Round Table 2:
Instrument and Need

Victor Hazlewood

July 18, 2024

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

$4M NSF MRI “AI Instrument”

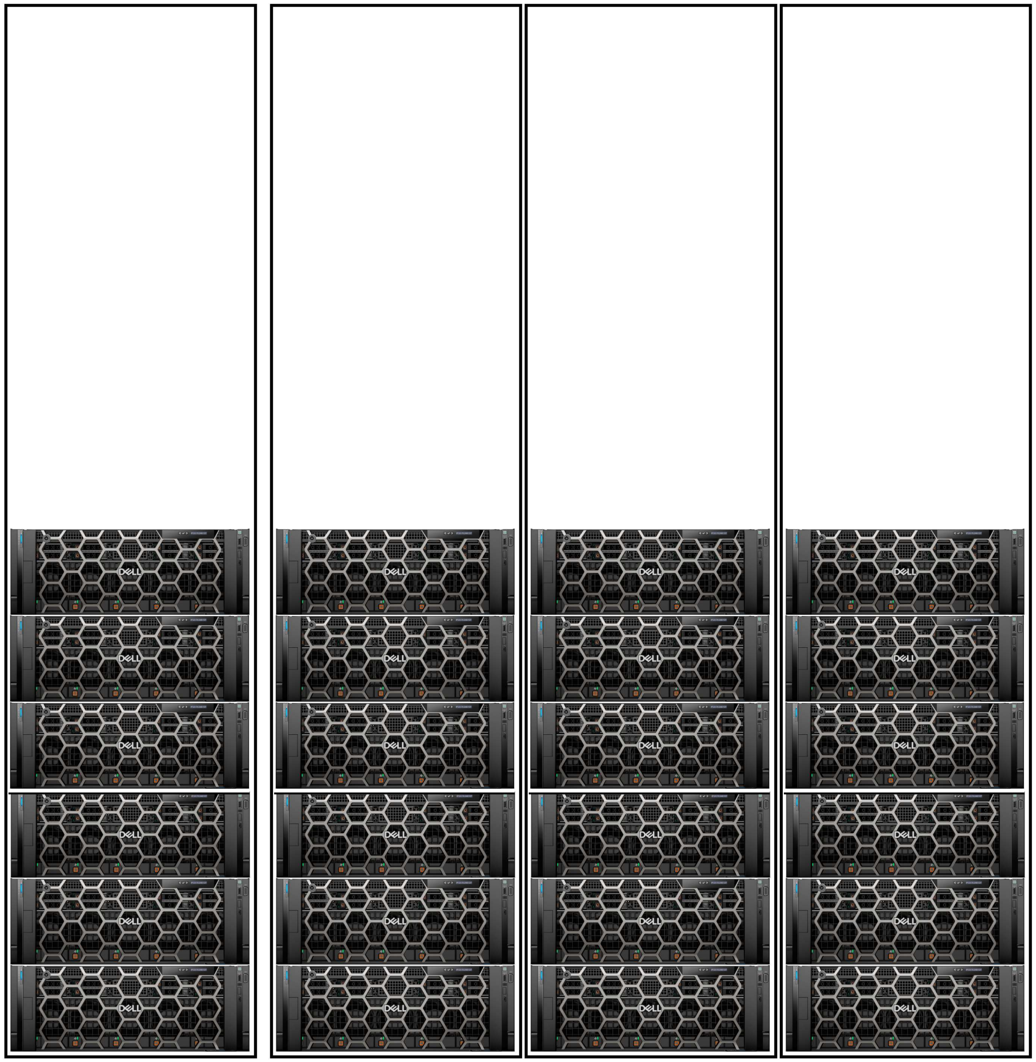
- A $4M AI computational resource to be placed into the ISAAC NG cluster at UTK*
- Two major components:
    - NVIDIA Blackwell based GPU servers
    - High Performance Storage

$4M NSF MRI “AI Instrument”

- NVIDIA Blackwell based GPU servers
  - Next generation Dell air-cooled GPU servers with NVIDIA Blackwell GPUs (like XE8640)
    - Pros: cheaper installation costs
    - Cons: takes up more rack space
  - Liquid cooled NVIDIA GB200 NVL72
    - Pros: the Ferrari configuration
    - Cons: higher installation costs due to liquid cooling and all in one rack power requirements (~200kW)

Solutions Rack Configuration

NVIDIA NVL72



Dell XE8640 type

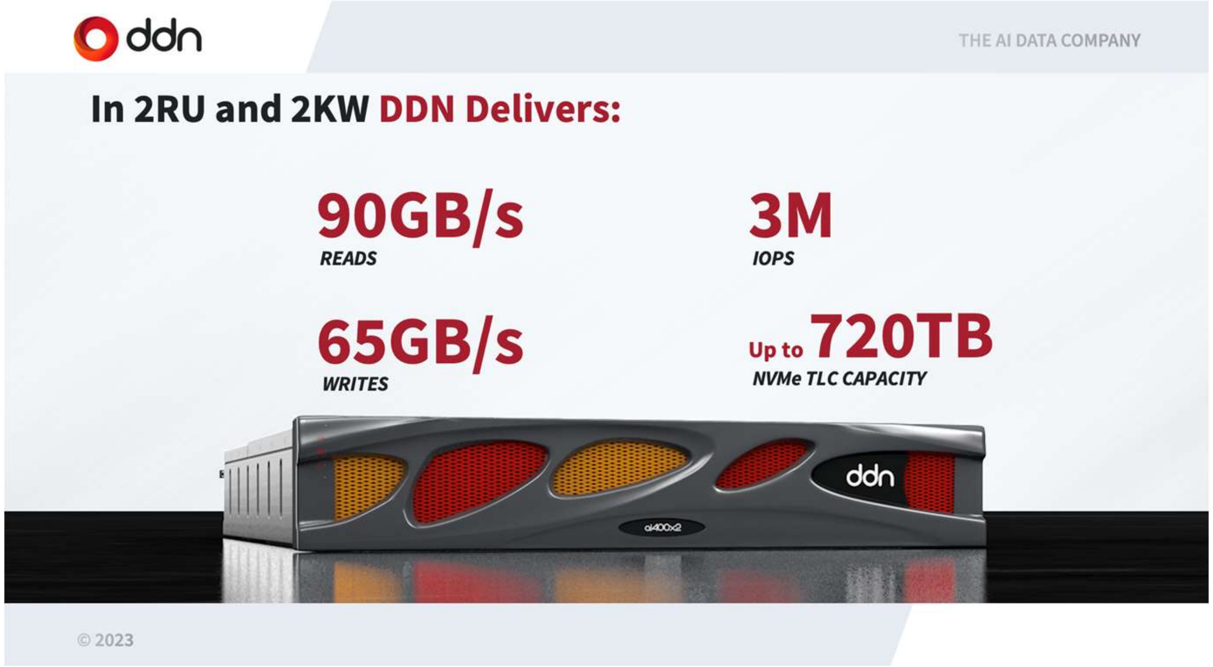
$4M NSF MRI “AI Instrument”

- High Performance Storage
  GPU I/O performance can be demanding
  - DDN AI400X2 1.1 PB TLC NVMe $797k
    TLC: lower density, better performance, fewer errors
  - DDN ES400NVX3 2.2 PB QLC NVMe $800k
    QLC: higher density, less performance, do not last as long
  - VAST 600 TB 5x Cbox, 2 IB switches, 5x Dbox
    $1M 250GB/s read, 27.5 GB write

DDN Storage

AI400X2 (TLC NVMe)

ES400NVX3 (QLC NVMe)



ddn

THE AI DATA COMPANY

In 2RU and 2KW DDN Delivers:

90GB/s *READS*

3M *IOPS*

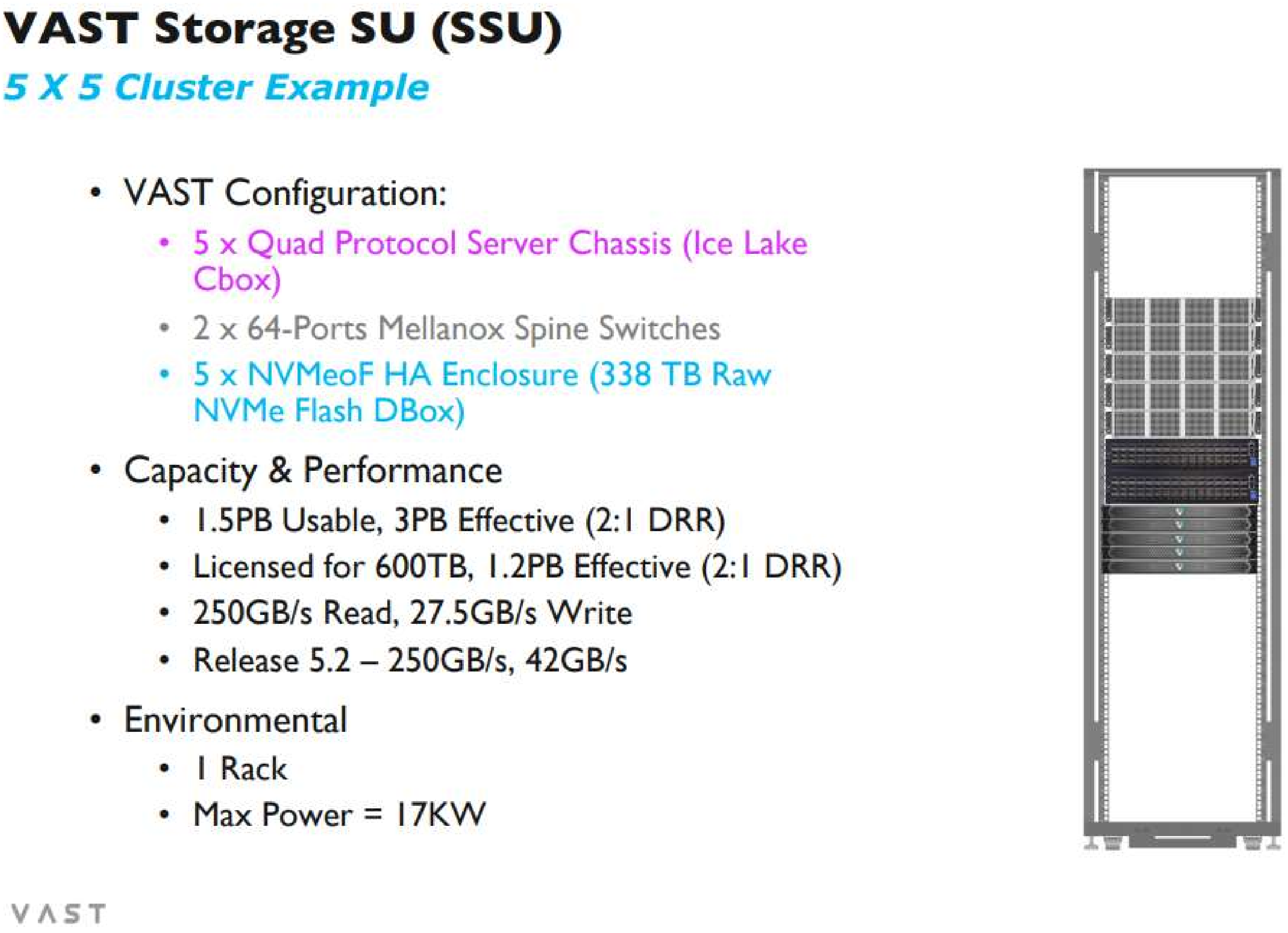
65GB/s *WRITES*

Up to 720TB *NVMe TLC CAPACITY*

© 2023

TLC Quoted solution is 2x with performance:
180 GB/s reads
130 GB/s writes

VAST Storage



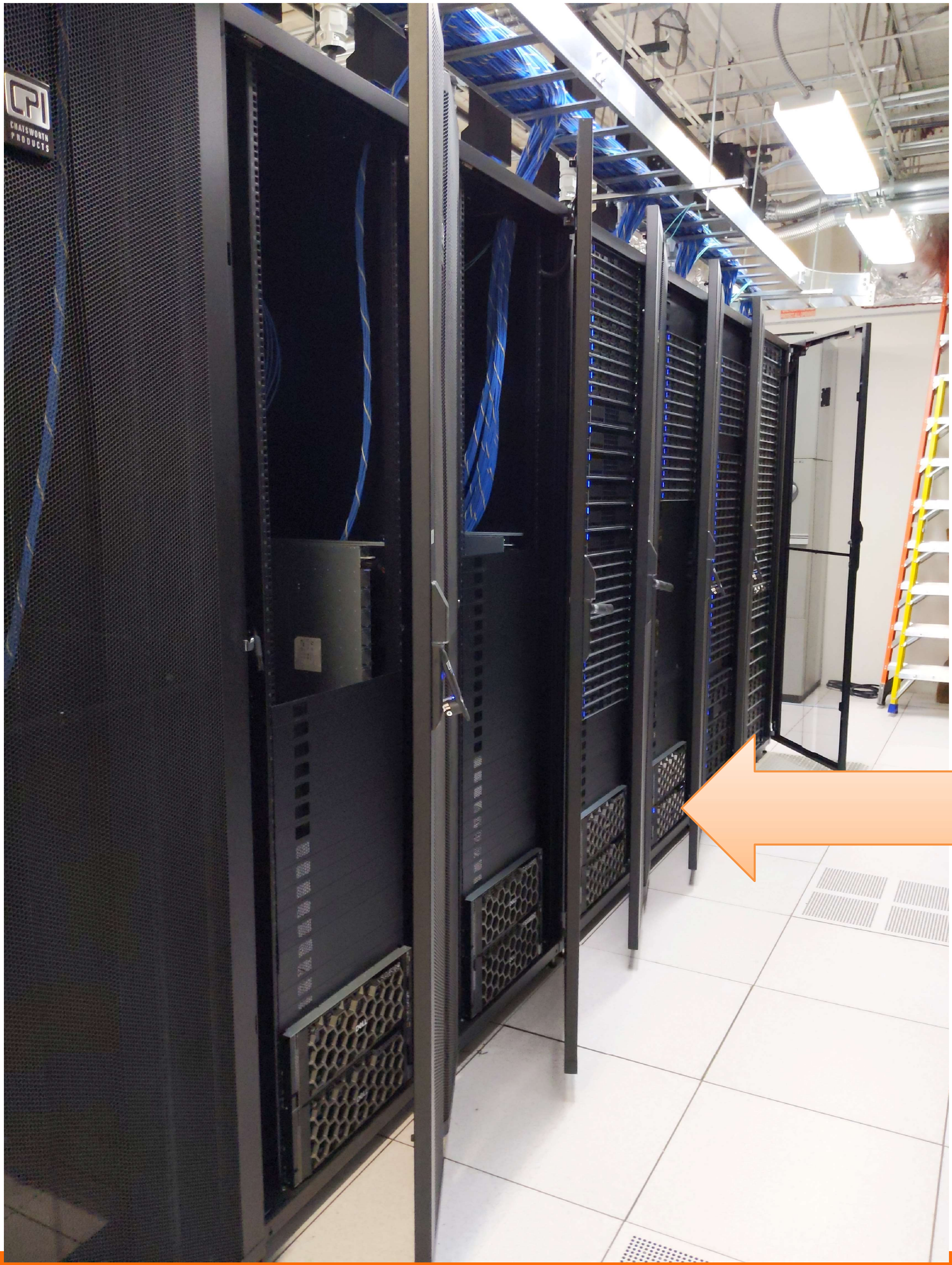
VAST Storage SU (SSU)

5 X 5 Cluster Example

- VAST Configuration:
  - 5 x Quad Protocol Server Chassis (Ice Lake Cbox)
  - 2 x 64-Ports Mellanox Spine Switches
  - 5 x NVMeoF HA Enclosure (338 TB Raw NVMe Flash DBox)
- Capacity & Performance
  - 1.5PB Usable, 3PB Effective (2:1 DRR)
  - Licensed for 600TB, 1.2PB Effective (2:1 DRR)
  - 250GB/s Read, 27.5GB/s Write
  - Release 5.2 – 250GB/s, 42GB/s
- Environmental
  - 1 Rack
  - Max Power = 17KW

VAST

Racks for ISAAC NG cluster at KPB data center



AI Tennessee Initiative
Dell XE8640 Servers

Bottom two 4U servers
In photo to the left

Instrument Need

- Need current facilities information on available resources for AI research and AI research training from each partner University
- From the researchers what do you have available currently?
- What type of research is not possible today that this instrument would make possible?
- RAG, LLMs for specific tasks or disciplines, others, ML capabilities?