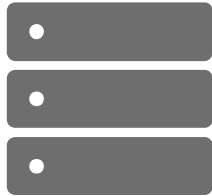# Fixity in the Cloud

Buzz Hays
Global Lead
Media & Entertainment Solutions
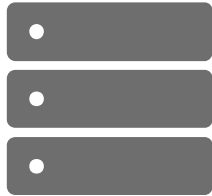
How to automate fixity checks on GCP
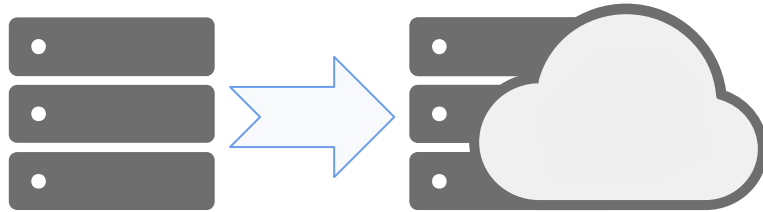
```
$ openssl dgst -md5 -binary ArchiveFile.mov openssl enc -base64
```

Google Cloud

How to automate fixity checks on GCP

```
$ openssl dgst -md5 -binary ArchiveFile.mov openssl enc -base64
$ apnFdauH+MfR7R5S5+NJzg==
```
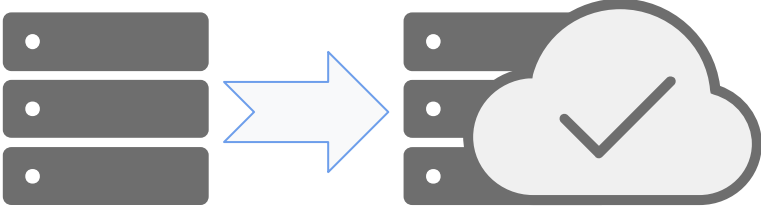
Google Cloud

How to automate fixity checks on GCP

apnFdauH+MfR7R5S5+NJzg==

Google Cloud

How to automate fixity checks on GCP

apnFdauH+MfR7R5S5+NJzg==   apnFdauH+MfR7R5S5+NJzg==

Google Cloud

How to automate fixity checks on GCP

```
$ gsutil stat gs://my-archive-bucket/ArchiveFile.mov
```

Google Cloud

How to automate fixity checks on GCP

```
$ gsutil stat gs://my-archive-bucket/ArchiveFile.mov

gs://my-archive-bucket/ArchiveFile.mov:
    Creation time:              Mon, 09 Sep 2019 11:22:13 GMT
    Update time:                Mon, 09 Sep 2019 11:22:13 GMT
    Storage class:              COLDLINE
    ...
    Hash (crc32c):              GRht3A==
    Hash (md5):                 apnFdauH+MfR7R5S5+NJzg==
    ...
```

Google Cloud

How to automate fixity checks on GCP

```
$ bin/verify_fixity_obj.sh
Verifying fixity for: gs://fixity-bucket/ArchiveFile.mov
 ObjectMeta MD5 Hash:    apnFdauH+MfR7R5S5+NJzg==
 Last Check MD5 Hash:    apnFdauH+MfR7R5S5+NJzg==
 Downloaded MD5 Hash:    apnFdauH+MfR7R5S5+NJzg==
   Hashes match, or last hash is empty.
   Storing new fixity check record.
```

Google Cloud

Where to store hashes?

Cloud SQL

BigQuery

Google Cloud

NDSA Levels of Digital Preservation

	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
	Protect your data	Know your data	Monitor your data	Repair your data
Storage Location	2 copies	3 copies (min)	DR on 1 copy (min)	DR all copies
Fixity	On ingest	On ingest, write-block, virus	Periodic checks, logs, detect corrupts, virus	Check on events, repair/replace, write access +1
Security	Authorized Access	Access restrictions	Logs of access	Audit Logs
Metadata	Inventory & Location	Store admin & transform, logs	Standard, technical, descriptive metadata	Preservation Metadata
File Formats	Limit formats	Inventory of formats	Monitor obsolescence	Format migrations

NDSA Levels of Digital Preservation

	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
	Protect your data	Know your data	Monitor your data	Repair your data
Storage Location	2 copies	3 copies (min)	DR on 1 copy (min)	DR all copies
Fixity	On ingest	On ingest, write-block, virus	Periodic checks, logs, detect corrupts, virus	Check on events, repair/replace, write access +1
Security	Authorized Access	Access restrictions	Logs of access	Audit Logs
Metadata	Inventory & Location	Store admin & transform, logs	Standard, technical, descriptive metadata	Preservation Metadata
File Formats	Limit formats	Inventory of formats	Monitor obsolescence	Format migrations

Google Cloud

How are cloud costs calculated for fixity?

Standard (baseline):

12 * $0.02 GB/month storage fee = $0.24 GB/year

0 * $0.00 GB annual fixity read = $0.00 GB/year

Total cost of NDSA Level 3+: $0.24 GB/year

Nearline:

12 * $0.01 GB/month storage fee = $0.12 GB/year

1 * $0.01 GB annual fixity read = $0.01 GB/year

Total cost of NDSA Level 3+: $0.13 GB/year

Coldline:

12 * $0.004 GB/month storage fee = $0.048 GB/year

1 * $0.05 GB annual fixity read = $0.05 GB/year

Total cost of NDSA Level 3+: $0.098 GB/year

Google Cloud

Thank you

Google Cloud