

INTRODUCING OPTANE DC PERSISTENT MEMORY & DAOS

DCPMM Technical Solution Specialist

Fumiyasu Ishibashi

fumiyasu.ishibashi@intel.com

intel OPTANE™ DC 
PERSISTENT MEMORY

✓ **BIG AND AFFORDABLE MEMORY**

128, 256, 512GB MODULES
DDR4 PIN COMPATIBLE

✓ **BYTE ADDRESSABLE**

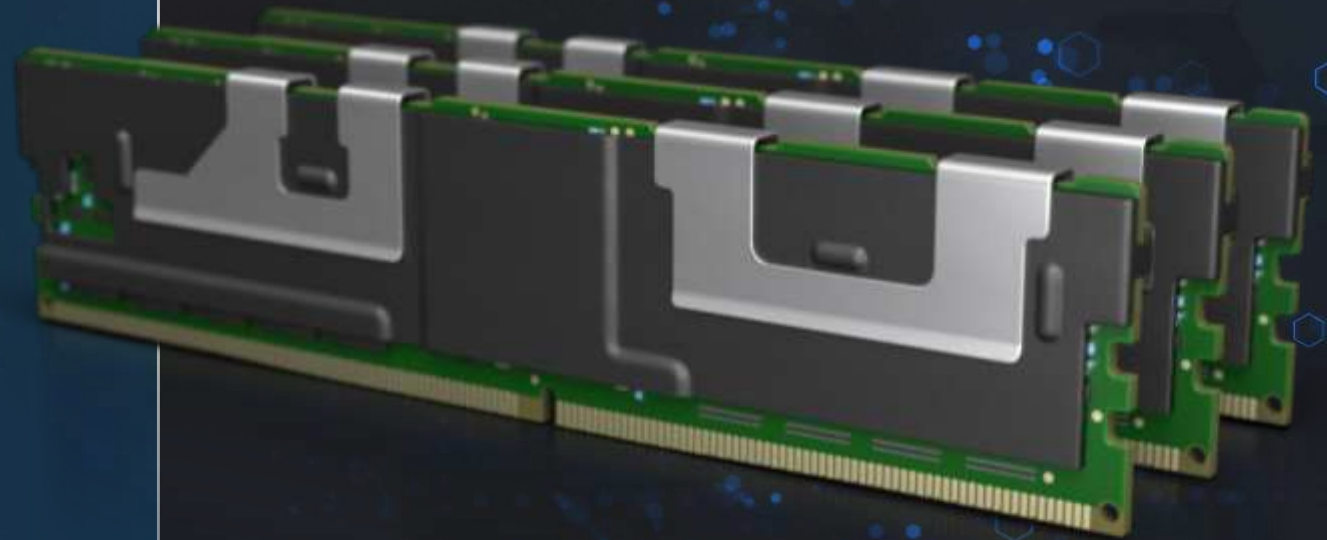
DIRECT LOAD/STORE ACCESS

✓ **HIGH PERFORMANCE STORAGE**

NATIVE PERSISTENCE

✓ **HIGH RELIABILITY AND SECURITY**

✓ **TWO OPERATIONAL MODES**



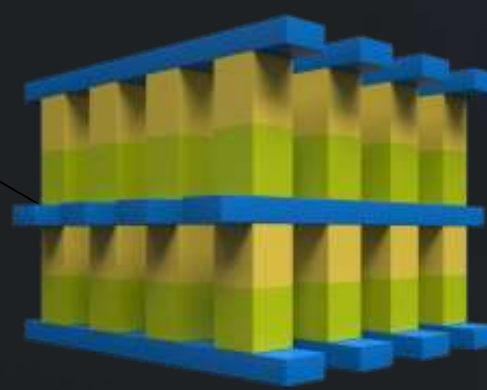
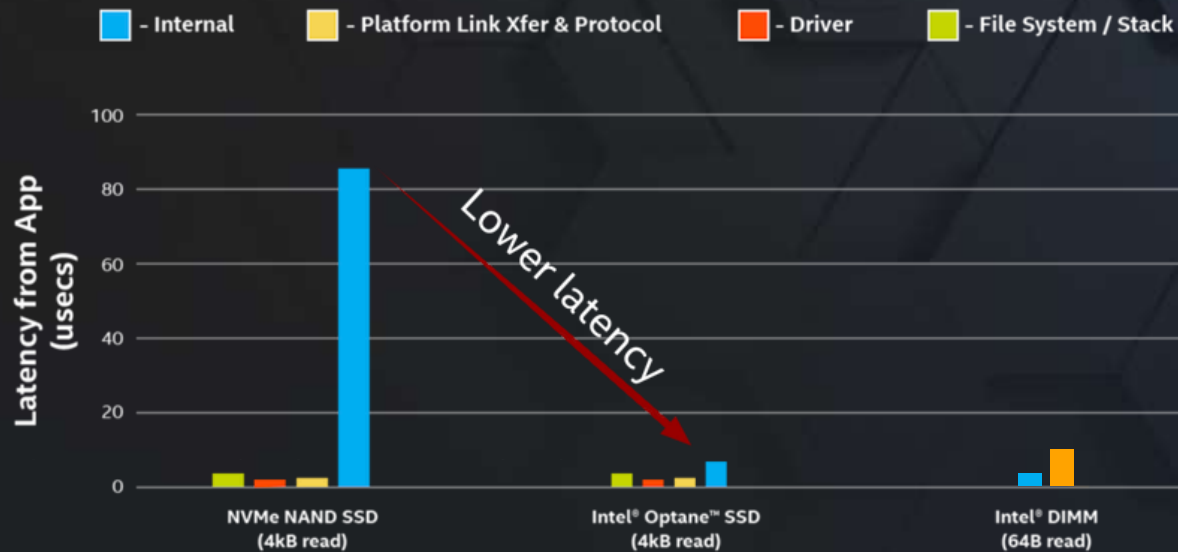
INTEL® OPTANE™ DC PERSISTENT MEMORY

DRAM IS **COSTLY**
SLOWING CAPACITY
GROWTH²

INTEL® OPTANE™ DC
PERSISTENT MEMORY
CLOSE MEMORY GAP

INTEL® OPTANE™ MEMORY
HIGH THROUGHPUT, LOW
LATENCY, HIGH ENDURANCE

NAND-BASED SSDs
NOT DESIGNED FOR
MEMORY FUNCTION



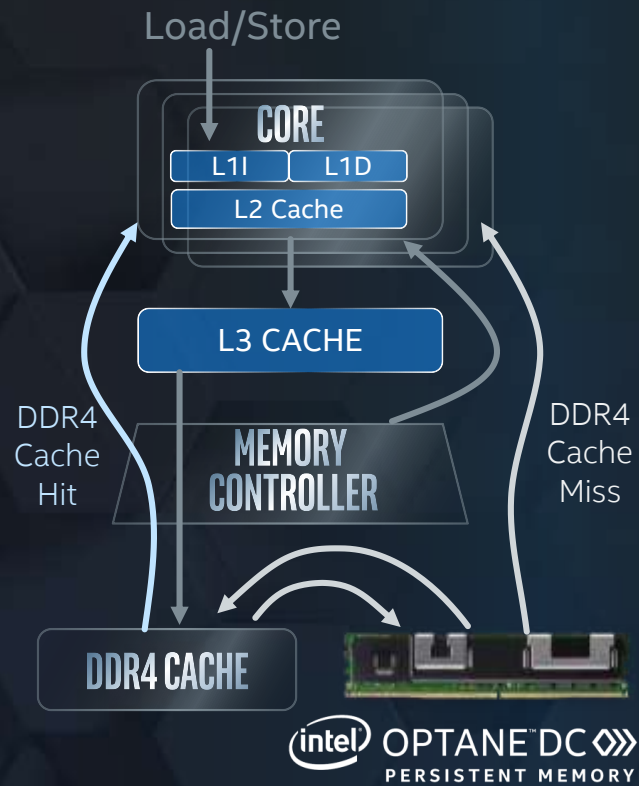
PERFORMANCE DETAILS

- Intel® Optane™ DC persistent memory is programmable for different power limits for power/performance optimization
- 12W – 18W, in 0.25 watt granularity - for example: 12.25W, 14.75W, 18W
- Higher power settings give best performance
- Performance varies based on traffic pattern
 - Contiguous 4 cacheline (256B) granularity vs. single random cacheline (64B) granularity
 - Read vs. writes

Granularity	Traffic	Module	Bandwidth
256B (4x64B)	Read	256GB, 18W	8.3 GB/s
256B (4x64B)	Write		3.0 GB/s
256B (4x64B)	2 Read/1 Write		5.4 GB/s
64B	Read		2.13 GB/s
64B	Write		0.73 GB/s
64B	2 Read/1 Write		1.35 GB/s

TWO OPERATIONAL MODES

MEMORY MODE



HW MANAGED

DATA PLACEMENT

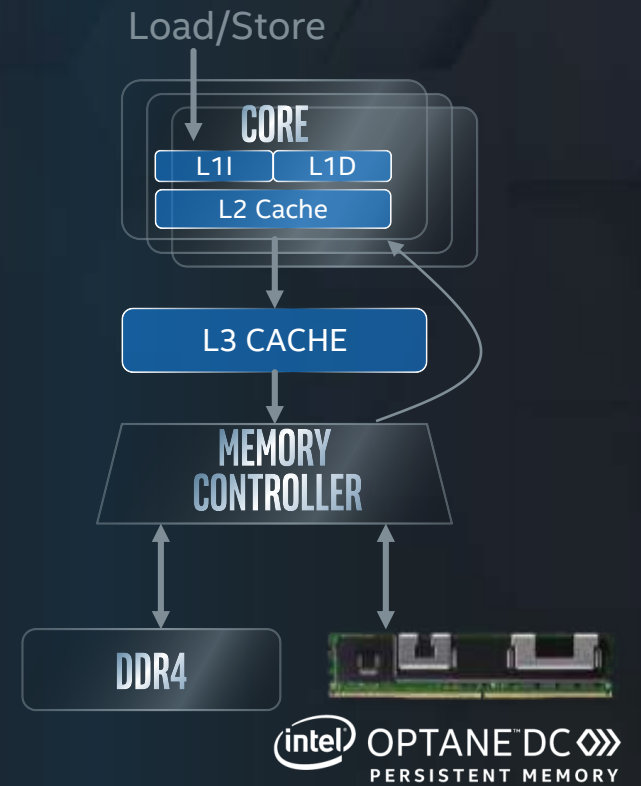
SW MANAGED

VOLATILE

PERSISTENCY

PERSISTENT

APPLICATION DIRECT MODE



APPLICATIONS ENABLED FOR INTEL® OPTANE™ DC PERSISTENT MEMORY - APP DIRECT MODE

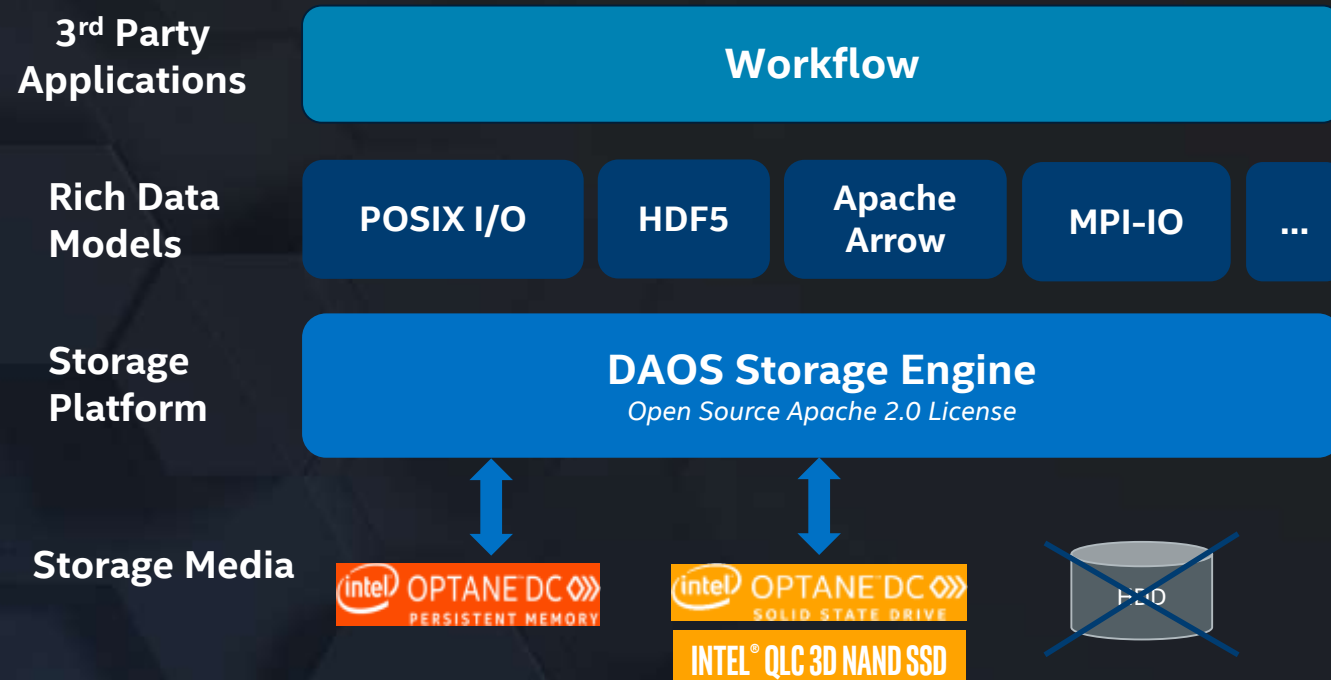
Intel® Optane™ DC persistent memory used in App Direct mode provides persistent performance & the maximum memory capacity. These are the list of applications with support announced by ISV or open source enabled by Intel for App Direct (AD) use cases.

Intel/Partner
sales collateral available

Application Type	Application Version	SW Type	Operating Mode
AI/Analytics	Panzura Vizion.ai v1.0	ISV	App Direct
Database	Aerospike* Enterprise Edition 4.8	ISV	App Direct
Database	Microsoft SQL Server 2019	ISV	App Direct
Database	SAP* HANA 2.0 SPS 03	ISV	App Direct
Infrastructure & Storage	NetApp* Memory Accelerated Data (MAX Data) 1.3	ISV	App Direct
AI/Analytics	Baosight* xInsight v2.0	ISV	App Direct
AI/Analytics	Cloudera* Apache Hbase* - CDH 6.2	ISV	App Direct
AI/Analytics	Gigaspace* Insight Edge Platform & XAP v14.0	ISV	App Direct
AI/Analytics	Optimized Analytics Package for Apache Spark* SQL version 2.3.2 (open source on github)	Open Source	App Direct
Database	Apache Cassandra* 4.x (open source on github)	Open Source	App Direct
Database	Apache HBase v3.0 (open source on github)	Open Source	App Direct
Database	Kingbase* Analytics Database (KADB V3R2)	ISV	App Direct
Database	KX* kdb+ 3.7t	ISV	App Direct
Database	Hazelcast IMDG 4.0	ISV	App Direct
Infrastructure & Storage	Apache Hadoop* 3.1 HDFS Cache (patch available)	Open Source	App Direct



DISTRIBUTED ASINCHRONOUS OBJECT STORAGE



Benefits

- Built natively over **new userspace** PMEM/NVMe software stack
- **Rich** storage semantics
- High **throughput/IOPS @arbitrary** alignment/size
- **Fine-grained, low-latency & True zero-copy** I/Os
- **Scalable** communications
- **Software-managed redundancy**
- Rely on **COTS** hardware

DAOS ARCHITECTURE



- High-latency communications
- P2P operations
- No HW acceleration

Conventional Storage Systems

Data & Metadata

Block Interface

Linux Kernel I/O

Intel® 3D-XPoint Storage

Intel® 3D-NAND Storage



HDD



- Low-latency, high-message-rate communications
- Collective operations & in-storage computing

DAOS Storage Engine

Metadata, low-latency I/Os & indexing/query

Bulk data

Memory Interface

PMDK

NVMe Interface

SPDK

Intel® 3D-XPoint Storage

3D-NAND/XPoint Storage

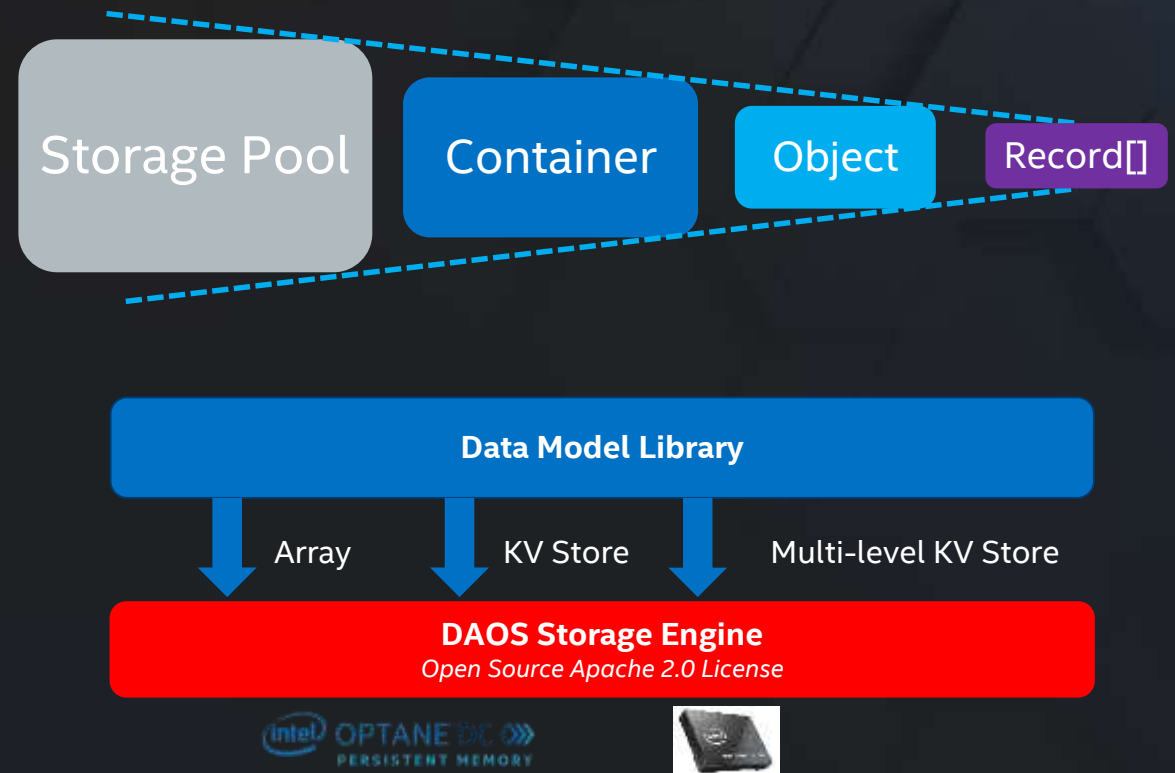


HDD



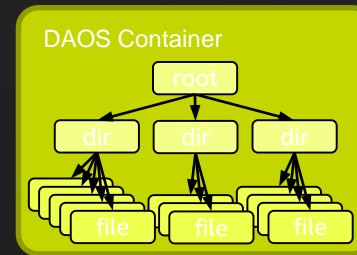
DAOS DATA MODEL

- Non-POSIX rich storage API as the new foundation
 - Scalable storage model suitable for both **structured & unstructured** data
 - key-value stores, multi-dimensional arrays, columnar databases, ...
 - Accelerate data analytic/AI frameworks
 - **Non-blocking** data & metadata operations
 - **Extendable** through microservice architecture

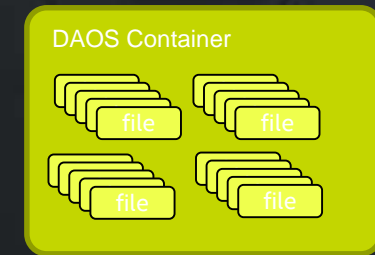


DATASET MANAGEMENT

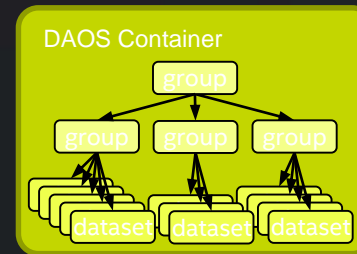
- Aggregate related datasets into manageable and coherent entities
 - Distributed consistency & automated recovery
 - Full Versioning
 - Simplified data management
 - Snapshot
 - Cross-tier Migration
 - Indexing



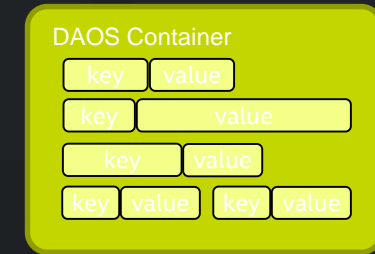
Encapsulated POSIX Namespace



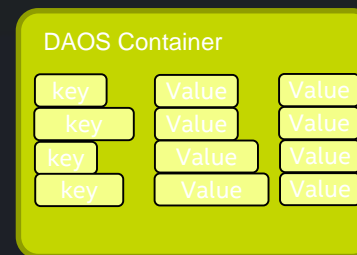
File-per-process



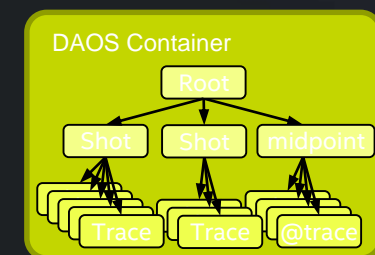
HDF5 « File »



Key-value store

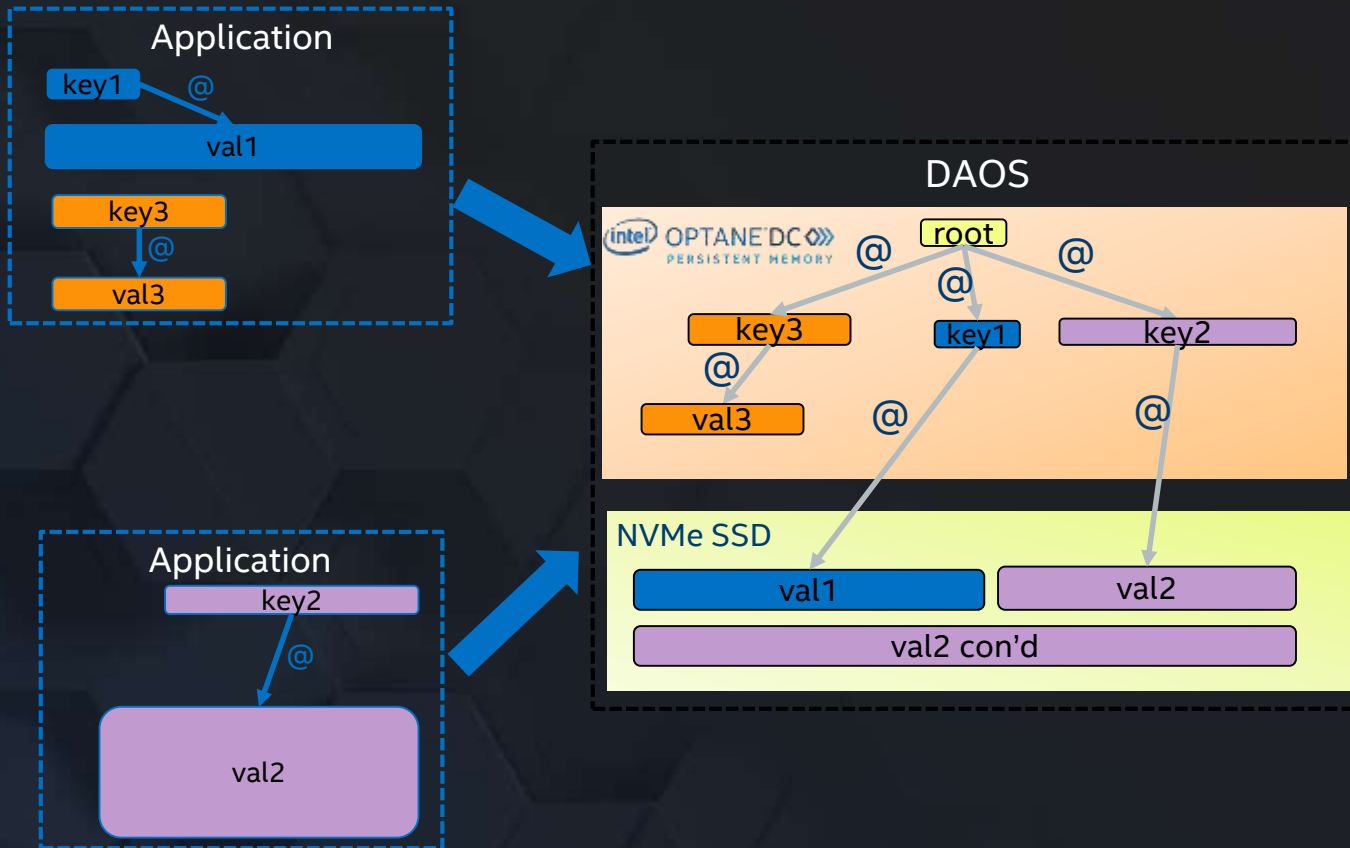


Columnar Database



SEGY

ADVANCED STORAGE API



Fast data retrieval

- Avoid file serialization and offset management
- Keys can be of any size/type
- Keys can be ordered with range query support

Scalable Insert & Fetch

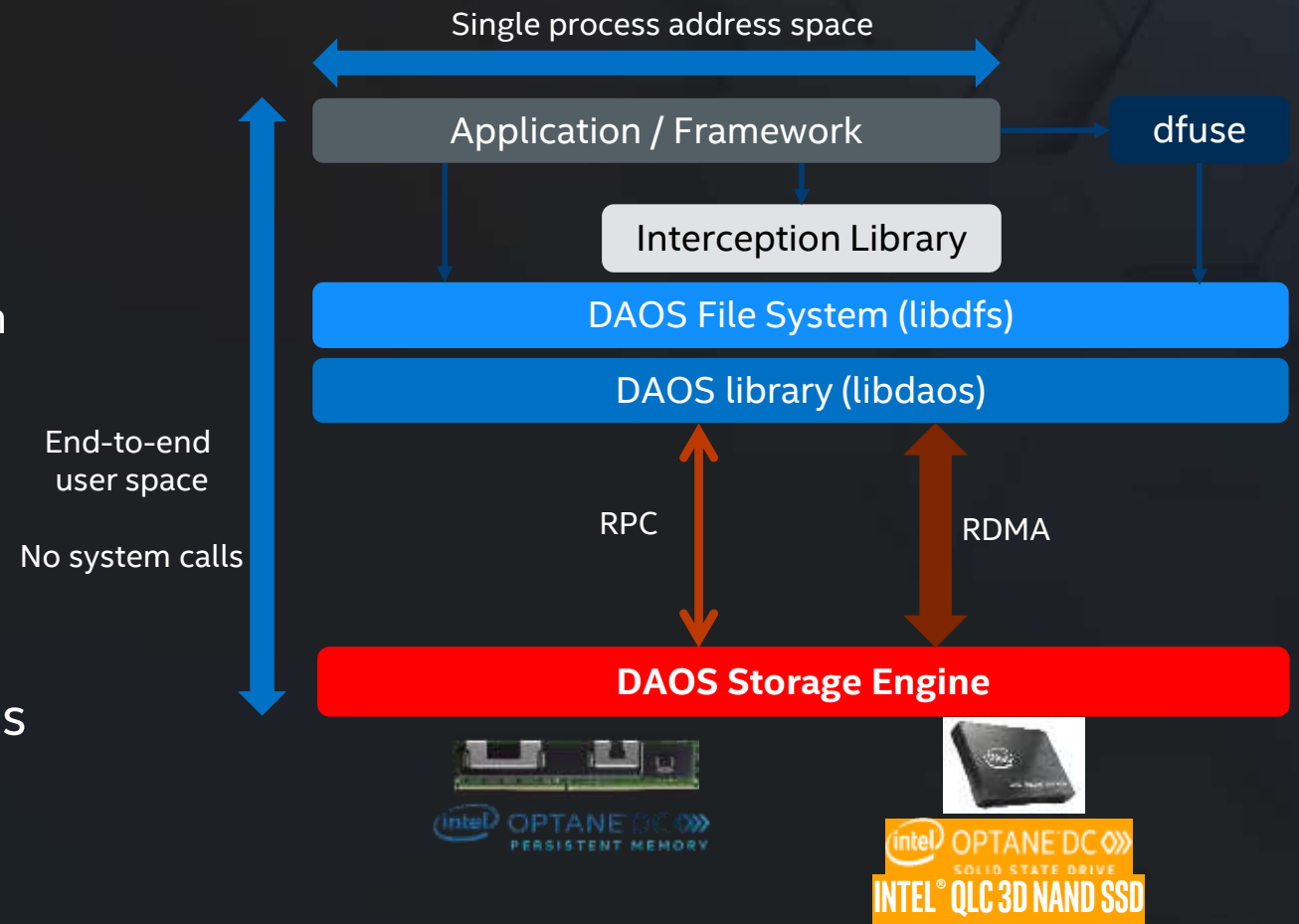
- Allow concurrent access/update
- Unconstrained by POSIX serialization
- Non-blocking
- Distributed transactions keep KV store always consistent

Data indexing

- Enable in-storage computing
- Query & custom index
- Data provenance

POSIX I/O SUPPORT

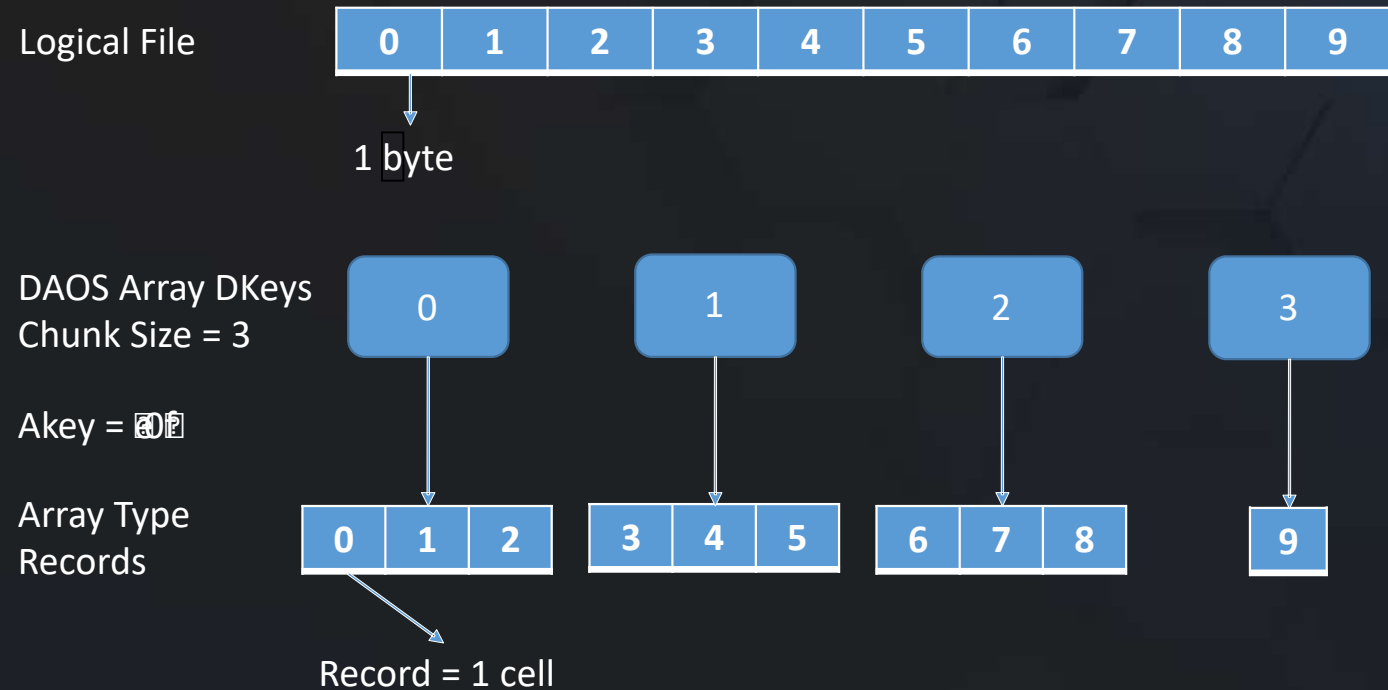
- DAOS File System (libdfs)
 - Encapsulated POSIX namespace
 - Application/framework can link directly with libdfs
 - ior/mdtest backend provided
 - MPI-IO driver leveraging collective open
 - TensorFlow, ...
- FUSE Daemon (dfuse)
 - Transparent access to DAOS
 - Involves system calls
- I/O interception library
 - OS bypass for read/write operations



MPI-IO DRIVER FOR DAOS

The DAOS MPI-IO driver is implemented within the I/O library in MPICH (ROMIO).

- Added as an ADIO driver
- Portable to Open-MPI, Intel MPI, etc.
- <https://github.com/daos-stack/mpich>
- daos_adio branch
- PR to mpich master in review
- 1 MPI File = 1 DAOS Array Object



Application works seamlessly by just specifying the use of the driver by appending "daos:" to the path.

DAOS: PRIMARY STORAGE ON AURORA



Aurora DAOS configuration

- Capacity: 230PB
- Bandwidth: >25TB/s

"Combined in Aurora, the Intel compute system, Cray Slingshot network, and the Intel DAOS storage open new possibilities for accelerating the scientific research needed to solve critical human challenges such as cancer and disease. DAOS enables the creation of new storage data models tailored specifically to applications like the Cancer Distributed Learning Environment (CANDLE) which provide a powerful platform to advance a wide array of scientific challenges using deep learning."

– Rick Stevens, Associate Laboratory Director for Computing, Environment and Life Sciences

"The Argonne Leadership Computing Facility is excited to be the first major production deployment of the DAOS storage system as part of Aurora, an US exascale system coming in 2021. As designed, it will provide us unprecedented levels of metadata operation rates and extremely high bandwidth for I/O intensive workloads."

– Susan Coghlan, ALCF-X Project Director/Exascale Computing Systems Deputy Director

DAOS COMMUNITY ROADMAP



DAOS:

- End-to-end data integrity
- Per-container ACL
- Improved control plane
- Lustre/UNS integration
- Replication & self-healing
- Erasure code (preview)
- Conditional updates

Application Interface:

- POSIX I/O with conditional update support

DAOS:

- Erasure code
- Telemetry & per-job statistics
- Distributed transactions

Application Interface:

- POSIX I/O with distributed transaction support
- HDF5 data mover
- Container parking/serialization

DAOS:

- Catastrophic recovery tools

DAOS:

- NVMe & DCPMM support
- Per-pool ACL
- UNS in DAOS via dfuse
- Replication & self-healing (preview)

Application Interface:

- MPI-IO Driver
- HDF5 DAOS Connector
- Basic POSIX I/O support
- Spark

DAOS:

- Online server addition
- Advanced control plane
- Multi OFI provider support

Application Interface:

- POSIX data mover
- Async HDF5 operations over DAOS

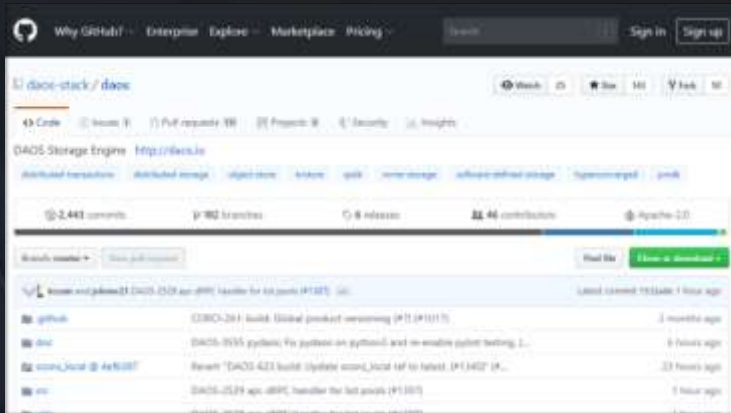
DAOS:

- Progressive layout
- Placement optimizations
- Checksum scrubbing

NOTE: All information provided in this roadmap is subject to change without notice.



RESOURCES



[Source Code on GitHub](https://github.com/daos-stack/daos)



[Admin Guide](#)



[DAOS Solution Brief](#)

Community mailing list on Groups.io

daos@daos.groups.io

Support

<https://jira.hpdd.intel.com>



LEGAL DISCLAIMER

All information provided here is subject to change without notice.

The products described in this document 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 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 intel.com.

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 www.intel.com/benchmarks.

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 data. You should review this content, consult other sources, and confirm whether referenced data are accurate.

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.



