services@sdsc.edu

Accelerating Genomics Research Machine Learning with Slurm

Willy Markuske
SDSC - Research Data Services
Slurm User Group Conference 2023

Slurm at SDSC HPC

The San Diego Supercomputer Center at the University of California, San Diego was founded in 1985 as one of the five original NSF Supercomputing Centers.

SDSC has made use of numerous scheduler systems throughout the years but Slurm usage is fairly recent and expanding.

First major system to use Slurm was the Expanse system in 2020.

The Triton Shared Computing Cluster currently uses TORQUE with plans to transfer to Slurm.

SDSC
RESEARCH
DATA
SERVICES

SDSC Research Data Services

Research Data Services (RDS) provides a number of colocation and administration services to groups across the UC system and externally:

- Colocation in a 19,000-square foot datacenter
- Cloud Compute and Storage (Openstack/Ceph)
- Universal Scale Storage (Qumulo)
- Enterprise Networking
- System Administration
- *Custom Compute Cluster Development and Management*

SDSC RESEARCH DATA SERVICES

RDS Custom Compute Clusters

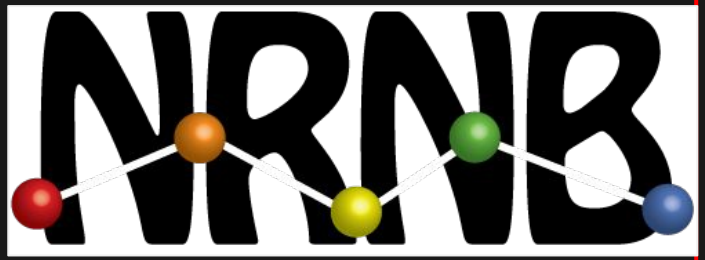
- Relatively new service to provide tailored cluster computing using Slurm to UCSD research groups that have a need not met by larger offerings at SDSC
- All aspects of the cluster can be customized
  - compute nodes
  - accelerators
  - backend networking
  - storage
- Currently support groups in
  - genomics and network biology
  - medical imaging
  - marketing and business
  - weather prediction

SDSC RESEARCH DATA SERVICES

NRNB Compute Cluster

National Resource for Network Biology (NRNB) sponsored by the Ideker Lab at UCSD School of Medicine through NSF grants.

- Complete cluster refresh in 2020 through 2025
- Supports genomics research for 3 PIs and 200+ users
- Slurm based workload management
- Hardware
  - 12 AMD Epyc 7002 based standard compute nodes
  - 4 AMD Epyc 7002 based high memory nodes
  - 8 Nvidia RTX 2080Ti across two nodes
  - 8 Nvidia RTX 5000 across two nodes
  - 20 Nvidia V100 across five nodes
  - 20 Nvidia A30 across five nodes
  - 2 PBs of BeeGFS storage



SDSC RESEARCH DATA SERVICES

Researcher Needs

Cluster design and features are driven by direct researcher feedback. Slurm is used as the resource manager to handle various workloads with a singular technology for researchers.

Key Needs:

- Jupyter based development environments
- High throughput job management
- GPU management for AI workloads
- Automated data ingestion and dissemination
- Resource availability monitoring

SDSC RESEARCH DATA SERVICES

Python/R Jupyter Environment

What is the best way to provide support and consistency for mixed Python and R code development?

SDSC RESEARCH DATA SERVICES

How to provide Jupyter notebooks?

Jupyter Notebooks

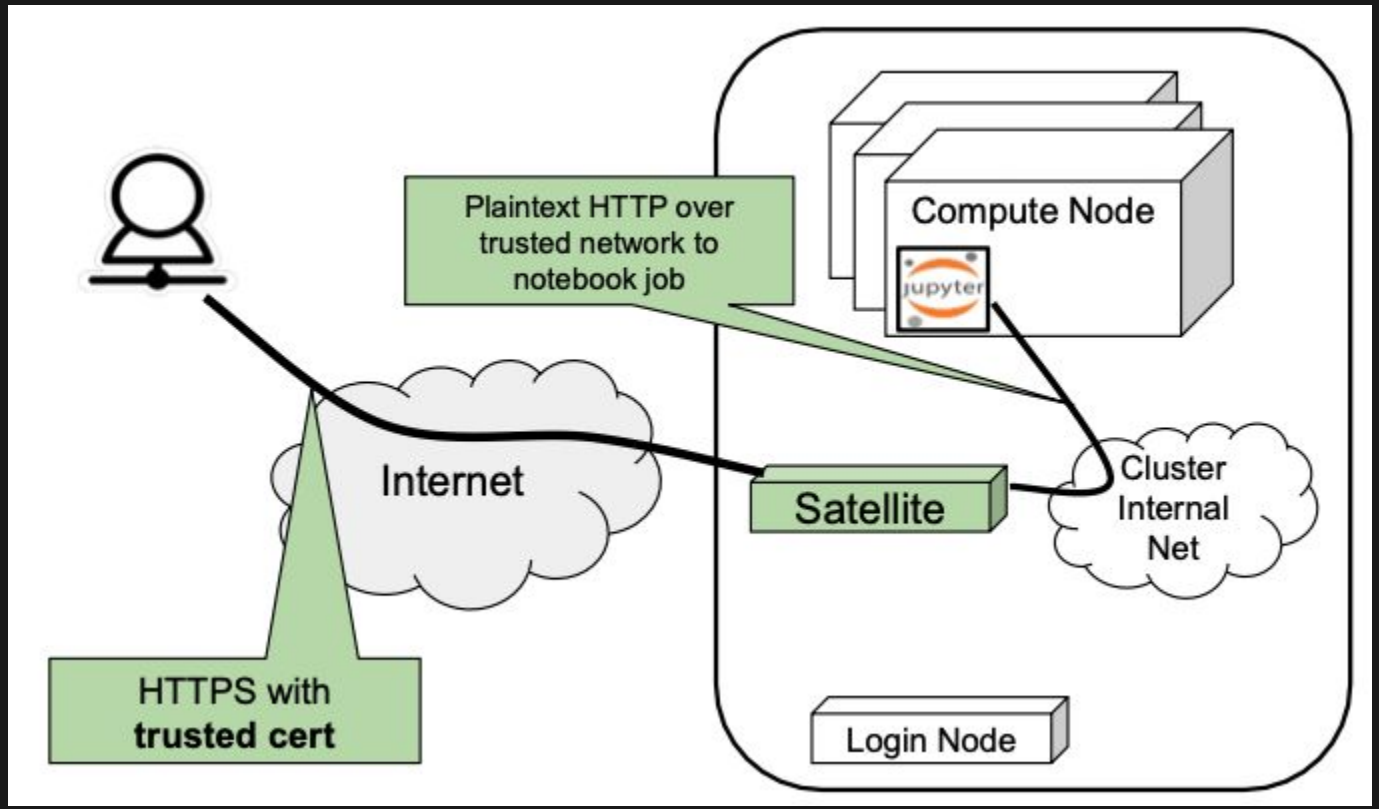
- Users define their own environments
- Resource allocation and sizing can be done with Slurm job submission
- Requires manual port forwarding and ssh tunneling
- No central login site
- Move management responsibility to user

JupyterHub

- Predefined system environments
- Resources allocated through Slurm spawners of defined sizes
- Handles HTTP proxy
- Centralized login with cluster credentials
- Another service that needs to be maintained

SDSC RESEARCH DATA SERVICES

SDSC Satellite

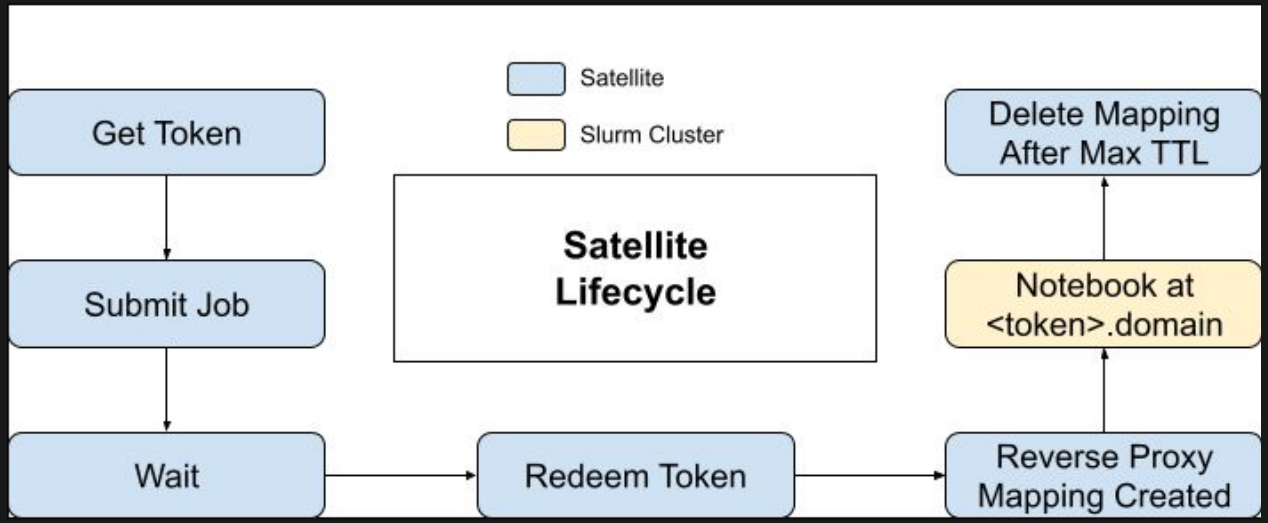


SDSC RESEARCH DATA SERVICES

SDSC Satellite

Single command from users that uses subset of sbatch flags

```
jupyter-submit -p partition -A account -t time -c ncpus -m memory -g ngpus -J jobname
```



SDSC RESEARCH DATA SERVICES

High Throughput Adjustments

Primary jobs are genome wide association studies and AI training which lend themselves to embarrassingly parallel parallelization through Slurm array jobs. Tasks are run over different genomics data files with no MPI support and often limited MP support.

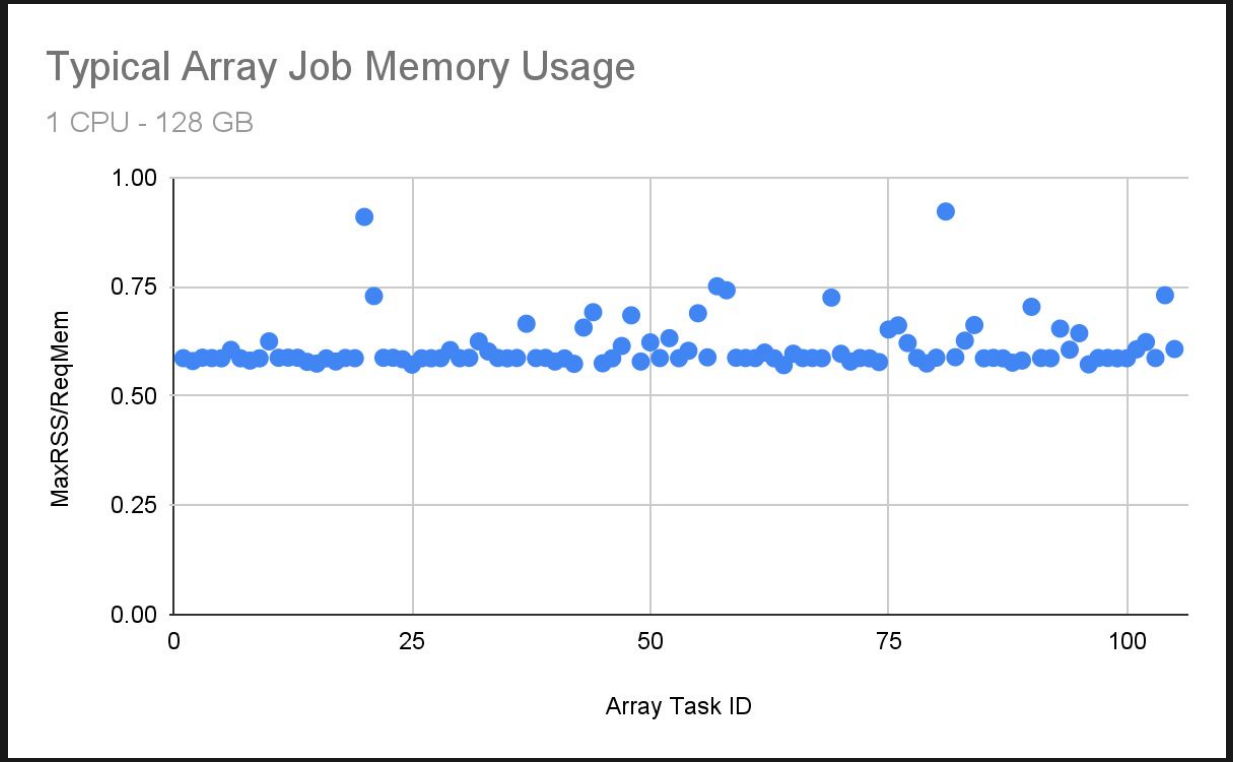
Used guidance from High Throughput Computing, Broderick Gardner, SchedMD SLUG 2019:

- Prioritized CPU Freq and Memory for slurmctld operation and state, Epyc 7F32/ 128 GB
- StateSaveLocation and SlurmdSpoolDir saved to local NVMe
  - StateSaveLocation periodically copied to persistent storage
- MaxArraySize=50001, MaxJobCount=500000
- Train use arrays and limit simultaneous jobs via %
- Prioritize via partitions

SDSC RESEARCH DATA SERVICES

High Throughput Adjustments

Should we allow oversubscription? Probably better user training.



SDSC RESEARCH DATA SERVICES

GPU Resources

How GPUS are configured
Initial investments in MIG and slurm MIG support

- GPUs are defined using gres resource with autodetect=nvml
- Each GPU also manually defined with core affinities corresponding to it's cpu NUMA node and GPU type
- Tested MPS with V100 systems but found that all MPS jobs would be applied to device0 regardless of requested mps count
- Tested MIG support will allow for better gpu utilization

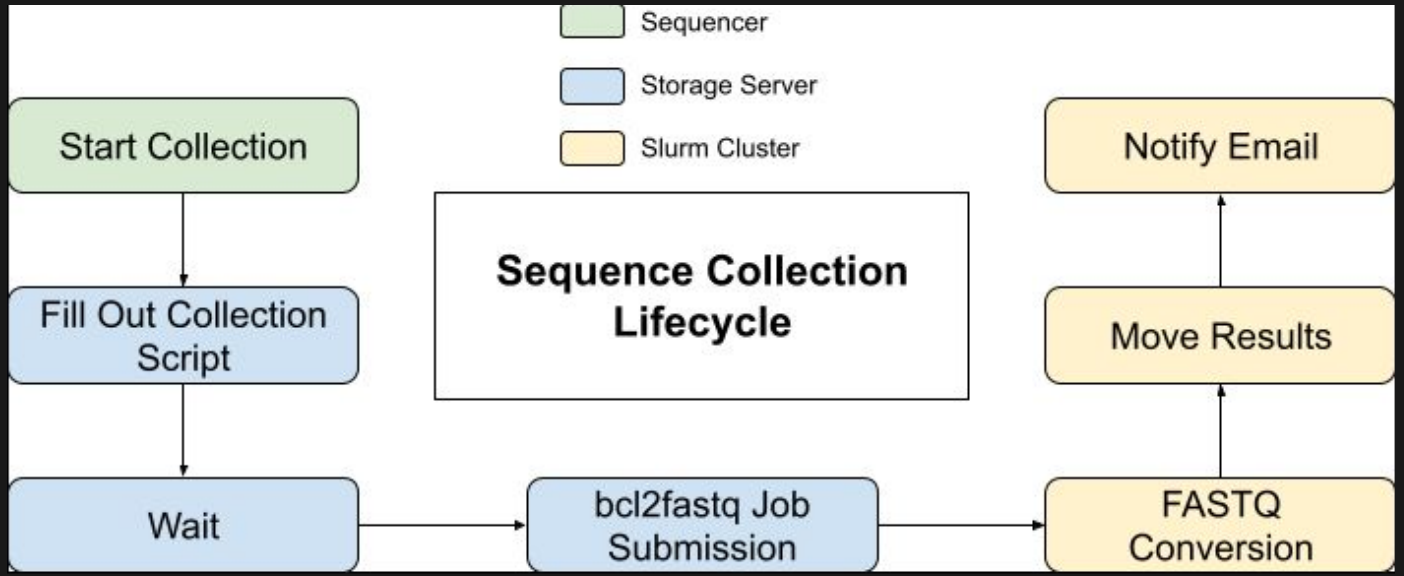
```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 525.85.12    Driver Version: 525.85.12    CUDA Version: 12.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  NVIDIA A30          On   | 00000000:17:00.0 Off |                    0 |
| N/A   31C    P0    30W / 165W |   4575MiB / 24576MiB |      0%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+
|   1  NVIDIA A30          On   | 00000000:65:00.0 Off |                    0 |
| N/A   31C    P0    31W / 165W |   1415MiB / 24576MiB |      0%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+
|   2  NVIDIA A30          On   | 00000000:CA:00.0 Off |                    0 |
| N/A   33C    P0    32W / 165W |   1415MiB / 24576MiB |      0%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+
|   3  NVIDIA A30          On   | 00000000:E3:00.0 Off |                    0 |
| N/A   35C    P0    32W / 165W |   1415MiB / 24576MiB |      0%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|    0   N/A  N/A    853697      C   ...nda3/envs/G2PT/bin/python     1428MiB |
|    0   N/A  N/A    853828      C   ...nda3/envs/G2PT/bin/python     1048MiB |
|    0   N/A  N/A    853846      C   ...nda3/envs/G2PT/bin/python     1048MiB |
|    0   N/A  N/A    853861      C   ...nda3/envs/G2PT/bin/python     1048MiB |
|    1   N/A  N/A    853828      C   ...nda3/envs/G2PT/bin/python     1412MiB |
|    2   N/A  N/A    853846      C   ...nda3/envs/G2PT/bin/python     1412MiB |
|    3   N/A  N/A    853861      C   ...nda3/envs/G2PT/bin/python     1412MiB |
+-----------------------------------------------------------------------------+
```

SDSC RESEARCH DATA SERVICES

Automating Sequencing

Wetlab provides sequencing collection for researchers to use as genomics datasets.

Leverage Slurm to provide a more streamlined way to convert sequence collections and provide files to researcher.

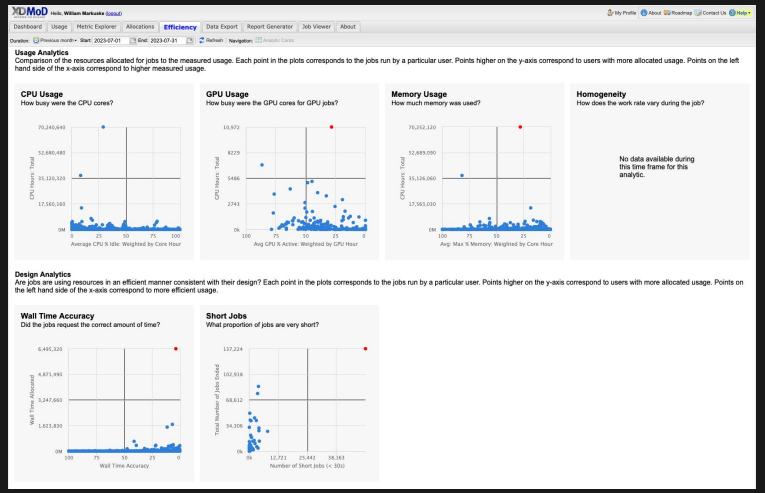


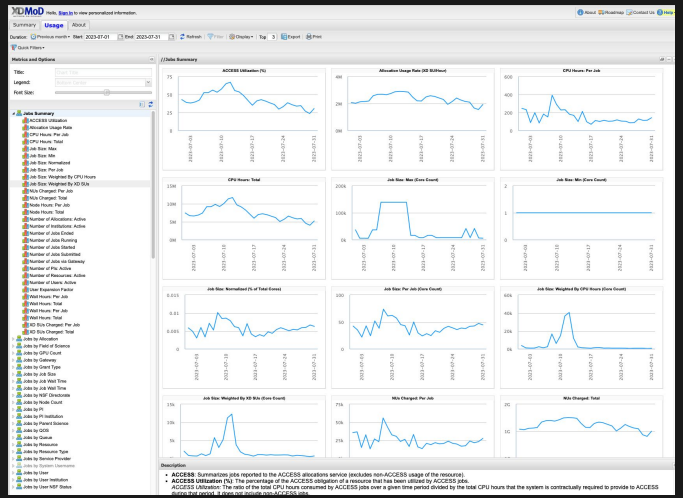
SDSC RESEARCH DATA SERVICES

User Focused Monitoring

Received user feedback on XDMoD monitoring:

- Too many options as a user
- Want immediate node status clarification based on resources
- Historical data not as useful as users
- Want linked plots to easily determine node status





User Focused Monitoring

User Focused Monitoring

- Simplified dashboard via Dash
- Directly shows node status and resource allocation
- Users can sort and select nodes across all plots at once
- Job tab in development
  - Show all user job efficiencies similar to XDMoD over a given time period
  - Selectable jobs by number and graph selection to show detailed information

SDSC RESEARCH DATA SERVICES

Questions?

consult@sdsc.edu

services@sdsc.edu

SDSC RESEARCH DATA SERVICES