Results of 2013 Survey of Parallel Computing Needs Focusing on NSF-funded Researchers

Craig A. Stewart
Andrew Arenson
Jeremy Fischer
Matthew R. Link
Scott A. Michael
Julie Wernert

Indiana University

PTI Technical Report PTI-TR15-04A
15 May 2015

Please cite as:

Stewart, C.A., Arenson, A., Fischer, J., Link, M.R., Michael, S.A., & Wernert, J. (2015). *Results of 2013 Survey of Parallel Computing Needs Focusing on NSF-funded Researchers*. PTI Technical Report PTI-TR15-04A. Indiana University, Bloomington, IN. http://hdl.handle.net/2022/19906

This report is based upon work supported in part by the National Science Foundation under Award Number 1205518 to Indiana University. This report is also supported by the Indiana University Pervasive Technology Institute. IUPTI has been supported in part by a major grant award from the Lilly Endowment, Inc. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or other funding agencies.

Table of Contents

Table of Tables

1. Introduction

The field of supercomputing is experiencing a rapid change in system structure, programming models, and software environments in response to advances in application requirements and in underlying enabling technologies. Traditional parallel programming approaches have relied on static resource allocation and task scheduling through programming interfaces such as MPI and OpenMP. These methods are reaching their efficiency and scalability limits on the emerging classes of systems, spurring the creation of innovative and dynamic strategies and software tools, including advanced runtime system software and programming interfaces that use them. To accelerate adoption of these next-generation methods, Indiana University is investigating the creation of a single supported Reconfigurable Execution Framework Testbed (REFT) to be used by parallel application algorithm developers as well as researchers with advanced tools for parallel computing. These investigations are funded by National Science Foundation Award Number 1205518 to Indiana University with Thomas Sterling as Principal Investigator, and Maciej Brodowicz, Matthew R. Link, Andrew Lumsdaine, and Craig Stewart as Co-Principal Investigators.

As a starting point in this research, we proposed to assess needs in parallel computing, in general, and needs for software tools and test beds, in particular, within the NSF-funded research community. As one set of data toward understanding these needs, we conducted a survey of researchers funded by the National Science Foundation. Because of the strong possibility of distinct needs of researchers funded by what is now the Division of Advanced Cyberinfrastructure, researchers funded by the other divisions of the Computer and Information Sciences and Engineering (CISE) Directorate, and researchers funded by the remainder of the NSF, we surveyed these populations separately.

We report here the methods and summarize the results we obtained in this survey.

2. Materials and Methods

Appendix 1 contains a copy of the survey as administered online. The survey instrument and associated recruitment materials were submitted to the Indiana University Human Subjects Office to ensure compliance with university guidelines and federal regulations, including that all principal investigators and personnel had completed human subjects research training and certification through the Collaborative Institutional Training Initiative (CITI). After some suggested clarifications and refinements, the survey, under protocol 1205008730, was approved as “exempt” research on May 22, 2012, as documented in Appendix 2 and Appendix 3.

The names and email addresses representing the populations from which random samples were drawn from the National Science Foundation’s (NSF) database of awards utilizing the criteria on the Awards Advanced Search page (http://www.nsf.gov/awardsearch/advancedSearch.jsp).

For the period 1 September 2008 through 31 December 2012 there were a total of:

- 508 unique PIs funded by the CISE Division of Advanced Cyberinfrastructure (ACI) or its predecessor, the Office of Cyberinfrastructure (OCI)
- 2,359 unique PIs funded by CISE Divisions other than ACI
- 2,489 unique PIs funded by NSF Divisions other than CISE.

Our belief regarding PIs was that those who had been funded by ACI/OCI would likely have the most specific needs relative to computing testbeds, that researchers funded by CISE and not funded by ACI or OCI would have the next most specific needs, and that researchers funded by NSF Divisions other than CISE would overall have the most general needs. This hierarchy influenced the creation of sample populations.
The MS(r) Excel function RAND() was used to select 500 of the 508 PIs funded by ACI/OCI. We then

selected a random sample of 2,000 of the 2,359 PIs funded by a CISE Division other than ACI (or its predecessor OCI). There were some individuals who had been funded both by ACI/OCI and also funded by other Divisions of CISE. We eliminated duplicates by removing such individuals from the list of potential CISE invitees, leaving 1,948 individuals asked to take the survey and categorized as "CISE" respondents. We went through the same process with the 2,489 individuals who were funded by NSF Divisions other than CISE, removing duplicates of people who had been funded by NSF Directorates other than CISE from the "NSF" sample. This process left us with:

- OCI lists – 500 individuals invited to take the survey
- CISE list –1,948 invitees
- NSF list – 1,813 invitees

That totals 4,261 individuals invited to participate in the survey.

The survey was open from March 25, 2013, to April 24, 2013. Three reminders were sent at intervals of two weeks remaining, one week remaining, and one day remaining.

Table 1. Response rates to questions about access to current resources.

Cohort	Number of invitees	Number of respondents	Percent participation rate
NSF OCI awardees	500	104	20.8%
Other CISE awardees	1,948	195	10.1%
Other NSF awardees	1,813	183	10.1%

3. Summary of Results

The results of the summary are shown here in tabular form. Text responses are shown with identifying references removed. No other changes were made in text responses; e.g. typographical errors are shown as they were entered, as well as criticisms of the survey.

The data sets and copies of SPSS descriptive statistics describing the data are available online at http://hdl.handle.net/2022/19924, so anyone who wants to pursue further analyses may do so.

Table 2. Responses to questions about access to current resources.

Access to current resources	**Yes - all of the time**	**Most of the time**	**Some of the time**	**No - Never**	**Total responses**
Do you currently have access to sufficient cyberinfrastructure facilities (computation, data storage and management, visualization, software tools) to address your current research needs?					
NSF OCI awardees					
Frequencies	26	51	22	5	104
Percentages	24.0%	49.0%	21.2%	4.8%	
Other CISE awardees					
Frequencies	55	92	45	3	195
Percentages	28.2%	47.2%	23.1%	1.5%	
Other NSF awardees					
Frequencies	63	79	36	4	182
Percentages	34.6%	43.4%	19.8%	2.2%	
Considering just computational resources, do you have access to sufficient computational resources to address your current research needs?					
NSF OCI awardees					
Frequencies	37	44	15	6	102
Percentages	36.3%	43.1%	14.7%	5.9%	
Other CISE awardees					
Frequencies	73	92	29	3	197
Percentages	37.1%	46.7%	14.7%	1.5%	
Other NSF awardees					
Frequencies	73	70	36	3	
Percentages	38.8%	37.2%	19.1%	1.6%	182

Access to current resources	Yes - all of the time	Most of the time	Some of the time	No - Never	Total responses
Considering just data storage and management resources, do you have access to sufficient computational resources to address your current research needs?					
NSF OCI awardees					
Frequencies	37	40	23	2	102
Percentages	36.6%	39.2%	22.5%	2.0%	
Other CISE awardees					
Frequencies	78	63	47	6	194
Percentages	40.2%	32.5%	24.2%	3.1%	
Other NSF awardees					
Frequencies	76	63	27	3	179
Percentages	40.4%	33.5%	20.7%	1.7%	
Considering just visualization resources, do you have access to sufficient computational resources to address your current research needs?					
NSF OCI awardees					
Frequencies	36	37	21	8	102
Percentages	35.3%	36.3%	20.6%	7.8%	
Other CISE awardees					
Frequencies	69	64	46	13	192
Percentages	35.9%	33.3%	24.0%	6.8%	
Other NSF awardees					
Frequencies	62	64	42	10	178
Percentages	34.8%	36.0%	23.6%	5.6%	
Mean (+ 95% Confidence Intervals)					
Considering just software tools for your research, do you have access to sufficient computational resources to address your current research needs?					
NSF OCI awardees					
Frequencies	15	39	42	7	103
Percentages	14.6%	37.9%	40.8%	6.8%	
Other CISE awardees					
Frequencies	56	88	46	5	195
Percentages	28.7%	45.1%	23.6%	2.6%	
Other NSF awardees					
Frequencies	48	79	51	4	
Percentages	26.4%	43.4%	28.0%	2.2%	

Table 3. Responses to questions about expected availability of future resources.

Expected availability of future resources	Highly confident (> 90% likelihood in my estimation)	Confident (> 50% likelihood in my estimation)	Not confident (< 50% likelihood in my estimation)	Highly pessimistic (< 10% likelihood in my estimation)	Total responses
Looking five years ahead, do you expect that the growth in availability of cyberinfrastructure facilities (computation, data storage and management, visualization, software tools) will be sufficient to address your research needs at that time?					
NSF OCI awardees					
Frequencies	15	39	42	7	103
Percentages	14.6%	37.9%	40.8%	6.8%	
Other CISE awardees					
Frequencies	54	77	45	7	183
Percentages	29.5%	42.1%	24.6%	3.8%	
Other NSF awardees					
Frequencies	53	74	35	5	167
Percentages	31.7%	44.3%	21.0%	3.0%	
Mean (+ 95% Confidence Intervals)					
Looking five years ahead, do you expect that the growth in availability of computational resources will be sufficient to address your research needs at that time?					
NSF OCI awardees					
Frequencies	31	38	26	3	98
Percentages	31.6%	38.8%	26.5%	3.1%	
Other CISE awardees					
Frequencies	64	75	39	4	182
Percentages	35.2%	41.2%	21.4%	2.2%	
Other NSF awardees					
Frequencies	61	69	34	2	166
Percentages	36.7%	41.6%	20.5%	1.2%	

Expected availability of future resources	Highly confident (> 90% likelihood in my estimation)	Confident (> 50% likelihood in my estimation)	Not confident (< 50% likelihood in my estimation)	Highly pessimistic (< 10% likelihood in my estimation)	Total responses
Looking five years ahead, do you expect that the growth in availability of data storage and management resources will be sufficient to address your research needs at that time?					
NSF OCI awardees					
Frequencies	29	34	33	3	99
Percentages	29.3%	34.3%	33.3%	3.0%	
Other CISE awardees					
Frequencies	64	75	39	4	182
Percentages	25.2%	41.2%	21.4%	2.2%	
Other NSF awardees					
Frequencies	63	69	28	5	165
Percentages	38.2%	41.8%	17.0%	3.0%	
Looking five years ahead, do you expect that the growth in availability and functionality and capability of visualization resources will be sufficient to address your research needs at that time?					
NSF OCI awardees					
Frequencies	21	46	28	3	98
Percentages	21.4%	46.9%	28.6%	2.8%	
Other CISE awardees					
Frequencies	46	81	44	8	179
Percentages	25.7%	45.3%	24.6%	4.5%	
Other NSF awardees					
Frequencies	60	65	35	4	164
Percentages	36.6%	39.6%	21.3%	2.4%	

Expected availability of future resources	**Highly confident (> 90% likelihood in my estimation)**	**Confident (> 50% likelihood in my estimation)**	**Not confident (< 50% likelihood in my estimation)**	**Highly pessimistic (< 10% likelihood in my estimation)**	**Total responses**
Looking five years ahead, do you expect that the growth in quality and functionality of software tools for your research will be sufficient to address your research needs at that time?					
NSF OCI awardees					
Frequencies	16	33	37	13	99
Percentages	16.2%	33.3%	37.4%	13.1%	
Other CISE awardees					
Frequencies	41	87	48	7	183
Percentages	22.4%	47.5%	26.2%	3.8%	
Other NSF awardees					
Frequencies	48	78	36	5	167
Percentages	28.7%	46.7%	21.6%	3.0%	

Table 4. Responses to Yes/No question about whether or not respondents use parallel or high throughput computing.

Software Tools	**Yes**	**No**	**Total responses**
Do you use parallel or high throughput computing in your research?			
NSF OCI awardees			
Frequencies	22	80	102
Percentages	21.6%	78.4%	
Other CISE awardees			
Frequencies	94	91	185
Percentages	50.8%	49.2%	
Other NSF awardees			
Frequencies	55	112	167
Percentages	32.9%	67.1%	

Table 5. Responses to questions about use of parallel and high throughput computing tools. Mean values are reported with a 95% confidence interval.

Expected availability of future resources	**% of respondents**						**Total responses (count)**	**Mean values**
	1	**2**	**3**	**4**	**5**	**n/a**		
[Logic: this question presented only to those who answered "Yes" to use of parallel or high throughput computing tools] Please rate the importance of the following parallel tools in your current research (from 1 = Not at all Important to 5 = Very Important)								
NSF OCI awardees								
MPI (Message Passing Interface)	1.3	1.3	0	11.8	80.3	5.3	76	4.78±0.15
CUDA or other library for GPUs	10.5	17.1	23.7	11.8	23.7	13.2	76	3.24±0.24
OpenMP	10.7	8.9	13.3	33.3	29.3	5.4	75	3.66±0.31
Cloud/Data Parallel (e.g. Hadoop)	40.5	10.8	18.9	8.1	10.8	10.8	74	2.30±0.36
Condor (High Throughput Computing)	43.2	13.5	13.5	8.1	5.4	16.2	74	2.03±0.33
BOINC (Berkeley Open Infrastructure for Network Computing – High Throughput Computing)	50.0	14.9	9.5	1.4	0	24.3	74	1.50±0.21
I'm not sure – I use programs provided by someone else	12.0	2.0	6.0	2.0	2.0	76.0	50	2.17±0.89
Other	4.8	0.0	2.4	0.0	19.0	29.2	42	
Other CISE awardees								
MPI (Message Passing Interface)	25.5	9.9	5.5	17.6	24.2	17.6	91	3.07±0.38
CUDA or other library for GPUs	19.8	15.4	11.0	17.6	20.9	15.4	91	3.05±0.35
OpenMP	22.5	12.4	13.5	14.6	19.1	18.0	89	2.95±0.37
Cloud/Data Parallel (e.g. Hadoop)	15.6	10.0	17.8	24.4	23.3	8.9	90	3.33±0.31
Condor (High Throughput Computing)	23.3	11.6	17.4	9.3	17.4	20.9	86	2.82±0.37
BOINC (Berkeley Open Infrastructure for Network Computing – High Throughput Computing)	34.1	13.6	12.5	3.4	5.7	10.7	88	0.33
I'm not sure – I use programs provided by someone else	14.1	2.8	14.1	4.2	11.3	53.5	71	2.03±0.55
Other								
Other NSF awardees								
MPI (Message Passing Interface)	4.4	0.0	13.3	15.6	45.7	20.0	45	4.25±0.37
CUDA or other library for GPUs	15.2	6.5	10.9	15.2	21.7	30.4	46	3.31±0.56
OpenMP	6.8	0.0	15.9	20.5	22.7	34.1	44	3.79±0.47
	1	**2**	**3**	**4**	**5**	**n/a**		

Expected availability of future resources	**% of respondents**						**Total responses (count)**	**Mean values**
Cloud/Data Parallel (e.g. Hadoop)	25.5	12.8	10.6	10.6	12.8	27.7	47	2.62
Condor (High Throughput Computing)	15.6	11.1	6.7	8.9	13.3	44.4	45	2.88
BOINC (Berkeley Open Infrastructure for Network Computing – High Throughput Computing)	26.7	11.1	8.9	2.2	4.4	46.7	45	2.00
I'm not sure – I use programs provided by someone else	6.7	4.4	8.9	0.0	17.8	62.2		3.47
Other								

Table 6. Open text field responses for "other" in question on parallel software tools

Please specify what other parallel or high throughput computing software tools you use
NSF OCI awardees
variety of libraries
scientific workflows
Pthreads and TBB; low-level RDMA e.g. DMAPP, PAMI, (Open)SHMEM
pthreads
parallel databases
Matlab
Intel's Threading Building Blocks
Eclipse Parallel Tools Platform
Crystallographic data (Crystallographic Open Database and others), USPEX, homegrown scripts/code
Amazon EC2
Other CISE awardees
VisIt and paraview
Storm, DDS
slurm
shared memory in multi-core, lots of it
sesc
SELFE Hydrodynamics code
Seattle testbed
Pthreads, HPX
Our own distributed computing software
large shared-memory systems programming
Java concurrency package and homebrew software
IBM Cluster - Cloud (like Amazon Service)
I just need many jobs managed by something - doesn't matter what.
I do parallel simulations - this is an embarrassingly parallel application
Custom
Currently I only use the distributed algorithm designed by myself
Cilk, Java, bare pthreads
Cilk Plus
Both open-source and in-house computational genomics tools
Architecture and Memory simulation tools such as SiMICS
Other NSF awardees
Sun Grid Engine
Self developed / FFTW
multiple serial jobs for "embarrassingly parallel" problems
CIPRES Web Portal
CIPRES

Table 7. Responses to questions about parallel computing facilities used. Mean values are reported with a 95% confidence interval.

Facilities	% of respondents						Total responses	Mean	+/-95% CI
	1	2	3	4	5	n/a			
Please rate the importance of the following facilities in your current research (from 1 = Not at all Important to 5 = Very Important) [Logic: this question presented only to those who answered "Yes" to use of parallel or high throughput computing tools]									
NSF OCI awardees									
Computing systems in my own lab or dept.	3.8	9.0	10.3	20.5	52.6	2.8	78	4.13±0.27	
Computing systems campus (or at my own institution in multi-campus research institutions)	9.1	9.1	11.7	16.9	48.1	5.2	77	3.90±0.33	
State or regional campus facilities (e.g. SURAGrid, DiaGrid)	40.8	11.8	13.2	13.2	11.8	9.2	77	2.38±0.36	
XSEDE – the eXtreme Science and Engineering Discovery Environment	18.2	7.8	13.0	19.5	37.7	3.9	77	3.53±0.35	
DOE supercomputer systems or services - e.g. DOE INCITE program	32.9	3.9	10.5	10.5	36.8	5.3	76	3.15±0.42	
DOD supercomputers or facilities	52.9	6.6	6.6	5.3	15.8	13.2	76	2.14±0.39	
Other									
Other CISE awardees									
Computing systems in my own lab or dept.	3.3	1.1	4.4	11.0	80.2	0.0	91	4.64±0.18	
Computing systems campus (or at my own institution in multi-campus research institutions)	10.2	8.0	12.5	29.5	34.1	5.7	88	3.73±0.29	
State or regional campus facilities (e.g. SURAGrid, DiaGrid)	40.4	14.6	19.1	11.2	5.6	9.0	89	2.20±0.28	
XSEDE – the eXtreme Science and Engineering Discovery Environment	44.9	11.2	13.5	7.9	7.9	14.6	89	2.09±0.32	
DOE supercomputer systems or services - e.g. DOE INCITE program	50.6	12.4	7.9	9.0	4.5	15.7	89	1.87±0.29	
DOD supercomputers or facilities								1.71±0.28	
Other									
Other NSF Awardees									
Computing systems in my own lab or dept.	3.8	7.5	7.5	17.0	64.2		53	4.30±0.32	
Computing systems campus (or at my own institution in multi-campus research institutions)	5.7	13.2	13.2	18.9	49.1		53	3.92±0.36	

Facilities	% of respondents						Total responses	Mean	+/-95% CI
	1	2	3	4	5	n/a			
State or regional campus facilities (e.g. SURAGrid, DiaGrid)	39.6	7.5	11.3	13.2	11.3	17.0	53	2.39±0.47	
XSEDE – the eXtreme Science and Engineering Discovery Environment	35.8	9.4	7.5	11.3	26.4	9.4	53	2.81±0.51	
DOE supercomputer systems or services - e.g. DOE INCITE program	37.7	5.7	7.5	13.2	18.9	17.0	53	2.64±0.51	
DOD supercomputers or facilities	41.5	15.1	3.8	11.3	11.3	17.0	53	2.23±0.46	

Table 8. Responses to questions about strong vs. weak scaling and dynamic and static data

Scaling and Dynamic Data	% of respondents				Total responses
	Strong Scaling	**Weak Scaling**	**Both**	**I'm not sure**	
NSF OCI awardees					
If you do any sort of parallel computing, do your applications involve "strong scaling" or "weak scaling" (Strong scaling means that with a fixed problem size, the number of processors is increased. Weak scaling involves increasing the size of the problem with the number of processors).	17.9	11.5	64.1	6.4	78
Other CISE awardees					
If you do any sort of parallel computing, do your applications involve "strong scaling" or "weak scaling" (Strong scaling means that with a fixed problem size, the number of processors is increased. Weak scaling involves increasing the size of the problem with the number of processors).	30.0	11.1	48.9	10	90
Other NSF Awardees					
If you do any sort of parallel computing, do your applications involve "strong scaling" or "weak scaling" (Strong scaling means that with a fixed problem size, the number of processors is increased. Weak scaling involves increasing the size of the problem with the number of processors).	37.7	17.0	26.4	18.9	53
	Static	**Dynamic**	**Both**		
NSF OCI Awardees					
Do your analyses involve dynamic data - e.g. data that change in real time – or data that are static (not changing)	45.5	7.8	46.8		77
Other CISE awardees					
Do your analyses involve dynamic data - e.g. data that change in real time – or data that are static (not changing)	39.6	8.8	51.6		
Other NSF awardees					
Do your analyses involve dynamic data - e.g. data that change in real time – or data that are static (not changing)	47.2	15.1	37.7		

Table 9. Responses to Yes/No question about whether or not software tools allow respondents to create software applications that satisfy their current research needs

Software Tools	**No**	**Yes**	**Total responses**
Do the software tools you use now allow you to create software applications that satisfy your current research needs?			
NSF OCI awardees			
Frequencies	43	25	78
Percentages	55.1%	44.9%	
Other CISE awardees			
Frequencies	33	58	91
Percentages	36.3%	63.7%	
Other NSF Awardees			
Frequencies	14	38	52
Percentages	26.9%	73.1%	

Table 10. Responses to questions about obstacles presented in research by current parallel computing tools. Mean values are reported with a 95% confidence interval.

Future Research Needs	% of respondents					Total responses	Mean value
	1	2	3	4	5		
As regards your current research needs, please indicate the obstacles presented to your research by the parallel computing tools you currently use (rank each from 1= not at all important, to 5= extremely important). [Logic: this question presented only to those who answered "No" to the question "Do the software tools you use now allow you to create software applications that satisfy your current research needs?"].							
NSF OCI Awardees							
Current software tools do not allow applications to scale due to latency they induce	18.6	7.0	25.6	37.2	11.6	43	3.16±0.40
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	12.2	9.8	24.4	31.7	22.0	41	3.41±0.41
Programming heterogeneous systems is too difficult	9.3	9.3	18.6	16.3	45.6	43	3.81±0.42
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	37.2	18.6	23.3	14.0	7.0	43	2.35±0.40
Current software tools do not allow global addressing of a sufficient amount of memory	34.9	11.6	25.6	14.0	14.0	43	2.60±0.45
Data transport among processes is inadequate	14.0	14.0	30.2	23.3	18.6	43	3.19±0.39
My applications involve algorithms that are inherently scale-limited	36.5	19.5	19.5	19.5	4.9	41	2.37±0.41
I am a computer science researcher and development of new tools is part of my research	33.3	0.0	19.0	11.9	35.7	42	3.17±0.53
Other CISE awardees							
Current software tools do not allow applications to scale due to latency they induce	7.1	7.1	21.4	39.3	25.0	28	3.68±0.45
Current software tools do not allow applications to scale to a	3.6	7.1	21.4	35.7	32.1	28	3.69±0.42

Future Research Needs	% of respondents					Total responses	Mean value
	1	2	3	4	5		
large enough number of processors for other reasons							
Programming heterogeneous systems is too difficult	0.0	3.7	7.4	33.3	56.6	27	4.41±0.31
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	22.2	14.8	14.8	29.6	18.5	27	3.07±0.58
Current software tools do not allow global addressing of a sufficient amount of memory	18.5	14.8	25.9	22.2	18.5	27	3.07±0.55
Data transport among processes is inadequate	11.5	7.7	26.9	26.9	26.9	26	3.50±0.53
My applications involve algorithms that are inherently scale-limited	7.4	18.5	37.0	18.5	18.5	27	3.22±0.47
I am a computer science researcher and development of new tools is part of my research	17.2	3.4	3.4	13.8	62.1	29	4.00±0.59
Other NSF awardees							
Current software tools do not allow applications to scale due to latency they induce	7.1	14.3	28.6	28.6	21.4	14	3.43±0.70
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	16.7	0.0	16.7	33.3	33.3	12	3.67±0.91
Programming heterogeneous systems is too difficult	7.1	0.0	28.6	50.0	14.3	14	3.64±0.58
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	25.0	8.3	16.7	33.3	16.7	12	3.08±0.96
Current software tools do not allow global addressing of a sufficient amount of memory	23.1	15.4	7.7	23.1	30.8	13	3.23±0.99
Data transport among	0.0	8.3	33.3	16.7	41.7		3.92±0.69

Future Research Needs	**% of respondents**					**Total responses**	**Mean value**
	1	**2**	**3**	**4**	**5**		
processes is inadequate							
My applications involve algorithms that are inherently scale-limited	15.4	23.1	23.1	23.1	15.4	13	3.00±0.82
I am a computer science researcher and development of new tools is part of my research	61.5	0.0	15.4	7.7	15.4	13	2.15±0.99

Table 11. Open text field responses to question about characteristics that a parallel computing software environment would have in order to meet current research needs – from the user view.

If you are primarily a software application user, please describe in your own terms the characteristics that a parallel computing software environment would have in order to meet your current research needs
NSF OCI Awardees
we need to be able to track how students/researchers learn to use parallel tools
we are in need of efficient priority queues, load balancing schemes, and fault tolerance algorithms.
Similar architecture to the new Intel phi coprocessor -- combination of vectorization, threading, MPI across multiple types of processors. Tools are still primitive/raw, though.
Should be more elegant, high-level ways to implement multithreading.
need scalable environments that can deal with high dimensional data, that can have multichannel, n dimensional components with dynamic components
My needs are bifurcated -- I have some needs that the computational environment support a rather generic 'stack' of software packages. And I have other needs that require simply raw compute cycles to run discipline-specific apps developed by others. In essence, I want an environment that is more fully featured and supported and another that is more of a compute-sandbox.
Memory per processor. Very fast communication between processors. We do not use Global Arrays
Many user applications were intended to run on desktop computers. They are not designed for efficiency, much less parallelism. What's missing here is the lack CI STAFF to help these applications software users to pick the best available codes and how to use them.
High bandwidth, low latency asynchronous communication, including among the heterogeneous hardware devices. Amdahl's law still dictates the need for very fast single-thread processing.
Ease of use. Needs to compile “out of the box”; needs to be compatible with existing software; Needs to scale on single node with many cores (up to 64). Must also have good application performance for smaller problem sizes.
ease of use - I need a cs-trained staff person to launch our Condor jobs - We are a [DEIDENTIFIED: what kind of lab] lab.
configurable runtime environment
An ability to better manage the working set
Affordable, flexible computational materials development software application platforms do not exist. Many groups have developed key components, e.g. USPEX (Oganov, SUNY), ASE (CAMd, Danish Technical University), Materials Project and pymatgen (Ceder, MIT), etc., but utilization of these resources requires the work of a software developer at the level of high level programming languages, e.g. python.
Adequate access to systems at scale.
Other CISE awardees
Hadoop & Hive processing for very large (Facebook-scale) datasets
Execute and manage many parallel instances of program
Other NSF awardees
More core resources. I have access to a super-computer, but not the control language facility to create user spaces with imported software functions. GPUs are nice but have limited functionality. Better random number resources for parallel computing would also be nice.
It should be open source so that it is free. There should be no licensing or mutual compatibility issues (current FFT package not running with python for example).
I'm in [DEIDENTIFIED: field of work/research] --- I build things, but I'm not a computer scientist by training. I'd mainly look for maximal transparency. I'd like to use parallelization without having to think about it.

If you are primarily a software application user, please describe in your own terms the characteristics that a parallel computing software environment would have in order to meet your current research needs
Better tools, graphics generation, gui
an easier user interface - GUI, rather than command line.
Ability to allocate nodes dynamically as needed and to continue, or easily restart jobs, if one node fails. Currently combining MPI and Open MP in jobs that require both large memory per threat and a large number of threads is difficult. MPI has improved over the years but it also seems deficient at times when managing memory issues in large jobs.

Table 12. Open text field responses to question about characteristics needed in a hardware testbed for software development and computer science research considering current needs.

If you are primarily a software developer or computer scientist, please describe the characteristics that you need in a hardware testbed for computing tool development in order to advance your current research
NSF OCI awardees
see above (above answer added by CSR here: we need to be able to track how students/researchers learn to use parallel tools)
reduced cost access to commercial elastic cloud resources; no specialized hardware needed. The public cloud, exemplified by AWS and now MS and Google, is on the right track and covers most use cases. More transparency/control over IO is desirable, but not a showstopper.
Primarily, I need sandboxed systems that allow developers to work with experimental kernels.
Hardware should have high-bandwidth, low-latency asynchronous communication paths.
overcome the heterogeneity, improve the time-to-complete for parallel apps Need better tools to link, run multiple codes and simplify development. of hpc capability, and make codes more user friendly.
Multiple OSes and platforms for tests, but none need to be large.
More transparent access to data movement and assembly. I am looking for heterogeneous hardware or architecture combinations along with improved hardware-level monitoring for data transfer (networking), memory usage (caches) and instruction bandwidth (vectorization).
Easy means to distribute processing and, perhaps more importantly, tools to monitor bottlenecks.
Better parallel debugging tools and simpler high-performance I/O APIs would go a long way towards improving my research. as a [DEIDENTIFIED: type of work done], I would like a platform where the kernel allows access to all hardware monitoring capabilities (e.g. network counters, CPU capabilities like instruction-based sampling and lightweight profiling, energy consumption), support application study with and without interference by other jobs, measurements of I/O and communication load induced by other jobs.
Adequate access to systems at scale.
Other CISE awardees
we need dedicated testbed for experiment important hardware features, push to the limit of the hardware capabilities and more aggressively develop software tools
State of the art debugger development requires heterogeneous computing platform hardware with libraries that are ultra reliable. Ideally there must be nationwide cooperation on this, say through NSF's SI2

If you are primarily a software developer or computer scientist, please describe the characteristics that you need in a hardware testbed for computing tool development in order to advance your current research
Raw computer, very large memory footprint (terabytes), transparent access to large SSD storage (tens of TB) and introspective tools for post mortem of memory, compute, and communication performance/
Multicore, high speed interconnects, large-scale parallelism, GPU, storage hierarchy with different types of storage media.
massively multicore machine for concurrent, but non-distributed, applications. Anything that minimizes resource contention (memory, cache, bus, etc.) on the same machine lack of real data, sensor network is still a vision not reality, it has many factors (energy, storage, failure, data correlation) to consider, making problem solving highly difficult,
It should be configurable to model as many different realistic systems as possible, including likely future configurations.
I research [DEIDENTIFIED: subject researched], via simulation. I need the ability to run jobs for a long time (weeks) because of the inherent slowdowns. Automatic checkpointing would be terrific. The generate trace output files are often 100 Gb or more (compressed).
I need good profiling and debugging tools.
I need direct access to the network stack as much of my research involves [DEIDENTIFIED: research topic].
highly parallel; representative of broad range of current computing architectures and processing hardware; configurable in terms of computational, network, and I/O characteristics; permitting configuration changes on a short notice; allowing integration of heterogeneous processing hardware, including accelerators
heterogeneous nodes
Heterogeneous computing systems (CPU+FPGA, +GPU). Easy to use integrated tools are inexistent. Development of tools for heterogeneous platforms is VERY hard AND very expensive ($ and time). It is beyond the scope of academic research: small research delta for a huge time investment.
Good abstractions
Extended storage for terabytes of data.
Easy access to processors, such as GPUs either as separate processors, or potentially as one large GPU.
Other NSF awardees
I need much, much faster memory access. I would like data flow (rather than current instruction flow) hardware. Current hardware is poorly designed for the numerical solution of partial differential equations
benchmarks, processing heuristics, test suites

Table 13. Responses to Yes/No question about whether future needs for software tools are the same as current needs

Future Research Needs	No	Yes	Responses
Do you anticipate that the software tools you use now will allow you to create software applications that satisfy your future research needs?			
NSF OCI awardees			
Frequencies	41	36	77
Percentages	53.2%	46.8%	
Other CISE awardees			
Frequencies	31	58	89
Percentages	34.8%	65.2%	
Other NSF awardees			
Frequencies	23	29	52
Percentages	44.2%	55.8	
Are your future needs for parallel software tools the same as your current needs?			
NSF OCI Awardees			
Frequencies	13	18	31
Percentages	41.9%	58.1%	
Other CISE awardees			
Frequencies	8	14	22
Percentages	36.4%	63.6%	
Other NSF awardees			
Frequencies	7	6	13
Percentages	53.8%	46.2%	

Table 14. Responses to questions about obstacles anticipated in the future related to parallel computing tools. Mean values are reported with a 95% confidence interval.

Future Research Needs	% of respondents					Total responses	Mean values
	1	2	3	4	5		
As regards future research needs, please indicate the obstacles presented to your research by the parallel computing tools you currently use (rank each from 1= not at all important, to 5= extremely important). [Logic: this question presented only to those who answered "No" to both questions that future software tools will not meet needs and the needs will be different than current needs]							
NSF OCI Awardees							
Current software tools do not allow applications to scale due to latency they induce	15.8	15.8	21.1	10.5	36.8	19	3.37±0.74
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	21.1	10.5	10.5	31.6	26.3	19	3.32±0.73
Programming heterogeneous systems is too difficult	15.8	5.3	15.8	5.3	57.9	19	3.84±0.76
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	42.1	10.5	21.1	15.8	10.5	19	2.42±0.71
Current software tools do not allow global addressing of a sufficient amount of memory	22.2	22.2	27.8	16.7	11.1	18	2.72±0.66
Data transport among processes is inadequate	15.0	10.0	15.0	15.0	45.0	20	3.65±0.72
My applications involve algorithms that are inherently scale-limited	42.1	21.1	26.3	0.0	10.5	19	2.16±0.63
I am a computer science researcher and development of new tools is part of my research	16.7	5.6	27.8	5.6	44.4	18	3.56±0.76
Other CISE awardees							
Current software tools do not allow applications to scale due to latency they induce	6.7	20.0	26.7	33.3	13.3	15	3.27±0.64
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	0.0	13.3	20.0	26.7	40.0	15	3.93±0.61

Future Research Needs	% of respondents					Total responses	Mean values
	1	2	3	4	5		
Programming heterogeneous systems is too difficult	6.3	6.3	18.8	25.0	43.8	16	3.94±0.66
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	26.7	33.3	6.7	33.3	0.0	15	2.47±0.69
Current software tools do not allow global addressing of a sufficient amount of memory	20.0	20.0	26.7	26.7	6.7	15	2.80±0.70
Data transport among processes is inadequate	26.7	13.3	26.7	20.0	13.3	15	2.80±0.79
My applications involve algorithms that are inherently scale-limited	23.1	46.2	15.4	7.7	7.7	13	2.31±0.71
I am a computer science researcher and development of new tools is part of my research	12.5	12.5	0.0	25.0	50.0	19	3.88±0.80
Other NSF Awardees							
Current software tools do not allow applications to scale due to latency they induce	7.1	0.0	42.9	42.9	7.1	14	3.43±0.54
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	0.0	16.7	16.7	41.7	25.0	12	3.75±0.67
Programming heterogeneous systems is too difficult	0.0	7.1	42.9	21.4	28.6	14	3.71±0.58
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	23.1	0.0	30.8	38.5	7.7	13	3.08±0.79
Current software tools do not allow global addressing of a sufficient amount of memory	0.0	14.3	21.4	57.1	7.1	14	3.57±0.49
Data transport among processes is inadequate	0.0	30.8	30.8	23.1	15.4	13	3.23±0.66
My applications involve algorithms that are inherently scale-limited	0.0	7.1	35.7	42.9	14.3	14	3.64±0.49

Future Research Needs	% of respondents					Total responses	Mean values	
	1	2	3	4	5			
I am a computer science researcher and development of new tools is part of my research	66.7	8.3	8.3	16.7	0.0	12	1.75±0.77	

Table 15. Open text field responses to question about characteristics that a parallel computing software environment would have in order to meet future research needs – from the user view.

If you are primarily a software application user, please describe in your own terms the characteristics that a parallel computing software environment would have in order to meet your future research needs
NSF OCI awardees
More software needs to be developed with proper support by experts for users to use effectively.
Many processors, fast communication between them, large fast cache.
I design [DEIDENTIFIED: what is designed].
Higher-level multithreading without having to dig through a cryptic library.
computational fluid dynamics using finite difference scheme. Inherently parallel except for radiation transport which is non-local. Parallel transport schemes are diffusive. Need fast reshape data tools.
Other CISE awardees
Very-low latency shared memory machines
Seamlessly use the available cores both on CPU and GPU, and also at the same time help with sustainability issues.
data-centric large scale natural language processing with support for rapid experimentation and visualization
Other NSF awardees
Simple to use. Both numerically stable as well as some basic stability against hardware failure. The later would allow longer wall times for typical queues which we need for algorithms that exhibit low scaling.
I'd like to be able to distribute my model output files on many computers, so that operations which needed to address all the files (without interaction) could be done in real time.
Handling visualization of large data sets e.g. replacement for netcdf
a GUI

Table 16. Open text question regarding transformative scientific challenges that could be pursued if researchers were not limited by cyberinfrastructure resources

Imagine that availability of computing processors and computing tools was not a limiting factor. What transformative scientific challenges would you be able to pursue if you were not limited by cyberinfrastructure resources?
NSF OCI Awardees
Visualization and data analysis remain a challenge for our very large multi-variate 3D data sets. High time resolution data dumps are also not possible and use of techniques such as in-situ viz are required.
The next big challenge is reliability in very large ensembles of computations.
Processing across multiple classes of devices, from iOS (iPads, iPhones) to supercomputing clusters. Merging numerical model execution with synchronized visualization of output in a video game loop. Fly through the output as it is being generated.
Multiscale combustion problems Multiscale materials problems
modeling the entire solar convection zone plus atmosphere to understand the solar dynamo and the impact of magnetic fields it produces on the solar atmosphere.
Memory and disk space are the big issues for us, all the time. That's the deal-breaker!
Fresh design of large scale computational science and engineering applications to make full use of exascale architectures.
Existing resources (nr cores, amount of RAM, interconnect) are much better than the existing software is able to use. What is needed is motivation and funding to redesign algorithms and program them to use machines better for most applications, not just the handful of 'pretty super model applications'.
Development of energy materials -- both storage and production -- capable of powering next generation electric vehicles and enabling renewable energy through grid-level storage. Explore the role of defects on the electronic properties of photovoltaics or on the kinetics of catalysis. Combine materials data into a searchable framework to identify correlations in calculable materials properties (fast) and experimentally measured materials properties (slow), reducing the scope of the latter. Design devices and materials at the nano-scale for direct usage in the semiconductor industry. Design of semiconductor quantum bits.
Compete with Google/MS/Facebook/Twitter in research building and deploying massive scale services. Academia is at risk of being shut out of the game in building these systems; we can't extrapolate from experiments on 20 machines and expect to be relevant.
As memory goes up with number of processors that is not an issue. Problems inherently involve wide range of temporal and spatial scales. Need some easy way of static mesh decomposition to let different regions run at own pace and communicate as needed.
Other CISE awardees
We would be able to put the output of my collaborators [DEIDENTIFIED: what kind of models] models on line, and let other researchers perform their analyses locally on the data. Currently, reuse of the model is limited. It is not generally feasible for other researchers to download the long term model simulations (13TB growing to 65TB), and they do not have the computing resources to allow ad hoc analysis by outsiders. The cost for hosting on a 3rd-party platform, such as Amazon, is prohibitive.
Understanding the extent to which computer *performance* (not output values) is chaotic, the factors that influence it, the distribution of performance, how different hardware features affect chaos, and how we might control performance variation better.
Real-time computation.
large scale experimentation with (human) language variation and change via information theoretic analysis of ambiguity management

Imagine that availability of computing processors and computing tools was not a limiting factor. What transformative scientific challenges would you be able to pursue if you were not limited by cyberinfrastructure resources?
Improve program correctness. Investigate code for bugs.
Address bigger and harder problems and test more alternatives.
Other NSF awardees
We would be able to do interactive nesting - like performing a high-resolution nested simulation in real time.
Much larger, more complex problems; more physics, parameter sweeps, finer grids multiphysics, domain interactions, combining models/systems at different time and resolution (e.g., organism behaviors coupled with molecular dynamics, membranes, gene circuits)
i could have better visualization of processes, time series, big data
I am primarily a [DEIDENTIFIED: what kind of scientist] scientist. The main limitation facing my research is that of current theories/algorithms. From my perspective theoretical/algorithmic progress is more crucial than cyber-infrastructure, although the latter is very important too. Keeping the current pace of progress in the development of the cyber-infrastructure should provide computational resources in the future that will allow to implement/apply algorithms/theories to tackle problems that cannot be solved today.
Energy research. Memory capacity is a serious issue for part of our algorithms.
And what if you had all the money you needed? You could cure cancer and invent time travel and figure out how to live forever, and ... Not a useful question really, is it.

Table 17. Open text field responses to question about characteristics needed in a hardware testbed for software development and computer science research considering future needs.

If you are primarily a software developer or computer scientist, please describe the characteristics that you need in a hardware testbed for computing tool development in order to advance your future research
NSF OCI Awardees
See previous answer on this topic (previous answer added by CSR here: The next big challenge is reliability in very large ensembles of computations.)
Access to a range of systems from medium scale to largest
Basically I need high-throughput access to large numbers of CPUs/cores/whatever on an actual production system. If I'm trying to implement something new, the limiting issue is being able to test at scale. If I have to wait a week between tests, that really slows down the development process.
Systems like Titan, BlueWaters, BlueGene/Q are great and could be used well if there was a more collaborative support structure for rewriting old applications to use modern approaches that have been proven to work.
Need a better approach to heterogeneous parallel programming approaches for real end-to-end applications. Not some additional language concepts described in CS papers. Commercial cloud at reduced cost. We need a single-payer system for access to commercial cloud offerings.
Similar to my previous answer - need access to next-generation standards and foundational runtimes.
I need the latest in processors (manycore and accelerator) along with the development environments and support from vendor developers who administer the system.
Other CISE awardees
The most helpful would be very-low latency shared memory computers.
Scalable storage infrastructure
Need access to open source tools so that we can change architecture or memory configurations and recompile applications easily.
Need a third-party platform that allows affordable storage of 100+ TB of data either online or near-line. Offerings such as Amazon S3 are too expensive. Amazon Glacier has 3-5 hour restore time and charges for large restores
MPI very many processors reliable and extensive access
Large core count with hardware-supported shared address space. Cores may or may not be homogeneous. Memory may or may not be coherently cached.
Heterogeneous nodes
Capacity: number of cores, amount of memory, communication bandwidth, disk space for traces. I/O *bandwidth* is less of an issue.
Be able to play with the infrastructure-level details, i.e., physical machines as well as hypervisors.
As stated in the previous section, plus low level instrumentation support (including hardware level signal capture, storage, and analysis)
(as answered before) easily configured hardware with reliable libraries, and ideally collaborative ventures championed by say NSF's SI2.
Other NSF awardees
We need 100x times more memory channels per processor to get past the "memory wall". More in the future. We need hardware that is designed to be scalable. That is - it knows how to handle (and buffer) very different memory access times.
We build [DEIDENTIFIED: what is built] and visualization and large memory are essential requirements.
Simple implementation, debugging and profiling of serial and parallel code with low overhead. Sufficiently high performance per core.

If you are primarily a software developer or computer scientist, please describe the characteristics that you need in a hardware testbed for computing tool development in order to advance your future research
Robust, scalable, easily modified by users.
- consistent with current/emerging architectures - scalable - ability to play with architectural parameters (e.g., cache/memory sizes, block/line sizes, speeds, topologies, link behaviors) - access to counters - connections to RAS behavior

4. Appendix 1 – Survey Invitation Letter and Survey Instrument

Survey of Parallel Computing Needs

I am writing to ask for your participation in a survey being conducted as part of activities funded by the National Science Foundation via NSF award 1205518 (REFT - A Reconfigurable Execution Framework Testbed for data-driven and extreme scale computing). The purpose of this grant award from the NSF is to fund needs analysis and planning for a community environment testbed that will enable development and application of new parallel software technologies and new parallel algorithm design.

The purpose of this survey is to assess your scalable application needs and determine what constraints you face, and determine what solutions to those constraints may be of interest. This will help the Indiana University research team identify needs for a reconfigurable computing testbed that will be used to host and support innovative parallel programming models, environments, and operating and runtime systems to meet those needs NOT adequately met by current conventional parallel computing software technologies.

This survey is done under the auspices of the Indiana University Center for Survey Research (CSR), which assures that your responses will remain completely confidential. Neither your name nor your organization will be associated with any data or included in any reports. This survey has been approved (protocol #1205008730) by the Indiana University Institutional Review Board (IRB).

If you have any questions about this survey or how the results will be used, please feel free to contact Rebecca Schmitt, Chief of Staff, Center for Research in Extreme Scale Technologies, Pervasive Technology Institute, Indiana University, at raslowe@iu.edu, or call (812)-856-0501.

Thank you for your time and help with this important effort that will impact future decisions related to environments and tools that will enable the development of new software supporting scientific research.

Yours Truly,

Thomas Sterling
Principal Investigator, REFT - A Reconfigurable Execution Framework Testbed for data-driven and extreme scale computing
Chief Scientist, Center for Research in Extreme Scale Technologies
Indiana University

&

Craig A. Stewart
Co-Principal Investigator
Associate Director, Center for Research in Extreme Scale Technologies
Indiana University

Access to current resources

Do you currently have access to sufficient cyberinfrastructure facilities (computation, data storage and management, visualization, software tools) to address your current research needs?

- ☐ Yes - all of the time
- ☐ Most of the time
- ☐ Some of the time (< 50%)
- ☐ No - Never

Considering just computational resources, do you have access to sufficient computational resources to address your current research needs?

- ☐ Yes - all of the time
- ☐ Most of the time
- ☐ Some of the time (< 50%)
- ☐ No - Never

Considering just data storage and management resources, do you have access to sufficient resources to address your current research needs?

- ☐ Yes - all of the time
- ☐ Most of the time
- ☐ Some of the time (< 50%)
- ☐ No - Never

Considering just visualization resources, do you have access to resources sufficient in functionality and capability to address your current research needs?

- ☐ Yes - all of the time
- ☐ Most of the time
- ☐ Some of the time (< 50%)
- ☐ No - Never

Considering just software tools for your research, do you have access to sufficient and sufficiently high quality / properly functioning tools to address your current research needs?

- ☐ Yes - all of the time
- ☐ Most of the time
- ☐ Some of the time (< 50%)
- ☐ No - Never

Expected availability of future resources

growth in availability of cyberinfrastructure facilities (computation, data storage and management, visualization, software tools) will be sufficient to address your research needs at that time?

- ☐ Highly confident (> 90% likelihood in my estimation)
- ☐ Confident (> 50% likelihood in my estimation)
- ☐ Not confident (< 50% likelihood in my estimation)
- ☐ Highly pessimistic (< 10% likelihood in my estimation)

Looking five years ahead, do you expect that the growth in availability of computational resources will be sufficient to to address your research needs at that time?

- ☐ Highly confident (> 90% likelihood in my estimation)
- ☐ Confident (> 50% likelihood in my estimation)
- ☐ Not confident (< 50% likelihood in my estimation)
- ☐ Highly pessimistic (< 10% likelihood in my estimation)

Looking five years ahead, do you expect that the growth in availability of just data storage and management resources will be sufficient to address your research needs at that time?

- ☐ Highly confident (> 90% likelihood in my estimation)
- ☐ Confident (> 50% likelihood in my estimation)
- ☐ Not confident (< 50% likelihood in my estimation)
- ☐ Highly pessimistic (< 10% likelihood in my estimation)

Looking five years ahead, do you expect that the growth in availability and functionality and capability of visualization resources will be sufficient to address your research needs at that time?

- ☐ Highly confident (> 90% likelihood in my estimation)
- ☐ Confident (> 50% likelihood in my estimation)
- ☐ Not confident (< 50% likelihood in my estimation)
- ☐ Highly pessimistic (< 10% likelihood in my estimation)

Looking five years ahead, do you expect that the growth in quality and functionality of software tools for your research will be sufficient to address your research needs at that time?

- ☐ Highly confident (> 90% likelihood in my estimation)
- ☐ Confident (> 50% likelihood in my estimation)
- ☐ Not confident (< 50% likelihood in my estimation)
- ☐ Highly pessimistic (< 10% likelihood in my estimation)

Software Tools

Do you use parallel or high throughput computing in your research?

☐ Yes
☐ No

Thank you. Please continue pressing 'Next Page >>' until you reach the end of the survey.

Please rate the importance of the following parallel tools in your current research (from 1 = Not at all Important to 5 = Very Important)

	1	2	3	4	5	n/a
MPI (Message Passing Interface)	☐	☐	☐	☐	☐	☐
CUDA or other library for GPUs	☐	☐	☐	☐	☐	☐
OpenMP	☐	☐	☐	☐	☐	☐
Cloud/Data Parallel (e.g. Hadoop)	☐	☐	☐	☐	☐	☐
Condor (High Throughput Computing)	☐	☐	☐	☐	☐	☐
BOINC (Berkeley Open Infrastructure for Network Computing - High Throughput Computing)	☐	☐	☐	☐	☐	☐
I'm not sure - I use programs provided by someone else	☐	☐	☐	☐	☐	☐
Other	☐	☐	☐	☐	☐	☐

Please specify what other parallel or high throughput computing software tools you use ______________________

Facilities

Thank you. Please continue pressing 'Next Page >>' until you reach the end of the survey.

Please rate the importance of the following facilities in your current research (from 1 = Not at all Important to 5 = Very Important)

	1	2	3	4	5	n/a
Computing systems in my own lab or department	☐	☐	☐	☐	☐	☐
Computing systems on my own campus (or at my own institution in multi-campus research institutions)	☐	☐	☐	☐	☐	☐
State or regional campus facilities (e.g. SURAGrid, DiaGrid)	☐	☐	☐	☐	☐	☐
XSEDE - the eXtreme Science and Engineering Discovery Environment	☐	☐	☐	☐	☐	☐
DOE supercomputer systems or services - e.g. DOE INCITE program	☐	☐	☐	☐	☐	☐
DOD supercomputers or facilities	☐	☐	☐	☐	☐	☐

Scaling and Dynamic Data

Thank you. Please continue pressing 'Next Page >>' until you reach the end of the survey.

If you do any sort of parallel computing, do your applications involve "strong scaling" or "weak scaling." (Strong scaling means that with a fixed problem size, the number of processors is increased and the time to completion is decreased. Weak scaling involves increasing the size of the problem with the number of processors).

- ☐ Strong Scaling
- ☐ Weak Scaling
- ☐ Both
- ☐ I'm not sure

Do your analyses involve dynamic data - e.g. data that change in real time - or data that are static (not changing)

- ☐ Static
- ☐ Dynamic
- ☐ Both

Current Research Needs

Thank you. Please continue pressing 'Next Page >>' until you reach the end of the survey.

In this section, we ask about software tools and your current research needs

Do the software tools you use now allow you to create software applications that satisfy your current research needs?	☐ Yes ☐ No

Thank you. Please press 'Next Page >>' to continue the survey.

As regards your current research needs, please indicate the obstacles presented to your research by the parallel computing tools you currently use (rank each from 1= not at all important, to 5= extremely important)

Current software tools do not allow applications to scale due to latency they induce	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Programming heterogeneous systems is too difficult	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Current software tools do not allow global addressing of a sufficient amount of memory	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Data transport among processes is inadequate	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
My applications involve algorithms that are inherently scale-limited	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
I am a computer science researcher and development of new tools is part of my research	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
If you are primarily a software application user, please describe in your own terms the characteristics that a parallel computing software environment would have in order to meet your current research needs	________________________________
If you are primarily a software developer or computer scientist, please describe in your own terms the characteristics that you need in a hardware testbed for computing tool development in order to advance your current research	________________________________

Future research needs

Thank you. Please press 'Submit' to finish the survey.

In this section, we ask about software tools and your future research needs

Do you anticipate that the software tools you use now will allow you to create software applications that satisfy your future research needs?	☐ Yes ☐ No

Thank you. Please press 'Submit' to finish the survey.

Are your future needs for parallel software tools the same as your current needs?	☐ Yes ☐ No

Thank you. Please press 'Submit' to finish the survey.

As regards your future research needs, please indicate the obstacles presented to your research by the parallel computing tools you currently use (rank each from 1= not at all important, to 5= extremely important)

Current software tools do not allow applications to scale due to latency they induce	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Current software tools do not allow applications to scale to a large enough number of processors for other reasons	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Programming heterogeneous systems is too difficult	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
My applications require dynamic allocation of processors and the tools I use require static allocation of processors before analysis begins	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Current software tools do not allow global addressing of a sufficient amount of memory	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
Data transport among processes is inadequate	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
My applications involve algorithms that are inherently scale-limited	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
I am a computer science researcher and development of new tools is part of my research	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

If you are primarily a software application user, please describe in your own terms the characteristics that a parallel computing software environment would have in order to meet your future research needs	____________________
Imagine that availability of computing processors and computing tools was not a limiting factor (not processors, not data management tools, not communications, not programming tools). What transformative scientific challenges would you be able to pursue if you were not limited by cyberinfrastructure resources? (what about memory capacity)	____________________
If you are primarily a software developer or computer scientist, please describe in your own terms the characteristics that you need in a hardware testbed for computing tool development in order to advance your future research	____________________

5. Appendix 2 – Documentation of Review and Approval (DRA)

INDIANA UNIVERSITY INSTITUTIONAL REVIEW BOARD (IRB)

DOCUMENTATION OF REVIEW AND APPROVAL (DRA)

Reviewing IRB (please choose one): IRB STUDY NUMBER: 1205008730

Biomedical: ☐ IRB-02 ☐ IRB-03 ☐ IRB-04 ☐ IRB-05

Behavioral: ☐ IRB-01 ☒ IUB IRB

Please type only in the gray boxes. To mark a box as checked, double-click the box, select "checked", and click "OK".

Section I: Investigator Information

Principal Investigator (advisor in the case of student/fellow/resident research):

Name *(Last, First, Middle Initial)*: **Sterling, Thomas**

Department: PTI Phone: 812.856-4597 E-Mail: tron@indiana.edu

Fax: 812.856.6872 Address: 2711 E. 10th Street, Wrubel Computing Center, Room 109

Co-Principal Investigator (for student/fellow/resident research):

Name:	Phone: _____	E-Mail: _____
Andrew Lumsdaine	812.855.7071	lums@cs.indiana.edu
Craig Stewart	812.325.3452	stewart@iu.edu
Matt Link	812.855.6339	mrlink@iu.edu
Maciej Brodowicz	225.334.2457	mbrodowi@indiana.edu

☐ Student: ☐ Fellow ☐ Resident

☐ Undergraduate

☐ Graduate

Additional Study Contact:

Name: Phone: E-Mail:

Project Title:

Survey for REFT: A Reconfigurable Execution Framework Testbed for Data-driven and Extreme Scale Computing

Anticipated Project Completion Date: June 30, 2013

Sponsor/Funding Agency: NSF PI on Grant: Thomas Sterling

Sponsor Protocol# /Grant #: 1053575* Period: from: 7/1/2012 to 6/30/2013*

Sponsor Type: ☒ Federal ☐ State ☐ Industry ☐ Not-for-Profit ☐ Unfunded ☐ Internally Funded

Funding Status: ☒ Pending ☐ Funded ☐ N/A

Grant Title (if different from project title): CI-P: REFT - A Reconfigurable Execution Framework Testbed for data-driven and extreme scale computing

Section II: Type of Review

☒ Exempt Review

☐ Expedited Review

☐ Full Board Review (*Choose One*) → ☐ Behavioral: ☐ IRB-01 ☐ IU Bloomington IRB
☐ Biomedical: ☐ IRB-02 ☐ IRB-03 ☐ IRB-04 ☐ IRB-05

Section III: Documents Included with Research Submission

☐ Assent, dated: _____
Number of assent documents: _____

☐ Authorization, dated: _____
Number of authorizations: _____

☐ Clinical Investigator's Brochure, dated: _____

☐ Expedited Research Checklist, dated:

1 IRB Form v01/01/2012

☒ Exempt Research Checklist, dated: 05/18/2012

☐ HIPAA & Recruitment Checklist, dated: _____

☐ Informed Consent, dated: _____
Number of consent documents: _____

☒ Investigator List, dated: 05/14/2012

☐ Protocol, dated: _____

☐ Recruitment materials (please list and date): _____

☐ Request form(s) for vulnerable population(s) (please list and date); _____

☒ Surveys, questionnaires (please list and date):
Current as of 05/14/2012; see link to survey:

☐ Summary Safeguard Statement or HUD Form, dated:

☒ Study Information Sheet, dated 05/18/2012

☐ Other (please list and date):

Section IV: Investigator Statement of Compliance

By submitting this form, the Principal Investigator assures that all information provided is accurate. He/she assures that procedures performed under this project will be conducted in strict accordance with federal regulations and Indiana University policies and procedures that govern research involving human subjects. He/she acknowledges that he/she has the resources required to conduct research in a way that will protect the rights and welfare of participants, and that he/she will employ sound study design which minimizes risks to subjects. He/she agrees to submit *any* change to the project (e.g. change in principal investigator, research methodology, subject recruitment procedures, etc.) to the Board in the form of an amendment for IRB approval prior to implementation.

Section V: IRB Approval

This research project, including all documents included with the submission (e.g., informed consent statement, authorization, and/or waiver of authorization) has been reviewed and approved by the Indiana University IRB for a maximum of a one year period unless otherwise indicated as follows:_______________

☒ Exempt Category(ies), if applicable: 2

☐ Expedited Category(ies), if applicable: _____

Authorized IRB Signature: Sara Benken
Digitally signed by Sara Benken
DN: cn=Sara Benken, o=HSO, ou=ORA, email=slbenken@iu.edu, c=US
Date: 2012.05.22 08:38:28 -04'00'

IRB Approval Date: 05/22/2012

Printed Name of IRB Member: Sara Benken

6. Appendix 3 – Letter Classifying Study as Exempt Research

INDIANA UNIVERSITY

OFFICE OF RESEARCH ADMINISTRATION

To: THOMAS STERLING
COMPUTER SCIENCE

From: IU Human Subjects Office
Office of Research Administration – Indiana University

Date: May 22, 2012

RE: **EXEMPTION GRANTED**

Protocol Title: Survey for REFT - A Reconfigurable Execution Framework Testbed for Data-driven and Extreme Scale Computing

Protocol #: 1205008730

Funding Agency/Sponsor: NATIONAL SCIENCE FOUNDATION

IRB: IRB-IUB, IRB00000222

Your study named above was accepted on May 22, 2012 as meeting the criteria of exempt research as described in the Federal regulations at 45 CFR 46.101(b), paragraph(s) (2) . This approval does not replace any departmental or other approvals that may be required.

As the principal investigator (or faculty sponsor in the case of a student protocol) of this study, you assume the following responsibilities:

Amendments: Any proposed changes to the research study must be reported to the IRB prior to implementation. To request approval, please complete an Amendment form and submit it, along with any revised study documents, to irb@iu.edu. Only after approval has been granted by the IRB can these changes be implemented.

Completion: Although a continuing review is not required for an exempt study, you are required to notify the IRB when this project is completed. In some cases, you will receive a request for current project status from our office. If we are unsuccessful at in our attempts to confirm the status of the project, we will consider the project closed. It is your responsibility to inform us of any address changes to ensure our records are kept current.

Per federal regulations, there is no requirement for the use of an informed consent document or study information sheet for exempt research, although one may be used if it is felt to be appropriate for the research being conducted. As such, these documents are returned without an IRB-approval stamp. Please note that if your submission included an informed consent statement or a study information sheet, the IRB requires the investigational team to use these documents.

You should retain a copy of this letter and any associated approved study documents for your records. Please refer to the project title and number in future correspondence with our office. Additional information is available on our website at http://researchadmin.iu.edu/HumanSubjects/index.html.

If you have any questions, please contact our office at the below address.