Improved Cluster Resource Tracking

DSE 241 Project Proposal Willy Markuske

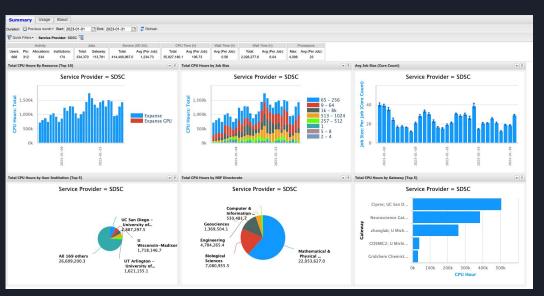
NRNB Compute Cluster Slurm Database

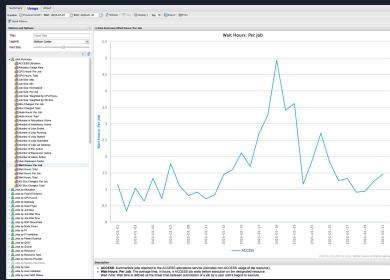
Information is collected for all submitted jobs in a MySQL database including:

- State
- Submitted/Start/End Times
- CPUs Requested
- Memory Requested
- GPUs Requested
- Actual Resources Used
 - o CPU Hour/Req CPU Hour
- Nodes used

JobID	State	Start	End	ReqCPUS	ReqMem	ReqTRES
370824	RUNNING	2022-08-08T15:10:14	Unknowr	16	256G	billing=1+
400803	RUNNING	2022-08-18T12:29:17	Unknowr	1	256G	billing=1+
419930	RUNNING	2022-08-23T15:25:11	Unknown	4	32G	billing=4+
421886	RUNNING	2022-08-24T23:50:51	Unknowr	ı 5	205G	billing=5+
422379_131	RUNNING	2022-08-31T22:21:22	Unknowr	1 1	40G	billing=1+
422379_132	RUNNING	2022-08-31T23:32:17	Unknowr	1 1	40G	billing=1+
422379_133	RUNNING	2022-09-01T01:30:42	Unknowr	1 1	40G	billing=1+
422379_134	RUNNING	2022-09-01T02:33:21	Unknowr	1 1	40G	billing=1+
422379_135	RUNNING	2022-09-01T03:42:07	Unknowr	1 1	40G	billing=1+
422379_136	RUNNING	2022-09-01T04:31:18	Unknowr	1 1	40G	billing=1+
422379_137	RUNNING	2022-09-01T06:30:40	Unknowr	1 1	40G	billing=1+
422379_138	RUNNING	2022-09-01T08:53:51	Unknowr	1 1	40G	billing=1+
422379_139	RUNNING	2022-09-01T09:29:31	Unknowr	1 1	40G	billing=1+
422379_140	RUNNING	2022-09-01T09:54:12	Unknowr	1 1	40G	billing=1+
425379	RUNNING	2022-08-25T13:02:43	Unknowr	1 1	100G	billing=1+

Current Standard Software - XDMoD





What Can be Made Better?

Users:

- How many resources are currently being used by the cluster?
- How long can I expect my job to take to start running if submitted now?
- How many resources have I consumed in a given date range?

Administrators:

- Is the cluster currently oversubscribed?
- What trends can be seen between high and low usage across different types of resources?
- Compare resource usage patterns of different users.

Visualization Implementation

- Python backend to scrape data from the job database
- D3JS site with 2 main pages
 - Current Summary
 - Historical Data
- Users can see and interact with current summary page to see the resource allocation of running jobs and pending jobs
 - Ex: View a current plot of GPUs in use by system normalized by GPUs available in each system
- Users/Admins can plot trends for completed jobs using the historical data page to view resource utilization in conjunction with key analysis indicators
 - Ex: Number of GPUs requested per day indicating days of the week to find trends with submissions based on weekdays

