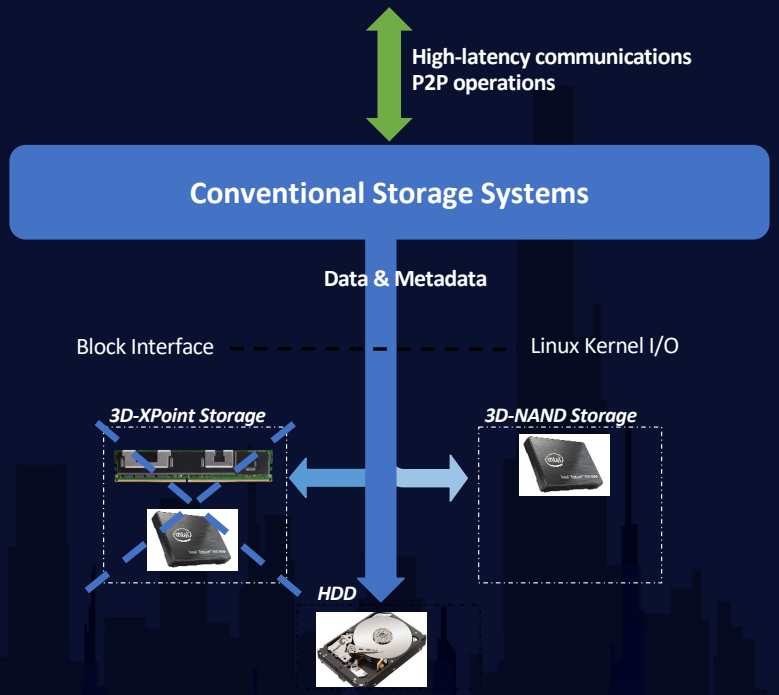
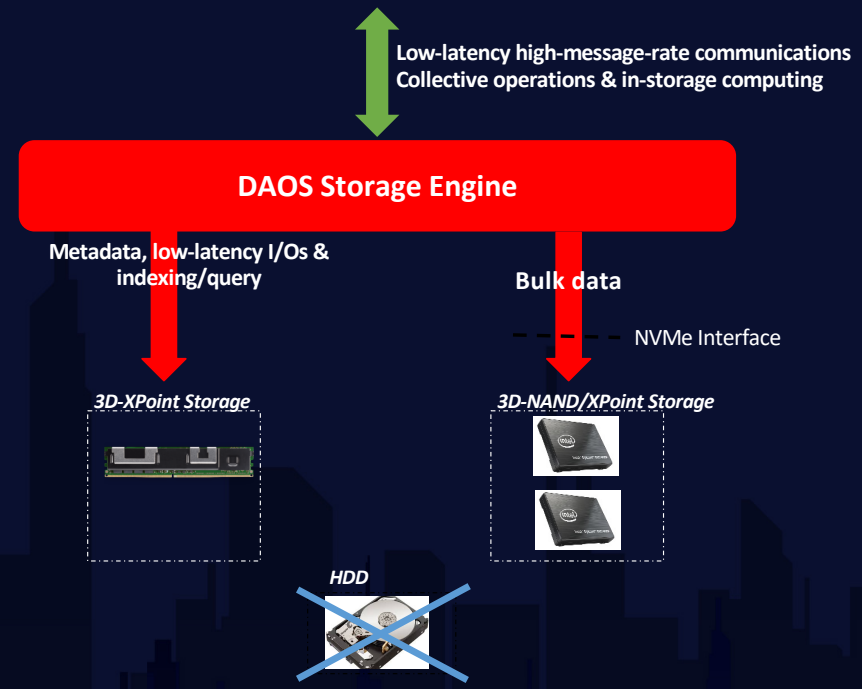
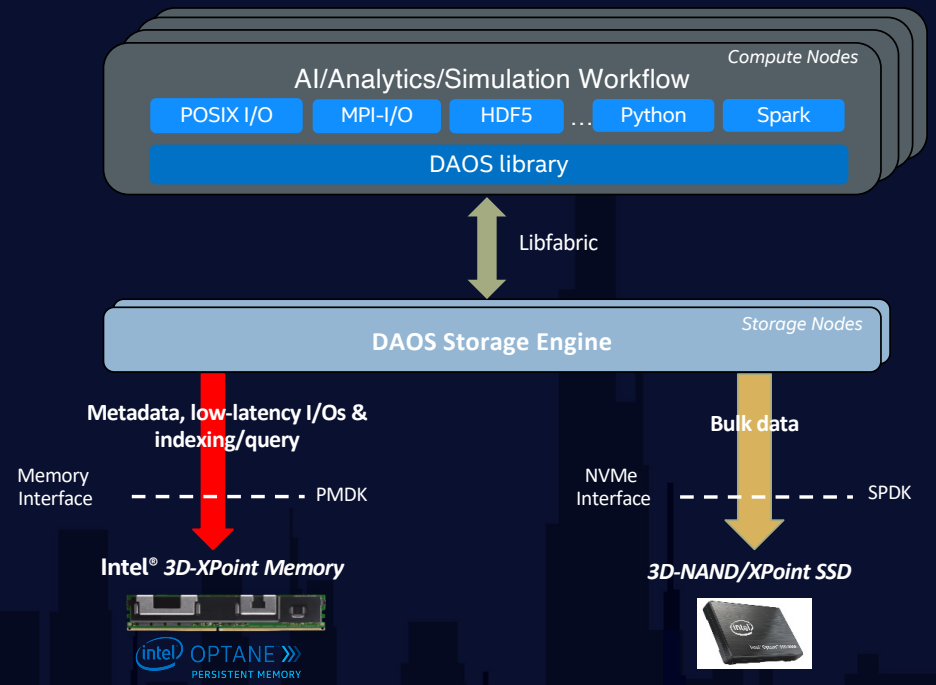

sodacon 2020
DATA CONNECTED
DISTRIBUTED ASYNCHRONOUS OBJECT STORE (DAOS) OVEWVIEW
ZHEN LIANG
Technical architect
Intel
soda
foundation

DAOS ARCHITECTURE





DAOS STACK OVERVIEW

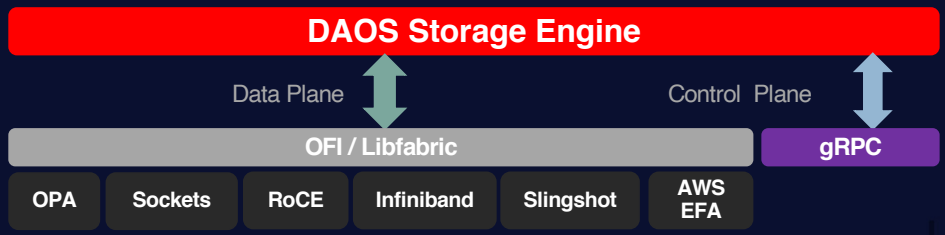


Deliver high-IOPS, high-bandwidth and low-latency storage with advanced features in a single tier

- DAOS library directly linked with the applications
- No need for dedicated cores
- Low memory/CPU footprint
- End-to-end OS bypass
- Non-blocking, lockless, snapshot support, …
- Low-latency & high-message-rate communications
- Native support for RDMA & scalable collective operations
- Support for OPA, Infiniband, OPA, Slingshot, RoCE, …
- Fine-grained I/O with media selection strategy
- Only application data on SSD to maximize throughput
- Small I/Os aggregated in pmem & migrated to SSD in large chunks
- Full userspace model with no system calls on I/O path
- Built-in storage management infrastructure (control plane)
- NFSv4-like ACL

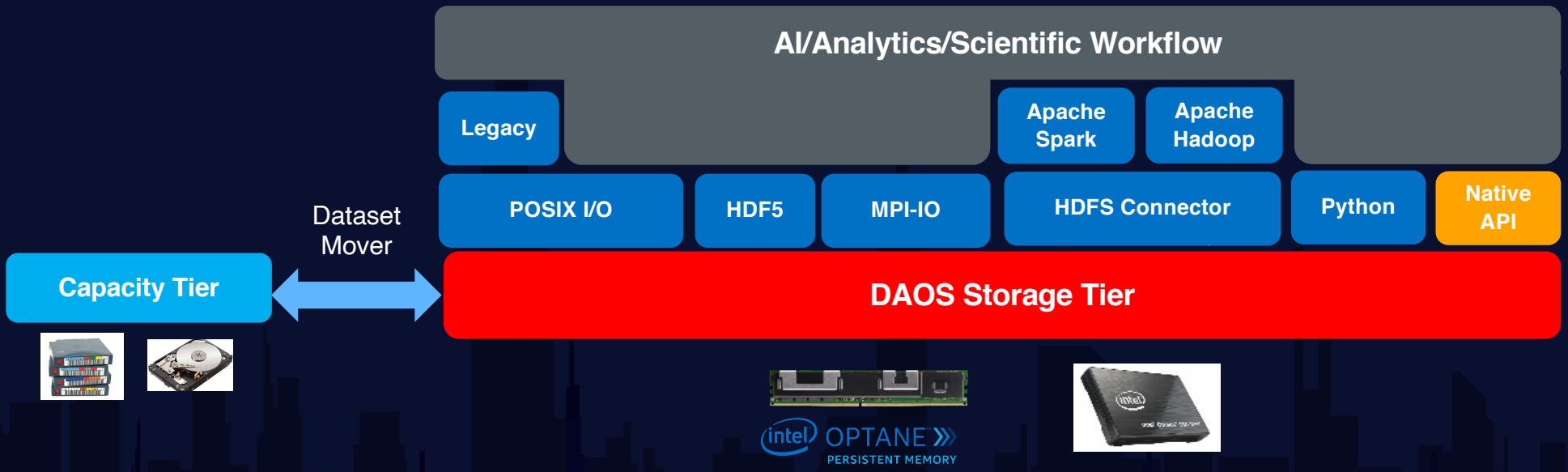
DAOS FEATURE SUPPORT

- Storage management
  - Integrated control plane
    - Deployment, firmware upgrade, …
    - Monitoring, telemetry & per-job stats
    - RAS events
  - Elastic storage
    - Storage node/SSD drain/reintegration
    - Online server addition
    - Online rebalancing
  - Security
    - Certificates & Access Control List (ACL)
  - DAOS-aware parallel data mover



- Networking
  - Native support for many interconnects
  - Optimized data placement
- Data protection & self-healing
  - Replication & Erasure code
  - End-to-end data integrity
  - Catastrophic recovery
- Advanced storage features
  - Dataset snapshot
  - Distributed serializable transactions

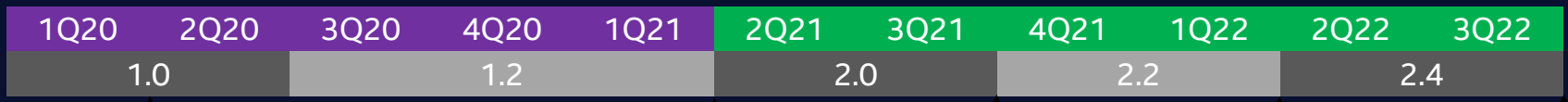
APPLICATION INTERFACE



BE PART OF SODA FOUNDATION

- Open source
  - Apache 2.0 license
  - github
- Distributed object storage
  - Key-value/array interface
  - Support different I/O middleware
  - Telemetry data
  - Data mover
- Vendor neutral
  - PDMK: SNIA standard, a vendor-neutral, growing collection of libraries started by Intel,
  - SPDK: vendor neutral
  - Libfabric: support TCP, Intel OPA, Mellanox, Cray GNI…
- Lead technology
  - Persistent memory, NVMe SSD, userspace RDMA
- High value use cases
  - Deliver high-bandwidth file I/O and high metadata rate to scientific simulations
  - Provide unmatched IOPS to speed up big data applications
  - Accelerate AI applications with orders-of-magnitude lower latency I/O to boost inference and high-bandwidth reads to train AI models faster
  - Bolster performance and scalability of distributed (No)SQL databases with native serializable transaction support

DAOS COMMUNITY ROADMAP



Released on June 18

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

DAOS:
- End-to-end data integrity
- Per-container ACL
- Improved control plane
- Replication & self-healing
- Conditional updates
- Erasure Code (Preview)
- Online server addition
- Lustre/UNS integration

Application Interface:
- HDF5 vol w/o async
- POSIX I/O with conditional update support
- POSIX data mover
- Async HDF5 operations over DAOS
- Spark

DAOS:
- Erasure code
- Telemetry& per-job statistics
- Advanced control plane
- Distributed transactions
- Multi OFI provider support

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

DAOS:
- Progressive layout
- Placement optimizations
- Checksum scrubbing

DAOS:
- Catastrophic recovery tools