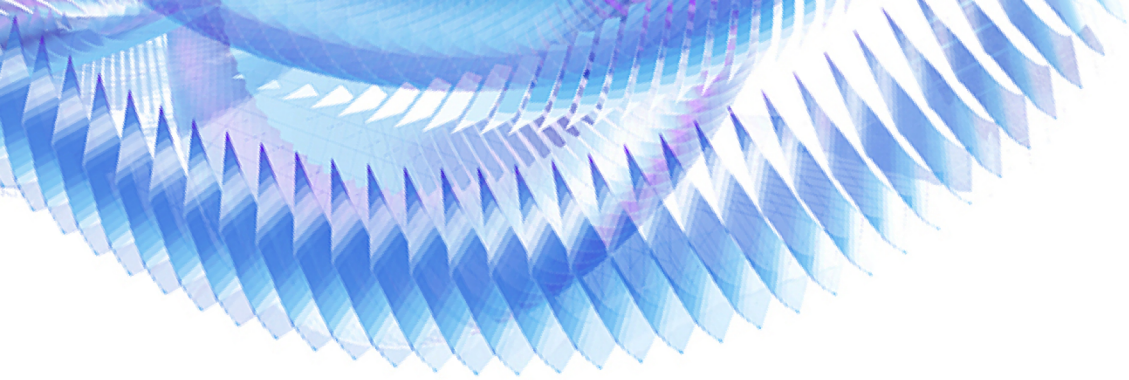


SC19
Denver, CO | hpc is now.

OVERVIEW OF FORTHCOMING DAOS FEATURES

Kristin Jacque, Intel
Liang Zhen, Intel

November, 2019

NOTICES AND DISCLAIMERS

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 product or component can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. For more complete information about performance and benchmark results, visit http://www.intel.com/benchmarks .

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/benchmarks .

Intel Advanced Vector Extensions (Intel AVX) provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at http://www.intel.com/go/turbo.

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

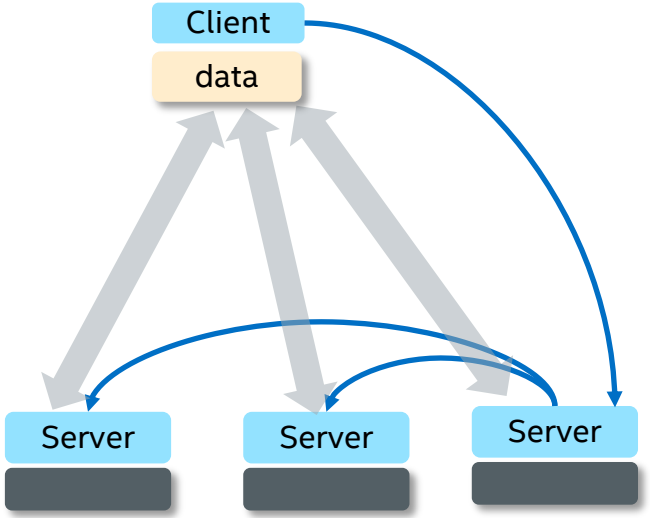
Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Forthcoming DAOS features

- Data protection and self-healing
- End-to-end data integrity
- Data aggregation
- Storage target reintegration and addition
- Security
- Long-term features

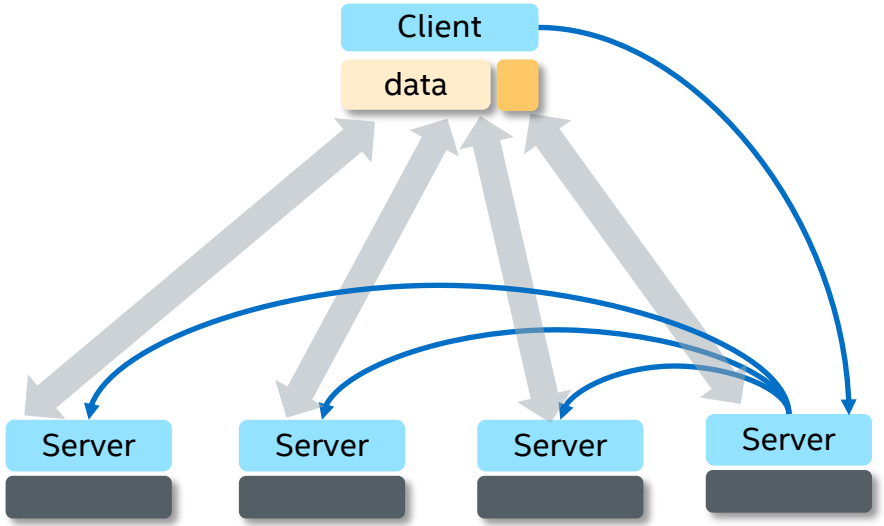
Data protection and self-healing - Replication

- Data replication in DAOS
  - Primary-slave replication
  - Distributed transaction for atomicity
- Degraded mode
  - Non-blocking protocol for server fail-out
- Online data recovery
  - Low impact on ongoing I/O
  - Declustered data reconstruction



Data protection and self-healing – Erasure code

- Erasure code in DAOS
  - Computed by client on write
  - Distributed transaction for atomicity
  - Replication for partial write
    - Merge and encode by server
- Degraded mode
  - Non-blocking protocol for server fail-out
  - Client side inflight data reconstructing
- Online data recovery
  - Server side data exchange and reconstruction

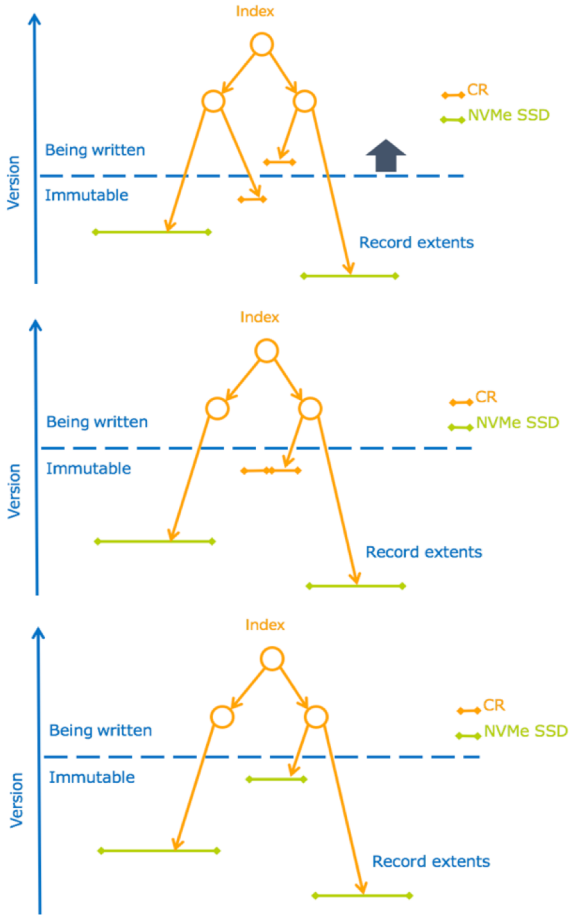


End-to-end data integrity

- Detect silent data corruption
  - Computed by client on write
  - Stored in persistent memory
  - Verified by server/client during I/O
- Correct data corruption
  - Data is protected by replication or Erasure coding
- Degraded mode
  - Disable I/O to corrupted object shard

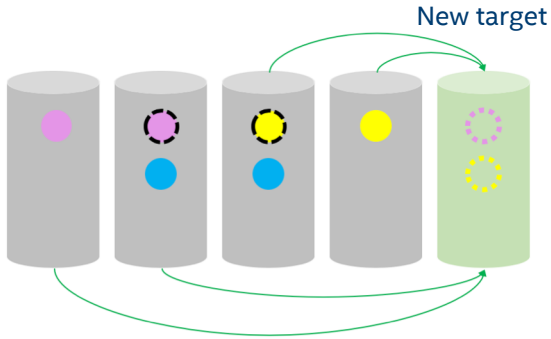
Data aggregation

- Merge small extents in SCM
  - Migrate merged extent to NVMe SSD
- Merage extents in NVMe SSD to larger extent
- Reclaim old snapshots
  - Overwrites: delete old version
  - Punch/delete: delete whole subtree
- EC aggregation
  - Compute parities for partial writes

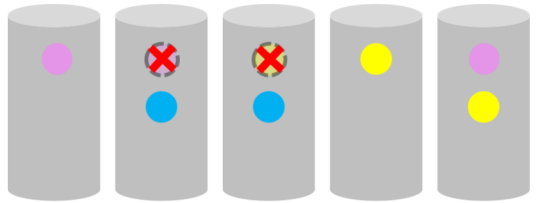


Storage target reintegration and addition

- Reintegrate recovered target to the pool
  - Add temporarily excluded storage targets back to the pool
    - Replaced: empty storage target
    - Not replaced: retained data but lagging behind
  - Migrate data back to the reintegrated targets
- Expand the pool size
  - Add more nodes/devices to the system
  - Rebalance data within the pool
- Online data rebalance



Rebalancing



After rebalancing

Security

Pool Access Control Lists

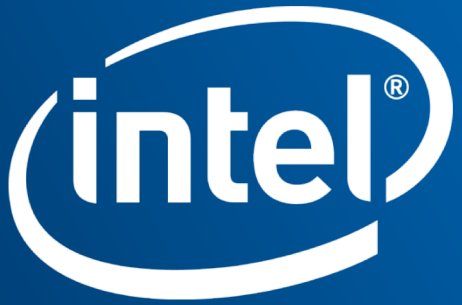
- Read/write permissions checked on pool connect
- Principals: Owner, named user, owner-group, named group, everyone
- DMG tool support
  - Pass in custom ACL on pool create
  - Get ACL
  - Overwrite ACL
  - Change/remove entries in ACL (coming soon!)

Certificates

- To identify servers, agent, admin (DMG tool)
- gRPC wire security (DMG to servers)
- “Insecure mode” – for development only

Long-term features

- Progressive layout
- Checksum scrubbing
- Telemetry & per-job statistics
- Advanced POSIX I/O
- Distributed transaction
- Data mover
- Catastrophic recovery


intel®