# IBM COS → Domino Dataset Synchronization

\*Page owner:\* {{@your-team}}

## 1 Overview

This script incrementally synchronises objects stored in an IBM Cloud Object Storage (COS) bucket to a Domino dataset. It is designed for \*one-way\* mirroring (COS ➜ Domino) and can be called manually or via \*cron\*. Only new or updated objects are downloaded, minimising bandwidth and execution time.

* Current dataset: \*\*cos-edd-sync\*\*
* Default target path inside a workspace: \*/mnt/dataset/XXX\* (override with `DATASET\_DIR`).

## 2 Quick start

1. Ensure the dataset \*\*cos-edd-sync\*\* is mounted in your workspace.
2. Install requirements once:

`{% raw %}`pip install ibm-cos-sdk tqdm{% endraw %}`

1. Set the mandatory environment variables and run:

export COS\_ENDPOINT="https://s3.eu-de.cloud-object-storage.appdomain.cloud"  
export COS\_ACCESS\_KEY\_ID="<access>" # HMAC key  
export COS\_SECRET\_ACCESS\_KEY="<secret>" # HMAC secret  
export COS\_BUCKET\_NAME="<bucket>"  
export COS\_PREFIX="domino/" # or leave blank  
python3 cos\_sync.py --mode sync

## 3 Environment variables

|  |  |  |
| --- | --- | --- |
| Variable | Required | Purpose |
| ---------- | ---------- | --------- |
| `COS\_ENDPOINT` | ✔ | COS endpoint URL (region-specific) |
| `COS\_ACCESS\_KEY\_ID` | ✔ | HMAC access key |
| `COS\_SECRET\_ACCESS\_KEY` | ✔ | HMAC secret key |
| `COS\_BUCKET\_NAME` | ✔ | Name of the bucket to mirror |
| `COS\_PREFIX` |  | Limit sync to this prefix (e.g. `domino/`) |
| `COS\_OBJECT\_KEYS` |  | Comma-separated list of individual keys |
| `COS\_KEYS\_FILE` |  | File containing one key per line |
| `DATASET\_DIR` |  | Local destination folder (default `/mnt/dataset/XXX`) |

`COS\_OBJECT\_KEYS`, `COS\_KEYS\_FILE` and `COS\_PREFIX` are \*mutually exclusive\* – set \*\*one\*\* only.

## 4 Modes

* \*\*sync\*\* (default) – mirror COS to the dataset.
* \*\*tree\*\* – print an indented object tree (useful to inspect prefixes).
* \*\*map\*\* – print a \*source ➜ destination\* mapping table.

Run with, for example:

`python3 cos\_sync.py --mode tree`

## 5 Scheduling with cron

Synchronise every hour:

0 \* \* \* \* COS\_ENDPOINT=… COS\_ACCESS\_KEY\_ID=… COS\_SECRET\_ACCESS\_KEY=… \  
 COS\_BUCKET\_NAME=… COS\_PREFIX=domino/ \  
 /usr/bin/python3 /opt/scripts/cos\_sync.py --mode sync >> /var/log/cos\_sync.log 2>&1

Logs are appended in \*/var/log/cos\_sync.log\*.

## 6 Extending / Re-using

* Download \*selected\* keys only – define `COS\_OBJECT\_KEYS` / `COS\_KEYS\_FILE`.
* Change destination folder (`DATASET\_DIR`) to create multiple mirrors.
* Combine with \*\*Vault\*\* or \*\*Domino environment variables\*\* to keep credentials out of source control.

## 7 Troubleshooting

|  |  |  |
| --- | --- | --- |
| Symptom | Likely cause | Fix |
| --------- | -------------- | ----- |
| `404 NoSuchKey` | Wrong prefix / key list includes bucket name | Use `--mode tree` to inspect exact keys and adjust `COS\_PREFIX` or key list |
| No files downloaded | `COS\_PREFIX` too restrictive or credentials lack \*List\* permission | Verify IAM/HVault policy and prefix |
| Slow sync | Very large objects; first run downloads everything | Subsequent runs transfer only deltas |

## 8 Repository

The script lives in GitLab: \*{{link\_to\_repo}}\*

Feel free to open merge requests for improvements or additional features.

## 9 Maintainers

\*Primary contact:\* {{@your-name}}

\*Last updated:\* {date}