RELIABLE DATA DISASTER RECOVERY WITH LYVE CLOUD

With Lyve Cloud, the University of Virginia has a secure primary storage and disaster recovery site for its extensive faculty, student, and research data.

SEAGATE
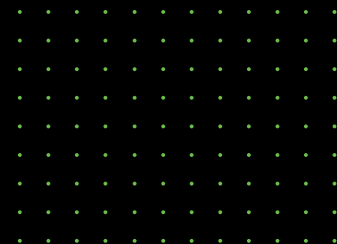**LYVE**™Cloud

SEAGATE

CONTENTS

Introduction

The University of Virginia (UVA)—a public research university and major academic medical center founded in 1819 and located in Charlottesville, Virginia—created its Department of Applied Math and Computer Science in 1965 and was granted one of the first computer science PhDs in the country in 1968. The UVA Department of Computer Science became an independent entity in 1984. With significant strengths in the three core areas—computer systems, theory and artificial intelligence—the department fosters highly productive collaborations that have led to breakthroughs in cyber-physical systems, cybersecurity, machine learning and high-performance computing among other areas. UVA applies these discoveries to address some of the most critical societal challenges of the 21st century, in fields such as healthcare, neuroscience, and sustainability. These innovations are the result of extensive research in computer systems, networks, the internet of things (IoT), and security.

The data-heavy demands of this innovative work had UVA searching for a solution that would support the computer science department's expansive storage and recovery needs. The offering had to be reliable, secure, easy to use, and budget friendly (meaning no hidden fees). With Seagate® Lyve™ Cloud, the UVA team experienced another highly productive partnership.

University of Virginia Storage and Recovery Needs

The UVA Department of Computer Science is a nationally recognized leader in computer science research and education, and includes 60 faculty members, 240 graduate students and 3,961 undergraduates. Faculty members are active in essential computer science and engineering research and engaged in interdisciplinary initiatives with scientists, humanists, engineers in other disciplines, and medical personnel. The graduate students transition from students to collaborators by participating in world-class research projects. And the school's undergraduates are involved in research and enriched by an integrated and challenging curriculum. To maintain its focus on all of these projects, UVA requires a highly reliable data storage and recovery solution with the following features:

- A dependable, easy-to-use offsite storage and disaster recovery system for data used by students and faculty, as well as stored research data

- A data-loss prevention plan for all data types, no matter the location (on site, on-premise and/or with the user)

- The ability to recover and restore damaged or destroyed data without slowing down access to it

- A way to accomplish all of this without breaking the department's budget

Seagate Lyve Cloud Benefits

Losing data to human errors, disaster events, or ransomware can destroy entire organizations. With Lyve Cloud, the UVA's mass data is protected should the university experience a data loss incident. Seagate's storage-only cloud is the ideal repository for the computer science department's backup and disaster recovery because it offers multi-region availability, replication, and object immutability. Additional benefits include:

- **Data Analytics:** The UVA team can access their research and project data via a reliable, secure, highly available and flexible solution that complements the university's existing multicloud strategy.

- **No Hidden Fees:** Lyve Cloud allows the seamless transfer of data across public and private cloud environments without extra charges for egress or API calls. This predictable capacity-based pricing helps lower cloud storage costs up to by as much as 70% with industry-leading total cost of ownership (TCO).

- **World-Class Security:** Lyve Cloud adheres to the most stringent, globally recognized data security standards and is ISO 27001:2013 and SOC2® certified. Data is safeguarded with ransomware protection, object immutability, enterprise-grade identity management support, data encryption at rest and in flight, and automatic data replication.

- **Script Run Support:** Local data is mirrored to remote storage in Lyve Cloud exactly as it exists when a copy or script is run, ensuring the latest data is accessed and the team members can manipulate local and remote data, and transfer between Lyve Cloud and other systems.

UVA Data Backup Overview

To establish its disaster recovery plan, the university adopted a multi-step backup process that includes data authentication and snapshots, and transferring data. The team employs the following software and file systems to backup, recover, and work with their data.

Rclone

The computer science department uses rclone as the primary software application for interacting with the bucket in Lyve Cloud. It allows the team to carry out data operations and shift the data between the university and Lyve Cloud systems. Rclone uses two primary commands to accomplish this:

1. Copy: This initially creates the remote copy of local data and syncs it.

2. Sync: When this runs, remote data is altered to it is an exact match of what is found locally.

Zeta File System (ZFS)

ZFS is the university's file system choice for users' home and project directories. It creates snapshots of the data, providing a complete view of the stored data at the time the capture. Additional storage is not required since the ZFS saves its pointers to blocks of data when these snapshots are taken.

Process Details

Rclone Authentication

Users load a server to run the rclone commands to transfer data into the Lyve Cloud bucket. After creating a bucket, permission and service account—and receiving the access and secret key within Lyve Cloud—the configuration run through begins. Below the various steps are outlined.

```
$ rclone config
No remotes found, make a new one?
n) New remote
s) Set configuration password
q) Quit config
n/s/q> n
name> remote
```

1. Choose the preferred S3 backend

```
Type of storage to configure.
Choose a number from below, or type in your own value.
[snip]
XX / Amazon S3 Compliant Storage Providers including AWS, Alibaba, Ceph,
ChinaMobile, ArvanCloud, Digital Ocean, Dreamhost, Huawei OBS, IBM COS, Lyve
Cloud, Minio, RackCorp, SeaweedFS, and Tencent COS
    \ (s3)
[snip]
Storage> s3
```

2. Select Lyve Cloud as the S3 provider

```
Choose your S3 provider.
Choose a number from below, or type in your own value.
Press Enter to leave empty.
[snip]
XX / Seagate Lyve Cloud
    \ (LyveCloud)
[snip]
provider> LyveCloud
```
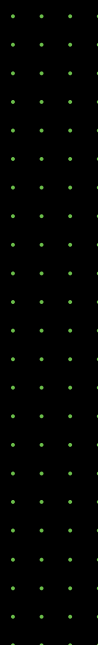
3. Press enter to load the access key and secret key (taken from Lyve Cloud) in the configuration file

```
Get AWS credentials from runtime (environment variables or EC2/ECS meta
data)
Only applies if access_key_id and secret_access_key is blank.
Choose a number from below, or type in your own boolean value (true or
false).
Press Enter for the default (false).
 1 / Enter AWS credentials in the next step.
   \ (false)
 2 / Get AWS credentials from the environment (env vars or IAM).
   \ (true)
env_auth>
```

4. Note the AWS access key ID

```
Leave blank for anonymous access or runtime credentials.
Enter a value. Press Enter to leave empty.
access_key_id> XXX
```

5. Complete AWS secret access key (password)

```
Leave blank for anonymous access or runtime credentials.
Enter a value. Press Enter to leave empty.
secret_access_key> YYY
```

6. Leave region section blank

```
Region to connect to.
Leave blank if you are using an S3 clone and you don't have a region.
Choose a number from below, or type in your own value.
Press Enter to leave empty.
   / Use this if unsure.
 1 | Will use v4 signatures and an empty region.
   \ ()
   / Use this only if v4 signatures don't work.
 2 | E.g. pre Jewel/v10 CEPH.
   \ (other-v2-signature)
region>
```

7. Select preferred endpoint from the list

```
Endpoint for S3 API.
Required when using an S3 clone.
Choose a number from below, or type in your own value.
Press Enter to leave empty.
 1 / Seagate Lyve Cloud US East 1 (Virginia)
   \ (s3.us-east-1.lyvecloud.seagate.com)
 2 / Seagate Lyve Cloud US West 1 (California)
   \ (s3.us-west-1.lyvecloud.seagate.com)
 3 / Seagate Lyve Cloud AP Southeast 1 (Singapore)
   \ (s3.ap-southeast-1.lyvecloud.seagate.com)
endpoint> 1
```

### 8. Leave location constraint blank

```
Location constraint - must be set to match the Region.
Leave blank if not sure. Used when creating buckets only.
Enter a value. Press Enter to leave empty.
location_constraint>
```

9. Choose default ACL

```
Canned ACL used when creating buckets and storing or copying objects.
This ACL is used for creating objects and if bucket_acl isn't set, for
creating buckets too.
For more info visit https://docs.aws.amazon.com/AmazonS3/latest/dev/acl-
overview.html#canned-acl
Note that this ACL is applied when server-side copying objects as S3
doesn't copy the ACL from the source but rather writes a fresh one.
Choose a number from below, or type in your own value.
Press Enter to leave empty.
   / Owner gets FULL_CONTROL.
 1 | No one else has access rights (default).
   \ (private)
[snip]
acl>
```
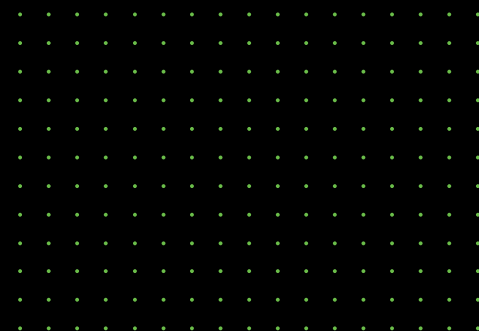
10.   Once configured, the file should look like following:

```
[remote]
type = s3
provider = LyveCloud
access_key_id = XXX
secret_access_key = YYY
endpoint = s3.us-east-1.lyvecloud.seagate.com
```

Snapshots and Data Transfer

Once authenticated, ZFS begins taking snapshots of the user directories and uses rclone to move that data. This allows users to work simultaneously without interfering with rclone's processes. Rclone will direct all data through its two commands into the Lyve Cloud bucket. In other words, when an rclone sync/copy command is executed, the remote data acts as a mirror of the local file system at the time of the snapshot.

Rclone itself provides several options for adjusting transfers, such as modifying the maximum duration and amount allowed to transfer, disabling symbolic links by converting them into a different file type (while maintaining the pointer), and adjusting the number of parallel transfers that can occur. Testing within the UVA computer science department revealed 32 parallel transfers were possible, with speeds reaching 10-12MB per second, per transfer. After completing this testing, the solution used at most 16 parallel transfers because

anything greater was unnecessary.

This first pseudocode outlines the overall process for running rclone. ZFS snapshots are used, and directories are listed and iterated one at a time, allowing equal time for each user's home directory to be allocated. Alternatively, this can be run at the top level of a directory.

```
// initialize a list of directories to copy/sync.
// note, exact syntax is not used and the trailing '/' refers to the contents of
// of a given directory.
dirs_list = zfs_snapshot/

// iterate over each directory running an individual rclone copy/sync command
// for each.
for (dir_to_backup in dirs_list):
    rclone <copy or sync> dir_to_backup/ remote-bucket:remote-path
    --transfers=8
    --checkers=8
    --max-duration=60m
```

It should be noted that the destination within the remote bucket has the same name as the local directory for clarity purposes. The first two of the three flags used (--transfers=8, --checkers=8) causes rclone to use eight CPU cores in parallel for transferring files and eight cores for running hash sums. Additionally, transfers and checks can occur simultaneously.

This iteration across directories means uniform time allocations for each. The drawback to this approach is that once local directories are deleted, they will not be removed remotely. However, this can be remedied by running rclone sync on the parent directory instead of the individual ones. Since the sync command causes the remote directory to match the local exactly as it exists, this will result in the automatic removal of the directories. The tradeoff is not being able to specify a time allocation for each individual directory.
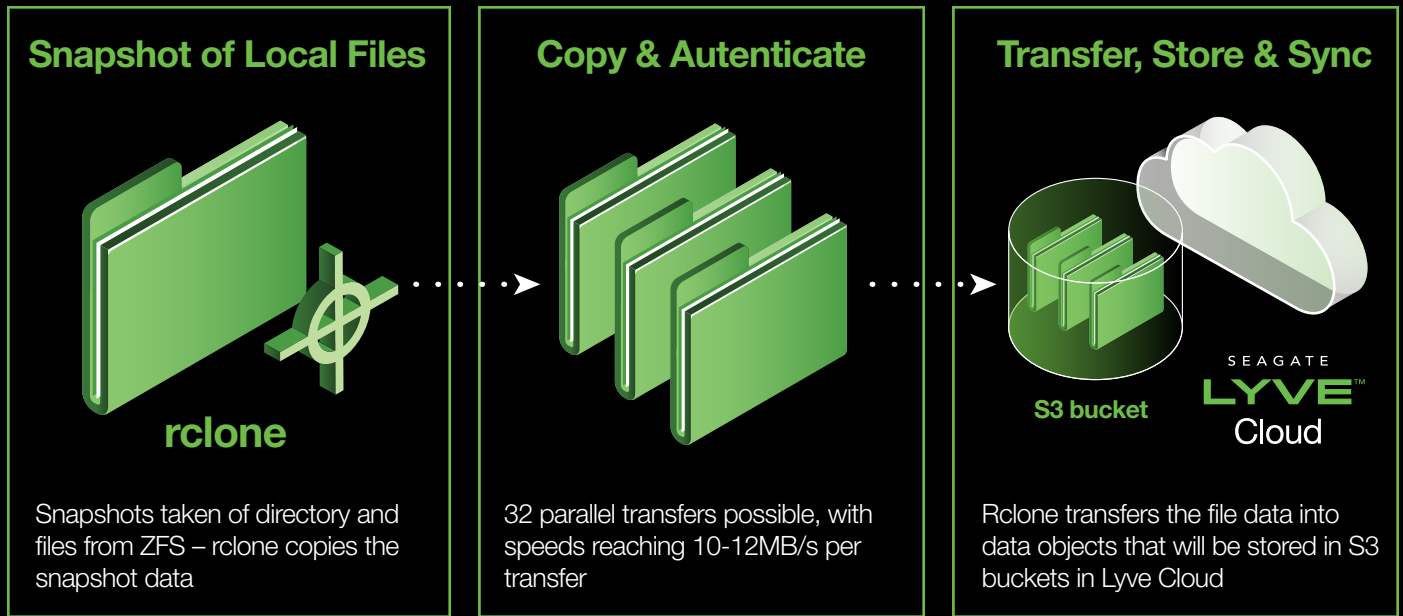
A simple solution is to compare the remote structure against the local. If a directory exists remotely but not locally, then it is deleted, as shown in pseudocode example 2:

```
// create a list of remote directories and local
// note, '>' describes redirecting output to a file
rclone lsd remote-bucket:remote-path > remote_list

for (remote_dir in remote_list):
    if (remote_dir not in local_filesystem):
        rclone delete remote-bucket:remote-path/remote_dir
```

University of Virgina Process

Snapshot of Local Files

rclone

Snapshots taken of directory and files from ZFS – rclone copies the snapshot data

Copy & Autenticate

32 parallel transfers possible, with speeds reaching 10-12MB/s per transfer

Transfer, Store & Sync

S3 bucket

SEAGATE
LYVE™
Cloud

Rclone transfers the file data into data objects that will be stored in S3 buckets in Lyve Cloud

Conclusion

The University of Virginia relies on Seagate Lyve Cloud as its disaster recovery site to securely duplicate data in the event of data damage or destruction. With this solution in place, the computer science department can employ software and file systems to backup, recover, and work with their data no matter where it resides. With Lyve Cloud, UVA has a highly dependable, easy-to-use, secure offsite storage and disaster recovery system... all without hidden fees.

Ready to Learn More? Visit us at seagate.com

Learn more about the steps to authenticate rclone with Lyve Cloud at
https://rclone.org/s3/#lyve

Read more about the UVA computer science department at
https://engineering.virginia.edu/departments/computer-science

seagate.com

SEAGATE