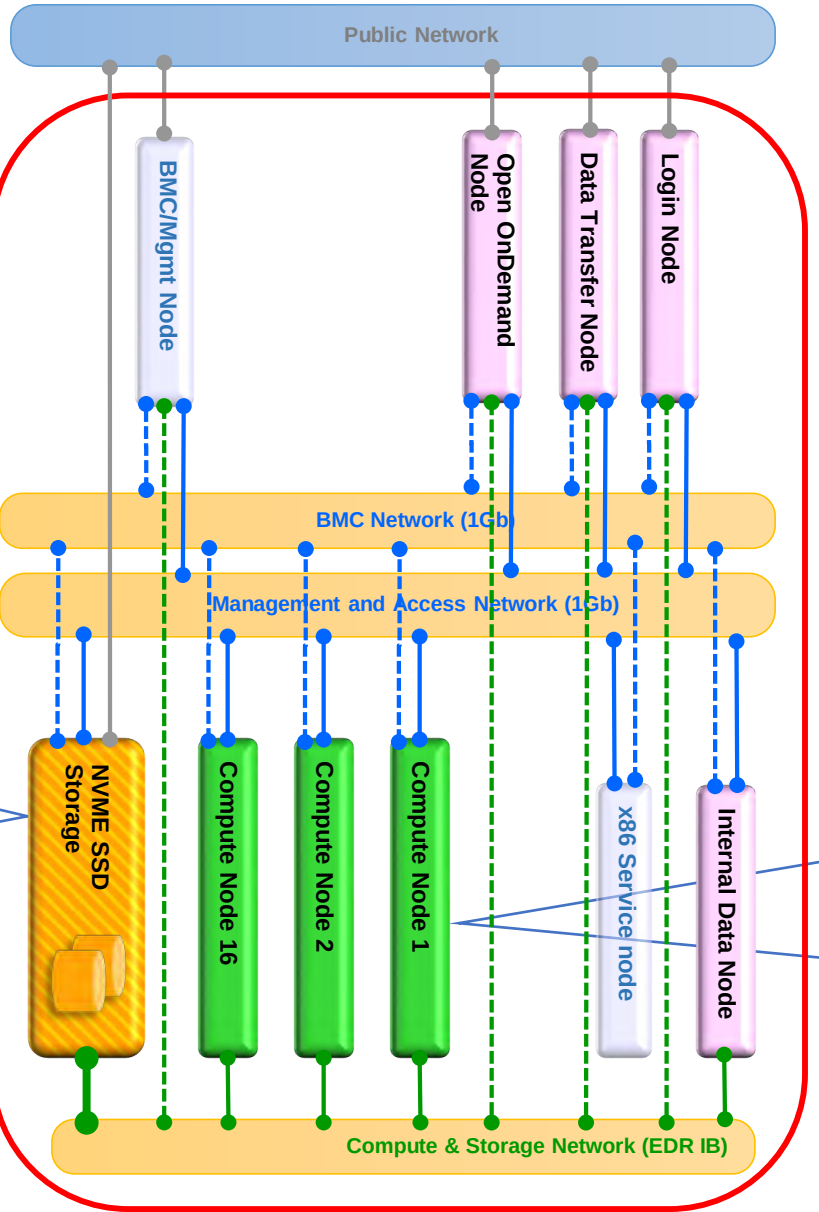
Computer System for Scalable Deep Learning

V. Kindratenko, **Dawei Mu**, Yan Zhan, J.D. Maloney, National Center for Supercomputing Applications, Urbana, IL, USA

Main software stack

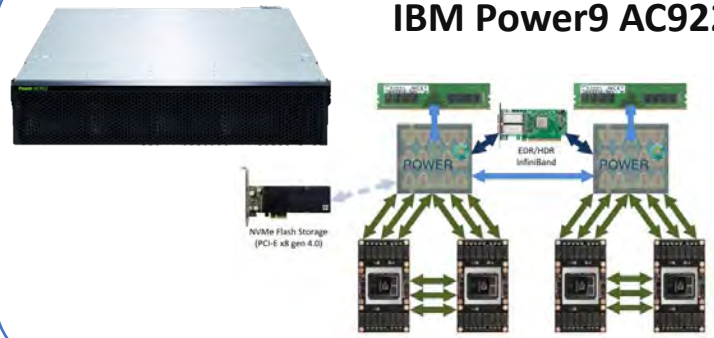
- RHEL 8.4
- CUDA 11.1.2, cuDNN 8.1.1, NCCL 2.8.3
- NVIDIA HPC-SDK 21.5
- IBM XLC, IBM XLFORTRAN
- Advance toolchain for Linux on Power
- WMLCE 1.7.0
- OpenCE 1.3.1
- SLURM & Open OnDemand



- 16 IBM AC922 nodes
  - IBM 8335-GTH AC922 server
    - 2x 20-core POWER9 CPU @ 2.4GHz
    - 256 GB DDR4
    - 4x NVIDIA V100 GPUs
      - 5120 cores
    - 16 GB HBM 2
- 2-Port EDR 100 GB IB ConnectX-5 Adapter
- DDN GS400NVE Flash Array
  - 360 TB usable NVME SSD-based storage
  - Spectrum Scale File System
- Login node
- Data Transfer Node
- OnDemand web interface node
  - VS Code
  - Jupyter Notebook
  - Jupyter Lab
  - TensorBoard
  - H2O-AI

DDN GS400NVE Flash Array

IBM Power9 AC922



Computer System for Scalable Deep Learning

V. Kindratenko, **Dawei Mu**, Yan Zhan, J.D. Maloney, National Center for Supercomputing Applications, Urbana, IL, USA

HOROVOD

NVIDIA

Open Cognitive Environment

PYTORCH

slurm workload manager

Wrapper Suite

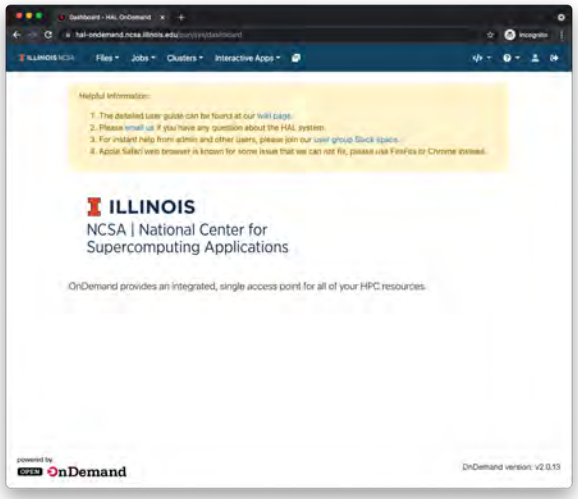
jupyter

Watson Machine Learning

TensorFlow

RAPIDS

∞



Visual Studio Code

julia

SINGULARITYCE

OPEN MPI

Traditional Terminal via SSH

- **mainstream ML/DL frameworks**
- **HPC essential software stack**
- **container solution**
- **tailored apps SWSuite and tar2h5**

Web Portal with Open OnDemand

- **build-in terminal**
- **build-in file manager**
- **build-in IDE VS Code**
- **Jupyter notebook/lab**
- **ML/DL UI H2O.ai**

H2O.ai

HDF

2022 Australian Computer Science Week