Evaluation Report

1st Annual Advancing Research Computing on Campuses (ARCC) Workshop

Prepared by
Lori Tanner
Clemson University

For
Jim Bottum
Clemson University

Martin Biernat
National Center for Supercomputing Applications

Table of Contents

Summary

Abstract:

Advancing Research Computing on Campuses: Best Practices Workshop featured a range of relevant topics in the emerging distributed infrastructure, spanning discussions of best practices for operating and supporting a campus shared research computing infrastructure to integrating business models for advanced research computing resources and services. The workshop successfully increased collaboration between attendees who have since extended their knowledge to their home institutions.

Executive summary

The evolving role of computing and data in scientific discovery has created new challenges for campus IT organizations. Infrastructure—both technical and human—has become increasingly generalized, making it difficult for a campus to effectively support the complex and diverse requirements of today's computer and data intensive research-driven agenda. Compounded by the increasingly competitive nature of access to national resources, this has forced campuses to take on a larger role in supporting the increasing demands for research computing resources. As campus IT organizations build and expand infrastructures components to address the needs of their users, coordination between campuses is critical in order to created successful, sustainable, scalable, and flexible models.

The sustainability models for campus shared research computing resources and services cannot be discussed without addressing the ever-increasing "long tail" of science. That is, researchers who have outgrown their desktops and now need access to advanced computing resources but may not have experience using such resources. Key to addressing these challenges is collaborative discussions with the IT experts and stakeholders across the country. Partnerships among institutions have emerged to build the robust infrastructure needed to meet the advanced computing needs of researchers. With these collaborative endeavors, comes the need to develop and hone best practices.

This report summarizes and evaluates the Advancing Research Computing on Campuses: Best Practices Workshop held on January 16-17, 2014 at the NCSA facility on the Urbana-Champaign campus of the University of Illinois. Participation in the workshop was kept to a small group of approximately 90 attendees from 30 different institutions/organizations to encourage collaborative discussions.

Funding for the workshop came from industry sponsors and a minimal workshop fee (primarily to encourage commitment). Sponsors included:

- Dell
- EMC^{2}®
- Omnibond
- DataDirect Networks™

There was also support from our partner Internet2

Experts in the field shared specific models and practices emerging to support and address the advance computing environment at their respective facilities. Topics discussed included (not exhaustive):

- The evolution of research computing at UC San Diego
- Shared Cluster program UC Los Angeles
- Implementing campus bridging
- Business model for affordable and sustainable archival storage
- High performance computing environment and sustainability
- Service platforms
- The condo model

Researchers also shared different perspectives of clusters and how to integrate with campus infrastructures. The discussions moved from the institutional level to a national and global perspective of advanced computing, to making sound business and policy decisions.

Participants were surveyed as to their satisfaction with each of the events/presentations. Responses suggest that the workshop was successful in meeting its primary objective to bring together professionals in the community and discuss ways to leverage the summation of experience and expertise for the overall community's benefit. There was support for continuing the workshop in the future as well as suggestions for additional topics for discussion. Overall, the workshop was informative and invited discussions and collaborative exchanges with the attendees. However, a more diverse participant group should be encouraged to attend/present for future workshops (i.e. women and underrepresented groups).

Background

Needs Addressed

The 1st Annual Advancing Research Computing on Campuses Workshop was an effort to bring together professionals in the field of advanced research computing to discuss best practices for operating and supporting a campus shared research computing infrastructure. The evolving role of computing and data in scientific discovery has created new challenges for campus IT organizations.

Infrastructure has become increasingly generalized, making it difficult for a campus to support effectively the complex and diverse requirements of contemporary computing and data intensive research. Compounded by the increasingly competitive nature of access to national resources, campuses have been forced to take on a larger role in supporting the increasing demands for research computing resources. As campus IT organizations build and expand infrastructure components to address the needs of their users, coordination between campuses is critical in order to create successful, sustainable, scalable, and flexible models.

Literature

There have been great strides in high performance computing processor performance that began 1980s, (Graham 2005; Mills 2008). Especially now, the demand for high performance computing remains high, particularly for national defense/security and also biotechnology, weather pattern studies, and other areas of research benefiting from high- tech computer model simulation (Bernard 2001; Graham 2005; Helbing 2010; Kettinger 2001; Pezzoli 2007). As such, the National Science Foundation's (NSF) report to the federal government in 2001 emphasized the need to allocate a greater pool of money towards the creation and development of more high performance computing facilities (Graham 2005). The NSF recommended that this money be spread among university institutions for the purposes of high performance computing development and research, and recommended policies that would open university doors as much as possible to researchers in computer science and information technology fields (Graham 2005; Helbing 2010).

Research in academic fields such as climate change and biotechnology are also cited as motivations (Graham 2005). As the U.S. and nations across the globe continue to advance their information super highways, improvements with respect to the speed of information transfer for text, video, images, etc. will not necessarily be the central focus of technological advancement (Helbing 2010; Bernard 2001). Instead, some researchers assert that, with each major advance in communication has also come the advancement and evolution of society's conception of knowledge, its relationship to information, and its use of facts (Bernard 2001). Researchers assert, technical capabilities have exceeded an ability to obtain information and complete basic calculations; currently, supercomputers can process and depict

interrelations between data from a variety of sources and interdisciplinary fields (Bernard 2001; Graham 2005; Helbing 2010; Kettinger 2001; Pezzoli 2007). The advancement of technology to depict highly conditional situations has great import for issues such as climate change, the international economy, the advent of 'virtual' political borders and revolutions, health epidemics, and other areas in which local realities are greatly influenced by a complex web of global dynamics and relationships (Bernard 2001). Envisioning future living with high performance computing technology, researchers foresee a 'smartly' functioning society with better-organized economies, well-planned cities, and efficient systems (Helbing 2010).

More specifically, this translates to a society of highly customized information services; personalized education in which individuals and collective groups are closely linked to libraries, archives, and other sources of knowledge; well-ordered cities with smart traffic planning and logistics; 'smart' energy production and consumption in which generation, delivery, grid structure and supply/demand are better matched; and greater capacity to anticipate and provide efficient solutions in the face of shifting migration patterns, criminal networks, and revolutionary uprisings (Helbing 2010).

While supercomputers appear to be humanity's saving grace in the face of issues such as climate change, the facilities and equipment required involve heavy reliance on energy sources. Operating 365 days, or 8,760 hours, per year, high performance computing centers must be able to generate several megawatts of power at any given time, with obvious implications for carbon dioxide emissions (Mills 2008; Singer 2011). This has both environmental and economic implications. As a result, researchers and political interests are now stressing best practices for computing center design through efficient cooling methods in high performance computing centers and efficient design of IT equipment within these centers (Graham 2005; Mills 2008). Key points include efficient approaches to cooling facilities (e.g. spray-cooling technology and 'free cooling' methods that take advantage of cool air from outside the facility); utilization of parallelism and multi-core processors (i.e. using simpler, less powerful processors to achieve performance); leveraging a thriving consumer electronics industry which capitalizes on low-power embedded processor technology, enabling high performance computing facilities to "consume a tiny fraction of the power of existing computing centers without unduly compromising performance"; and locating closer to less expensive renewable energy sources (Graham 2005; Mills 2008). Some researchers are also pushing for the development of energy benchmarks in order to gauge energy efficiency in high performance computing centers (Mills 2008).

It is generally agreed upon the high performance computing cyberinfrastructure currently promotes and will continue to build upon society's knowledge-base (CIF21 2012); to affect our understanding of reality in the context of globalization; and will be a tool in seeking solutions and finding connections in regard to dilemmas such as climate change, security, health biotechnology, traffic bottleneck, city planning, etc. While most literature is in agreement on the importance of high performance computing technology and its integral role in

society, a core difference emerges in terms of each researcher or author's specific concerns with the technology.

Cyberinfrastructure science includes not only high performance computing systems, but also data storage systems, repositories, advanced instruments, and visualization systems that enable researchers to gain new insights and investigate broader and more complex research challenges (CIF21 2012)

With government investment clearly devoted to the development of high performance computing facilities and the research produced by these centers, it is clear that university institutions will most likely continue to receive federal grants to pursue research involving the use of supercomputers. Having key players in the high performance computing realm come together and share case studies of their individual centers would be of value to the field and allow for documentation of any socioeconomic changes; challenges; new ideas and/or issues that arise. "Fostering innovation requires cyberinfrastructure spanning multiple levels (national, community, campus)…[it must] be managed efficiently and equitably, particularly in the allocation of resources…and with well-established mechanism to ensure sustainability," (CIF21 2012 p. 3)

References

Bernard, Kathleen C. "Ordering Chaos: Supercomputing at the Edge." *Technology 2001: The Future of Computing and Communications*. 6th ed. Cambridge: Massachusetts Institute of Technology, 1994. 71-94.

Chandler, David. "MIT Teams Win U.S. Department of Defense Grants."*MITnews*. Massachusetts Institute of Technology, 22 May 2009.

Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21). May 2012. Web. http://www.nsf.gov/cise/aci/cif21/CIF21Vision2012current.pdf

Graham, Susan L., Marc Snir, and Cynthia A. Patterson. *Getting up to Speed the Future of Supercomputing*. Washington, DC: National Academies, 2005.

Helbing, Dirk. "The FuturICT Knowledge Accelerator: Unleashing the Power of Information for a Sustainable Future by Dirk Helbing :: SSRN." *The FuturICT Knowledge Accelerator: Unleashing the Power of Information for a Sustainable Future* 10 (2010): 1-21. *Social Science Research Network*. Swiss Federal Institute of Technology.

Kettinger, William. "National Infrastructure Diffusion and the U.S. Information Super Highway." *Computers and Security* 20 (2001): 43-52. *ScienceDirect*. SciVerse.

Mills, Evan, William Tschudi, John Shalf, and Horst Simon. "Supercomputers: Super-polluters?" *Data Center Journal* 7 (2008): 16 18. *Http://hightech.lbl.gov/documents/DATA_CENTERS/DCJ-V7-16-18.pdf*. Laurence Berkeley National Laboratory.

Pezzoli, Keith, Robert Tukey, Hiram Sarabia, Ilya Zaslavsky, Marie Lynn Miranda, William A. Suk, Abel Lin, and Mark Ellisman. "The NIEHS Environmental Health Sciences Data Resource Portal: Placing Advanced Technologies in Service to Vulnerable Communities." *PubMed Central* 15 (2007): 564-71. *Environmental Health Perspectives*.

Singer, Stephen. "Mass. Schools Team up for Supercomputer Center." *Stock Market Opinion and Analysis, Breaking News, Stocks Quotes, Earnings Estimates*. Oct.-Nov. 2011.

Stakeholders

Institutions

Brandeis University
Brown University
Clemson University
Cornel
Georgia Tech
Harvard University
Indiana University
Kansas State University
Marshall University
Northwestern University
Ohio State University
Purdue
Rice University
Stanford University
Temple University
Texas A&M University
The Pennsylvania State University
The University of Iowa
University of Arkansas
University of California-LA
University of California-SD
University of Chicago
University of Connecticut
University of Florida
University of Hawaii
University of Illinois
University of Kansas
University of Kentucky
University of Notre Dame
University of Oklahoma
University of Pittsburgh
University of Southern California
University of Toronto
University of Utah
University of Wisconsin
West Virginia University
Wharton
Yale University

Organizations

DDN
Fermilab
International Science Grid This Week
MGHTCC
NCSA
National Science Foundation
Open Science Grid
XSEDE

Sponsors

Dell
EMC^{2}®
Omnibond
DataDirect Networks™

Partners

Internet2

Participants

The target audience for the workshop included campus administrators supporting (or needing to support) research computing resources and services and academic research computing professionals. There were 81 participants representing 50 different facilities, organizations, industry, and institutions.

Objectives

The objective of the workshop is to bring together professionals in the community and discuss ways to leverage the summation of experience and expertise for the overall community's benefit.

Activities

January 15, 2014

Registration/informal Reception

January 16, 2014

- Welcome: John Towns, and James Bottum
- Opening: Ed Seidel
- Lightening Talks: Moderator—Chuck Thompson
  - Ron Hawkins
  - Richard Moore
  - Tajendra Vir Singh
  - John Towns
  - Richard Knepper
  - Henry Neeman
  - Daniel Andresen
  - Rob Gardner
  - Andrew Sherman
  - Kiran Keshav
  - Jeff Pummill
  - Jarek Nabrzyski
  - Paul Brenner
- Managing Programs of Condo-style Resources: Moderator—Jim Pepin
  - Preston Smith
  - Ruth Marinshaw
  - Julia Harrison
  - Maureen Dougherty
- Dell HPC Solutions: Furnished Condos
  - Mark R Fernandez
- Researcher Perspectives: Of Condo of Condos/Clusters: Moderator—Barr von Oehsen
  - Alex Feltus
  - Mark Neubauer
  - Brian Haymore
  - Julia Harrison
- Integrating with Campus Infrastructures: Moderator—Ruth Pordes
  - Tracy Smith
  - Guy Almes
  - Dan Schmiedt
  - Wendy Huntoon

January 17, 2014

- Regional/National Consortia: Moderator—John Towns
  - Eric Deumens
  - James Cuff
  - James Bottum
  - Miron Livny
- NSF Perspective
  - Irene Qualters
- National Perspectives: Moderator—James Cuff
  - James Cuff
  - Bill Kramer
  - Chris Loken
  - Ruth Pordes
  - John Towns
- Business Practice and Policy: Moderator—David Richardson
  - James Bottum
  - John Goodhue

    - Carie Lee Kennedy
- Blue Waters Tour

Location

NCSA on the Urbana-Champaign campus of the University of Illinois
1205 West Clark Street,
Urbana, IL

INSERT BEST PRACTICE NOTES HERE

Budget

NOT SURE IF THIS IS ACCURATE

A/V Staff	$2,600.00	2 Days @ $1300 for Auditorium (holds 192 people)
Office Supplies	$0.00	JT paying
Wednesday Night Reception - Food/Drinks	$3,813.75	$44.87
Thursday Breakfast	$1,097.50	$12.91
Thursday AM Break	$882.50	$10.38
Thursday Lunch	$1,360.00	$16.00
Thursday PM Break	$572.25	$6.73
Thursday Dinner	$2,405.00	$28.29
Friday Breakfast	$1,097.50	$12.91
Friday AM Break	$858.60	$10.10
Friday Lunch	$1,360.00	$16.00
Transportation to NPCF	$0.00	JT Paying
Total	$16,047.10	
INCOME	**AMOUNT**	**DESCRIPTION**
Registration Fee	$3,350.00	$50 registration fee/Qty. 67 registrants
Sponsors	$4,000.00	DELL
	$1,000.00	Omnibond
	$1,000.00	DDN
	$2,500.00	Internet2 (paying Clemson?)
	$2,500.00	EMC
Total	$14,350.00	
BALANCE (INCOME - EXPENSES)	-$1,697.10	JT Paying

Expected Outcomes

Trends in approaches will be examined in order to look at the changing landscape and how it is transforming campus interactions and sustainability models. In discussing best practices, it is expected the workshop will help increase collaboration between attendees, eventually extending to their current and future bases. New products and services were announced and showcased. Attendees were able to network and recruit graduate students and PhD candidates

Constraints

The size of the facility was the major constraint. The planning committee chose to keep the number of attendees minimal so as to foster dialog and networking. The facility could accommodate approximately 200 guests. The aim was to have close to 100 attend inclusive of presenters.

Evaluation

Procedures

At the conclusion of the workshop a sixteen-question survey was given to the participants asking for feedback on the success of the event. The questions were designed for primarily Likert-scale type responses, with a few open-ended for suggestions. Questions ranged from general, overall satisfaction with the workshop, to a level of satisfaction with a particular sponsor's presentation.

Sample

The demographics of the workshop attendees included:)

- 20 Women
- 61 Men
- #? IT executive leaders
- #? Researchers/Scientists
- #? Industry representatives
- with #? to #? years of experience in advance computing

Of the 81 attendees (inclusive of presenters) 45 completed the survey. Not all 45 completed all sixteen of the survey questions.

Methodology

The survey was a paper version. Only attendees at the closing event were given the survey. The surveys were completed anonymously. Completed surveys were collected as attendees exited the final workshop presentation.
Survey Questions:

1. How satisfied are you with the event?
2. Was the duration of the event: too long, just right, too short?
3. How satisfied are you with the Lightning Talks?
4. How satisfied are you with the "Managing Programs of Condo-style Resources" session?
5. How satisfied are you with the "Dell HPC Solutions—"Furnished" Condos: sponsor talk session?
6. How satisfied are you with the "Researchers Perspectives: of Condo of Condos/Clusters" session?
7. How satisfied are you with the "Integrating with Campus Infrastructures" session?
8. How satisfied are you with the "Regional/National Consortia" session?
9. How satisfied are you with the "NSF Perspective" session?
10. How satisfied are you with the "National Perspectives" session?
11. How satisfied are you with the "Business Practice and Policy" session?
12. What topics would you like covered next year?
13. What are the strengths of the ARCC Workshop?
14. What improvements would you suggest for the ARCC Workshop?
15. Will you be attending the 2nd Annual ARCC Workshop in February?
16. Please provide any additional comments

Findings

Overall those completing the survey were either *very satisfied* or *somewhat satisfied* with the workshop and most of the events. This was also true for the length of the event. This suggests that overall the best practices discussions and presentations met the objective of the workshop to bring together professionals in the community and discuss ways to leverage the summation of experience and expertise for the overall community's benefit. However, the majority of the responses were less than satisfied, *neutral* or even *somewhat dissatisfied* with the sponsor presentation by the Dell representative. This can suggest that the presentation was not following the expected agenda and/or that the presentation was a product promotion rather than best practices strategy.

Questions 12-14 encouraged suggestions for future workshops. Frequently there were responses with the suggestion of more lightning talks. This format seemed to be popular and engaging venue. However, participants requested more discussion time after the lightning talks

> *Favorite part was the lightning talks. It's great to hear so many examples of ways to do things that may apply to my situation*
>
> *More lightning talks, less long presentations, some time for questions/discussions after lighting talks*
>
> *The lightning talks were very efficient. I thought that in most cases the lightning talks had as much content as the longer talks.*
>
> *I would like to see some opportunity for discuss or questions in response to lightning talks*
>
> *Spread lightning talks throughout the days as a break from panel discussions*
>
> *Maybe more lightning talks with more time for discussion with them*

Participants also wanted more information about data management, cloud computing and security.

Surprisingly, 49% of the respondents would not commit to attending a 2nd Annual ARCC Workshop in February (questions #15). The question was not designed to encourage explanation as to why; rather it was a Likert-type response with choices of *yes, no, maybe*. Perhaps the time of year is the concern.

Conclusions

With the exception of the Dell presentation, the workshop met the objectives and outcomes to help increase collaboration between attendees, eventually extending to their current and future bases. This being the overall consensus, there could have been more time designated for networking and small group discussion. Without more open-ended responses with each survey question, it is difficult to do more than speculate as to the reason for some of the outliner responses.

Also problematic is that the survey was only distributed at the end of the last session on the 2nd day. To encourage a higher response rate, a short 3 or 4 question survey after each session with open-ended responses for clarification could give more insight.

Of particular importance would be the impact the workshop has had on individual attendees several months after the event. A follow-up survey or better yet, a personal interview with a random sample would allow for measuring the added value of the workshop and prove helpful for planning future events.

Recommendations

Successes

The workshop drew a national representation of Advanced Computing leaders, experts, researchers, and scientists to collaborate on best practices. Accommodations were comfortable and convenient.

Challenges

Time of year the event was held seemed to be the most difficult challenge to overcome. January is difficult for planning personal schedules, making travel arrangements, and university semester changes. Concerted effort went into recruiting attendance and securing presenters.

The planning stages were difficult in getting together via teleconferencing. High profile leaders have already busy schedules so coordination of meetings in different time zones was challenging.

Minor issues arose with the collection of presentation materials and uniformity in presentation slides. It was unclear who the point of contact was for uploading presenter artifacts.

Modifications

- Collect data
    - Demographics on survey-data analysis
    - Information about attendees: years experience, gender, affiliation, underrepresented groups, role in IT,
    - Head count in each event—compare to survey results
    - Follow-up with attendees—implementation of best practices
    - Give out survey after each event
    - Incentive to complete survey
- More diverse attendees/presenters-underrepresented groups
- Broader audience to include students
- Use social media to advertise before, during, and after
- Different time of the year for 2nd event.