

# GCS 操作中的一些Best Practice

## 一、批量删除文件

后台多进程递归删除：

```
nohup gsutil -m rm -r gs://bucket/* >> output.log 2>&1 &  
nohup gsutil -m acl ch -u AllUsers:R gs://bucket/* >> pub.log 2>&1 &
```

## 二、如果文件量太大，建议采用lifecycle的方式

生成lifecycle.json文件：

```
{  
  "rule": [  
    {  
      "action": {"type": "Delete"}, "condition": {"age": 1}  
    }  
  ]  
}
```

设置lifecycle：

```
gsutil lifecycle set lifecycle.json gs://bucket
```

这样一天之后，系统会自动删除文件夹。

## 三、删除Bucket

使用gsutil删除bucket时，bucket要是空时才能删除。

使用程序删除bucket，可以指定force参数，强行删除：

```
# -*- coding: UTF-8 -*-  
from google.cloud import storage  
bucket_name="4funvideo"  
storage_client = storage.Client()  
bucket = storage_client.get_bucket(bucket_name)  
bucket.delete(force=True)  
print('Bucket {} deleted'.format(bucket.name))
```

但如果bucket文件超过256个，删除将会不成功。

```
ValueError: Refusing to delete bucket with more than 256 objects. If you actually want to delete this bucket, please delete the objects yourself before calling Bucket.delete().
```

## 四、 批量设置bucket内文件的属性

在把GCS的bucket做为HTTP LB的backend，并开启CDN功能的时候，要求每个bucket里的object是public的。根据文档，如果object的metadata标示为private的，这个object将不能被CDN cache。

这时，要把所有object的属性改成公开。

如果数据量非常大，建议采用：

- 创建新bucket
- 修改bucket的默认acl - defacl
- 通过file transfer，把老bucket的object复制到新文件夹
- 复制的过程，object将获得新bucket定义的属性

具体实现方法：

```
cao_haibin@cloudshell:~ (haibincao)$ gsutil mb gs://haibinsync
Creating gs://haibinsync/...
cao_haibin@cloudshell:~ (haibincao)$ gsutil defacl get gs://haibinsync >
defacl.json
```

修改这个json文件，添加allusers的Reader权限：

```
[
  {
    "entity": "project-owners-273428662757",
    "projectTeam": {
      "projectNumber": "273428662757",
      "team": "owners"
    },
    "role": "OWNER"
  },
  {
    "entity": "project-editors-273428662757",
    "projectTeam": {
      "projectNumber": "273428662757",
      "team": "editors"
    },
    "role": "OWNER"
  },
]
```

```
{
  "entity": "project-viewers-273428662757",
  "projectTeam": {
    "projectNumber": "273428662757",
    "team": "viewers"
  },
  "role": "READER"
},
{
  "entity": "allUsers",
  "role": "READER"
}
]
```

设置defacl:

```
cao_haibin@cloudshell:~ (haibincao)$ gsutil defacl set defacl.json
gs://haibinsync
Setting default object ACL on gs://haibinsync/...
```

查看defacl:

```
cao_haibin@cloudshell:~ (haibincao)$ gsutil defacl get gs://haibinsync
```

设置好后，在console中创建file transfer：

Google Cloud Platform

haibincao

Data Transfer

Transfer Service | cloud

Transfer Service | on-premises

Transfer Appliance

Create a transfer job

Transfer data to a Cloud Storage bucket from a source you specify here.

Requirements:

- Transfer Service service account must have permission to read from your source and write to your destination.
- If you're deleting the source files, the Transfer Service's service account will need delete access to the source.
- If your service account doesn't have these permissions yet, a bucket owner must grant them.

If you have 'storage.buckets.setIamPolicy' permission for the source and destination buckets, creating a transfer job will grant that service account the required source and destination permissions to complete the transfer.

[Learn more](#)

1 Select source

☒ Google Cloud Storage bucket

☐ Amazon S3 bucket

☐ Microsoft Azure Storage container BETA

☐ List of object URLs

You must have read access to the source bucket.

Cloud Storage bucket

☒ bucket\_haibin\_test

Browse

[Specify file filters](#)

Continue

2 Select destination

3 Configure transfer

Cancel

Creating a transfer job grants Cloud Storage Transfer Service's service account the necessary source and destination permissions to complete the transfer. Your permissions will update to reflect this change.

Google Cloud Platform

haibincao

Data Transfer

Transfer Service | cloud

Transfer Service | on-premises

Transfer Appliance

Create a transfer job

Transfer data to a Cloud Storage bucket from a source you specify here.

Requirements:

- Transfer Service service account must have permission to read from your source and write to your destination.
- If you're deleting the source files, the Transfer Service's service account will need delete access to the source.
- If your service account doesn't have these permissions yet, a bucket owner must grant them.

If you have 'storage.buckets.setIamPolicy' permission for the source and destination buckets, creating a transfer job will grant that service account the required source and destination permissions to complete the transfer.

Learn more

1 Select source

2 Select destination

3 Configure transfer

Cloud Storage bucket

haibinsync

Browse

Transfer options

You can set additional rules for how your transfer handles overwrites and deletions. By default, your transfer only overwrites an object when the source version is different from the destination version. No other objects are overwritten or deleted.

☐ Overwrite destination with source, even when identical

☐ Delete objects from source once they are transferred

☐ Delete object from destination if there is no version in source

Continue

Cancel

Creating a transfer job grants Cloud Storage Transfer Service's service account the necessary source and destination permissions to complete the transfer. Your permissions will update to reflect this change.

Google Cloud Platform

haibincao

Data Transfer

Create a transfer job

Transfer Service | cloud

Transfer Service | on-premises

Transfer Appliance

Transfer data to a Cloud Storage bucket from a source you specify here.

Requirements:

- Transfer Service service account must have permission to read from your source and write to your destination.
- If you're deleting the source files, the Transfer Service's service account will need delete access to the source.
- If your service account doesn't have these permissions yet, a bucket owner must grant them.

If you have 'storage.buckets.setIamPolicy' permission for the source and destination buckets, creating a transfer job will grant that service account the required source and destination permissions to complete the transfer.

[Learn more](#)

✓ Select source

▼

✓ Select destination

▼

3 Configure transfer

⬆

Schedule ?

☒ Run now

☐ Run daily at 3:56:12 PM

Description

bucket\_haibin\_test-haibinsync\_once

Choose a unique description to help identify your transfer.

Create

Cancel

Creating a transfer job grants Cloud Storage Transfer Service's service account the necessary source and destination permissions to complete the transfer. Your permissions will update to reflect this change.

|  |                      |                    |                    |                     |                           |  |
|--|----------------------|--------------------|--------------------|---------------------|---------------------------|--|
| Google Cloud Platform  |                      |                    |                    |                     |                           |  |
| Transfer Service for cloud data  |                      |                    |                    |                     |                           |  |
| <a href="#">CREATE TRANSFER JOB</a> <a href="#">REFRESH</a> <a href="#">DELETE</a> |                      |                    |                    |                     |                           |  |
| <input type="checkbox"/> Description   | Source               | Source bucket      | Destination bucket | Operation frequency | Latest transfer operation |  |
| <input type="checkbox"/> bucket_haibin_test-haibinsync_once                        | Google Cloud Storage | bucket_haibin_test | haibinsync         | Only once           | Completed                 |  |

## bucket\_haibin\_test-haibinsync\_once

Operations [Configuration](#)

### Transfer configuration

All scheduled transfer operations will use the following transfer configuration. Any changes made to the configuration will be applied to subsequent transfer operations only.

|   |                                  |
|---|----------------------------------|
| <b>Name</b> ?                             | 16478863598240266215             |
| <b>Source</b>                             | Google Cloud Storage             |
| <b>Source bucket</b>                      | bucket_haibin_test               |
| <b>Destination bucket</b>                 | haibinsync                       |
| <b>Schedule</b> ?                         | Once on Apr 13, 2020, 4:21:18 PM |
| <b>Transfer files with these prefixes</b> |                                  |
| <b>Exclude files with these prefixes</b>  |                                  |
| <b>Synchronization options</b>            | —                                |
| <b>Configuration last modified on</b>     | Apr 13, 2020, 4:21:18 PM         |
| <b>Time-based filtering</b>               |                                  |
| <b>Maximum age</b>                        | —                                |
| <b>Minimum age</b>                        | —                                |

再检查一下GCS里的对象属性：

| <input type="checkbox"/> | Name                                 | Size     | Type       | Storage class | Last modified             | Public access      | Encryption         | Retention expiration date | Holds |
|--------------------------|--------------------------------------|----------|------------|---------------|---------------------------|--------------------|--------------------|---------------------------|-------|
| <input type="checkbox"/> | Od1b78a1d7b97c0e819b73d50e07bc05.mp4 | 2.68 MB  | video/mp4  | Standard      | 4/13/20, 4:20:54 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | 0dc0c7405199a092071784b54bb605ec.jpg | 8.64 KB  | image/jpeg | Standard      | 4/13/20, 4:20:50 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | 0dc0c7405199a092071784b54bb605ec.mp4 | 3.41 MB  | video/mp4  | Standard      | 4/13/20, 4:20:53 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | 1f5bf069cc4f9d3627ed6c981f256156.jpg | 9.54 KB  | image/jpeg | Standard      | 4/13/20, 4:20:49 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | 1f5bf069cc4f9d3627ed6c981f256156.mp4 | 2.63 MB  | video/mp4  | Standard      | 4/13/20, 4:20:51 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | 32668228544b5b25a0cd29c074738364.jpg | 5.41 KB  | image/jpeg | Standard      | 4/13/20, 4:20:50 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | 4903e9533057f13ac80a4c6ea441ad73.mp4 | 11.76 MB | video/mp4  | Standard      | 4/13/20, 4:20:52 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | be33e4173762f4868409251911d7a5d6.jpg | 8.31 KB  | image/jpeg | Standard      | 4/13/20, 4:20:53 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | be33e4173762f4868409251911d7a5d6.mp4 | 2.23 MB  | video/mp4  | Standard      | 4/13/20, 4:20:51 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | d2d1804989b903b5e15b27e7784863e2.jpg | 8.61 KB  | image/jpeg | Standard      | 4/13/20, 4:20:54 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | d2d1804989b903b5e15b27e7784863e2.mp4 | 2.56 MB  | video/mp4  | Standard      | 4/13/20, 4:20:53 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | neural_network_eager_api.txt         | 19.98 KB | text/plain | Standard      | 4/13/20, 4:20:51 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | test0.webm                           | 4.12 MB  | video/webm | Standard      | 4/13/20, 4:20:53 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |
| <input type="checkbox"/> | xxxx.txt                             | 18.12 KB | text/plain | Standard      | 4/13/20, 4:20:49 PM UTC+8 | Public to internet | Google-managed key | –                         | None  |

都是Public啦。