



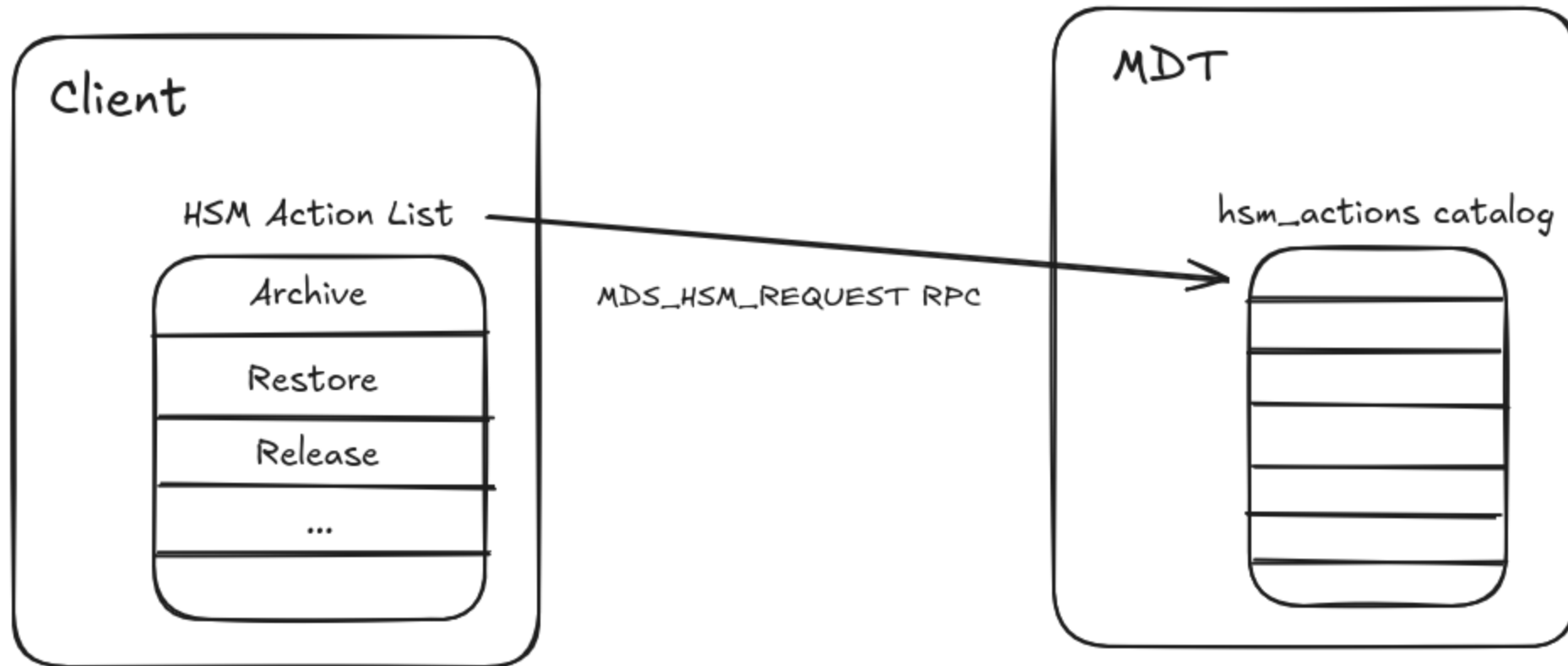
LUG 25: developer day

HSM development and FLR integration

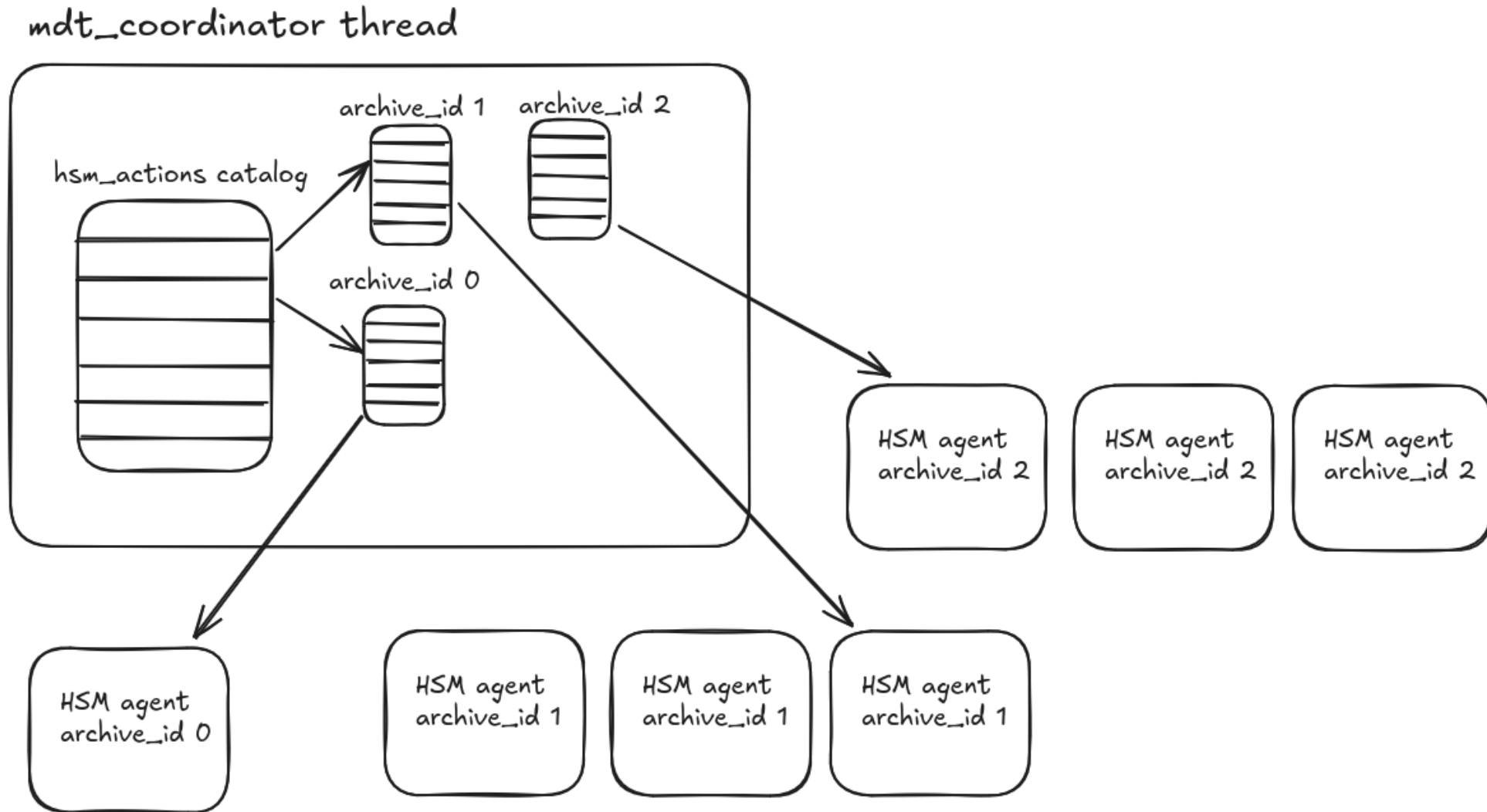
Guillaume Courrier <guillaume.courrier@cea.fr>

Etienne Aujames <eaujames@ddn.com>

Lustre HSM: architecture

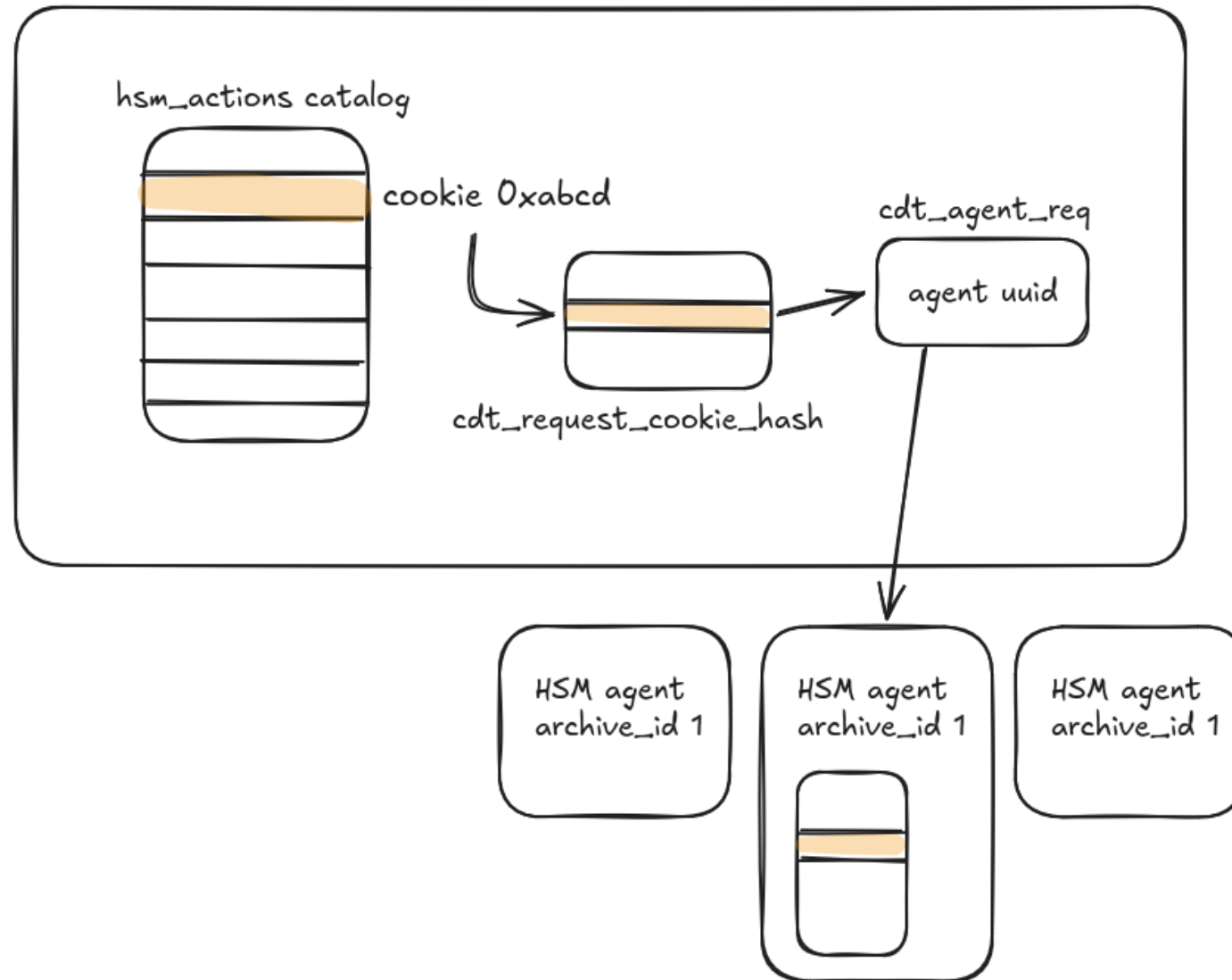


Lustre HSM: architecture

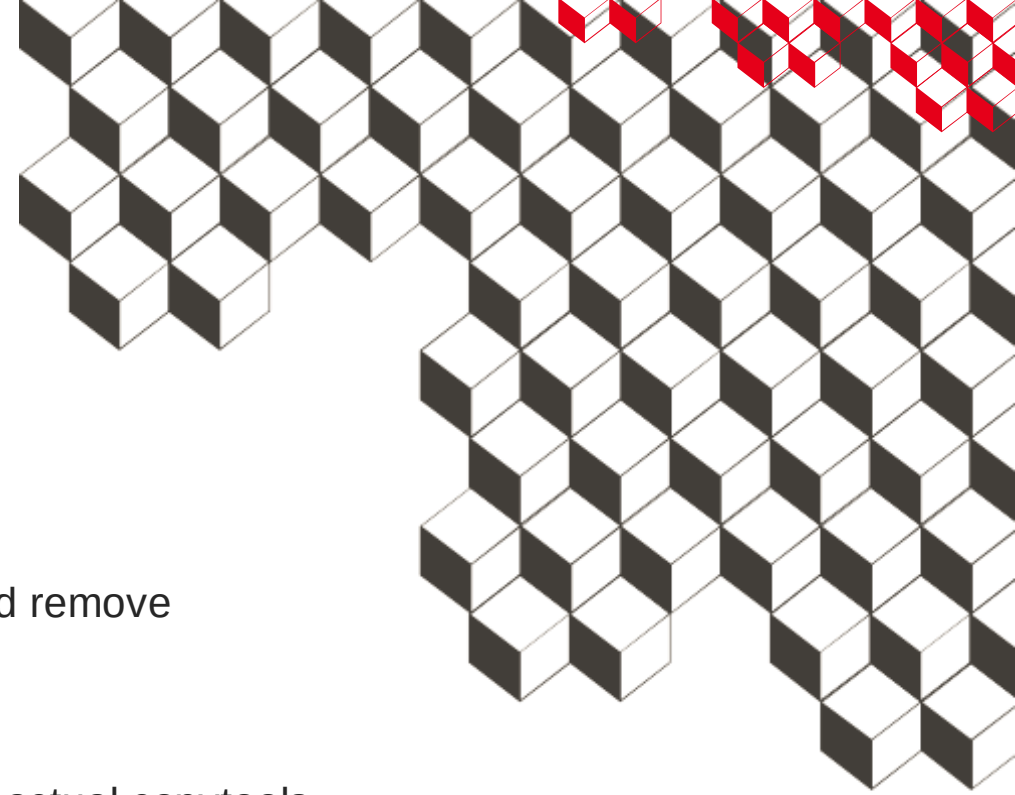


Lustre HSM: architecture

HSM cancel

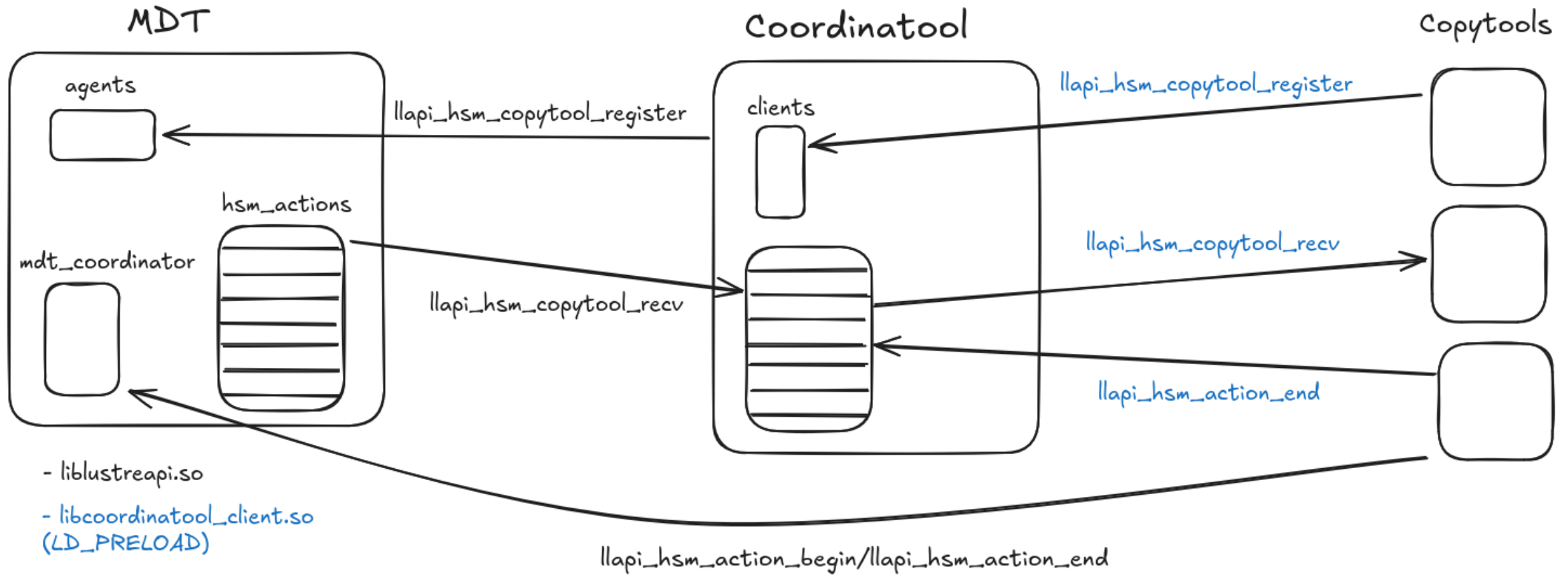


External coordinator

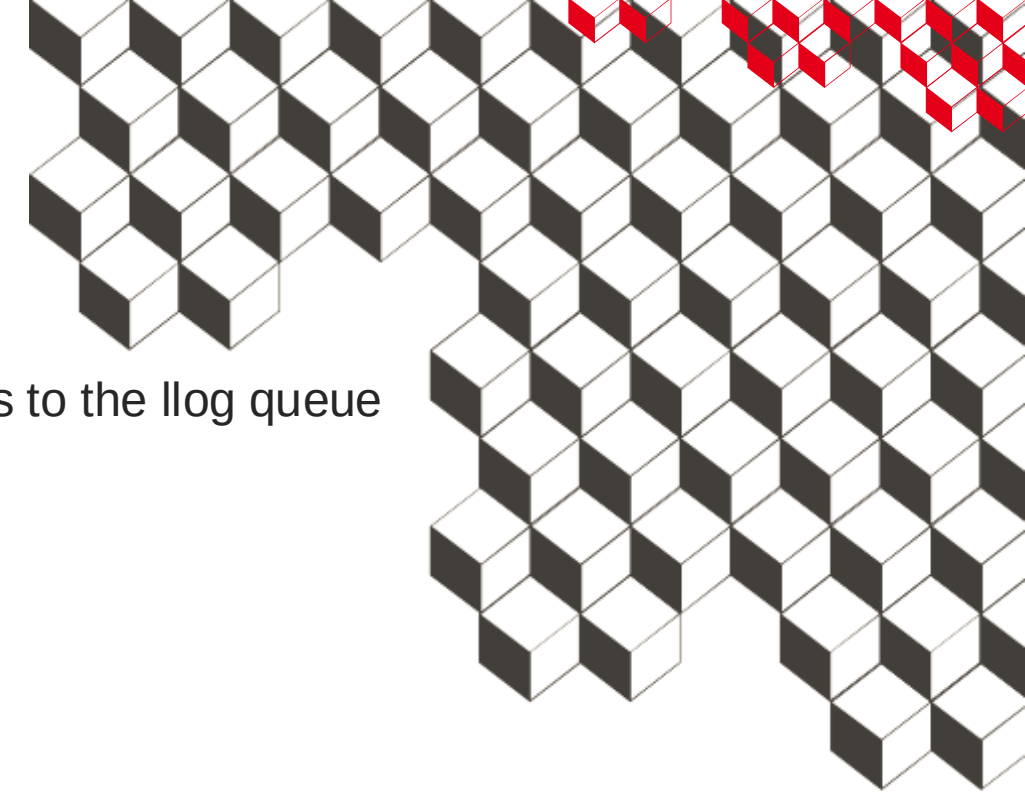


- The project seems to be abandoned by HPE
 - [LU-13384](#): HSM copytool API for external coordinator
 - [LU-10968](#): add coordinator bypass upcalls for HSM archive and remove
- The CEA used a special copytool [coordinatool](#)
 - It acts as an external coordinator to re-balance the requests to actual copytools
 - It uses LD_PRELOAD to overload llapi HSM. This is used to intercept calls from the real copytools.

External coordinator



Coordinator scalability

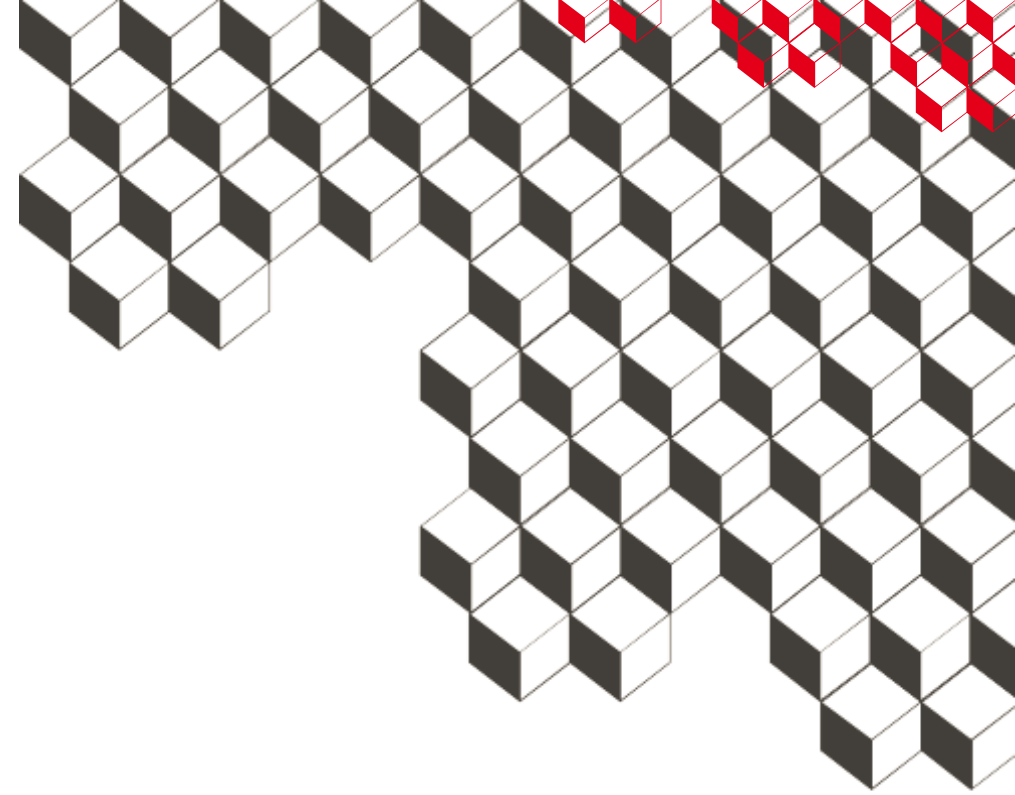


- Improve coordinator scalability == reduce/optimize accesses to the llog queue
- Merged patch
 - LU-13651 hsm: call `hsm_find_compatible_cb()` only for cancel
 - LU-14793 hsm: record index for further HSM action scanning
 - LU-16356 hsm: store `crh` in `rhashtable` instead of list
 - LU-18498 hsm: don't lock llog at `cdt`
- On-going
 - [57428](#) :LU-18556 hsm: optimize llog record modification
 - Cache the llog record for direct modification for started actions
 - [57396](#) :LU-18498 hsm: optimize `mdt_agent_record_update()`
 - Avoid scanning the whole llog for not contiguous records

HSM cancel

- The cancel request is stored on the catalog with the same cookie
 - This creates a lot of exceptions in the code
 - We need to process the catalog to find the original request's cookie
 - Request is not marked canceled when first received
- A few known bugs:
 - If 2 cancels or more end up in the same HAL for sending to a copytool, the HAL will be sent to the copytool holding the first request to be canceled regardless of which copytools have received the requests for the other requests
 - Original request is marked cancel only when it is scheduled for sending
 - Requests are not canceled until the cancel can be scheduled
 - If copytools are already busy, no active request can be canceled

HSM FLR mirror



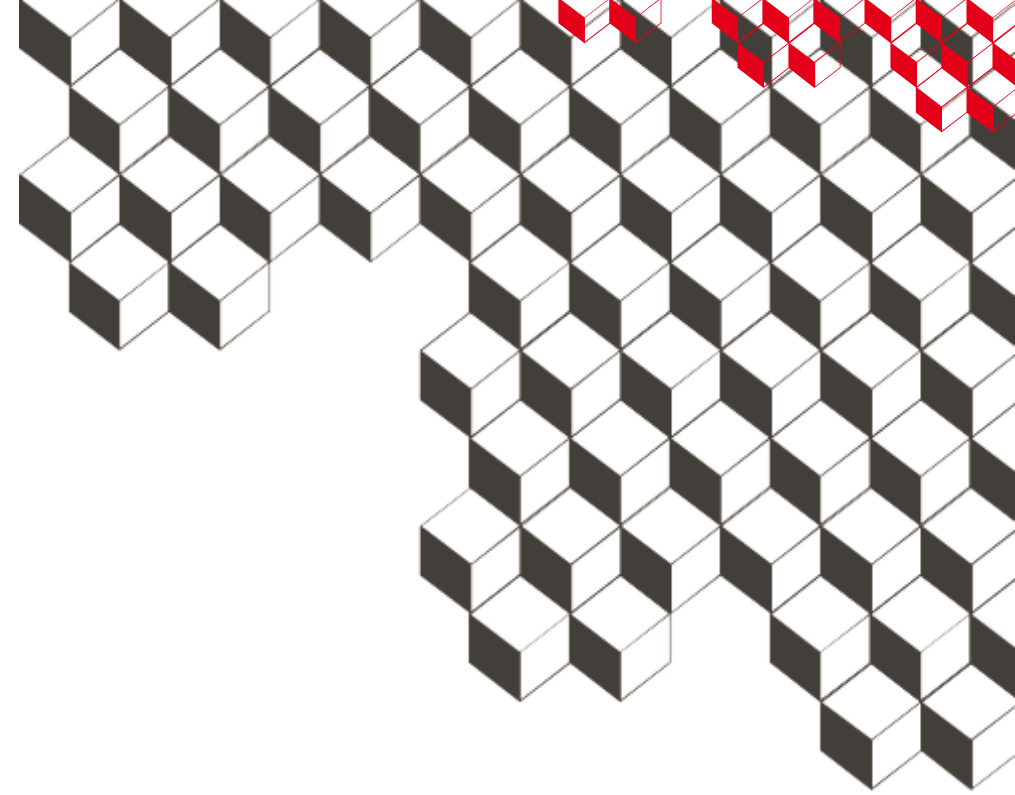
- Several tickets mention that:
 - [LU-13637](#): Combining RO-PCC/RW-PCC/HSM with FLR HLD
 - [LU-10606](#): HSM info as part of LOV layout xattr
 - [LU-10092](#): PCC: Lustre Persistent Client Cache
- Pro:
 - Support several archives for a file
 - Component/extent archive
 - Only 1 xattr needs to be updated (e.g: swap_layout)
 - Avoid workarounds like the: [LU-13048](#) mdd: allow release after a non-blocking migrate
 - It can use the same mechanism that FLR --> code deduplication
- Cons
 - Complex to implement
 - Backward compatibility

HSM FLR mirror

- 39387: LU-10606 hsm: store HSM xattr as a basic layout
 - Store the hsm xattr inside a FLR component
 - HSM state and FLR are redundant and could mismatch
e.g: HS_DIRTY without LCME_FL_STALE
- HSM states could be translated to FLR state:
 - HS_DIRTY = LCME_FL_STALE
 - HS_EXISTS at least one HSM mirror exist with LCME_FL_INIT
 - HS_RELEASED means that the layout contains only HSM mirrors
 - Etc...

HSM FLR mirror

- Backward compatibility:
 - trusted.hsm must still be accessible:
 - This can be reconstructed with the layout
 - All the hsm commands should work



Patch reviews

