

Cloud Object Storage

Developer Tools

Product Introduction



Tencent  
Cloud

## Copyright Notice

©2013-2017 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

## Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

## Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Documentation Legal Notice ..... 2

Developer Tools ..... 4

    Migration Tool..... 4

    Local Synchronization Tool ..... 10

    Hadoop Tool ..... 16

Developer Tools

## Migration Tool

### Feature Description

This migration tool is used to migrate files from services such as AWS S3, Aliyun OSS, Qiniu, to COS. It can also migrate file lists, which means migrating files to COS based on given URLs. Only applicable to COS 4.0

### Operating Environment

#### System Environment

Linux, Mac OS

#### Required Software

Operation platform for this tool is \*nix. An environment of Python2.6 or above is required, and pip, gcc, python-dev should be installed on the machine. You may use the package manager (included in the system) to install relevant depended resources.

In centos, use the following command to install:

```
sudo yum install python-pip python-devel gcc
```

In ubuntu/debian, use the following command to install:

```
sudo apt-get install python-pip python-dev gcc
```

### How to Use

## Acquire the Tool

Github project: [https://github.com/tencentyun/cos\\_migrate\\_tool](https://github.com/tencentyun/cos_migrate_tool)

## How to Install

It is recommended to use pip for installation. Refer to the [Official Site](#) to learn about how to install pip, or use package managers (such as apt, yum) to install python-pip package.

```
pip install -U cos_migrate_tool
```

After executing the command above, you can check whether the installation is successful by using the following command.

```
cos_migrate_tool -h
```

## Configuration File

Configure file template. "common" section is for basic configurations, "workspace" is the workspace directory mentioned above. "source" section is for configuring data source information. Configure oss attributes if you wish to migrate oss to cos. "destination" section is for configuring cos attributes.

Please delete the annotation texts in configuration files (

```
# Annotation
```

```
).
```

 Refer to [Link](#) for a blank template

```
[common]
```

```
workspace=/tmp/tmp6 # Workspace directory
```

```
threads=20 # Number of working threads. 10 threads if not configured
```

[source]

type=oss

accesskeyid=

accesskeysecret=

bucket=

endpoint=

[destination]

type=cosv4

region=shanghai

accesskeyid=

appid=

accesskeysecret=

bucket=sdctest

## Migrate Files in OSS

[common]

workspace=/tmp/tmp6

[source]

type=oss

accesskeyid= # oss accesskey id

accesskeysecret= # oss accesskey secret

bucket= # Name of the bucket to be migrated

endpoint= # endpoint of oss, for example: oss-cn-beijing.aliyuncs.com

[destination]

type=cosv4

region=shanghai # region of cos, such as shanghai, guangzhou

accesskeyid= # accesskeyid of cos

appid= # appid of cos

accesskeysecret= # accesskeysecret of cos

bucket=sdctest # bucket of cos

prefix\_dir=/dir21 # Directory of cos. Migrated files will be placed under this directory (root directory if not configured)

### Migrate Files in Qiniu

[common]

workspace=/tmp/tmp11

[source]

type=qiniu

accesskeyid= # accesskeyid of qiniu

accesskeysecret= # accesskeysecret of qiniu

bucket= # qiniu bucket to be migrated

domain\_url= # Download domain of qiniu

[destination]

type=cosv4

region=shanghai # region of cos, such as shanghai, guangzhou

accesskeyid= # accesskeyid of cos

appid= # appid of cos

accesskeysecret= # accesskeysecret of cos

bucket=sdctest # bucket of cos

prefix\_dir=/dir21 # Directory of cos. Migrated files will be placed under this directory (root directory if not configured)

### Migrate Files to S3

[common]

workspace=/tmp/tmp21

[source]

type=s3

```
accesskeyid= # accesskey id of s3
accesskeysecret= # accesskey secret of s3
bucket= # Name of the s3 bucket to be migrated
```

```
[destination]
type=cosv4
region=shanghai
accesskeyid=
appid=
accesskeysecret=
bucket=
```

### Migrate Lists of Files

```
[common]
workspace=
```

```
[source]
type=url
url_list_file=/tmp/urllist.txt # The list file containing the URLs of files to be migrated. Each line in the
file contains a complete URL
timeout=3 # Timeout for HTTP requests
```

```
[destination]
type=cosv4
region=
accesskeyid=
appid=
accesskeysecret=
bucket=
```

### Run the Tool

Once installed, there will be an executable command

```
cos_migrate_tool
```

in the system, which will be used for all subsequent migration operations. How to execute this command:

```
cos_migrate_tool -c /path/to/your/conf
```

Modify the configuration file according to the templates mentioned above. You need to configure a workspace directory in the configuration file. Temporary files generated in migration operations will be stored in this directory, so please make sure there is enough storage space for this directory. It is recommended to use different directories if there are multiple concurrent migration tasks.

During the migration process, you can check the fail\_file.txt (located in the workspace directory you configured) to view the list of files that weren't migrated successfully.

## How to Uninstall

Run the following command:

```
pip uninstall cos_migrate_tool
```

## FAQs

1. pip command does not exist. Use "apt install python-pip" or "yum install python-pip" to install pip.
2. Failed to use pip to install migration tool. Try executing "sudo pip install cos\_migrate\_tool".

# Local Synchronization Tool

## Feature Description

Synchronize sub files and sub directories in local directories to COS. Only applicable to COS 4.0  
Implementation mechanism: The COS local synchronization tool will acquire the user's local file list, upload the files and log the result of the upload operation locally. The tool will re-acquire local file list each time is it executed, and perform comparison and synchronization operations with the local database (upload or delete).

## Operating Environment

### System Environment

Linux/Windows system

### Required Software

JDK 1.7 or 1.8

## How to Use

### Acquire the Tool Package

Download Link: [Tool Package cos\\_sync.zip](#)

Decompress the package and go to the path containing the tool:

```
unzip cos_sync.zip && cd cos_sync
```

## Configuration Instructions

Configuration file is located at /conf/config.json in the tool package directory:

```
{
  "appid" : "xxxxxx",
  "secret_id" : "xxxxxxxxxxxxxxxxxxxxxxxxxxxx",
  "secret_key" : "xxxxxxxxxxxxxxxxxxxxxxxxxxxx",
  "bucket" : "xxxxxx",
  "timeout" : "60",
  "thread_num" : "20",
  "delete_sync" : "1",
  "daemon_mode" : "0",
  "daemon_interval" : "60",
  "enable_https" : "0",
  "region" : "gz",

  "local_path" : "/home/test/data",
  "cos_path" : "/mysyncfolder"
}
```

Name	Description	Valid Value
app_id	APPID that needs to perform the operation (you can acquire it from the console)	APPID number
secret_id	The private key ID that corresponds to the APPID (you can acquire it from the console)	String
secret_key	The private key that corresponds to the APPID (you can acquire it from the console)	String
bucket	Name of the bucket to be synchronized. You need to create the bucket in the console beforehand. Refer to <a href="#">Create</a>	String

Name	Description	Valid Value
	<a href="#">Bucket</a> .	
timeout	Timeout for COS connections. You can increase this value when the network is poor. Unit: second	Number
thread_num	Number of concurrent threads. Increase this value for a higher concurrent level and upload speed. Lowering this value will reduce upload speed.	Number
delete_sync	Delete COS file in sync when deleting local file. 1: Delete remote file when deleting local file; 0: Ignore deleted local file	Number
daemon_mode	Run in daemon mode. 1: Run the synchronization tool in cycles; 0: Exit the tool after one execution.	Number
daemon_interval	(In daemon mode) Time interval for checking local file changes. Unit: second	Number
enable_https	Enable HTTPS for transmission. 1: Use HTTPS &verbar; 0: Use HTTP	Number
region	Region in which the bucket resides. For example: Tianjin, North China (tj), Shanghai, East China (sh), Guangzhou, South China (gz)	String
local_path	The absolute local path to be synchronized. In Windows, paths are separated using double	String

Name	Description	Valid Value
	backslashes "\". Linux example: /home/user/dir Windows example: C:\\document\\dir	
cos_path	Target COS path to which the files are to be synchronized. The path needs to end with "/" to indicate that it's a directory. The root directory is "/".	String

### Use the Software

Execute the synchronization tool (For Windows, double click start\_cos\_sync.bat)

```
sh start_cos_sync.sh
```

When the execution is completed, the tool will display the statistical information of successful/failed file creation/deletion operations as well as how long they took.

### Q&A

#### FAQs

I accidentally deleted files on COS after synchronization is completed. Will the files be uploaded again if I run the tool again?

No. The tool keeps the list of synchronized files locally. It does not acquire file list from COS.

Where is the database for storing synchronization records located? What will happen if I delete the records and then execute the tool again?

Synchronization result records are stored in the data file under the db directory. If you execute the

tool again after deleting the records, the tool will try to upload all local files onto COS, and existing files on the COS will be overwritten.

Does it support Chinese file names or directories?

Yes. Currently, all paths and files in UTF-8 encoding are supported.

## Common Errors

If synchronization fails, check the error log under the log directory first. Common error codes are shown below.

code: -3, connection timeout

The connection to COS service has timed out, please check whether resolution and port are normal.

How to check DNS: Suppose your region is "sh", execute "dig sh.file.myqcloud.com" locally and see if the IP received is an IP from Tencent Cloud. You can verify this by using external ping tools, such as [chinaz - Ping](#). An IP from Tencent Cloud data center should be resolved as 10.\.\*.\*. The same can be applied for other regions \*

How to check port: If IP is resolved correctly, execute

```
telnet sh.file.myqcloud.com 80
```

and see if

Escape character is '^['.

is returned. If there is no return, check your local firewall configuration, and whether your network functions normally.

code:-133, ERROR\_CMD\_BUCKET\_NOTEXIST

Check if "region" is configured correctly in your configuration file. Refer to configuration instruction for detailed relationships between regions and configurations.

## Other Errors

Please submit a ticket. Tell us the relevant configurations of the config.json for your synchronization tool (private key is not required) as well as your packaged log directory

## Version History

- 2016.11.17 4.x JAVA synchronization tool became available

## Hadoop Tool

### Feature Description

You can use Hadoop big data processing engines (such as MapReduce, Hive) to process data that is stored on Tencent Cloud Object Storage (COS). Only applicable to COS 4.0

Implementation mechanism: The Hadoop cosn plugin is used to execute high-layer computing tasks (such as MapReduce, Hive, Spark, Tez) on file systems that use Tencent COS as underlying storage. With cosn, Hadoop is able to process data that has been uploaded to COS using any method.

### Operating Environment

#### System Environment

Linux/Windows system

#### Required Software

Hadoop-2.7.2 or above

### How to Use

#### Acquire cos-java-sdk

Download link: <https://github.com/tencentyun/cos-java-sdk-hadoop-v4>

Execute the following command from the download path and compile to acquire cos\_api-4.2.jar which is located in the target directory

```
mvn clean package -Dmaven.test.skip=true
```

## Acquire hadoop-cos Plugin

Download link: <https://github.com/tencentyun/hadoop-cosn-v4>

As cosn relies on SDK, please copy the cos\_api-4.2.jar which was compiled in the previous step to src/main/resources and execute the following command to compile, to acquire the hadoop-cos-2.7.2.jar under the target directory

```
mvn clean package -Dmaven.test.skip=true
```

## Plugin Installation

Modify hadoop\_env.sh

Add the following content and add the cosn relevant jar packages to Hadoop environment variable

```
for f in $HADOOP_HOME/share/hadoop/tools/lib/*.jar; do
  if [ "$HADOOP_CLASSPATH" ]; then
    export HADOOP_CLASSPATH=$HADOOP_CLASSPATH:$f
  else
    export HADOOP_CLASSPATH=$f
  fi
done
```

Copy cos\_api-4.2.jar and hadoop-cos-2.7.2.jar (acquired in the previous two steps) to \$HADOOP\_HOME/share/hadoop/tools/lib, and copy json-20140107.jar and httpmime-4.2.5.jar (required by SDK. You can find these two JAR packages in the local maven repository) to this directory.

Modify Configuration File to Use the Plugin

Modify \$HADOOP\_HOME/etc/hadoop/core-site.xml, add COS relevant user and implementation

class information. For example:

```
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/data/rabbitliu/work/hadoop/hadoop_test</value>
  </property>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
  <property>
    <name>dfs.name.dir</name>
    <value>/data/rabbitliu/work/hadoop/hadoop_test/name</value>
  </property>
  <property>
    <name>fs.cos.userinfo.appid</name>
    <value>1252448703</value>
  </property>
  <property>
    <name>fs.cos.userinfo.secretId</name>
    <value>xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx</value>
  </property>
  <property>
    <name>fs.cos.userinfo.secretKey</name>
    <value>xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx</value>
  </property>
  <property>
    <name>fs.cosn.impl</name>
    <value>org.apache.hadoop.fs.cosnative.NativeCosFileSystem</value>
  </property>
  <property>
    <name>fs.cos.buffer.dir</name>
    <value>/data/rabbitliu/work/hadoop/hadoop_test/cos_buf</value>
```

```
</property>  
</configuration>
```

Note the attributes with "COS" in them:

- fs.cos.userinfo.appid: Enter the APPID of the COS account you use
- fs.cos.userinfo.secretId/secretKey: Enter the private key information of your account
- fs.cosn.impl is the implementation class of cosn, it is always `org.apache.hadoop.fs.cosnative.NativeCosFileSystem`
- fs.cos.buffer.dir: Please configure a directory that actually exists. Temporary files generated during operations will be stored here

## Use the Software

### Use hadoop fs Common Commands

```
[rabbitliu@VM_83_1_centos bin]$ hadoop fs -ls cosn://rabbitliu/
```

Found 7 items

```
-rw-rw-rw- 1 rabbitliu rabbitliu 3669 2016-10-25 21:23 cosn://rabbitliu/b.txt  
drwxrwxrwx - rabbitliu rabbitliu 0 1970-01-01 08:00 cosn://rabbitliu/dir1  
drwxrwxrwx - rabbitliu rabbitliu 0 1970-01-01 08:00 cosn://rabbitliu/mr  
-rw-rw-rw- 1 rabbitliu rabbitliu 16952 2016-10-25 21:37 cosn://rabbitliu/qcloud_sign.proto  
-rw-rw-rw- 1 rabbitliu rabbitliu 2048 2016-10-25 21:48 cosn://rabbitliu/rabbit_test2K.txt  
-rw-rw-rw- 1 rabbitliu rabbitliu 52428800 2016-10-27 16:40 cosn://rabbitliu/rabbit_test50MB.txt  
drwxrwxrwx - rabbitliu rabbitliu 0 1970-01-01 08:00 cosn://rabbitliu/xx1
```

### Use MapReduce's wordcount Application

```
bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount  
cosn://rabbitliu/mr/input cosn://rabbitliu/mr/output3
```

When successfully executed, it will return the following statistical information:

#### File System Counters

COSN: Number of bytes read=72  
COSN: Number of bytes written=40  
COSN: Number of read operations=0  
COSN: Number of large read operations=0  
COSN: Number of write operations=0  
FILE: Number of bytes read=547350  
FILE: Number of bytes written=1155616  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=0  
HDFS: Number of bytes written=0  
HDFS: Number of read operations=0  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=0

#### Map-Reduce Framework

Map input records=5  
Map output records=7  
Map output bytes=59  
Map output materialized bytes=70  
Input split bytes=99  
Combine input records=7  
Combine output records=6  
Reduce input groups=6  
Reduce shuffle bytes=70  
Reduce input records=6  
Reduce output records=6  
Spilled Records=12  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1

GC time elapsed (ms)=0

Total committed heap usage (bytes)=653262848

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=36

File Output Format Counters

Bytes Written=40