对象存储数据迁移工具使用手册

文档变更记录

序号	变更 (+/-) 说明	作者	版本号	日期	批准
1	+	孟祥瑞	V1.0	2021/09/21	
2	+	李卓	V1.1	2023/01/04	

1、介绍

ZOS 数据迁移工具是一款将数据迁移至 ZOS 对象存储的工具。您可以在本地服务器使用迁移工具，将其它云存储包括亚马逊、阿里云、腾讯云、华为云、天翼云 OOS 和 XOS 的数据迁移到 ZOS 进行存储。

ZOS 数据迁移工具具有以下特点：

- 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，根据 `conflictMode` 参数设置进行覆盖、忽略或按修改时间比较。
- 支持整桶文件迁移
- 支持指定前缀进行文件迁移
- 支持指定文件夹进行文件迁移
- 支持指定某些文件进行迁移
- 支持对象并行下载和上传
- 支持迁移文件多版本
- 支持迁移失败对象的记录
- 支持指定时间来选择迁移的文件
- 支持处理同名文件的迁移选项

2、运行环境

迁移工具支持在 Windows 或 Linux 上运行，运行环境要求如下：

- python3 运行环境，适用于 python3.3 及其以上版本
- 迁移工具所在的服务器可以访问源云存储资源池和 ZOS 资源池

3、迁移步骤

迁移需要的步骤：

- 1) 迁移准备；
- 2) 修改配置文件；
- 3) 执行迁移

3.1、迁移准备

拥有 ZOS 账户，获取 ZOS 的 AccessKey 和 SecretKey，且在 ZOS 中存在目标存储桶 Bucket

3.2、修改配置文件

名称	描述	类型	是否必须
srcType	迁移源类型 <ul style="list-style-type: none">● S3: 亚马逊(默认)● OSS: 阿里云● COS: 腾讯云● OBS: 华为云● OOS: 天翼云经典版 I 型● XOS: 天翼云融合版	string	否
srcUrl	源资源池 URL 地址	string	是
srcAccessKey	源资源池账户 AccessKey	string	是
srcSecretKey	源资源池账户 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">● Bucket: 整桶迁移● Folder: 文件夹迁移● Files: 文件迁移● Prefix: 前缀迁移	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
MigratePrefix	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)账户 AccessKey	string	是
destSecretKey	目的资源池(ZOS)账户 SecretKey	string	是
destBucket	目的资源池(ZOS)Bucket	string	是
SyncObjectVersioning	文件多版本迁移开关，默认为 False <ul style="list-style-type: none"> ● True: 若源资源池存储桶开启了多版本功能，则在目的资源池的存储桶会开启多版本功能，迁移源存储桶中所有对象版本至目的存储桶，若源存储桶未开启多版本，则只迁移最新版本。 ● False: 无论源资源池存储桶是否开启了多版本功能，则只迁移最新版本。 	string	否
RecordFailedObject	迁移或删除失败记录选项，默认为 False <ul style="list-style-type: none"> ● 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。 ● False: 不会记录迁移失败和扫描失败的情况。 	string	否
MigrationLastModified	设置起始时间，仅迁移 LastModified 大于设置时间的对象，可以设置两种格式： <ul style="list-style-type: none"> ● "year-month-day", 默认时间为"1970-01-01" ● "year-month-day hour:minute:second", 默认时间为"1970-01-01 08:00:00" 	string	否
ConflictMode	同名文件处理选项，默认为 OVERWRITE。当	string	否

	<p>SyncObjectVersioning 配置项为 True 时，忽略该配置，直接进行同名覆盖</p> <ul style="list-style-type: none"> ● OVERWRITE: 同名文件进行覆盖 ● IGNORE: 同名文件进行忽略 ● 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 有其一不相等，则执行覆盖，否则，文件(size 和 etag 都相等)，文件将被执行跳过。 		
logLevel	<p>设置运行时日志打印级别，默认为 ERROR。</p> <ul style="list-style-type: none"> ● ERROR ● INFO ● DEBUG 	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

配置文件示例：

```
{
  "srcType":"OSS",
  "srcUrl":"https://oss-cn-beijing.aliyuncs.com",
  "srcAccessKey":"LTAI5t5xkw92wAynxgaFUKjX",
```

```
"srcSecretKey":"vvhfpDefDzPSQW7Eh6sPDSKD15WyZe",
"srcBucket":"oss-migration-bucket",
"srcMigrationType":"Bucket",
"srcMigrateFolder":["first_dir", "first_direct", "oss_dir"],
"srcMigrateFiles":["hebei.txt", "sichuan.txt", "gansu.txt"],
"srcMigratePrefix":["h"],
"destUrl":"http://192.168.56.103:7480",
"destAccessKey":"RW0VXS7X5EXTS3E0J0YN",
"destSecretKey":"AoG3J5N1b5d0wHEAZVb32Xk8SEpDxujOOUBm0Mku",
"destBucket":"zos-migration-bucket ",
"syncObjectVersioning":"False",
"recordFailedObject":"True",
"migrationLastModified":"2021-09-19",
"conflictMode":"COMPARE",
"logLevel":"INFO",
"processNums":12,
"threadNums":16
}
```

3.3、执行迁移

在 Windows 平台，运行文件 zsync.exe 即可；在 Linux 平台，运行 ./zsync 可执行文件。

3.4、日志

使用数据迁移工具，若 recordFailedObject 开关打开，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，则在目的存储桶中会存在相应记录错误的文件。

- 1) 文件 srcType + "_ZOS_Migration_Failure_" + 时间 + ".txt"，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；
- 2) 文件 srcType + "_ZOS_Delete_Failure_" + 时间 + ".txt"，该文件仅在开启迁移文件多版本(syncObjectVersioning)时才有意义，该文件记录了删除过程中删除失败的文件名、对应版本号以及删除失败原因；
- 3) 文件 srcType + "_ZOS_Scan_Failure_" + 时间 + ".txt"，该文件记录了在开始

迁移文件前，扫描源存储桶和目的存储桶的过程中出现的错误。

运行日志生成在 logs/migration.log 文件中，根据实际需要可以配置 debug、info 和 error 日志级别。

4、注意事项

- 1) 在使用同步源资源池多版本对象功能时，建议同步到 ZOS 资源池的新存储桶中。若选择同步到 ZOS 资源池中已有数据的存储桶，则会对存储桶中的原始数据先进行清理再形成覆盖，可能会造成原有数据丢失。
- 2) 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，此参数适用于未开启多版本对象迁移功能的场景。若开启多版本对象迁移功能，中断重启会先清理桶中已迁数据，再从头开始迁移各个对象及其所有版本。
- 3) 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	多版本	访问控制	自定义元数据
S3	√	√	√	√	√
OSS	√	√	√	√	√
COS	√	√	√	√	√
OBS	√	×	√	√	√
OOS	√	×	×	√	×

√ 表示支持，× 表示不支持

- 4) 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

5) 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 或其他级别	STANDARD
OSS	STANDARD 或其他级别	STANDARD
COS	STANDARD 或其他级别	STANDARD
OBS	STANDARD 或其他级别	STANDARD
OOS	STANDARD 或其他级别	STANDARD

- 6) 除 COS 外，其余源厂商在使用迁移对象多版本功能时，若某个对象版本存在 deletemarker，且无论该 deletemarker 是否为最新版本，则依旧会迁移该对象及其历史版本，且在目的存储桶中可显示存在。在源端为 COS，且 deletemarker 是最新版本时，该对象不再迁移至目的端。当存在 deletemarker 且不是最新版本时，迁移该对象所有版本至目的端。
- 7) COS 作为源端时，直接使用 ap-beijing 作为访问地址，而其他云厂商使用 https://<endpoint>作为访问地址。
- 8) OBS 在指定前缀迁移时，前缀为空格，会全部迁移。