

Alibaba Cloud

Data Online Migration

Migrate data from HTTP/HTTPS
sources to OSS

Document Version: 20220519

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 because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, 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 this document 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 directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
 Danger	A danger notice 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.
 Warning	A warning notice 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 restart an instance.
 Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: If the weight is set to 0, the server no longer receives new requests.
 Note	A note indicates supplemental instructions, best practices, tips, and other content.	 Note: You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click Settings> Network> Set network type .
Bold	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click OK .
<code>Courier font</code>	Courier font is used for commands	Run the <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

Table of Contents

1. Background information	05
2. Prepare for a migration job	06
3. Migrate data	08
4. Manage migration jobs	14

1. Background information

This tutorial describes how to migrate network resources to Alibaba Cloud Object Storage Service (OSS).

Alibaba Cloud Data Online Migration is a data channel between different storage services. You can use Data Online Migration to migrate data from third-party data stores to OSS or between OSS buckets.

When you use Data Online Migration, you need only to log on to the console, specify a source data address and a destination OSS data address, and then create a migration job. After the migration job starts, you can manage the job in the console. For example, you can view the progress and traffic information of the job. You can also generate a migration report to check the list of migrated files and the list of files that fail to be migrated.

This tutorial includes the following topics:

- [Preparations](#)
- [Implement migration](#)
- [Subsequent operations](#)


2.Prepare for a migration job

This topic describes how to prepare for a migration job.

Network resources

- Estimate the amount of data to be migrated.

Estimate the size and the number of files that you want to migrate.

 **Note** To make sure that your data can be migrated as expected, you must enter an appropriate size and number of files when you [create a migration job](#).

- Load network resources.

- i. Create a list file on your on-premises device.

The list file consists of two columns. Separate the columns with a tab character (`\t`). Each line represents one file. Separate files with line feeds (`\n`).

- The first column includes a list of HTTP or HTTPS URLs. Data Transport uses HTTP GET requests to download files from the HTTP or HTTPS URLs and uses HTTP HEAD requests to obtain metadata of the files.
- The second column includes a list of file names. After a file is migrated, the name of the object that corresponds to the file consists of a prefix and the file name. If a file name contains Chinese characters or special characters such as spaces and tab characters, you must encode the URL.
- If a file name contains symbols, you must encode the URL and the file name. For example, a file is named `# ¥.jpg`. After the file name is encoded, it is displayed as `##EF%BF%A5.jpg`. Use the following information to specify the file in the list file:

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

- If a file name contains Chinese characters, you must encode only the URL. Do not change the file name. For example, a file is named `<a Chinese file name>.jpg`. After the file name is encoded, it is displayed as `%e5%9b%be%e7%89%87.jpg`. Use the following information to specify the file in the list file:

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

Notice

- You must specify specific files instead of file folders in the list file.
- If your device runs on a Windows operating system, we recommend that you use a text editor such as EditPlus or Sublime Text instead of Notepad to edit the list file. If your device runs on a Linux operating system, we recommend that you use **Vi Improved (Vim)** to edit the list file.

The following example shows the content of a list file 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
http://docs-aliyun.cn-hangzhou.oss.aliyun-inc.com/pics/﻿#%EF%BF%A5.jpg      pics/﻿#%EF%BF
%A5.jpg
http://docs-aliyun.cn-hangzhou.oss.aliyun-inc.com/pics/%e5%9b%be%e7%89%87.jpg      pics
/<a Chinese file name>.jpg
```

- ii. Upload the list file to Object Storage Service (OSS). For more information, see [Simple upload](#).

Note

- After the list file is uploaded to OSS, the online migration service downloads the list file and migrates the files based on the specified URLs.
- When you create a migration job, the OSS URL of a list file is in the `oss://{bucket name}/{list file name}` format, such as `oss://mybucket/list.txt`.

Alibaba Cloud OSS

- Create a destination OSS bucket.

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

- Create a Resource Access Management (RAM) user and grant required permissions to the RAM user.
 - i. Log on to the [RAM console](#).
 - ii. In the left-side navigation pane, choose **Identities > Users**.
 - iii. On the **Users** page, click **Create User**.
 - iv. On the **Create User** page, specify Logon Name and Display Name in the User Account Information section.
 - v. In the Access Mode section, select Console Access and OpenAPI Access. Then, save the generated logon name, password, AccessKey ID, and AccessKey secret.
 - a. Console Access: If you select this option, you must configure the console password, password reset settings, and multi-factor authentication settings.
 - b. OpenAPI Access: If you select this option, an AccessKey pair is automatically created for the RAM user. The RAM user can call API operations or use other development tools to access Alibaba Cloud resources.
 - vi. After the RAM user is created, return to the **Users** page. Find the RAM user and click **Add Permissions** in the Actions column. In the panel that appears, select the AliyunOSSFullAccess and AliyunMGWFullAccess policies and click OK. This way, the RAM user is granted the permissions to read and write files from OSS and perform online data migrations.
 - vii. In the left-side navigation pane, click **Overview**.
 - viii. On the page that appears, navigate to the Account Management section and click the link under RAM user logon. On the page that appears, enter the logon name and password of the RAM user to log on to the Alibaba Cloud Management Console.

3. Migrate data

This topic describes how to migrate data by using Data Transport.

Precautions

When you perform online migration, take note of the following items:

-
-

Step 1: Create a source data address

- 1.
- 2.
3. In the **Create Data Address** panel, configure the following parameters and click **OK**.



Parameter	Required	Description
Data Type	Yes	Select Http/Https .
Data Name	Yes	Enter a name. The name must be 3 to 63 characters in length. Special characters are not supported, except for hyphens (-) and underscores (_).
File Path	Yes	Specify the file path in the format of <code>oss://{bucket}/{Name of a list file}</code> . For more information, see Create a list file .
List Access Endpoint	Yes	Enter a suitable endpoint based on the region where the source bucket resides. For more information, see Regions and endpoints .
List Access AK	Yes	Specify the AccessKey pair that is used to migrate data. The AccessKey pair is automatically generated for the RAM user in the preparation process. For more information about the preparation process, see Prepare for a migration job .
List Access SK	Yes	

- 4.
- 5.

Step 2: Create a destination data address

- 1.
2. In the **Create Data Address** panel, configure the following parameters and click **OK**.

Parameter	Required	Description
Data Type	Yes	Select OSS .


Parameter	Required	Description
Data Name	Yes	Enter a name. The name must be 3 to 63 characters in length. Special characters are not supported, except for hyphens (-) and underscores (_).
Data Region	Yes	Select the region where the destination data address is located.
Enable Transfer Acceleration	No	<p>For Data Online Migration to use the transfer acceleration feature provided by Object Storage Service (OSS), you must enable the transfer acceleration feature for buckets. The transfer acceleration feature takes effect within 30 minutes after it is enabled. After you enable the feature, wait for 30 minutes before you create a migration job.</p> <div>  Notice If you use the transfer acceleration feature to access your OSS buckets, you are charged additional transfer acceleration fees. For more information, see Transfer acceleration. </div>
OSS Endpoint	Yes	<p>Select an endpoint of the region where your data is stored. For more information, see Regions and endpoints.</p> <ul style="list-style-type: none"> ◦ http://oss-cn-endpoint.aliyuncs.com: specifies that you use an HTTP-based endpoint to access OSS over the Internet. ◦ https://oss-cn-endpoint.aliyuncs.com: specifies that you use an HTTPS-based endpoint to access OSS over the Internet. ◦ http://oss-cn-qingdao-internal.aliyuncs.com: specifies that you use a private HTTP-based endpoint to access OSS. ◦ https://oss-cn-qingdao-internal.aliyuncs.com: specifies that you use a private HTTPS-based endpoint to access OSS. <div>  Notice For example, a private OSS endpoint is used as the source data address when you create a migration job. In this case, the destination data address must be an Apsara File Storage NAS or OSS data address in the same region. </div>

Parameter	Required	Description
AccessKey Id	Yes	Specify the AccessKey pair that is used to migrate data. For more information, see Prepare for a migration job .
AccessKey Secret	Yes	
OSS Bucket	Yes	Select a bucket to store the migrated data.
OSS Prefix	No	The value cannot start with a forward slash (/) but must end with a forward slash (/). For example, you can specify <code>data/to/oss/</code> as the OSS prefix. Do not specify this parameter if you want to migrate data to the root directory of the specified bucket.

Step 3: Create a migration job


-
-
- In the **Job Config** step, configure the following parameters and click **Next**.

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

Parameter	Required	Description
Migration Type	Yes	<p>Full: The system runs a full migration job. After all data is migrated, the migration job ends. If the data at the source data address changes after the migration job is complete, you can submit another full migration job. In this case, the system migrates only the data that has changed since the last migration job.</p> <div> Notice Before you start a full migration job, the system compares files at the source data address with those at the destination data address. If a source file has the same name as a destination file, the destination file is overwritten if one of the following conditions is met:</div> <ul style="list-style-type: none">◦ The Content-Type property values of the source file and the destination file are different.◦ The last modified time of the source file is later than that of the destination file.◦ The size of the source file is different from that of the destination file.

Parameter	Required	Description
File Overwrite Method	Yes	<p>The method that is used to process the source file whose name is the same as that of a destination file. Valid values:</p> <ul style="list-style-type: none">◦ LastModified: If a source file has the same name as a destination file, the system compares the last modified time indicated by the LastModified property of the files.<ul style="list-style-type: none">■ If the last modified time of the source file is later than that of the destination file, the source file is migrated and the destination file is overwritten.■ If the last modified time of the source file is earlier than that of the destination file, the source file is not migrated.■ If the last modified time of the two files are the same, the system compares the Size and Content-Type property values of the files. If the two files have the same property values, the source file is not migrated. If the values of at least one property are different, the source file is migrated and the destination file is overwritten.◦ Condition: If a source file has the same name as a destination file, the system compares the values of the LastModified, Size, and Content-Type properties of the files.<ul style="list-style-type: none">■ If all property values of the two files are the same, the source file is not migrated.■ If the two files have different values of at least one property, the source file is migrated and the destination file is overwritten.◦ All: If a source file has the same name as a destination file, the system performs no comparison and directly overwrites the destination file with the source file.◦ No: If a source file has the same name as a destination file, the system performs no comparison and directly skips the source file during the migration.

4. In the **Performance** step, navigate to the **Data Prediction** section and configure the **Data Size** and **File Count** parameters.

 **Note** To ensure a successful migration, we recommend that you estimate the amount of data that the system will migrate as accurately as possible. For more information, see [Estimate the amount of data that you want to migrate](#).

5.

6.

Success Stories

For examples of migrating data from an HTTP or HTTPS address to an OSS bucket, see [Migrate data from a local IDC to OSS](#).

4. Manage migration jobs

This topic describes several subsequent operations that you can perform after you create a migration job.


View the status of a migration job

After you create a migration job, one of the following states is displayed for the job:

- **Migrating:** Data is being migrated. Wait until the migration job is completed.
- **Create Failed:** The migration job has failed to be created. You can view the cause of the failure and create another migration job.
- **Finished:** The migration job is completed. You can view the migration report.
- **Failed:** The migration job has failed. You can view the migration report and migrate the files that have failed to be migrated.

Modify the throttling settings

During migration, you can modify the throttling settings of a migration job based on your needs.

1. Log on to the [Data Transport console](#).
2. Choose **Data Online Migration > Migration Jobs**. On the migration job list, find a migration job and click **Manage**.
3. Click **Stop** and verify that the job is stopped.
4. On the Flow Control Time Schedule chart, click **Reset**.
 - To add a throttling setting, select an appropriate time range and maximum flow and then click **Add**.
 - To delete a throttling setting, click  next to the throttling setting.
 - To modify a throttling setting, delete the current setting and then add a new throttling setting.
5. Click **Submit** and then click **Start** to rerun the migration job.

View a migration report


1. On the migration job list, find the migration job and click **Manage**.
2. Click **Generate Migration Report**. After a report is generated, click **Export** to export the migration report.

In a migration report, the following files appear in the **File Lists** section:


- The file whose name ends with `_total_list`. This file lists all files to be migrated.
 - The file whose name ends with `_completed_list`. This file lists the files that have been migrated.
 - The file whose name ends with `_error_list`. This file lists the files that have failed to be migrated.
3. In the [Object Storage Service \(OSS\) console](#), find the automatically generated folder `aliyun_mgw_import_report/`. The folder includes the three files that appear in the migration report. You can download these files and view the detailed list that is included in each file. We recommend that you use the [ossbrowser](#) tool to view these files.

Each list consists of the following columns:

- The columns of the file whose name ends with `_total_list` include the source data address, file name, source version ID, and destination version ID. The columns also display whether the job version is the latest at the source data address and whether the job version is deleted at the source data address. The last two columns display the file size (measured in bytes) and the time when the file was last modified. The format of a source data address is `<vendor>://<bucketName>/<prefix>/<objectName>`. The following string provides an example: `oss://bucket-test1022/myprefix/testfile.txt`.
- The columns of the file whose name ends with `_completed_list` include the file name, the source version ID, and the destination version ID. The columns also display whether the job version is the latest at the source data address and whether the job version is deleted at the source data address. The last four columns display the file size (measured in bytes), CRC-64 checksum, time when the migration job was completed, and whether the migration was skipped.

 **Note** Skip migration: indicates whether the file that has the same name at the source data address and the destination data address has been skipped during migration. Valid values: 1 and 0. The value 1 indicates that the file has been skipped during migration. The value 0 indicates that the file has not been skipped during migration. The value of this field depends on the value of the **File Overwrite Method** parameter that you specify when you create the migration job.

- The columns of the file whose name ends with `_error_list` include the file name, the source version ID, and the destination version ID. The columns also display whether the job version is the latest at the source data address and whether the job version is deleted at the source data address. The last three columns display the time when the migration job started, the time when the migration job ended, and the error description.

 **Note** The following four fields apply only if multi-version migration is enabled:

- Source version ID: the version ID of the file at the source data address.
- Destination version ID: the version ID of the file at the destination data address.
- Latest at source data address: indicates whether the current version is the latest version of the file at the source data address. Valid values: true and false.
- Deleted at source data address: indicates whether the current version is deleted at the source data address. Valid values: true and false.

Rerun a failed migration job

If a migration job fails, you can check the generated file whose name ends with `_error_list` to identify the cause of the failure and resolve the issue. Then, on the migration job list, find the failed migration job, click **Manage**, and click **Retry** to migrate the failed files.