Integrated Summary of

NIH Virtual Workshop on Broadening Cloud Computing Usage in Biomedical Research

and

Request for Information (RFI): NIH Programs to Increase Access to Cloud Computing to Diverse Biomedical Research Institutions

Background and Purpose of the Workshop and RFI

The NIH Strategic Plan for Data Science aims to support an efficient, effective biomedical research data infrastructure for NIH researchers. To achieve this goal, NIH launched the Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative, expanding available cloud resources for data science by partnering with commercial cloud service providers including Google, AWS, and Microsoft Azure. Many well-funded institutions have begun to transition and use cloud-based resources to enable research, however it has been unclear how less research-intensive institutions perceive accessibility and desirability of cloud computing.

To better understand current and future needs and uses of cloud computing for biomedical, clinical, behavioral, and social sciences, the Office of Data Science Strategy (ODSS), the National Institute of General Medical Sciences (NIGMS), and the National Institute of Minority Health and Health Disparities (NIMHD) published the "Request for Information (RFI): NIH Programs to Increase Access to Cloud Computing to Diverse Biomedical Research Institutions" on July 8, 2021. These agencies also coordinated with the Center for Information Technology (CIT) to hold the "NIH Virtual Workshop on Broadening Cloud Computing Usage in Biomedical Research" September 13-14, 2021. The workshop sought to identify challenges and strategies to broadly adopting cloud computing among universities and colleges within Institutional Development Award (IDeA) eligible states, institutions with Research Centers in Minority Institutions (RCMIs), Tribal Colleges and Universities (TCUs), and other Minority Serving Institutions (MSIs), identified as "target communities" in this summary.

Content of the Workshop and RFI

Outreach for both the workshop and RFI focused on target communities. Plenary sessions were open to registrants and NIH staff observers, whereas breakout sessions were limited to participants from targeted communities.

The workshop consisted of two half-day sessions.

Dr. Susan Gregurick, NIH Associate Director for Data Science and Director of ODSS, opened the workshop by presenting NIH's efforts to create a connected, equitable data ecosystem. Mr. Nick Weber, Program Manager for Cloud Services in (CIT), introduced basic information about cloud computing and described some of NIH's cloud-based research platforms. Participants were divided into groups based on institution type (IDeA Networks of Biomedical Research Excellence [INBRE], RCMI, TCU, and MSI) to discuss the needs and challenges of cloud computing in their own research areas and institutions.

Dr. Lakshmi Kumar Matukumalli, the NIH program officer for INBRE, opened the second day by introducing NIGMS cloud sandbox pilot projects, followed by three presentations from NIH awardees: Dr. Benjamin King from University of Maine, Dr. Stephanie Byrum from University of Arkansas for Medical Sciences, and Dr. Gloria Washington from Howard University. These presentations demonstrated the application of cloud computing for their research and training. Workshop participants attended breakout sessions according to institution type to discuss their research opportunities and access to cloud computing.

Key Findings

Information was collected from three sources: workshop breakout sessions, live polls during workshop plenary session, and RFI responses. Findings were highly consistent. The key findings represent conclusions from all three sources, unless noted otherwise. Although attendance in some breakout sessions and the number of RFI responses were lower than expected, the consistency of the findings suggests that the barriers and opportunities expressed largely reflect the experiences of broader populations represented by the respondents.

Cloud computing needs

- Participants expressed the need for and interest in adopting cloud computing. Some participants also expressed the need for information on how to start the process of adapting to a cloud computing environment.

Challenges to adopting cloud computing

- Many participants use on-site computing infrastructure, including high-performance computing (HPC) clusters, and expressed reluctance to give up investments in local computing. They expressed the desire to use cloud computing to complement the existing on-site resources.

- Participants expressed the need for a greater understanding of cloud computing, including terminology, resources, tools, and infrastructure. Participants from RCMIs and TCUs consistently raised this concern in workshop breakout sessions. Concerns include:

    - What are the resources available in the cloud (e.g., what types of computing and storage resources are available in the cloud and what are the differences)? What types of data and software resources are available in the cloud and what are the benefits of using them in the cloud? What are the differences between the various tools in the cloud, or between tools in the cloud and those on local servers? How does one decide which of these resources are needed?

    - What is the cost model of using cloud computing? For example, the cost of using some cloud resources is straightforward while the cost of using other resources might be hard to estimate. Do resources and/or cost estimation approaches exist to help increase accuracy? Why is the cost structure so complicated? How does one learn to use cloud most cost effectively?

    - What kinds of new research can be conducted in the cloud? How does one begin asking questions that cloud computing can offer?

    - What change in research structures can foster collaborative teams of researchers?
    - How can cloud be useful for areas such as social sciences when more common applications tend to be in the genomics area? While some participants are excited to change the way health disparities research is approached if they have access to cloud, they were unsure how to start the transition.
    - Is the data stored in the cloud secure? Why is it secure or not secure? What kinds of security standards are available for cloud-based systems and how can an "appropriate" level of security be reached?
- Participants stated a need for cloud computing training. They asked for a wide range of training from principal investigator (PI) through the pipeline of undergraduate students and IT support. The participants expressed training needs: for both Computer Science (CS) and non-CS students; on both cloud IT skills and biomedical applications; and for beginner, intermediate, and advanced users.
- Participants were uncertain about cloud cost and long-term sustainability (how to pay monthly cloud usage bills after their grants expire). This is not an issue for on-site computing; however, it is a major concern for the cloud that may prevent adoption.
- Participants were concerned about cloud IT support and needs related to governance.

Opportunities in cloud computing

- Participants said cloud environments could enhance data/software sharing and reuse, which can improve research efficiency and reproducibility.
- Some participants stated that collaboration between cloud users, especially collaboration with cloud-resource-rich institutions, could address many perceived challenges. Partnerships could also help to establish research clusters and conduct advanced collaborative research.
- Participants noted the direct benefits of using cloud, including access to resources such as scalable computing power, storage, and cloud-based big data and software applications.
- Many participants stated cloud training could help the career development of various groups including students, faculty, researchers, and IT professionals in data science and cloud computing fields.
- Many participants noted they could conduct innovative research driven by cloud-based big data, interoperable software, and scalable computing and storage. This approach could foster multidisciplinary collaborations and innovative research as seen in the National COVID Cohort Collaborative (N3C).

Conclusion

NIH ODSS, NIGMS, NIMHD and CIT collaboratively issued a RFI and held a workshop to solicit information on needs, challenges, and opportunities of cloud computing in target communities. The information from RFI responses and workshop discussions are consistent. The need and desire for cloud computing in target communities was confirmed, though a broad range of challenges were raised. The greatest challenge is the need for training on cloud computing. Training is needed for all levels of users from cloud novice to experienced users. Cost, cloud IT support, and data security were also major concerns. Additional challenges were identified, including a need to understand cloud resources and how to utilize them to conduct innovative research, as well as how to use the cloud to complement on-site computing. Participants identified several opportunities including new research, collaboration, career development, data sharing, and reuse brought by the cloud as exciting opportunities.