Lustre* HSM in the Cloud

Robert Read, Intel HPDD

Overview

Lustre in the Cloud

HSM for Cloud

- Importing from Amazon Simple Storage Service* (S3)
- General archive with S3
- Crazy snapshot idea

Lustre in the Cloud

Lustre is not intended to be used for long-term storage in AWS

- (though HA support makes this possible)

S3 is more cost effective than Lustre (in AWS) for long term storage

Data needs to be managed outside the filesystem

- Data outlives filesystems
- Datasets can be shared

HSM for Cloud Data

Could HSM help with cloud data management?

- Supports “on demand” use case
- Existing tools don’t support S3
- Other challenges?

Much of the rest of this is speculative and “what iffy”

Not a Roadmap

Using FIDs in the Archive - Not a Good Idea

Works fine for initial import in an empty filesystem (no FID clashes)

But importing into a live file system requires remapping FIDs

- No rename in S3
- Don’t want to force a database

Don’t use FIDs when archive will outlast the filesystem

How about a GUID?

GUID can be stored in file's extended attributes

- Database lookup not required
- Could support multiple archives with different identifiers

Disadvantage: Mapping is lost when file is deleted

- Include xattr in changelog entry?
- Retain unlinked, archived inodes?

Extended Attributes can be Anything…

How about a URL?

- hsm-url=s3://my-bucket/my/big/file.log

Automated Import from S3

Import data from an S3 bucket on demand.

- Provide POSIX access to large dataset stored on S3
- Allow processing to begin before data has been imported
- Retrieve data from S3 as it is needed
- Remains in Lustre* until released

New HSM Tools with S3 support

S3 Import Tool

- Scan keys in S3 and create “stub” files in Lustre
- Save the object URL as extended attribute
- Create other metadata based on defaults (uid, gid)

S3 Copy Tool

- Copy file data from S3 links

Sample Import to Lustre from S3 Bucket

```
[root@test1]# lhsm mirror --url s3://my-bucket/tools  --dest /mnt/lustre/tools --uid 500 --gid 500

[root@test1]# tree /mnt/lustre
/mnt/lustre
└── tools
    └── emacs-21.4
        ├── aclocal.m4
        ├── AUTHORS
        ├── BUGS
        ├── ChangeLog
        ├── config.bat
        ├── config.guess
        ├── config.sub
. . .

[root@test1]# lhsm status -r  /mnt/lustre/tools
100 released /mnt/lustre/tools/emacs-21.4/vpath.sed
100 released /mnt/lustre/tools/emacs-21.4/aclocal.m4
100 released /mnt/lustre/tools/emacs-21.4/install-sh
100 released /mnt/lustre/tools/emacs-21.4/configure
100 released /mnt/lustre/tools/emacs-21.4/config.sub
```

Sample Import to Lustre (part 2)

```
[root@test1]# lhsm status -r -l  /mnt/lustre/tools
100 released (s3://my-bucket/tools/emacs-21.4/vpath.sed) /mnt/lustre/tools/emacs-21.4/vpath.sed
100 released (s3://my-bucket/tools/emacs-21.4/aclocal.m4) /mnt/lustre/tools/emacs-21.4/aclocal.m4
100 released (s3://my-bucket/tools/emacs-21.4/install-sh) /mnt/lustre/tools/emacs-21.4/install-sh
100 released (s3://my-bucket/tools/emacs-21.4/configure) /mnt/lustre/tools/emacs-21.4/configure
100 released (s3://my-bucket/tools/emacs-21.4/config.sub) /mnt/lustre/tools/emacs-21.4/config.sub
. . .

[root@test1]# lhsm restore /mnt/lustre/tools/emacs-21.4/vpath.sed
restore: /mnt/lustre/tools/emacs-21.4/vpath.sed

[root@test1t]# lhsm status /mnt/lustre/tools/emacs-21.4/vpath.sed
100 archived /mnt/lustre/tools/emacs-21.4/vpath.sed


[root@test1]# file /mnt/lustre/tools/emacs-21.4/aclocal.m4
/mnt/lustre/tools/emacs-21.4/aclocal.m4: ASCII English text

[root@test1]# lhsm status /mnt/lustre/tools/emacs-21.4/vpath.sed
100 archived /mnt/lustre/tools/emacs-21.4/vpath.sed
```

General Purpose S3 Archive

Archive

- Generate GUID and save in xattrs
- Store data in s3://archive-bucket/objects/GUID
- Supports hardlinks and transparent to renames

Restore

- Retrieve GUID from xattrs
- Fetch data by GUID

Namespace Archive

Store Metadata in S3

- JSON format
- Include special files
- Store entire directory in one key
  - Single fetch for complete directory
  - Updates are expensive
- One key per directory entry
  - Fetch per key
  - Updates easy


```
{
  "name": "work/important.txt",
  "xattr": {
    "user.hsm_guid": "e819d8fe-e969-40a5-af8a-0956da5c2e8c"
  },
  "stat": {
    "ino": 180144002274692600,
    "mode": 33188,
    "gid": 0,
    "uid": 0,
    "nlink": 1,
    "mtime": 1427217574,
    "atime": 1427217756,
    "size": 38,
    "ctime": 1427217756,
    "rdev": 0,
    "dev": 743766374
  },
  "type": "reg"
}
```

HSM Snaphots

Create stub file in .hsmsnap/ referencing the GUID in archive

Generate a new GUID every time file is archived

Multiple "snapshots" available directly to the user

Could be created by copytool or based on policy

HSM Snapshot Demo

Create a “snapshot” of a file when archived

```
[root@test1]# lhsm archive -a 2 important.txt

[root@test1]# lhsm status -l important.txt
2 archived (66b2d482-7d1d-462a-b76e-bdd2c50891f8) important.txt


[root@test1]# tree .hsmsnap/
.hsmsnap/
└── important.txt^2015-03-24T10:19:34-07:00

[root@test1]# lhsm status -l .hsmsnap/
2 released (66b2d482-7d1d-462a-b76e-bdd2c50891f8) .hsmsnap/important.txt^2015-03-24T15:12:41-07:00
```

HSM Snapshot Demo (part 2)

Create a new snapshot when file is archived again

```
[root@test1 work]# lhsm status -l important.txt
2 dirty (66b2d482-7d1d-462a-b76e-bdd2c50891f8) important.txt


[root@test1 work]# lhsm archive -a 2 important.txt
archive: important.txt


[root@test1 work]# lhsm status -l important.txt
2 archived (a326deae-2024-4ce1-9999-817a78fd51be) important.txt


[root@test1 work]# lhsm status -l .hsmsnap/
2 released (66b2d482-7d1d-462a-b76e-bdd2c50891f8) .hsmsnap/important.txt^2015-03-24T15:12:41-07:00
2 released (a326deae-2024-4ce1-9999-817a78fd51be) .hsmsnap/important.txt^2015-03-24T15:14:18-07:00
```

Summary

- Lustre is being used on the cloud
- Need to manage datasets that outlive the filesystem
- Using extended attributes adds flexibility without a database
  - Importing existing S3 data straightforward
  - General archive to S3 needs metadata storage decision
  - Creating approximate snapshots on HSM (CASH)

Legal Information

- No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.
- Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.
- This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.
- The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.
- Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.
- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at http://www.intel.com/content/www/us/en/software/intel-solutions-for-lustre-software.html.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase.
- Test and System Configurations: Demonstrations used single virtual machine with CentOS 6.6 and recent build of Lustre master branch.
- For more complete information about performance and benchmark results, visithttp://www.intel.com/performance.
- Intel and the Intel logo, are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others


intel
Software