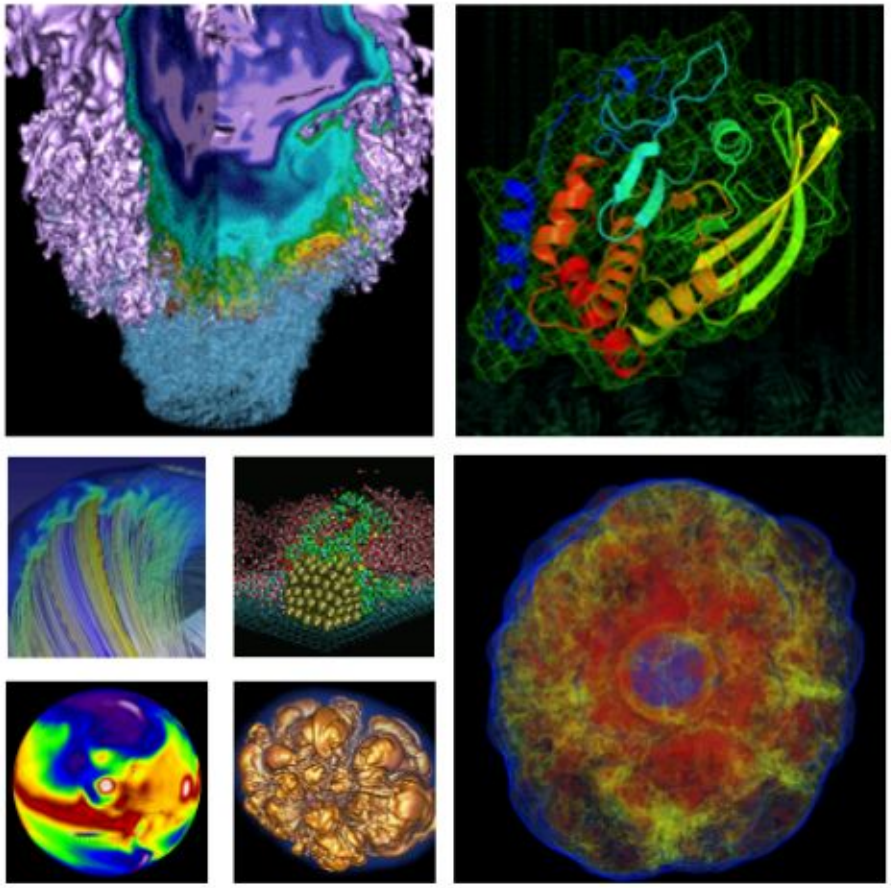
NERSC Users Group Monthly Meeting



NERSC

August 17, 2017

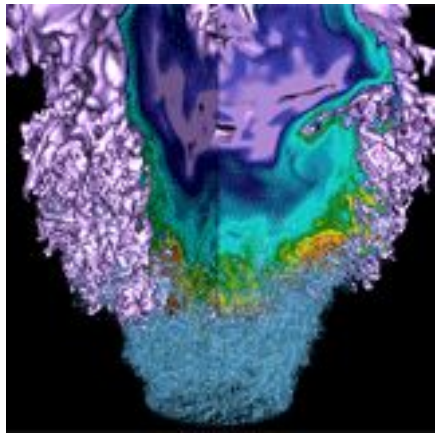
U.S. DEPARTMENT OF ENERGY

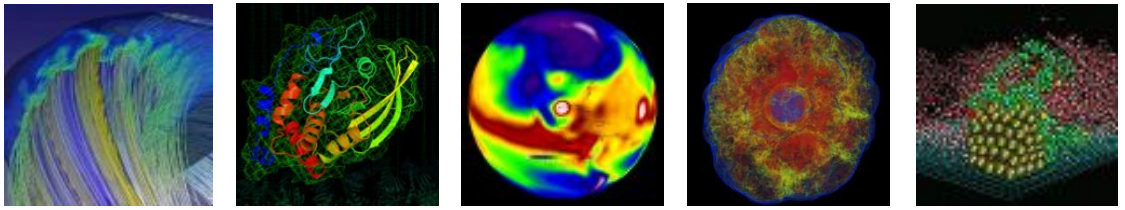
BERKELEY LAB
Lawrence Berkeley National Laboratory

Agenda

- Plans for Data Day/NUG 2017
- NERSC Achievement Awards
- Best Practices for Avoiding Performance Variability
- NERSC's New ERCAP System

Data Day/NUG 2017





https://www.nersc.gov/users/NUG/annual-meetings/nersc-data-day-and-nug2017

Data Day Updates

- **Data Day Topics**
  - Machine learning
  - Burst buffer
  - Shifter
  - Data management
  - Python
  - Esnet
- **Data Data Schedule**
  - Demo and Science Talk, Sep 19
  - Hackathon, Sep 20 AM

Data Competition

- **Wed AM, Sept 20**
- **A combined Data Day and NUG event**
- **Teams will be given data to analyze and present meaning results for an award**
- **Details to come**

NUG 2017, Wed PM, Sept 20

- **NERSC Status, Updates, & Future Plans**
  - **The view from DOE Advanced Computing Research**
  - **NERSC Status and Initiatives**
  - **Innovations Past and Future on Cori and Edison NERSC Storage 2020 Vision and Roadmap**
  - **NERSC in the Era of Big Data**
  - **Allocations and Accounting Modernization**
  - **High Performance Computing and Cybersecurity Listening to Users: Exascale Requirement Reviews and the NERSC Annual Survey**

NUG 2017, Thurs, Sept 21

- **NERSC Science and Technology Day**
  - Welcome from NERSC (Sudip Dosanjh)
  - NERSC High Impact Science (Richard Gerber)
  - Preparing Users for Cori KNL (Helen He)
  - NESAP II: Next Steps in Application Performance and Portability (Jack Deslippe)
  - ECP Science Applications (lunch time Plenary talk, Doug Kothe, ORNL)
  - 4 NERSC Achievement Awards Presentation and Talks
  - Lightning talks from large-scale Cori runs
  - UPC++ with the new V1.0 code -- Brian Van Straalen

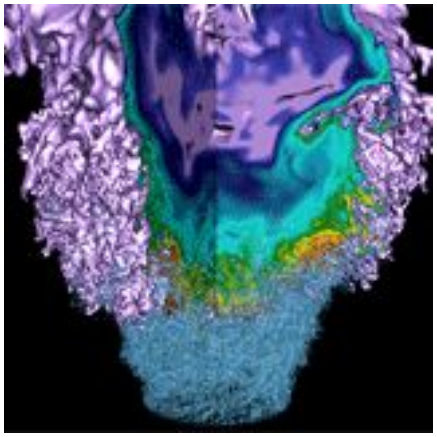
NUG 2017, Thurs, Sept 21

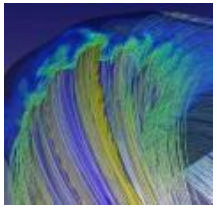
- **Lightning talks from large-scale Cori runs**
  - **HPX (Thomas Heller)**
  - **Deep Learning at 15 PFlops (Thorsten Kurth)**
  - **Galactos: Computing the 3-pt Anisotropic Correlation Function for 2 Billion Galaxies (Debbie Bard)**
  - **Large-Scale GW Calculations on Pre-Exascale HPC Systems (Mauro Del Ben)**
  - **Ultrahigh-order Maxwell solver with extreme scalability for electromagnetic PIC simulations of plasma (Jean-Luc Vay)**
  - **NWChem (Bert de Jong)**
  - **Hipmer (Rob Egan)**
  - **Julia and Celeste (Jeff Regier)**

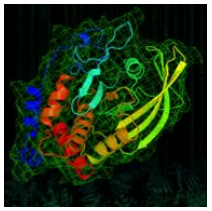
NUG 2017, Thurs, Sept 21

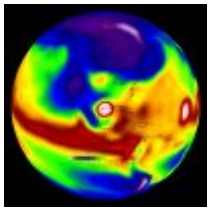
- **Data Day and NUG2017 info and registration at: https://www.nersc.gov/users/NUG/annual-meetings/nersc-data-day-and-nug2017**

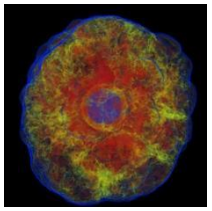
NERSC Achievement Awards

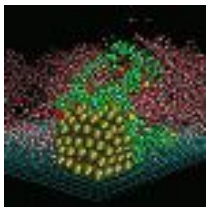












U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

NERSC HPC Achievement Awards

- **Annual awards to recognize extraordinary users of NERSC HPC resources.**
  - NERSC Award for Innovative Use of High Performance Computing
  - NERSC Award for High Impact Scientific Achievement
  - Each award has Open Category and Early Career Category
  - Total of 4 awards
- **Awards to be presented at annual NUG meeting in Sept**
- **Help us with nominations please! Details and nominations forms are available at**
  - http://www.nersc.gov/science/nersc-hpc-achievement-awards/

Early Career Award Alumni

David Cohen-Tanugi
Cofounder at EMBR labs (MIT)

Tanmoy Das
Assistant Professor of Physics, Indian Institute of Science (LANL)

Edgar Solomonik
Assistant Professor of Computer Science, University of Illinois (UC Berkeley)

Victor Ovchinnikov
Research Fellow of Computational Chemistry and Biophysics, Harvard University (Harvard)

Anubav Jain
Research Scientist in Chemistry, Lawrence Berkeley National Lab (LBNL)

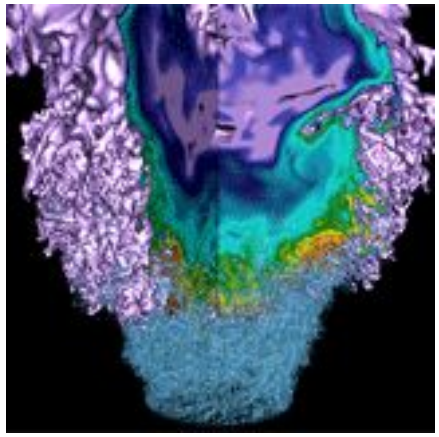
Ken Chen
Research Fellow, National Astronomical Observatory of Japan (UC Santa Cruz)

Taylor Barnes
Grace Hopper Postdoctoral Fellow, Lawrence Berkeley National Laboratory (Caltech)

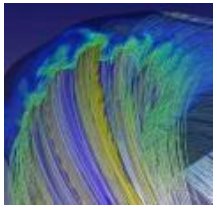
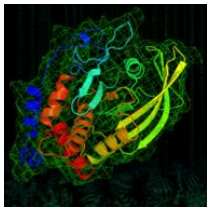
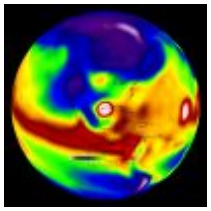
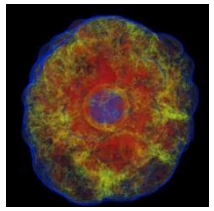
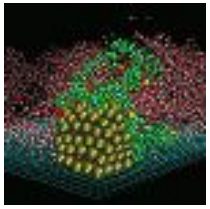
Nathan Howard
Research Scientist, MIT Plasma Science and Fusion Center (MIT)

Scott French
Google, Inc. (UC Berkeley)

Best Practices for Avoiding Performance Variability



NERSC

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

Many potential sources

- **Many potential sources of variability on HPC systems**
- **Will cover some best practices NERSC has identified to mitigate variability and improve performance**
  - Hugepages
  - Executable location
  - Topology
  - KNL modes
  - Process and thread affinity
  - Core specialization
  - I/O

hugepages

- Can reduce cost of accessing memory
- Especially important for codes with many MPI_Alltoall operations

1. `module load craype-hugepages2M`
2. Recompile code
3. Add `module load craype-hugepages2M` to script

See: `man intro_hugepages` on cori/edison for details

Executable location

Copy executables to compute node memory

- **Greatly improve job startup time**
- **Can help variability**

```
sbcast -f  --compress ./my_program.x /tmp/my_program.x
srun -n 1024 -c 2 --cpu_bind=cores /tmp/my_program.x
```

For applications with dynamic executables and many shared libraries shifter can help.

Job topology

- Edison/Cori use a Cray Aries network with dragonfly topology
- SLURM has some options to control job placement
- Can increase wait to start job
- Most effective for NNODES < 300

```
#SBATCH -N 256
#SBATCH --switches=2
```

1 switch = 384 nodes = 1 Aries group

KNL cache mode

- Quadrant cache mode is NERSC default on Cori
- The MCDRAM cache is direct mapped
- Cache size is 16GB
- Best and most consistent performance:

memory used per node less than 16GB

Running jobs

- **Use at least 8 MPI ranks per node**
  - Read `man intro_mpi`
- **Check job script generator**
  - https://my.nersc.gov/script_generator.php
- **Check MPI and OpenMP affinity**
  - `check-mpi.*` test codes
- **Add `#SBATCH -S 1` to dedicate a core to OS**
- **Check your I/O practices**

http://www.nersc.gov/users/storage-and-file-systems/i-o-resources-for-scientific-applications/

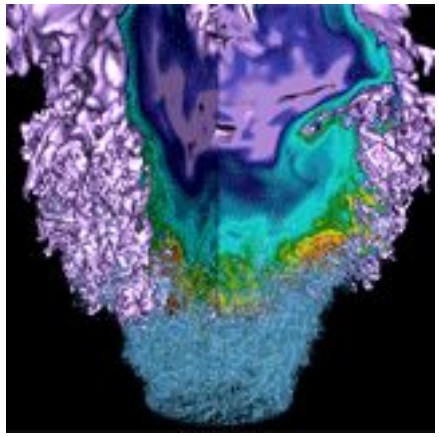
Summary

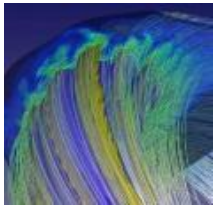
```
#!/bin/bash
#SBATCH -N 2
#SBATCH -C knl,quad,cache
#SBATCH -p regular
#SBATCH -t 60
#SBATCH -S 4
#SBATCH --switches=1
export OMP_NUM_THREADS=8
export OMP_PLACES=threads
export OMP_PROC_BIND=spread
module load craype-hugepages2M
sbcast -f  --compress ./my_program.x /tmp/my_program.x
srun -n 16 -c 32 --cpu_bind=cores /tmp/my_program.x
```

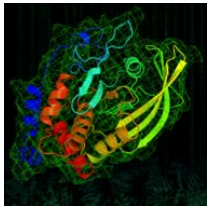
https://www.nersc.gov/users/application-performance/performance-variability/

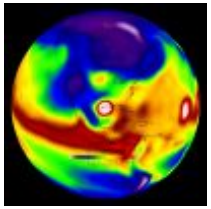
https://my.nersc.gov/script_generator.php

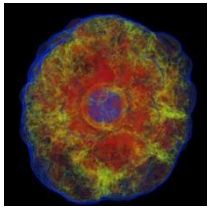
NERSC's New ERCAP System

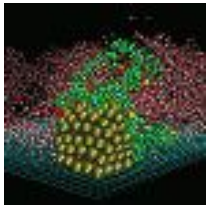












U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

Outline

- What's changed?
- What's the same?
- Brief Demo

What's Changed?

- **ERCAP has been moved to our Help Desk system**
- **ERCAP system was rewritten in ServiceNow**
  - Cloud based system
  - Modern technology compared to NIM
  - Easier maintenance and upgrade capabilities
- **Automates the End-to-End submission, review, approval, allocation process**
  - Users, DOE & NERSC managers all have access to the ERCAP requests through the same system
  - Provides a "Status" bar to show where the request is in the review/approval process

What's the Same?

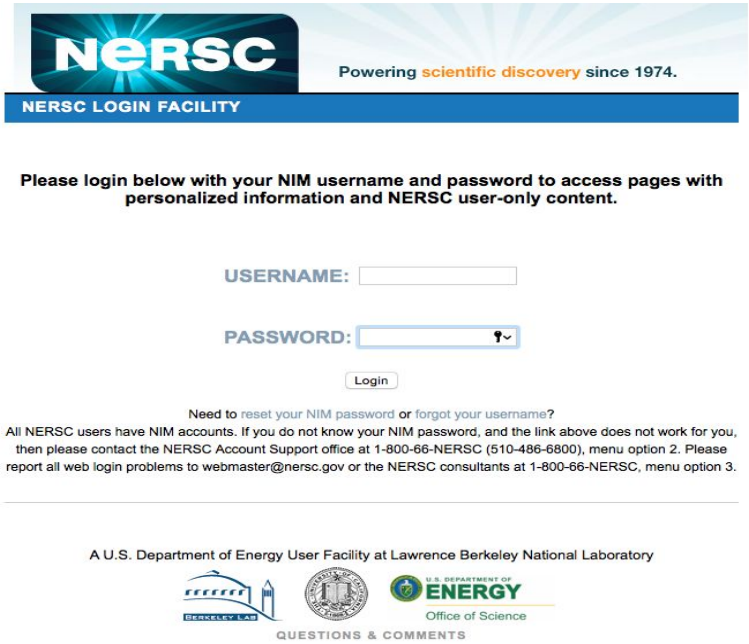
- **The questions are the same**
  - Eliminated the "tabs" format
  - More "completeness" checking
  - Only need to provide info on "Top 5" codes
- **The process remains the same**
  - PI or Proxy can initiate/complete the request
  - Initial review by NERSC before approval by DOE
  -

Brief Demo

Log into NERSC Help Desk using:

- **https://ercap.nersc.gov**

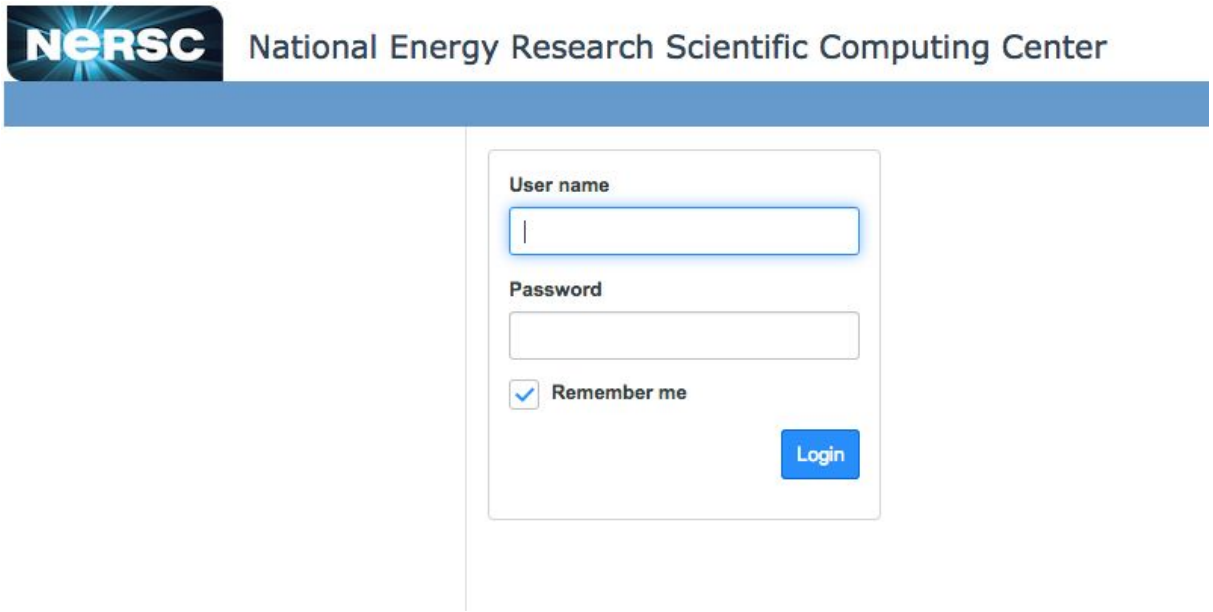
Users with NIM accounts will use shibboleth:



Demo

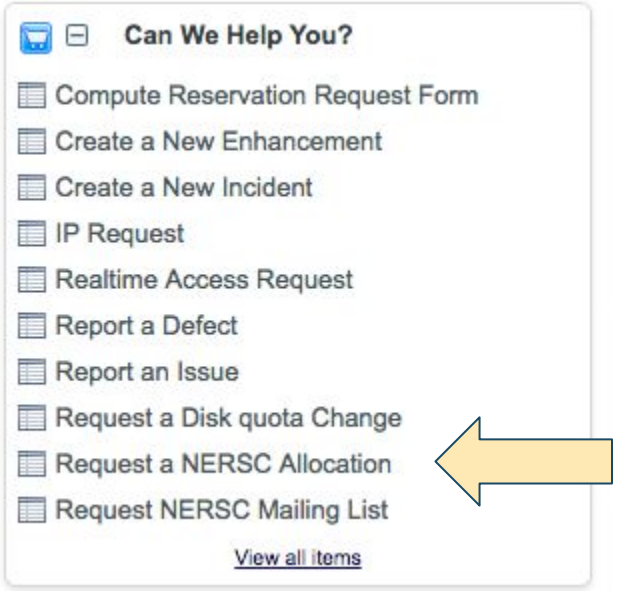
New PIs/Proxies who do not yet have NIM accounts will sign on with a "local" account

- **Login info will be provided by NERSC Account Support**



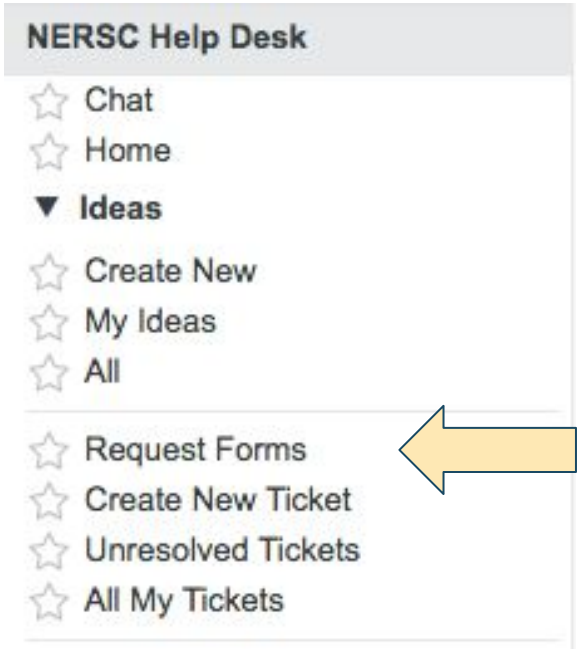
Demo

The link to submit an ERCAP request is in the "How Can We Help You?"menu on the Self Service Homepage:



Demo

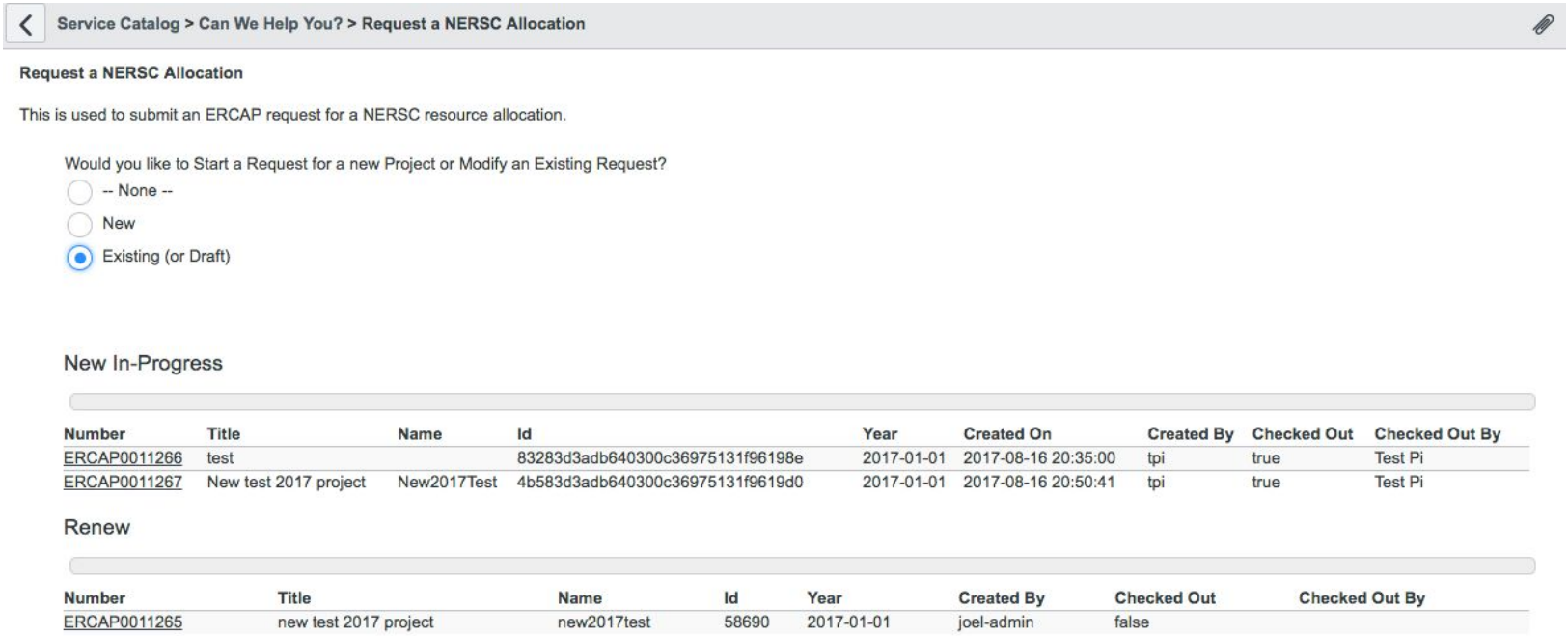
You can also get to this menu from the "Request Forms" link on the Modules list in the left-hand panel:



DEMO

"Existing" displays requests in progress and requests for current projects to be renewed

- Can be checked out by one person at a time




NERSC