Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-03-26 23:01:11
REFERENCE NO: 195

Author Names & Affiliations

- Thomas Furlani - Director, Center for Computational Research, University at Buffalo, SUNY
- Venu Govindaraju - Vice President, Office of Research and Economic Development, University at Buffalo, SUNY
- Liesl Folks - Dean, School of Engineering and Applied Sciences, University at Buffalo, SUNY

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Furlani: High performance computing, Computational Chemistry; Govindaraju: Biometrics, machine learning; Folks: Physics, materials informatics

Title of Submission

Improving national CI capacity through moderate sized campus systems

Abstract (maximum ~200 words).

We propose that NSF improve the nation’s CI capacity and capability through increased investment in moderate sized campus based CI systems. The arguments in favor of this strategy are several. First, many of today’s scientific application codes do not scale well beyond a few hundred cores or the problems being addressed by faculty are not large enough for the application to run at larger node sizes. Second, many of the large national facilities are designed/optimized for the most demanding computational problems with expensive internode networks and filesystems designed for large I/O rates, making jobs comprising tens to hundreds of processors not a cost effective use of these resources. Third, the operating costs (power, cooling, and staffing) of campus systems are typically borne by the host institution thereby maximizing NSF’s investment. Finally, campus CI plays a crucial role in the training of a next generation who will lead the creation of the new HPC tools and methodologies.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

As a result of a decade of targeted investments, the University at Buffalo (UB) today has a remarkably strong assembly of faculty with expertise in a wide range of computational and data intensive areas. In fact, many of these faculty members specifically specialize in computational methods to the extent that our computational facilities are the primary factor that enables their research. These faculty are distributed over a range of disciplines, including biostatistics, chemistry, economics, finance, geology, pharmacy, physics, mathematics, materials, medicine, and the engineering disciplines. In addition, many of UB’s high-profile research centers are heavily dependent on advanced computing and data analytics, including the NYS Center of Excellence in Materials Informatics (CMI), the Institute for Research and Education in Energy, Environment, and Water (RENEW), the NYS Center of Excellence in Bioinformatics and Life Sciences, the Institute for Lasers, Photonics, and Biophotonics (ILPB), the Buffalo Institute for Genomics and Data Analytics (BIG), and the Buffalo Clinical and Translational Science (CTSA) center. UB's new innovative PhD program in Computational and Data-Enabled Science and Engineering (CDSE), which was established in response to the need for such specialized training, is another such program. The continued success of these efforts depends critically on our ability to provide adequate computing and data analysis capability.

Further, there is growing interest at UB and elsewhere in the area of Digital Humanities - the name given to a vast field of research projects at the intersection of Humanities disciplines and computing that provide both new archival resources for Humanities scholars and opportunities for asking new kinds of questions. Through its interdisciplinarity, Digital Humanities can transform research in the humanities while offering new conceptual and practical challenges to researchers in the social sciences as well as those in STEM fields, such as computer, physical, and cognitive sciences.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Given the continually increasing demand in advanced computing and data infrastructure to support the university's research and educational initiatives described in question 1 (this is not a situation unique to UB), the university's advanced cyberinfrastructure is unable to keep pace with demand through current external funding opportunities. While NSF has helped significantly by earmarking national resources to meet the needs at the high end, it has not been sufficient to keep pace with campus demand. In addition, much of the computing done at the campus level involves research and applications that are well suited (optimized) for moderate size parallel jobs on the order of tens to hundreds of processors or large memory nodes; jobs that are best handled locally. Indeed, most of UB's CI users run simulations in the range of tens to hundreds of processors (cores).

To support the need for moderate size CI systems, we reference a recent workload analysis carried out on Blue Waters (https://drive.google.com/open?id=0B_EbrAJu57PsQi1rV0xPU1kwT1U) in which Blue Waters utilization by job size was studied (see Figure 3.0-1 Variation in job size over time on Blue Waters - page 35) and note that the node hours delivered for 1 – 64 node jobs are roughly equal to the other job size distributions (since late 2015). While this is partially driven by the allocation process, it speaks to the importance of these "small" jobs to advance research and education. It is our position that jobs in the 1 – 64 node range would best be accommodated on campus level systems, freeing up valuable cycles on the national systems to solve the computationally challenging problems requiring much larger resources than can be provided at the campus level.

By increasing the funding available for campus based systems (either through MRI's or another program), NSF could solve two problems simultaneously, namely freeing up cycles on the large national research for grand challenge class problems, and greatly improving the campus based CI capacity. Another advantage to increased support for campus based systems is that campus MRI's typically don't include personnel or operational support costs, while national scale systems do, and therefore the dollars invested in a campus system have the potential to provide more compute power per dollar spent. Campus HPC facilities also leverage additional (often large) local university investments.

An additional role that we emphasize here in support of improved campus CI is the training of a new generation of computational and data scientists who will lead the creation of the new tools and methodologies. This training requires access to diverse and a broad range of cutting edge hardware locally so students may learn and innovate quickly while using national capability resources for production usage. Continued investment in campus CI (coupled to local campus support, both financial and through multidisciplinary training efforts such as UB's CDSE program) is critical to maintain this pipeline of next generation computational and data scientists. Computational and data intensive activities are continuing to gain in importance for fields not traditionally active in these areas. Strong evidence in support of the increasing usage by fields not historically invested in CI is found in Figure 2.1-2 (page 15 of the Blue Waters Workload Analysis Report), which shows the breakdown in use of Blue Waters by NSF directorate over time.

The availability and flexibility of local CI resources (both hardware and faculty/staff expertise) is crucial to developing and maintaining outreach activities, particularly for non-traditional areas in the humanities and social sciences (indeed, XSEDE's campus champion outreach activities are largely focused on investing in, and leveraging local campus resources, both personnel and infrastructure, for exactly this goal).

Finally, in many academic settings, faculty in the Humanities and other “non-traditional HPC areas” are poorly connected to and poorly positioned to leverage National CI facilities to help them with their computing problems. They typically have higher support infrastructure needs that are best handled by university teams. Increasing local CI capacity, with the support personnel necessary to fully leverage the infrastructure at the campus level, will have the added benefit of creating a local resource for researchers who to date have been “left behind” in the adoption of cyberinfrastructure to advance their research.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

In addition to having the necessary hardware, realizing the full potential of high performance computing (HPC) to positively impact the discovery and innovation in the United States will require that we expand the community of domain specialists and computing users who can take advantage of HPC to support their work. This will mean attracting new graduates from STEM programs who know how to apply HPC. Thus it is important to redouble workforce development efforts on training through university graduate and undergraduate degree programs.

Universities have been successful in establishing degree programs and adding HPC modules into existing courses. However, faculty and students need access to meaningful HPC problems to work on that foster the skills needed by academia, federal agencies and industry. Moreover, graduate students and post-doctoral fellows and other trainees need funded research opportunities to partner with the HPC community in solving the research challenges we face. We believe that a complete approach to addressing the higher education portion of our workforce development mission should also include a variety of strategies, such as internships, scholarships, graduate and post-doctoral fellowships, and funded research in addition to collaborative curriculum development in order to develop a robust university environment that produces the workforce ready to apply HPC to problems of national interest.

Consent Statement

Author Names and Affiliations

- Thomas Furlani, Director, Center for Computational Research, University at Buffalo, SUNY; furlani@buffalo.edu
- Venu Govindaraju, Vice President, Office of Research and Economic Development, University at Buffalo, SUNY
- Liesl Folks, Dean, School of Engineering and Applied Sciences, University at Buffalo, SUNY

Research domains, disciples/subdisciplines of authors

- Furlani: High performance computing, Computational Chemistry
-]Govindaraju: Biometrics, machine learning,
- Folks: Physics, materials informatics

Title

Improving national CI capacity through moderate sized campus systems

Abstract
We propose that NSF improve the nation's CI capacity and capability through increased investment in **moderate sized campus based CI systems**. The arguments in favor of this strategy are several. First, many of today's scientific application codes do not scale well beyond a few hundred cores or the problems being addressed by faculty are not large enough for the application to run at larger node sizes. Second, many of the large national facilities are designed/optimized for the most demanding computational problems with expensive internode networks and filesystems designed for large I/O rates, making jobs comprising tens to hundreds of processors not a cost effective use of these resources. Third, the operating costs (power, cooling, and staffing) of campus systems are typically borne by the host institution thereby maximizing NSF's investment. Finally, campus CI plays a crucial role in the training of a next generation who will lead the creation of the new HPC tools and methodologies.

Question 1 (maximum 1200 words) – Research Challenge(s). ***Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.***

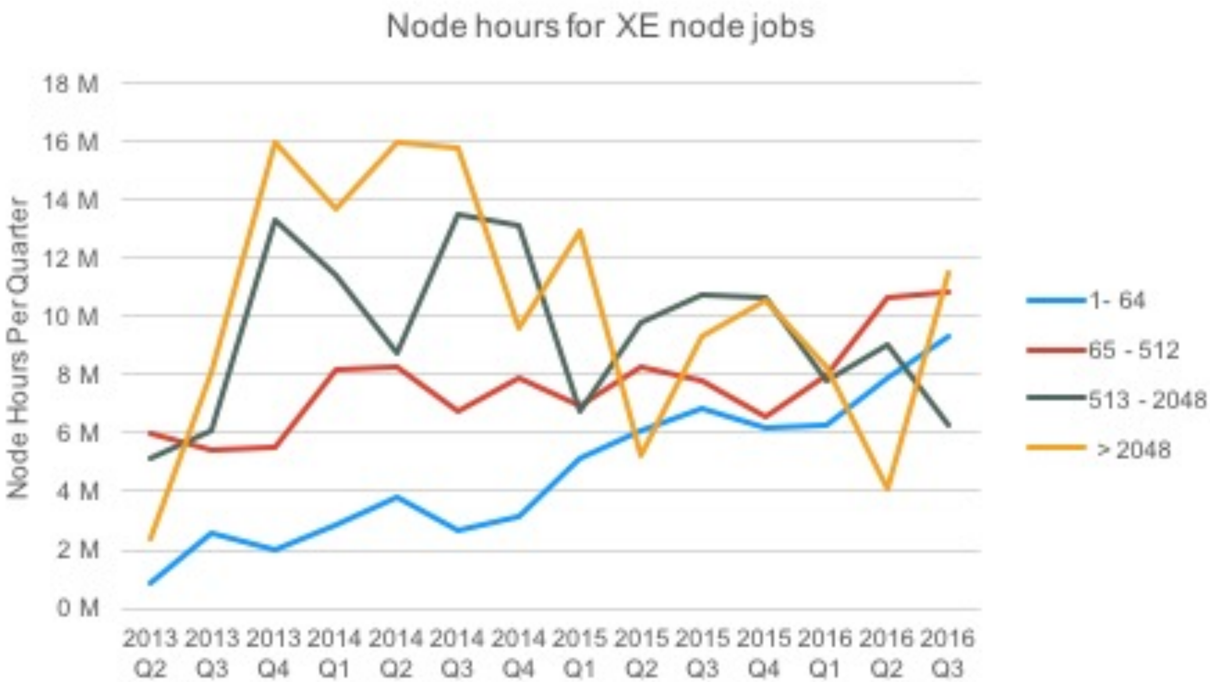
As a result of a decade of targeted investments, the University at Buffalo (UB) today has a remarkably strong assembly of faculty with expertise in a wide range of computational and data intensive areas. In fact, many of these faculty members specifically specialize in computational methods to the extent that our computational facilities are the primary factor that enables their research. These faculty are distributed over a range of disciplines, including biostatistics, chemistry, economics, finance, geology, pharmacy, physics, mathematics, materials, medicine, and the engineering disciplines. In addition, many of UB's high-profile research centers are heavily dependent on advanced computing and data analytics, including the NYS Center of Excellence in Materials Informatics (CMI), the Institute for Research and Education in Energy, Environment, and Water (RENEW), the NYS Center of Excellence in Bioinformatics and Life Sciences, the Institute for Lasers, Photonics, and Biophotonics (ILPB), the Buffalo Institute for Genomics and Data Analytics (BIG), and the Buffalo Clinical and Translational Science (CTSA) center. UB's new innovative PhD program in Computational and Data-Enabled Science and Engineering (CDSE), which was established in response to the need for such specialized training, is another such program. The continued success of these efforts depends critically on our ability to provide adequate computing and data analysis capability.

Further, there is growing interest at UB and elsewhere in the area of Digital Humanities - the name given to a vast field of research projects at the intersection of Humanities disciplines and computing that provide both new archival resources for Humanities scholars and opportunities for asking new kinds of questions. Through its interdisciplinarity, Digital Humanities can transform research in the humanities while offering new conceptual and practical challenges to researchers in the social sciences as well as those in STEM fields, such as computer, physical, and cognitive sciences.

Question 2 (maximum 1200 words) – Cyberinfrastructure Needed to Address the Research Challenge(s). Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

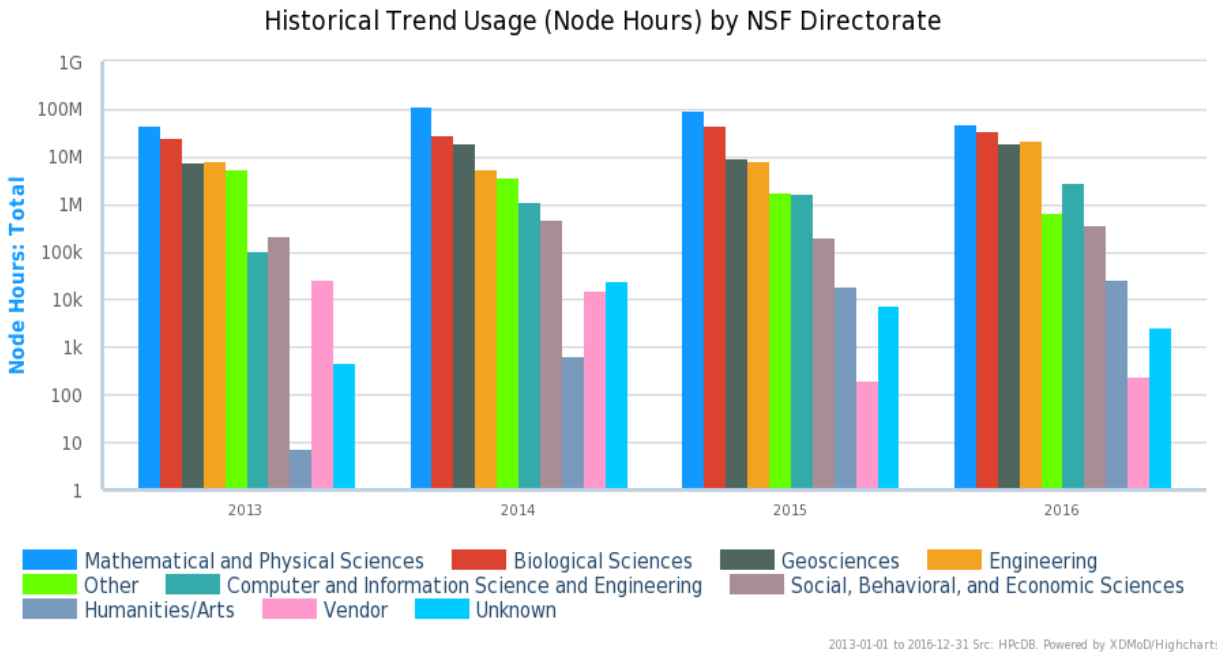
Given the continually increasing demand in advanced computing and data infrastructure to support the university's research and educational initiatives described in question 1 (this is not a situation unique to UB), the university's advanced cyberinfrastructure is unable to keep pace with demand through current external funding opportunities. While NSF has helped significantly by earmarking national resources to meet the needs at the high end, it has not been sufficient to keep pace with campus demand. In addition, much of the computing done at the campus level involves research and applications that are well suited (optimized) for moderate size parallel jobs on the order of tens to hundreds of processors or large memory nodes; jobs that are best handled locally. Indeed, most of UB's CI users run simulations in the range of tens to hundreds of processors (cores).

To support the need for moderate size CI systems, consider the following plot of Blue Waters utilization by job size and note that the node hours delivered for 1 – 64 node jobs are roughly equal to the other job size distributions (since late 2015). While this is partially driven by the allocation process, it speaks to the importance of these "small" jobs to advance research and education. It is our position that jobs in the 1 – 64 node range would best be accommodated on campus level systems, freeing up valuable cycles on the national systems to solve the computationally challenging problems requiring much larger resources than can be provided at the campus level.



By increasing the funding available for campus based systems (either through MRI's or another program), NSF could solve two problems simultaneously, namely freeing up cycles on the large national research for grand challenge class problems, and greatly improving the campus based CI capacity. Another advantage to increased support for campus based systems is that campus MRI's typically don't include personnel or operational support costs, while national scale systems do, and therefore the dollars invested in a campus system have the potential to provide more compute power per dollar spent. Campus HPC facilities also leverage additional (often large) local university investments.

An additional role that we emphasize here in support of improved campus CI is the training of a new generation of computational and data scientists who will lead the creation of the new tools and methodologies. This training requires access to diverse and a broad range of cutting edge hardware locally so students may learn and innovate quickly while using national capability resources for production usage. Continued investment in campus CI (coupled to local campus support, both financial and through multidisciplinary training efforts such as UB's CDSE program) is critical to maintain this pipeline of next generation computational and data scientists. Computational and data intensive activities are continuing to gain in importance for fields not traditionally active in these areas. The following figure shows the breakdown in use of Blue Waters by NSF directorate, showing the increasing usage by fields not historically invested in CI.



The availability and flexibility of local CI resources (both hardware and faculty/staff expertise) is crucial to developing and maintaining outreach activities, particularly for non-traditional areas in the humanities and social sciences (indeed, XSEDE's campus champion outreach activities are largely focused on investing in, and leveraging local campus resources, both personnel and infrastructure, for exactly this goal).

Finally, in many academic settings, faculty in the Humanities and other "non-traditional HPC areas" are poorly connected to and poorly positioned to leverage National CI facilities to help them with their computing problems. They typically have higher support infrastructure needs that are best handled by university teams. Increasing local CI capacity, with the support personnel necessary to fully leverage the infrastructure at the campus level, will have the added benefit of creating a local resource for researchers who to date have been "left behind" in the adoption of cyberinfrastructure to advance their research.

Question 3 (maximum 1200 words, optional) – Other considerations. ***Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.***

In addition to having the necessary hardware, realizing the full potential of high performance computing (HPC) to positively impact the discovery and innovation in the United States will require that we expand the community of domain specialists and computing users who can take advantage of HPC to support their work. This will mean attracting new graduates from

STEM programs who know how to apply HPC. Thus it is important to redouble workforce development efforts on training through university graduate and undergraduate degree programs.

Universities have been successful in establishing degree programs and adding HPC modules into existing courses. However, faculty and students need access to meaningful HPC problems to work on that foster the skills needed by academia, federal agencies and industry. Moreover, graduate students and post-doctoral fellows and other trainees need funded research opportunities to partner with the HPC community in solving the research challenges we face. We believe that a complete approach to addressing the higher education portion of our workforce development mission should also include a variety of strategies, such as internships, scholarships, graduate and post-doctoral fellowships, and funded research in addition to collaborative curriculum development in order to develop a robust university environment that produces the workforce ready to apply HPC to problems of national interest.