OAK RIDGE National Laboratory

Summit Architecture Overview

Tom Papatheodore

Oak Ridge Leadership Computing Facility (OLCF)

SC19 Hands-On with Summit

Denver, CO

November 22, 2019

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

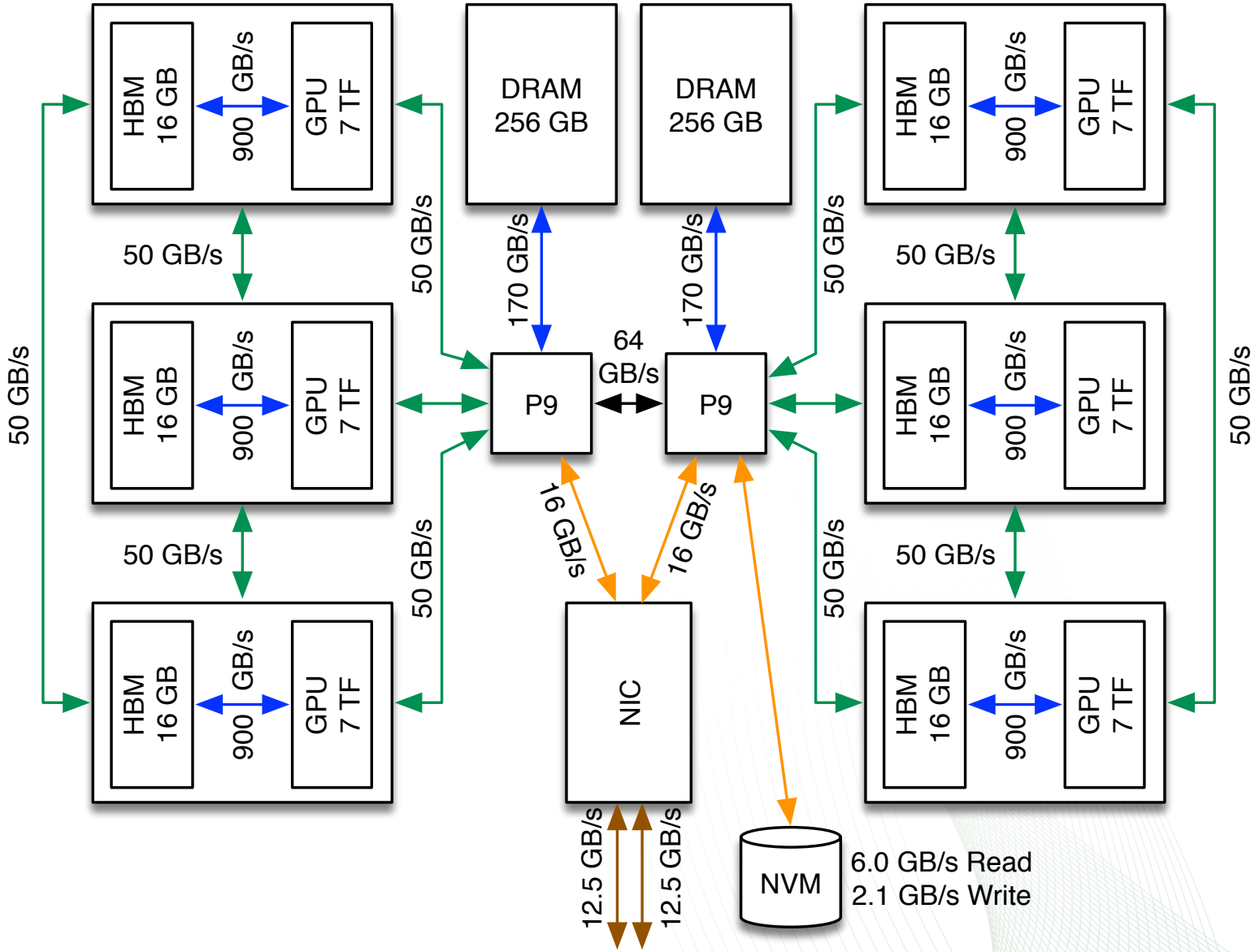
Each node has

- 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

149 PF HPL
#1 raw performance

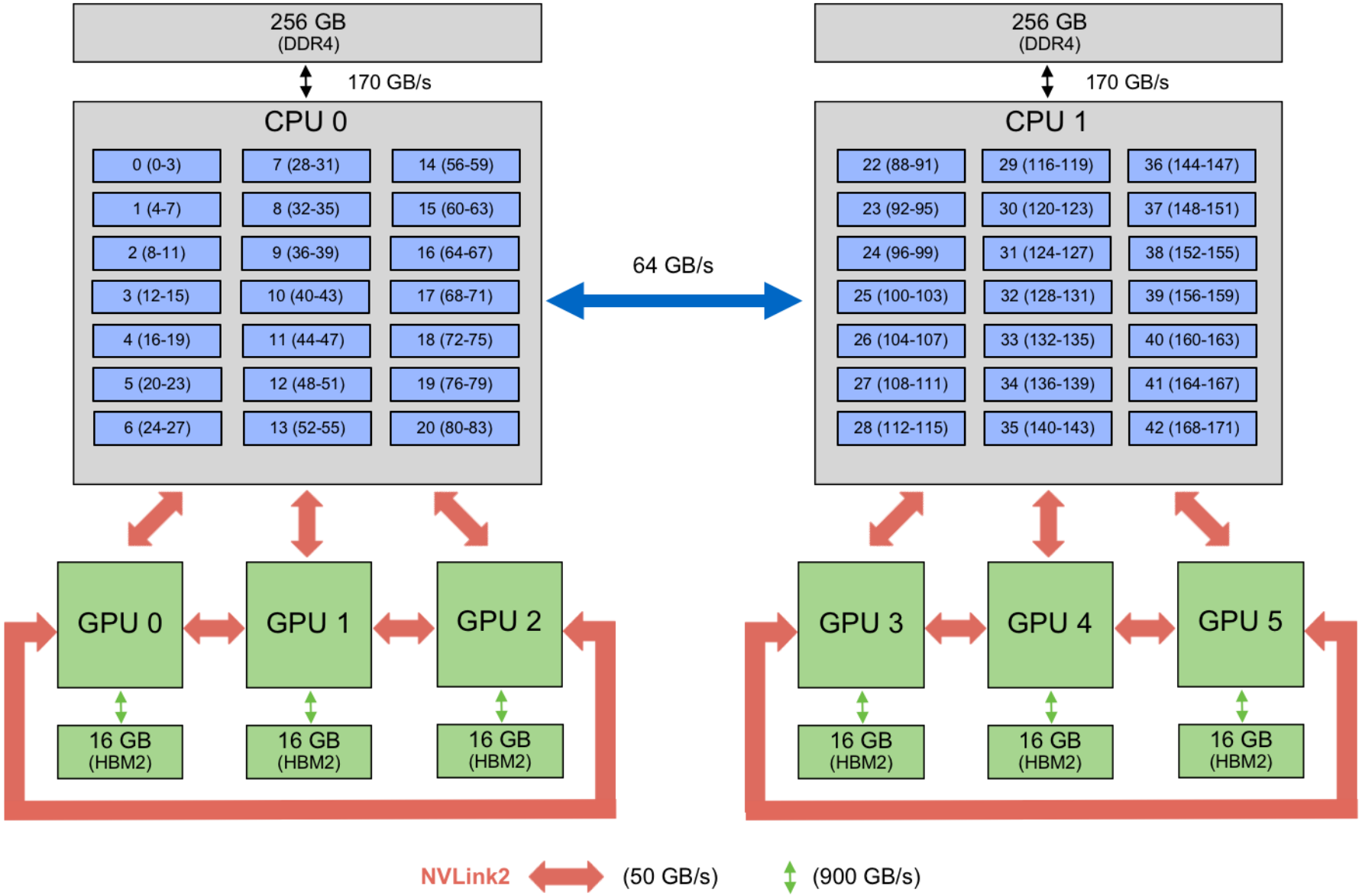
Summit Node Schematic

- 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
- 21 CPU cores on each P9 (each with 4 hw threads)



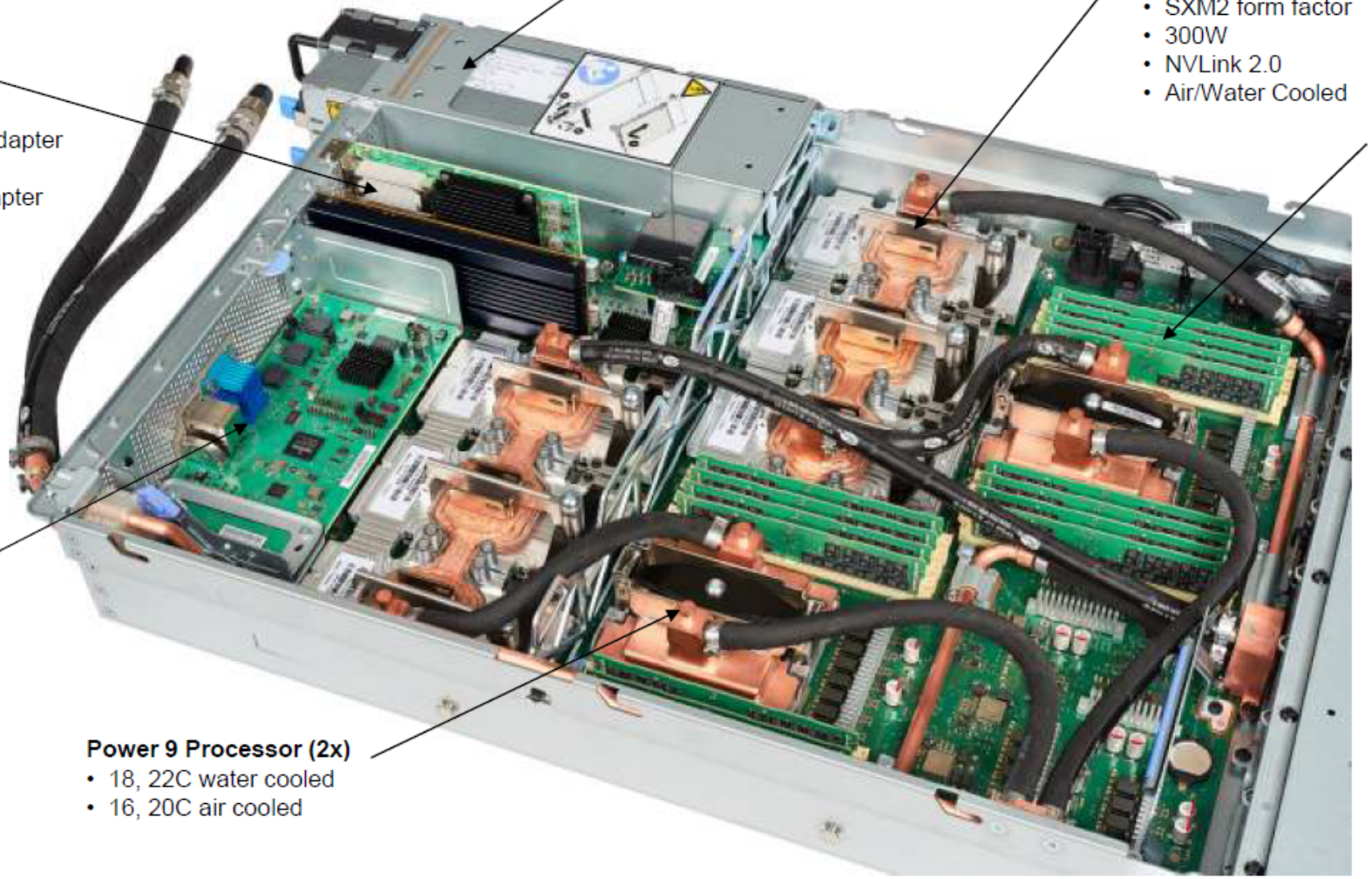
Summit Node

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



OAK RIDGE National Laboratory

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

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

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

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

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

Available File Systems / Storage Areas on Summit

NFS Directories – This is where you would keep source code and build your application.

NOTE: These directories are read-only from the compute nodes!

`/ccs/home/userid`

- Your personal home directory

`/ccs/proj/trn001`

- Can be accessed by all participants of this event
- This area is shared with all members of the project (source code, scripts, etc.)

GPFS Directories (parallel file system) – This is where you should write data when running on Summit's compute nodes.

`/gpfs/alpine/trn001/scratch/userid (or $MEMBERWORK/trn001)`

- Your personal GPFS scratch directory

`/gpfs/alpine/trn001/proj-shared (or $PROJWORK/trn001)`

- Can be accessed by all participants of the event
- This area is shared with all members of the project (data written from compute nodes)

Must use `jsrun` to run on compute nodes

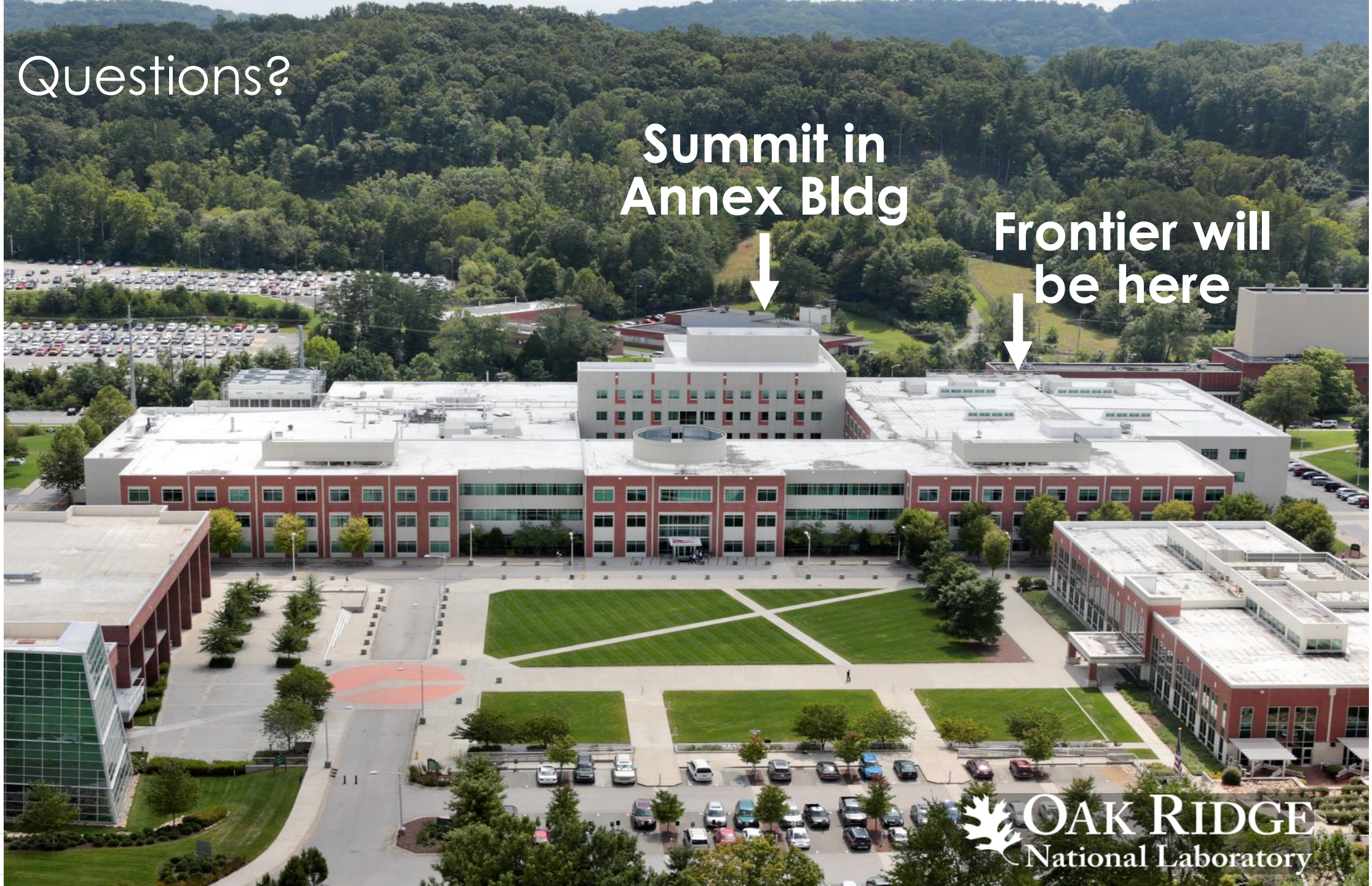
```
[tpapathe@login1: ~]$ hostname
login1

[tpapathe@login1: ~]$ bsub -P TRN001 -nnodes 1 -W 10 -Is /bin/bash
Job <15167> is submitted to default queue <batch>.
<<Waiting for dispatch ...>>
<<Starting on batch1>>

[tpapathe@batch1: ~]$ hostname
batch1

[tpapathe@batch1: ~]$ jsrun -n1 hostname
h49n16
```

Questions?


Summit in Annex Bldg
Frontier will be here
OAK RIDGE National Laboratory