High Performance Computing System Acquisition: Jetstream – A Self-Provisioned, Scalable Science and Engineering Cloud Environment (Year 1 Annual Report)

Craig A. Stewart
Ian Foster
Nirav C. Merchant
James Taylor
Matthew W. Vaughn

Indiana University

PTI Technical Report PTI-TR16-001

Annual Report for Award 1445604: 12/01/2014 – 11/30/2015

February 15, 2016

Citation:

Stewart, C.A., Foster, I., Merchant, N.C., Taylor, J., & Vaughn, M.W. "High Performance Computing System Acquisition: Jetstream – A Self-Provisioned, Scalable Science and Engineering Cloud Environment (Year 1 Annual Report)," Indiana University, Bloomington, IN. PTI Technical Report PTI-TR16-001, February 2016.

Table of Contents

High Performance Computing System Acquisition: Jetstream – A Self-Provisioned, Scalable Science and Engineering Cloud Environment (Year 1 Annual Report) i

1. Accomplishments 1

1.1. What are the major goals of the project? 1

1.2. What was accomplished under these goals? 2

1.2.1. Major Activities 2

1.2.2. Specific Objectives: 2

1.2.3. Significant results 4

1.2.4. Key outcomes or Other achievements 5

1.3. What opportunities for training and professional development has the project provided? 5

1.4. How have the results been disseminated to communities of interest? 6

1.5. What do you plan to do during the next reporting period to accomplish the goals? 8

2. Products 9

2.1. Products resulting from this project during the specified reporting period 9

2.1.1. Journals 9

2.1.2. Conference Papers and Presentations 9

2.1.3. Other Publications 10

3. Participants 11

3.1. Individuals 11

3.1.1. Full details of individuals who have worked on the project: 12

3.2. Partner organizations 16

3.2.1. Full details on partner organizations 16

3.3. Have other collaborators or contacts been involved? 16

4. Impact 17

4.1. What is the impact on the development of the principal discipline(s) of the project? 17

4.2. What is the impact on other disciplines? 17

4.3. What is the impact on the development of human resources? 17

4.4. What is the impact on physical resources that form infrastructure? 18

4.5. What is the impact on institutional resources that form infrastructure? 18

4.6. What is the impact on information resources that form infrastructure? 18

4.7. What is the impact on technology transfer? 18

4.8. What is the impact on society beyond science and technology?..... 18

5. Changes/ Problems..... 19

5.1. Changes in approach and reasons for change 19

5.2. Actual or Anticipated problems or delays and actions or plans to resolve them 19

5.3. Changes that have significant impact on expenditures..... 19

5.4. Significant changes in use or care of human subjects 19

5.5. Significant changes in the use or care of vertebrate animals..... 19

5.6. Significant changes in the use or care of biohazards 20

1. Accomplishments

1.1. What are the major goals of the project?

As a first of a kind cloud resources funded by the NSF, Jetstream has already had significant impact on community research plans and expectations. Our extensive outreach and information dissemination activities have made many communities in science and engineering aware of the coming availability of Jetstream. This has caused several researchers to plan on it as a resource to use as part of their overall research activities, viewing Jetstream as an expected and essential part of research cyberinfrastructure supporting science, particularly research done "in the long tail of science." So far, Jetstream PI Craig Stewart has written seven letters of support for research proposals explaining the fit between Jetstream resources and the proposed research, promising to support requests for Jetstream resources to support the research, and committing resources from the PI's discretionary allocation in the unlikely event that researchers are unable to get the resources they need through the normal allocation processes. Jet stream's design characteristics and plans for mode of use have met with widespread approval, and we have every reason to believe that Jetstream as proposed and described in the PEP will prove a widely used and valuable resource for open US science and engineering research.

The purpose of the Jetstream computational resource is to ensure that the science and engineering community has ready access to the advanced computational and data-driven capabilities required to tackle today's most complex problems and issues. Jetstream will in particular complement previous NSF investments in advanced computational infrastructure by adding its first cloud environment for use in science and engineering research across all areas of research and education supported by the NSF. Jetstream will be a new type of computational research resource for the national open (unclassified) research community - a data analysis and computational resource that US scientists and engineers will use interactively. This system will enable many US researchers and engineers to make new discoveries that are important to understanding the world around us and will help researchers make new discoveries that improve the quality of life of American citizens.

The project deliverable will be the Jetstream system. Jetstream will be a configurable large-scale computing resource that leverages both on-demand and persistent virtual machine technology to support a much wider array of software environments and services than current NSF resources can accommodate. As a fully configurable "cloud" resource, Jetstream bridges the obvious major gap in the current ecosystem, which has machines targeted at large-scale High- Performance Computing, high memory, large data, high-throughput, and visualization resources. As the open cloud for science, Jetstream will:

- Provide "self-serve" academic cloud services, enabling researchers or students to select a VM image from a published library, or alternatively to create or customize their own virtual environment for discipline- or task-specific personalized research computing. Authentication to this "self-serve" environment will be via Globus.
- Host persistent Science Gateways. Jetstream will support persistent science gateways, including the capability of hosting persistent science gateways within a VM when the nature of the gateway is consistent with operation within a VM. Galaxy will be one of the initial science gateways supported.
    - Data movement, storage and dissemination.

    - Jetstream will support data transfer with Globus Connect.
    - Users will be able to store VMs in the Indiana University persistent digital repository, IUScholarWorks (scholarworks.iu.edu) and obtain a Digital Object Identifier (DOI) that is associated with the VM stored.
- Provide virtual Linux desktop services delivered from Jetstream to tablet devices. This service is aimed to increase access to Jetstream for users at institutions with limited resources including small schools, schools in EPSCoR states, and Minority Serving Institutions.

1.2. What was accomplished under these goals?

1.2.1. Major Activities

Major activities accomplished during the first program year in terms of project implementation include:

- Filled all staff positions funded during the initial construction award for Jetstream.
- Established necessary financial and management controls by appropriately starting.

1.2.2. Specific Objectives:

- Entries and ledgers in IU's standard financial systems, the standard financial systems of subcontractors, and by installing required software to enable collaboration.
- Completed a Program Execution Plan which was submitted to the National Science Foundation, sent out for scientific peer review, revised, and then formally submitted to the NSF as material to amend the Cooperative Services Agreement. This proposed amendment has been accepted and incorporated into the Continuing Service Agreement by the NSF.
- Completed purchase contracts with the major hardware vendor for Jetstream - Dell, Inc.
- Ordered, received, installed, and tested one of three components of the Jetstream computational hardware - the system that will be known as Jetstream-Arizona. An acceptance report was submitted to the National Science Foundation which has declared acceptance of this component of the overall Jetstream hardware.
- Ordered and received the two main processing components of the Jetstream computational and storage hardware, which will eventually be known as Jetstream- Indiana and Jetstream-Texas. Preliminary acceptance tests (burn-in) have been run and OpenStack software deployment began in November. Operations staff decided in November to begin adoption of the latest OpenStack release "Liberty" after evaluating its merits for use by the project. This will enable us to more easily keep pace with future OpenStack releases, setting us up for easier in-place upgrades.
- Formally submitted two requests for supplemental management and operations funding to the NSF.
- Held periodic team management meetings including a minimum of one "all-site" management meeting per month, and one "all-site" technical meetings per month. Communicated regularly with relevant NSF staff and submitted monthly and quarterly reports in a timely fashion.
- Developed an Allocations Position Paper that documents Jetstream's allocations policies that are unique to the XSEDE environment, and these were then introduced to the XSEDE XRAC members to assist them with properly allocating Jetstream.
- Initially, the project has committed to allocate primarily on the basis of vCPU hours, and memory will be tied to vCPU hours. Public IPs will not be allocable, but will be available in limited

numbers for user VMs. While there has not been a formal decision on a storage allocation policy, this will be developed and implemented in early 2016.

- Developed an effective collaborative organization. During the initial implementation phase of Jetstream the parties with key operational responsibilities have included the lead institution, the Indiana University Pervasive Technology Institute, the Texas Advanced Computing Center, and strong and active involvement by the University of Arizona, Chicago University (Computational Institute), and Johns Hopkins University. We have also had ongoing engagement with partner institutions that are slated for funding during the Management and Operations phase or are unfunded participants, so that as the Jetstream project moves forward we will be able to do so effectively and efficiently.

Accomplishments relevant to achievement of goals for this project are described below:

- Provide "self-serve" academic cloud services, enabling researchers or students to select a VM image from a published library, or alternatively to create or customize their own virtual environment for discipline- or task-specific personalized research computing. Authentication to this “self-serve” environment will be via Globus-Auth.
- We have created an operational instance of the Jetstream software stack (CentOS, OpenStack, KVM, Atmosphere Interface) operating on the Jetstream-Arizona system.
- The University of Arizona (UA) deployed Atmosphere onto Indiana University's infrastructure and integrated with Jetstream's OpenStack Kilo test cluster and Globus' OAuth2 services. The integration required changes to multiple Atmosphere components, including authentication services, Atmosphere's provisioning services, and configuration management. Globus integration enabled Atmosphere authentication using XSEDE identities. In addition to Atmosphere-specific components, UA modified the code for existing open source projects, namely Apache Libcloud and Liftoff Software's Gate One web-based terminal software. These code changes were submitted back to the open source community.
- The publically accessible VM library is so far populated with a small set of preconfigured virtual machine (VM) images. These VM images have come from a variety of sources: VMs re-used from iPlant, VMs contributed by students funded by Indiana University, and VMs developed around scientific codes contributed by the scientific community.
- TACC, the University of Chicago, and University of Arizona have been engaged in the development of an Identity Management capability that meets both the local requirements/policies of each IU and UT-Austin/TACC, while providing the user experience described in the proposal and PEP. We have an LDAP-replication process that will allow IU to clone TACC’s accounting information to their own local copy. A work plan for implementing a ‘Request Access to Jetstream’ button in the XSEDE user portal was devised in collaboration with the XUP team.
- One key project goal was to have live demonstrations of Jetstream in operation at the IEEE/ACM SC15 conference. This goal was accomplished. Staff and students of the University of Arizona, Indiana University, and Texas Advanced Computing Center gave live demos repeatedly during the week of SC15. Perhaps most impressively a graduate student from SUNY Binghamton demonstrated a VM that he had packaged and implemented on Jetstream under the supervision of his advisor. Neither had had any funding from the Jetstream project, and the research done is contributing toward the Pd.D. research of the graduate student.
- Host persistent Science Gateways. Jetstream will support persistent science gateways, including the capability of hosting persistent science gateways within a VM when the nature of the gateway is consistent with operation within a VM. Galaxy will be one of the initial science gateways supported.
- We have operated test instances of Science Gateways on Jetstream, and have also installed and operated Galaxy on other IU and TACC resources so that when sufficient computational

capability is available for use running the Jetstream software stack we will be able to install and operate Galaxy in relatively short order.

- Data movement, storage and dissemination. Jetstream will support data transfer with Globus Connect, and users will be able to store VMs in the Indiana University persistent digital repository, IUScholarWorks (scholarworks.iu.edu) and obtain a Digital Object Identifier (DOI) that is associated with the VM stored.
- TACC, U. Chicago, and University of Arizona have led work to manage data transfer with Globus Connect. IU has established a workflow for submission of a VM to the IU
- Digital repository and receiving a DOI associated with that digital object.
- Provide virtual Linux desktop services delivered from Jetstream to tablet devices. This service is aimed to increase access to Jetstream for users at institutions with limited resources including small schools, schools in EPSCoR states, and Minority Serving Institutions.
- IU has prepared and demonstrated a virtual desktop accessible via "pad" devices over wireless networks, and has demonstrated that its performance is appropriate for student use.

1.2.3. Significant results

While Jetstream is still in its implementation phase, we have the following significant technical results so far:

- We have actually implemented and demonstrated features that demonstrate in action three of the four major objectives. (The fourth objective can be demonstrated fully only when multiple sites of Jetstream are in operation).
- The computational efficiency as measured by the High Performance Linpack (HPL) benchmark of the Jetstream software stack is quite high for a cloud resource; only a 3% loss in computational efficiency as compared to operating on a "bare metal" Linux kernal. Other benchmarks are not this efficient; especially those that involve memory or network traffic, as shown below:

Table 1. HPCC results

Benchmark	**Computational Efficiency**
HPL	97%
Ptrans	64%
MPIFFT	67%
Random	80%
Stream	77%
DGEMM	98%
Bandwidth	88%
Latency	97%

1.2.4. Key outcomes or Other achievements

A virtual desktop delivered to a PDA via wireless connection has been implemented and proved good enough for use in research education.

In addition to the technical results described we have achieved and observed the other key outcomes:

- A cloud resource operated by the NSF has been well received by the community of researchers working in the areas of science and engineering central to the NSF's charge, as well as researchers supported by the NIH and USDA.
- Within the context of many services and functionalities that could be offered in a first-of-a-kind cloud resource funded by the NSF, the particular choices and priorities set in Jetstream functionality have resonated well with the user community, particularly scientists, engineers, and students whose work falls into the "long tail of science" and scientists, engineers, and students at institutions with limited cyberinfrastructure budgets, HVCUs, institutions serving large populations of students from Traditionally Underserved Groups (as defined by the NSF), and institutions in EPSCoR states.

1.3. What opportunities for training and professional development has the project provided?

Training activities have been thus far been primarily for individuals involved in the project and students generally.

Training activities for Jetstream project staff have included:

- During Q1 of the acquisition phase of the project, a technical team kickoff meeting was held December 16 and 17 in Tucson, AZ. This included brief tutorials on several of the technical aspects of Jetstream software integration.
- Also during Q1, an "all partner" meeting was held January 20-22 in Tucson, AZ. The two and a half day agenda allowed for technical discussion on the first day. The second day included presentations by all major software partners including brief tutorials on many of the key software elements, as well as demonstrations of the software to be included in the system integration. Project collaborators each discussed their anticipated use of the system and what research activities they hoped to achieve.
- Time has been allocated and expended for technical staff to study about and experiment with the software components required to create Jetstream as proposed to and funded by the National Science Foundation.

Training activities related to students have included:

- Cloud software technology. During the summer Joseph Cavazos, a Hispanic undergraduate computer science student, conducted a feasibility study on container virtualization and application micro-services architecture as a method for packaging and deploying engineering and scientific applications on multiple OpenStack Clouds using Docker container and container cluster managers such as Swarm and Kubernetes. The study also included the packaging and deployment of several applications such as web technologies, machine learning, and data analytic programs. A report documenting best practices and caveats that were discovered or encountered during the experimental study is available online at https://github.com/cloudandbigdatalab/cloudandbigdatalab.github.io/blob/master/pdf/docker-kubernetes.pdf. Since NSF indicated a lack of interest in funding any REU projects associated

with Jetstream during calendar 2015, this activity was funded by Indiana University as part of its facilities commitment to making this project overall a success.

- Network Workbench implementation within a VM. One of the applications that we initially proposed be made available via Jetstream was the Network WorkBench (http://nwb.cns.iu.edu(Du8ring the summer two graduate students in the lab of Professor Katy Boerner - Kartik Adur and David Ebenezer – implemented the NWB tools in VMs running on the Jetstream test system. Since NSF indicated a lack of interest in funding any student work associated with Jetstream during calendar 2015, this activity was funded by Indiana University as part of its facilities commitment to making this project overall a success.
- Graduate Student Renan DelValle (SUNY-Binghamton) implemented Apache Airavata, a multi-tenanted science gateway framework, Docker, and Apache Mesos to support data analysis software created by the National Snow and Ice Data Center. Analysis of NSIDC data was one of the use cases driving the design of Jetstream.
- Co-PI Vaughn identified a specific early user who will access the system even before it is open to formal Early User access to develop a Docker-based data science classroom curriculum.

1.4. How have the results been disseminated to communities of interest?

Results have been communicated to communities of interest in the following ways:

Published papers in peer-reviewed technical conferences:

Jeremy Fischer, Steven Tuecke, Ian Foster, and Craig A. Stewart. 2015. Jetstream: A Distributed Cloud Infrastructure for Under-resourced higher education communities. In Proceedings of the 1st Workshop on The Science of Cyberinfrastructure: Research, Experience, Applications and Models (SCREAM '15). ACM, New York, NY, USA, 53- 61. DOI=http://dx.doi.org/10.1145/2753524.2753530

Craig A. Stewart, Timothy M. Cockerill, Ian Foster, David Hancock, Nirav Merchant, Edwin Skidmore, Daniel Stanzione, James Taylor, Steven Tuecke, George Turner, Matthew Vaughn, and Niall I. Gaffney. 2015. Jetstream: a self-provisioned, scalable science and engineering cloud environment. In Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure (XSEDE '15). ACM, New York, NY, USA, , Article 29 , 8 pages. DOI=10.1145/2792745.2792774 http://doi.acm.org/10.1145/2792745.2792774 Hancock, D.Y., M.R. Link, and C.A. Stewart. "Acceptance Test for Jetstream Test Cluster – Jetstream-Arizona (JA)

Dell PowerEdge Test and Development Cluster," Indiana University, Bloomington, IN. PTI Technical Report PTI-TR15- 007, Jan 2015. Available at http://hdl.handle.net/2022/20355

Presentations:

Stewart, C.A. 2015. Jetstream Overview – XSEDE '15 Panel - New and emerging US cyberinfrastructure resources. Presentation. XSEDE'15, July 26 - 30, 2015, St. Louis, MO, USA. http://hdl.handle.net/2022/20338. Jetstream is supported by NSF award 1445604 (Craig Stewart, PI)

Stewart, C.A. 2015. Jetstream Overview – XSEDE '15 Panel - New and emerging US cyberinfrastructure resources. Presentation. XSEDE'15, July 26 - 30, 2015, St. Louis, MO, USA. http://hdl.handle.net/2022/20322

Stewart, Craig A. (2015). Great National Cyberinfrastructure Resources for Great Kentucky Researchers and Students.. Presentation to the EPSCoR Kentucky Science and Technology Corporation Conference

Vasiliadis, Vas, Stewart, Craig A. (2015). Jetstream A national science & engineering cloud: Jetstream Overview - EGI 2015. Presentation on Jetstream system to EGI 2015 Conference in Lisbon, Portugal.

Stewart, Craig A. (2015). Jetstream Overview. Jetstream overview presentation for XRAC (Large Resource Allocations Committee (LRAC) and Medium Resource Allocations Committee (MRAC).

Seiffert, Kurt A. (2015). Jetstream: A national science & engineering cloud. Globus World Lightning Talk.. Presentation highlighting the capabilities and purpose for the upcoming Jetstream system.

Hancock, David Y.; Stewart, Craig A. (2015). Jetstream: A national science & engineering cloud. HPC User Forum. The presentation describes Jetstream capabilities and use applications across science domains.

Stewart, Craig A (2015). Jetstream: A science & engineering cloud Prepared for XSEDE SP Forum. Presentation to the XSEDE SP Forum

Stewart, Craig A (2015). "Big Data: Where can EPSCoR states use big data and what tools do EPSCoR states need to thrive?" Presentation to EPSCoR meeting - http://hdl.handle.net/2022/19210

Stewart, Craig A (2015). Jetstream: A science & engineering cloud. Prepared for Polar Workshop. http://hdl.handle.net/2022/19703

Stewart, Craig A (2015). Jetstream: A national science & engineering cloud - Prepared for the IUB Faculty Council Information Technology Subcommittee. http://hdl.handle.net/2022/19700

Stewart, Craig A (2015). Jetstream: A national science and engineering cloud (University of Vermont). Presentation to EPSCoR meeting. http://hdl.handle.net/2022/20439

Fischer, Jeremy L (2015). Jetstream: A new national research and education cloud - Prepared for Statewide 15.

Presentation to Indiana University system-wide IT. http://hdl.handle.net/2022/20431

Fischer, Jeremy L (2015). Jetstream: A new national research and education cloud - Prepared for SPAN Workshop 2.

Presentation to Southern Partnership in Advanced Networking group. http://hdl.handle.net/2022/20446
Stewart, Craig A. (2015). Jetstream status and updates: It's alive! Presentation at SC15. http://hdl.handle.net/2022/20508

Meetings with research groups interested in being future users of Jetstream Network Workbench Tools (http://nwb.cns.iu.edu/)

BrownDog project (http://browndog.ncsa.illinois.edu/) Galaxy (https://usegalaxy.org/)

Unidata (http://www.unidata.ucar.edu/) IRIS (http://www.iris.edu/)

Polar HPDC / National Snow and Ice Data Center (http://nsidc.org/) EPSCoR Bioinformatics Advisory Committee

University of Vermont – Dr. David Krag lab. EarthCube/CHORDS - http://earthcube.org/group/chords Montana State University

Monash University BTVIgnite

SciGaP

Southern Partnership in Advanced Networking Live demonstrations at SC15

Scheduled presentations in IU display

Jump on Jetstream (Craig Stewart, IU) – Monday, November 16, 2015

Jetstream UI demo: science and engineering in the cloud made simple (Edwin Skidmore, UA) – Tuesday, November 17, 2015

Cloud infrastructure for science gateways and science applications on Jetstream (Madhusudhan Govindaraju and Renan DelValle, SUNY-Binghamton, Marlon Pierce and Suresh Marru, IU) – Tuesday, November 17, 2015 Jetstream UI demo: science and engineering in the cloud made simple (Craig Stewart, IU) – Wednesday, November 18, 2015

Cloud infrastructure for science gateways and science applications on Jetstream (Madhusudhan Govindaraju and Renan DelValle, SUNY-Binghamton, Marlon Pierce and Suresh Marru, IU) – Wednesday, November 18, 2015 Jetstream UI demo: science and engineering in the cloud made simple (Sangeeta Kuchimanchi, UA) – Wednesday, November 18, 2015

PI Vaughn provided seven ad hoc demonstrations of launching a VM via the Jetstream User Interface to interested parties, including the technical leads of the Open Commons Consortium.

Birds of a Feather sessions

XSEDE'15 Conference July 2015, St. Louis, MO

SC15 Conference November 2015, Austin, TX

1.5. What do you plan to do during the next reporting period to accomplish the goals?

Our key goal during Program Year 2 is to transition Jetstream into fully operational mode.

Particular technical highlights of this work are as follows:

ID Management: Verify the trust relationship functions for OpenStack Liberty. Identify and coordinate the changes to the

XSEDE User Portal needed to enable easy-on access to Jetstream. Hardware: Finish hardware acceptance tests, verifying all functionality.

Software: Deploy OpenStack cloud software such that it may be used by scientific users, and integrate iRods and Globus Connect software movement as needed.

Allocations management: Determine a policy for cycles and storage for the project PI to take to the XRAC, thus enabling allocations in Q12016.

Administrative aspects of this work include:

Completing acceptance reports and having an on site review as input to the NSF preparatory to an NSF decision to accept the Jetstream resource.

Work with NSF to complete action on two supplemental funding requests already pending

Submit for NSF consideration a supplemental request for Research Experiences for Undergraduates related to Jetstream for the summer of 2016

Training aspects of this work include:

Train XSEDE staff to be able to provide effective front-line user support and extended collaborative services related to Jetstream

Continue to train staff and collaborators associated with the Jetstream project itself Train new users about capabilities of and effective use of Jetstream.

User service and access elements of this work include:

Make Jetstream available for community use in "Early Operations Mode" on or about 20 January 2016

Make Jetstream available for community use in fully operational mode sometime during Q1 or early Q2 of calendar 2016.

2. Products

2.1. Products resulting from this project during the specified reporting period

2.1.1. Journals

2.1.2. Conference Papers and Presentations

Craig A. Stewart (2014). "Jetstream: A national science and engineering cloud". IEEE/ACM Supercomputing Conference. New Orleans, LA. Status = OTHER; Acknowledgement of Federal Support = Yes

Fischer, Jeremy and Tuecke, Steven and Foster, Ian and Stewart, Craig A. (2015). Jetstream: A Distributed Cloud Infrastructure for Underresourced Higher Education Communities. Proceedings of the 1st Workshop on The Science of Cyberinfrastructure: Research, Experience, Applications and Models. New York, NY, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, Craig A. and Cockerill, Timothy M. and Foster, Ian and Hancock, David and Merchant, Nirav and Skidmore, Edwin and Stanzione, Daniel and Taylor, James and Tuecke, Steven and Turner, George and Vaughn, Matthew and Gaffney, Niall

I. (2015). Jetstream: A Self-provisioned, Scalable Science and Engineering Cloud Environment. Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure. New York, NY, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). XSEDE '15 Panel - New and emerging US cyberinfrastructure resources. XSEDE '15. St. Louis, MO. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). XSEDE ’15 - New and emerging US cyberinfrastructure resources. XSEDE '15. St. Louis, MO. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

2.1.3. Other Publications

Hancock, D.Y., M.R. Link, and C.A. Stewart (2015). Acceptance Test for Jetstream Test Cluster – Jetstream-Arizona (JA) Dell

PowerEdge Test and Development Cluster,. PTI Technical Report PTI-TR15-007. Available at http://hdl.handle.net/2022/20355. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Big Data: Where can EPSCoR states use big data and what tools do EPSCoR states need to thrive?. Presentation to EPSCoR meeting. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Great National Cyberinfrastructure Resources for Great Kentucky Researchers and Students. Presentation to the EPSCoR Kentucky Science and Technology Corporation Conference. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Vasiliadis, V., Stewart, C. A. (2015). Jetstream A national science & engineering cloud: Jetstream Overview for EGI 2015. Presentation on Jetstream system to EGI 2015 Conference in Lisbon, Portugal. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Jetstream Overview. Jetstream overview presentation for XRAC (Large Resource Allocations Committee (LRAC) and Medium Resource Allocations Committee (MRAC). Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Jetstream status and updates: It’s alive!. Presentation at SC15. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Jetstream: A national science & engineering cloud. Prepared for the IUB Faculty Council Information Technology Subcommittee. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Seiffert, Kurt A. (2015). Jetstream: A national science & engineering cloud. Globus World Lightning Talk. Presentation highlighting the capabilities and purpose for the upcoming Jetstream system. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Hancock, D.Y, Stewart, C.A. (2015). Jetstream: A national science & engineering cloud. HPC User Forum. The presentation describes Jetstream capabilities and use applications across science domains. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Jetstream: A national science and engineering cloud. Presentation to EPSCoR meeting at University of Vermont. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Fischer, J.L. (2015). Jetstream: A new national research and education cloud - Prepared for SPAN Workshop 2. Presentation to Southern Partnership in Advanced Networking group. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Fischer, J.L., Miller, T.M. (2015). Jetstream: A new national research and education cloud - Prepared for Statewide 15. Presentation to Indiana University system-wide IT. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Stewart, C.A. (2015). Jetstream: A science & engineering cloud ¬ Prepared for XSEDE SP Forum. Presentation to the XSEDE SP Forum. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Craig A. Stewart (2014). "Big Data: Where can EPSCoR states use big data and what tools do EPSCoR states need to thrive?". This invited presentation was given to the EPSCoR/IDeA Foundation on December 4, 2014.. Status = OTHER; Acknowledgement of Federal Support = Yes

3. Participants

3.1. Individuals

Name	Most Senior Project Role	Nearest Person Month Worked
Stewart, Craig	PD/PI	4
Foster, Ian	Co PD/PI	1
Merchant, Nirav	Co PD/PI	1
Taylor, James	Co PD/PI	1
Vaughn, Matthew	Co PD/PI	3
Hammond, Bret	Other Professional	7
Lowe, John Michael	Other Professional	6
Miller, Therese	Other Professional	1
Carlson, Tharon	Other	1

Cockerill, Tim	Other	2
Fischer, Jeremy	Other	3
Gregory, Steve	Other	3
Hancock, Dave	Other	4
Hanlon, Matt	Other	1
Lenards, Andy	Other	1
Liming, Lee	Other	1
Packard, Mike	Other	4
Seth, Akhill	Other	1
Skidmore, Edwin	Other	2
Tuecke, Steve	Other Professional	1
Turner, George	Other Professional	11

3.1.1. Full details of individuals who have worked on the project:

Craig A Stewart Email: stewart@iu.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 4
Contribution to the Project: Principal Investigator and project manager responsible for system acquisition, integration and implementation.
Funding Support: NSF Award 1445604
International Collaboration: No
International Travel: No

Ian Foster
Email: foster@uchicago.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1
Contribution to the Project: Co-Principal Investigator responsible for system software integration.
Funding Support: Subaward on NSF Award 1445604 to Indiana University
International Collaboration: No
International Travel: No

Nirav C Merchant
Email: nirav@email.arizona.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1
Contribution to the Project: Co-Principal Investigator responsible for software testing and integration for the Jetstream system.
Funding Support: Subaward on NSF Award 1445604 to Indiana University
International Collaboration: No

International Travel: No

James Taylor
Email: james@taylorlab.org
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1
Contribution to the Project: Co-Principal Investigator responsible for software integration during the operations and maintenance phase of the project.
Funding Support: Subaward on NSF Award 1445604 to Indiana University
International Collaboration: No
International Travel: No

Matthew W Vaughn
Email: vaughn@tacc.utexas.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 3
Contribution to the Project: Co-PI on the project responsible for system integration and implementation.
Funding Support: Subaward on NSF award 1445604 to Indiana University
International Collaboration: No
International Travel: No

Bret Hammond
Email: bret@iu.edu
Most Senior Project Role: Other Professional
Nearest Person Month Worked: 7
Contribution to the Project: System Administration
Funding Support: Indiana University
International Collaboration: No
International Travel: No

John Michael Lowe
Email: jomlowe@iu.edu
Most Senior Project Role: Other Professional
Nearest Person Month Worked: 6
Contribution to the Project: System Administration
Funding Support: Indiana University
International Collaboration: No
International Travel: No

Therese Miller
Email: millertm@iu.edu
Most Senior Project Role: Other Professional
Nearest Person Month Worked: 1
Contribution to the Project: Project Manager
Funding Support: Indiana University
International Collaboration: No
International Travel: No

Tharon Carlson
Email: tharon@iplantcollaborative.org
Most Senior Project Role: Other

Nearest Person Month Worked: 1
Contribution to the Project: Developer
Funding Support: University of Arizona
International Collaboration: No

Tim Cockerill
Email: cockerill@tacc.utexas.edu Most Senior Project Role: Other
Nearest Person Month Worked: 2
Contribution to the Project: Project management
Funding Support: University of Texas Austin
International Collaboration: No
International Travel: No

Jeremy Fischer Email: jeremy@iu.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 3
Contribution to the Project: Outreach, XSEDE integration, Administrative
Funding Support: Indiana University
International Collaboration: No
International Travel: No

Steve Gregory
Email: esteve@iplantcollaborative.org
Most Senior Project Role: Other
Nearest Person Month Worked: 3
Contribution to the Project: Senior software engineer
Funding Support: University of Arizona
International Collaboration: No
International Travel: No

Dave Hancock
Email: dyhancoc@iu.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 4
Contribution to the Project: Senior Investigator, Technical Lead

Matt Hanlon
Email: mrhanlon@tacc.utexas.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 1
Contribution to the Project: Developer
Funding Support: University of Texas Austin
International Collaboration: No
International Travel: No

Andy Lenards
Email: lenards@iplantcollaborative.org
Most Senior Project Role: Other
Nearest Person Month Worked: 1
Contribution to the Project: Software Developer
Funding Support: University of Arizona

International Collaboration: No
International Travel: No

Lee Liming
Email: lliming@uchicago.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 1
Contribution to the Project: Project management
Funding Support: University of Chicago
International Collaboration: No
International Travel: No

Mike Packard
Email: mpackard@tacc.utexas.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 4
Contribution to the Project: System Administration
Funding Support: University of Texas Austin
International Collaboration: No
International Travel: No

Akhil Seth
Email: akhil@tacc.utexas.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 1
Contribution to the Project: Developer
Funding Support: University of Texas Austin
International Collaboration: No
International Travel: No

Edwin Skidmore
Email: edwin@iplantcollaborative.org
Most Senior Project Role: Other
Nearest Person Month Worked: 2
Contribution to the Project: Project lead responsible for coordination of iPlant and Atmosphere software stack
Funding Support: University of Arizona
International Collaboration: No
International Travel: No

Steve Tuecke
Email: tuecke@uchicago.edu
Most Senior Project Role: Other
Nearest Person Month Worked: 1
Contribution to the Project: System software implementation
Funding Support: University of Chicago
International Collaboration: No
International Travel: No

George Turner
Email: turnerg@iu.edu

Most Senior Project Role: Other
Nearest Person Month Worked: 11
Contribution to the Project: Chief Architect
Funding Support: Indiana University
International Collaboration: No
International Travel: No

3.2. Partner organizations

Name	Type of Partner Organization	Location
University of Colorado/CIRES (Brian Johnson)	Academic Institution	Boulder, CO
University of Texas, San Antonio (Paul Rad)	Academic Institution	San Antonio, TX

3.2.1. Full details on partner organizations

3.2.1.1. University of Colorado/CIRES

Partner's Contribution to the Project:

- Facilities
- Collaborative Research

More Detail on Partner and Contribution:

Brian Johnson (University of Colorado-National Snow and Ice Data Center) has agreed to participate as a research collaborator using the Jetstream system for research in the polar science area.

3.2.1.2. University of Texas at San Antonio

Partner's Contribution to the Project:

- Collaborative Research

More Detail on Partner and Contribution:

Paul Rad (UTSA and Rackspace) will contribute software expertise for system integration.

3.3. Have other collaborators or contacts been involved?

No

4. Impact

4.1. What is the impact on the development of the principal discipline(s) of the project?

Within computational science and the science of cyberinfrastructure the impact so far has been primarily in orienting community discussion around Jetstream as a "first-of-a-kind" federally funded cloud resource. This has led to considerable community discussion on the following issues:

- Best practices in cloud environment software stacks
- Best practices in accounting and allocation in management of federally-funded software (resulting in a position paper on this topic)
- Considerable discussion within XSEDE and other projects such as the Open Science Grid and AQCI-REF about priorities and need in cloud computing
- Consensus agreement that as published and demonstrated so far, Jetstream appears to be designed in a way that will be a significant addition to the current suite of NSF-funded advanced cyberinfrastructure resources available to support open science and engineering research in the US
- Technical papers about Jetstream have been widely read (more than 100 downloads on the primary XSEDE Jetstream paper according to ResearchGate statistics).

4.2. What is the impact on other disciplines?

The primary impact on disciplines other than computational science and the science of cyberinfrastructure is to change the community dialog about cloud resources from "when oh when will the NSF fund a cloud resource" to "The NSF has funded a cloud resource and it's on the way; let's get ready." There has been significant interest in many disciplines of science and engineering in the upcoming availability of Jetstream. Researchers in many disciplines are actively working with the Jetstream team so as to be positioned to take advantage of Jetstream as soon as it becomes available for use.

4.3. What is the impact on the development of human resources?

The Jetstream project has provided opportunities for research, teaching and mentoring in science and engineering areas as follows:

Joseph Cavazos, a Hispanic undergraduate computer science student, did research and education in summer of 2015 related to cloud technology.

IU masters students Kartik Adur and David Ebenezer – implemented software within VMs during the summer of 2015. Graduate Student Renan DelValle (SUNY-Binghamton) implemented novel applications in Docker and Apache cloud environment tools under the supervision of his graduate studies advisor Madhusudhan Govindaraju (associate professor of computer science at SUNY-Binghamton).

The Jetstream project implementation under NSF funding and facilities provided by participating organizations improved the performance, skills, or attitudes of members of underrepresented groups that will improve their access to or retention in research, teaching, or other related professions as follows:

The Jetstream staff include members of traditionally underserved groups and all staff have had their expertise and breadth of knowledge.

The Jetstream project has new educational materials related to use of cloud computing and published these materials on the web as part of the Jetstream-cloud web site.

The Jetstream team has provided exposure to science and technology for practitioners, teachers, young people, or other members of the public via several articles in Science Node (formerly international Science Grid this Week) (science node)

4.4. What is the impact on physical resources that form infrastructure?

Not applicable.

4.5. What is the impact on institutional resources that form infrastructure?

Institutional resources in the form of licensed software and trained staff have been enhanced at all five of the partners involved in the initial construction of Jetstream.

4.6. What is the impact on information resources that form infrastructure?

Jetstream is at this point operational (although not widely available) and constitutes a new information resource for the US. Were there a civil emergency of a nature such that the use of Jetstream's computational and data analysis capabilities could be put to use to aid response to such an emergency, the Jetstream team could have the system operational and aiding the US government in 24 hours.

4.7. What is the impact on technology transfer?

There have been no formal invention disclosures thus far. There has been significant transfer of information between the Jetstream team and our corporate partners – Primarily Dell, but also DDN and other computational system integrators. We have met with partners and non-partners alike to help them understand the use cases and reasoning behind the system design, and help them understand our planned use cases and modalities. Every vendor we have met with (Dell, DDN, HP, IBM, Lenovo, Cray, Adaptive Systems, and Microsoft) has stated that our interactions with them have helped them better understand current and future needs and technology solutions in cloud technology.

4.8. What is the impact on society beyond science and technology?

Nothing to report

5. Changes/ Problems

5.1. Changes in approach and reasons for change

None

5.2. Actual or Anticipated problems or delays and actions or plans to resolve them

Unanticipated additions to the administrative processes implemented since the award was initially executed have somewhat slowed initial activities during PY1.

The amount of time it has taken to have documents and award modifications processed through the Division of Grants and Awards has been problematic. If DGA is in the future not able to act on paperwork in no more than approximately 3 months this will constitute a significant risk for successful project completion, one that was not anticipated when the risk registry was initially created.

The current budgetary uncertainty regarding XSEDE2 is at present the largest risk to success of this project overall. At NSF direction and not subject to the proposing organization's discretion, the Management and Operations budget assumes the availability of ample assistance from human and technical resources funded by XSEDE and XSEDE2 (the proposed successor to XSEDE). If the funding for the XSEDE2 project is cut to an extent that makes it impossible for XSEDE2 to be effective in delivering the services the proposing institutions have been instructed to depend upon, then the Jetstream budget will have to be increased or the project will ultimately fail to achieve its stated objectives for lack of NSF funding support.

5.3. Changes that have significant impact on expenditures

None

5.4. Significant changes in use or care of human subjects

Not applicable

5.5. Significant changes in the use or care of vertebrate animals

Not applicable

5.6. Significant changes in the use or care of biohazards

Not applicable