



# 文件上云迁移工具

用户使用指南

天翼云科技有限公司

---

## 目录

1. 简介 .....	3
1.1 术语和缩略语.....	3
1.2 功能介绍.....	3
1.3 使用环境.....	4
2. 安装与部署 .....	5
2.1 硬件需求.....	5
2.2 软件需求.....	5
2.3 网络需求.....	5
2.4 安装步骤.....	6
2.4.1 Linux.....	6
3. 使用方法 .....	7
3.1 配置.....	7
3.1.1 本地上云.....	8
3.2 执行.....	9
3.3 查看日志.....	10

# 1.简介

## 1.1 术语和缩略语

术语	缩写	释义
文件上云迁移工具	CDM	文件上云迁移工具，主要用于文件上云，以及文件到对象存储的迁移；支持全量迁移、增量迁移、迁移校验等功能。
对象存储	ZOS	对象存储（原生版）I型（简称ZOS）是天翼云为客户提供的一种海量、弹性、高可靠、高性价比的存储产品，是专门针对云计算、大数据和非结构化数据的海量存储形态，通过S3协议和标准的服务接口，提供非结构化数据（图片、音视频、文本等格式文件）的无限存储服务

## 1.2 功能介绍

文件迁移工具结合 ZOS 使用，主要提供将文件传输到云上 ZOS 以及将云上 ZOS 中的对象文件下载到本地指定文件夹中两种功能。

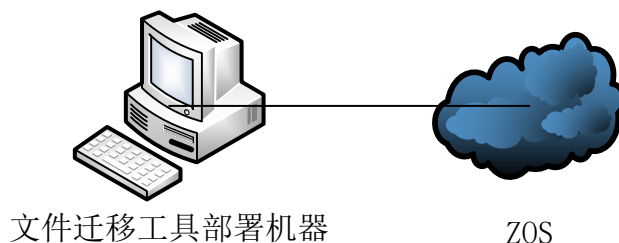
### 简介：

- 支持断点续传。迁移过程中，如果出现中断，重新启动工具后，可以继续执行迁移工作。已经迁移完成的部分数据，将会进行比对，如果一致则不进行迁移动作。

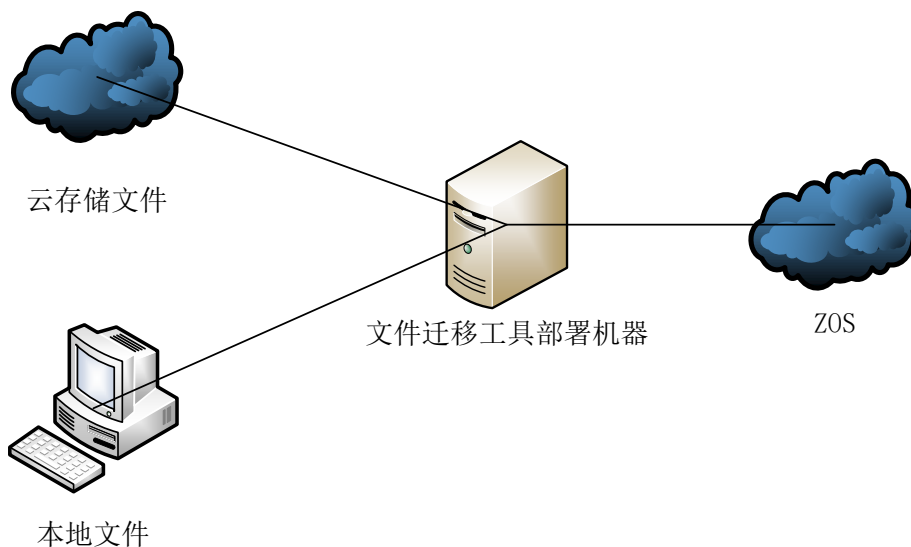
- 支持指定文件夹进行文件迁移；
- 文件上云后可以指定对应的前缀进行存储；

## 常见用法:

1. 将文件迁移工具部署在需要上云的文件所在的机器上,使用文件迁移工具直接上传到 ZOS。



2. 需要上云的文件（可以是其他机器的文件，也可以是云存储上的文件）先通过挂载等方式连接到部署了文件迁移工具的机器上，机器可以连通 ZOS，迁移工具可以通过机器上的挂载点，将文件上传到 ZOS



## 1.3 使用环境

迁移工具目前支持部署在本地设备端，会占用部分设备资源；

## 3.安装与部署

### 2.1 硬件需求

名称	版本	说明
CPU	2C	该参数为推荐配置
内存	8GB	该参数为推荐配置
系统盘大小	50GB	该参数为推荐配置

### 2.2 软件需求

名称	版本	说明
操作系统	Centos	-
python3	3.6+	Python 运行环境
ZOS_PythonSDK	1.0+	对接 ZOS

### 2.3 网络需求

名称	说明
公网	可公网部署亦可内网部署,看是否存在迁移公网文件/对象的需求。

带宽	按需（和迁移速度正相关）
内网带宽	按需（和内网迁移速度正相关）

## 2.4 安装步骤

### 2.4.1 Linux

首先提供 Ctyun-data-migration.tar.gz 压缩文件，在确认前置软件环境已经安装完成的情况下，执行压缩包中的 install.sh，会自动完成所有的安装动作；在执行相关操作时建议使用 root 权限，包括后续在迁移工具的使用过程中也建议使用 root 用户进行。

```
sh install.sh
```

```
[root@cstor-test home]#  
[root@cstor-test home]# ls  
Ctyun-data-migration  Ctyun-data-migration-1.0-0.tar.gz  
[root@cstor-test home]#  
[root@cstor-test home]#  
[root@cstor-test home]#  
[root@cstor-test home]#  
[root@cstor-test home]#  
[root@cstor-test Ctyun-data-migration]#  
[root@cstor-test Ctyun-data-migration]# sh install.sh  
WARNING: Running pip install with root privileges is generally not a good idea. Try  
Processing ./urllib3-1.26.4-py2.py3-none-any.whl  
Processing ./six-1.16.0-py2.py3-none-any.whl  
Processing ./s3transfer-0.4.2-py2.py3-none-any.whl  
Processing ./python_dateutil-2.8.1-py2.py3-none-any.whl  
Processing ./jmespath-0.10.0-py2.py3-none-any.whl  
Processing ./botocore-1.20.73-py2.py3-none-any.whl  
Processing ./boto3-1.17.73-py2.py3-none-any.whl  
Installing collected packages: urllib3, six, python-dateutil, jmespath, botocore, s  
Successfully installed boto3-1.17.73 botocore-1.20.73 jmespath-0.10.0 python-dateuti  
.4  
Preparing... ##### [100%]  
Updating / installing...  
  1:rc1one-1.62.2-1 ##### [100%]  
Preparing... ##### [100%]  
Updating / installing...  
  1:Ctyun-Migration-1.0-0 ##### [100%]  
[root@cstor-test Ctyun-data-migration]#
```

卸载时可以使用 uninstall.sh 完成卸载：

```
[root@ctestor-test Ctyun-data-migration]#
[root@ctestor-test Ctyun-data-migration]# sh uninstall.sh
Uninstalling urllib3-1.26.4:
  Successfully uninstalled urllib3-1.26.4
Uninstalling six-1.16.0:
  Successfully uninstalled six-1.16.0
Uninstalling s3transfer-0.4.2:
  Successfully uninstalled s3transfer-0.4.2
Uninstalling python-dateutil-2.8.1:
  Successfully uninstalled python-dateutil-2.8.1
Uninstalling jmespath-0.10.0:
  Successfully uninstalled jmespath-0.10.0
Uninstalling botocore-1.20.73:
  Successfully uninstalled botocore-1.20.73
Uninstalling boto3-1.17.73:
  Successfully uninstalled boto3-1.17.73
error: Failed dependencies:
  rclone >= 1.61.0 is needed by (installed) Ctyun-Migration-1.0-0.noarch
Preparing... ##### [100%]
Cleaning up / removing...
  1:Ctyun-Migration-1.0-0 ##### [100%]
[root@ctestor-test Ctyun-data-migration]#
```

## 4.使用方法

### 3.1 配置

#### 3.1.1 基础配置信息

安装完成后，在系统的 /root/.config/rclone/ 目录下存在配置文件

migrations.conf，在配置文件中需要包含以下内容：

参数名称	释义
[task1]	section 名称，每个任务一个，当前是固定为 task1、task2、task3、task4
enabled	当前任务是否启用：true 和 false
type	once 【当前版本固定】
name	任务名称，主要用于在日志中进行隔离（限制字符（数字、大小写字母、'-','_')）
access_key	ZOS 对应账户的 access_id
secret_key	ZOS 对应账户的 secret_key
endpoint	ZOS 的访问地址

src_path	本地准备上云的文件所在的路径
dest_bucket	云上 ZOS 的桶名称
dest_prefix	预期将文件存储在云上桶中的位置，如果为空就是以桶为根目录（限制字符（数字、大小写字母、'-'、'_'））
reverse	是否反向传输，本地上云应为 false 或空，对于从云上迁移到本地情况，填写为 true

### 3.1.2 本地上云

在[src\_path]中需要填写准备上云的本地目录，在[dest\_bucket]中需要填写云端的桶名 bucket，以及预期对这个上云目录名称前缀[dest\_prefix]（可选，如果不填写 prefix 将会以桶作为迁移的目的目录）；

```
1 [task1]
2 enabled=false
3 type=once
4 name=task1
5 access_key=123
6 secret_key=456
7 endpoint=https://192.168.0.1
8 src_path=/mnt/test
9 dest_bucket=bc-test
10 dest_prefix=task1
11 reverse=false
```

### 3.1.3 云上迁移到本地

云上迁移到本地的操作与本地上云的配置类似，只需要将 migrate.conf 中



的 `reverse` 参数配置为 `true`。

示例：

```
1  [task1]
2  enabled=false
3  type=once
4  name=task1
5  access_key=123
6  secret_key=456
7  endpoint=https://192.168.0.1
8  src_path=/mnt/test
9  dest_bucket=bc-test
10 dest_prefix=task1
11 reverse=true
```

注意，此时 `migrate.conf` 中 `src_path` 指的是将云上文件迁移到本地后存放的文件夹，如果其不存在，将会默认创建。

## 3.2 执行

手动执行迁移动作，首先需要完成配置，然后执行对应的命令行 `migration`；如果需要切换任务目录或者是 AKSK 等信息，需要修改对应的配置文件然后重新执行 `cli`：

示例：

```
[root@cstor-test Ctyun-data-migration]# migration
There is no one task enabled!
[root@cstor-test Ctyun-data-migration]#
[root@cstor-test Ctyun-data-migration]#
```

在执行迁移过程中，根据预期的任务内容进行粗粒度的判断，如果迁移的时间可能很长的话，建议搭配 `nohup` 与 `&` 后台执行 `migration` 指令：

```
nohup migration &
```

因为如果长时间执行的话，可能会因为命令执行窗口异常退出或者是关闭导致 `migration` 进程被终止。

如果任务正常执行完成，将会打印对应的任务名称以及执行所消耗的时间：

```
[root@localhost rclone]#  
[root@localhost rclone]#  
[root@localhost rclone]# migration  
[task567]work Done! [2.0314779959999214]  
[root@localhost rclone]#  
[root@localhost rclone]#  
[root@localhost rclone]#
```

### 3.3 查看任务运行状态及错误日志

在迁移工具中提供了命令 `migtool` 用于查询任务进度以及执行异常的异常日志信息：

```
migtool -h  
Please check if the input parameters are standardized!  
Usage:  
eg:  
    migtool status [taskname] # Obtain the progress and status of the  
specified task.  
    migtool errout [taskname] # Obtain error information for the  
specified task.
```

其中 `migtool status [taskname]` 提供查询指定任务当前执行进度的能力（不包括任务执行是否成功的结果）：

```
[root@localhost Ctyun-migration]# migtool status task567  
status  
task567  
get status!  
Transferred:      1000 KiB / 1000 KiB, 100%, 0 B/s, ETA -  
Transferred:      10 / 10, 100%  
Elapsed time:     0.3s  
[root@localhost Ctyun-migration]#
```

`migtool errout [taskname]`将查询到的错误日志打印到屏幕上：

```
[root@localhost Ctyun-migration]# migtool errout task567
errout
task567
get errlog!
2023/07/28 17:22:36 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:23:41 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:23:54 ERROR : : file_999: Failed to copy: expected element type <Error> but have <html>
2023/07/28 17:23:54 ERROR : : Attempt 1/3 failed with 1 errors and: expected element type <Error> but have <html>
2023/07/28 17:23:54 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:23:59 ERROR : : Attempt 2/3 succeeded
2023/07/28 17:24:40 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:27:42 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:28:18 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:43:35 ERROR : : file_549: Failed to copy: expected element type <Error> but have <html>
2023/07/28 17:43:39 ERROR : : Attempt 1/3 failed with 1 errors and: expected element type <Error> but have <html>
2023/07/28 17:43:39 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:43:44 ERROR : : Attempt 2/3 succeeded
2023/07/28 17:44:05 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/28 17:46:05 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/31 16:01:05 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/31 16:01:45 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/31 17:34:50 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
2023/07/31 17:35:09 ERROR : : Entry doesn't belong in directory "" (same as directory) - ignoring
[root@localhost Ctyun-migration]#
```

提示：部分情况下错误日志会过多，可以重定向输出到指定的文件中再查看。

```
[root@localhost Ctyun-migration]# migtool status task5
[task5] is not a task in the current configuration!
Usage:
eg:
  migtool status [taskname] # Obtain the progress and status of the specified task.
  migtool errout [taskname] # Obtain error information for the specified task.
```

当输入的 taskname 不存在，或者是 taskname 之前没有执行过，将会产生对应的错误提示信息：

```
[root@localhost ctyun_migration]# migtool status task5
[task5] is not a task in the current configuration!
Usage:
eg:
  migtool status [taskname] # Obtain the progress and status of the specified task.
  migtool errout [taskname] # Obtain error information for the specified task.
[root@localhost ctyun_migration]#

[root@localhost ctyun_migration]# migtool errout task4
[task4]The specified task has not been initiated yet!
Usage:
eg:
  migtool status [taskname] # Obtain the progress and status of the specified task.
  migtool errout [taskname] # Obtain error information for the specified task.
[root@localhost ctyun_migration]#
```

任务执行的日志信息保存在/var/log/ctyun-migration/migration.log 中，日志信息中可以查看整体任务的执行结果，执行的速度，任务消耗的时间等信息。

### 3.4 异常信息

迁移任务开始前，系统会校验配置文件 migrate.conf 中的信息是否有误。不同的参数配置将产生不同的异常信息。具体包含如下几个参数。

#### 3.4.1 src\_path

该异常信息是有 src\_path 参数引起的。在本地上云阶段，src\_path 所代表的

是本地路径，如果该路径不存在，则会产生错误信息：

```
Error MESSAGE: The src_path is not exist. ---task name:
[taskname]
ADVISE: Please check the src_path of [taskname]
```

在从云上迁移到本地的任务类型中，src\_path 指的是将云上文件迁移到本地后存放的文件夹，因此工具将不会对该路径进行检测，如果指定的路径不存在，将会默认创建。

### 3.4.2 end\_point

end\_point 作为云端服务器地址，如果其配置错误，任务将会连接超时，超时时间为 1 分钟，并产生异常信息。在本地上云与云上迁移至本地两种任务类型的执行过程中出现异常都会产生错误信息。

```
ERROR MESSAGE: Could not connect to the endpoint
URL:http://***** ---task name: [taskname]
ADVISE: Please check the endpoint of [taskname]
```

### 3.4.3 access\_key 与 secret\_key

access\_key 与 secret\_key 作为连接云端的密钥，其中一个出现错误都会产生异常信息。作为建立连接的基础，该异常信息在本地上云与云上迁移至本地两个阶段都有可能出现。

```
ERROR MESSAGE: 403 Forbidden when calling the HeadBucket
operation. ---task name: [taskname]
ADVISE: Please check the access_key or secret_key of
[taskname]
```

### 3.4.4 dest\_bucket

上述错误信息的产生是由于 dest\_bucket 也就是云端的目标桶不存在产生的。

```
ERROR MESSAGE: 404 Not Found when calling the HeadBucket
operation. ---task name: [taskname]
```

```
ADVISE: Please check whether dest_bucket exists of [taskname]
```

### 3.4.5 dest\_prefix

对于 dest\_prefix 存在限制字符范围（数字、大小写字母、'-'、'\_'），如果出现范围外的字符，将会产生错误信息。如果 prefix 路径中存在前置"/"，在实际传输过程中，工具会将前置的"/"删除掉：

```
ERROR MESSAGE: Illegal characters in dest_prefix. May only contains numbers, letters and '-', '_'. ---task name: [taskname]
ADVISE: Please check the dest_prefix of [taskname]
```

如果出现部分特殊字符（例如%，如果不熟悉 configparser 的配置文件语法）可能导致进程 crash。

### 3.4.6 迁移进程异常中止

迁移过程中，某些情况下，由于系统资源不足或者误操作，或者是参数错误等原因，会导致迁移文件的进程 rclone 异常退出，这个时候在屏幕上会打印导致对应退出的可能原因，日志也会同时记录到 migration.log 中，可以方便后续提供给天翼云的技术人员进行相关问题分析。

```
[root@localhost ctyun_migration]# migration
[child][task567][ret=-9] RC may have been destroyed!
[root@localhost ctyun_migration]#

[root@localhost ctyun-migration]# tail migration.log
2023-08-02 10:47:03,922 - root - INFO - /root/.config/rclone/migrations.conf
2023-08-02 10:47:03,923 - root - INFO - /root/.config/rclone/last_task.conf
2023-08-02 10:47:03,924 - root - INFO - find the section
2023-08-02 10:47:03,957 - root - INFO - [Main]task:task567 is enabled, now start work!
2023-08-02 10:47:04,198 - root - INFO - now start a rclone task:task567
2023-08-02 10:47:04,199 - root - DEBUG - [task567]command is:rclone --checkers 8 --transfers 4 --progress --stats=2s --log-file /var/log/ctyun-migration/rclone_task567.log copy /mnt/test task567://rctest/tmp
2023-08-02 10:47:04,220 - root - INFO - [task567][0.020385920999615337]num of dirs is:1
2023-08-02 10:47:04,582 - root - INFO - [task567][0.36196246899999096]Dircopy Done!T2:5592.199908787 T3:5592.561871256
2023-08-02 10:47:12,219 - root - ERROR - [child][task567][ret=-9] RC may have been destroyed!
2023-08-02 10:47:12,220 - root - DEBUG - [task567]work Done![8.020248799]
```