RESULTS of January 2019 Survey: Campus Champion On-premise and Off-premise HPC

The purpose of the survey is to gather input from researchers at Campus Champion institutions on their use of on-premise and off-premise resources for high performance computing.

This survey was conducted by Academic and Research Computing Systems at New Jersey Institute of Technology.

Participants
48 invitees from 44 institutions completed the survey. Institutions and departments are listed in the Appendix at the end of this document.

SECTION 1: On-premise high performance computing environment
Overview of question content:
- How does your institution manage on-premise HPC resources?
- Estimate the quantities of clusters available to researchers, number of CPU cores in the clusters, and number of GPUs in the clusters.
- Do your on-premise HPC resources utilize a parallel file system?
- What type(s) of node interconnect are you using?
- How important is providing on-premise HPC resources in attracting new research faculty at your institution?

Brief summary of Section 1 responses:
- Most (77%) institutions use centralized management of on-premise HPC resources
- More than half (54%) of participants report 2 to 5 clusters as resources available at their institutions; over one third (37%) report availability of 5001- 20,000 CPU cores in the clusters; one third (33%) report availability of 11 - 50 GPUs in the clusters
- Most (71%) participants report using a parallel file system in on-premise HPC resources
- Most (81%) participants report using Infiniband for node interconnect
- Most (71%) participants consider the role of on-premise HPC resources in attracting new research faculty as either very important (33%) or essential (38%)

Complete Section 1 questions and responses are below.

Question: *How does your institution manage on-premise HPC resources?*
Response choices (multiple responses allowed):
- Centralized (provided and managed by central IT)
- Departmental (provided and managed by individual departments)
- Condominium (you own resources in a centralized or departmental resource)
- Research computing colocation
- Individually owned
- Other – please specify

Responses:

n = 48
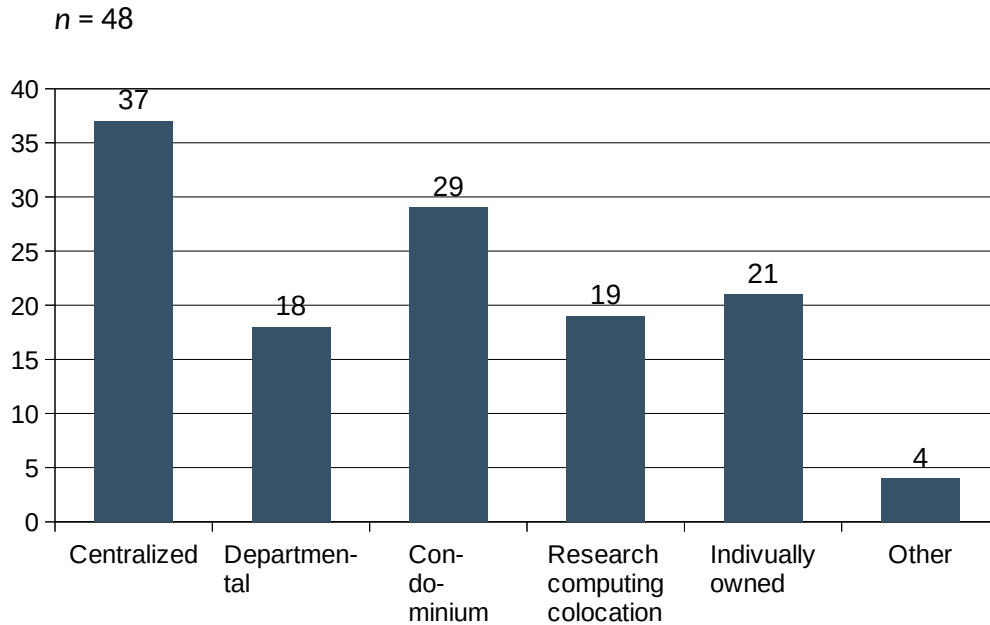
Other specifications (write-in):
    Central, but not operated by campus central IT org
    Application support from outside IT
    Jointly between multiple research units
    All of the above plus AWS and probably others

Question: *For the next three questions, please estimate the quantities of the indicated resources available to researchers at your institution.*
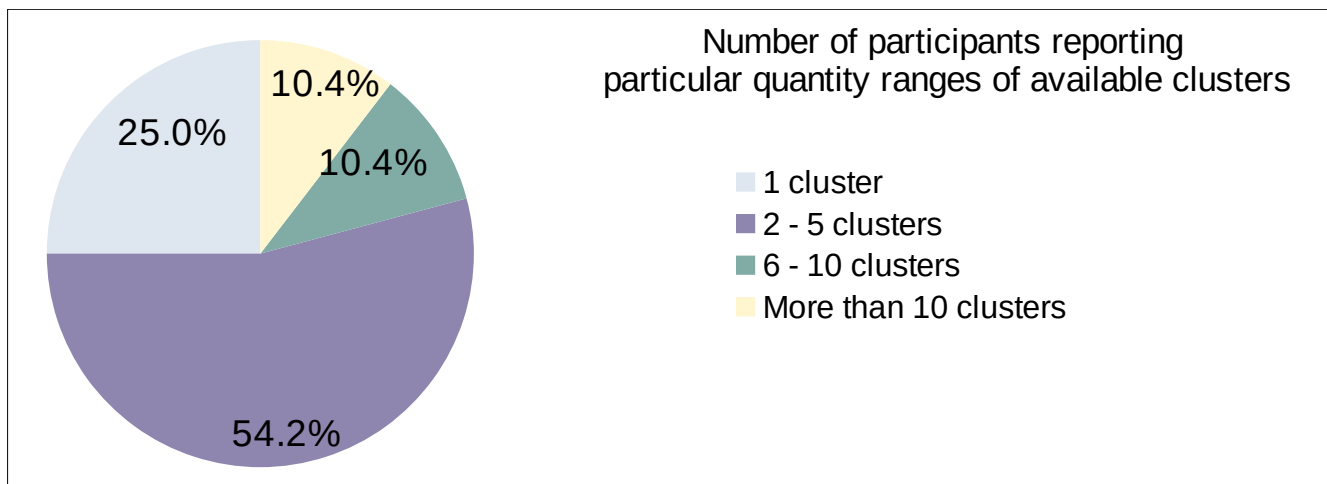*Number of clusters; Total number of CPU cores in the cluster(s) selected; Total number of GPUs in the cluster(s)*

Response choices (single response only):
- Clusters: 1; 2 to 5; 6 to 10; More than 10; Don't know
- CPU cores: Less than 1000; Between 1001 and 5000; Between 5001 and 20,000; More than 20,000; Don't know
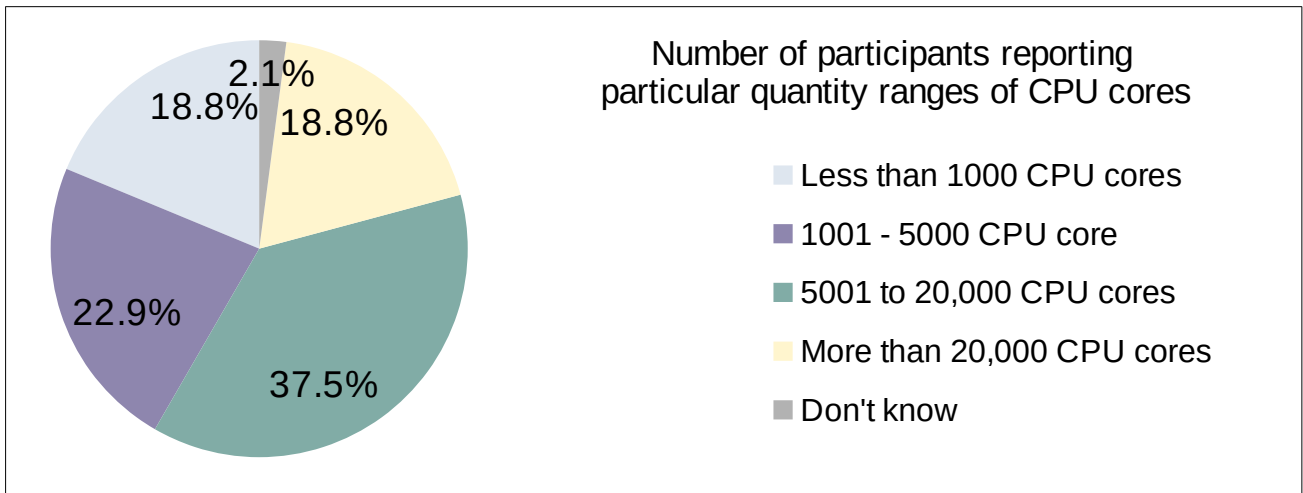- GPUs: None; 1 to 10; 11 to 20; 21 to 50; More than 50; Don't know

Responses:
Note: Finer-grained responses linking the quantity of reported CPU cores and GPUs to the reported quantity of clusters are tabulated in the Appendix.
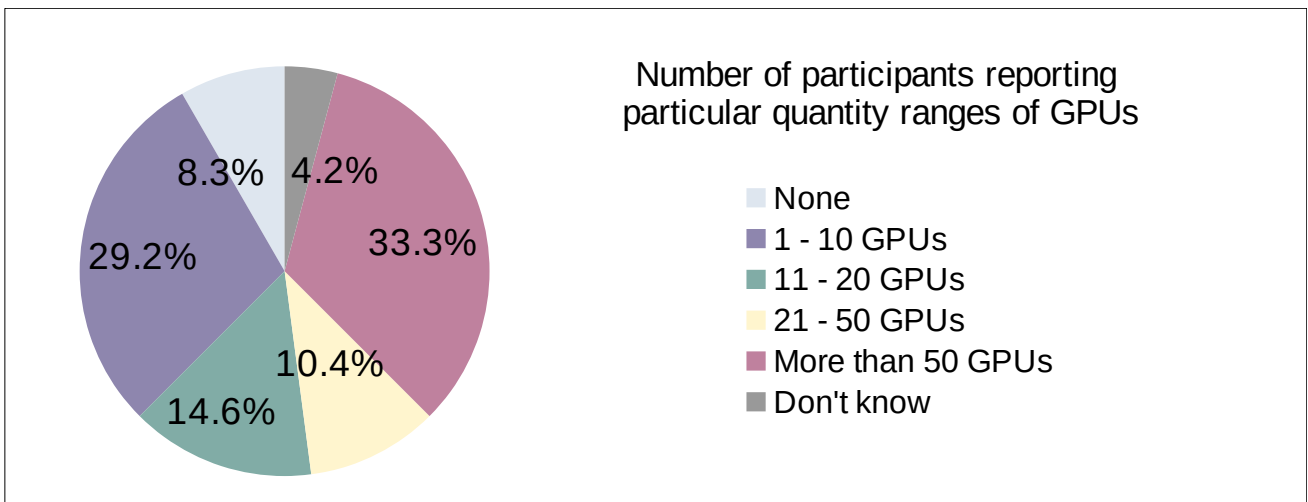
n = 48

n = 48

Number of participants reporting
particular quantity ranges of CPU cores

- Less than 1000 CPU cores
- 1001 - 5000 CPU core
- 5001 to 20,000 CPU cores
- More than 20,000 CPU cores
- Don't know

2.1%
18.8%
18.8%
22.9%
37.5%

n = 48

Number of participants reporting
particular quantity ranges of GPUs

- None
- 1 - 10 GPUs
- 11 - 20 GPUs
- 21 - 50 GPUs
- More than 50 GPUs
- Don't know

8.3%
4.2%
29.2%
33.3%
14.6%
10.4%

Question: *Do any on-premise HPC resources that you use utilize a parallel file system?*
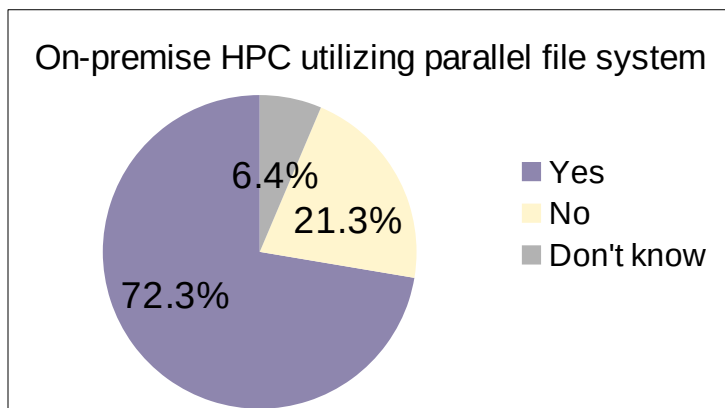**Response choices** (single response only):
- Yes
- No
- Don't know

Responses:
Note: 1 response is omitted due to technical malfunction; hence n = 47 rather than 48

n = 47

On-premise HPC utilizing parallel file system

- Yes
- No
- Don't know

6.4%
21.3%
72.3%

Question: *What type(s) of node interconnect are you using?*
Response choices (multiple responses allowed):
- 1GigE (1 Gb/sec Ethernet)
- 10GigE (1 Gb/sec Ethernet)
- Infiniband
- OmniPath
- Don't know
- Other – please specify

n = 48
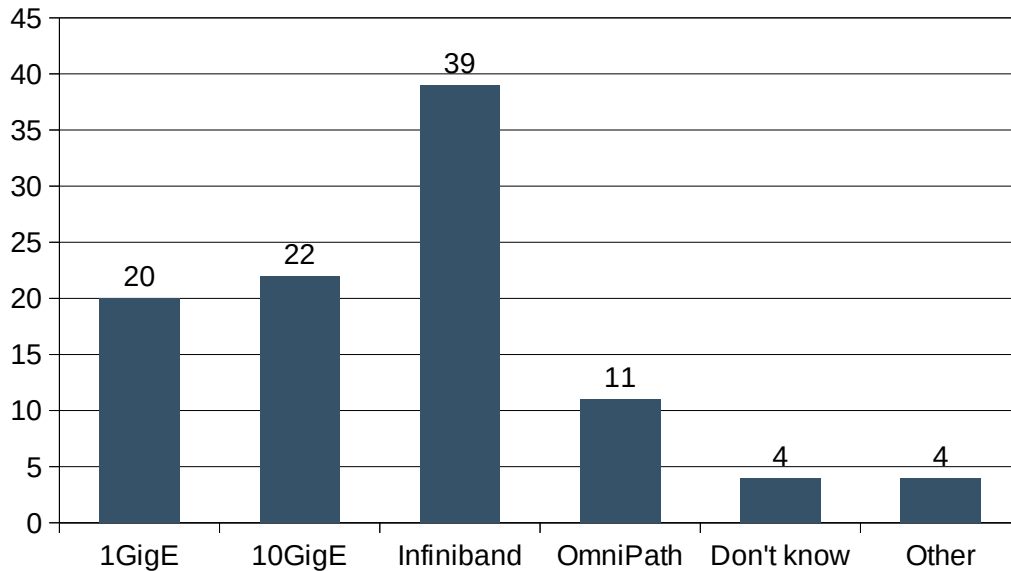
Other responses (write-in):
    25GigE, 100GigE
    40GigE
    100/25/10 GbE
    40 / 100 GigE

Question: *How important is providing on-premise HPC resources in attracting new research faculty at your institution?*
Response choices (single response only):
- Not important
- Somewhat important
- Very important
- Essential
- Don't know

Importance of HPC resources for research faculty

- Not important
- Somewhat important
- Very important
- Essential
- Don't know

4.2% 6.3%
18.8%
37.5%
33.3%
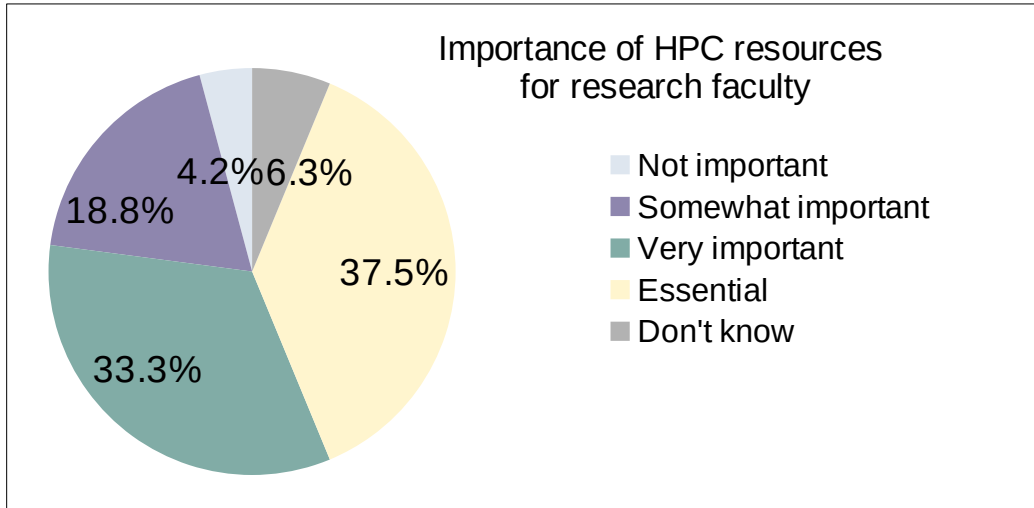
SECTION 2: Use of commercial cloud providers and non-commercial resources

Overview of question content:
- Are you using commercial cloud providers for HPC?
- Which commercial cloud providers did you consider, and which are you using?
- Did you perform benchmarking, and do you have shareable documentation?
- Commercial vs non-commercial: What is your proportional use of commercial cloud providers and non-commercial resources?
- Non-commercial, on-premise vs off-premise: What is your proportional use of on-premise and off-premise non-commercial resources?
- If you do not use commercial cloud providers, why not?
- Will you re-evaluate your use of commercial cloud providers and if so, when?

Brief summary of Section 2 responses:
- Nearly one third (31%) of participants use commercial cloud providers; most (80%) of these also use non-commercial resources
- Nearly all commercial cloud users considered using AWS and GCP; most (86%) use AWS, somewhat more than half (57%) use GCP
- One third (33%) of commercial cloud users performed benchmarking; none can share documentation
- Commercial vs non-commercial: Most (80%) participants who use both commercial cloud and non-commercial resources predominantly use non-commercial resources
- Non-commercial; on-premise vs off-premise: Most (81%) non-commercial resource use is on-premise
- Most (81%) participants who do not use commercial cloud providers cite expense as the reason
- About two thirds (63%) of participants plan on re-evaluating their use/non-use of commercial cloud providers; of these, nearly three quarters (73%) will do so within 18 months

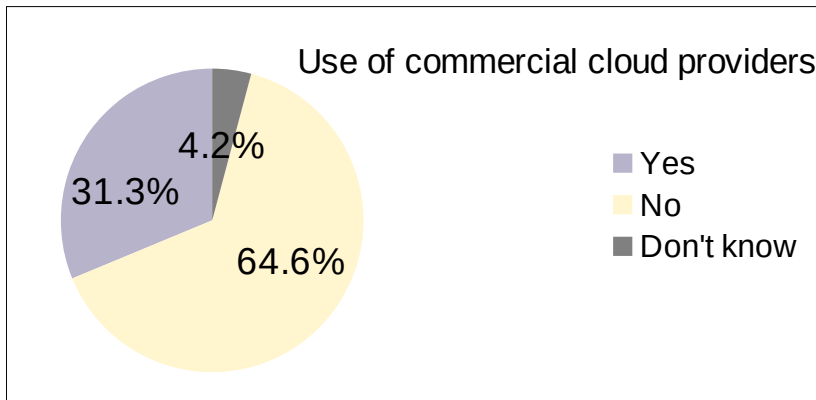Complete Section 2 questions and responses are below.

Question: *Are you using commercial cloud providers for HPC?*
Response choices (single response only):

- Yes
- No
- Don't know

Responses:

n = 48

Use of commercial cloud providers

4.2%
31.3%
64.6%

- Yes
- No
- Don't know

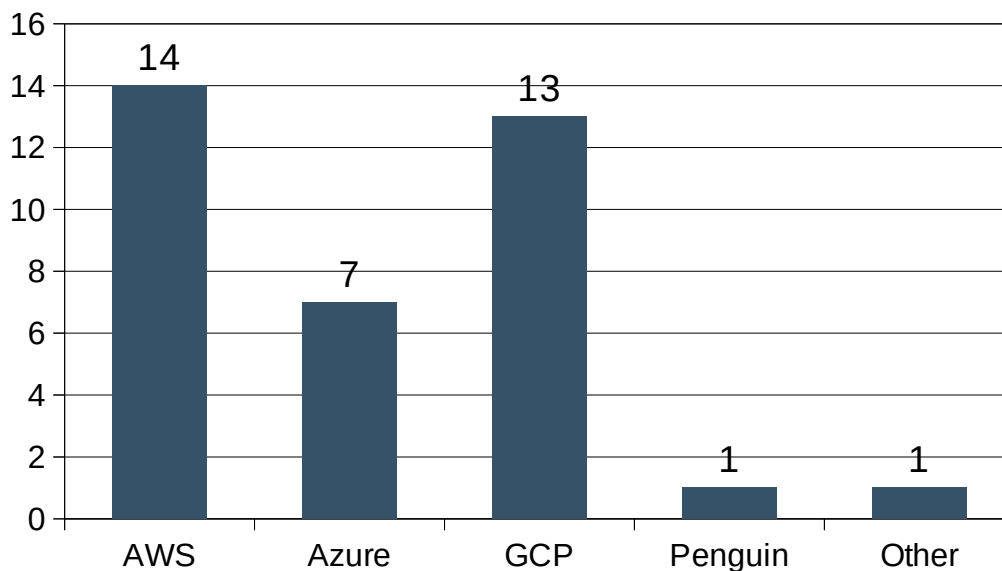Of the 15 participants who reported using commercial cloud providers:
Question: *In choosing a commercial cloud provider(s), which did you initially consider?*
Response choices (multiple responses allowed):
- Amazon Web Services (AWS)
- Azure
- Google Cloud Platform (GCP)
- Penguin Computing
- Other – please specify

Responses:

n = 15

Other responses (write-in):
     HIPPA needs led Amazon GovCloud

Question: *Did you perform benchmarking on any commercial cloud providers?*
Follow-up question: *Did you perform benchmarking on any commercial cloud providers?*
Follow-up question: *Do you have any documentation on your benchmarking?*
Follow-up question: *Can you share the documentation with us?*
Response choices (single response only):
- Yes
- No
- Don't know

Responses:

Of the 15 participants who reported using commercial cloud providers, 5 performed benchmarking, 7 did not, and 3 don't know.
Of the 5 participants who performed benchmarking, 3 have documentation, 1 does not, and 1 does not know.
Of the 3 participants who have benchmarking documentation, 1 cannot share the documentation, and 2 don't know.

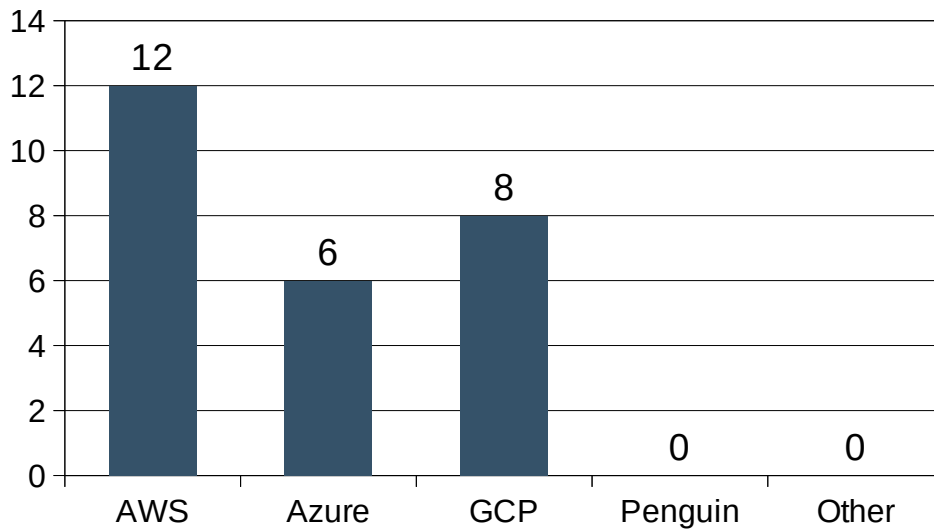Of the 15 participants who reported using commercial cloud providers:
Question: *Which commercial cloud provider(s) are did you using?*
Response choices (multiple responses allowed):
- Amazon Web Services (AWS)
- Azure
- Google Cloud Platform (GCP)
- Penguin Computing
- Other – please specify

Responses:

n = 15

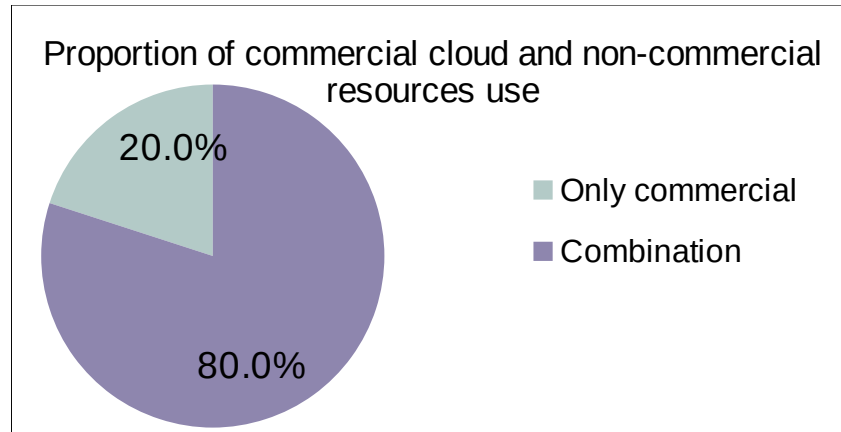For the 15 participants who reported using commercial cloud providers:

Question: *What is your proportionate use of commercial cloud providers and non-commercial resources? Do you 1) exclusively use commercial cloud providers, or 2) use some combination of commercial and non-commercial resources?*

Response choices (single responses only):
- I only use commercial cloud providers
- I use a combination of commercial and non-commercial resources

Responses:

n = 15

Proportion of commercial cloud and non-commercial resources use

- Only commercial
- Combination

20.0%

80.0%

For the 12 participants who reported using a combination of commercial and non-commercial resources:

Question: *In what approximate proportions do you use commercial providers and non-commercial resources?*

Response choices (single responses only):
- Mostly commercial: more than 60% commercial, less than 40% non-commercial
- About equal: 40-60% commercial, 40-60% non-commercial
- Mostly non-commercial: more than 60% non-commercial, less than 40% commercial
- Don't know

Responses:

n = 12

Proportion of commercial and non-commercial providers/ resources
Note: No participant chose "About equal"

- Mostly commercial
- Mostly non-com-mercial

16.7%

83.3%
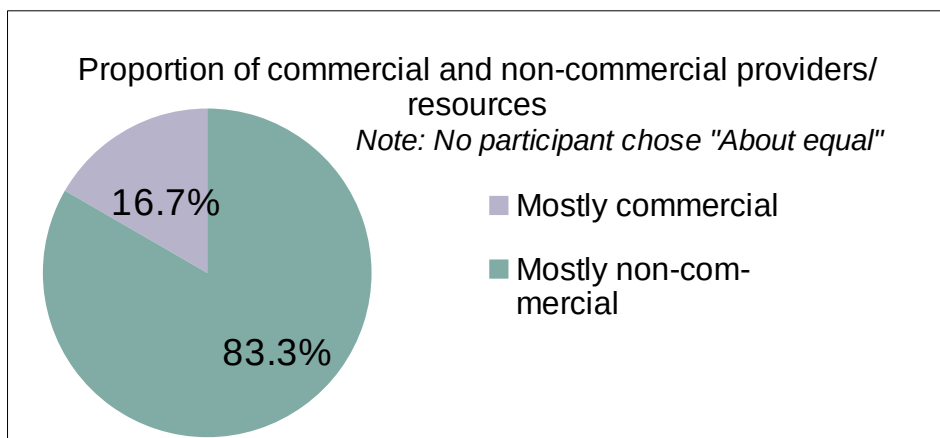
For 27 participants who reported using non-commercial

resources, either exclusively or in combination with commercial cloud providers (Note: 47 participants used non-commercial resources; data on 20 were not obtained due to data collection error):

Question: *In what approximate proportions do you use on-premise resources and non-commercial resources such as XSEDE, OSG, or resources at other institutions?*
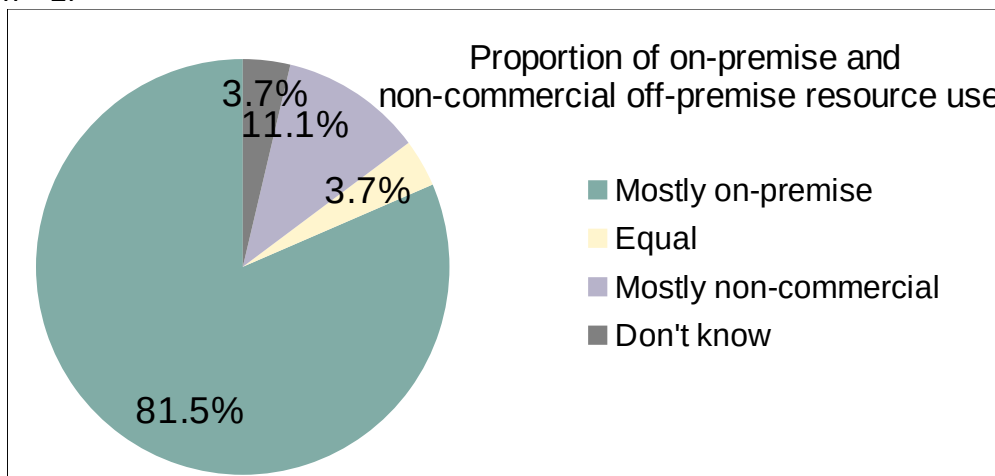**Response choices** (single responses only):
- Mostly on-premise: more than 60% on-premise, less than 40% non-commercial HPC
- About equal: 40-60% on-premise, 40-60% non-commercial
- Mostly non-commercial: more than 60% non-commercial HPC, less than 40% on-premise
- Don't know

Responses:
Note: Participants overwhelmingly use on-premise resources rather than other non-commercial resources (e.g. XSEDE, OSG, and resources at other institutions) regardless of their use/non-use of commercial cloud providers; responses were essentially identical across participants who use a combination of commercial and non-commercial providers/resources and participants who exclusively use non-commercial resources.
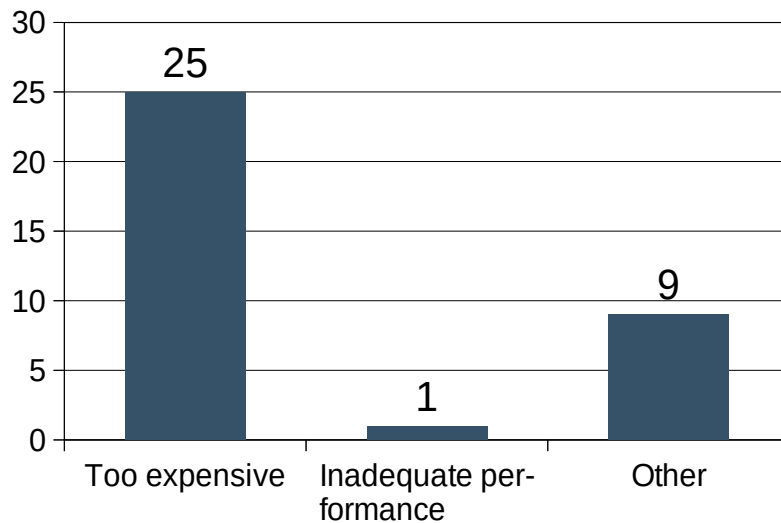
n = 27

For the 31 participants who reported not using commercial cloud providers:
Question: *Please provide the reason(s) you are not using commercial cloud providers*
Response choices (multiple responses allowed):
- Too expensive
- Inadequate performance – please describe
- Other – please specify

n = 31

Inadequate performance descriptions (write-in):
- Not sure about the performance and am always fearful of paying for it if we or the students are not using it properly or forget to shutdown an active cloud server.

Other specifications (write-in):
- Difficult procurement path, support for rebilling.
- We do help with XSEDE access if our resources cannot meet researcher needs
- Core users are happy with on-prem HPC; people in CS and data science use cloud providers on their own
- Skill gap … although this is slowly going away.
- Capacity met locally or via XSEDE
- Demand being met on-prem
- I don't know
- We have a centralized cloud system for such HPC resource utilization.
- More difficult to set up than using existing campus clusters.

For all 48 participants:
Question: *Do you plan to re-evaluate your use/non-use of commercial cloud providers?*
Follow-up question: *If so, when?*
Response choices (single response only):
- Yes
  - Within 18 months
  - After 18 months
  - Don't know
- No
- Don't know

Responses:

30 participants plan to re-evaluate their use of commercial cloud providers; 7 do not; 11 don't know. Of the 30 participants who plan to re-evaluate their use of commercial cloud providers, 22 plan to re-evaluate within 18 months, 4 plan to re-evaluate after 18 months, and 4 do not know.

For all 48 participants:
Question: *Please provide any comments that you think would be useful in the context of this survey. (optional write-in)*

Responses:

- Please post a summary of this data to the CC list or as a paper to PEARC with a post to CC list that it's out there.

- We run primarily HPC on premise. At 90+% utilization 24x7, cloud is prohibitively expensive. Enterprise had a contract in place with Azure and Amazon, and we send our HIPPA users to Amazon Gov Cloud as a result.

- Our users can submit on local systems and have jobs automatically use their cloud accounts in addition to running on free resources. Very few people ever buy hardware in our centralized cluster. Using HTCondor allows our various campus units to share computing capacity and users to backfill other clusters. We have few "HPC" clusters (MPI-optimized), as most user work can fit on single servers at a time and benefits from HTC scalability.

- I don't think an average HPC user has the knowledge to answer a lot of questions in this survey - it's better directed to the staff that run the resources :)

- Our Google Cloud work is concentrated in one very large lab for a single genomics use case in the 40-50k core range
  We provide a large campus resource that supports both condo and rented cores
  Several units still run their own clusters and several large faculty led labs run their own clusters, but the campus cluster probably has 70% if I had to guess of the HPC/HTC cores on campus.

- Our HPC resources are centralized but under the Vice President for Research rather than Central IT (but I checked the centralized button.)

- Two points:
  1] As I said in the 'other' field, we do assist researchers who need to access XSEDE resources, but we try to keep them on our on-campus resources if we can meet their needs.
  2] We constantly evaluate the best combination of HPC resources, from user and institution standpoints.

- We have debated the on-prem vs. cloud HPC question a lot and we are mostly sticking to on-prem HPC due to cost and ease of use.
  1) cost - research funding agencies won't cover cloud HPC costs. A few of our researchers whose cloud HPC costs are covered by collaborators or other parties do happily use commercial cloud providers. They tend to be in CS and data analytics than traditional HPC users even though those lines are blurred in the era of "convergence of ai and hpc". I've explored whether our users can do HPC in the commercial clouds for free and the answer is generally no unless they do cutting edge research vendors would be willing to throw some research credit at. (http://temelsob.people.cofc.edu/cloud-hpc.php)
  2) ease of use - most of our users have used traditional HPC systems and they don't feel the need to migrate to the cloud yet. Most of their needs are met by our on-prem HPC and XSEDE resources when necessary.

- Our HPC cluster is under-utilized due to lack of support and promotion by IT, which has no personnel to support research computing and comprised by the enterprise IT culture. Frustrated researchers gain little by going to the cloud if they still get no research computing support.

- Currently evaluating Azure for CUI workloads, and Google for cloudbursting.

- We evaluate commercial cloud resources periodically as pricing models and needs change.

- Thanks for conducting this!

- We have a diverse set of centralized HPC resources of which I have contributed in their acquisition. The support at our institution, guided by Dana Brunson, has been exceptional. While the current state is good, there is uncertainty in how it will develop in 5 years time.

- I am not a consumer of HPC resources except for various HPC related class time and exercises. I assist others with issues on local HPC systems.

APPENDIX

The table below provides finer-grained results of two related questions in Section 1: *Please estimate the quantities of the indicated resources available to researchers at your institution*
Number of clusters; Total number of CPU cores in the cluster(s) selected

This table shows the quantities of reported CPU cores per reported quantity of available clusters.

Quantity of available clusters	Number of CPU cores in clusters	Number of participants reporting specified quantity of CPU cores
1 cluster 12 participants	Less than 1000	6
	1001 - 5000	4
	5001 – 20,000	2
2 to 5 clusters 26 participants	Less than 1000	3
	1001 - 5000	7
	5001 – 20,000	13
	More than 20,000	2
	Don't know	1
6 to 10 clusters 5 participants	5001 – 20,000	2
	More than 20,000	3
More than 10 clusters 5 participants	5001 – 20,000	1
	More than 20,000	4

The table below provides finer-grained results of two related questions in Section 1: *Please estimate the quantities of the indicated resources available to researchers at your institution. Number of clusters; Total number of GPUs in the cluster(s)*

This table shows the quantities of reported GPUs per reported quantity of available clusters.

Quantity of available clusters	Number of GPUs in clusters	Number of participants reporting specified quantity of GPUs
1 cluster 12 participants	None	3
	1 - 10	5
	11 - 20	1
	21 - 50	1
	More than 50	2
2 to 5 clusters 26 participants	None	1
	1 - 10	8
	11 - 20	5
	21 - 50	3
	More than 50	7
	Don't know	2
6 to 10 clusters 5 participants	1 - 10	1
	11 - 20	1
	21 - 50	1
	More than 50	2
More than 10 clusters 5 participants	More than 50	5

The table below lists each participant's reported institution and department.

INSTITUTION	DEPARTMENT	INSTITUTION	DEPARTMENT
Arizona State University	Research Computing	Pomona College	ITS
Boise State University	Research Computing	Portland State University	Research Computing
Case Western Reserve U.	UTech	Purdue University	Research Computing
Clarkson University	Office of Information Technology	Rowan University	ECE
Cleveland Clinic Foundation	Lerner Research	Rutgers university	OARC

	Institute - Genomic Medicine Institute		
College of Charleston	IT	Stanford University	Research Computing Center
College of William and Mary	IT/HPC	University of Arizona	Research Technologies/High Performance Computing
			UITS
Colorado School of Mines	Research Computing	University of Alaska Fairbanks	Geophysical Institute Research Computing Systems
Dominican University of California	Health and Natural Sciences	University of California, Berkeley	Research IT
Einstein College of Medicine	Medicine/endo	University of Delaware	Research Computing
George Mason University	Office of Research Computing	University of Illinois at Chicago	Mechanical and Industrial Engineering
Georgetown University	University Information Services	University of Illinois, Urbana-Champaign	Beckman Institute
Georgia Tech	PACE	University of Maine System	IT
Johns Hopkins University	Center for Imaging Science	University of Massachusetts Boston	Research Computing
Kansas State University	Computer Science	University of Michigan	ARC-TS
Lehigh University	Library & Technology Services	University of Missouri	Research Computing Support Services
Marshall University	Center for Environmental, Geotechnical and Applied Sciences	University of Nevada, Reno	Office of Information Technology
Mississippi State University	High Performance Computing Collaboratory	University of Texas, Dallas	IT
New York University	HPC, IT	University of Texas, El Paso	Mechanical Engineering
Notre Dame	Center for Research Computing	University of Wisconsin-Madison	Center for High Throughput Computing
Oklahoma State University	High Performance Computing Center	West Chester University	Computer Science
	Chemistry		
Penn State	Institute for CyberScience	West Virginia University	Research Computing