2023-05-08

Alliance Researcher Council Cloud Survey Report

Table of Contents

Contributors

The cloud survey was designed and developed by the Researcher Council Cloud Survey Steering Group:

Marie-Jean Meurs (Université du Québec à Montréal) – Chair

Jeff Albert (University of Victoria)

Jeff Moon (Compute Ontario)

Liseanne Cadieux (the Alliance)

Matthew Smith (University of British Columbia)

Randall Sobie (University of Victoria)

Authors of the report on the Researcher Council Cloud Survey Steering Group:

Fares Dhane (the Alliance)

Felipe Pérez-Jvostov (the Alliance)

Support:

Angela Law (the Alliance)

Caitilín O'Hare (the Alliance)

Chantale Blais (the Alliance)

Key findings

- Alliance Community Cloud is not widely used but is very important for those researchers that do–particularly in the Sciences and Engineering.
- Canadian researchers increasingly adopt cloud resources.
- Cost and ease-of-use are the most important decision-making factors for researchers' choice of cloud service.
- The Alliance Community Cloud is missing several service offerings, including containerization.
- Many researchers use cloud to store (back-up) and share data with collaborators.
- Researchers share privacy and security concerns when using commercial cloud.
- Researchers use commercial and Alliance Community Cloud to access additional storage and compute.

Background and Context

Recent requirements for greater computing power have driven many academic researchers to adopt cloud technologies, due to their scalability and flexibility. Cloud technologies have several added benefits such as optimized resource utilization based on performance, the ability to share information resources (e.g., database, software, network) with other cloud users and the ability to use these resources on demand and with minimal management effort. Despite the growing adoption and use of cloud technologies, to date, it is still unknown how Canadian researchers use them, or for what purpose. In this regard, the Digital Research Alliance of Canada (the Alliance) conducted a survey to gain insights into how researchers are using cloud technologies (e.g., storage and compute), be they commercial or community based. The survey was designed to identify patterns, preferences and experiences of researchers from a range of disciplines while identifying gaps in service provision (i.e., equity and accessibility) to improve the functionality of the Alliance Community Cloud (i.e., ease of use and navigability).

This report presents the findings of the survey, which was conducted between January 16 and February 13, 2023, and collected responses from over 500 researchers across Canada and academic disciplines. The report aims to provide a first insight for future efforts related to cloud technologies in Canadian research.

Respondents

A total of 511 researchers responded to the survey. Quebec (36%) and Ontario (30%) were highly represented, compared to British Columbia (13.7%), Alberta (8.8%) and Nova Scotia (5.1%). The remaining provinces accounted for only ~6% of respondents. The majority of respondents identified themselves as professors (44.46%), with representation of research administrators (10.28%), research support staff (9%) and students (graduate and undergraduate; 8.5%). A small proportion identified themselves as librarians, government officials, postdoctoral researchers and research associates – combined these accounted for ~7.60% of all respondents. Almost a quarter of all respondents had two or more roles within their institution (e.g., professor and researcher).

Organized by the Canadian Research and Development Classification (CRDC) 2020, researchers from the Natural Sciences and Medical, Health and Life Sciences represented half of all respondents with *c.* 25% each, followed by Engineering and Technology with 23%. Researchers in the Social Sciences (13.4%) and Humanities and Arts (11.6%) accounted for a quarter of all respondents. For simplicity, the six Divisions were distilled to three: *Sciences and Engineering* (SE) which includes Natural Sciences, Engineering and Technology, and Agriculture and Veterinary Sciences; *Social Sciences and Humanities* (SSH) which includes Humanities and the Arts, and Social Sciences; and *Health Research* (HR) which includes Medical, Health and Life Sciences. A distribution of respondents by these three groups is presented in Figure 1.
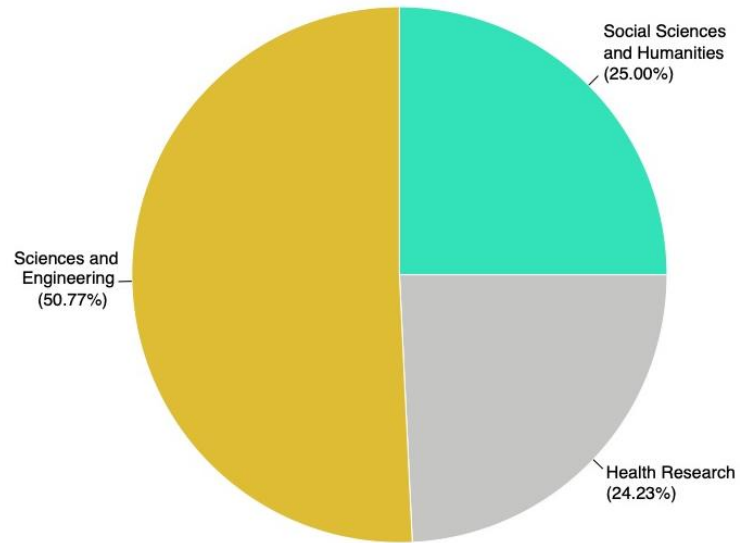
Figure 1. Distribution of respondents within the three broad research domain categories. **Note:** Some multi-disciplinary researchers are represented in more than one domain.
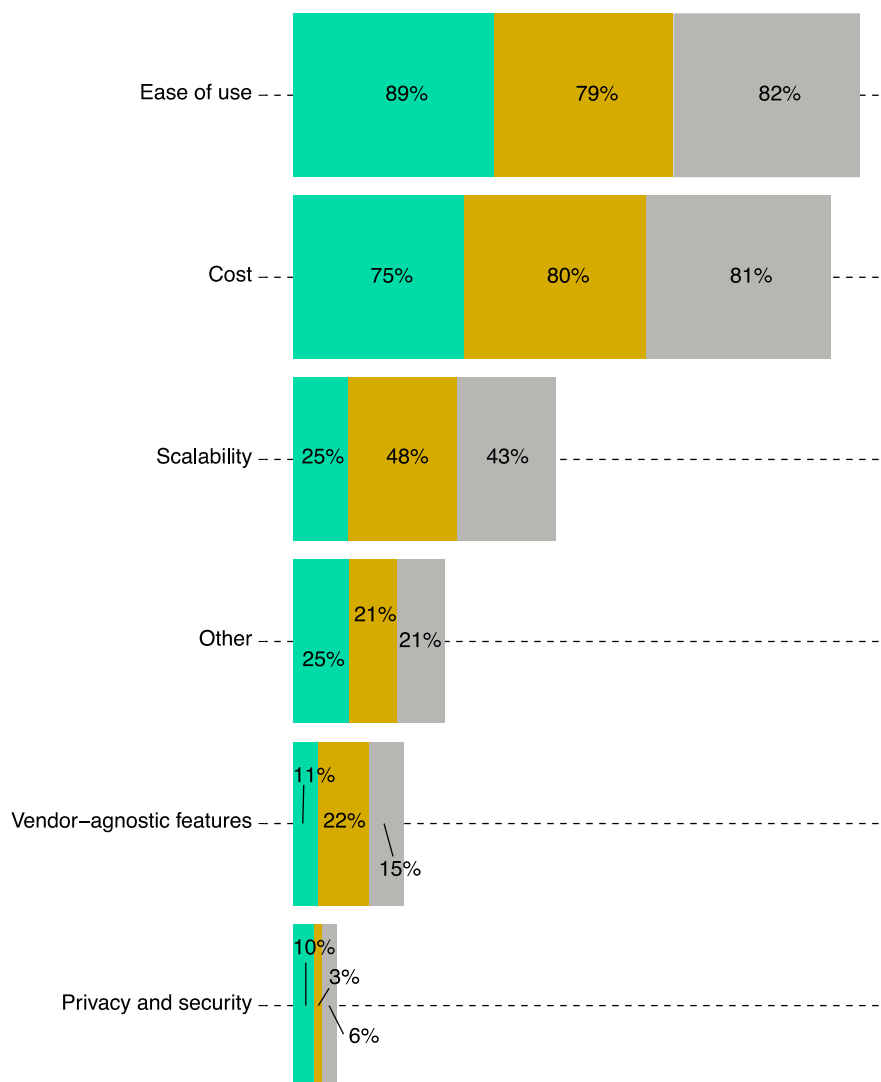
General Use of Cloud

Cloud technologies are being widely adopted by Canadian researchers. The majority of respondents are active users of cloud technologies (83.2%), while only 11.8% do not use such resources or were not sure (4.9%). Those researchers that did use cloud technologies mainly use them for storage solutions (SSH – 89%; SE – 78; HR – 84%), while researchers in Sciences and Engineering also use cloud for additional compute resources (61%, Figure 2). Indeed, many respondents identified cloud as an excellent means to store and back up data in collaborative projects, with specific mentions of applications like Dropbox, Google Drive and OneDrive. Most researchers also identified accessing cloud directly through software applications (SSH – 86%; SE – 74; HR – 83%) compared to web browsers (SSH – 78%; SE – 67; HR – 69%) or secured shell connections (SSH –43%; SE – 74; HR – 54%) – although the latter was extensively used in the Sciences and Engineering, with 72% respondents connecting to the cloud via secured shell. While these results may reflect some differences in technological proficiency between disciplines or different computational workflows, they may also reflect a wider adoption of user-friendly applications, as ease-of-use was just as important a factor as cost in choosing cloud services in the Sciences and Engineering (both 79%), and Health Research (82% vs. 81% respectively), and even more important in Social Sciences and Humanities (90% vs. 76%; Figure 3).
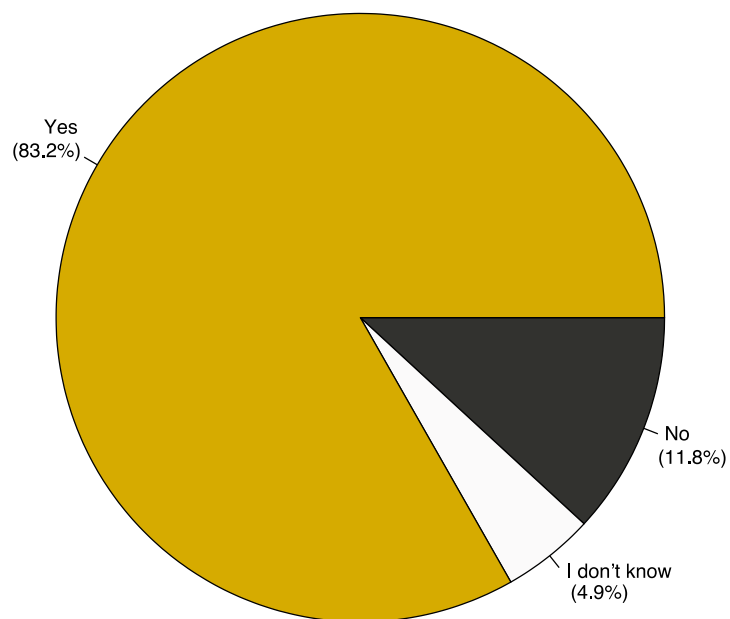
Figure 2. **Question**: Do you currently, or have you in the past, use(d) cloud resources (see options in Question 5) to support your research? Yes, n = 288; No, n = 41; Not sure, n = 17; Total = 346.

Ease of use
- SSH: 89%
- SE: 79%
- HR: 82%

Cost
- SSH: 75%
- SE: 80%
- HR: 81%

Scalability
- SSH: 25%
- SE: 48%
- HR: 43%

Other
- SSH: 25%
- SE: 21%
- HR: 21%

Vendor–agnostic features
- SSH: 11%
- SE: 22%
- HR: 15%
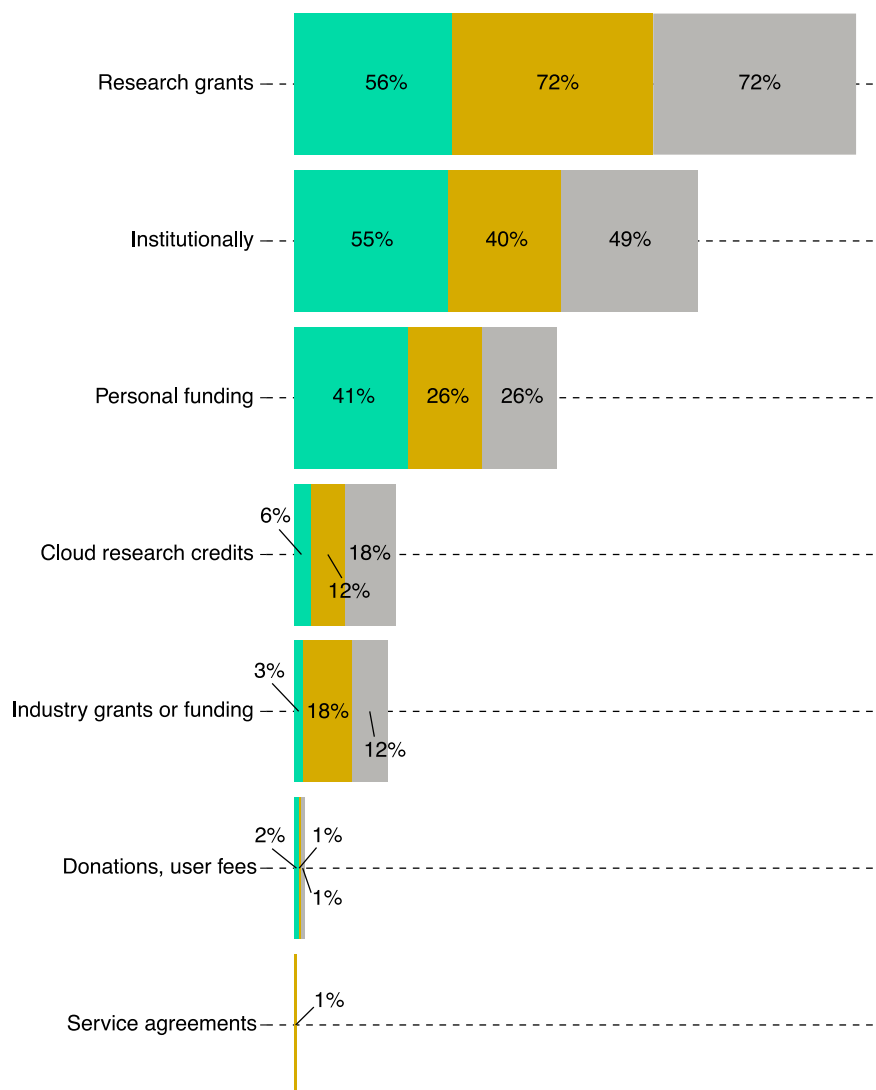
Privacy and security
- SSH: 10%
- SE: 3%
- HR: 6%

■ Social Sciences and Humanities (SSH); ■ Health Research (HR); ■ Sciences and Engineering (SE).

Figure 3. Question: What factors do you consider in choosing which cloud platform to use (Alliance or commercial)? Social Sciences and Humanities, n = 114; Health Research, n = 111; Science and Engineering, n = 191; Unique respondents = 357.

Commercial Cloud

Close to 60% of respondents acknowledged using commercial cloud (e.g., Microsoft Azure, Google Cloud, Amazon Web Services), covering these expenses through research grants (SSH – 56%; SE – 71%; HR – 72%), while also relying on institutional funds (SSH – 54%; SE – 42%; HR – 49%). Very few respondents use cloud credits or industry funding to cover such expenses (Figure 4). An important observation, however, is that many researchers use their own personal funding to access cloud services, and this is particularly evident in the Social Sciences and Humanities where 40% of respondents identified using their personal funding to cover the costs associated with using commercial cloud – an issue that certainly requires further exploration.
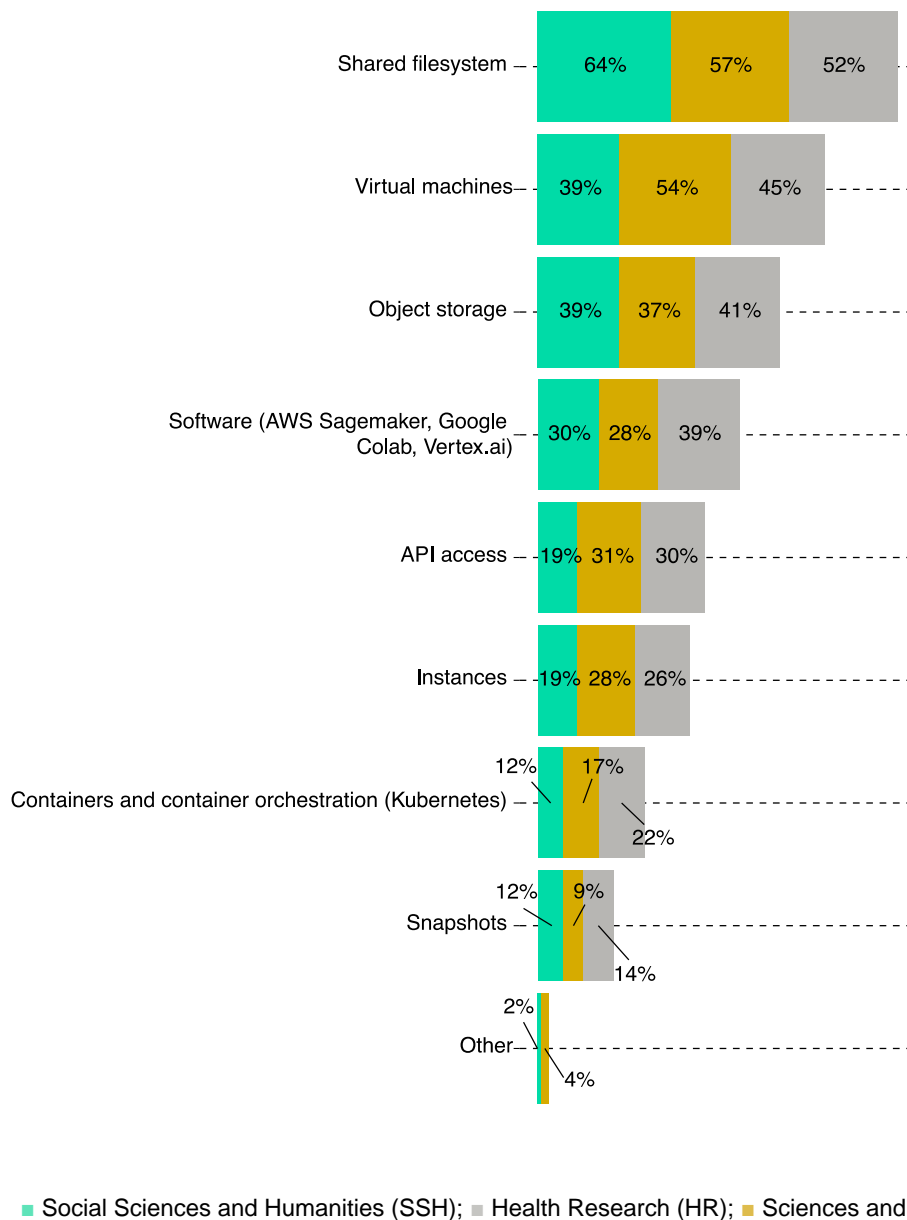
Research grants	56%	72%	72%
Institutionally	55%	40%	49%
Personal funding	41%	26%	26%
Cloud research credits	6%	18%	12%
Industry grants or funding	3%	18%	12%
Donations, user fees	2%	1%	1%
Service agreements		1%	

■ Social Sciences and Humanities (SSH); ■ Health Research (HR); ■ Sciences and Engineering (SE).

Figure 4. Question: How is your commercial cloud budget funded? Social Sciences and Humanities, n = 64; Health Research, n = 72; Science and Engineering, n = 102; Unique respondents = 197.

The current survey does not support quantifying how much individual researchers spend on commercial cloud services, nor the impact that these expenditures may have on grant or institutional budgets. Nonetheless, with the growing adoption of cloud technologies, as well as the sources of funding used to cover them, commercial cloud may represent growing costs to be absorbed by publicly funded research grants and higher education institutions. While some researchers access commercial cloud services at no cost (25.6%) – likely through limited free instances or small storage solutions – 30.77% of the respondents spend less than $1000/year on commercial cloud services, with only a small percentage (1.8%) reporting spending over 100 times this amount.

Despite the potential costs, researchers turn to commercial cloud due to its ease-of-use, and to access additional and flexible storage resources (Figure 5). Researchers in the Sciences and Engineering, in particular, also rely on commercial cloud for additional computational resources, and in some instances to access specialized hardware such as Tensor Processing Units (TPUs) and Graphic Processing Units (GPUs) which are often optimized for deep learning and machine learning applications. While the reasons for researchers to use commercial cloud were evident, what aspects of it are being used, and how, are less so. For example, shared filesystems and virtual machines are both important elements across disciplines (SSH – 64% and 39%; SE – 57% and 54%; HR – 52% and 45% respectively), but so is object storage (~40% across disciplines) and to a lesser extent specific software platforms such as JupyterHub, NextCloud and the Canadian Advanced Network for Astronomical Research Open Stack (SSH – 30%; SE – 28%; HR – 39%).

Shared filesystem — 64% | 57% | 52%

Virtual machines — 39% | 54% | 45%

Object storage — 39% | 37% | 41%

Software (AWS Sagemaker, Google Colab, Vertex.ai) — 30% | 28% | 39%

API access — 19% | 31% | 30%

Instances — 19% | 28% | 26%

Containers and container orchestration (Kubernetes) — 12% | 17% | 22%

Snapshots — 12% | 9% | 14%

Other — 2% | 4%

■ Social Sciences and Humanities (SSH); ■ Health Research (HR); ■ Sciences and Engineering (SE).

Figure 5. **Question**: What components of the commercial cloud do you use? Social Sciences and Humanities, n = 64; Health Research, n = 69; Science and Engineering, n = 106; Unique = 196.

Alliance Community Cloud

The surveyed revealed that 24% of researchers identified using the Alliance Community Cloud while 65% did not or were unsure (11%). Several researchers specifically commented they were unaware of the distinction between the Alliance Community Cloud, their institutional provision and or that of a regional organization (e.g., Calcul Québec, BC DRI, etc.). In some cases, researchers were unaware that the services the Alliance provides are open to all academic researchers. Those that do use Alliance Community Cloud services, however, do so for a variety of reasons such as the addition of, and/or faster computing resources as well as storage resources, web hosting and specialized hardware (Figure 6). In the Social Sciences and Humanities, granting requirements also played an important role for using Alliance Community Cloud.
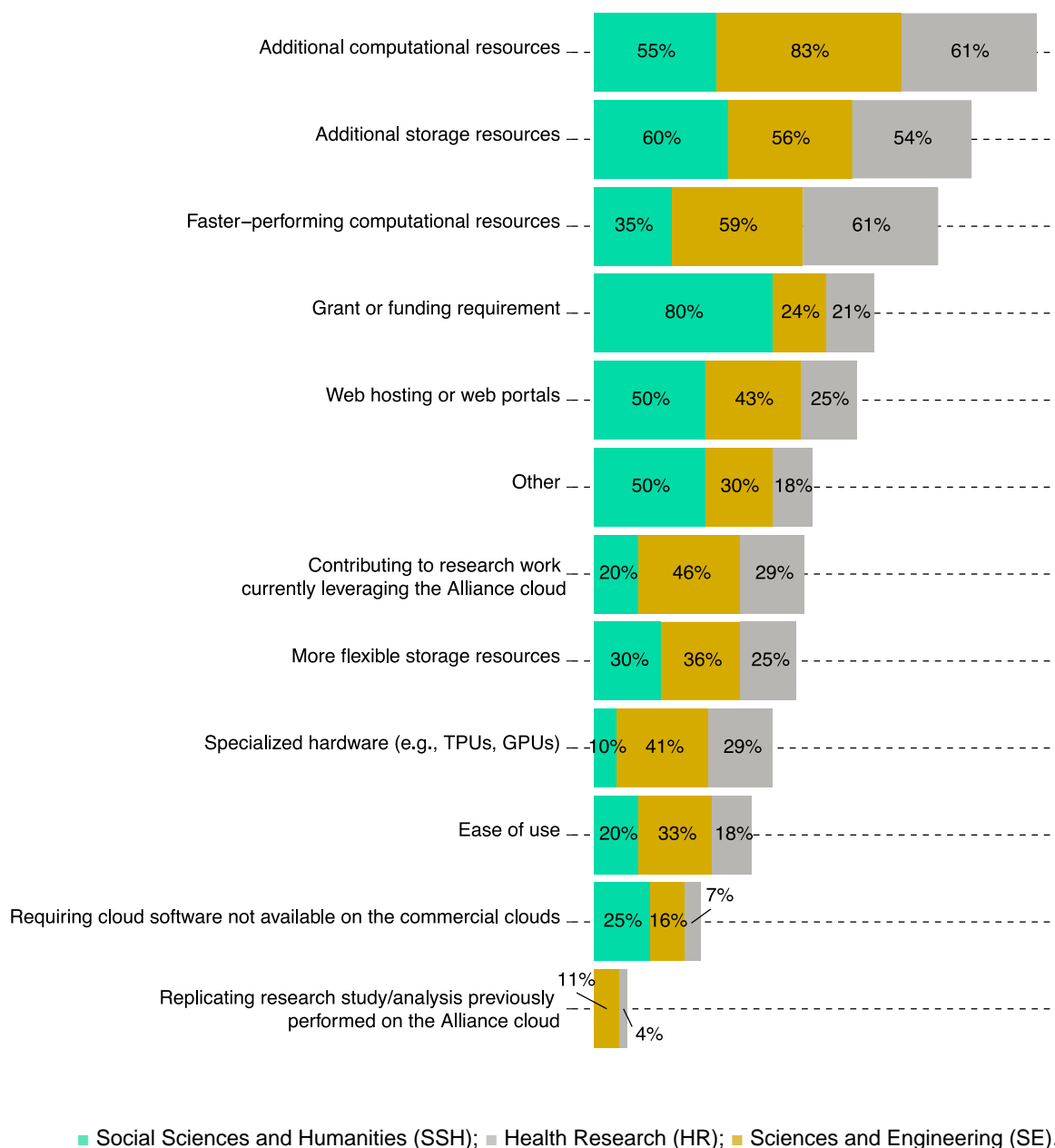
Figure 6. Question: What is/was your primary reason for using the Alliance Community Cloud? Social Sciences and Humanities, n = 20; Health Research, n = 28; Science and Engineering, n = 70; Unique respondents = 100.

When asked what elements or components of the Alliance Community Cloud researchers used specifically, virtual machines were by far the most used (SSH – 84%; SE – 83%; HR – 74%), with far fewer respondents reporting use of object storage or API access (SSH – 32% and 26%; SE – 23% and 26%; HR – 26% and 13% respectively). Many users of the Alliance Community Cloud (44%) further commented on the lack of specific services within the Alliance offering, with particular emphasis on Docker and managed Kubernetes, as well as reminders for content migration and tracking. Many researchers provided further recommendations such as better and more accessible manuals, video tutorials, real-time chat cloud support and specific cloud training. Requests for Virtual office hours were also commonplace.

Commercial vs. Community Cloud

Beyond the need for some containerization services lacking in the Alliance Community Cloud, there were few remarkable differences in how researchers access and use cloud technologies. For example, both are equally used to store research data as backup, or to share it with collaborators, and these data are usually transferred via Secure Shell (SSH) and Secure File Transfer Protocols (SFTP) – although automated data transfers via cloud tools are more prevalent in commercial offerings. In terms of user support, almost 75% of respondents mentioned not getting appropriate vendor support for the use commercial cloud, with a number of researchers acknowledging relying on institutional and Alliance user support for issues related to commercial cloud. Data security and privacy concerns were the main issues many researchers identified when using cloud, and this was ever more important for commercial offerings – more than 50% of respondents acknowledged having concerns about storing data in commercial cloud.

Conclusion

Adoption of Cloud technologies is growing in higher education research worldwide, and Canada is no exception. Many researchers across disciplines use cloud as complementary compute and storage resources – particularly to back up and share their data with collaborators and peers. While the adoption of commercial cloud is increasing, Alliance Community Cloud plays an important role for those researchers that do use these services. When asked about the general importance of different cloud services, for example, the Alliance Community Cloud offering was the most important cloud service for researchers in Sciences and Engineering, while also playing an important role in Health Research (Figure 7). Researchers also heavily rely on institutional cloud offerings, although it is currently difficult to parse out whether this is a true reliability or lack of clarity between service providers, as both are supported by local support staff. Cost and accessibility play major roles in researchers' decisions. Affordable and easy-to-use solutions that do not compromise data privacy and security are the most important features for many researchers, particularly because many use institutional or grant funds to cover such expenses.
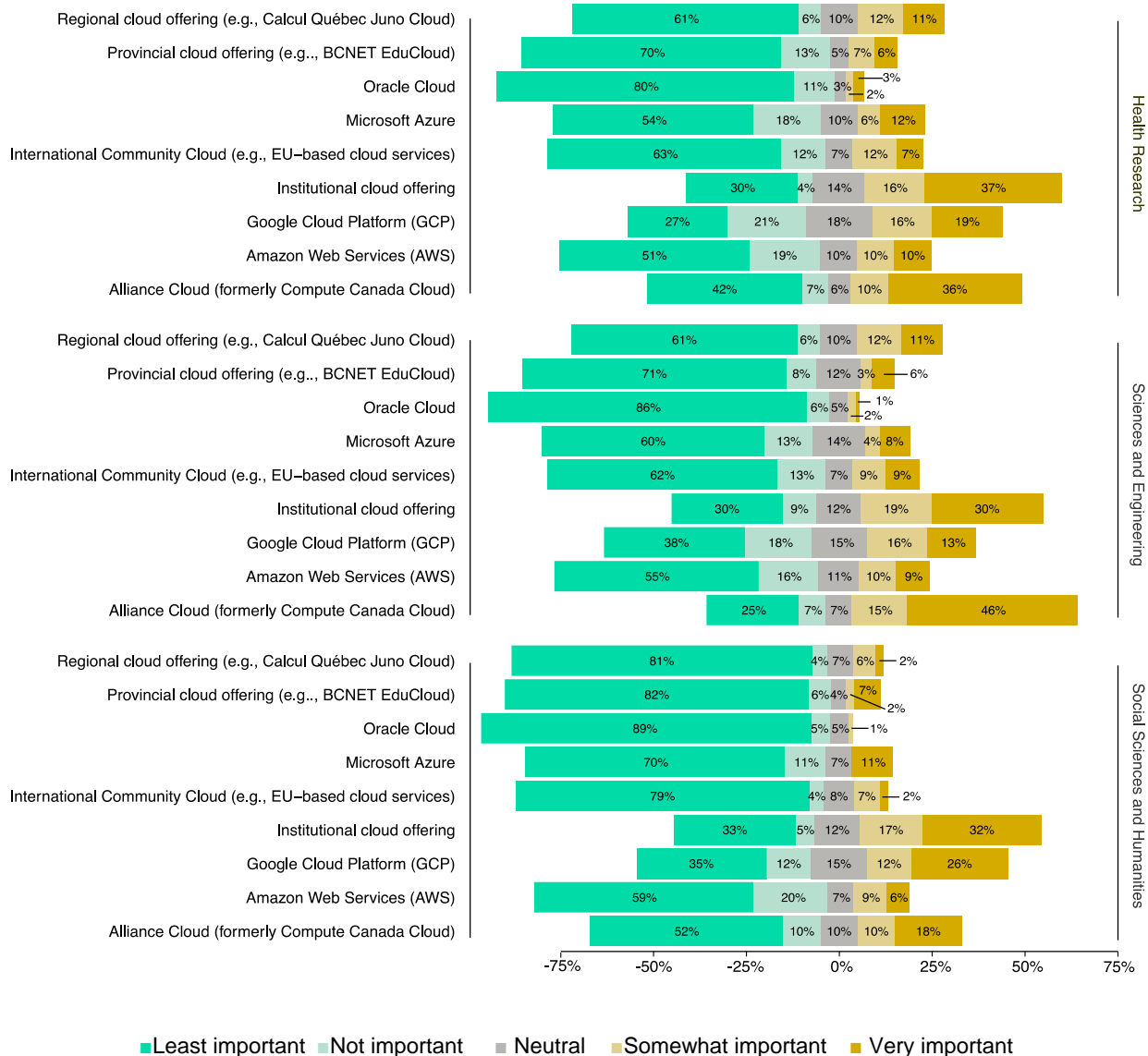
Figure 7. Question: Please indicate how important these cloud services are to support your research. Social Sciences and Humanities, n = 111; Health Research, n = 113; Science and Engineering, n = 196; Unique respondents = 363.

There is a clear need for increased functionality in the Alliance Community Cloud, including managed Kubernetes, containerization like Docker and software platforms such as JupyterHub. Despite the smaller user base, the Alliance Community Cloud would also benefit from revamped user support, training and documentation. Considering the growing cost of commercial cloud and

broad adoption of cloud technologies, it is conceivable that more researchers will rely on Alliance offerings in the future.