

2023 Minerva User Satisfaction Survey Results

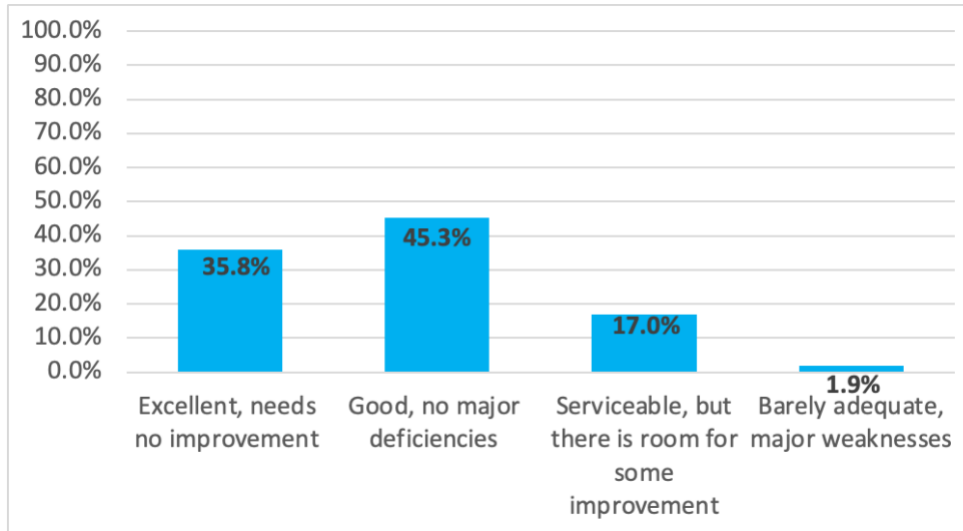
Scientific Computing and Data

April 1, 2024

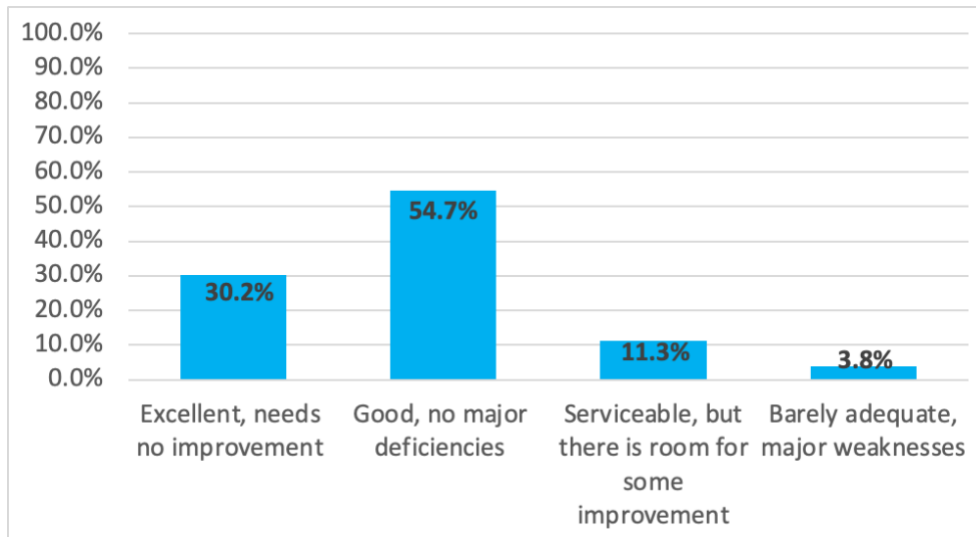
The 2023 Minerva User Satisfaction Survey—distributed in January 2023—solicited feedback from 1,100 active Minerva users. Of these, 54 users responded (5% response rate).

We asked five questions:

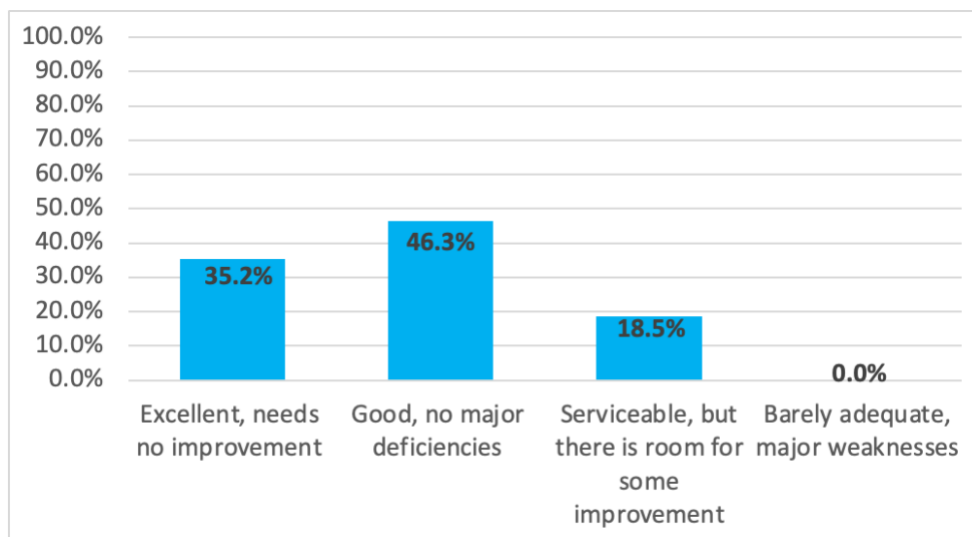
Q1: Overall, how satisfied are you with LSF queue structure, storage, and compute resources (file system, GPUs, high-memory nodes, SM, etc)?



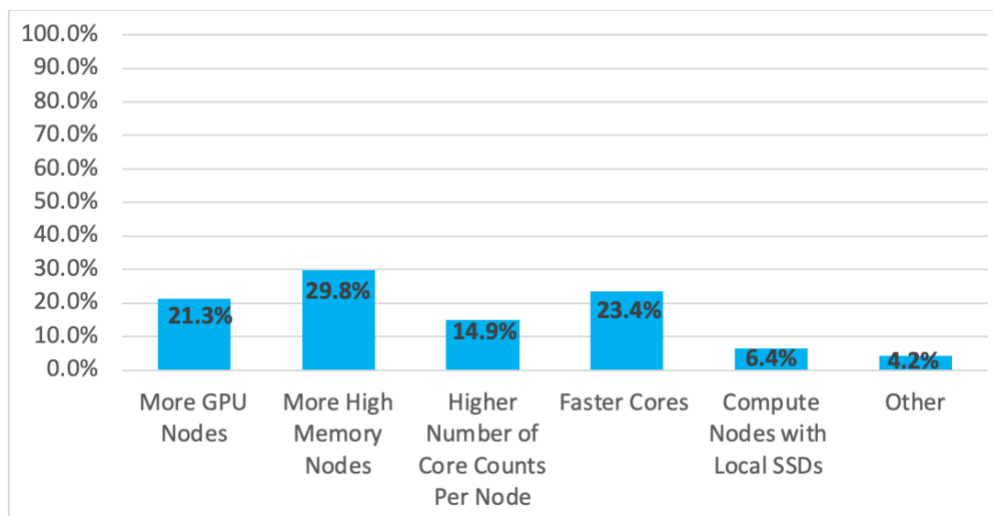
Q2: Please rate current software environment (packages and services such as database, data transfer, container etc).



Q3: Please rate your satisfaction with operations (ticket system, responsiveness of staff, documentation, user support, etc).



Q4: Which of the following would you most prefer for future Minerva expansion?



Q5: What suggestions do you have for improving our service?

All comments and responses are as follows:

System Related Issues:

- not enough GPUs, this is limiting a great deal of my research. I understand Nvidia GPUs may be expensive but there are many new companies such as Intel and AMD providing deep learning processing. I would be happy to help evaluate if this could work for our purposes. • Wait times for GPU resources can be very long, to the point it is shorter to simply run programs on CPU and wait longer for them to run • The gpu queue is often not efficient. The wait time is long even for a single job requiring few resources. • Please more GPUs, they are constantly being used and it is limiting the research that I am doing. Does not have to be NVIDIA, other companies like AMD and Intel have respectable deep learning processors. Worth looking at these if it is more cost efficient.
 - We added 2 more GPU nodes (0.5TB RAM & 4 *80GB H100 on each) in 2023 to serve more runs.
 - We plan to expand and update our compute resource this year with new hardware including about 200 extra GPU cards to provide more compute capacity and boost productivity for users.
- I haven't had any issues with the software but it good to keep the CUDA drivers up to date. I would like to use CUDA 12.2 for some analysis but am not able to find it. would be good to
 - Our system admins are working on upgrading the GPU driver. Currently, the GPU drivers on some nodes have been upgraded to a version that is compatible with CUDA 12.2.

Job Scheduler & Queues:

- After the LSF update, there appears to be a longer lag in launching jobs on some queues (interactive most notably). During this lag, the status shows as "new job waiting to be scheduled" (or similar). • I find that sometimes I will submit jobs and they will be pending for a long while, and other times they will run right away. • wait times in queues can be high than on other hpcs • Sometimes it's very hard to get high-mem nodes even if my "priority" is high (meaning that I didn't calculate anything for weeks). • More nodes in general, jobs can take a long time in queue at this current time.
 - There could be many reasons for a job to stay in the queue for longer than usual; The whole cluster or the compute partition your job is supposed to be dispatched to may be all busy; Your job requests a large amount of compute resources such as amount of memory and compute cores so that no compute node can satisfy the resource requirement at the moment; Your job would overlap with a scheduled PM. To see the pending reason use the "bjobs"

command with the “-l” flag:

```
$ bjobs -l
```

- You may want to use the LSF command “bqueues” to check the load status of the queues, which would be helpful for you to determine which queue is less crowded and where your jobs should be submitted to.
 - We have set an extra global limit on the total amount of memory a user can request for all their jobs on Minerva effective August 25th, 2023, which is 15TB per user. This change is implemented to further avoid the whole cluster taken by the jobs of few users in response to recent users' feedback. We will keep monitoring and fine tune this limit as needed.
- The only thing is that I often need working interactively with internet connection and, sometimes, it is difficult to find available interactive nodes.
 - Compute nodes with internet access.
 - Users can also use other queues (e.g., premium) for interactive jobs. Internet connection can be enabled with proxies loaded in module:


```
$ module load proxies
```
 - If you still have trouble with accessing internet, please contact us at hpchelp@hpc.mssm.edu
 - Change time limit of a job while it is running
 - We do not support extension of the walltime limit for running jobs as it would disrupt the scheduling of other users' jobs on Minerva. You can modify the walltime limit while the job is waiting in the queue in PEND state using the “bmod” command.
 - The process of submitting jobs, collecting successful results, checking for the inevitable failed jobs, resubmitting failed jobs, and repeating until all jobs finally succeed all feels much more laborious than it should. I suppose addressing this would require some kind of higher abstraction layer on top of LSF (or SLURM, or any other job queue system), and I'm not even sure if such a thing exists.
 - Retrieving job stats and managing log files is significantly more complicated on Minerva than HPCs I've used at previous institutions
 - Unfortunately, LSF doesn't provide a user-friendly interface other than the standard command-line tools such as “bjobs”, “bhist”, “bhosts”, and “bqueues” to check the status of user jobs. Our latest slides on the basic LSF commands are at https://labs.icaahn.mssm.edu/minervalab/wp-content/uploads/sites/342/2023/10/Minerva_LSF_2023-10-04.pdf
 - We have a wrapper script authored by Harm van Bakel, which will make it easier to interact with the LSF job scheduler on Minerva, as suggested.


```
To load them up, $ml LSFqueue
To get more info on the module, $module help LSFqueue
And a detailed readme file at
/hpc/packages/minerva-centos7/LSFqueue/1.0/README.txt
```

- You can also get quick commands listed by running “minerva_help” on login nodes
- (Plans to) remove services, such as LSF to use SLURM instead, needs an extensive discussion first, comparing the benefits of the new service with the cost (e.g. to rewrite all the scripts using LSF) removing the old service. It cannot be simply done after a group is displeased with the old service and complains about it.
 - SLURM offers basically the same fundamental functionalities as LSF and is widely used on many other HPC clusters. We are open to either option, but it seems that the majority of our users want to stay with LSF based on the recent survey feedback. Please see the survey results at <https://labs.icaahn.mssm.edu/minervalab/wp-content/uploads/sites/342/2023/12/Minerva-Job-Scheduler-User-Survey-2023-2.pdf>
- I also think that there should be a dedicated node for rstudio sessions - I often will try to grab a session and it may take hours to get it with the amount of memory that I require. I need these sessions for my main project and this significantly delays my progress.
 - In case you need a significant amount of memory for your session, you would have better chances to get it launched when it is scheduled to run on any compute node available at the moment. It would be much harder for LSF to make room for a high-memory job on a dedicated node unless the node itself is dedicated only to your jobs. Please contact us at hpchelp@hpc.mssm.edu with more details about your workflow so that we can take a look and see if there's a way to improve it.

Software/Packages:

- I would suggest removing the R module configuration that forces all R versions to share installed packages in ~/.Rlib. This causes problems with any R package with compiled code (e.g. C code), because such packages fail to load in any version of R other than the one used to install them. However, undoing that configuration at this point will be problematic, because everyone will have to reinstall any R packages they installed in their home directories. • For some reason the different module versions of R are all configured to install packages in the same place, which causes major issues with using multiple versions.
 - Users can install R packages locally into a desired location as required, not into the default path “~/.Rlib”, by specifying the environment variable “R_LIBS_USER” accordingly.
- Notable limitations (e.g., docker)
 - Because of security concerns we do not support Docker. Instead, you can use the Singularity tool on Minerva. Docker container images can easily be pulled and run as Singularity container images, which are safe to run on the shared cluster. See <https://labs.icaahn.mssm.edu/minervalab/documentation/running-container-singularity/> for more information.

- If you have questions about running a docker image with singularity on Minerva, please contact us at hpchelp@hpc.mssm.edu
- RStudio web server is extremely useful but it seems like only one person is managing it and it can only be used on li03c03. • RStudio being able to run with a conda environment.
 - Our RStudio on-the-fly script can now be used from all the login nodes. We will keep improving the script and more functions will be added in the future.
- Files tab on OnDemand needs to have bookmarked folder locations so we can navigate faster. • OnDemand needs a dark mode.
 - Thank you for your suggestions. We will explore the possibilities.
- I face some connection issues. The visual studio code connection with Minerva get disconnected sometimes.
 - Newer versions of visual studio code do not work with the current operating system on Minerva, please check: https://code.visualstudio.com/docs/remote/faq#_can-i-run-vs-code-server-on-older-linux-distributions
We are planning to upgrade our operating system.
 - We are exploring other ways to support visual studio code on Minerva

Tickets:

- Due to staffing shortages, tickets take a little longer to clear, but this is completely understandable at the moment. • Sometimes helpdesk is understaffed and it takes a while to get an answer • There is a clear need for more staff. They are very helpful and responsive but take a lot of time and I know it's because there's only maybe 2 staff managing this right now. It is not reasonably possible for only 2-3 people to manage the whole system and be responsible for HPC tech support at the same time. • Usually staff responds to inquiries, but there are also instances where no responses are given more than 24h later, which has an impact on working progress. • Recently, my tickets response has been quite slow. • More staff!!! Any useful expansion will take time to integrate and as of now I don't think they can handle what we already have.
 - The HPC team is currently short-staffed. We sent an announcement “HPC Ticket Updates: responses will be largely delayed due to being short-staffed” to all HPC users on November 6th, 2023. The responses to your tickets at hpchelp@hpc.mssm.edu will be delayed, which may be up to 72 hours due to being short-staffed.
 - We are actively hiring new staff members.
- I have more than once received cryptic or no answers to questions or issues I submitted through the system
 - We handled more than 3,500 tickets in 2023 and tried our best to reduce the time to respond to users' tickets. If you do not hear from us in a couple of days since you opened a ticket at hpchelp@hpc.mssm.edu it is most likely

slipped out mistakenly. Sorry about that. We must admit we make mistakes sometimes. Please send us a note and remind us about your open ticket.

Documentation on website:

- Better and more up-to-date documentation (e.g., for Jupyter server, VS Code connection, etc.) needed; a lot of the documented methods don't work anymore, it seems • The documentation is still quite challenging to work with. For example, getting jupyter notebooks to work with non-conda environments was extremely challenging. Additionally, whether I use conda or venv, I find that until I manually specify the location of where to install packages (vs. the main minerva python location, which is where it tries and fails by default) • Improvements on documentation on how to access Minerva outside the campus from a Linux PC
 - We are continuously working to improve our documentation. We will try adding more details about the usage instructions and updating them in a timely manner.

Others:

- I think it would be good if there were Zoom drop in hours
 - Scientific Computing and Data & DTP are holding Digital Concierge sessions via Zoom every Wednesday 3:30-4:30 PM. Announcement with registration link is sent out every Wednesday. HPC and Data-Ark staff will be there helping users.
- Integrate the data warehouses better
 - The MSDW OMOP database is already hosted and accessible on the Minerva High-Performance Computing (HPC) cluster alongside other research data sets. To provide better user experience and functionality we have worked closely with the MSDW team.
 - Please reach us at hpchelp@hpc.mssm.edu for more detailed suggestion on the operation if there still is.
 - If you have any concerns or questions about their operations, please reach out to them at MSDW Support. We understand that there still is room to improve on our end and will continue to work to provide better service to our users.

Thank you for your kind words!!!

- Everyone is generally very responsive
- My concerns are always answered promptly and with professionalism.
- Staff is great, got back on all my issues with a very fast turnaround and helpful information
- with the addition of OnDemand, data transfer has become so much easier.

- No major suggestions. The Minerva HPC team continue to be excellent to work with. The team's responsiveness and ability to accommodate our team has been great. Thank you for all you do.
- I would like to express my sincere appreciation for the exceptional service provided by your team. Your accommodating nature, especially when it comes to handling out-of-the-ordinary requests, has been incredibly impressive and has significantly enhanced my experience.