
intel
Leveraging Heterogeneity
Aurora as a Critical Step
Dr. Robert W. Wisniewski
CTO and Chief Architect HPC, Intel
Aurora Technical Lead and PI

Notices and Disclaimers

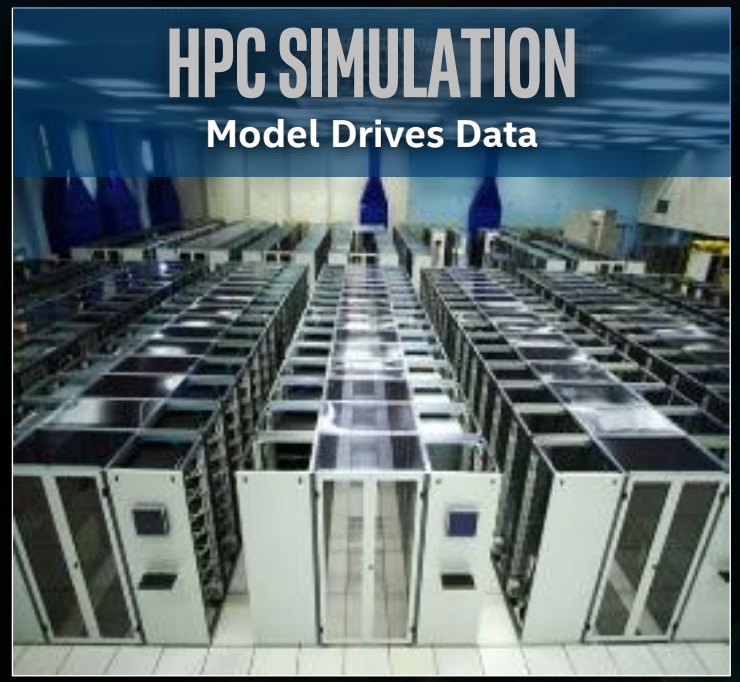
Intel technologies may require enabled hardware, software or service activation.

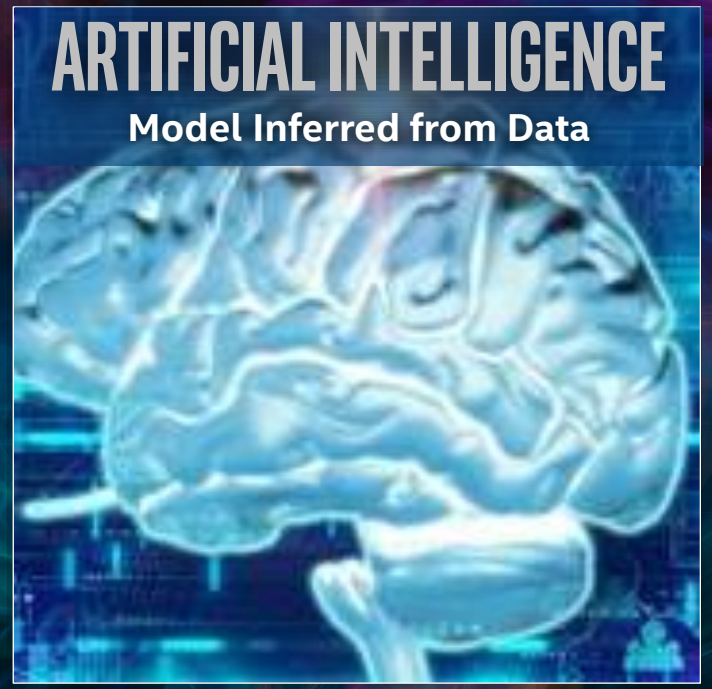
No product or component can be absolutely secure.

Your costs and results may vary.

intel®

Three Pillars of the Exascale Era





Data Store

Visualization

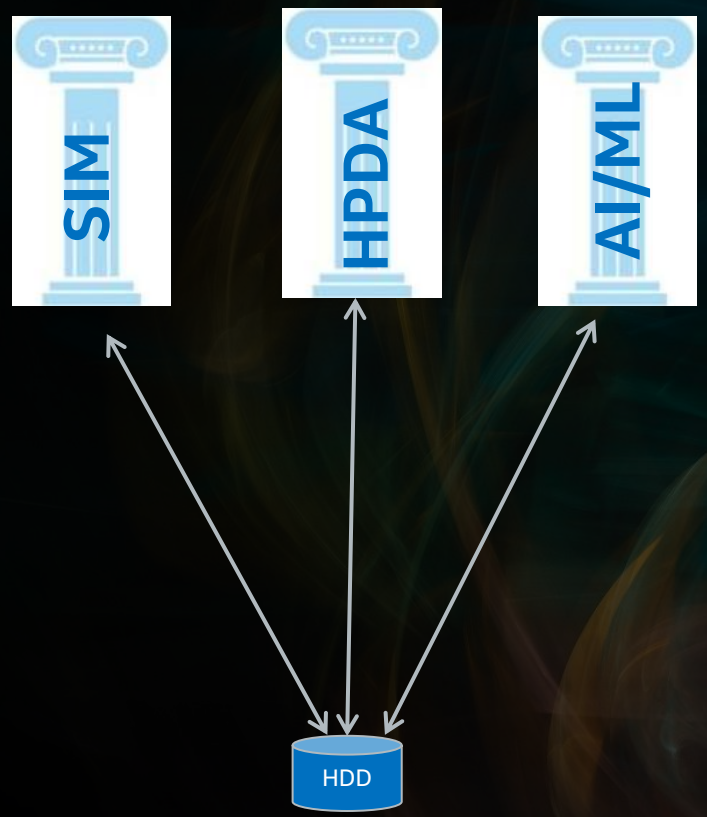
Key Trends Affecting HPC

- Heterogeneity is all around us
  - Compute: Scale, Vector, Spatial, Matrix (SVSM)
  - Software
    - Classical HPC stack, AI/ML frameworks, Big Data
  - Memory, I/O
- New types of compute requirements
  - AI, Cloud, Big Data, Edge
  - AI and Cloud are large markets and key drivers of requirements
- HPC is more complex than ever and fundamental shifts are occurring

Converged workloads benefit from a tightly-coupled “Data Centric” architecture

Today

(communication through thin linearized pipe to filesystem)

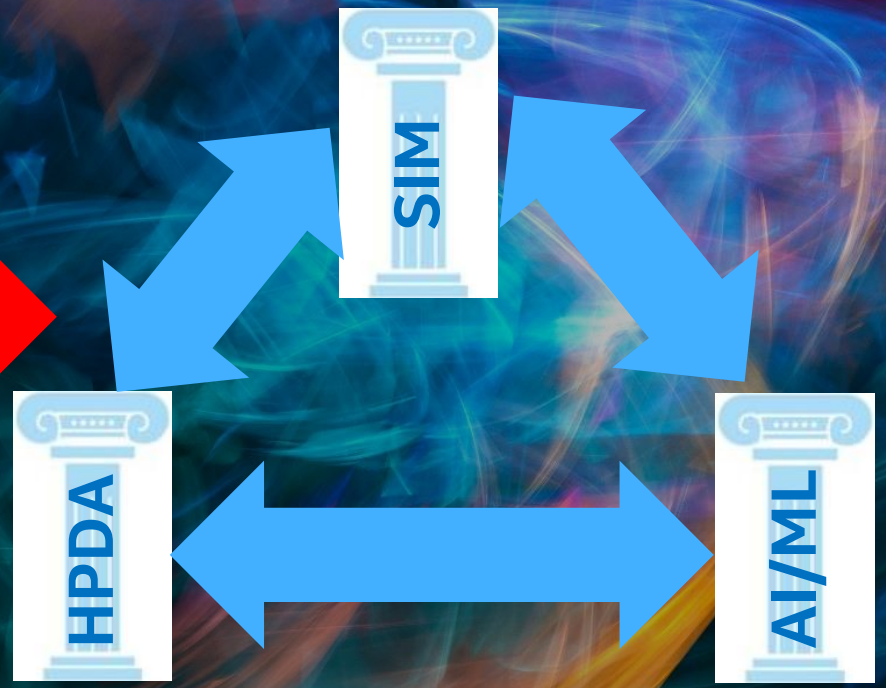


DAOS, NVM,

New Architecture

Tomorrow

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

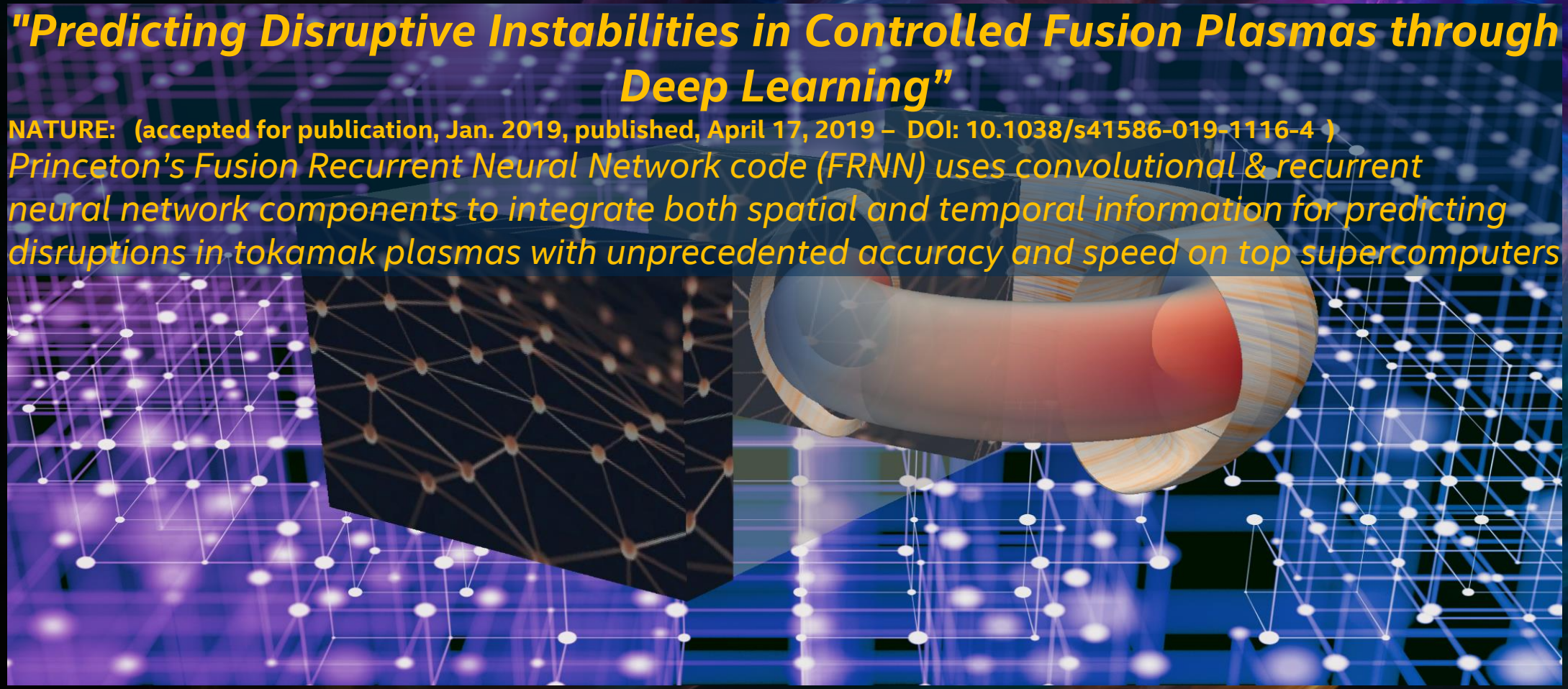


Deep learning and using surrogates bring exciting new technology to accelerate progress

"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



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.**

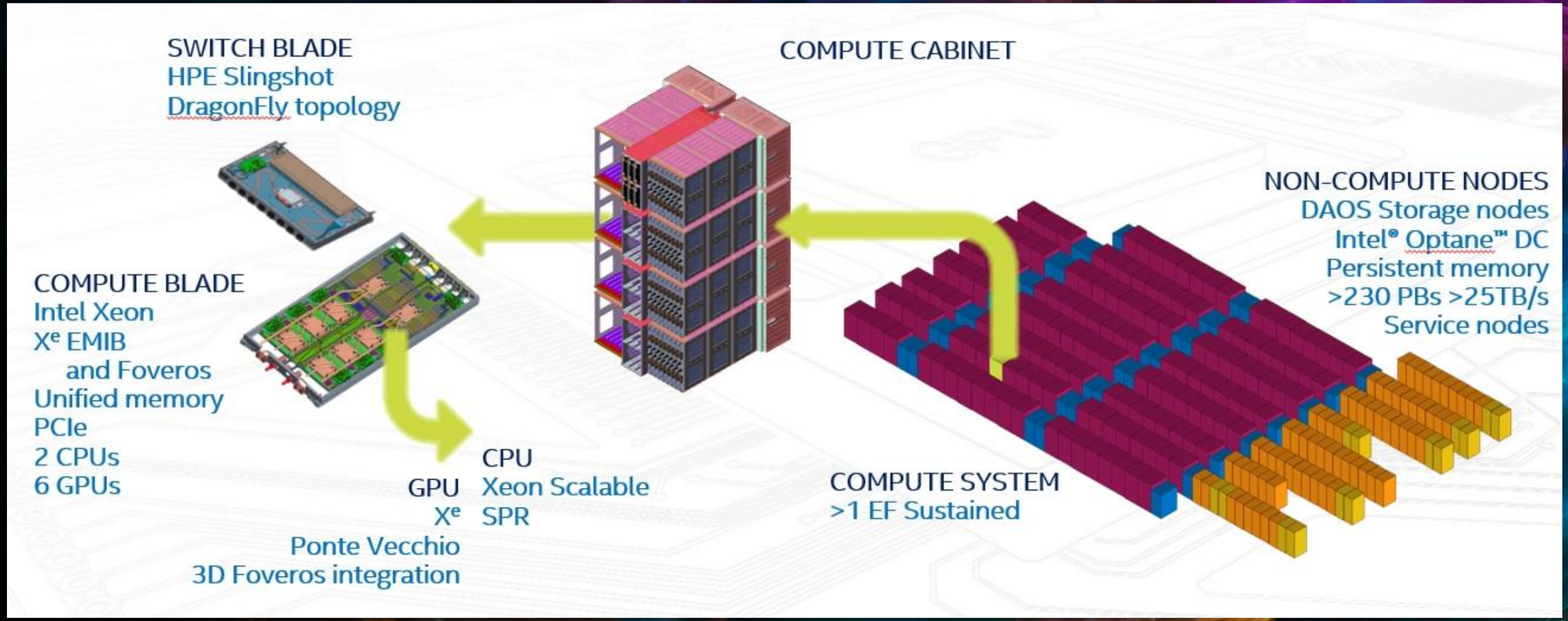
10,000 square feet
The amount of floor space for Aurora, which **equals to 4 tennis courts.**

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

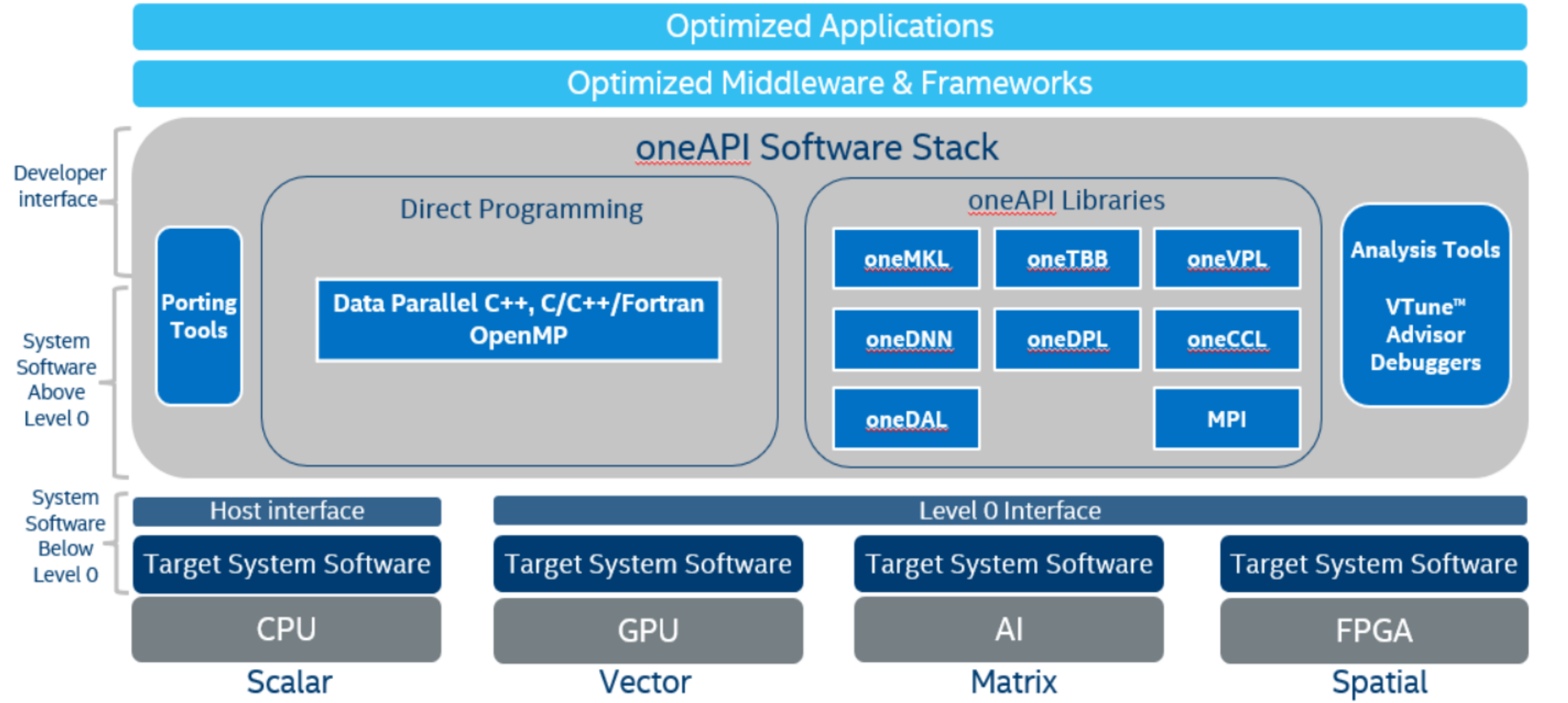
34,000 gallons per minute
The rate of water moving through the **cooling loop.**

Data Source: https://www.exascaleproject.org/wp-content/uploads/2020/10/Exascale-day-Aurora-Exascale-System.pdf

Aurora System Architecture

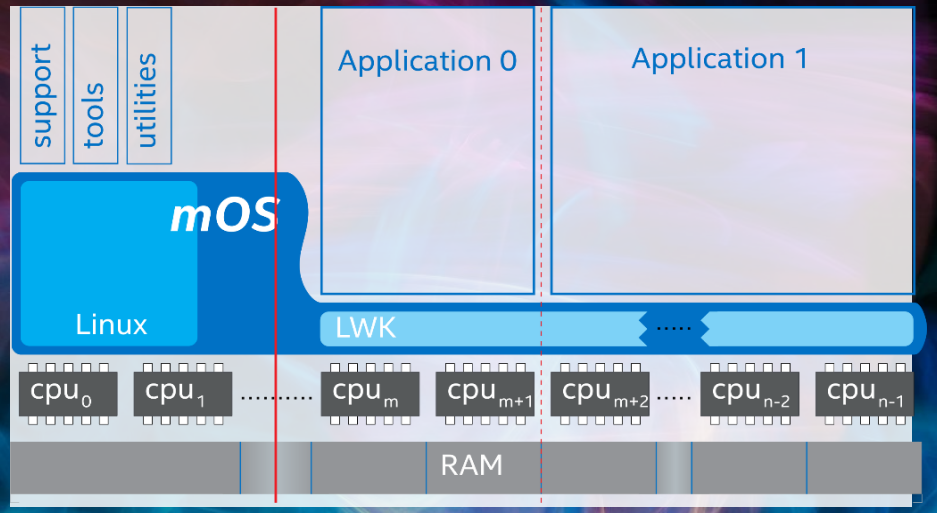


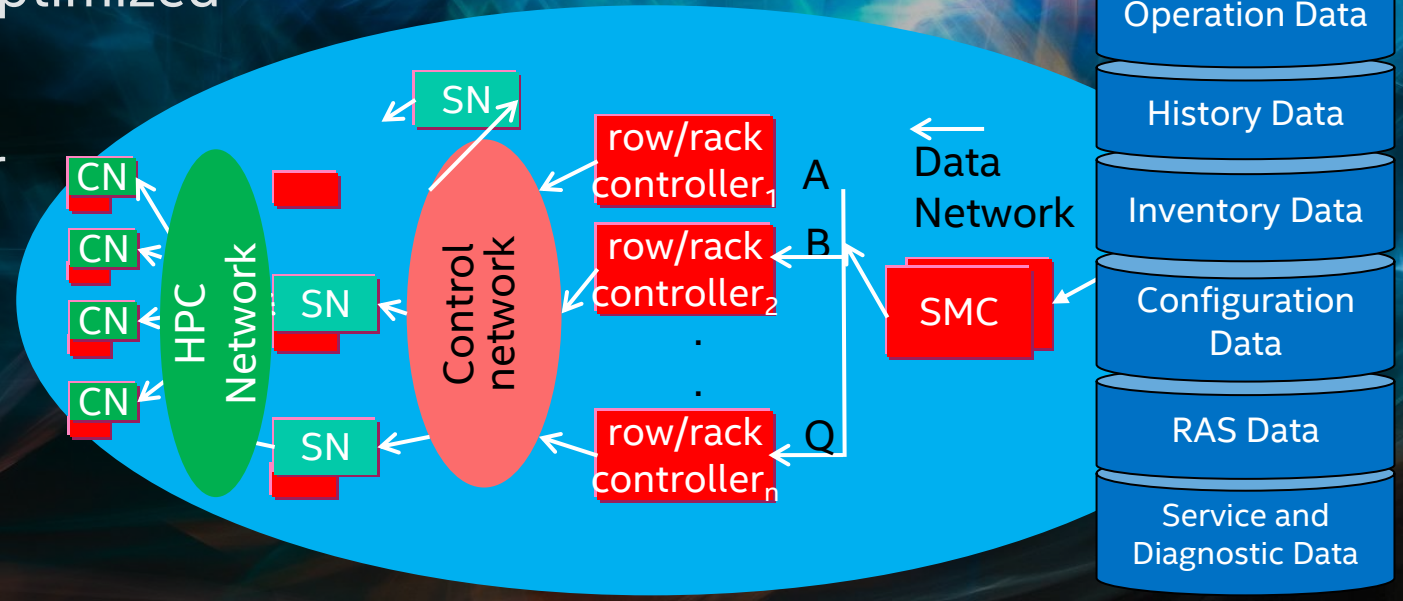
oneAPI Software Stack for HPC and AI Applications



Core Software HPC Components

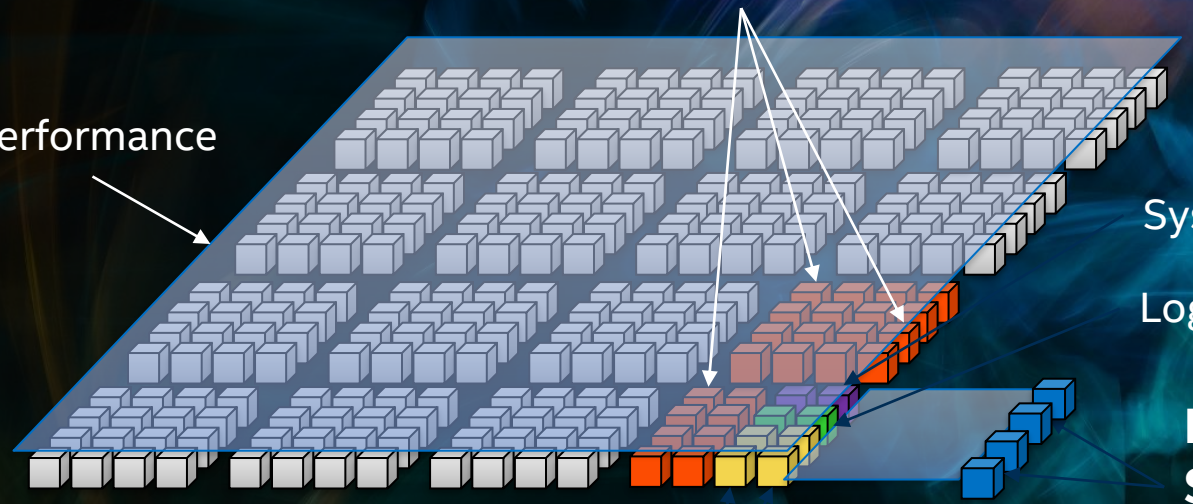
- **oneAPI**
  - **Developer environment program once run everywhere**
- **mOS**
  - Scalable operating system
- **Unified Control System**
  - Unified, Productive (single pane of glass), Reliable
- **MPI**
  - Scalable, high performance, topology optimized
- **GEOPM**
  - Global Extensible Open Power 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



High-Performance Fabric

System Service Nodes

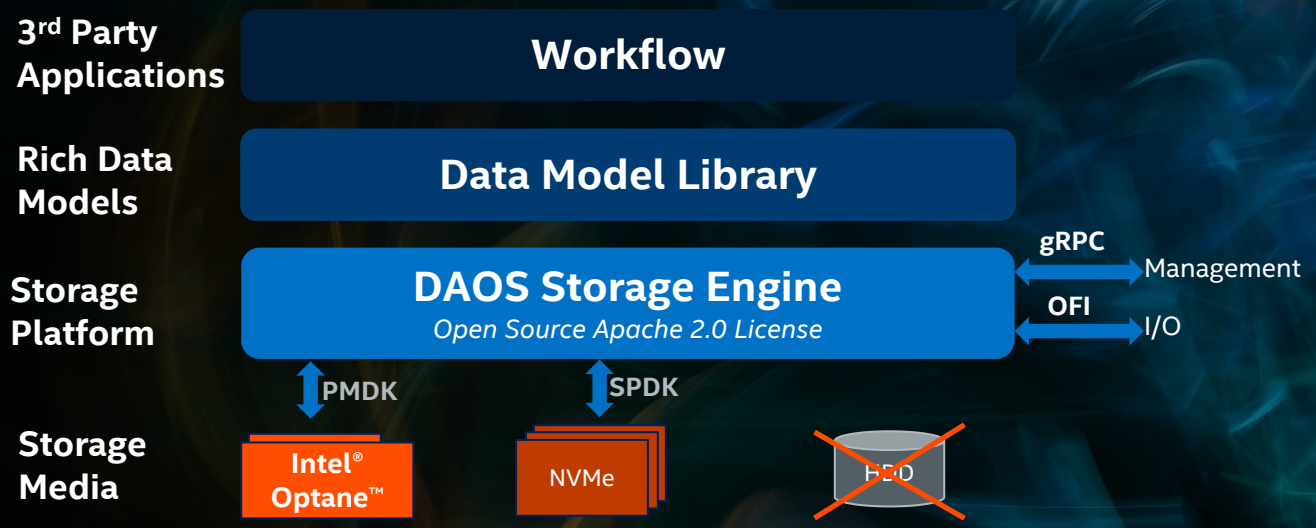
Login Nodes

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

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

DAOS: Distributed Asynchronous Object Storage

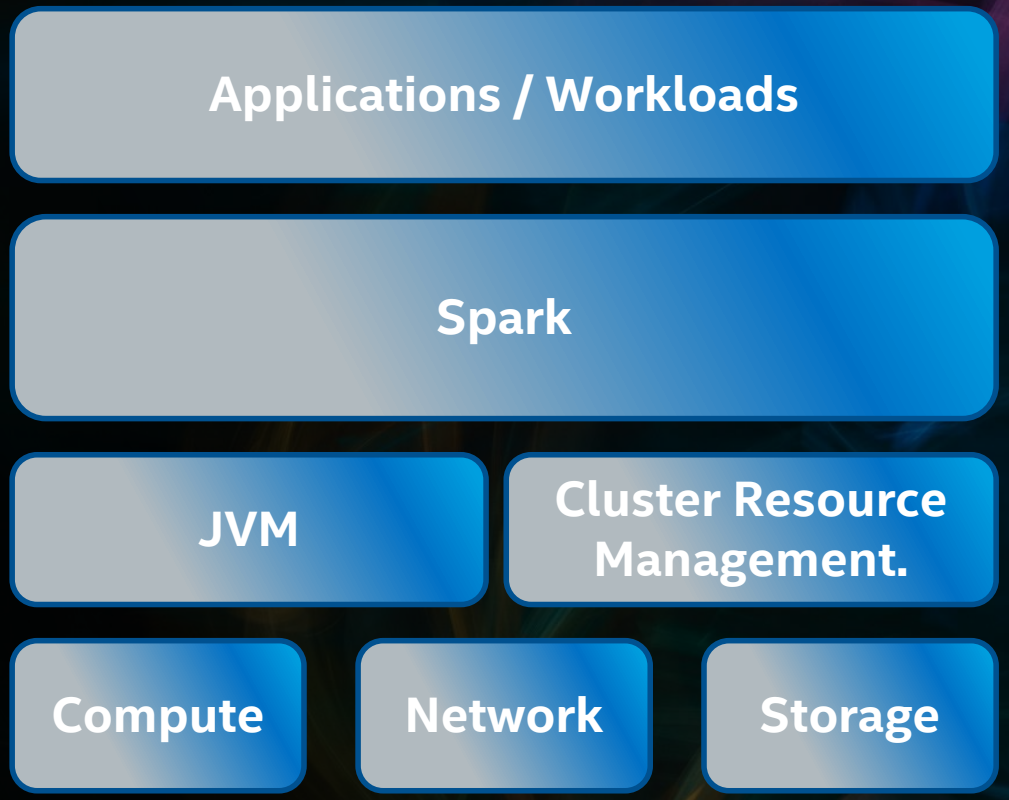
Scale-out object store built from the ground up for massively distributed NVM storage



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

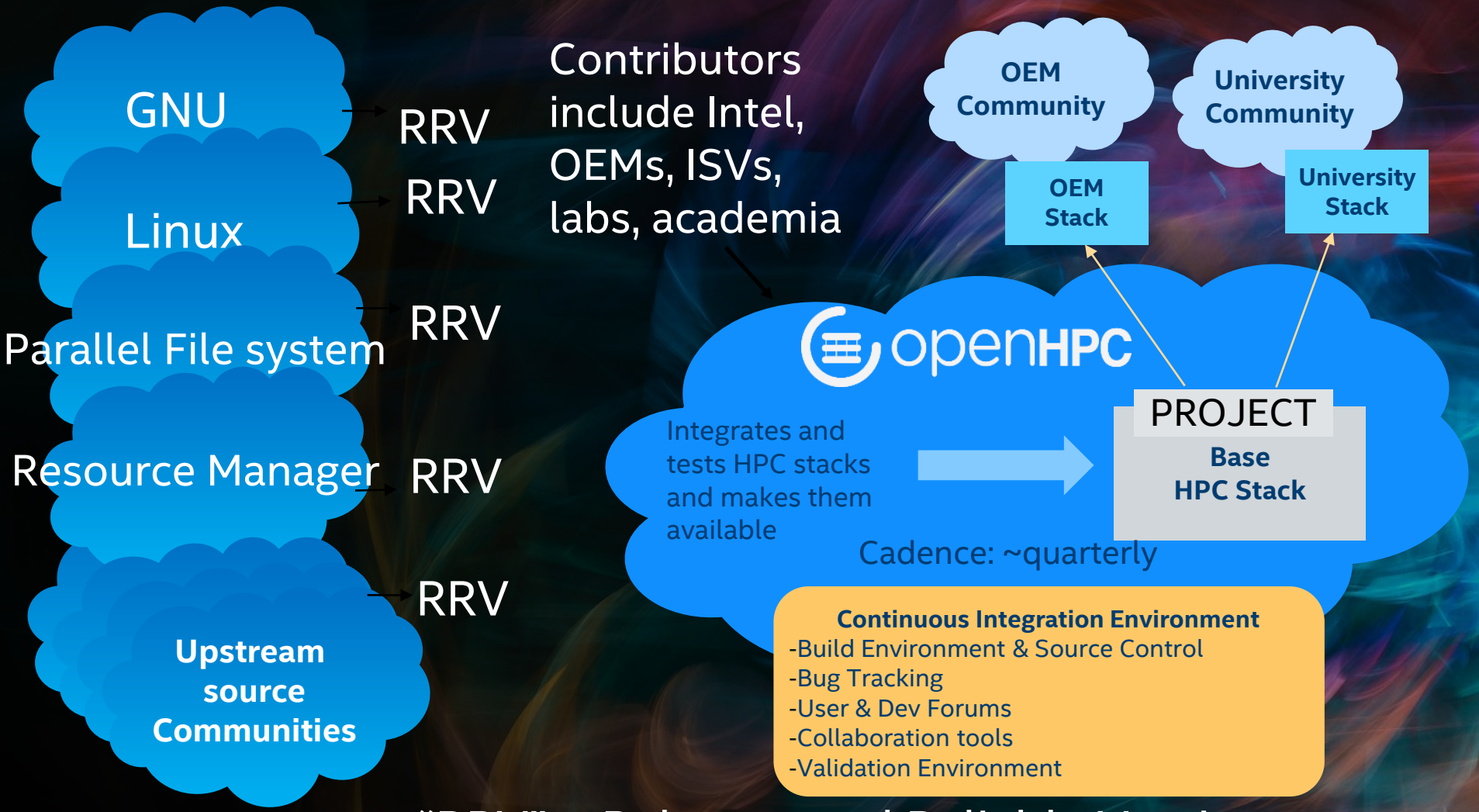
Bringing Spark Analytics to Exascale



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

OpenHPC



- 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

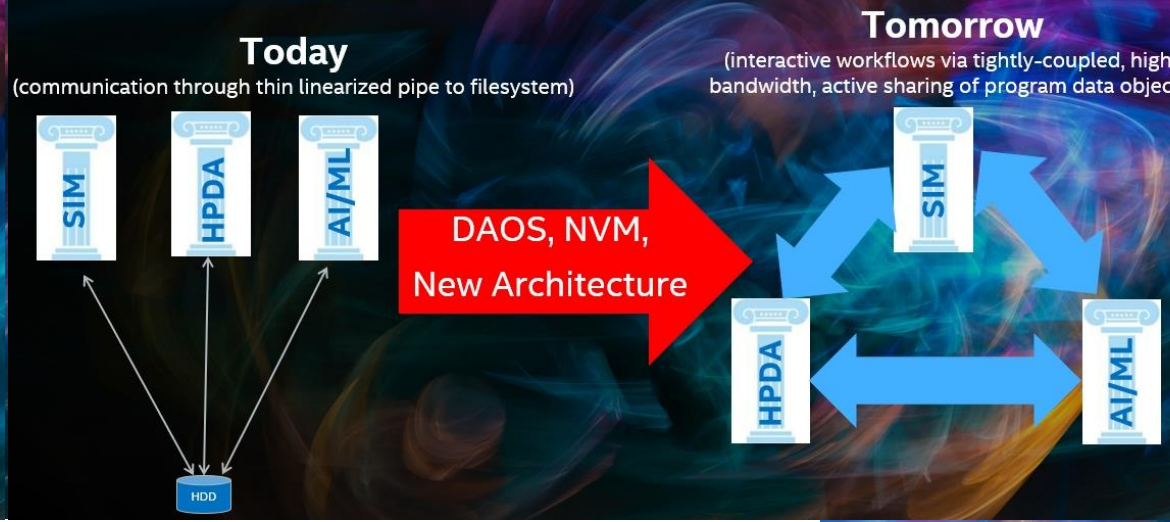
"RRV" = Relevant and Reliable Version

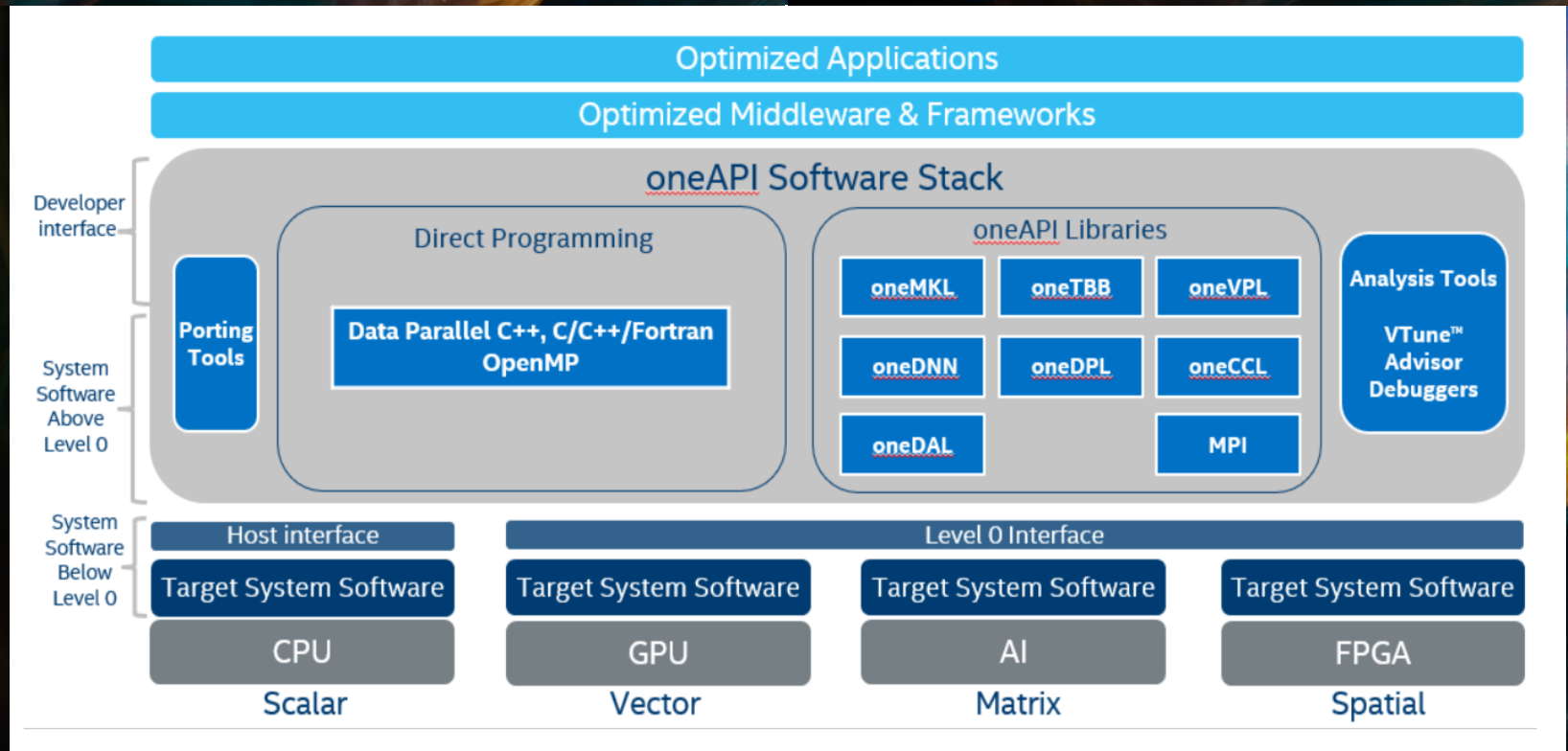
intel

Key Trends Affecting HPC

- Heterogeneity is all around us
  - Compute: Scale, Vector, Spatial, Matrix (SVSM)
  - Software
    - Classical HPC stack, AI/ML frameworks, Big Data
  - Memory, I/O
- New types of compute requirements
  - AI, Cloud, Big Data, Edge
  - AI and Cloud are large markets and key drivers of requirements
- HPC is more complex than ever and fundamental shifts are occurring

Converged workloads benefit from a tightly-coupled "Data Centric" architecture



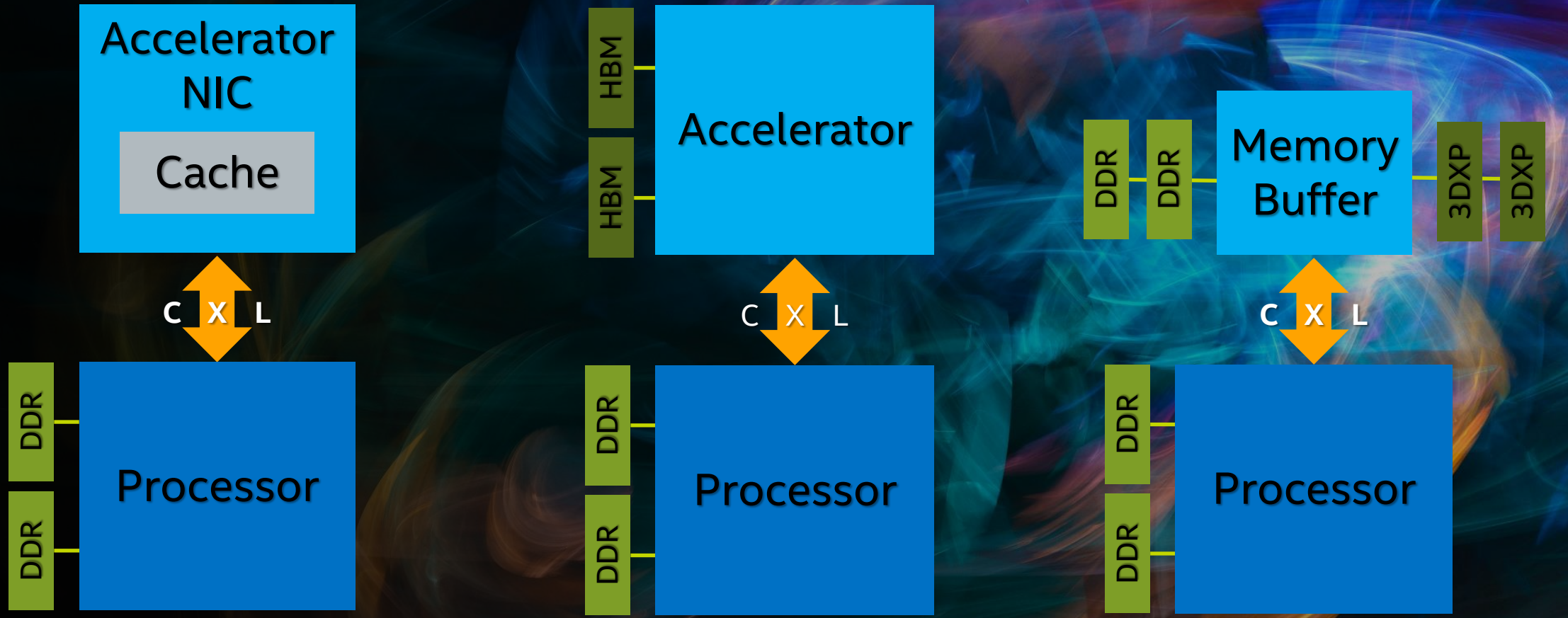


Addressing Challenges Raised by Trends

- 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

Tight Coupled Components

- Using EMIB and Foveros allows tighter coupling of components
- CXL technology allows for a productive shared memory software environment



Thank You

intel®