

# S3 Buckets

S3 is a simple object storage service initially provided by Amazon. There are numerous datasets that you can access through S3 as well as store your own objects. EPFL is providing an official S3 endpoint as well as other faculties at EPFL.

- [Ask for an S3 bucket at epfl central IT](#)
- [IC-IT hosted datasets](#)

Here are a couple of ways of accessing S3 buckets in the SCITAS clusters.

- [s3fs](#)
  - [Configure access to your existing S3 bucket](#)
  - [Mount the bucket as a filesystem in your home directory](#)
- [boto \(python\)](#)
- [rclone](#)

## s3fs

Allows you to use an S3 bucket as a Filesystem. It is installed on the cluster login nodes (must be mounted before running the jobs on the login node).

### Configure access to your existing S3 bucket

```
echo S3_ACCESS_KEY:S3_SECRET_KEY > ${HOME}/.passwd-s3fs
chmod 600 ${HOME}/.passwd-s3fs
```

### Mount the bucket as a filesystem in your home directory

For EPFL s3 service, if you use an other endpoint you have to modify the url (or remove it for Amazon S3)

```
mkdir ${HOME}/mybucket
s3fs BUCKET_ID ${HOME}/mybucket -o url=https://s3.epfl.ch/ -o passwd_file=${HOME}/.passwd-s3fs
```

Unmount the filesystem

```
fusermount -u ${HOME}/mybucket
```

## boto (python)

Install the library

```
module load gcc python
pip3 install --user boto
```

Using the library (source: <https://icitdocs.epfl.ch/display/clusterdocs/Accessing+Datasets>)

```
#!/usr/bin/env python3

import boto
import boto.s3.connection

endpoint = 's3.epfl.ch'
access_key = 'put your access key here!'
secret_key = 'put your secret key here!'
bucket_id = 'put your bucket id here!'

conn = boto.connect_s3(
    aws_access_key_id = access_key,
    aws_secret_access_key = secret_key,
    host = endpoint,
    calling_format = boto.s3.connection.OrdinaryCallingFormat(),
)

bucket = conn.get_bucket(bucket_id)

# listing objects in bucket
for key in bucket.list():
    print ("{name}\t{size}\t{modified}".format(
        name = key.name,
        size = key.size,
        modified = key.last_modified,
    ))
```

More information in the [boto documentation](#).

```
chmod u+x s3.py
./s3.py
test      6      2021-06-23T08:29:48.409Z
```

## rclone

Edit ~/.rclone.conf

```
[private]
type = s3
access_key_id = put_your_access_key_here
secret_access_key = put_your_secret_key_here
region = other-v2-signature
endpoint = https://s3.epfl.ch/
```

List a bucket content

```
rclone ls private:<bucket_id>
```

More information in the [rclone documentation](#).