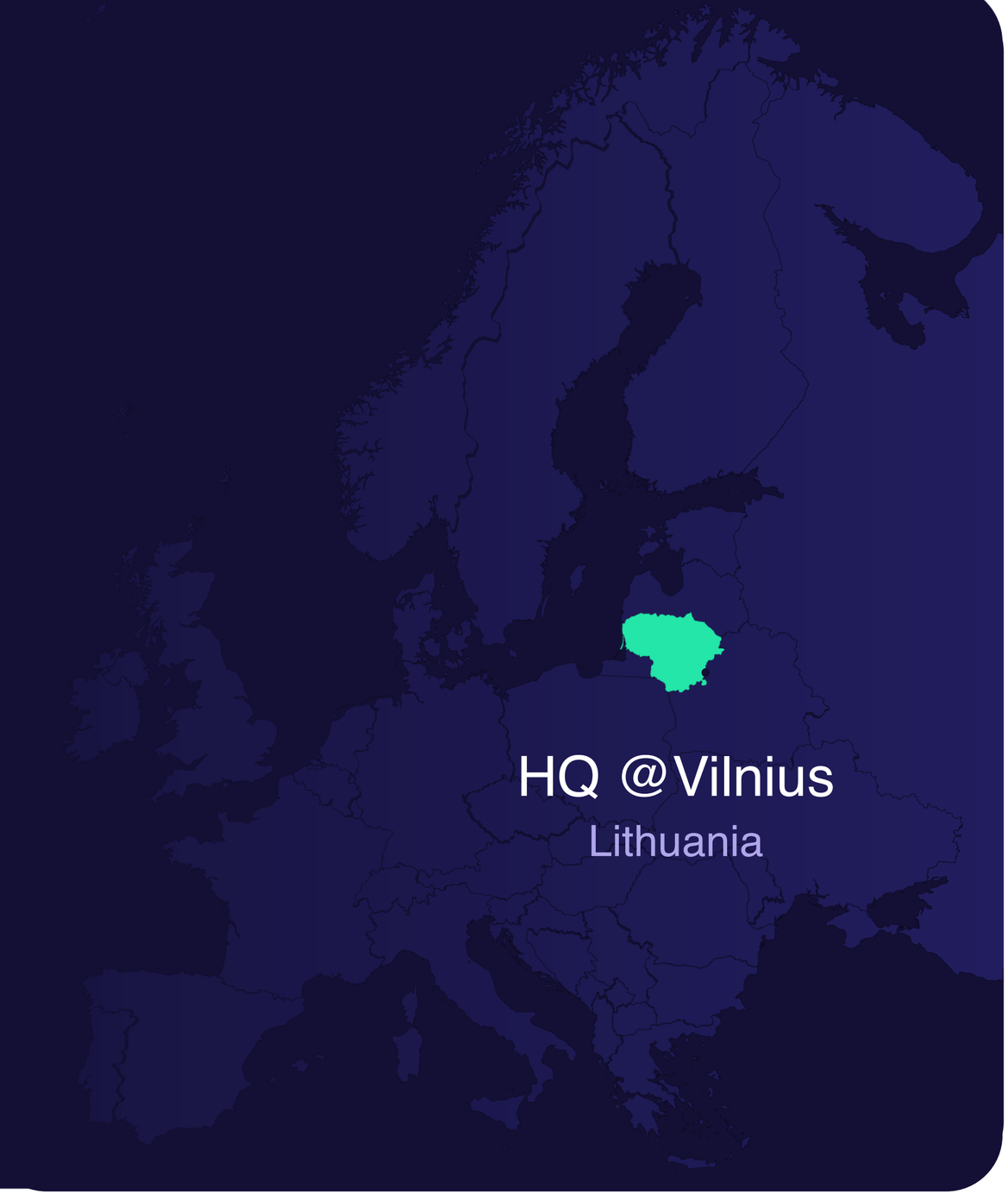
oxylabs®

Repurposing Ceph Cloud-S3 Module for Internal Bucket Transfers

Laimis Juzeliūnas
DevOps Engineer

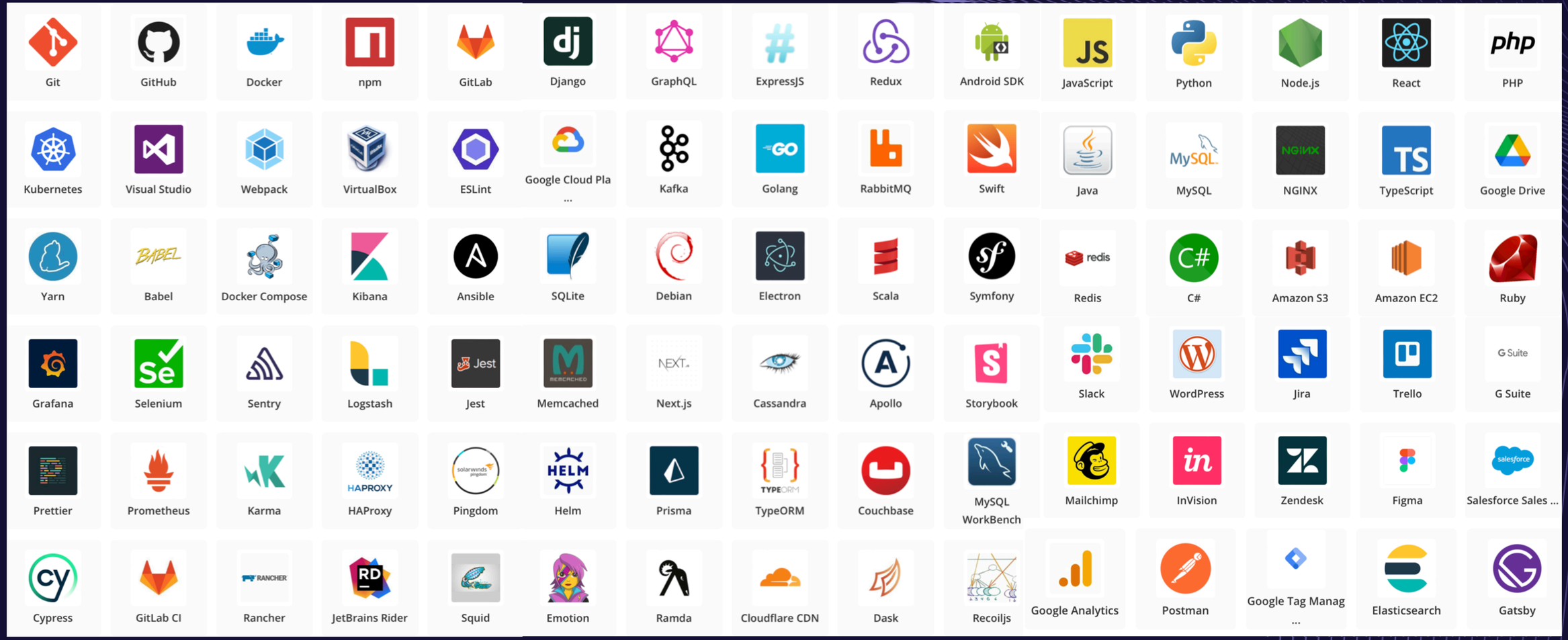
What do we do?

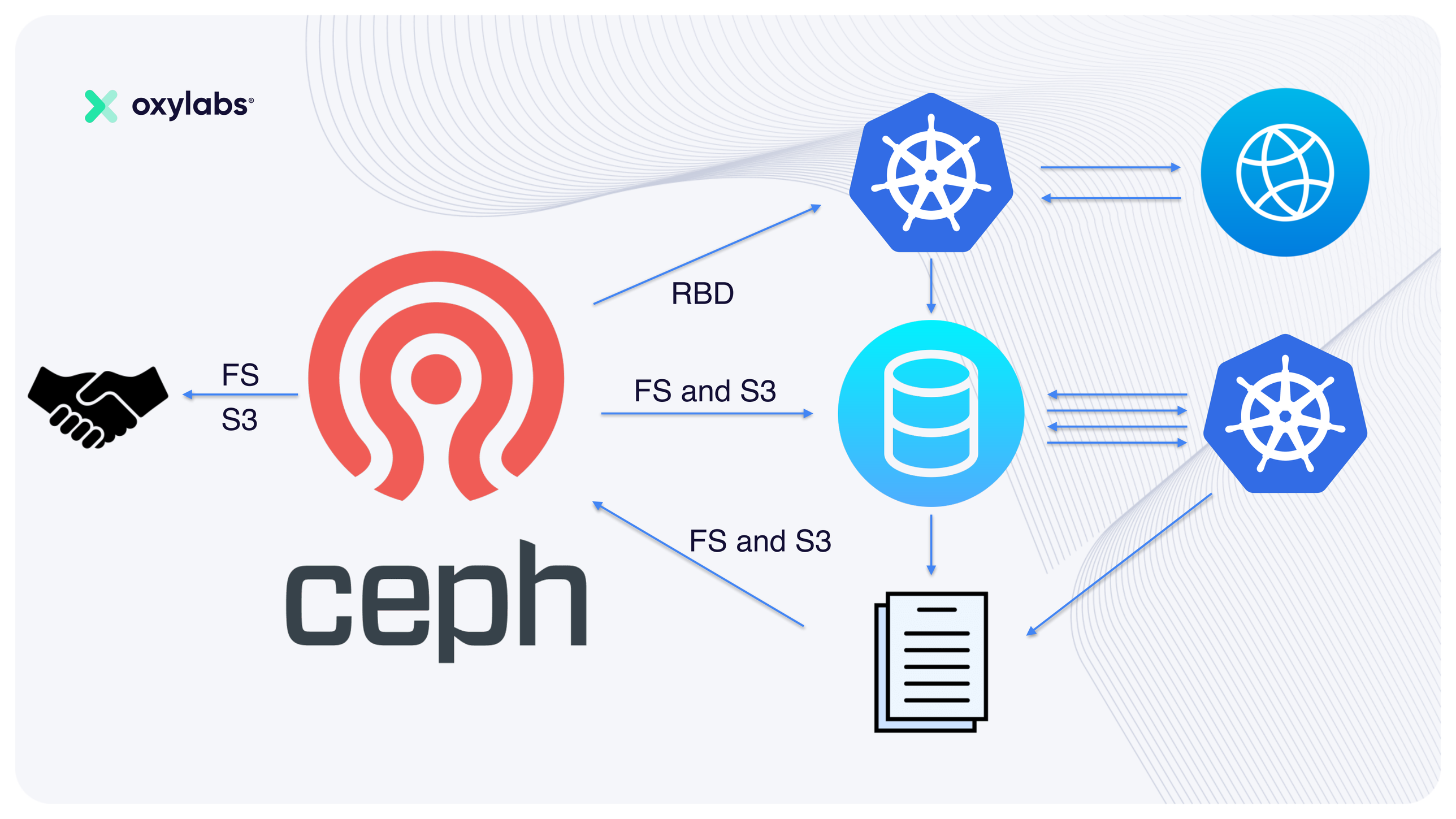
- Open source web intelligence gathering
- Proxy and API solutions
- Datasets for text, images, audio and video



Our technology landscape

Or at least part of it…




oxylabs
RBD
FS
S3
FS and S3
FS and S3
ceph

Ceph usage timeline

2021
Ceph RBD for Kubernetes

2022
CephFS for shared storage

2023
4 PiB Ceph cluster

2023
CephFS for exposed datasets

2023
Data Platform on S3 (RGW)

2024
6 PiB Ceph cluster

2026
Private Cloud

2025
8 PiB Ceph cluster

2024
Ceph Rebuild

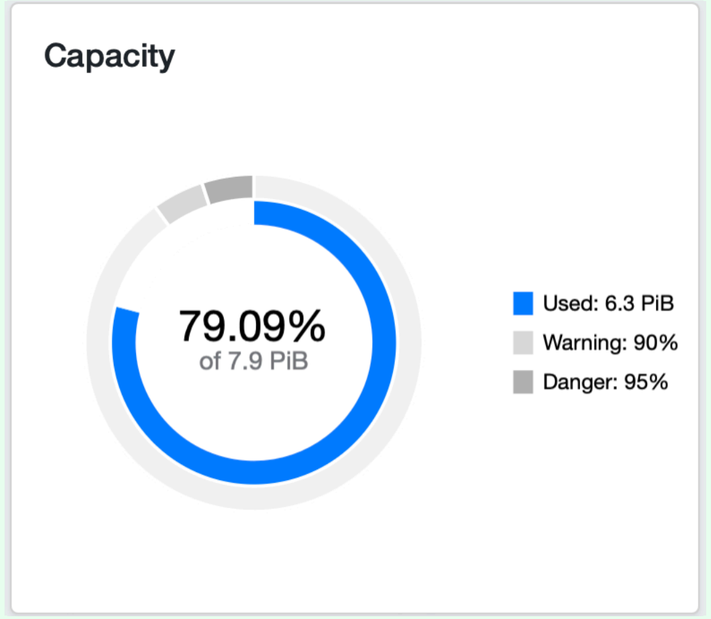
oxylabs®

Our Data Cluster

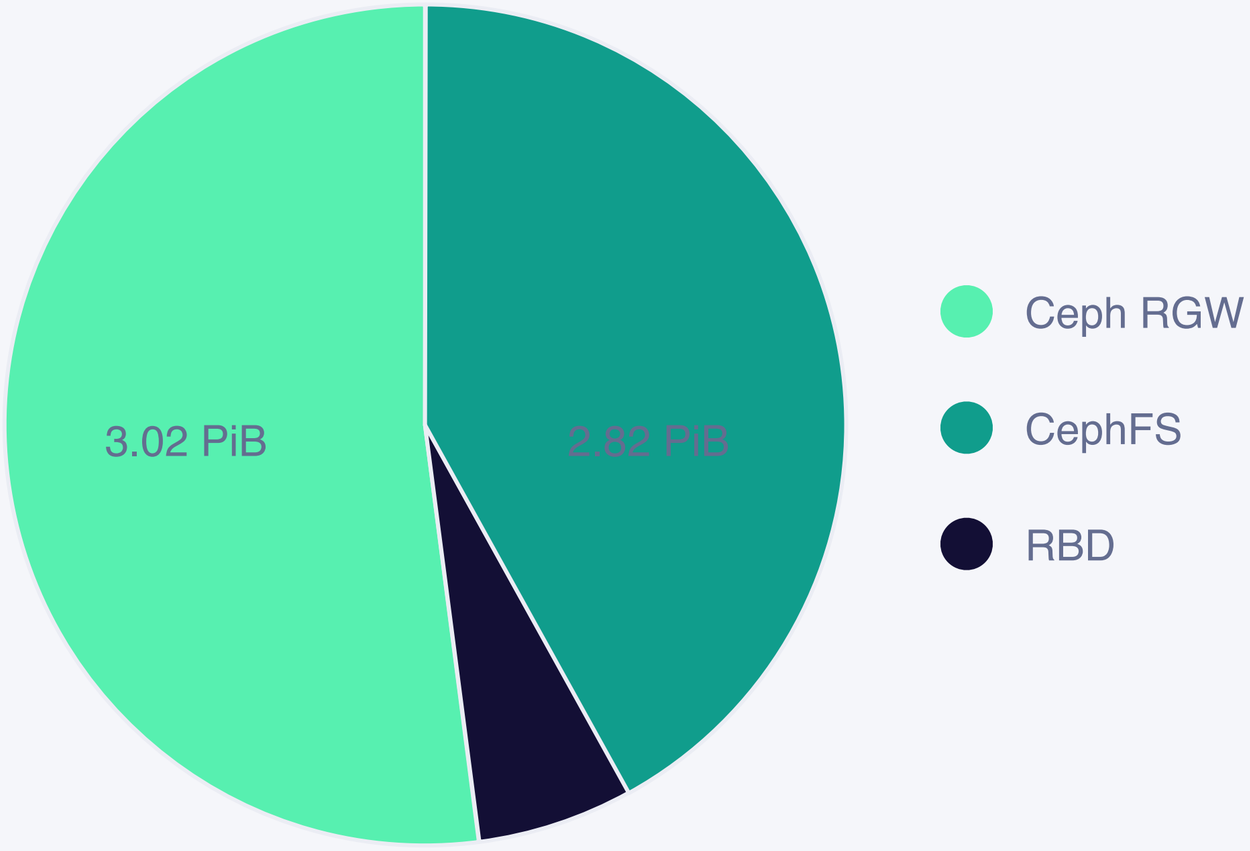
OSDs 384

Objects 1.05 G

Size

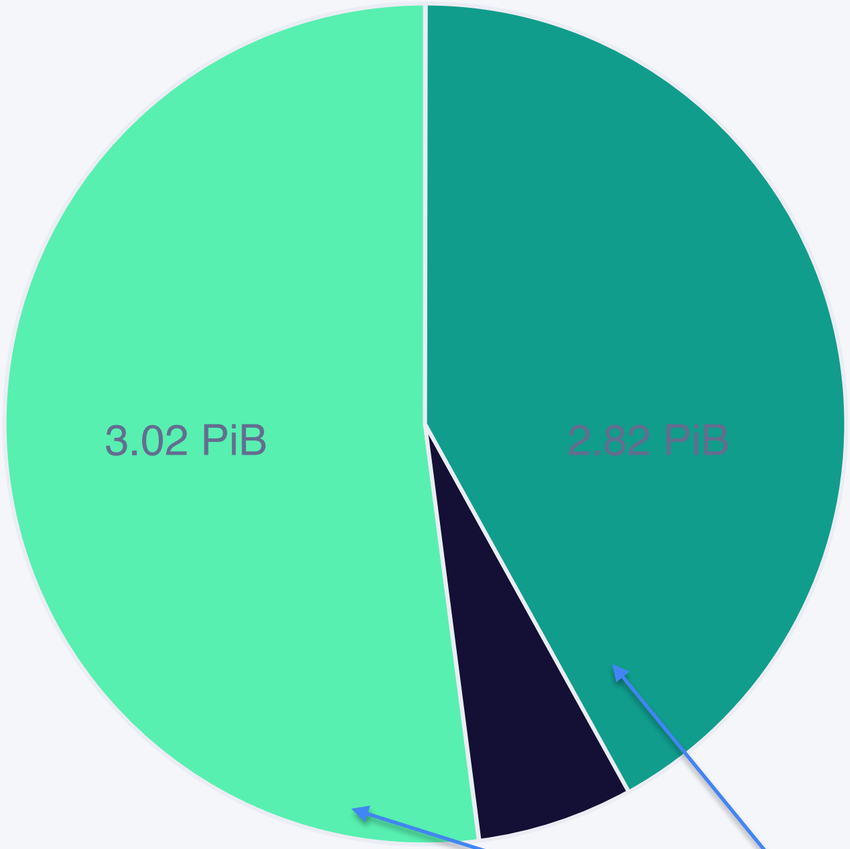


Storage usage



oxylabs®

Storage usage - Data Exposure Problem



Ceph RGW - data generated

copy

CephFS - data exposed to clients

RBD

Same datasets - 6x replication

200TB extra storage per month

oxylabs®

oxylabs®

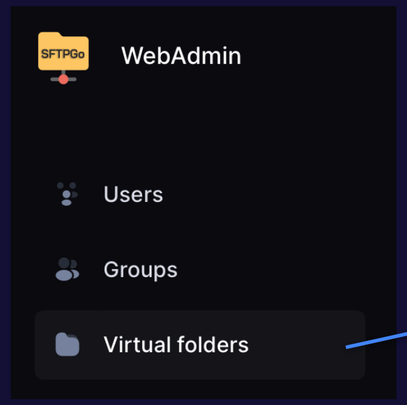
Solutions

SFTPGo
For RGW data exposure

ceph

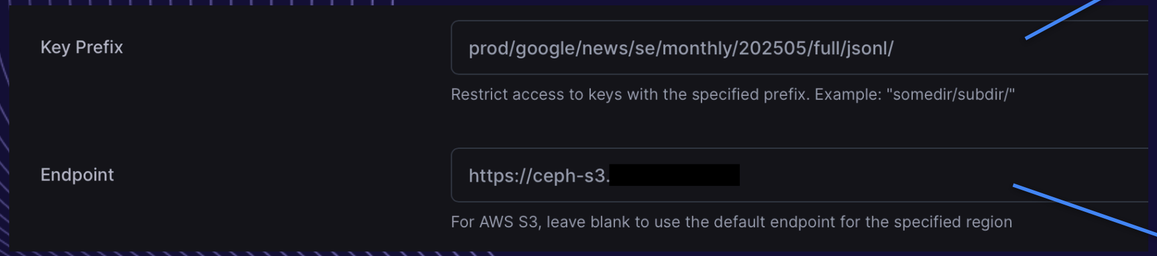
S3 Lifecycle
For consumed space

oxylabs®



SFTPGo

For RGW data exposure
Virtual S3 folders
Access controls



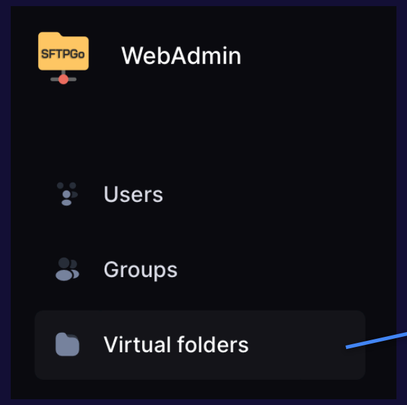
ceph

S3 Lifecycle

For consumed space
Storage classes
Compressed pools
Erasure coded pools

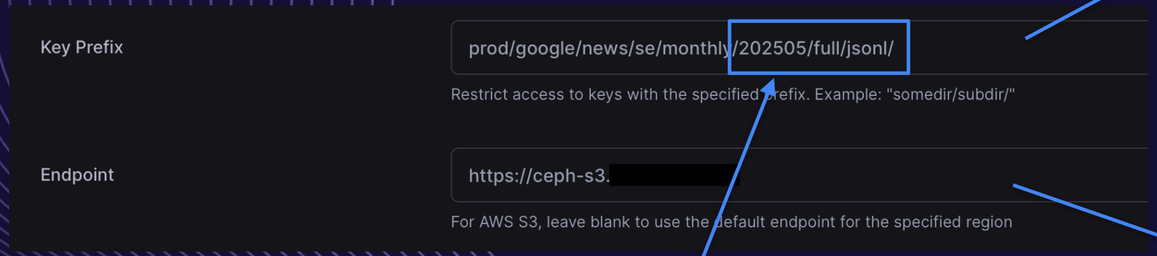
Solutions

oxylabs®



SFTPGo

For RGW data exposure
Virtual S3 folders
Access controls



New Problems

ceph

S3 Lifecycle

For consumed space
Storage classes
Compressed pools
Erasure coded pools

Problem statement

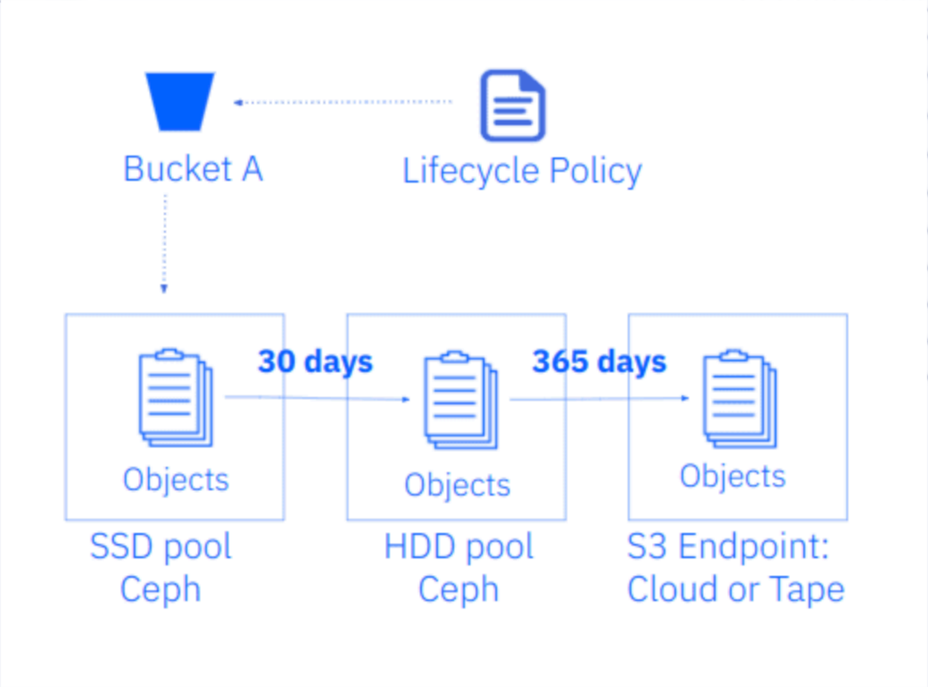
“Obsolete and old data, that needs to be kept for archival purposes,
Is exposed and visible to users,
When using storage classes.”

Problem background

“Too many tools already in
Use within our technology stack
Causing maintenance overhead.
We just don’t want any more of them
for data migrations.”

oxylabs®

Ceph cloud-s3 module



Ceph Blog: https://ceph.io/en/news/blog/2025/rgw-tiering-enhancements-part1/
Ceph Blog: https://ceph.io/en/news/blog/2025/rgw-tiering-enhancements-part2/
Authors: Daniel Parkes, Anthony D'Atri

oxylabs®

CLOUD TRANSITION

This feature enables transitioning S3 objects to a remote cloud service as part of object lifecycle via Storage Classes. The transition is unidirectional: data cannot be transitioned back from the remote zone. The goal of this feature is to enable data transition to multiple cloud providers. The currently supported cloud providers are those that are compatible with AWS (S3).

A special storage class of tier type `cloud-s3` or `cloud-s3-glacier` is used to configure the remote cloud S3 object store service to which data is transitioned. These are defined in terms of zonegroup placement targets and, unlike regular storage classes, do not need a data pool.

User credentials for the remote cloud object store service must be configured. Note that source ACLs will not be preserved. It is possible to map permissions of specific source users to specific destination users.

CLOUD STORAGE CLASS TIER TYPE

- `tier-type` (string)

  The type of remote cloud service that will be used to transition objects. The below tier types are supported:

  - `cloud-s3` : Regular S3 compatible object store service.
  - `cloud-s3-glacier` : S3 Glacier or Tape storage services.

Ceph Docs: https://docs.ceph.com/en/latest/radosgw/cloud-transition/

oxylabs®

CLOUD TRANSITION

This feature enables transitioning S3 objects to a remote cloud service as part of object lifecycle via Storage Classes. The transition is unidirectional: data cannot be transitioned back from the remote zone. The goal of this feature is to enable data transition to multiple cloud providers. The currently supported cloud providers are those that are compatible with AWS (S3).

A special storage class of tier type `cloud-s3` or `cloud-s3-glacier` is used to configure the remote cloud S3 object store service to which data is transitioned. These are defined in terms of zonegroup placement targets and, unlike regular storage classes, do not need a data pool.

User credentials for the remote cloud object store service must be configured. Note that source ACLs will not be preserved. It is possible to map permissions of specific source users to specific destination users.

CLOUD STORAGE CLASS TIER TYPE

- `tier-type` (string)

  The type of remote cloud service that will be used to transition objects. The below tier types are supported:

  - `cloud-s3` : Regular S3 compatible object store service.
  - `cloud-s3-glacier` : S3 Glacier or Tape storage services.

Ceph Docs: https://docs.ceph.com/en/latest/radosgw/cloud-transition/

oxylabs®

Create reduced storage pool (erasure coded, can be compressed or any other):
ceph osd pool create europe-1.rgw.buckets.erasure erasure

Initial setup

oxylabs®

Create reduced storage pool (erasure coded, can be compressed or any other):
ceph osd pool create europe-1.rgw.buckets.erasure erasure

Create a placement target (bog) within zonegroup:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id bog

Modify placement target within zone pointing to a reduced storage pool:
radosgw-admin zone placement add --rgw-zone europe-1 --placement-id bog --data-pool europe-1.rgw.buckets.erasure --index-pool europe-1.rgw.buckets.index

Initial setup

oxylabs®

Create reduced storage pool (erasure coded, can be compressed or any other):
ceph osd pool create europe-1.rgw.buckets.erasure erasure

Create a placement target (bog) within zonegroup:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id bog

Modify placement target within zone pointing to a reduced storage pool:
radosgw-admin zone placement add --rgw-zone europe-1 --placement-id bog --data-pool europe-1.rgw.buckets.erasure --index-pool europe-1.rgw.buckets.index

Create target S3 buckets (can be done through Ceph Dashboard):
aws s3api create-bucket \
    --bucket **bog-archive** \
    --endpoint-url=https://ceph-s3.your.endpoint \
    --create-bucket-configuration '{"LocationConstraint": "europe:bog"}' \
    --profile your-s3-profile-with-keys

Initial setup

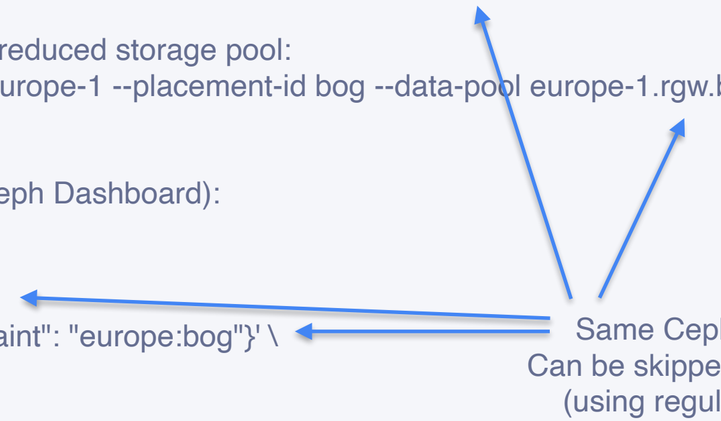
oxylabs®

Create reduced storage pool (erasure coded, can be compressed or any other):
ceph osd pool create europe-1.rgw.buckets.erasure erasure

Create a placement target (bog) within zonegroup:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id bog

Modify placement target within zone pointing to a reduced storage pool:
radosgw-admin zone placement add --rgw-zone europe-1 --placement-id bog --data-pool europe-1.rgw.buckets.erasure --index-pool europe-1.rgw.buckets.index

Create target S3 buckets (can be done through Ceph Dashboard):
aws s3api create-bucket \
    --bucket **bog-archive** \
    --endpoint-url=https://ceph-s3.your.endpoint \
    --create-bucket-configuration '{"LocationConstraint": "europe:bog"}' \
    --profile your-s3-profile-with-keys



Same Ceph cluster, erasure coded pool
Can be skipped if no storage reduction needed
(using regular pools in default-placement)

Initial setup

oxylabs®

Add a new Storage Class with cloud-s3 as the tier type:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG --tier-type cloud-s3**

Modify the Storage Class with tier configurations that point back to the same cluster and the new bucket:
radosgw-admin zonegroup placement modify --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG** --tier-config=**endpoint=https://ceph-s3.your.endpoint**,access_key=access-key,secret=secret-key,**target_path="bog-archive"**,retain_head_object=false

cloud-s3 setup

oxylabs®

Add a new Storage Class with cloud-s3 as the tier type:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG --tier-type cloud-s3**

Modify the Storage Class with tier configurations that point back to the same cluster and the new bucket:
radosgw-admin zonegroup placement modify --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG** --tier-config=**endpoint=https://ceph-s3.your.endpoint**,access_key=access-key,secret=secret-key,**target_path="bog-archive"**,retain_head_object=false

In stead of AWS S3 (or other) - point it back to the same Ceph cluster RGW endpoint

cloud-s3 setup

oxylabs®

Add a new Storage Class with cloud-s3 as the tier type:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG --tier-type cloud-s3**

Modify the Storage Class with tier configurations that point back to the same cluster and the new bucket:
radosgw-admin zonegroup placement modify --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG** --tier-config=**endpoint=https://ceph-s3.your.endpoint**,access_key=access-key,secret=secret-key,**target_path="bog-archive"**,retain_head_object=false

Set to **false** to remove metadata visibility

In stead of AWS S3 (or other) - point it back to the same Ceph cluster RGW endpoint

cloud-s3 setup

oxylabs®

Add a new Storage Class with cloud-s3 as the tier type:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG --tier-type cloud-s3**

Modify the Storage Class with tier configurations that point back to the same cluster and the new bucket:
radosgw-admin zonegroup placement modify --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG** --tier-config=**endpoint=https://ceph-s3.your.endpoint**,access_key=access-key,secret=secret-key,**target_path="bog-archive"**,retain_head_object=false

Set to **false** to remove metadata visibility

Bucket in reduced storage pool from Initial setup
Use regular pools and buckets for data movement only

In stead of AWS S3 (or other) - point it back to the same Ceph cluster RGW endpoint

cloud-s3 setup

Add a new Storage Class with cloud-s3 as the tier type:
radosgw-admin zonegroup placement add --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG --tier-type cloud-s3**

Modify the Storage Class with tier configurations that point back to the same cluster and the new bucket:
radosgw-admin zonegroup placement modify --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG** --tier-config=**endpoint=https://ceph-s3.your.endpoint**,access_key=access-key,secret=secret-key,**target_path="bog-archive"**,retain_head_object=false

Create a S3 bucket lifecycle rule
pointing to the new Storage Class:

```
<LifecycleConfiguration>
   <Rule>
      <ID>TransitionCSVToBOG</ID>
      <Status>Enabled</Status>
      <Filter>
         <Tag>
            <Key>file_type</Key>
            <Value>csv</Value>
         </Tag>
      </Filter>
      <Transition>
         <Days>30</Days>
         <StorageClass>BOG</StorageClass>
      </Transition>
   </Rule>
</LifecycleConfiguration>
```

Set to **false** to remove metadata visibility

In stead of AWS S3 (or other) - point it back to the same Ceph cluster RGW endpoint

cloud-s3 setup

radosgw-admin zonegroup placement modify --rgw-zonegroup europe --placement-id default-placement --storage-class **BOG** --tier-config=**endpoint=https://ceph-s3.your.endpoint**,access_key=access-key,secret=secret-key,**target_path="bog-archive"**,retain_head_object=false

Create a S3 bucket lifecycle rule pointing to the new Storage Class:

```
<LifecycleConfiguration>
   <Rule>
      <ID>TransitionCSVToBOG</ID>
      <Status>Enabled</Status>
      <Filter>
         <Tag>
            <Key>file_type</Key>
            <Value>csv</Value>
         </Tag>
      </Filter>
      <Transition>
         <Days>30</Days>
         <StorageClass>BOG</StorageClass>
      </Transition>
   </Rule>
</LifecycleConfiguration>
```

```
T14:37:53.508+0000 7f97e66786c0  0 lifecycle: Transitioning object(X4H7BH_custom_member_id.csv) to the cloud endpoint(https://ceph-s3.          )
T14:38:01.080+0000 7f97e767a6c0  0 lifecycle: Transitioning object(XLWWKP_multisource_ids_websites2.csv) to the cloud endpoint(https://ceph-s3.          )
T14:38:01.368+0000 7f97e767a6c0  0 lifecycle: Transitioning object(TOYCIT_profile_ids.csv) to the cloud endpoint(https://ceph-s3.          )
T14:38:01.448+0000 7f97e767a6c0  0 lifecycle: Transitioning object(OVYENE_clean_members2.csv) to the cloud endpoint(https://ceph-s3.          )
T14:38:01.732+0000 7f97e767a6c0  0 lifecycle: Transitioning object(M3CHXK_clean_member_id_10k.csv) to the cloud endpoint(https://ceph-s3.          )
T14:38:01.856+0000 7f97e767a6c0  0 lifecycle: Transitioning object(HO5WNE_profile_results_member_not_found_ids.csv) to the cloud endpoint(https://ceph-s3.          )
T14:38:01.916+0000 7f97e767a6c0  0 lifecycle: Transitioning object(AJVYYX_member_id_hash_matching.csv) to the cloud endpoint(https://ceph-s3.          )
```

In stead of AWS S3 (or other) - point it back to the same Ceph cluster RGW endpoint

cloud-s3 setup

oxylabs®

```
{
  "Rules": [
    {
      "ID": "AbortIncompleteMultipartUploads",
      "Status": "Enabled",
      "AbortIncompleteMultipartUpload": {
        "DaysAfterInitiation": 3
      },
      "Filter": {}
    },
    {
      "ID": "ExpireJSONLFiles",
      "Status": "Enabled",
      "Expiration": {
        "Days": 1
      },
      "Filter": {
        "Tag": {
          "Key": "dump_file_type",
          "Value": "jsonl"
        }
      }
    },
    {
      "ID": "ExpireCSVFiles",
      "Status": "Enabled",
      "Expiration": {
        "Days": 1
      },
      "Filter": {
        "Tag": {
          "Key": "dump_file_type",
          "Value": "csv"
        }
      }
    },
    {
      "ID": "TransitionParquetToBOG",
      "Status": "Enabled",
      "Transitions": [
        {
          "Days": 1,
          "StorageClass": "BOG"
        }
      ],
      "Filter": {
        "Tag": {
          "Key": "dump_file_type",
          "Value": "parquet"
        }
      }
    }
  ]
}
```


```
root@analyzer ~ #
root@analyzer ~ # date
Wed Apr 16 01:03:37 PM UTC 2025
root@analyzer ~ # mc ls dp/dp-laimis-data-raw/
[2025-04-16 13:04:02 UTC]     0B 202504/
root@analyzer ~ # mc ls dp/dp-laimis-data-raw/202504/
[2025-04-16 13:04:16 UTC]     0B csv/
[2025-04-16 13:04:16 UTC]     0B jsonl/
[2025-04-16 13:04:16 UTC]     0B parquet/
root@analyzer ~ # mc du dp/dp-laimis-data-raw/202504/csv/
93GiB   152 objects     dp-laimis-data-raw/202504/csv
root@analyzer ~ # mc du dp/dp-laimis-data-raw/202504/jsonl/
84GiB   719 objects     dp-laimis-data-raw/202504/jsonl
root@analyzer ~ # mc du dp/dp-laimis-data-raw/202504/parquet/
67GiB   721 objects     dp-laimis-data-raw/202504/parquet
root@analyzer ~ #
root@analyzer ~ # mc ls dp/dp-laimis-data-raw/202504/csv/                  | tail -1
[2025-04-15 14:10:51 UTC] 415MiB STANDARD part-00000-2215b3ac-ecad-4b05-ac96-030d6147c69d-c000.csv.gz
root@analyzer ~ # mc ls dp/dp-laimis-data-raw/202504/jsonl/partition_by_column=lithuania/ | tail -1
[2025-04-15 14:11:54 UTC] 109MiB STANDARD part-00081-633c9e97-9893-496c-acb9-82c20c11a489.c000.json.gz
root@analyzer ~ # mc ls dp/dp-laimis-data-raw/202504/parquet/partition_by_column=lithuania/ | tail -1
[2025-04-15 14:14:42 UTC]  88MiB STANDARD part-00081-7c0eaf61-ac6a-45d5-94b8-bae69b2146e4.c000.gz.parquet
root@analyzer ~ #
root@analyzer ~ # mc ls dp/dp-laimis-data-bog/
root@analyzer ~ # mc du dp/dp-laimis-data-bog/
0B      0 objects       dp-laimis-data-bog
root@analyzer ~ #
```

Source bucket with data in csv, jsnol and parquet formats

Objects are older than 1 day - lifecycle should pick up next midnight

Target bucket empty

Source bucket dp-laimis-data-raw bucket with csv, json and parquet files
Target bucket dp-laimis-data-bog empty

oxylabs®

```
{
  "Rules": [
    {
      "ID": "AbortIncompleteMultipartUploads",
      "Status": "Enabled",
      "AbortIncompleteMultipartUpload": {
        "DaysAfterInitiation": 3
      },
      "Filter": {}
    },
    {
      "ID": "ExpireJSONLFiles",
      "Status": "Enabled",
      "Expiration": {
        "Days": 1
      },
      "Filter": {
        "Tag": {
          "Key": "dump_file_type",
          "Value": "jsonl"
        }
      }
    },
    {
      "ID": "ExpireCSVFiles",
      "Status": "Enabled",
      "Expiration": {
        "Days": 1
      },
      "Filter": {
        "Tag": {
          "Key": "dump_file_type",
          "Value": "csv"
        }
      }
    },
    {
      "ID": "TransitionParquetToBOG",
      "Status": "Enabled",
      "Transitions": [
        {
          "Days": 1,
          "StorageClass": "BOG"
        }
      ],
      "Filter": {
        "Tag": {
          "Key": "dump_file_type",
          "Value": "parquet"
        }
      }
    }
  ]
}
```


```
root@analyzer ~ #
root@analyzer ~ # date
Thu Apr 17 12:59:54 PM UTC 2025
root@analyzer ~ # mc ls dp/dp-laimis-data-raw
root@analyzer ~ # mc du dp/dp-laimis-data-raw
0B      0 objects       dp-laimis-data-raw
root@analyzer ~ #
root@analyzer ~ # mc ls dp/dp-laimis-data-bog/
[2025-04-17 13:00:24 UTC]     0B dp-laimis-data-raw/
root@analyzer ~ # mc du dp/dp-laimis-data-bog/
67GiB   721 objects     dp-laimis-data-bog
root@analyzer ~ # mc ls dp/dp-laimis-data-bog/dp-laimis-data-raw/
[2025-04-17 13:00:44 UTC]     0B 202504/
root@analyzer ~ # mc ls dp/dp-laimis-data-bog/dp-laimis-data-raw/202504/
[2025-04-17 13:00:56 UTC]     0B parquet/
root@analyzer ~ #
root@analyzer ~ #
root@analyzer ~ #
```

Source bucket empty the day after

Target (BOG) bucket containing source bucket folder with only parquet objects/files

Source bucket dp-laimis-data-raw bucket empty
Target bucket dp-laimis-data-bog containing parquet files after Ceph RGW lifecycle processing

Note: here dp-laimis-data-bog bucket is set as BOG StorageClass

oxylabs®

Problem solved

“Old and obsolete data

No longer exposed,

And moved to archive buckets

By Ceph internals.”

oxylabs®

Considerations

oxylabs®

1. Adjust RGW lifecycle processing aggressiveness
   rgw_lc_max_worker for large number of buckets
   rgw_lc_max_wp_worker for large number of objects

Considerations

oxylabs®

1. Adjust RGW lifecycle processing aggressiveness
   rgw_lc_max_worker for large number of buckets
   rgw_lc_max_wp_worker for large number of objects

2. Note RGW lifecycle processing hours
   rgw_lifecycle_work_time changes require RGW daemon restarts for the whole cluster

Considerations

oxylabs®

1. Adjust RGW lifecycle processing aggressiveness
   rgw_lc_max_worker for large number of buckets
   rgw_lc_max_wp_worker for large number of objects

2. Note RGW lifecycle processing hours
   rgw_lifecycle_work_time changes require RGW daemon restarts for the whole cluster

3. Adjust incomplete mp cleanup for target buckets
   Regular expiry values can interfere with large amounts of data for unfinished lifecycle processes

Considerations

oxylabs®

1. Adjust RGW lifecycle processing aggressiveness
   rgw_lc_max_worker for large number of buckets
   rgw_lc_max_wp_worker for large number of objects

2. Note RGW lifecycle processing hours
   rgw_lifecycle_work_time changes require RGW daemon restarts for the whole cluster

3. Adjust incomplete mp cleanup for target buckets
   Regular expiry values can interfere with large amounts of data for unfinished lifecycle processes

4. Every target bucket is a StorageClass
   tier_targets list within zonegroup placement can grow rapidly when using a large amount of buckets

Considerations

oxylabs®

Thank you!