OAK RIDGE National Laboratory

Summit Architecture Overview

Tom Papatheodore

Oak Ridge National Laboratory

Introduction to Summit

Oak Ridge, TN

May 20, 2019

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

ORNL Summit System Overview

System Performance

- Peak of 200 Petaflops (FP_{64}) for modeling & simulation
- Peak of 3.3 ExaOps (FP_{16}) for data analytics and artificial intelligence

The system includes

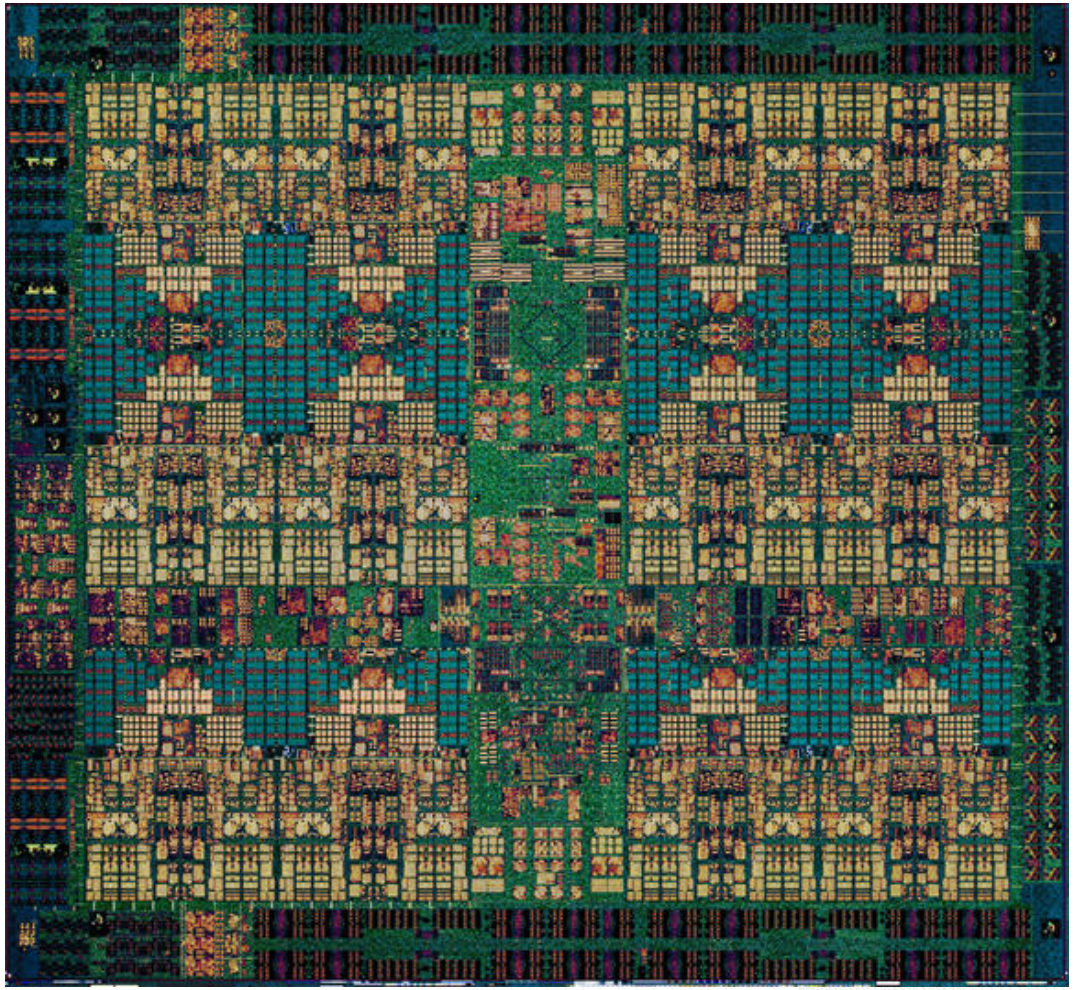
- 4,608 nodes
- Dual-port Mellanox EDR InfiniBand network
- 250 PB IBM file system transferring data at 2.5 TB/s

Each node has

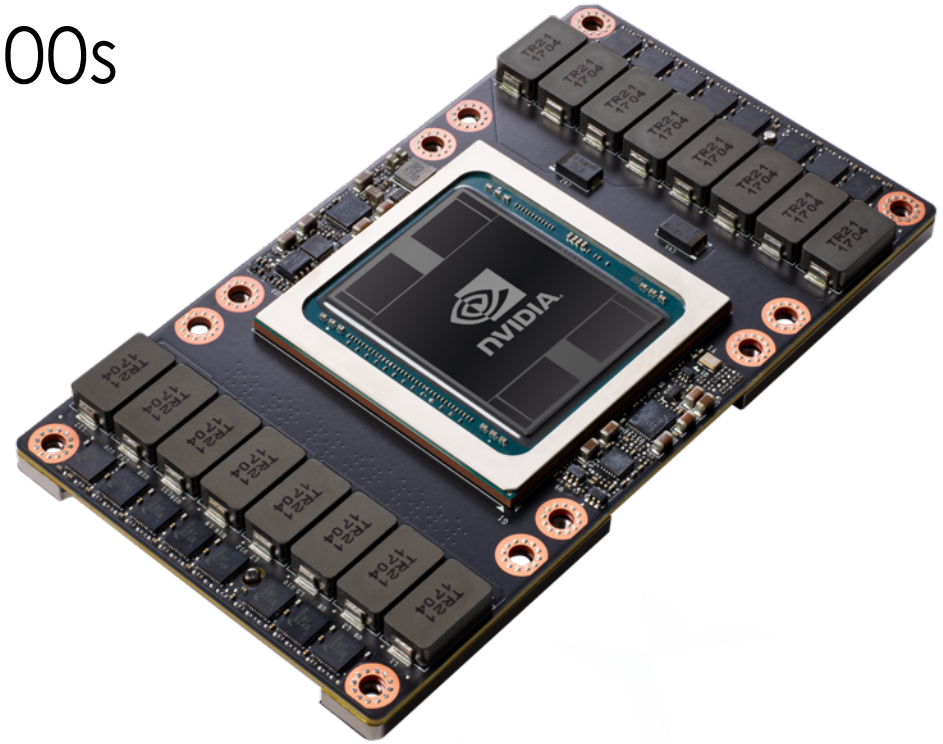
- 2 IBM POWER9 processors
- 6 NVIDIA Tesla V100 GPUs
- 608 GB of fast memory (96 GB HBM2 + 512 GB DDR4)
- 1.6 TB of NV memory

IBM Power9 Processor

- Summit's P9s: 22 cores (4 hwthreads/core)
- PCI-Express 4.0
  - Twice as fast as PCIe 3.0
- NVLink 2.0
  - Coherent, high-bandwidth links to GPUs
- 14nm FinFET SOI technology
  - 8 billion transistors
- Cache
  - L1I: 32 KiB (per core, 8-way set associative)
  - L1D: 32 KiB (per core, 8-way)
  - L2: 512 KiB (per pair of cores)
  - L3: 120 MiB eDRAM, 20-way (shared by all cores)



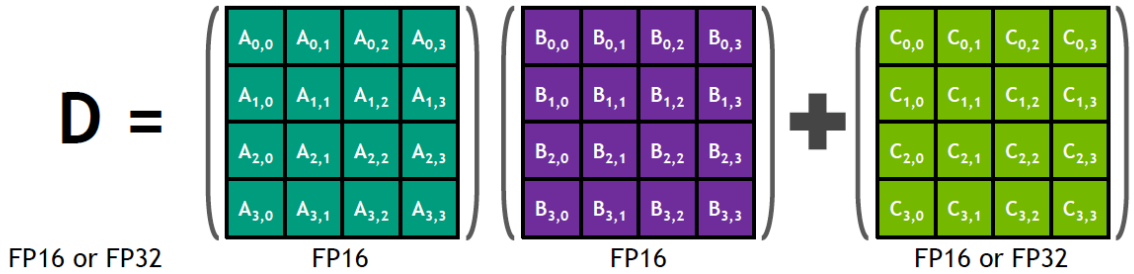
Summit Contains 27,648 NVIDIA Tesla v100s



Each Tesla v100 GPU has:

- 150+150 GB/s total BW (NVLink v2.0)
- 5,120 CUDA cores (64 on each of 80 SMs)
- 640 Tensor cores (8 on each of 80 SMs)
- 20MB Registers | 16MB Cache | 16GB HBM2 @ 900 GB/s
- 7.5 DP TFLOPS | 15 SP TFLOPS | 120 FP_{16} TOPS

- Tensor cores do mixed precision multiply-add of 4x4



$$D = AB + C$$

Type	Size	Range	$u = 2^{-t}$
half	16 bits	$10^{\pm 5}$	$2^{-11} \approx 4.9 \times 10^{-4}$
single	32 bits	$10^{\pm 38}$	$2^{-24} \approx 6.0 \times 10^{-8}$
double	64 bits	$10^{\pm 308}$	$2^{-53} \approx 1.1 \times 10^{-16}$
quadruple	128 bits	$10^{\pm 4932}$	$2^{-113} \approx 9.6 \times 10^{-35}$

In 2018 Summit Demonstrated Its Balanced Design

Achieves #1 on TOP500, #1 on HPCG, #1 Green500, and #1 on I/O 500

144 PF HPL
#1 raw performance

2.9 PF HPCG
#1 fast data movement

14.668 GF/W
#1 energy efficiency

#1 HPC storage performance

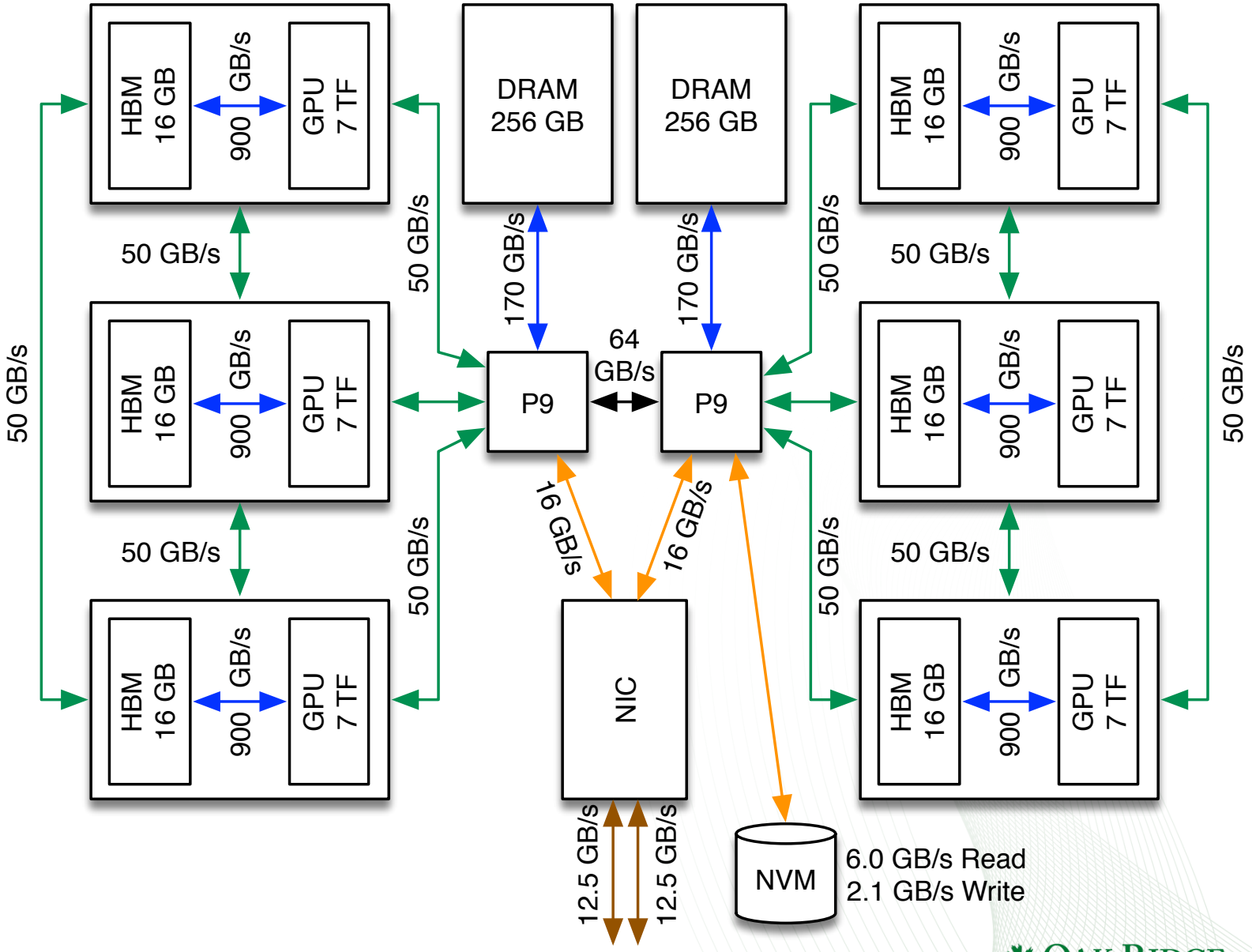
How is Summit Architecture different from Titan? ORNL's leadership supercomputer

- Many fewer nodes
- Much more powerful nodes
- Much more memory per node and higher memory bandwidth
- Much higher bandwidth between CPUs and GPUs
- Faster interconnect
- Much larger and faster file system
- 7x more performance for only slightly more power (Summit's 9.8 MW vs Titan's 8.2)

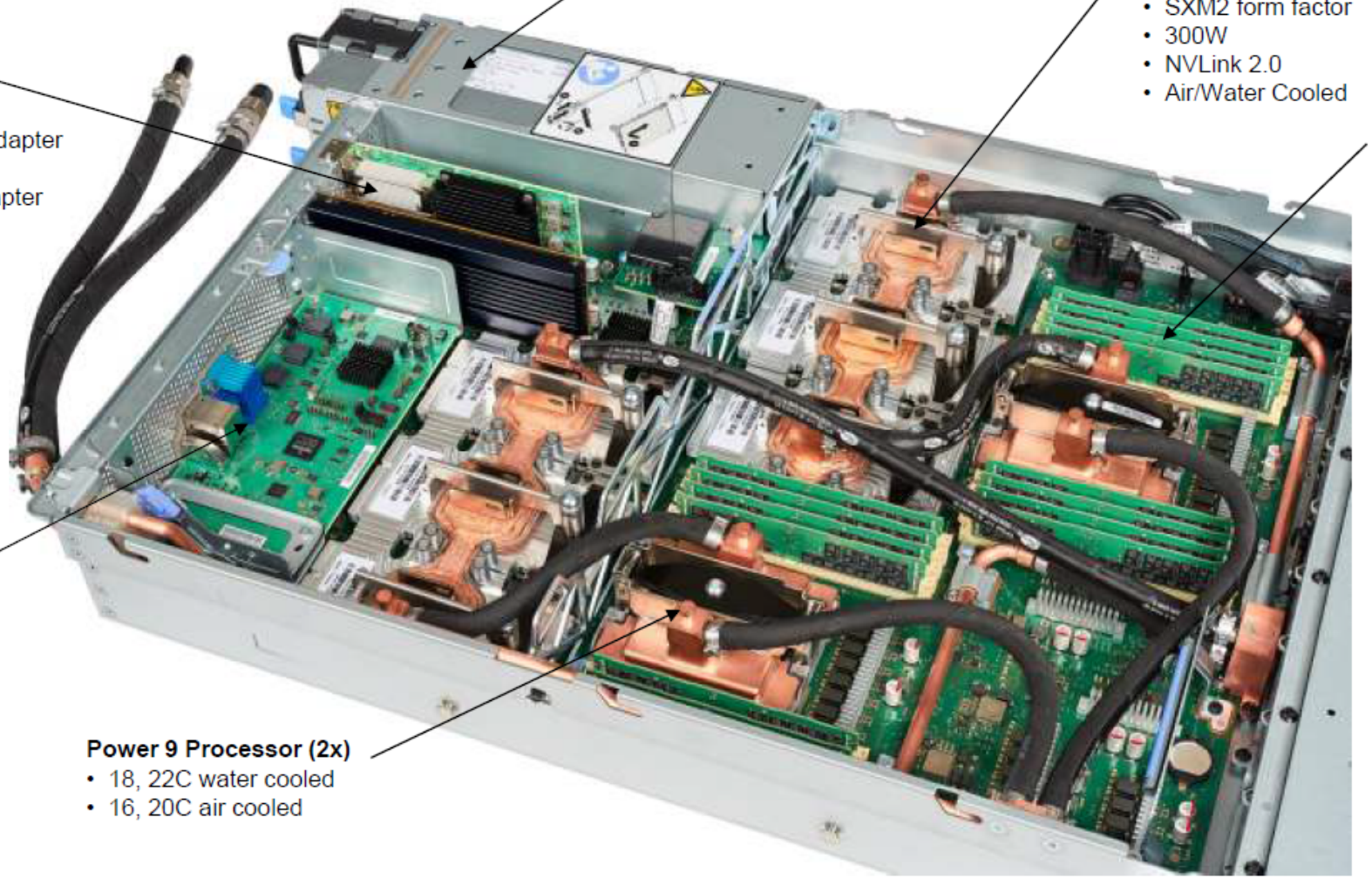
Feature	Titan	Summit
Peak FLOPS	27 PF	200 PF
Max possible Power	9 MW	13 MW
Number of Nodes	18,688	4,608
Node performance	1.4 TF	42 TF
Memory per Node	32 GB DDR3 + 6 GB GDDR5	512 GB DDR4 + 96 GB HBM2
NV memory per Node	0	1.6 TB
Total System Memory	0.7 PB	2.8 PB + 7.4 PB NVM
System Interconnect	Gemini (6.4 GB/s)	Dual Port EDR-IB (25 GB/s)
Interconnect Topology	3D Torus	Non-blocking Fat Tree
Bi-Section Bandwidth	15.6 TB/s	115.2 TB/s
Processors on node	1 AMD Opteron™ 1 NVIDIA Kepler™	2 IBM POWER9™ 6 NVIDIA Volta™
File System	32 PB, 1 TB/s, Lustre®	250 PB, 2.5 TB/s, GPFS™

Summit Node Schematic

- Coherent memory across entire node
- NVLink v2 fully interconnects three GPUs and one CPU on each side of node
- PCIe Gen 4 connects NVM and NIC
- Single shared NIC with dual EDR ports



Summit Board (1 node) showing the Water Cooling



Power Supplies (2x)
- 2200W
- 200VAC, 277VAC, 400VDC input

NVidia Volta GPU
- 3 per socket
- SXM2 form factor
- 300W
- NVLink 2.0
- Air/Water Cooled

Memory DIMM's (16x)
- 8 DDR4 IS DIMMs per socket
- 8, 16, 32,64, 128GB DIMMs

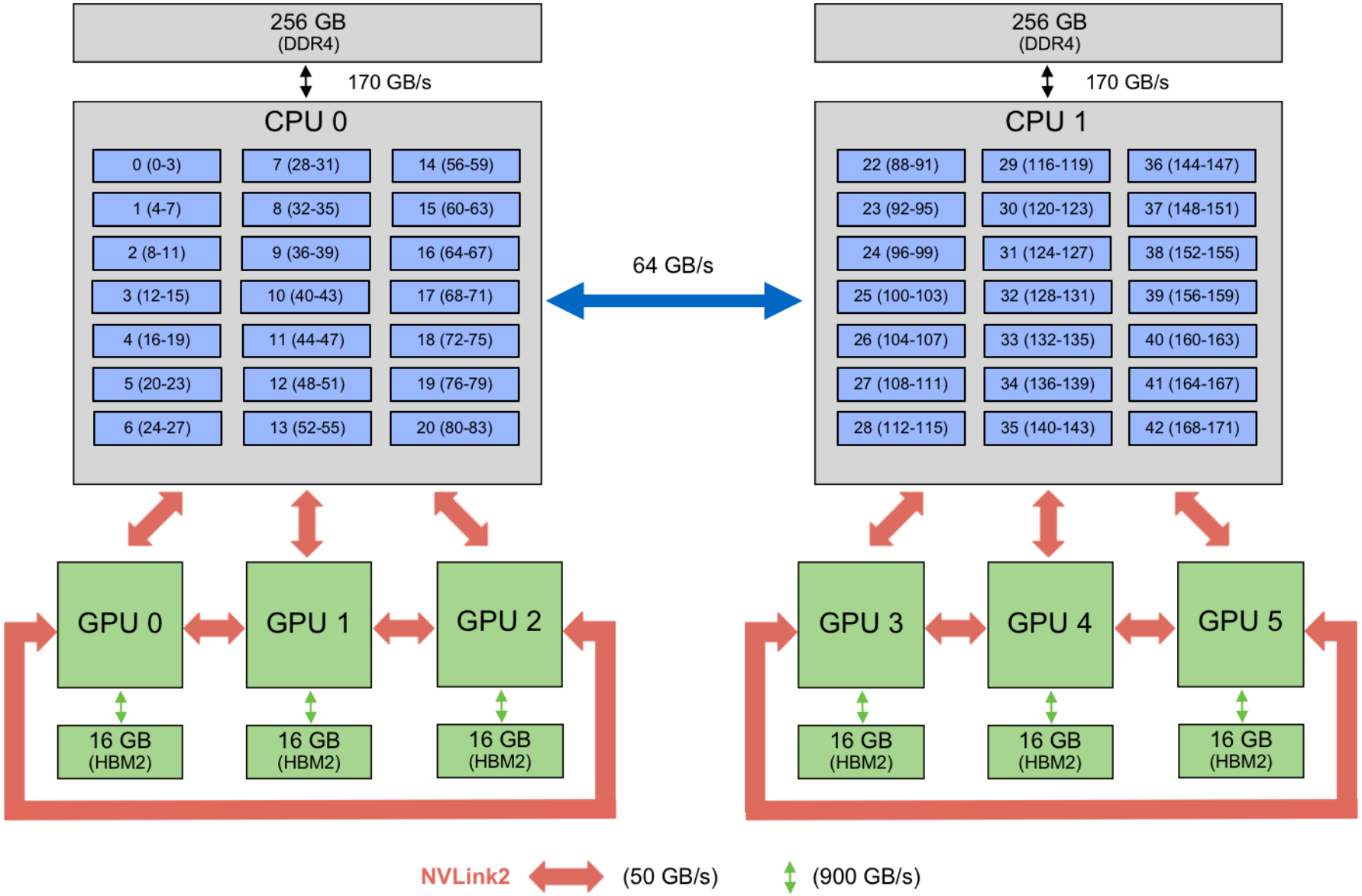
PCIe slot (4x)
- Gen4 PCIe
- 2, x16 HHHL Adapter
- 1, Shared slot
- 1 x8 HHHL Adapter

BMC Card
- IPMI
- 1 Gb Ethernet
- VGA
- 1 USB 3.0

Power 9 Processor (2x)
- 18, 22C water cooled
- 16, 20C air cooled

Summit Node

(2) IBM Power9 + (6) NVIDIA Volta V100



OAK RIDGE National Laboratory

Questions?