



Humans Advancing Research in the Cloud (HARC)

Craig A. Stewart (<https://orcid.org/0000-0002-3906-1594>)

stewart@iu.edu

Executive Director, IU Pervasive Technology Institute

Brian D. Voss

Research Engagement Manager, IU Pervasive Technology Institute

bvoss@iu.edu

pti.iu.edu @iu_pti



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

I, and the IU Pervasive Technology Institute,
respect the rights and dignity of all people

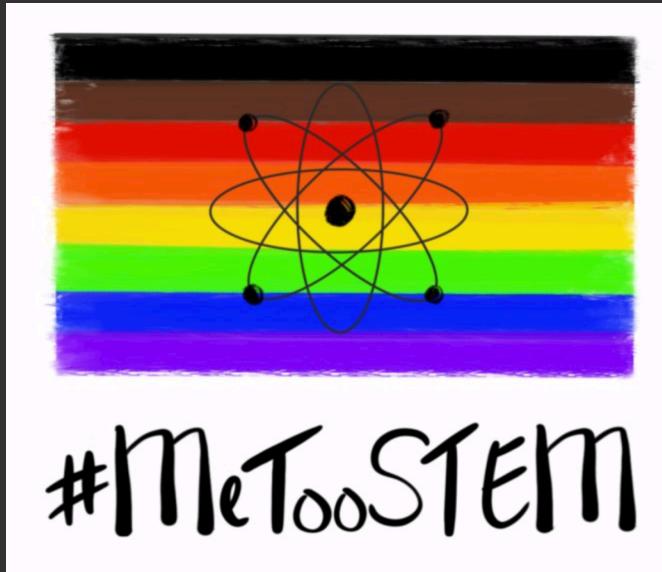


Image courtesy metoostem.com



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

There are multiple projects going now about people, research, and clouds

- “
- NIH STRIDES (<https://datascience.nih.gov/strides>)
 - NSF Cloudbank (<https://www.cloudbank.org>)
 - Internet2 "Exploring Clouds for Acceleration of Science (E-CAS)" (<https://www.internet2.edu/vision-initiatives/initiatives/exploring-clouds-acceleration-science/>)
 - HARC (<https://harc.iu.edu>)

- Please cite as: Stewart, C.A. and B.D. Voss. 2019. Humans Advancing Research in the Cloud (HARC). Presentation. Presented at fall meeting of the Coalition for Advanced Scientific Computing (CASC), The Arlington Hotel, Arlington, VA, 25 September 2019. [http###_____](http://#/#/)
- Copyright Trustees of Indiana University
- Released under a CC by 4.0 license (details at <https://creativecommons.org/licenses/by/4.0/legalcode.txt>). You may reuse and remix the contents of this talk with attribution save the one image used by permission #metoostem; for use of that please contact metoostem.com





200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

But we're the people who put people in the definition of Cyberinfrastructure

“ *Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked by high speed networks to make possible scholarly innovation and discoveries not otherwise possible.*

Indiana University has focused strategically on the role of people in supporting information technology and research since the 1990s and what was then called a distributed support model developed by Brian Voss, then an Associate Director with University Computing Services at IU



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

HARC – About the Project

“ Objectives: *To study and improve the understanding within the higher education community of how people in support roles advance the use of cloud-based cyberinfrastructure (CI) in the advancement of research at universities. To provide detail in the form of ‘return on investment’ (ROI) of the use of cloud-based CI by offering case studies on its use as compared to campus-based or existing national CI resources*

Funding: *\$1,001,000 in a contract-based award from Microsoft Corporation. While the award is provided by Microsoft, the project entails examination and use examples of all cloud CI vendors.*



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

Project Phases

- “ Phase 1: Provide support to campus-based research support engineers (post-docs, research associates) so that they may explore use of cloud services in conducting a diverse set of CI-enabled research endeavors.
- Our goal was to support case-based analysis of use of the cloud for research and establish credibility of our efforts in the broader academic community by engaging in research and producing publications
 - Findings were presented at a workshop at PEARC19 in Chicago July 29th, featuring 13 presentations (8 by project participants and 5 by non-project researchers via solicited and peer-reviewed paper contributions.

Project Phases

“ *Phase 2: In the next phase of the project (underway now through July 2020) we will focus on evolving the project to achieve the following goals:*

- Increase Awareness** by further building the community and continuing support for further publication and presentation of activities in the community
- Support the development of an engaged community** sharing experiences and detailed use-case information that will increase broader technical understanding
- Support further research endeavors** under a project-based solicitation format and advance experiences with cloud CI
- Support the development of skills** via creation of classroom delivered curriculum
- Provide feedback to vendors** on the challenges faced by researchers attempting to make use of the cloud to replace or enhance premise-based or other available CI resources.

Phase 1 Detail

“ Supported projects by people who enable research using CI.
Solicited proposals and funded eight (8) Cloud Research
Support Engineers (CRSEs) across diverse disciplines.

Support commenced in January 2019 and was provided through July 2019.

Efforts were reported on at a workshop at PEARC19 at the end of July 2019 (more on this later)

Phase 1 Detail

Project participants and their areas of research engagement

“

- Dan Sholler, UC-Berkeley, *Astronomy*
- Kris Ezra, Purdue University, *Systems of Systems*
- Derek Weitzel, University of Nebraska Lincoln, *SciToken Condor Integration*
- Eletheria Kontou, UNC-Chapel Hill, *Transportation Systems*
- Josiah Leong, Stanford U./Indiana U., *Neuroscience*
- Yongwook Song, University of Kentucky, *Machine Learning*
- Nuyun Zhang, Georgia Institute of Technology, *CI Use In Campus Research*
- John Mulligan, Rice University, *Digital Humanities*





200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

HARC '19 Workshop

“

Peer-to-Peer style format

13 presentations on relevant projects

9 peer-reviewed papers published in the ACM digital library with open access for all

>30 in attendance

Agenda, Abstracts & Presentations available





200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

Phase 2 Plans

Deliver more HARC workshops around the country

- “
 - **Use a format similar to the PEARC19 emphasizing researcher-to-researcher presentation**
 - **Provide opportunities for exposure to vendor services and use (Azure in particular)**
 - **Provide opportunities for hands-on experiences**

Support the development of additional use-case studies – papers, presentations, sharing of use-specific knowledge of both the importance of human support and ROI of cloud use

Support the development of curriculum on cloud use

Provide community engagement resources (via project website and a GitHub Repository)

Increase vendor understanding of the challenges faced by researchers and the opportunities for improving their services and support

What we've learned so far

ROI for use of the cloud, where premise-based CI exists, is still not financially competitive

“

- Especially true for ‘capacity’ use, less true for ‘capability’ use

The burden of learning how to make use of cloud services makes it challenging, and none of the vendors are doing an acceptable job of marketing and supporting research use

- Lack of documentation and vendor support with pre-built tools
- Lack of focus on researchers – more vendor focus on enterprise

The importance of people is even more critical when using the cloud, due to lack of available support resources

AWS has a stronger foot-hold due to early adoption; Google seems to be more aggressively entering the market; Microsoft has a corporate interest but is really challenged to understand higher ed and basic research so far



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

Things we've told Microsoft

As they have funded this initiative we have given them very direct feedback

“

- The process to get Azure credits is ponderous
- Billing and payment for credits is a major impediment; a commercial model won't work for higher education
- Microsoft does not have a sufficient a strategy for skill development relative to their competitors
- Support specifically for researchers lags competitors
- The core tools have great capabilities and would be much better used with better support resources



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

Direct CRSE Feedback to MS(1)

“ CRSEs were asked to “name one thing”

- Work to provide templates for common applications for researchers, but also provide a sandbox environment to try things out without signing up or being charged to see if one can make a workable solution.
- Invest in presenting more scientific use cases and setup a comprehensive set of actions for attracting researchers’ attention to enable them to test Azure and transition their work there.
- The web interface for Azure is not intuitive.
- Training for Azure on research computing related topics, not just enterprise software development.
- It is very difficult to manage billing for a given project let alone multiple projects

Direct CRSE Feedback to MS (2)

“ CRSEs were asked to “name one thing”

- Much of the documentation for the services is hard to find and based in screenshots. The standard, of course, is copy-and-paste code snippets and commands in GitHub-like documentation.
- Microsoft provides a more detailed (step-by-step) video tutorials (Including a whole series of videos that well-known machine learning algorithms such as DNN, CNN, LSTM, and RNN using also well-known data set for non-computational experts
- There is a large barrier of entry for the cloud, which does not exist for managed systems like on premises-Cl. But the elimination of that barrier of entry will include both training and technology improvements.

So what's the positive with MS? (lots)

- “➤ *Significant success with projects that involved a great deal of imagery, “cloud native” capabilities, and interactivity.*
 - This is consistent with our general observation that the major advantage in use of commercial cloud resources in research has to do with cloud-native capabilities and “capability plays” in general
- *Lots of cred to Microsoft for their stewardship of and policies around Github*
- *Significant interest in AI capabilities of MS cloud services*



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

For more information

Project related materials (papers and presentations), personnel, and workshops will be communicated via the project website:

- <https://harc.iu.edu>

A GitHub Repo will be established (by mid-October) to share papers, presentations, use-cases, pre-built tools, and other community-supplied experiences and discussion. This will be an ecumenical resource for information supporting research in the cloud – all clouds:

- <https://github.com/HARC-PTI>

For more information contact the HARC project leader for PTI, Brian Voss bvoss@iu.edu

We plan to expand HARC activities specifically to extend in depth and detail beyond our current IU-funded activities and contract from Microsoft, Inc.

Acknowledgements

Support for this work is gratefully acknowledged:

- NSF Awards 1445604, 1053575, 1548562, 1362134, 1405767, 1243436, 1148996 and 0946726.
- The Indiana University Pervasive Technology Institute
- Microsoft Inc. through support of the Humans Advancing Research in the Cloud project at the Indiana University Pervasive Technology Institute



Utterly Shameless Advertisements: Follow @scinode on Twitter (it's free and it's great)



[Home](#) [Archive](#) [Contribute](#) [Sponsor](#) [About](#) [Give Now](#)



Measuring the storm

Switzerland works hard to predict weather despite extreme topography.

[Read more](#)

And follow @iu_pti too

Any Questions?



INDIANA UNIVERSITY
BICENTENNIAL