Archiving and Retrieving Data with the Lustre TSM Copytool and LTSM

Thomas Stibor
t.stibor@gsi.de

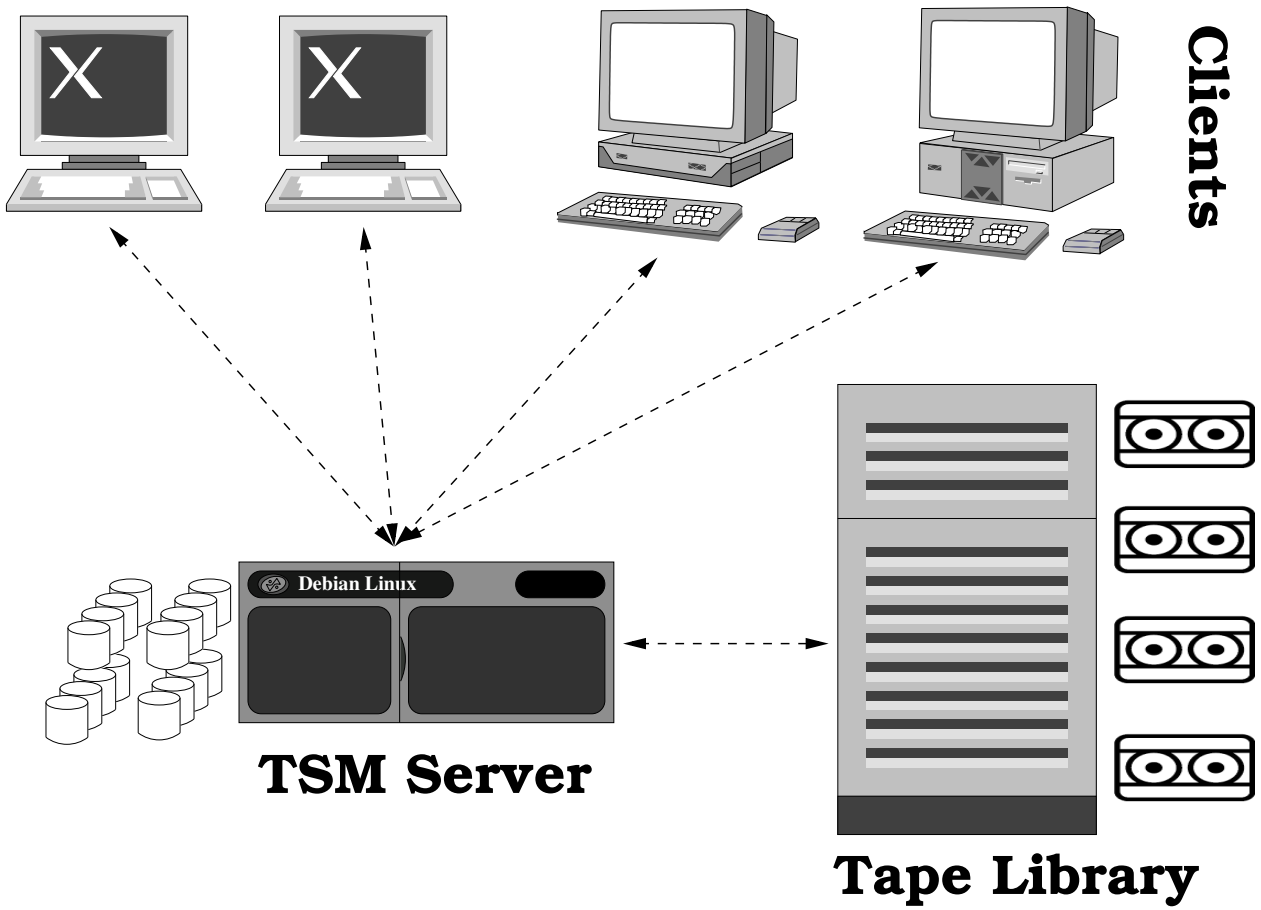
High Performance Computing
GSI Helmholtz Centre for Heavy Ion Research
Darmstadt, Germany

Tuesday 5th December, 2017

Funded by Intel® through GSI's Intel Parallel Computing Center

TSM Overview

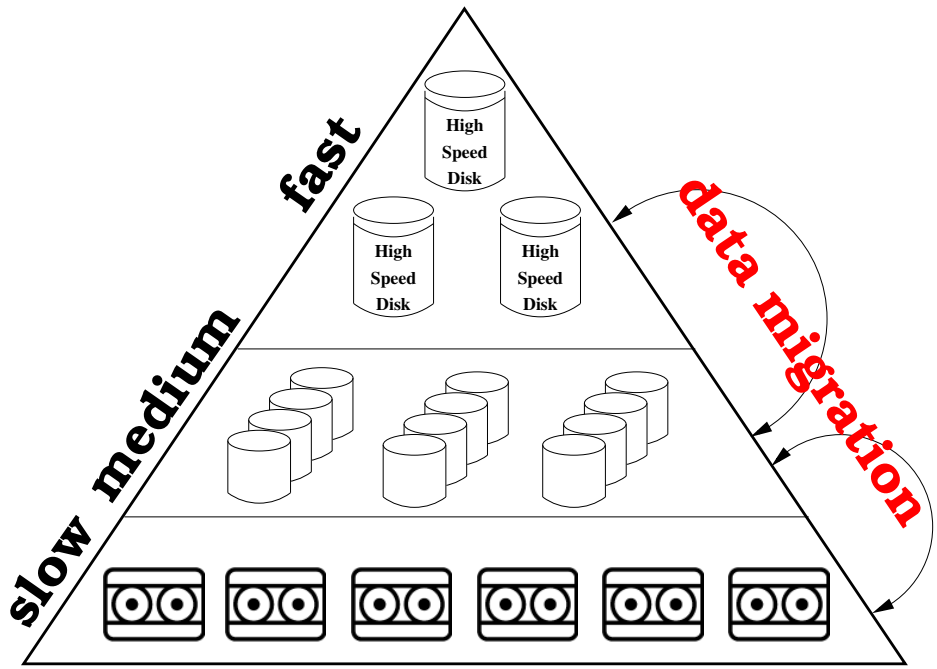
Tivoli Storage Manager[1] (TSM) is a client/server software from IBM employed in heterogeneous distributed environments to *backup* and *archive* data.



[1] Now renamed to IBM Spectrum Protect.

Some TSM Features

Compression:	Compress data stream seamless either on client or server side.
Deduplication:	Eliminating duplicate copies of repeating data.
Collocation:	Store and pack data of a client in few number of tapes as much as possible to reduce the number of media mounts and for minimizing tape drive movements.
Storage hierarchies:	Automatically move data from faster devices to slower devices based on characteristics such as file size or storage capacity.



Meta data is stored in a DB2 database (part of TSM server).

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server

The TSM server is an object storage server and is developed for storing and retrieving *named* objects. The object name is used for accessing objects and is composed of:

- `fs`: File space name (mount point),
- `hl`: High level name (directory name),
- `ll`: Low level name (file name),

```
fs: /lustre, hl: /doc, ll: /pub.tex
object id (hi,lo)                              : (0,30375)
object info length                             : 48
object info size (hi,lo)                       : (0,32768) (32768 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 71147
crc32                                          : 0x41ac6a53 (1101818451)
archive description                            : GECCO 2004 publication
owner                                          : tstibor
insert date                                    : 2017/09/19 10:55:07
expiration date                                : 2018/09/19 10:55:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (44,0,161,0,0)
estimated size (hi,lo)                         : (0,32768) (32768 bytes)
lustre fid                                     : [0x200000401:0x7c:0x0]
lustre stripe size                             : 65536
lustre stripe count                            : 1
```

Data Organized on TSM Server (cont.)

Note: All archived data is stored in that object named `/fs/hl/ll` format independent of gStore or whatever. That is, data can *always* be retrieved[2] with e.g. IBM's tool `dsmc`.

```
tsm> q filespace
  #     Last Incr Date          Type    File Space Name
--------------------------------------------------------------------------------
  1     00/00/0   00:00:00     API:/           /
  2     00/00/0   00:00:00     API:ltsm        /
  3     00/00/0   00:00:00     API:API-Server  /adamczew
  4     00/00/0   00:00:00     API-Server      /adsmcli
  5     00/00/0   00:00:00     API:API-Server  /agatadaq
  6     00/00/0   00:00:00     API-Server      /aladin
  7     00/00/0   00:00:00     API-Server      /alice
  8     00/00/0   00:00:00     API:API-Server  /alice2009backup
  9     00/00/0   00:00:00     API:API-Server  /aliceraw
 10     00/00/0   00:00:00     API:API-Server  /alicetest
 11     00/00/0   00:00:00     API:API-Server  /aliprod
 12     00/00/0   00:00:00     API-Server      /andreas
 13     00/00/0   00:00:00     API:API-Server  /andronic
 ...    ...       ...           ...             ...
192     00/00/0   00:00:00     API:API-Server  /xray_esr
193     00/00/0   00:00:00     API:API-Server  /xsmcnpphits
194     00/00/0   00:00:00     API:API-Server  /yushman
195     00/00/0   00:00:00     API-Server      /z311
196     00/00/0   00:00:00     API:API-Server  /zumbruch
```

[2] although size fields are messed up.

Data Organized on TSM Server (cont.)

```
> dsmc query archive -subdir=yes -se=aixtsm3 /hadesmay14raw/prod01/*
Session established with server AIXTSM3: AIX
  Server Version 6, Release 3, Level 4.0
  Server date/time: 12/04/2017 16:54:10  Last access: 12/04/2017 16:53:59

             Size  Archive Date - Time    File - Expires on - Description
             ----  -------------------    ------------------------------
API    305,070 KB  05/05/2014 14:32:15    /hadesmay14raw/prod01/be1412019584607.hld ...
API  1,559,206 KB  05/05/2014 14:32:13    /hadesmay14raw/prod01/pt1411416565807.hld ...
API         32  B  05/05/2014 15:28:22    /hadesmay14raw/prod01/be1412018522114.hld ...
API         32  B  05/05/2014 15:33:48    /hadesmay14raw/prod01/be1412020041611.hld ...
API  1,619,011 KB  05/05/2014 14:41:38    /hadesmay14raw/prod01/pt1411418013204.hld ...
...
> dsmc retrieve -se=aixtsm3 \
/hadesmay14raw/prod01/be1412019584607.hld /tmp/hadesmay14raw/prod01/be1412019584607.hld
Retrieving /hadesmay14raw/prod01/be1412019584607.hld -->
           /tmp/hadesmay14raw/prod01/be1412019584607.hld [Done]
```

How about some older archived data?

```
> dsmc query archive -subdir=yes -se=aixtsm3 /kaos/lmdv/may5.raw0443
             Size  Archive Date - Time    File - Expires on - Description
             ----  -------------------    ------------------------------
API       UNKNOWN  08/25/1996 21:18:11    /kaos/lmdv/may5.raw0443 Never *

> dsmc retrieve -se=aixtsm3 /kaos/lmdv/may5.raw0443 /tmp/may5.raw0443
Retrieving /kaos/lmdv/may5.raw0443 --> /tmp/may5.raw0443 [Done]
```

This works on Windows, Linux, Mac, Solaris (Sparc/X86), HP-UX and AIX machines.

Data Organized on TSM Server (cont.)

```
>./src/ltsmc -v debug --query -f /kaos -n XXXXXX -p YYYYYY -s aixtsm3 "/kaos/lmdv/may5.raw0443"
fs: /kaos, hl: /lmdv, ll: /may5.raw0443
object id (hi,lo)                              : (0,134474)
object info length                             : 25
object info size (hi,lo)                       : (809120824,875634746) (3475147478468206650 bytes)
object type                                    : DSM_OBJ_FILE
object magic id                                : 859189536
crc32                                          : 0x3a343037 (0976498743)
archive description                            : *
owner                                          : kaosuser
insert date                                    : 1996/08/25 21:18:11
expiration date                                : 65535/00/00 00:00:00
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo):  (24735,0,11599,0,0)
estimated size (hi,lo)                         : (4294967295,4294967295) (18446744073709551615 bytes)
lustre fid                                     : [0x32:0x0:0x0]
lustre stripe size                             : 0
lustre stripe count                            : 0
```

Even when you mess up the fields, data can be retrieved, so no worry, your data is not doomed !!

Backup vs Archive

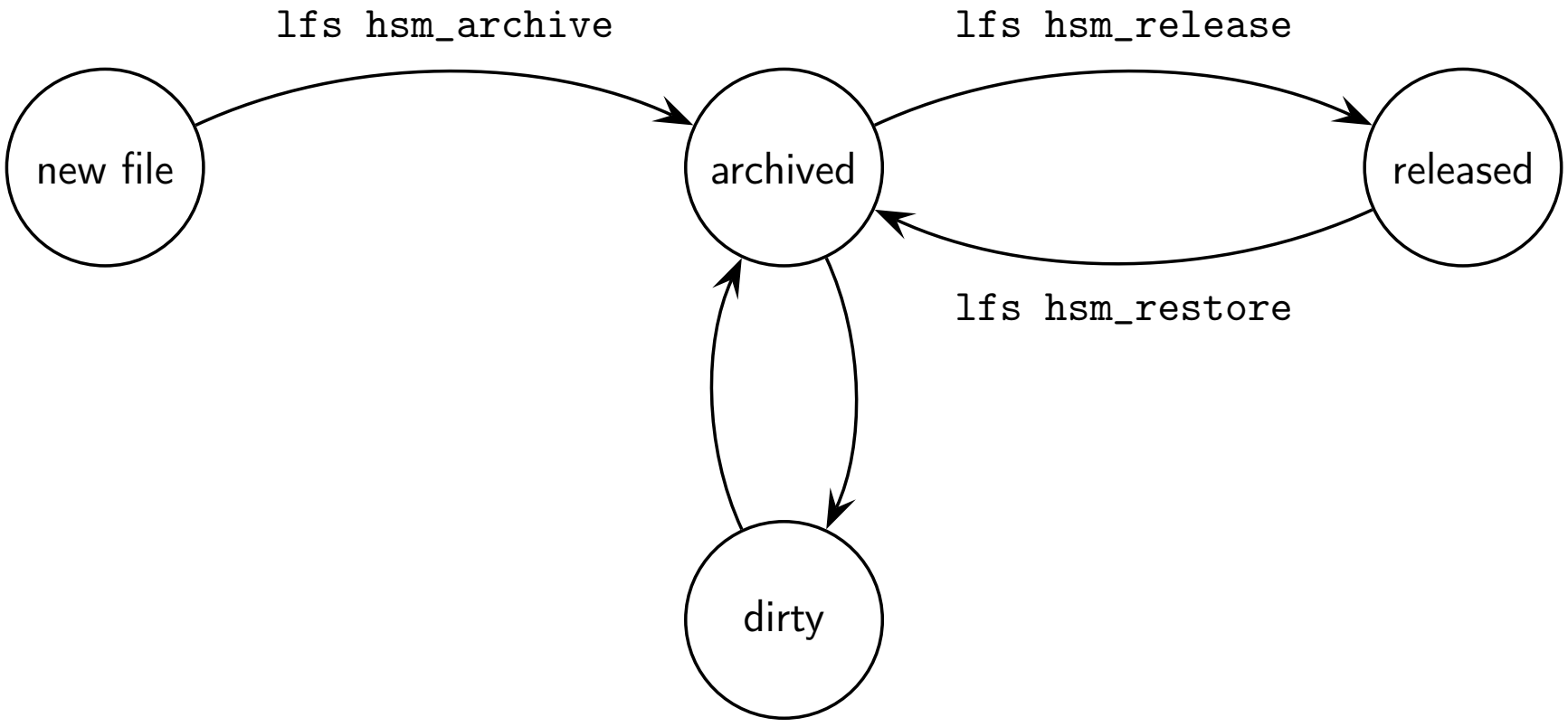
Backup: A copy of the data is stored in the event the original becomes lost or damaged. Typically an incremental (forever) backup strategy is performed.

Archive: Remove from an on-line system those data no longer in day to day use, and place them into a long term retrievable storage (such as tape drives).

Lustre has since version 2.5 hierarchical storage management (HSM) capabilities, that is, data can be automatically *archived* to low-cost storage media such as tape storage systems and seamlessly *retrieved* when accessing the data on Lustre clients.

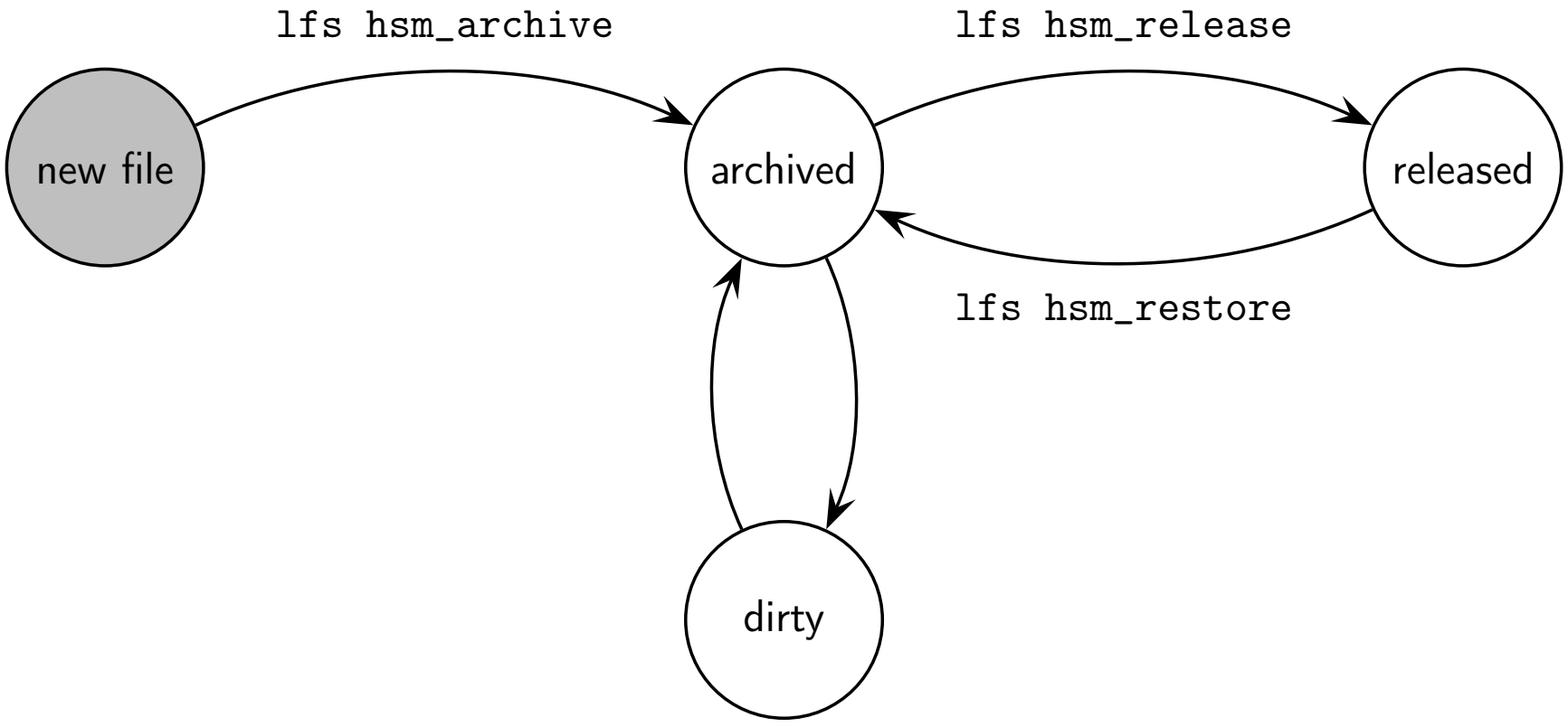
Lustre HSM $\equiv$ seamlessly *archiving*, *retrieving* and *deleting* data.

Overview of Lustre HSM State Diagram



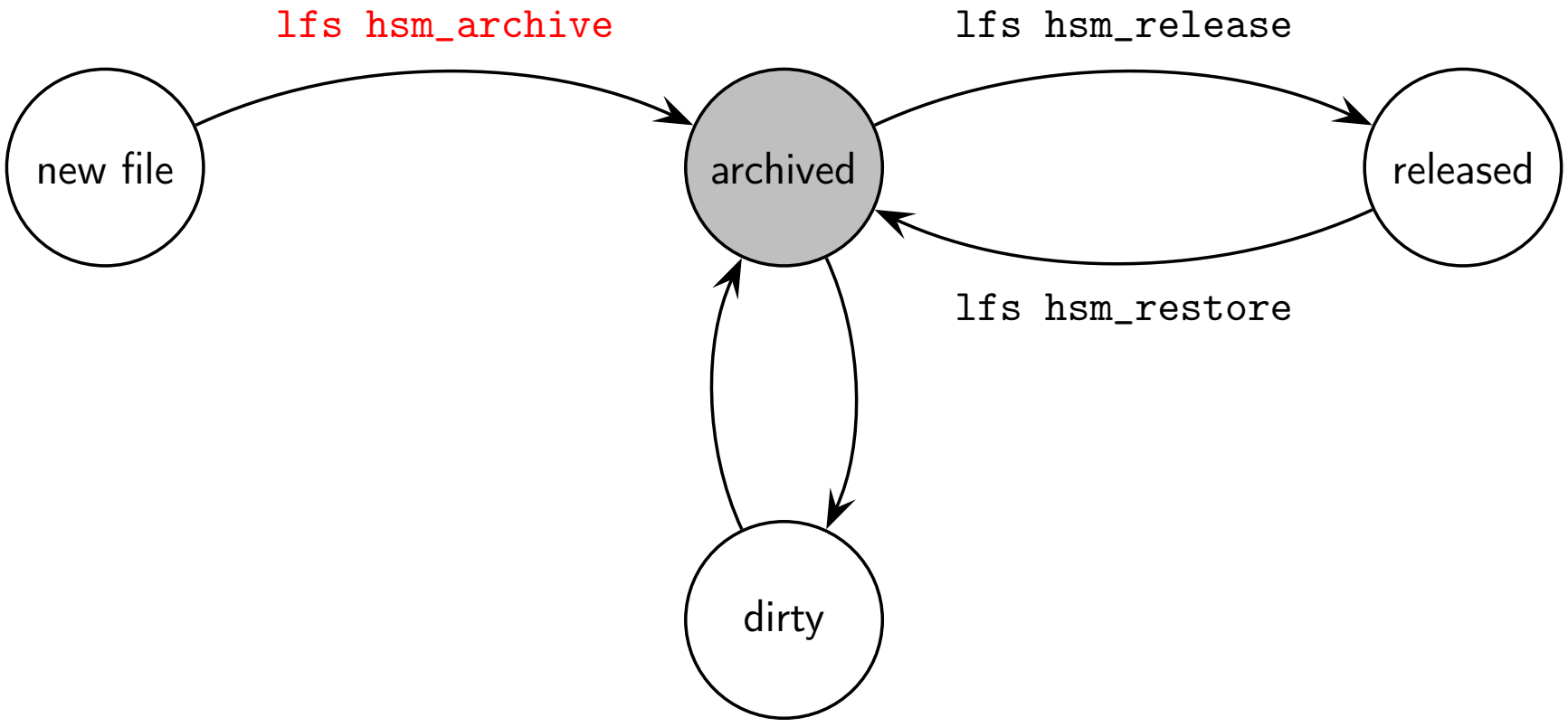

```
>dd if=/dev/zero of=zeros bs=1MiB count=32 conv=sync
32+0 records in
32+0 records out
33554432 bytes (34 MB) copied, 0.401738 s, 83.5 MB/s
```

Overview of Lustre HSM State Diagram



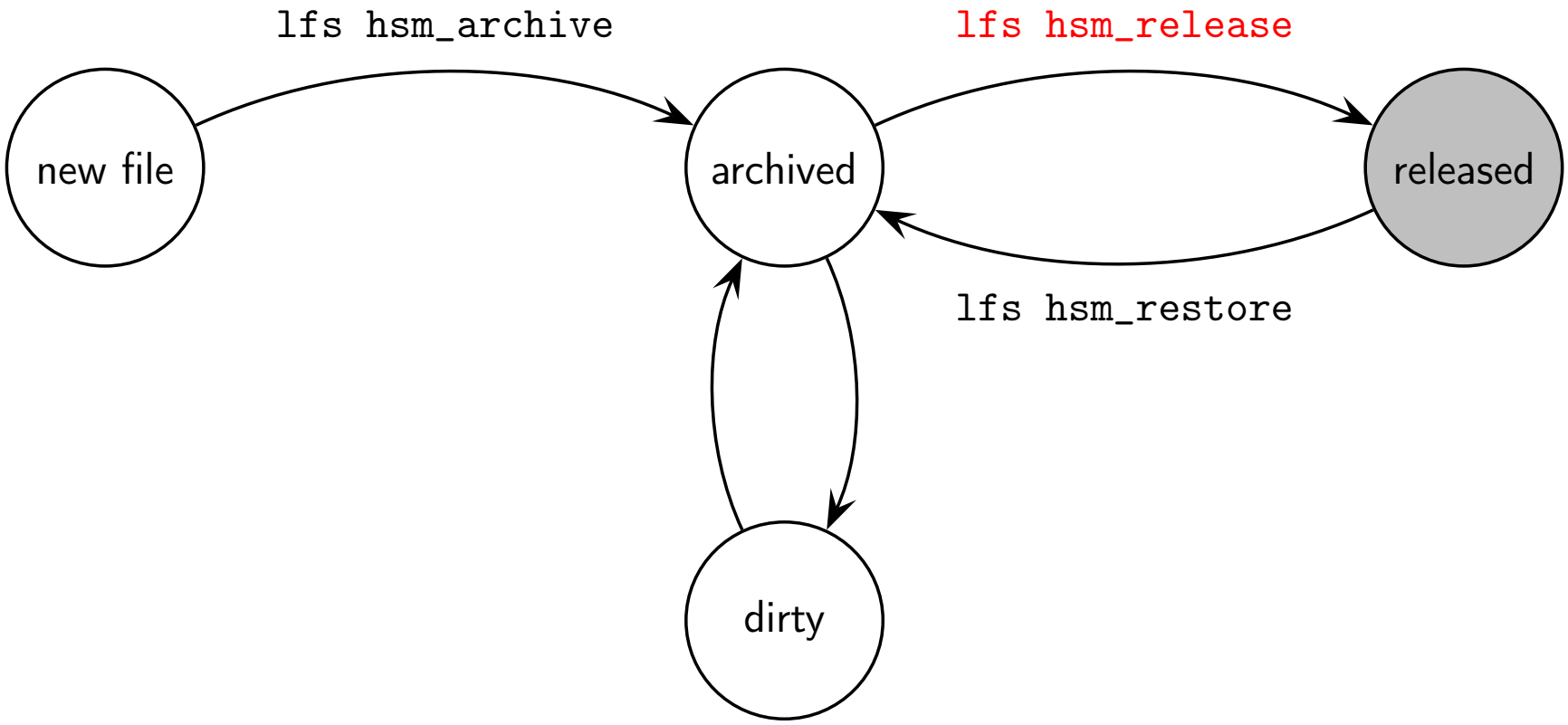

```
>lfs hsm_state ./zeros && ll -h zeros && du -h ./zeros
./zeros: (0x00000000)
-rw-r--r-- 1 root root 32M Sep  6 13:55 zeros
32M ./zeros
```

Overview of Lustre HSM State Diagram



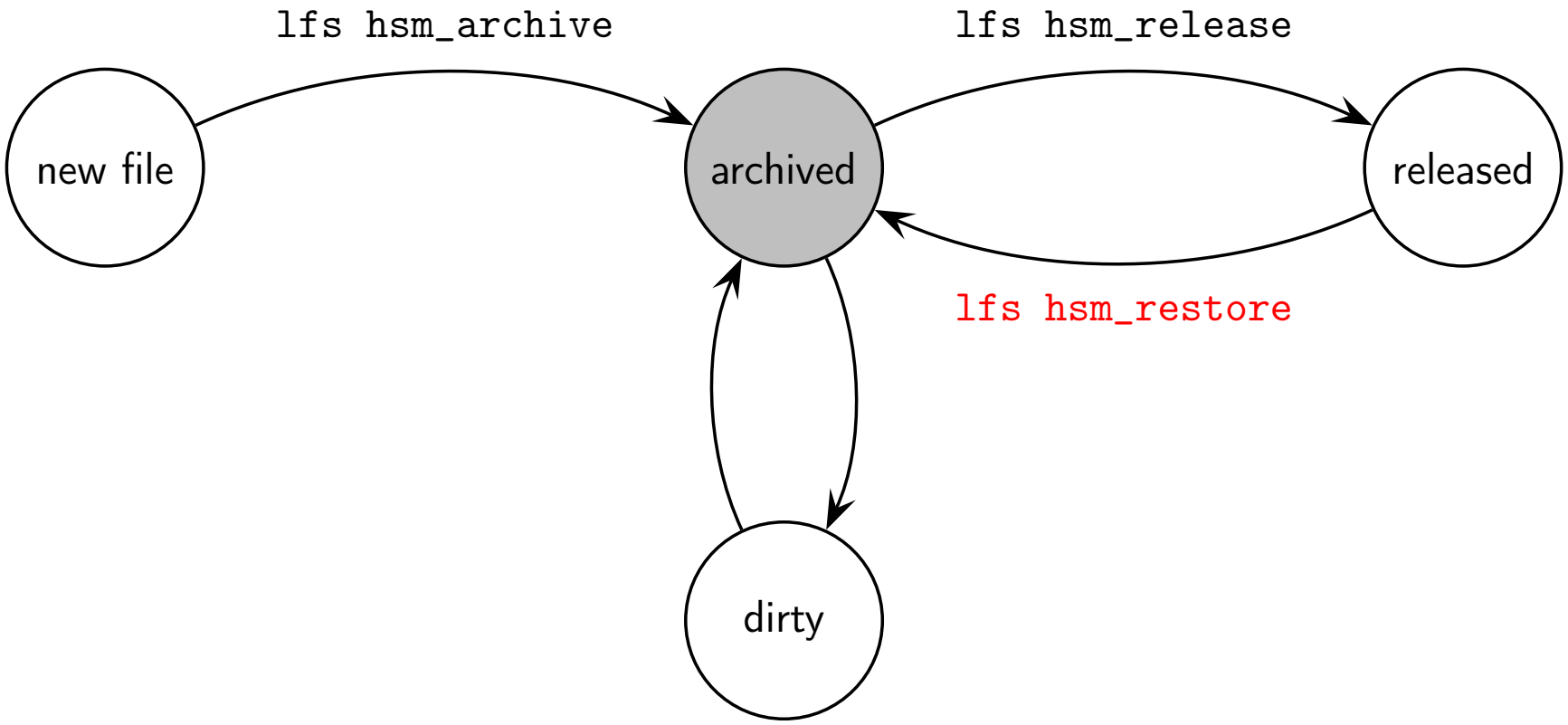

```
>lfs hsm_state ./zeros && ll -h zeros && du -h ./zeros
./zeros: (0x00000009) exists archived, archive_id:1
-rw-r--r-- 1 root root 32M Sep  6 13:55 zeros
32M ./zeros
```

Overview of Lustre HSM State Diagram



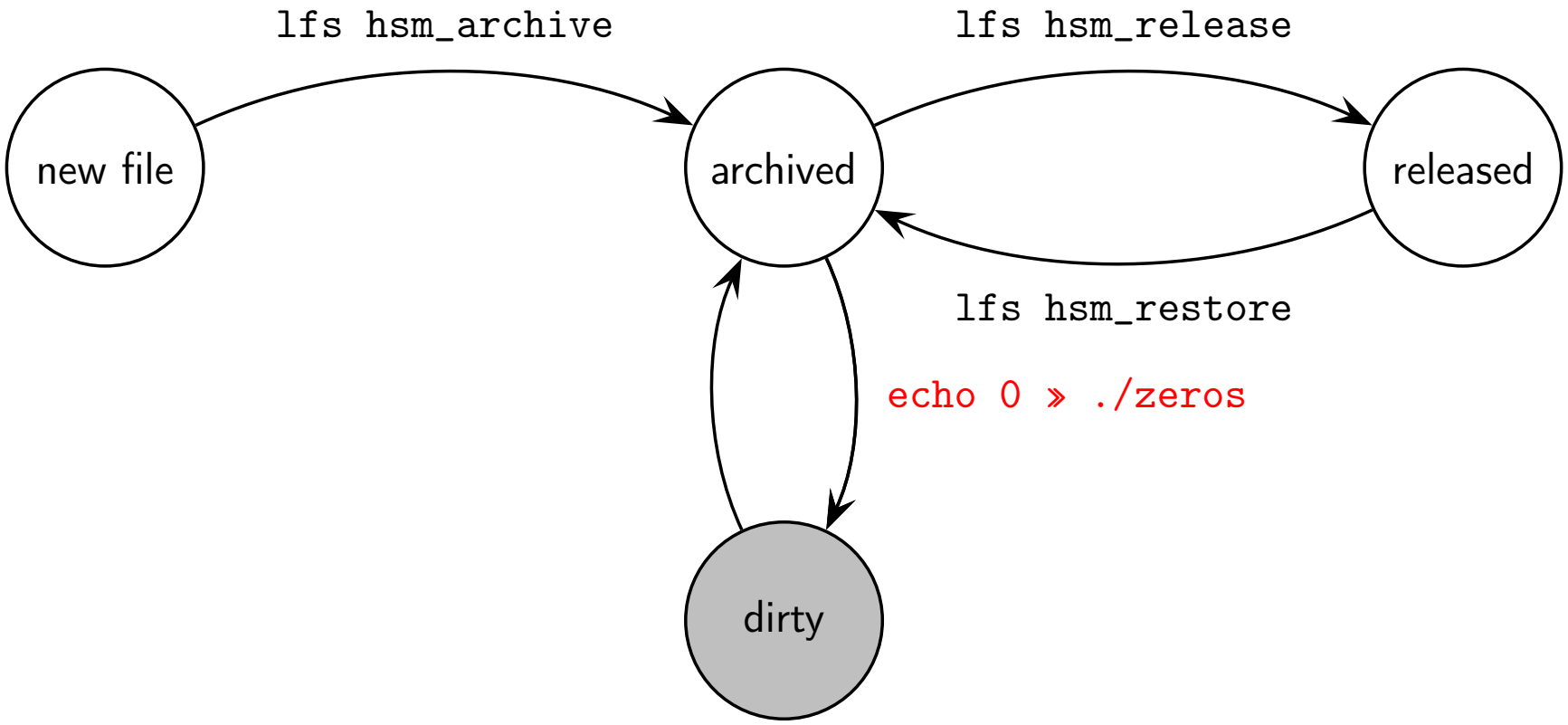

```
>lfs hsm_state ./zeros && ll -h zeros && du -h ./zeros
./zeros: (0x0000000d) released exists archived, archive_id:1
-rw-r--r-- 1 root root 32M Sep  6 13:55 zeros
512 ./zeros
```

Overview of Lustre HSM State Diagram



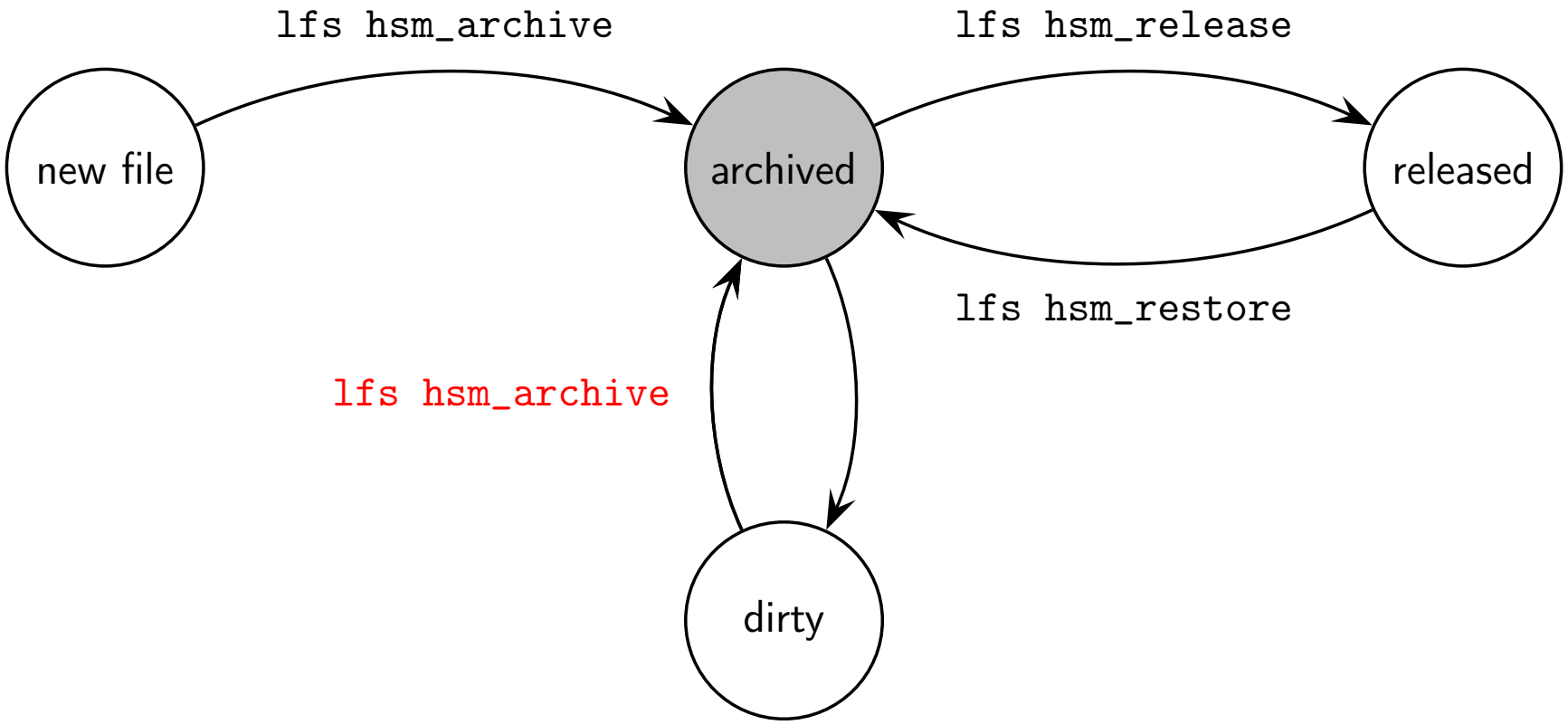

```
>lfs hsm_state ./zeros && ll -h zeros && du -h ./zeros
./zeros: (0x00000009) exists archived, archive_id:1
-rw-r--r-- 1 root root 32M Sep  6 13:55 zeros
32M ./zeros
```

Overview of Lustre HSM State Diagram



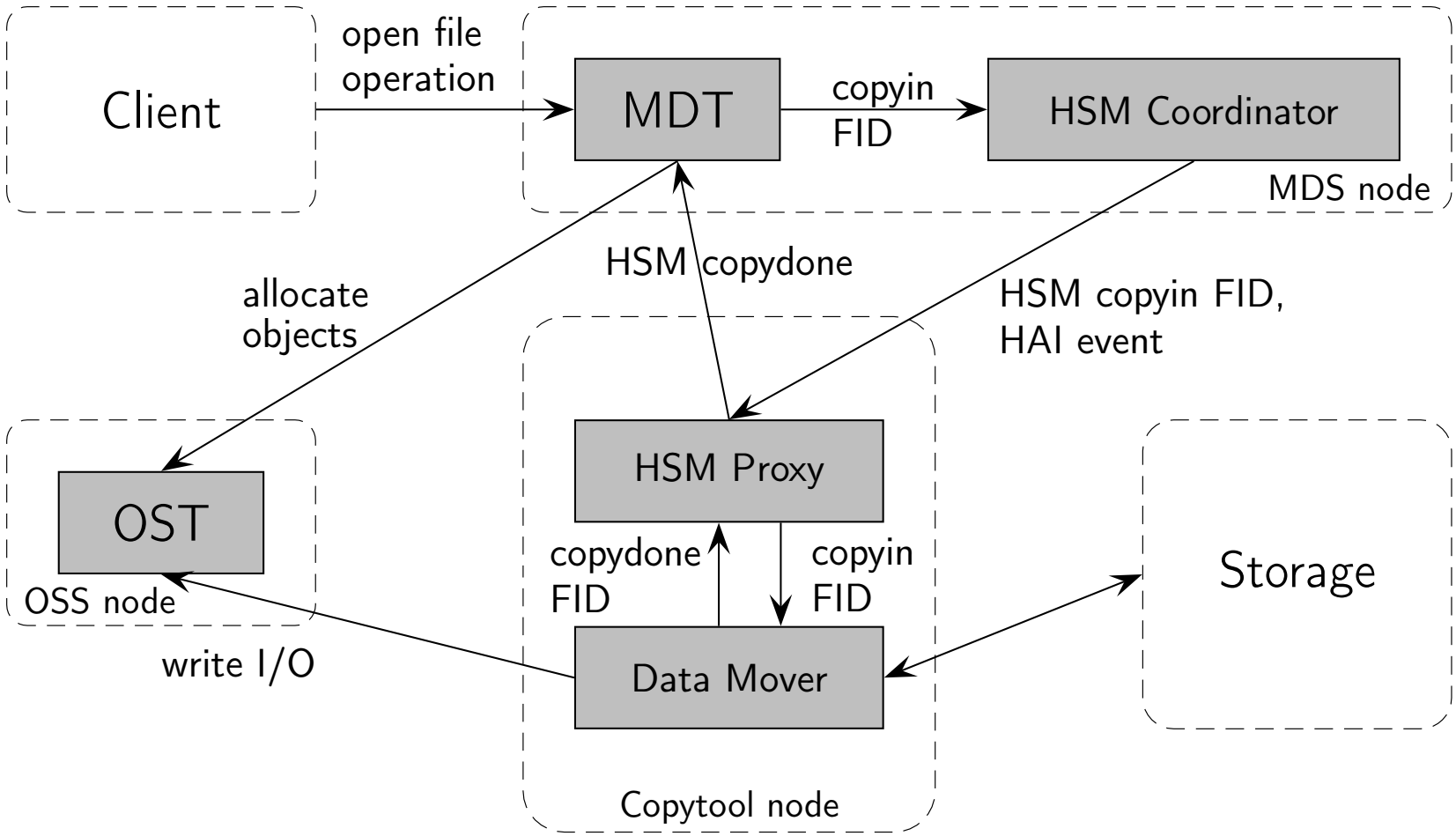

```
>echo 0 >> ./zero && lfs hsm_state && ll -h zeros && du -h ./zeros
./zeros: (0x0000000b) exists dirty archived, archive_id:1
-rw-r--r-- 1 root root 33M Sep 19 09:16 ./zeros
32M ./zeros
```

Overview of Lustre HSM State Diagram




```
>lfs hsm_archive ./zero &&  lfs hsm_state && ll -h zeros && du -h ./zeros
./zeros: (0x00000009) exists archived, archive_id:1
-rw-r--r-- 1 root root 33M Sep 19 09:16 ./zeros
33M ./zeros
```

Lustre HSM Framework



Data flow and triggered actions for *retrieving* data.

Lustre Copytool

Moves data between a Lustre mount point and HSM backend. The copytool receives *archive*, *retrieve* and *delete* actions from MDT node and triggers data moving operations

```
switch (session->hai->hai_action) {
      case HSMA_ARCHIVE:
              rc = ct_archive(session);
              break;
      case HSMA_RESTORE:
              rc = ct_restore(session);
              break;
      case HSMA_REMOVE:
              rc = ct_remove(session);
              break;
      case HSMA_CANCEL:
              ...
              ...
}
```

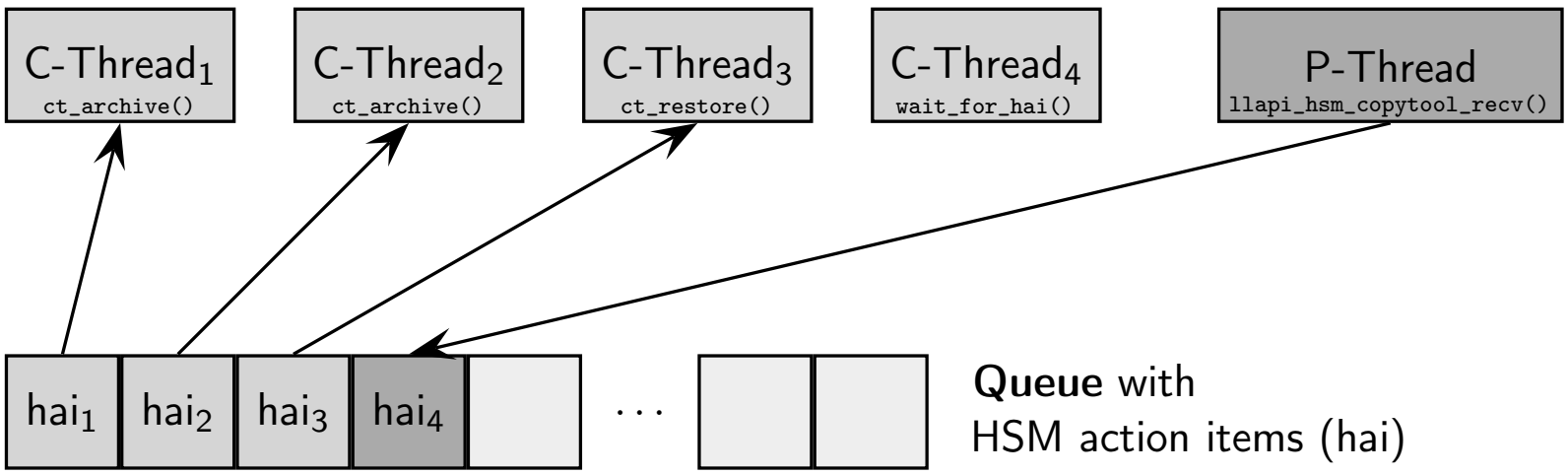
```
static int ct_archive(struct session_t *session)
{
    rc = fid_realpath(opt.o_mnt, &session->hai->hai_fid,
                      fpath, sizeof(fpath));
    if (rc < 0) {
        CT_ERROR(rc, "fid_realpath failed");
        goto cleanup;
    }

    rc = ct_hsm_action_begin(session, mdt_index, open_flags,
                             false);
    if (rc < 0) {
        CT_ERROR(rc, "ct_hsm_action_begin on '%s' failed",
                 fpath);
        goto cleanup;
    }
    ...
    ...
    rc = tsm_archive_fpath(opt.o_fsname, fpath, NULL,
                           fd, &lustre_info, session);
    ...
}
```

Implementation Details and Scaling

For achieving high data throughput by means of parallelism the copytool employs the *producer-consumer* model.

- Data structure is a concurrent *queue*.
- Producer thread receives HSM action items from the MDS's and enqueues them.
- Multiple consumer threads, each having a session opened to the TSM server, which are dequeueing items and executing the HSM actions.



Lustre TSM Copytool

```
>./src/lhsmtool_tsm --help
usage: ./src/lhsmtool_tsm [options] <lustre_mount_point>
        -a, --archive-id <int> [default: 0]
                archive id number
        -t, --threads <int>
                number of processing threads [default: 2]
        -n, --node <string>
                node name registered on tsm server
        -p, --password <string>
                password of tsm node/owner
        -o, --owner <string>
                owner of tsm node
        -s, --servername <string>
                hostname of tsm server
        -v, --verbose {error, warn, message, info, debug} [default: message]
                produce more verbose output
        --abort-on-error
                abort operation on major error
        --daemon
                daemon mode run in background
        --dry-run
                don't run, just show what would be done
        --restore-stripe
                restore stripe information
        -h, --help
                show this help

IBM API library version: 7.1.6.0, IBM API application client version: 7.1.6.0
version: 0.7.0-4 © 2017 by GSI Helmholtz Centre for Heavy Ion Research
```

Lustre TSM Copytool Usage Tips

I am still archiving a very large file and want to know the *progress*?

```
>lfs hsm_action ./zeros
./zeros: NOOP
>sudo lfs hsm_archive ./zeros
>lfs hsm_action ./zeros
./zeros: ARCHIVE running (32764 bytes moved)
>lfs hsm_action ./zeros
./zeros: ARCHIVE running (65528 bytes moved)
...
...
...
>lfs hsm_action ./zeros
./zeros: NOOP
```

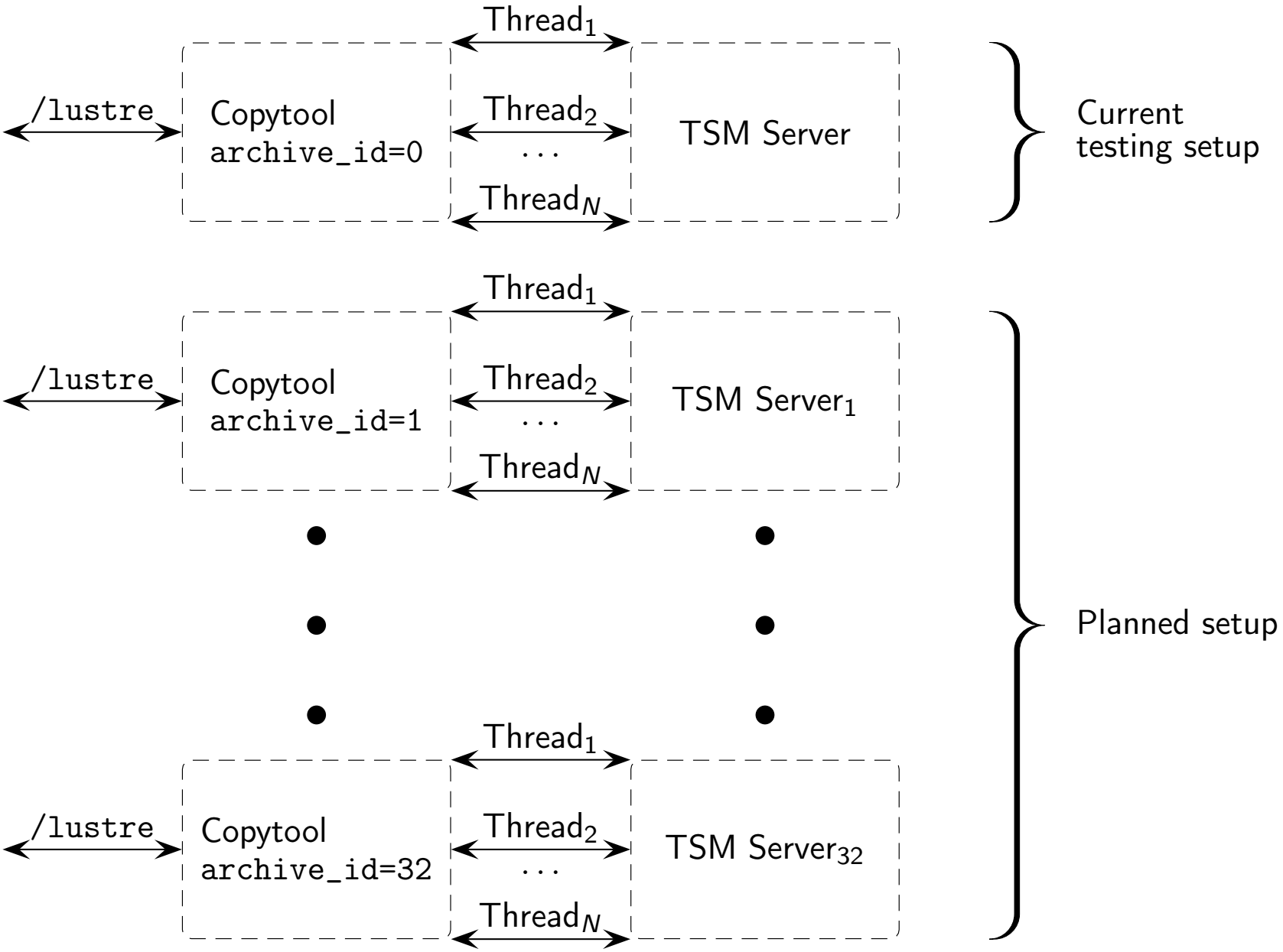
After a progress of 25% I decided to cancel the archive operation!

```
>sudo lfs hsm_cancel ./zeros
```

Before executing any *archive*,*retrieve*,*release*,... commands check state of file with

```
>lfs hsm_state <FILE>
```

Lustre TSM Copytool Scaling



Note: LL_HSM_MAX_ARCHIVE (sizeof(__u32) * 8) = 32

Lustre TSM Copytool Access Control

Access to TSM server is granted by registering nodes, that is,

```
register node <NODE> <PASSWD> domain=<DOMAIN>
```

Each copytool with `archive_id`$=a$ thus connects as `<NODE>`$=\text{node}_a$ and `<PASSWD>` to the TSM server.

In addition, we set

```
update node <NODE> archdelete=no
```

so that the node cannot delete archived files. Note, *archive* operation can never overwrite data. If a file is archived multiple times, then it exists exactly multiple times on the TSM server.

Lustre TSM Copytool Access Control (cont.)

Identical file stored twice. Incremental archive does not exist.

```
[INFO] 1512039747.265481 [16059] tsmapi.c:708 [query] object # 0
fs: /, hl: /home/tstibor/dev/tsm/github/ltsm, ll: /README.md
object id (hi,lo)                           : (0,10701)
object info length                          : 48
object info size (hi,lo)                    : (0,25437) (25437 bytes)
object type                                 : DSM_OBJ_FILE
object magic id                             : 71147
crc32                                       : 0x9770c6a5 (2540750501)
archive description                         :
owner                                       :
insert date                                 : 2017/11/30 11:16:04
expiration date                             : 2018/11/30 11:16:04
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (2,0,429,0,0)
estimated size (hi,lo)                      : (0,25437) (25437 bytes)

[INFO] 1512039747.265508 [16059] tsmapi.c:708 [query] object # 1
fs: /, hl: /home/tstibor/dev/tsm/github/ltsm, ll: /README.md
object id (hi,lo)                           : (0,10638)
object info length                          : 48
object info size (hi,lo)                    : (0,25437) (25437 bytes)
object type                                 : DSM_OBJ_FILE
object magic id                             : 71147
crc32                                       : 0x9770c6a5 (2540750501)
archive description                         :
owner                                       :
insert date                                 : 2017/11/27 15:21:14
expiration date                             : 2018/11/27 15:21:14
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (2,0,398,0,0)
estimated size (hi,lo)                      : (0,25437) (25437 bytes)
```

Lustre TSM Copytool Access Control (cont.)

Try to delete TSM object `fs:/hl:/home/tstibor/dev/tsm/github/ltsm ll:/README.md`

```
>./src/ltsmc -v debug -n polaris -p polaris -s polaris-kvm-tsm-server \
             --delete /home/tstibor/dev/tsm/github/ltsm/README.md
...
[DEBUG] 1512040096.741356 [16096] tsmapi.c:1236 [rc:0] get_qra: 1
[DEBUG] 1512040096.741745 [16096] tsmapi.c:1201 dsmBeginTxn: handle: 1 ANS0302I (RC0) Successfully done.
[DEBUG] 1512040096.741771 [16096] tsmapi.c:1211 dsmDeleteObj: handle: 1 ANS0302I (RC0) Successfully done.
[ERROR] 1512040096.743250 [16096] tsmapi.c:1222 dsmEndTxn: handle: 1 ANS0266I (RC2302)
                                                The dsmEndTxn vote is ABORT, so check the reason field
[ERROR] 1512040096.743280 [16096] tsmapi.c:1223 dsmEndTxn reason: handle: 1 ANS1126E (RC27)
                                                The file space cannot be deleted because this node
                                                does not have permission to delete archived or backed
                                                up data.
[DEBUG] 1512040096.743288 [16096] tsmapi.c:1243 [rc:2302] tsm_del_obj: 1
[WARN] 1512040096.743294 [16096] tsmapi.c:1245 tsm_del_obj failed, object not deleted
...
```

Overview of LTSM (https://github.com/tstibor/ltsm)

LTSM - Lightweight TSM API, Lustre TSM Copytool for Archiving Data and TSM Console Client

This project consists of *four* parts:

1. Lightweight TSM API/library (called *tsmapi*) supporting operations (*archiving*, *retrieving*, *deleting*, *querying*).
2. Lustre TSM Copytool.
3. TSM console client.
4. Benchmark and test suite.

```
>./src/ltsmc --help
usage: ./src/ltsmc [options] <files|directories|wildcards>
--archive
--retrieve
--query
--delete
--pipe
-l, --latest [retrieve object with latest timestamp when multiple exists]
-x, --prefix [retrieve prefix directory]
-r, --recursive [archive directory and all sub-directories]
-t, --sort={ascending, descending, restore} [sort query in date or restore order]
-f, --fsname <string> [default: '/']
-d, --description <string>
-n, --node <string>
-o, --owner <string>
-p, --password <string>
-s, --servername <string>
-v, --verbose {error, warn, message, info, debug} [default: message]
-c, --checksum <file>
-h, --help

IBM API library version: 7.1.6.2, IBM API application client version: 7.1.6.0
version: 0.7.0-5 © 2017 by GSI Helmholtz Centre for Heavy Ion Research
```

Workflow by means of tsm_f* Functions in LTSM

```
int tsm_fconnect(struct login_t *login, struct session_t *session);
void tsm_fdisconnect(struct session_t *session);
int tsm_fopen(const char *fs, const char *fpath, const char *desc, struct session_t *session);
ssize_t tsm_fwrite(const void *ptr, size_t size, size_t nmemb, struct session_t *session);
int tsm_fclose(struct session_t *session);
```

Let's use these function calls via *ltsmc*

```
>wget -O - -o /dev/null http://google.com | ./src/ltsmc --pipe --description "Google website snapshot" \
--owner "it's me" -v warn -n polaris -p polaris -s polaris-kvm-tsm-server \
-f /lustre /lustre/tstibor/google.com
>./src/ltsmc -v info -n polaris -p polaris -s polaris-kvm-tsm-server -f /lustre --query /lustre/tstibor/google.com
[INFO] 1512045728.384295 [21427] tsmapi.c:708 [query] object # 0
fs: /lustre, hl: /tstibor, ll: /google.com
object id (hi,lo)                          : (0,10707)
object info length                         : 48
object info size (hi,lo)                   : (0,11223) (11223 bytes)
object type                                : DSM_OBJ_FILE
object magic id                            : 71147
crc32                                      : 0xb0b5824a (2964685386)
archive description                        : Google website snapshot
owner                                      : it's me
insert date                                : 2017/11/30 13:42:07
expiration date                            : 2018/11/30 13:42:07
restore order (top,hi_hi,hi_lo,lo_hi,lo_lo): (2,0,434,0,0)
estimated size (hi,lo)                     : (0,11223) (11223 bytes)
lustre fid                                 : [0:0x0:0x0]
lustre stripe size                         : 0
lustre stripe count                        : 0
```

How to transfer the website snapshot which is located on the TSM server to a Lustre filesystem (i.e `/lustre/tstibor/google.com`)?

Workflow by means of tsm_f* Functions in LTSM (cont.)

```
./src/ltsmc -v info -n polaris -p polaris -s polaris-kvm-tsm-server -f /lustre \
            --retrieve /lustre/tstibor/google.com && cat /lustre/tstibor/google.com

<!doctype html><html itemscope="" itemtype="http://schema.org/WebPage" lang="de"><head>
<meta content="text/html; charset=UTF-8" http-equiv="Content-Type">
<meta content="/images/branding/googleg/1x/googleg_standard_color_128dp.png"
itemprop="image"><title>Google</title><script>(function(){window.google=
{kEI:'n_wfWoLxMdHRkwWT-pKgCg',kEXPI:'1352553 ......
```

Note: This workflow is independent of Lustre and works with any Linux kernel supported filesystem and of course with IBM's *dsmc*.

On Lustre we can touch an empty file `/lustre/tstibor/google.com` and set the Lustre HSM flags:

```
touch /lustre/tstibor/google.com && sudo lfs hsm_set --exists --archived /lustre/tstibor/google.com

>ls -la && lfs hsm_state ./google.com
total 21
drwxr-xr-x 2 tstibor tstibor 10752 Nov 30 14:12 .
drwxr-xr-x 6 root    root     9728 Nov 30 13:39 ..
-rw-r--r-- 1 tstibor tstibor     0 Nov 30 14:12 google.com
./google.com: (0x00000009) exists archived
```

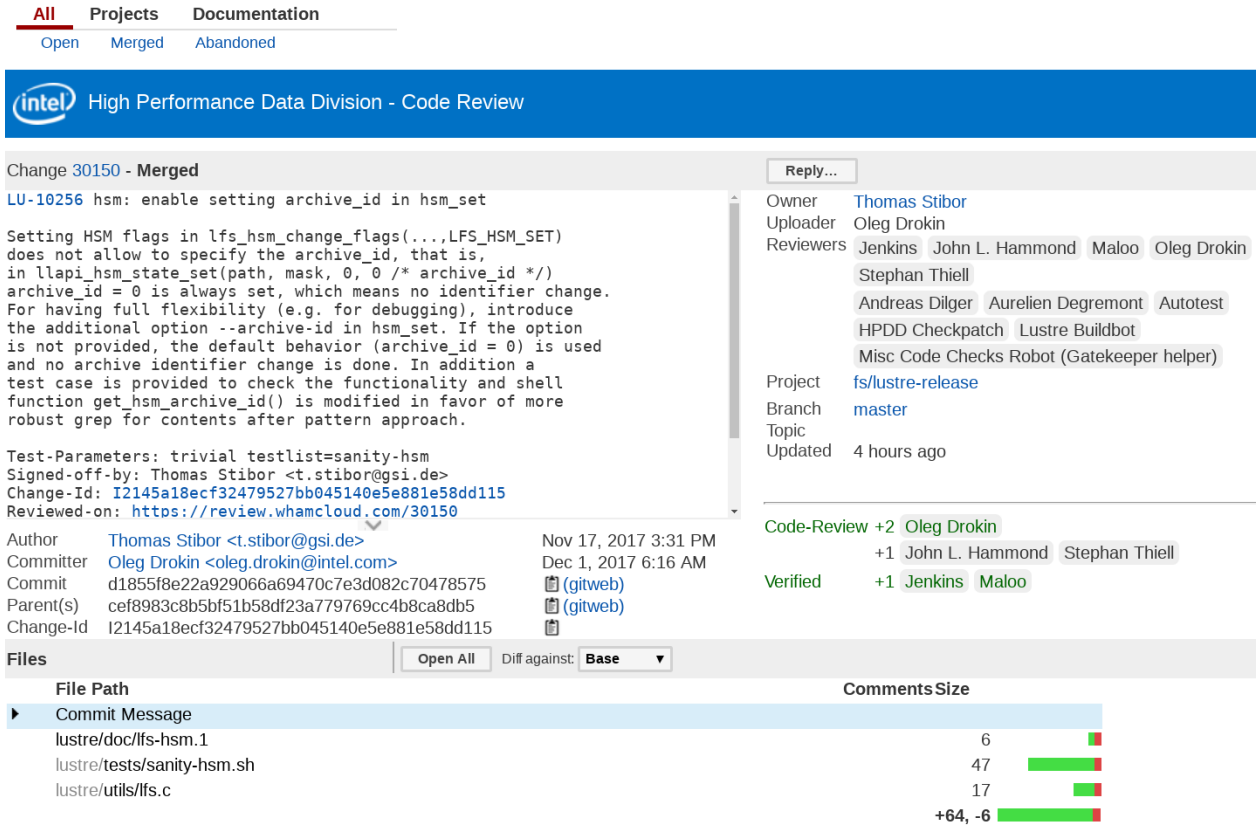
When the file is accessed. e.g. `cat /lustre/tstibor/google.com` it is seamlessly retrieved from the top of the TSM storage hierarchy.

Workflow by means of `tsm_f*` Functions in LTSM (cont.)

Note:

```
lfs hsm_set
Set HSM user flag on specified files.
usage: hsm_set [--norelease] [--noarchive] [--dirty] [--exists] [--archived] [--lost] <file> ...
```

has no `archive-id` option! So let's file a patch and send it upstream



Note, this workflow is encapsulated in the script ltsmsync.sh.

Summary & and Handover to Jörn's Talk

- Project available at https://github.com/tstibor/ltsm
- Manual pages and more documentation.
- CentOS 7.X RPM package + Debian 8.X DEB package.
- Autoconf build system, can build LTSM also without Lustre support, and with IBM'S TSM Library (6.X)/7.X/8.X support.
- For own experiments and how to setup (in KVM) a TSM server see TSM Server Installation Guide.

Questions please after Jörn's talk