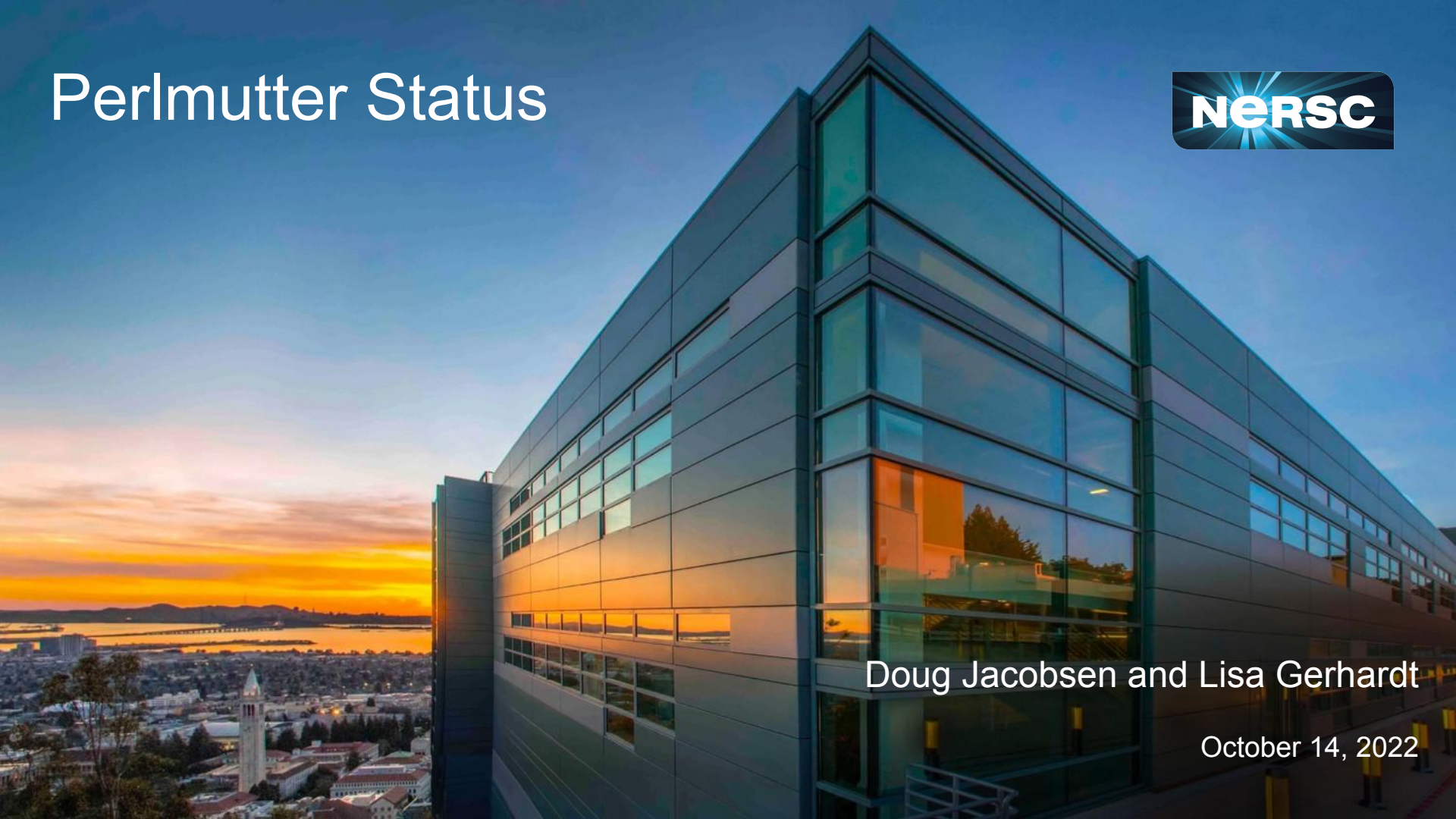
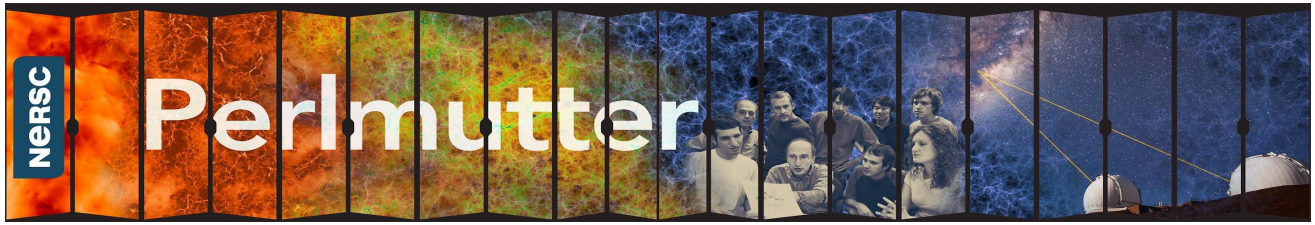
Perlmutter Status

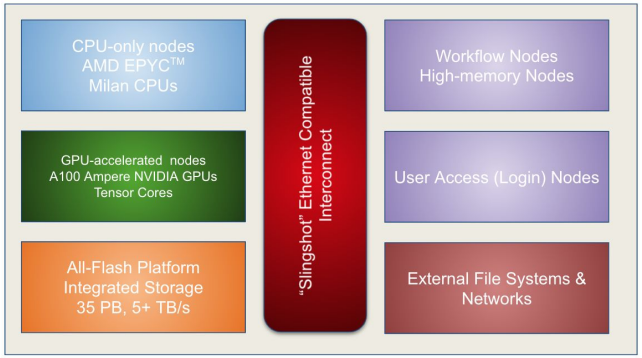


Doug Jacobsen and Lisa Gerhardt

October 14, 2022



- HPE Cray System with 4x capability of Cori
- GPU-accelerated (GPU/CPU) and CPU-only nodes
- HPE Cray Slingshot high-performance network
- All-Flash filesystem



Phase I: Arrived Spring 2021

- 1,536 GPU-accelerated nodes
- 1 AMD "Milan" CPU + 4 NVIDIA A100 GPUs per node
- 256 GB CPU memory and 40 GB GPU high BW memory
- 35 PB FLASH scratch file system
- User access and system management nodes

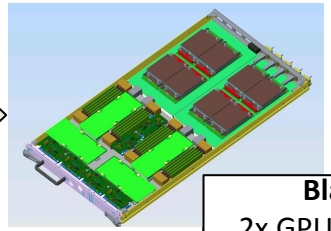
Phase II Addition - arrives last Winter

- 3,072 CPU only nodes
- 2 AMD "Milan" CPUs per node
- 512 GB memory per node
- Upgraded high speed network
- CPU partition will match or exceed performance of entire Cori system

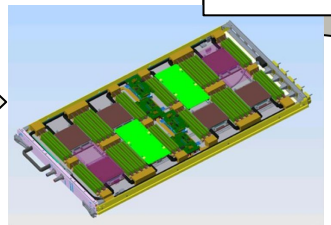
Perlmutter at a glance

35 PB All Flash File System

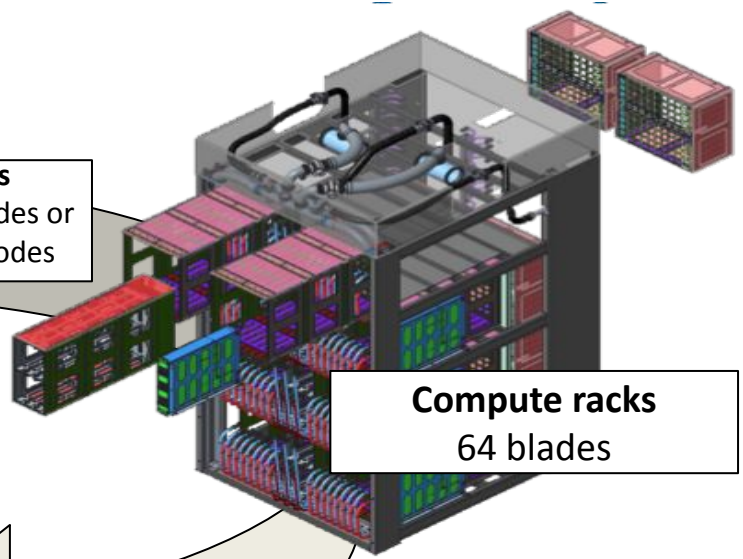
1536 NVIDIA A100 GPU Nodes
4x GPU + 1x CPU
160 GiB HBM + DDR
4x 200G "Slingshot" NICs



AMD EPYC 7003 CPU Node
2x CPUs
> 256 GiB DDR4
1x 200G "Slingshot" NIC



Blades
2x GPU nodes or
4x CPU nodes

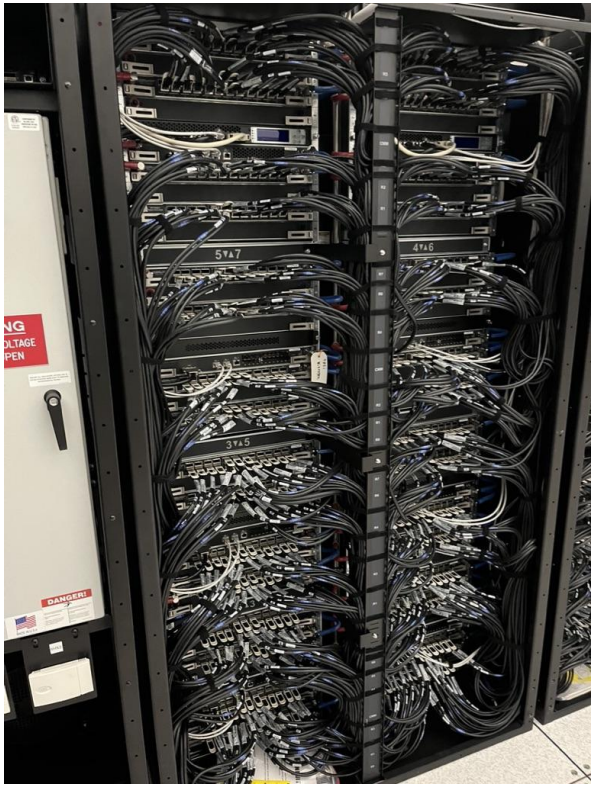


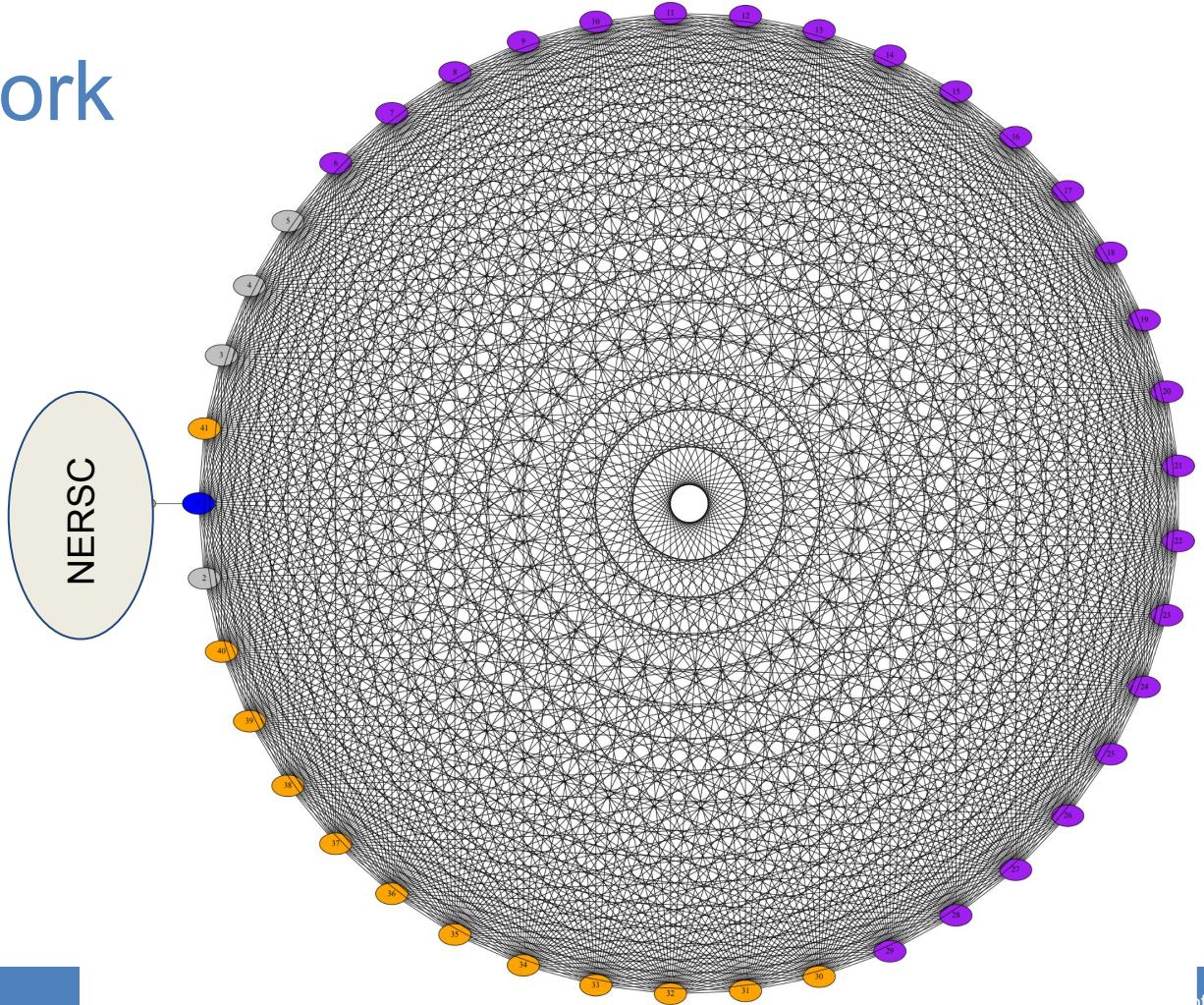
Compute racks
64 blades

Centers of Excellence
Network
Storage
App. Readiness
System SW

Perlmutter system
12 GPU racks
12 CPU racks
~6 MW

Slingshot Network





New, Non-Compute Functionality

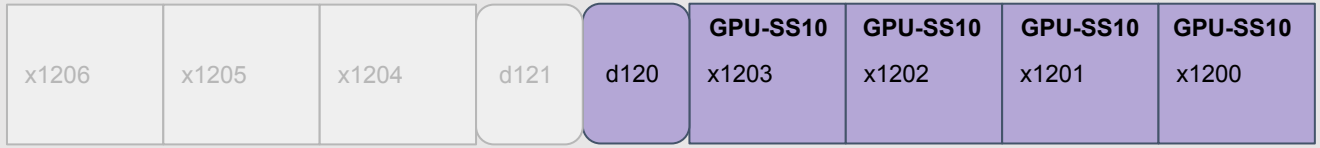
- Direct network connectivity from NERSC network to System High Speed Network
- Quota is enforced by the file system instead of the batch system
  - 20TB “soft” quota, 30TB “hard” quota, grace period 24 hours
- "Native" GPFS Clients on Computes
- Lmod: a more sophisticated program for handling modules
- Crontabs replaced with “scrontab”
  - Same functionality, but scheduled via the batch system
- Dynamic linking by default to avoid needing to recompile every time the system is updated
- Kubernetes-based control plane enhances system manageability, reliability, and provides API-based interactions
  - In the future, will expand user-interaction options
- Rolling reboots to minimize user disruption

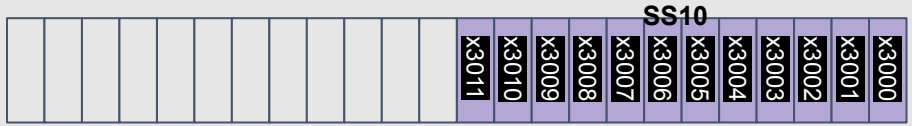
Why Phases?

- Technology delays in system components
- Went with a phased delivery to maximize the availability of Perlmutter resources to users
- Phase I: GPU nodes, SSD Lustre, Slingshot10 (Cray Network / Mellanox NIC)
- Phase II: CPU nodes, Rebuild GPU Nodes, Slingshot11 (Cray Network / Cray NIC)

Phase 1

January 2022





testing

not present

perlmutter

muller

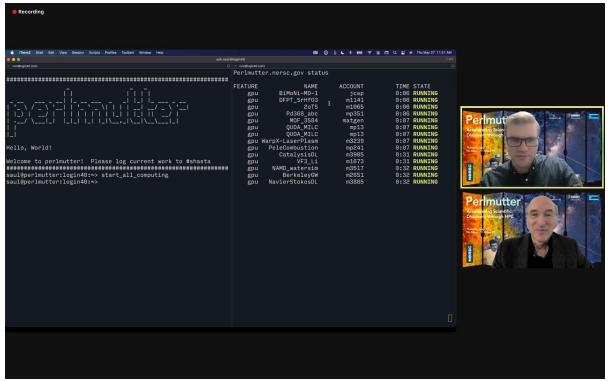
alvarez

x storage/gw

NERSC

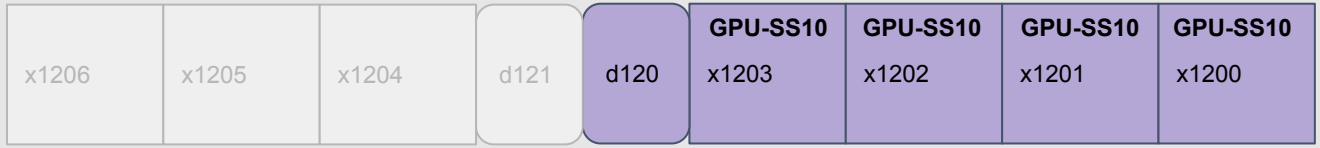
Deployment Timeline

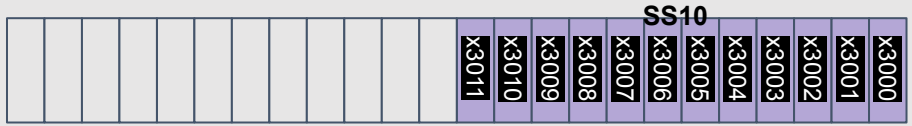
- December, 2020: Filesystem Delivery
- February, 2021: GPU Cabinet Delivery
- May, 2021: Perlmutter dedication ceremony
- July 16, 2021: first users on Perlmutter
- July - October, 2021 : Initial users added in waves (NESAP, ECP, SF)
- February - April, 2022: GPU nodes cycled out for cleaning



Phase 1

January 2022





testing

not present

perlmutter

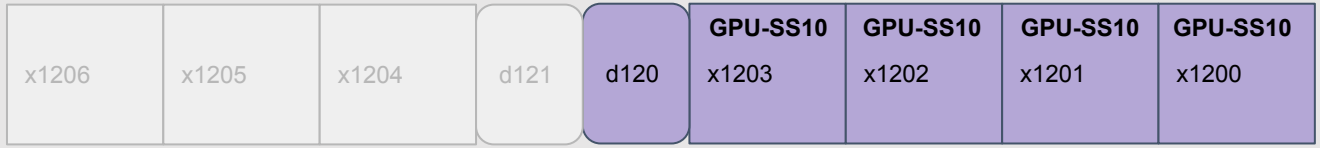
muller

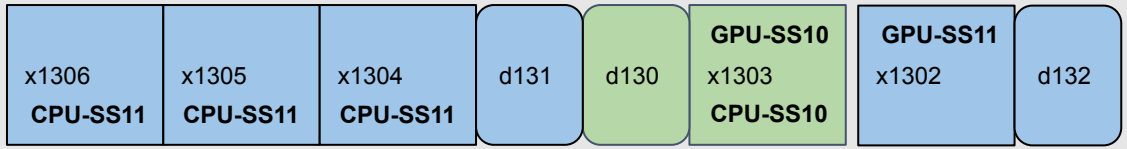
alvarez

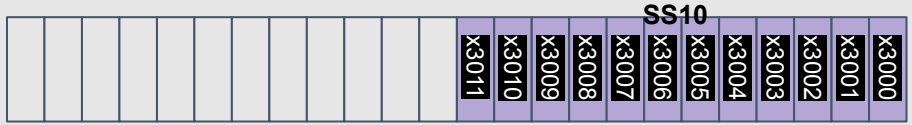
x storage/gw

NERSC

February 2022







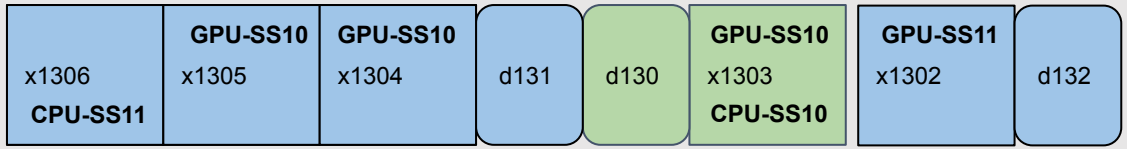
Test Systems Keep Things Stable for Users

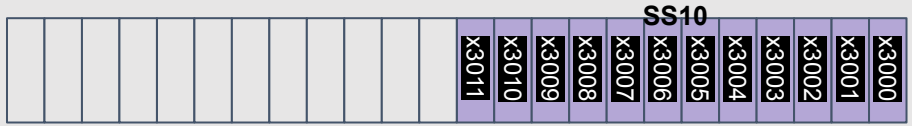
- NERSC has deployed two test systems for Perlmutter
  - Muller: pre-production, test new software and system configurations
  - Alvarez: test far-field changes that require a lot of development
- A major question before phase 2 integration was “Can SS10 and SS11 co-exist?”
  - Needed new procedures for managing hybrid system (no manual covers this because NERSC created it)
  - Hoes lustre perform with TCP/IP communication (RDMA wasn't possible between SS10 and SS11)?
  - Does the system function at all this way?!?
- Added SS10 GPU and SS11 CPU cabinets to Alvarez
  - Gave systems group a platform to set up configuration without disrupting Perlmutter
  - Early user testing to ensure that SS11 will work and can co-exist with SS10
  - Thanks to all the users who helped with testing!

Deployment Timeline Con't

- February - May, 2022: CPU cabinets delivered
- May 11, 2022: First CPU nodes opened to users
- May 17, 2022: All NERSC users added to Perlmutter
- June - October, 2022: **remanufacturing**

March 2022





testing

not present

perlmutter

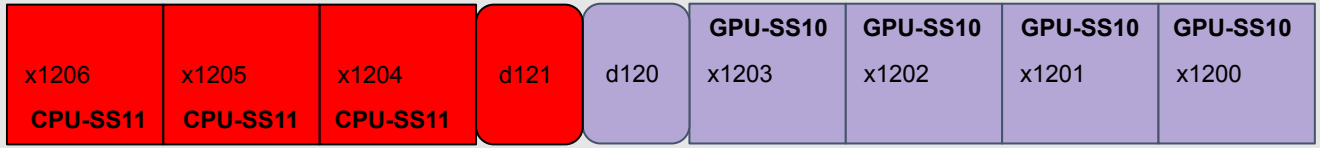
muller

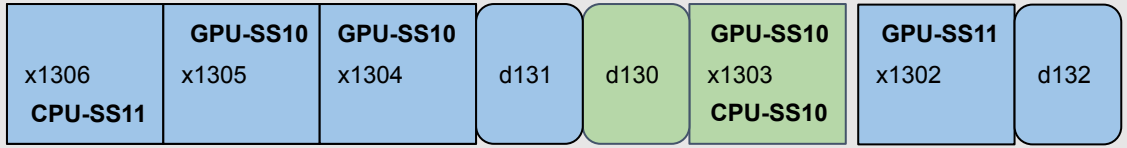
alvarez

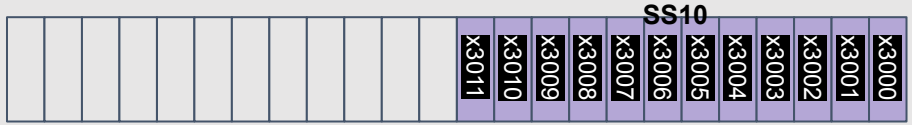
x storage/gw

NERSC

April 2022







testing

not present

perlmutter

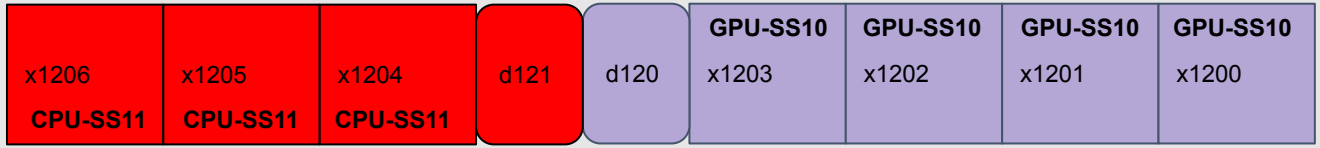
muller

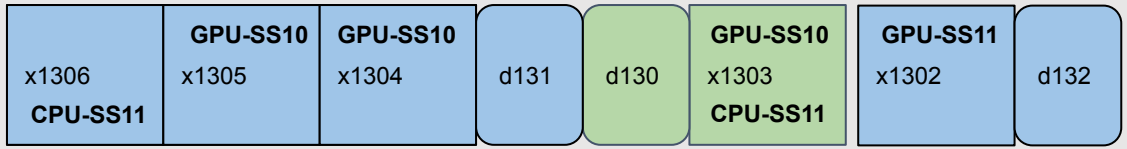
alvarez

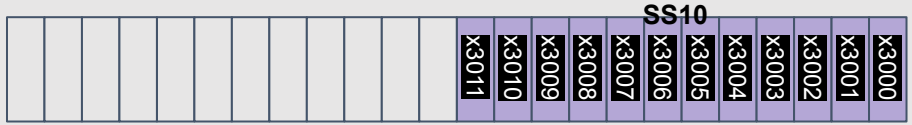
x storage/gw

NERSC

May 2022







testing

not present

perlmutter

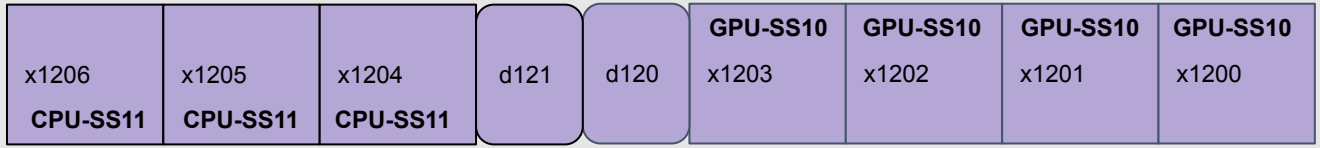
muller

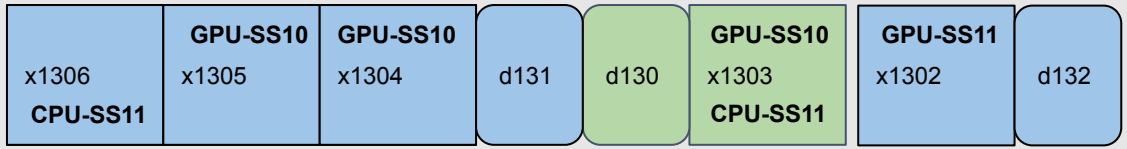
alvarez

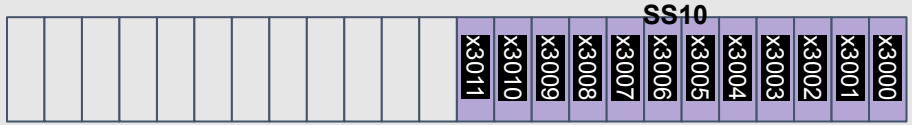
x storage/gw

NERSC

May 2022







testing

not present

perlmutter

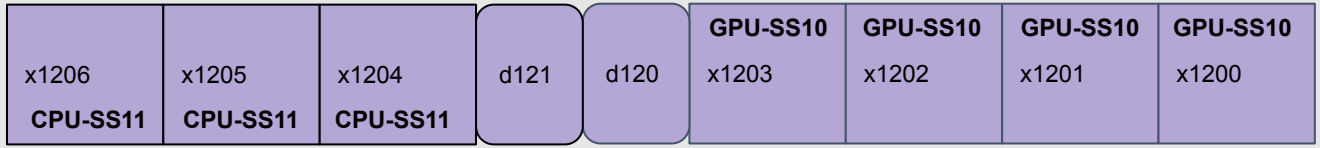
muller

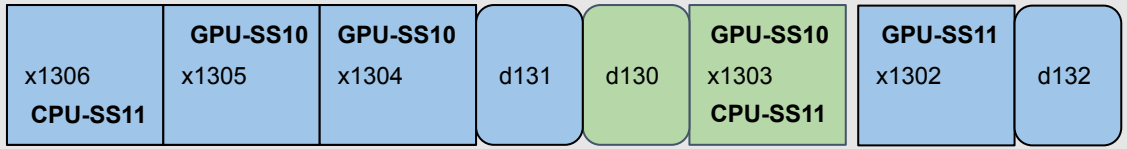
alvarez

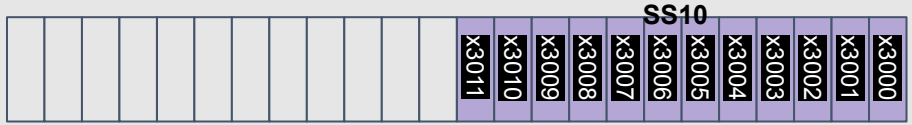
x storage/gw

NERSC

June 2022







testing

not present

perlmutter

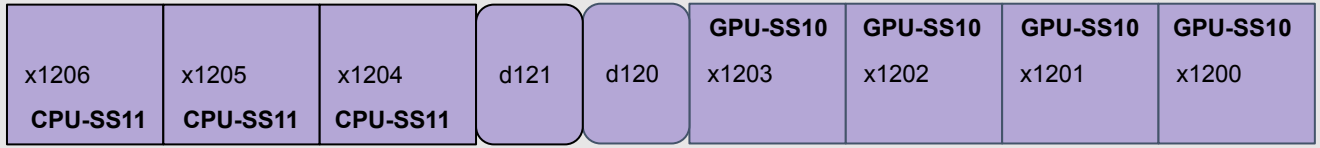
muller

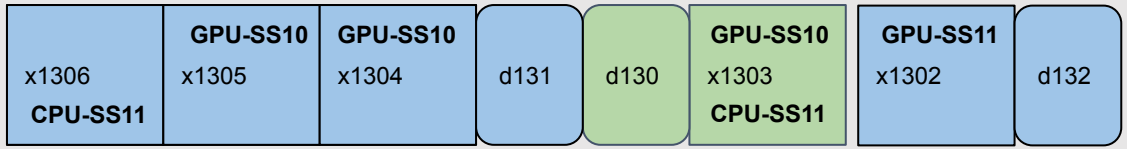
alvarez

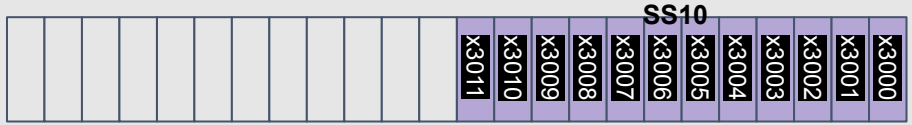
x storage/gw

NERSC

June 2022







testing

not present

perlmutter

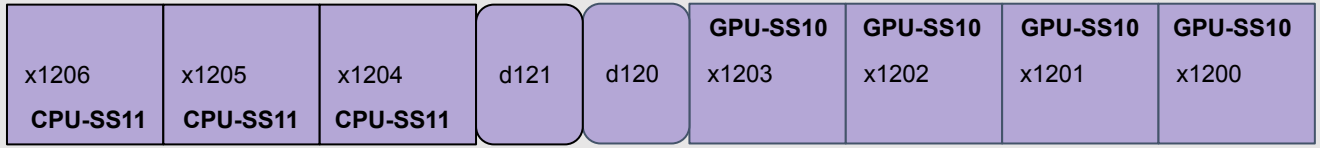
muller

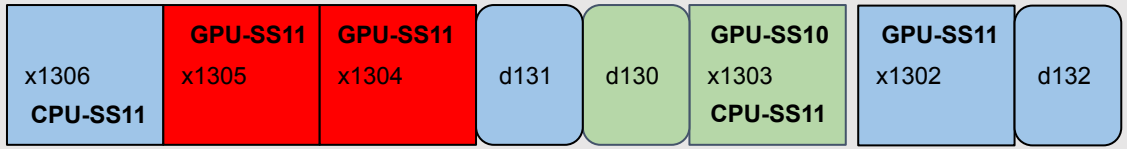
alvarez

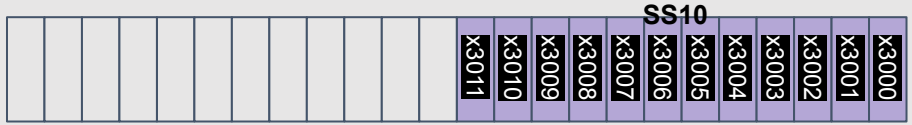
x storage/gw

NERSC

June 2022







testing

not present

perlmutter

muller

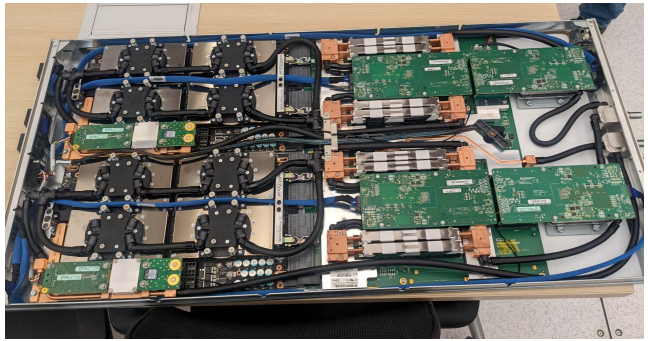
alvarez

x storage/gw

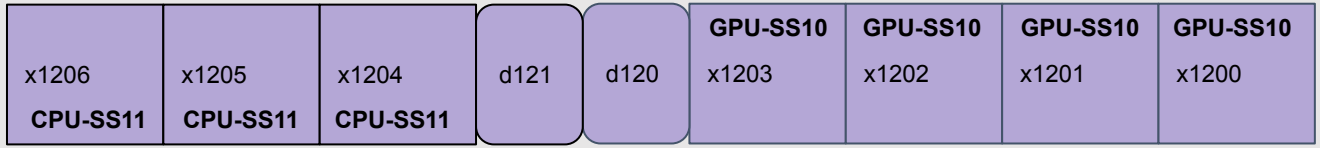
NERSC

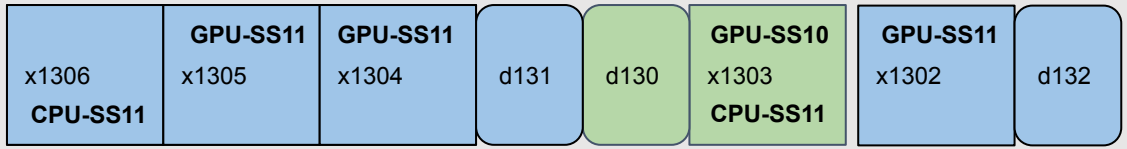
Rebuilding a Supercomputer

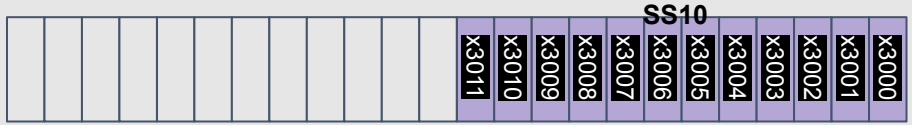
- The Slingshot 11 NIC was delayed, system was delivered with Slingshot 10
- To get to phase II, each GPU node had to be physically upgraded to the hardware configuration needed for Slingshot 11, along with a major software upgrade to fully utilize the new hardware
- The remanufacturing was done over a period of five months by a team of HPE experts
  - 788 Blades
  - 300 Storage Servers
  - 48 Login Nodes
  - 28 I/O Gateways
  - 36 Mgmt Servers



July 2022







testing

not present

perlmutter

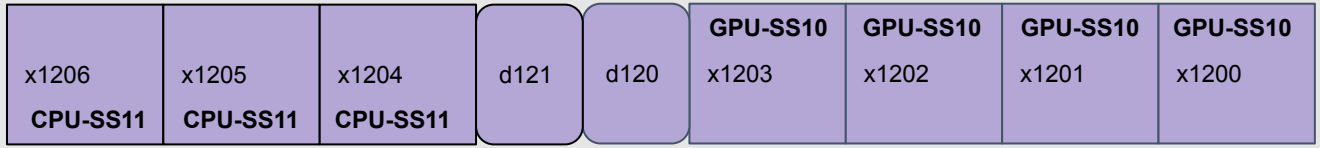
muller

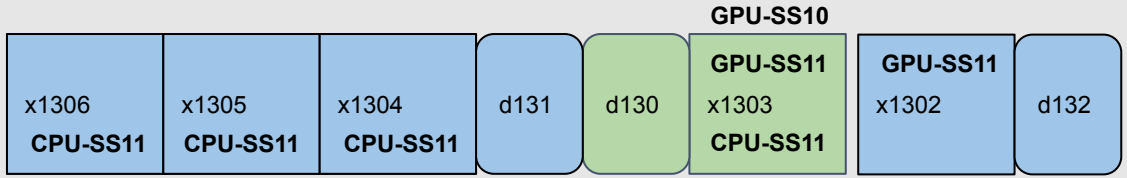
alvarez

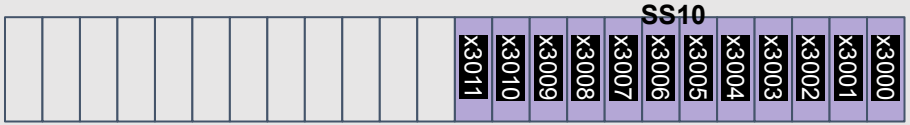
x storage/gw

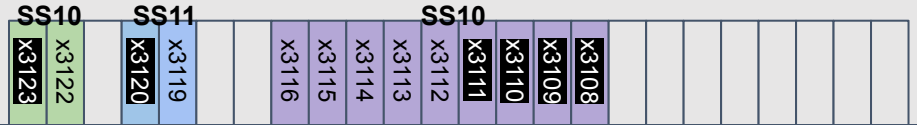
NERSC

July 2022









testing

not present

perlmutter

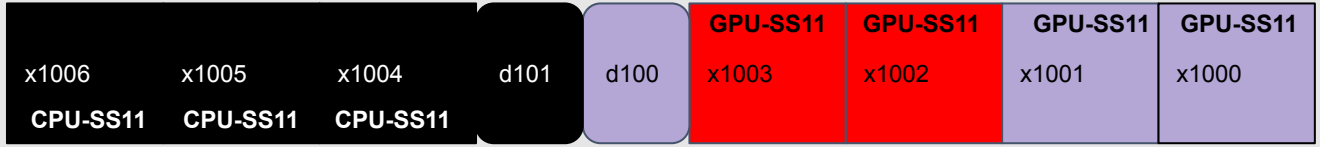
muller

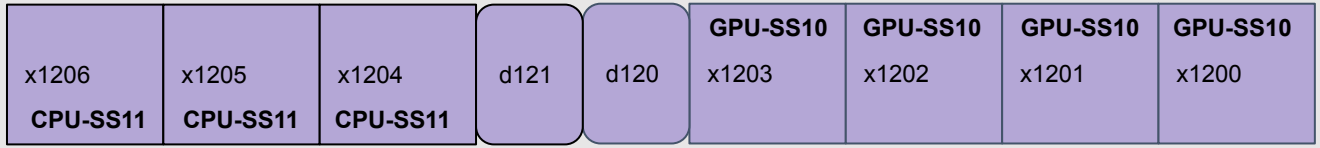
alvarez

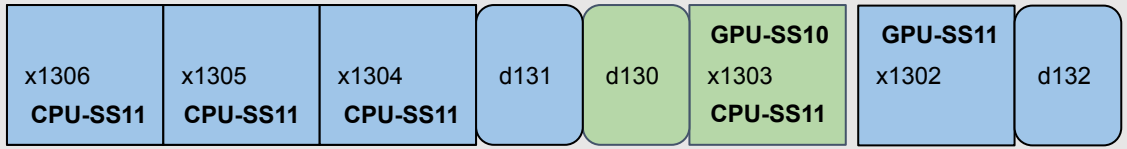
x storage/gw

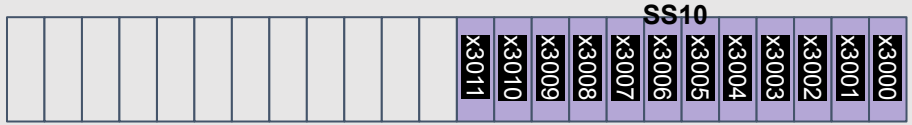
NERSC

July 2022









testing

not present

perlmutter

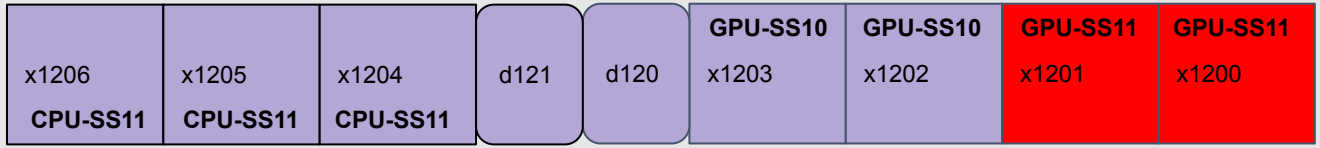
muller

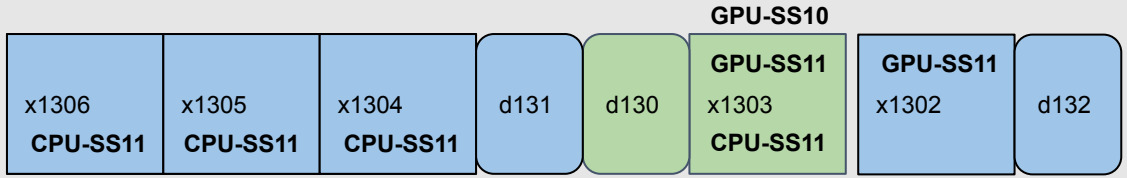
alvarez

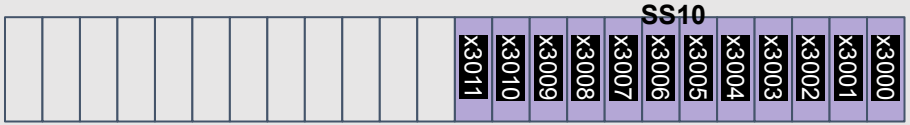
x storage/gw

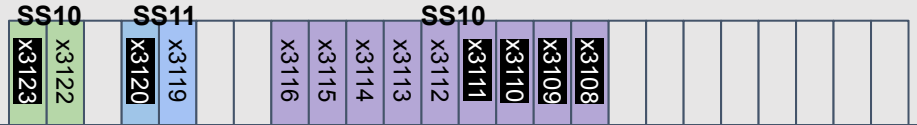
NERSC

July 2022









testing

not present

perlmutter

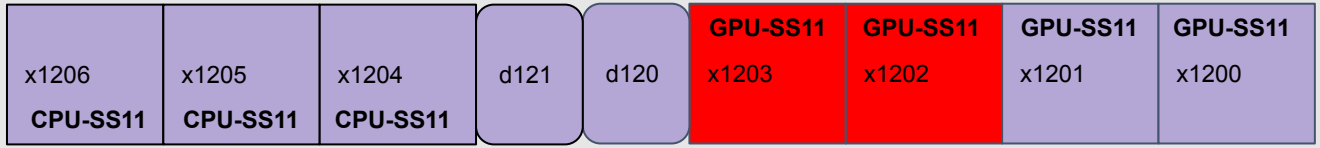
muller

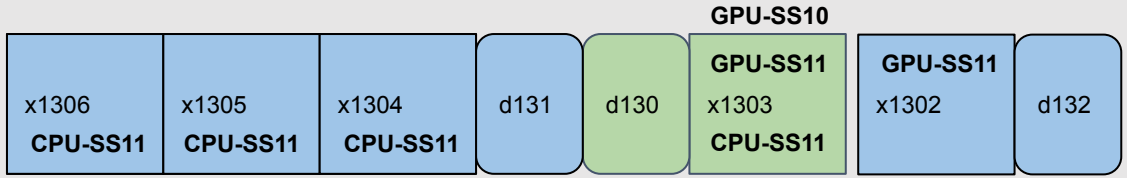
alvarez

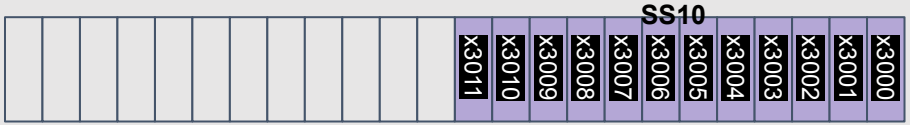
x storage/gw

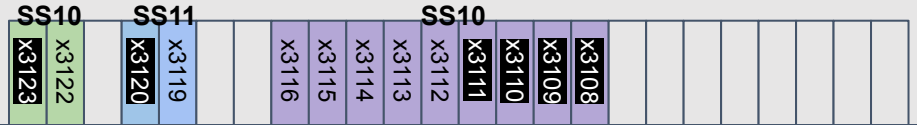
NERSC

July 2022









testing

not present

perlmutter

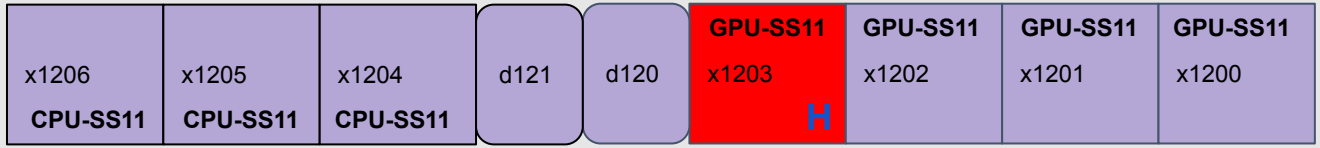
muller

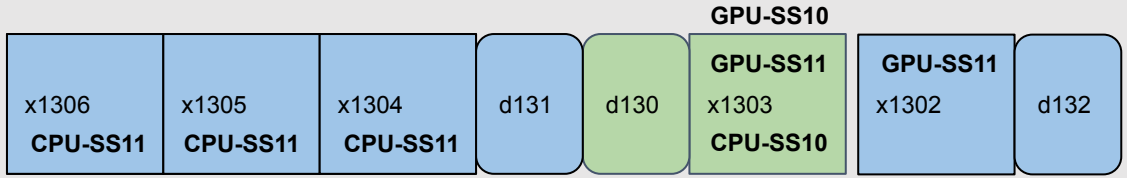
alvarez

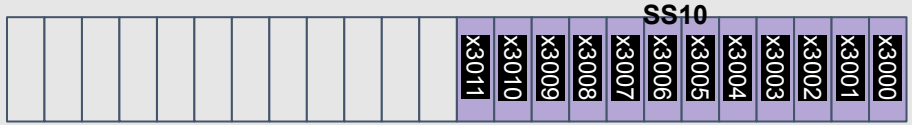
x storage/gw

NERSC

August 2022







testing

not present

perlmutter

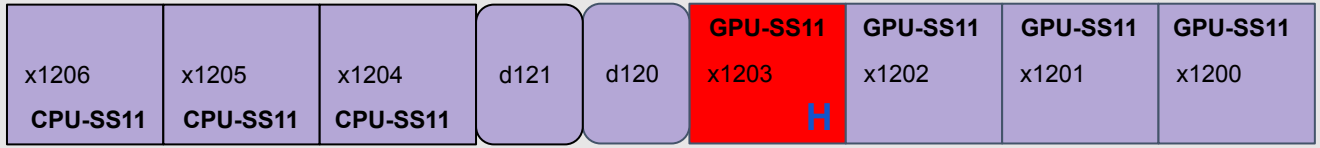
muller

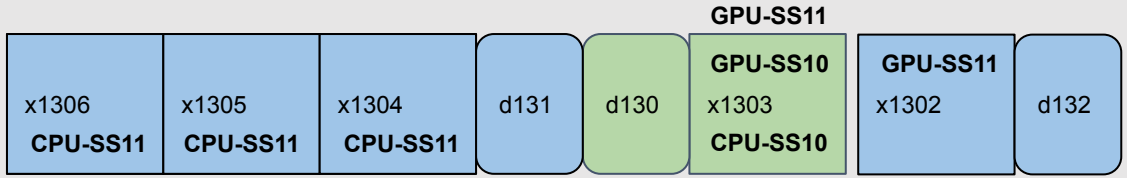
alvarez

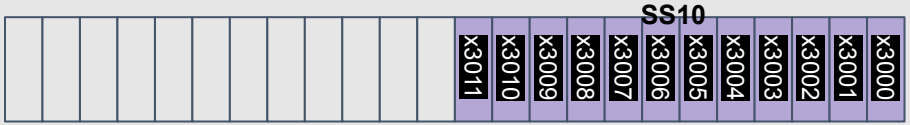
x storage/gw

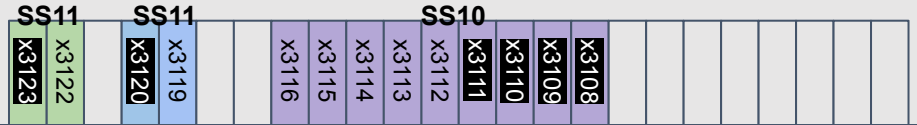
NERSC

August 2022









testing

not present

perlmutter

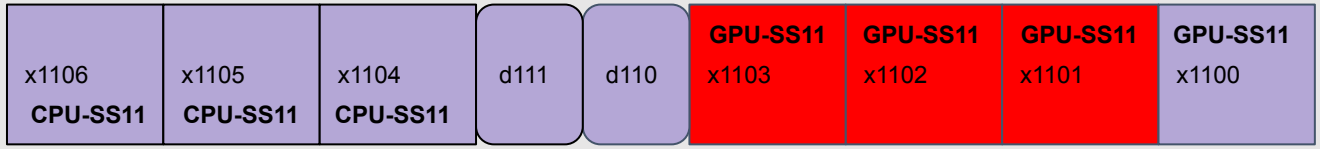
muller

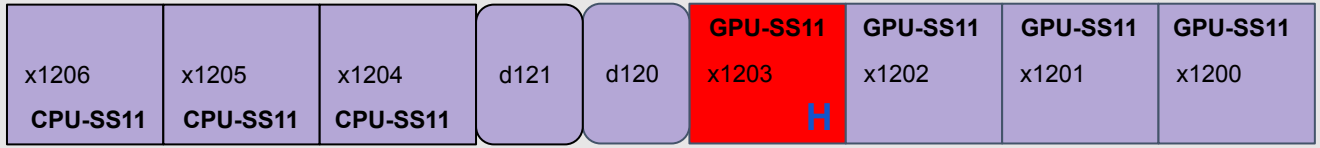
alvarez

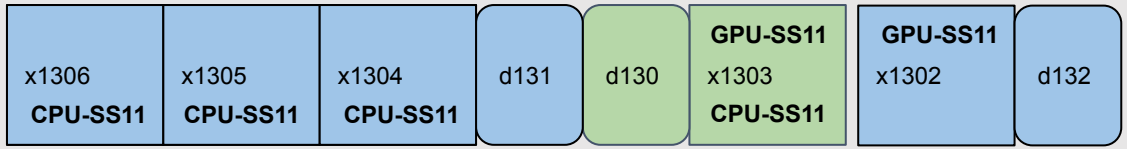
x storage/gw

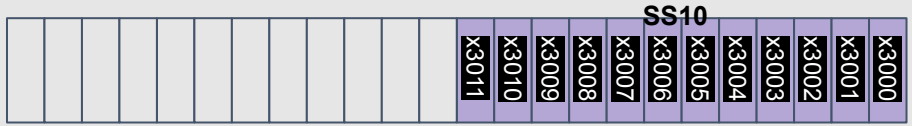
NERSC

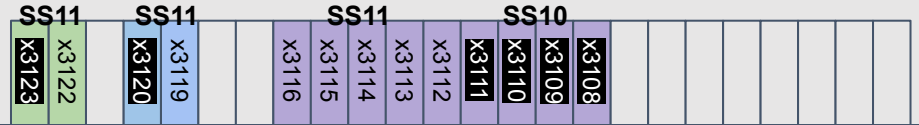
August 2022





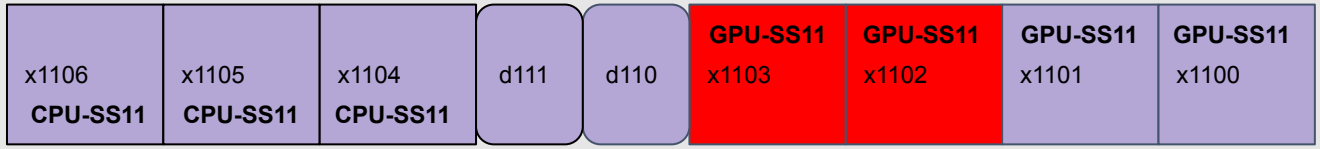


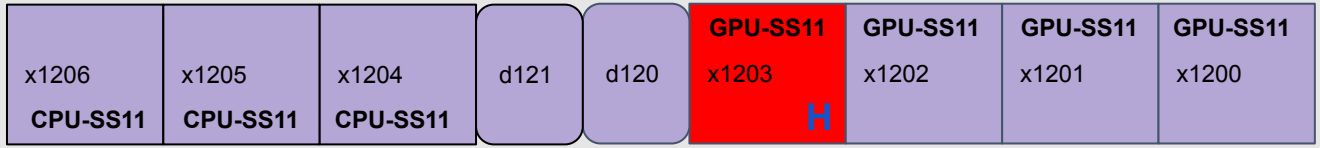


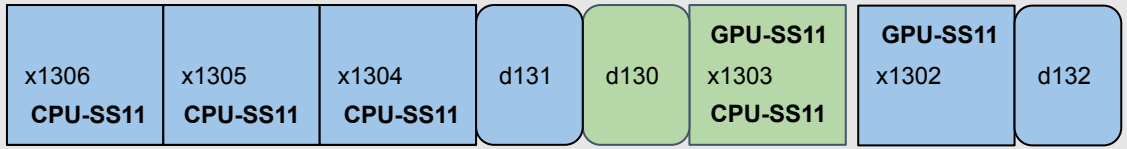


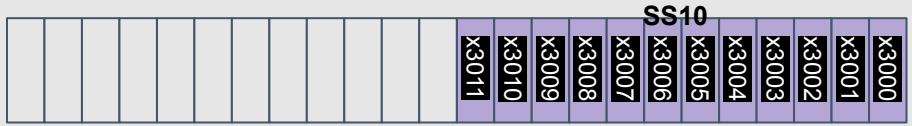
NERSC

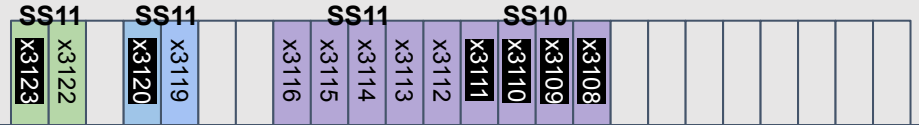
August 2022











testing

not present

perlmutter

muller

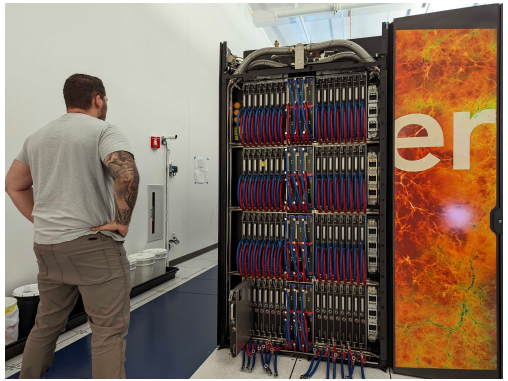
alvarez

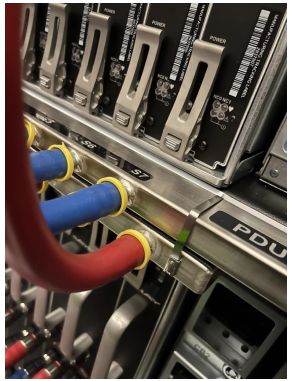
x storage/gw

NERSC

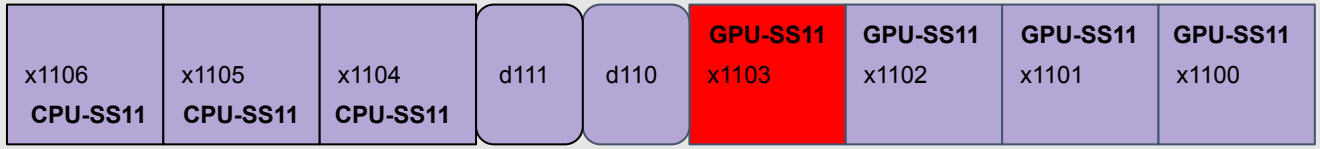
Deployment Timeline Con't

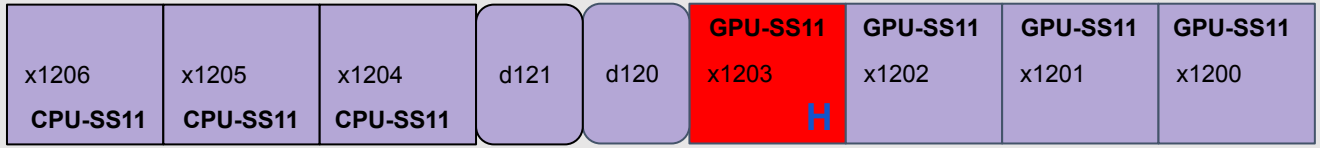
- July 11, 2022: First SS11 GPU nodes opened to users
- August 15, 2022: All SS10 GPU nodes are removed from the system
- August 15 - ~~August 23~~ September 15, 2022: All Lustre Servers upgraded from SS10 to SS11
- October 11, 2022 (now): Remanufacturing finished, I/O network protocol changed to the final phase II configuration (kflind), lustre upgraded to 2.15

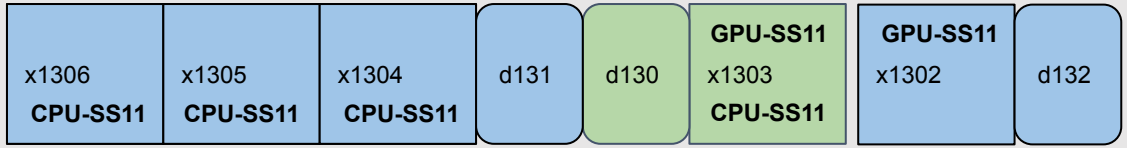


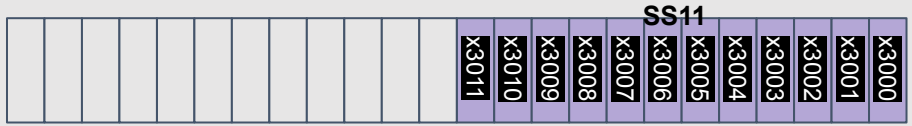


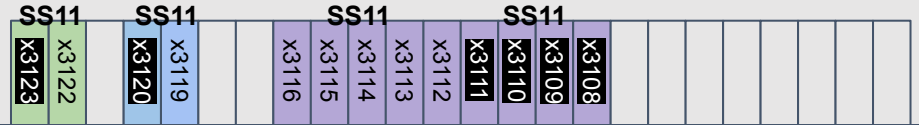
September 2022











testing

not present

perlmutter

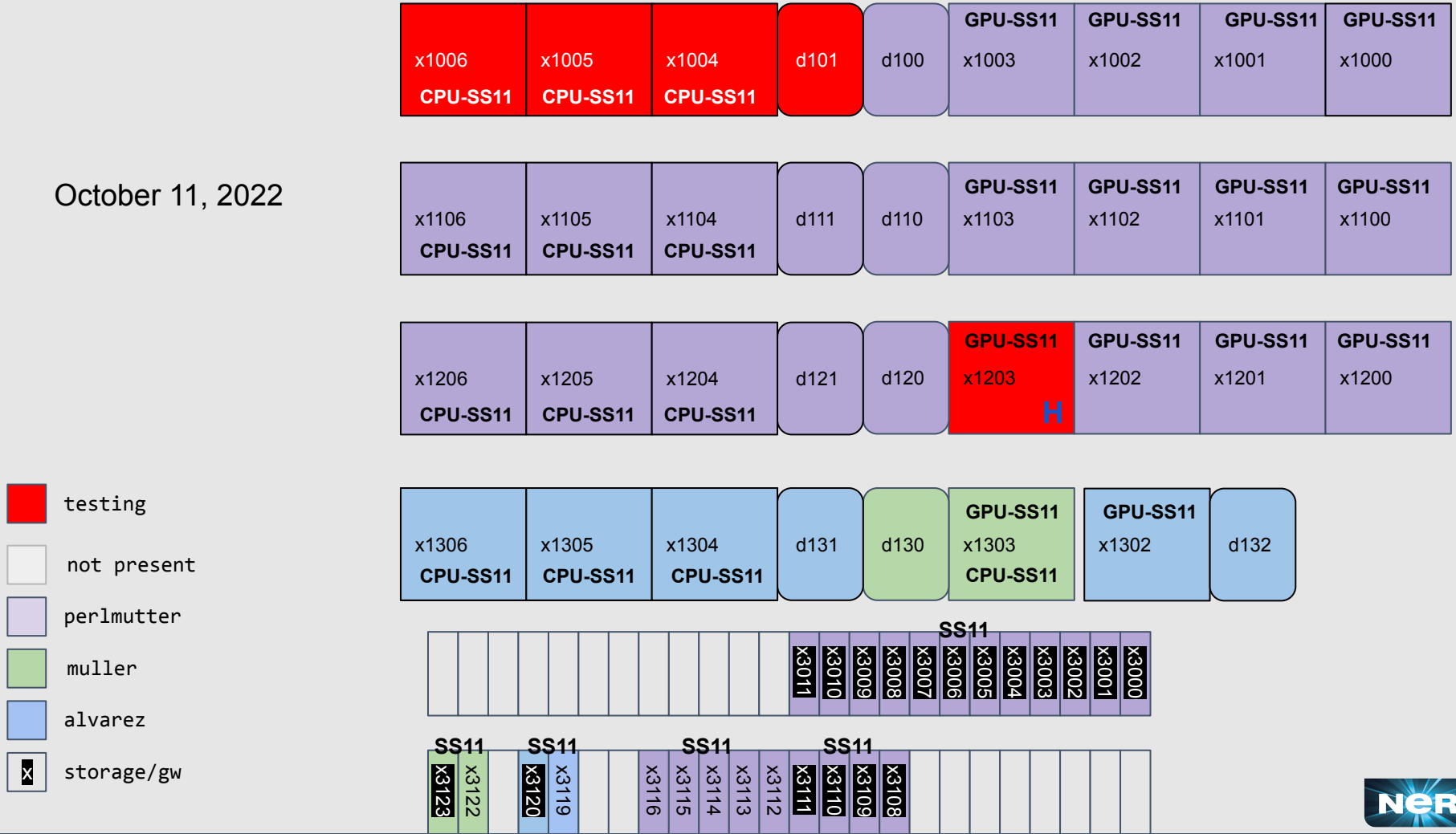
muller

alvarez

x storage/gw

NERSC

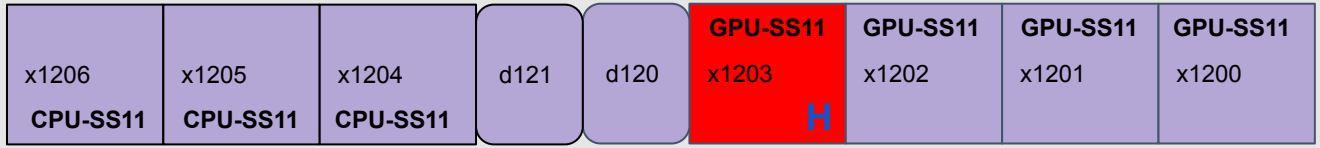
October 11, 2022

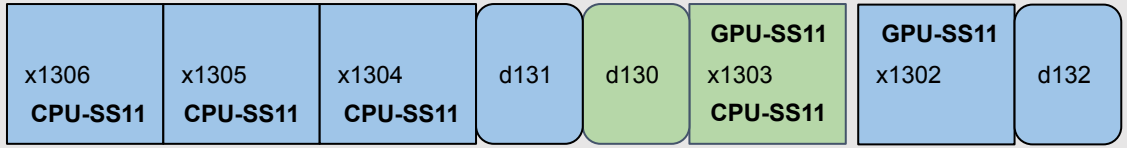


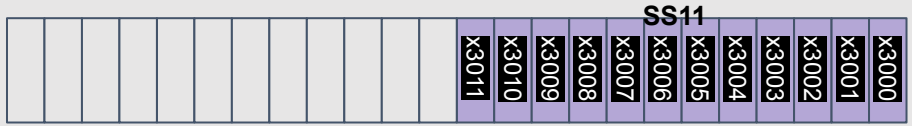
Upcoming Changes (Near Term)

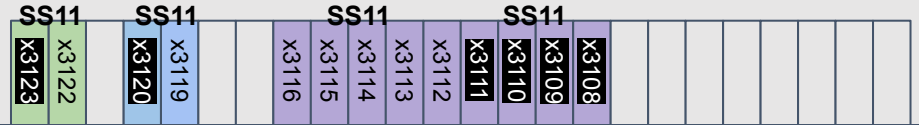
- Repair work - A lot of remanufacturing dust is beginning to settle
- Three more CPU racks (October 25, 2022)
- All login and GPU compute node repairs completed (est November, 2022)
- Fully Upgraded Software Stack (November, 2022)
  - Slingshot software upgrade to v2.0
    - Resolves many system stability and performance issues
  - SLES 15sp4 / cos 2.4
    - Major kernel update allowing multiple jobs/GPU node (new code + cgroupv2)
    - Updated NVIDIA driver/cuda
- Three more CPU racks (again) (November, 2022)
- Tuning to improve performance and stability
- Final acceptance (several interruptions for full system testing)
  - Acceptance-related disruptions primarily in December and January (estimated)

October 25, 2022









testing

not present

perlmutter

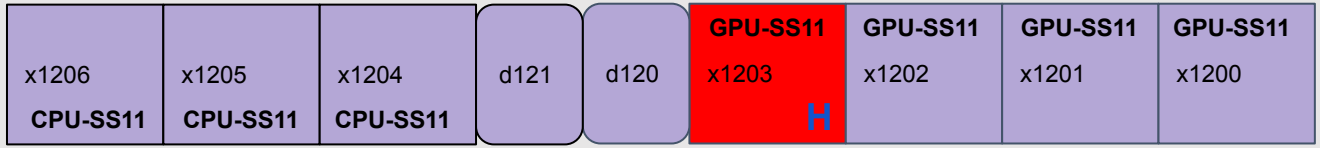
muller

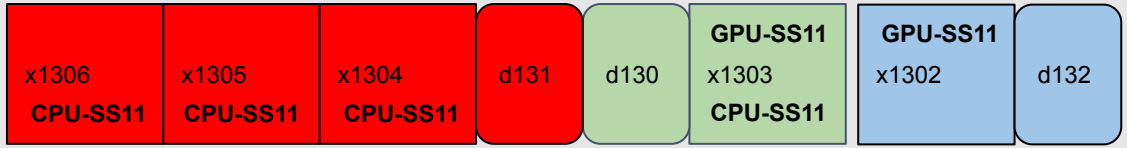
alvarez

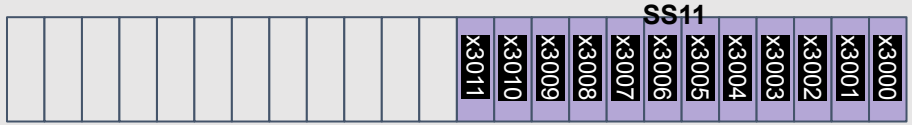
x storage/gw

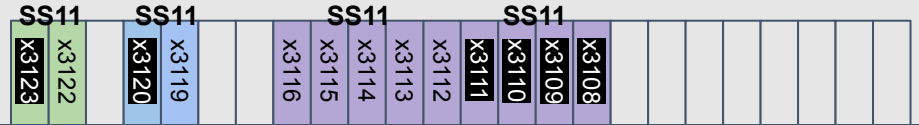
NERSC

~November 5, 2022









testing

not present

perlmutter

muller

alvarez

x storage/gw

NERSC

~November 15, 2022

testing

not present

perlmutter

muller

alvarez

x storage/gw

x1006 CPU-SS11	x1005 CPU-SS11	x1004 CPU-SS11	d101	d100	GPU-SS11 x1003	GPU-SS11 x1002	GPU-SS11 x1001	GPU-SS11 x1000
x1106 CPU-SS11	x1105 CPU-SS11	x1104 CPU-SS11	d111	d110	GPU-SS11 x1103	GPU-SS11 x1102	GPU-SS11 x1101	GPU-SS11 x1100
x1206 CPU-SS11	x1205 CPU-SS11	x1204 CPU-SS11	d121	d120	GPU-SS11 x1203	GPU-SS11 x1202	GPU-SS11 x1201	GPU-SS11 x1200
x1306 CPU-SS11	x1305 CPU-SS11	x1304 CPU-SS11	d131	d130	GPU-SS11 x1303 CPU-SS11	GPU-SS11 x1302	d132	

SS11: x3011 x3010 x3009 x3008 x3007 x3006 x3005 x3004 x3003 x3002 x3001 x3000

SS11: x3123 x3122

SS11: x3120 x3119

SS11: x3116 x3115 x3114 x3113 x3112

SS11: x3111 x3110 x3109 x3108

NERSC

Upcoming Changes (Long Term)

- Continuous Operations
  - Expect most maintenances / software updates to be done non-disruptively
  - Changes to network and Lustre are the key limiting factors right now
- User Access Instance (UAIs)
  - Each user logs in directly to a dedicated container
  - Will minimize user collisions (“The login node is slow”)
  - NERSC will provide standard images, but user customized images could also be used
  - Also capability for long-running k8s managed user services
  - Earliest likely user availability in Q2 or Q3 2023
- API-driven interactions
  - RESTful interface to Slurm
  - Start/manage/maintain gitlab runners
  - Data movement operations

So many acknowledgments

- Everybody at NERSC, but especially:
  - Computational Systems Group
  - N9 User Integration Team
  - N9 Project Management Team
  - Early User Testers
- HPE/Cray
  - Site Support Team
  - Extended-on-site Installation Team
  - N9 Project Management Team
  - HPE/Cray Engineers on Slack and Zoom
  - Cray EX Leadership Team (Hardware and Software)

Thank You