



# 对象存储数据迁移工具

用户使用指南

天翼云科技有限公司

## 目 录

1 介绍.....	3
2 工具使用流程.....	3
2.1 工具安装.....	3
2.2 修改配置文件.....	4
2.3 执行迁移.....	8
2.3.1 Linux 系统.....	8
2.3.2 Windows 系统.....	8
2.4 迁移进度查询.....	8
2.4.1 Linux 系统.....	9
2.4.2 Windows 系统.....	9
2.5 迁移结果查看.....	9
2.6 迁移日志查看.....	10
3 常见场景.....	10
3.1 全量/增量迁移.....	10
3.2 失败重试.....	10
3.3 断点续传.....	11
4 注意事项.....	11

# 1 介绍

对象存储（CT-ZOS，Zettabyte Object Storage）数据迁移工具是一款用于将存储在其他对象存储中的数据迁移至 ZOS 的产品。

ZOS 数据迁移工具支持的特性如下：

- 支持多个数据源，包括：亚马逊 S3、阿里云 OSS、华为云 OBS、腾讯云 COS、天翼云 ZOS、天翼云 OOS 和 XOS。
- 支持整桶文件迁移。
- 支持指定文件夹、文件、前缀进行文件迁移。
- 支持迁移源桶内指定时间区间内的文件。
- 支持记录迁移失败对象以及重试失败对象的迁移。
- 支持处理同名文件的迁移选项。
- 支持对象并行下载和上传。
- 支持断点续传。
- 支持实时查询迁移进度。
- 支持设置对象的存储类型和访问控制。
- 支持迁移过程中的一致性校验。

## 2 工具使用流程

### 2.1 工具安装

迁移工具运行环境要求如表 2.1、2.2 和 2.3 所示：

表 2.1 硬件要求

名称	版本	说明
CPU	8C	可按需调整
内存	16GB	可按需调整
系统盘大小	50GB	可按需调整

表 2.2 软件要求

名称	版本	说明
操作系统	Centos7.8/Kylin/Windows	—

表 2.3 网络要求

名称	说明
公网/内网	可公网部署亦可内网部署，迁移工具所在的服务器可访问源云存储资源池和 ZOS 资源池即可。
带宽	按需（和迁移速度正相关）

天翼云官网提供迁移工具压缩包：ZOS\_Migration\_Tool.zip，在上述运行环境确定后，下载并解压迁移工具压缩包，完成迁移工具的安装。

## 2.2 修改配置文件

迁移工具包 ZOS\_Migration\_Tool.zip 解压完成后，文件夹内存在配置文件 migration.conf，该配置文件包含字段如表 2.4。

表 2.4 migration.conf

参数名称	说明	类型	必填
srcType	迁移源类型，默认 S3 <ul style="list-style-type: none"><li>● S3：亚马逊、ZOS、XOS</li><li>● OSS：阿里云</li><li>● COS：腾讯云</li><li>● OBS：华为云</li><li>● OOS：天翼云经典版 I 型</li></ul>	string	是
srcUrl	源资源池 URL 地址	string	是
srcAccessKey	源资源池 AccessKey	string	是
srcSecretKey	源资源池 SecretKey	string	是
srcBucket	源资源池 Bucket	string	是
srcMigrationType	源资源池迁移模式，默认为 Bucket <ul style="list-style-type: none"><li>● Bucket：整桶迁移</li><li>● Folder：文件夹迁移</li><li>● Files：文件迁移</li><li>● Prefix：前缀迁移</li></ul>	string	否
srcMigrateFolder	指定源资源池迁移的文件夹列表，仅当 srcMigrationType 为 Folder 时有效	list	否

srcMigrateFiles	指定源资源池迁移的文件名列表，仅当 srcMigrationType 为 Files 时有效	list	否
srcMigratePrefix	指定源资源池迁移文件的前缀列表，仅当 srcMigrationType 为 Prefix 时有效	list	否
destUrl	目的资源池(ZOS)URL 地址	string	是
destAccessKey	目的资源池(ZOS)AccessKey	string	是
destSecretKey	目的资源池(ZOS)SecretKey	string	是
destBucket	目的资源池(ZOS)Bucket	string	是
enableConsistencyCheck	一致性校验开关，默认为 False <ul style="list-style-type: none"> <li>● True: 迁移过程开启一致性校验，校验不通过的对象会记录在 Possible_Inconsistent_Objects_xxx.txt 文件中</li> <li>● False: 迁移过程关闭一致性校验</li> </ul>	string	否
objectStorageClass	对象存储类型，默认适配源端对象的存储类型 <ul style="list-style-type: none"> <li>● STANDARD: 标准存储</li> <li>● STANDARD_IA: 低频存储</li> <li>● GLACIER: 归档存储</li> </ul>	string	否
objectAcl	对象访问权限，默认适配源端对象的访问权限 <ul style="list-style-type: none"> <li>● private: 私有</li> <li>● public-read: 公共读</li> </ul>	string	否
recordFailedObject	迁移或删除失败记录选项，默认为 False <ul style="list-style-type: none"> <li>● True: 记录迁移失败的对象和扫描失败的对象，形成记录文件放入目标存储桶中</li> <li>● False: 不会记录迁移失败和扫描失败的情况</li> </ul>	string	否
migrationAfterModified	设置起始时间，仅迁移 LastModified 大于或等于设置时间的对象，可以设置两种格式： <ul style="list-style-type: none"> <li>● "year-month-day", 默认时间为</li> </ul>	string	否

	<p>"1970-01-01"</p> <ul style="list-style-type: none"> <li>● "year-month-day hour:minute:second", 默认时间为 "1970-01-01 08:00:00"</li> </ul>		
migrationBeforeModified	<p>设置终止时间，仅迁移 LastModified 小于设置时间的对象，可以设置两种格式：</p> <ul style="list-style-type: none"> <li>● "year-month-day", 默认时间为当前时间+3650 Days</li> <li>● "year-month-day hour:minute:second", 默认时间为当前时间+3650 Days</li> </ul>	string	否
conflictMode	<p>同名文件处理选项，默认为 OVERWRITE</p> <ul style="list-style-type: none"> <li>● OVERWRITE: 同名文件进行覆盖</li> <li>● IGNORE: 同名文件进行忽略</li> <li>● COMPARE: 同名文件按最后修改时间比较，如果源 LastModified 小于目的 LastModified，则此文件被执行跳过；源 LastModified 大于目的 LastModified，则执行覆盖；源 LastModified 等于目的 LastModified，则继续进行判断，若二者文件 size 或 etag 有一不相等，则执行覆盖，否则，文件(size 和 etag 都相等)，文件将被执行跳过。</li> </ul>	string	否
logLevel	<p>设置运行时日志打印级别，默认为 ERROR</p> <ul style="list-style-type: none"> <li>● ERROR</li> <li>● INFO</li> <li>● DEBUG</li> </ul>	string	否

processNums	迁移任务的进程并发数量,默认为 12	int	否
threadNums	迁移任务的线程并发数量,默认为 16	int	否
failRetryMode	是否使用重试模式, 默认为 False。 为 True 时与 retryFailFiles 配合使用, 对 retryFailFiles 配置的文件列表进行重 试。	string	是
retryFailFiles	迁移过程中自动生成的错误日志列表, 只支持对 _ZOS_Migration_Failure_ 命名 的日志文件进行重试。	string	否

配置文件示例:

```
{
  "srcType": "S3",
  "srcUrl": "http://sichuan-2.zos.ctyun.cn",
  "srcAccessKey": "xxxxxxxxxxxxxxxx",
  "srcSecretKey": "xxxxxxxxxxxxxxxxxxxxxxxx",
  "srcBucket": "bucket-name",
  "srcMigrationType": "Bucket",
  "srcMigrateFolder": [],
  "srcMigrateFiles": [],
  "srcMigratePrefix": [],
  "destUrl": "https://jiangsu-10.zos.ctyun.cn",
  "destAccessKey": "xxxxxxxxxxxxxxxx",
  "destSecretKey": "xxxxxxxxxxxxxxxxxxxxxxxx",
  "destBucket": "bucket-name",
  "enableConsistencyCheck": "False",
  "objectStorageClass": "",
  "objectAcl": "",
  "recordFailedObject": "True",
  "migrationAfterModified": "",
  "migrationBeforeModified": "",
  "conflictMode": "IGNORE",
  "logLevel": "DEBUG",
  "processNums": 12,
  "threadNums": 16,
```

```
"failRetryMode": "False",  
  "retryFailFiles": []  
}
```

## 2.3 执行迁移

### 2.3.1 Linux 系统

在 Linux 环境使用迁移工具前，需要先赋予工具可执行权限，执行如下指令：

```
chmod +x zsync
```

上述操作建议在管理员权限下执行。

在 Linux 平台，在配置文件 migration.conf 填写完毕后，执行指令：

```
./zsync start
```

开始执行迁移。

为避免关闭终端导致 zsync 进程中断，可使用指令：

```
nohup ./zsync start &
```

将 zsync 进程放在后台执行，不受终端关闭影响。

### 2.3.2 Windows 系统

在 Windows 平台，在配置文件 migration.conf 填写完毕后，cmd 执行指令：

```
zsync start
```

开始执行迁移。

## 2.4 迁移进度查询

表 2.5 迁移进度参数

参数	说明
Query time	查询用时，单位秒
Migration time	迁移的总用时，单位秒
Scanned nums	源桶扫描的对象个数
Ignored nums	迁移忽略的对象个数
Finished nums	迁移完成的对象个数



Failed nums	迁移失败的对象个数
Download rate	当前下载速率
Upload rate	当前上传速率

## 2.4.1 Linux 系统

迁移过程中，执行如下指令：

```
./zsync info
```

查询实时的迁移进度。

## 2.4.2 Windows 系统

迁移过程中，cmd 执行指令：

```
zsync info
```

即可查询实时的迁移进度。

## 2.5 迁移结果查看

迁移完成后，与 zsync 可执行文件同级目录下会生成 migration\_results.txt，即迁移结果文件。该文件记录了 zsync 工具执行迁移的结果信息，包括：ZOS 桶名、迁移起止时间、源桶扫描对象个数、迁移忽略个数、迁移完成个数和迁移失败个数等信息。

表 2.6 migration\_results.txt

参数	说明
bucket	目的桶名
migration time	迁移起止时间
scanned nums	源桶扫描的对象个数
ignored nums	迁移忽略的对象个数
finished nums	迁移完成的对象个数
failed nums	迁移失败的对象个数

## 2.6 迁移日志查看

Windows 平台迁移产生的日志存储在 `zsync.exe` 同级 `log/` 目录下；Linux 平台迁移产生的日志存储在 `/var/log/zsync/`。根据实际需要可在 `migration.conf` 中配置 `debug`、`info` 和 `error` 等不同的日志级别。

# 3 常见场景

## 3.1 全量/增量迁移

全量迁移：通常指将源端对象存储桶中的数据全部迁移到天翼云 ZOS 桶中。

增量迁移：在全量迁移后，继续将新增的数据进行迁移。

用户想配置全量迁移或增量迁移任务时，通过修改配置文件 `migration.conf` 中的 `conflictMode` 字段来实现。当 `conflictMode` 字段参数配置为 `OVERWRITE`，则对同名文件进行覆盖，即实现全量迁移；当 `conflictMode` 字段参数配置为 `IGNORE`，则不对同名文件进行迁移，实现增量迁移。

**注意：**若您迁移的目的桶存有存量数据，配置为 `OVERWRITE` 进行全量迁移可能会导致目的桶存量数据被覆盖，造成数据丢失。

```
1 {
2   "srcType": "S3",
3   "srcUrl": "http://sichuan-2.zos.ctyun.cn",
4   "srcAccessKey": "xxxxxxxxxxxx",
5   "srcSecretKey": "xxxxxxxxxxxx",
6   "srcBucket": "bucket-name",
7   "srcMigrationType": "Bucket",
8   "srcMigrateFolder": [],
9   "srcMigrateFiles": [],
10  "srcMigratePrefix": [],
11  "destUrl": "http://jiangsu-10.zos.ctyun.cn",
12  "destAccessKey": "xxxxxxxxxxxx",
13  "destSecretKey": "xxxxxxxxxxxx",
14  "destBucket": "bucket-name",
15  "enableConsistencyCheck": "False",
16  "objectStorageClass": "",
17  "objectAcl": "",
18  "recordFailedObject": "True",
19  "migrationAfterModified": "",
20  "migrationBeforeModified": "",
21  "conflictMode": "IGNORE",
22  "LogLevel": "DEBUG",
23  "processNums": 12,
24  "threadNums": 16,
25  "failRetryMode": "False",
26  "retryFailFiles": []
27 }
```

## 3.2 失败重试

用户若想对迁移任务中迁移失败的文件开启失败重试，需要先对失败文件进行记录，再开启失败重试。

### 1.配置失败文件记录:

可以通过修改配置文件 `migration.conf` 中 `recordFailedObject` 字段进行配置。`True` 表示开启失败记录，在扫描完源存储桶和目的存储桶中所有对象以及迁移过程执行完成后，若出现错误，则在本地和目的存储桶中会存在相应记录错误的文件；`False` 表示不打开失败记录，工具不会记录迁移失败和扫描失败的情况。

### 2.配置失败文件重试:

进行失败文件的重试需要用户配置 `failRetryMode` 和 `retryFailFiles` 两个字段。失败文件的记录保存在文件 `srcType + "_ZOS_Migration_Failure_" + 时间 + ".txt"` 中，该文件记录了迁移过程中迁移失败的对象名、对应版本号以及迁移错误原因；用户可以将 `failRetryMode` 字段配置为 `True`，将错误文件的名称填写到配置文件的 `retryFailFiles` 字段，重新运行工具即可重试迁移失败的对象。

## 3.3 断点续传

工具迁移过程发生中断（如机器重启，手动 `kill` 工具，扫描桶列表失败），重启工具即可按中断前迁移进度继续传输。

**注意：**若配置文件 `migration.conf` 中关键信息修改（源桶相关信息、目的桶相关信息、存储类型、访问控制、迁移时间范围），则重启工具不支持断点续传。

## 4 注意事项

- 1) 暂不支持多版本迁移。
- 2) 最大支持迁移单个对象为 48.8GB。
- 3) 迁移过程中，如果出现中断，重新启动工具前，可以置 `conflictMode` 参数为 `IGNORE`，已迁移过的同名文件会进行忽略，节省迁移时间。
- 4) 迁移工具并发数调节：与服务器性能相关，通常配置文件 `migration.conf` 进程数设置不超过服务器 CPU 的核心数。
- 5) 迁移速度：迁移速度和源端网络带宽、服务器网络带宽、文件大小均有关。对于小文件，由于 `IOPS` 较高的原因，会导致迁移速度较低。
- 6) 一致性校验：仅校验迁移前后对象的 `lastmodified` 和 `size`，无法保证迁移结果的正确性，故在迁移任务完成后，请您务必自行做好源端和目的端数据的正确性核验。同时，开启一致性校验会导致迁移效率降低，以及 `API` 调用次数增加。

7) 增量迁移无法实现动态的增量，需要源桶和目的桶进行对比从而获得增量数据。

8) 各个源端云厂商对于同步迁移对象元数据的支持情况

云厂商	存储级别	对象标签	访问控制	自定义元数据
S3	√	√	√	√
OSS	√	√	√	√
COS	√	√	√	√
OBS	√	×	√	√
OOS	√	×	√	×

√ 表示支持，× 表示不支持

9) 访问控制权限，源端对象迁移至 ZOS 目的端的情况

云厂商	源端对象访问控制	(ZOS)目的端对象访问控制
S3	public-read	public-read
	public-read-write	public-read-write
	private 或其他	private
OSS	public-read	public-read
	public-read-write	public-read-write
	private 或其他	private
COS	public-read	public-read
	public-read-write	public-read-write
	private 或其他	private
OBS	public-read	public-read
	public-read-write	public-read-write
	private 或其他	private
OOS	public-read	public-read
	public-read-write	public-read-write
	private 或其他	private

10) 存储级别方面，源端对象迁移至 ZOS 目的端的情况

云厂商	源端对象存储级别	(ZOS)目的端对象存储级别
S3	STANDARD	STANDARD
	STANDARD_IA	STANDARD_IA
OSS	STANDARD	STANDARD
	STANDARD_IA	STANDARD_IA

COS	STARDARD	STARDARD
	STARDARD_IA	STARDARD_IA
OBS	STARDARD	STARDARD
	STARDARD_IA	STARDARD_IA
OOS	STARDARD	STARDARD
	STARDARD_IA	STARDARD_IA

- 11) COS 作为源端时，直接使用 ap-beijing 类型的参数作为访问地址，而其他云厂商使用 https://<endpoint>作为访问地址。
- 12) OBS 在指定前缀迁移时，前缀为空格，会全部迁移。