Survey form and methods for second CASC survey of academic research computing and data center usage

Please cite as:

Authors:
Sharon Broude Geva
sgeva@umich.edu
Orcid ID: 0000-0001-5953-0434
University of Michigan
Ann Arbor
Michigan, USA

Alan Chalker
alanc@osc.edu
Orcid ID: 0000-0002-5475-8779}
Ohio Supercomputer Center
Columbus
Ohio, USA

Curtis W. Hillegas
curt@princeton.edu}
Orcid ID: 0000-0003-3647-9255
Princeton University
Princeton
New Jersey, USA

Donald Petravick
petravic@illinois.edu
University of Illinois
 Champaign-Urbana
 Illinois, USA

Alan Sill
Alan.Sill@ttu.edu}
Orcid ID: 0000-0003-2527-764X}
Texas Tech University
Lubbock
Texas, USA

Craig A. Stewart
stewart@iu.edu
ORCID ID: 0000-0003-2423-9019}

Department of Computer Science, Indiana University;
National Center for Supercomputing Applications, University of Illinois Urbana Champaign
Urbana-Champaign
Illinois, USA

Abstract

Availability of cloud-based resource delivery modes is transforming many areas of computing. Many academic institutions that support research computing facilities are considering and changing their mix of on-premise and remote facilities (including in particular use of commercial cloud facilities). A working group of the Coalition for Academic Scientific Computation (an educational nonprofit 501(c)(3) organization) has conducted an annual survey of higher education institutions now for two years running, with intentions of continuing. This survey asks a number of questions of academic institutions regarding their investments in research and data-oriented computing facilities, the extent of those facilities, and institutional activities. This technical report includes the full text of the survey instrument itself and describes the methods and survey population.

Introduction

Availability of cloud-based resource delivery modes is transforming many areas of computing. Many academic institutions that support research computing facilities are considering and changing their mix of on-premise and remote facilities (including in particular use of commercial cloud facilities).

The Coalition for Academic Scientific Computation [1] is an educational nonprofit 501(c)(3) organization with 93 member institutions representing many of the nation's most forward-thinking universities and computing centers. CASC is dedicated to advocating for the use of the most advanced computing technology to accelerate scientific discovery for national competitiveness, global security, and economic success, as well as develop a diverse and well-prepared 21st century workforce.

CASC supports the operation of a number of working groups, including a working group focused on Return on Investment in advanced academic use of research cyberinfrastructure (CI) resources. Starting in 2020 the CASC ROI working group began surveying academic institutions regarding their investments in research and data-oriented computing facilities, the extent of those facilities, and institutional activities. Published reports on these activities are available online [2,3].

This paper reports describes the methods used in the 2021 survey – the second annual community-wide survey conducted by the Coalition for Advanced Scientific Computation, and presents the contents of the survey instrument itself.

Methods

The survey was administered via the Qualtrics online survey system. Respondents were assured that individual responses would remain anonymous, but identifying information was requested in order to assure unique responses by institution. We issued a broad appeal for participation to the email lists of several academic computing groups and organizations within the US and internationally. In so doing we put priority on and specifically asked for assistance in obtaining broad dissemination to improve the breadth and diversity of representation of participating organizations.

We put out our initial requests for participation on February 12, 2021, and closed the survey on March 16, 2021. Invitations to participate in the survey were sent to the email lists of several organizations, as follows:

- CASC member representative mailing list
- EDUCAUSE Research Computing and Data Community Group (https://www.educause.edu/community/research-computing-and-data-community-group)
- PRACE – Partnership for Advanced Computing in Europe (https://prace-ri.eu)
- United Kingdom HPC-SIG (https://hpc-sig.org.uk)
- AIHEC – American Indian Higher Education Consortium (http://www.aihec.org)
- Minority-Serving Cyberinfrastructure Consortium (https://www.aamu.edu/about/administrative-offices/information-technology-services/initiatives/minority-serving-cyberinfrastructure-consortium/index.html)
- WHPC – Women in High Performance Computing (https://womeninhpc.org)
- SWEETER - SWEETER–SouthWest Expertise in Expanding, Training, Education and Research (https://hprc.tamu.edu/sweeter/)

Because CASC members represent both the 'sponsors' of this work, in some ways, and in other ways represent the core population from which sample from year to year, most attention was devoted to encouraging people to submit information related to the activities of CASC member institutions. CASC member institutions were asked to indicate their membership, but this was not required to participate in the survey. (We retain the ability to track for example the usage patterns of CASC members over time). With an initial invitation to the CASC member email list, we sent reminders on February 23 and March 10.

We received a total of 71 responses, of which 68 were complete and 52 were from CASC members.

As compared with the 2020 survey, the 2021 survey was simplified to make the survey easier to fill out and more compact. When asking about usage of various forms of computing and data resources, for example, we typically asked respondents to characterize their use of a particular type of resource as one of: "predominant"; "significant"; or "minimal/none." The majority of the questions only required the respondent to select from a short list of possible answers. A few questions asked for short text answers, and only the first few demographic questions were required to be answered. This simplified both responding to the survey and analysis of the results.

The Survey Instrument

The full text of the survey instrument is shown below:

Default Question Block

CASC Annual Survey
on Resource Funding and Usage
for Academic Research Computing and Data

Overview:
The Coalition for Academic Scientific Computation (CASC) regularly conducts an Annual
Survey on Resource Funding and Usage for Academic Research Computing and Data. The
purpose of this survey is to gather information related to use and funding of commercial and
academic clouds and other forms of shared distributed computing, on-premises data centers,
and other physical and virtual resources in academic research computing and data settings.
The summarized results of the survey will be made available to be shared freely with the broad
community, including institutional and federal stakeholders.

Participant Qualifications:
All institutions are encouraged to complete this survey even if they are not CASC members.
This will ensure that the results of the survey broadly represent the needs and landscape of
institutions providing research computing and data services to support academic research.
We are especially interested in receiving responses from institutions serving historically
underrepresented communities. Please note that we encourage you to collaborate with your
colleagues on a unified response from your institution.

About CASC:
CASC is a 501(c)(3) non-profit organization with about 100 members representing scientific
computing centers of higher-education institutions, national labs and academic research
centers. CASC is conducting this survey as part of its mission to advocate for the importance
of and need for public and private investment in research computing and data services to
support academic research and to serve as a trusted advisor to federal agencies on the
direction of relevant funding programs.

Survey Logistics:

- This survey consists of 6 sections, each with 5-7 questions.

- Only a few initial questions require answers and are marked as such. The rest are optional.
- All questions intrinsically apply to your institution in aggregate unless otherwise stated.
- If you don't know any answer, or are unable to provide one, please feel free to skip a question.
- Please help by completing the survey no later than March 9, 2021.

For questions about this survey, please contact the CASC survey team (annualsurvey@casc.org)

Definition of Terms:

"Research Computing and Data (RCD)": Services and resources not provided for general enterprise computing (e.g. email servers, web hosting, financial systems, etc), but rather to directly facilitate research efforts (e.g. high-performance or throughput computing, data analysis, machine learning, etc.)

"Resources": Hardware and/or software related to RCD (e.g. computer servers, data storage, research networks, computational software, etc)

"Researchers": Anyone making use of RCD resources at the institution, including faculty, postdoctoral fellows, research staff, students, etc

"Access": The ability to use RCD resources based on institutional login credentials or accounts that are tied to funding that is associated with or managed by the institution

"Support Staff": Anyone whose primary job responsibilities are to operate, maintain, or support the use of RCD resources

"Core Hour": Equal to 1 CPU core running for 1 hour. Note 1 individual core over a year is 8760 core-hours, so a 10k-core cluster would have a theoretical 87.6M core-hours available.

"Return on Investment (ROI)": The relative value realized per dollar invested in something

For the purpose of this survey, all RCD Resources are considered exclusively in one of the four following categories:

"On-premises": RCD Resources residing in facilities under direct control of an administrative group within an institution and available for use by many researchers (e.g. CIO office, Colleges, etc)

"Commercial Cloud": RCD Resources not residing in facilities under direct control of an institution (e.g. commercial service providers such as Amazon, Microsoft, Google, etc.)

"National / Regional": RCD Resources provided by non-commercial entities such as XSEDE or DOE labs, etc. These can include fixed-location facilities as well as clouds (e.g., Jetstream)

"Individual": RCD Resources directly under the control by and only available for use of individual researchers or research groups (e.g. small lab-based clusters, powerful workstations, etc.)

All RCD Resources can also be considered exclusively as one of the three following types:

"Computational": RCD Resources whose primary purpose is to analyze data

"Storage": RCD Resources whose primary purpose is to store data

"Network": RCD Resources whose primary purpose is to transfer data

Section A (1 out of 6 total sections): Demographics

Is your institution a CASC member? (Required)

Yes

No

Are you responding authoritatively as or on behalf of the CASC voting member representative? (Required)

Yes

No

Are you responding authoritatively on behalf of your institution? (Required)

Yes

No

In which country is your institution based? (Required)

What is the name of your institution?

Did you respond to this survey last year?

Yes

No

Please provide your name (in case we have multiple responses from your institution or need clarification about any of your answers)

[   ]

Section B (2 out of 6 total sections): RCD Resources

If significant **RCD resources** are not being utilized by researchers, what would help change this? (Select all that apply)

Increased availability of federal funding for resources

Educational materials / awareness campaigns targeted at researchers

Technical support assistance for support staff

[   ]Other

What is the relative scale of each **computational resources** category utilized by researchers?

On-premises	Predominant	Significant	Minimal / None
Commercial Cloud	Predominant	Significant	Minimal / None
National/Regional	Predominant	Significant	Minimal / None
Individual	Predominant	Singificant	Minimal / None

What is the relative scale of each **storage resources** category utilized by researchers?

On-premises	Predominant	Significant	Minimal / None
Commercial Cloud	Predominant	Significant	Minimal / None
National/Regional	Predominant	Significant	Minimal / None
Individual	Predominant	Significant	Minimal / None

What is the total number of CPU cores available across all **on-premises computational resources**?

0 - 100 cores

100 - 1,000 cores

1,000 - 10,000 cores

10,000 - 100,000 cores

Above 100,000 cores

What is the total capacity available across all **on-premises storage resources**?

0 - 100 TB

100 - 1,000 TB (0.1 - 1.0 PB)

1,000 - 10,000 TB (1.0 - 10 PB)

10,000 - 100,000 TB (10 - 100 PB)

Above 100,000 TB (100 PB)

What is the total number of CPU core hours utilized by researchers on **commercial cloud computational resources**?

0 - 100,000 core-hours

100,000 - 1 million core-hours

1 - 10 million core-hours

10 - 100 million core-hours

Above 100 million core-hours

What is the total number of TB utilized by researchers on **commercial cloud storage resources**?

0 - 100 TB

100 - 1,000 TB (0.1 - 1.0 PB)

1,000 - 10,000 TB (1.0 - 10 PB)

10,000 - 100,000 TB (10 - 100 PB)

Above 100,000 TB (100 PB)

What is the total **off-campus network (internet) bandwidth** available to researchers?

0 - 1 Gbps

1 - 10 Gbps

10 - 100 Gbps

100 - 400 Gbps

Above 400 Gbps

Section C (3 out of 6 total sections) - RCD Investments

What is the average total annual capital expenditures for **on-premises computational resources**?

$0K - $100K

$100K - $1M

$1M- $10M

Above $10M

What is the average total annual operating expenditures for **on-premises computational resources**? (e.g. power, maintenance contracts, etc , but **not support staff** costs)

$0K - $100K

$100K - $1M

$1M- $10M

Above $10M

What is the average total annual support staff operating expenditures for **on-premises computational resources**? (e.g. salaries and other support costs not included in the previous question)

$0 - $100K
$100K - $1M
$1M- $10M
Above $10M

What funding models are used for capital expenses for **RCD resources** (Select all that apply)

Researchers or research groups write grant requests to purchase individual resources
Colleges or departments fund on-premises resources that are offered to researchers within their own organizational unit
The institution makes one-time or intermittent investments in on-premises resources
The institution provides a regular budget for on-premises resources

What funding models are used for operating and staff support expenses for **RCD resources**? (Select all that apply)

Researchers or research groups billed after the fact based of their usage
Colleges or departments provide annual or other regular contributions
Ongoing institutional budget

What rough fraction of the total **institutional RCD budget** (capital and operating) is expended on each of the resource types below? (Total must sum to 100%)

Commercial cloud computational resources	0
Commercial cloud data storage	0
Commercial cloud network services	0
On-premises resources	0
Total	0

Section D (4 out of 6 total sections) - RCD Return on Investments

At what levels are data on the below **RCD related** topics compiled? (Select all that apply)

Resource usage	Researchers / groups	Colleges / departments	Institution as a whole
Academic outcomes	Researchers / groups	Colleges / departments	Institution as a whole
Return on Investment (ROI)	Researchers / groups	Colleges / departments	Institution as a whole

How often is such data compiled?

Resource usage	Less frequently than annually	Annually	More frequently than annually
Academic outcomes	Less frequently than annually	Annually	More frequently than annually
Return on Investment (ROI)	Less frequently than annually	Annually	More frequently than annually

Who has access to reports on such data? (Select all that apply)

Resource usage	Institutional Administration	Researchers	General Public
Academic outcomes	Institutional Administration	Researchers	General Public
Return on Investment (ROI)	Institutional Administration	Researchers	General Public

What is the average total annual **institutional research income** for all researchers that utilize on-premises resources? (e.g. grants, awards, contracts, etc)

$0 - $100K
$100K - $1M
$1M - $10M
$10M- $100M
Above $100M

Which of the following best describes when **on-premises versus commercial cloud resources** provides better return on investment for your research computing needs?

On-premises is better for nearly all needs
On-premises is better for more situations than commercial cloud
Commercial cloud is better for more situations than On-premises
Commercial cloud is better for nearly all needs

If you are utilizing **commercial cloud resources**, what are the major reasons? (Select all that apply)

Need for cloud-native capabilities
Resilience and 'always-on' functionality
Overall cost reduction
Experimentation with utility for research
Autonomy for researchers to select resources to utilize
Inability to provide sufficient on-premises resources
Other

Section E (5 out of 6 total sections) - RCD Policies and Procedures

What is the level at which **on-premises and/or individual computational resources** are managed? (Select all that apply)

Researchers and/or research groups
College and/or department level
Institutional level

What is the level at which **on-premises and/or individual storage resources** are managed? (Select all that apply)

Researchers and/or research groups
College and/or department level
Institutional level

Are there institutional policies and/or procedures that impact the use of these **RCD resource categories**? (e.g. purchasing restrictions, cybersecurity requirements, energy efficiency goals, etc.)

On-premises	Required to use	Encouraged to use	Discouraged to use	Prohibited to use
Commercial Cloud	Required to use	Encouraged to use	Discouraged to use	Prohibited to use
National / Regional	Required to use	Encouraged to use	Discouraged to use	Prohibited to use
Individual	Required to use	Encouraged to use	Discouraged to use	Prohibited to use

Who is the highest-level person with **direct day-to-day supervisory** responsibility for RCD resources?

Student or Postdoc
Faculty or Professional Staff
Center Director or similar
CIO or VP or similar
President or Provost or similar

Which **RCD related organizations** are researchers engaged with? (Select all that apply.)

InCommon
Internet2
Open Grid Forum (OGF)
Educause

Does your institution have a designated **campus cyberinfrastructure champion**?

Yes
No

Section F (6 out of 6 total sections) - Misc Questions

Do researchers make use of or participate in any of these **National / Regional RCD resources**? (Select all that apply.)

US NSF XSEDE

Open Science Grid

Campus Compute Cooperative

Aristotle Federated Cloud

Internet2 E-CAS

US DOE National Labs

US Department of Defense HPCMP

US NASA Centers

NIH STRIDES Initiative

If the COVID-19 pandemic has resulted in **significant changes** to your institution's investment levels for RCD resources and/or policies compared to previous years, can you please provide a brief description?

Please categorize your current strategy with respect to **Commercial Cloud computational resources** (Select all that apply)

We have not yet assessed them

We have assessed them and decided not to utilize them

We are and/or plan to utilize them for pilot projects

We are and/or plan to utilize them for production projects

Please categorize your current strategy with respect to **Commercial Cloud storage resources** (Select all that apply)

We have not yet assessed them

We have assessed them and decided not to utilize them

We are and/or plan to utilize them for pilot projects

We are and/or plan to utilize them for production projects

Please provide any studies / whitepapers / resources you recommend regarding **RCD resource category** comparisons

What component of the topic of **ROI for RCD** has not been covered in this survey, from your point of view?

References

[1] Coalition for Academic Scientific Computation. 2021. Home page. www.casc.org

[2] Chalker, A. C. W. Hillegas, A. Sill, Sharon B. Geva, and C.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(Port-land, OR, USA)(PEARC '20). Association for Computing Machinery, New York, NY, USA, 26–33. https://doi.org/10.1145/3311790.3396642

[3] Geva, S.B., A. Chalker, C.W. Hillegas, D. Petravick, A. Sill, C,.A,. Stewart. 2021. Results from a second longitudinal survey of academic research computing and data center usage: expenditures, utilization patterns, and approaches to return on investment. *In* Practice and Experience in Advanced Research Computing (PEARC '21), July 18-22, 2021, Boston, MA, USA. https://doi.org/10.1145/3437359.3465589}