

Alibaba Cloud Lightning Cube

Migrate data from HTTP/HTTPS sources to OSS

Issue: 20190509

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus, page names, and other UI elements.	Click OK .
Courier font	It is used for commands.	Run the <code>cd / d C :/ windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[] or [a b]	It indicates that it is an optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>

Style	Description	Example
<code>{}</code> or <code>{a b}</code>	It indicates that it is a required value, and only one item can be selected.	<code>swich {stand slave}</code>

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 Background information.....	1
2 Prerequisites.....	2
3 Create a migration job.....	5
4 Manage migration jobs.....	9

1 Background information

This section describes how to migrate network resources to Object Storage Service (OSS).

Alibaba Cloud Data Transport is used as a data channel between various data stores . With Data Transport, you can migrate data from third-party data stores to OSS or between OSS buckets.

To migrate data, you only need to enter the data addresses for both the data source and OSS destination, and then create a migration job. After starting a migration job , you can perform management tasks for the job such as viewing the process and status of the job. Additionally, you can generate the migration report to view a list of migrated files and a list of failed files.

This guide includes the following sections:

- [Prerequisites](#)
- [Create a migration job](#)
- [Manage migration jobs](#)

2 Prerequisites

This section describes what you need to do before creating a migration job.

Network resources

- Estimate the amount of data to be migrated.

Estimate the size and number of files to be migrated.



Note:

To ensure a successful migration, you need to enter the appropriate size and number of files when [creating a migration job](#).

- Load network resources

1. Create a list file on the local PC.

The list file has two columns.

- The first column includes a list of HTTP/HTTPS addresses. The migration service uses the HTTP GET method to download a file from an HTTP/HTTPS address and the HTTP HEAD method to obtain metadata of a file.
- The second column includes a list of file names. After a file is migrated, the object name of the files includes a prefix and a file name. Separate two columns with a tab (`\ t`).



Notice:

You must specify a file in a list file rather than a file folder.

Each line includes a file name. Separate two lines with a line feed `\ n`.

Assume that a list file is named `list.txt`.

```
http :// docs - aliyun . cn - hangzhou . oss . aliyun - inc . com  
/ docs / my . doc          docs / my . doc  
http :// docs - aliyun . cn - hangzhou . oss . aliyun - inc . com  
/ pics / my . jpg          pics / my . jpg
```



Notice:

If a file name includes special characters, such as Chinese characters, spaces, and tabs, you need to perform URL encoding.

- When a file name includes signs, you must transcode the link and file name of the file. For example, a file is named `##. jpg` . After the file name is transcoded, the file name is displayed as `## EF % BF % A5 . jpg` and you need to change the link and file name of a row in a list file as follows.

```
http :// docs - aliyun . cn - hangzhou . oss . aliyun - inc .  
com / pics /## EF % BF % A5 . jpg      pics /## EF % BF % A5 .  
jpg
```

- When a file name includes Chinese characters, you must transcode the link of the file and keep the file name. For example, a file is named `< a Chinese file name >. jpg` . After the file name is transcoded, the file name is displayed as `% e5 % 9b % be % e7 % 89 % 87 . jpg` and you need to change the link of a row in a list file as follows.

```
http :// docs - aliyun . cn - hangzhou . oss . aliyun - inc .  
com / pics /% e5 % 9b % be % e7 % 89 % 87 . jpg      pics /< a  
Chinese file name >. jpg
```

2. Upload the list file to OSS.

The address of the list file has the following format: `oss ://{ bucket }/{
list file name }`.

Alibaba Cloud Object Storage Service

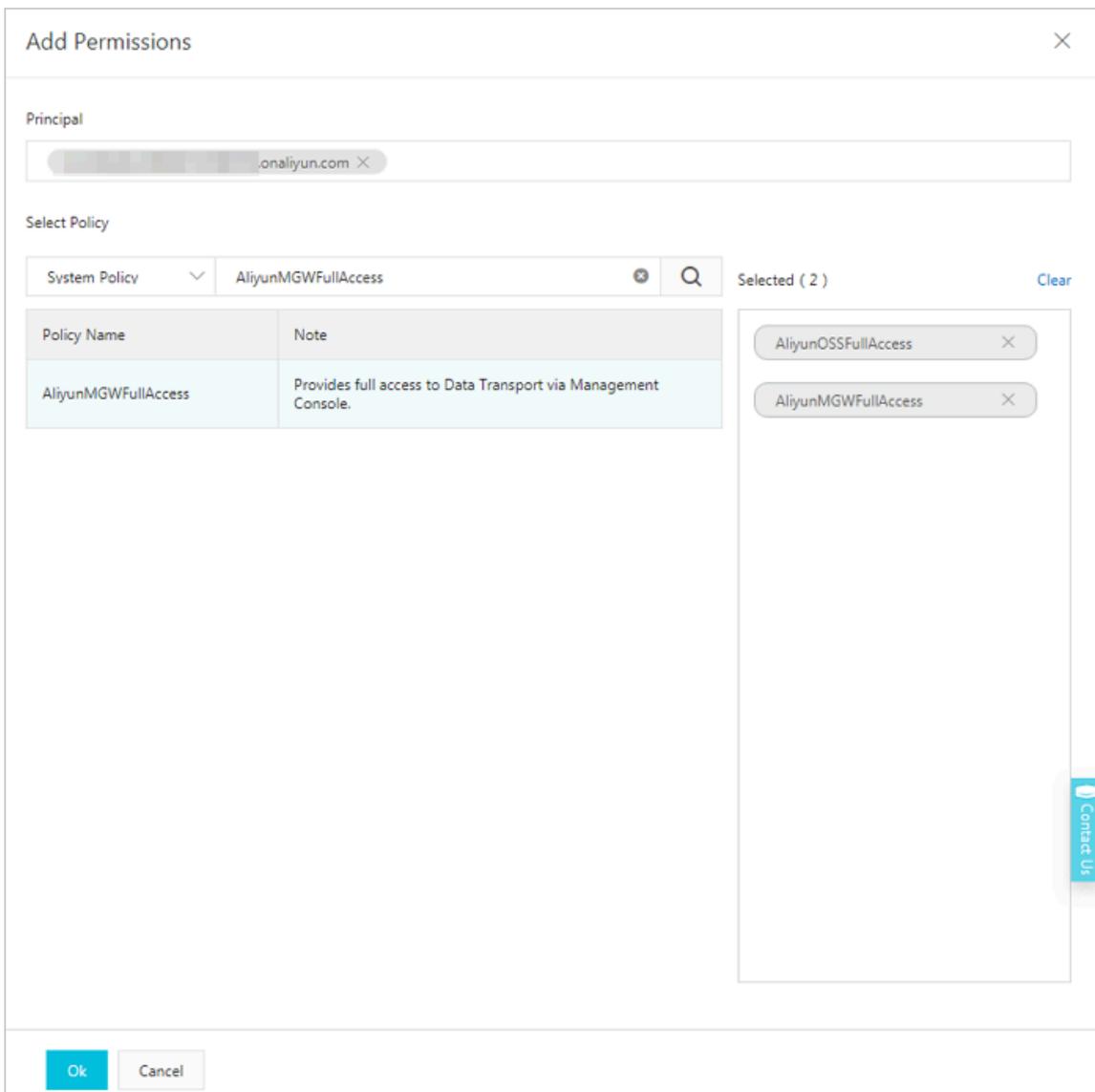
- Create a destination bucket.

Create a destination bucket, which is used to store the migrated data. For more information, see [Create a bucket](#).

- Create and authorize a RAM user

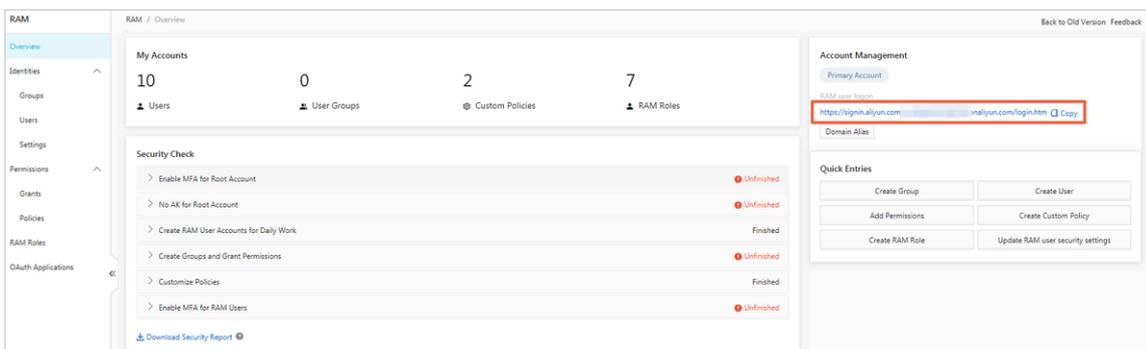
1. Log on to the [RAM console](#).
2. Choose Identities > Users > Create User.
3. Select Console Password Logon and Programmatic Access and then enter the required User Account Information.
4. Click OK to save the generated account, password, AccessKeyID, and AccessKeySecret.
5. Select the required user account, click Add Permissions to grant the read/write permission (AliyunOSSFullAccess) and migration permission

(AliyunMGWFullAccess) for the RAM user. The Add Permissions dialog is shown in the following figure.



6. Choose OK > Finished.

7. In the left-side navigation pane, select Overview, click the link in the RAM user logon section, and enter the username and password of the newly created RAM user to log on to the console.



3 Create a migration job

This section describes the operations and considerations for data migration.

Precautions

When creating a migration job, you need to note the following issues:

- A migration job occupies the network resources of the source data address and destination data address. To ensure business continuity, we recommend that you specify a speed limit for a migration task or perform the migration task during off-peak hours.
- Before a migration task is performed, files at both the source data address and the destination data address are checked. If files at the source data address have the same names as those at the destination data address and have a later update time than those at the destination data address, the files at the destination data address are overwritten during a migration task. If the content for two files is different, you must change the name of one of the files and back up these files.

Step 1: Create a source data address

1. Log on to the [Data Transport console](#).
2. Choose Data Online Migration > Data Address, and then click Create Data Address.
3. In the Create Data Address dialog box, set the required options and click OK. The options are described as follows:

Option	Required	Description
Data Type	Yes	Select Http/Https.
Data Name	Yes	The data name can be 3 to 63 characters in length. Special characters are not supported, except for hyphens (-) and underscores (_).
File Path	Yes	Enter oss://{bucket}/{the name of a list file}. For more information, see Create a list file .

Option	Required	Description
List Access Endpoint	Yes	Enter the appropriate endpoint based on the comparison table for regions and endpoints. For more information, see Regions and endpoints .
List Access AK and List Access SK	Yes	Enter an Accesskey that is used to migrate data.

- You are required to apply for whitelist permissions because this feature is in the beta testing phase. Click Application.
- Enter the required information and submit the beta testing application for migration. After the application has been approved, you will receive an SMS notification.

Step 2: Create a destination data address

- Select Data Online Migration > Data Address and click Create Data Address.
- In the Create Data Adress dialog box, set the required options and click OK. The options are described as follows.

Option	Required	Description
Data Type	Yes	Select OSS.
Data Region	Yes	Select a region where the destination data address is located.
Data Name	Yes	The data name can be 3 to 63 characters in length. Special characters are not supported, except for hyphens (-) and underscores (_).
OSS Endpoint	Yes	Select an endpoint based on the region where data is located. For more information, see Endpoints .
AccessKeyId and AccessKeySecret	Yes	Enter an AccessKey to migrate data. For more information, see Create an AccessKey .
OSS Bucket	Yes	Select a bucket to store migration data.

Option	Required	Description
OSS Prefix	No	An OSS prefix cannot start with a forward slash (/) and must end with a forward slash (/). For example: data / to / oss /. If you want to store data to the root directory of a bucket, you can leave the OSS Prefix field blank.

Step 3: Create a migration job

1. Select Data Online Migration > Migration Jobs and click Create Job.
2. In the Create Job dialog box, read the Terms of Migration Service, select I understand the above terms and conditions, and activate Data Transport, and then click Next.
3. In the Create Job dialog box, set the required options and click Next.

The options are described as follows:

Option	Required	Description
Job Name	Yes	The job name can 3 to 63 characters in length and can contain lowercase letters, numbers, and hyphens (-). A job name cannot start or end with a hyphen (-).
Source Data Address	Yes	Select the created source data address.
Destination Data Address	Yes	Select a destination data address that you have created.

Option	Required	Description
Migration Type	Yes	<p>Before you start a migration job, Data Transport compares files of the source data address with those of the destination data address. The files at the source data address are disregarded during migration. This occurs if the source files with an earlier update time have the same name, ContentType, and size as the destination files. However, all the other files are migrated.</p> <ul style="list-style-type: none"> · Full: performs a full migration. After all of the files are migrated, a migration job is closed. When you perform a full migration job again, the migration service only migrates files that have been changed after the last full migration job.

4. On the Performance tab, navigate to the Data Prediction section, and enter the Data Size and File Count.



Note:

To ensure a successful migration, you must estimate the amount of data to be migrated. For more information, see [Estimate the amount of data to be migrated](#).

5. This step is optional. On the Performance tab, navigate to the Flow Control section, set the Time Range and Max Flow, and then click Add.



Note:

To ensure business continuity, we recommend that you set the Time Range and Max Flow based on the fluctuation of visits. The default value of the Time Range is 06 : 00 - 12 : 00 . The default value of the Max Flow is 5 MB / s .

6. Click Create. Wait until a migration job is complete.

4 Manage migration jobs

This section describes several subsequent operations after you create a migration job.

View the status of a migration job

After you create a migration job, only one migration job status is displayed. The status can be one of the following:

- **Migrating:** indicates that data is migrating. Wait.
- **Create Failed:** indicates that you failed to create a migration job. You can view the cause of the failure and recreate a migration job.
- **Completed:** indicates that a migration job is complete. You can view a migration report.
- **Failed:** indicates that a migration job failed. You can view the migration report and migrate failed files.

Modify flow control settings

During a migration job, you can modify flow control settings at any time based on your needs.

1. In the [Data Transport console](#), choose **Data Online Migration > Migration Jobs**. On the **Migration Jobs** page, locate a migration job and click **Manage** next to the job.
2. Click **Stop** and ensure that the job is stopped.
3. On the **Flow Control Time Schedule** chart, click **Reset**.
 - To add a flow control setting, select the appropriate **Time Range** and **Max Flow**, and click **Add**.
 - To delete a flow control setting, click  next to the flow control setting.
 - To modify a flow control setting, you must first delete the previous setting and add a new flow control setting.
4. Click **OK**, and click **Start** to restart the job.

View a migration report

1. On the **Migration Jobs** page, locate a job and click **Manage** next to the job.

2. Click **Generate Migration Report**. After a report is generated, click **Export** to export the report.

In the migration report, the following file names appear in the **File list** section:

- The file name ends with `_total_list`. This file contains a list of total migration files.
 - The file name ends with `_completed_list`. This file contains a list of successful migration files.
 - The file name ends with `_error_list`. This file contains a list of failed migration files.
3. In the [OSS console](#), locate the automatically generated folder `aliyun_mgw_import_report`. The three files that appear in the migration report are included in this folder. You can download and view the detailed list of files. We recommend that you use the [ossbrowser](#) tool to view these files.

The file formats are as follows:

- The file name includes the source data address, file name, file size (measured in bytes), and last modified time. This file contains a list of total migration files. The format of the data source address is: `< vendor >://< bucketName >/< prefix >/< objectName >`. For example, `oss :// bucket - test1022 / myprefix / testfile . txt`.
- The file name includes the file name, file size (measured in bytes), checksum (CRC64), and migration completion time. This file contains a list of successful migration files.
- The file name includes the file name, migration start time, migration end time, and error description. This file contains a list of failed migration files.

Retry after a migration failure

If a migration job failed, you can view the generated file whose name ends with `_error_list` to find the cause of the failure and troubleshoot the issue. On the **Migration Jobs** page, locate a failed job, click **Manage** next to the job, and click **Retry** to migrate failed files.

More information

For more information, see the following sections:

- *Migrate data between Alibaba Cloud Object Storage Service (OSS) buckets*
- *Migrate data from Tencent Cloud Object Service (COS) to OSS*
- *Migrate data from Amazon Simple Storage Service (Amazon S3) to OSS*
- *Migrate data from Azure Blob to OSS*
- *Migrate data from Qiniu Cloud-Object Storage (KODO) to OSS*
- *Migrate data from Baidu Object Storage (BOS) to OSS*
- *Migrate data from Kingsoft Standard Storage Service (KS3) to OSS*
- *Migrate data from UPYUN Storage Service (USS) to OSS*
- *Migrate data from Google Cloud Storage to OSS*
- *Migrate data between NAS file systems*
- *Migrate data from NAS to OSS*
- *Migrate data from ECS instances to OSS*