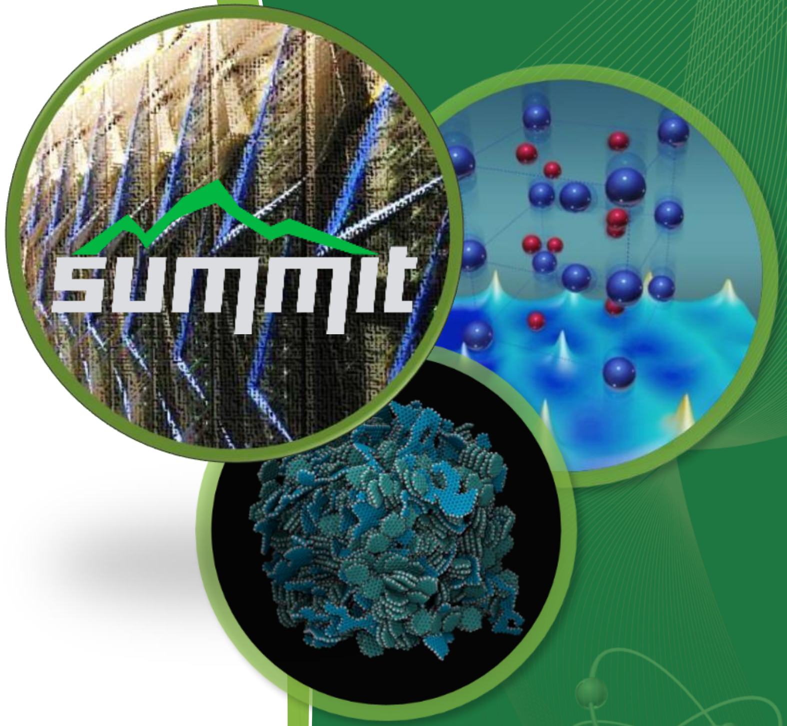
Summit Update

Buddy Bland

Oak Ridge National Laboratory

ASCAC Meeting

September 17-18, 2018



ORNL is managed by UT-Battelle
for the US 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.

OAK RIDGE
National Laboratory

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-rail 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 non-volatile memory

What makes Summit so powerful?

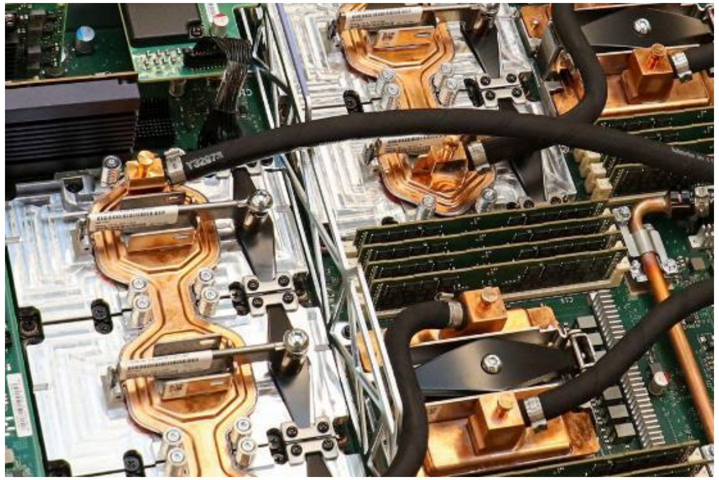
Summit's architecture is a sweet spot that has broad capability across:

- Traditional HPC modeling and simulation
- High performance data analytics
- Artificial Intelligence

- Powerful CPUs for scalar operations and data analytics
- Accelerators that address high-performance arithmetic in FP_{64}, FP_{32}, and FP_{16} precision
- Exceptional performance on artificial intelligence and machine learning
- High speed interconnect with switch based collective operations
- High-performance file system

Non-Recurring Engineering (NRE) has made a difference

- Design, packaging, and cooling of Summit nodes
  - 6 GPU and 2 CPU node connected with NVLink
  - Water cooled nodes for high density
- Switch based collectives and HW based tag matching improve latency in the interconnect
- Compilers
  - LLVM based, open-source compilers for Power+GPUs
  - OpenACC and CUDA Fortran for Power+GPUs



SHARP Scalable Hierarchical Aggregation and Reduction Protocol

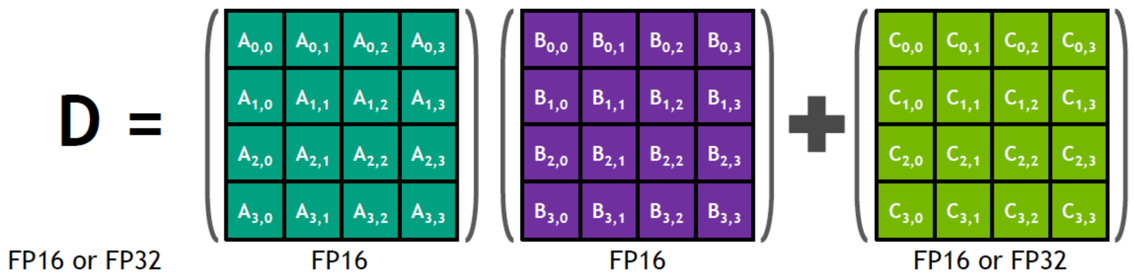
OAK RIDGE National Laboratory

Summit Contains 27,648 NVIDIA Tesla V100s

Each Tesla v100 GPU has:

- 300 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} TFLOPS

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

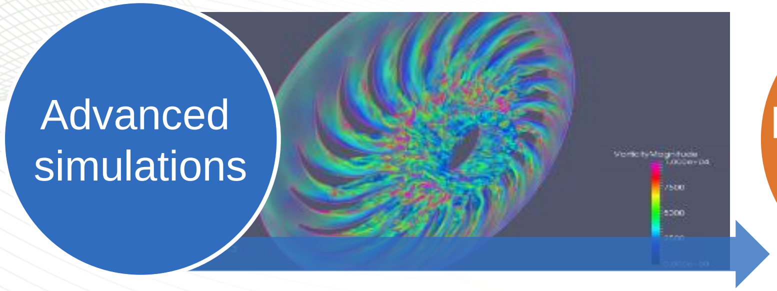


$$D = AB + C$$

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	$10^{\pm 4932}$	$2^{-113} \approx 9.6 \times 10^{-35}$

- The Modeling & Simulation community can benefit from better utilizing mixed / reduced precision
- Eg: Possible to achieve 4x FP64 peak for 64bit LU on V100 with iterative mixed precision (Dongarra et al.)

Summit Excels Across Simulation, Analytics, AI



- Data analytics – CoMet bioinformatics application for comparative genomics. Used to find sets of genes that are related to a trait or disease in a population. Exploits cuBLAS and Volta tensor cores to solve this problem 5 orders of magnitude faster than previous state-of-art code.
  - **Has achieved 2.36 ExaOps** mixed precision (FP_{16}-FP_{32}) on Summit
- Deep Learning – global climate simulations use a half-precision version of the DeepLabv3+ neural network to learn to detect extreme weather patterns in the output
  - **Has achieved a sustained throughput of 1.0 ExaOps (FP_{16})** on Summit
- Nonlinear dynamic low-order unstructured finite-element solver accelerated using mixed precision (FP_{16} thru FP_{64}) and AI generated preconditioner. Answer in FP_{64}
  - **Has achieved 25.3 fold speedup** on Japan earthquake – city structures simulation
- **Half-dozen Early Science codes are reporting >25x speedup on Summit vs. Titan**

Summit Displays Its Balanced Design
Achieves #1 on TOP500, #1 on HPCG, and #1 Green500 (level 3)

122 PF HPL
Shows DP performance

2.9 PF HPCG
Shows fast data movement

13.889 GF/W
Shows energy efficiency

OAK RIDGE National Laboratory

Five Gordon Bell Finalists Credit Summit Supercomputer

The finalists—representing Oak Ridge, Lawrence Berkeley, and Lawrence Livermore National Laboratories and the University of Tokyo—leveraged Summit's unprecedented computational capabilities to tackle a broad range of science challenges and produced innovations in machine learning, data science, and traditional modeling and simulation to maximize application performance. The Gordon Bell Prize winner will be announced at SC18 in Dallas in November. Finalists include:

- An ORNL team led by computational systems biologist Dan Jacobson and OLCF computational scientist Wayne Joubert that developed a genomics algorithm capable of using mixed-precision arithmetic to attain exascale speeds.
- A team from the University of Tokyo led by associate professor Tsuyoshi Ichimura that applied AI and mixed-precision arithmetic to accelerate the simulation of earthquake physics in urban environments.
- A Lawrence Berkeley National Laboratory-led collaboration that trained a deep neural network to identify extreme weather patterns from high-resolution climate simulations.
- An ORNL team led by data scientist Robert Patton that scaled a deep learning technique on Summit to produce intelligent software that can automatically identify materials' atomic-level information from electron microscopy data.
- A LBNL and Lawrence Livermore National Laboratory team led by physicists André Walker-Loud and Pavlos Vranas that developed improved algorithms to help scientists predict the lifetime of neutrons and answer fundamental questions about the universe.

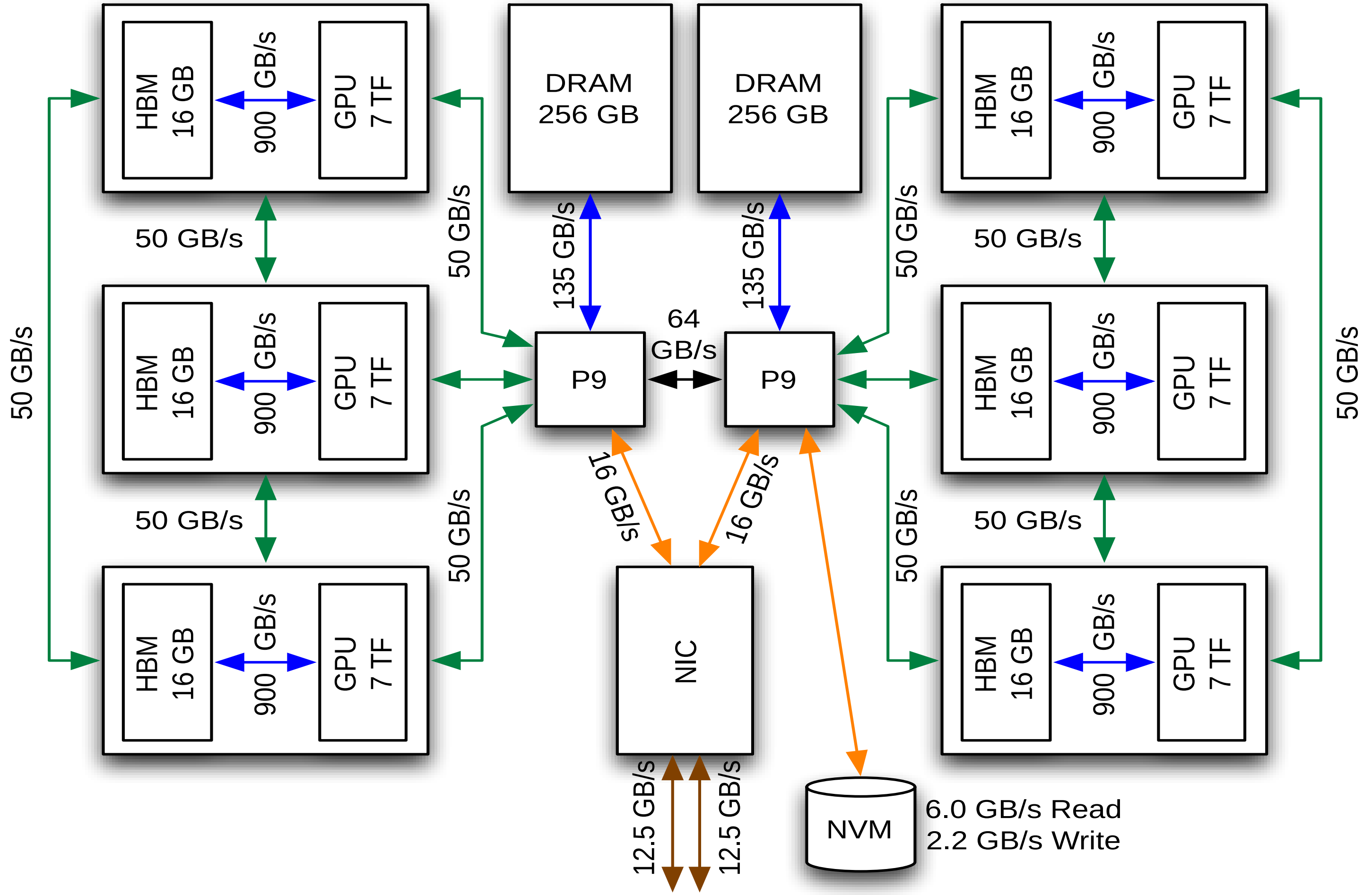
How is Summit Architecture different from Titan?

- Many fewer nodes
- Much more powerful nodes
- Much more memory per node and higher memory bandwidth
- Faster interconnect
- Much higher bandwidth between CPUs and GPUs
- Much larger and faster file system
- 7x more performance for only slightly more power
  (HPL 122 PF run was 8.8 MW)

Feature	Titan	Summit
Peak $FLOPS_{64}$	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 Rail 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



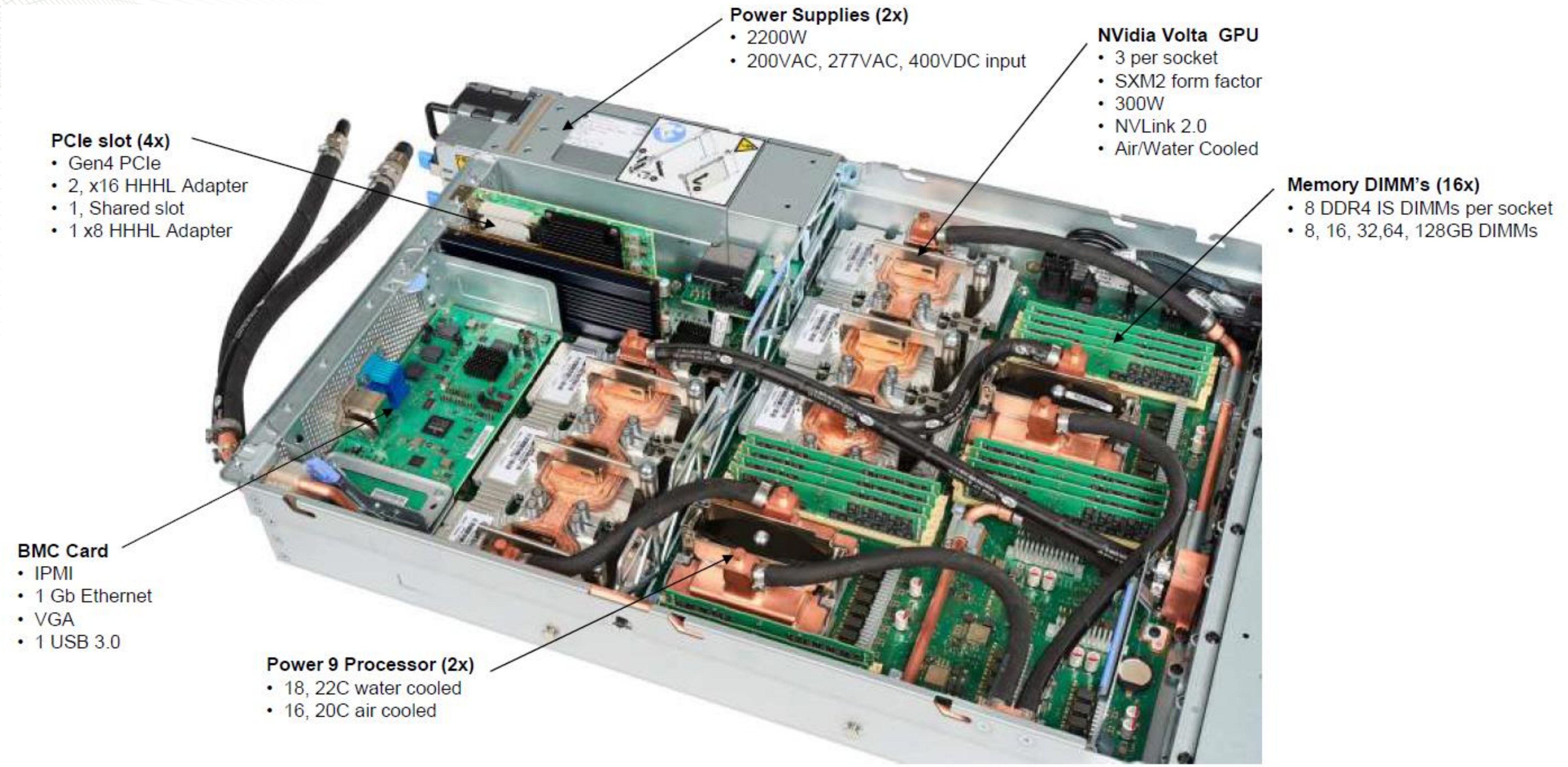
TF	42 TF (6x7 TF)
HBM	96 GB (6x16 GB)
DRAM	512 GB (2x16x16 GB)
NET	25 GB/s (2x12.5 GB/s)
MMsg/s	83

HBM/DRAM Bus (aggregate B/W)
NVLINK
X-Bus (SMP)
PCIe Gen4
EDR IB

HBM & DRAM speeds are aggregate (Read+Write).
All other speeds (X-Bus, NVLink, PCIe, IB) are bi-directional.

RIDGE Laboratory

Summit Board (1 node) showing the Water Cooling



Summit Status

- Hardware is stable and performs well
  - All nodes functional
  - InfiniBand hardware performs as expected
  - 30,000 disks are running
- But there are a few glitches
  - Disk cable management arm design was not strong enough to support all the cables. Will be replaced.
  - IBM has agreed to replace copper SAS cables with Fiber Optic cables. Should happen in November/December 2018.
  - Some early-life node failures which have been repaired or replaced
- We are working our way through the acceptance tests. We are finding and resolving bugs.

How to Get Time on Summit

U.S. DEPARTMENT OF ENERGY
INCITE
LEADERSHIP COMPUTING

ALCC

- The US Department of Energy provides access to leadership supercomputers to facilitate capability-limited research for significant advances in science and engineering through the INCITE and ALCC User Programs
- Access is awarded via annual competitive-proposal calls
  - Innovative and Novel Computational Impact on Theory and Experiment (INCITE) aims to accelerate scientific discoveries and technological innovations by **awarding, on a competitive basis**, time on supercomputers to researchers with large-scale, computationally intensive projects that address “grand challenges” in science and engineering (http://www.doeleadershipcomputing.org/incite-program).
  - ASCR Leadership Computing Challenge (ALCC) supports projects of interest to the Department of Energy (DOE) with an emphasis on high-risk, high-payoff simulations in areas directly related to the DOE mission and for broadening the community of researchers capable of using leadership computing resources (https://science.energy.gov/ascr/facilities/accessing-ascr-facilities/alcc/)
- Summit will be available to user programs starting in 2019

Questions?