Return on Investment for three cyberinfrastructure facilities: a local campus supercomputer; the NSF-funded Jetstream cloud system; and XSEDE

IEEE/ACM Utility and Cloud Computing, Zurich, Switzerland, 20 December 2018

Presented by Craig A. Stewart

Executive Director, Indiana University Pervasive Technology Institute

ORCID ID 0000-0003-2423-9019

stewart@iu.edu

If you would like copies of my slides, just take a picture of this one!
Slides are online at http://hdl.handle.net/2022/22590

PERVASIVE TECHNOLOGY INSTITUTE

Please cite this talk as:

Stewart, C.A., Hancock, D.Y., Wernert, J., Link, M.R., Wilkins-Diehr, N., Miller, T. and Gaither, K. 2018. *Return on Investment for three cyberinfrastructure facilities: a local campus supercomputer; the NSF-funded Jetstream cloud system; and XSEDE (the eXtreme Science and Engineering Discovery Environment).* Presented at IEEE Utility and Cloud Computing, 17-20 December, Zurich, Switzerland. Available from http://hdl.handle.net/2022/22590

There is a version of the conference paper with live web links in it online at http://hdl.handle.net/2022/22590

PERVASIVE TECHNOLOGY INSTITUTE

What is cyberinfrastructure?

- Term popularized by US National Science Foundation, defined by IU as:

  *"Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible." [1]**

- Very similar to the European concept of e-Science but explicitly includes people and human expertise as part of the infrastructure
- Cloud and utility computing are specific forms of computing and storage resources included in the general term "cyberinfrastructure."

*References at end of slides

The Indiana University Pervasive Technology Institute

- @IU_PTI is Indiana University's initiative for advanced information technology research, development, and delivery in support of scientific discovery, scholarly investigation, and artistic creation.
- In context of this talk @IU_PTI provides CI services in three contexts:
  - With funding from Indiana University, we provide the IU community with computational resources (supercomputing, cluster, cloud), storage, and visualization resources. (pti.iu.edu).
  - With funding from the US National Science Foundation, we lead delivery of cloud resources to the US open (nonclassified) research community (jetstream-cloud.org).
  - With funding from the US NSF we participate in the delivery of consulting, information, software, and CI resources to the US national research community via XSEDE.
  - **We provide resources at scale (total budget > $10M/year). Those who pay for these services want to know that their money is well used!**

Defining financial ROI for our purposes

- Return on investment: “ratio that relates income generated …to the resources (or asset base) used to produce that income.”
- ROI > 1 = good
- *ROI for CI* = $\dfrac{\textit{Value of a similar commercial service}}{\textit{Cost that was actually paid for the “locally owned” service}}$
- Why is it important? Money for research is a limiting factor in many places. We must be able to cost-justify our choices to people who care about costs first and foremost
- Our analyses are imperfect and evolving over time. We’re doing the best we can with the data we can get.

Introduction and outline

- What is cyberinfrastructure (CI) ✗
- Why should we care about its Return on Investment? ✗
- ROI estimates for 3 Cyberinfrastructure facilities in the US
  - Computational resources:
    - A traditional supercomputer (owned by a US university)
    - A federally-funded cloud system (Jetstream)
  - A unique support organization:
    - XSEDE (eXtreme Science and Engineering Discovery Environment) – the human side of CI
- What other factors impact "buy or lease" decisions?
- Where are CI services (in support of research) in evolution toward a utility service, and where are things headed in the future?

ROI for two US-based Computational Resources

- Big Red II: 1 PetaFLOPS Cray supercomputer – paid for and owned entirely by Indiana University. Mix of dual CPU and CPU-GPU nodes [2]. Calculated for 2013-2017

- Jetstream: a cloud resource used by the US national research community, funded by NSF, hosted at IU and at Texas Advanced Computing Center. Based on Dell hardware, OpenStack cloud software, University of Arizona's Atmosphere interface [3]

- ROI = $\frac{\text{(what would have been spent to buy computer time from AWS)}}{\text{(what was actually spent on local systems)}}$

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total Investment (Local)

$$TI_{local} = AC + MC + TS + TPC + TCC$$

- *AC* = Acquisition Cost
- *MC* = Maintenance Cost
- *TS* = Total Systems Admin. Costs
- *TPC* = Total Power Cost
- *TCC* = Total Co-Location Cost

Total Local Investment (in USD)

TI_local

```
=     $7,500,000
    +$  300,000
    +$  517,429
    +$1,528,468
    +$  286,200
=   $10,132,097
```

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total Value of AWS as alternative

$$TV_{AWS} = CH \times IC$$

CH is the actual core hours utilized during this period on Big Red II, not the total of available hours.

IC=Instance Cost

Total Value of AWS (3-year reserved rate)

TV_{AWS} = 637,874,648 x $0.039

= ***$24,877,111***

Total Value of AWS (1-year reserved rate)

TV_{AWS} = 637,874,648 x $0.059

= ***$37,634,604***

PERVASIVE
TECHNOLOGY INSTITUTE

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Entity	"Investment"	"Return" (Value)	ROI
Big Red II not including value of GPU resources	$10,132,097	$24,877,111 to $37,634,604	2.5 to 3.7
Add value of GPUs to TV_{AWS}		Around $3M	2.8 to 4.0
Add storage		A good bit more	Even higher

PERVASIVE TECHNOLOGY INSTITUTE

Calculating ROI on a federally-funded cloud system: Jetstream – 2018

Return On Investment for Jetstream

$$ROI_J = \frac{NVM_{avg} \times 24 \times 365.25 \times IC}{AIC \times 0.2 + OM}$$

Instance Cost (IC) determined based on comparable instance type in AWS. For Jetstream

- Average vCPUs per instance 16.9–18.5
- Average memory size 44.6–48.4 GB

Return on Investment for Jetstream (1-year reserved pricing)

$$\boldsymbol{ROI_J} = \frac{1152 \times 24 \times 365.25 \times 0.504}{6{,}576{,}101 \times 0.2 + 1{,}315{,}220}$$

$$= \boldsymbol{1.94}$$

Return on Investment for Jetstream (On demand pricing)

$$\boldsymbol{ROI_J} = \frac{1152 \times 24 \times 365.25 \times 0.796}{6{,}576{,}101 \times 0.2 + 1{,}315{,}220}$$

$$\boldsymbol{= 3.06}$$

Calculating ROI for XSEDE (eXtreme Science and Engineering Discovery Environment)

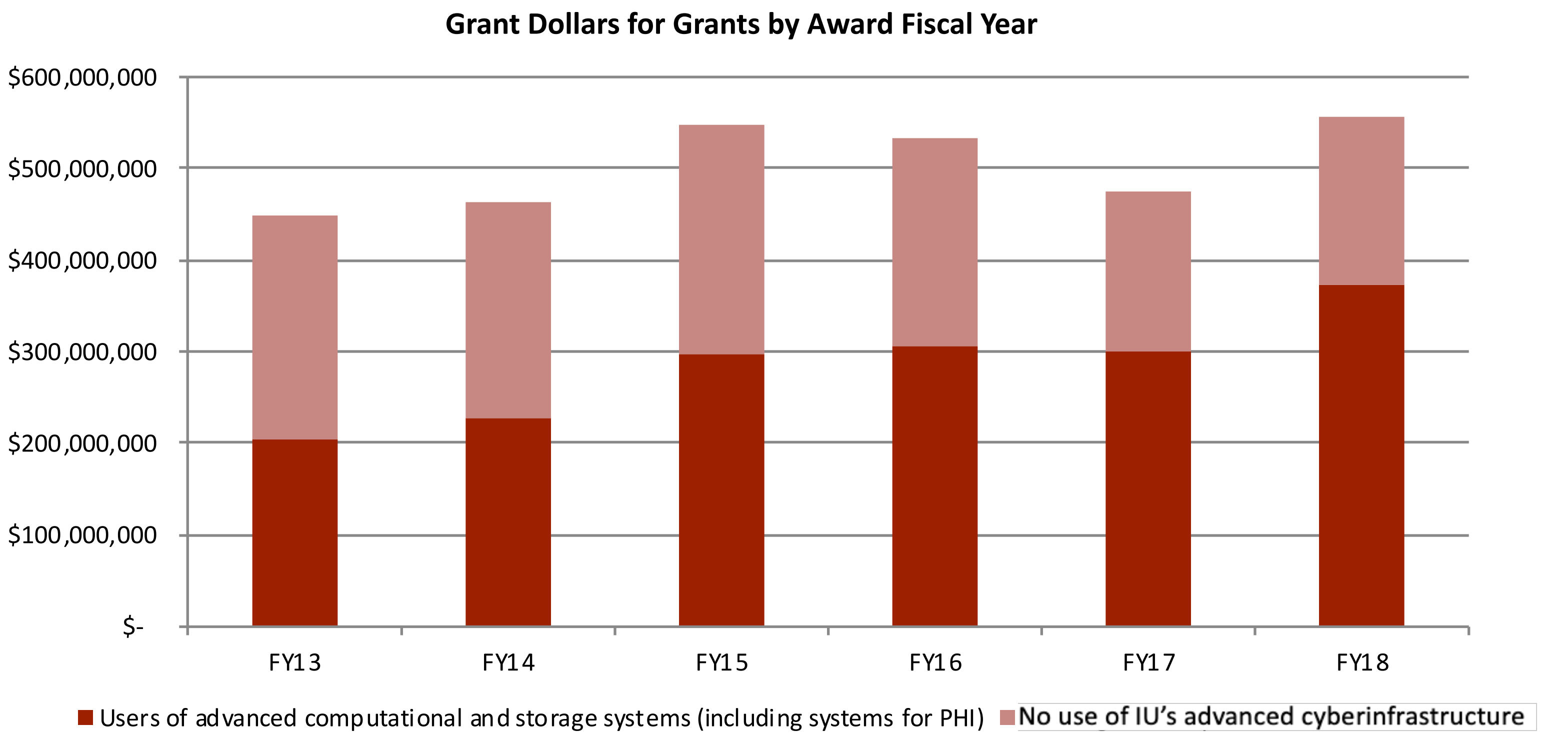
Calculating ROI on a federally-funded support and management organization: XSEDE 2016-2018

Value	PY4 (~FY2016)	PY5 (~FY2017)	PY6 (~FY 2018)
System operators	$18,343,739	$16,013,453	$15,224,653
Coding & Optimization	$1,700,00	$7,153,333	$7,358,333
End User Support	--	$3,889,800	-
Training	--	--	$1,084,724
Quantified value of XSEDE	$20,043,739 + value to stock market	$27,056,586	$24,752,435
XSEDE Budget	$23,562,931	$23,067,000	$18,442,569
ROI	**> 1**	**1.17**	**1.34**

Summary of ROI analyses

Entity	Cost	ROI
Big Red II not counting GPUs	$10,132,097	2.5 to 3.7
Big Red II counting value of GPUs	$10,132,097	2.8 to 4.0
Jetstream	$2,630,440	1.9 to 3.1
For Jetstream and Big Red II add storage		Also increases ROI!
XSEDE	$18,442,569 to $23,562,931	At least 1 to 1.3

Other views of ROI and impact for Big Red II



- Computing resources are delivered to the IU community as a free good. We are very sensitive to user community needs.
- Researchers from 134 departments and 214 disciplines/sub-disciplines

PERVASIVE TECHNOLOGY INSTITUTE

Other views of ROI and impact for Jetstream

- Focus of NSF solicitation was to increase breadth of use of NSF-funded cyberinfrastructure
- > 80% of users of Jetstream had never before used any of the NSF-funded cyberinfrastructure supported by XSEDE or its predecessors.
- We traded reliability of responsiveness for efficiency and scalability.
- But real limits on bursting.
- We built Jetstream from the ground up, focused on researchers who are not computational experts. Because of that we do things with Jetstream and deliver services to users in ways that cannot be bought from a commercial cloud provider.

Other views of ROI and impact for XSEDE

- XSEDE and predecessors have supported tens of thousands of users and 3 Nobel prizes
  - 2013 Nobel Prize in Chemistry "for the development of multiscale models for complex chemical systems"
  - 2013 Nobel prize in Physics for the discovery of the Higgs Boson
  - 2017 Nobel Prize in Physics "for decisive contributions to the LIGO detector and the observation of gravitational waves"
  - while ALSO supporting a lot of "start-up" style research
- XSEDE provides services that are unique and which accelerate research & discovery in the United States.

This talk in the context of this conference

- From the conference web site:
  - “Cloud Computing delivers computational resources on-demand as services that are commoditized and delivered comfortably analogous to traditional utilities such as electricity, gas, water and telephony.”
  - “It is essential to understand how to effectively transform these services into Utilities that provide value to both users and providers.”
  - “There is also increasing interest from commercial providers to offer business and revenue models around the services they offer.”

Where is cloud computing in terms of being a commodity today in support of research?

- "Commodities are vital components of commerce that are standardized and hence easy to exchange for goods of the same type, have a fairly uniform price around the world (excluding transport costs and taxes) and help make other products." (*The Economist*)
- Something can become a commodity through a transition:
  - It begins as a very specialized product.
  - Then, it becomes widely available.
  - At some point, standardization becomes sufficient to meet the vast majority of consumer needs, and pricing becomes a more critical factor driving consumers toward the product provided as a utility.

Progress toward Commoditization

Resource	Level of "differentiation"	Approximation to a utility?
Big Red II	A LOT	Very little so far but you CAN actually buy supercomputer time (and GPUs and FPGAs and….) from cloud providers
Jetstream	A good deal for naïve users	Closer to a utility. For the expert user much of what is done could be done on a commercial cloud system
XSEDE	Aggregates many services – approximates undifferentiated	Among the three examples discussed perhaps the closest to being a utility. •Regulated like an essential utility •Reduces differentiation among services

Resource	Freedom for users	Local Voice	Exit from use
Big Red II (& locally owned clusters & super computers)	• Complete freedom to specify system: architecture, interconnect, scale, upgrade timing • Better deals from vendors • Set level of investment –gather money, buy, run.	Local voice of users through surveys and advisory committees	Easy to quit having local resources someday if it makes sense
Jetstream (or other private cloud)	• Some freedom to set operating policies • Not free from OpenStack progress!	"Local voice" greater than w commercial clouds	Not particularly difficult
Commercial clouds	• Complete freedom in your VMs • Otherwise not much • Converts expenses from fixed capital investment into operating cost • Cloud-native benefits (apps, bursting, time-varying loads, spot markets, entry cost)	Little	Exit can be VERY difficult if you have large amounts of data stored

Why is XSEDE more like a utility?

Conclusions and predictions

- Statements like "cloud computing reduces costs" are too vague to be useful
  - It's NOT true that use of commercial clouds reduces per-unit costs we documented.
- In general, the economics still favor "buy" rather than "lease," unless:
  - You are using commercial clouds for cloud-native capabilities or risk management.
  - You are very clever about spot pricing.
  - Entry cost of running your own systems is prohibitive.
- We can see significant progress towards making research a commodity:
  - 5 years ago, GPU and HPC cycles from commercial clouds were a novelty at best
- Predictions:
  - The cost of power and cooling, and needs for cybersecurity, will drive much of the research computing we do to be cloud-based and more like a utility.
  - There will be far fewer organizations running their own hardware in 5 years. This will in part be thanks to the work represented done by attendees of this conference. Scientists will have more "voice" in setting up resources in a commercial cloud
  - Many "smaller research groups" will have access to state-of-the-art data analysis resources than today because of advances in cloud and utility computing.

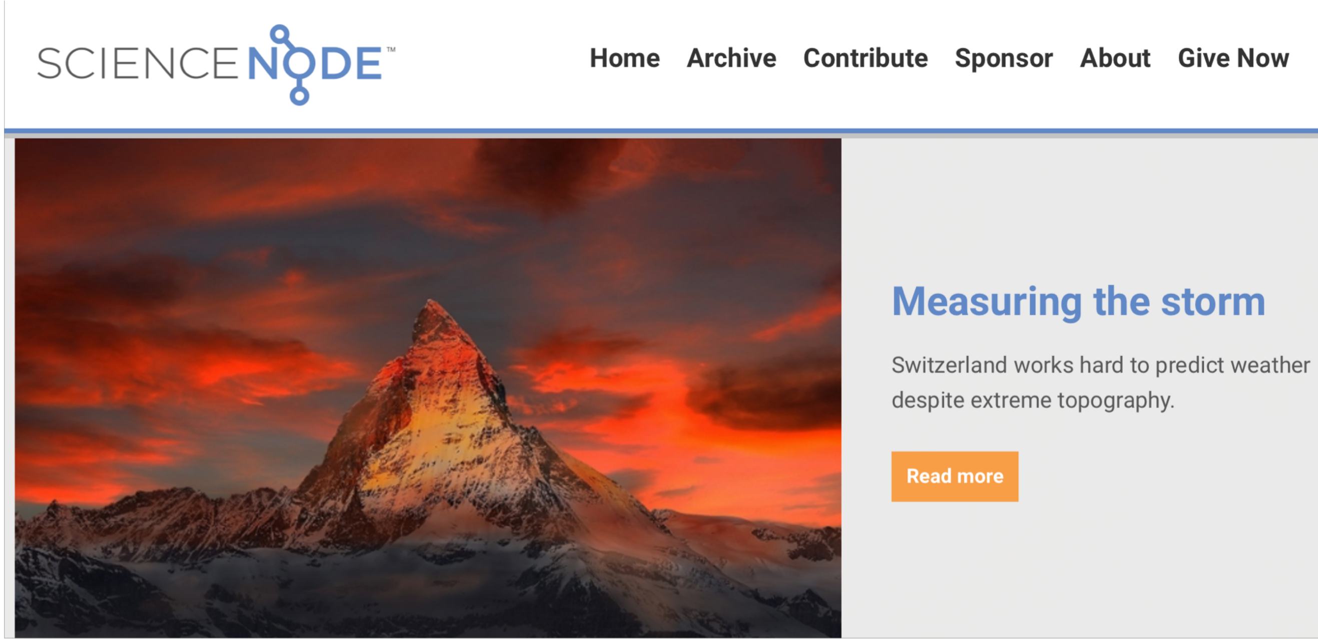
Acknowledgments

- The Indiana University Pervasive Technology Institute was created in 1999 by a major gift from the Lilly Endowment and persists today through a combination of competitively obtained federal funding, donations, and IU support.
- **IU PTI is a collaborative organization. The vast majority of the work reported here was done by staff of the Research Technologies Division of University Information Technology Services. Their wonderful skills and work are gratefully acknowledged.**
- Thanks to all my colleagues at IU PTI and XSEDE.
- Thanks to Harmony Jankowski for editing the paper and for editing and design of the talk.
- Thanks to anonymous reviewers for helpful comments on earlier versions of our paper.
- Jetstream is supported by NSF award 1445604 (David Y. Hancock, IU, PI; Craig Stewart, founding PI).
- XSEDE is supported by NSF award ACI-1548562.
- Thanks to my Swiss colleague Rudolph Eigenmann for first suggesting ROI analysis of XSEDE.
- Opinions presented here are those of the author(s) and do not necessarily represent the views of the NSF, IUPTI, IU, or the Lilly Endowment, Inc. Any and all mistakes are the sole responsibility of the senior author.

THANK YOU!

- Danke fuer Ihre Aufmerksamkeit
- Merci de votre attention
- Grazie per l'attenzione
- Questions?

- Follow @scinode on Twitter
- It's free and it's great



PERVASIVE TECHNOLOGY INSTITUTE

More about the Indiana University Pervasive Technology Institute (@IU_PTI on twitter)

- The Indiana University Pervasive Technology Institute (IUPTI) transforms new innovations in cyberinfrastructure and computer science into robust tools and supports the use of such tools in academic and private sector research and development. IUPTI does this while bolstering the Indiana Economy and building Indiana's 21st century workforce.
- IU_PTI is Indiana University's initiative for advanced information technology research, development, and delivery in support of scientific discovery, scholarly investigation, and artistic creation.
- Information technology today pervades scholarly discovery in the humanities, research in all areas of the sciences, and the processes of artistic creation. The "Pervasive" in the name IU Pervasive Technology Institute reflects the foundational importance of computer science, informatics, cyberinfrastructure, and information technology research to most of what is done in academia and industry today.
- The Indiana University Pervasive Technology Institute was created in 1999 by a major gift from the Lilly Endowment and persists today through a combination of competitively obtained federal funding, donations, and IU support.

References

[1] Stewart, C.A., R. Knepper, M.R. Link, M.A. Pierce, E. Wernert, N. Wilkins-Diehr. 2018. Cyberinfrastructure, Cloud Computing, Science Gateways, Visualization, and Cyberinfrastructure Ease of Use. Book Chapter. Pp. 1063-1073 In: Encyclopedia of Information Science and Technology, 4th Edition, Vol II. M. Khosrow-Pour (ed.). IGI Global, Hershey, PA. http://hdl.handle.net/2022/21589

[2] More details about Big Red II online at https://kb.iu.edu/d/bcqt

[3] More details about Jetstream online at Hancock*, D.Y., C.A. Stewart*, M. Vaughn, J. Fischer, J.M. Lowe, G. Turner, T.L. Swetnam, T.K. Chafin, E. Afgan, M.E. Pierce, W. Snapp-Childs. 2018. Jetstream—Early operations performance, adoption, and impacts. Concurrency and Computation Practice and Experience. 2018. e4683. https://doi.org/10.1002/cpe.4683

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total Investment (Local)

$$TI_{local} = AC + MC + TS + TPC + TCC$$

- *AC* = Acquisition Cost
- *MC* = Maintenance Cost
- *TS* = Total Systems Administration Costs
- *TPC* = Total Power Cost
- *TCC* = Total Co-Location Cost
- *NOTE: System PROGRAMMING time not included, as it must be done for clouds or for local systems.*

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total Local Investment (in USD)

$$TI_{local} = AC + MC + TS + TPC + TCC$$

TI_{local}

= $7,500,000 + $300,000

+ $517,429 + $1,5328,468

+ $286,200

= ***$10,132,097***

AC (Acquisition Cost)

= ***$7,500,000***

MC (Maintenance Cost)

= ***$300,000***

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total System Administration Salary

$$TS = (YS_1 + YS_1 \times FBR_1) \times 0.75 + \sum_{n=2}^{5} YS_n + YS_n \times FBR_n$$

TS

= (75K + 75K x 0.4334) x 0.75

+ 76K + 76K x 0.3976

+ 77K + 77K x 0.391

+ 79K + 79K x 0.391

+ 81K + 81K x 0.4023

= ***$517,429***

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total Power Cost

$$TPC = P_{avg} \times PUE_{avg} \times kWh_{avg} \times 24 \times 365.25 \times 4.75$$

TPC

= 379kW x 1.65 x $0.0587
x 24 x 365.25 (days / year)
x 4.75 (years)

= ***$1,528,468***

Total Colocation Cost

$$TCC = RCC \times NR \times n$$

TCC

= $4620 x 12 x 5 + $9000

= ***$286,200***

Calculating ROI on a locally-owned supercomputer: Big Red II - 2013-2017

Total Value of AWS as alternative

$$TV_{AWS} = CH \times IC$$

CH is the actual core hours utilized during this period on Big Red II, not the total of available hours.

IC=Instance Cost

PERVASIVE TECHNOLOGY INSTITUTE

Total Value of AWS (3-year reserved rate)

$\mathbf{TV_{AWS}}$ = 637,874,648 x $0.039

= ***$24,877,111***

Total Value of AWS (1-year reserved rate)

$\mathbf{TV_{AWS}}$ = 637,874,648 x $0.059

= ***$37,634,604***

Total Value of available GPU hours (3-year reserved rate)

$\mathbf{TV_{AWS}}$ = 25,675,766 x $0.399

= ***$10,244,635***