intel

AURORA
LEADING HPC INTO THE FUTURE

Dr. Robert W. Wisniewski
Chief Architect, HPC, Intel
Aurora Technical Lead and PI

NOTICES & 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.

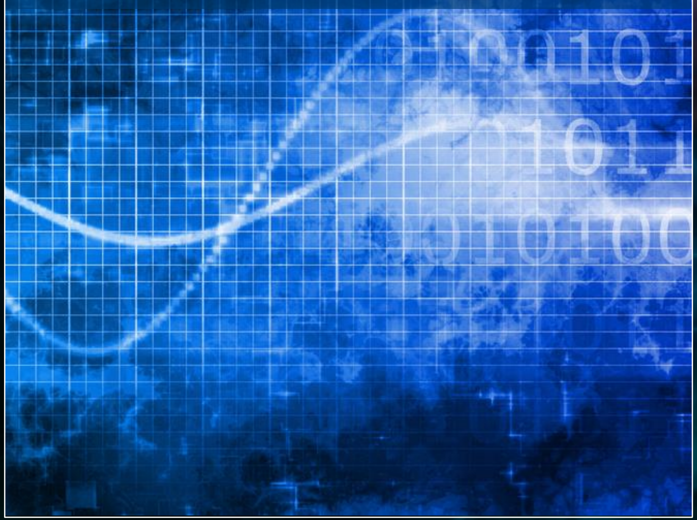
THREE PILLARS OF THE EXASCALE ERA

HPC SIMULATION

Model Drives Data

DATA ANALYTICS

Data Drives Insight



ARTIFICIAL INTELLIGENCE

Model Inferred from Data

DATA STORE

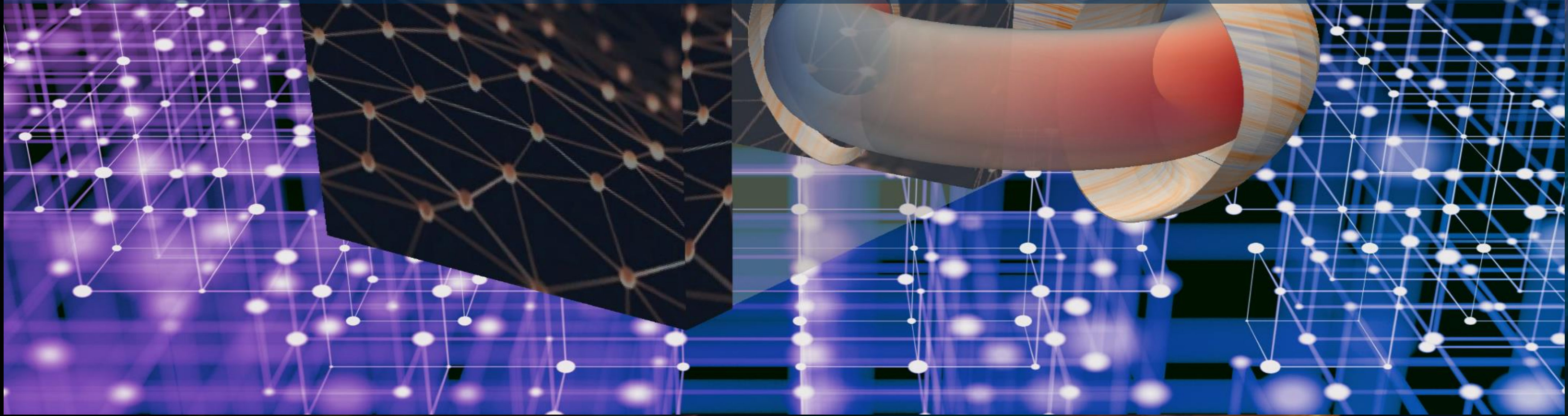
VISUALIZATION

ARTIFICIAL INTELLIGENCE/DEEP LEARNING BRINGS EXCITING NEW TECHNOLOGY TO ACCELERATE PROGRESS

"Predicting Disruptive Instabilities in Controlled Fusion Plasmas through Deep Learning"

NATURE: (accepted for publication, Jan. 2019, published, April 17, 2019 – DOI: 10.1038/s41586-019-1116-4)

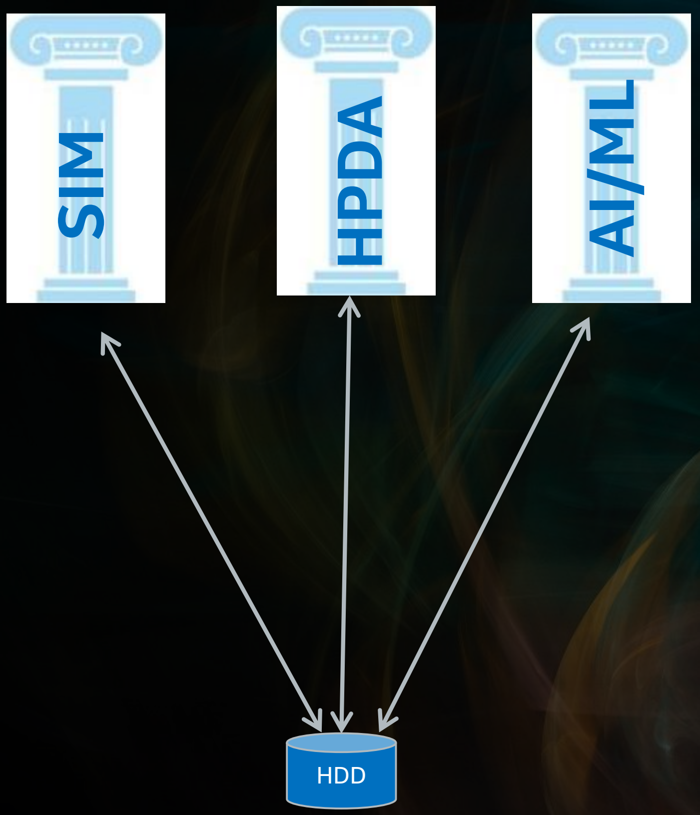
Princeton's Fusion Recurrent Neural Network code (FRNN) uses convolutional & recurrent neural network components to integrate both spatial and temporal information for predicting disruptions in tokamak plasmas with unprecedented accuracy and speed on top supercomputers



CONVERGED WORKLOADS BENEFIT FROM A TIGHTLY-COUPLED "DATA CENTRIC" ARCHITECTURE

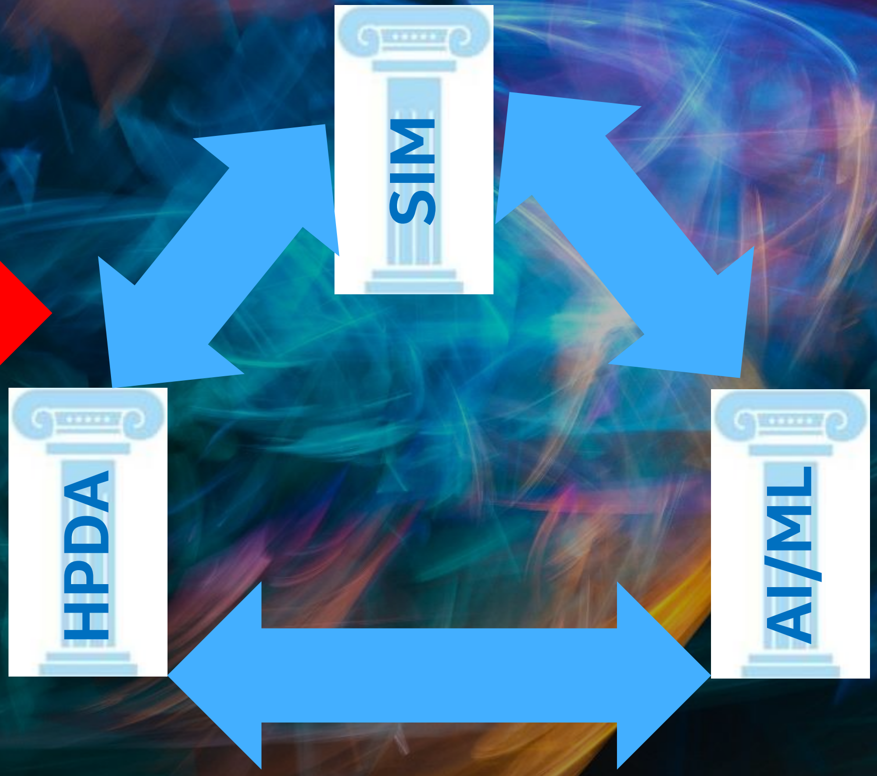
Today

(communication through thin linearized pipe to filesystem)



Tomorrow

(interactive workflows via tightly-coupled, high-bandwidth, active sharing of program data objects)



AURORA AT A GLANCE

Exascale = a billion billion (a quintillion) operations per second

Artificial Intelligence

Analytics

HPC Simulation

1 second
The time it takes Aurora to solve a math problem that would take 40 years if all the people on Earth each did **one calculation every 10 seconds.**

600 tons
The weight of Aurora, which equals that of an **Airbus 380.**

300 miles
The length of optical cable used in Aurora could reach **from Los Angeles to San Jose, California.**

10,000 square feet
The amount of floor space for Aurora, which **equals to 4 tennis courts.**

8 minutes
The time it takes Aurora to store enough characters to write **a stack of books that could reach the moon.**

34,000 gallons per minute
The rate of water moving through the **cooling loop.**

AURORA AT A GLANCE

Building the Foundation for Exascale Computing

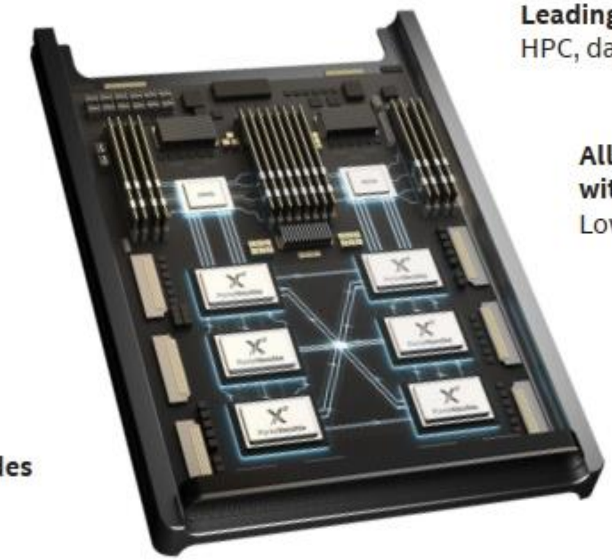
Aurora Node Architecture

2 Future Intel® Xeon™ Scalable Processors
"Sapphire Rapids"

6 X^e Architecture Based GPUs
"Ponte Vecchio"

oneAPI
Unified programming model

Unparalleled I/O Scalability across Nodes
8 fabric endpoints per node, DAOS



Leading Performance
HPC, data analytics, AI

All-to-All Connectivity within Node
Low latency, high bandwidth

Unified Memory Architecture
Across CPUs and GPUs

Packaging
Foveros and EMIB

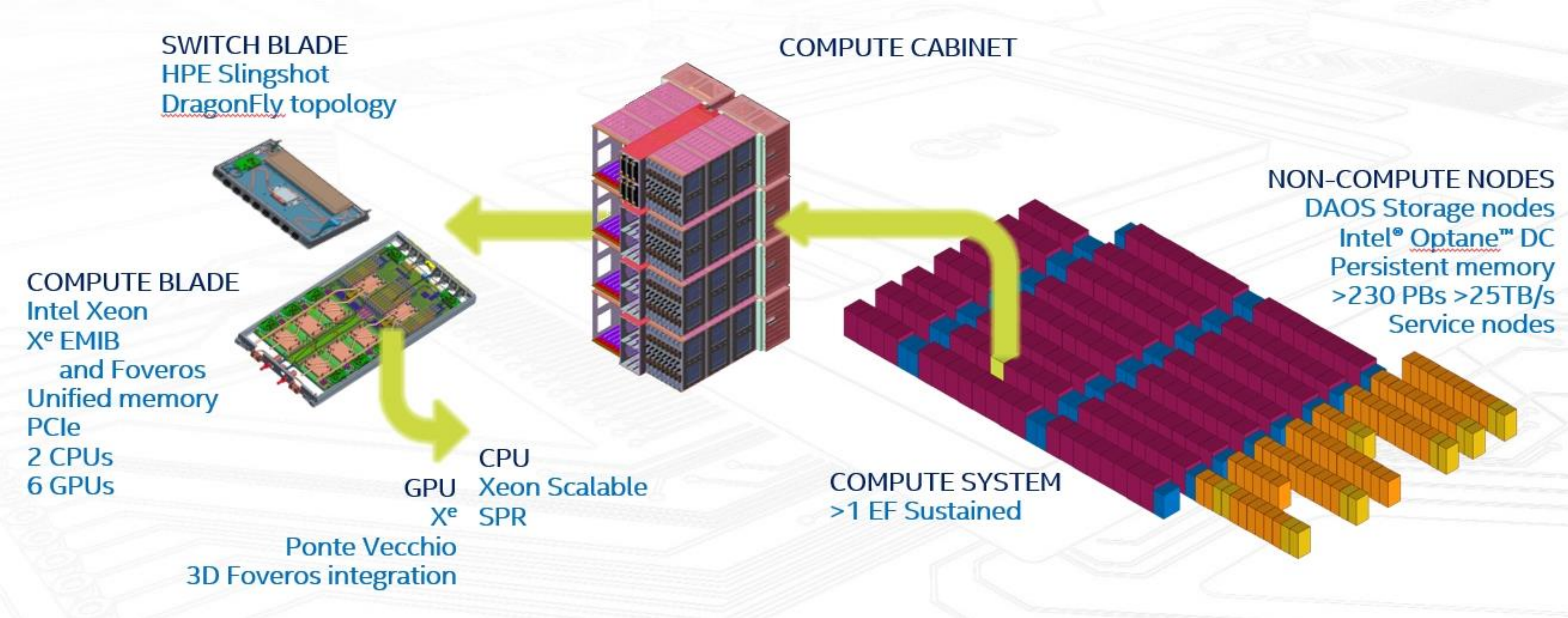
Unmatched Exascale-Class Storage Performance

Exascale systems require a completely rearchitected storage infrastructure. Aurora will benefit from the fastest High Performance Computing (HPC) storage on the planet – based on Intel® Optane™ persistent memory and the open source Distributed Asynchronous Object Storage (DAOS) framework, which together have enabled systems to achieve #1 ranking on the IO500 list.

Additional Details

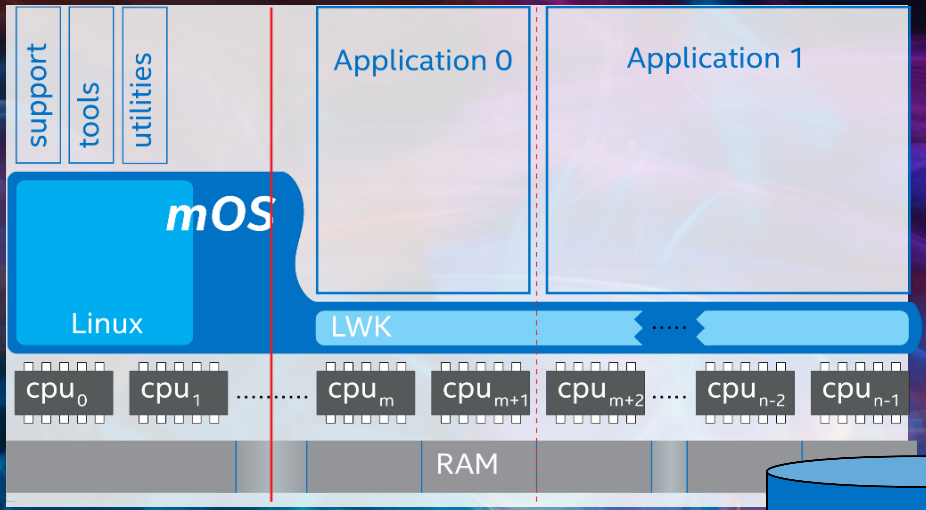
- Aurora will have more than 230 petabytes of storage with 25TB/s access rates
- Interconnect: HPE Slingshot
- Topology: Dragonfly
- Network switch: 64-port switch, 25GB/s per direction

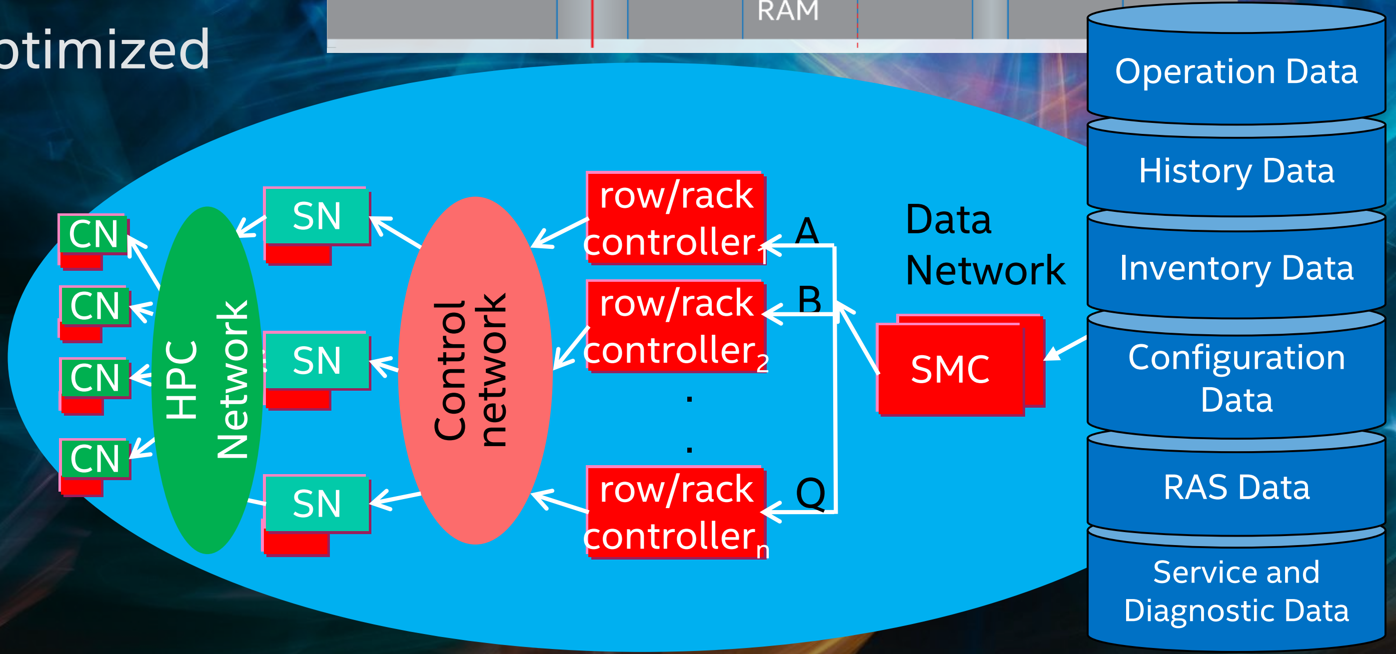
AURORA SYSTEM ARCHITECTURE



CORE SYSTEM SOFTWARE HPC COMPONENTS

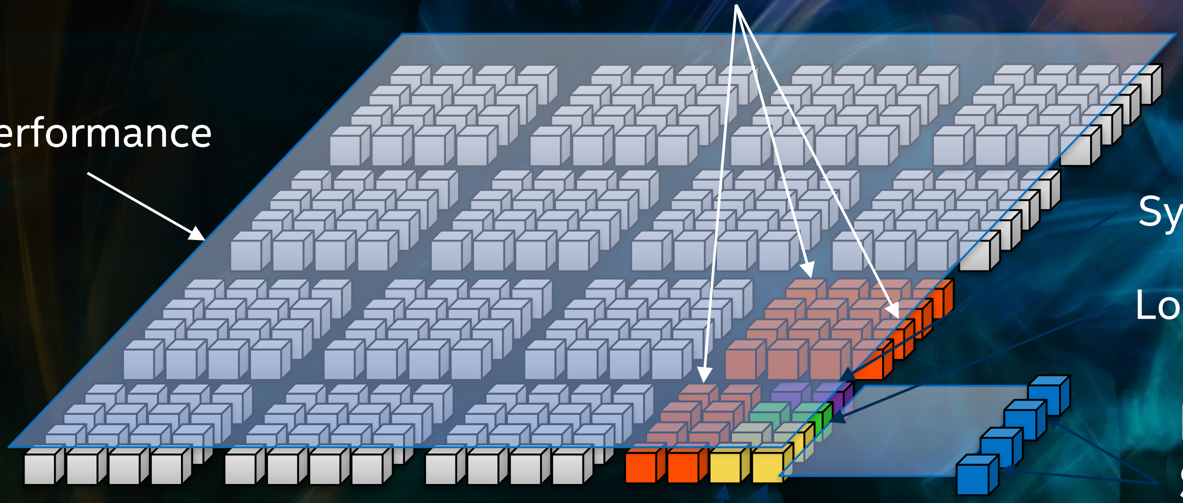
- **mOS**
  - Scalable operating system
- **Unified Control System**
  - Unified, Productive (single pane of glass), Reliable
- **MPI**
  - Scalable, high performance, topology optimized
- **GEOPM**
  - Global Extensible Open Power Manager
- **PMIx**
  - Process management with “Instant On”
- **DAOS**
  - Distributed Asynchronous Object Store





DISAGGREGATED HIGH-PERFORMANCE STORAGE USING DAOS

DAOS Nodes (DNs)
Xeon® servers
Storage-class memory and NVMe attached storage
DAOS service



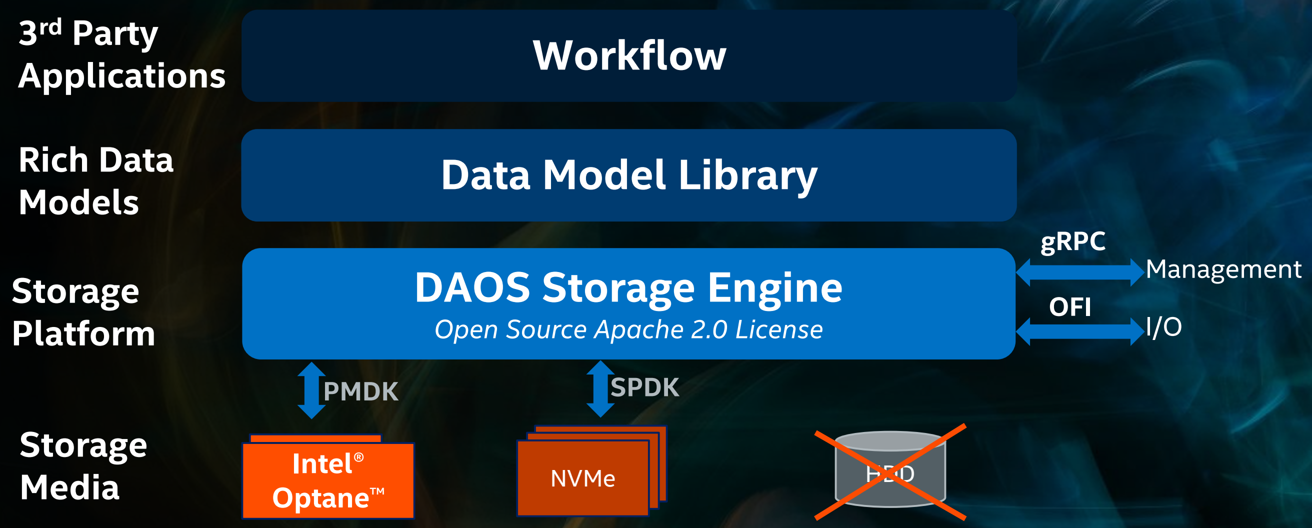
External Parallel File System(s)
Lustre, GPFS, …

Gateway Nodes (GNs)
Xeon servers with no local storage
IO forwarding service and data mover

intel.

DAOS: DISTRIBUTED ASYNCHRONOUS OBJECT STORAGE

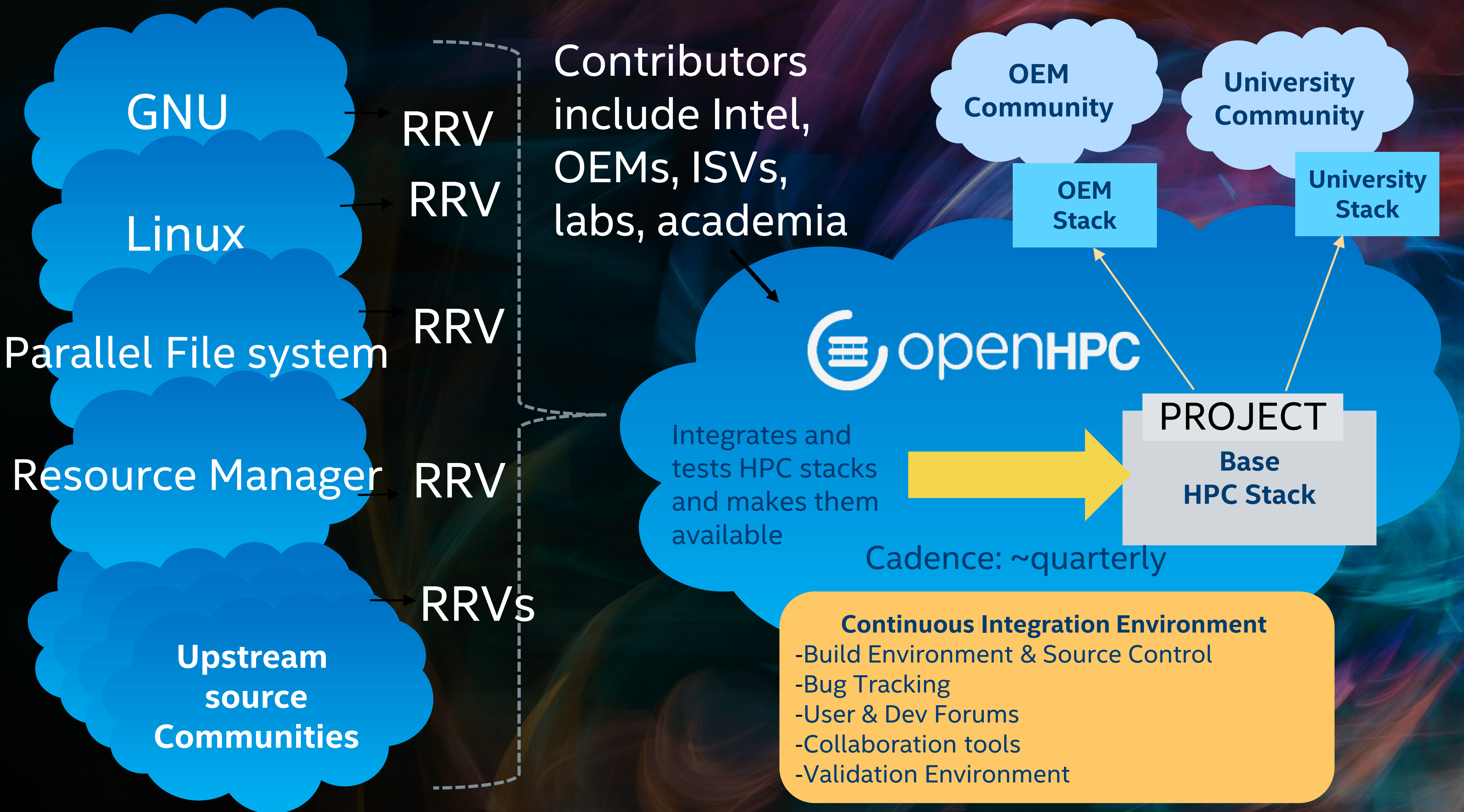
Scale-out object store built from the ground up for massively distributed NVM storage



DAOS Benefits

- Built over **new user space** PMEM/NVMe software stack
- High **throughput/IOPS** at **arbitrary** alignment/size
- **Ultra-fine grained** I/O
- **Scalable** communication and I/O over homogenous, shared-nothing servers
- **Software**-managed **redundancy**
  - Declustered replication and erasure code with self healing

OPENHPC



"RRV" = Relevant and Reliable Version

- Facilitates a vibrant and efficient software ecosystem
- Eases HPC application development
- Simplifies system administration and maintenance
- Extends to new workloads (AI and BD)
- Allows users to quickly take advantage of hardware innovation

CONVERGED WORKLOADS BENEFIT FROM A TIGHTLY-COUPLED "DATA CENTRIC" ARCHITECTURE

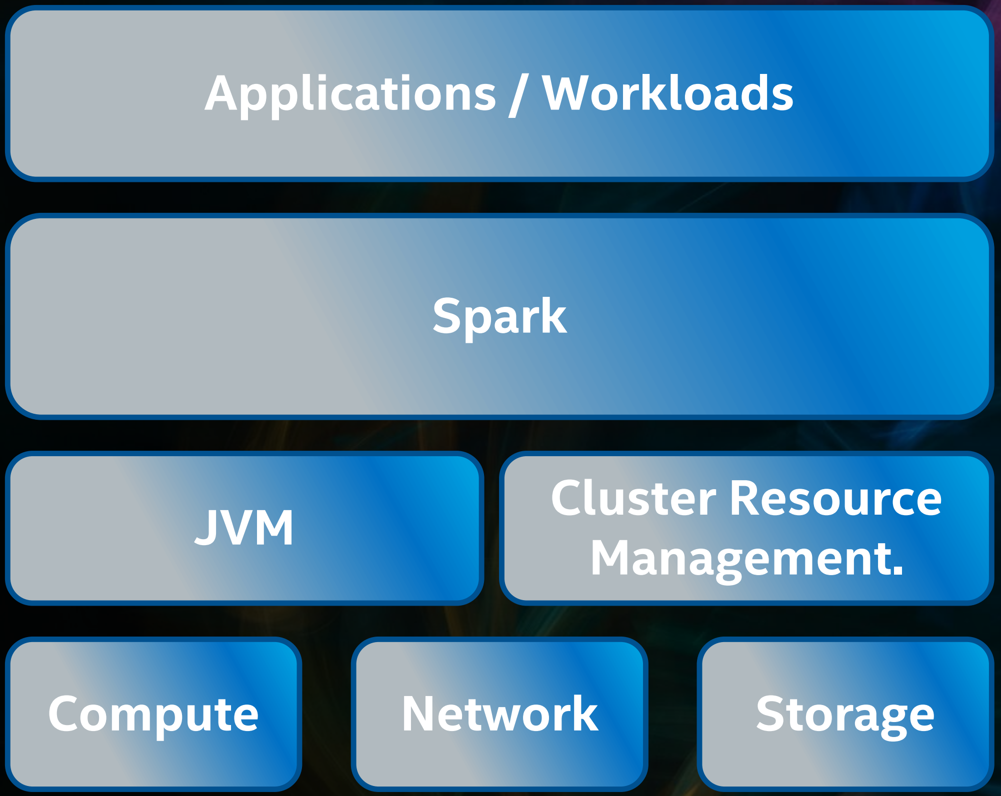
Today

(communication through thin linearized pipe to filesystem)

Tomorrow

(interactive workflows via tightly-coupled, high-bandwidth, active sharing of program data objects)

BRINGING SPARK ANALYTICS TO EXASCALE



1. Port workloads to Spark
2. Integrate with cluster resource management (in Spark Job Scheduler)
3. Support NUMA Aware Task Scheduling (in Spark Task Scheduler)
4. Support DAOS as intermediate data storage (in Spark BlockManager)
5. Support high performance fabric (in Spark Shuffle)
6. Support kernel offloading to new hardware (in Spark MLlib
7. Support DAOS as input/output storage (in Spark DataSource)

Bring Spark analytics capability to Exascale
Leverage new hardware and high performance fabric to achieve great performance

SUMMARY OF HETEROGENEITY TRENDS

- Heterogeneity is all around us
  - Compute, memory, I/O, software ecosystem
- New types of compute requirements
  - AI, Big Data, Edge
- AI and Cloud are large markets and thus primary drivers of requirements
- HPC is more complex than ever and fundamental shifts are occurring

ADDRESSING CHALLENGES RAISED BY TRENDS

- System design methodology needed
- Leverage massive investment by cloud and AI, but optimize for HPC
  - ex: GPUs
- Integrate heterogeneous components at the right level
- Provide a programming model encompassing expanding compute
  - Scalar, Vector, Matrix, Spatial, Mixed Precision, and Edge ←→ HPC machines
- Provide scalable software that supports new data models
- Facilitate platforms for converged HPC, AI, and Big Data computing

Arigato Gozaimashita