OAK RIDGE National Laboratory

Summit Updates

Presented to
2021 OLCF User Meeting

Don Maxwell
HPC Scalable Systems Group Leader

June 22, 2021

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

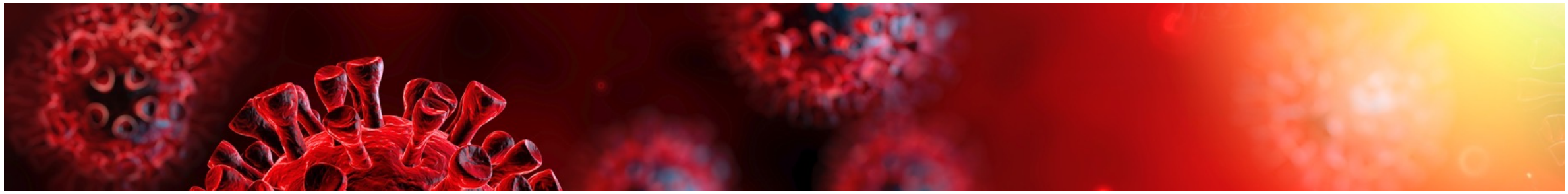
U.S. DEPARTMENT OF ENERGY

Summit IBM AC922

Specifications and Features

- Processor: IBM Power9™ (2/node)
- GPUs: ~~27,648~~ 28,080 NVIDIA Volta V100s (6/node)
- Nodes: ~~4,608~~ 4,680
- Node Performance: 42TF
- Memory/node: 512GB DDR4 + 96GB HBM2
- NV Memory/node: 1.6TB
- Total System Memory: >10PB DDR4 + HBM + Non-volatile
- Interconnect Topology: Mellanox EDR 100G InfiniBand, Non-blocking Fat Tree
- Peak Power Consumption: 13MW

Summit COVID Hardware



- Add 54 nodes, each with double the HBM and 4x the DDR4 and NVMe
- Allows jobs that need larger on-node memory to run on up to 54 nodes
- Applications that benefit
  - Computational Chemistry
  - AI/DL for medical image analysis

Feature	Current Summit Nodes	Large Memory Nodes
Peak $FLOPS_{64}$	200 PF	203 PF
Number of Nodes	4,680	54
Node performance	43 TF	43 TF
Memory per Node	512 GB DDR4 + 96 GB HBM2	2048 GB DDR4 + 192 GB HBM2
NV memory per Node	1.6 TB	6.4 TB
Total System Memory	2.8 PB + 7.4 PB NVM	2.9 PB + 7.7 PB NVM
System Interconnect	Dual Rail EDR-IB (25 GB/s)	Dual Rail EDR-IB (25 GB/s)
Interconnect Topology	Non-blocking Fat Tree	Non-blocking Fat Tree
Bi-Section Bandwidth	115.2 TB/s	115.2 TB/s
Processors on node	2 IBM POWER9™ 6 NVIDIA Volta™	2 IBM POWER9™ 6 NVIDIA Volta™
File System	250 PB, 2.5 TB/s, GPFS™	250 PB, 2.5 TB/s, GPFS™

Summit Contingency Hardware

3 New Summit Cabinets

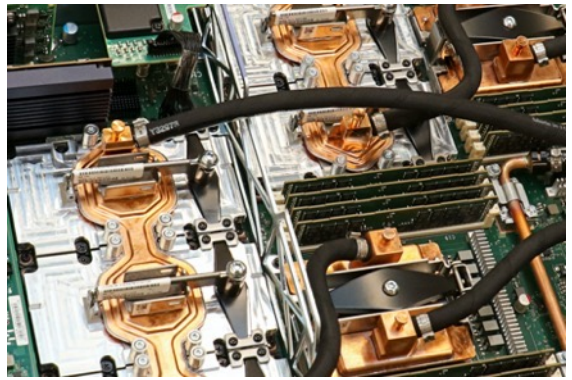
54 additional nodes added in March 2021

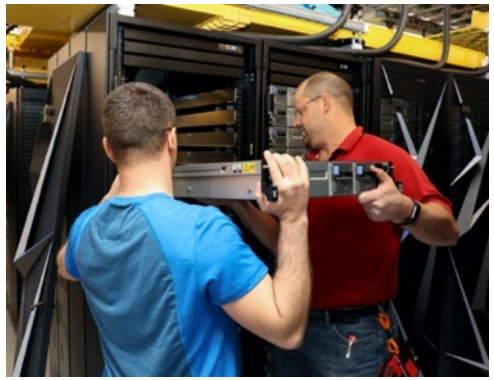
Summit Specs

Effectively same specifications as original Summit hardware

6th Year Spares

Would be used as spare parts if Summit extended into a 6th year





How do I access the new nodes?

- COVID nodes are a separate partition due to heterogeneity
  - Batch queue name has a trailing –hm (high memory)
    - batch-hm
  - All users have access

- Contingency nodes are included in regular queues
  - Simply increased the normal machine size providing the ability to submit jobs up to 4662 nodes

- Queues for longer walltimes for small jobs
  - COVID nodes (**batch-hm**) provides longer walltime flexibility for small jobs (i.e., a single-node job can run for 24 hours)
  - **killable** queue provides this functionality to regular Summit partition
    - Caveats
      - Preemptable with the following guarantees

Bin	Min Nodes	Max Nodes	Max Walltime (Hours)	Guaranteed Walltime
4	46	91	24.0	6.0 (hours)
5	1	45	24.0	2.0 (hours)

https://docs.olcf.ornl.gov/systems/summit_user_guide.html#killable-queue-policy

Summit Software

- Currently
  - RHEL 7.6
  - CUDA 10.1
  - MOFED 4.7
  - LSF 10.1.0.9
  - xCAT 2.15

- Summit IBM HPC Software Stack Upgrade
  - Software Updates
    - RHEL 8.2
    - CUDA 11
    - MOFED 4.9
    - LSF 10.1.0.11
    - xCAT 2.16
    - Updates to LAPACK, LAPACKE, and LAPACK GPU support in ESSL
    - GDRCopy support in Spectrum MPI
    - JSM
      - Improved GPU Binding options
      - Support for multiple SMPI installs with a single JSM

Questions?