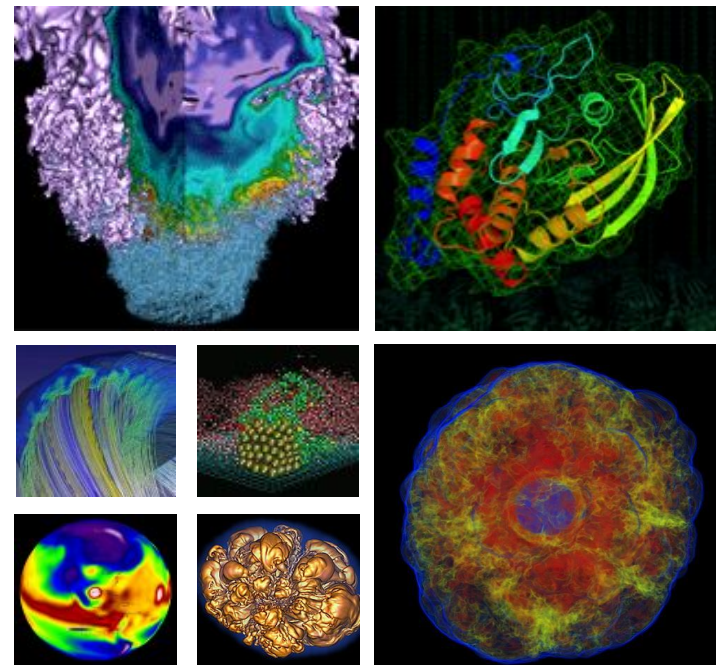


# NUG Monthly Meeting



21 January, 2021



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Today's plan

---



- Interactive - please participate!
  - Raise hand or just speak up
  - **NERSC User Slack** (link in chat), **#webinars** channel
- Agenda:
  - Win-of-the-month
  - Today-I-learned
  - Announcements/CFPs
  - Topic of the day: **Perlmutter**
  - Coming meetings: topic suggestions/requests?
  - Last month's numbers

# Win of the month

---



Show off an achievement, or shout out someone else's achievement, e.g.:

- Had a paper accepted
- Solved a bug
- A scientific achievement (maybe candidate for Science highlight, or High Impact Scientific Achievement award)
- An Innovative Use of High Performance Computing (also a candidate for an award) (<https://www.nersc.gov/science/nersc-hpc-achievement-awards/>)

Tell us what you did, and what was the key insight?

# Today I learned



What surprised you that might benefit other users to hear about?  
(and might help NERSC identify documentation improvements!)

Eg:

- Something you got stuck on, hit a dead end, or turned out to be wrong about
  - Give others the benefit of your experience!
  - Opportunity to improve NERSC documentation
- A tip for using NERSC
- Something you learned that might benefit other NERSC users

"If we knew what it was we were doing, it would not be called research, would it?" - Einstein

# Announcements and CFPs

---



Check latest weekly email

New Allocation Year starts today!

# New: NERSC 101 Appointment Type



- New to NERSC?  
Mentoring Folks that or new at NERSC?  
Questions that you were too afraid to ask?
- We added a “NERSC 101” appointment type
- To schedule: <https://nersc.as.me>

A screenshot of the NERSC appointment scheduling web interface. At the top right is the NERSC logo. Below it, a blue circle with the number "1" is followed by the text "Choose Appointment". There are four rectangular buttons stacked vertically, each with a title and "30 minutes" below it. The buttons are: "KNL Optimization", "Cori File Systems", "Containers", and "NERSC 101".

Appointment Type	Duration
KNL Optimization	30 minutes
Cori File Systems	30 minutes
Containers	30 minutes
NERSC 101	30 minutes

# AY Transition

---



- Today is AY2021!

(things users/PIs might need to do?)

# Changes to Premium QOS for 2021



- When to use premium
  - For a scientific emergency (i.e. not expiring allocation)
    - Unexpected scientific event, review that needs results right away
  - Infrequently
- Starting in AY2021, projects can use premium at **x2 charge factor until they've spent 20%** of their allocation on premium jobs
- **Then charge increases to x4**
  - NERSC may change this in response to premium usage
- **“premium” QOS is not enabled for all users by default.** PIs will be able to toggle permissions for users in their project to charge a job to premium in iris
  - Instructions:
  - <https://docs.nersc.gov/iris/iris-for-pis/#enabling-the-premium-qos-available-from-ay21>



# Topic of the day - Perlmutter



- **When will I be able to use it?**
  - What are the installation stages?
- **What will it look like?**
  - Architecture
  - Environment
- **How can I prepare?**



# Perlmutter: A System Optimized for Science



- HPE Shasta System (will be largest Shasta system when deployed)
- HPE Slingshot interconnect (first large Slingshot system)
  - High-performance, scalable, low-latency, Ethernet-compatible
- 35PB all-flash Lustre SCRATCH
- First NERSC system designed to meet needs of both:
  - Large-scale simulation
  - Data analysis from experimental facilities
- 3-4x capability of Cori
- What does all this mean?
  - Many new and unique features
  - Development uncertainties, timelines are estimates!

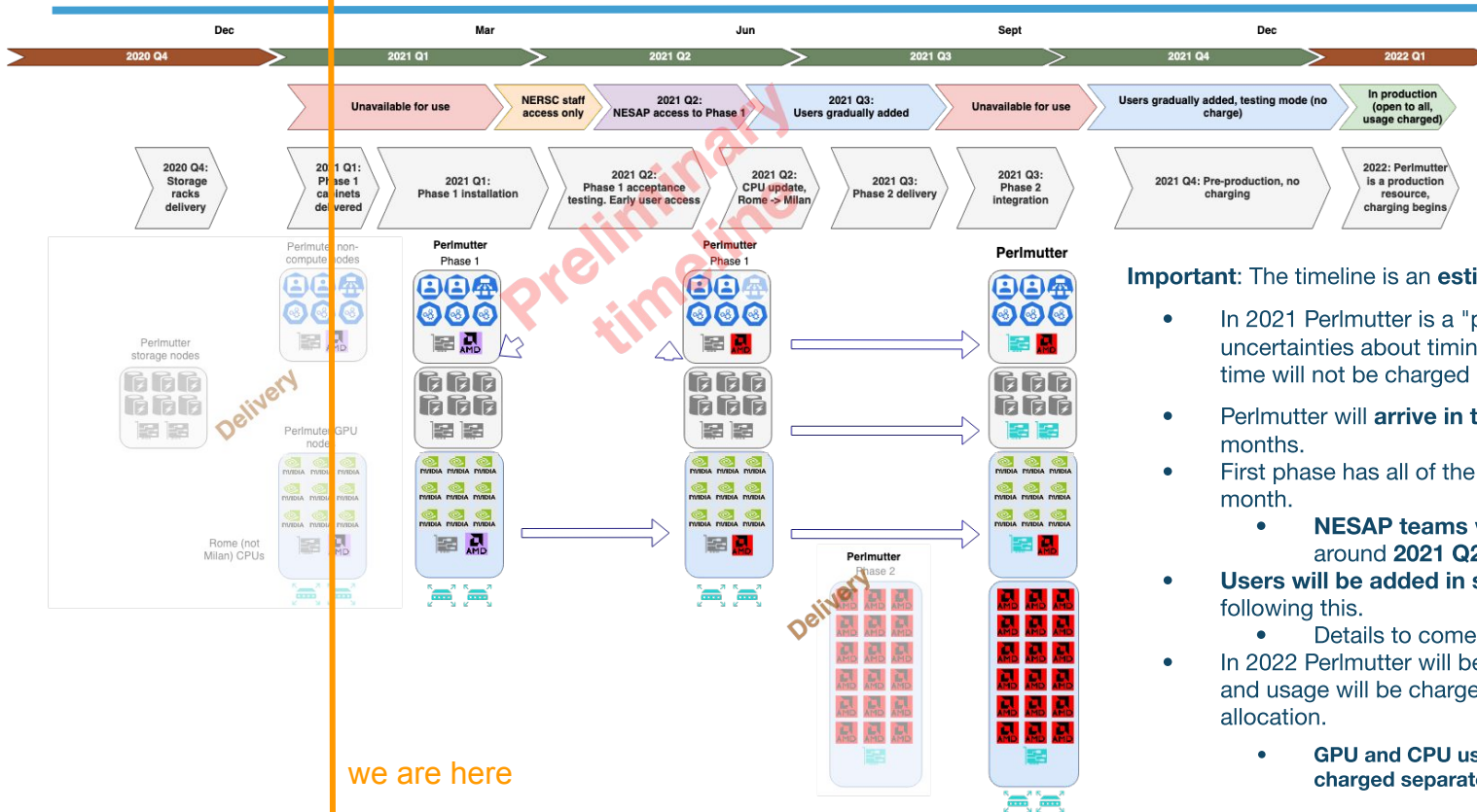
# When will I be able to use it?



**Important:** The timeline is an **estimate!**

- In 2021 Perlmutter is a "pre-production" system: uncertainties about timing and availability. Compute time will not be charged in this period.
- Perlmutter will arrive in **two phases** over the next 12 months.
- First phase has all of the GPU cabinets, arrives next month.
  - **NESAP teams** will get access to **phase 1** around **2021 Q2**.
- **Users will be added in stages** as the system matures following this.
  - Details to come
- In 2022 Perlmutter will become a production system, and usage will be charged against the project's allocation.
  - **GPU and CPU usage will be allocated and charged separately**

# Perlmutter installation stages



**Important:** The timeline is an **estimate!**

- In 2021 Perlmutter is a "pre-production" system: uncertainties about timing and availability. Compute time will not be charged in this period.
- Perlmutter will **arrive in two phases** over the next 12 months.
- First phase has all of the GPU cabinets, arrives next month.
  - **NESAP teams** will get access to **phase 1** around **2021 Q2**.
- **Users will be added in stages** as the system matures following this.
  - Details to come
- In 2022 Perlmutter will become a production system, and usage will be charged against the project's allocation.
  - **GPU and CPU usage** will be allocated and charged separately

# Perlmutter vs Cori



Slingshot allows Perlmutter to be less complex than Cori (one network with everything)

SCRATCH on Perlmutter is on the same high-speed network - smaller distance between compute and storage

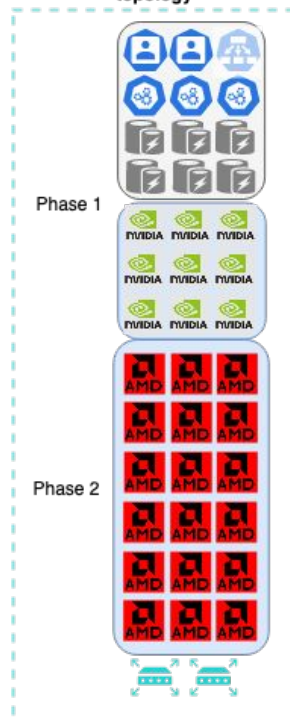
SCRATCH on Perlmutter is **internal** to Perlmutter (Cori's can be accessed from outside)

Both use dragonfly topology

Perlmutter uses containerization for better orchestration of services

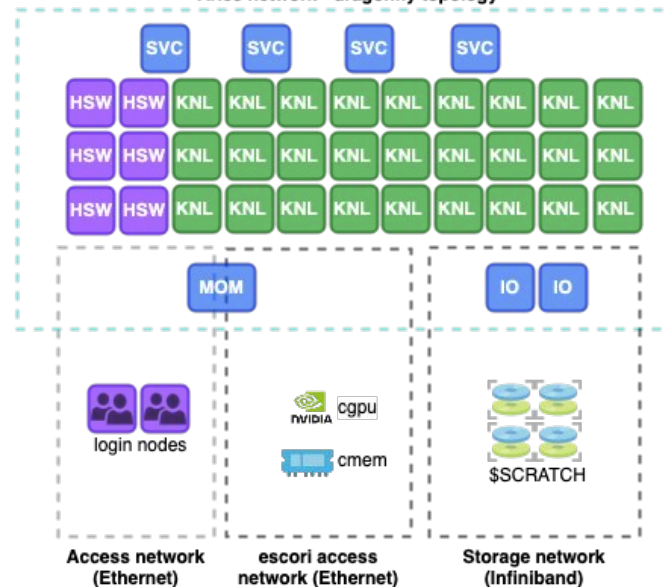
## Perlmutter

Slingshot network - dragonfly topology



## Cori

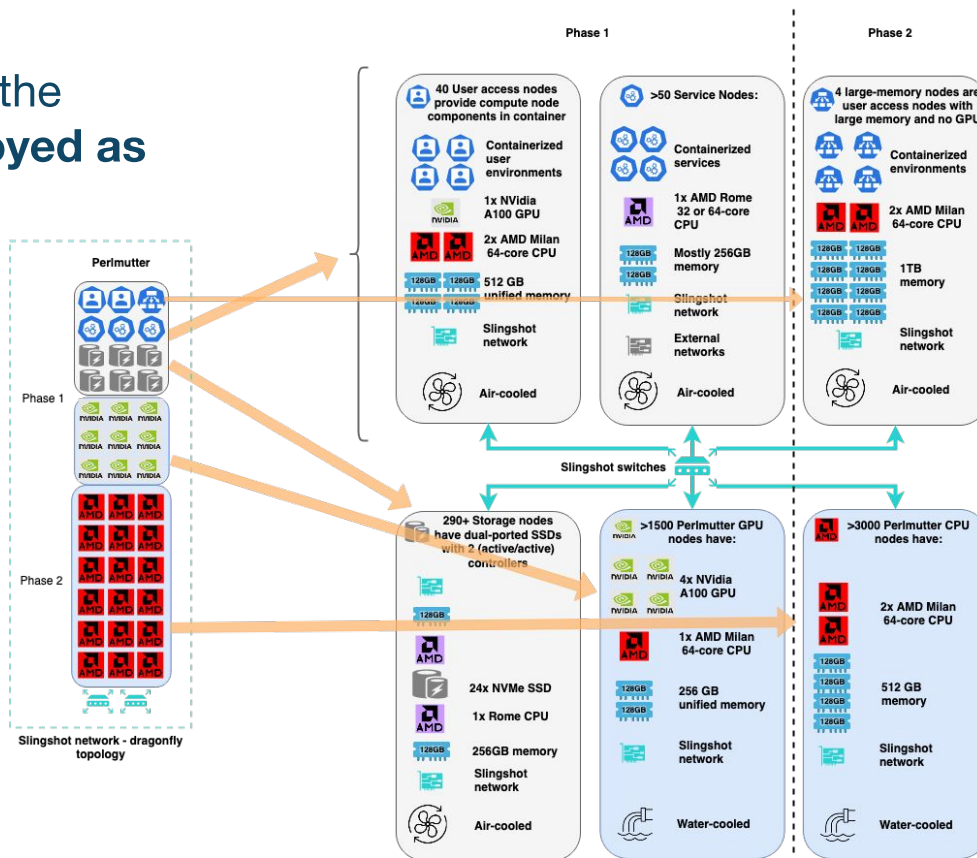
Aries network - dragonfly topology



# Perlmutter Architecture Overview



Perlmutter is a "**Shasta**" system: the non-compute **services are deployed as containers** using Kubernetes for orchestration.





# Perlmutter Architecture Overview



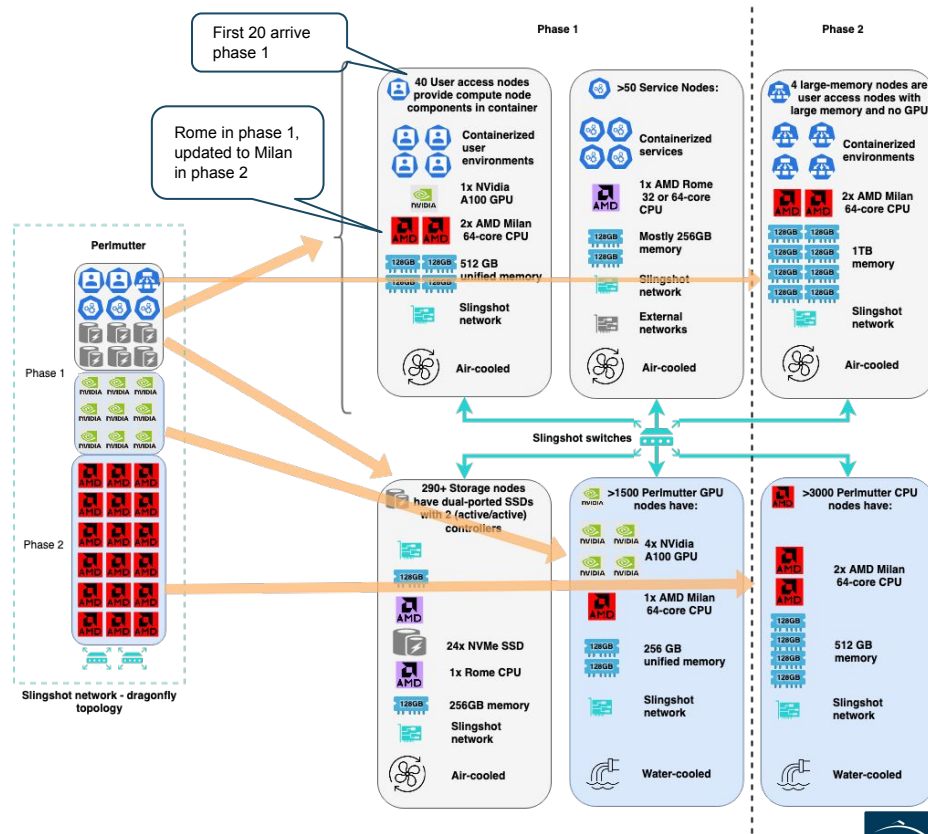
"Slingshot" network - improvements over Aries: faster, better traffic control, Ethernet-compatible

## Phase 1:

- Non-compute nodes (First 20 login nodes, service nodes)
- Storage (35 PB Lustre, all flash)
- GPU compute: 1500+ nodes with 4x NVidia A100 GPUs

## Phase 2:

- 4 Large memory nodes
- 20 more login nodes
- CPU compute: 3000+ nodes with 2x AMD 64-core CPU



# Perlmutter User Environment



Very similar to Cori!

- Access via ssh
- Same \$HOME, CFS filesystems
- Same \$SCRATCH usage model (quotas etc)
- Jupyter
- Familiar compilers and programming environments

**module load foo**

**module restore PrgEnv-cray**



# Programming Environments



	GPU Support	FORTRAN/C /C++	OpenACC 2.x	OpenMP 5.x	CUDA	Kokkos / Raja	Cray MPI
PGI							
CCE							
GNU				(Comm. Effort)			
LLVM				(Comm. Effort)			

Vendor  
Supported

NERSC  
Supported

# Preparing for Perlmutter



## 1. Read more about it! <https://docs.nersc.gov/performance/readiness/>

NERSC

NERSC Documentation

Search

GitHub/NERSC/docs  
8 Stars · 28 Forks

NERSC Documentation

Home

Getting Started

Accounts >

Iris >

Systems >

Storage Systems >

Connecting >

Environment >

Policies >

Development >

Running Jobs >

Applications >

Analytics >

Machine Learning >

Performance >

Perlmutter Readiness

Getting started on KNL

Vectorization

Parallelism

Memory Bandwidth

Arithmetic Intensity

Compiler Diagnostics

I/O >

Portability

Variability

Network

Perlmutter Readiness

✖ Write a validation test

Performance doesn't matter if you get the wrong answer!

- <https://doi.org/10.1086/342267>
- <https://doi.org/10.1109/MCSE.2017.3971169>

- Define benchmarks for performance. These should represent the science cases you want to run on Perlmutter.
- Use optimized libraries when possible (FFT, BLAS, etc).
- Start with 1 MPI rank per GPU.
- Start with UVM and add explicit data movement control as needed.

Table of contents

Perlmutter GPU information

Accelerator porting recommendations

Fortran

C++

std

CUDA

Directives

Proxy hardware platforms

Cloud providers

HPC systems

CPU

Software environment

Programming Models

Fortran

C

C++

Python

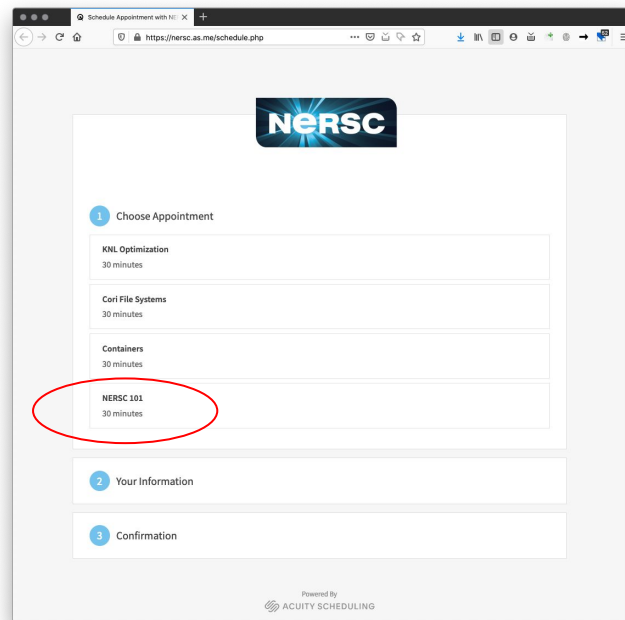
Algorithms

References and Events

# Preparing for Perlmutter



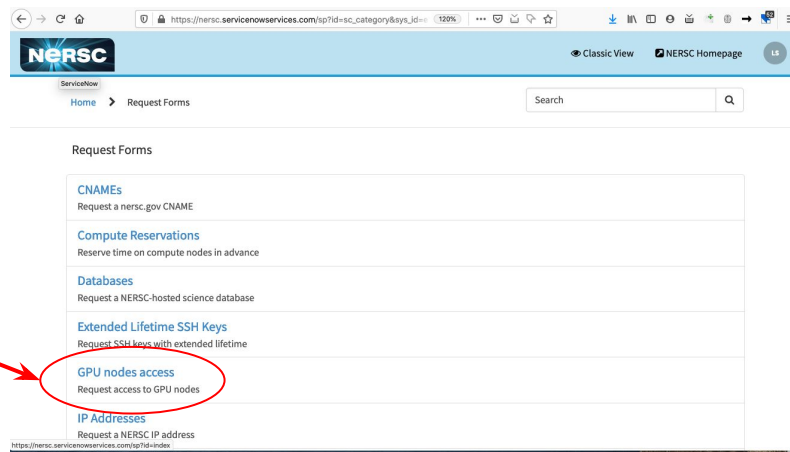
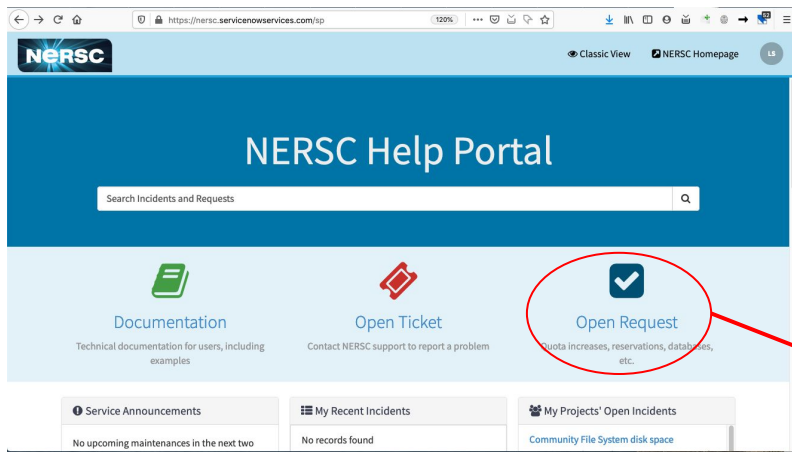
1. Read more about it! <https://docs.nersc.gov/performance/readiness/>
2. **Get training assistance:** NERSC 101 office hours  
<https://nersc.as.me/schedule.php>



# Preparing for Perlmutter



1. Read more about it! <https://docs.nersc.gov/performance/readiness/>
2. Get training assistance: NERSC 101 office hours  
<https://nersc.as.me/schedule.php>
3. **Request access to the NERSC GPU development platform**



# Coming up

---



Topic requests/suggestions?

- Eg:
  - Deep(ish) dive into NERSC's Slurm setup

We'd love to hear some lightning talks **from NERSC users** about the research you use NERSC for!

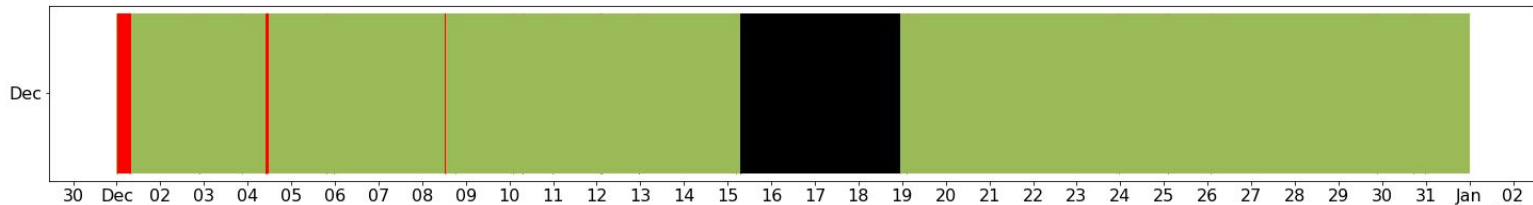
# Last month's numbers - December



Scheduled and overall availability:	Scheduled	Overall
	Cori	99.7%
	HPSS	100.0%
	CFS	100.0%

## Cori:

- 29 Unscheduled incidents (either unavailable or degraded) - total 11hrs 3 minutes
  - Mostly too short to be visible in chart
- Scheduled power work at NERSC facility - long outage



# Last month's numbers - December

---



Cori Utilization: 95.6%

Large jobs: 55.2%

New Tickets: **427**

Closed Tickets: **460**

Backlog at 1 Oct: **491**



# NERSC

**Thank You**



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

