2022 Cloud Mini-Survey Summary

Submitted December 2022

Scientific Computing Advisory Committee

Summary

As part of the call for white papers in support of the CfA strategic planning exercise (November 2022), the Scientific Computing Advisory Committee (SCAC) submitted two white papers: one on High Performance and Cloud Computing, and another on Data Management and IT Support. These white papers drew heavily from the SCAC's June 2022 memo to the Director's Office, but were augmented with the results from a cloud computing survey that the SCAC conducted in October 2022, in light of the potential benefits of further investment in this resource that have been identified throughout the community.

The cloud computing survey results reinforced the previous conclusions of the SCAC in its 2019 survey:
1. There is broad interest in cloud computing at CfA with 80% of respondents expressing interest and with about 40% of respondents actually using it.
2. 80% of respondents wished to learn more about the cloud, speaking to the fact that most people are aware of it but not fluent in its potential uses.
3. Most of this uncertainty surrounds lack of knowledge (80%) and costs (73%).

The results of the 2019 and 2022 surveys indicate that investment is needed in personnel who can understand cloud tools and guide users at the CfA. A better understanding of the costs of cloud computing relative to on-premises computing or HPC will be critical as well to ensure the most efficient and effective use of limited funds. Full results of the 2022 cloud computing survey are appended to this summary.
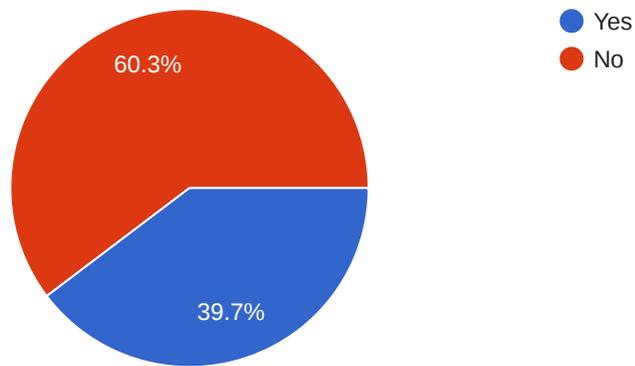
CfA SCAC Cloud Services Mini-Survey

63 responses
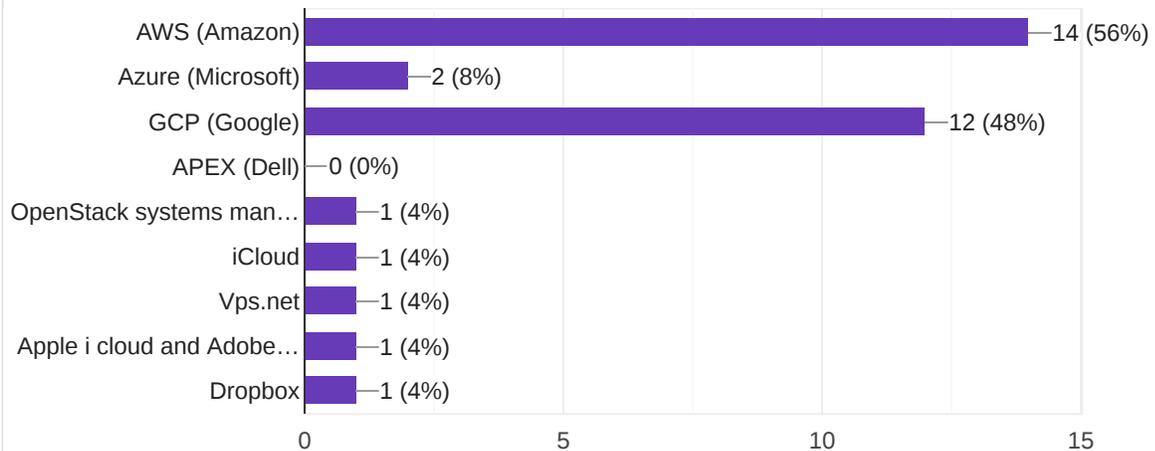
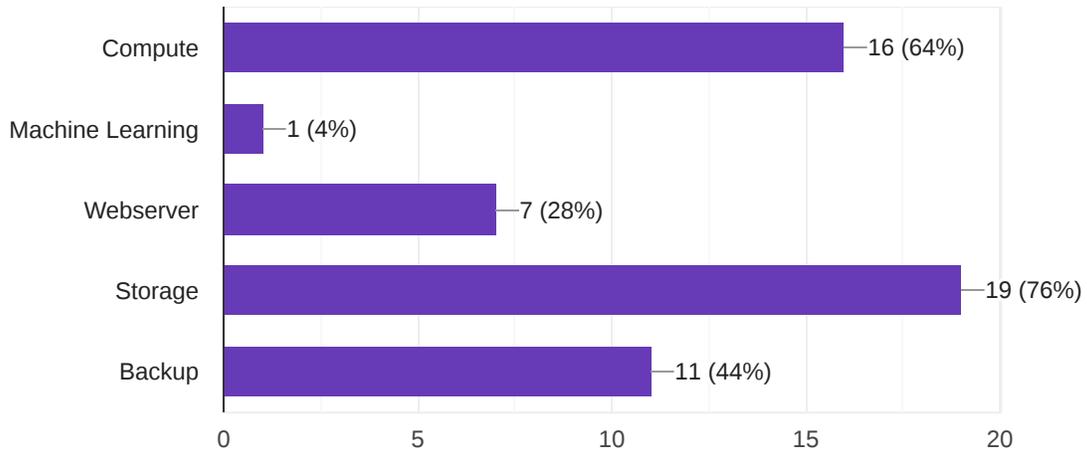## Have you or are you using cloud services?

63 responses

- 🔵 Yes
- 🔴 No

60.3%

39.7%

Cloud users

Which cloud(s)?

25 responses

Option	Count
AWS (Amazon)	14 (56%)
Azure (Microsoft)	2 (8%)
GCP (Google)	12 (48%)
APEX (Dell)	0 (0%)
OpenStack systems man...	1 (4%)
iCloud	1 (4%)
Vps.net	1 (4%)
Apple i cloud and Adobe...	1 (4%)
Dropbox	1 (4%)

What service(s)

25 responses

Service	Count
Compute	16 (64%)
Machine Learning	1 (4%)
Webserver	7 (28%)
Storage	19 (76%)
Backup	11 (44%)

How easy/difficult was/is it to use (1- easy, 5 - very difficult)

25 responses

Rating	Count
1	4 (16%)
2	4 (16%)
3	10 (40%)
4	6 (24%)
5	1 (4%)
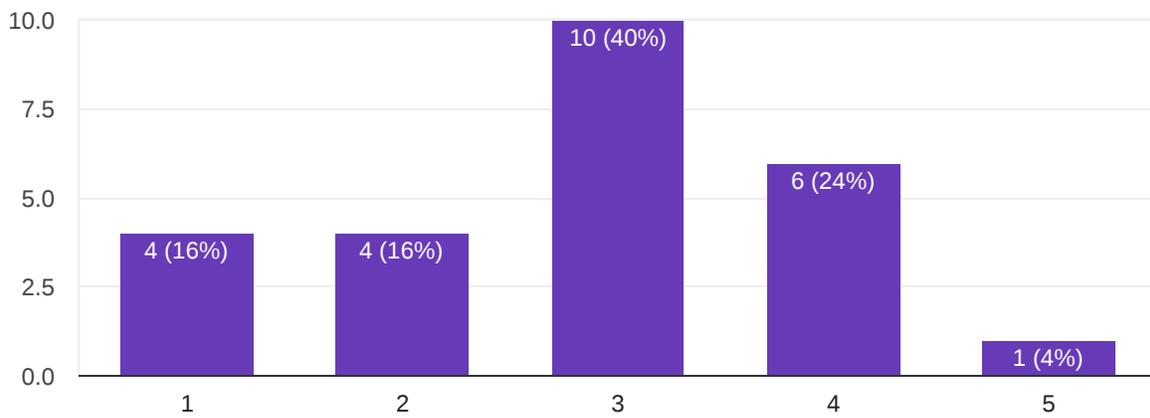
How about cost? (0 -free, 1- very reasonable, 5 - exorbitant)

24 responses

Rating	Count
0	4 (16.7%)
1	1 (4.2%)
2	6 (25%)
3	8 (33.3%)
4	3 (12.5%)
5	2 (8.3%)
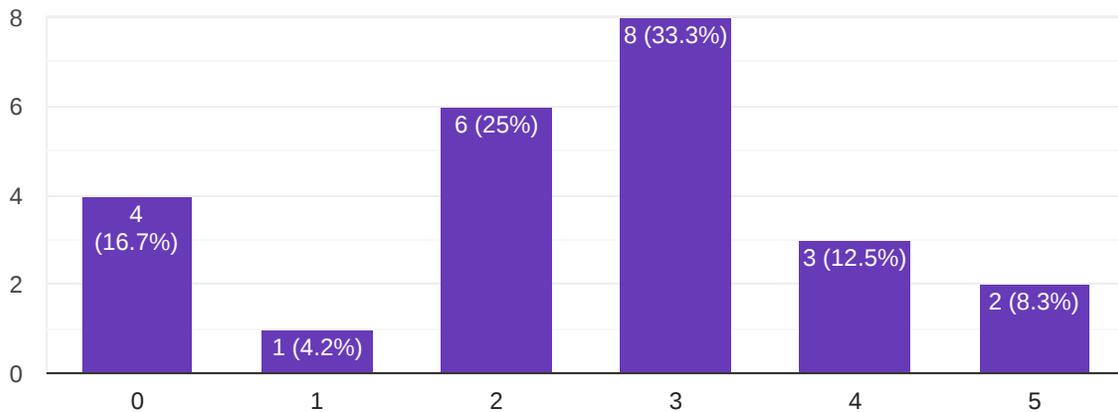
If you have stopped using it, why?

7 responses

It was a short-term need.

Previous Job.

Because of georedundancy it showballed from 50 USD a month to 500 in just over a year

Didn't need it long-term

Miscommunication between Google account sync and recent MacOS capabilities

I used it heavily while at Yale for my heavy compute needs, but there was no clear way to reimburse the cost from e.g. my Hubble grant. I had been paying out of pocket, but the costs started to become annoying (in the hundreds of dollars), so I just stopped
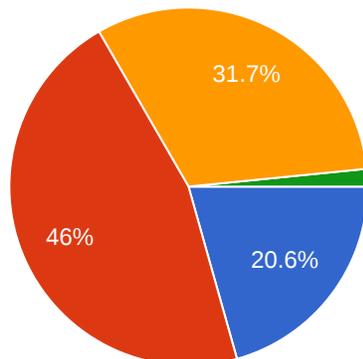
probably need to downscope because it costs too much staff time to manage GCP costs effectively

Everyone

Broadly, how would you characterize your personal understanding of the services provided by cloud platforms like Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure?

Copy

63 responses

- 🔵 I barely know what you're talking about!
- 🔴 I think I understand the basics of what they offer
- 🟠 I have a good understanding of some of their services, but not all
- 🟢 I understand them very well

Pie chart values: 31.7%, 46%, 20.6%

Imagine that the CfA provided some kind of high-quality training in the cloud (e.g. general concepts and/or specific technologies). What kind of time investment in training would make sense for you?
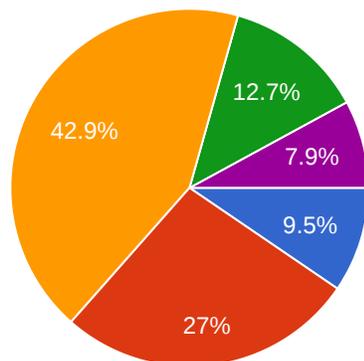
63 responses
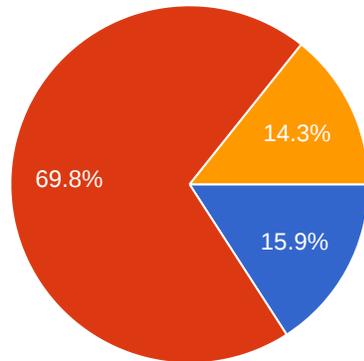
- ● None — my current level of understanding is fine for my role
- ● Hours, but not weeks — it would be useful to learn some more
- ● Weeks or more — it would be useful to learn lots more

14.3%
69.8%
15.9%

Consider the project(s) on which you spend most of your work time. Would the use of cloud services be helpful for these efforts?
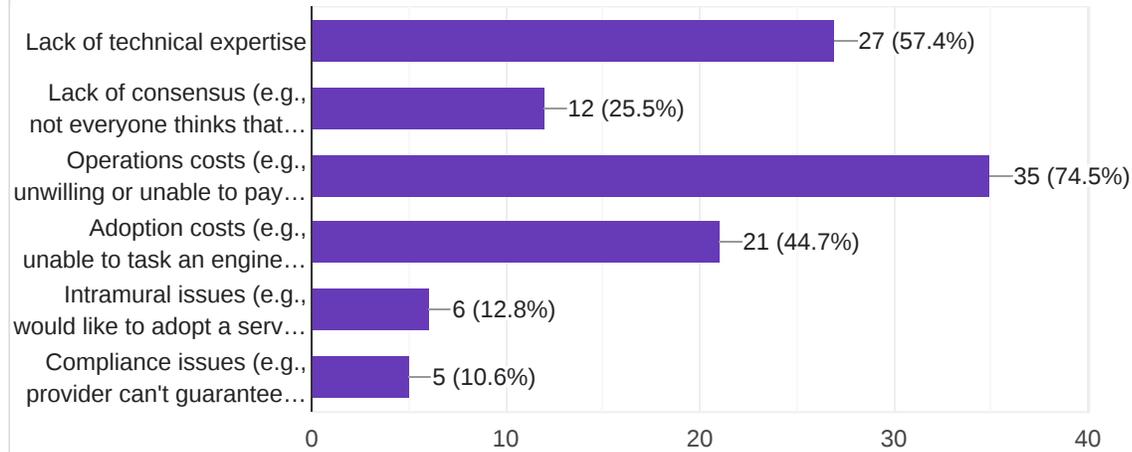
63 responses

- ● No — we have seized opportunities; or we don't nee…
- ● Possibly — there might be some benefits to be gained
- ● Probably — there are likely worthwhile benefits to be gained
- ● Yes — we are missing important opportunities; benef…
- ● Unsure — I am not aware of what opportunites we are mis…

42.9%
12.7%
7.9%
9.5%
27%

In cases where you think that cloud usage would be beneficial to your project(s), what key barriers are at play? If unsure or there are no important opportunities or barriers, leave everything unchecked.

Copy

47 responses

Please rate the importance of adopting the following kinds of cloud services to the long-term success of your project(s). If a service is important but would better be provided on-premises rather than in the cloud, rate it low. If a service is currently free but you would pay for it if you had to, rate it high.
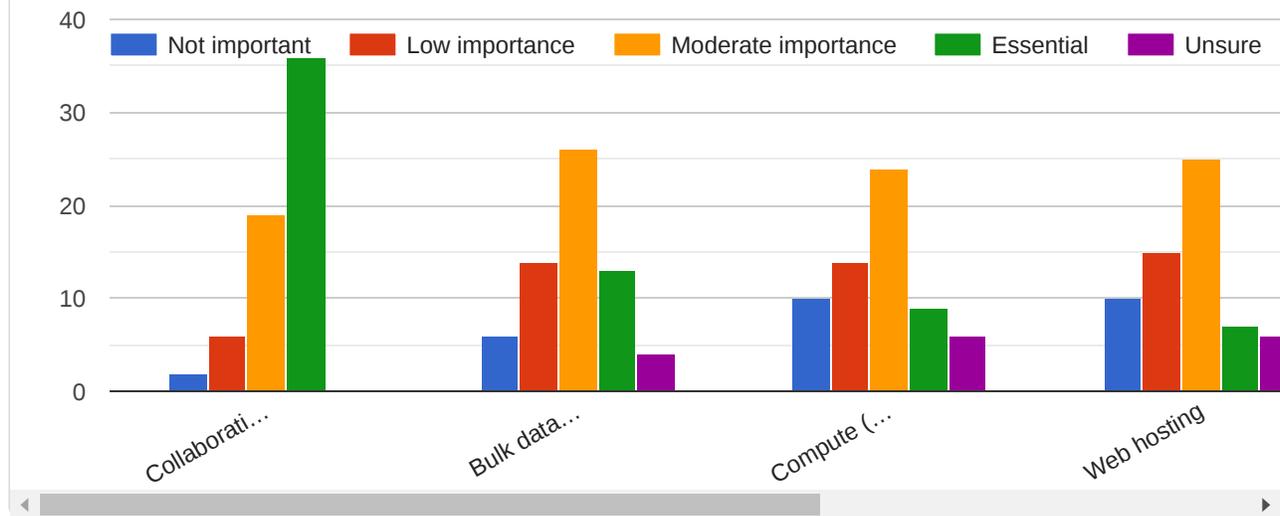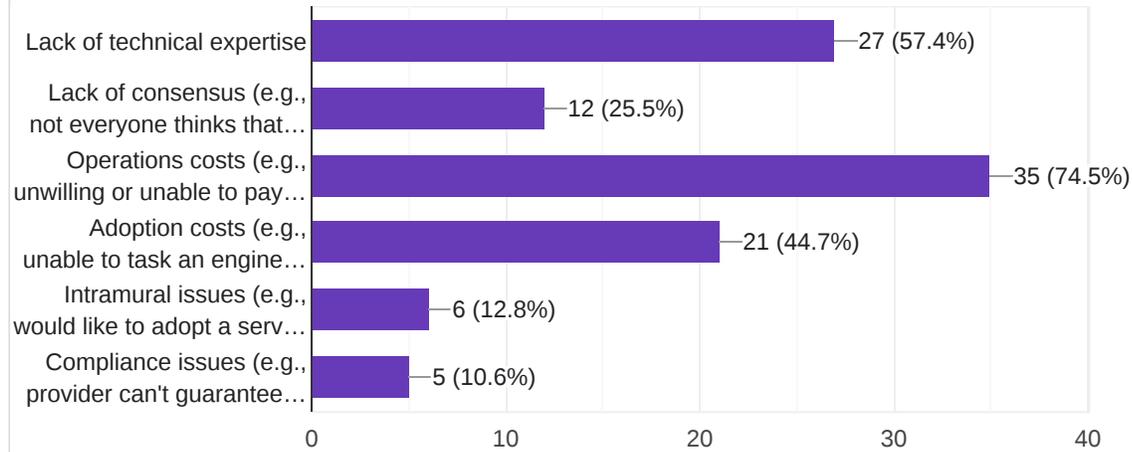
Copy

In cases where you think that cloud usage would be beneficial to your project(s), what key barriers are at play? If unsure or there are no important opportunities or barriers, leave everything unchecked.

Copy

47 responses

Barrier	Count
Lack of technical expertise	27 (57.4%)
Lack of consensus (e.g., not everyone thinks that…)	12 (25.5%)
Operations costs (e.g., unwilling or unable to pay…)	35 (74.5%)
Adoption costs (e.g., unable to task an engine…)	21 (44.7%)
Intramural issues (e.g., would like to adopt a serv…)	6 (12.8%)
Compliance issues (e.g., provider can't guarantee…)	5 (10.6%)

Please rate the importance of adopting the following kinds of cloud services to the long-term success of your project(s). If a service is important but would better be provided on-premises rather than in the cloud, rate it low. If a service is currently free but you would pay for it if you had to, rate it high.

Copy

Legend: …derate importance | Essential | Unsure

Categories: Compute (…, Web hosting, Machine l…, Databases
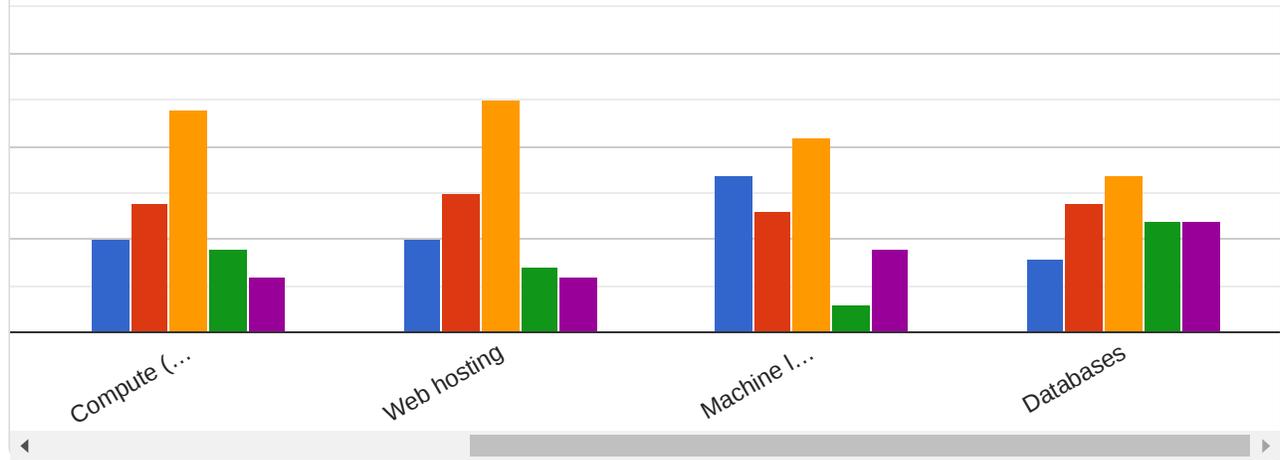
Any other comment about using cloud services?

12 responses

An option I don't see listed here is an institute-run cluster running OpenStack (or similar) that could provide a base level of computing/storage. Even with the difference between computing clusters for SAO vs. HCO, a portion of an existing cluster or new cluster could provide a significant amount of flexibility for the average user, and pre-made instance images with installed software/etc could be maintained with the added flexibility for the user to make custom changes building on that base. This was an extremely popular and successful model at a previous institute I was at, and vastly reduced our use of other pay-by-time cloud platforms.

I'm wary of making any kind of long-term commitments to a cloud services vendor. It feels safer to have an in-house HPC and storage solution, especially given ITAR and external collaborations.

My particular use case for compute services is for acceleration of and increased memory for electromagnetic simulations using commercial engineering software. This can be done with GPUs etc. on cloud services, but needs specific software licenses in place, and has very specific hardware and OS requirements.

We need a secure file sharing option for receiving documents containing PII from prospective fellows and employees.

Cloud is where services are headed. One roadblock is that services offered may need to be FedRAMP approved if ITAR or EAR are involved. In the software group, Alassian products (Jira, Confluence) are going to the cloud soon and we're watching the approval process on the website. (https://www.fedramp.gov)

I have many unknown unknowns about cloud computing, so it's difficult for me to assess at this point how useful it may or may not be. A "what can cloud computing do for you" seminar or FAQ would be helpful, and may lead to me wanting much more. I feel like I can do my job just fine without it, but perhaps I just don't know what I'm missing.

Perhaps an inexpensive option to explore would be utilizing unused local CPU time using a cluster management tool such as Condor (free).

https://htcondor.org/htcondor/overview/

Our mission's data will be soon be provided on the cloud by NASA. I think 95% of our users have no idea how to take advantage of this, budget for computation time, etc., including me (the data provider).

Better integration of the Google drive storage sync and mirroring system with on-campus computing would greatly improve reliability and stability of data processing across users and

platforms.

We should move forward on this, as soon as possible.

I find the useful for exceptional circumstance which come up too rarely to support purchasing permanent resources. As well as communal data.

Commercial cloud is too expensive. There is fear of vendor lock-in, or getting stuck into an uneconomical mode of operation. It is unclear how to relate these efforts to grid computing and "local cloud" and use of "free" computational resources from national and institutional research clusters. A consultant resource to help navigate the ecosystem would be helpful.

Google Forms