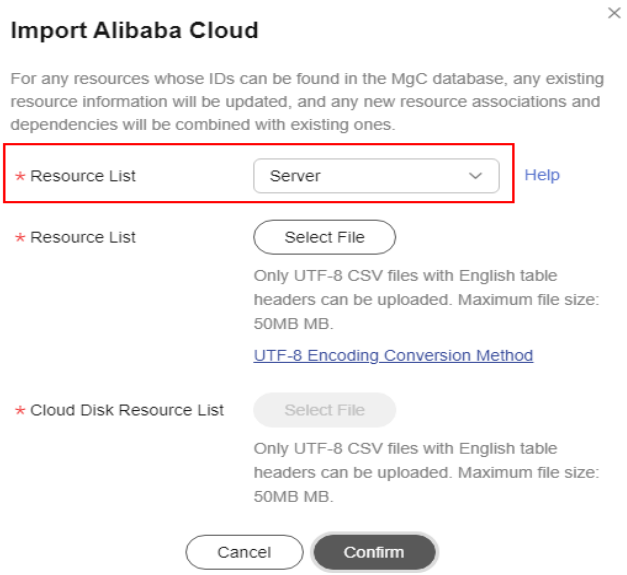
Figure 4-10 Importing details of Alibaba Cloud resources



NOTICE

Only CSV files encoded in UTF-8 format with English table headers can be imported. The size of a single file cannot exceed 15 MB. For details about how to convert the encoding format of a CSV file to UTF-8, see **How Do I Convert the Encoding Format of a CSV File to UTF-8?**

Step 4 In the **Resource Type** drop-down list, select the types of resources to be imported and upload the corresponding file.



Resource Type	Procedure
Servers	1. Click **Select File** next to **Resource List** and select the **exported ECS list**. 2. Click **Select File** next to **Cloud Disk List** and select the **exported disk list**.
Redis	Click **Select File** next to **Resource List** and select the **exported Redis instance list**.
Object storage	Click **Select File** next to **Resource List** and select the **exported OSS bucket list**.
Databases	Click **Select File** next to **Resource List** and select the **exported MySQL instance list**.

Step 5 Click **Confirm**. After the resource list is uploaded successfully, select the corresponding resource categories to view the imported resource information.

Step 6 (Optional) Group imported Alibaba Cloud resources as applications. For details, see **Grouping Resources as Applications**. Then assess the applications to obtain recommendations for Huawei Cloud resources. For details, see **Getting Target Recommendations**.

----End

4.1.6 Importing RVTools Data

Notes and Constraints

- **Supported RVTools versions**

  The supported RVTools versions include:

  - 4.4.1
  - 4.4.2

  - 4.4.3
  - 4.4.4
  - 4.4.5
  - 4.5.0
  - 4.5.1
  - 4.6.1
- **File format requirements**

  Only the Excel (.xlsx) format is supported.
- **File size and compression ratio requirements**

  The file to be imported cannot be larger than 100 MB, and the compression ratio cannot be lower than 5%.

Step 1: Exporting RVTools Data

Step 1 Start the RVTools application.

Step 2 On the login page, enter the following information:

- In the **IP address/Name** text box, enter IP address of the vCenter server.
- In the **User name** text box, enter the username for connecting to the vCenter server.
- In the **Password** text box, enter the password corresponding to the username.

Step 3 Click **Login**. In the top menu bar, choose **File** > **Export all to Excel**.

Step 4 Select a path for saving the file that contains details about your resource.

----End

Step 2: Uploading the Exported File to OBS

Step 1 Sign in to Huawei Cloud **OBS console** and create a Standard bucket for storing the file exported from RVTools. For details, see **Creating a Bucket**.

Step 2 Upload the **file exported from RVTools** to the Standard OBS bucket. For details, see **Uploading an Object**.

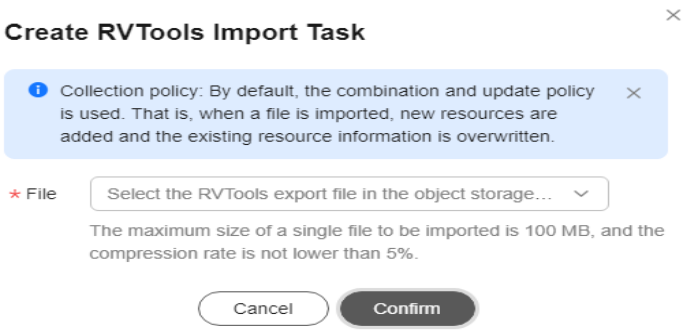
----End

Step 3: Importing RVTools Data

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Import** > **RVTools data**.

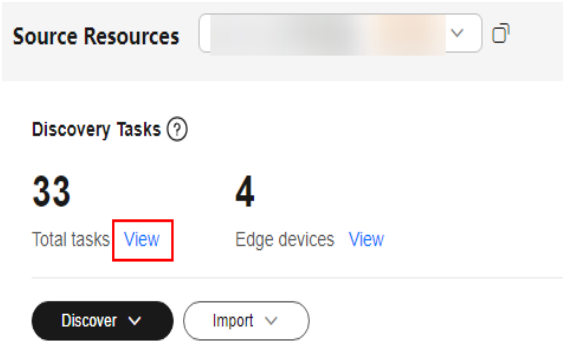


NOTE

If you are using this function for the first time, click **OK** in the **Authorization Required** dialog box to agree to the agency-based authorization.

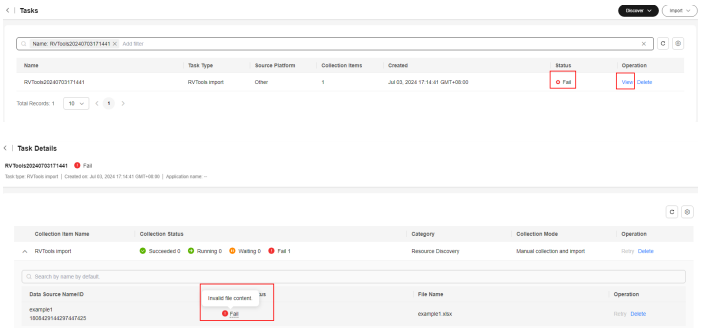
Step 4 In the **File** drop-down list, locate the OBS bucket you used to store the file exported from RVTools in **step 2**, select the exported file, and click **Confirm**.

Step 5 After the upload is successful, in the **Discovery Tasks** card, click **View** next to **Total tasks**. In the task list, view the task status.



- If the task status is **Failed**, click **View** in the **Operation** column to view the data source that failed to be collected. You can move the cursor to the collection status to view the failure cause. For details about how to rectify common failure causes, see **What Can I Do If Importing RVTools Data Fails?**

  After handling the failure cause, you need to delete the failed task and perform a second import.



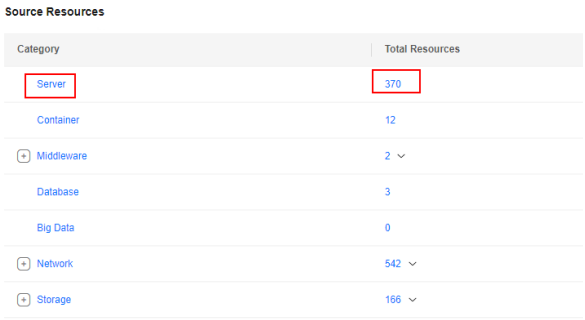
- If the task status is **Succeeded**, you can view the imported **resource details** and **resource statistics**.

----End

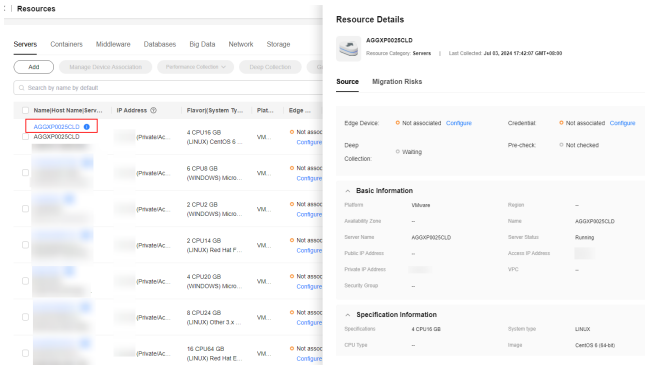
Viewing Resource Details

If the status of the RVTools import task is **Succeeded**, you can view the details of VMware servers, including the hostname, server ID, IP address, specifications, OS type, and platform.

Step 1 On the **Source Resources** page, in the resource list, click **Server** in the **Category** column or the number in the **Total Resources** column.



Step 2 In the server list, you can view the imported server information. You can click a resource name to view more details.



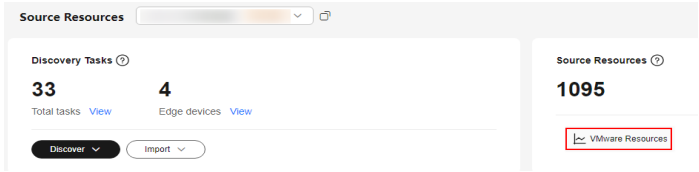
----End

Viewing Resource Statistics

If the status of the RVTools import task is **Succeeded**, you can view the VMware resource statistics under the corresponding tab.

- **Host**: CPU model, host quantity, total memory, allocated VM memory, memory allocation rate, VM quantity, quantity of running VMs, storage usage, and host list.
- VMs: the number of Windows OS versions, number of Linux OS distributions, total memory, total number of vCPU, number of disks, disk capacity, number of VMs, and VM list

Step 1 On the **Source Resources** page, click **VMware Resources** in the **Source Resources** card.



Step 2 On the displayed **Resource Statistics** page, click the **Host** or **VM** tab to view resource statistics.

----End

4.2 Simple Project

4.2.1 Discovering Resources over the Internet

Before migrating, you need to discover your resources and collect their details. This section describes how to discover resources provisioned on cloud platforms such as Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Google Cloud, Azure, Qiniu Cloud, and Kingsoft Cloud over the Internet. The resource types MgC can discover depends on the cloud platform.

- Alibaba Cloud and Huawei Cloud: servers, containers, middleware, databases, networks, and storage
- Tencent Cloud: servers, databases, and storage
- AWS , Google Cloud, and Azure: servers, containers, middleware, databases, storage, and networks
- Qiniu Cloud and Kingsoft Cloud: object storage

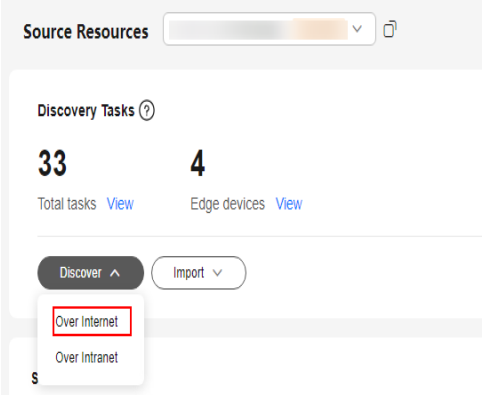
Creating a Discovery Task

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Discover** > **Over Internet**.

Figure 4-11 Discovery over Internet



Step 4 Configure a discovery task based on **Table 4-8**.

Table 4-8 Parameters for creating an Internet-based discovery task

Area	Parameter	Description	Mandatory
Basic Settings	Task Name	Specify a name for the task.	Yes
	Task Description	Describe the task.	No
Task Settings	Source Platform	Select the source cloud platform. Currently, Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Google Cloud, Azure, Qiniu Cloud, and Kingsoft Cloud are supported.	Yes
	Credential	Select the credential for accessing the source cloud platform. If no credential is available, choose **Create** to add one. For details, see **Managing Credentials**. • If the source cloud platform is Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Qiniu Cloud, or Kingsoft Cloud, select **AK/SK** for **Authentication** and enter the AK/SK pair of your source cloud account. • If the source cloud platform is Google Cloud, select **Configuration File** for **Authentication** and upload the configuration file that contains your Google Cloud service account credentials. The file must be in JSON format and cannot exceed 4 KB. • If the source cloud platform is Azure, Select **ID/Secret** for **Authentication**. To learn how to obtain Azure credentials, see **How Do I Obtain Azure Credentials?**	Yes
	Region	Select the region where your source environment is located. Multiple regions can be selected.	Yes

Step 5 Choose the resource types (collection items) to be collected from the **Resource Type** drop-down list. For resource types supported for each cloud platform, see **Table 4-9**.

Figure 4-12 Selecting the categories of resources to be discovered

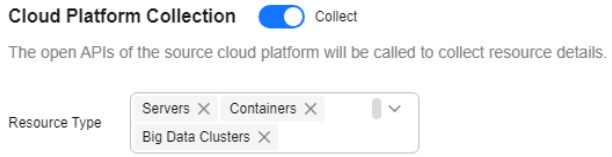


Table 4-9 Supported resource types

Cloud Platform	Resource Type	Subtype
Alibaba Cloud	• Servers • Containers • Big data clusters • Databases	-
	Middleware	• Redis • Kafka
	Storage	• Object storage • File storage
	Networks	• Cloud connections • Load balancers (ALB and CLB) • Private lines • Public domain names • Private domain names • EIPs • Public NAT gateways • Route tables • Security groups • VPCs • VPN gateways
Huawei Cloud	• Servers • Containers • Big data clusters • Databases	-
	Middleware	• Redis • Kafka

Cloud Platform	Resource Type	Subtype
	Storage	• Object storage • File storage
	Networks	• Load balancers (ELB) • Public domain names • Private domain names • EIPs • Public NAT gateways • Route tables • Security groups • VPCs
AWS	• Servers • Containers • Databases	-
	Storage	• Object storage • File storage
	Networks	• Load balancers (ELB) (Currently, only CLB balancers are supported.) • Public domain names • Private domain names • EIPs • Public NAT gateways • Route tables • Security groups • VPCs
	Storage	• Object storage • File storage
Tencent Cloud	• Servers • Databases	-
	Storage	• Object storage • File storage
Google Cloud	• Servers • Containers • Databases	-
	Middleware	Redis

Cloud Platform	Resource Type	Subtype
	Storage	• Object storage • File storage
	Networks	• EIPs • Route tables • Security groups • VPCs
Azure	• Servers • Containers • Databases	-
	Storage	• Object storage • File storage
	Middleware	• Redis • Kafka
	Networks	• EIPs • Route tables • Security groups • Public NAT gateways • VPCs • Load balancers
Qiniu Cloud	Storage	Object storage
Kingsoft Cloud	Storage	Object storage

Step 6 (Optional) Group the servers to be discovered as an application.

- If an **application** is available, select the application from the **Application** drop-down list.
- If no applications are available, click **Create Application**. In the displayed dialog box, enter an application name and description, select the business scenario, environment, and region, and click **OK**.

Step 7 Click **Confirm**. After this Internet-based discovery task is created, MgC starts discovering details about resources and applications.

- You can **view the list and details of discovered resources** on the **Resource Discovery** page.
- You can **view the task status and details** on the task list page.

Step 8 Wait until the task status changes to **Succeeded**. Then perform subsequent operations for different types of resources.

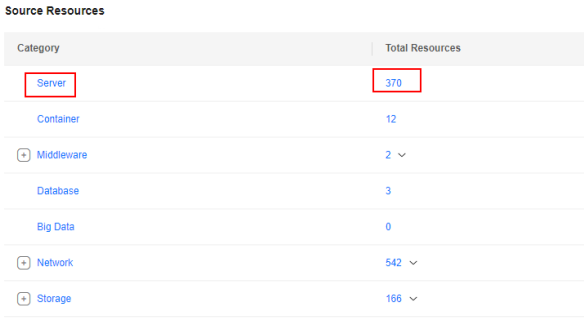
- For discovered servers, **perform a deep collection** or **check their migration readiness** and then design a **migration solution** or **migration plan**.
- For discovered containers, **perform a deep collection**and **design a migration solution**.
- **Design migration solutions** for middleware, database, and storage resources. **Deep collection** is available for AWS RDS (for MySQL, MariaDB, Aurora, PostgreSQL, SQL Server, and Oracle) and AWS DocumentDB databases.
- Perform a **deep collection** and design migration plans for object storage resources. For details, see **Migration Plans**.

----End

Check Migration Readiness for Servers

Perform the following steps to check whether your source servers meet the migration requirements:

1. **Install Edge** in the source intranet environment and connect Edge to MgC.
2. In the **Resources** area, click **Server** in the **Category** column or the number in the **Total Resources** column.



3. Select the servers to migrate together and click **Group as Application** above the list.
   - If you have **created an application**, select the application from the drop-down list and click **OK**.
   - If you have not created an application, click **Create Application** in the displayed dialog box. Then enter an application name and description, select a business scenario, environment, and target region, and click **Create**. Then click **OK**.
4. On the top of the server list, choose **Migration Scenario** > **Server migration**.
5. For each source server, click **Configure** in the **Migration Readiness** column.
6. Configure the parameters listed in **Table 4-10**.

Table 4-10 Parameters for configuring migration readiness

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.

Parameter	Configuration
Access IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration.
Port	Enter the port on the source server that allows access from the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC. **NOTICE** The account provided in the credential must have sufficient permissions, so Edge can collect necessary server details. To perform a deep collection for the server to collect as much as details, the credential you provided must meet the following requirements: • Linux: **root** and its password • Windows: **administrator** and its password

7. Click **Confirm**. The system checks whether the source server can be accessed using the information you provided and whether the server can be migrated. If **Ready** shows up in the **Migration Readiness** column, the source server can be migrated. Then you can **design a migration solution** or **design a migration plan** to migrate the server.

Performing a Deep Collection for Servers

Perform the following steps to perform a deep collection for your source servers:

1. Ensure that **Edge** has been installed in the source intranet environment and has been connected to MgC.
2. In the **Resources** area, click **Server** in the **Category** column or the number in the **Total Resources** column.
3. For each server whose details need to be collected, in the **Edge Device** column, click **Configure**. Configure the parameters listed in **Table 4-11**.

Table 4-11 Parameters for configuring a deep collection

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.

Parameter	Configuration
Access IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration.
Port	Enter the port on the source server that allows access from the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC. **NOTICE** The account provided in the credential must have sufficient permissions, so Edge can collect necessary server details. To perform a deep collection for the server to collect as much as details, the credential you provided must meet the following requirements: • Linux: **root** and its password • Windows: **administrator** and its password

4. Click **Confirm**. Then the system automatically starts executing a deep collection. When **Collected** shows up in the **Deep Collection** column, the collection is complete. You can **design a migration solution** or **design a migration plan** to migrate your servers.

Performing a Deep Collection for Containers

Perform the following steps to perform a deep collection for your containers:

1. Ensure that **Edge** has been installed in the source intranet environment and has been connected to MgC.
2. In the **Resources** area, click **Container** in the **Category** column or the number in the **Total Resources** column.
3. For each container resource whose details need to be collected, click **Associate** in the **Edge Device** column.

   To associate multiple container resources with one Edge device, select them and choose **Manage Device Association** above the list.
4. Select the Edge device in the source environment and click **OK**. Wait until **Associated** appears in the **Edge Device** column.
5. For each container resource, in the **Credential** column, click **Associate** to configure their credentials.
6. Select the resource credential. If the credential has not been added to MgC, go to the Edge console, **add the credential** to Edge, and synchronize it to MgC.

7. Click **OK**. Then MgC checks whether the resource can be accessed using the associated credential. When **Ready** appears in the **Deep Collection** column, click **Collect** to perform a deep collection. You can click **Collect Again** in the **Deep Collection** column to perform a second deep collection if needed. After the collection is complete, click the resource name to view the collected details.

Performing a Deep Collection for Object Storage Resources

Follow the steps below to perform a deep collection for your object storage resources, so the system can use the collected details to recommend rightsized migration clusters.

1. Ensure that **Edge** has been installed in the source intranet environment and has been registered with MgC.
2. On the **Resource Discovery** page, in the **Resources** area, click **Storage** in the **Category** column or the number in the **Total Resources** column.
3. Locate an object storage resource and click **Configure** in the **Deep Collection** column.
4. Select the Edge device in the source environment and the credential used for accessing the resource, and click **Confirm**. If the credential has not been added to MgC, go to the Edge console and **add the resource credential** to Edge and synchronize it to MgC.

   NOTICE

   To perform a deep collection for your Azure object storage resources, you need to provide your storage account and access key. For details about how to obtain the credentials, see **How Do I Obtain the Required Credentials Before Using MgC to Perform a Deep Collection for My Azure Object Storage Resources?**

5. Click **Add Prefix** in the **Operation** column.
6. Enter a prefix to filter the objects whose details need to be collected. If this parameter is not specified, all objects in the bucket are collected by default. Click **OK** to save the prefix settings.
7. Click **Deep Collection** in the **Operation** column. The system starts collecting path details. You can perform a deep collection on a resource for multiple times. When **Collection Status** changes to **Completed**, click the resource name to view the collected information.

Performing a Deep Collection for Databases

Deep collection is available for AWS RDS (for MySQL, MariaDB, Aurora, PostgreSQL, SQL Server, and Oracle) and AWS DocumentDB databases. The data collected includes database version, engine, server character set, average transactions per second (TPS), and queries per second (QPS), and other key performance metrics (KPIs). The collected data depends on the database type.

1. Ensure that **Edge** has been installed in the source intranet environment, can access source databases, and has been registered with MgC.

2. On the **Resource Discovery** page, in the **Resources** area, click **Database** in the **Category** column or the number in the **Total Resources** column.
3. In the database list, filter all discovered AWS databases by applying the **Vendor** filter. Locate a supported database and in the **Edge Device** column, click **Associate**.

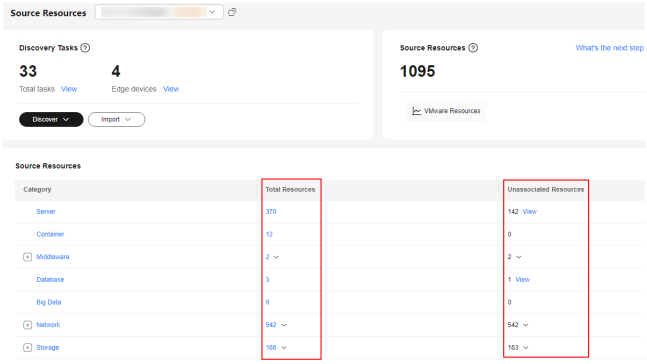
   To associate multiple databases with one Edge device, select them and choose **Manage Device Association** above the list.
4. Select your Edge device. For **Access Setting**, if the selected resource is in the same VPC as the Edge device, select **Private access**. Otherwise, select **Public access**. Then click **OK**. Wait until **Associated** appears in the **Edge Device** column.
5. In the **Credential** column, click **Associate** to associate the credential for accessing the resource.
6. Select the database credential. If the credential has not been added to MgC, go to the Edge console, **add the credential** to Edge, and synchronize it to MgC.
7. Click **OK**. Then MgC checks whether the resource can be accessed using the associated credential. When **Ready** appears in the **Deep Collection** column, click **Collect** to perform a deep collection. You can click **Collect** in the **Deep Collection** column to perform a deep collection again if needed.
8. Wait for the deep collection to complete. Then click the database name to go to the database details page. In the **Database Information** area, you can view the collected details.

Viewing Discovered Source Resources

A discovery task can collect only basic information about source resources. More detailed information needs to be collected through a deep collection.

1. On the **Resource Discovery** page, in the **Resources** area, view the number of discovered resources by category and the number of resources that are not associated with applications.

   Figure 4-13 Viewing the resource list

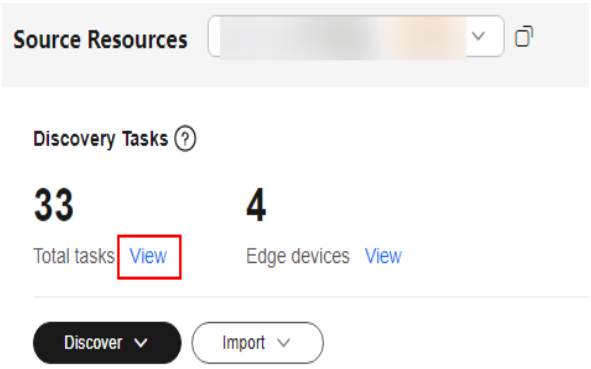


2. Click a resource category or number to view the resource list.
3. Click a resource name to view its details.

Viewing Discovery Tasks

Step 1 On the **Resource Discovery** page, click **View** next to **Total tasks** to go to the task list.

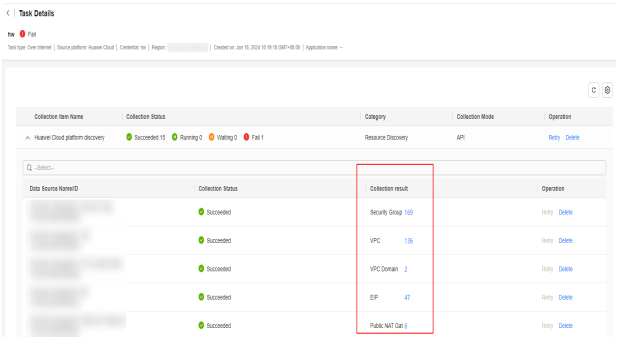
Figure 4-14 Viewing the task list



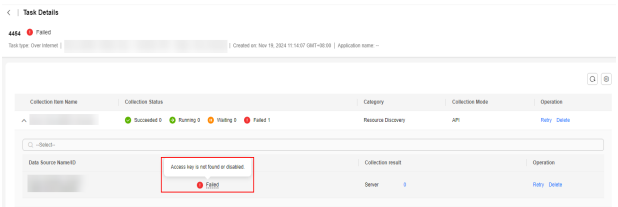
Step 2 Locate a discovery task, and click **View** in the **Operation** column.

Step 3 Click ⌄ before a collection item to view the data sources contained in the collection item and the collection results.

Figure 4-15 Viewing collection results



If a data source fails to be collected, you can move the cursor to the collection status of the data source to view the failure cause.



Step 4 In the **Collection Result** column, you can click the number next to a resource category to view the corresponding resource list.

----End

4.2.2 Discovering Resources over an Intranet

This section describes how to discover servers in on-premises environments. Before getting started, you need to install Edge in the source intranet. Then you can discover servers by network range or VMware host.

Precautions

- Only VMs in VMware vSphere 5.0 to 7.0 can be discovered.
- When the system scans VMware VMs or scans servers on specified network ranges, it uses the servers' private IP addresses and the ID of the used Edge device to identify discovered servers. If a server's private IP address changes after a collection is complete, the server will be identified as a new one during the next collection, and the total number of discovered servers will increase. To avoid this, you are advised not to change private IP addresses of source servers before the migration is complete.

Prerequisites

- You have **installed Edge** in the source intranet environment and have connected the Edge device to MgC.
- You have **added source server credentials** to Edge.

NOTICE

To perform a deep collection for your source servers to collect as much as details, provide server credentials that meet the following requirements:

- Linux: **root** and its password
- Windows: **administrator** and its password

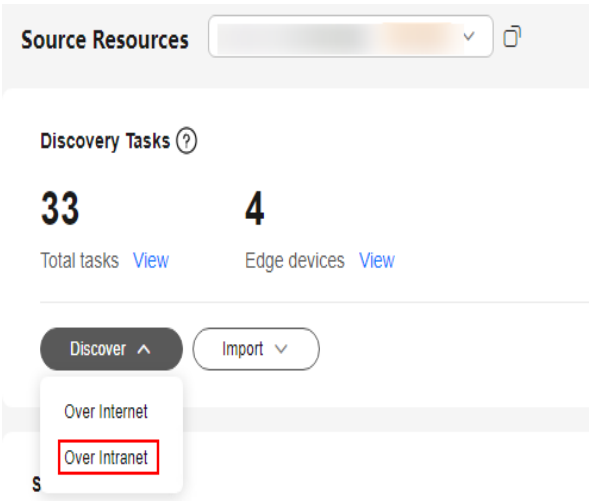
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Discover** > **Over Intranet**.

Figure 4-16 Discovery over an intranet



Step 4 Configure the parameters listed in **Table 4-12**.

Table 4-12 Parameters for creating an intranet-based discovery task

Parameter	Description
Task Name	Enter a task name.
Task Description	Describe the task.
Edge Device	Select the **Edge device** in the source intranet environment.

Step 5 Enable **Scan Network Range** or **Scan VMware VMs** to discover servers as needed.

- If **Scan Network Range** is enabled, configure parameters listed in **Table 4-13**.

  Table 4-13 Parameters for scanning a network range

Parameter	Description
Protocol	Select the communication protocol **TCP** or **ICMP**.
Network Range	Enter an IP address range which must fall within: – 10.0.0.0 – 10.255.255.255 – 172.16.0.0 – 172.31.255.255 – 192.168.0.0 - 192.168.255.255
Linux	Enter the port for scanning Linux servers. This parameter is available only if you choose the TCP protocol. If you need to skip Linux servers during the scan, set this parameter to **0**.

Parameter	Description
Windows	Enter the port for scanning Windows servers. This parameter is available only if you choose the TCP protocol. If you need to skip Windows servers during the scan, set this parameter to **0**.

- If **Scan VMware VMs** is enabled, in the **IP Address** text box, enter the IP address of the vCenter Server where the VMs will be discovered, and select the credential for accessing the vCenter Server. All VMs managed by the vCenter Server will be discovered. If the vCenter Server's credential has not been added to Edge, **add the resource** to Edge. When adding the credential, enter the username and password for logging in to the vCenter Server.

Step 6 Click **Confirm**. The intranet-based discovery task is created, and MgC starts collecting details about source servers.

On the **Source Resources** page, in the **Discovery Tasks** card, click **View** next to **Total tasks**.

Step 7 Wait until the task status changes to **Succeeded**, and perform a deep collection.

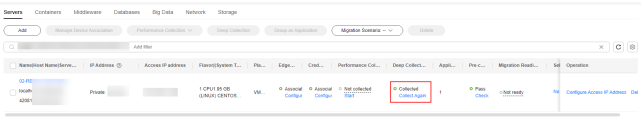
1. On the **Source Resources** page, in the resource list, click **Server** in the **Category** column or the number in the **Total Resources** column.
2. In the **Edge Device** or **Credential** column, click **Configure**.
3. Configure the parameters listed in **Table 4-14**.

Table 4-14 Parameters for configuring a deep collection

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.
Access IP Address	Select the IP address for accessing the source server. It can be a public or private IP address. After the pre-migration check is passed, the IP address you select here will be used for migration.
Port	Enter the port on the source server that allows access from the Edge device. – By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. – By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to Edge, go to the Edge console and **add the server credential** to Edge and synchronize it to MgC.

4. Click **Confirm**. Then the system automatically starts executing a deep collection. When **Collected** shows up in the **Deep Collection** column, the collection is complete. You can **design a migration solution** or **design a migration plan** to migrate your servers.

Figure 4-17 Discovery completed



----End

4.2.3 Importing Details of Alibaba Cloud Resources

You can import CSV-formatted resource details from Alibaba Cloud directly into MgC. You first need to export the resource details from Alibaba Cloud. After the import is complete, you can view the resource details on the MgC console. With the imported details, you can get recommendations for suitable Huawei Cloud resources.

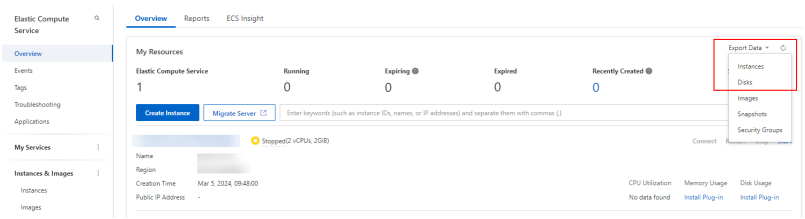
The supported Alibaba Cloud resources are:

- Elastic Compute Service (ECS) instances and cloud disks
- Object Storage Service (OSS) buckets
- ApsaraDB RDS for MySQL instances
- ApsaraDB RDS for GaussDB instances

Exporting Data of Alibaba Cloud ECS Instances and Disks

Step 1 Sign in to the Alibaba Cloud ECS console.

Step 2 On the **Overview** page, choose **Export Data** > **Instances** to export a list of all ECS instances.



Step 3 Choose **Export Data** > **Disks** to export a list of all cloud disks.

Step 4 Open the instance list and the disk list and adjust the resource information if needed. Ensure that the lists contain the necessary information.

List	Mandatory Fields
Instance list	Instance ID, OS, instance specifications, CPU, and memory
Disk list	Instance ID, disk ID, capacity (GiB), and disk attributes

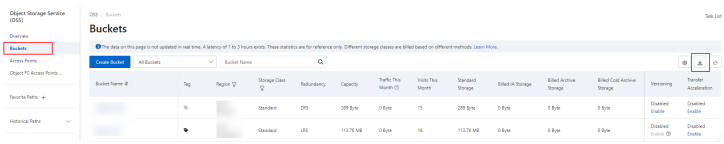
----End

Exporting Data of Alibaba Cloud OSS Buckets

Step 1 Sign in to the OSS console.

Step 2 In the left navigation pane, choose **Buckets**.

Step 3 Click the  icon in the upper right corner of the list. The **Export Bucket List** dialog box is displayed.



Step 4 Select the information to be exported and click **OK**.

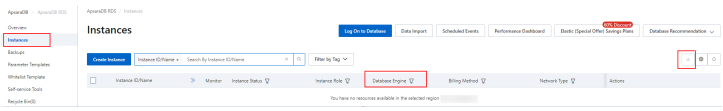
----End

Exporting Data of Alibaba Cloud ApsaraDB RDS for MySQL Instances

Step 1 Sign in to the Alibaba Cloud ApsaraDB RDS for MySQL console.

Step 2 In the navigation pane on the left, choose **Instances**. Select a region in the upper part of the page.

Step 3 Filter all MySQL instances in the **Database Engine** column and click the  icon in the upper right corner of the list.



Step 4 In the displayed dialog box, select the information to be exported. Click **OK**. The system exports the instance information to a CSV file and automatically downloads the file.

Step 5 Modify the exported CSV file so MgC can parse the file content.

1. Use a file editor to open the CSV file and search for **PrivatePort(A network port,)** in the table header.
2. Delete the comma (,) from this field and save the file. The correct format is **PrivatePort(A network port)**.

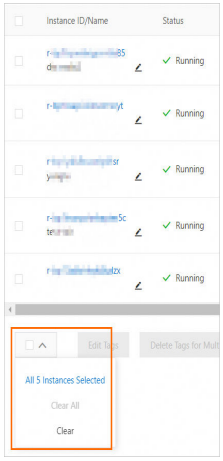
----End

Exporting Data of Alibaba Cloud ApsaraDB RDS for Redis Instances

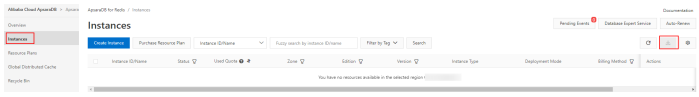
Step 1 Sign in to the Alibaba Cloud ApsaraDB for Redis console.

Step 2 In the navigation pane on the left, choose **Instances**. Select a region in the upper part of the page.

Step 3 In the lower left corner of the page, expand and click **Select All**.



Step 4 Click the  icon in the upper right corner of the list. In the dialog box that is displayed, select the information to be exported.



Step 5 Click **OK**. The system exports the instance information to a CSV file and automatically downloads the file.

Step 6 Modify the exported CSV file so MgC can parse the file content.

1. Use a file editor to open the CSV file.
2. Add a comma (,) at the end of the table heading line and save the file.

   The following shows an example table header before modification:

   ```
   ID,Name,Quantity
   ```

   The following shows the table header after modification:

   ```
   ID,Name,Quantity,
   ```

----End

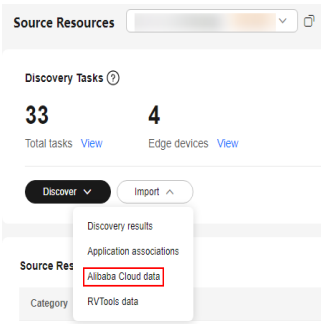
Importing Details of Alibaba Cloud Resources

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Import** > **Alibaba Cloud Data**.

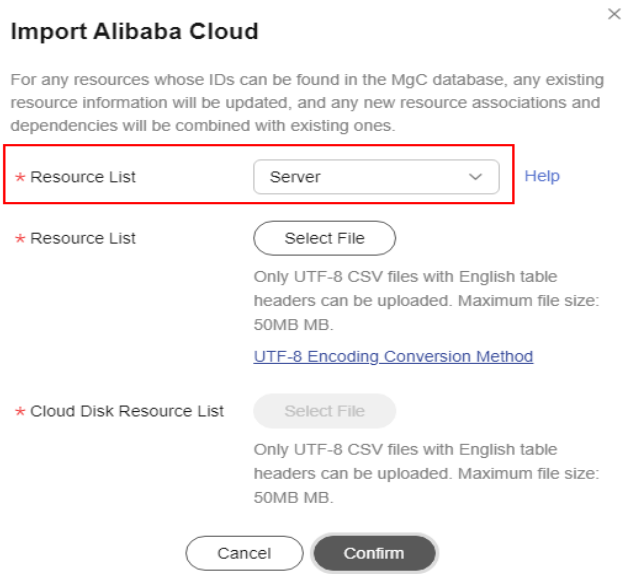
Figure 4-18 Importing details of Alibaba Cloud resources



NOTICE

Only CSV files encoded in UTF-8 format with English table headers can be imported. The size of a single file cannot exceed 15 MB. For details about how to convert the encoding format of a CSV file to UTF-8, see **How Do I Convert the Encoding Format of a CSV File to UTF-8?**

Step 4 In the **Resource Type** drop-down list, select the types of resources to be imported and upload the corresponding file.



Resource Type	Procedure
Servers	1. Click **Select File** next to **Resource List** and select the **exported ECS list**. 2. Click **Select File** next to **Cloud Disk List** and select the **exported disk list**.
Redis	Click **Select File** next to **Resource List** and select the **exported Redis instance list**.

Resource Type	Procedure
Object storage	Click **Select File** next to **Resource List** and select the **exported OSS bucket list**.
Databases	Click **Select File** next to **Resource List** and select the **exported MySQL instance list**.

Step 5 Click **Confirm**. After the resource list is uploaded successfully, select the corresponding resource categories to view the imported resource information.

Step 6 (Optional) Group imported Alibaba Cloud resources as applications. For details, see **Grouping Resources as Applications**. Then assess the applications to obtain recommendations for Huawei Cloud resources. For details, see **Getting Target Recommendations**.

----End

4.2.4 Importing RVTools Data

Notes and Constraints

- **Supported RVTools versions**

  The supported RVTools versions include:

  - 4.4.1
  - 4.4.2
  - 4.4.3
  - 4.4.4
  - 4.4.5
  - 4.5.0
  - 4.5.1
  - 4.6.1
- **File format requirements**

  Only the Excel (.xlsx) format is supported.
- **File size and compression ratio requirements**

  The file to be imported cannot be larger than 100 MB, and the compression ratio cannot be lower than 5%.

Step 1: Exporting RVTools Data

Step 1 Start the RVTools application.

Step 2 On the login page, enter the following information:

- In the **IP address/Name** text box, enter IP address of the vCenter server.
- In the **User name** text box, enter the username for connecting to the vCenter server.
- In the **Password** text box, enter the password corresponding to the username.

Step 3 Click **Login**. In the top menu bar, choose **File** > **Export all to Excel**.

Step 4 Select a path for saving the file that contains details about your resource.

----End

Step 2: Uploading the Exported File to OBS

Step 1 Sign in to Huawei Cloud **OBS console** and create a Standard bucket for storing the file exported from RVTools. For details, see **Creating a Bucket**.

Step 2 Upload the **file exported from RVTools** to the Standard OBS bucket. For details, see **Uploading an Object**.

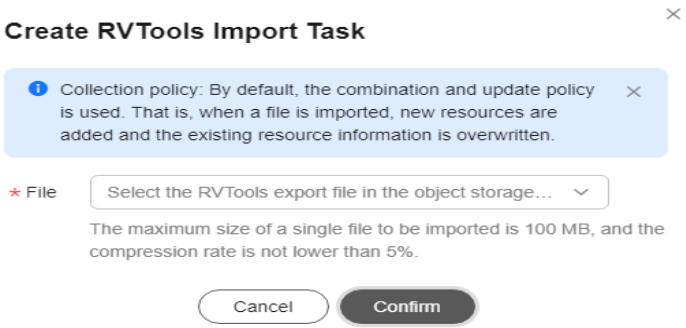
----End

Step 3: Importing RVTools Data

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Discovery Tasks** card, choose **Import** > **RVTools data**.

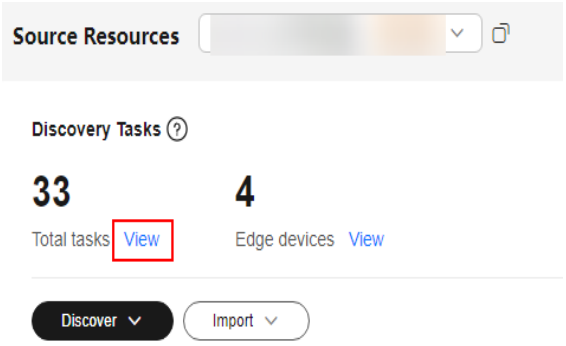


NOTE

If you are using this function for the first time, click **OK** in the **Authorization Required** dialog box to agree to the agency-based authorization.

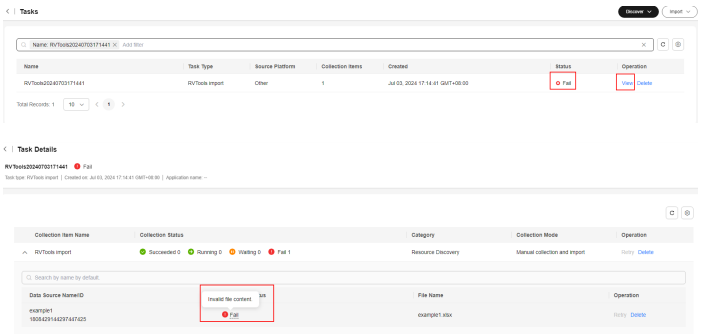
Step 4 In the **File** drop-down list, locate the OBS bucket you used to store the file exported from RVTools in **step 2**, select the exported file, and click **Confirm**.

Step 5 After the upload is successful, in the **Discovery Tasks** card, click **View** next to **Total tasks**. In the task list, view the task status.



- If the task status is **Failed**, click **View** in the **Operation** column to view the data source that failed to be collected. You can move the cursor to the collection status to view the failure cause. For details about how to rectify common failure causes, see **What Can I Do If Importing RVTools Data Fails?**

  After handling the failure cause, you need to delete the failed task and perform a second import.



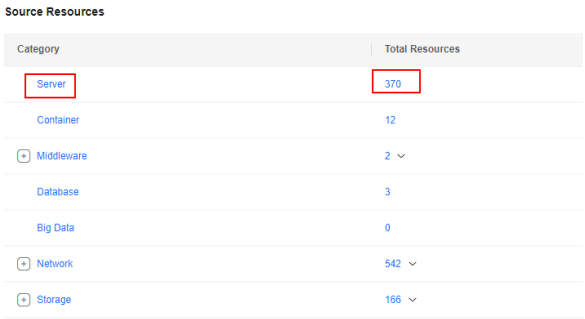
- If the task status is **Succeeded**, you can view the imported **resource details** and **resource statistics**.

----End

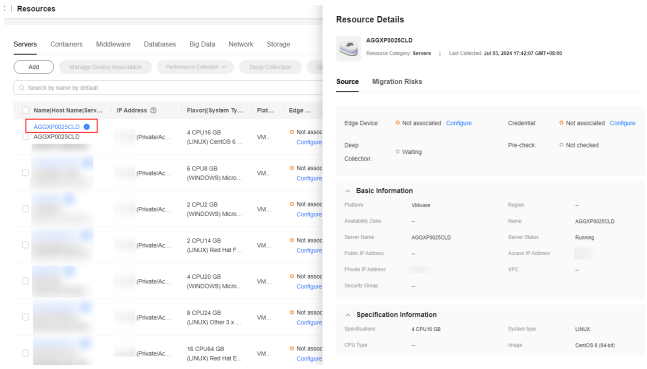
Viewing Resource Details

If the status of the RVTools import task is **Succeeded**, you can view the details of VMware servers, including the hostname, server ID, IP address, specifications, OS type, and platform.

Step 1 On the **Source Resources** page, in the resource list, click **Server** in the **Category** column or the number in the **Total Resources** column.



Step 2 In the server list, you can view the imported server information. You can click a resource name to view more details.



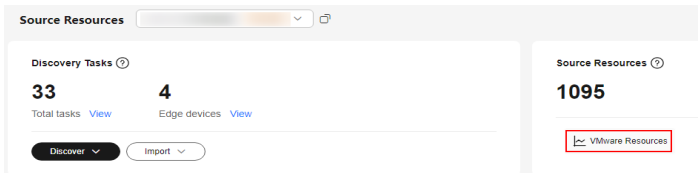
----End

Viewing Resource Statistics

If the status of the RVTools import task is **Succeeded**, you can view the VMware resource statistics under the corresponding tab.

- **Host**: CPU model, host quantity, total memory, allocated VM memory, memory allocation rate, VM quantity, quantity of running VMs, storage usage, and host list.
- VMs: the number of Windows OS versions, number of Linux OS distributions, total memory, total number of vCPU, number of disks, disk capacity, number of VMs, and VM list

Step 1 On the **Source Resources** page, click **VMware Resources** in the **Source Resources** card.



Step 2 On the displayed **Resource Statistics** page, click the **Host** or **VM** tab to view resource statistics.

----End

4.3 Manually Adding Resources

You can manually add servers and Oracle databases to MgC. To collect details about these resources, you need to install Edge in the source environment.

Preparations

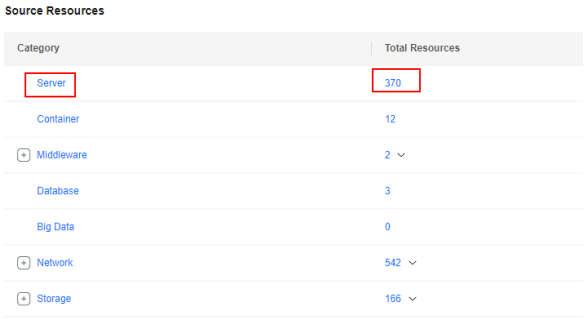
- **Create a migration project** on the MgC console.
- **Install Edge** in the source intranet environment and **connect Edge to MgC**.
- **Add resource credentials** to Edge.

Adding a Server

Step 1 Sign in to the **MgC console**.

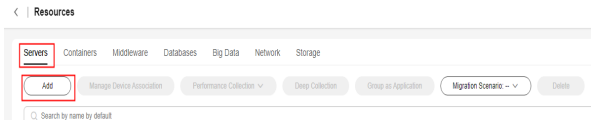
Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Resources** area, click **Server** in the **Category** column or the number in the **Total Resources** column.



Step 4 Click **Add** above the list.

Figure 4-19 Adding a server



Step 5 In the displayed dialog box, configure parameters listed in **Table 4-15** and click **Confirm**. The system automatically checks the credential status and starts collecting resource details.

Table 4-15 Parameters for adding a server

Parameter	Description
Name	Enter a server name.
Edge Device	Select the Edge device installed in the source environment.
Type	Select the OS type of the source server.
IP Address	Enter the IP address of the source server. If the source server is in the same VPC as the Edge device, you can enter the private IP address of the server. Otherwise, you have to enter its public IP address.

Parameter	Description
Port	Enter the port on the source server opened to the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC.

Step 6 View the added server on the **Servers** tab.

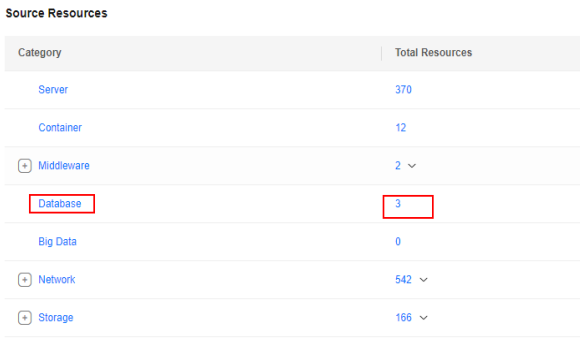
----End

Adding a Database

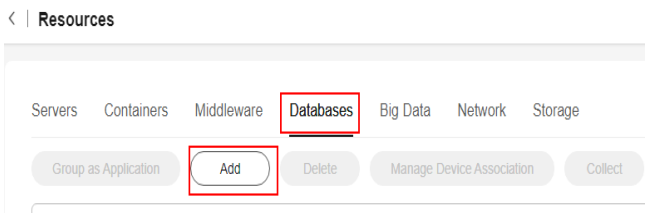
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Resources** area, click **Database** in the **Category** column or the number in the **Total Resources** column.



Step 4 Click **Add**.



Step 5 Set the parameters listed in **Table 4-16**.

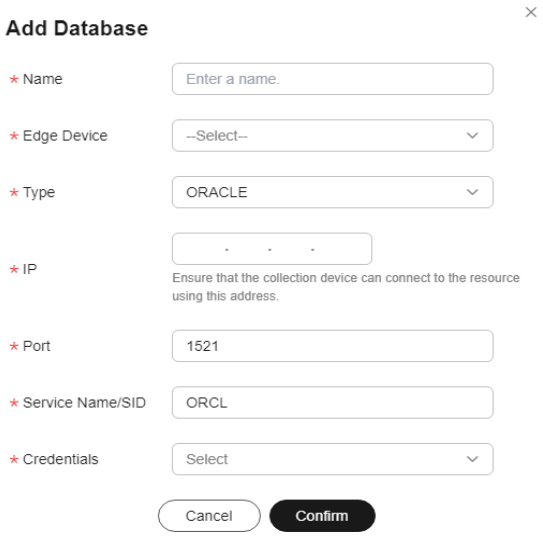
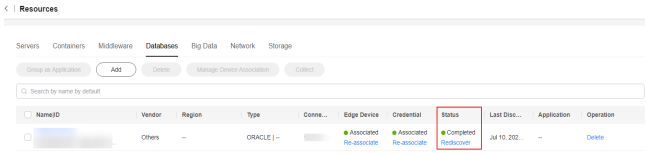


Table 4-16 Parameters for adding a database

Parameter	Description
Name	Enter a server name.
Edge Device	Select the Edge device that is connected to MgC.
Type	Select **Oracle**.
IP Address	Enter the IP address for connecting to the source Oracle database server.
Port	Enter the listening port of the source Oracle database. The default value is 1521.
Service Name/SID	Enter the service name of the source database. The default value is **ORCL**. You can log in to the Oracle database server and run the following command to view the service name of the database you want to add to MgC: `SELECT instance_name FROM v$instance;`
Credential	Select the database credential. If the credential has not been added to MgC, go to the Edge console, **add the credential** to Edge, and synchronize it to MgC. When you add the credential to Edge, set **Resource Type** to **Database** and **Authentication** to **User name/Password**.

Step 6 Click **Confirm**. The system automatically starts collecting details about the database. When **Collected** is displayed in the **Deep Collection** column, the collection is complete. Click the database name. On the displayed database details page, you can view the collected information.



----End

4.4 Collecting Server Performance Data

MgC can collect server performance data, including the CPU usage, memory usage, disk IOPS, inbound and outbound traffic, inbound and outbound packets, and the number of network connections. The collected performance data will be used by MgC to recommend rightsized servers.

Notes

- After a collection starts, performance data is collected every 5 minutes by default.
- The collection duration should be at least 1 hour. If the collection duration is too short, the maximum and average values in 7- and 30-day history cannot be calculated.
- A performance collection runs for seven days by default and stops after the period ends. You can stop or restart the collection at any time. After the collection is restarted, the collection period is recalculated.
- Collected performance data will be retained for 180 days and automatically deleted after that.
- It is recommended that Edge be installed on an independent server to ensure collection stability and efficiency.
- Edge can collect performance data from up to 1,000 Linux servers at the same time. If there are 1,000 or more Linux servers to be collected, the server where Edge is installed must have at least 8 CPUs and 16 GB of memory with at least 8 GB available.
- Edge can collect performance data from up to 500 Windows servers at the same time. If there are 500 or more Windows servers to be collected, the server where Edge is installed must have at least 16 CPUs and 32 GB of memory with at least 8 GB available.

Prerequisites

- You have completed a discovery task over the Internet or intranet, or you have manually added servers to MgC.
- You have **installed Edge** in the source intranet environment and have connected the Edge device to MgC.
- You have **added source server credentials** to Edge.

Procedure

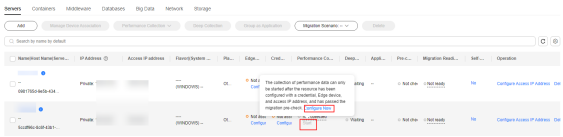
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Resources** area, click **Server** in the **Category** column or the number in the **Total Resources** column.

Step 4 Configure an Edge device and credentials for the servers whose performance data needs to be collected.

- If the **Start** button in the **Performance Collection** column is dimmed, move the cursor to the **Start** button. In the dialog box that is displayed, click **Configure Now**.



- If the **Start** button in the **Performance Collection** column is available, the configuration is complete. In this case, skip this step and go to **step 7**.

Step 5 Configure the parameters listed in **Table 4-17**.

Table 4-17 Parameters for configuring a migration pre-check

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.
IP Address	Select the IP address used for accessing the source server. It can be a public or private IP address. If you need to use a proxy to access the server, enter the proxy IP address. After the migration pre-check passes, the IP address will be used for subsequent migration.
Port	Enter the port on the source server opened to the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC.

Step 6 Click **Confirm**. The system will verify the correctness of the configuration.

Step 7 After the configuration is verified, click **Start** in the **Performance Collection** column to start collecting server performance data. The collection status will change to **Collecting**.

After performance data collection is complete, you can perform the following operations:

- **Viewing Collected Data**
  a. Click a server name to go to its details page. In the **Performance Info** area, you can view the collected performance data. The maximum values in 7 days and 30 days are hourly aggregated data. To view these maximum values, ensure the collection duration exceeds 1 hour.
  b. Click **View** in the **Operation** column to view the details and line chart about the performance metric in the collection period.
- Stoping Collection
  Click **Stop** to pause performance data collection.

----End

4.5 Grouping Resources as Applications

Overview

You can divide discovered resources into groups, which we refer to as applications. Subsequent recommendation and migration processes are executed based on these applications.

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Source Resources**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Resources** area, choose a resource category to view the resource list.

Step 4 Select the resources to be added to an application and choose **Group as Application** in the upper left corner of the page.

Step 5 Select the application from the drop-down list. If no applications are available, click **Create Application** in the displayed dialog box. Then enter an application name and description, select the business scenario, environment, and region, and click **Create**.

Step 6 Click **OK**. You can view the application name in the **Application** column of the resources.

----End

5 Application Management

5.1 Creating an Application

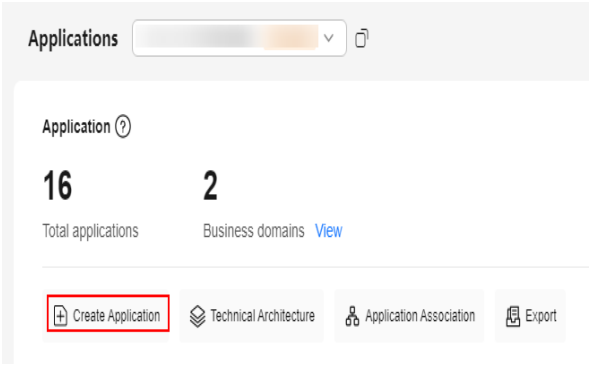
You can group resources with a shared business purpose as an application. These applications will be used for getting target resource recommendations and creating migration workflows.

Creating an Application

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Applications**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Create Application**.



Step 4 Enter an application name and description, select a service scenario and environment, select the region you are migrating to, and click **Create Application**. The application is successfully created, and the page for adding resources to the application is displayed.

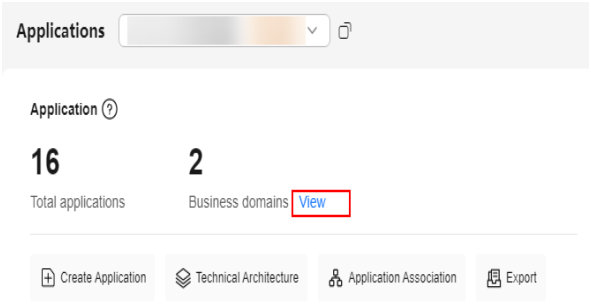
- If source resources have been discovered, and you want to add the discovered resources to the created application, select the resources and click **Add Now**.
- If no source resources have been discovered, click **Add Later**. You can add resources to the application later.

----End

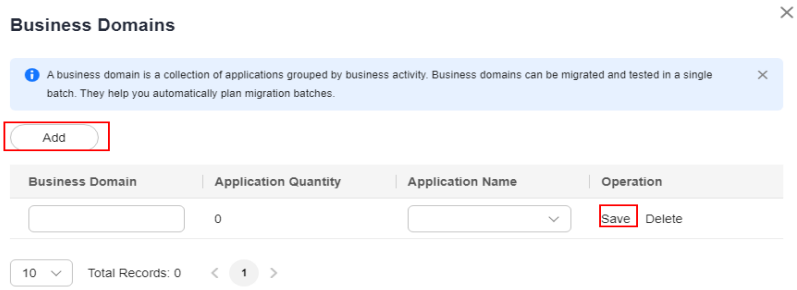
Managing Applications

You can manage applications by business domain.

Step 1 On the **Applications** page, click **View** next to **Business domains**.



Step 2 Click **Add**, enter a business domain name, select the applications you want to add to this domain, and click **Save**.



----End

Modifying an Application

This operation is only supported for simple projects.

Step 1 In the application list, locate the application you want to modify and click **Modify** in the **Operation** column.

Step 2 Modify application parameters and click **OK**.

----End

5.2 Adding Discovered Resources to an Application

The following procedure describes how to group discovered source servers as applications. These applications will be used for subsequent resource assessment and migration.

Prerequisites

Source servers have been discovered through Internet- and intranet-based discovery tasks or manual addition.

Procedure

Step 1 Sign in to the **MgC console**.

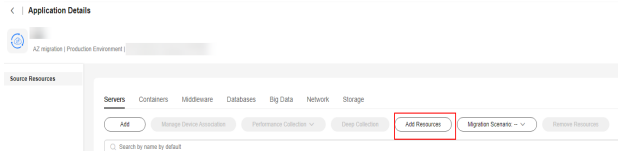
Step 2 In the navigation pane on the left, choose **Research** > **Applications**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the application list, click the application name.

Step 4 Click the tab based on the category of resources you want to add. The following describes how to add source servers to the application. The procedure also applies to other categories of resources.

Step 5 On the **Servers** tab, click **Add Resources** above the list.

Figure 5-1 Add discovered resources to an application



Step 6 Select the source servers to be added to the application and click **Add Now** in the lower right corner. You can see the source servers added to the application in the list.

Step 7 Click **Assess** in the upper right corner of the page. Assess the application resources to obtain recommendations for the best-fit Huawei Cloud resources. For more information, see **Getting Target Recommendations**.

----End

5.3 Manually Adding Custom Resources to an Application

You can add custom servers and Oracle databases to an application. To get the data about these resources to MgC, you need to install Edge in the source environment.

Preparations

- **Create a migration project** on the MgC console.
- **Install Edge** in the source intranet environment and **connect Edge to MgC**.
- **Add resource credentials** to Edge.
- **Created an application** on the application management page.

Add Servers to the Application

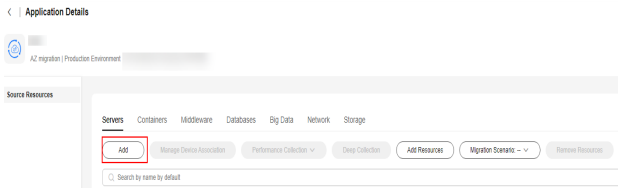
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Applications**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the application list, click the application name.

Step 4 On the **Servers** tab, click **Add Custom Resources** above the list.

Figure 5-2 Adding custom servers to the application



Step 5 In the displayed dialog box, configure parameters listed in **Table 5-1** and click **Confirm**. The system automatically checks the credential status and starts collecting resource details.

Table 5-1 Parameters for adding a server

Parameter	Description
Name	Enter a server name.
Edge Device	Select the Edge device installed in the source environment.
Type	Select the OS type of the source server.
IP Address	Enter the IP address of the source server. If the source server is in the same VPC as the Edge device, you can enter the private IP address of the server. Otherwise, you have to enter its public IP address.
Port	Enter the port on the source server opened to the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC.

Step 6 View the added server on the **Servers** tab.

----End

Adding Databases to the Application

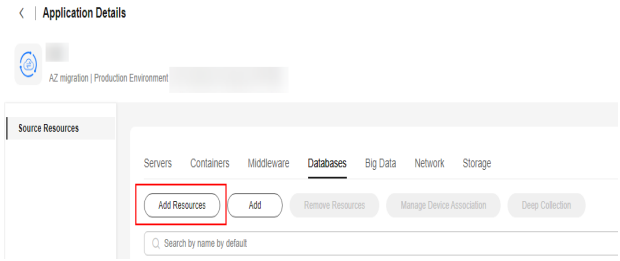
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Applications**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the application list, click the application name.

Step 4 On the **Databases** tab, click **Add** above the list.

Figure 5-3 Adding custom databases



Step 5 Set the parameters listed in **Table 5-2**.

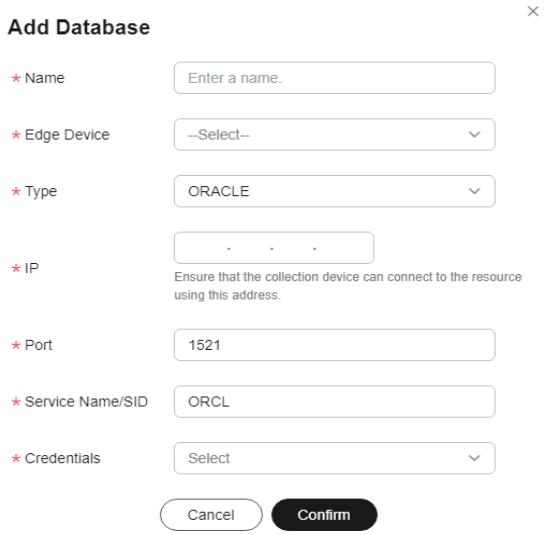
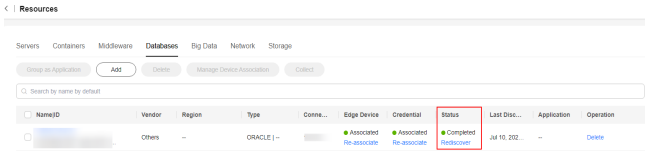


Table 5-2 Parameters for adding a database

Parameter	Description
Name	Enter a server name.
Edge Device	Select the Edge device that is connected to MgC.
Type	Select **Oracle**.
IP Address	Enter the IP address for connecting to the source Oracle database server.
Port	Enter the listening port of the source Oracle database. The default value is 1521.

Parameter	Description
Service Name/SID	Enter the service name of the source database. The default value is **ORCL**. You can log in to the Oracle database server and run the following command to view the service name of the database you want to add to MgC: `SELECT instance_name FROM v$instance;`
Credential	Select the database credential. If the credential has not been added to MgC, go to the Edge console, **add the credential** to Edge, and synchronize it to MgC. When you add the credential to Edge, set **Resource Type** to **Database** and **Authentication** to **User name/Password**.

Step 6 Click **Confirm**. The system automatically starts collecting details about the database. When **Collected** is displayed in the **Deep Collection** column, the collection is complete. Click the database name. On the displayed database details page, you can view the collected information.



----End

5.4 Collecting Server Performance Data

MgC can collect server performance data, including the CPU usage, memory usage, disk IOPS, inbound and outbound traffic, inbound and outbound packets, and the number of network connections. The collected performance data will be used by MgC to recommend rightsized servers.

Collect performance data of host resources in an application on the application details page.

Notes

- After a collection starts, performance data is collected every 5 minutes by default.
- The collection duration should be at least 1 hour. If the collection duration is too short, the maximum and average values in 7- and 30-day history cannot be calculated.
- A performance collection runs for seven days by default and stops after the period ends. You can stop or restart the collection at any time. After the collection is restarted, the collection period is recalculated.
- Collected performance data will be retained for 180 days and automatically deleted after that.
- It is recommended that Edge be installed on an independent server to ensure collection stability and efficiency.

- Edge can collect performance data from up to 1,000 Linux servers at the same time. If there are 1,000 or more Linux servers to be collected, the server where Edge is installed must have at least 8 CPUs and 16 GB of memory with at least 8 GB available.
- Edge can collect performance data from up to 500 Windows servers at the same time. If there are 500 or more Windows servers to be collected, the server where Edge is installed must have at least 16 CPUs and 32 GB of memory with at least 8 GB available.

Prerequisites

- You have **manually added custom source resources** or **added discovered source resources**.
- You have **installed Edge** in the source intranet environment and have connected Edge to MgC.
- You have **added source server credentials** to Edge.

Procedure

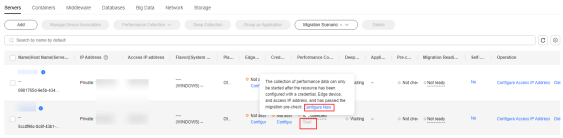
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Applications**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the application list, click the application name.

Step 4 Configure an Edge device and credentials for the servers whose performance data needs to be collected.

- If the **Start** button in the **Performance Collection** column is dimmed, move the cursor to the **Start** button. In the dialog box that is displayed, click **Configure Now**.



- If the **Start** button in the **Performance Collection** column is available, the configuration is complete. In this case, skip this step and go to **step 7**.

Step 5 Configure the parameters listed in **Table 5-3**.

Table 5-3 Parameters for configuring a migration pre-check

Parameter	Configuration
Type	Set this parameter based on the source server OS type.
Edge Device	Select the Edge device in the source environment.
IP Address	Select the IP address used for accessing the source server. It can be a public or private IP address. If you need to use a proxy to access the server, enter the proxy IP address. After the migration pre-check passes, the IP address will be used for subsequent migration.

Parameter	Configuration
Port	Enter the port on the source server opened to the Edge device. • By default, port 5985 on Windows source servers must be opened to the Edge device. The port cannot be changed. • By default, port 22 on Linux source servers must be opened to the Edge device. You can specify a different port if needed.
Credential	Select the server credential. If the credential has not been added to MgC, go to the Edge console, **add the server credential** to Edge, and synchronize it to MgC.

Step 6 Click **Confirm**. The system will verify the correctness of the configuration.

Step 7 After the configuration is verified, click **Start** in the **Performance Collection** column to start collecting server performance data. The collection status will change to **Collecting**.

After performance data collection is complete, you can perform the following operations:

- **Viewing Collected Data**
  a. Click a server name to go to its details page. In the **Performance Info** area, you can view the collected performance data. The maximum values in 7 days and 30 days are hourly aggregated data. To view these maximum values, ensure the collection duration exceeds 1 hour.
  b. Click **View** in the **Operation** column to view the details and line chart about the performance metric in the collection period.
- Stoping Collection
  Click **Stop** to pause performance data collection.

----End

5.5 Viewing Application Dependencies

To assist you in creating migration plans and designing architectures for the target environment, MgC organizes the relationships between your resources and applications. It then represents these technical architectures and dependencies into graphs.

Prerequisites

You have **imported application associations** or **imported discovery results**.

Viewing the Technical Architecture

Step 1 On the **Applications** page, click **Technical Architecture** in the **Application** card, and view a matrix of all discovered services, microservices, and resources.

Step 2 Right-click a service and view its technical architecture, dependencies, and microservices.

----End

Viewing Applications

On the **Applications** page, under **Applications**, you can view all applications in the current migration project. In the application list, you can view the application name, type, business domain as well as the number of involved microservices, downstream applications, and upstream applications.

- To view all microservices contained in an application, click the number in the **Microservices** column.
- To view all applications that an application relies on, click the number in the **Upstream Applications** column.
- To view all applications that an application is relied on, click the number in the **Downstream Applications** column.
- To view what microservices and resources are used by an application as well as their layers and associations, click **Technical Architecture** in the **Operation** column. In the displayed technical architecture, click a microservice or resource node to view its details.
- To view the associations between an application and its upstream and downstream applications, you can click **Dependency Map** in the **Operation** column. To view the details about this service and its upstream and downstream services, click its box.

Viewing Resource Candidates

Step 1 On the **Application Discovery** page, click **Candidates** in the **Source Resources** card.

Step 2 On the **Resources** tab, view the list of all collected resources and their details.

- To view the associations between a service and its upstream and downstream services, you can click **Service Dependencies** in the **Operation** column.
- To delete a service, click **Delete** in the **Operation** column.

----End

Viewing Microservices

Step 1 On the **Applications** page, click **Candidates** in the **Source Resources** pane.

Step 2 Click the **Microservices** tab to view the list of all collected microservices.

- To view the layer of a microservice and its associations with other resources, click **Technical Architecture** in the **Operation** column. In the displayed technical architecture, click a microservice or resource node to view its details.

- To view the associations between a microservice and its upstream and downstream services, choose **More** > **Service Dependencies** in the **Operation** column. To view the details about this microservice and its upstream and downstream services, click its box.
- To delete a microservice, click **More** > **Delete** in the **Operation** column.

----End

Exporting Application Dependencies

On the **Applications** page, click **Export** in the **Applications** card to export the access-layer call chains, database dependencies, middleware dependencies, and application dependencies for further analysis.

6 Big Data Lineage

6.1 Installing Edge

Before using the big data lineage function, you need to deploy Edge on a server that can access your metadata source.

> ⚠ CAUTION
>
> Edge does not support automatic restart. Do not restart Edge during task execution, or tasks will fail.

Procedure

Step 1 **Install Edge**.

Step 2 **Connect Edge to MgC**.

Step 3 **Add credentials for your big data resources** to Edge.

----End

6.2 Collecting Metadata

6.2.1 Creating a Metadata Connection

Create a connection to your metadata source, so you can collect source metadata to extract the list of databases and tables.

You can create connections to big data cloud services and components listed in **Table 6-1**.

Table 6-1 Supported metadata sources

Type	Metadata Source
Big data components	• Hive Metastore • Doris • HBase • ClickHouse • Delta Lake (with metadata) • Delta Lake (without metadata) • Hudi (with metadata) • Hudi (without metadata)
Big data cloud services	• MaxCompute • ApsaraDB for ClickHouse

NOTICE

- Before creating a connection to Hive Metastore, you need to obtain the IP address and port required for connecting to Hive Metastore
- Before creating a connection to MaxCompute, you need to obtain the endpoint for the region where your MaxCompute resources are hosted.
- Before creating a connection to Doris, you need to obtain the IP address and port required for accessing the Doris cluster
- Before creating a connection to HBase, you need to obtain the IP address and connection port of at least one ZooKeeper node in the HBase cluster.

Prerequisites

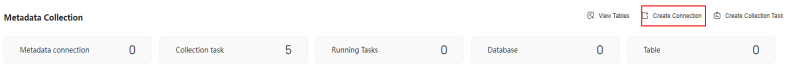
You have **installed Edge** on a server in the source environment.

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Metadata Collection** area, click **Create Connection**.



Step 4 Select a connection type and click **Next**.

Step 5 Based on the selected connection type, configure connection parameters.

- **Parameters for creating a connection to Hive Metastore**

- **Parameters for creating a connection to Doris**
- **Parameters for creating a connection to HBase**
- **Parameters for creating a ClickHouse connection**
- **Parameters for creating a connection to Delta Lake (with metadata)**
- **Parameters for creating a connection to Delta Lake (without metadata)**
- **Parameters for creating a connection to Hudi (with metadata)**
- **Parameters for creating a connection to Hudi (without metadata)**
- **Parameters for creating a connection to MaxCompute**
- **Parameters for creating a connection to ApsaraDB for ClickHouse**

Table 6-2 Parameters for creating a connection to Hive Metastore

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **Hive-Metastore-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Secure Connection	Determine whether to enable secure connection. • If Hive Metastore is deployed in an unsecured cluster, do not enable secure connection. • If Hive Metastore is deployed in a secured cluster, enable secure connection and provide access credentials. For details about how to obtain and add credentials to MgC, see "Big data - Hive Metastore" in **Adding Resource Credentials**.
Hive Version	Select the Hive version at the source. **CAUTION** If the source Hive version is 2.1.1, select **1.x**.
Hive Metastore IP Address	Enter the IP address for connecting to the Hive Metastore node.
Hive Metastore Thrift Port	Enter the port for connecting to the Hive Metastore Thrift service. The default port is **9083**.

Table 6-3 Parameters for creating a connection to Doris

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.

Parameter	Configuration
Connection Name	The default name is **Doris-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Doris Credential	Select the Doris credential added to the Edge device. For details about how to add credentials, see "Big data - Doris" in **Adding Resource Credentials**.
Database IP Address	Enter the IP address for accessing the Doris cluster.
Database Port	Enter the port used for accessing the Doris cluster. The default value is 3306.
Database Name	Enter the name of the Doris database.

Table 6-4 Parameters for creating a connection to HBase

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **HBase-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment. **CAUTION** You need to configure **/etc/hosts** on the server where the Edge is installed, so the HBase metadata connection can be established. For more information, see **Preparing for Development and Operating Environments**.
HBase Credential	Select the HBase credential added to the Edge device. For details about how to add credentials, see "Big data - HBase" in **Adding Resource Credentials**.
Secured Cluster	Choose whether the cluster is secured.
ZooKeeper IP Address	Enter the IP address for connecting to the ZooKeeper node. You can enter the public or private IP address of the ZooKeeper node.
ZooKeeper Connection Port	Enter the port for connecting to the ZooKeeper node.
HBase Version	Select the HBase version.

Table 6-5 Parameters for creating a connection to ClickHouse

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
ClickHouse Credential (Optional)	Select the ClickHouse credential added to Edge. For details about how to add credentials, see "Big data - ClickHouse" in **Adding Resource Credentials**.
Secured Cluster	Choose whether the cluster is secured.
ClickHouse IP Address	Enter the IP address of the server that hosts ClickHouse.
HTTP Port	If the ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. The default value is 8123.
HTTP SSL/TLS Port	If the ClickHouse cluster is secured, enter the HTTPS port for communicating with the ClickHouse server.

Table 6-6 Parameters for creating a connection to Delta Lake (with metadata)

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **Delta-Lake-with-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.

Parameter	Configuration
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 6-7 Parameters for creating a connection to Delta Lake (without metadata)

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **Delta-Lake-without-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 6-8 Parameters for creating a connection to Hudi (with metadata)

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.

Parameter	Configuration
Connection Name	The default name is **Hudi-with-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 6-9 Parameters for creating a connection to Hudi (without metadata)

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **Hudi-without-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.

Parameter	Configuration
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 6-10 Parameters for creating a connection to MaxCompute

Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **MaxCompute-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the **Edge device** in the source environment.
Alibaba Cloud Credential	Select the credential of the source Alibaba Cloud account. For details about how to add credentials, see "Big data - MaxCompute" in **Adding Resource Credentials**.
MaxCompute Project	Enter the name of the MaxCompute project to be collected. You can obtain the project name from the MaxCompute console.
Endpoint	Enter the endpoint of the region where the MaxCompute project is located. For details about the MaxCompute endpoints in different regions, see **MaxCompute Endpoints**.

Table 6-11 Parameters for creating a connection to ApsaraDB for ClickHouse

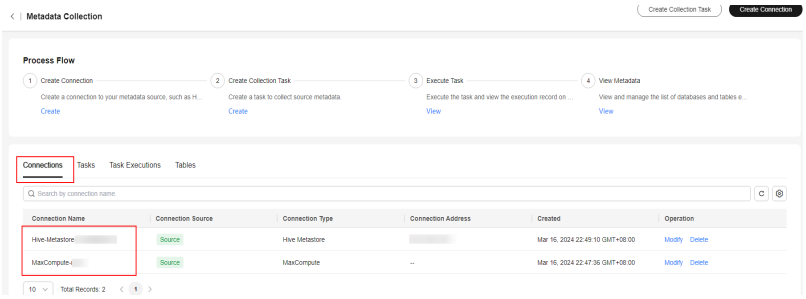
Parameter	Configuration
Connection To	You only need to create a metadata connection to the source.
Connection Name	The default name is **ApsaraDB for ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.

Parameter	Configuration
Edge Device	Select the **Edge device** in the source environment.
ClickHouse Credential	Select the ApsaraDB for ClickHouse credential added to Edge. For details about how to add credentials, see "Big data - ClickHouse" in **Adding Resource Credentials**.
Database URL	Enter the public address of the source ClickHouse cluster. You can view the IP address in the cluster details.

Step 6 Click **Test**. MgC verifies whether the connection can be established using the configuration information. If the test is successful, the connection can be established.

Step 7 After the connection test is successful, click **Confirm**. The metadata connection is created.

Step 8 Click **Metadata connections** in the **Metadata Collection** area. On the **Connections** tab, you can view the list of created metadata connections and their basic information. Locate the connection you created and in the **Operation** column, click **Modify** to modify the connection settings.



----End

6.2.2 Creating a Metadata Collection Task

Use the created metadata connection to create a collection task that gathers the details about source databases, tables, and fields to MgC.

You can create metadata collection tasks for:

- Hive Metastore
- Doris
- HBase
- ClickHouse
- Hudi (with metadata)
- Hudi (without metadata)
- MaxCompute
- ApsaraDB for ClickHouse

You can also add metadata about tables manually or by importing a file. For details, see **Viewing Metadata**.

Prerequisites

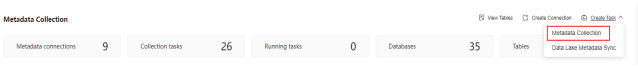
You have **created a metadata connection**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Metadata Collection** area, choose **Create Task** > **Metadata Collection**.



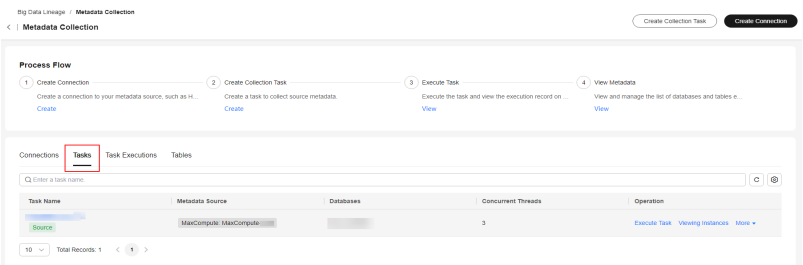
Step 4 Set task parameters based on **Table 6-12**.

Table 6-12 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Metadata-Collection-***4 random characters* (including letters and numbers). You can also specify a name.
Metadata Connection	Select the **created metadata connection**.
Databases	Enter the names of the databases whose metadata needs to be collected. **NOTICE** This parameter is mandatory only if a **MaxCompute** metadata connection is selected. In other cases, it is optional. If it is left blank, the metadata of all databases is collected.
Concurrent Threads	Set the number of concurrent threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. More concurrent threads means more efficient collection, but more connection and Edge device resources will be consumed.

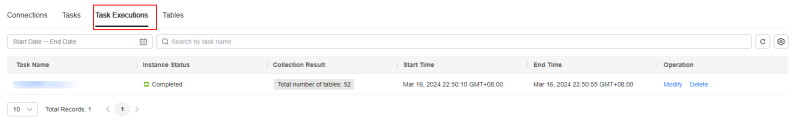
Step 5 Click **Confirm**. The metadata collection task is created.

Step 6 Click **Collection tasks**. Under **Tasks**, you can review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

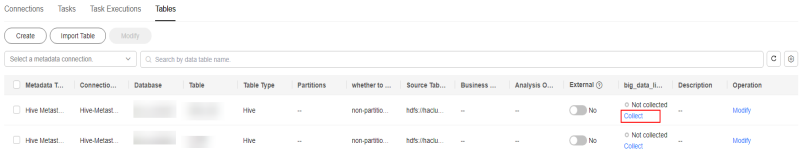


Step 7 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 8 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection result of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.



Step 9 Locate a table and click **Collect** in the **Lineage** column to **create a lineage collection task**.



----End

6.2.3 Creating a Data Lake Metadata Collection Task

Use the created metadata connection to create a collection task that gathers the details about source databases, tables, and fields to MgC. Only Delta lakes with metadata are supported.

For Delta lakes without metadata, you can manually add information about database tables one by one or import them in a batch. For details, see **Viewing Metadata**.

Prerequisites

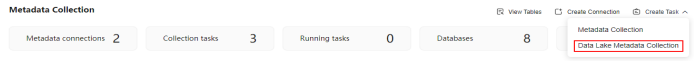
You have **created a connection to Delta Lake (with metadata)**.

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Metadata Collection** area, choose **Create Task** > **Data Lake Metadata Collection**.



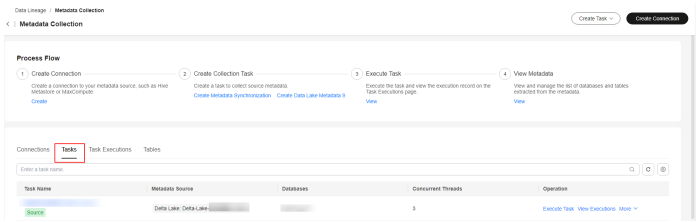
Step 4 Set task parameters based on **Table 6-13**.

Table 6-13 Parameters for configuring a metadata collection task

Parameter	Configuration
Task Name	The default name is **Data-Lake-Metadata-Sync-Task-***4 random characters* (including letters and numbers). You can also specify a name.
Metadata Connection	Select the created **connection to Delta Lake (with metadata)**.
Databases	Enter the names of the databases whose metadata needs to be collected.
Concurrent Threads	Set the maximum number of threads for executing the collection. The default value is **3**. The value ranges from **1** to **10**. More concurrent threads means more efficient collection, but more connection and Edge device resources will be consumed.
Custom Parameters	You can configure custom parameters to specify the tables and partitions to collect or set criteria to filter tables and partitions. • If the metadata source is Alibaba Cloud EMR, add the following parameter: – **Parameter**: **conf** – **Value**: **spark.sql.catalogImplementation=hive** • If the source is Alibaba Cloud EMR Delta Lake 2.2 and is accessed through Delta Lake 2.3, add the following parameters: – **Parameter**: **master** – **Value**: **local** • If the source is Alibaba Cloud EMR Delta Lake 2.1.0 and Spark 2.4.8 is configured to process Delta Lake data, add the following parameters: – Parameter: **mgc.delta.spark.version** – Value: **2** • If the source is Alibaba Cloud EMR and Spark 3 is configured to process Delta Lake data, add the following parameters: – Parameter: **jars** – **Value**: **'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'** **CAUTION** Replace the parameter values with the actual environment directory and Delta Lake version.

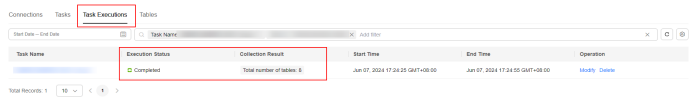
Step 5 Click **Confirm**. The metadata collection task is created.

Step 6 Click **Collection tasks**. Under **Tasks**, you can review the created metadata collection task and its settings. You can modify the task by choosing **More** > **Modify** in the **Operation** column.

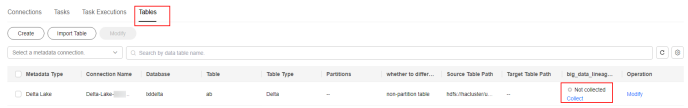


Step 7 Click **Execute Task** in the **Operation** column to run the task. Each time the task is executed, a task execution is generated.

Step 8 Click **View Executions** in the **Operation** column. Under **Task Executions**, you can view the execution records of the task and the status and collection results of each task execution. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.



Step 9 Locate a table and click **Collect** in the **Lineage** column to **create a lineage collection task**.



----End

6.2.4 Viewing Metadata

You can collect metadata about databases and tables by creating a collection task or manually adding or importing metadata to MgC.

Creating a Collection Task

For details about how to create a collection task, see **Creating a Metadata Collection Task**. When a task execution enters a **Completed** status and the collection results are displayed, you can view the list of databases and tables extracted from collected metadata on the **Tables** tab.

Manually Adding a Table

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Metadata Collection** area, click **View Tables**.

Step 4 On the displayed page, click **Add**.

Step 5 Configure the parameters listed in **Table 6-14**.

Table 6-14 Parameters for adding a table

Parameter	Description
Metadata Connection	Select the **created metadata connection**.
Database	Enter a name.
Table	Enter a name.
Source Table Path (Optional) It is mandatory only if a Delta Lake connection is selected for **Metadata Connection**.	Enter the path for storing the data table in the source database.
Target Table Path (Optional) It is mandatory only if a Delta Lake connection (without metadata) is selected for **Metadata Connection**.	Enter the path for storing the data table in the target database.
Partition Info (Optional)	User-defined
Is Partitioned (Optional)	Specify whether the table is partitioned.
Business Owner (Optional)	You can specify the owner of the business that the data table belongs to.
Analysis Owner (Optional)	Specify the analysis owner of the table.
External	Set it as required. If the table is external, the consistency verification result of the table is not displayed.
Description (Optional)	Enter a table description.

Step 6 Click **Confirm**. Then you can view the table in the table list.

----End

Importing Tables

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Metadata Collection** area, click **View Tables**.

Step 4 On the displayed page, click **Import**.

Step 5 Select the **created metadata connection**.

Step 6 Click **Download** to download the import template to the local PC. Open the import template and fill in the template with the information of the tables to be added.

NOTICE

- A maximum of 10,000 tables can be imported at a time.
- The tables to be imported must come from the same metadata source.
- If a connection to Delta Lake (with metadata) is selected for **Metadata Connection**, the **source_path** parameter in the template is mandatory.
- If a connection to Delta Lake (without metadata) is selected for **Metadata Connection**, the **source_path** and **target_path** parameters in the template are mandatory.

Step 7 Go back to the console and click **Select File** to upload the populated template to MgC.

Step 8 Click **Confirm**. The system starts the import. A task execution record is generated for this import. The execution status represents the import status.

----End

6.3 Collecting Data Lineage

6.3.1 Overview

Collecting data lineage can capture relationships and dependencies between data elements. MgC can virtualize the collected data lineage into graphs.

The following describes the types of jobs data lineage can be collected from and the collection precautions.

Job Type	Collection Precaution
Lineage template	You need to download the lineage template and fill it with your data.

6.3.2 Creating a Lineage Collection Task

Prerequisites

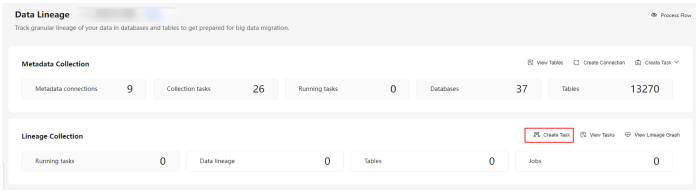
You have **collected metadata**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Lineage Collection** area, click **Create Task**.

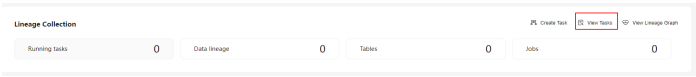


Step 4 Select a job type and configure the parameters shown.

Type	Parameter	Configuration
Lineage template	File	Download the lineage template to the local PC and set parameters in the template. The following fields are mandatory: • Target Database (TargetDataset) • Target Table (TargetTable) • Target Connection Name (TargetConnection-Name) • Target Component Type (TargetComponent-Type) • Upstream Database (SourceDataset) • Upstream Table (SourceTable) • Upstream Connection Name (SourceConnectionName) • Upstream Component Type (SourceComponentType) • Job ID (JobId) **NOTICE** • The value of **Target Component Type** and **Upstream Component Type** in the template can be **Hive SQL** or **MaxCompute**. • Cells in the template cannot contain formulas. Otherwise, the parsing will fail. Go back to the console and click **Select File** to upload the saved file to MgC. **CAUTION** The file size cannot exceed 100 MB.

Step 5 Click **Confirm**. The data lineage collection task is created. The system automatically starts collecting data lineage.

Step 6 Click **View Tasks**. On the displayed page, you can view the collection task in the task list.



Step 7 Wait until the task status changes to **Completed**. Then click **View Lineage** in the upper right corner of the page to view the **Lineage Graph**.

----End

6.3.3 Viewing Data Lineage

Prerequisites

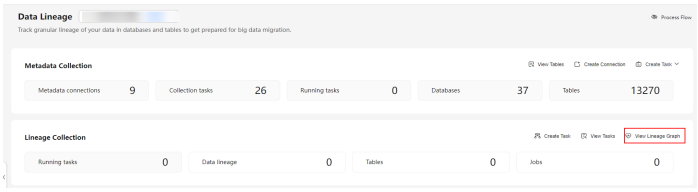
A **lineage collection task** has been completed.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Research** > **Data Lineage**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Lineage Collection** area, click **View Lineage Graph**.



Step 4 On the displayed page, view the lineage graph by database and table. If big data verification is performed, you can also filter data by verification batch.

----End

7 Migration Solutions

7.1 Associating Source Servers with Target Servers

You can associate source servers with existing servers on Huawei Cloud. Then your workloads can be migrated to these Huawei Cloud servers.

For source servers with target servers associated, there is no need to **assess** them before you create workflows to execute migrations.

Prerequisites

You have finished resource discovery and collection by referring to **Simple Project**.

Notes

A target server must meet the following requirements:

- The target server must be stopped.
- During the migration, disks on the target server are formatted and re-partitioned based on the source disk settings for receiving data migrated from the source server.
- To migrate over the Internet, the target server must be able to access the Internet.
- The target server must be in the same region as the **application** to which the source server is added.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Design** > **Migration Solutions**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **View** in the **Target Configuration** card.

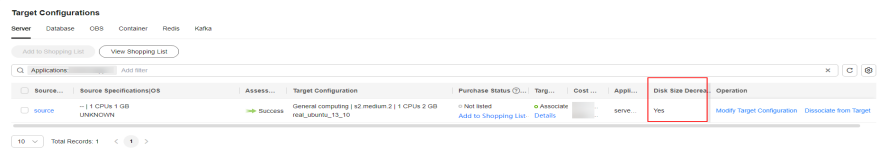
Step 4 On the displayed **Servers** tab, locate a source server and click **Associate** in the **Target Association** column.

Step 5 In the displayed dialog box, select the region of the **application** and select a project. You can select a target server based on the disk usage of the source server. This allows you to choose a target server with a smaller disk capacity or fewer disks than the source server, reducing disk capacity, optimizing resource utilization, and avoiding unnecessary costs.

NOTICE

- Disk downsizing is only available for Linux, and the decreased sizes must be larger than the used sizes of the source disks.
- In the cross-AZ migration scenario, only disk upsizing is supported. Even if you choose to downsize disks here, the settings will not be applied, and the system will create target disks as large as source disks.

Step 6 Click **OK**. After the association is complete, **Associated** appears in the **Target Association** column. You can click **Details** to view the specifications of the associated target server. The system will automatically check whether the associated target server has smaller disks than the source server. If it does, **Yes** will appear in the **Disk Downsized** column. If it does not, **No** will show up.



----End

Dissociating a Source Server from the Target Server

Step 1 On the **Servers** tab, locate the source server and click **Dissociate from Target** in the **Operation** column.

Step 2 In the displayed dialog box, click **Yes**. Please note that this operation cannot be undone.

----End

7.2 Getting Target Recommendations

Assessing an application can generate recommendations for rightsized Huawei Cloud resources based on the specifications, performance, and business purpose of the source resources added to the application. These recommendations also take into account your requirements for cost, availability, and compliance.

Target recommendations can be generated for servers, databases, object storage, containers, and middleware.

NOTE

If your source servers have been **associated with existing servers** on Huawei Cloud, you can skip this section and directly create a migration workflow to migrate them.

Prerequisites

You have **discovered resources**, and **grouped the resources as applications**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Design** > **Migration Solutions**. Select a **migration project** in the upper left corner of the page.

On the **Migration Solutions** page, you can view how many source resources and applications are managed in the current migration project, as well as whether the source resources have been configured with target resources.

Step 3 Click **Assess** in the **Target Configuration** card.

Step 4 In the **Select Application** drop-down list, select the application you want to assess.

Step 5 In the **Select Resources** area, select the application resources to be assessed.

Step 6 Configure the policy used to compute target recommendations based on **Table 7-1**.

Table 7-1 Settings used for computing target recommendations

Parameter	Description
Target Region	Select the region where you want to purchase resources on Huawei Cloud. You are advised to select a region close to your target users for lower network latency and quick access.
Assessment Policy	• Match source configuration MgC recommends the right Huawei Cloud resources based on source resource specifications. • Match business scenario MgC recommends the right Huawei Cloud resources based on the business scenario of source resources and Huawei Cloud best practices. • Cross-AZ migration This policy only applies to migration of ECSs between AZs on Huawei Cloud, and MgC only assesses servers in the application. You need to select the target AZ you want to migrate to. For details about how the system recommends appropriate target resources, see **How Does MgC Generate Target Recommendations?**

Parameter	Description
Priority	• High performance MgC recommends target resources with optimal performance. • Low cost MgC recommends the most cost-effective target resources that meet your demands.
Preferences	Configure your preferences for target servers. Your preferences will be first attempted to be matched during the assessment. For details about how the system recommends appropriate target resources, see **How Does MgC Generate Target Recommendations?** • Server Types (Optional) Select the server types you prefer. • Server Series (Optional) Select the server series you prefer. The system will generate recommendations based on your preferred server types and series. **NOTICE** If you select **Display only series allowed on DeHs**, **Server Types** will be dimmed, and the server series allowed on DeHs in the target region will be listed. • System Disk (Optional) Select the system disk type you prefer. • Data Disk (Optional) Select the data disk type you prefer. • Sizing Criteria Choose the criteria that the system will use to generate server recommendations. – If you select **As-in on source**, the system will recommend target servers with the same or as close CPU and memory capacity as the source servers. – If you select **Performance-based**, you need to perform a **performance collection** for the source servers, and then set assessment parameters. The system will then recommend target servers with your desired CPU and memory capacity. **NOTICE** The more performance data is collected, the more accurate the recommendations are. The collection of server performance data should take no less than seven days.
	For the container assessment, configure parameters such as **Cluster Type**, **Cluster Version**, and **Container Network Model** for getting recommendations for container resources.

Step 7 Click **Create Assessment**. After the assessment is complete, you can **review the target recommendations**. You can also **view the performance data of source servers**.

Step 8 (Optional) Perform the following operations:

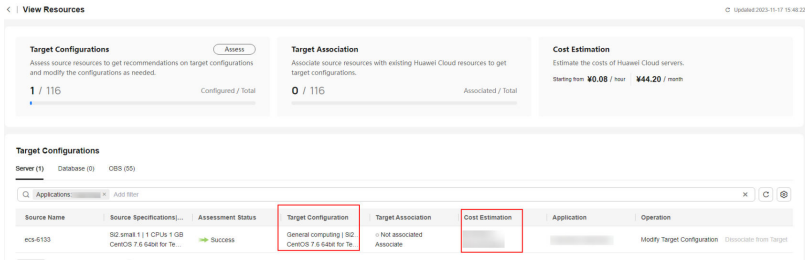
- **Modify target recommendations**. You can modify the recommended specifications for target servers and their disks.
- **Associate source servers with target servers**. If you find that you already have servers that match your requirements on Huawei Cloud based on the target server recommendations, you can associate them with source servers.
- **Associate source servers with DeHs** if you want to migrate to Huawei Cloud DeH.

----End

Viewing Target Recommendations

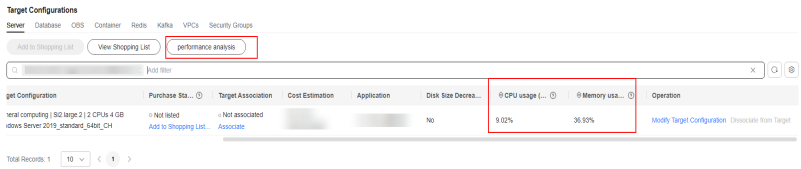
In the application list on the **Migration Solutions** page, click **View Target Configurations** in the **Operation** column.

In the **Target Configurations** area, you can view the specifications of Huawei Cloud resources recommended based on the source resource specifications and your preferences. It also gives you the ability to estimate what it will cost to run your services on Huawei Cloud.



Viewing Server Performance Data

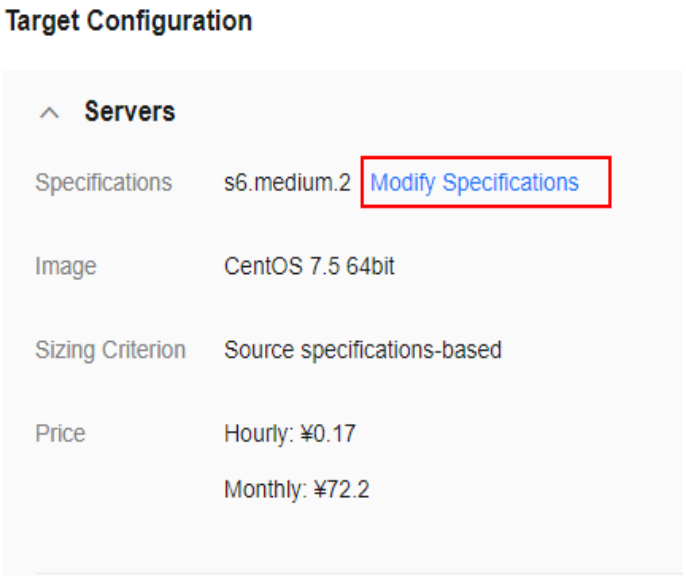
On the **Target Configurations** page, in the server list, you can view the average CPU and memory usage of each server over the last 7 or 30 days. Click **Performance Analysis** to view the performance statistics of all servers.



Modifying Target Recommendations

Step 1 In the **Target Configurations** area, locate the server that you want to modify the recommended target configurations for and click **Modify Target Configuration** in the **Operation** column.

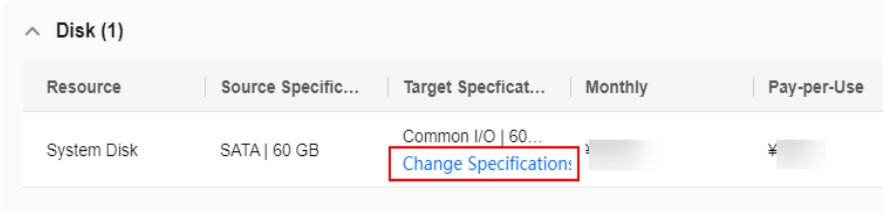
Step 2 Modify the specifications and image for the target server.



Step 3 In the disk area, locate a disk and click **Modify Specifications** in the **Target Specifications** column. You can modify the disk type and capacity. You can downsize disks for Linux target servers based on the disk usage of the source servers. If you downsize a disk for the target server, the system will set **Disk Downsized** to **Yes**. The reverse also applies.

NOTICE

- The system disk capacity ranges from 40 GB to 1,024 GB.
- The data disk capacity ranges from 10 GB to 32,768 GB.
- Disk downsizing is only available for Linux, and the decreased sizes must be larger than the used sizes of the source disks.
- In the cross-AZ migration scenario, only disk upsizing is supported. Even if you choose to downsize disks here, the settings will not be applied, and the system will create target disks as large as source disks.



----End

Associating Source Servers with DeHs

After you associate source servers with DeHs, the system will automatically provision the recommended target servers on the DeHs if you:

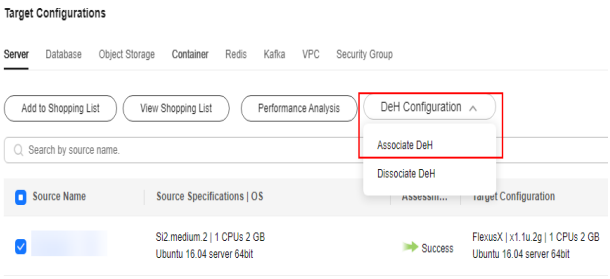
- Buy these recommended target servers using a shopping list.

- Create migration workflows to migrate these source servers.

Step 1 In the **Target Configurations** area, under **Server**, select the source server to be associated with the same DeH.

Step 2 Choose **DeH Configuration** > **Associate DeH** above the list.

Figure 7-1 Associating source servers with a DeH



Step 3 Select the target DeH to be associated from the drop-down list and click **Confirm**. **Associated** will show up in the **DeH** column. You can click **Details** to view the associated DeH.

----End

7.3 Purchasing Resources

After the specifications of target resources are confirmed, you can directly purchase these target resources in batches using a shopping list on MgC. Currently, only servers can be purchased this way. The purchased target servers will be automatically associated with the paired source servers. You can then manually redeploy your services on these target servers. Alternatively, you can create server migration workflows from the beginning to automate the entire purchase and service migration process.

Prerequisites

You have **got target recommendations**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Design** > **Migration Solutions**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **View** in the **Target Configuration** card.

Step 4 In the **Target Configurations** area, under **Server**, select a server whose **Assessment Status** is **Success**, and click **Add to Shopping List** in the **Purchase Status** column.

Step 5 On the top of the page, click **View** in the **Shopping List** card.

Step 6 In the **Process Flow** area, click **View Templates**. In the **My Templates** window that is displayed on the right, click **Create Template**.

Step 7 In the **Basic Info** area, select a template type (only server is supported currently) and specify a template name. In the **Configuration Info** area, set template parameters based on **Table 7-2**.

Table 7-2 Parameters for configuring a shopping template

Parameter	Configuration
Region	Select the region where you want to deploy target servers.
Project	Select a project in the target region.
AZ	**Random** is the default setting. You can also select an AZ.
VPC	Select a VPC in the target region.
Subnet	Select a VPC subnet. The subnet CIDR block must be within the selected VPC CIDR block.
Security Group	Select a security group that meet the following requirements: • Inbound ports 8899, 8900, and 22 are allowed for Windows migrations. • Inbound port 22 is allowed for Linux migrations
Enterprise Project	Select a commercial enterprise project that you want add purchased target servers to. The default enterprise project is **default**. An enterprise project makes it easy to manage projects and group cloud resources and users. For details about creating and managing enterprise projects, see the **Enterprise Management User Guide**.

Step 8 Click **Save**. You can view the created template in the template list.

Step 9 Return to the shopping list page, and click **Attach Shopping Template** in the **Shopping Template** column. In the displayed dialog box, select the created template from the drop-down list and click **Confirm**.

Step 10 Click **Buy** in the **Operation** column. Wait until the purchase status changes to **Purchased**, and view the ID of the purchased resource.

----End

7.4 Assessing Migration Risks

You can assess potential migration risks before performing a migration. The assessment results will indicate whether your source servers are compatible with the target cloud in terms of OS, firmware, and specifications.

NOTICE

Only servers can be assessed for migration risks.

Prerequisites

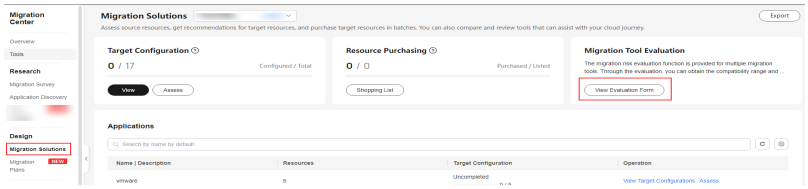
You have added source server information to MgC manually or through an Internet- or intranet- based discovery task.

Procedure

Step 1 Sign in to the **MgC console**.

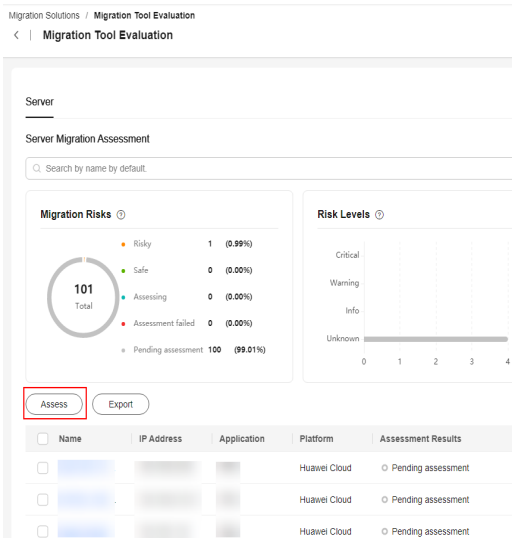
Step 2 In the navigation pane on the left, choose **Design** > **Migration Solutions**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **View Evaluation List** in the **Migration Tool Evaluation** card.

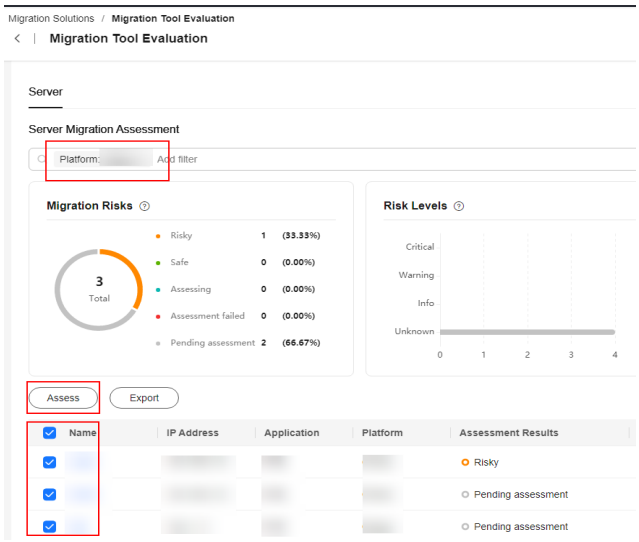


Step 4 On the **Server** tab, you can view the list of all source servers.

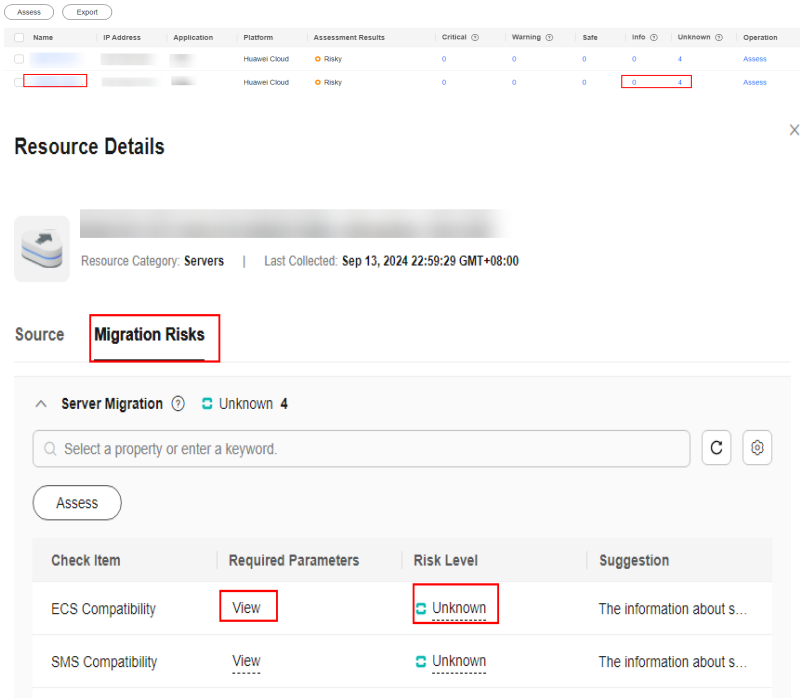
- To assess all servers in the **Pending assessment** state, click **Assess**. In the dialog box that is displayed, click **OK**.



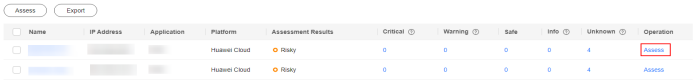
- To assess certain servers, search for them by name, IP address, platform, or application and select them. Click **Assess**. In the displayed dialog box, confirm the number of servers to be assessment and click **OK**.



Step 5 In the assessment list, view the assessment status and the number of migration risks at each level. Click the server name or a number in the row. On the **Migration Risks** tab displayed on the right, view all check items, server information required for checking each item, and risk levels.



Step 6 Rectify the migration risks on a server based on the suggestions provided by the system. Then click **Assess** in the **Operation** column to assess the server again.



----End

8 Migration Plans

8.1 Creating a Server Migration Plan

A server migration plan outlines the source servers to be migrated and the paired target servers. You can create a migration workflow on the top of a migration plan to implement batch migration. The following content describes how to create a server migration plan.

Preparations

- **Create a migration project**.
- Discover source servers and collect server details. MgC provides three collection methods to meet your requirements in different scenarios.
  - If your source servers are on a cloud platform, such as Alibaba Cloud, Huawei Cloud, AWS, Tencent Cloud, Google Cloud, or Azure, you can **collect the data about your servers over the Internet**. You can also **manually add the server details to MgC**.
  - If your source servers are in an on-premises IDC, you can **collect the server details over the intranet**. MgC enables you to scan for servers by network range or VMware vCenter Server. You can also **manually add the server details to MgC**.
- Install Edge in the source intranet environment and register an account. For details, see **Installing Edge for Windows**. Then **connect Edge to MgC**.

Notes and Constraints

- Only one migration workflow can be created for a migration plan.
- A server can only be added to one migration plan.
- A maximum of 100 servers can be added to a migration plan.

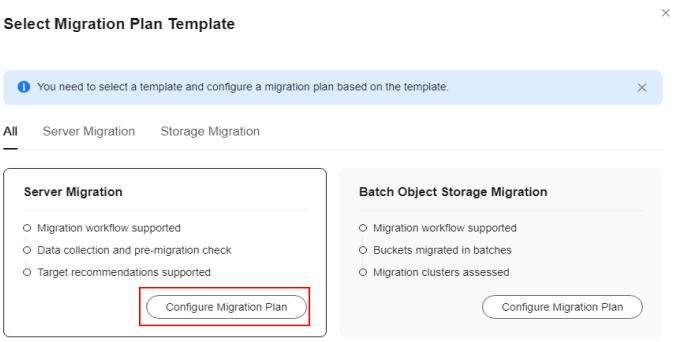
Procedure

Step 1 Sign in to the **MgC console**. In the upper left corner of the page, select the migration project you created.

Step 2 In the navigation pane, choose **Design** > **Migration Plans**.

Step 3 Click **Create Migration Plan** in the upper right corner of the page.

Step 4 In the **Server Migration** card, click **Configure Migration Plan**.

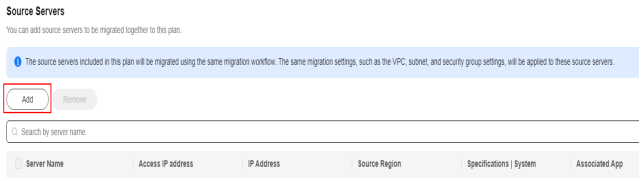


Step 5 In the **Basic Settings** area, set parameters listed in **Table 8-1**.

Table 8-1 Basic settings

Parameter	Configuration
Migration Plan Name	User-defined
Description (Optional)	
Source Platform	Select the cloud platform where the source servers are located. **NOTICE** A migration plan can only include servers from the same platform.
Target Region	Select the region you want to migrate to.

Step 6 In the **Source Servers** area, click **Add**.



Step 7 Select the servers to be migrated together and click **Confirm**.

CAUTION

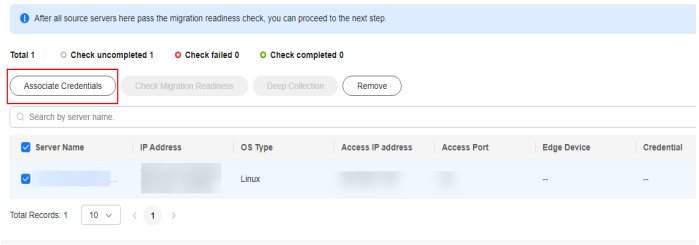
- The source servers included in this plan will be migrated using the same migration workflow. The same migration settings, such as the VPC, subnet, and security group settings, will be applied to these source servers.
- The selected servers must come from the source platform selected in **step 4**.

Step 8 Click **Save and Next** to go to the **Check Source Servers** phase.

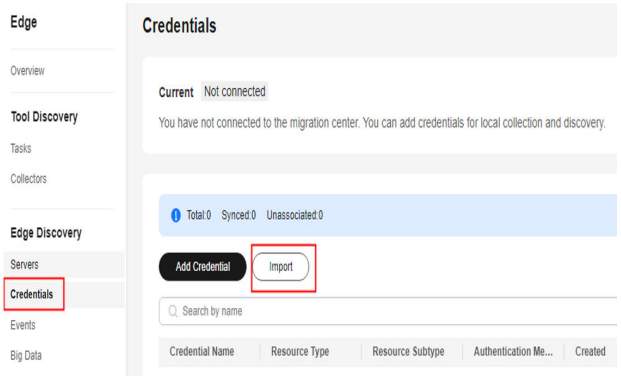
Step 9 Configure server credentials. There are two methods for you to do this:

- **Manual addition**: You can choose this method if you configure credentials for a small number of servers. For details, see **Adding Resource Credentials**.
- **Batch addition**: You can choose this method if you configure credentials for a large number of servers.

  a. Click **Associate Credentials** above the list of source servers to download the template for batch configuring credentials. Open the template file, enter the credential information (username and password) of all servers in the current migration plan, and save the file.

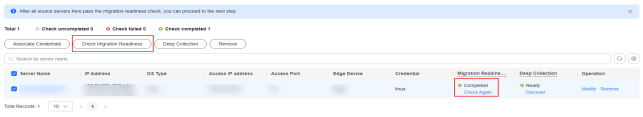


  b. Import the saved template file to the **Edge device you connect to MgC**. The system automatically configures the Edge device and credentials for the servers.



You can use **Modify** in the **Operation** column to update settings, such as **Access IP Address**, **Access Port**, **Edge Device**, and **Credential**.

Step 10 Select all the servers and click **Check Migration Readiness** above the list. The system checks if the source servers are ready for migration.



- If the **Migration Readiness Check** column displays a **Completed** status, the source server passes the check and is ready for migration.
- If the **Migration Readiness Check** column displays a **Check failed** status and the failure cause is that certain source server information is missing, click **Collect** in the **Deep Collection** column to collect the server details.
- If the **Migration Readiness Check** column displays a **Check failed** status and the failure cause is that the source server is unreachable, check whether the

access information you provided is correct. If it is correct, check the migration readiness again.

Step 11 After **Completed** shows up in the **Migration Readiness Check** column for all the source servers, click **Next** to go to the optional phase **Conduct Performance Collection**.

Step 12 (Optional) Collect server performance data. The collected data can be used to get target server recommendations in the next step. A collection takes some time because performance data is sampled from source servers once every 5 minutes. You can determine whether to perform performance collection. For more constraints, see **Precautions for performance collection**. If you do not need to collect server performance data, click **Next** to go to the **Configure Target** phase.

Select the source servers whose performance data needs to be collected and choose **Performance Collection** > **Start** above the list. **Collecting** will be displayed in the **Performance Collection** column

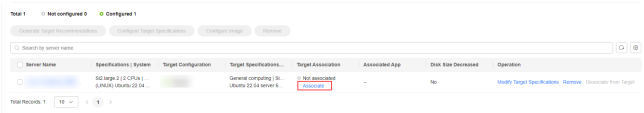
After the performance data collection is complete, you can:

- Click **View Performance** in the **Operation** column to view the collected server performance data.
- Choose **Performance Collection** > **Stop** above the list to suspend performance data collection for servers.

Step 13 Configure target servers. You can use the following methods to complete this job:

- **Associating source servers with target servers:** If you have cloud servers on Huawei Cloud, you can associate them with the source servers. Then in the migration workflow, these cloud servers will receive data and workloads migrated from the source servers. For requirements about your existing cloud servers (target servers), see **Associating Source Servers with Target Servers**.

  a. Locate a source server in the list, click **Associate** in the **Target Association** column.

  b. In the displayed dialog box, select the region and project where the existing cloud servers are located. If the source server's storage disks are overprovisioned, regardless of the quantity or space, you can choose an existing cloud server that precisely fits your needs to optimize resource utilization and save costs.

  NOTICE

  - Ensure that the target server has an EIP bound.
  - Disk downsizing is only available for Linux, and the decreased sizes must be larger than the used sizes of the source disks.
  - In the cross-AZ migration scenario, only disk upsizing is supported. Even if you choose to downsize disks here, the settings will not be applied, and the system will create target disks as large as source disks.

c. Click **Confirm**. After the association is complete, **Associated** shows up in the **Target Association** column. You can click **Details** to view the details of the associated target server. The system will automatically check whether the associated target server has downsized disks compared with the source server. If it is, **Yes** will be displayed in the **Disk Downsized** column. If it is not, **No** will be displayed.

- **Generating target server recommendations:** You can choose to let the system recommend you rightsized Huawei Cloud servers based on the source server specifications. You can also manually adjust the recommended server specifications. In the subsequent migration workflow, the target servers will be automatically created based on the specifications configured here. Source servers with target servers associated will be skipped during the recommendation process.

  a. In the source server list, select the ones that you want to generate target server recommendations for and click **Generate Target Recommendations** above the list.

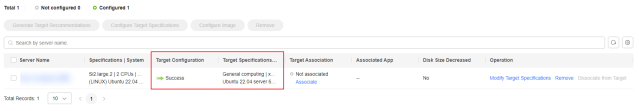
  b. Configure the policy used to compute target recommendations based on **Table 8-2**.

  Table 8-2 Settings used for computing target recommendations

Parameter	Description
Priority	▪ High performance MgC recommends target servers with optimal performance. ▪ Low cost MgC recommends the most cost-effective target servers that meet your demands.

Parameter	Description
Preferences	Configure your preferences for target servers. Your preferences will be the first attempted to be matched during the recommendation. For details about how the system recommends appropriate target resources, see **How Does MgC Generate Target Recommendations?** ■ Server Types (Optional) Select the server types you prefer. ■ Server Series (Optional) Select the server series you prefer. The system will generate recommendations based on your preferred server types and series. **NOTICE** If you select **Display only series allowed on DeHs**, **Server Types** will be dimmed, and the server series allowed on DeHs in the target region will be listed. ■ System Disk (Optional) Select the system disk type you prefer. ■ Data Disk (Optional) Select the data disk type you prefer. ■ Sizing Criteria Choose the criteria that the system will use to generate server recommendations. ○ If you select **As-in on source**, the system will recommend target servers with the same or as close CPU and memory capacity as the source servers. ○ If you select **Performance-based**, you need to perform a **performance collection** for the source servers, and then set assessment parameters. The system will then recommend target servers with your desired CPU and memory capacity. **NOTICE** The more performance data is collected, the more accurate the recommendations are. The collection of server performance data should take no less than seven days.

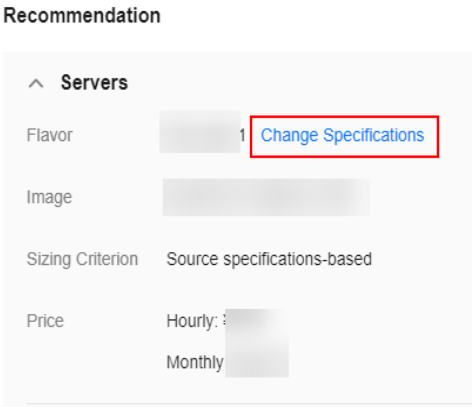
c. Click **Confirm**. If **Success** shows up in the **Target Configuration** column, the target recommendation is generated. You can check the recommended configuration in the next column



d. (Optional) Modify target server specifications recommended by the system if needed.

i. Locate a source server and click **Modify Target Specifications** in the **Operation** column.

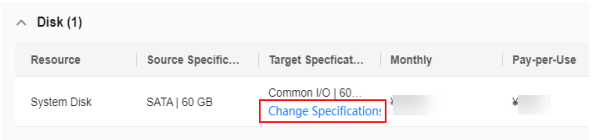
ii. Modify the specifications and image for the target server.



iii. In the disk area, locate a disk and click **Modify** in the **Target Specifications** column to modify the disk type and capacity. Only disks on Linux target servers can be downsized if the paired source servers have over-provisioned storage resources. If you downsize a disk for the target server, the system will set **Disk Downsized** to **Yes**. The reverse also applies.

NOTICE

- The system disk capacity ranges from 40 GB to 1,024 GB.
- The data disk capacity ranges from 10 GB to 32,768 GB.
- Disk downsizing is only available for Linux, and the decreased sizes must be larger than the used sizes of the source disks.
- In the cross-AZ migration scenario, only disk upsizing is supported. Even if you choose to downsize disks here, the settings will not be applied, and the system will create target disks as large as source disks.



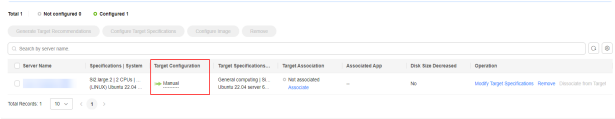
- **Configuring target server specifications:** You can manually configure the same target server specifications for the source servers all at once.

CAUTION

When selecting server specifications, it is crucial to ensure they meet your requirements. The system does not check whether they match the source servers. MgC does not assume responsibility for any issues arising from inappropriate server specifications.

a. In the server list, select the source servers and click **Configure Target Specifications** above the list.

b. Select the target server specifications and click **OK**. Then you can see **Target Configuration** changes to a **Manual** status.



- **Configuring an image for target servers**: You can select an image to apply to all the target servers.

CAUTION

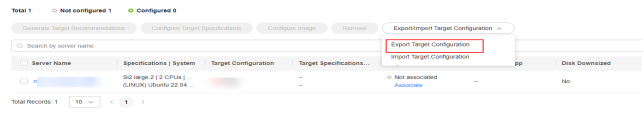
When selecting an image, it is crucial to ensure it meets your requirements. The system does not check whether it matches the source servers. MgC does not assume responsibility for any issues arising from inappropriate images.

a. In the server list, select the source servers and click **Configure Image** above the list.

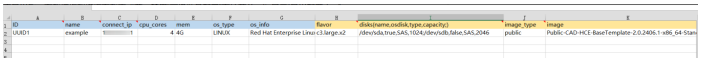
b. On the displayed page, choose an image type, an OS, and an OS version and click **Confirm**. If a target server has a system disk smaller than what defined in the image you choose, configuring the image will fail for that server.

- **Exporting and then importing target server configurations**: You can export the target server configurations, modify or configure the server, disk, and image settings, and then import the updated configurations back. Some technical expertise is needed for this process.

a. In the list of source servers, choose **Export/Import Target Configuration** > **Export Target Configuration** to download the server information to the local PC as an Excel file.



b. Open the exported file. On the first sheet named **servers**, the blue columns (columns A to G) indicate the source server information, which **does not need to be modified**. The yellow columns (columns H to K) indicate the target server information, which **needs to be configured**.

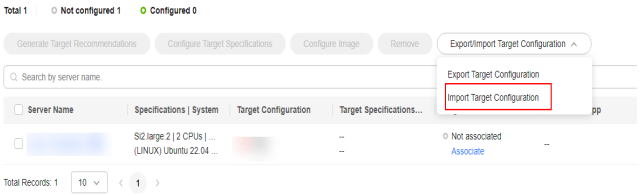


c. Configure the target server information (in columns H to K) according to **Table 8-3**.

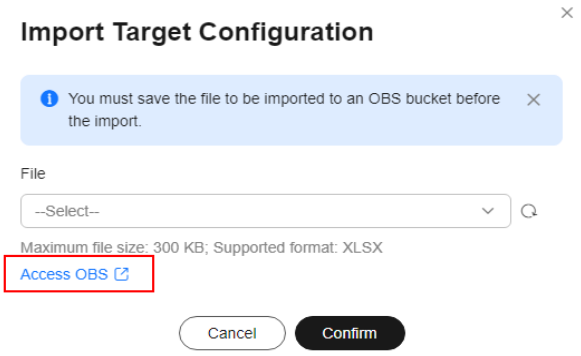
Table 8-3 Parameters in columns H to K

Parameter	Configuration
flavor	Enter the flavor name for the target server. The second sheet **flavors** lists the flavors of S- (general-purpose) and C- (computing-plus) series on sale in the target region. Choose a proper flavor and copy the flavor name to this cell. **NOTE** If a source server has been assessed, the recommended flavor is included in the exported file. If a source server has not been assessed, its **flavor** field is empty.
disks(name,osDisk,type,capacity;)	Enter the information about all disks. Separate multiple disks with semicolons (;). You must specify the name, function (system or not), type, and size for each disk. Separate these items with commas (,). Please keep in mind that the disk size is a number without unit. The default unit is GiB. Example: /dev/sda,true,SAS,1024;/dev/sdb,false,SAS,2046 **NOTE** If a source server has been assessed, the recommended disk specifications show up in the exported file. If a source server has not been assessed, the source disk specifications are used, except the disk type. The target disk type defaults to **SAS**.
image_type	Enter the image type. The options are: ■ **public**: public image ■ **private**: private image ■ **shared**: shared image
image	Enter the image name. You can view all available images in the target region on the third sheet **images**.

d. After configuring the target server information for all source servers, save the Excel file. Return to the MgC console. Choose **Export/Import Target Configuration** > **Import Target Configuration** above the list of source servers.



e. In the displayed dialog box, click **Access OBS** to go to the OBS page.



f. Select a bucket and upload the Excel file to the bucket.

g. After the upload is successful, return to the MgC console. In the **Import Target Configuration** dialog box, select the file stored in the OBS bucket from the drop-down list.

h. Click **Confirm** to import the target server configurations.

NOTE

- If the **flavor** parameter is empty in a row, this row will be ignored during the import.
- If the **disk** or **image** parameter is left empty in a row, the source server's disk and image information is used during the import.
- If the number in the **disk** field of a row is inconsistent with what the source server actually has, an error message is reported, indicating that the disk numbers are inconsistent.
- If in a row, the disk type conflicts with the flavor, an error message is displayed, indicating that the disk type is incorrect.

- **Associating source servers with DeHs** if you want to migrate to Huawei Cloud DeH. Then after you create a migration workflow for this plan, the system will provision the recommended target servers on the DeHs.

  a. In the source server list, select the ones to be associated with a DeH.

  b. Choose **DeH Configuration** > **Associate DeH** above the list.

  c. Select the target DeH to be associated from the drop-down list and click **Confirm**. **Associated** will show up in the **DeH** column. You can click **Details** to view the associated DeH.

Step 14 Perform subsequent operations as needed.

- Create a server migration workflow.

  Click **Create Workflow** to create a migration workflow. For more information, see **Step 5**.

- Only save the migration plan.

  Click **Confirm** to save the migration plan. Then you can:

  - Click **Design** in the **Operation** column to modify the migration plan.
  - Click **Create Workflow** in the **Operation** column to create a server migration workflow to migrate the source servers included in the migration plan. For more information, see **Step 5**.

----End

8.2 Creating a Batch Storage Migration Plan

You can configure migration plans using templates designed for different resources and create migration workflows on the top of these plans. You can also view the migration settings for resources by application. Follow this section to configure a migration plan for batch migrating object storage buckets.

If only one object storage bucket needs to be migrated, you can **create a migration workflow** directly.

Prerequisites

- Source buckets have been discovered over the Internet and their details have been collected.
- (Optional) A deep collection has been performed for the buckets to be migrated.

Notes

- A migration plan can be used to migrate buckets from the same source platform and to the same target region.
- A source bucket can be included in multiple migration plans.
- Source buckets in a migration plan may have different prefixes. You can give them the same prefix after they are migrated to the target bucket.

Procedure

Step 1 Sign in to the **MgC console**. In the navigation pane, choose **Design** > **Migration Plans**.

Step 2 Click **Create Migration Plan** in the upper right corner of the page.

Step 3 In the **Batch Object Storage Migration** card, click **Configure Migration Plan**.

Step 4 In the **Basic Settings** area, set parameters listed in **Table 8-4**.

Table 8-4 Basic settings

Parameter	Configuration
Migration Plan Name	User-defined
Description (Optional)	User-defined

Parameter	Configuration
Source Platform	Select the cloud platform where the source resources are located.
Target Region	Select the region where the target buckets are located.

Step 5 Above the source bucket list, click **Add**.

Step 6 Select the buckets to be migrated in a batch. For each of the selected buckets, click **Modify** in the **Operation** column to specify the migration method. For details about the migration methods, see **Table 8-5**. Then click **Confirm**.

CAUTION

- The selected resources must come from the source platform selected in **Basic Settings**.
- A maximum of 100 buckets can be added.

Table 8-5 Migration methods

Migration Method	Description
Full migration	Migrates all data in the source bucket or specified directories.
List migration	Migrates objects recorded in the list files. The list files must be stored in the target bucket.
Prefix migration	Migrates objects matched with the specified names or prefixes.

Step 7 In the list of buckets to be migrated, click **Modify** in the **Operation** column.

Step 8 Select the resource credential. It is best to enter the total number and size of objects to be migrated. Then click **Confirm**.

- If the migration method configured for the bucket is **List migration**, in the **List Path** box, enter the directory that stores the lists of objects to be migrated. The directory must end with a slash (/).
- If the migration method configured for the bucket is **Prefix migration**, add the names or prefixes of objects to be migrated. Note that you need to click **Save** each time you add an object name or prefix.

  NOTE

  If the files to be migrated are stored in the root directory of the source bucket, add their name prefixes directly. If the files are stored in a non-root directory, add their directories and name prefixes in the format of *Directory*|*Prefix*.

Step 9 After you configure the migration settings for all buckets to be migrated, click **Next**.

Step 10 For each source bucket, click **Modify** in the **Operation** column. Then select the credential used for accessing the target bucket, specify the target bucket, enter a prefix to rename or relocate migrated objects, and click **Save**.

Step 11 After you configure the target bucket settings for all buckets to be migrated, click **Next**. You can assess how large of a migration cluster is required for the migration and then create a migration cluster with the recommended specifications. Alternatively, you can skip this step and use an existing migration cluster. For details about how to create a cluster, see **Creating a Migration Cluster**.

Step 12 Click **Next** to select an existing migration cluster. All resources in this plan will be migrated using this cluster.

Step 13 Click **Select Cluster**. The **Select Cluster** dialog box is displayed on the right.

Step 14 In the cluster list, select the cluster you want to use and click **Confirm**. If you want to modify an existing cluster to meet your requirements, see **Managing a Migration Cluster**.

Step 15 Click **OK**. After migration plan is created, you can see it in the list.

- If you need to modify the plan settings, click **Design** in the **Operation** column.
- After **Completed** appears in the **Progress** column, click **Create Workflow** in the **Operation** column to **create a migration workflow** to migrate all buckets in the plan in a batch.

----End

8.3 Creating a Batch File Storage Migration Plan

Follow this section to configure a migration plan for batch migrating file systems.

The Batch Object Storage Migration template can be used to migrate multiple file systems in a batch. If only one file system needs to be migrated, **create a migration workflow** directly.

Preparations

- Creating a migration project

  Create a migration project on the MgC console. For details, see **Managing Migration Projects**.
- Creating a migration cluster

  Create a dedicated migration cluster for this migration. A cluster consists of a master node and several list and migration nodes. For details about how to create a cluster, see **Creating a Migration Cluster**.
- Creating target file systems

  Create target file systems for receiving migrated source data. For details, see **Creating a General Purpose File System** or **Creating an SFS Turbo File System**.

Notes

- Only one migration workflow can be created for a migration plan.

- A maximum of 100 file systems can be added to a migration plan.
- Target file systems included in a migration plan must be in the same region.

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane, choose **Design** > **Migration Plans**.

Step 3 Click **Create Migration Plan** in the upper right corner of the page.

Step 4 In the **Batch File Storage Migration** card, click **Configure Migration Plan**.

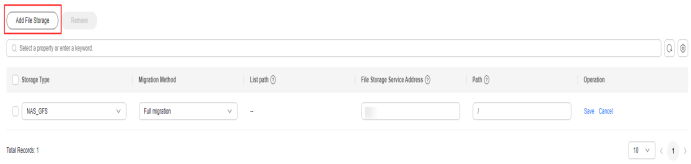
Step 5 In the **Basic Settings** area, set parameters listed in **Table 8-6**.

Table 8-6 Basic settings

Parameter	Configuration
Migration Plan Name	User-defined
Description (Optional)	User-defined
Target Region	Select the region where the target file storages are located.

Step 6 Click **Add**. A new row will be added below, allowing you to configure a source file system. Each click of this button will add another row, enabling you to include multiple source file systems in this plan.

Figure 8-1 Adding a source file system



Step 7 Configure a source file system based on **Table 8-7**.

Table 8-7 Parameters for configuring a source file system

Parameter	Description	Remarks
Storage Type	Supported source storage types include NAS_GFS, NAS_NFS_V3_PROTOCOL, and NAS_NFS_V3_MOUNT.	-

Parameter	Description	Remarks
Migration Method	Select a migration method. • Full migration: This method is used to migrate all data in a specific path in the source file system. • List migration: This method is used to migrate files and directories recorded in lists.	-
List Path	If the list migration method is used, enter the path to the list files. If the full migration method is used, leave it blank.	A list file must meet the following requirements: • A list file cannot exceed 30 MB. • A list file must be a .txt file. • A list file must be in UTF-8 without BOM. • Each line in a list file can contain only one file or directory name, and the name must be URL encoded. • Each line in a list file cannot exceed 16,384 (16 × 1,024) bytes, or the migration will fail. • A list file can contain a maximum of 10,000 lines.

Parameter	Description	Remarks
File System Address	Enter the mount address of the source file system. • If the storage type is **NAS_GFS**, the file system address can include a host name, domain name, (optional) subdomain, port, and path. • If the storage type is **NAS_NFS_V3_PROTOCOL** or **NAS_NFS_V3_MOUNT**, the file system address must match any of the following formats: – *<IP-address>:/* – *<IP-address>:/<path>* – *<domain-name>:/* – *<domain-name>:/<path>*	For a full migration, the combination of a source file system address and the path to be migrated must be unique. For a list migration, the combination of a source file system address and the list path must be unique. In both cases, the prefix matching rule is applied. For example, if you have added the file system address **192.168.0.1:/a** with a migration path **/abc/a**, an error will be reported if you add the file system address **192.168.0.1:/a** with a migration path **/abc**. To overwrite an existing path, you need to delete the existing address and path combination first.
Path	Enter the directory where files and directories to be migrated are located. The format is **/** *Folder name*.	If the list migration method is selected, enter **/** in the **Path** text box.

Step 8 Each time after you configure a source file system, click **Save** in the **Operation** column. After all source file systems to be migrated are configured, click **Next**.

Step 9 In the list, click **Modify** in the **Operation** column to configure a target files system for each source file system.

Table 8-8 Parameters for configuring a target file system

Parameter	Description	Remarks
Source Path	This path is automatically generated by combining the source file system address with the path to be migrated.	-
Storage Type	Only **NAS_NFS_V3_MOUNT** is supported.	

Parameter	Description	Remarks
File System Address	• If a General Purpose file system is used as the target, enter the mount address of the file system. To obtain the mount address, go to the file system list and click the icon next to the address in the **Mount Point** column. • If an SFS Turbo file system is used as the target, enter the shared path of the file system. To obtain the shared path, go to the SFS Turbo file system list, click next to the path in **Shared Path** column.	Combinations of target file system addresses and paths can be duplicated.
Path	Enter the directory for storing files and directories migrated. The format is */Folder name*.	-

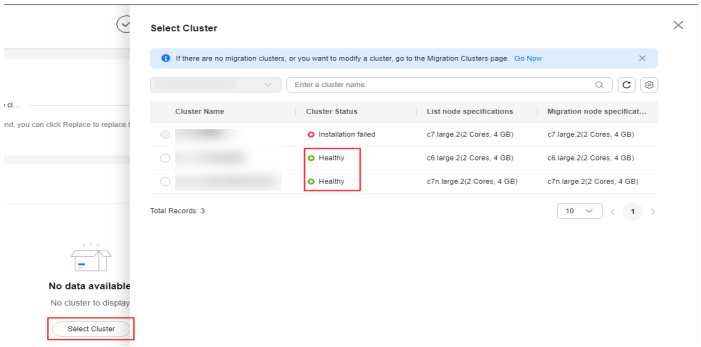
Step 10 Each time after you configure a target file system, click **Save** in the **Operation** column. After you configure a target file system for each source file system, click **Next**. You can assess how large of a migration cluster is required for the migration and then create a migration cluster with the recommended specifications. Alternatively, you can skip this step and use an existing migration cluster. For details about how to create a cluster, see **Creating a Migration Cluster**.

Step 11 Click **Next** to select an existing migration cluster. All resources in this plan will be migrated using this cluster.

Step 12 Click **Select Cluster**. The **Select Cluster** page is displayed on the right.

Step 13 In the cluster list, select the cluster to be used and click **OK**. The cluster must be in a healthy state. If you want to modify an existing cluster to meet your requirements, see **Managing a Migration Cluster**.

Figure 8-2 Selecting a migration cluster



Step 14 Click **OK**. After migration plan is created, you can see it in the list.

- If you need to modify the plan settings, click **Design** in the **Operation** column.
- After **Completed** appears in the **Progress** column, click **Create Workflow** in the **Operation** column to **create a migration workflow** to migrate all source file systems in the plan in a batch.

----End

9 Migration Clusters

9.1 Creating a Migration Cluster

You can create dedicated migration clusters to accelerate storage migration. A migration cluster consists of nodes that have the migration plugin installed and are used to execute migration and listing tasks.

Precautions

- A maximum of two migration clusters can be created in each region in an account.
- To ensure migration stability and data security, you are not allowed to log in to nodes in migration clusters. If you indeed need to log in to the nodes, contact technical support.
- If a migration cluster is idle for more than 30 days, its status will change to **Discarded**. The cluster will not be deleted but cannot be restored.
- You can see the details about resources contained in a migration cluster and the security group settings of all cluster nodes in **Migration Cluster Resources and Settings**.

Billing

You are billed at standards rates for migration clusters and additional resources used during migrations.

- Master, migration, and list nodes in migration clusters are all ECSs, and you pay for these ECSs. For details, see **ECS Pay-per-Use Billing** or **ECS Price Calculator**.
- If you migrate over the Internet, you need to pay for NAT gateways used by migration clusters. For details, see **NAT Gateway Billing** or **NAT Gateway Price Calculator**.
- If you choose to enable log collection, you need to pay Log Tank Service (LTS) for resources you consume. For details, see **LTS Billing** or **LTS Price Calculator**.

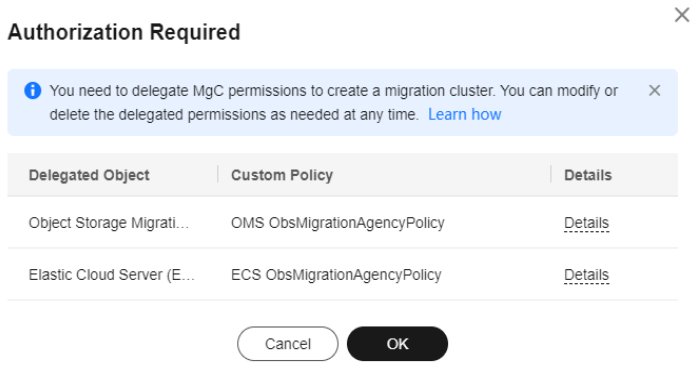
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the left navigation pane, choose **Deploy** > **Migration Clusters**.

Step 3 Click **Create Cluster** in the upper right corner of the page.

If this is your first time to create a migration cluster, you need to delegate MgC the required permissions. For more information about the required permissions, see **Agency Permissions**.



Step 4 Configure the parameters listed in **Table 9-1**.

Table 9-1 Parameters for creating a cluster

Area	Parameter	Configuration	Constraints
Basic Settings	Cluster Name	Enter a name.	The cluster name must be unique in the same account.
	Region	Select the region to provision the cluster.	The cluster must be provisioned in the target region you are migrating to.
	Cluster Type	Select what the cluster will be used for.	Currently, only storage migration is supported.
Node Settings	Master Node	It is used to manage migration nodes and list nodes.	A cluster can only have one master node.

Area	Parameter	Configuration	Constraints
	Migration Node	Migration nodes are used for executing migration and verification tasks. The recommended specifications are 8 vCPUs and 16 GB of memory.	• The node specifications cannot be modified after the cluster is created. • The number of nodes must meet the following requirements: – Number of migration nodes + Number of list nodes + 1 ≤ 100 – Number of migration nodes + Number of list nodes + 1 ≤ Number of unused IP addresses in the subnet
	List Node	List nodes are used for listing tasks. The recommended specifications are 8 vCPUs and 16 GB of memory.	
Network Settings	VPC	Select a VPC from the drop-down list.	-
	Subnet	Make sure that there are enough unused IP addresses for the migration and list nodes in this cluster.	Number of unused IP addresses in the subnet ≥ Number of migration nodes + Number of list nodes + 1

Area	Parameter	Configuration	Constraints
	Network Type	• **Internet**: You need to select a public NAT gateway. If there is no gateway available, choose **Buy Gateway** from the drop-down list and select the gateway specifications and EIPs you want to associate with the gateway. A maximum of 20 EIPs can be selected at a time. • **Intranet**: This option is suitable for data migration within a region. • **Private line**: Source data is directly accessed through the private line. For details about Direct Connect, see **Direct Connect**.	-
Advanced Settings	DNS Configuration (Optional)	Enter the IP address of the DNS server to update the value of **nameserver** in the **/etc/resolv.conf** file. Use commas (,) to separate multiple DNS server addresses, for example, 192.0.2.1,192.0.2.2.	A maximum of three DNS IP addresses can be specified.
	Domain Mapping (Optional)	Add mappings between domain names and IP addresses to update the **/etc/hosts** file.	A maximum of 500 mappings can be added.

Area	Parameter	Configuration	Constraints
-	Traffic Limiting	Allocate the maximum bandwidth to be used by the workflow during a specified period. • If you do not select this option, migration traffic is not limited. • If you select this option, limit the migration traffic by setting the start time, end time, and bandwidth limit. **NOTICE** For example, if you set **Start Time** to **08:00**, **End Time** to **12:00**, and **Maximum Bandwidth** to **20 MB/s**, the maximum migration speed is limited to 20 MB/s when the migration task is running in the period from 08:00 to 12:00. The migration speed is not limited beyond this period.	• A maximum of five traffic limiting rules can be added. • The time is the local standard time of the region you are migrating to.
	Log Collection	• If this option is enabled, logs generated during the migration are collected for possible troubleshooting later. • If this option is disabled, logs generated during storage migrations are not collected.	-

Step 5 Click **Confirm**. Then you can view the cluster in the list. For details about cluster statuses, see **Cluster Statuses**. If the cluster status is **Creation failed**, move the cursor to the status to view the failure cause. After the fault is rectified, choose **More** > **Retry** to try to create the cluster again.

Step 6 After the migration cluster is created, perform the following operations:

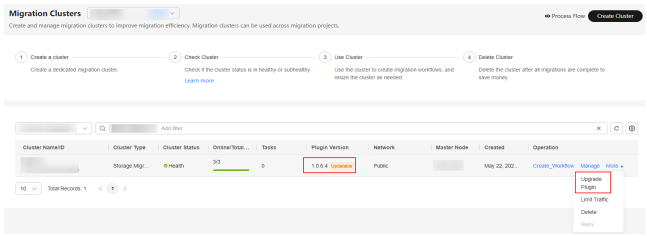
- If the cluster status is **Healthy** or **Subhealthy**, click **Create Workflow** in the **Operation** column to .

- Click **Manage** in the **Operation** column to add or delete nodes and traffic limiting rules for the cluster. For details, see **Managing a Migration Cluster**.

----End

Upgrading Plug-ins

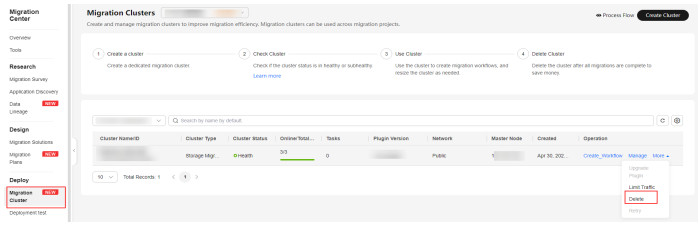
You can upgrade the migration plugin to the latest version for a cluster by choosing **More** > **Upgrade Plugin** in the **Operation** column.



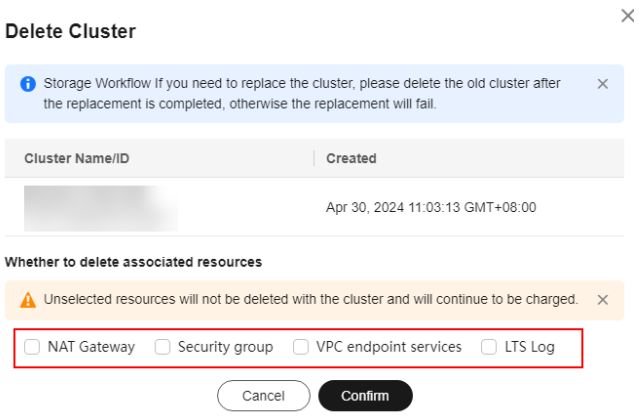
Deleting a Migration Cluster

If a migration cluster is no longer needed after your data migration is complete, you can delete the cluster and the associated resources.

Step 1 In the cluster list, locate the cluster to be deleted and choose **More** > **Delete** in the **Operation** column.



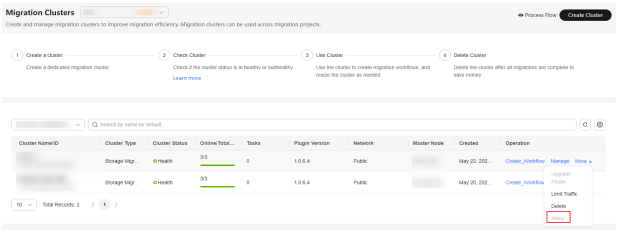
Step 2 Select whether to delete associated resources and click **OK**.



----End

Retrying to Create a Migration Cluster

If a migration cluster fails to be created, after you rectify the fault, you can try to create the cluster again by choosing **More** > **Retry** in the **Operation** column.



9.2 Managing a Migration Cluster

For a migration cluster, you can configure traffic limiting rules and add or remove nodes as needed.

Prerequisites

You have **created a cluster**.

Adding a Traffic Limiting Rule

Step 1 Sign in to the **MgC** console. In the left navigation pane, choose **Deploy** > **Migration Clusters**.

Step 2 In the cluster list, locate a migration cluster and click **Manage** in the **Operation** column.

Step 3 In the **Traffic Limiting** area, click **Add** to add a rule.

Step 4 Configure the start time, end time, and bandwidth limit.

For example, if you set **Start Time** to **08:00**, **End Time** to **12:00**, and **Maximum Bandwidth** to **20 MB/s**, the maximum migration speed is limited to 20 MB/s when the migration task is running in the period from 08:00 to 12:00. The migration speed is not limited beyond this period.

NOTICE

- The time is the local standard time of the region you are migrating to.
- A maximum of five rules can be added.

Step 5 Click **Confirm**.

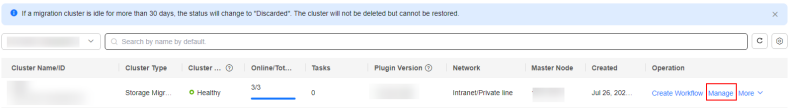
----End

Adding Nodes

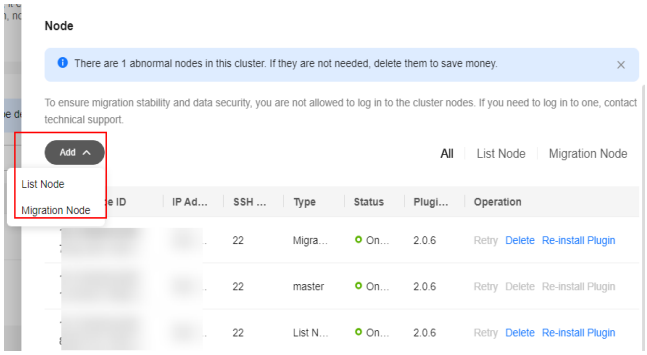
The procedure for adding a migration node is the same as that for adding a list node.

Step 1 Sign in to the **MgC** console. In the left navigation pane, choose **Deploy** > **Migration Clusters**.

Step 2 In the cluster list, click **Manage** in the **Operation** column.

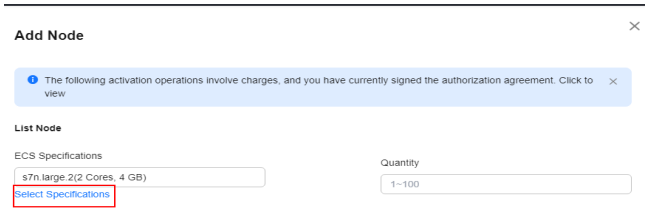


Step 3 In the node information area, choose **Add** > **Migration Node**.



Step 4 Select the node specifications and enter the number of nodes to be added. The number of nodes in the cluster must meet the following requirements:

- Number of migration nodes + Number of list nodes + 1 ≤ 100
- Number of migration nodes + Number of list nodes + 1 ≤ Number of unused IP addresses in the subnet



Step 5 Click **Confirm**. After the nodes are added, MgC starts to install the migration plug-in on these nodes, and you can view the added nodes in the node list.

----End

9.3 Billing

You are billed at standards rates for migration clusters and additional resources used during migrations.

- Master, migration, and list nodes in migration clusters are all ECSs, and you pay for these ECSs. For details, see **ECS Pay-per-Use Billing** or **ECS Price Calculator**.
- If you migrate over the Internet, you need to pay for NAT gateways used by migration clusters. For details, see **NAT Gateway Billing** or **NAT Gateway Price Calculator**.
- If you choose to enable log collection, you need to pay Log Tank Service (LTS) for resources you consume. For details, see **LTS Billing Description** or **LTS Price Calculator**.

9.4 Cluster Statuses

For details about cluster statuses, see **Table 9-2**.

Table 9-2 Cluster statuses

Status	Description
Creating	Cluster resources are being created.
Creation failed	Cluster resources fail to be created.
Connecting	The master node is waiting for going online, and other nodes are waiting for creation.
Healthy	All nodes are online.
Subhealthy	At least one migration node and one list node are online.
Unavailable	All migration nodes or all list nodes are offline.
Offline	The master node is offline. The cause may be that the network is interrupted or the ECS is deleted. Check the VPCEP service and ECS resources.
Upgrading	Plugins are being upgraded in the cluster.
Upgrade failed	Plugins failed to be upgraded in the cluster.
Discarding	The cluster is being discarded because no tasks have run in the cluster within 30 days.
Discard failed	The cluster failed to be discarded.
Deleting	The cluster is being deleted.
Deletion failed	The cluster failed to be deleted.
Discarded	The cluster is discarded since its VPCEP resources are deleted.
Pending creation	Cluster resources are waiting for creation.
Installing	Plugins are being installed on the master node.
Installation failed	Plugins failed to be installed on the master node.
Pending installation	The master node is waiting for plugin installation.
Pending upgrade	The cluster is waiting for plugin upgrade.
Pending deletion	The cluster is waiting for deletion.

9.5 Migration Cluster Resources and Settings

This section describes the resources and their settings of a migration cluster.

Migration Cluster Resources

Resource	Configurable During Cluster Creation	Mandatory	Quantity	Naming Rule	Description
ECSs	Yes	Yes	• Master node: 1 • Migration nodes: created based on the settings specified during cluster creation • List nodes: created based on the settings specified during cluster creation	oms_cluster_ecs_*Node ID*	These ECSs have the migration plug-in installed and are used to schedule and perform migration tasks.

Resource	Configurable During Cluster Creation	Mandatory	Quantity	Naming Rule	Description
NAT gateway	Yes	No	1	oms_cluster_nat_*Cluster ID*	This NAT gateway enables list and migration nodes to communicate with migration sources over the public network, so that source data can be migrated.

Resource	Configurable During Cluster Creation	Mandatory	Quantity	Naming Rule	Description
LTS	Yes	No	1	Host group name: oms_lts_host_group_*Cluster ID*	LTS is used to save and search for logs of the migration cluster. If **Log Collection** is not enabled during cluster creation, these LTS resources will not be created.
				Host group name: oms_lts_host_group_label_*Cluster ID*	
				Log group name: oms_lts_log_group_*Cluster ID*	
				Log stream name: oms_lts_stream_group_*Cluster ID*	
				Log access name: oms_lts_access_config_*Cluster ID*	

Resource	Configurable During Cluster Creation	Mandatory	Quantity	Naming Rule	Description
VPC endpoint service	No	Yes	1	*Region name*.oms-cluster.*Service ID*	This VPC endpoint service allows OMS to communicate with the master node in the migration cluster to manage migration tasks. For details, see **VPC Endpoint Service Configuration**.

Resource	Configurable During Cluster Creation	Mandatory	Quantity	Naming Rule	Description
Security groups	No	Yes	2	oms_cluster_sg_master_*Cluster ID*	This is the security group of the master node in the migration cluster. It enables the communication between OMS and the migration cluster. For details about the security group settings, see **Security Group Settings for the Master Node**.

Resource	Configurable During Cluster Creation	Mandatory	Quantity	Naming Rule	Description
				oms_cluster_sg_agent_*Cluster ID*	This is the security group of the list and migration nodes in the migration cluster. It enables communication between nodes in the migration cluster. For details about the security group settings, see **Security Group Settings for List and Migration Nodes**.

VPC Endpoint Service Configuration

Item	Configuration	Description
VPC	The VPC where the migration cluster works is selected.	-
Service Type	Interface	-
Connection Approval	This option is enabled.	VPC endpoints need to be approved before they can connect to this endpoint service.
Port Mapping	Ports 22 and 9002 are mapped.	They are used for plugin installation, task delivery, and task management.
Backend Resource Type	ECS	The ECS used as the master node.
Connection Management	The VPC endpoint of OMS is allowed to access this VPC endpoint service.	Allows OMS to communicate with the master node in the migration cluster.
Permission Management	OMS is added to the whitelist.	Allows access from the VPC endpoint of OMS.

Security Group Settings for the Master Node

Direction	Priority	Action	Type	Protocol & Port	Source or Destination	Description
Inbound	1	Allow	IPv4	All	Security group of the mater node	Allows VMs in the security group to communicate with each other.
Inbound	1	Allow	IPv6	All	Security group of the mater node	Allows VMs in the security group to communicate with each other.
Inbound	1	Allow	IPv4	TCP: 22	198.19.128.0/17	Allows OMS to communicate with the master node to install the migration plugin.

Direction	Priority	Action	Type	Protocol & Port	Source or Destination	Description
Inbound	1	Allow	IPv4	TCP: 9002	198.19.128.0/17	Allows OMS to communicate with the master node to deliver and manage tasks.
Inbound	1	Allow	IPv6	All	The security group of the list and migration nodes	Allows the list and migration nodes to communicate with the master node.
Inbound	1	Allow	IPv4	All	The security group of the list and migration nodes	Allows the list and migration nodes to communicate with the master node.
Outbound	1	Allow	IPv6	All	::/128	Allows outbound traffic to communicate with the migration source.
Outbound	1	Allow	IPv4	All	0.0.0.0/0	Allows outbound traffic to communicate with the migration source.
Outbound	1	Allow	IPv6	All	The security group of the list and migration nodes	Allows the list and migration nodes to communicate with the master node.
Outbound	1	Allow	IPv4	All	The security group of the list and migration nodes	Allows the list and migration nodes to communicate with the master node.

Security Group Settings for List and Migration Nodes

Direction	Priority	Action	Type	Protocol & Port	Source or Destination	Description
Inbound	1	Allow	IPv6	All	Security group of the mater node	Allows the list and migration nodes to communicate with the master node.
Inbound	1	Allow	IPv4	All	Security group of the mater node	Allows the list and migration nodes to communicate with the master node.
Inbound	1	Allow	IPv4	All	The security group of the list and migration nodes	Allows VMs in the security group to communicate with each other.
Inbound	1	Allow	IPv6	All	The security group of the list and migration nodes	Allows VMs in the security group to communicate with each other.
Outbound	1	Allow	IPv4	All	0.0.0.0/0	Allows outbound traffic to communicate with the migration source.
Outbound	1	Allow	IPv6	All	::/128	Allows outbound traffic to communicate with the migration source.
Outbound	1	Allow	IPv6	All	Security group of the mater node	Allows the list and migration nodes to communicate with the master node.
Outbound	1	Allow	IPv4	All	Security group of the mater node	Allows the list and migration nodes to communicate with the master node.

10 Migration Workflows

10.1 Workflow Quotas

To ensure smooth migrations, MgC has the following limits on migration workflows.

- A maximum of 50 migration workflows can be created per day in a project.
- A maximum of 1,000 servers can be concurrently. For any servers beyond this number, the workflows will pause at the first step and put these servers in a pending state until other servers complete their migration. The workflows will then automatically start on these servers in the order the workflows were created.
- A maximum of 100 resources can be included in a workflow.

10.2 Creating a Server Migration Workflow

This section describes how to create a server migration workflow using the standard template.

CAUTION

- A workflow can contain a maximum of 100 servers.
- You can migrate a maximum of 1,000 servers concurrently. For any servers beyond this number, the workflows will pause at the first step and put these servers in a pending state until other servers complete their migration. The workflows will then automatically start on these servers in the order the workflows were created.
- If this is your first time to create a server migration workflow, you need to delegate MgC the required permissions. For more information about the required permissions, see **Agency Permissions**.

Precautions

Before creating a server migration workflow, read and understand the following precautions.

Item	Precaution
Source download bandwidth	Used to download the migration Agent to source servers. • If each source server uses a dedicated bandwidth, the bandwidth must be at least 30 Mbit/s. • If source servers share a bandwidth, the average bandwidth must be at least 50 Mbit/s.
Migration bandwidth	• Used to migrate data. It affects the migration speed and duration. • For details about how to estimate the migration duration, see **How Long Does a Migration Take?**
CPU and memory	• At least 520 MB of available memory • At least 0.3 CPUs available in Linux and at least 1 CPU available in Windows
OS compatibility	For details about what OSs are supported, see **Supported OSs**.
Server migration statements	For details about the important statements you need to understand before the migration, see **What Are the Important Statements of SMS?**
Notes and constraints	For details about the notes and constraints for server migration, see **Notes and Constraints**.
Billing	For details about the fees that may be incurred during the migration, see **Billing**.
Permissions configuration	For details about the permissions the target account must have, see **Permissions Management**.
Migration network and ports	For details about the requirements for the migration network and ports, see **How Do I Set Up a Secure Migration Network for Using SMS?**

Prerequisites

- You have **discovered source servers**.
- Servers to be migrated have been added to an application, and the application servers have been assessed. For details, see **Grouping Resources as Applications** and **Getting Target Recommendations**. After the assessment is complete, you can purchase recommended target servers in batches. For details, see **Purchasing Resources**. Source servers with target servers associated do not need to be assessed. For details, see **Associating Source Servers with Target Servers**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Workflows**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Create Workflow** in the upper right corner of the page.

Step 4 In the **Server Migration** card, click **Preview Steps** to view the steps predefined in the template and the detailed description of each step. If the type of a step is **Automated**, the step is automatically performed by MgC. If the type of a step is **Manual**, you need to perform the step manually. Click **Next: Configure Workflow** in the lower right corner.

Step 5 Configure the workflow parameters based on **Table 10-1**.

Table 10-1 Parameters for configuring a server migration workflow

Area	Parameter	Description
Workflow Details	Name	User-defined
	Description	
	Select Source Resources	• **By application**: Servers in the selected application will be migrated together. • **By migration plan**: Servers in the selected migration plan will be migrated together.
Application	Application	If **Select Source Resources** is set to **By application**, select the application that contains the servers to be migrated.
Migration Plan	Migration Plan	If **Select Source Resources** is set to **By migration plan**, select the plan that contains the servers to be migrated. For details about how to create a migration plan, see **Creating a Server Migration Plan**.
Migration Network	Network Type	If you select **Public**, ensure that all target servers have EIPs bound. These EIPs will be used for the migration.

Area	Parameter	Description
		If you select **Private**, configure Direct Connect connections, VPN connections, VPC peering connections, or subnets in the target VPC in advance to ensure that the source environment can access the target environment. ● If the source environment cannot access the Internet, you need to enter the private IP address of the source proxy server and the port used by the proxy software. ● If the source proxy server cannot access the Internet, put the SMS-Agent installation package at a location where the source servers can access directly or over a proxy. You can download the SMS-Agent installation package from the SMS console.
Target Environment	Region	The target region. It defaults to the one you selected when you assessed the application.
	Project	A project in the target region.
	VPC	● If the source IP address is 192.168.*X.X*, you are advised to create a VPC and a subnet that both belong to network range 192.168.0.0/16. ● If the source IP address is 172.16.*X.X*, you are advised to create a VPC and a subnet that both belong to network range 172.16.0.0/12. ● If the source IP address is 10.*X.X*, you are advised to create a VPC and a subnet that both belong to network range 10.0.0.0/8.
	Subnet	The subnet has to be in the same network segment as the VPC.
	Security Group	● If there are Windows source servers, the security group must allow access on ports 8899, 8900, and 22. ● If there are Linux source servers, the security group must allow access on port 22. **CAUTION** – For security purposes, you are advised to only allow traffic from the source servers on these ports. – The firewall of the target servers must allow traffic to these ports.

Area	Parameter	Description
Advanced Settings	Start Target After Migration	• If you select **No**, the target servers will be stopped after the migration is complete. • If you select **Yes**, the target servers will be started after the migration is complete.
	Set Bandwidth Limit	• If you select **No**, the migration traffic is not limited. • If you select **Yes**, you can limit the bandwidth that can be used for migration based on the source bandwidth and service requirements. **CAUTION** Consider the migration scale to set an appropriate bandwidth limit. If the bandwidth allocated to the workflow is too small, migration tasks in the workflow may preempt the limited bandwidth resource, and some servers may fail to be migrated.
	Install rsync on Source	• If you select **No**, the rsync component will not be installed on the source servers. • If you select **Yes**, the rsync component will be automatically installed on the source servers. **CAUTION** Linux migrations depend on rsync. If rsync is not installed on a source server, the server will fail to be migrated.
	Retain IP Address	Enable this option to retain the private IP addresses of the source servers on the target servers. Enabling it may cause risks. You need to evaluate and take responsibility for any potential risks.
	Enable Quick Mode	Enable this option if incremental synchronization is not required. This option is disabled by default. If it is enabled, incremental synchronization is skipped and then subsequent steps are performed after full replication is complete in the workflow. Set this option based on your requirements.
	Enterprise Project	Select the enterprise project you want to migrate to. The enterprise project **default** is selected by default.

Step 6 Click **Next: Confirm**.

Step 7 Confirm the workflow settings, and click **Confirm**. The **Run Workflow** dialog box is displayed, which indicates that the workflow has been created.

- If you want to start the migration immediately, click **Confirm** to run the workflow.
- If you want to **add a stage or step** to the workflow, click **Cancel**. The workflow enters a **Waiting** state, and the migration is not started. To start the migration, click **Run** in the **Operation** column.

Step 8 On the migration workflow details page, view the workflow settings and the migration progress. After the step for starting the migration Agent is completed, a migration task is automatically created on the SMS console. For details about the server information mapping between MgC and SMS, see **What Are the Information Mappings Between MgC and SMS?**

- Move the cursor to the migration progress bar. In the box that is displayed, view more migration details.
- When the migration progress bar reaches a step that requires manual confirmation, move the cursor to the progress bar and click **Confirm** next to the step status in the displayed window, so that the subsequent migration steps can be executed.
- When the workflow reaches the **ResizeDiskPartition** step, the system identifies whether disk capacity reduction has been performed on the target server.
  - If yes, go to **SMS console** and resize disks and partitions for the target server. For details, see the **Partition Resizing** parameter in **Configuring a Target Server**. After the adjustment is complete, go back to the MgC console and click **Confirm** next to the step status so that the workflow can continue.
  - If no, skip this step.
- The **StartSynchronization** step is repeated before you verify your services on the target server.
- When the progress bar reaches **Cutover**, the migration is complete. You need check whether your service systems are running properly on the target server. If they are, manually switch services to the target server. After the switchover is complete, click **Confirm** in the workflow. The system automatically performs the following steps **SourceClear** and **MigrationTaskClear**.

----End

10.3 Creating a Cross-AZ Migration Workflow

This section describes how to create a cross-AZ migration workflow using the standard template.

Prerequisites

- You have discovered servers by referring to **Simple Project**.
- Servers to be migrated have been added to an application, and the application servers have been assessed or associated with existing target servers. For details, see **Grouping Resources as Applications** and **Getting Target Recommendations**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Workflows**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Create Workflow** in the upper right corner of the page.

Step 4 In the **Cross-AZ Migration** card, click **Preview Steps** to view the steps predefined in the template and the detailed description of each step. If the type of a step is **Automated**, the step is automatically performed by MgC. If the type of a step is **Manual**, you need to perform the step manually. Click **Configure Workflow** in the lower right corner.

Step 5 Configure the workflow parameters listed in **Table 10-2**.

Table 10-2 Parameters required for creating a cross-AZ migration workflow

Region	Parameter	Description
Workflow Details	Name	Enter a workflow name.
	Description	Describe the workflow.
Application	Application	Select the application which contains the servers to be migrated.
Migration Settings	Target Region	Select the region where the source AZ is located. The region configured in the application is populated by default.
	Target AZ	Select the AZ you want to migrate to. The configuration must be the same as that of the created application.
	Target Network	Only **Retain original** is available.
	Target Server	Create now. MgC creates backups and images for source servers, and uses the images to create target servers immediately after the workflow runs.
	Stop Target Server	• If you select **Yes**, target servers will be stopped after being created. • If you select **No**, target servers will be started after being created.
	Stop Source Server	• If you select **Yes**, source servers will be stopped before incremental backups are created for them. This ensures data consistency as high as possible. • If you select **No**, source servers remain running when incremental backups are created for them.

Region	Parameter	Description
	Create System Disk Image	• If you select **Yes**, a system disk image will be created for each of the source servers. The images can be used to reinstall the OS for the paired target servers. • If you select **No**, the system will not create system disk images for the source servers.
Advanced Settings	Delete Intermediate Resources	If this function is enabled, intermediate resources generated during the migration, such as backups, snapshots, and images, will be deleted after the service cutover is complete.
	Retain Primary NIC IP Addresses	If this function is enabled, the private and public IP addresses of the primary NIC on source servers will be retained on target servers, and random private IP addresses will be allocated to source servers. If a rollback is needed, it has to be performed manually.

Step 6 Click **Next: Confirm**.

Step 7 Confirm the workflow settings, and click **Confirm**. The **Run Workflow** dialog box is displayed, which indicates that the workflow has been created.

- If you want to start the migration immediately, click **Confirm** to run the workflow.
- If you want to **add a stage or step** to the workflow, click **Cancel**. The workflow enters a **Waiting** state, and the migration is not started. To start the migration, click **Run** in the **Operation** column.

Step 8 On the migration workflow details page, view the workflow settings and the migration progress.

- Move the cursor to the migration progress bar. In the box that is displayed, view more migration details.
- When the migration progress bar reaches a step that requires manual confirmation, move the cursor to the progress bar and click **Confirm** next to the step status in the displayed window, so that the subsequent migration steps can be executed.

----End

10.4 Creating a Storage Migration Workflow

This section describes how to create a storage migration workflow from the predefined template.

The following regions are supported:

- LA-Santiago
- LA-Sao Paulo

- TR-Istanbul
- AP-Bangkok
- AP-Singapore
- AP-Jakarta
- ME-Riyadh
- CN North-Beijing4
- CN East-Shanghai1

Notes and Constraints

- For more information, see **What Are the Restrictions on Using MgC for Storage Migration?**
- Data in the Deep Cold Archive storage of Alibaba Cloud OSS cannot be restored using the restoration function of MgC. You need to manually restore the data before migration.

Warnings

When you create a workflow, there are three options for **Overwrite Existing**: **Never**, **Always**, and **If older or different size**. If you choose **Never**, restarting the migration task after interruptions or pauses may cause incomplete data migration, but the task may still be displayed as successful, which affects data integrity. Use the **Never** option with caution.

Prerequisites

- You have created an **OBS bucket** or **SFS file system** on Huawei Cloud.
- You have **created a migration cluster**.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Workflows**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Create Workflow** in the upper right corner of the page.

Step 4 In the **Storage Migration** card, click **Preview Steps** to view the stages and steps predefined in the template and the detailed description of each stage and step. Steps of the **Automated** type will be automatically performed by MgC. Click **Configure Workflow** in the lower right corner.

Step 5 Set workflow basics based on **Table 10-3**.

Table 10-3 Basic parameters

Parameter	Description
Name	Enter a workflow name.
Region	Select a region you are migrating to.

Parameter	Description
Description	Enter a description.
Cluster	Select a migration cluster. The cluster consists of a master node and several migration and list nodes. If no cluster is available, **create a cluster**. **NOTICE** A master node is created by the system in a migration cluster by default. You do not need to configure it.

Step 6 Configure the migration source and target based on **Table 10-4** and **Table 10-5**.

Table 10-4 Parameters for configuring a migration source

Parameter	Description	Remarks
Location Type	The supported migration sources include: • Huawei Cloud OBS • Alibaba Cloud OSS • Baidu Cloud BOS • Tencent Cloud COS • Kingsoft Cloud KS3 • Qiniu Cloud KODO • UCloud US3 • Amazon S3 • Azure Blob Storage • NAS_GFS • NAS_SMB • NAS_NFS_V3_MOUNT • NAS_NFS_V3_PROTOCOL • HTTP/HTTPS data source	-
AK	Enter the AK of the source cloud account.	These parameters are available when cloud storage is selected for **Location Type**.
SK	Enter the SK of the source cloud account.	
Bucket	Enter the name of the source bucket to be migrated.	
Endpoint	Enter the endpoint of the region where the source bucket is located. For example, if **Location Type** is set to **Alibaba Cloud OSS** and the source bucket is located in **CN East 1 (Hangzhou)**, enter **oss-cn-hangzhou.aliyuncs.com**.	

Parameter	Description	Remarks
Type	Set this parameter based on the source bucket type. You can **view the bucket type in the basic information**.	This parameter is available when **Huawei Cloud OBS** is selected for **Location Type**.
APPID	Enter the APPID of your Tencent Cloud account. **NOTE** You can view the APPID on the account information page of the Tencent Cloud console.	This parameter is available when **Tencent Cloud COS** is selected for **Location Type**.
Authentication Type	The following authentication types are supported: • Connection string • Storage account name and key For details about how to obtain the credentials, see **How Do I Obtain Credentials for Accessing Microsoft Azure?**	This parameter is available when **Azure Blob Storage** is selected for **Location Type**.

Parameter	Description	Remarks
List Path	Enter the path where the lists of files to be migrated are stored. These lists must be stored in the same region as the target bucket. You need to write the URLs of files to be migrated and their new names at the target into the lists. Each line in the list can contain only one URL and one file name. Restrictions on list files are: • The files must be in .txt format, and their metadata **Content-Type** must be **text/plain**. • A single file can contain a maximum of 100,000 rows. • A single file cannot exceed 300 MB. • A maximum of 10,000 list files can be stored in the folder. • The files must be in UTF-8 without BOM. • The length of each line in a file cannot exceed 65,535 characters, or the migration will fail. • The **Content-Encoding** metadata of the files must be left empty, or the migration will fail. • In the files, a tab character (\t) must be used to separate the URL and new file name in each line. The format is [URL][Tab character][New file name]. Only the Chinese and special characters in the names must be URL encoded. • Spaces are not allowed in each line in a file. Spaces may cause migration failures because they may be mistakenly identified as object names.	These parameters are available when **HTTP/HTTPS data source** is selected for **Location Type**.
File System Address	Enter the mount address of the source file system. The format is *IP address*:/ or *IP address*:/*xxx*, for example, 192.1.1.1:/0001.	These parameters are available when **Location Type** is set to **NAS_SMB**, **NAS_NFS_V3_MOUNT**, or **NAS_NFS_V3_PROTOCOL**.
Path	Enter the directory where files to be migrated are located. The format is /*Folder name*.	

Parameter	Description	Remarks
Username	Enter the username of the account that can access all files in the source file system, for example, **administrator**.	These parameters are available when **Location Type** is set to **NAS_SMB**.
Password	Enter the password of the account.	
Domain on Windows	Enter the domain of the source node. **NOTE** You only need to enter the content before .com. For example, if the domain is test.com, enter test.	

Table 10-5 Parameters for configuring a migration target

Parameter	Description	Remarks
Location Type	Select Huawei Cloud storage based on the source storage type.	-
AK	Enter the AK of the Huawei Cloud account you are migrating to.	These parameters are available when **Location Type** is set to **Huawei Cloud OBS**.
SK	Enter the SK of the Huawei Cloud account you are migrating to.	
Bucket	Select the OBS bucket you are migrating your data to.	
Endpoint	Enter the endpoint of the region where the target OBS bucket is located. For example, if the target bucket is located in the **CN North-Beijing4** region of Huawei Cloud, enter **obs.cn-north-4.myhuaweicloud.com**. **NOTE** If the migration source is an OBS bucket, you can view the endpoint in the OBS bucket overview.	
Specify Prefix	Specify a prefix to rename or relocate objects migrated to the target bucket. For example, if you specify the prefix **/D**, source file **/A/B/C.txt** will be relocated to **/D/A/B/C.txt** after being migrated to the target bucket. For details, see: **Adding a Name Prefix or Path Prefix to Migrated Objects**	

Parameter	Description	Remarks
File System Address	Enter the mount address of the target file system. To obtain the mount address, go to the SFS file system list and click the  icon next to the address in the **Mount Point** column.	These parameters are available when **Location Type** is set to **NAS_SMB** or **NAS_NFS_V3_MOUNT**.
Path	Enter the directory for storing files migrated. The format is */Folder name*.	
Username	Enter the username of the account that can access all files in the target file system, for example, **administrator**.	These parameters are available when **Location Type** is set to **NAS_SMB**.
Password	Enter the password of the account.	
Domain on Windows	Enter the domain of the target node. **NOTE** You only need to enter the content before .com. For example, if the domain is test.com, enter test.	

Step 7 Configure the migration task based on **Table 10-6**.

Table 10-6 Parameters for configuring a migration task

Parameter	Value	Description
Task Type	Full migration	Migrates all data in the source bucket or specified paths.

Parameter	Value	Description
	List migration	Migrates files recorded in the list files. In **List Path** box, enter the path of the object lists stored in the target bucket. Restrictions on an object list file vary with the target location. ● Target location: Huawei Cloud OBS – An object list file cannot exceed 30 MB. – An object list file must be a .txt file, and the Content-Type metadata must be text/plain. – An object list file must be in UTF-8 without BOM. – Each line in an object list file can contain only one object name, and the object name must be URL encoded. – Each line in an object list file cannot exceed 16 KB, or the migration will fail. – The Content-Encoding metadata of an object list file must be left empty, or the migration will fail. – An object list file can contain a maximum of 10,000 lines. ● Target location: NAS – An object list file cannot exceed 30 MB. – An object list file must be a .txt file. – An object list file must be in UTF-8 without BOM. – Each line in an object list file can contain only one object name, and the object name must be URL encoded. – Each line in an object list file cannot exceed 16 KB, or the migration will fail. – An object list file can contain a maximum of 10,000 lines.
	Prefix migration	This option is only available for migration from cloud storage. If you enter a file name or name prefix in the **Prefix** box, only the objects that exactly match the specified name or prefix are migrated. **NOTICE** ● If the files to be migrated are stored in the root directory of the source bucket, add their name prefixes directly. If the files are stored in a non-root directory, add their directories and name prefixes in the format of *Directory/Prefix*. ● Use commas (,) to separate multiple prefixes.

Parameter	Value	Description
Concurrent Subtasks	-	Specify the maximum number of concurrent subtasks. There cannot be more than 10 concurrent subtasks for each online migration node. For example, if there are 2 online migration nodes, the maximum number of subtasks can be 20 or any number below.
Overwrite Existing	Never	Files existing at the target will never be overwritten. **WARNING** • If you choose **Never** for the initial migration, the attributes of involved parent folders at the source will not be migrated to the target. As a result, the folder attributes may be incomplete at the target. To avoid this issue, use the **Never** option with caution for the initial migration. • If a migration task is paused or interrupted and then restarted or resumed, the **Never** option will cause the system to skip files that were not completely migrated earlier, but the task may still be marked as successful. This affects data integrity. To avoid this issue, use the **Never** option with caution.
	Always	Files existing at the migration target will always be overwritten.
	If older or different size	• Files that already exist at the target will be overwritten if they are older than or have different sizes from the paired files at the source. • Verification will be performed for folders after their contents are migrated. Folders that already exist at the target will be overwritten if they have different last modification times, sizes, or permissions from the paired folders at the source. **NOTE** For empty folders, the overwrite policy is the same as that for files.
Migrate Metadata	-	Decide whether to migrate metadata. • If you select this option, object metadata will be migrated. • If you do not select this option, only the Content-Type and Content-Encoding metadata will be migrated.
Clear Cluster	-	Determine whether to clear the migration cluster after the migration is complete. • If you select this option, a step for clearing the migration cluster will be created in the workflow. You can also choose whether to clear resources used by the cluster, such as NAT gateways, security groups, and VPCEP resources. • If you do not select this option, a step for clearing the migration cluster will not be created in the workflow.

Step 8 (Optional) Configure advanced options based on **Table 10-7**.

Table 10-7 Advanced options

Parameter	Description	Remarks
Target Storage Class	Choose the storage class that your data will be migrated to in the target bucket. For details about storage classes, see **Introduction to Storage Classes**.	-
Enable KMS Encryption	• If you do not select this option, whether migrated data will be encrypted in the target bucket depends on the server-side encryption setting of the bucket. • If you select this option, all migrated objects will be encrypted before they are stored in the target bucket. **NOTE** • Using KMS to encrypt migrated data may slow down the migration speed by about 10%. • This option is only available when KMS is supported in the region you are migrating to.	This parameter is only available for migrations to Huawei Cloud OBS.
Restore Archive Data	• If you do not select this option, the system directly records archive objects in the list of objects that failed to be migrated and continues to migrate other objects in the migration task. • If you select this option, the system automatically restores and migrates archive objects in the migration task. If an archive object fails to be restored, the system skips it and records it in the list of objects that failed to be migrated and continues to migrate other objects in the migration task. **NOTE** The system will restore archive data before migrating it, and you pay the source cloud platform for the API requests and storage space generated accordingly.	-
Filter Source Data	Filter files to be migrated by applying filters. For details about the filters, see **Source Data Filters**.	

Parameter	Description	Remarks
Download Data from CDN	If the default domain name cannot meet your migration requirements, then as long as the source cloud service provider supports custom domain names, you can bind a custom domain name to the source bucket, and enable the CDN service on the source platform to reduce data download expenses. Enter a custom domain name in the **Domain Name** text box and select a transmission protocol. HTTPS is more secure than HTTP and is recommended. If the migration source is the Alibaba Cloud OSS or Tencent Cloud COS, you also need to select an authentication type and enter an authentication key.	
Send SMN Notification	Determine whether to use SMN to get notifications about migration results. ● If you do not select this option, no SMN messages are sent after the migration. ● If you select this option, after the migration, SMN messages are sent to the subscribers of the selected topic. You can select the language and trigger conditions for sending messages.	
Limit Traffic	Allocate the maximum bandwidth to be used by the workflow during a specified period. ● If you do not select this option, migration traffic is not limited. ● If you select this option, limit the migration traffic by setting **Start Time**, **End Time**, and **Bandwidth Limit**. For example, if you set **Start Time** to **08:00**, **End Time** to **12:00**, and **Bandwidth Limit** to **20 MB/s**, the maximum migration speed is limited to 20 MB/s when the migration task is running from 08:00 to 12:00. The migration speed is not limited beyond this period. **NOTE** – The rate limit ranges from 0 MB/s to 1,048,576 MB/s. – A maximum of five rules can be added. – The time is the local standard time of the region you are migrating to.	-

Parameter	Description	Remarks
Schedule Migration	Schedule the migration to automatically run during a period. • If you do not select this option, you need to manually start or stop the migration. • If you select this option, the migration runs during the specified period and stops beyond that period. For example: – If you set **Start Time** to **08:00** and **End Time** to **12:00**, the migration task runs from 08:00 to 12:00 every day. The migration stops beyond that period. – If you set **Start Time** to **12:00** and **End Time** to **08:00**, the migration runs from 12:00 of the current day to 08:00 of the next day. The migration stops beyond that period.	-

Step 9 Click **Next: Confirm**.

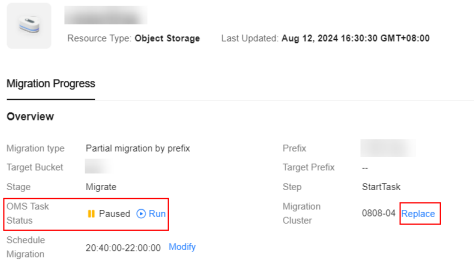
Step 10 Confirm the workflow settings, and click **Confirm**. The **Run Workflow** dialog box is displayed, which indicates that the workflow has been created.

- If you want to start the migration immediately, click **Confirm** to run the workflow.
- If you want to **add a stage or step** to the workflow, click **Cancel**. The workflow enters a **Waiting** state, and the migration is not started. To start the migration, click **Run** in the **Operation** column.

Step 11 On the migration workflow details page, view the workflow settings and the migration progress. You can also perform the following operations:

- Move the cursor to the migration progress bar of a resource. In the displayed window, view the migration details about the resource.
- When a migration reaches a step that requires manual confirmation, place the cursor on the progress bar and click **Confirm** next to the step status in the displayed window. The migration can continue only after you confirm.
- In the **Basic Information** area, click **Manage** next to the cluster name. The cluster details page is displayed on the right. On the displayed page, you can:
  - Add, edit, or delete traffic limiting rules to control cluster traffic based on your requirements.
  - Add or delete migration nodes or list nodes, or upgrade plug-ins for existing nodes as required.

Step 12 (Optional) Click the migration progress bar or click **Migration Progress** in the window displayed when you move the cursor to the progress bar. The migration details page is displayed on the right. You can view the task overview and progress details. You can also perform the following operations:

Operation	Description
Change the migration cluster.	You can change the migration cluster only when the migration task (workflow) is **Paused**.  1. In the **Overview** area, click **Replace** next to the migration cluster name. 2. In the displayed dialog box, select a new cluster from the drop-down list and click **Confirm**. After the cluster is changed, the workflow starts to run automatically.
Modify the migration schedule.	1. In the **Overview** area, click **Modify** next to **Schedule Migration**. 2. Set **Start Time** and **End Time**, and click **Confirm**.
Modify the number of concurrent subtasks.	1. In the **Progress** area, click **Modify** under **Expected Concurrent Subtasks** to change the expected number of concurrent subtasks. There cannot be more than 10 concurrent subtasks for each online migration node. For example, if there are 2 online migration nodes, the maximum number of subtasks can be 20 or any number below. 2. Click **Confirm**.
Add traffic limiting rules.	1. In the **Migration Speed** area, click **Add** to add a rule to limit the bandwidth the migration can use in a specified period. **NOTICE** – The bandwidth limit ranges from 1 MB to 1,024 GB. – Time periods in different rules cannot overlap. For example, if there is a rule added for the period from 8:00 to 12:00, you cannot configure rules for any overlapped periods, such as from 7:00 to 13:00, 7:00 to 8:00, and 9:00 to 12:00. – The start time of a rule cannot be later than the end time. For example, the time period from 23:00 to 01:00 is not allowed. 2. Click **Save**.

Operation	Description
Obtain the list of files that fail to be migrated, skipped or migrated.	In the **File Statistics** area, view the path of the list of files that failed to be migrated, skipped, or migrated. Click a file path, and it will take you to the OBS bucket where the list is stored. You can download the list from the bucket.
View traffic statistics.	In the **Traffic Statistics** area, view the migration traffic in the last hour, last 6 hours, last 24 hours, or the entire migration period.

----End

Source Data Filters

The following table describes the rules and restrictions for setting source data filters.

Table 10-8 Filter options

Option	Description	Patten Rule	Constraint
Exclude Patterns	If a file matches any excluded pattern, the file will not be migrated or compared for consistency. Both exact match and fuzzy match are supported.	• Exact match You need to specify absolute paths and use slashes (\) to escape special characters in the paths.	• Except for {}, consecutive characters specified in pattern rules are not allowed, for example, ***, *?, **?, ?*, ?**, *{*, *}*, *}?, ?{*, {*}, {,}, {*,,, *}, and ,*, • Only asterisks (*) can be used as wildcard characters in {}. • {1} cannot be nested in {0}. • Excluded patterns take precedence over included patterns. • Semicolons (;) are used to separate patterns outside {}.
Include Patterns	• If no included patterns are specified, all files in the source will be migrated. • If included patterns are specified, only the files whose absolute paths match the specified patterns will be migrated or compared for consistency.		

Option	Description	Patten Rule	Constraint
		CAUTION Precautions for configuring exclude and include patterns: • If the file system address ends with :/, when you configure the paths to be excluded or included, enter their absolute paths relative to the mount point. For example, if the file system address is **192.1.1.1:/** and the mount point is **/mnt/turbo**, enter absolute paths relative to **/mnt/turbo**. For example: [root@oms-cluster-ecs filter_test]# pwd /mnt/sts_turbo/mgc/filter_test [root@oms-cluster-ecs-filter_test]# ll drwxr-xr-x 2 root root 0 Aug 16 15:27 test2 -rw-r--r-- 1 root root 5 Aug 16 15:27 test2.log To exclude the **test2.log** file from the migration, you can enter its absolute path **/mgc/filter_test/test2.log** in the **Exclude**	

Option	Description	Patten Rule	Constraint
		• Fuzzy match – An asterisk (*) matches zero or more characters except for slashes (/). – A pair of asterisks (**) matches zero or more characters including slashes (/). – A question mark (?) matches exactly one character, but not slashes (/). – Commas (,) are used to separate patterns in {}. Patterns in {} are in an OR relationship. – Wildcard characters asterisk (*) and question mark (?) are escaped by backslashes (\). In other cases, a backslash (\) means itself.	

Option	Description	Patten Rule	Constraint
Time Range	Filters files and directories to be migrated based on when they were last modified. Only files and directories whose last modification times fall in the configured time range will be migrated. The start time and end time can be left empty. If they are left empty, the system will not filter out source files by time. The time can be precise to the minute.		

The following table lists example pattern rules for different scenarios.

NOTICE

Assume that you want to migrate the directory **test** in the source storage system.

- If the source storage system is a NAS device, enter **/test** as an "include" pattern.
- If the source storage system is an object storage system, enter **test** as an "include" pattern.

Scenario	Example Pattern for NAS	Example Pattern for Object Storage	Description
File paths that end with *xx*	/xx /**xx	xx /**xx	*xx* can be an expression containing asterisks (*) and question marks (?).
Files whose names start with *xx* in the root directory	/xx* • **/testssss** matches the pattern. • **/test/xx** does not matches the pattern.	/xx* • **testssss** matches the pattern. • **/test/xx** does not matches the pattern.	
File paths that start with *xx*	/xx** /xx/**	xx** xx/**	
Files whose names contain *xx*	**xx*	**xx*	
File paths that contain *xx*	**xx**	**xx**	

Scenario	Example Pattern for NAS	Example Pattern for Object Storage	Description
File paths that start with *xx* and end with *yy*.	/xx**yy	xx**yy	*xx* and *yy* can be expressions containing asterisks (*) and question marks (?).
File paths that end with *xx* or *yy*	**{xx,yy}	**{xx,yy}	
Files whose names contain *xx* or *yy*	**{xx,yy}*	**{xx,yy}*	
Files paths that contain *xx* or *yy*	**{xx,yy}**	**{xx,yy}**	

10.5 Creating a Batch Object Storage Migration Workflow

This section describes how to create a workflow to efficiently migrate buckets in batches.

Notes and Constraints

- For more information, see **What Are the Restrictions on Using MgC for Storage Migration?**
- Data in the Deep Cold Archive storage of Alibaba Cloud OSS cannot be restored using the restoration function of MgC. You need to manually restore the data before migration.

Warnings

When you create a workflow, there are three options for **Overwrite Existing**: **Never**, **Always**, and **If older or different size**. If you choose **Never**, restarting the migration task after interruptions or pauses may cause incomplete data migration, but the task may still be displayed as successful, which affects data integrity. Use the **Never** option with caution.

Prerequisites

You have **created a batch object storage migration plan**.

Procedure

Step 1 Sign in to the **MgC console**.

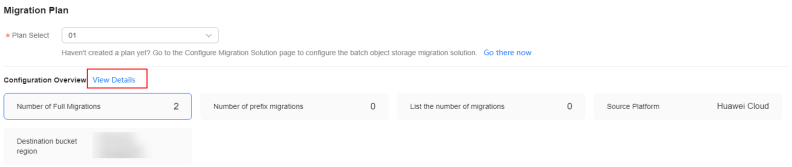
Step 2 In the navigation pane on the left, choose **Migrate** > **Workflows**. Select a **migration project** in the upper left corner of the page.

Step 3 Click **Create Workflow** in the upper right corner of the page.

Step 4 In the **Batch Object Storage Migration** card, click **Preview Steps** to view the stages and steps predefined in the template and the detailed description of each stage and step. Steps of the **Automated** type will be automatically performed by MgC. Click **Configure Workflow** in the lower right corner.

Step 5 In the **Basic Information** area, enter a name and description for the workflow.

Step 6 In the **Migration Plan** area, select the created **migration plan**. Then you should view the overview of the migration plan. Click **View Details** to view more information about the plan.



Step 7 In the **Migration Cluster** area, select the cluster used for the migration. The cluster selected in the migration plan is used by default, but you can select another one if needed. The modification is applied to the current workflow but not to the migration plan.

Step 8 Configure the migration task based on **Table 10-9**.

Table 10-9 Parameters for configuring a migration task

Parameter	Value	Description
Concurrent Subtasks	-	Set the maximum number of concurrent subtasks. There cannot be more than 10 concurrent subtasks for each online migration node. For example, if there are 2 online migration nodes, the maximum number of subtasks can be 20 or any number below.
Overwrite Existing	Never	Files existing at the migration target are never overwritten. **WARNING** • If you choose **Never** for the initial migration, the attributes of involved parent folders at the source will not be migrated to the target. As a result, the folder attributes may be incomplete at the target. To avoid this issue, use the **Never** option with caution for the initial migration. • If a migration task is paused or interrupted and then restarted or resumed, the **Never** option will cause the system to skip files that were not completely migrated earlier, but the task may still be marked as successful. This affects data integrity. To avoid this issue, use the **Never** option with caution.
	Always	Files existing at the migration target are always overwritten.

Parameter	Value	Description
	If older or different size	• Files that already exist at the target will be overwritten if they are older than or have different sizes from the paired files at the source. • Verification will be performed for folders after their contents are migrated. Folders that already exist at the target will be overwritten if they have different last modification times, sizes,or permissions from the paired folders at the source. **NOTE** For empty folders, the overwrite policy is the same as that for files.
Migrate Metadata	-	Decide whether to migrate metadata. • If you select this option, object metadata will be migrated. • If you do not select this option, only the Content-Type and Content-Encoding metadata will be migrated.

Step 9 (Optional) Configure advanced options based on **Table 10-10**.

Table 10-10 Advanced options

Parameter	Description
Target Storage Class	Choose the storage class that your data will be migrated to in the target bucket. For details about storage classes, see **Introduction to Storage Classes**.
Enable KMS Encryption	• If you do not select this option, whether migrated data will be encrypted in the target bucket depends on the server-side encryption setting of the bucket. • If you select this option, all migrated objects will be encrypted before they are stored in the target bucket. **NOTE** • Using KMS to encrypt migrated data may slow down the migration speed by about 10%. • This option is only available when KMS is supported in the region you are migrating to.

Parameter	Description
Restore Archive Data	• If you do not select this option, the system records archive objects in the list of objects that failed to be migrated and continues to migrate other objects in the migration task. • If you select this option, the system automatically restores and migrates archive objects in the migration task. If an archive object fails to be restored, the system skips it and records it in the list of objects that failed to be migrated and continues to migrate other objects in the migration task. **NOTE** The system will restore archive data before migrating it, and you pay the source cloud platform for the API requests and storage space generated accordingly.
Filter Source Data	Filter files to be migrated by applying filters. For details about the filters, see **Source Data Filters**.
Download Data from CDN	If the default domain name cannot meet your migration requirements, then as long as the source cloud service provider supports custom domain names, you can bind a custom domain name to the source bucket, and enable the CDN service on the source platform to reduce data download expenses. Enter a custom domain name in the **Domain Name** text box and select a transmission protocol. HTTPS is more secure than HTTP and is recommended. If the migration source is the Alibaba Cloud OSS or Tencent Cloud COS, you also need to select an authentication type and enter an authentication key.
Send SMN Notification	Determine whether to use SMN to get notifications about migration results. • If you do not select this option, no SMN messages are sent after the migration. • If you select this option, after the migration, SMN messages are sent to the subscribers of the selected topic. You can select the language and trigger conditions for sending messages.

Parameter	Description
Limit Traffic	Allocate the maximum bandwidth to be used by the workflow during a specified period. • If you do not select this option, migration traffic is not limited. • If you select this option, limit the migration traffic by setting **Start Time**, **End Time**, and **Bandwidth Limit**. For example, if you set **Start Time** to **08:00**, **End Time** to **12:00**, and **Bandwidth Limit** to **20 MB/s**, the maximum migration speed is limited to 20 MB/s when the migration task is running from 08:00 to 12:00. The migration speed is not limited beyond this period. **NOTE** – The rate limit ranges from 0 MB/s to 1,048,576 MB/s. – A maximum of five rules can be added. – The time is the local standard time of the region you are migrating to.

Step 10 Click **Next: Confirm**.

Step 11 Confirm the workflow settings, and click **Confirm**. The **Run Workflow** dialog box is displayed, which indicates that the workflow has been created.

- If you want to start the migration immediately, click **Confirm** to run the workflow.
- If you want to **add a stage or step** to the workflow, click **Cancel**. The workflow enters a **Waiting** state, and the migration is not started. To start the migration, click **Run** in the **Operation** column.

Step 12 On the migration workflow details page, view the workflow settings and the migration progress. You can also perform the following operations:

- Move the cursor to the migration progress bar of a resource. In the displayed window, view the migration details about the resource.
- When a migration reaches a step that requires manual confirmation, place the cursor on the progress bar and click **Confirm** next to the step status in the displayed window. The migration can continue only after you confirm.
- In the **Basic Information** area, click **Manage** next to the cluster name. The cluster details page is displayed on the right. On the displayed page, you can:
  - Add, edit, or delete traffic limiting rules to control cluster traffic based on your requirements.
  - Add or delete migration nodes or list nodes, or upgrade plug-ins for existing nodes as required.

----End

10.6 Creating a Batch File Storage Migration Workflow

This section describes how to create a workflow to efficiently migrate file systems in batches.

Notes and Constraints

For more information, see **What Are the Restrictions on Using MgC for Storage Migration?**

Warnings

When you create a workflow, there are three options for **Overwrite Existing**: **Never**, **Always**, and **If older or different size**. If you choose **Never**, restarting the migration task after interruptions or pauses may cause incomplete data migration, but the task may still be displayed as successful, which affects data integrity. Use the **Never** option with caution.

Prerequisites

You have **created a batch file storage migration plan**.

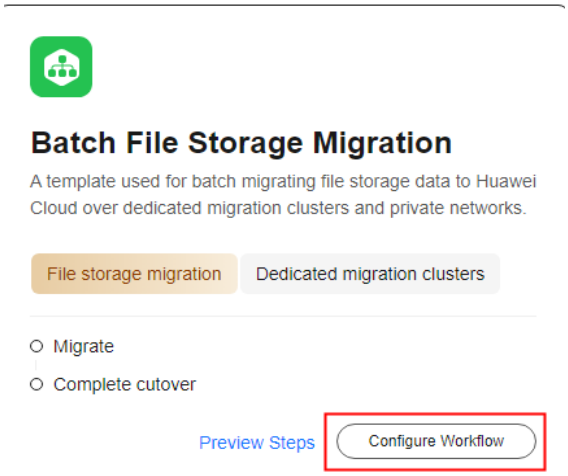
Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Workflows**. Select a **migration project** in the upper left corner of the page.

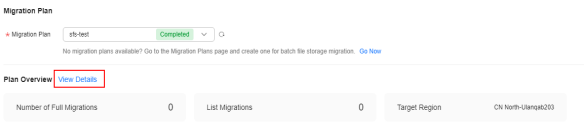
Step 3 Click **Create Workflow** in the upper right corner of the page.

Step 4 In the **Batch File Storage Migration** card, click **Preview Steps** to view the stages and steps predefined in the template and the detailed description of each stage and step. Steps of the **Automated** type will be automatically performed by MgC. Click **Configure Workflow** in the lower right corner.



Step 5 In the **Basic Information** area, enter a name and description for the workflow.

Step 6 In the **Migration Plan** area, select the created **migration plan**. Then you should view the overview of the migration plan. Click **View Details** to view more information about the plan.



Step 7 In the **Migration Cluster** area, select the cluster used for the migration. The cluster specified in the migration plan is preselected by default, but you can select another one if needed. The modification is applied to the current workflow but not to the migration plan.

Step 8 Configure the migration task based on **Table 10-11**.

Table 10-11 Parameters for configuring a migration task

Parameter	Value	Description
Concurrent Subtasks	-	Set the maximum number of concurrent subtasks. There cannot be more than 10 concurrent subtasks for each online migration node. For example, if there are 2 online migration nodes, the maximum number of subtasks can be 20 or any number below.
Overwrite Existing	Never	Files existing at the migration target are never overwritten. **WARNING** • If you choose **Never** for the initial migration, the attributes of involved parent folders at the source will not be migrated to the target. As a result, the folder attributes may be incomplete at the target. To avoid this issue, use the **Never** option with caution for the initial migration. • If a migration task is paused or interrupted and then restarted or resumed, the **Never** option will cause the system to skip files that were not completely migrated earlier, but the task may still be marked as successful. This affects data integrity. To avoid this issue, use the **Never** option with caution.
	Always	Files existing at the migration target are always overwritten.
	If older or different size	• Files that already exist at the target will be overwritten if they are older than or have different sizes from the paired files at the source. • Verification will be performed for folders after their contents are migrated. Folders that already exist at the target will be overwritten if they have different last modification times, sizes,or permissions from the paired folders at the source. **NOTE** For empty folders, the overwrite policy is the same as that for files.
Migrate Metadata	-	Decide whether to migrate metadata. • If you select this option, object metadata will be migrated. • If you do not select this option, only the Content-Type and Content-Encoding metadata will be migrated.

Step 9 (Optional) Configure advanced options based on **Table 10-12**.

Table 10-12 Advanced options

Parameter	Description
Sync Incremental Before Cutover	• If you select **Yes**, an incremental migration step will be automatically added after the migration verification step in the migration phase. • If you select **No**, the service cutover phase starts after the migration verification step in the migration phase is complete.
Filter Source Data	Filter files and directories to be migrated by applying filters. For details about the filters, see **Source Data Filters**.
Send SMN Notification	Determine whether to use SMN to get notifications about migration results. • If you do not select this option, no SMN messages will be sent after the migration is complete. • If you select this option, after the migration is complete, SMN messages are sent to the subscribers of the selected topic. You can select the language and trigger conditions for sending messages.
Limit Traffic	Allocate the maximum bandwidth to be used by the workflow during a specified period. • If you do not select this option, migration traffic is not limited. • If you select this option, limit the migration traffic by setting **Start Time**, **End Time**, and **Bandwidth Limit**. For example, if you set **Start Time** to **08:00**, **End Time** to **12:00**, and **Bandwidth Limit** to **20 MB/s**, the maximum migration speed is limited to 20 MB/s when the migration runs from 08:00 to 12:00. The migration speed is not limited beyond this period. **NOTE** – The bandwidth limit ranges from 0 MB/s to 1,048,576 MB/s. – A maximum of five rules can be added. – The time is the local standard time of the region you are migrating to.

Step 10 Click **Next: Confirm**.

Step 11 Confirm the workflow settings, and click **Confirm**. The **Run Workflow** dialog box is displayed, which indicates that the workflow has been created.

- If you want to start the migration immediately, click **Confirm** to run the workflow.

- If you want to **add a stage or step** to the workflow, click **Cancel**. The workflow enters a **Waiting** state, and the migration is not started. To start the migration, click **Run** in the **Operation** column.

Step 12 In the workflow list, click the workflow name to go to its details page. You can view the configuration information and migration progress of the workflow. You can also perform the following operations:

- Move the cursor to the migration progress bar of a resource. In the displayed window, view the migration details about the resource.
- When a migration reaches a step that requires manual confirmation, place the cursor on the progress bar and click **Confirm** next to the step status in the displayed window. The migration can continue only after you confirm.
- In the **Basic Information** area, click **Manage** next to the migration cluster name. The cluster details page is displayed on the right. On the displayed page, you can:
  - Add, edit, or delete traffic limiting rules to control cluster traffic based on your requirements.
  - Add or delete migration nodes or list nodes, or upgrade plug-ins for existing nodes as required.

----End

10.7 Adding a Stage or Step

You can customize a migration workflow to fit your requirements better by adding stages or steps to the workflow.

CAUTION

- You can only add a stage before or after an existing stage when the existing stage is in **Waiting** or **Paused** state. This is true for step additions.
- If the previous stage is in **Running**, **Paused**, or **Complete** state, you can only add a stage after the existing stage. This is also true for step additions.
- You cannot add steps after repeatable steps.

Adding a Stage

Step 1 On the migration workflow details page, move the cursor to the migration stage before or after which you want to add a stage. In the displayed window, choose **Add Stage Before** or **Add Stage After**.

Step 2 Enter a stage name and description, click **Add Step**, select a step type, enter a step name and description, and click **Confirm**. Multiple steps can be added.

Step 3 Click **Confirm**.

NOTICE

Manually added stages can be modified or deleted, but pre-defined stages cannot.

----End

Adding a Step

Step 1 On the migration workflow details page, move the cursor to the step before or after which you want to add a step. In the displayed window, choose **Add Step Before** or **Add Step After**.

Step 2 Select a step type based on **Table 10-13**, enter a step name and description, and click **Confirm**.

Table 10-13 Step types

Type	Description
Checkpoint	You need to manually confirm this type of steps, so that the workflows can continue.

Step 3 Go back to the migration stage and view the added step.

NOTICE

Manually added steps can be modified or deleted, but pre-defined steps cannot.

----End

11 Big Data Migration

11.1 Precautions

Notes and Constraints

- MgC cannot migrate the DATETIME type of data from MaxCompute. Before the migration, you need to convert such data to the STRING format.
- MgC cannot migrate the VARCHAR or CHAR type of data from MaxCompute. Before the migration, you need to convert such data to the STRING format.
- MgC cannot migrate the metadata of transaction tables (those with the **transactional=true** field in DDL statements).
- The TIMESTAMP_NTZ type of data in MaxCompute is not supported by the Alibaba Cloud SDK. As a result, the metadata of involved tables will fail to be migrated.

Notes

- Editing a CSV template file

  When using the CSV template provided by MgC to filter data, do not use Excel to edit the CSV file. The file edited and saved in Excel cannot be identified by MgC.

- Selecting a DLI connection

  Metadata migration and data migration require different DLI connections. Metadata migration requires a DLI connection with a SQL queue configured, and data migration requires a DLI connection with a general-purpose queue configured. If an incorrect connection is configured for the migration task, the migration task will fail.

- Constraints on the number of databases

  During the metadata migration, MgC automatically creates the same number of databases identified in metadata on DLI. A maximum of 10 DLI databases can be created in an account. If the number of databases you want to migrate exceeds the quota, the metadata migration may fail.

11.2 Preparations

To ensure a smooth migration, you need to make the following preparations.

Preparing a Huawei Account

Before using MgC, prepare a HUAWEI ID or an IAM user that can access MgC and obtain an AK/SK pair for the account or IAM user. For details, see **Making Preparations**.

Obtaining an AK/SK Pair for Your Alibaba Cloud Account

Obtain an AK/SK pair for your Alibaba Cloud account. For more information, see **Viewing the AccessKey Pairs of a RAM User**.

Ensure that the AK/SK pair has the following permissions:

- **AliyunReadOnlyAccess**: read-only permission for OSS
- **AliyunMaxComputeReadOnlyAccess**: read-only permission for MaxCompute

For details about how to obtain these permissions, see **Granting Permissions to RAM Users**.

(Optional) If there are partitioned tables to be migrated, grant the **Information Schema** permission to the source account. For details, see **Authorization for RAM Users**.

Creating a Migration Project

On the MgC console, create a migration project. For details, see **Managing Migration Projects**.

Configuring an Agency

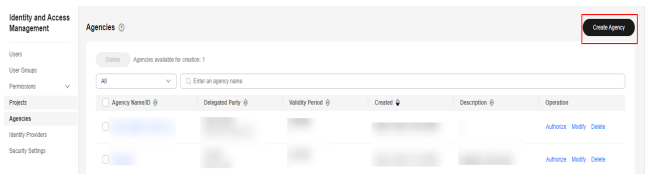
To ensure that DLI functions can be used properly, you need to configure an agency with DLI and OBS permissions.

Step 1 Sign in to the Huawei Cloud management console.

Step 2 Hover the mouse over the username in the upper right corner and choose **Identity and Access Management** from the drop-down list.

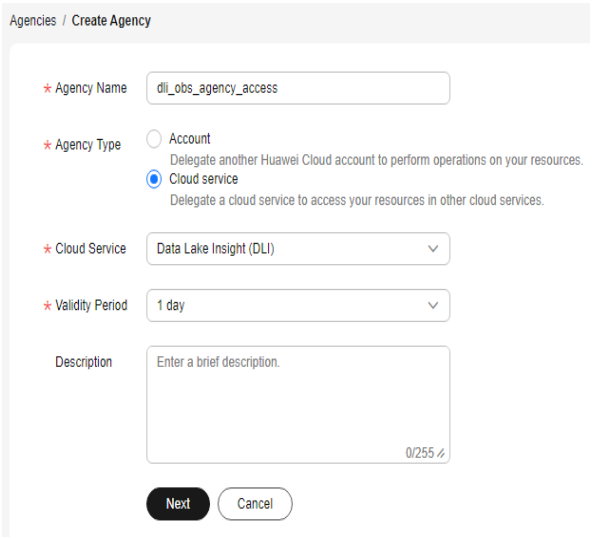
Step 3 In the navigation pane, choose **Agencies**.

Step 4 Click **Create Agency**.



Step 5 On the **Create Agency** page, set the following parameters:

- **Agency Name**: Set a name, for example, **dli_obs_agency_access**.
- **Agency Type**: Select **Cloud service**.
- **Cloud Service**: Select **(Data Lake Insight) DLI** from the drop-down list.
- **Validity Period**: Set a period as needed.
- **Description**: This parameter is optional.



Step 6 Click **Next**. The **Select Policy/Role** tab is displayed.

Step 7 Click **Create Policy** in the upper right corner. Create two policies (one for OBS and one for DLI) by referring to **Step 8** and **Step 9**. If there are existing policies containing the required permissions, you can use them and skip this step, as well as steps 8 and 9.

Step 8 Configure policy information.

1. **Policy Name**: Set a name, for example, **dli-obs-agency**.
2. **Policy View**: Select **JSON**.
3. Copy and paste the following content to the **Policy Content** box.

> **CAUTION**
>
> Replace **bucketName** with the name of the bucket where the JAR packages are stored.

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "obs:bucket:GetBucketPolicy",
                "obs:bucket:GetLifecycleConfiguration",
                "obs:bucket:GetBucketLocation",
                "obs:bucket:ListBucketMultipartUploads",
                "obs:bucket:GetBucketLogging",
                "obs:object:GetObjectVersion",
                "obs:bucket:GetBucketStorage",
                "obs:bucket:GetBucketVersioning",
```



```
                "obs:object:GetObject",
                "obs:object:GetObjectVersionAcl",
                "obs:object:DeleteObject",
                "obs:object:ListMultipartUploadParts",
                "obs:bucket:HeadBucket",
                "obs:bucket:GetBucketAcl",
                "obs:bucket:GetBucketStoragePolicy",
                "obs:object:AbortMultipartUpload",
                "obs:object:DeleteObjectVersion",
                "obs:object:GetObjectAcl",
                "obs:bucket:ListBucketVersions",
                "obs:bucket:ListBucket",
                "obs:object:PutObject"
            ],
            "Resource": [
                "OBS:*:*:bucket:bucketName",//Replace bucketName with the name of the bucket where the JAR packages are stored.
                "OBS:*:*:object:*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "obs:bucket:ListAllMyBuckets"
            ]
        }
    ]
}
```

Step 9 Configure policy information.

1. **Policy Name**: Set a name, for example, **dli-agency**.
2. **Policy View**: Select **JSON**.
3. Copy and paste the following content to the **Policy Content** box.

```
{
    "Version": "1.1",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "dli:table:showPartitions",
                "dli:table:alterTableAddPartition",
                "dli:table:alterTableAddColumns",
                "dli:table:alterTableRenamePartition",
                "dli:table:delete",
                "dli:column:select",
                "dli:database:dropFunction",
                "dli:table:insertOverwriteTable",
                "dli:table:describeTable",
                "dli:database:explain",
                "dli:table:insertIntoTable",
                "dli:database:createDatabase",
                "dli:table:alterView",
                "dli:table:showCreateTable",
                "dli:table:alterTableRename",
                "dli:table:compaction",
                "dli:database:displayAllDatabases",
                "dli:database:dropDatabase",
                "dli:table:truncateTable",
                "dli:table:select",
                "dli:table:alterTableDropColumns",
                "dli:table:alterTableSetProperties",
                "dli:database:displayAllTables",
                "dli:database:createFunction",
                "dli:table:alterTableChangeColumn",
                "dli:database:describeFunction",
                "dli:table:showSegments",
                "dli:database:createView",
```



```
            "dli:database:createTable",
            "dli:table:showTableProperties",
            "dli:database:showFunctions",
            "dli:database:displayDatabase",
            "dli:table:alterTableRecoverPartition",
            "dli:table:dropTable",
            "dli:table:update",
            "dli:table:alterTableDropPartition"
        ]
    }
  ]
}
```

Step 10 Click **Next**.

Step 11 Select the created custom policies with OBS and DLI permissions and click **Next**. Set **Scope** to **All resources**.

Step 12 Click **OK**. It takes 15 to 30 minutes for the authorization to be in effect.

Step 13 Update the agency permissions by referring to **Updating DLI Agency Permissions**.

----End

Creating a VPC

Before purchasing an ECS, you need to create a VPC and subnet for it. For details, see **Creating a VPC and Subnet**.

> ⚠ CAUTION
>
> The VPC of the ECS must not conflict with the CIDR block used by the DLI elastic resource pool you prepare below. When you create a DLI queue, the preset CIDR block is 172.16.0.0/18.

Purchasing an ECS

- Purchase a Linux ECS. The ECS and the DLI resources you prepare later must be in the same region. For details, see **Purchasing an ECS**. Select the **created VPC and subnet** for the ECS. The ECS must:
  - Be able to access the Internet and the domain names of MgC, IoTDA, and other cloud services. For details about the domain names to be accessed, see **Domain Names**.
  - Allow outbound traffic on 8883 if the ECS is in a security group.
  - Run CentOS 8.*X*.
  - Have at least 8 vCPUs and 16 GB of memory.
- Configure an EIP for the ECS, so that the ECS can access the Internet. For details, see **Assigning an EIP** and **Binding an EIP to an ECS**.
  - **Billing Mode**: Select **Pay-per-use**.
  - **Bandwidth**: **5 Mbit/s** is recommended.

Installing Edge and Connecting Edge to MgC

- On the purchased ECS, install Edge, which will be used for data verification. For details, see **Installing Edge for Linux**.

- Register an Edge account. In the address box of a browser, enter the NIC IP address of the Linux ECS and port 27080, for example, https://x.x.x.x:27080. After the first login, the registration page is displayed. Enter a username and password, confirm the password, and click **Privacy Statement**. Read the statement carefully, select **I have read and agree to the Privacy Statement,** and click **OK** to complete the registration.
- **Connect Edge to MgC**.

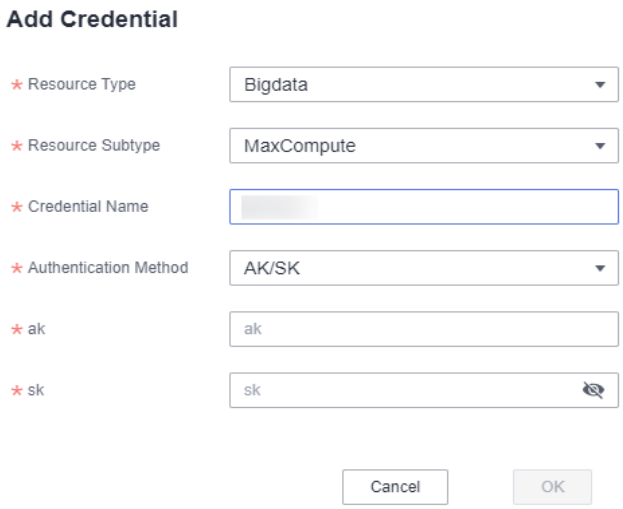
CAUTION

Edge does not support automatic restart. Do not restart Edge during task execution, or tasks will fail.

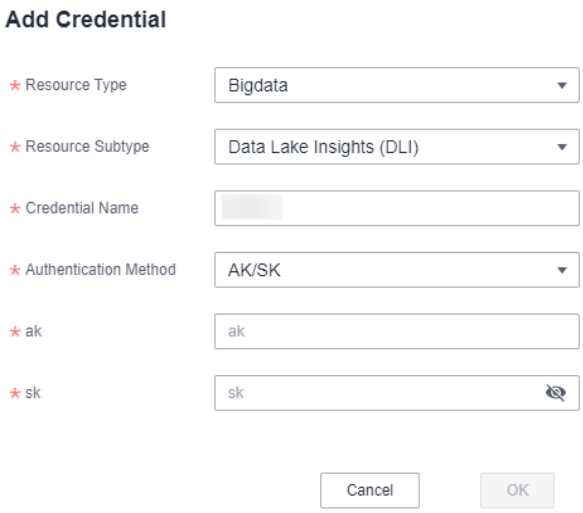
Adding Credentials

On the Edge console, add the AK/SK pairs of your Alibaba Cloud account and Huawei Cloud account. For details, see **Adding Resource Credentials**.

- Enter the AK/SK pair of your source Alibaba Cloud account. This key pair will be used to access your MaxCompute resources.



- Enter the AK/SK pair of your Huawei Cloud account. This key pair will be used to access your DLI resources.



Creating an OBS Bucket and Uploading JAR Packages

Create a bucket on Huawei Cloud OBS and upload the Java files (in JAR packages) that data migration depends on to the bucket. For details about how to create an OBS bucket, see **Creating a Bucket**. For details about how to upload files, see **Uploading an Object**.

The JAR packages that data migration depends on are **migration-dli-spark-1.0.0.jar**, **fastjson-1.2.54.jar**, and **datasource.jar**. Below, you'll find a description of what each package is used for and how to obtain them.

- **migration-dli-spark-1.0.0.jar**
  - This package is used to create Spark sessions and submit SQL statements.
  - You can obtain the package from the **/opt/cloud/Edge/tools/plugins/collectors/bigdata-migration/dliSpark** directory on the **ECS where Edge is installed**.
- **fastjson-1.2.54.jar**
  - This package provides fast JSON conversion.
  - You can obtain the package from the **/opt/cloud/Edge/tools/plugins/collectors/bigdata-migration/deltaSpark** directory on the **ECS where Edge is installed**.
- **datasource.jar**
  - This package contains the configuration and connection logic of data sources and allows Edge to connect to different databases or data storage systems.
  - You need to obtain and compile files in **datasource.jar** as required..

Purchasing an Elastic Resource Pool and Adding a Queue on DLI

An elastic resource pool provides compute resources (CPU and memory) required for running DLI jobs. For details about how to buy a resource pool, see **Buying an Elastic Resource Pool**.

You can add a general-purpose queue and an SQL queue to the pool for running jobs. For more information, see **Creating Queues in an Elastic Resource Pool**.

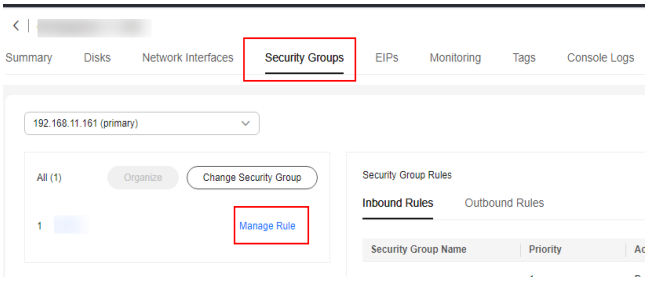
Creating an Enhanced Datasource Connection on DLI

Step 1 Sign in to the DLI console. In the navigation pane on the left, choose **Resources** > **Resource Pool**.

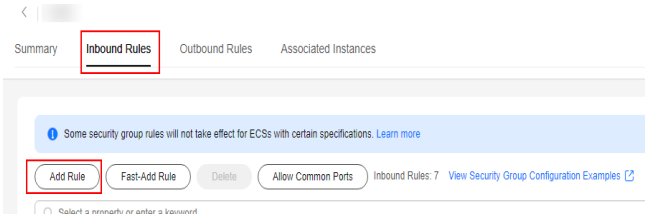
Step 2 Click next to the name of the elastic resource pool. In the expanded information, obtain the CIDR block of the elastic resource pool.

Step 3 In the security group of the ECS with Edge installed, add an inbound rule to allow access from the CIDR block of the elastic resource pool.

1. Sign in to the ECS console.
2. In the ECS list, click the name of the Linux ECS with Edge installed.
3. Click the **Security Groups** tab and click **Manage Rule**.



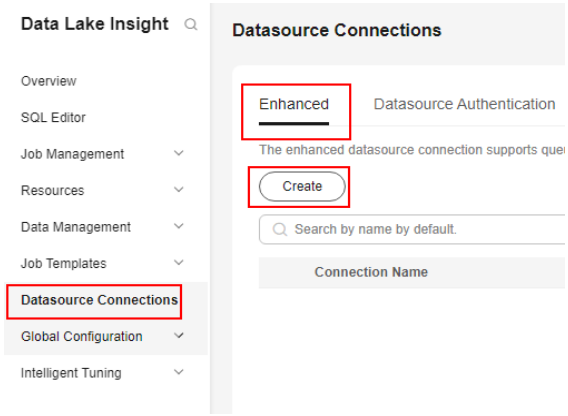
4. Under **Inbound Rules**, click **Add Rule**.



5. Configure parameters for the inbound rule.
   - **Priority**: Set it to **1**.
   - **Action**: Select **Allow**.
   - **Type**: Select **IPv4**.
   - **Protocol & Port**: Select **Protocols/All**.
   - Source: Select **IP Address** and enter the CIDR block of the elastic resource pool.
6. Click **OK**.

Step 4 Switch to the DLI console. In the navigation pane on the left, click **Datasource Connections**.

Step 5 Under **Enhanced**, click **Create**.



Step 6 Configure enhanced datasource connection information based on **Table 11-1**.

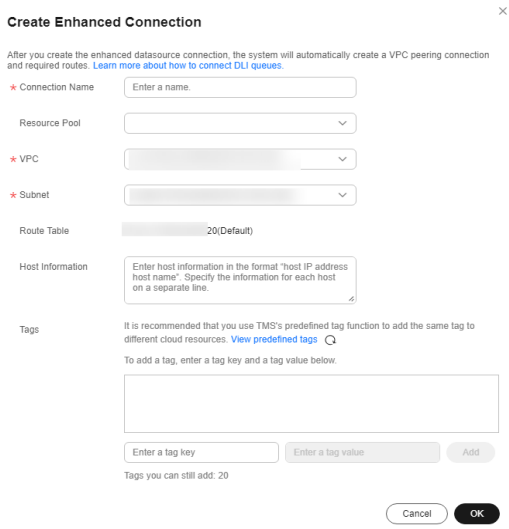


Table 11-1 Parameters required for creating an enhanced connection

Parameter	Configuration
Connection Name	Enter a name.
Resource Pool	Select the **purchased elastic resource pool**.
VPC	Select the **created VPC and subnet**.
Subnet	
Route Table	Retain the preset value.

Parameter	Configuration
Host Information	Add the information about the source MaxCompute hosts in the following format: **EndpointIP Endpoint** (with a space in between) **TunnelEndpointIP TunnelEndpoint** (with a space in between) Separate two pieces of information by pressing **Enter**. For example: 118.178.xxx.xx service.cn-hangzhou.maxcompute.aliyun.com.vipgds.alibabadns.com 47.97.xxx.xx dt.cn-hangzhou.maxcompute.aliyun.com To obtain the values of **EndpointIP** and **TunnelEndpointIP**, use any server with a public IP address to ping the endpoint and tunnel endpoint of the region where the source MaxCompute cluster is located. The corresponding IP addresses will display in the command output. For details about MaxCompute endpoints and Tunnel endpoints, see **Endpoints in different regions**.

Step 7 Click **OK**. After the creation process is finished, the **Connection Status** of the newly created connection will be **Active**, indicating that the connection has been successfully created.

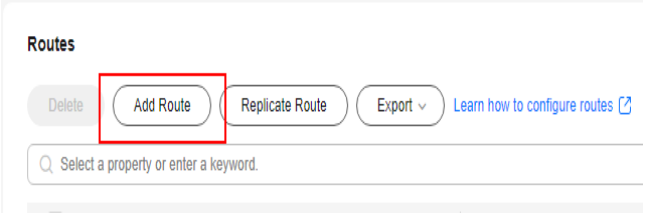
----End

Adding and Configuring Routes

- Add routes.

  Add two routes for the **DLI enhanced datasource connection**. For details, see **Adding a Route**. In the routes, the IP address must be the same as the host IP addresses configured during datasource connection creation.
- Configure routes.

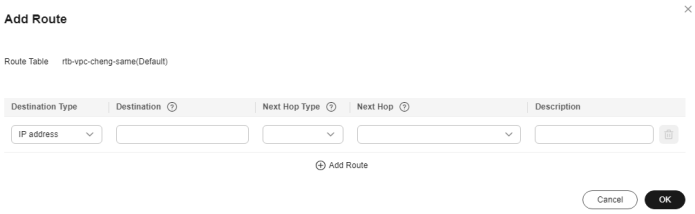
1. Sign in to the VPC console.
2. In the navigation pane on the left, choose **Virtual Private Cloud** > **Route Tables**.
3. In the route table list, click the route table used for creating a datasource connection (that is, the route table of the VPC where the ECS resides).
4. Click **Add Route**.



5. Set the parameters by following the instructions. You need to click

to add two routes.

- **Destination Type**: Select **IP address**.
- **Destination**: Enter the host IP address configured during datasource connection creation.
- **Next Hop Type**: Select **Server**.
- **Next Hop**: Select the **purchased ECS**.



Configuring SNAT Rules

Configure SNAT rules for the ECS by following the instructions below. Restarting the ECS will clear the rules, so you will need to reconfigure them after a restart.

Step 1 Log in to the **purchased ECS**.

Step 2 Run the following commands in sequence:

```
sysctl net.ipv4.ip_forward=1
```

This command is used to enable IP forwarding in Linux.

```
iptables -t nat -A POSTROUTING -o eth0 -s {CIDR block where the DLI elastic resource pool resides} -j SNAT --to {Private IP address of the ECS}
```

This command is used to set an iptables rule for network address translation.

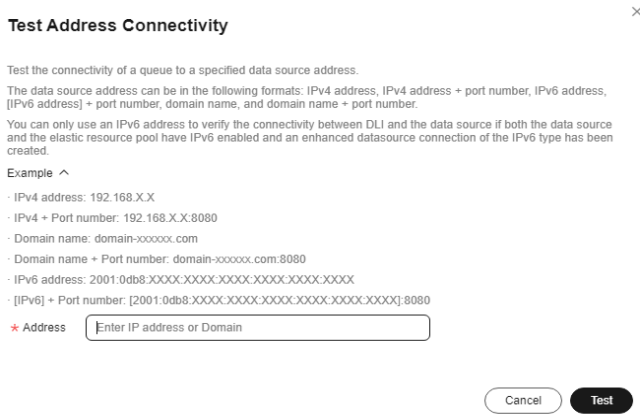
----End

Testing the Connectivity Between the DLI Queue and the Data Source

Step 1 Sign in to the DLI console. In the navigation pane on the left, choose **Resources** > **Queue Management**.

Step 2 Locate the elastic resource pool that **the DLI queue was added** and choose **More** > **Test Address Connectivity** in the **Operation** column.

Step 3 Enter the endpoint, endpoint IP address, Tunnel endpoint, and Tunnel endpoint IP address of MaxCompute to perform four connectivity tests. You are advised to use port 443.



Step 4 Click **Test**.

If the test address is reachable, you will receive a message.

If the test address is unreachable, you will also receive a message. Check the network configurations and try again. Network configurations include the VPC peering and the datasource connection. Check whether they have been activated.

----End

Adding to the DLI Spark 3.3 Whitelist and the JAR Program Whitelist for Metadata Access

Contact the DLI technical support to whitelist you to use the DLI Spark 3.3 feature and allow Jar access to DLI metadata.

(Optional) Enabling DLI Spark Lifecycle Whitelist

If the metadata to be migrated has a lifecycle (that is, the DDL contains the LIFECYCLE field), contact DLI technical support to enable the Spark lifecycle feature whitelist.

(Optional) Enabling the CIDR Block 100 Whitelist

If you use Direct Connect to migrate data, you need to request VPC support to enable the whitelist of the 100.100.x.x segment.

Submit a service ticket to the VPC service and provide the following information:

- Huawei Cloud account and project ID of the region where your DLI resources reside. For details about how to obtain them, see **API Credentials**.
- DLI tenant name and tenant project ID: Contact DLI technical support to obtain them.

11.3 Creating a Source Connection

A source connection enables Edge to access and read data in Alibaba Cloud MaxCompute.

Prerequisites

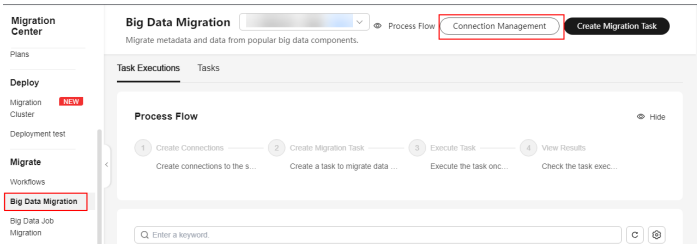
You have completed all **preparations**.

Procedure

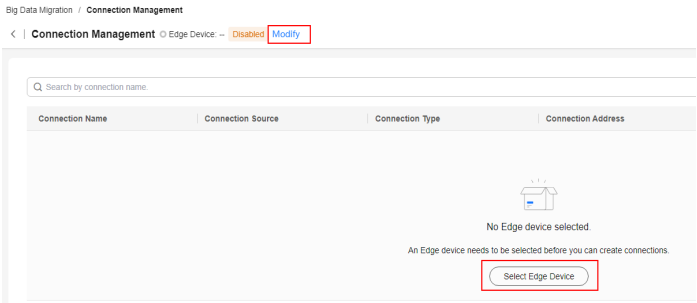
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Migration**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 Click **Connection Management** in the upper right corner of the page.



Step 4 If you are performing a big data migration on MgC for the first time, you need to select an Edge device to enable this migration feature. Click **Modify** in the upper left corner of the page or click **Select Edge Device**. In the drop-down list, select the Edge device you connected to MgC in **Making Preparations**.

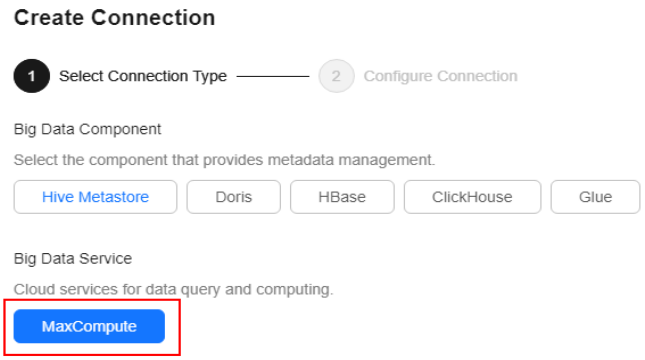


CAUTION

Ensure that the selected Edge device is always **Online** and **Enabled** until your migration is complete.

Step 5 Click **Create Source Connection** in the upper right corner of the page.

Step 6 On the **Select Connection Type** page, select **MaxCompute** and click **Next**.

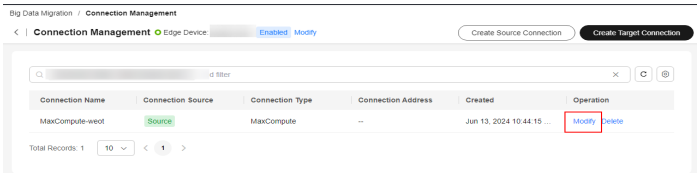


Step 7 Configure the **parameters for creating a MaxCompute connection**, and click **Test**. If the test is successful, the connection is set up.

Table 11-2 Parameters for creating a MaxCompute connection

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **MaxCompute-***4 random characters* (including letters and numbers). You can also customize a name.
Edge Device	Select the Edge device you connected to MgC in **Preparations**.
Alibaba Cloud Credential	Select the Alibaba Cloud credential (an AK/SK pair) you added on Edge in **Making Preparations**.
MaxCompute Project	Enter the name of the MaxCompute project where the data to be migrated is managed. You can obtain the project name from the **Projects** page on the MaxCompute console.
Endpoint	Enter the VPC endpoint of the region where the MaxCompute project is located. For details about the MaxCompute VPC endpoint in each region, see **Endpoints in different regions (VPC)**.

After the source connection is created, you can view it in the list on the **Connection Management** page. If necessary, click **Modify** in the **Operation** column to modify the connection configuration.



----End

11.4 Creating a Target Connection

A target connection is used to write source data to DLI.

Precautions

For a metadata migration, a DLI connection with an SQL queue configured is required.

For a data migration, two DLI connections are required: one with an SQL queue configured and the other with a general queue.

Prerequisites

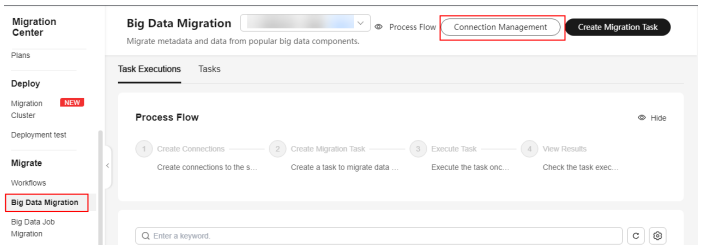
You have completed all **preparations**.

Procedure

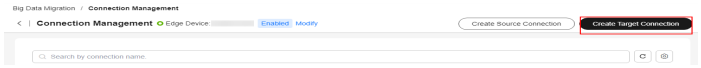
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Migration**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 Click **Connection Management** in the upper right corner of the page.



Step 4 In the upper right corner of the page, click **Create Target Connection**.



Step 5 On the **Select Connection Type** page, select **Data Lake Insight (DLI)** and click **Next**.

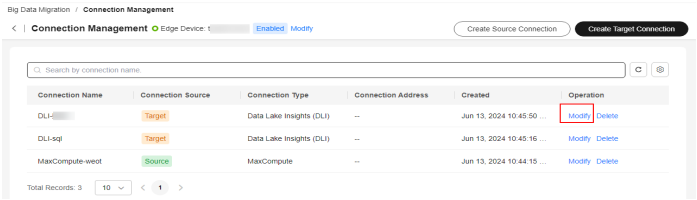
Step 6 Configure the parameters listed in **Table 11-3**, and click **Test**. If the test is successful, the connection is set up.

Table 11-3 Parameters for creating a DLI connection

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **DLI-***4 random characters* (including letters and numbers). You can also customize a name.

Parameter	Configuration
DLI Credential	Select the Huawei Cloud credential (an AK/SK pair) you added on Edge in **Making Preparations**. If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed.
Region	Select the region where your DLI resources are located. If you do not select **This is my MgC credential**, you need to manually enter the region code, for example, **cn-north-1**.
Project	Select the project where your DLI resources are managed. Your source data will be managed in this project after being migrated. If you do not select **This is my MgC credential**, you need to manually enter the project ID. For details about how to obtain the project ID, see **Obtaining Project Information**. 
Queue	Enter the name of the SQL queue or general queue created in **Purchasing an Elastic Resource Pool and Adding a Queue on DLI**. • If you are creating a connection using an SQL queue, enter the name of the SQL queue. • If you are creating a connection using a general queue, enter the name of the general queue.

After the target connection is created, you can view it in the list on the **Connection Management** page. If necessary, click **Modify** in the **Operation** column to modify the connection configuration.



----End

11.5 Migrating Big Data

11.5.1 Migrating Full Metadata

Migrate all metadata from source databases to Huawei Cloud DLI.

NOTICE

The system cannot migrate the metadata of transaction tables (those with the **transactional=true** field in DDL statements).

Prerequisites

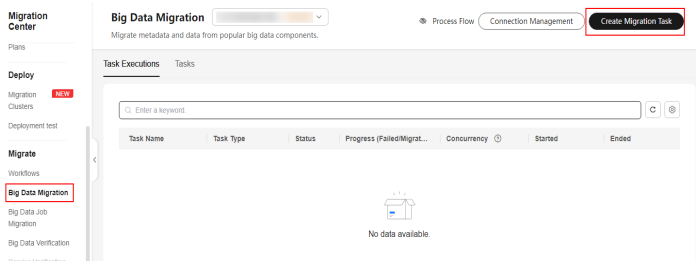
- You have completed all **preparations**.
- **A source connection has been created**.
- **Target connections have been created**.

Procedure

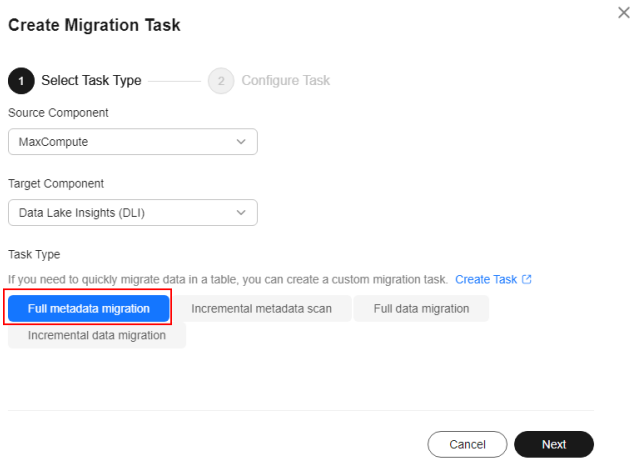
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Migration**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 In the upper right corner of the page, click **Create Migration Task**.



Step 4 Select **MaxCompute** for **Source Component**, **Data Lake Insight (DLI)** for **Target Component**, **Full metadata migration** for **Task Type**, and click **Next**.



Step 5 Set parameters required for creating a full metadata migration task based on **Table 11-4**.

Table 11-4 Parameters required for creating a full metadata migration task

Area	Parameter	Configuration
Basic Settings	Task Name	The default name is **Full-metadata-migration-from-MaxCompute-to-DLI-***4 random characters* (including letters and numbers). You can also customize a name.
	Edge Device	Select the Edge device you connected to MgC in **Preparations**.
Source Settings	Source Connection	Select the **source connection you created**.
	MaxCompute Parameters	The parameters are optional and usually left blank. If needed, you can configure the parameters by referring to **MaxCompute Documentation**.
Migration Scope	By database	Enter the names of databases (projects) to be migrated in the **Include Databases** text box. If there are tables you do not want to migrate, download the template in CSV format, add information about these tables to the template, and upload the template to MgC. For details, see steps **Step 5.2** to **Step 5.5**.
	By table	1. Download the template in CSV format. 2. Open the downloaded CSV template file with Notepad. **CAUTION** Do not use Excel to edit the CSV template file. The template file edited and saved in Excel cannot be identified by MgC. 3. Retain the first line in the CSV template file. From the second line onwards, enter the information about tables to be migrated in the format of {MaxComute project name}, {Table name}. *MaxComute project name* refers to the name of the MaxCompute project to be migrated. *Table name* refers to the data table to be migrated. **NOTICE** – Use commas (,) to separate the MaxCompute project name and the table name in each line. Do not use spaces or other separators. – After adding the information about a table, press **Enter** to start a new line. 4. After all table information is added, save the changes to the CSV file. 5. Upload the edited and saved CSV file to MgC.

Area	Parameter	Configuration
Target Settings	Overwriting Policy	Select the policy for processing tables that already exist at the target. • **Skip**: Such tables will be skipped during the migration, and no operation will be performed on them. • **Rebuild**: Such tables and the data in them will be deleted from the target, and new empty tables will be created at the target based on the source table settings.
	Target Connection	Select the **DLI connection with a SQL queue** you created. **CAUTION** Do not select the connection with a general queue configured.
	Custom Parameters	Configure the parameters as needed. For details, see **Configuration parameter description** and **Custom Parameters**.
	Table Type	Select an option as required.
	File Format	Select an option as required.
	Compression Format	Select an option as required.
Migration Settings	Concurrency	Set the number of concurrent migration subtasks. The default value is **3**. The value ranges from **1** to **10**.

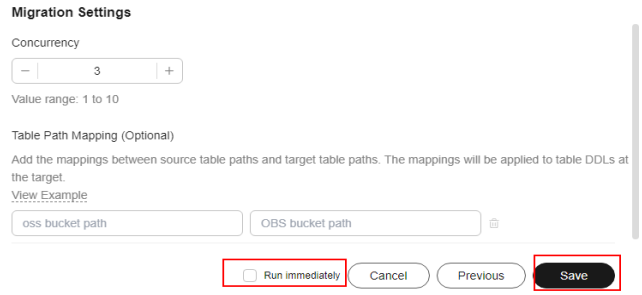
Area	Parameter	Configuration
	(Optional) Table Path Mapping	Map storage paths for metadata of external tables between the source and the target. This can help smoothly move the metadata of MaxCompute external tables from Alibaba OSS to Huawei Cloud OBS. The mapped paths in Huawei Cloud OBS must be created in advance. In the paths, the bucket names can be different, and other parts must be the same. Assume that at the source, external tables are stored in **oss://oss-cn-hangzhou.aliyuncs.com/ali-test/documents/example_table**. In the path, **oss-cn-hangzhou.aliyuncs.com** is the endpoint of the region where the involved bucket resides, **ali-test** is the bucket name, **documents** is the root directory of the bucket, and **example_table** is the directory that stores metadata of external tables (called table directory below). When you configure a mapped target path, the root directory and the table directory must remain unchanged, for example, **obs://hw-test/documents/example_table**. Example: ● Source: **oss://oss-cn-hangzhou.aliyuncs.com/ali-test/documents/example_table** ● Target: **obs://hw-test/documents/example_table** **NOTE** The preceding path mapping is only an example.

Step 6 After the configuration is complete, execute the task.

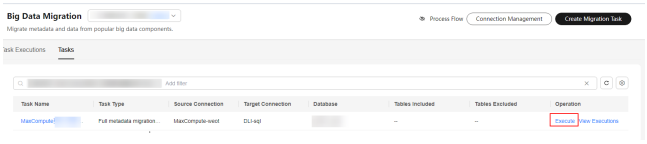
NOTICE

- A migration task can be executed repeatedly. Each time a migration task is executed, a task execution is generated.
- You can click the task name to modify the task configuration.

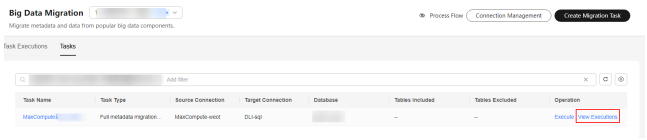
- You can select **Run immediately** and click **Save** to create the task and execute it immediately. You can view the created task on the **Tasks** page.



- You can also click **Save** to just create the task. You can view the created task on the **Tasks** page. To execute the task, click **Execute** in the **Operation** column.



Step 7 After the migration task is executed, click **View Executions** in the **Operation** column. On the **Task Executions** tab, you can view the details of the running task execution and all historical executions.



Click **View** in the **Progress** column. On the displayed page, **view and export the task execution results**.

----End

11.5.2 Scanning for Incremental Metadata

Identify the metadata that has been changed in the source databases after the previous migration is complete, and synchronize the incremental metadata to Huawei Cloud CLI.

Prerequisites

- **A source connection has been created**.
- **Target connections have been created**.
- At least a **full metadata migration** has been completed.

Preparations

- Gain whitelist access to the Spark 3.3.1 feature.

  Contact technical support to whitelist you to use the Spark 3.3.1 feature.
- Configure a DLI job bucket.

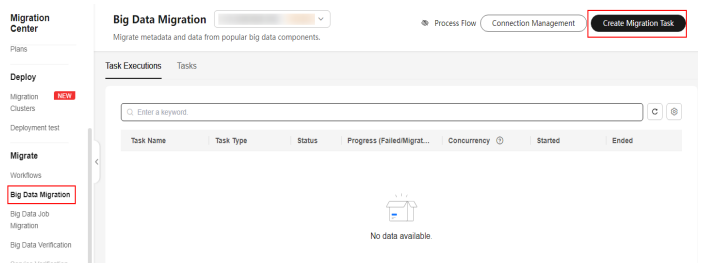
  You need to purchase a bucket or parallel file system on OBS. The bucket is used to store temporary data generated by DLI. For details, see **Configuring a DLI Job Bucket**.

Procedure

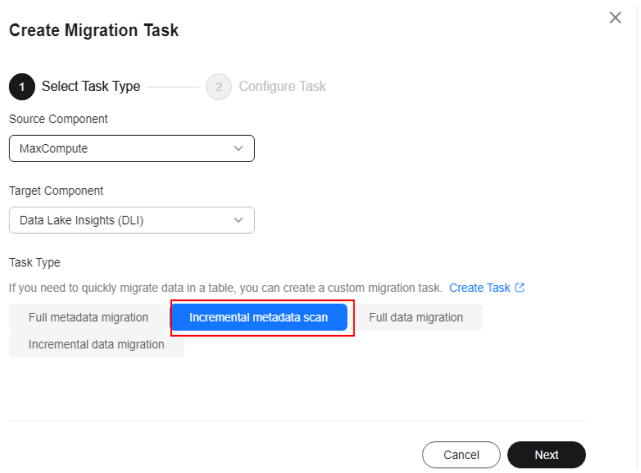
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Migration**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 In the upper right corner of the page, click **Create Migration Task**.



Step 4 Select **MaxCompute** for **Source Component**, **Data Lake Insight (DLI)** for **Target Component**, **Incremental metadata scan** for **Task Type**, and click **Next**.



Step 5 Configure parameters required for creating an incremental metadata scan based on **Table 11-5**.

Table 11-5 Parameters required for creating an incremental metadata scan

Area	Parameter	Configuration
Basic Settings	Task Name	The default name is **Incremental-data-scan-of-MaxCompute-and-DLI-***4 random characters* (including letters and numbers). You can also customize a name.
	Edge Device	Select the Edge device you connected to MgC in **Preparations**.
Source Settings	Source Connection	Select the **source connection you created**.

Area	Parameter	Configuration
	MaxCompute Parameters	The parameters are optional and usually left blank. If needed, you can configure the parameters by referring to **MaxCompute Documentation**.
Migration Scope	By database	Enter the names of databases (projects) to be migrated in the **Include Databases** text box. If there are tables you do not want to migrate, download the template in CSV format, add information about these tables to the template, and upload the template to MgC. For details, see steps **Step 5.2** to **Step 5.5**.
	By table	1. Download the template in CSV format. 2. Open the downloaded CSV template file with Notepad. **CAUTION** Do not use Excel to edit the CSV template file. The template file edited and saved in Excel cannot be identified by MgC. 3. Retain the first line in the CSV template file. From the second line onwards, enter the information about tables to be migrated in the format of {MaxComute project name}, {Table name}. *MaxComute project name* refers to the name of the MaxCompute project to be migrated. *Table name* refers to the data table to be migrated. **NOTICE** – Use commas (,) to separate the MaxCompute project name and the table name in each line. Do not use spaces or other separators. – After adding the information about a table, press **Enter** to start a new line. 4. After all table information is added, save the changes to the CSV file. 5. Upload the edited and saved CSV file to MgC.
Target Settings	Target Connection	Select the DLI connection with a SQL queue created in **Creating a Target Connection**. **CAUTION** Do not select the connection with a general queue configured.
	Custom Parameters	Configure the parameters as needed. For details, see **Configuration parameter description** and **Custom Parameters**.

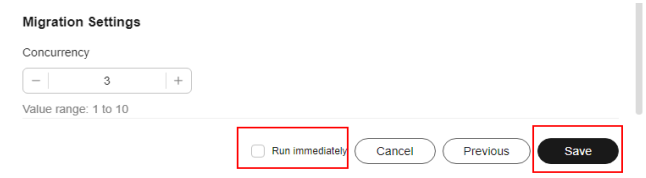
Area	Parameter	Configuration
Migration Settings	Concurrency	Set the number of concurrent migration subtasks. The default value is **3**. The value ranges from **1** to **10**.

Step 6 After the configuration is complete, execute the task.

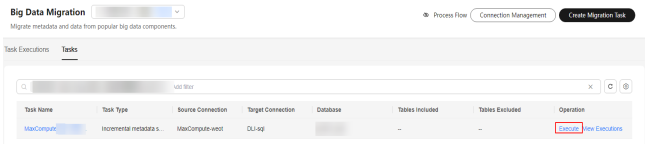
NOTICE

- A migration task can be executed repeatedly. Each time a migration task is executed, a task execution is generated.
- You can click the task name to modify the task configuration.

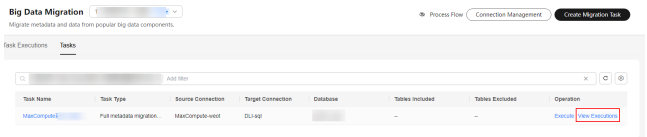
- You can select **Run immediately** and click **Save** to create the task and execute it immediately. You can view the created task on the **Tasks** page.



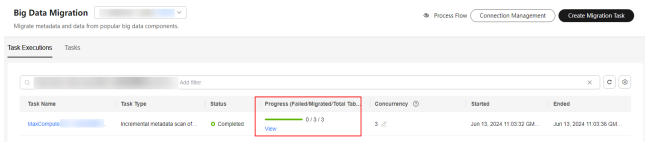
- You can also click **Save** to just create the task. You can view the created task on the **Tasks** page. To execute the task, click **Execute** in the **Operation** column.

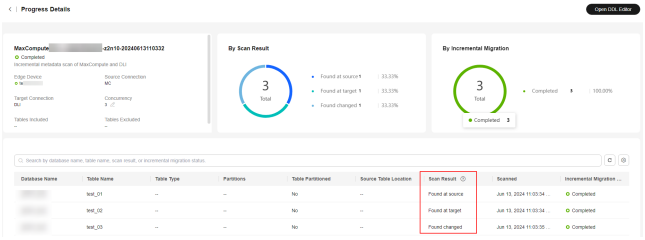


Step 7 After the migration task is executed, click **View Executions** in the **Operation** column. On the **Task Executions** tab, you can view the details of the running task execution and all historical executions.

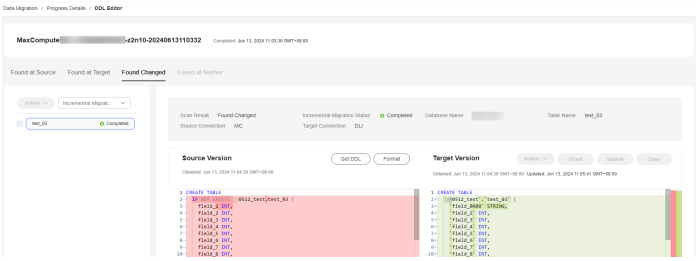


Click **View** in the **Progress** column. On the displayed **Progress Details** page, **view and export the incremental metadata scan results**.





Step 8 In the upper right corner of the progress details page, click **Open DDL Editor** to compare and edit the structures of incremental tables.



----End

11.5.3 Migrating Full Data

Migrate all data from source databases to Huawei Cloud DLI.

Prerequisites

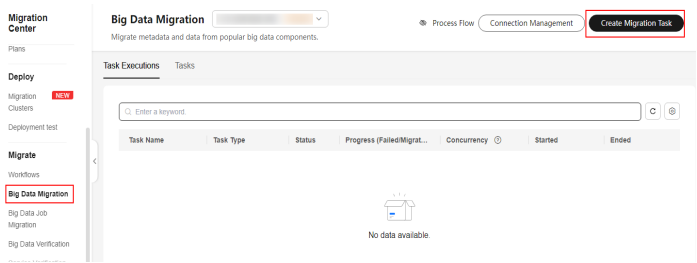
- You have completed all **preparations**.
- **A source connection has been created**.
- **Target connections have been created**.
- You have been added to the whitelist that allows JAR programs to access DLI metadata. If you have not, contact technical support.

Procedure

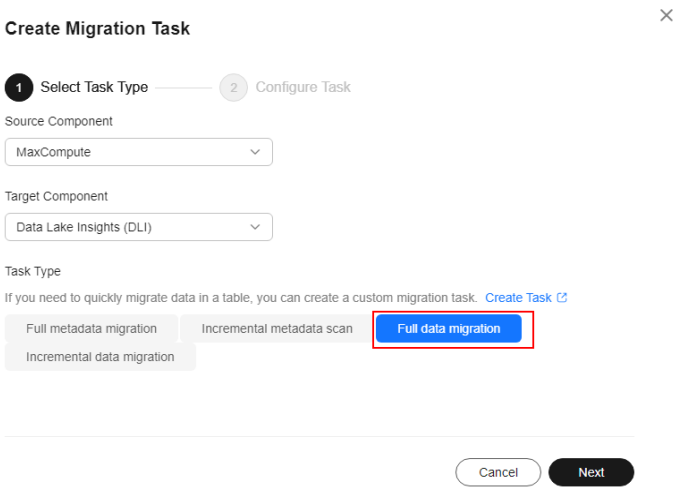
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Migration**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 In the upper right corner of the page, click **Create Migration Task**.



Step 4 Select **MaxCompute** for **Source Component**, **Data Lake Insight (DLI)** for **Target Component**, **Full data migration** for **Task Type**, and click **Next**.

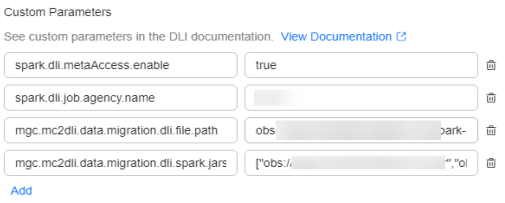
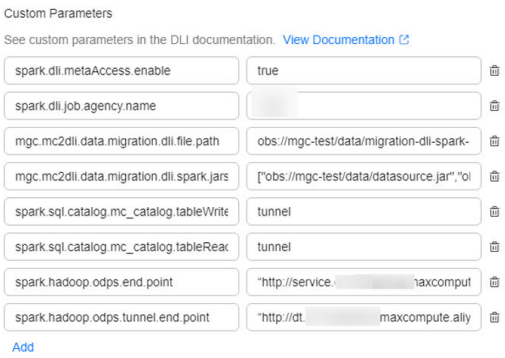


Step 5 Configure parameters required for creating a full data migration task based on **Table 11-6**.

Table 11-6 Parameters required for creating a full data migration task

Area	Parameter	Configuration
Basic Settings	Task Name	The default name is **Full-data-migration-from-MaxCompute-to-DLI-***4 random characters* (including letters and numbers). You can also customize a name.
	Edge Device	Select the Edge device you connected to MgC in **Making Preparations**.
Source Settings	Source Connection	Select the **source connection you created**.
	Estimated Project Period (Day) (Optional)	If this parameter is set, the system checks table lifecycle during the migration. If the lifecycle of a table ends before the expected end time of the project, the table will be skipped. If this parameter is not set, all tables are migrated by default.
	MaxCompute Parameters	The parameters are optional and usually left blank. If needed, you can configure the parameters by referring to **MaxCompute Documentation**.
Migration Scope	By database	Enter the names of databases (projects) to be migrated in the **Include Databases** text box. If there are tables you do not want to migrate, download the template in CSV format, add information about these tables to the template, and upload the template to MgC. For details, see steps **Step 5.2** to **Step 5.5**.

Area	Parameter	Configuration
	By table	1. Download the template in CSV format. 2. Open the downloaded CSV template file with Notepad. **CAUTION** Do not use Excel to edit the CSV template file. The template file edited and saved in Excel cannot be identified by MgC. 3. Retain the first line in the CSV template file. From the second line onwards, enter the information about tables to be migrated in the format of {MaxComute project name}, {Table name}. *MaxComute project name* refers to the name of the MaxCompute project to be migrated. *Table name* refers to the data table to be migrated. **NOTICE** – Use commas (,) to separate the MaxCompute project name and the table name in each line. Do not use spaces or other separators. – After adding the information about a table, press **Enter** to start a new line. 4. After all table information is added, save the changes to the CSV file. 5. Upload the edited and saved CSV file to MgC.
Target Settings	Target Connection	Select the DLI connection with a general queue created in **Creating a Target Connection**. **CAUTION** Do not a DLI connection with a SQL queue configured.

Area	Parameter	Configuration
	Custom Parameters	Configure the parameters as needed. For details, see **Configuration parameter description** and **Custom Parameters**. • If the migration is performed over the Internet, set the following four parameters:  – **spark.dli.metaAccess.enable**: Enter **true**. – **spark.dli.job.agency.name**: Enter the name of the **DLI agency you configured**. – **mgc.mc2dli.data.migration.dli.file.path**: Enter the **OBS path for storing the migration-dli-spark-1.0.0.jar package**. For example, obs://mgc-test/data/migration-dli-spark-1.0.0.jar – **mgc.mc2dli.data.migration.dli.spark.jars**: Enter the **OBS path for storing the fastjson-1.2.54.jar and datasource.jar packages**. The value is transferred in array format. Package names must be enclosed using double quotation marks and be separated with commas (,) For example: ["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] • If the migration is performed over a private network, set the following four parameters:  – **spark.dli.metaAccess.enable**: Enter **true**. – **spark.dli.job.agency.name**: Enter the name of the **DLI agency you configured**.

Area	Parameter	Configuration
		– **mgc.mc2dli.data.migration.dli.file.path**: Enter the **OBS path for storing the migration-dli-spark-1.0.0.jar package**. For example, obs://mgc-test/data/migration-dli-spark-1.0.0.jar – **mgc.mc2dli.data.migration.dli.spark.jars**: Enter the **OBS path for storing the fastjson-1.2.54.jar and datasource.jar packages**. The value is transferred in array format. Package names must be enclosed using double quotation marks and be separated with commas (,) For example: ["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] – **spark.sql.catalog.mc_catalog.tableWriteProvider**: Enter **tunnel**. – **spark.sql.catalog.mc_catalog.tableReadProvider**: Enter **tunnel**. – **spark.hadoop.odps.end.point**: Enter the VPC endpoint of the region where the source MaxCompute service is provisioned. For details about the MaxCompute VPC endpoint in each region, see **Endpoints in different regions (VPC)**. For example, if the source MaxCompute service is located in Hong Kong, China, enter **http://service.cn-hongkong.maxcompute.aliyun-inc.com/api**. – **spark.hadoop.odps.tunnel.end.point**: Enter the VPC Tunnel endpoint of the region where the source MaxCompute service is located. For details about the MaxCompute VPC Tunnel endpoint in each region, see **Endpoints in different regions (VPC)**. For example, if the source MaxCompute service is located in Hong Kong, China, enter **http://dt.cn-hongkong.maxcompute.aliyun-inc.com**.
Migration Settings	Large Table Migration Rules	Control how large a table will be split into multiple migration subtasks. You are advised to retain the default settings. You can also change the settings as needed.

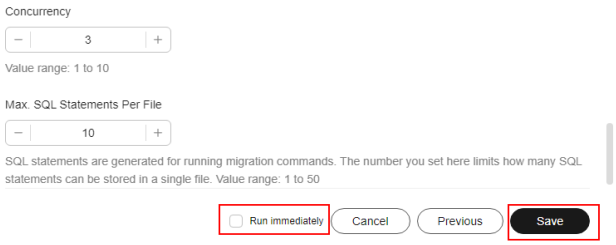
Area	Parameter	Configuration
	Small Table Migration Rules	Control how small a table will be merged into one migration subtask along with other small tables. This can accelerate your migration. You are advised to retain the default settings. You can also change the settings as needed.
	Concurrency	Set the number of concurrent migration subtasks. The default value is **3**. The value ranges from **1** to **10**.
	Max. SQL Statements Per File	SQL statements are generated for running migration commands. The number you set here limits how many SQL statements can be stored in a single file. The default value is **3**. The value ranges from **1** to **50**.

Step 6 After the configuration is complete, execute the task.

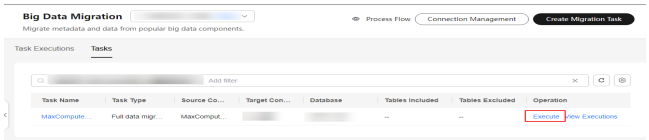
NOTICE

- A migration task can be executed repeatedly. Each time a migration task is executed, a task execution is generated.
- You can click the task name to modify the task configuration.

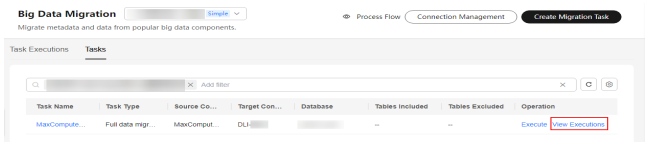
- You can select **Run immediately** and click **Save** to create the task and execute it immediately. You can view the created task on the **Tasks** page.



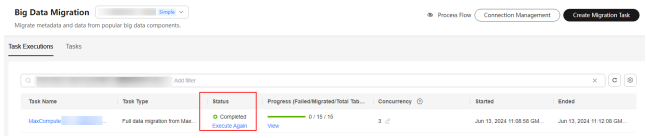
- You can also click **Save** to just create the task. You can view the created task on the **Tasks** page. To execute the task, click **Execute** in the **Operation** column.



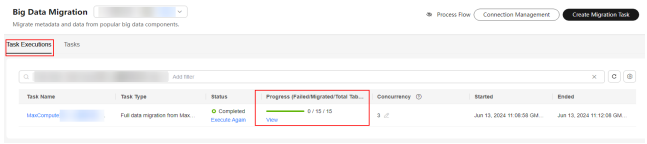
Step 7 After the migration task is executed, click **View Executions** in the **Operation** column. On the **Task Executions** tab, you can view the details of the running task execution and all historical executions.

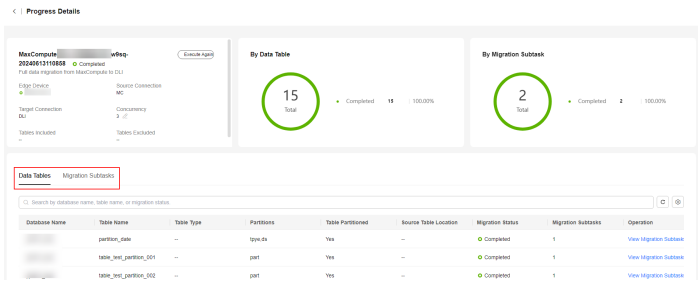


Click **Execute Again** in the **Status** column to run the execution again.



Click **View** in the **Progress** column. On the displayed **Progress Details** page, **view and export the task execution results**.





Step 8 (Optional) After the data migration is complete, verify data consistency between the source and the target databases. For details, see **Verifying the Consistency of Data Migrated from MaxCompute to DLI**.

----End

11.5.4 Migrating Incremental Data

Synchronize data that has been added, modified, or deleted after the previous migration is complete from source databases to Huawei Cloud DLI.

Prerequisites

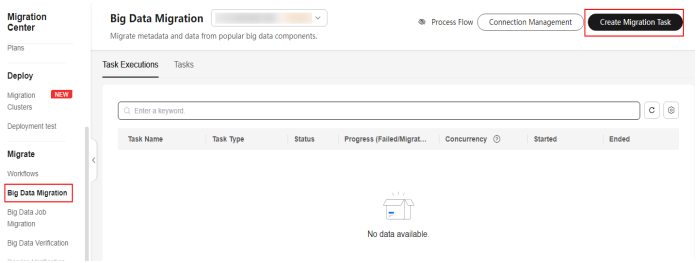
- You have completed all **preparations**.
- **A source connection has been created**.
- **Target connections have been created**.
- At least one **full data migration** has been completed.
- You have been added to the whitelist that allows JAR programs to access DLI metadata. If you have not, contact technical support.

Procedure

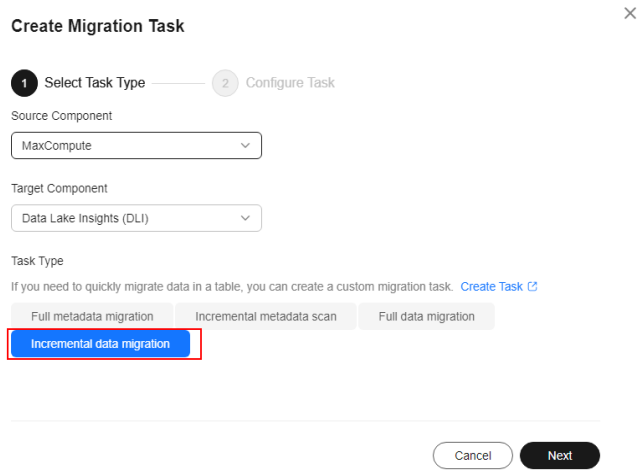
Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Migration**. In the upper left corner of the page, select the migration project created in **Preparations**.

Step 3 In the upper right corner of the page, click **Create Migration Task**.



Step 4 Select **MaxCompute** for **Source Component**, **Data Lake Insight (DLI)** for **Target Component**, **Incremental data migration** for **Task Type**, and click **Next**.



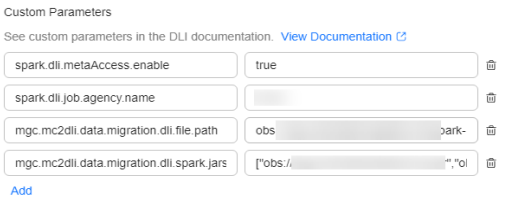
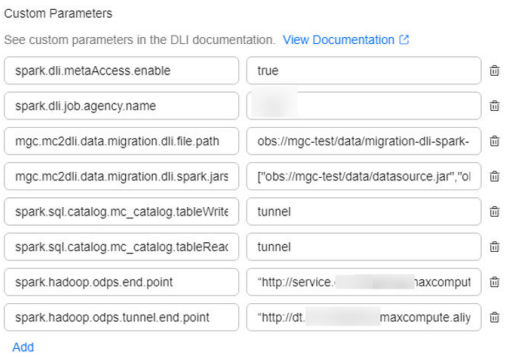
Step 5 Configure parameters required for creating an incremental data migration task based on **Table 11-7**.

Table 11-7 Parameters required for creating an incremental data migration task

Area	Parameter	Configuration
Basic Settings	Task Name	The default name is **Incremental-data-migration-from-MaxCompute-to-DLI-***4 random characters* (including letters and numbers). You can also customize a name.
	Edge Device	Select the Edge device you connected to MgC in **Preparations**.
Source Settings	Source Connection	Select the **source connection you created**.

Area	Parameter	Configuration
	Incremental Scope	Select a T-*N* option to limit the migration to the incremental data generated within a specific time period (24**N* hours). **T** refers to the time when the task is executed. Assume that you select **T-1** and the task is executed at 14:50 on June 6, 2024. The system migrates incremental data generated from 14:50 on June 5, 2024 to 14:50 on June 6, 2024. If you select **Specified Date**, only incremental data generated on the specified date is migrated.
	Filter Partitions	Decide whether to filter partitions to be migrated by update time or creation time. The default value is **By update time**. Select **By update time** to migrate the data that was changed in the specified period. Select **By creation time** to migrate the data that was created in a specified period.
	MaxCompute Parameters	The parameters are optional and usually left blank. If needed, you can configure the parameters by referring to **MaxCompute Documentation**.
Migration Scope	By database	Enter the names of databases (projects) to be migrated in the **Include Databases** text box. If there are tables you do not want to migrate, download the template in CSV format, add information about these tables to the template, and upload the template to MgC. For details, see steps **Step 5.2** to **Step 5.5**.

Area	Parameter	Configuration
	By table	1. Download the template in CSV format. 2. Open the downloaded CSV template file with Notepad. **CAUTION** Do not use Excel to edit the CSV template file. The template file edited and saved in Excel cannot be identified by MgC. 3. Retain the first line in the CSV template file. From the second line onwards, enter the information about tables to be migrated in the format of {MaxComute project name}, {Table name}. *MaxComute project name* refers to the name of the MaxCompute project to be migrated. *Table name* refers to the data table to be migrated. **NOTICE** – Use commas (,) to separate the MaxCompute project name and the table name in each line. Do not use spaces or other separators. – After adding the information about a table, press **Enter** to start a new line. 4. After all table information is added, save the changes to the CSV file. 5. Upload the edited and saved CSV file to MgC.
Target Settings	Target Connection	Select the DLI connection with a general queue created in **Creating a Target Connection**. **CAUTION** Do not a DLI connection with a SQL queue configured.

Area	Parameter	Configuration
	Custom Parameters	Configure the parameters as needed. For details, see **Configuration parameter description** and **Custom Parameters**. • If the migration is performed over the Internet, set the following four parameters:  – **spark.dli.metaAccess.enable**: Enter **true**. – **spark.dli.job.agency.name**: Enter the name of the **DLI agency you configured**. – **mgc.mc2dli.data.migration.dli.file.path**: Enter the **OBS path for storing the migration-dli-spark-1.0.0.jar package**. For example, obs://mgc-test/data/migration-dli-spark-1.0.0.jar – **mgc.mc2dli.data.migration.dli.spark.jars**: Enter the **OBS path for storing the fastjson-1.2.54.jar and datasource.jar packages**. The value is transferred in array format. Package names must be enclosed using double quotation marks and be separated with commas (,) For example: ["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] • If the migration is performed over a private network, set the following four parameters:  – **spark.dli.metaAccess.enable**: Enter **true**. – **spark.dli.job.agency.name**: Enter the name of the **DLI agency you configured**.

Area	Parameter	Configuration
		– **mgc.mc2dli.data.migration.dli.file.path**: Enter the **OBS path for storing the migration-dli-spark-1.0.0.jar package**. For example, obs://mgc-test/data/migration-dli-spark-1.0.0.jar – **mgc.mc2dli.data.migration.dli.spark.jars**: Enter the **OBS path for storing the fastjson-1.2.54.jar and datasource.jar packages**. The value is transferred in array format. Package names must be enclosed using double quotation marks and be separated with commas (,) For example, ["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"] – **spark.sql.catalog.mc_catalog.tableWriteProvider**: Enter **tunnel**. – **spark.sql.catalog.mc_catalog.tableReadProvider**: Enter **tunnel**. – **spark.hadoop.odps.end.point**: Enter the VPC endpoint of the region where the source MaxCompute service is provisioned. For details about the MaxCompute VPC endpoint in each region, see **Endpoints in different regions (VPC)**. For example, if the source MaxCompute service is located in Hong Kong, China, enter **http://service.cn-hongkong.maxcompute.aliyun-inc.com/api**. – **spark.hadoop.odps.tunnel.end.point**: Enter the VPC Tunnel endpoint of the region where the source MaxCompute service is located. For details about the MaxCompute VPC Tunnel endpoint in each region, see **Endpoints in different regions (VPC)**. For example, if the source MaxCompute service is located in Hong Kong, China, enter **http://dt.cn-hongkong.maxcompute.aliyun-inc.com**.
Migration Settings	Large Table Migration Rules	Control how large a table will be split into multiple migration subtasks. You are advised to retain the default settings. You can also change the settings as needed.

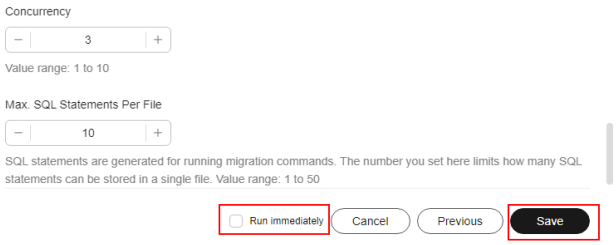
Area	Parameter	Configuration
	Small Table Migration Rules	Control how small a table will be merged into one migration subtask along with other small tables. This can accelerate your migration. You are advised to retain the default settings. You can also change the settings as needed.
	Concurrency	Set the number of concurrent migration subtasks. The default value is **3**. The value ranges from **1** to **10**.
	Max. SQL Statements Per File	SQL statements are generated for running migration commands. The number you set here limits how many SQL statements can be stored in a single file. The default value is **3**. The value ranges from **1** to **50**.

Step 6 After the configuration is complete, execute the task.

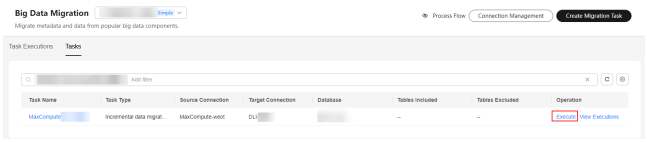
NOTICE

- A migration task can be executed repeatedly. Each time a migration task is executed, a task execution is generated.
- You can click the task name to modify the task configuration.

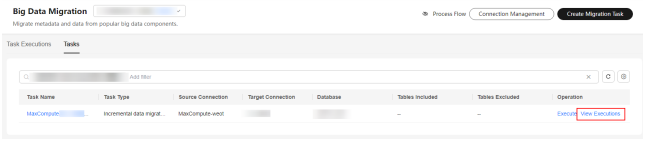
- You can select **Run immediately** and click **Save** to create the task and execute it immediately. You can view the created task on the **Tasks** page.



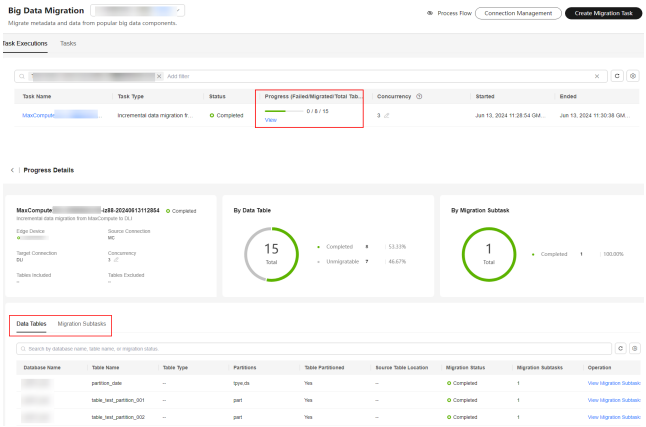
- You can also click **Save** to just create the task. You can view the created task on the **Tasks** page. To execute the task, click **Execute** in the **Operation** column.



Step 7 After the migration task is executed, click **View Executions** in the **Operation** column. On the **Task Executions** tab, you can view the details of the running task execution and all historical executions.



Locate the running execution and click **View** in the **Progress** column. On the displayed **Progress Details** page, **view and export the task execution results**.



Step 8 (Optional) After the data migration is complete, verify data consistency between the source and the target databases. For details, see **Verifying the Consistency of Data Migrated from MaxCompute to DLI**.

----End

11.6 Viewing and Exporting Task Execution Results

Each time a migration task is executed, a task execution is generated. On the task execution details page, you can view and export the task execution results.

NOTICE

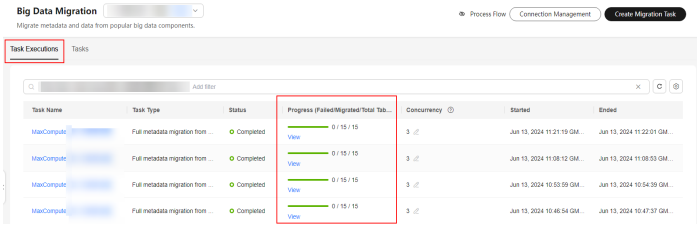
The exported files can be downloaded within 15 days of when they are generated but will be automatically deleted after that.

Prerequisites

A migration task has been created and executed.

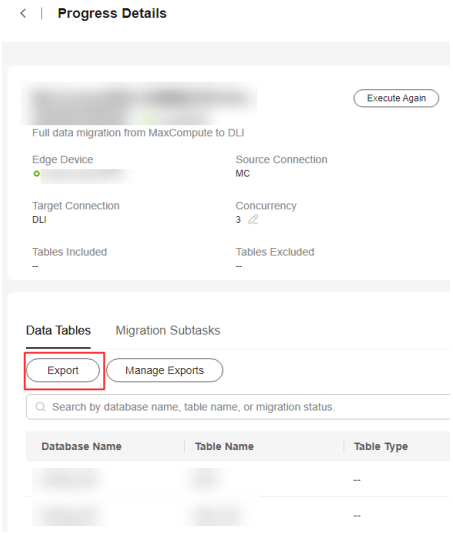
Procedure

Step 1 In the **Task Executions** page, locate a task execution and click **View** in the **Progress** column.



Step 2 Under **Tables**, click **Export**. The system starts to export the task execution results. You can filter the task execution results and export only the filtered results.

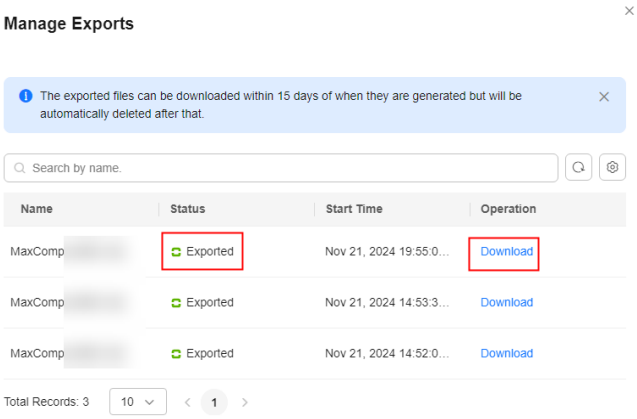
Figure 11-1 Exporting execution results



Step 3 After the export is complete, click **Manage Exports**.

Step 4 After the export is complete, click **Download** in the **Operation** column to download the exported results.

Figure 11-2 Downloading exported results



----End

11.7 Custom Parameters

For details about the custom parameters that can be configured for the target, see **Table 11-8** and **configuration parameter description**.

Table 11-8 Supported custom parameters

Parameter	Value Range	Description	Mandatory
spark.dli.metaAccess.enable	-	Enter **true**.	Yes
spark.dli.job.agency.name		Enter the **DLI agency name**.	
mgc.mc2dli.data.migration.dli.file.path		Enter the **OBS path for storing the migration-dli-spark-1.0.0.jar package**. For example, obs://mgc-test/data/migration-dli-spark-1.0.0.jar	
mgc.mc2dli.data.migration.dli.spark.jars		Enter the **OBS path for storing the fastjson-1.2.54.jar and datasource.jar packages**. The value is transferred in array format. Package names must be enclosed using double quotation marks and be separated with commas (,) For example, ["obs://mgc-test/data/datasource.jar","obs://mgc-test/data/fastjson-1.2.54.jar"]	
spark.sql.catalog.mc_catalog.tableWriteProvider		Enter **tunnel**.	Mandatory for migration over private lines
spark.sql.catalog.mc_catalog.tableReadProvider		Enter **tunnel**.	
spark.hadoop.odps.end.point		Enter the VPC endpoint of the region where the source MaxCompute service is provisioned. For details about the MaxCompute VPC endpoint in each region, see **Endpoints in different regions (VPC)**. For example, if the source MaxCompute service is located in Hong Kong, China, enter **http://service.cn-hongkong.maxcompute.aliyun-inc.com/api**.	

Parameter	Value Range	Description	Mandatory
spark.hadoop.odps.tunnel.endpoint		Enter the VPC Tunnel endpoint of the region where the source MaxCompute service is located. For details about the MaxCompute VPC Tunnel endpoint in each region, see **Endpoints in different regions (VPC)**. For example, if the source MaxCompute service is located in Hong Kong, China, enter **http://dt.cn-hongkong.maxcompute.aliyun-inc.com**.	
sc_type	A, B, and C	Compute resource type. Currently, resource types A, B, and C are available. If this parameter is not specified, the minimum configuration (type A) is used. • A: (vCPUs: 8; memory: 32 GB; driverCores: 2; executorCores: 1; driverMemory: 7 GB; executorMemory: 4 GB; and numExecutor: 6) • B: (vCPUs: 16; memory: 64 GB; driverCores: 2; executorCores: 2; driverMemory: 7 GB; executorMemory: 8 GB; and numExecutor: 7) • C: (vCPUs: 32; memory: 128 GB; driverCores: 4; executorCores: 2; driverMemory: 15 GB; executorMemory: 8 GB; and numExecutor: 14)	No
executorCores	1- 4	Number of CPU cores of each Executor in the Spark application. This configuration will replace the default setting in **sc_type**.	
numExecutors	1-100	Number of Executors in a Spark application. This configuration will replace the default setting in **sc_type**.	
executorMemory	1-16 G	Executor memory of the Spark application, for example, **2 GB** or **2048 MB**. This configuration will replace the default setting in **sc_type**. The unit must be provided. Otherwise, the startup fails.	
driverCores	1-4	Number of CPU cores of the Spark application driver. This configuration will replace the default setting in **sc_type**.	

Parameter	Value Range	Description	Mandatory
driverMemory	1-16 G	Driver memory of the Spark application, for example, **2 GB** or **2048 MB**. This configuration will replace the default setting in **sc_type**. The unit must be provided. Otherwise, the startup fails.	

12 Big Data Verification

12.1 Overview

MgC allows you to verify the consistency of data migrated from various big data computing and storage engines, such as Hive, HBase, Doris, and MaxCompute. Consistency verification ensures data accuracy and reliability and enables to you migrate big data to Huawei Cloud with confidence.

Verification Methods

- **Full verification**: The consistency of all inventory data is verified.
- **Daily incremental verification**: The consistency of incremental data is verified based on the creation or update time. You can choose to verify incremental data for one day or several consecutive days.
- **Hourly incremental verification**: Data consistency is verified based on the creation time or update time multiple times within 24 hours. The verification automatically stops at 00:00 on the next day.
- **Date-based verification**: This method applies only to tables partitioned by date in the year, month, and day format. You can choose to verify consistency of such tables for one day or several consecutive days. Tables that are not partitioned by date are not verified.
- **Selective verification**: This method can be used to verify the consistency of the data within a specified time period. You can only select a period going backward from the current time for verification.

Supported Source and Target Components

Source Component	Target Component
• Hive • HBase • Doris • MaxCompute • ClickHouse • Delta Lake • Hudi	• Hive • DLI • MRS (Doris) • MRS (Hbase) • MRS (ClickHouse) • CloudTable (ClickHouse) • CloudTable (HBase) • Delta • Hudi

Verification Methods Available for Each Component

Component	Verification Method
Hive	• Full verification • Daily incremental verification • Hourly incremental verification • Date-based verification
DLI	
MaxCompute	• Full verification • Daily incremental verification • Hourly incremental verification • Date-based verification
Doris	• Full verification • Daily incremental verification • Hourly incremental verification
HBase	• Full verification • Selective verification
ClickHouse	Full verification
ApsaraDB for ClickHouse	Full verification
CloudTable (HBase)	• Full verification • Selective verification
CloudTable (ClickHouse)	Full verification

Component	Verification Method
Delta	• Full verification • Daily incremental verification • Hourly incremental verification • Date-based verification
Hudi	• Full verification • Daily incremental verification • Hourly incremental verification • Date-based verification

12.2 Preparations

Before using big data verification, complete the following preparations:

Creating a Migration Project

Migration projects allow you to group, isolate, and manage migration resources. For details, see **Managing Migration Projects**.

Preparing Edge

To verify big data consistency, you need to install Edge for Linux in your source environment. For details, see **Installing Edge for Linux**. After Edge is installed, connect Edge to MgC. For details, see **Connecting the Edge Device to MgC**.

CAUTION

Edge does not support automatic restart. Do not restart Edge during task execution, or tasks will fail.

Enabling Big Data Verification

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. In the upper left corner of the page, choose the **migration project you created** from the project drop-down list

Step 3 Click **Select Edge Device**.

Step 4 In the displayed dialog box, select the Edge device you connected to MgC from the drop-down list.

- If an Edge device is **Online** and **Enabled**, the device has been used for big data verification in a project. You can click **Confirm** to continue using the tool.

- If an Edge device is **Online** and **Disabled**, the device has not been used for big data verification in any project, and you can click **Enable** to use this device for big data verification.
- If an Edge device is **Offline** and **Deleted**, the device has been deleted, and it cannot be used for big data verification. However, if this device has been used for big data verification before being deleted, you can select it and click **View** to view the verification tasks and results.
- If an Edge device is **Offline** and **Disabled**, the device cannot be used for big data verification, and all operation buttons are disabled even if you select it and click **View**.

> ⚠ CAUTION
>
> Before performing big data verification, ensure that the selected Edge device is **Online** and **Enabled**.

----End

12.3 Creating a Connection to the Source

12.3.1 Creating a Connection to a Source Cloud Service

To verify the consistency of data stored using big data cloud services, you need to establish connections between MgC and the source and target cloud services.

The supported source big data cloud services include:

- MaxCompute
- ApsaraDB for ClickHouse

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Connection Management**.

Step 4 Click **Create Connection** in the upper right corner of the page.

Step 5 Select the big data cloud service you use and click **Next**.

Step 6 Configure the connection parameters based on the big data cloud service you selected.

- **Parameters for creating a connection to MaxCompute**
- **Parameters for creating a connection to ApsaraDB for ClickHouse**

Table 12-1 Parameters for creating a connection to MaxCompute

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **MaxCompute-***4 random characters* (including letters and numbers). You can also customize a name.
Alibaba Cloud Credential	Select the credential of the source Alibaba Cloud account. For details about how to add credentials, see "Big data - MaxCompute" in **Adding Resource Credentials**.
MaxCompute Project	Enter the name of your MaxCompute project. You can obtain the project name from the MaxCompute console.
Endpoint	Enter the endpoint of the region where the MaxCompute project is located. For details about MaxCompute endpoints in different regions, see **MaxCompute Endpoints**.

Table 12-2 Parameters for creating a connection to ApsaraDB for ClickHouse

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **ApsaraDB for ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
ClickHouse Credential	Select the ApsaraDB for ClickHouse credential added to Edge. For details about how to add credentials, see "Big data - ClickHouse" in **Adding Resource Credentials**.
Database URL	Enter the public address of the source ClickHouse cluster. You can view the IP address in the cluster details.

Step 7 Click **Test**. MgC verifies whether the cloud service can be connected using the information you provided. If the test is successful, the connection can be set up.

Step 8 After the connection test is successful, click **Confirm**. The cloud service connection is created.

Step 9 On the **Connection Management** page, view the created connection and its basic information. In the **Operation** column, click **Modify** to modify the connection settings.

----End

12.3.2 Creating a Connection to a Source Component

To verify the consistency of data stored using big data components, you need to establish connections between MgC and the big data components.

The supported source big data components include:

- Doris
- HBase
- ClickHouse
- Hive Metastore
- Delta Lake (with metadata)
- Delta Lake (without metadata)
- Hudi (with metadata)
- Hudi (without metadata)

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Connection Management**.

Step 4 Click **Create Connection** in the upper right corner of the page.

Step 5 Select a big data component and click **Next**.

Step 6 Set parameters based on the big data component you selected.

- **Parameters for creating a connection to Doris**
- **Parameters for creating a connection to HBase**
- **Parameters for creating a connection to ClickHouse**
- **Parameters for creating a connection to Hive Metastore**
- **Parameters for creating a connection to Delta Lake (with metadata)**
- **Parameters for creating a connection to Delta Lake (without metadata)**
- **Parameters for creating a connection to Hudi (with metadata)**
- **Parameters for creating a connection to Hudi (without metadata)**

Table 12-3 Parameters for creating a connection to Doris

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Doris-***4 random characters* (including letters and numbers). You can also customize a name.
Doris Credential	Select the source Doris credential added to the Edge device. For details about how to add credentials, see "Big data - Doris" in **Adding Resource Credentials**.

Parameter	Configuration
Database IP Address	Enter the IP address for accessing the source Doris cluster.
Database Port	Enter the port used for accessing the source Doris cluster. The default value is 3306.
Database Name	Enter the name of the source Doris database.

Table 12-4 Parameters for creating a connection to HBase

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **HBase-***4 random characters* (including letters and numbers). You can also customize a name.
HBase Credential	Select the source HBase credential added to the Edge device. For details about how to add credentials, see "Big data - HBase" in **Adding Resource Credentials**.
Secured Cluster	Choose whether the cluster is secured.
ZooKeeper IP Address	Enter the IP address for connecting to the source ZooKeeper node. You can enter the public or private IP address of the ZooKeeper node.
ZooKeeper Port	Enter the port for connecting to the source ZooKeeper node.
HBase Version	Select the source HBase version.

Table 12-5 Parameters for creating a connection to ClickHouse

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
ClickHouse Credential (Optional)	Select the source ClickHouse credential added to the Edge device. For details about how to add credentials, see "Big data - ClickHouse" in **Adding Resource Credentials**.
Secured Cluster	Choose whether the cluster is secured.

Parameter	Configuration
ClickHouse Server IP Address	Enter the IP address for accessing the source ClickHouse server. Generally, the IP address refers to that of the server where ClickHouse is hosted.
HTTP Port	If the source ClickHouse cluster is unsecured, enter the HTTP port for communicating with the ClickHouse server. The default value is 8123.
HTTP SSL/TLS Port	If the source ClickHouse cluster is secured, enter the HTTPS port for communicating with the ClickHouse server.

Table 12-6 Parameters for creating a connection to Hive Metastore

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Hive-Metastore-***4 random characters* (including letters and numbers). You can also customize a name.
Secure Connection	Choose whether to enable secure connection. • If Hive Metastore is deployed in an unsecured cluster, do not enable secure connection. • If Hive Metastore is deployed in a secured cluster, enable secure connection and provide access credentials. For details about how to obtain and add credentials to MgC, see "Big data - Hive Metastore" in **Adding Resource Credentials**.
Hive Version	Select the source Hive version. **CAUTION** If the source Hive version is 2.1.1, select **1.x**.
Hive Metastore IP Address	Enter the IP address for connecting to the Hive Metastore node.
Hive Metastore Thrift Port	Enter the port for connecting to the Hive Metastore Thrift service. The default port is 9083.

Table 12-7 Parameters for creating a connection to Delta Lake (with metadata)

Parameter	Configuration
Connection To	Select **Source**.

Parameter	Configuration
Connection Name	The default name is **Delta-Lake-with-metadata-*4 random characters*** (including letters and numbers). You can also customize a name.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 12-8 Parameters for creating a connection to Delta Lake (without metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Delta-Lake-without-metadata-*4 random characters*** (including letters and numbers). You can also customize a name.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.

Parameter	Configuration
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 12-9 Parameters for creating a connection to Hudi (with metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Hudi-with-metadata-***4 random characters* (including letters and numbers). You can also customize a name.
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Table 12-10 Parameters for creating a connection to Hudi (without metadata)

Parameter	Configuration
Connection To	Select **Source**.
Connection Name	The default name is **Hudi-without-metadata-***4 random characters* (including letters and numbers). You can also customize a name.

Parameter	Configuration
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor.
Spark Client Directory	Enter the installation directory of the Spark client.
Environment Variable Address	Enter the absolute path of the environment variable file (configuration file), for example, **/opt/bigdata/client/bigdata_env**.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the directory. **NOTICE** After the migration is complete, you need to manually clear the folders generated at this location to release storage space.

Step 7 Click **Test**. MgC verifies whether the component can be connected using the information you provided. If the test is successful, the connection can be set up.

Step 8 After the connection test is successful, click **Confirm**. The connection is created.

Step 9 On the **Connection Management** page, view the created connection and its basic information. In the **Operation** column, click **Modify** to modify the connection settings.

----End

12.4 Creating a Table Group and Adding Tables to the Group

Group source tables before verifying their consistency.

NOTICE

- A maximum of 10,000 tables can be imported at a time.
- The tables to be imported must come from the same metadata source.
- When a table is imported, the system does not verify how many table groups the table is added to or what verification rules are configured for those groups. You are advised to import a table into a maximum of three table groups with different verification rules.

Prerequisites

- You have completed all **preparations**.
- **A source connection has been created**.

Creating a Table Group

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Table Management**.

Step 4 Under **Table Groups**, click **Create**.

Step 5 Configure the parameters listed in **Table 12-11**.

Table 12-11 Parameters for creating a table group

Parameter	Description
Table Group	Enter a name.
Metadata Source	Select the **created source connection**. **CAUTION** A table group can only contain tables coming from the same metadata source.
Verification Rule	Select the method for verifying data consistency and the inconsistency tolerance. MgC provides multiple verification rules for you to choose. For details about these rules, click **View More**.
Description (Optional)	Enter a description to identify the table group.

Step 6 Click **Confirm**. The table group is created. In the table group list, you can view information about the created table group.

After the table group is created, you can add tables to the table group by referring to **Adding Tables to a Table Group**.

----End

Adding Tables to a Table Group

Step 1 On the **Table Management** page, click the **Tables** tab.

Step 2 Choose **Table Management** > **Import** above the list.

Step 3 Select a metadata connection and the table groups that tables are added to.

Step 4 Click **Download** to download the import template to the local PC. Open the import template and fill in the information of the tables to be added.

NOTICE

- A maximum of 10,000 tables can be imported at a time.
- Tables in a table group must come from the same metadata source.
- Cells in the template must not contain formulas and must be in text format. Otherwise, the table parsing will fail.
- If a Delta Lake (with metadata) or Hudi (with metadata) connection is selected for **Metadata Connection**, the **source_path** parameter in the template is mandatory.
- If a Delta Lake (without metadata) or Hudi (without metadata) connection is selected for **Metadata Connection**, the **source_path** and **target_path** parameters in the template are mandatory.

Step 5 Go back to the console and click **Select File** to upload the populated template to MgC.

Step 6 Click **Confirm**. Then you can view the tables in the table list.

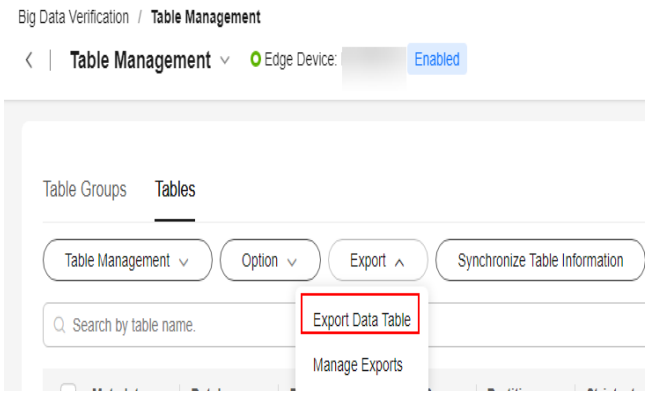
----End

Exporting Data Tables

You can export information about tables from Delta Lake (with metadata) and Hudi (with metadata) to CSV files. Information about data tables from sources without metadata cannot be exported.

Step 1 On the **Table Management** page, click the **Tables** tab.

Step 2 Choose **Export** > **Export Data Table** above the list.



Step 3 Select the table groups that contain the data tables to be exported and click **Confirm**.

Step 4 After the export is complete, choose **Export** > **Manage Exports**.

Step 5 Click **Download** in the **Operation** column to download the data table information to the local PC.

----End

12.5 Creating a Connection to the Target

12.5.1 Creating an Executor Connection

Executors are used to obtain data for consistency verification from the source and the target.

You need to connect the source and target executors to MgC, so that they can receive and execute commands from MgC. The supported big data components include:

- Hive Metastore
- Delta Lake (with metadata)
- Delta Lake (without metadata)
- Hudi (with metadata)
- Hudi (without metadata)

CAUTION

- Ensure that CLI clients are installed on both the source and target executors.
- You need to create connections to both the source and target executors.

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select the created **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Connection Management**.

Step 4 Click **Create Connection** in the upper right corner of the page.

Step 5 Select an executor, click **Next**, and set connection parameters based on **Table 12-12**.

Table 12-12 Parameters for creating an executor connection

Parameter	Configuration
Connection To	Select an option as required. **NOTICE** The CDH executor supports only the migration source. The MRS executor supports both the migration source and target.
Connection Name	The default name is *Executor type*-4 random characters (including letters and numbers). You can also customize a name.

Parameter	Configuration
Executor Credential	Select the login credential of the executor. For details about how to add credentials, see "Big data - Executor" in **Adding Resource Credentials**.
Executor IP Address	Enter the IP address for connecting to the executor.
Executor Port	Enter the port for connecting to the executor. The default port is **22**.
Installation Directory	Enter the installation directory of the MRS or CDH client. That is, the directory where **./install.sh** is installed.
SQL File Location	Enter a directory for storing the SQL files generated for consistency verification. You must have the read and write permissions for the folder. **NOTICE** After the verification is complete, you need to manually clear the folders generated at this location to release storage space.

Step 6 Click **Test**. MgC verifies whether the executor can be connected using the configuration information. If the test is successful, the executor can be connected.

Step 7 After the connection test is successful, click **Confirm**. The executor connection is created.

Step 8 On the **Connection Management** page, view the created connection and its basic information. In the **Operation** column, click **Modify** to modify the connection settings.

----End

12.5.2 Creating a Connection to a Target Cloud Service

To verify the consistency of data stored using big data cloud services, you need to establish connections between MgC and the source and target cloud services.

The supported target big data cloud services include:

- Data Lake Insight (DLI)
- CloudTable (ClickHouse)
- CloudTable (HBase)

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Connection Management**.

Step 4 Click **Create Connection** in the upper right corner of the page.

Step 5 Select a big data cloud service and click **Next**.

Step 6 Configure the connection parameters based on the big data service you selected.

- **Parameters for creating a connection to DLI**
- **Parameters for creating a connection to CloudTable (ClickHouse)**
- **Parameters for creating a connection to CloudTable (HBase)**

Table 12-13 Parameters for creating a connection to DLI

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **DLI-***4 random characters* (including letters and numbers). You can also customize a name.
DLI Credential	Select the credential used for accessing DLI. For details about how to add credentials, see "Big data - DLI" in **Adding Resource Credentials**. If the selected credential is the one you currently use to access MgC, you can select **This is my MgC credential**, and the projects in the region you choose will be listed.
Region/Project	Select the region and project where the data to be verified is located.
Queue	Enter the name of the DLI queue to be verified. The queue must be a SQL queue.

Table 12-14 Parameters required for creating a connection to CloudTable (ClickHouse)

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **CloudTable-ClickHouse***4 random characters* (including letters and numbers). You can also customize a name.
CloudTable (ClickHouse) Credential	Select the credential of the CloudTable (ClickHouse) cluster. For details about how to add credentials, see "Big data - ClickHouse" in **Adding Resource Credentials**. When you add the credential, select **Username/Password** for **Authentication** and enter the username and password for logging in to the ClickHouse cluster created on CloudTable.
Database URL	Enter the URL to access the CloudTable (ClickHouse) database. You can obtain the access address from the **basic information of the ClickHouse cluster**.

Table 12-15 Parameters required for creating a connection to CloudTable (HBase)

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **CloudTable-HBase***4 random characters* (including letters and numbers). You can also customize a name.
CloudTable (HBase) Credential	No credential is required for unsecured clusters.
Secured Cluster	Currently, only unsecured clusters (common clusters) are supported.
Database URL	Enter the URL to access the CloudTable (HBase) database. You can obtain the access address from the**basic information** of the HBase cluster. The address displayed after **ZK Link (Intranet)** is one you need.

Step 7 Click **Test**. MgC verifies whether the cloud service can be connected using the configuration information. If the test is successful, the cloud services can be connected.

Step 8 After the connection test is successful, click **Confirm**. The cloud service connection is created.

Step 9 On the **Connection Management** page, view the created connection and its basic information. In the **Operation** column, click **Modify** to modify the connection settings.

----End

12.5.3 Creating a Connection to a Target Component

To verify the consistency of data stored using big data components, you need to establish connections between MgC and the big data components.

The supported big data components include:

- Doris
- HBase
- ClickHouse
- Hive Metastore

Procedure

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Connection Management**.

Step 4 Click **Create Connection** in the upper right corner of the page.

Step 5 Select a big data component and click **Next**.

Step 6 Set parameters based on the big data component you selected.

- **Parameters for creating a connection to Doris**
- **Parameters for creating a connection to HBase**
- **Parameters for creating a connection to ClickHouse**
- **Parameters for creating a connection to Hive Metastore**

Table 12-16 Parameters for creating a connection to Doris

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **Doris-***4 random characters* (including letters and numbers). You can also customize a name.
Doris Credential	Select the target Doris credential added to Edge. For details about how to add credentials, see "Big data - Doris" in **Adding Resource Credentials**.
Database IP Address	Enter the IP address for accessing the target Doris cluster. To obtain the address, log in to **FusionInsight Manager**, choose **Cluster** > **Services** > **Doris**, and check **Host Where Leader Locates**.
Database Port	Enter the port for accessing the target Doris cluster. To obtain the port, log in to **FusionInsight Manager**, choose **Cluster** > **Services** > **Doris** > **Configurations**, and search for **query_port**.
Database Name	Enter the name of the target Doris database.

Table 12-17 Parameters for creating a connection to HBase

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **HBase-***4 random characters* (including letters and numbers). You can also customize a name.
HBase Credential	Select the target HBase credential added to Edge. For details about how to add credentials, see **Big data - HBase** in **Adding Resource Credentials**.
Secured Cluster	Choose whether the cluster is secured.

Parameter	Configuration
ZooKeeper IP Address	Enter the IP address for connecting to the target ZooKeeper node. You can enter the public or private IP address of the ZooKeeper server.
ZooKeeper Port	Enter the port number for connecting to the target ZooKeeper node. The default value is 2181.
HBase Version	Select the target HBase version.

Table 12-18 Parameters for creating a connection to ClickHouse

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **ClickHouse-***4 random characters* (including letters and numbers). You can also customize a name.
ClickHouse Credential (Optional)	Select the target ClickHouse credential added to Edge. For details about how to add credentials, see "Big data - ClickHouse" in **Adding Resource Credentials**.
Secured Cluster	Choose whether the cluster is secured.
ClickHouse IP Address	Enter the IP address of the target ClickHouse server.
HTTP Port	If the ClickHouse cluster is unsecured, enter the HTTP port for communicating with the target ClickHouse server. The default value is 8123.
HTTP SSL/TLS Port	If the ClickHouse cluster is secured, enter the HTTPS port for communicating with the target ClickHouse server.

Table 12-19 Parameters for creating a connection to Hive Metastore

Parameter	Configuration
Connection To	Select **Target**.
Connection Name	The default name is **Hive-Metastore-***4 random characters* (including letters and numbers). You can also customize a name.

Parameter	Configuration
Secure Connection	Choose whether to enable secure connection. • If Hive Metastore is deployed in an unsecured cluster, do not enable secure connection. • If Hive Metastore is deployed in a secured cluster, enable secure connection and provide access credentials. For details about how to obtain and add credentials to MgC, see "Big data - Hive Metastore" in **Adding Resource Credentials**.
Hive Version	Select the target Hive version. **CAUTION** Only version 3.x is available.
Hive Metastore IP Address	Enter the IP address for connecting to the Hive Metastore node.
Hive Metastore Thrift Port	Enter the port for connecting to the Hive Metastore Thrift service. The default port is 9083.

Step 7 Click **Test**. MgC verifies whether the component can be connected using the configuration information. If the test is successful, the cloud services can be connected.

Step 8 After the connection test is successful, click **Confirm**. The connection is created.

Step 9 On the **Connection Management** page, view the created connection and its basic information. In the **Operation** column, click **Modify** to modify the connection settings.

----End

12.6 Creating and Executing Verification Tasks

You can use the created source and target connections to create verification tasks.

For details about the supported big data components and verification methods, see **Overview**.

Precautions

- A pair of verification tasks for the source and the target must use the same verification method.
- If the source and target HBase clusters use different security authentication modes, the verification tasks cannot be executed at the same time, or they will fail to be executed. This is because the authentication information must be handled differently in each cluster. The secured cluster requires authentication information to be loaded, whereas the non-secured cluster needs that information cleared.

- A verification task must be completed within the same day. If the task execution extends past midnight (00:00), the verification results may be inaccurate. Plan verification tasks carefully to avoid executing across days.

Notes and Constraints

- Before verifying data migrated from EMR Delta Lake to MRS Delta Lake, please note:
  - If the source EMR cluster uses Spark 3.3.1, data verification is supported regardless of whether the source cluster contains metadata storage.
  - If the source EMR cluster uses Spark 2.4.8, data verification is supported only when the source cluster contains metadata storage.
- Verification is not available for Lindorm and HBase tables that only store cold data.
- The verification results of data migrated between Hive 2.x and Hive 3.x may be inaccurate. In Hive 2.x, when you query the fixed-length type CHAR (N) of data, if the actual data length does not meet the specified length N, Hive will pad the string with spaces to reach the required length. However, in Hive 3.x, this padding operation does not occur during queries. To avoid this issue, you are advised to use Beeline to perform the verification.

Prerequisites

- A table group has been created, and tables to be verified have been added to the group. For details, see **Creating a Table Group and Adding Tables to the Group**.
- Connections to the source and the target have been created. For details, see **Creating Connections**

Procedure

Step 1 Sign in to the **MgC console**.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Task Management**.

Step 4 Click **Create Task** in the upper right corner of the page.

Step 5 Select a big data component and verification method as needed and click **Next**.

Step 6 Configure task parameters based on the selected big data component and verification method.

NOTE

The task parameters vary with the big data component.

- **Parameters for creating a full verification task**
- **Parameters for creating a daily incremental verification task**
- **Parameters for creating an hourly incremental verification task**
- **Parameters for creating a date-based verification task**
- **Parameters for creating a selective verification task**

Table 12-20 Parameters for creating a full verification task

Area	Parameter	Configuration
Basic Info	Task Name	The default name is *Component*-**full-verification**-*4 random characters* (including letters and numbers). You can also customize a name.
Task Settings	Table Groups	Select the table groups that contain the tables to be verified.
	Executor Connection	This parameter is available for Hive, Delta Lake, and Hudi. • To create a verification task for the source, select the source executor connection. • To create a verification task for the target, select the target executor connection.
	DLI Connection	If DLI is selected in the previous step, the task can only be created for the target. You need to select the **created DLI connection**.
	Doris Connection	If Doris is selected in the previous step: • To create a verification task for the source, select the source Doris connection. • To create a verification task for the target, select the target Doris connection.
	HBase Connection	This parameter is available for HBase and CloudTable (HBase). • To create a verification task for the source, select the source HBase connection. • To create a verification task for the target, select the target HBase or CloudTable (HBase) connection.
	ClickHouse Connection	This parameter is available for ClickHouse, Alibaba Cloud ApsaraDB for ClickHouse, and CloudTable (ClickHouse) • To create a verification task for the source, select the source MRS ClickHouse or Alibaba Cloud ApsaraDB for ClickHouse connection. • To create a verification task for the target, select the target MRS ClickHouse or CloudTable (ClickHouse) connection.
	Execution Time	Specify when the task will be executed. After the task is activated, it will be automatically executed at the specified time every day. **NOTE** You are advised to run the verification during off-peak hours.

Area	Parameter	Configuration
	Advanced Options	• **Concurrency**: Specify the number of concurrent threads on an executor for the verification task. The default value is **3**. The value ranges from **1** to **10**. **CAUTION** If you are creating a verification task for an Alibaba Cloud EMR Hive cluster, set this parameter based on the source data volume and master node specifications. Consider the following rules: – The total number of concurrent threads for verification tasks running simultaneously in the source cluster cannot exceed 70% of the total number of cores on the metadata node. – The total resources allocated to verification tasks cannot exceed the resources of the execution queue. The total resources allocated to tasks can be calculated as follows: Allocated memory = Number of executors × Memory on an executor × Concurrency Allocated cores = Number of executors × Cores on an executor × Concurrency Assume that the total source data volume is 500 GB spread across 10,000 tables, in which there are 8 large tables with 50 GB data and 100,000 partitions. The master node has 8 vCPUs and 32 GB of memory. – According to rule 1, the maximum number of concurrent requests is 5, which is the rounded down value of 5.6 (0.7 × 8). – According to rule 2, you need to select **spark-sql** in the **Execution Command** area and set the following parameters:  executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G • **Max. SQL Statements Per File**: Each time the task is executed, files are created for storing the SQL statements generated for querying tables. You can control how many SQL statements can be stored in a single file. The default value is **10**. The recommended value ranges from **1** to **50**.

Area	Parameter	Configuration
		• **Timeout (s)**: indicates the maximum time allowed for an SQL statement to end normally. The unit is second (s). The default value is **600**. The value ranges from **600** to **7,200**. • **Send SMN Notifications**: Determine whether to use SMN to notify you of the task status in a timely manner through emails, SMS messages, or customized URLs. **NOTICE** – Before enabling this function, you need to create a topic on the SMN console. For details, see **Creating a Topic**. – Using this function may incur a small amount of fees, which are billed by SMN. For details, see **SMN Billing**.
Data Filtering	Non-partitioned Table Verification	Decide how to verify non-partitioned tables. • **Verify all**: All non-partition tables are verified as consistent. • **Skip all**: All non-partitioned tables are skipped during consistency verification. • **Filter by update time**: Only non-partitioned tables whose update time falls within the specified time range are verified as consistent. The update time of a non-partitioned table may not be accurate if it contains non-inserted data, and it may be excluded from verification.
	Advanced Options	**Max Fields Per SQL Statement**: Limit how many fields can be queried by one SQL statement. Too many or few fields in a SQL statement results in low query efficiency. The default value **0** means no limit is set. The value ranges from **100** to **500**.
OBS Bucket Check	-	• If you need to upload task logs and content verification results to an OBS bucket for management and analysis, **configure an OBS bucket**. After the bucket is configured, the task logs and content verification results will be automatically uploaded to the specified OBS bucket. • If you do not need to upload the task logs and content verification results to OBS, select **I confirm that I only need to view logs and data verification results on Edge and do not need to upload them to OBS**.

Area	Parameter	Configuration
Execution Script (This parameter is only available for Hive.)	Security Authentication	If security authentication (for example, Kerberos authentication) is enabled for the big data cluster, select this option and configure the security authentication command. You must first manually upload the **.keytab** file that contains the authentication key to the executor. • **Keytab Path**: Enter the path where the **.keytab** file is stored on the executor. • **Keytab Principal**: Enter the principal of the **.keytab** file.
	Execution Command	You can configure Beeline or Spark SQL command parameters to run SQL statements for verifying data consistency. • **Beeline**: a command line tool used to interact with Hive. • **Spark SQL**: used to execute SQL statements to query and analyze data.
MaxCompute Parameters (This item is only available for MaxCompute.)	-	Add MaxCompute parameters as needed. For details about the parameters, see **MaxCompute Documentation**.
Execution Settings (This item is only available for DLI.)	Parameter	Set the parameters as required. For details about the supported custom parameters, see **Custom Parameters**.
Execution Settings (This item is only available for HBase.)	Run Mode	The supported run modes include: **Yarn**: This mode applies to large-scale distributed environments. It can make full use of cluster resources and improve task concurrency and efficiency. **Local**: This mode applies to small-scale datasets or development and test environments for quick debugging and verification.
	Parameters	Add command parameters based on the selected run mode and requirements.

Area	Parameter	Configuration
Command Parameters (This item is only available for Delta Lake and Hudi.)	spark-sql	Spark SQL is a module used in Spark for structured data processing. In Spark applications, you can seamlessly use SQL statements or DataSet APIs to query structured data. For more information, see **SparkSQL Principles**. Retain the default settings.

	spark-submit	This is a basic Spark shell command used to submit Spark applications. The command is as follows:
		`./bin/spark-submit \`
		`  --class <main-class> \`
		`  --master <master-url> \`
		`  ... # other options`
		`  <application-jar> \`
		`  [application-arguments]`
		Parameter description:
		• **--class**: indicates the name of the class of a Spark application.
		• **--master**: indicates the master to which the Spark application links, such as Yarn-client and Yarn-cluster.
		• **application-jar**: indicates the path of the JAR file of the Spark application.
		• **application-arguments**: indicates the parameter required to submit the Spark application. (This parameter can be empty.)
		The parameters that need to be added depend on the scenario:
		• If you are creating a verification task for an Alibaba Cloud EMR Delta Lake cluster that uses Spark 3, add the following parameters:
		– Parameter: **jars**
		– Value: **'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'**
		CAUTION
		Replace the parameter values with the actual environment directory and Delta Lake version.
		• If you are creating a verification task for an Alibaba Cloud EMR Delta Lake 2.1.0 cluster that uses Spark 2.4.8, add the following parameter:
		– Parameter: **mgc.delta.spark.version**
		– Value: **2**

Table 12-21 Parameters for creating a daily incremental verification task

Area	Parameter	Configuration
Basic Info	Task Name	The default name is *Component*-**daily-incremental-verification-***4 random characters* (including letters and numbers). You can also customize a name.
Task Settings	Table Groups	Select the table groups that contain the tables to be verified.
	Executor Connection	This parameter is available for Hive, Delta Lake, and Hudi. • To create a verification task for the source, select the source executor connection. • To create a verification task for the target, select the target executor connection.
	DLI Connection	If DLI is selected in the previous step, the task can only be created for the target. You need to select the **created DLI connection**.
	Doris Connection	If Doris is selected in the previous step: • To create a verification task for the source, select the source Doris connection. • To create a verification task for the target, select the target Doris connection.
	HBase Connection	If HBase is selected in the previous step: • To create a verification task for the source, select the source HBase connection. • To create a verification task for the target, select the target HBase or CloudTable (HBase) connection.
	(Optional) Metadata Connection	This parameter is available only if you are creating a verification task for a target Hive cluster. The connection is used to check whether the partitions to be verified can be found in the target cluster. Select the metadata connection to the target Hive cluster.
	Execution Time	Specify when the task will be executed. After the task is activated, it will be automatically executed at the specified time every day.

Area	Parameter	Configuration
	Advanced Options	• **Concurrency**: Specify the number of concurrent threads on an executor for the verification task. The default value is **3**. The value ranges from **1** to **10**. **CAUTION** If you are creating a verification task for an Alibaba Cloud EMR Hive cluster, set this parameter based on the source data volume and master node specifications. Consider the following rules: – The total number of concurrent threads for verification tasks running simultaneously in the source cluster cannot exceed 70% of the total number of cores on the metadata node. – The total resources allocated to verification tasks cannot exceed the resources of the execution queue. The total resources allocated to tasks can be calculated as follows: Allocated memory = Number of executors × Memory on an executor × Concurrency Allocated cores = Number of executors × Cores on an executor × Concurrency Assume that the total source data volume is 500 GB spread across 10,000 tables, in which there are 8 large tables with 50 GB data and 100,000 partitions. The master node has 8 vCPUs and 32 GB of memory. – According to rule 1, the maximum number of concurrent requests is 5, which is the rounded down value of 5.6 (0.7 × 8). – According to rule 2, you need to select **spark-sql** in the **Execution Command** area and set the following parameters:  executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G • **Max. SQL Statements Per File**: Each time the task is executed, files are created for storing the SQL statements generated for querying tables. You can control how many SQL statements can be stored in a single file. The default value is **10**. The recommended value ranges from **1** to **50**.

Area	Parameter	Configuration
		• **Timeout (s)**: indicates the maximum time allowed for an SQL statement to end normally. The unit is second (s). The default value is **600**. The value ranges from **600** to **7,200**. • **Send SMN Notifications**: Determine whether to use SMN to notify you of the task status in a timely manner through emails, SMS messages, or customized URLs. **NOTICE** – Before enabling this function, you need to create a topic on the SMN console. For details, see **Creating a Topic**. – Using this function may incur a small amount of fees, which are billed by SMN. For details, see **SMN Billing**.
Data Filtering	Incremental Scope	Select the time period in which the incremental data needs to be verified as consistent. By default, a 24-hour period is selected. **T** indicates the execution time of the task, and **T-n** indicates a time n × 24 hours before the execution time. If you select **Consecutive days**, the system verifies consistency of the incremental data generated during these consecutive days.
	Non-partitioned Table Verification	Decide how to verify non-partitioned tables. • **Verify all**: All non-partition tables are verified. • **Skip all**: All non-partitioned tables are skipped during consistency verification. • **Filter by update time**: Only non-partitioned tables whose update time falls within the specified time range are verified as consistent. The update time of a non-partitioned table may not be accurate if it contains non-inserted data, and it may be excluded from verification.

Area	Parameter	Configuration
	Advanced Options	• **Partition Filtering**: Determine to filter table partitions by creation time or update time. – **By update time**: An update time indicates the timestamp a table partition was last modified or updated. Choose this option if you are concerned about the latest status or changes of data in a partition. – **By creation time**: A creation time indicates the timestamp when a partition was created. Choose this option if you are concerned about the data generated from the time when the partition is created to a certain time point. • **Max. Partitions**: Limit how many partitions in a table are verified. The default value is **3**. The value ranges from **1** to **50**. For example, if this parameter is set to **3**, the system verifies the consistency of only the first three partitions that are sorted by ID in descending order. • **Max Fields Per SQL Statement**: Limit how many fields can be queried by one SQL statement. Too many or few fields in a SQL statement results in low query efficiency. The default value **0** means no limit is set. The value ranges from **100** to **500**.
OBS Bucket Check	-	• If you need to upload task logs and content verification results to an OBS bucket for management and analysis, **configure an OBS bucket**. After the bucket is configured, the task logs and content verification results will be automatically uploaded to the specified OBS bucket. • If you do not need to upload the task logs and content verification results to OBS, select **I confirm that I only need to view logs and data verification results on Edge and do not need to upload them to OBS**.

Area	Parameter	Configuration
Execution Script (This parameter is only available for Hive.)	Security Authentication	If security authentication (for example, Kerberos authentication) is enabled for the big data cluster, select this option and configure the security authentication command. You must first manually upload the **.keytab** file that contains the authentication key to the executor. • **Keytab Path**: Enter the path where the **.keytab** file is stored on the executor. • **Keytab Principal**: Enter the principal of the **.keytab** file.
	Execution Command	You can configure Beeline or Spark SQL command parameters to run SQL statements for verifying data consistency. • **Beeline**: a command line tool used to interact with Hive. • **Spark SQL**: a command line tool used to execute SQL statements to query and analyze Hive data.
MaxCompute Parameters (MaxCompute)	-	Add MaxCompute parameters as needed. For details about the parameters, see **MaxCompute Documentation**.
Execution Settings (This item is only available for DLI.)	Parameter	Set the parameters as required. For details about the supported custom parameters, see **Custom Parameters**.
Statistics Configuration (HBase)	Run Mode	The supported run modes include: **Yarn**: This mode applies to large-scale distributed environments. It can make full use of cluster resources and improve task concurrency and efficiency. **Local**: This mode applies to small-scale datasets or testing and development environments. It enables rapid debugging and verification.
	Parameter	Add command parameters based on the selected run mode and requirements.
Command Parameters (Delta Lake and Hudi)	spark-sql	Spark SQL is a module used in Spark for structured data processing. In Spark applications, you can seamlessly use SQL statements or DataSet APIs to query structured data. For more information, see **SparkSQL Principles**. Retain the default settings.

Area	Parameter	Configuration
	spark-submit	This is a basic Spark shell command used to submit Spark applications. The command is as follows: ./bin/spark-submit \ --class <main-class> \ --master <master-url> \ ... # other options <application-jar> \ [application-arguments] Parameter description: ● **--class**: indicates the name of the class of a Spark application. ● **--master**: indicates the master to which the Spark application links, such as Yarn-client and Yarn-cluster. ● **application-jar**: indicates the path of the JAR file of the Spark application. ● **application-arguments**: indicates the parameter required to submit the Spark application. (This parameter can be empty.) The parameters that need to be added depend on the scenario: ● If you are creating a verification task for an Alibaba Cloud EMR Delta Lake cluster that uses Spark 3, add the following parameters: – Parameter: **jars** – Value: **'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'** **CAUTION** Replace the parameter values with the actual environment directory and Delta Lake version. ● If you are creating a verification task for an Alibaba Cloud EMR Delta Lake 2.1.0 cluster that uses Spark 2.4.8, add the following parameter: – Parameter: **mgc.delta.spark.version** – Value: **2**

Table 12-22 Parameters for creating an hourly incremental verification task

Area	Parameter	Configuration
Basic Info	Task Name	The default name is *Component*-**hourly incremental verification-***4 random characters* (including letters and numbers). You can also customize a name.
Task Settings	Table Groups	Select the table groups that contain the tables to be verified.
	Executor Connection	This parameter is available for Hive, Delta Lake, and Hudi. • To create a verification task for the source, select the source executor connection. • To create a verification task for the target, select the target executor connection.
	DLI Connection	If DLI is selected in the previous step, the task can only be created for the target. You need to select the **created DLI connection**.
	Doris Connection	If Doris is selected in the previous step: • To create a verification task for the source, select the source Doris connection. • To create a verification task for the target, select the target Doris connection.
	(Optional) Metadata Connection	This parameter is available only if you are creating a verification task for a target Hive cluster. The connection is used to check whether the partitions to be verified can be found in the target cluster. Select the metadata connection to the target Hive cluster.
	Start Time	Specify when the task will be executed. After the task is activated, it will be automatically executed at the specified time.

Area	Parameter	Configuration
	Advanced Options	• **Concurrency**: Specify the number of concurrent threads on an executor for the verification task. The default value is **3**. The value ranges from **1** to **10**. **CAUTION** If you are creating a verification task for an Alibaba Cloud EMR Hive cluster, set this parameter based on the source data volume and master node specifications. Consider the following rules: – The total number of concurrent threads for verification tasks running simultaneously in the source cluster cannot exceed 70% of the total number of cores on the metadata node. – The total resources allocated to verification tasks cannot exceed the resources of the execution queue. The total resources allocated to tasks can be calculated as follows: Allocated memory = Number of executors × Memory on an executor × Concurrency Allocated cores = Number of executors × Cores on an executor × Concurrency Assume that the total source data volume is 500 GB spread across 10,000 tables, in which there are 8 large tables with 50 GB data and 100,000 partitions. The master node has 8 vCPUs and 32 GB of memory. – According to rule 1, the maximum number of concurrent requests is 5, which is the rounded down value of 5.6 (0.7 × 8). – According to rule 2, you need to select **spark-sql** in the **Execution Command** area and set the following parameters:  executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G • **Max. SQL Statements Per File**: Each time the task is executed, files are created for storing the SQL statements generated for querying tables. You can control how many SQL statements can be stored in a single file. The default value is **10**. The recommended value ranges from **1** to **50**. • **Timeout (s)**: indicates the maximum time allowed for an SQL statement to end normally.

Area	Parameter	Configuration
		The unit is second (s). The default value is **600**. The value ranges from **600** to **7,200**. • **Send SMN Notifications**: Determine whether to use SMN to notify you of the task status in a timely manner through emails, SMS messages, or customized URLs. **NOTICE** – Before enabling this function, you need to create a topic on the SMN console. For details, see **Creating a Topic**. – Using this function may incur a small amount of fees, which are billed by SMN. For details, see **SMN Billing**.
Data Filtering	Execution Interval	Control how frequent the task will be executed.
	Non-partitioned Table Verification	Decide how to verify non-partitioned tables. • **Verify all**: All non-partition tables are verified as consistent. • **Skip all**: All non-partitioned tables are skipped during consistency verification. • **Filter by update time**: Only non-partitioned tables whose update time falls within the specified time range are verified as consistent. The update time of a non-partitioned table may not be accurate if it contains non-inserted data, and it may be excluded from verification.

Area	Parameter	Configuration
	Advanced Options	• **Partition Filtering**: Determine to filter table partitions by creation time or update time. – **By update time**: An update time indicates the last time a table partition was modified or updated. Choose this option if you are concerned about the latest status or changes of data in a partition. – **By creation time**: A creation time indicates the timestamp when a partition was created. Choose this option if you are concerned about the data generated from the time when the partition is created to a certain time point. • **Max. Partitions**: Limit how many partitions in a table are verified. The default value is **3**. The value ranges from **1** to **50**. For example, if this parameter is set to **3**, the system verifies the consistency of only the first three partitions that are sorted by ID in descending order. • **Max Fields Per SQL Statement**: Limit how many fields can be queried by one SQL statement. Too many or few fields in a SQL statement results in low query efficiency. The default value **0** means no limit is set. The value ranges from **100** to **500**.
OBS Bucket Check	-	• If you need to upload task logs and content verification results to an OBS bucket for management and analysis, **configure an OBS bucket**. After the bucket is configured, the task logs and content verification results will be automatically uploaded to the specified OBS bucket. • If you do not need to upload the task logs and content verification results to OBS, select **I confirm that I only need to view logs and data verification results on Edge and do not need to upload them to OBS**.

Area	Parameter	Configuration
Execution Script (This parameter is only available for Hive.)	Security Authentication	If security authentication (for example, Kerberos authentication) is enabled for the big data cluster, select this option and configure the security authentication command. You must first manually upload the **.keytab** file that contains the authentication key to the executor. • **Keytab Path**: Enter the path where the **.keytab** file is stored on the executor. • **Keytab Principal**: Enter the user name corresponding to the **.keytab** file.
	Execution Command	You can configure Beeline or Spark SQL command parameters to run SQL statements for consistency verification. • **Beeline**: a command line tool used to interact with Hive. • **Spark SQL**: a command line tool used to execute SQL statements to query and analyze Hive data.
MaxCompute Parameters (MaxCompute)	-	Add MaxCompute parameters as needed. For details about the parameters, see **MaxCompute Documentation**.
Execution Settings (This item is only available for DLI.)	Parameter	Set the parameters as required. For details about the supported custom parameters, see **Custom Parameters**.
Command Parameters (Delta Lake and Hudi)	spark-sql	Spark SQL is a module used in Spark for structured data processing. In Spark applications, you can seamlessly use SQL statements or DataSet APIs to query structured data. For more information, see **SparkSQL Principles**. Retain the default settings.

Area	Parameter	Configuration
	spark-submit	This is a basic Spark shell command used to submit Spark applications. The command is as follows: ./bin/spark-submit \ --class <main-class> \ --master <master-url> \ ... # other options <application-jar> \ [application-arguments] Parameter description: ● **--class**: indicates the name of the class of a Spark application. ● **--master**: indicates the master to which the Spark application links, such as Yarn-client and Yarn-cluster. ● **application-jar**: indicates the path of the JAR file of the Spark application. ● **application-arguments**: indicates the parameter required to submit the Spark application. (This parameter can be empty.) The parameters that need to be added depend on the scenario: ● If you are creating a verification task for an Alibaba Cloud EMR Delta Lake cluster that uses Spark 3, add the following parameter: – Parameter: **jars** – Value: **'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'** **CAUTION** Replace the parameter values with the actual environment directory and Delta Lake version. ● If you are creating a verification task for an Alibaba Cloud EMR Delta Lake 2.1.0 cluster that uses Spark 2.4.8, add the following parameter: – Parameter: **mgc.delta.spark.version** – Value: **2**

Table 12-23 Parameters for creating a date-based verification task

Area	Parameter	Configuration
Basic Info	Task Name	The default name is *Component*-**date-based-Verification-***4 random characters* (including letters and numbers). You can also customize a name.

Area	Parameter	Configuration
Task Settings	Table Groups	Select the table groups that contain the tables to be verified.
	Executor Connection	This parameter is available for Hive, Delta Lake, and Hudi. • To create a verification task for the source, select the source executor connection. • To create a verification task for the target, select the target executor connection.
	DLI Connection	If DLI is selected in the previous step, the task can only be created for the target. You need to select the **created DLI connection**.
	(Optional) Metadata Connection	This parameter is available only if you are creating a verification task for a target Hive cluster. The connection is used to check whether the partitions to be verified can be found in the target cluster. Select the metadata connection to the target Hive cluster.
	Execution Time	Specify when the task will be executed. After the task is activated, it will be automatically executed at the specified time every day.

Area	Parameter	Configuration
	Advanced Options	• **Concurrency**: Specify the number of concurrent threads on an executor for the verification task. The default value is **3**. The value ranges from **1** to **10**. **CAUTION** If you are creating a verification task for an Alibaba Cloud EMR Hive cluster, set this parameter based on the source data volume and master node specifications. Consider the following rules: – The total number of concurrent threads for verification tasks running simultaneously in the source cluster cannot exceed 70% of the total number of cores on the metadata node. – The total resources allocated to verification tasks cannot exceed the resources of the execution queue. The total resources allocated to tasks can be calculated as follows: Allocated memory = Number of executors × Memory on an executor × Concurrency Allocated cores = Number of executors × Cores on an executor × Concurrency Assume that the total source data volume is 500 GB spread across 10,000 tables, in which there are 8 large tables with 50 GB data and 100,000 partitions. The master node has 8 vCPUs and 32 GB of memory. – According to rule 1, the maximum number of concurrent requests is 5, which is the rounded down value of 5.6 (0.7 × 8). – According to rule 2, you need to select **spark-sql** in the **Execution Command** area and set the following parameters:  executor-memory = 4G master = yarn num-executors = 20 executor-cores = 2 driver-memory = 10G • **Max. SQL Statements Per File**: Each time the task is executed, files are created for storing the SQL statements generated for querying tables. You can control how many SQL statements can be stored in a single file. The default value is **10**. The recommended value ranges from **1** to **50**.

Area	Parameter	Configuration
		• **Timeout (s)**: indicates the maximum time allowed for an SQL statement to end normally. The unit is second (s). The default value is **600**. The value ranges from **600** to **7,200**. • **Send SMN Notifications**: Determine whether to use SMN to notify you of the task status in a timely manner through emails, SMS messages, or customized URLs. **NOTICE** – Before enabling this function, you need to create a topic on the SMN console. For details, see **Creating a Topic**. – Using this function may incur a small amount of fees, which are billed by SMN. For details, see **SMN Billing**.
Data Filtering	Time Range	Select the time period in which the incremental data needs to be verified as consistent. By default, a 24-hour period is selected. **T** indicates the execution time of the task, and **T-n** indicates a time n × 24 hours before the execution time. If you select **Consecutive days**, the system verifies consistency of the incremental data generated during the specified days.
	Non-partitioned Table Verification	Decide how to verify non-partitioned tables. • **Verify all**: All non-partition tables are verified as consistent. • **Skip all**: All non-partitioned tables are skipped during consistency verification. • **Filter by update time**: Only non-partitioned tables whose update time falls within the specified time range are verified as consistent. The update time of a non-partitioned table may not be accurate if it contains non-inserted data, and it may be excluded from verification.
	Advanced Options	**Max Fields Per SQL Statement**: Limit how many fields can be queried by one SQL statement. Too many or few fields in a SQL statement results in low query efficiency. The default value **0** means no limit is set. The value ranges from **100** to **500**.

Area	Parameter	Configuration
OBS Bucket Check	-	• If you need to upload task logs and content verification results to an OBS bucket for management and analysis, **configure an OBS bucket**. After the bucket is configured, the task logs and content verification results will be automatically uploaded to the specified OBS bucket. • If you do not need to upload the task logs and content verification results to OBS, select **I confirm that I only need to view logs and data verification results on Edge and do not need to upload them to OBS**.
Execution Script (Hive)	Security Authentication	If security authentication (for example, Kerberos authentication) is enabled for the big data cluster, select this option and configure the security authentication command. You must first manually upload the **.keytab** file that contains the authentication key to the executor. • **Keytab Path**: Enter the path where the **.keytab** file is stored on the executor. • **Keytab Principal**: Enter the user name corresponding to the **.keytab** file.
	Execution Command	You can configure Beeline or Spark SQL command parameters to run SQL statements for consistency verification. • **Beeline**: a command line tool used to interact with Hive. • **Spark SQL**: a command line tool used to execute SQL statements to query and analyze Hive data.
MaxCompute Parameters (MaxCompute)	-	Add MaxCompute parameters as needed. For details about the parameters, see **MaxCompute Documentation**.
Execution Settings (This item is only available for DLI.)	Parameter	Set the parameters as required. For details about the supported custom parameters, see **Custom Parameters**.

Area	Parameter	Configuration
Command Parameters (Delta Lake and Hudi)	spark-sql	Spark SQL is a module used in Spark for structured data processing. In Spark applications, you can seamlessly use SQL statements or DataSet APIs to query structured data. For more information, see **SparkSQL Principles**. Retain the default settings.

	spark-submit	This is a basic Spark shell command used to submit Spark applications. The command is as follows:
		`./bin/spark-submit \`
		`--class <main-class> \`
		`--master <master-url> \`
		`... # other options`
		`<application-jar> \`
		`[application-arguments]`
		Parameter description:
		• **--class**: indicates the name of the class of a Spark application.
		• **--master**: indicates the master to which the Spark application links, such as Yarn-client and Yarn-cluster.
		• **application-jar**: indicates the path of the JAR file of the Spark application.
		• **application-arguments**: indicates the parameter required to submit the Spark application. (This parameter can be empty.)
		The parameters that need to be added depend on the scenario:
		• If you are creating a verification task for an Alibaba Cloud EMR Delta Lake cluster that uses Spark 3, add the following parameter:
		– Parameter: **jars**
		– Value: **'/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-core_2.12-*.jar,/opt/apps/DELTALAKE/deltalake-current/spark3-delta/delta-storage-*.jar'**
		CAUTION
		Replace the parameter values with the actual environment directory and Delta Lake version.
		• If you are creating a verification task for an Alibaba Cloud EMR Delta Lake 2.1.0 cluster that uses Spark 2.4.8, add the following parameter:
		– Parameter: **mgc.delta.spark.version**
		– Value: **2**

Table 12-24 Parameters for creating a selective verification task

Area	Parameter	Configuration
Basic Info	Task Name	The default name is *Component*-**selective verification**-*4 random characters* (including letters and numbers). You can also customize a name.
Task Settings	Table Groups	Select the table groups that contain the tables to be verified.
	HBase Connection	This parameter is available for HBase and CloudTable (HBase). • To create a verification task for the source, select the source HBase connection. • To create a verification task for the target, select the target HBase or CloudTable (HBase) connection.
	Advanced Options	• **Concurrency**: Specify the number of concurrent threads on an executor for the verification task. The default value is **3**. The value ranges from **1** to **10**. • **Timeout (s)**: indicates the maximum time allowed for an SQL statement to end normally. The unit is second (s). The default value is **600**. The value ranges from **600** to **7,200**. • **Send SMN Notifications**: Determine whether to use SMN to notify you of the task status in a timely manner through emails, SMS messages, or customized URLs. **NOTICE** – Before enabling this function, you need to create a topic on the SMN console. For details, see **Creating a Topic**. – Using this function may incur a small amount of fees, which are billed by SMN. For details, see **SMN Billing**.
Data Filtering	Time Range	Select the time period in which the data needs to be verified as consistent.

Area	Parameter	Configuration
OBS Bucket Check	-	• If you need to upload task logs and content verification results to an OBS bucket for management and analysis, **configure an OBS bucket**. After the bucket is configured, the task logs and content verification results will be automatically uploaded to the specified OBS bucket. • If you do not need to upload the task logs and content verification results to OBS, select **I confirm that I only need to view logs and data verification results on Edge and do not need to upload them to OBS**.
Execution Settings (HBase)	Run Mode	The supported run modes include: • **Yarn**: For large-scale distributed environments. It can make full use of cluster resources and improve verification concurrency and efficiency. • **Local**: For small-scale datasets or testing and development environments. It enables rapid debugging and verification.
	Parameter	Add command parameters based on the selected run mode and requirements.

Step 7 Click **Save**. After the creation is successful, the system automatically synchronizes the task settings to the Edge device. Then in the task list, you can view the created task and its settings synchronization status.

Step 8 After the settings synchronization is complete, execute the task using either of the following methods:

- **Automatic execution**: The task will be executed at the specified time automatically.
  a. In the task list, locate the task and click **Activate** in the **Schedule Status** column.
  b. In the displayed dialog box, click **OK** to activate the task.
- **Manual execution**: You can manually execute the task immediately.
  a. In the task list, locate the task and click **Execute** in the **Operation** column.
  b. In the displayed dialog box, click **OK** to execute the task immediately.

Step 9 Click **View Executions** in the **Operation** column. On the executions list page, you can:

- View the status, progress statistics, and execution start time and end time of each task execution.

> **CAUTION**
>
> If a task execution takes a long time or the page is incorrectly displayed, set the log level of the executor's Driver to **ERROR**.

- Upload the logs of a task execution to your OBS bucket for review and analysis by clicking **Upload Log** in the **Operation** column. Before uploading logs, you need to configure a log bucket on the Edge console. For details, see **Configuring a Log Bucket**.
- Cancel a running execution or terminate an execution using the **Cancel/Terminate** button.
- If there are tables whose verification results are not obtained, obtain the results again by clicking **Path** in the **Statistics** column.

----End

12.7 Viewing and Exporting Verification Results

After the verification tasks on the source and the target are executed, you can view and export the verification results of tables by batch. A batch is a group of verification tasks executed on the same day. If a task is executed multiple times on a single day, only the latest verification results are included in the statistics. Its previous results will be overwritten.

Prerequisites

You have **created and executed verification tasks**.

Precautions

When sum operations (like sum or allsum) are performed, if the results exceed the valid digit limit of the Double type, the precision is lost. Additionally, when very large floating-point numbers are verified, the actual differences may not be detected if they exceed the significant digits that the Double type can represent. In such cases, the calculated differences may erroneously appear as 0.

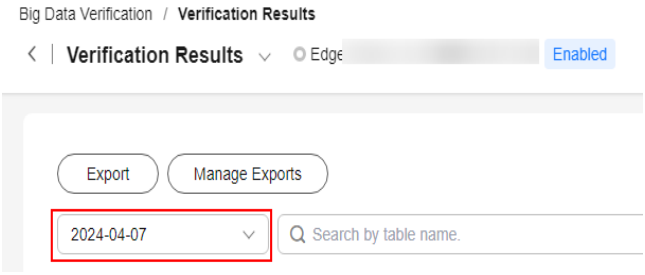
Viewing Verification Results

Step 1 Sign in to the **MgC** console.

Step 2 In the navigation pane on the left, choose **Migrate** > **Big Data Verification**. Select a **migration project** in the upper left corner of the page.

Step 3 In the **Features** area, click **Verification Results**.

Step 4 Select a task batch (date).



Step 5 In the list, perform the following operations on a table:

- View the verification details of table partitions by clicking the **View** in the **Partitions** column. Then click **View Details** to view the verification results of a table partition.
- Viewing the historical verification results by clicking **View History** in the **Operation** column.

----End

Exporting Verification Results

Step 1 On the **Verification Results** page, click **Export**.

Step 2 Select a batch and table group, and then click **Confirm**.

Step 3 After the export is complete, click **Manage Exports**.

Step 4 Click **Download** in the **Operation** column to download the verification results.

----End

12.8 Custom Parameters

The following table lists the custom parameters supported when you create a DLI verification task.

Table 12-25 Custom parameters supported for DLI verification tasks

Parameter	Default Value	Description
mgc.mc2dli.table.partition.enable	true	Indicates whether to query the DLI metadata when a DLI table partition is empty or missing. • If this parameter is set to **true**, the verification status of any empty DLI table partitions will be succeeded, and that of any missing partitions will be failed. • If this parameter is set to **false**, the verification status of any empty or missing DLI table partitions will be succeeded.

Parameter	Default Value	Description
spark.sql.files.maxRecordsPerFile	0	The maximum number of records to be written into a single file. If the value is zero or negative, there is no limit.
spark.sql.autoBroadcastJoinThreshold	209715200	The maximum size of the table that displays all working nodes when a connection is executed. You can set this parameter to **-1** to disable the display. **NOTE** Currently, only the configuration unit metastore table that runs the **ANALYZE TABLE COMPUTE statistics noscan** command and the file-based data source table that directly calculates statistics based on data files are supported.
spark.sql.shuffle.partitions	200	The default number of partitions used to filter data for join or aggregation.
spark.sql.dynamicPartitionOverwrite.enabled	false	Whether DLI overwrites the partitions where data will be written into during runtime. If you set this parameter to **false**, all partitions that meet the specified condition will be deleted before data overwrite starts. For example, if you set this parameter to **false** and use INSERT OVERWRITE to write partition 2021-02 to a partitioned table that has the 2021-01 partition, this partition will be deleted. If you set this parameter to **true**, DLI does not delete partitions before overwrite starts.
spark.sql.files.maxPartitionBytes	134217728	The maximum number of bytes to be packed into a single partition when a file is read.
spark.sql.badRecordsPath	-	The path of bad records.

Parameter	Default Value	Description
spark.sql.legacy.correlated.scalar.query.enabled	false	• **If set to true:** – When there is no duplicate data in a subquery, executing a correlated subquery does not require deduplication from the subquery's result. – If there is duplicate data in a subquery, executing a correlated subquery will result in an error. To resolve this, the subquery's result must be deduplicated using functions such as **max()** or **min()**. • **If set to false:** Regardless of whether there is duplicate data in a subquery, executing a correlated subquery requires deduplicating the subquery's result using functions such as **max()** or **min()**. Otherwise, an error will occur.

