NHUG NWSC-3 Update

Irfan Elahi
Director of High-Performance Computing Division
&
Project Director NWSC-3

NCAR

February 2, 2021

NSF

Significant Stakeholder Involvement

- Myriad user engagements (user survey, workload analysis, etc)
- Meetings with Science Requirement Advisory Panel (SRAP)
- Monthly report to NSF and monthly meetings with NSF
- Monthly meetings with UCAR Contracts
- Quarterly meetings with NCAR Director
- Weekly NWSC-3 project management meeting
  - *Schedule, budget, system engineering, RMP, PEP, education & outreach, transition to operations, etc*
- Weekly Meetings/Engagement with Technical Evaluation Team (TET)
- Business Evaluation Team (BET) engagements
- Regular meeting/engagements with Benchmark Team
- As required meetings with the Working Groups (Subject Matter Experts)
  - *Storage Working Group, System & Architecture WG, User Environment WG, Workflow, Analytics & Visualization WG and Operational & Metrics WG*
- Engagement with Vendors and Contractors
- External red team review
- NERSC, NREL & Oakridge visits
  - *Engagements with ECMWF, Argonne, NASA, etc*

NWSC-3 Strategic Objectives

- Provide a highly productive, data-intensive HPC resource for NCAR users and applications that builds on the success of the Cheyenne supercomputer and GLADE file system
- Invest in ExaScale concepts, GPU architectures, and Cloud computing technologies that are anticipated to be prevalent in systems in the next 5-10 years
- Enable application optimization and refactoring efforts required to make use of these technologies
- Ensure that users remain productive throughout the transition to NWSC-3, and beyond

NWSC-3 Scope and Requirements

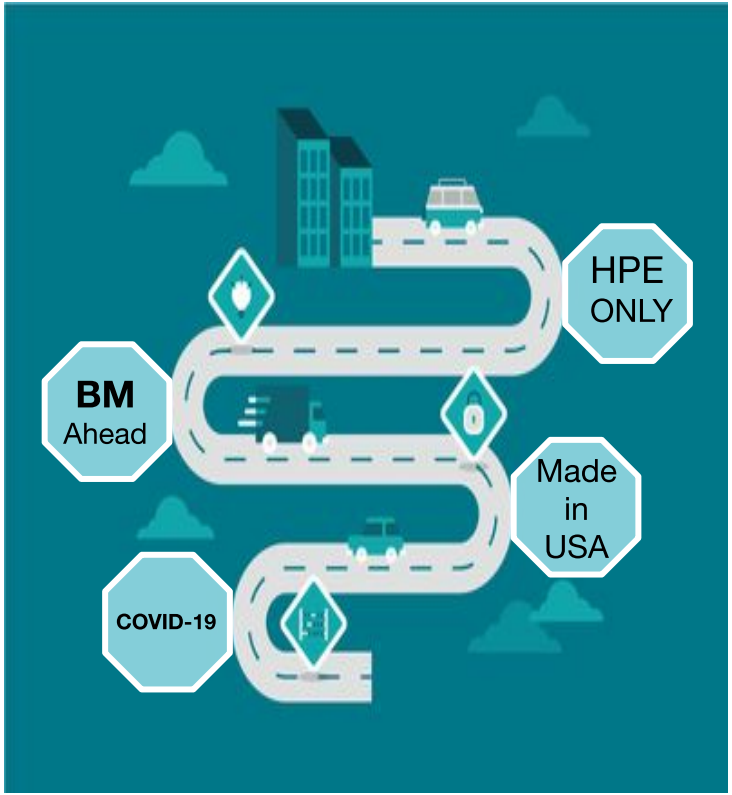
- High-Performance Computing (HPC) & Parallel File System (PFS)
- 3-fold sustained performance improvement over Cheyenne
  - 3 x Cheyenne Sustained Equivalent Performance (CSEP)
- 80/20 CPU/GPU Split or 80% = 2.4 CSEP and 20% = 0.6 CSEP
- 60 PB of usable file system storage
- Cloud bursting capability
- Must integrate into NCAR's Storage & Network environments
- Early production (ASD, porting, etc) - January 3, 2022
- Full production in April 2022
- Options
  - Option to add new racks with CPU &/or GPU partitions
  - Option to double the storage capacity
  - Option for out-year maintenance & support

NWSC-3 Co-Design Elements

- Co-design = working side by side
  - Not top-down
  - Not bottom-up
- C3 Approach
  - Collaborative
  - Collective
  - Consultative
- Intentional process to create solutions, innovation, and improvements to open up possibilities for better outcomes
- Co-design is dynamic and requires commitment to changes and feedback loop upto a certain point

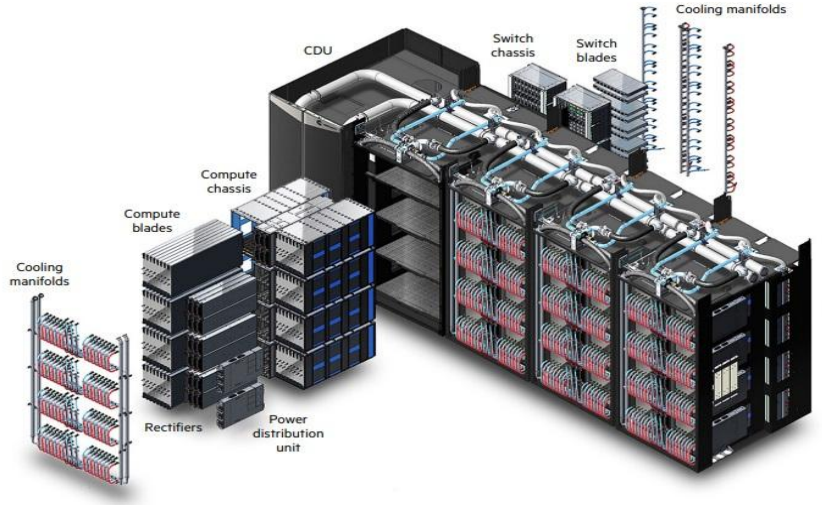
Procurement Challenges and Constraints

- **Limited domestic HPC vendor pool**
  - HPE has acquired SGI & Cray
  - IBM POWER was not cost competitive
- **Long time between RFP and delivery**
  - Facility Fit-Up requires significant lead time.
  - Dell will only quote for RTS (Ready To Ship) products
- **Benchmark suite requires resources**
  - Difficult for smaller vendors to bid
- **Vendors cited COVID-19 challenges**
  - Difficulty getting price warranty for 2021 delivery
  - Access to resources for RFP work
- **“Buy America” Requirement**



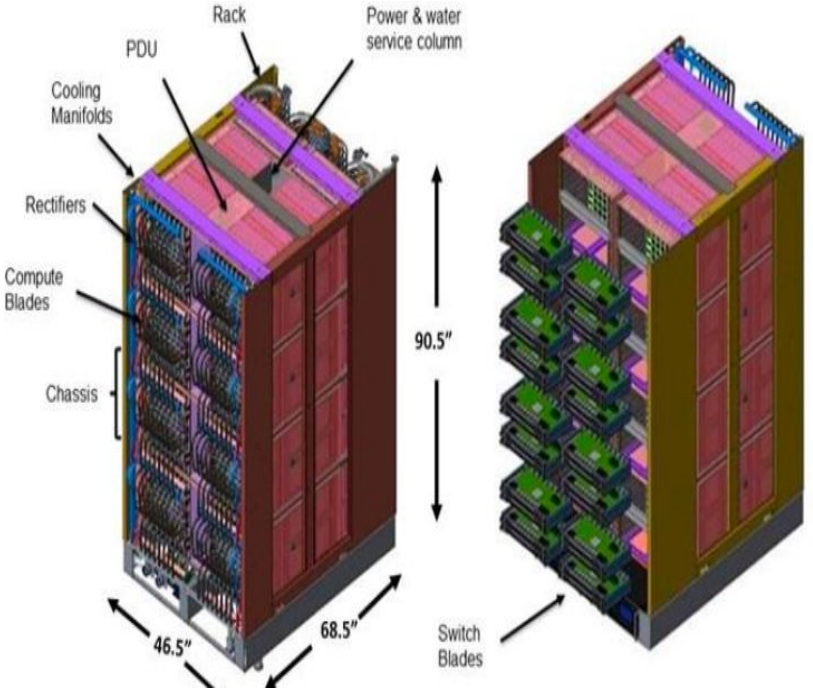
NWSC-3 HPE/Cray Solution

- Complete proposal received from HPE/Cray
  - Includes HPC and PFS
  - Peak: 19.87 PetaFlops
  - 60PB usable file system
- HPE CSEP Exceeds RFP requirement
  - 3.41 (3.51) CSEP proposed
    - CPU – 2.743 (2.84) CSEP
    - GPU – 0.67 CSEP
- Confidence in Crays' team
- Large installed base
- Includes onsite 1x FTE support
- HPE proposed PBS
  - Cheyenne currently uses PBS



HPE/Cray Solution - Key Facts

- Production HPC and PFS
  - Independent Test resources
- Cloud Bursting Capability
- 3.5-fold capacity improvement (3.5 CSEP)
  - 80/20 CPU/GPU Split or 80% = 2.8 CSEP and 20% = 0.67 CSEP
- 60 PB of usable file system storage
- Connectivity and interoperability with existing NWSC GLADE file systems
- Architected to easily augment the CPU and/or GPU partitions and storage



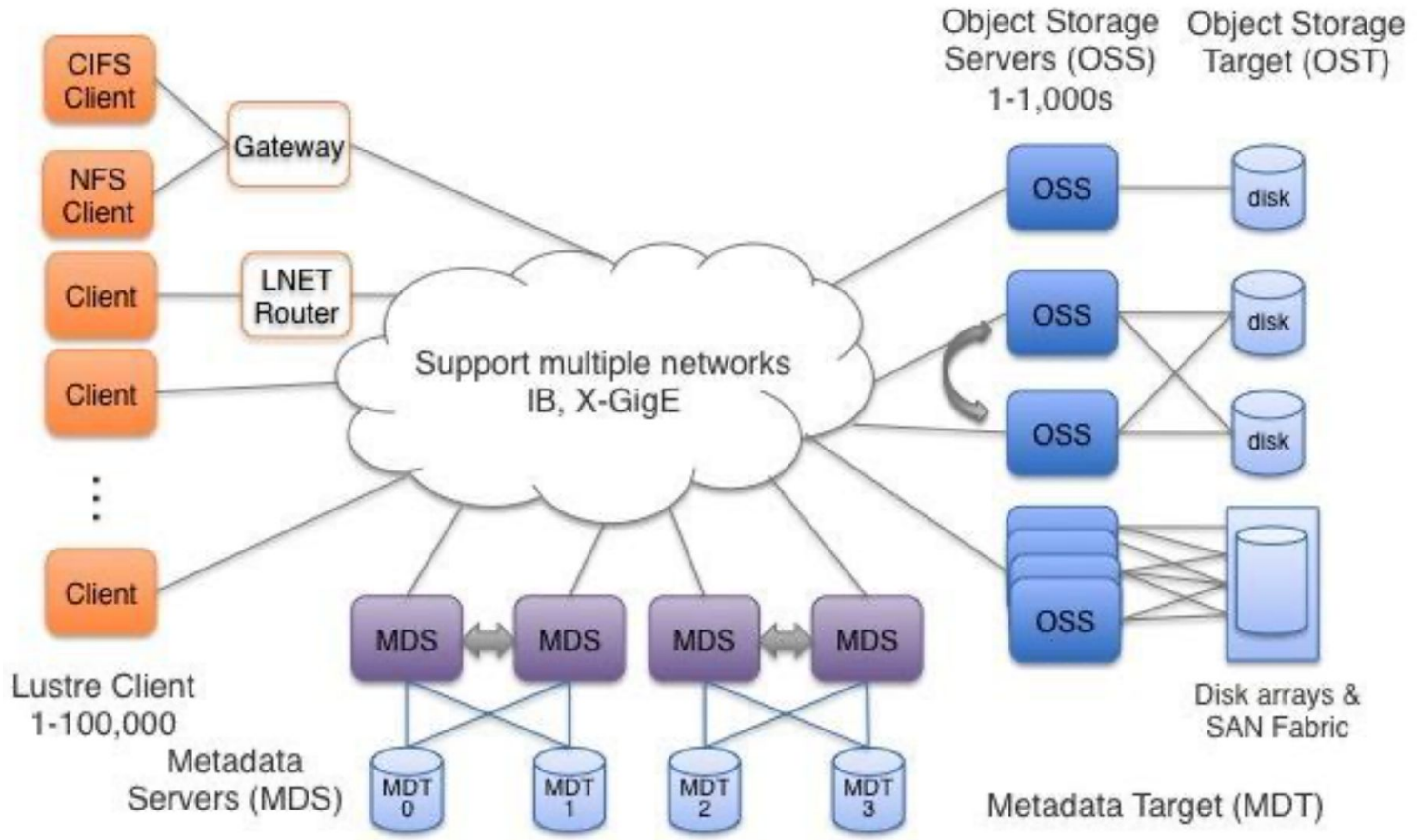
NWSC-3 Software Environment

- **Operating system is Cray Linux Environment (CLE)**
  - A tuned version of SUSE Linux
- **Altair PBS Professional Workload Manager**
  - With Accelerator Plus scheduler
- **Support for Docker containers, Singularity containers**
  - Supports the Open Container Initiative standard
- **Cray Programming Environment (CPE)**
  - Supports OpenMP 4.5 and 5.0, and MPI v3.1
- **CrayPAT: Cray Performance Analysis Tool**
- **NVIDIA (formerly PGI) Compiler Environment**
- **Intel Parallel Studio XE compiler suite**
- **Cray Lustre File System (based on 2.12 LTS)**

HPE/Cray Solution - Storage & File System

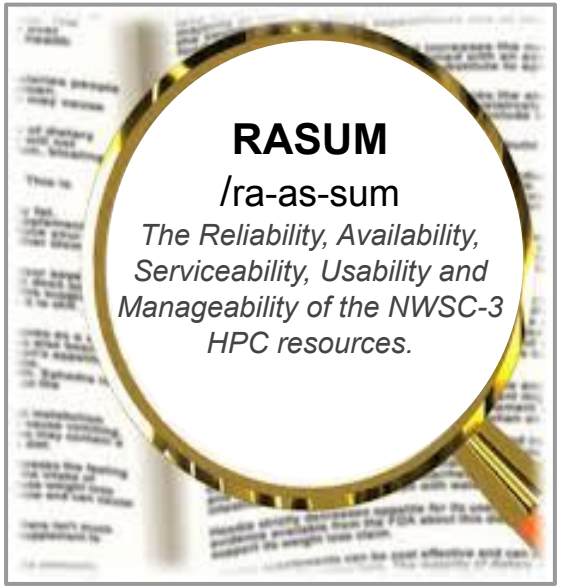
- Six HPE/Cray ClusterStor E1000 systems
- 60 petabytes of usable file system space
  - Can be expanded to 120 petabytes by exercising options. Requires additional racks.
- 300 GB per second aggregate I/O bandwidth to/from the NWSC-3 HPC system
- 5,088 × 16-TB drives
- 40TB SSD for Lustre file system metadata
- Two metadata management units (MDU) exporting four MDTs
  - One MDT exported per one MDS
  - Configured in highly available storage pairs
- Cray Lustre Parallel File System

Lustre Setup



Reliability, Availability, Serviceability, Usability & Manageability

Critical electrical and mechanical components on UPS



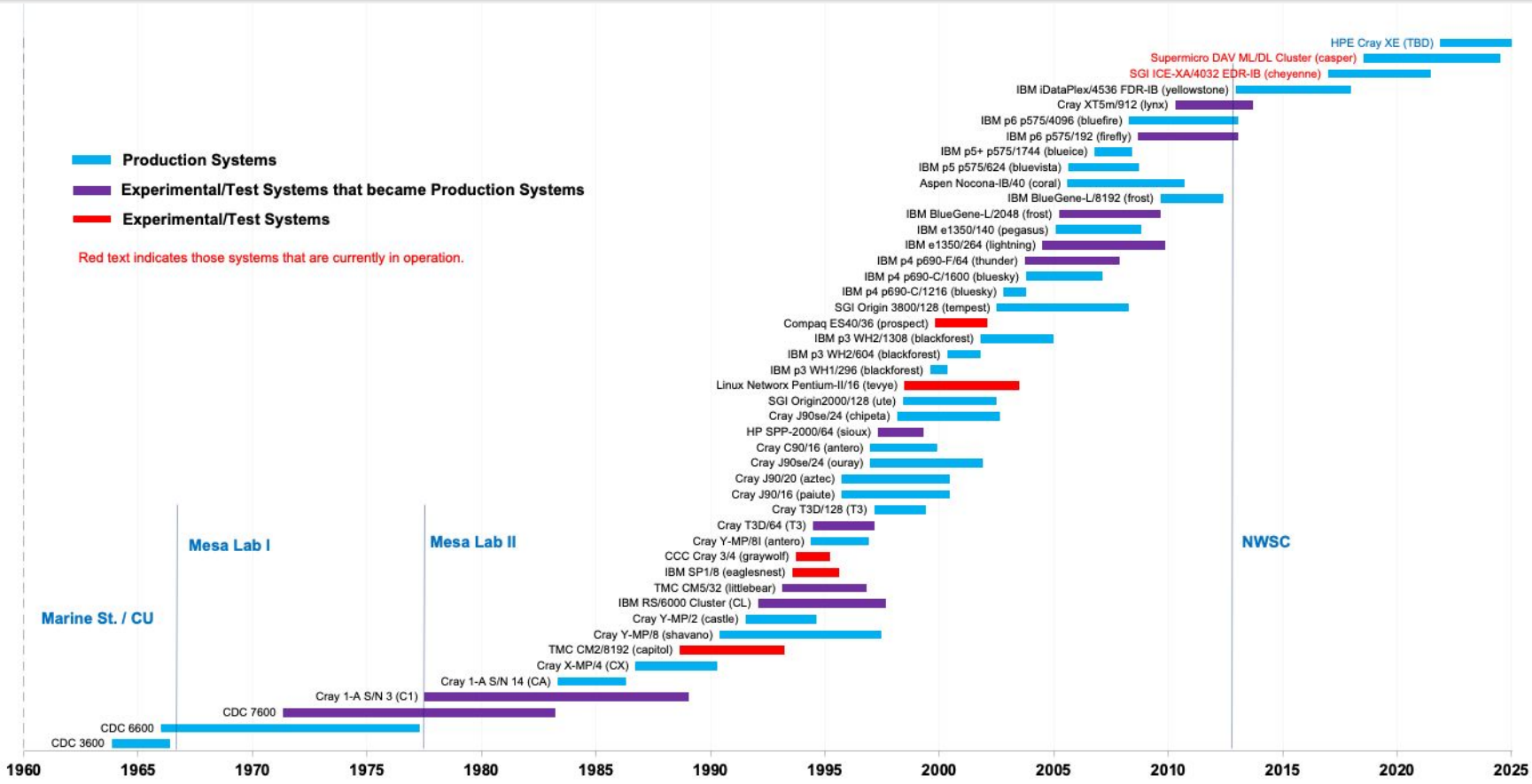
Architected with features for higher RASUM

Storage and file system will have 99% availability

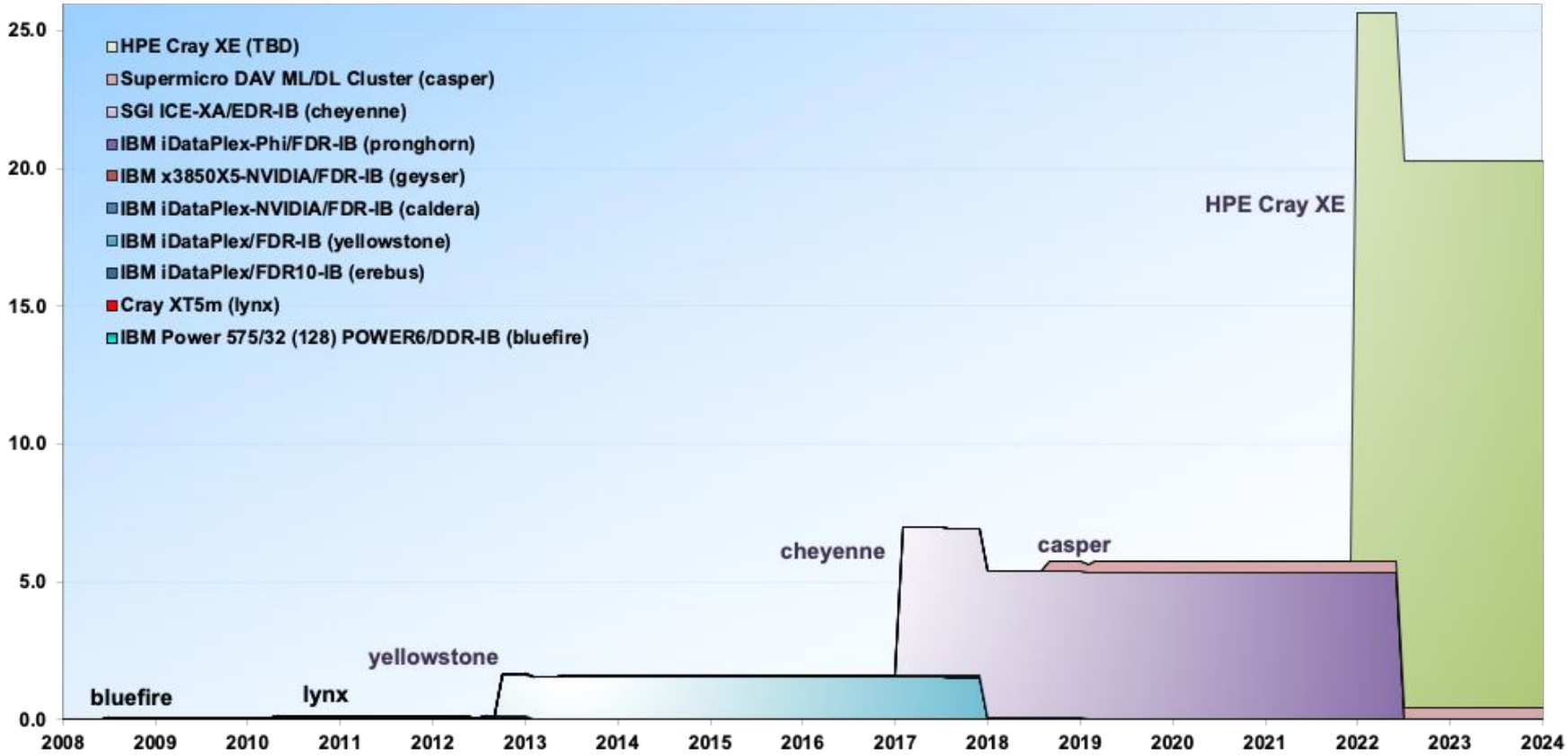
NWSC-3 HPC system will have 98% availability

Optimization of user environment to reduce failure

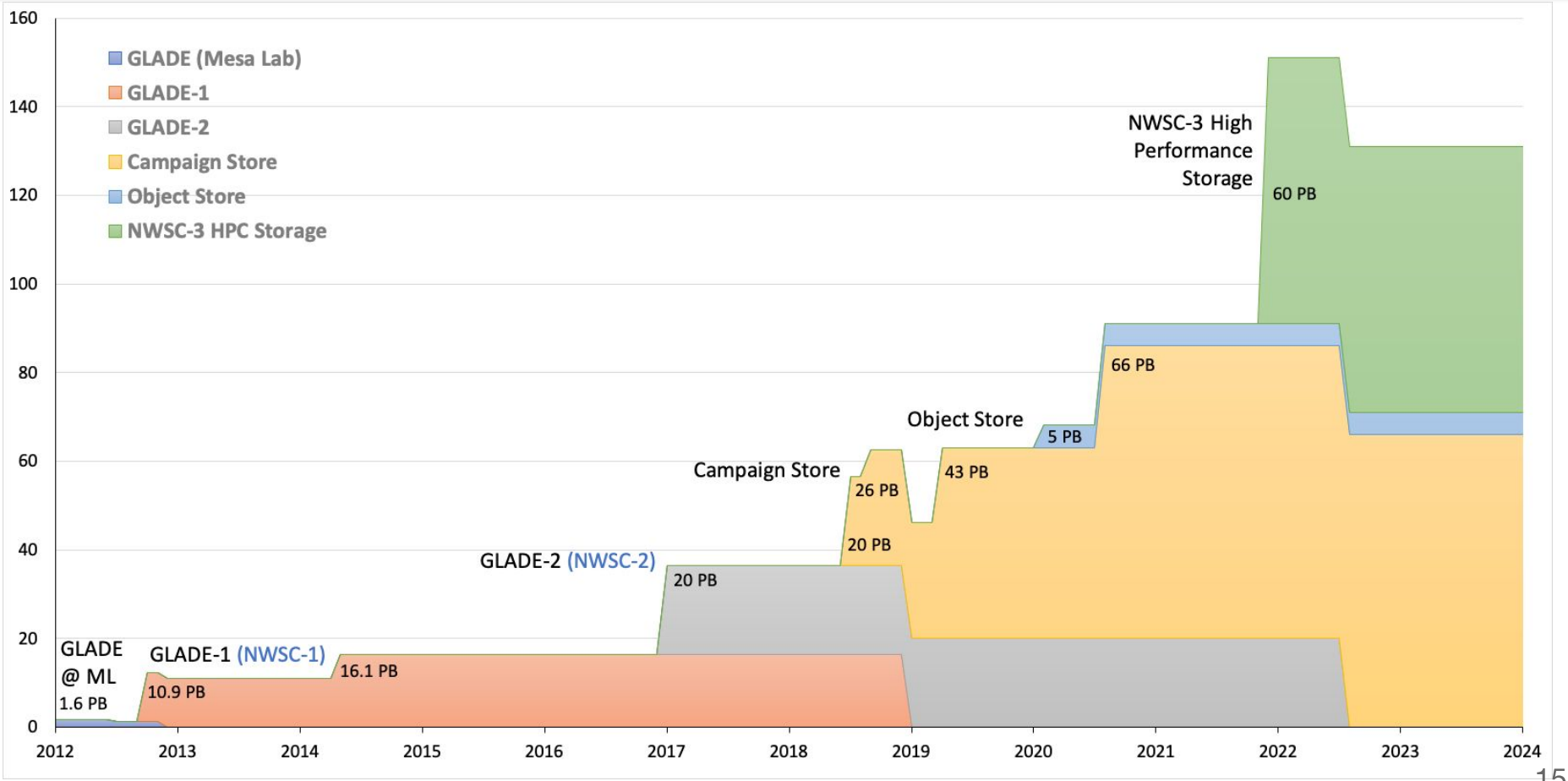
History of Supercomputing at NCAR



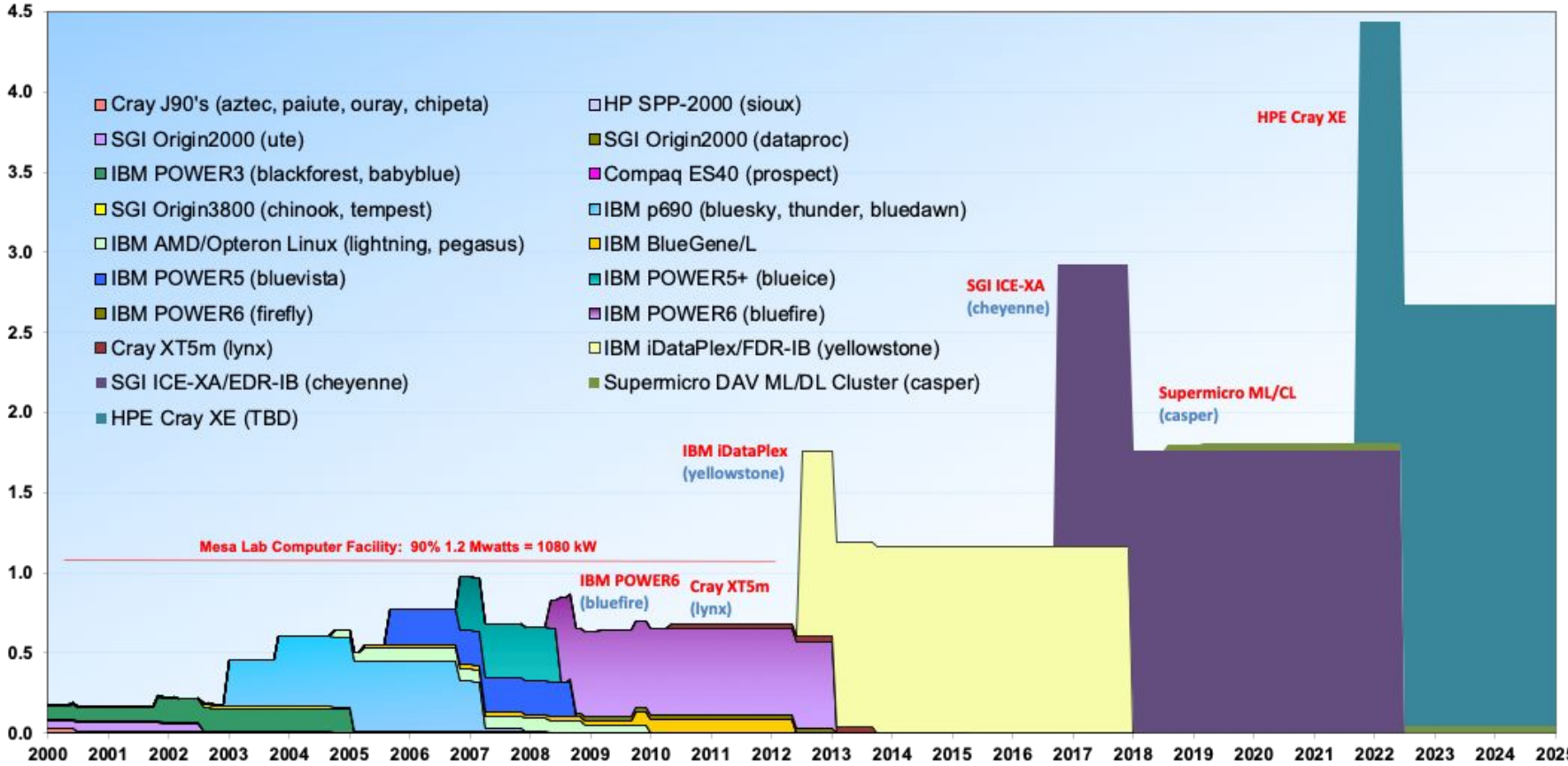
Peak PFLOPs at NCAR



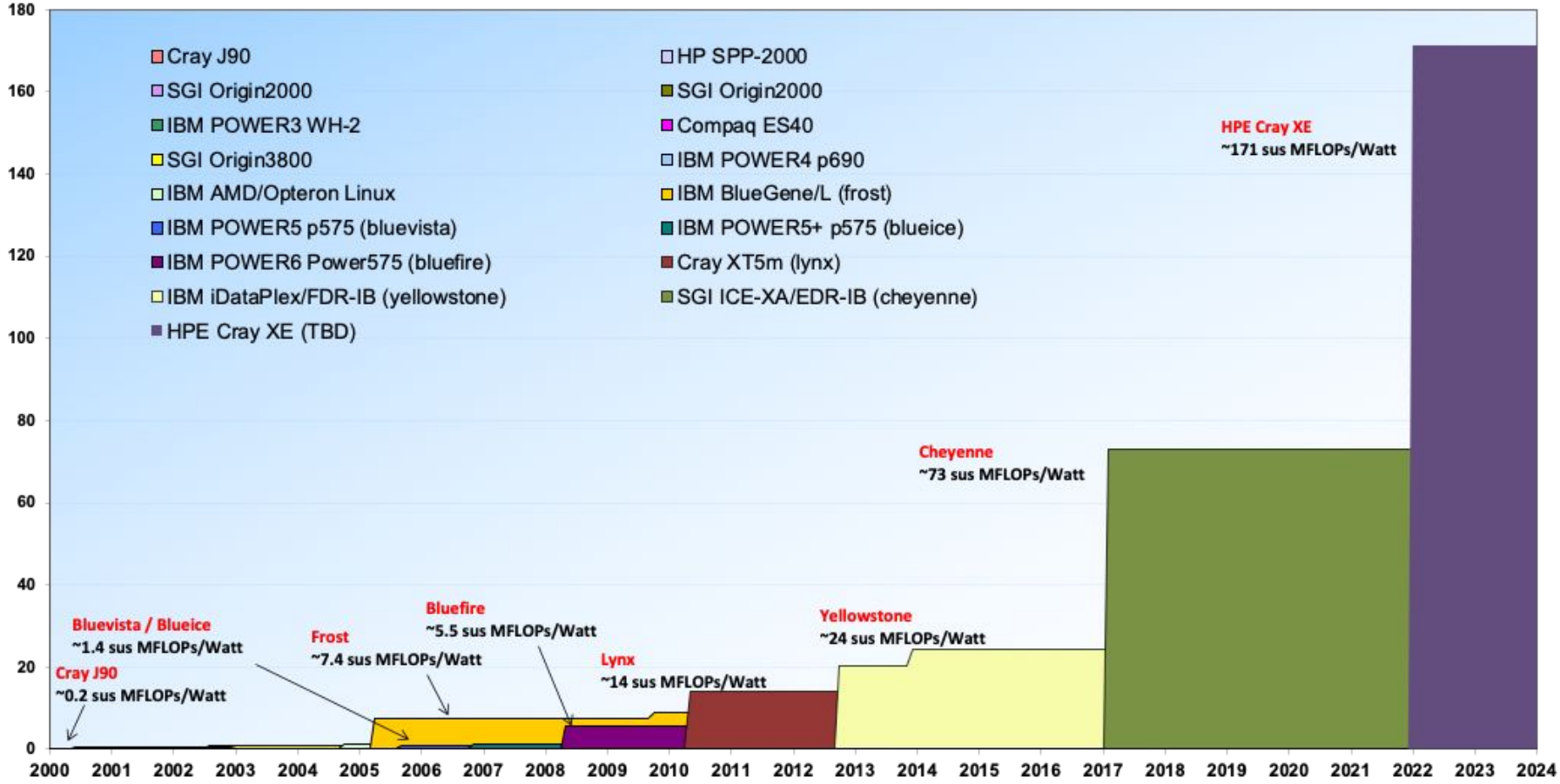
Total File System Storage Capacity at NCAR (PB)



Total Facility Power Consumption (Max MW)



System Power Consumption (sustained MFLOP/sec per Watt)



CY2021 Milestones and Deliverables

Quarter 1

- ~~NSF approval~~
- NWSC-3 contract award
- Begin NWSC-3 fit-up construction effort
- NWSC-3 Deployment Project Kick-Off

Quarter 2

- NWSC Visitor Center upgrade project begins with selected contractor
- Complete NWSC capacity upgrade

Quarter 3

- Pre-delivery factory trials
- NWSC-3 test systems delivery - August
- NWSC-3 fit-up nearing completion

Quarter 4

- NWSC-3 delivery - 10/01/2021
- NWSC-3 installation
- NWSC-3 conclude acceptance testing - 12/31/2021
- NWSC-3 ready for production use - **01/03/2022**

Casper - DAV-ML

Casper Augmentation Completed:

- New 8x NVIDIA V100 GPU Nodes
  - Dual IB & 100GbE
- New 4x NVIDIA V100 GPU Nodes
  - Dual IB & 100GbE
- Dedicated RDA nodes

Coming Soon:

- 64x *HTC* nodes
  - *HDR100 IB & 100GbE*
  - *384 GB RAM*
- New dedicated login nodes
- New wide & deep racks
- Common Scheduler across Cheyenne, Casper and NWSC-3