

# CloudLab4 User Needs

## Summary

Methodological notes

## Demographics and usage patterns

Q1. About how often have you visited CloudLab in the past month? (n = 52)

Q2. Are you a supervisor for any CloudLab users? (n = 49)

Q3. How do you use CloudLab? Select all that apply. (n = 52)

Q4. What topic do you research? (n = 48)

Q5a. Do you use cloud computing services other than CloudLab? (n = 48)

Q5b. Which other cloud computing services do you use? (n = 20)

## Needs

Q6 & Q7. How well has the CloudLab hardware you have used in recent research met your needs? (n = 42) and In your recent research, how well has the hardware you have accessed through other cloud computing services met your needs? (n = 21)

Q8. What hardware needs have not been met by CloudLab? (n = 25)

Q9. What hardware needs have not been met by the other cloud computing services you use? (n = 11)

Q10. What types of hardware would be useful to your research in the next three years? (n = 37)

Q11. What software packages would you like CloudLab to offer specialized support for? (n = 22)

Q12. What tools would you like to use to deploy, orchestrate, or monitor experiments on CloudLab? (n = 22)

## Priorities and concerns

Q13. Please rank the following topics from your greatest to least concern with respect to the cloud-enabled research you conduct (with or without CloudLab). (n = 44)

Q14. Please indicate how much you agree with the following statements about your use of CloudLab.

## Participant write in

## Summary

To identify user needs and prioritize updates to CloudLab, we distributed a survey to the testbed's users. Data were collected from 56 participants. This document presents results question by question.

Participants were frequent visitors to CloudLab who use it in their research and often their mentee's research. They study cloud computing, operating systems, networking, storage and data management, and security.

Results indicate that hardware availability is CloudLab users' top concern. Participants report that CloudLab has met their hardware needs, however, users want more nodes, newer CPUs, more GPUs, and more CXL devices. Most users find the number and type of nodes they need to be unavailable despite also believing that CloudLab offers the kind of hardware they need. These results indicate that an update to and increase in hardware is warranted.

Another area for improvement is the profile feature. Existing profiles meet many users' needs, but about one quarter of respondents disagreed that existing profiles met their needs. It was not uncommon for participants to have trouble creating new profiles—one user wrote in to say so in addition to their survey responses. Users further report that the accessibility of their research artifacts is highly important to them; by offering new profiles and emphasizing their utility in sharing research artifacts, CloudLab can better support users.

While the documentation was helpful to participants and users indicated that they can get useful help quickly, onboarding may be difficult for some people. One third of participants reported that they had to learn a lot to

1  
2  
2  
3  
3  
3  
4  
4  
4  
5  
5  
6  
6  
7  
7  
8  
8  
10  
16

get going with the platform. Further clarity could also be offered around reservation and extension policies—about a quarter of our participants (who are often supervisors and visit CloudLab frequently) do not fully understand how requests are approved or denied.

Chameleon, another free service, is CloudLab's top competitor within our sample, followed by AWS. Users report that other cloud providers do not offer bare metal access at a low cost like CloudLab does. Cost is of moderate concern to our participants. Maintaining a free service is likely important for continued success.

While the reservation system has been a key topic in prior user studies of CloudLab, participants in this survey report that they can identify available nodes and create reservations with ease. Moreover, they report being happy with the reservation system (though somewhat confused by procedures for approving reservations) and getting their approval quickly. While clarifying or emphasizing policy would be useful, results suggest changes to the reservation workflow are not as important as making new nodes available to users.

Energy usage was users' lowest priority. While a trending topic and important consideration for infrastructure providers, results suggest this is not an area CloudLab needs to prioritize right now. Likewise, security is a low priority to users and they are already confident in their experiments' security as is.

## Methodological notes

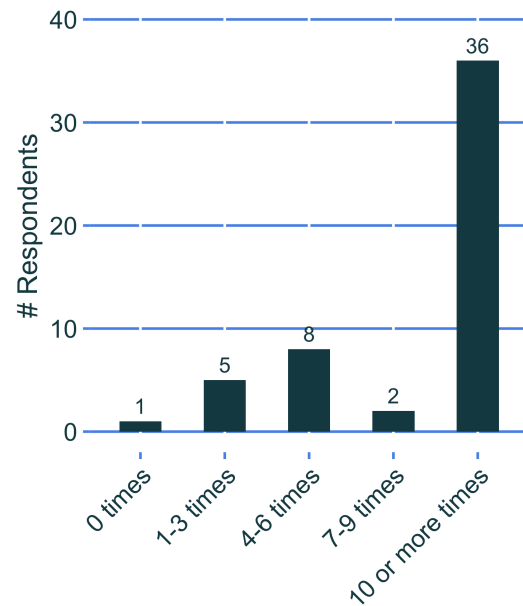
- The data presented is the data that has been collected as of 12:30PM March 21, 2024. Data from 56 participants are included. Not all participants answered all questions (see point below). The data includes surveys that were begun but not submitted ( $n = 14$ ). Because of the responses to text entry questions, the number of questions answered, the unique combinations of responses provided, and the unique IP addresses of participants, and the date of their entry, these unsubmitted responses were interpreted as legitimate data.
- The three sections of results in this document (*Demographics and usage patterns*, *Needs*, and *Priorities and concerns*) represent blocks of questions in the survey. The order of these blocks' presentation to participants was randomized to prevent the same block from going unanswered due to fatigue and to mitigate any bias that a question might introduce. Nevertheless, some questions proved more likely to go unanswered than others. The long table of Likert style questions in *Priorities and concerns* and the free text response questions in *Needs* likely dissuaded some potential respondents from completing the questionnaire. Questions in *Demographics and usage patterns* had 5 unexpected NAs on average. Questions in *Needs* had 16.1 unexpected NAs on average. Questions in *Priorities and concerns* had 17.05 unexpected NAs on average.
- The survey was distributed to the CloudLab User Feedback Group, a community of 43 people. Among that sample population, eight surveys were started and 4 surveys were finished—a 19% response rate. In May 2023 that group was comprised of only 13 people, eight of whom responded to a survey to evaluate a visualization that could inform users about their reservations. In that survey, all eight participants completed their surveys—a 62% response rate.
- One participant wrote an email to share additional thoughts because there was no free response option at the end of the survey. This email is included at the end of this document.
- As noted in Q2, of the 32 people who specified they are not supervisors for other CloudLab users, seven still report using CloudLab in their mentees' research (22%; see Q3). This somewhat contrarian result suggests the terms supervisors and mentees may not have been well understood and more piloting should be done to facilitate comprehension in for future surveys that cover these topics.

## Demographics and usage patterns

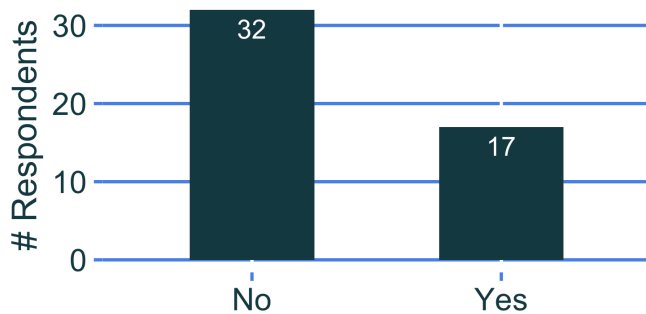
### Q1. About how often have you visited CloudLab in the past month? (n = 52)

Among our respondents, most had visited at least ten times in the past month.

The ranges that participants could choose from was meant to be informed by Google Analytics. However, with the transition from Universal Analytics to GA4, this data is no longer available. GA4 does report that the average CloudLab user made 3.7 visits to the testbed in February 2024. If true, this suggests our survey participants are especially frequent visitors to CloudLab.

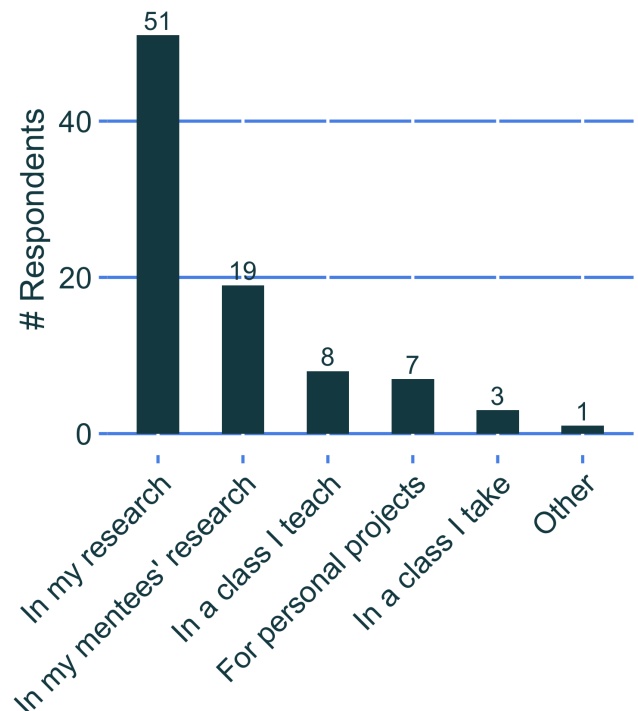


### Q2. Are you a supervisor for any CloudLab users? (n = 49)



Most participants are not supervisors of other CloudLab users (n = 32; see Q2). Of the 17 supervisors for other CloudLab users, ten use CloudLab in their mentees' research (59%; see Q3). Of the 32 people who specified they are not supervisors for other CloudLab users, seven still report using CloudLab in their mentees' research (22%; see Q3). The participant who reported using CloudLab for "other" purposes wrote that they use the testbed in their capacity as a research intern; they also responded that they use it in their research.

### Q3. How do you use CloudLab? Select all that apply. (n = 52)



#### Q4. What topic do you research? (n = 48)

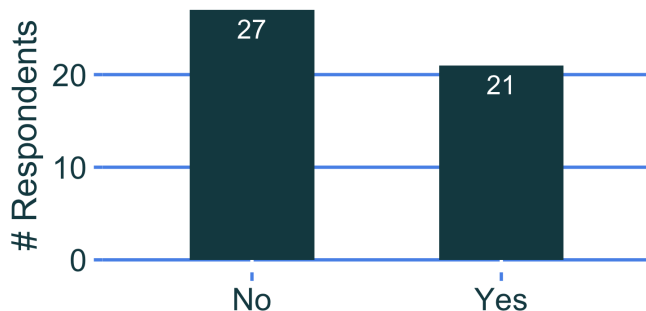
Common topics include cloud computing and infrastructure, operating systems, networking, storage and data management, and security.

All responses:

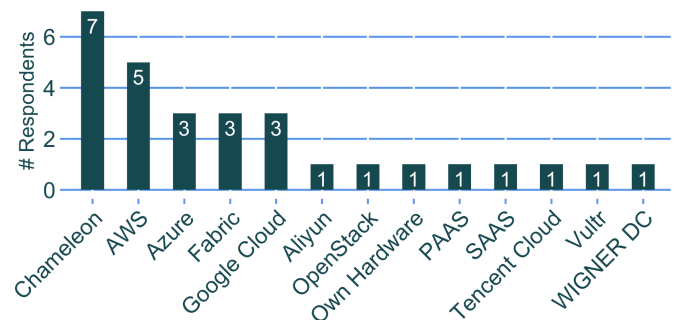
- low latency networking
- Clouds
- Cloud Edge Computing
- networking
- memory system
- Delegation Styled Lock
- Operating Systems and Computer Networks
- Operating Systems
- Cloud Computing
- programming languages
- System interconnect
- Distribute systems, in-memory storage
- Operating Systems, Security and Isolation, Distributed Systems
- Computer systems and networks
- RDMA storage
- DB transaction contention
- Network transport protocols, datacenter computing
- Operating Systems and Networking
- Serverless
- Distributed Storage
- Deployment of HPC inside Cloud infrastructures
- Related to OS and Architecture
- Distributed systems, OS, RDMA-based systems
- Operating System and Storage System
- Computer Architecture
- Operating systems, distributed systems
- Disaggregated Memory
- Network telemetry
- Low-level optimizations and networks
- approximate query processing
- Network Security solve using deep learning
- Operating Systems, Compilers , Disaggregated memory (Make Far memory efficient using OS/ compiler techniques)
- Distributed system, memory & storage system, network, virtualization, cloud
- WebAssembly, TEE
- Systems, storage
- System Security
- distributed computing
- OS/DB/Distributed system
- Distributed computing
- database, machine learning
- Cybersecurity Automation, Intrusion Detection, Intrusion Response, LLM
- Improve packet processing speed at the end-host
- RDMA/CXL-based disaggregated memory system
- Distributed databases
- data management using computational storage devices
- Atomistic simulations
- Virtualization, Cloud, 5G, 6G, Edge networking
- Cloud Computing

#### Q5a. Do you use cloud computing services other than CloudLab? (n = 48)

CloudLab is most participants' only cloud computing service. About 44% (n = 21) of participants do use another service, the most common of which is Chameleon. Six participants reported using multiple other cloud services.



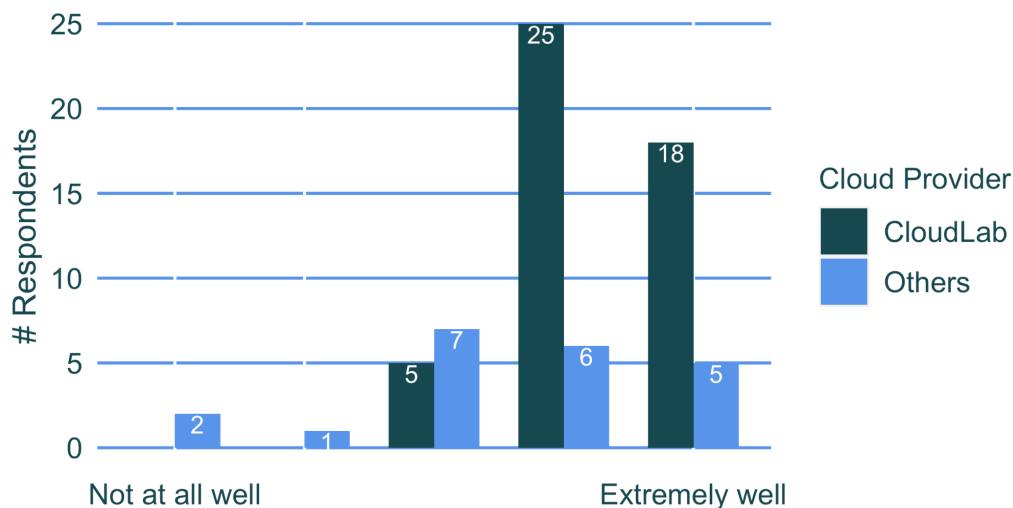
#### Q5b. Which other cloud computing services do you use? (n = 20)



## Needs

**Q6 & Q7. How well has the CloudLab hardware you have used in recent research met your needs? (n = 48) and In your recent research, how well has the hardware you have accessed through other cloud computing services met your needs? (n = 21)**

Using a 1-5 scale from “not well at all” to “extremely well,” participants reported that the CloudLab hardware met their needs very well (mean = 4.3). Participants who also use other hardware indicate that that hardware did not meet their needs as well as CloudLab; the mean response for other hardware was 3.5.



### Q8. What hardware needs have not been met by CloudLab? (n = 25)

Participants who did not indicate in Q6 that CloudLab met their needs “extremely well” were asked which needs had not been met. Two of those participants did not provide further detail (e.g., responding “NA”). Other responses indicate that users want more and updated hardware with more GPUs and CXL devices. The following is a list of unmet needs participants reported.

- One account added to two projects
- Lack of widespread GPUs
- More GPU nodes
- Large vMem GPU instance
- GPU availability in abundance
- More Intel machines, and more SmartNICs
- Not all NICs support the technologies we use, specifically XDP driver and HW timestamping. Also, it's impossible to test XDP offloading on cloudlab.
- Newer processors, NICs, GPUs
- 200 Gbps network, with more nodes
- Maybe newer cores
- CloudLab desperately needs a new large cluster with Intel processors and 100 GB networking. Think of this as an upgrade of the xl170 cluster. The c6525-100g cluster is inadequate both because of its small size and because its AMD processors are inferior to Intel processors for the work I do (Homa latency is about twice as high with AMD processors!)
- Better CPU and DRAM.
- small cpus
- Cloudlab has lots of very old CPU. The newer ones(skylake and later) are hard to reserve. Our research also requires to access machine with two or more sockets which has limited availability on CloudLab. We also need network card with SRIOV to run our simulation but the availability is low. Also the high bandwidth nic is not enough.
- Non NUMA machines
- Zone Namaspaced SSD, Remote DMA
- nvme, and large dram
- CXL-based memory

- CXL devices(e.g. memory and switch); PIM Process-in-memory
- I'd like to have CXL devices (although it's very rare nowadays so it's totally understandable that CloudLab doesn't have it). Besides specific hardware, sometimes it happened that I needed standard hardware (e.g. no GPUs/powerful NICs, just some regular CPUs and memory) and there weren't enough available machines. "
- Connectx4 network cards with higher throughput and latency, Hardware with features such as Intel MPX, Linear address masking.
- You have the hardware for AMD-SEV SNP, but its not enabled in the bios/uefi.
- I would be interested in various types of smart drives or configurations that can mimic such a device

### **Q9. What hardware needs have not been met by the other cloud computing services you use? (n = 11)**

Participants who indicated in Q7 that the other hardware they use does not meet their needs “extremely well” were asked Q8. One person responded “none.” Other participants reported that other hardware fails to meet some of the same needs as CloudLab; ConnectX-4 network cards, zoned namespaces, remote DMA, and desirable NICs are reportedly unavailable from all providers. One participant reported other services fail to meet some needs that CloudLab does: bare metal servers and a low cost. The following is a complete list of needs participants reported as unmet by services other than CloudLab:

- Zone Namespaced SSD, Remote DMA
- chameleon tutorials recommended a jupyter notebook interface. it was hard to use, kept losing connection and experiment status when i stepped away.
- Connect x4 cards, with Infiniband switches/ethernet switches.
- GPU; RDMA NIC; Bluefield 2
- bare metal servers, the expensive fee
- High bandwidth nic(100-200 Gbps nic) and smart nic.
- Cumbersome setup for ssh key access
- different machines and hardware
- Reachability from Internet
- RDMA NIC

### **Q10. What types of hardware would be useful to your research in the next three years? (n = 37)**

Overlapping with Q8 and Q9's results, participants report that their future research needs more powerful CPUs, new NICs, CXL devices, and more GPUs. One person responded NA. The following is a complete list of desired hardware:

- User interrupt supporting ones
- Remote DMA
- AMD Xilinx's Alveo series
- Hardware with features such as Intel MPX, Linear address masking, Any sort of memory extensions such as RISC capability hardware would be useful for disaggregated, operating system research.
- AMD SEV SNP
- Modern server class hardware (the kind of servers a public cloud would be built on), GPUs
- r650
- See above: Intel CPUs, 100 Gbps networking
- New CPU cores, including x86\_64, ARM-based, RISC-based
- Powerful CPU cores.
- More network-oriented hardware (i.e. more and better NICs). Also, some NICs like Netronome Agilio that allow specific offloading which can be useful for certain research areas.
- High-speed NICs (e.g., 200 or 400 Gbps)
- More machines with advanced RDMA NICs like CX6/7, especially more Intel based machines.
- Many-core processor, the more CPU cores it has, the better.
- SmartNICs such as Bf2/3, and more 100Gbps NICs
- types of smart drives or configurations with other co-processors
- Programmable Switch?
- CXL cards
- CXL-enabled hardware, preferably CXL3.0

- Possibly DDR5 and CXL devices.
- CXL memory devices.
- RDMA NIC; DPU like Bluefield; Different types of memory devices; CXL devices; GPU
- nodes with large memory and disk
- CXL-based tiered memory
- It would be great if Cloudlab can get some devices with CXL memory or more recent CPUs like Saphire Rapids and Zen 4 CPUs.
- GPU instances with Mass storage
- GPUs
- More GPUs, FPGAs, SmartNICs
- Multiple GPUs
- Some servers with GPU resources may help in the research.
- GPU nodes, IPMI
- GPUs, FPGAs, programmable hardware
- Hardware with with GPU
- GPU, edge devices
- GPU
- GPUs

### Q11. What software packages would you like CloudLab to offer specialized support for? (n = 22)

Participants requested support for a variety of software. However, 11 of the 22 people who responded to this question reported that they do not need further support in this respect, either commenting things like “none” or “Good to go with current packages.” The following is a complete list of participants’ requests:

- Xilinx Design Suite
- Quantum Espresso, LAMMPS.
- cluster management, slurm?
- More storage 16GB, mounting block stores can be time consuming, saving restoring state, if this process could improve it would help,
- More Linux distributions such as Arch or Gentoo
- Machine learning development and visualization tool
- I want cloudlab to give a larger root partition for ubuntu20.04
- Deep learning platforms, such as PyTorch, tensor flow
- Kernel level modification
- Kubernetes
- Driver for Mellanox ConnectX NICs, which is possible to be installed at initialization

### Q12. What tools would you like to use to deploy, orchestrate, or monitor experiments on CloudLab? (n = 22)

Participants request a variety of tools including DevOps tools like Kubernetes and Ansible. Better documentation for existing support infrastructure is also welcome. Nine participants responded that they do not need further support in this respect, either commenting things like “none” or “Nothing temporarily. Now everything is good.”

- Xilinx Design Suite
- slurm?
- OpenStack
- Network Simulation tools
- Mobile support,
- Qemu, docker
- kubernetes
- Kubernetes, Prometheus
- vscode+tmux; MobaXterm
- Ansible Playbooks
- It would be useful having some customizable provisioning using Ansible automatically applied to an experiment.
- I don't have a preference for particular tools, but better "cookbooks" or "recipes" for using them with cloudlab would be nice (I have had trouble internalizing how to use tools to manage cloudlab experiments).
- "Making it easier to build images (dependency configuration, compiling my binaries, etc)
- Better or more documented API for launching experiments from scripts (I didn't know this was even possible until Rob linked me on mastodon to a random Python script on github. Until then I had written a chromedriver script for the web UI)

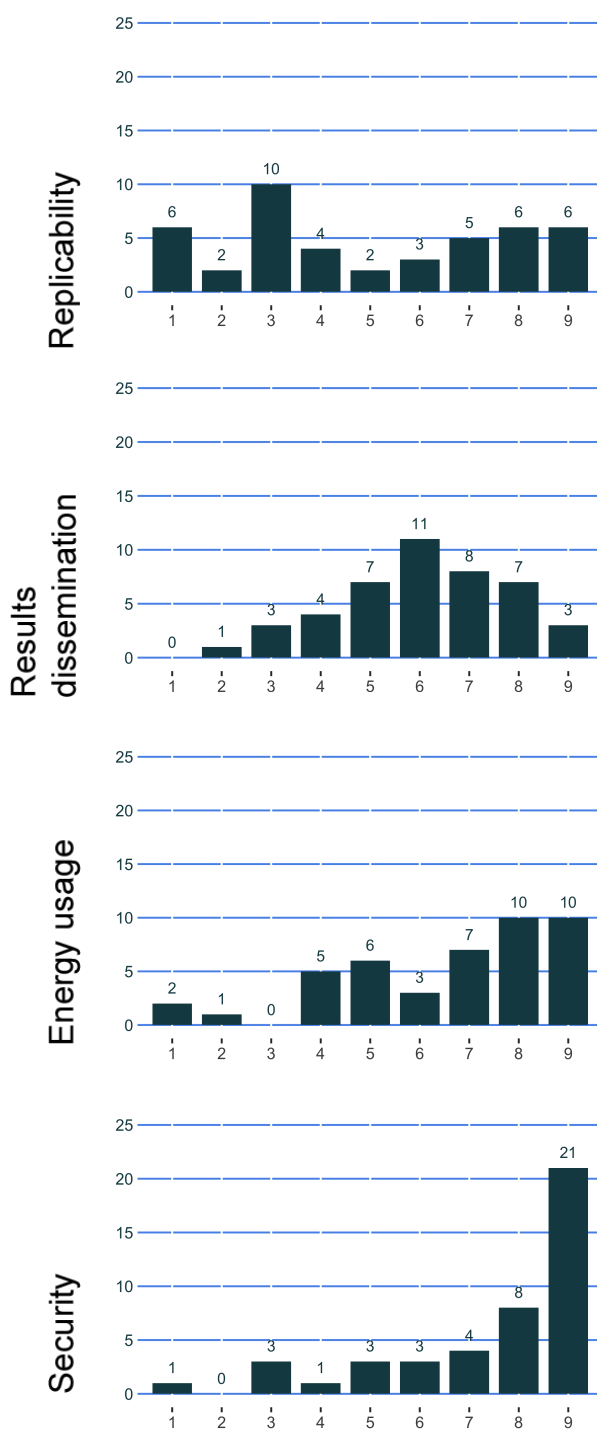
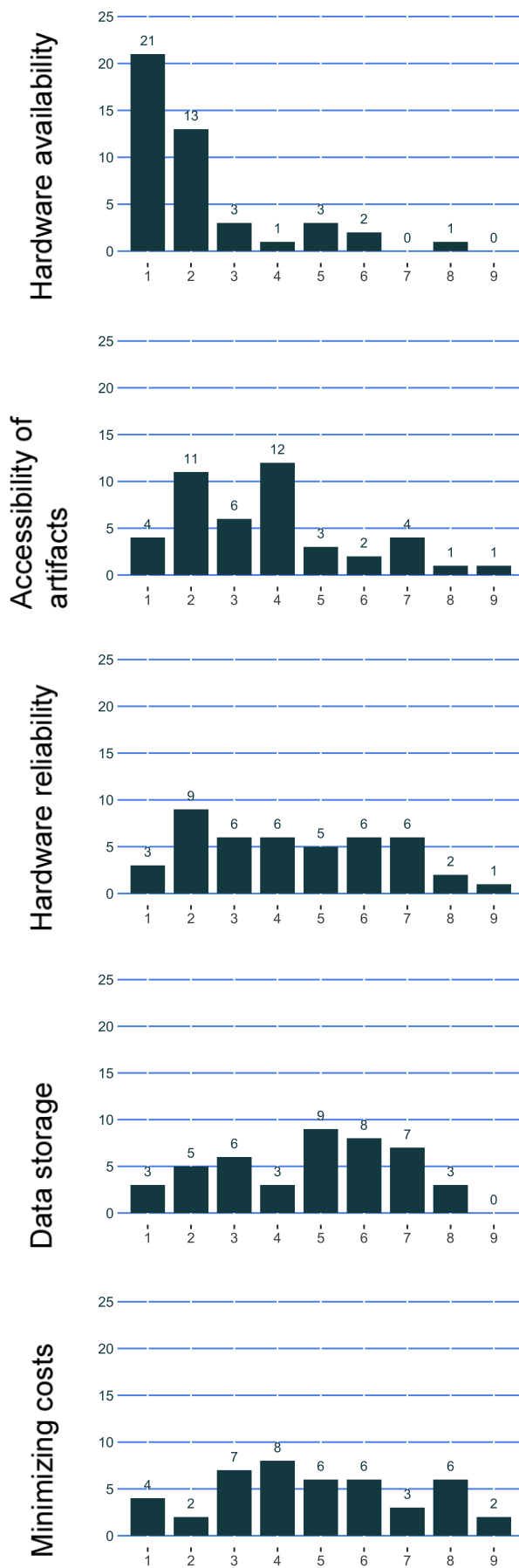
## Priorities and concerns

### Q13. Please rank the following topics from your greatest to least concern with respect to the cloud-enabled research you conduct (with or without CloudLab). (n = 44)

While some concerns were consistently of high priority or low priority, participants' rankings were non-normally distributed (see graphs below), indicating some segments have different concerns from others. Hardware availability is the top concern for nearly half of participants (n = 21, 48%) and security is often the lowest of our participants' concerns (n = 21, 48%). In contrast, as shown in the graphs below, the data distributions for artifact accessibility, data storage, costs, and replicability all show multiple peaks. The table below is ordered according to each concern's average ranking.

| Concern                                | Mean rank (low number is higher priority) |
|----------------------------------------|-------------------------------------------|
| Hardware availability                  | 2.16                                      |
| Accessibility of my research artifacts | 3.73                                      |
| Hardware reliability                   | 4.34                                      |
| Data storage                           | 4.75                                      |
| Minimizing costs                       | 4.86                                      |
| Replicability                          | 5.02                                      |
| Results dissemination                  | 6.07                                      |
| Energy usage                           | 6.61                                      |
| Security                               | 7.45                                      |

Q13. Distributions of rankings



Bars represent the number of participants who assigned the concern a given rank (1 - 9).  
x-axis: rank; y-axis: participants

**Q14. Please indicate how much you agree with the following statements about your use of CloudLab. Choose NA if the question is not applicable to you. (1, Strongly disagree - 5, Strongly agree)**

Using a scale of 1-5, participants reported their agreement with 31 statements about their experiences with CloudLab. Participants could respond NA if a question did not apply to them. Below, mean agreement scores, the number of responses, and graphs showing the distribution of participants' responses are provided for each statement.

Results show that participants are confident users of CloudLab, most of whom found they could get going with the system quickly. However, some participants felt the burden to entry was higher and would appreciate more support. Participants reported that the documentation answers their questions and is actionable. They also think they can get useful help quickly and easily.

Participants could accomplish key tasks like creating and monitoring experiments, determining when nodes are available, and creating reservations with ease. Most users report little confusion with respect to reservations, though some indicated they are indeed confused by the reservation system and do not understand how decisions about reservations are made. Users think reservation requests are approved quickly. They

are happy with the reservation system overall but a contingent of participants believe that some changes could help them work more efficiently.

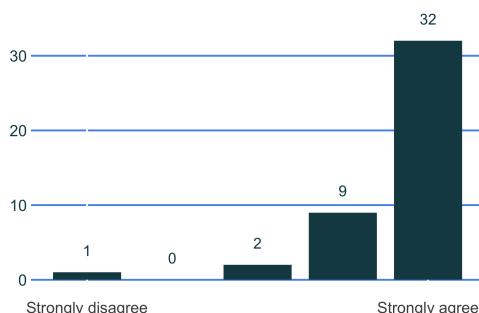
Response rates drop for the questions about getting help, suggesting that some participants have not sought help before (n = 27, 28, 29; see questions 23 through 27). Interestingly, response rates for the questions about documentation are high (n = 44; see questions 5 and 6). This is a hint that our participants are likely to reference the documentation.

Users' experiences working with datasets are varied—some find it easy to use and export data while others do not.

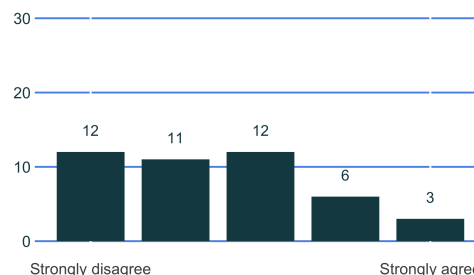
Results also show that existing profiles meet many users' needs, but about one third of respondents disagreed that existing profiles met their needs. It was not uncommon for participants to have trouble creating new profiles.

For some users the number of nodes they want are available but a majority report otherwise.

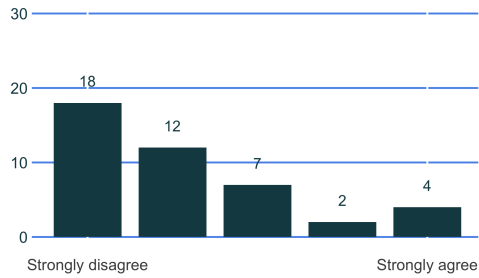
Responses were also negative with regard to the types of nodes—participants disagreed that the type they want are available when needed despite CloudLab having the kind of hardware desire.



1. I feel very confident using CloudLab.  
**mean = 4.61**  
n = 44



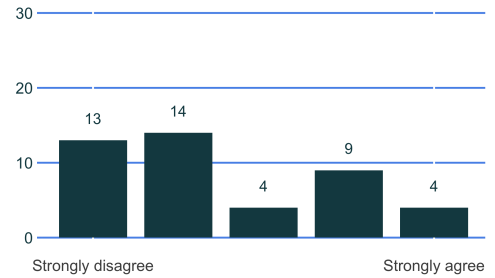
2. I want the support of a more technical person to be able to use CloudLab.  
**mean = 2.48**  
n = 44



3. I find CloudLab very cumbersome to use.

**mean = 2.12**

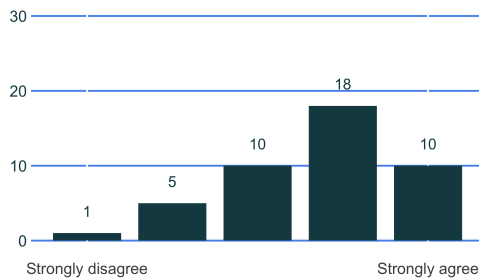
n = 43



4. I needed to learn a lot of things before I could get going with CloudLab.

**mean = 2.48**

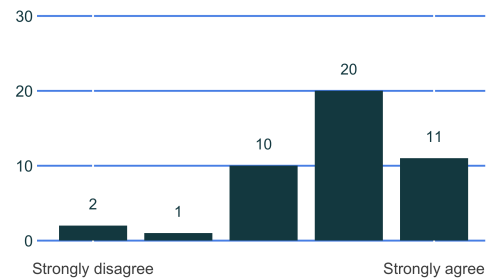
n = 44



5. The CloudLab documentation answers the questions I have.

**mean = 3.70**

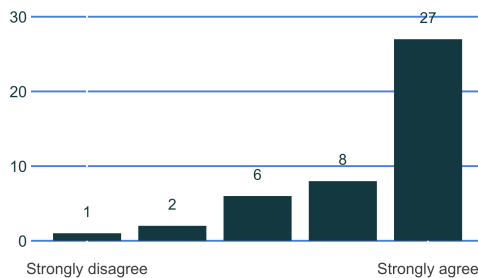
n = 44



6. I find it easy to act on what I have read in the CloudLab documentation.

**mean = 3.84**

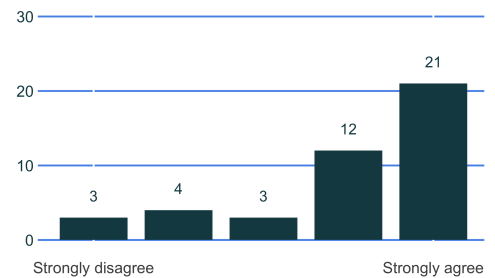
n = 44



7. By using CloudLab, I make efficient use of my time.

**mean = 4.32**

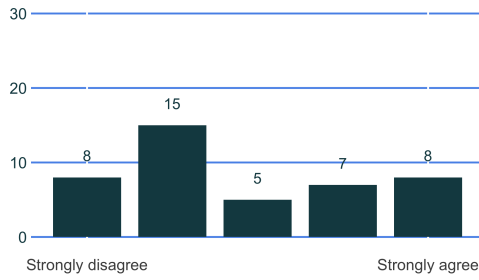
n = 44



8. I can easily determine which nodes are available and when.

**mean = 4.02**

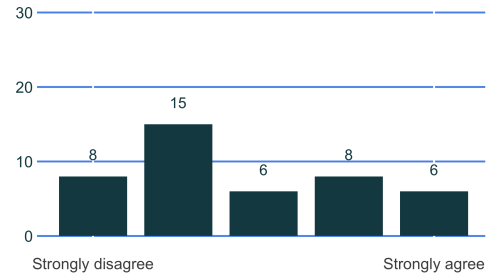
n = 43



9. The number of nodes I need are available when I want them.

**mean = 2.81**

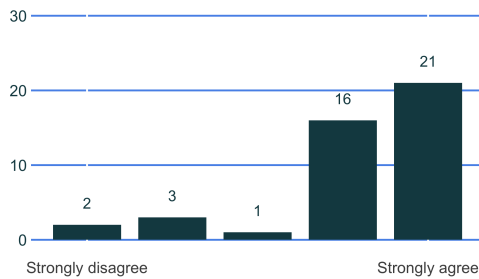
n = 43



10. The type of nodes I need are available when I want them.

**mean = 2.74**

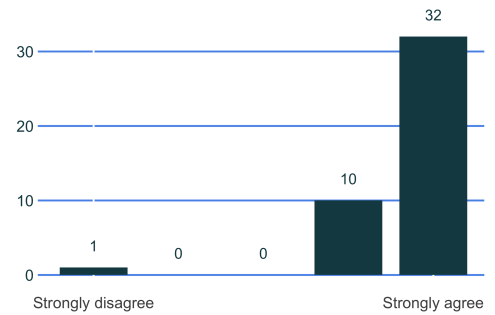
n = 43



11. CloudLab provides the kind of hardware I want to use in my research.

**mean = 4.19**

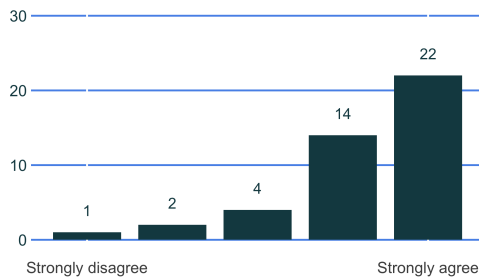
n = 43



12. I can easily create an experiment.

**mean = 4.67**

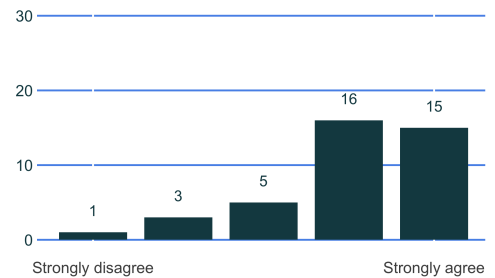
n = 43



13. I find monitoring my experiments easy.

**mean = 4.26**

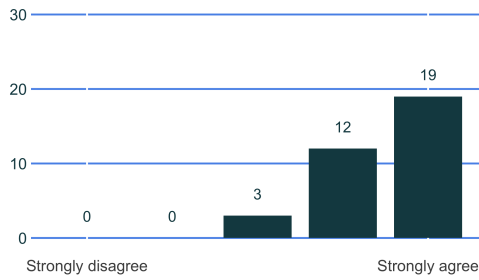
n = 43



14. I am confident in the security of my experiment.

**mean = 4.03**

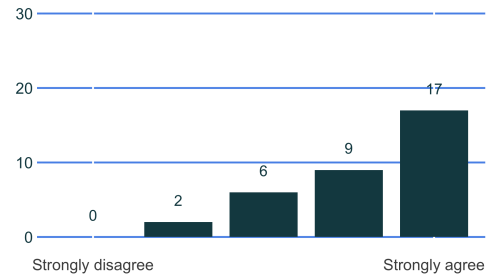
n = 40



**15.** I can easily create a reservation.

**mean = 4.47**

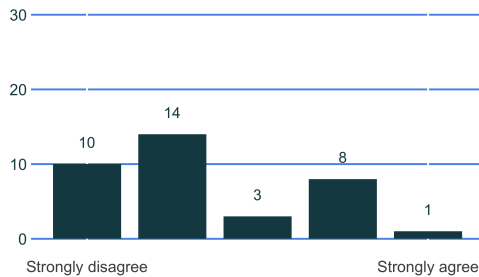
**n = 34**



**16.** My reservation requests are quickly approved.

**mean = 4.21**

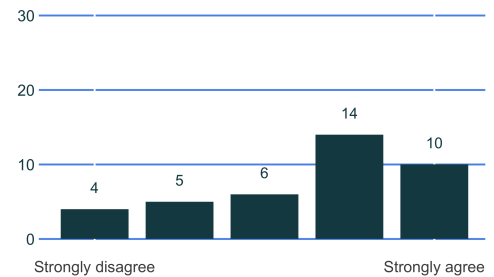
**n = 34**



**17.** I am confused by the reservation system.

**mean = 2.33**

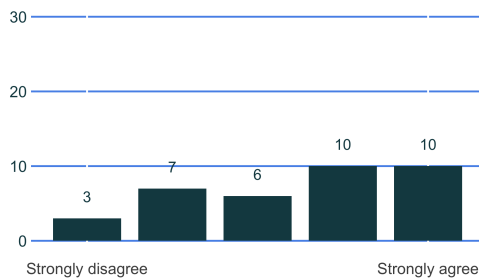
**n = 36**



**18.** I understand how decisions are made for extension requests.

**mean = 3.54**

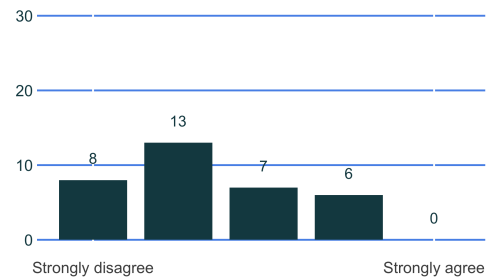
**n = 39**



**19.** I understand how decisions are made for reservation requests.

**mean = 3.47**

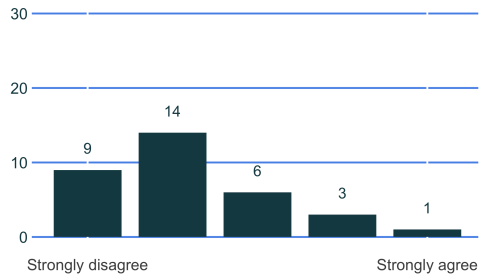
**n = 36**



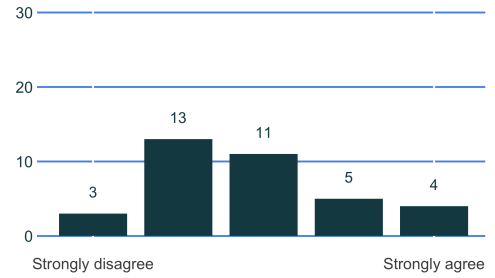
**20.** I have to wait a long time for my extension requests to be approved.

**mean = 2.32**

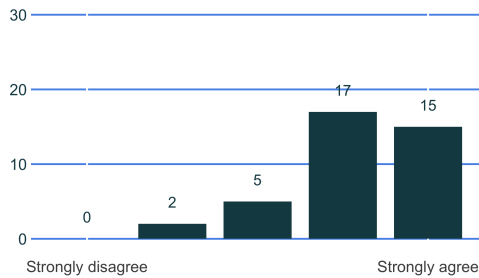
**n = 34**



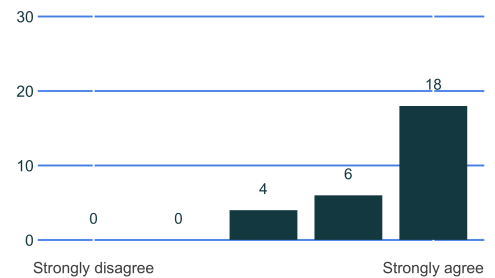
**21.** I have to wait a long time for my reservation requests to be approved.  
**mean = 2.18**  
 n = 33



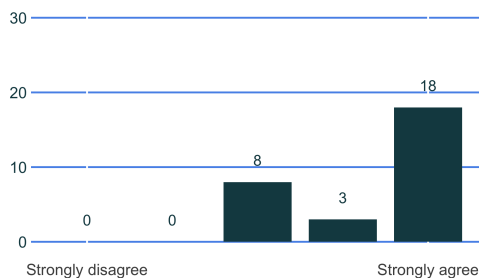
**22.** I have recently thought that if CloudLab made reasonable changes to their node allocation procedures, I could do my work more efficiently.  
**mean = 2.83; n = 36**



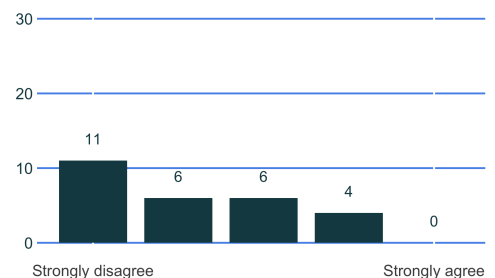
**23.** I am happy with CloudLab's system for reserving nodes.  
**mean = 4.15**  
 n = 39



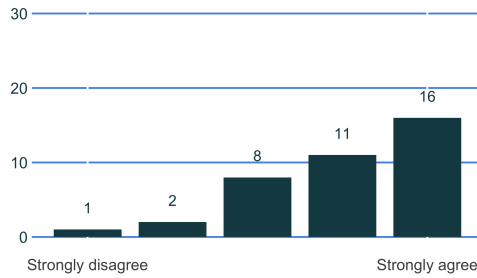
**24.** When I write in for help, I get a useful response.  
**mean = 4.5**  
 n = 28



**25.** It is easy to get help when I need it.  
**mean = 4.34**  
 n = 29



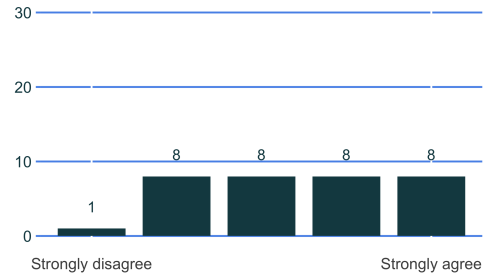
**26.** I have to wait a long time for my questions to be answered.  
**mean = 2.11**  
 n = 27



**27.** Adding or removing nodes from my experiment after it starts is a desirable feature.

**mean = 4.03**

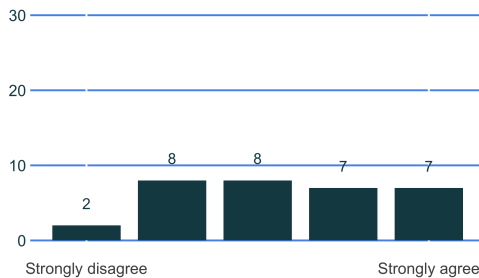
**n = 38**



**28.** I export experimental data with ease.

**mean = 3.42**

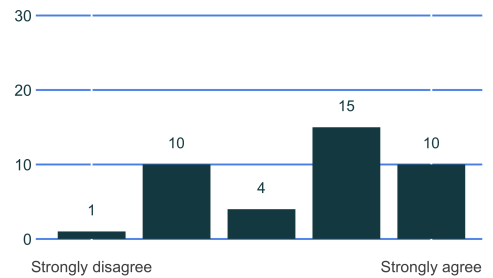
**n = 33**



**29.** I can use a dataset in my experiment with ease.

**mean = 3.28**

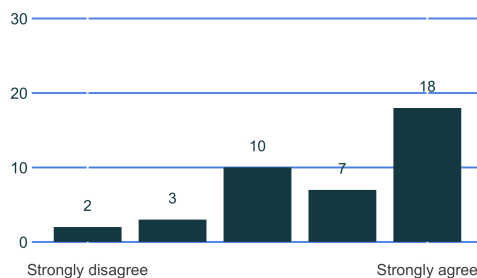
**n = 32**



**30.** The existing profiles from CloudLab meet the needs I have.

**mean = 3.58**

**n = 40**



**31.** I can easily create a new CloudLab profile.

**mean = 3.90**

**n = 40**

## Participant write in

One participant emailed a supplemental comment as an addendum to their survey. They report the difficulty of making a new profile and give an example from their experience. Terminology and workflows are an issue for this user:

[I]t is too hard to make a new profile!

""""

My goal was to make a new profile.

I tried logging in and clicking the "Profiles" tab that appears front and center, but there doesn't seem to be a way forward from there. Why not?!

What I ended up doing was click "Experiments" at the top (weird!) => Create Experiment Profile (why not "Create Profile"?). After that. success.

*End of document.*