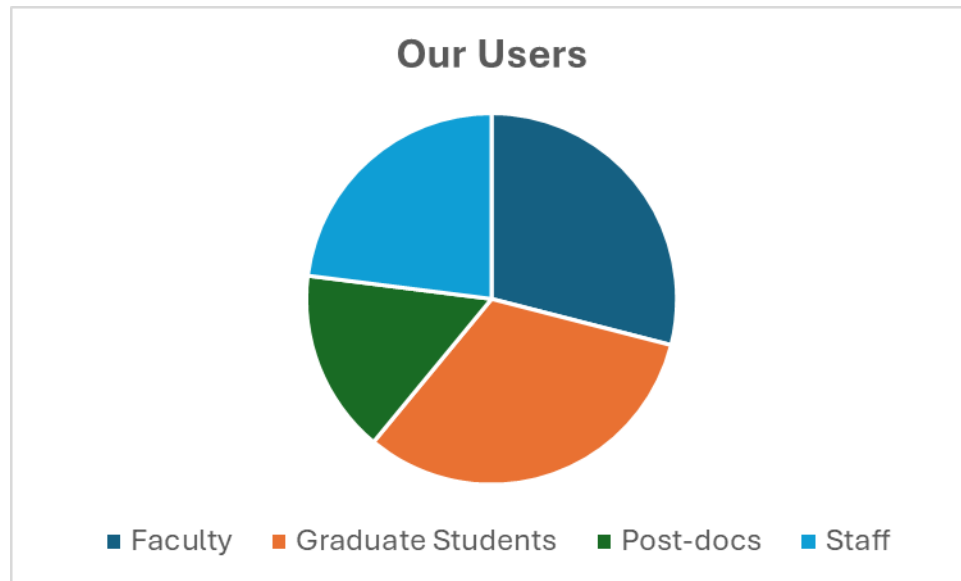


CRC 2023-2024 User Survey

We would like to thank all CRC's collaborators who took the time and effort to provide us with thoughtful and valuable feedback about our services. We invite you to review this summary of the highlights and responses.

Our Users



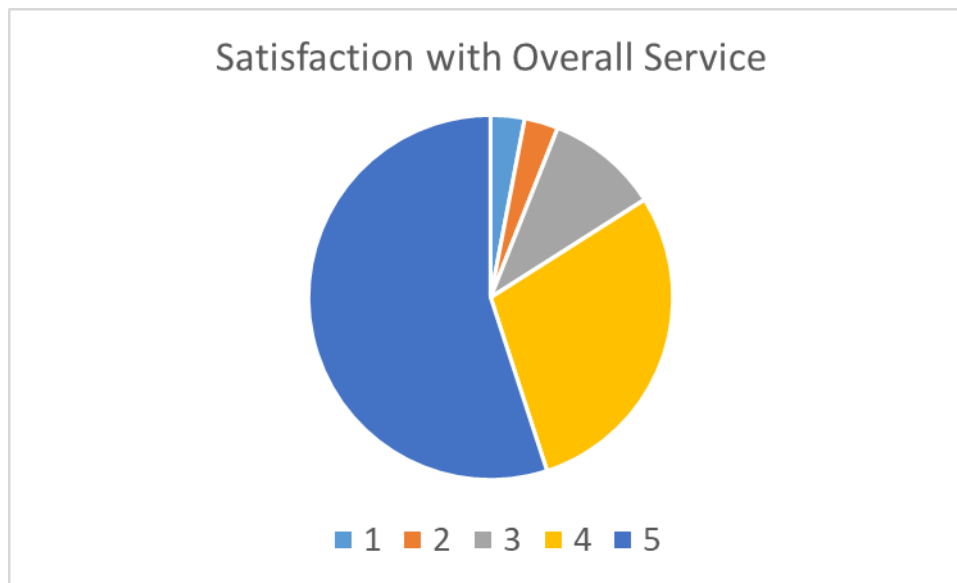
What Clusters They Use

mpi	26%
smp	35%
htc	55%
GPU	32%

(Please note: users may work on more than one cluster. These values represent relative usage; they do not add up to 100 percent of usage).

Are they satisfied with the overall service of CRC?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.

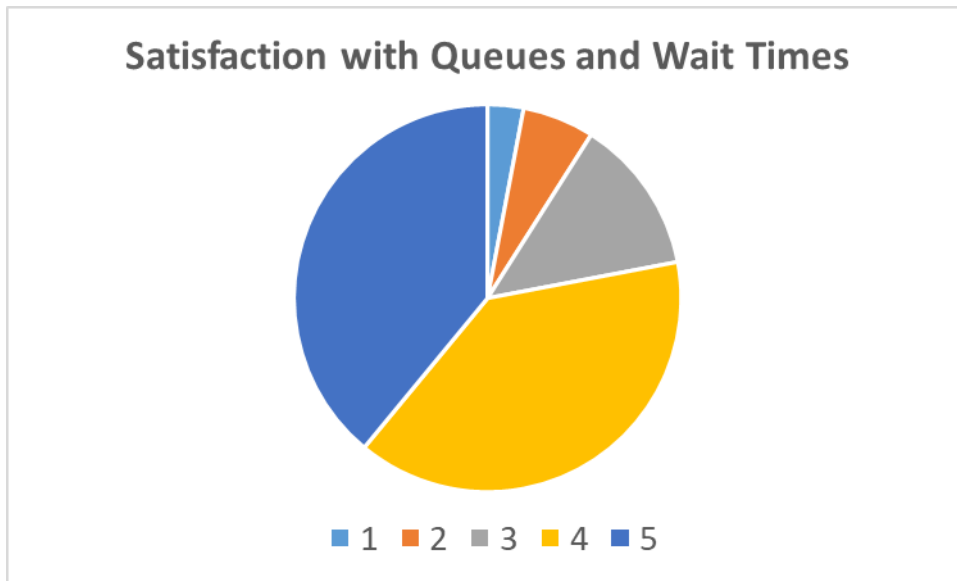


Some comments on issues we are addressing.

- Occasionally undergrads report having a hard time with a help ticket.
- Complexities of SLURM and lack of cluster availability make some users prefer working locally.
- Resources not available when needed.
- Cluster starts to break down a month before the scheduled maintenance.
- Access to storage on bgfs is very slow and makes many tasks frustratingly slow.
- Limited space on the home folder so many essential files (e.g. anaconda environment folders) have to live on bgfs.
- Long lag time on allocation requests
- Policies are not set appropriately to balance utilization and access.

How satisfied are users with access, queues and wait times?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.



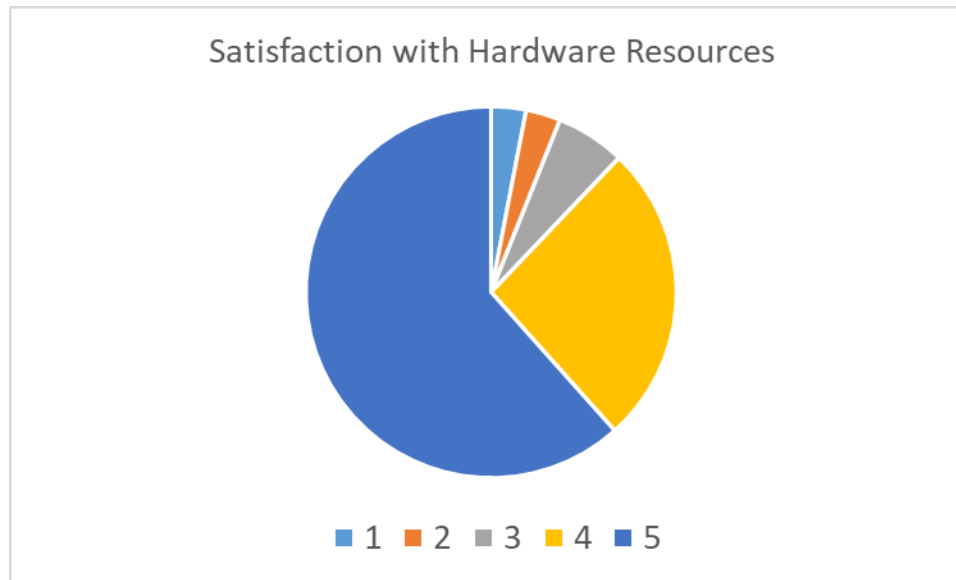
1

Some comments on issues we are addressing.

- The new A100-multi cluster doesn't allow job submission for more than 4 nodes. Allowing short jobs that request more nodes would be appropriate for that partition, since there are already plenty of GPUs allocated on the other partitions for single-node jobs.
- More GPU resources are likely needed.
- Access availability still has some limitations.
- Extremely low job allowing per user. Can't get enough ram.
- Wait times are opaque – there is no indication if it's going to open up in 5 seconds or 5 hours. It would be great to have an indicator (like experiencing high volume, expect wait times) or to be able to get an email alert when a session/node is ready if it's not able to be created right away.
- Larger GPU jobs often still take a long time to get allocated.
- Sometimes one person's submission (several hundred at one time)overwhelmed the system, and it took hours for my submission to start. This is becoming more frequent.
- Mpi is a busy cluster, and it is often tricky to access nodes.
- Sometimes users submit a ton of jobs all at once and it prevents others from running small interactive sessions or jobs.
- When job resources are waiting to run despite resources being available, the situation should be treated y as seriously as a node going down. There should be a target of a *minimum* of 80% average utilization. The average utilization of the a100s over the last 4 months is an abysmal 42%.

How satisfied are users with the hardware resources available at CRC?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.

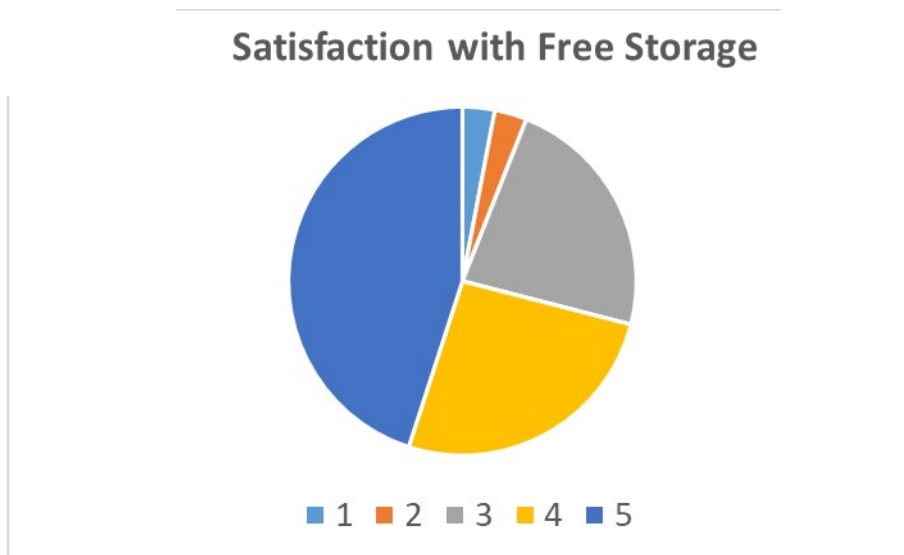


Some comments on issues we are addressing.

- CRC is great for small-scale GPU jobs, but it is missing a "middle tier." I say this as someone who has access to national resources for big jobs, and my own 20-GPU cluster at CRC; for me, the missing resource is something in between.
- GPUs are not the cutting edge.
- have had no issues with hardware limitations.
- additional accelerators
- PGA services could be improved.
- Need more gpu a100's.
- bgfs storage has slow read speed.
- More is always better.

How are users satisfied with CRC's free storage?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.

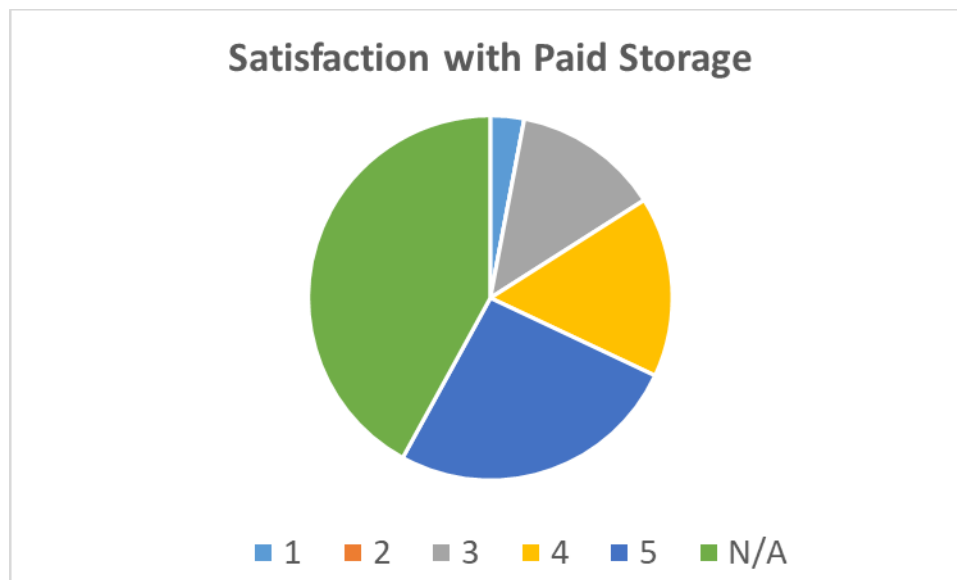


Some comments on issues we are addressing.

- Could be more generous (Our admin at DFCI Data Sciences believed storage limits to be un-academic).
- The type and amount of storage available to users could be better clarified.
- Should be per user than per group.
- Ended up purchasing more to be able to hold all the data we are working with."
- No clear way to see what storage is available in real time.
- If an account is full and needs to delete material, there's no clear way to see an updated version of how much storage is available without the system updating. if I delete large files, it isn't reflected immediately).
- Larger home drive size would relieve pressure to constantly remove conda environments and other files from home
- bgfs system is mostly full and the ix system seems to be slow.

How satisfied are users with the paid storage available at CRC?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.

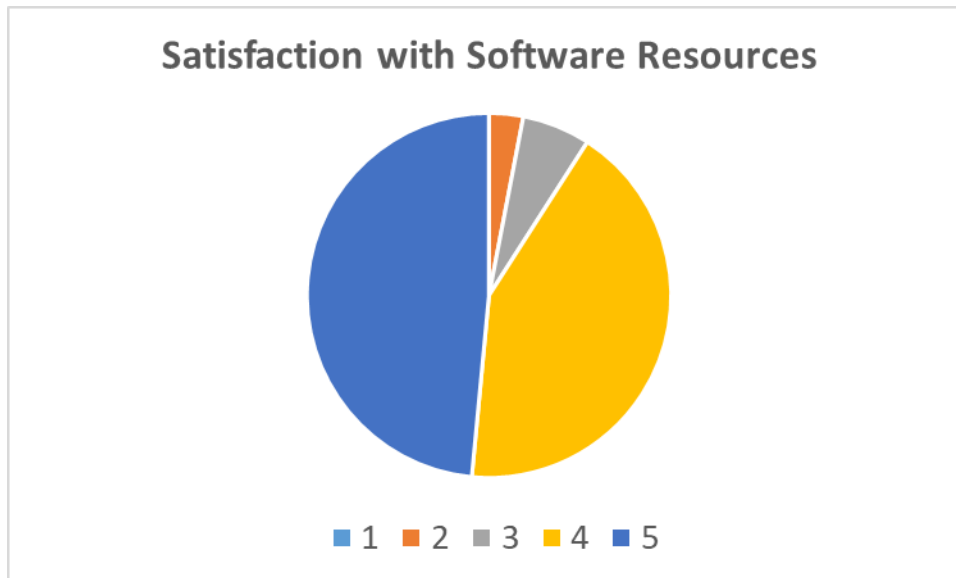


Some comments on issues we are addressing.

- Rates are a little high for our grant budgets.
- CRC does not implement a transparent or obvious data redundancy / data backup policy on /ix etc.
- BGFS is not fast enough (read speed) to support our machine learning workflows which access large data sets.
- Better if the paid storage was backed up.
- Couldn't purchase more storage on bgfs

How satisfied are users with the software resources available at CRC?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.

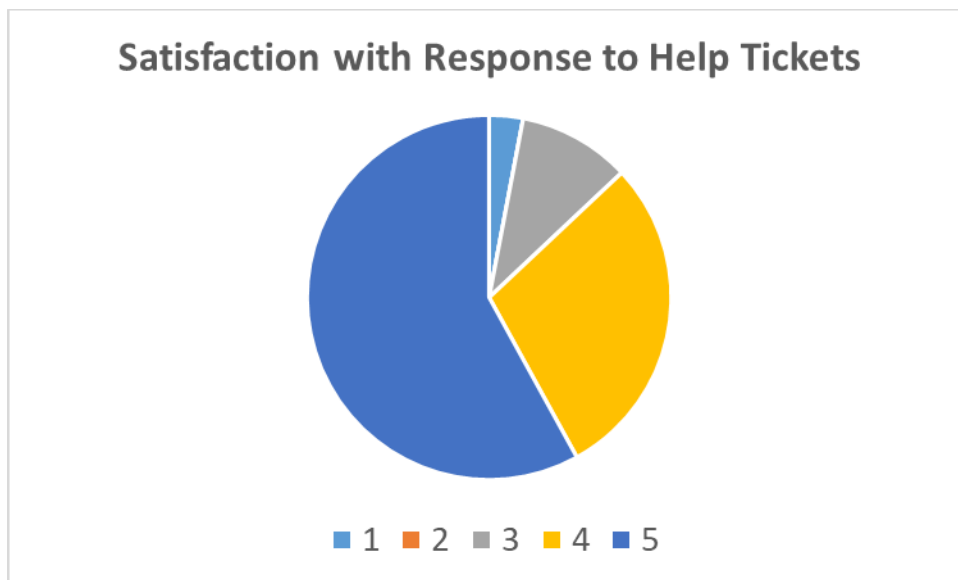


Some comments on issues we are addressing.

- Occasionally have to wait awhile for something to be installed.
- Need more tutorials on distributed computing on cluster. DASK/ Apache and also distributed computing on GPU's.
- "More information about tutorials/best practices/ FAQs especially for using Linux. Example: running R scripts as batch jobs.
- Not all software is listed in the module spider description and unless you know where to find it in the system, it isn't available. There are a lot of updated genomic softwares available in /ihome/crc/install and new people don't know to look there.

How satisfied are users with CRC support in responding to Help tickets?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.



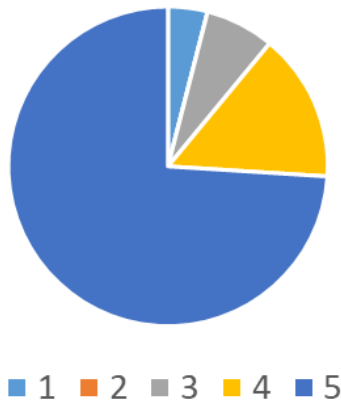
Some comments on issues we are addressing.

- A gentler approach with students who are learning the cluster ecosystem would be appreciated.
- The tone of response is often kind of terse and frustrated. Users asking “dumb questions” could be an opportunity to make better documentation.
- Sometimes condescending and terse
- Response is sometimes a bit slow.
- Response is a bit uneven: sometimes several days, sometimes only a few hours.
- Help tickets are usually responded to quickly and effectively.

How satisfied are users with CRC support by individual consultants?

Responses are rated from 1-5, with 5 as the highest satisfaction and 1 as the lowest.

Satisfaction with Support by Individual Consultants



What suggestions do users have for improving the services of CRC to meet the needs of your community? For example, are there gaps in our software offerings, technology, or workflow modalities that CRC should prioritize addressing?

Some comments on issues we are addressing.

- Data backup solution would be welcome. Paid options to host tools that can reach external audience would be worth investigating.
- Performing maintenance over weekends would minimize disruptions.
- Increase the tutorials as well as the hardwares.
- Default clusters are mainly busy. Will it be possible to have default to all the partitions and we can specify a particular one if required. For example on gpu, a100 are always busy, but gtx are free or a100 multi are free. A lot of people do not know this which leads to increase in the queue time of a100 for people who actually need that =).
<https://guix.gnu.org> package management; automated MFA
- Improve documentation/FAQ style support for common workflows, especially best practices for using linux in setting of the cluster. I
- Ability to look up a piece of information than commit to a 3 hour class.
- Large data storage (~200 Tb) with fast read speeds.
- Keep a record of who is using what software so that it might be possible to contact the users for help.
- Updating documentation: documentation is sometimes hard to find.

- Brief quick start guide for new students who can't spend the time attending the onboarding workshop and just need quick answers to simple questions.
- User manual should include the recommended commands to see the current allocation usage to check more regularly to calculate how much more allocation we might need to request.
- Scratch file system for medium term storage that is not backed up.
- Transition of PittIT to TeamDynamix is a giant leap forward because their previous system was absolute trash. The current CRC system serves its community well. My limited experience with TeamDynamix indicates there are still issues (couldn't access my tickets after submitting them, didn't consistently receive notifications). Please make sure these issues are resolved and we aren't losing features of the current system before switching.