

Cloud and on-premises data center usage, expenditures, and approaches to return on investment: A survey of academic research computing organizations

Alan Chalker
alanc@osc.edu
Ohio Supercomputer Center
Columbus, Ohio

Curtis W. Hillegas
curt@princeton.edu
Princeton University
Princeton, New Jersey

Alan Sill
Alan.Sill@ttu.edu
Texas Tech University
Lubbock, Texas

Sharon Broude Geva
sgeva@umich.edu
University of Michigan
Ann Arbor, Michigan

Craig A. Stewart
stewart@iu.edu
Indiana University Pervasive
Technology Institute
Bloomington, Indiana

ABSTRACT

The landscape of research in science and engineering is heavily reliant on computation and data processing. There is continued and expanded usage by disciplines that have historically used advanced computing resources, new usage by disciplines that have not traditionally used HPC, and new modalities of the usage in Data Science, Machine Learning, and other areas of AI. Along with these new patterns have come new advanced computing resource methods and approaches, including the availability of commercial cloud resources. The Coalition for Academic Scientific Computation (CASC) has long been an advocate representing the needs of academic researchers using computational resources, sharing best practices and offering advice to create a national cyberinfrastructure to meet US science, engineering, and other academic computing needs. CASC has completed the first of what we intend to be an annual survey of academic cloud and data center usage and practices in analyzing return on investment in cyberinfrastructure.

Critically important findings from this first survey include the following: many of the respondents are engaged in some form of analysis of return in research computing investments, but only a minority currently report the results of such analyses to their upper-level administration. Most respondents are experimenting with use of commercial cloud resources but no respondent indicated that they have found use of commercial cloud services to create financial benefits compared to their current methods. There is clear correlation between levels of investment in research cyberinfrastructure and the scale of both cpu core-hours delivered and the financial level of supported research grants. Also interesting is that almost every respondent indicated that they participate in some sort of national cooperative or nationally provided research

computing infrastructure project and most were involved in academic computing-related organizations, indicating a high degree of engagement by institutions of higher education in building and maintaining national research computing ecosystems. Institutions continue to evaluate cloud-based HPC service models, despite having generally concluded that so far cloud HPC is too expensive to use compared to their current methods.

CCS CONCEPTS

• **Social and professional topics** → **Computing / technology policy; Management of computing and information systems.**

KEYWORDS

HPC, cloud computing, return on investment

ACM Reference Format:

Alan Chalker, Curtis W. Hillegas, Alan Sill, Sharon Broude Geva, and Craig A. Stewart. 2020. Cloud and on-premises data center usage, expenditures, and approaches to return on investment: A survey of academic research computing organizations. In *Practice and Experience in Advanced Research Computing (PEARC '20)*, July 26–30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3311790.3396642>

1 INTRODUCTION

Science and Engineering research are now, in many disciplines, reliant on advanced research computing (ARC) resources such as:

- High Performance Computing (HPC) clusters
- Resources needed to store, manage, and analyze new sources and modes of ingestion of large data sets
- Human capabilities to operate and support their usage

The Coalition for Academic Scientific Computation (CASC) is a not-for-profit (501(c)3) organization founded in 1989 that currently includes 92 members from higher-education institutions and research organizations. Important elements of CASC's mission are to share best practices in provision and support of advanced computational resources for US researchers, advocate for the importance of and the need for public and private investment in research computing and data services to support academic research, to serve as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PEARC '20, July 26–30, 2020, Portland, OR, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6689-2/20/07...\$15.00
<https://doi.org/10.1145/3311790.3396642>

a trusted advisor to federal agencies on the direction of relevant funding programs, and to actively engage in discussions of policies related to research computing and data services.

In service of these aspects of the mission, CASC and its member organizations frequently analyze the broad picture of the current state of academic use of research cyberinfrastructure (CI) resources. (Cloud computing in particular is one form of, not something distinct from, cyberinfrastructure [9]). Knowing the current state allows the leadership of research CI on campuses across the country to advocate for, justify, and explain, through return on investment (ROI) data, the need for funding resource procurement and leasing, as well the hiring and retention of skilled staff. Campus stakeholders (including decision makers such as Provosts, Deans, Vice Presidents for Research, Chief Information Officers, and Chief Financial Officers) and federal funding agencies (including the National Science Foundation, the National Institutes of Health, and other federal funding agencies) must have and understand this information in order to make informed and prudent decisions about providing CI resources for research projects (through direct funding, allocations on agency systems, or commercial cloud credits) and campus-level CI facilities.

To aid CASC member organizations and the broader community, CASC created a working group to study the current state of academic cloud and data center usage to understand usage patterns, ways in which ROI is analyzed in making usage choices, and current best practices for measuring and reporting ROI values. This working group has initiated an Annual Academic Cloud and Data Center usage survey open to any academic research organization that is willing to provide information, either anonymously or not. This paper is the result of the work of that group to date, and describes conclusions drawn from its first year of survey results.

2 MOTIVATIONS AND GOALS

This current work was done under the auspices of a CASC working group titled the "Return on Investment in Academic Use of Data Centers and Clouds" (ROI-AUDCC), which was formed in early 2018 to pursue the topics discussed above. Prior to this group's formation in October and November 2017, there was an extensive email thread on the CASC mailing list regarding the general topic of ROI and the Cloud. As a result of significant interest from the CASC community, a special session about this topic was scheduled as part of the annual CASC meeting at the SC'17 conference. The many participants in this discussion agreed to develop a "state of the field" position paper. Discussion continued at the March 2018 CASC Spring Meeting, which resulted in the ROI-AUDCC working group being chartered, with two main purposes:

- (1) To collect and document the experience of communities that have already established methods of supporting academic computing, including data management systems, research computing clusters, and distributed research support systems such as data acquisition, visualization, storage, and bulk data processing.
- (2) To define profiles for patterns of usage of distributed resources, such as on-premises, remotely provided, or hybrid

cloud systems that can improve and/or optimize such resources in support of academic usage patterns.

The authors were the key members of this working group and agreed to develop and conduct a survey of the community in alignment with the first main purpose. After discussions, two types of surveys were developed: an initial short-form survey, and a subsequent long-form survey.

3 OVERVIEW, SAMPLE, AND PARTICIPANTS

An initial survey concentrating on organizational computing resource scope and ROI reporting practices was carried out among CASC member institutions in early 2019. The survey's six questions asked participants to select which of several institutional scenarios ranging from small to large-scale was the best match to their organizational computing scale, whether cloud computing resources were made available to users from their organizations through institutional credentials, and whether return on investment information was gathered and reported within the organization, and if so, at what level. Participants were also asked to suggest topics related to ROI for academic uses of data centers and clouds that had not been covered in this initial survey. 31 responses were received.

The results of this preliminary survey were reported to the CASC membership in fall 2019, and then used to design a larger, more comprehensive survey of the general community released at the SC'19 conference. We first describe the results of the short-form survey, and then consider in some detail the results of the longer survey.

4 SHORT-FORM SURVEY FORMAT AND RESPONSES

The short-form survey primarily consisted of a series of descriptive scenarios that the respondent was asked to read. They were then asked to identify a scenario that most closely matched their institution in each area or provide a short description of conditions at their institution. Here is an example scenario included in the survey: *This institution serves a variety of small-scale users who work using individual desktops and laptops, supplemented by up to several isolated laboratories and possibly a single common lab area for access to the web and a few common applications. It does not offer large-scale bulk computing yet, and available clusters if any are not shared with users across the entire institution. It has modest connectivity to the outside network not necessarily including a research-specific network connection.* Other scenarios were provided, ranging up to large-scale organizations that also provide services to national and international computing infrastructures. This survey also asked about cloud usage and the institutional level at which ROI information on investment in CI systems is reported.

The short form survey was announced to CASC members via the CASC mailing list in March 2019 and responses were collected through a customer relations management software facility that is part of the CASC web site. A total of 31 responses were received from the initial short-form survey. 25 of the responses were on behalf of an institution by the official CASC representative. Most of

the responses (26 of 31) were from institutions with significant centralized on-premises cluster resources with centralized file systems managed by a full-time professional staff coordinated by a director, and that offered software-installation support, and consulting in HPC and/or data analysis to users. A similar number (25 out of 31) of the institutions had agreements with commercial cloud vendors for cloud-based resources (e.g., email, data storage, cloud services) that can be accessed through institutional credentials.

Less than a quarter (7 out of 31 institutions) in the short survey reported having negotiated access for users with one or more commercial cloud vendors for access to cloud-based HPC or data storage that can be accessed through institutional credentials. In terms of ROI reporting, less than a quarter (7 of 31 responses) regularly reported ROI results at the institutional level. About half (16 of 31) indicated that return on investment information is gathered and analyzed by the institutional unit supplying HPC resources, but not reported at the overall institutional level, and about 30% (9 of 31) of the responses indicated that no specific ROI information is gathered, or such information is typically only gathered by, available to, and used by individual researchers or research groups.

5 LONG-FORM SURVEY FORMAT AND RESPONSE DEMOGRAPHICS

Based on what we learned through analysis of the short-form survey data, the committee developed a significantly longer survey. The longer form survey was made available to the high performance computing community generally, and not limited to representatives of CASC members. Responses were accepted until mid-January 2020. It was announced to the CASC community in November 2019 at the annual CASC luncheon at SC'19. Physical flyers encouraging participation in the survey were distributed widely to SC'19 attendees. In early December 2019, the survey was also advertised via several computational community mailing lists, including the Open Grid Forum, CaRCC, and Campus Champions lists.

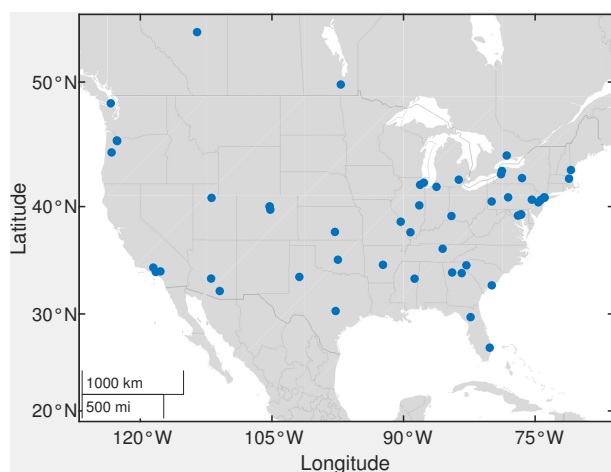


Figure 1: Deduplicated locations of the respondents to the CASC November 2019-January 2020 survey for fully completed responses selected as described in the text.

A total of 60 fully completed surveys were submitted from 53 institutions, of which 34 were received from CASC member institutions, out of a total of 101 responses initiated through the survey tool. The relatively low completion rate appears to be largely the result of the length of the survey. Geocoded locations from the IP information of the responses recorded by the survey tool, along with information in survey responses, were used to deduplicate responses from a given institution. In such cases, we selected the submission that contained the most comprehensive institutional response. The geographic distribution of responses included in the final sample is shown in Figure 1. As can be seen from this figure, the responses sampled institutions primarily located in the USA (48), with four from Canada, and in addition, one response from the United Kingdom.

This survey consisted of 49 questions, most of which had multiple choice answers, and was executed via a Qualtrics online survey tool. Responses were extracted in CSV format file and analyzed with OpenRefine (<https://openrefine.org>), Microsoft Excel, and Apple Numbers. MATLAB was also used for analysis and creation of figures. A copy of the long form survey question content is available on the CASC website. An example chart showing the results of a particular question is shown in Figure 2. Detailed analysis is provided in the next section.

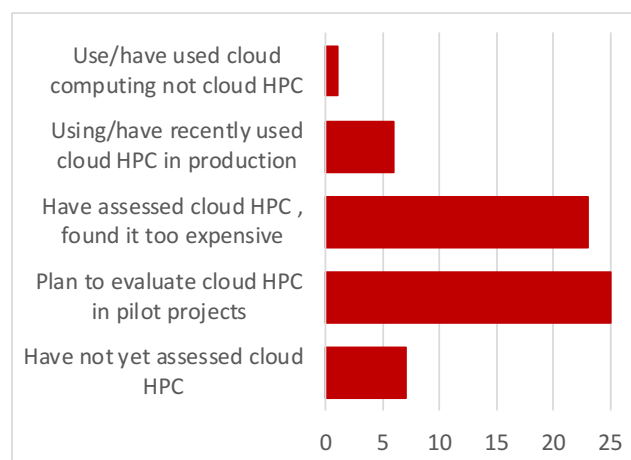


Figure 2: Question 47 - Please categorize your institution's current strategy with respect to cloud-based HPC. (More than one answer could be selected for this question, so responses may overlap.)

6 LONG FORMAT SURVEY RESULTS

Several qualitative conclusions stand out in the responses to both the shorter and the longer survey. Both surveys clearly indicate that practices for determining and reporting ROI are not at all uniform. Many institutions do not appear to report either financial or academic output non-financial ROI statistics to upper levels of their organizations. Where ROI reporting practices do exist, the details of such reporting seem to vary widely. While the responses generally indicate a highly positive financial impact for research

computing in the institutions represented, roughly two thirds of the respondents report that they are not aware of any analysis of ROI on research computing at the level of the entire institution, or that they know that such analysis and reporting do not exist. Notably, institutions with larger-scale research computing investments are more likely to have evaluated the use of cloud-based HPC quantitatively.

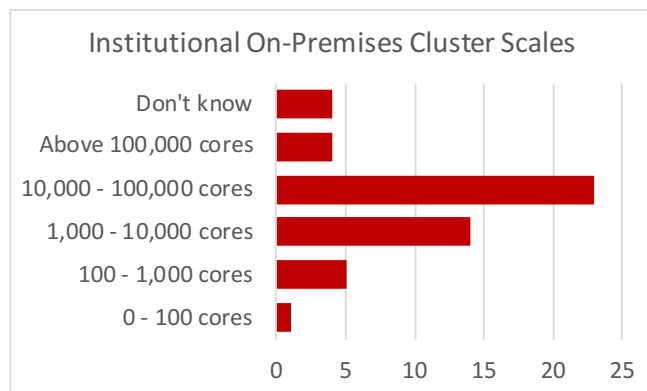


Figure 3: On-premises cluster total sizes per institution reported in the deduplicated CASC survey responses.

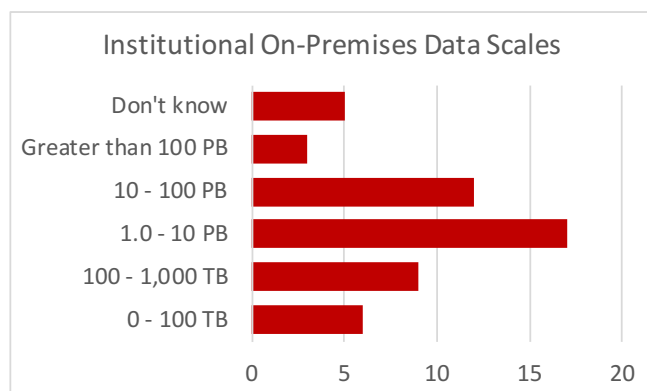


Figure 4: On-premises total data storage per institution reported in the deduplicated CASC survey responses.

Key quantitative results include the following:

- The primary mode of delivery for computational resources to their faculty, staff, and students was through on-premises HPC for 83% of the institutions, whereas 13% provide their clients with access to a hybrid intermix of both on-premises and commercial cloud resources. Data storage shifted the balance: Only 62% indicated they relied only on on-premises resources for data storage, and 30% used a mix of on-premises and cloud data storage.
- Institutions responding to the survey indicated they have access to a range of on-premises HPC clusters as shown in Figure 3, the most common scale falling in the range of 10K-100K total cores.

- Among institutions responding to the survey, on-premises data storage was most commonly in the 1-10 PB range, as shown in Figure 4.
- The most common annual budget was in the \$1M to \$10M range for on-premises advanced computing resources, as shown in Figure 5. Only one institution reported spending more than 10% of its total budget on cloud computing CPU resources, and only three institutions reported exceeding 10% of their total expenditures on cloud storage, although a significant fraction of responses (about 30% in each case) indicated they did not know the answer to these questions.
- Overall, 64% (34/53) of the survey participants indicated that they believe on-premises resources are a better value for most or all of their HPC tasks, and only 6% (3/53) indicated that cloud HPC was a better fit for some of their advanced research computing workloads. None (0/53) of the respondents indicated that cloud was a better value per dollar than on-premises HPC for most or all of their workloads.

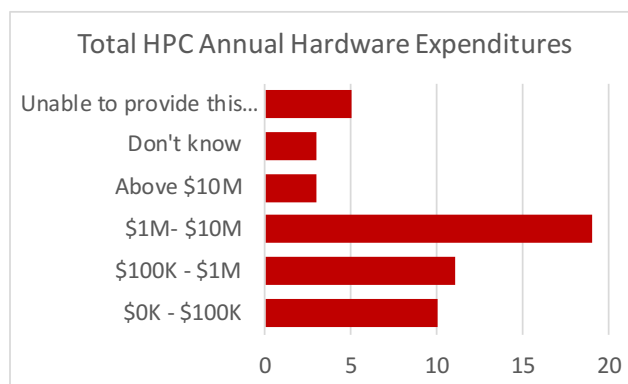


Figure 5: Average annual expenditures for on-premises advanced research computing resources per institution reported in the deduplicated CASC survey responses. Only a small portion of the responses indicated significant fractional expenditures on cloud-based CPU or data storage compared to these values.

Research income vs. engagement with other research computing communities: 49 of the 53 responding institutions are involved in a shared computational project (XSEDE, Open Science Grid, or other similar project). Of those that did not indicate such involvement, two are from outside the US (Canada) so the categories listed may not apply. 51 of the 53 institutions are members of an academic computing-related organization (InCommon, Internet2, Educause, or other similar organization). Of those that did not indicate such involvement, one is from Canada. With almost all respondents indicating involvement in such organizations and projects, it was not possible to determine any correlations with, for example, their research success, except to note that the vast majority of those who took the time to complete the survey clearly has decided that such involvement is important.

Percentage of institutions that have regular budgets for capital: 44 of 53 institutions (83%) indicated that either (a) the institution provides a regular budget for such resources that is used to make decisions on computing, data storage, or network purchases by the staff or administrators directly responsible for HPC facilities, or (b) that the institution makes one-time or intermittent investments in cluster, storage, or outside network bandwidth improvements. Of these, 21 indicated answer (a) and 36 indicated answer (b). Both answers may be true for a particular institution.

Percentage of institutions with regular budgets for operations and people: 45 of 53 institutions (85%) indicated they have a recurring ongoing institutional, college, or departmental support budget for operations and/or personnel maintenance of the resources not passed on to researchers.

Percentage of research computing expenditures on cloud computing, cloud-based storage, or cloud-based network: 38 of 53 answered in the 0%-10% category for cloud CPU expenses as a portion of the total research computing expenditures. 34 of 53 answered in the 0%-10% category for cloud storage expenses as a portion of the total research computing expenditures. 36 of 53 answered in the 0%-10% category for cloud data transfer or network expenses as a portion of the total research computing expenditures. Of note is that the allowable answer categories were not fine enough here, and failed to allow for a 0% response.

Assessments of cloud HPC cost-effectiveness: None of the institutions responding to the survey indicated that they had found cloud-based HPC to represent a cost savings, although most either had completed an assessment or had assessments in progress. (Note the responses in Figure 2 are non-exclusive.) Fully half of the responses indicated that the institution had assessed cloud-based HPC and found it too expensive to use compared to their existing on-premises resources. Nonetheless, a small percentage of responses indicated that the institution considers cloud methods to provide a better fit for some research computing workloads. It is not clear whether “cloud” was taken by these respondents to include things like cloud-based email, document sharing, and similar uses, although the questions in the survey were oriented towards cloud-based HPC applications.

More detail on those that have assessed cloud use and found it to be too expensive: Having the full set of responses allows analysis of how the answers to the questions in Figure 2 depend on each other and break into subsets. Of the 53 responses, 7 institutions indicated they had not yet assessed cloud use. Of the rest, 23 indicated they have assessed cloud use and found it to be too expensive. Of those who have not drawn this conclusion yet, 19 have a strategy and/or project plan to evaluate and/or demonstrate the use of cloud HPC in pilot projects. The large fraction of the responses indicating they were pursuing such pilot projects shows that some institutions continue to assess cloud HPC in spite of having drawn the initial conclusion that it is too expensive compared to their on-premises resources.

7 CORRELATION ANALYSES

In several cases, it is easier to see the relation between answers in different categories by charting the relationships between pairs of answers. Financial questions were phrased in terms of bins like the numerical ranges shown in Figure 5. For the charts in this section, bin indexes were used as follows: 1: \$0K-\$100K, 2: \$100K-\$1M, 3: \$1M-\$10M, 4: Either above \$10M if that was the highest range offered in the question, or \$10M-\$100M if that range was made available, and (where appropriate) 5: above \$100M. For Figures 8 and 9, the ranges for cpu core-hours were 1: 0-100K, 2: 100K-1M, 3: 1-10M, 4: 10-100M, 5: 100-M-1 billion, and 6: Above 1 billion. Colors are used in the 3D bar charts in this section as a guide to the eye to identify the data in each row.

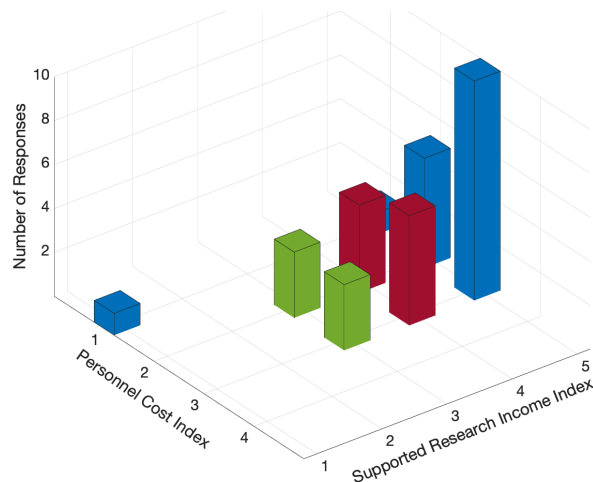


Figure 6: A cross analysis of personnel costs versus supported research.

Annual personnel expenditures versus annual research supported by the resources, regardless of technology (see Figure 6): This chart shows the expected effect that spending more on support gives you more return, and spending less gives you low supported research income. The zero results in the highest index personnel cost row mean that no institutions reported personnel expenditures above \$10M/year, although a significant number reported supporting research income levels above an order of magnitude higher.

Annual hardware expenditures for on-premises equipment versus annual supported research income (see Figure 7): This chart shows a flatter distribution at the high end in terms of results versus expenditure. In comparison with the previous figure, these results show that institutions report spending a bit more of their overall budget on equipment than on people, which may be a questionable assumption in the event that spending moves to the cloud.

Annual supported research income versus annual CPU hours delivered from on-premises resources (see Figure 8): This chart shows a compelling correlation between institutions

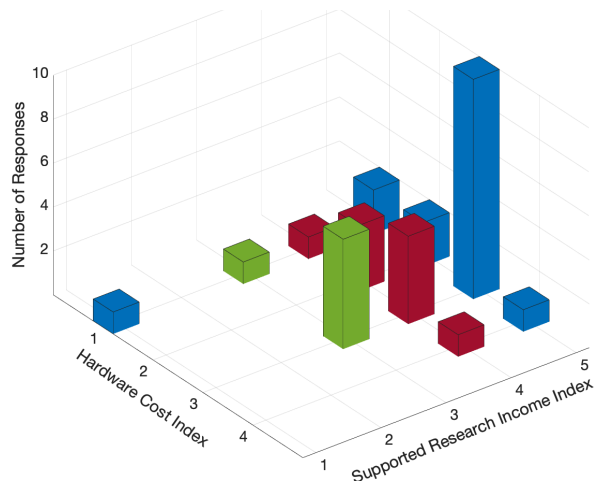


Figure 7: A cross analysis of hardware cost versus supported research.

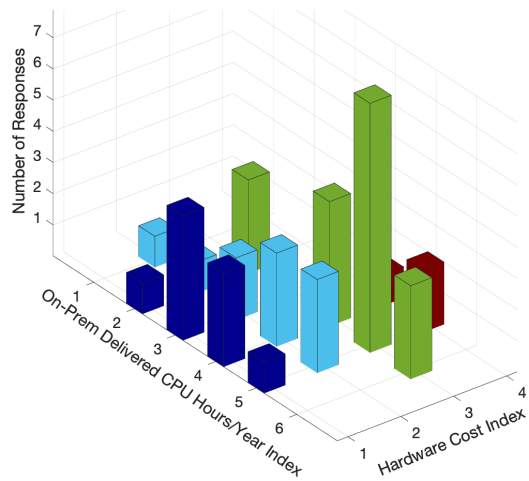


Figure 9: A cross analysis of on-premises delivered CPU hours versus hardware cost.

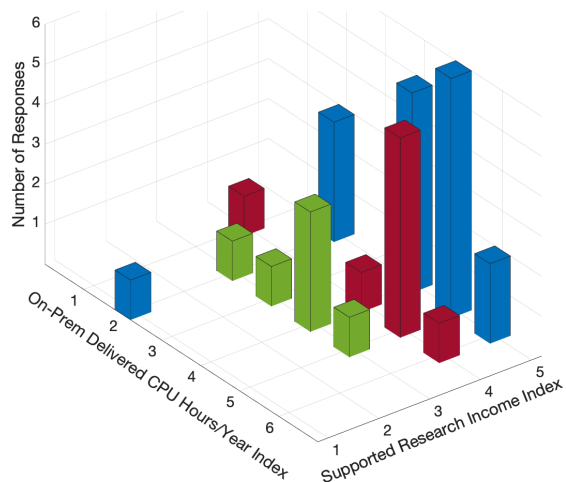


Figure 8: A cross analysis of on-premises delivered CPU hours versus supported research.

obtaining higher research income and delivering high amounts of CPU time.

Annual on-premises hardware expenditures vs. on-premises CPU hours delivered (see Figure 9): This chart is more scattershot compared to the others, but has a rough central trend line. It could reflect a variety of equipment configurations and ages at different institutions.

Annual on-premises hardware expenditures versus personnel expenditures (see Figure 10): This chart is broad in the hardware expenditure distribution but cuts out at the high personnel cost bin, with a strong central trend-line. The results support the previous observation that institutions responding to this survey tend to spend more on hardware than on personnel, although these quantities are roughly correlated, as one might expect. In

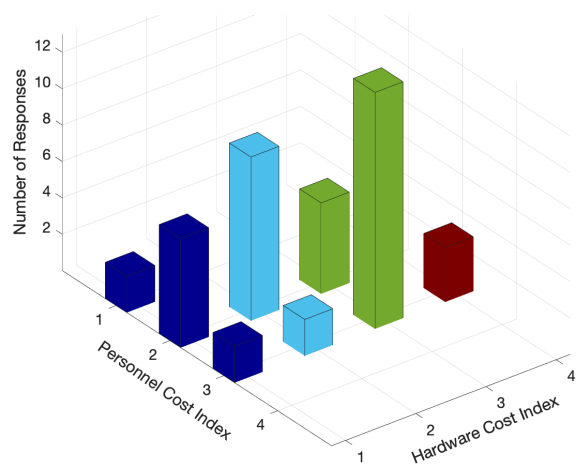


Figure 10: A cross analysis of personnel cost versus hardware cost.

the future, if more institutions spend significant amounts on cloud-based services, it may be interesting to study analogous spending relationships between cloud expenditures and personnel costs to support research.

Annual on-premises personnel expenditures versus operations annual expenditures (see Figure 11): Again, this chart has a rough central trend line, but it is noteworthy that despite typically reporting supported research incomes that were at least an order of magnitude higher than these expenditures as shown in Figure 6, the institutional responses show no data in the highest (“above \$10M”) bin for either variable.

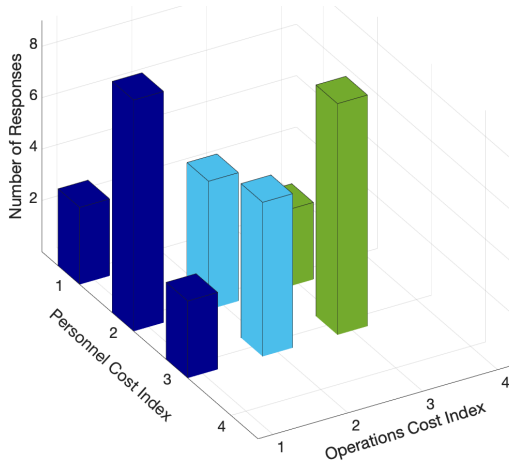


Figure 11: A cross analysis of personnel cost versus operations cost.

8 QUALITATIVE OBSERVATIONS

In addition to the statistical analysis in the previous section, the survey results allow for qualitative observations.

For example, over half of the completed deduplicated responses (28 out of 53) indicated that their institution does not have a defined procedure for determining ROI for infrastructure resources of this nature (i.e., research computing). Only 18 out of 53 indicated that ROI determinations for research computing are made at the level of the entire institution. The remaining responses indicated that a similar minority of institutions looked at ROI at the levels of academic departments or individual research groups. Hence, we can conclude that only about a third of the responses indicate that ROI for research computing is reported to and assessed by top management in the institution, and ROI determination practices vary widely between institutions.

Regarding available resources, most of the respondents who finished the survey were running small to modestly large clusters: fourteen with 1,000-10,000 cores and 28 with 10,000-100,000 cores. Of these, the numbers who stated they have not evaluated cloud versus on-premises HPC in terms of value were almost half in the lower category (six out of fourteen in the set that have 1,000-10,000 cores) but only one out of 28 in the 10,000-100,000 cores category. Thus, there is more interest in evaluating cloud versus on-premises HPC value for the institutions with larger infrastructures. This result, that larger institutions tend to give more detailed responses both in terms of evaluation of cloud versus on-premises data center HPC resources and in terms of practices for evaluation and reporting of ROI for either type of usage, carries through the entire set of survey responses and points in the direction of best practices for the field as a whole to follow.

Looking at the latter category of institutions with modestly large existing CPU resources, half (14 out of 28) answered that “On-premises computing provides better value per dollar for most HPC

tasks, but others are a better fit based on our experience to cloud-based services” and nearly half (13 out of 28) answered that “On-premises computing provides better value per dollar for our HPC needs.” No responses were received that indicated that cloud HPC provided better value per dollar for the institution’s needs, but interest clearly exists to continue exploring this question and to look for corner cases and particular use cases in which the total cost of bulk computing, data storage, or data transfer is not the main consideration.

The survey reported here overlaps substantially in terms of participants and content with a 2013 survey done under the auspices of XSEDE [4]. Comparison of the results of this survey shows that a noticeable gap remains between the expectations for cost savings and the expectations then that cloud technology would form a “democratizing force” for providing equal access to computational resources for institutions of higher education and the current reality. The expectations that cloud resources would be less expensive and more equally accessible do not seem to have been borne out so far.

9 RELATED WORK

Questions and analysis regarding ROI of cyberInfrastructure investments have become a common theme within the HPC community in recent years. Examples of such reports include:

- The Fermi National Lab HEPCloud project did a cost analysis in 2017 that compared the economics and operational trade-offs of cloud versus on-site HPC resources[1].
- Stewart et al. analyzed ROI for three different cyberinfrastructure resources. One of the resources was an on-premises HPC system, and another was an on-premises (OpenStack) cloud system. Both of these resources had a ROI > 1 as compared to the purchase of resources from commercial cloud providers[8].
- Stewart et al. published surveys of methods used to analyze financial and non-financial returns and impacts from investment in research CI. While no new results are presented in these papers, they represent the most current and most comprehensive analyses of value of investments. Perhaps most importantly, these papers outline methods for estimating proxy figures of merit for financial ROI values [6, 7].
- UCSD (University of California San Diego) has carried out recent work on cloud bursting through the Open Science Grid software stack for both CPU and GPU workloads that demonstrates the ability to aggregate large amounts of computation in relatively short amounts of time while attempting to minimize overall costs[5].

There is of course more to selection of computing environments than ROI, as is indicated in the following:

- Chris Downing from Red Oak Consulting in an HPCwire article said, “The last couple of years have seen cloud computing gradually build some legitimacy within the HPC world, but still the HPC industry lies far behind enterprise IT in its willingness to outsource computational power” [2].

- Dennis Gannon from Microsoft published a white paper in which he stated, “Observed from a distance, one might conclude the architectures of cloud data centers and the next generation of supercomputers are converging. However, it is important to keep in mind that the two are designed for different purposes” [3].

10 CONCLUSIONS AND FUTURE PLANS

The surveys reported here are intended to be the first instance of a survey repeated annually. Our intention is to carry out a longitudinal study to evaluate trends in return on investment for academic usage of on-premises data centers and both commercial and community-based academic research clouds. The greatest value will likely be realized through the analysis of trends in the data over a number of years.

With the experience of one round of the survey to rely upon, the authors intend to review and as appropriate revise the questions and survey to improve it and make the resulting data as useful as possible. This must be done in a manner that respects the longitudinal nature of the study allowing for improvements while preserving the ability to observe trends in the data. The working group intends to focus on the following areas:

- Refine questions based on current responses. Are the questions phrased in a way that makes them clear? Are the results consistent, or is it possible that the questions are not specific enough?
- Evolve questions into the future to keep them current with technology.
- Recruit broader participation to represent as diverse a set of institutions as possible. What can be done to result in more responses to the survey? How to get responses from more institutions that are not currently at the forefront of this field? Should the authors be soliciting more international responses, and if so, how to modify the survey to meet the broader needs?
- Allow the research computing and data (RCD) community to ask questions of the data without sacrificing anonymity of results and the corresponding willingness of respondents to be frank and to communicate the actual conditions at their institution.

The longitudinal study we plan is intended to be a community resource. If you are interested in performing further analyses of these data, please contact the authors and / or CASC at casc@casc.org. Suggestions about edits or additions to the questions in the long form survey are also welcome (the questions are available at the CASC website). The results of our initial surveys have already been enlightening and in some cases surprising. We look forward to learning more as this study evolves into the first of an annual series.

ACKNOWLEDGMENTS

The authors appreciate the support of the CASC Executive Officers and Director and the cooperation of those who filled out the surveys.

Stewart’s time on this project was supported by the NSF through grant ACI-1548562, and by the IU Pervasive Technology Institute, which was launched with major financial support from the Lilly Endowment, Inc. The authors would also like to acknowledge Dr. Donald F. (Rick) McMullen and Dr. Shawn Slavin for their assistance in preparing this manuscript.

REFERENCES

- [1] Brian Bockelman Dave Dykstra Ian Fisk Stuart Fuess Gabriele Garzoglio Maria Giron Oliver Gutsche Dirk Hufnagel Hyunwoo Kim Robert Kennedy Nicolo Magini David Mason Panagiotis Spentzouris Anthony Tiradani Steve Timm & Eric W. Vaandering Burt Holzman, Lothar A. T. Bauerdick. 2017. HEPCloud, A new Paradigm for HEP Facilities: CMS Amazon Web Services Investigation. *Comput Softw Big Sci* (2017). <https://doi.org/10.1007/s41781-017-0001-9>
- [2] Chris Downing. 2018. How the Cloud is Falling Short for HPC. *HPCwire* (2018). <https://www.hpcwire.com/2018/03/15/how-the-cloud-is-falling-short-for-research-computing/>
- [3] Dennis Gannon. 2018. The State of the Cloud for Science-2018. (2018). <https://doi.org/10.13140/RG.2.2.27664.30729>
- [4] David Lifka, Ian Foster, Susan Mehringer, Manish Parashar, Paul Redfern, Craig Stewart, and Steve Tuecke. [n.d.]. XSEDE Cloud Survey Report. <http://hdl.handle.net/2022/25307> Accepted: 2020-03-23.
- [5] University of California San Diego press release. 2019, accessed Feb. 17, 2020. Second GPU Cloudburst Experiment Yields New Findings. https://www.sdsu.edu/News%20Items/PR20200212_gpu_cloudburst.html
- [6] Craig A Stewart, Amy Apon, David Y Hancock, Thomas Furlani, Alan Sill, Julie Wernert, David Lifka, Nicholas Berente, Thomas Cheatham, and Shawn D Slavin. 2019. Assessment of Non-financial Returns on Cyberinfrastructure: A Survey of Current Methods. In *Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing* (Chicago, IL, USA) (HARC '19). ACM, New York, NY, USA, 2:1–2:10. <https://doi.org/10.1145/3355738.3355749>
- [7] Craig A. Stewart, David Y. Hancock, Julie Wernert, Thomas Furlani, David Lifka, Alan Sill, Nicholas Berente, Donald F. McMullen, Thomas Cheatham, Amy Apon, Ron Payne, and Shawn D. Slavin. 2019. Assessment of Financial Returns on Investments in Cyberinfrastructure Facilities: A Survey of Current Methods. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* (Chicago, IL, USA) (PEARC '19). ACM, New York, NY, USA, Article 33, 8 pages. <https://doi.org/10.1145/3332186.3332228>
- [8] Craig A Stewart, David Y Hancock, Julie Wernert, Matthew R Link, Nancy Wilkins-Diehr, Therese Miller, Kelly Gaither, and Winona Snapp-Childs. 2018. Return on Investment for Three Cyberinfrastructure Facilities: A Local Campus Supercomputer, the NSF-Funded Jetstream Cloud System, and XSEDE (the eXtreme Science and Engineering Discovery Environment). In *2018 IEEE/ACM 11th International Conference on Utility and Cloud Computing (UCC)*. 223–236. <https://doi.org/10.1109/UCC.2018.00031>
- [9] Craig A Stewart, Richard Knepper, Matthew R Link, Marlon Pierce, Eric Wernert, and Nancy Wilkins-Diehr. 2017. Cyberinfrastructure, Science Gateways, Campus Bridging, and Cloud Computing. In *Encyclopedia of Information Science and Technology*, M. Khosrow-Pour (Ed.). IGI Global, Hershey, PA, 1063–1074. <http://hdl.handle.net/2022/21589>