

Slurm job queue visualization system

Pavel Vasev, Alexander Igumnov, Andrei Popel, Alexander Bersenev

N.N. Krasovskii Institute of Mathematics and Mechanics
of the Ural Branch of the Russian Academy of Sciences

A supercomputer is a shared resource: many researchers run their parallel programs, which are placed on the supercomputer's nodes by a scheduler. Depending on the current load of the supercomputer and requirements for running a program (the number of required nodes, etc.), a new program run can be scheduled for a specific time, either earlier or later. If the researcher can vary the requirements for running the program, they may be able to run the program earlier, resulting in faster results. To achieve this, the researcher needs to understand the current and planned load of the supercomputer.

More specifically, *supercomputer* is a cluster system which consists of *nodes*. *Users* of supercomputer spawns *jobs*, where each job is a parallel program executed on nodes. Jobs are *scheduled* and places in *job queue*. Jobs that run now form a *current load*, and scheduled jobs form *planned load*.

The visualization task considered in the paper is as follows. The user of supercomputer needs to:

- See a general picture in time, which supercomputer nodes are busy and for how long they will be busy. In our practice, a time horizon of 2 days seems sufficient.
- See a general picture of the free nodes in time.
- Be able to highlight jobs of specific user within the set of all users jobs.

The URAN supercomputer of the Ural Branch of the Russian Academy of Sciences is managed by the Slurm [1] workload manager. Supercomputer software tools provide the following information about the current and planned load:

Text output in the format of the *squeue* command (see table 1).

Graphical output of the current load over nodes (see <https://umt.imm.uran.ru>).

These tools allow users to understand the current load of a supercomputer, including which nodes are free at the current time. However, they do not provide a graphical representation of how long the nodes will be busy or when the currently free nodes will become busy. This information is available in text form, but it is difficult for humans to analyze it. No other Slurm visualization tools have been found on the Internet that address the stated visualization task.

Based on the above, a new visualization system for the supercomputer's job queue has been developed. The chosen primary view is a table. The rows of the table represent the nodes. The columns of the table represent the time. Each cell in the table denotes 1 hour, and provides information about whether the node is busy, scheduled, or free at that hour. This view has the following features:

- The node load is displayed with an accuracy of one hour.
- With a relatively small number of nodes (50-100), it is possible to display them on a single screen without the need for scrolling or zooming. This allows for the analysis of the cluster's load without interactive intervention.

The system addresses the visualization challenge through the following features:

- Displaying the cluster's jobs queue in a table format is performed in text and web forms. The text form is convenient for quick analysis during a terminal session with the cluster, and since the user is logged in at that moment, they can view more information, such as their jobs. The web form is suitable for viewing over the Internet from computers and mobile devices.
- Highlighting underutilized nodes with color gradation based on their level of utilization.
- Features of the web form include allowing to click on a job name and highlight it on the overall view. The web page is periodically auto-updates to reflect changes. Additionally, the view may split into two columns for compact representation, if screen width is sufficient.
- Features of the text form include highlighting the jobs of the current user and displaying the id's of users jobs.

The program's output is shown in figure 1.

JOBID	PARTITION NAME		USER	ST	TIME	NODES	NODELIST
....							
19670313	apollo	id_start	user8	R	1:38:14	1	apollo25
19670150	apollo	int_bi_d	user8	R	2:46:42	1	apollo31
19669626	apollo	id_AC_pa	user8	R	18:09:17	2	apollo[17,30]
19669627	apollo	id_AC_pa	user8	R	18:09:17	2	apollo[30-31]
19669629	apollo	id_AC_pe	user8	R	18:09:17	2	apollo[22,31]
19669630	apollo	id_AC_pe	user8	R	18:09:17	1	apollo11
19669632	apollo	id_AC_pe	user8	R	18:09:17	3	apollo[11 apollo[17-19,21-29]
19670346	apollo	amulet.m	user9	R	14:34	12	
19670147	hiperf	monster_	user10	R	2:51:28	1	tesla-v100
19670158	hiperf	monster_	user10	R	31:07	1	tesla-a100
19668608	tesla	ExtNuc35	user4	R	2-14:56:54	1	tesla43
19668595	tesla	CytExt03	user4	R	2-15:45:29	1	tesla34
19668601	tesla	CytNuc03	user4	R	2-15:45:29	1	tesla44
19668602	tesla	CytNuc35	user4	R	2-15:45:29	1	tesla45
....							

Table 1. An example of output of Slurm's *squeue* command. Compare it with Figure 1.



Figure 1. An example of Slurm job queue visualization system output. The visualization program operates in a web environment. The vertical axis displays the node's name and status: [free cores/ total cores]. Status meaning: gray – node is fully loaded, white – node have few free cores, darkgreen – node is half loaded, lime – node is free. Each node is represented by a horizontal bar of cells. Each cell represents one hour. If a cell is yellow, it means that some job is running on that node during that hour. If a cell is darkyellow, it means that some job is scheduled on that hour. 1..9+ represents the number of jobs currently running on the node. The vertical purple bar represents boundary of a day (00:00 am).

The presented Slurm job queue visualization system is currently in the pilot phase. The source code is available online at <https://github.com/pavelvasev/mqvis>.

References

1. Slurm workload manager. URL: <https://slurm.schedmd.com>.