MAX PLANCK
COMPUTING & DATA FACILITY

MPCDF

HPSS S3 Interface Evaluation at MPCDF

Elena Summer

MAX PLANCK COMPUTING & DATA FACILITY

Backup & Archive Group

Karlsruhe September 2024

S3 (Simple Storage Service)

File Storage vs Object Storage:

- Structured hierarchical system vs flat address space
- Path vs ID
- Folders vs buckets
- Data and metadata

```
$ ls -l 006-64-mb
-rw-rw-r--  1 sthree sthree   67108864 Jul 16 13:52 006-64-mb

$ s3cmd info s3://sthree-test-bucket/006-64-mb-u7
s3://sthree-test-bucket/006-64-mb-u7 (object):
   File size: 67108864
   Last mod:  Wed, 21 Aug 2024 16:55:53 GMT
   MIME type: binary/octet-stream
   Storage:   INTELLIGENT_TIERING
   MD5 sum:   10f126317b6bd1bc01b72ffe233f50b3
   SSE:       none
   Policy:    none
   CORS:      none
   ACL:       CustomersName@amazon.com: FULL_CONTROL
   x-amz-meta-hpss.volume_name: PL004000
   x-amz-meta-hpss.relative_position: 77
   x-amz-meta-s3cmd-attrs:
atime:1722341286/ctime:1721130761/gid:2933/gname:sthree/md5:10f126317b6bd1bc
01b72ffe233f50b3/mode:33204/mtime:1721130761/uid:1933/uname:sthree
   x-amz-meta-pftp-multipart-upload-etag: 1f26db2f5204d0848d30db5ac11cc88e-5
   x-amz-meta-hpss.hash.creator: none
   x-amz-meta-hpss.hash.flags:
   x-amz-meta-x-s3proxy-meta-storage-class: STANDARD
   x-amz-meta-hpss.relative_position_offset: 0
```

HPSS S3 Interface

- Offered starting from HPSS version 10.3
- Is an HTTP server which receives S3 REST API requests and forwards them to HPSS
- Communicates with HPSS via HPSS PFTP server (using jclouds module)
- Stores S3 objects as files in HPSS

  Buckets become folders

  Metadata is kept
- Compatible with S3

```
scrub> ls -l /s3stuff/sthree-test-bucket/005-10-mb
Frw-rw-r--   1 sthree   sthree  10485760 Jul 16 2024
/s3stuff/sthree-test-bucket/005-10-mb


scrub> dump /s3stuff/sthree-test-bucket/005-10-mb
...
    Disk Segment 0:
...
            Allocated Length                2097152
            Create Time                     2024/07/16 13:23:45
            Update Time                     2024/07/16 13:23:45
...
Storage Level 1:
    Tape Segment 0:
            Written Length                  10485760
            Last Write Time                 2024/07/16 14:19:36
            Creation Time                   2024/07/16 14:17:19
...
        Physical Volumes:
            Physical Volume 0:
                Name                            PL004000
                Media Type                      3580 Gen8 Tape
...
```

HPSS S3 Interface Potential Usage at MPCDF

MPCDF offers:

- **“Nexus-S3** is a scaleable object storage service compatible with the Amazon S3 protocol. MPCDF users can opt-in to Nexus-S3 which provide a free 1TB (1M objects) quota (see opt-in below). Data can be accessed using standard S3 clients and libraries such as minio, s3cmd, rclone and python-boto3 as well as via Globus (MPCDF GO Nexus S3 Collection) or via a web browser/curl.”

  docs.mpcdf.mpg.de: Data/Nexus-S3: Object Storage for data Transfer and Sharing/Nexus-S3

- **“HPC Cloud** offers both block storage, in the form of disk volumes which can be directly attached to a server, and file storage, in the form of shared filesystems which can be NFS-mounted by the operating system. In addition, there is an object store providing containers (also called buckets) which allow data to be accessed from multiple clients though standard REST APIs.The block and object storage services are based on **Ceph**, while file storage offers a choice between CephFS and IBM Storage Scale (GPFS).”

  docs.mpcdf.mpg.de: HPC-Cloud/Technical and User Documentation/Storage

- Both services need a long-term archival solution.
- Nexus-S3 can sync data with rclone to HPSS S3 Proxy. Ceph can use Ceph Cloud Sync module as well as rclone to send data to HPSS S3 Proxy

HPSS S3 Interface potential usage at MPCDF

MPCDF offers:

- “For data publishing, the MPCDF recommends a **CKAN**-based data repository. CKAN is a software framework which allows to manage metadata as well as object data. Beside a web-based interface, CKAN offers a REST API for automation of common workflows. CKAN instances at the MPCDF are meant for Max Planck Institutes, groups or projects and not for individual users.”

  docs.mpcdf.mpg.de: Documentation/Data Data Publication and Metadata Management/Service: Data Repositories

- Storing the - handcrafted - metadata in CKAN directly and just keeping links to the "real" data. This could also be a link to a S3 object / bucket and can have an arbitrary size. A link is a predefined URL to a file (per default available for 7 days).

```
mc share download hpss/testbucket/mb/005.mb
```

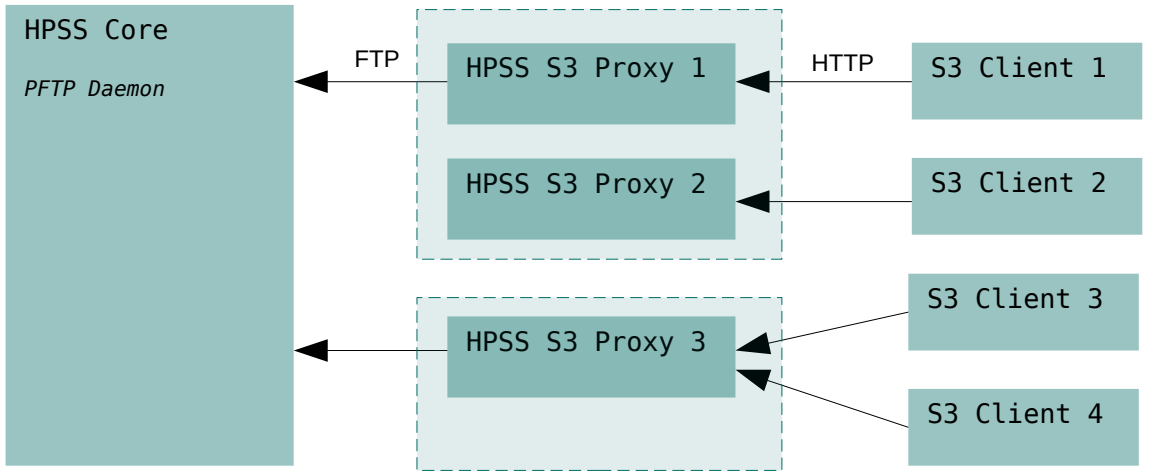
HPSS S3 Interface Installation

Our HPSS Core Machine:

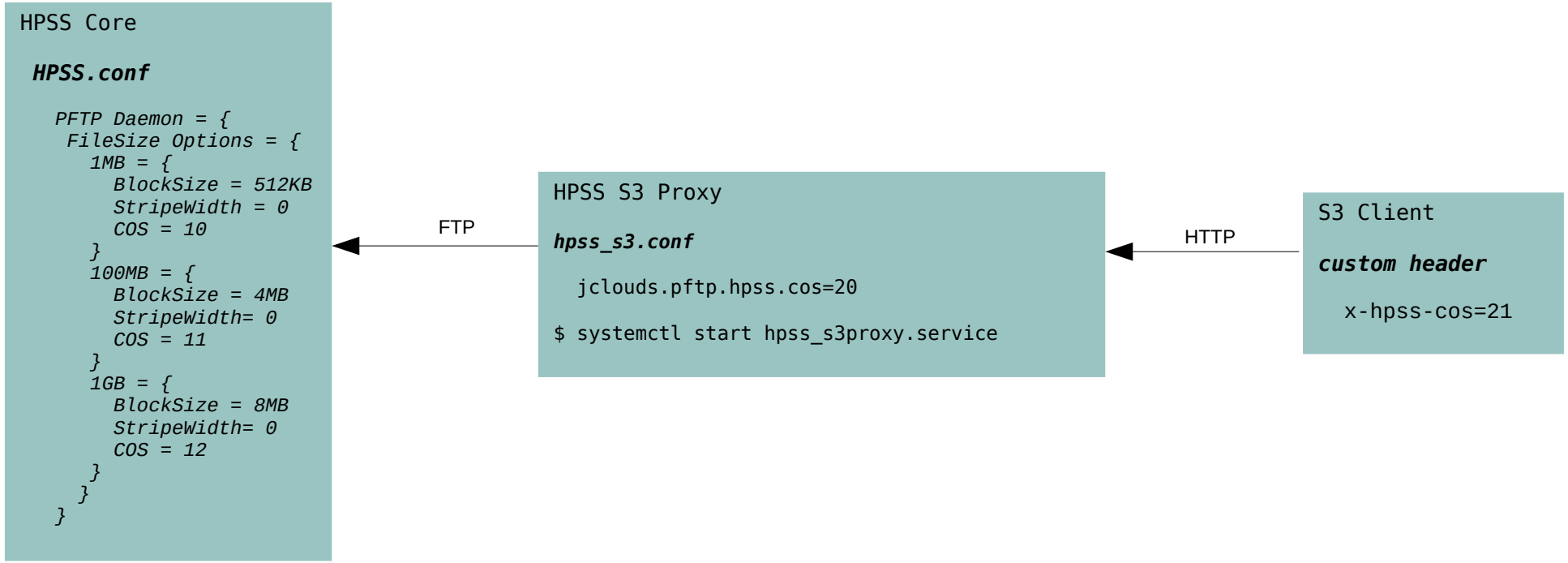
- Red Hat Enterprise Linux 8.4
- HPSS v 10.3u4

Our HPSS S3 Proxy Machine:

- Red Hat Enterprise Linux 8.4
- HPSS S3 Proxy v 10.3u5



HPSS S3 Interface Configuration



HPSS S3 Interface User Management

- HPSS S3 Proxy supports UNIX or PAM authentication. Kerberos is not supported yet.
- Create HPSS users with hpssuser on the HPSS core machine
- On S3 Proxy machine, create keystore for users to authenticate with an S3 client

  keytool -importpass -storetype pkcs12 -alias username -keystore /var/hpss/etc/hpsss3keystore
- On the S3 client, username and password are stored in a config file as access key and secret key

```
$ cat .s3cfg
[default]
access_key = username
secret_key = password
check_ssl_certificate = False
guess_mime_type = True
host_base = s3hpssproxy:8080
host_bucket = s3hpssproxy:8080/(%bucket)
```

```
$ cat .mc/config.json
{
        "version": "10",
        "aliases": {
                "s3hpss": {
                        "url": "s3hpssproxy:8080",
                        "accessKey": "username",
                        "secretKey": "password",
                        "api": "S3v4",
                        "path": "auto"
                }
        }
}
```

HPSS S3 Interface Tests

Tests with s3cmd client performed with HPSS S3 proxy ver 10.3 update 7:

```
#### s3cmd put and stat with multipart upload

$ ls -l 006-64-mb
-rw-rw-r--  1 sthree sthree   67108864 Jul 16 13:52 006-64-mb

$ s3cmd put 006-64-mb s3://sthree-test-bucket/006-64-mb-u7
upload: '006-64-mb' -> 's3://sthree-test-bucket/006-64-mb-u7'  [part 1 of 5, 15MB] [1 of 1]
 15728640 of 15728640   100% in    0s    37.15 MB/s  done
upload: '006-64-mb' -> 's3://sthree-test-bucket/006-64-mb-u7'  [part 2 of 5, 15MB] [1 of 1]
 15728640 of 15728640   100% in    0s    42.65 MB/s  done
upload: '006-64-mb' -> 's3://sthree-test-bucket/006-64-mb-u7'  [part 3 of 5, 15MB] [1 of 1]
 15728640 of 15728640   100% in    0s    44.01 MB/s  done
upload: '006-64-mb' -> 's3://sthree-test-bucket/006-64-mb-u7'  [part 4 of 5, 15MB] [1 of 1]
 15728640 of 15728640   100% in    0s    42.26 MB/s  done
upload: '006-64-mb' -> 's3://sthree-test-bucket/006-64-mb-u7'  [part 5 of 5, 4MB] [1 of 1]
 4194304 of 4194304   100% in    0s    26.19 MB/s  done

$ s3cmd info s3://sthree-test-bucket/006-64-mb-u7
s3://sthree-test-bucket/006-64-mb-u7 (object):
   File size: 67108864
   Last mod:  Wed, 21 Aug 2024 16:55:53 GMT
   MIME type: binary/octet-stream
   Storage:   INTELLIGENT_TIERING
   MD5 sum:   10f126317b6bd1bc01b72ffe233f50b3
   SSE:       none
   Policy:    none
   CORS:      none
   ACL:       CustomersName@amazon.com: FULL_CONTROL
   x-amz-meta-hpss.volume_name: PL004000
   x-amz-meta-hpss.relative_position: 77
   x-amz-meta-s3cmd-attrs:
atime:1722341286/ctime:1721130761/gid:2933/gname:sthree/md5:10f126317b6bd1bc01b72ffe233f50b3
/mode:33204/mtime:1721130761/uid:1933/uname:sthree
   x-amz-meta-pftp-multipart-upload-etag: 1f26db2f5204d0848d30db5ac11cc88e-5
   x-amz-meta-hpss.hash.creator: none
   x-amz-meta-hpss.hash.flags:
   x-amz-meta-x-s3proxy-meta-storage-class: STANDARD
   x-amz-meta-hpss.relative_position_offset: 0
```

```
Timestamps:

ctime:1721130761 = July 16, 2024 11:52:41 AM

mtime:1721130761 = July 16, 2024 11:52:41 AM
```

```
scrub> ls -l /s3stuff/sthree-test-bucket
Frw-r-----   1 sthree   sthree  67108864 Aug 21 2024  006-64-mb-u7
```

```
scrub> dump /s3stuff/sthree-test-bucket/006-64-mb-u7
Storage Level 1:
    Tape Segment 0:
        Storage Segment:
            Relative Start Address:
                Section                         77
                Offset                          0
        Physical Volumes:
            Physical Volume 0:
                Name                            PL004000
```

HPSS S3 Interface Tests

Tests with MinIO client performed with HPSS S3 proxy ver 10.3 update 7:

```
#### mc custom metadata

$  mc tag set s3hpss/sthree-test-bucket/006-64-mb-mc-u7 "elenatagkey1-64=elenatagval1-64"
Tags set for s3hpssproxy:8080/sthree-test-bucket/006-64-mb-mc-u7.

$ mc tag list s3hpss/sthree-test-bucket/006-64-mb-mc-u7
Name                       : s3hpssproxy:8080/sthree-test-bucket/006-64-mb-mc-u7
elenatagkey1-64            : elenatagval1-64
pftp-multipart-upload-etag : e8fd7b3cb88171149c496daece8d061d-4
```

HPSS S3 Interface Tests

Tests with s3cmd client performed with HPSS S3 proxy ver 10.3 update 7:

```
#### s3cmd get file from tape

scrub> purge /s3stuff/sthree-test-bucket/007-10-mb all 0
10 MB purged

$ s3cmd get s3://sthree-test-bucket/007-10-mb 007-10-mb-get
download: 's3://sthree-test-bucket/007-10-mb' -> '007-10-mb-get'  [1 of 1]
 10485760 of 10485760   100% in    1s     7.58 MB/s  done

[root@thpss ~]# rtmu

------------------------------------------------------------
Fri Aug 23 16:35:19 2024
------------------------------------------------------------
                       RequestID  Oldest Start Newest Update        State           Action       User                                  File
------------------------------- ------------- ------------- ------------ ---------------- ---------- ------------------------------------
dd73ad5ff73bf54a89776c4f56f9dce7 00:02:19 CORE 00:01:04  MVR  in progress             read     sthree /s3stuff/sthree-test-bucket/007-10-mb


                               PVNames (Read) PVLJobIdRead                                PVNames (Write) PVLJobIdWrite      Wait Reason
 ------------------------------------------- ------------ ------------------------------------------------ ------------- ----------------
                                     PL004000          454                                            n/a               device positioning
```

HPSS as a Tape Backend for Object Storage

- Using rclone

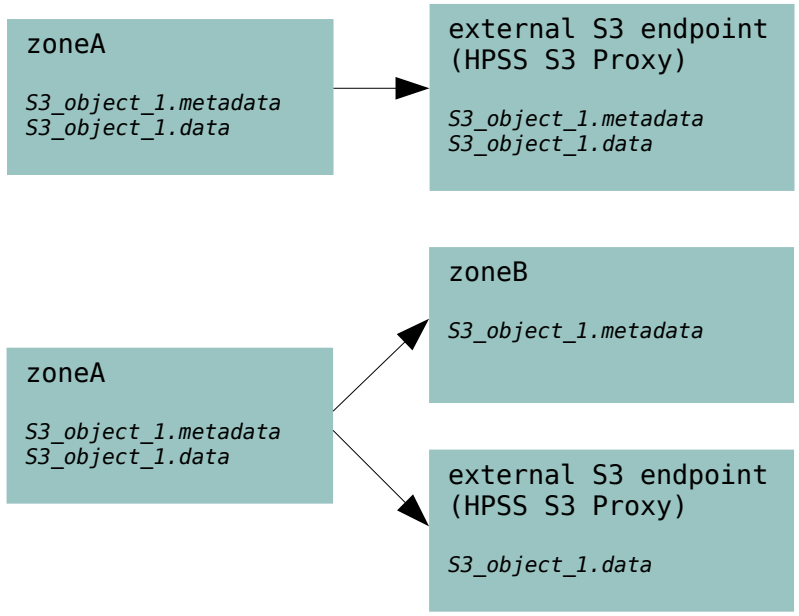
```
~/.config/rclone/rclone.conf:

[hpsss3-test]
type = s3
provider = Other
access_key_id = ...
secret_access_key = ...
endpoint = s3hpssproxy:8080
acl = private
no-check-certificate = true
upload_cutoff = 1G
chunk_size = 128M
upload_concurrency = 8

[nexuss3]
type = s3
provider = Ceph
access_key_id = ...
secret_access_key = ...
Endpoint = ...
upload_cutoff = 1G
chunk_size = 128M
upload_concurrency = 8
acl = private
```

```
rclone --immutable sync nexuss3:hpsss3-test hpsss3-test:mpl-test
```

- Using Ceph Cloud Sync module

  Sync object data and metadata to a remote zone

  One-way: the data is not synced back



Thank you!

For further questions please contact:

Elena Summer

MAX PLANCK COMPUTING & DATA FACILITY

Backup & Archive Group

Gießenbachstraße 2, 85748 Garching, Germany

Tel. +49 89 3299 2227

Email: elena.summer at mpcdf.mpg.de