Fixity in the Cloud

Buzz Hays
Global Lead
Media & Entertainment Solutions

Google Cloud
How to automate fixity checks on GCP

$ openssl dgst -md5 -binary ArchiveFile.mov openssl enc -base64
How to automate fixity checks on GCP

$ openssl dgst -md5 -binary ArchiveFile.mov openssl enc -base64
$ apnFdauH+MfR7R5S5+NJzg==
How to automate fixity checks on GCP
How to automate fixity checks on GCP
How to automate fixity checks on GCP

```bash
$ gsutil stat gs://my-archive-bucket/ArchiveFile.mov
```
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==
  ...
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.
Where to store hashes?

Cloud SQL  BigQuery
# NDSA Levels of Digital Preservation

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

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

---

**Storage Location**
- LEVEL 1: 2 copies
- LEVEL 2: 3 copies (min)
- LEVEL 3: DR on 1 copy (min)
- LEVEL 4: DR all copies

**Fixity**
- LEVEL 1: On ingest
- LEVEL 2: On ingest, write-block, virus
- LEVEL 3: Periodic checks, logs, detect corrupts, virus
- LEVEL 4: Check on events, repair/replace, write access +1

**Security**
- LEVEL 1: Authorized Access
- LEVEL 2: Access restrictions
- LEVEL 3: Logs of access
- LEVEL 4: Audit Logs

**Metadata**
- LEVEL 1: Inventory & Location
- LEVEL 2: Store admin & transform, logs
- LEVEL 3: Standard, technical, descriptive metadata
- LEVEL 4: Preservation Metadata

**File Formats**
- LEVEL 1: Limit formats
- LEVEL 2: Inventory of formats
- LEVEL 3: Monitor obsolescence
- LEVEL 4: Format migrations
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
Thank you