Lustre* HSM Proposals

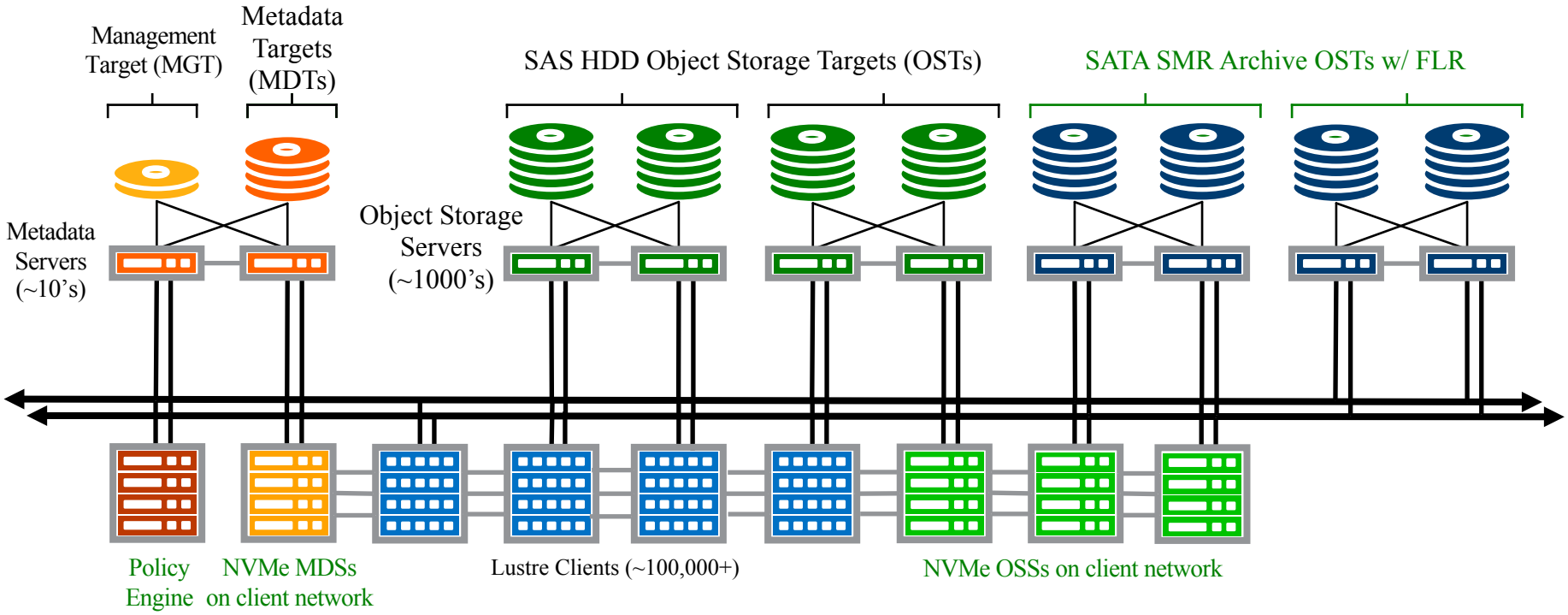
Andreas Dilger

High Performance Data Division

RUG 2016, Paris

Multi-Tiered Storage and File Level Redundancy

Full direct data access from clients to all storage classes



Multi-Tiered Storage and File Level Redundancy

Multi-Tiered Storage an upcoming requirement, with a single namespace

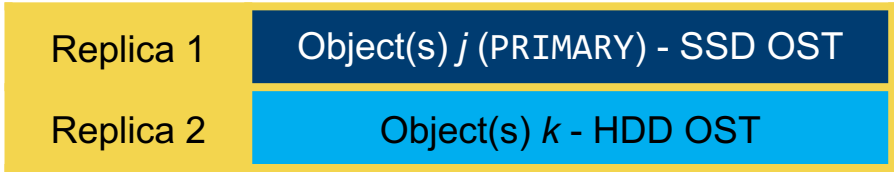
- Migrate NVM<->SSD<->HDD<->Archive, but allow direct access if needed
- Lustre OST Pool support improved in Lustre 2.9+ (filesystem/directory default pool)

File Level Redundancy brings significant value and functionality for **HPC**

- Can use lower-cost commodity single-port storage
- Availability better than HA failover, more reliable than any single device

Configure redundancy on a per-file/directory basis for flexibility/performance

- Can set default redundant layout on directory, or add redundancy after file is written
  - Mirror only 1 of 24 hourly checkpoints for recovery
  - 12+3 erasure code large striped files after write
  - Write to SSD, mirror to HDD, temporarily or not



FLR Phase 2: Integration with HSM Policy Engine

Leverage HSM Policy Engine, copytools to replicate/migrate across tiers

- Required functionality starting to appear in RobinHood v3
- Replicate/migrate by policy over tiers (path, extension, user, age, size, etc.)
- Release replica from fast storage tier(s) when space is needed/by age/by policy
- Run copytools directly on OSS nodes for fastest IO path
- Partial restore to allow data access before restore or migration completes

Migrate data directly by command-line, API, or job scheduler on demand

- Pre-stage input files, de-stage output files immediately at job completion

All storage classes in one namespace means data always directly usable

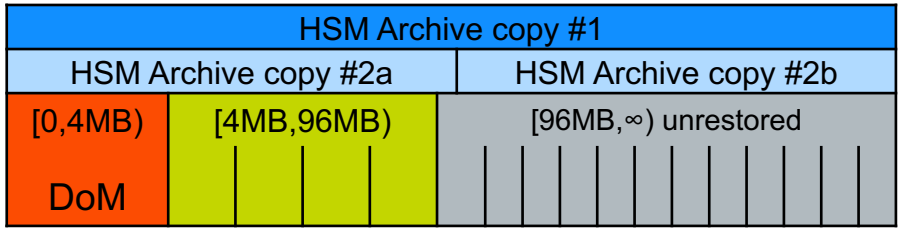
Partial File Restore for HSM

Composite File Layout allows different layout based on file offset

- Provides flexible layout infrastructure for multiple features, including:
  - Progressive File Layout (PFL)
  - File Level Redundancy (FLR)
  - Data-on-MDT (DoM)
- Layout components can be disjoint (e.g. PFL) or overlapping (e.g. FLR)

PFL could be used for HSM partial file restore

- HSM attrs could be layout components
- Archive copy is a special file replica
- Can have multiple replicas (e.g. offsite)



Legal Information

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.

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. For more complete information about performance and benchmark results, visit http://www.intel.com/performance.

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.

Intel technologies may require enabled hardware, specific software, or services activation. Check with your system manufacturer or retailer.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

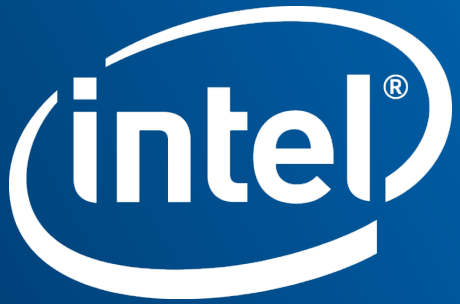
The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

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.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

Intel, the Intel logo and Intel® Omni-Path 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®