# NHUG HPCD Update

***Irfan Elahi***
*Director of High-Performance Computing Division*
*&*
*Project Director NWSC-3 (Derecho)*

NCAR

**March 1, 2022**

NSF

# Welcome NHUG Members



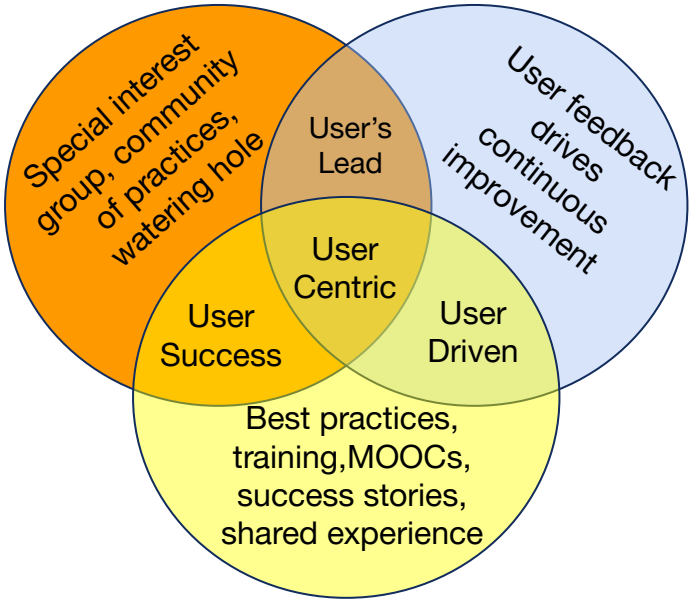





أَهْلاً و سَهْلاً


# NCAR HPC User Group (NHUG) - Scope

- User Group / Forum for HPC users
- Provide Feedback / Wishlist
  - Current / Future HPC systems / services
- Working Groups
  - Develop collaborations between subsets of user community aligned by interests
- Monthly NHUG Meetings



# NHUG communication Channels

wiki

Slack

Slido

# Derecho Status Update

# CY-2022 Milestones and Deliverables

## Quarter 1

- Continue Facility and Augmentation Work
  - 95% Completed.
- Delivery of GUST (Test System)
  - GUST Storage has already been deployed.
  - GUST Compute delivery expected March 2022.

## Quarter 2

- HPCD Project Team Planning
- ASD and Early User access to GUST
- HPE to Build Derecho and Storage Systems
  - At Chippewa Falls, WI

## Quarter 3

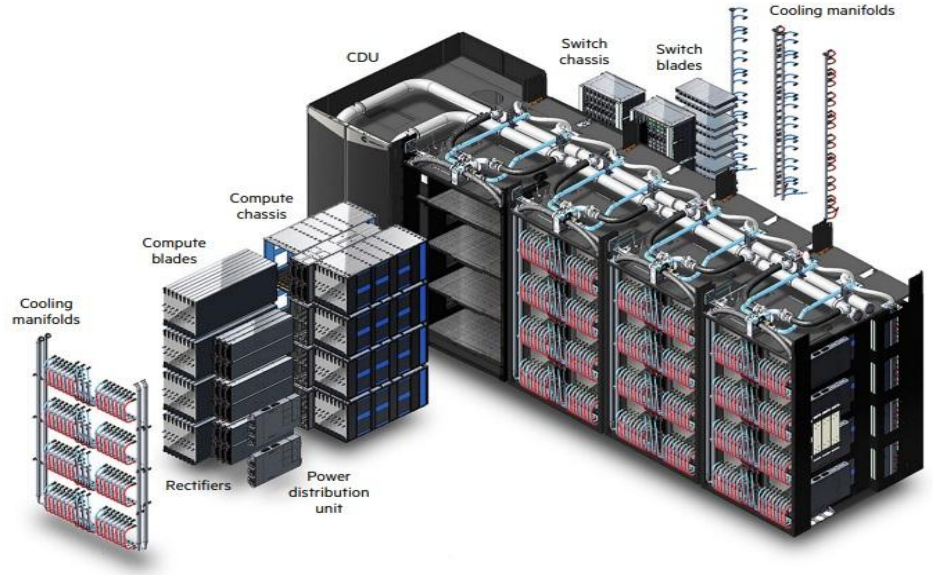
- Pre-delivery Factory Trials
- Delivery of Derecho and Storage System
- Commissioning, Functional Testing

## Quarter 4

- Commissioning and Functional Testing
- ATP and Benchmarking
- Solution Acceptance

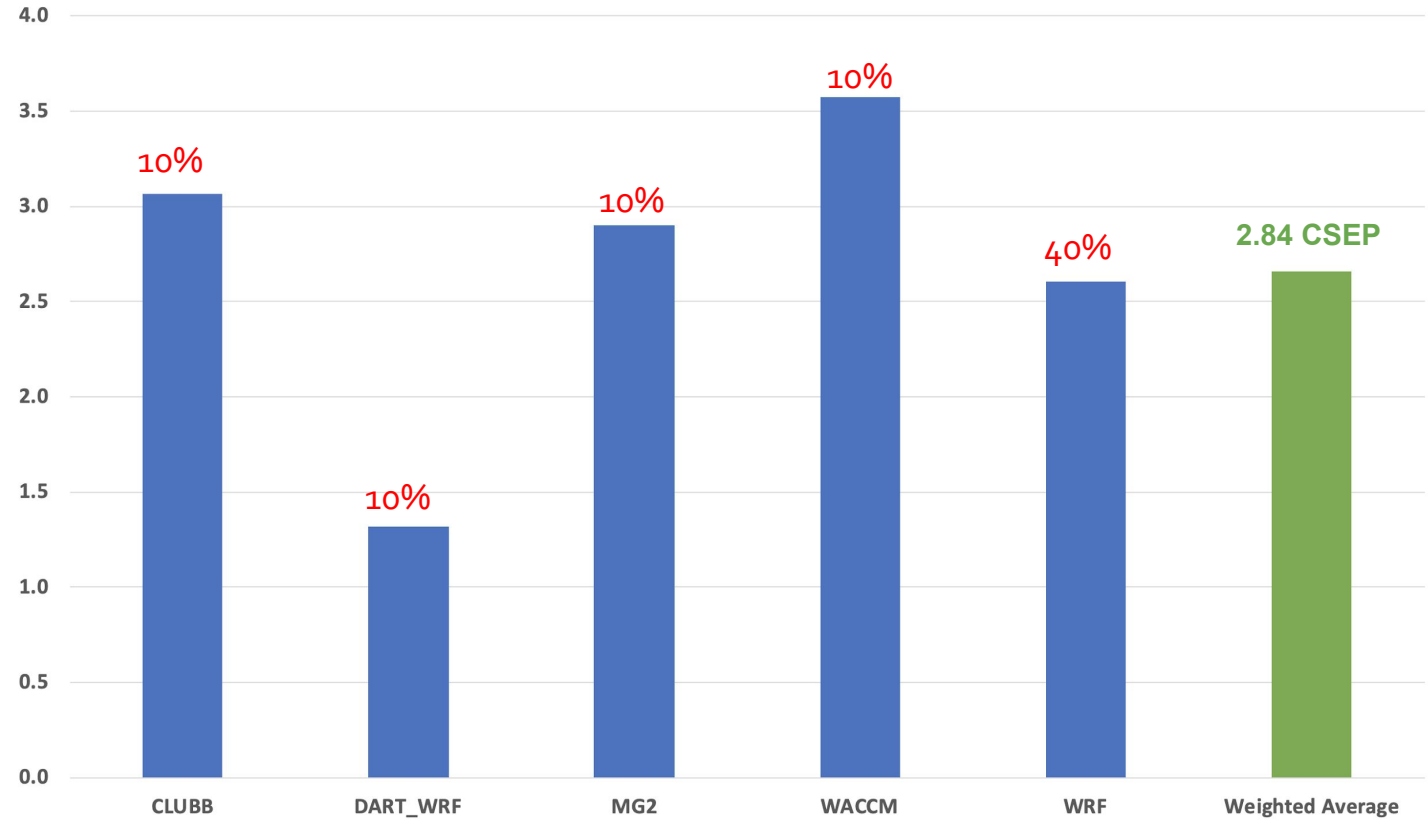
# Derecho (NWSC-3) HPE/Cray Solution

- Complete proposal received from HPE/Cray
  - Includes HPC and PFS
  - Peak: 19.87 PetaFlops
  - 60PB usable file system
- HPE CSEP Exceeds RFP requirement
  - 3.51 CSEP
    - CPU – 2.84 CSEP
    - GPU – 0.67 CSEP
- Large installed base
- Includes onsite 1x FTE support

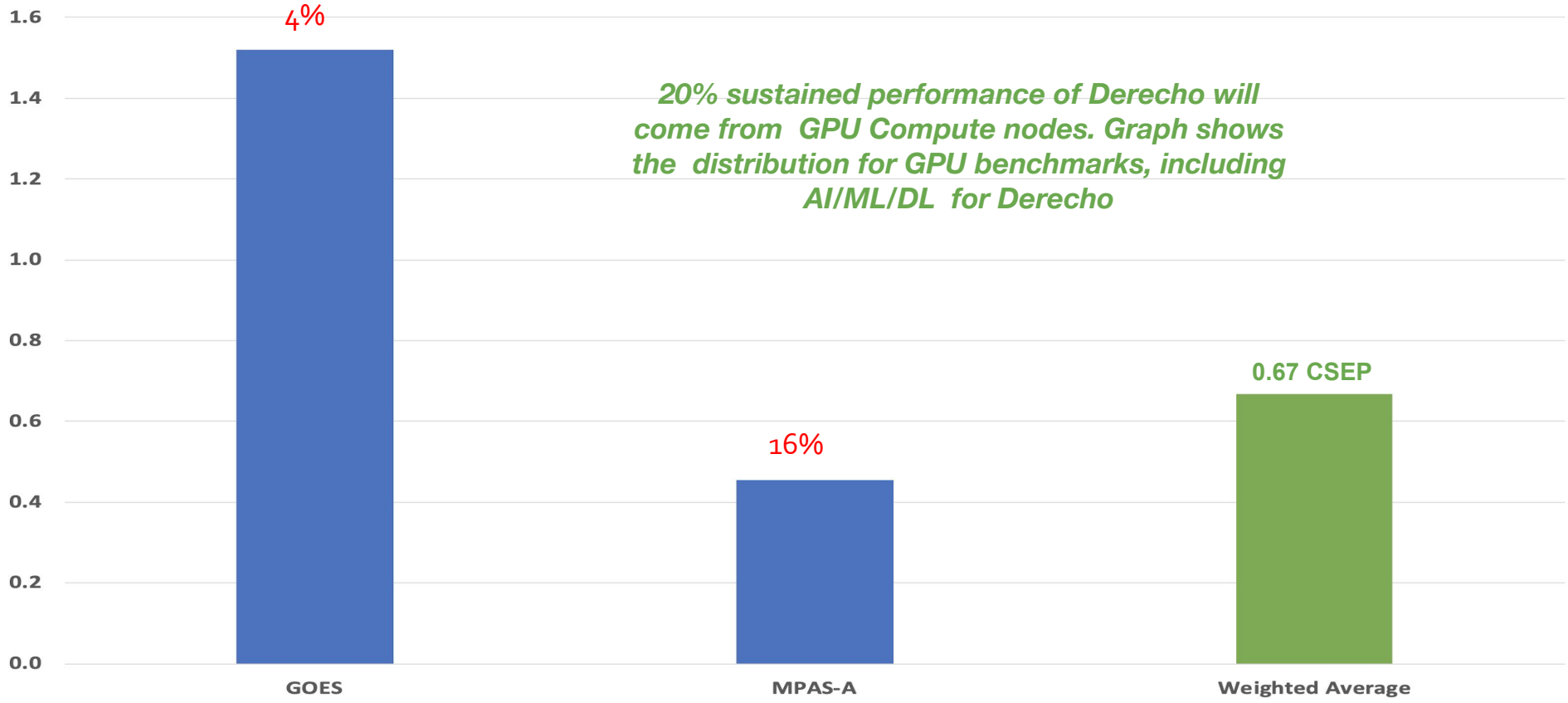


# CSEP (Cheyenne Sustained Equivalent Performance) - CPU Partition

***80% sustained performance of Derecho will come from CPU-Only Compute nodes. Graph shows the distribution for CPU-only benchmarks for Derecho***



# CSEP - GPU Partition




*20% sustained performance of Derecho will come from GPU Compute nodes. Graph shows the distribution for GPU benchmarks, including AI/ML/DL for Derecho*

# Derecho Nodes

## CPU Cabinet

4 nodes per compute blade

1 slingshot injection

64 blades per cabinet

256 nodes per cabinet

210.7 kW

0.65 tons of mech cooling

~1.3 PFLOPS

0.29 CSEP

## GPU Cabinet

2 nodes per compute blade

4 x GPU per node

4 Slingshot Injections

64 blades per cabinet

128 nodes per cabinet

190 kW

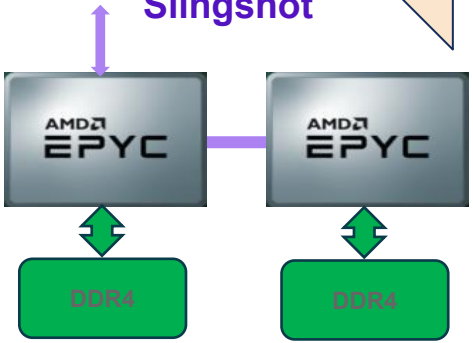
0.59 tons of mech cooling

~10.3 PFLOPS

1.04 CSEP

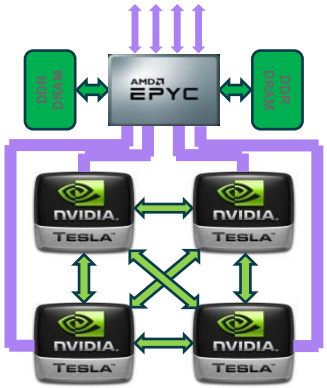

## CPU Node

To Slingshot




## GPU Node

To Slingshot



# Derecho - Storage & File System

- 6 x HPE/Cray ClusterStor E1000 systems
- 60 petabytes of usable file system space
  - Can be expanded to 120 petabytes
- 300 GB/s aggregate I/O bandwidth
- 5,088 × 16-TB drives
- 40TB SSD for Lustre file system metadata
- Two metadata management units (MDU) & 4 MDTs
  - One MDT exported per one MDS
  - Configured in highly available storage pairs
- Cray Lustre Parallel File System

# Derecho - Network Environment

## Derecho Production HPC System

11 Olympus Cabinets
Direct Water-cooled cabinets

2488 CPU-only Compute Nodes
82 GPU Compute Nodes

CPU-only Compute Nodes:
2 x 64c 2.45GHz AMD Milan
16x 16GB DIMMs (256GB Total)
1 x 200 Gb SS-11 NIC

GPU Compute Nodes:
1 x 64c 2.45GHz AMD Milan
8x 64GB DIMMs (512GB Total)
1x NVIDIA SXM4 A100 Redstone 4 GPU
4 x 200 Gb SS-11 NICs

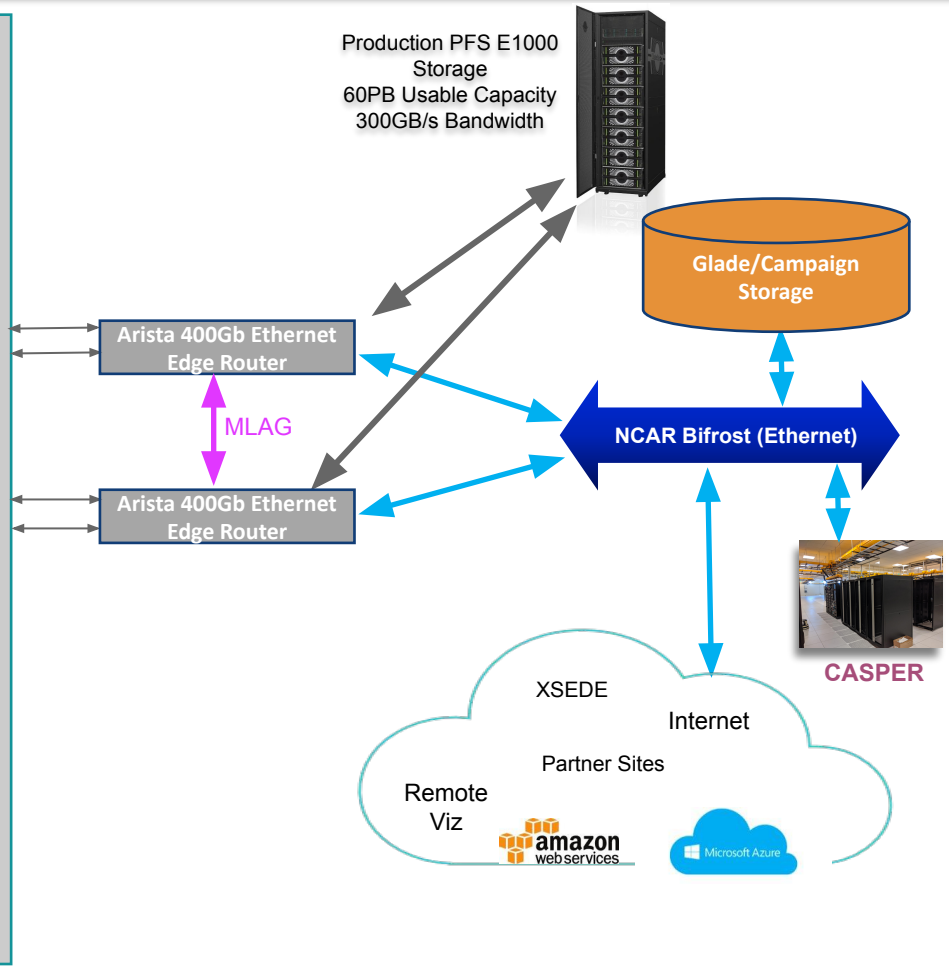
2 River Racks
Air-cooled 19" 42u Racks

20 Management Servers:
3 Cluster Managers, 9 Support, 2 Scheduler, 2 Fabric Managers

6 Login Nodes:
2 x 64c 2.45/3.5 GHz AMD Milan 7763
16x 32GB DIMMs (512GB Total)
1x 100Gb Ethernet adapter
1 x 200 Gb SS-11 NIC

2 GPU Login Nodes:
2 x 64c 2.45GHz AMD Milan
16x 32GB DIMMs (512GB Total)
2x NVIDIA GPU
1 x 200 Gb SS-11 NIC

Slingshot Interconnect Fabric



# Reliability, Availability, Serviceability, Usability & Manageability

Critical electrical and mechanical components on UPS

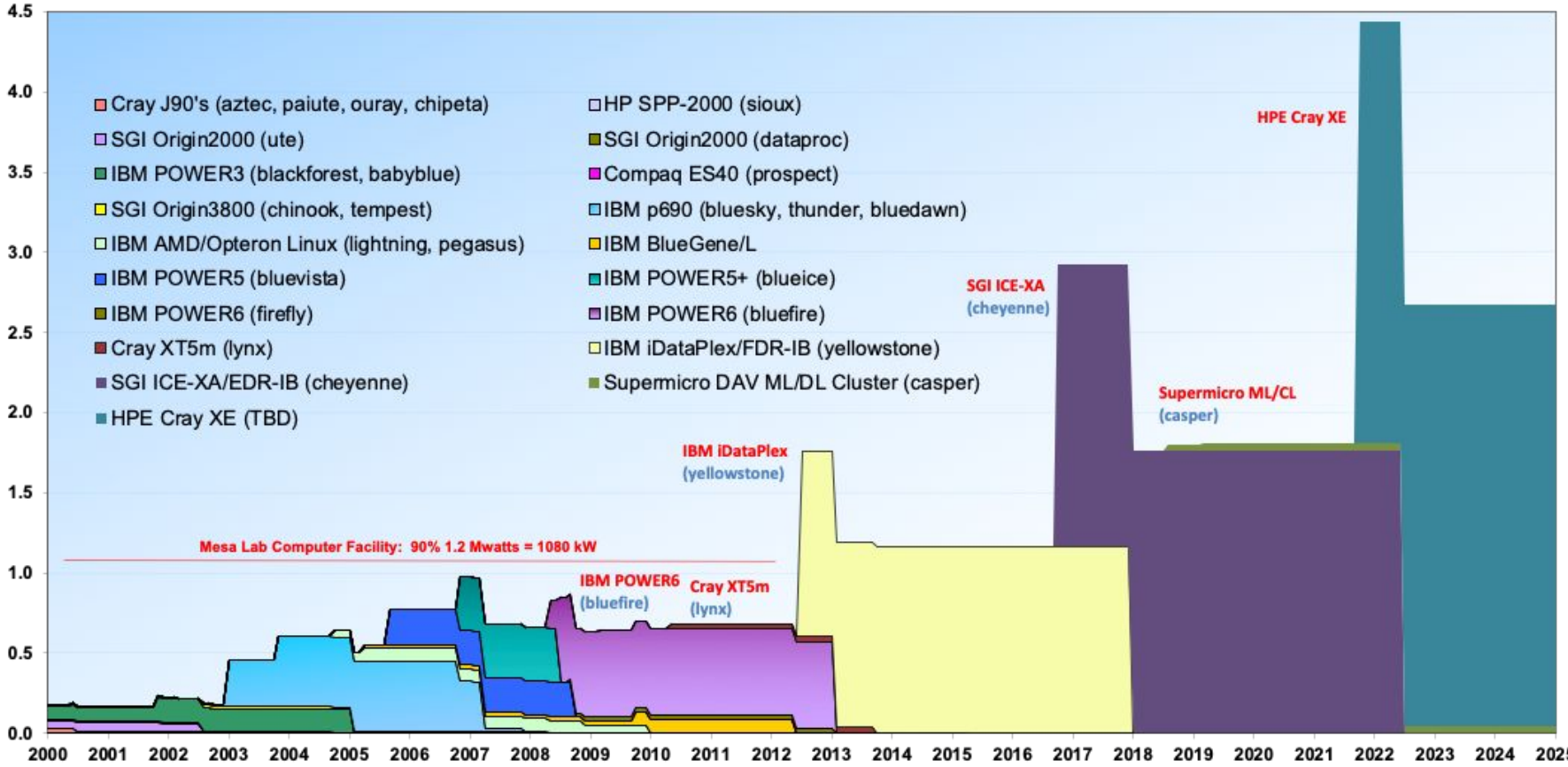

Architected with features for higher RASUM

Storage and file system will have 99% availability

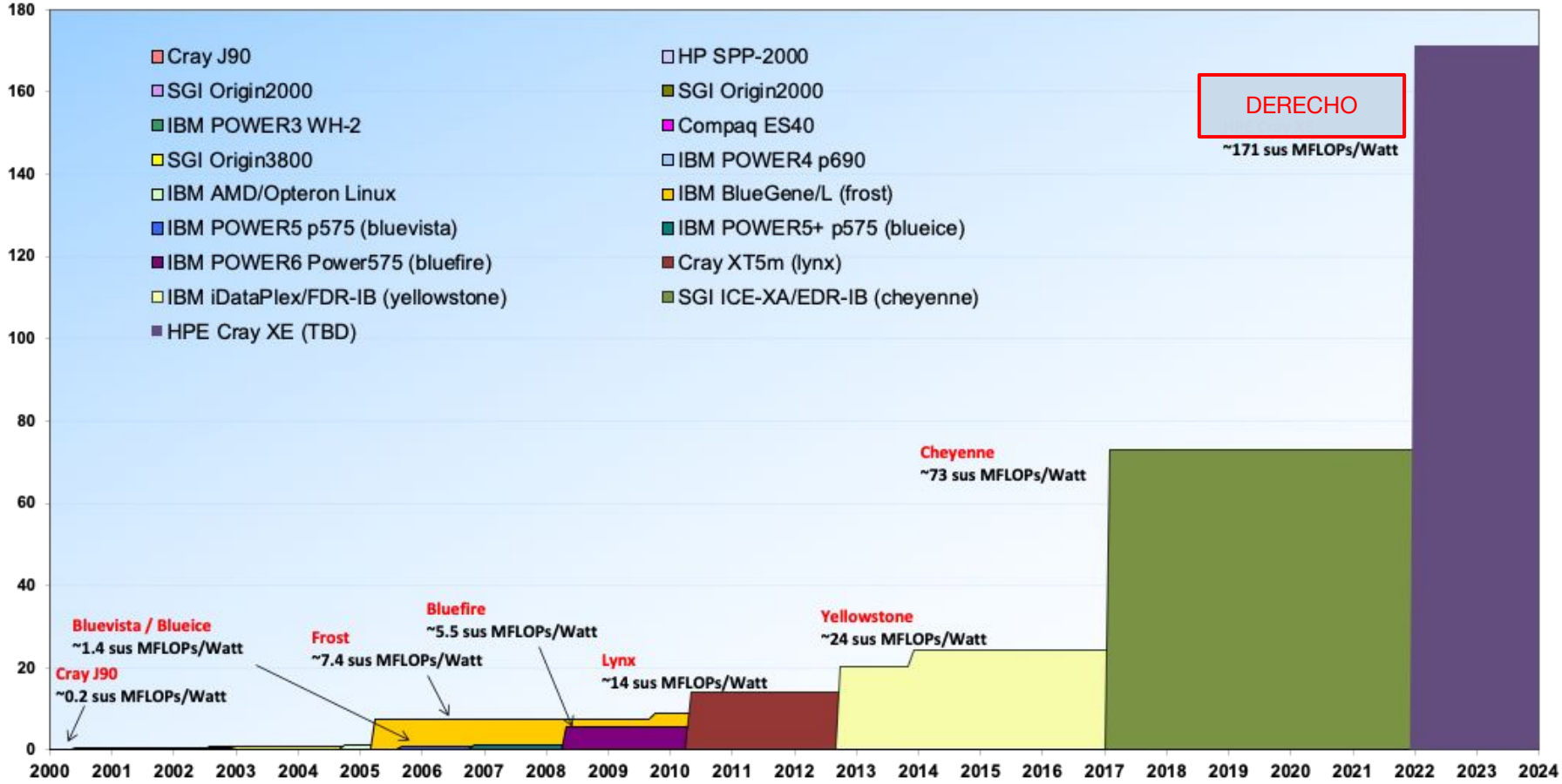
Derecho Compute nodes will have 98% availability

**Optimization of user environment to reduce failure**

# Total Facility Power Consumption (Max MW)



# System Power Consumption (sustained MFLOP/sec per Watt)



# Summary of University ASD Submissions

- 8 Submissions Considered
  - 322.4 million cpu-core-hours requested
  - 152,200 GPU-hours requested
- Each request reviewed by three panel members
- Each reviewer provided rating (excellent to reject)
  - Excellent=5 Reject=0
- Ratings averaged to rank order the proposals
- All proposals were rated good/very good (3.5) or better
  - This indicates that all proposals were of very high quality

# Summary of NCAR ASD Submissions

- Two phase process
- First Phase - Strategic Prioritization
  - 18 one page submissions
  - Reviewed by 15 member panel from NCAR Scientist Assembly (NSA)
  - 10 selected for 2nd phase
    - Based on strategic merit
    - Rough fit to available resources
    - 1 Proposal designated as an alternate in the event that other projects failed to make progress
- Second Phase
  - 10 full proposal prepared similar to CHAP proposals
  - Reviewed by NSC panel for computational efficiency and effectiveness
  - No issues identifies, all approved to move forward