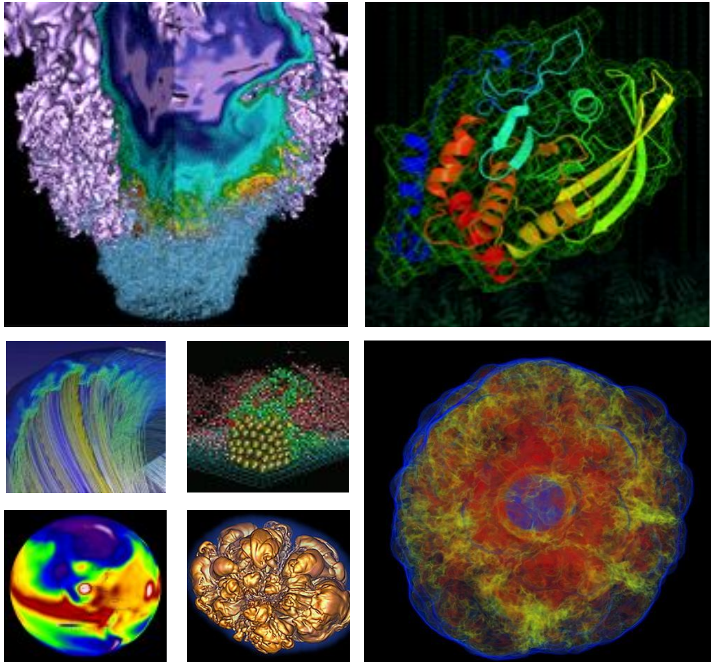
NUG Monthly Meeting



NERSC

25 May, 2023

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

Today's plan

- Interactive - please participate!
  - Raise hand or just speak up
  - NERSC User Slack (link in chat), **#webinars** channel

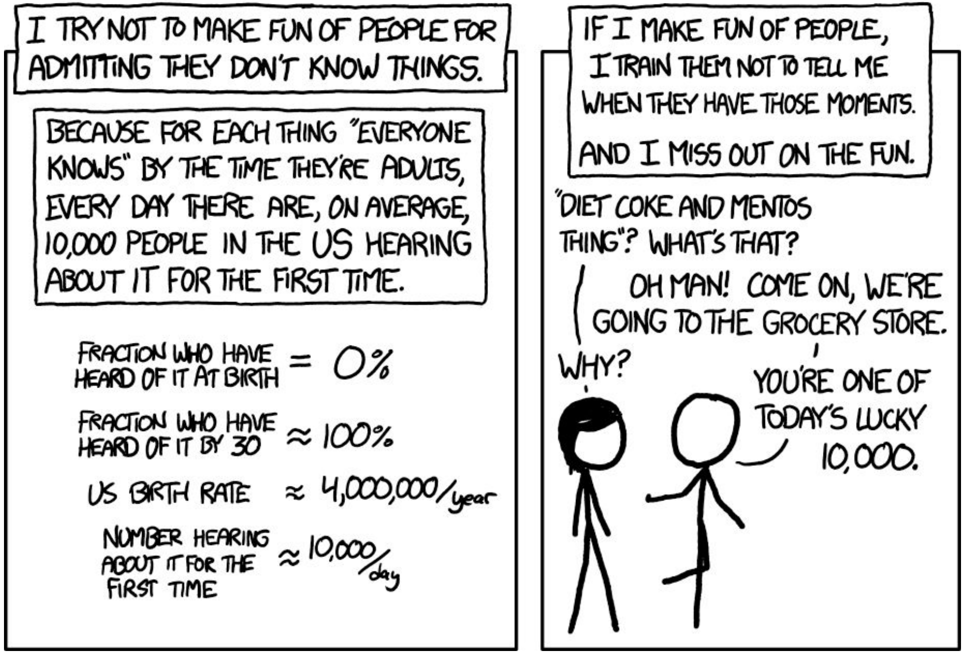
- Agenda:
  - Introductions
  - Win-of-the-month
  - Today-I-learned
  - Cori Retirement Update
  - Announcements and Calls for Participation
  - Topic of the day: **SLURM Magic** with Charles Lively
  - Coming meetings: topic suggestions/requests?

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/)

Today I learned


I TRY NOT TO MAKE FUN OF PEOPLE FOR ADMITTING THEY DON'T KNOW THINGS.
BECAUSE FOR EACH THING "EVERYONE KNOWS" BY THE TIME THEY'RE ADULTS, EVERY DAY THERE ARE, ON AVERAGE, 10,000 PEOPLE IN THE US HEARING ABOUT IT FOR THE FIRST TIME.
FRACTION WHO HAVE HEARD OF IT AT BIRTH = 0%
FRACTION WHO HAVE HEARD OF IT BY 30 ≈ 100%
US BIRTH RATE ≈ 4,000,000/year
NUMBER HEARING ABOUT IT FOR THE FIRST TIME ≈ 10,000/day
IF I MAKE FUN OF PEOPLE, I TRAIN THEM NOT TO TELL ME WHEN THEY HAVE THOSE MOMENTS.
AND I MISS OUT ON THE FUN.
"DIET COKE AND MENTOS THING"? WHAT'S THAT?
OH MAN! COME ON, WE'RE GOING TO THE GROCERY STORE.
WHY?
YOU'RE ONE OF TODAY'S LUCKY 10,000.

https://xkcd.com/1053/

What surprised you that might benefit other users to hear about?

Eg:

- Something you got stuck on, hit a dead end, or turned out to be wrong about
- 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

Cori Retirement Update

Cori is scheduled to be retired on May 31, 2023 at noon PDT

- Users be able to log into Cori login nodes and access Cori scratch file system until Wednesday, June 7, 2023.
- At 12:00 noon Wednesday, May 31, Cori queues will be turned off. A reservation has been put in place to prevent any jobs from running over, but any errant jobs and all jobs in the queue will be deleted at that time.
- Cori login nodes and scratch file systems are planned to be available until 12:00 PST (noon) on June 7, 2023, to provide a final opportunity for users to retrieve their files off the Cori-specific scratch file system.

To Move Data to Perlmutter:

- Cori Scratch to Perlmutter Scratch: Use Globus (https://docs.nersc.gov/services/globus/)
- Use Community File System for regularly retrieved data
- Use HPSS Tape Archive for infrequently (less frequently) accessed data

Scratch storage is not permanent!

Please back up irreplaceable data to CFS or HPSS!

Cori Retirement Update

For more information, submit a ticket through ServiceNow.

- Next Cori to Perlmutter Office Hours:
  - Friday May 26, 10am - 12 noon PST
  - Tuesday May 30, 10am - 12 noon PST
- See docs.nersc.gov for updated information about using Perlmutter
  - For more information, including connection information (login required for Zoom link)

    https://www.nersc.gov/users/training/events/migrating-from-cori-to-perlmutter-office-hours-may-2023/

Seminars, Events, and Trainings*

- Training on Advanced SYCL Techniques & Best Practices, May 30, 2023
  - See NERSC Training Page for information to register, please see https://www.nersc.gov/users/training/events/advanced-sycl-techniques-and-best-practices-may2023/.
- Introduction to NERSC Resources Training, June 8
  - Information and to register, please see https://www.nersc.gov/users/training/events/introduction-to-nersc-resources-jun2023/
- Crash Course in Supercomputing (for Novice Parallel Programmers, June 22
  - Information and to register, please see https://www.nersc.gov/users/training/events/crash-course-in-supercomputing-jun2023/.

*you can find even more info and events in the weekly emails!

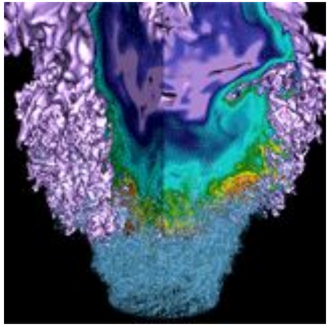
Seminars, Events, and Trainings*

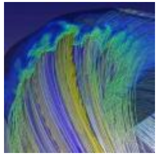
- Call for Submissions US Research Software Engineer Association Conference
  - Please see https://us-rse.org/usrse23/. Poster abstracts are due June 19.
- Call for Participation Open for RSE-eScience-2023 Workshop
  - Please see https://us-rse.org/rse-escience-2023/. Abstract submissions are due June 30
- Call for Papers: 6th Annual Parallel Applications Workshop, Alternatives To MPI+X (PAW-ATM) at SC23
  - Submissions close July 24, 2023. For more information and to submit a paper, please visit: https://go.lbl.gov/paw-atm.

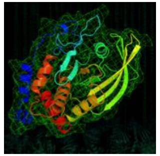
Calls for Participation

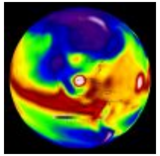
- IDEAS-ECP Webinar "The OpenSSF Best Practices Badge Program”, June 14,2023
  - No cost to attend, but registration is required. Please register at the event webpage.
- Fortran User Group LLVM Flang Survey
  - https://forms.gle/vz43AwQyLJ7T47Um7
- NERSC Message of the Day (MOTD) Survey
  - https://forms.gle/fKVHMuCM8wmqzTbDA

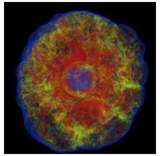
SLURM at NERSC

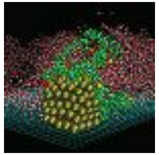












The Basics: What is SLURM?

- SLURM: Simple Linux Utility for Resource Management

  - Functionality: Manages jobs on supercomputing clusters
  - Role: Efficiently schedules and allocates resources for parallel computing
  - Popularity: Widely used in the HPC community
  - Trusted: Preferred choice for job management on large-scale supercomputers
  - Key Features: Job scheduling, resource management, fault tolerance
  - Scalability: Handles large-scale parallel computing with ease

SBATCH and SRUN Flags

- sbatch: Slurm batch and script srun: Slurm run command

  Some selected Flags:

  -n, --ntasks= <number>: sbatch does not launch tasks, it requests an allocation of resources and submits a batch script.

  --tasks, --ntasks-per-node= <ntasks>: Request that ntasks be invoked on each node. If used with the --ntasks option, the --ntasks option will take precedence and the --ntasksper-node will be treated as a maximum count of tasks per node.

  -c, --cpus-per-task= <ncpus>: Request that ncpus be allocated per process. This may be useful if the job is multithreaded and requires more than one CPU per task for optimal performance.

  -q, --qos=high: request a high quality of service for the job.

  -K, --kill-on-bad-exit[=0|1]: Controls whether or not to terminate a step if any task exits with a non-zero exit code. I

  -k, --no-kill: Do not automatically terminate a job if one of the nodes it has been allocated fails.

Essential Job Monitoring

- squeue is a powerful command used to view information about jobs located in the Slurm scheduling queue.
  - Provides real-time information about job status, resource allocation, and other job-related details.
  - Provides essential monitoring for Job Status, Real-time Updates, and Filtering and Sorting.

  sqs is a Slurm command used for querying the Slurm Queue System (SQS) database.

  - It provides historical information about job submissions, resource allocations, and job status.
  - Wrapped into squeue on perlmutter for Queue Status, Fairshare Usage, and Resource Utilization

SACCT Essential Flags

- sacct - Can report resource usage for running or terminated jobs including individual tasks, which can be useful to detect load imbalance between the tasks

Typical Flags Used:

- Filter by user: Use -u or --user option followed by the username(s).
- Filter by job ID: Use -j or --job option followed by the job ID(s).
- Filter by time range: Use -S or --starttime and -E or --endtime options.
- Customize output format with -o or --format option.
- Use various fields like job ID, user, start time, end time, state, exit code, and resource usage metrics.
- Helpful for analyzing job performance, resource utilization, and troubleshooting.

Refer to Slurm documentation or sacct --help for more options and filters.

SACCT Range/Specification Flags

- Use the --start and --end options to specify a time range (e.g., --start=2023-01-01 --end=2023-01-31).
- Use the --format option to customize the output format for desired fields.

  Example format: --format=account,jobid,submit,start,end,elapsed,nnodes,qos,exitcode,state.

  -X option shows a summary for each job, rather than multiple lines.

- Use the --state=COMPLETED flag to filter jobs that ran successfully and completed.
- Use the --format option to customize the output format for desired fields.
- Provides concise and meaningful information about job accounting.
- Useful for analyzing job submissions, start and end times, resource usage, and job status.

SACCT Flags and Options

- --format: Specifies the output format for the desired fields.
- qos%30 and qos%-30: Formatting options to adjust the output width for the QoS field.
- -p: Produces output delimited with the "|" symbol, suitable for importing into a spreadsheet.
- --duplicates: Helps identify all instances of a job, useful when a user requeues their jobs.
- --node: Filters jobs based on the specified node, helpful for investigating suspected node-related issues.
- Example command:

```
sacct --node=nid02930 -X --start=2022-04-01 -o state -n | sort | uniq -c
```

- Counts the number of jobs, since April 1, that ended in various states on the specified node.
- Useful for analyzing node-specific job statistics and identifying patterns.

- Refer to Slurm documentation or sacct --help for more options and filters.

SControl Command

- scontrol command is used by system administrators to make changes in Slurm, but it also offers functionalities for regular HPC users.
- scontrol show res: Shows all reservations on the system.
- scontrol show res=reservationname: Shows details of a specific reservation, identified by "reservationname".

Example usage: scontrol show res=debug.

- Provides information about the reservation's start and end time, duration, nodes, core count, features, partition, and more.
- Useful for checking reservation details to ensure your job aligns with the reservation constraints.
- Enables users to verify the reservation status and determine resource availability.
- Other scontrol show commands exist for viewing various aspects of Slurm.

SINFO Command

sinfo command provides information about Slurm partitions, giving insights into available resources.

- Useful for understanding the availability of nodes for specific job types.
- Does not directly map to queues but provides an idea of resource availability.
- Example formatting: sinfo -o "%.20P %.5a %.10D %.20F".
- -o flag specifies the output format using placeholders for partition, availability, nodes, and node status.
- Provides a concise and informative table displaying partition names, availability status, node counts, and node status breakdown.
- Node status breakdown format: allocated/idle/other/total.
- Helps identify the number of nodes currently available, in use, idle, or in other states.
- Enables users to determine resource availability for job submissions.

Creating An Alias in SLURM

In SLURM, an alias is a shorthand or shortcut that allows users to create custom commands or abbreviations for frequently used SLURM commands or command sequences.

- **Time-saving**: Aliases save time by reducing the need to type long and repetitive commands.
- **Ease of use**: Aliases make SLURM commands easier to remember and use. Instead of recalling and typing lengthy commands with numerous options, users can simply execute the defined aliases.
- **Customization**: Aliases allow users to customize their SLURM workflow according to their specific needs.
- **Customization**: Aliases allow users to customize their SLURM workflow according to their specific needs.
- **Consistency**: Aliases promote consistency in command execution.
- **Productivity**: Using aliases enhances productivity by streamlining command execution.

Alias Examples

alias sinfonodes='sinfo -o "%.18P %.10a %.15l %.20F %.20b"'

- sinfonodes alias based on sinfo displays detailed information about node availability and status in a formatted output, including partition, node state, and features.
- The -o flag specifies the format of the output.
- "%.18P %.10a %.15l %.20F %.20b" defines the format string for the output, specifying the columns to display.
- The output will show partition, alloc nodes, node state, features, and node names.

alias checkjob='sacct -o jobid,user,jobname,qos,submit,timelimit,nnodes,ncpus,start,end,elapsed,exitcode -j '

- This command creates an alias called checkjob for the sacct command.
- The -o flag specifies the format of the output.
- "jobid,user,jobname,qos,submit,timelimit,nnodes,ncpus,start,end,elapsed,exitcode" defines the format string for the output, specifying the job attributes to display.
- The -j flag is used to specify the job ID to check.

Alias Examples (2)

alias sallocpg='salloc -N 1 -C gpu -t 60 -c 10 -G 1 -A nstaff -q interactive'

- This command creates an alias called sallocpg for the salloc command.
- The -N flag specifies the number of nodes to allocate.
- The -C flag specifies the required node features or constraints, in this case, GPUs.
- The -t flag specifies the time limit for the allocation.
- The -c flag specifies the number of CPUs to allocate per task.
- The -G flag specifies the number of GPUs required.
- The -A flag specifies the account or project to charge the allocation.
- The -q flag specifies the queue to use.

sallocpg alias is used to allocate resources for a job, requesting 1 node with GPU support, a time limit of 60 minutes, 10 CPUs per task, 1 GPU, using the "nstaff" account, and running in the interactive queue.

Alias Examples (3)

```
alias sacctacc="sacctmgr show association user=clively -p"
```

- This command creates an alias called sacctacc for the sacctmgr command.
- The show association subcommand is used to display job association information.
- The user=yunhe flag specifies the user for which to display the association.
- The -p flag formats the output in a parsable format.
- sacctacc alias shows the job association information for a specific user, in this case, "clively", in a parsable format.

```
alias checkpm='sinfo --format='\''%b %t %D'\'' | grep gpu | grep ss11'
```

- This command creates an alias called checkpm.

- Refer to Slurm documentation or sacct --help for more options and filters.

Coming up

Upcoming topics:

- JupyterHub at NERSC
- NERSC Tips and Tricks
- Security at NERSC

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

Nominate a topic at: https://forms.gle/WjYx7zV7SAz2CaYz7

Science Highlights Submission:
https://docs.google.com/forms/d/e/1FAIpQLScP4bRCtcde43nqUx4Z_sz780G9HsXtpecQ_qIPKvGafDVVKQ/viewform

Lightning Talks

Highlights

NERSC

Thank You