# Jetstream (NSF Award 1445604) Year Program Year 2 Annual Report (Dec 1, 2015 – Nov 30, 2016)

*Craig A. Stewart*[1]
*David Y. Hancock*[1]
*Matthew Vaughn*[2]
*Nirav Merchant*[3]
*J. Michael Lowe*[1]
*Jeremy Fischer*[1]
*Lee Liming*[4]
*James Taylor*[5]
*Enis Afgan*[5]
*George Turner*[1]
*C. Bret Hammond*[1]
*Edwin Skidmore*[3]
*Michael Packard*[2]
*Ian Foster*[4]

[1]Indiana University Pervasive Technology Institute
[2]University of Texas at Austin, Texas Advanced Computing Center
[3]University of Arizona
[4]University of Chicago Computation Institute
[5]Johns Hopkins University

December 31, 2016

Stewart, C.A., Hancock, D.Y., Vaughn, M., Merchant, N., Lowe, J.M., Fischer, J., Liming, L., Taylor, J., Afgan, E., Hammond, C.B., Skidmore, E., Foster, I. (2016). Jetstream (NSF Award 1445604) Year Program Year 2 Annual Report (Dec 1, 2015 – Nov 30, 2016) (PTI Technical Report PTI-TR16-008). Bloomington, IN: Indiana University. Retrieved from http://hdl.handle.net/2022/21194

**Table of Contents**

**Jetstream (NSF Award 1445604) Year Program Year 2 Annual Report (Dec 1, 2015 – Nov 30, 2016)** ..... **i**
**1. Introduction and executive summary** ..... **1**
**2. Accomplishments** ..... **2**
***2.1. What are the major goals of the project?*** ..... ***2***
***2.2. What was accomplished under these goals?*** ..... ***3***
***2.3. What opportunities for training and professional development has the project provided?*** ..... ***14***
***2.4. How have the results been disseminated to communities of interest?*** ..... ***14***
***2.5. What do you plan to do during the next reporting period to accomplish the goals?*** ..... ***21***
**3. Products** ..... **23**
**4. Participants/Organizations** ..... **27**
***4.1. Individuals*** ..... ***27***
***4.2. Partner organizations*** ..... ***28***
***4.3. Have other collaborators or contacts been involved?*** ..... ***31***
**5. Impacts** ..... **31**
***5.1. What is the impact on the development of the principal discipline(s) of the project?*** ..... ***31***
***5.2. What is the impact on other disciplines?*** ..... ***32***
***5.3. What is the impact on the development of human resources?*** ..... ***32***
***5.4. What is the impact on physical resources that form infrastructure?*** ..... ***33***
***5.5. What is the impact on institutional resources that form infrastructure?*** ..... ***33***
***5.6. What is the impact on information resources that form infrastructure?*** ..... ***33***
***5.7. What is the impact on technology transfer?*** ..... ***33***
***5.8. What is the impact on society beyond science and technology?*** ..... ***33***
**6. Changes/Problems** ..... **34**
***6.1. Changes in approach and reasons for change*** ..... ***34***
***6.2. Actual or Anticipated problems or delays and actions or plans to resolve them 34***
***6.3. Changes that have significant impact on expenditures*** ..... ***35***
***6.4. Significant changes in use or care of human subjects*** ..... ***35***
***6.5. Significant changes in the use or care of vertebrate animals*** ..... ***35***
***6.6. Significant changes in the use or care of biohazards*** ..... ***35***
**7. References** ..... **36**

## 1. Introduction and executive summary

The National Science Foundation (NSF) has invested heavily in delivering supercomputer resources, the vast majority of which have always been and are today supercomputers and large scale computing clusters, for use by the national US open science and engineering research community for several decades. As a result, many innovations, discoveries, improvements in the quality of life for Americans, new technology spurring new businesses and economic growth, and two Nobel prizes, have been produced. In the past three years, however, the NSF made significant investments in interesting and novel cyberinfrastructure facilities including cloud-based systems. These new CI systems significantly broaden the base of resources funded by the NSF and available to the national research community. Specifically, as stated by the NSF, the goals are: to support the "entire spectrum of NSF-funded communities ... and to promote a more comprehensive and balanced portfolio that in turn supports the entire scientific, engineering and educational community."

The Jetstream system (funded in 2014) and Wrangler (funded in 2013) are novel cyberinfrastructure systems funded by the NSF and supported by XSEDE. Jetstream is in fact so novel that it was specifically labeled as a pilot project. Indiana University and its partners focused on the following key factors in designing and proposing the Jetstream system:

- Incorporation of current best practices in social science understanding of technology adoption. In particular, we took into account current understandings that suggest technology adoption is driven by: performance expectancy (perceived value), effort expectancy (perceived ease of use), social influence, and facilitating conditions (including knowledge of a technology and the belief that end users will find it accessible) [1].
- A built-in user base that was part of the project team focused largely on biological sciences (other than molecular dynamics), earth sciences, and social sciences. In particular, we included communities that were known to be operating on cyberinfrastructure facilities insufficient to meet current user demands, such as Galaxy [2].
- A user interface – Atmosphere – known to the biological research community and used extensively as part of the iPlant (now Cyverse) project [3].

During PY2 of the Jetstream project, the following was accomplished:

- The Jetstream system was accepted by the NSF and is in full production mode.
- There was a major public launch of Jetstream on September 1, 2016 that attracted significant interest among the user communities targeted as the reason for creating Jetstream, as well as among the computational science community.
- Jetstream has operated at or above the target levels for availability and delivery. Usage is increasing but as of yet below hoped for and targeted levels.
- Outreach activities by the Jetstream team have been highly effective. Dozens of presentations, talks, and meetings have been held to increase awareness of and interest in Jetstream and more than 2,000 people in aggregate have attended some sort of outreach event promoting awareness of Jetstream.
- Interest in Jetstream as measured by resource allocation requests and awards is rising sharply since the beginning of September 2016, primarily as a result of extra effort (not

funded by the NSF) to encourage the completion and submission of proposals.

- There are a number of products of this project already, including published articles in peer-reviewed conference proceedings about computational science, and papers accepted and in review in biology and bioinformatics.
- More than 2,000 people have used Jetstream in total (including hundreds of students)
- There has been important transfer of information from the Jetstream project to major open source technology projects, particularly OpenStack. There have been no formal invention disclosures yet, but Jetstream staff and research faculty are playing an important role in the development of OpenStack. OpenStack itself is continually transferred from its open source repository to anyone in the US that wishes to download, install, and use it.
- Through our impact on the OpenStack project, we are playing an important role in the ongoing improvement of the open source cloud technology that is the most widely used open source cloud environment for small business and industry. This aids the development of the cloud-based economy in the US generally.
- Jetstream is both a first of a kind production science cloud and a pilot project in delivering a production cloud with NSF funding. Jetstream is a managed science and engineering cloud – a cloud managed for science and engineering. While we are still in the early months of production availability of Jetstream to the national research community, the early signs are that Jetstream will be important to the research community of the US and have important benefits beyond that community.

## 2. Accomplishments

### *2.1.* What are the major goals of the project?

Jetstream is 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 a 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 provides:

- “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.
- Hosting of persistent VMs to provide services beyond the command line interface for science gateways and other science services.
- Enable data movement, storage, and dissemination
    - Jetstream supports data transfer with Globus Connect.
    - Users are able to store VMs in the Indiana University digital repository, IUScholarWorks [4], and make them discoverable with a Digital Object Identifier (DOI).

- Provide virtual desktop services to tablet devices, increasing CI access for users at resource-limited institutions (e.g., small schools, schools in EPSCoR states, and Minority Serving Institutions), thus expanding access to Jetstream and the NSF XSEDE-supported ecosystem (XSEDE is the eXtreme Science and Engineering Discovery Environment).

While Jetstream is a hardware acquisition, the system is both a pilot and production system that is made novel by the software components and interfaces [5]. The overarching goal of the Jetstream team is to continue to improve upon existing functionality throughout the duration of the award by increasing stability and adding new features. The ability at which the latter can be done will ultimately be determined by the management and operations budget.

## 2.2. What was accomplished under these goals?

### *2.2.1. Major Activities*

**Jetstream has been formally accepted by the NSF and moved into full production "operations and management" mode.** Jetstream successfully passed the acceptance review by a panel of experts appointed by the National Science Foundation, which visited Indiana University in early May 2016. The panel's review report, subsequent review by the NSF Computer and Information Sciences and Engineering directorate, and review by the NSF Division of Grants resulted in the NSF declaring Jetstream to be in production operations as of 27 May 2016. This review reached the following conclusions:

- The hardware components purchased and installed as Jetstream match the hardware component numbers and capabilities as specified in the system description in the final grant proposal revised scope of work (based on the final budget specified by the NSF).
- Jetstream meets the requirements set out in NSF solicitation 14-536 in terms of integration with XSEDE.
- Jetstream fulfills the tests specified in our peer-reviewed Project Execution Plan in ways that demonstrate that Jetstream as currently implemented is indeed the system proposed, and it operates successfully.
- Jetstream functions effectively as an integrated cloud resources for scientific research
- Jetstream was operated during an extended test period in ways consistent with operational resource availability metrics set by mutual agreement between the NSF and Indiana University and its subcontractors. This includes that Jetstream was being allocated as called for in NSF solicitation 14-536 as a resource at 90% of its capacity via NSF-specified allocation processes operated by XSEDE (the Extreme Science and Engineering Discovery Environment).

A complete description of the acceptance test results is available online in the formal acceptance report submitted by IU at http://hdl.handle.net/2022/20908. A paper from Stewart and collaborators (C. A. Stewart, D. Y. Hancock, M. Vaughn, J. Fischer, L. Liming, N. Merchant, T. Miller, J. M. Lowe, D. Stanzione, J. Taylor, and E. Skidmore, "Jetstream - Performance, Early Experiences, and Early Results," in *Proceedings of the XSEDE16 Conference*, 2016.) describes the implementation, acceptance process, and early experiences with XSEDE. Table 1 (from Stewart et al. 2016) demonstrates that Jetstream easily passed the hardware functionality tests.

**Table 1. Jetstream hardware performance functionality test results. In the outcome column G in a green circle indicates that a metric is being met successfully; Y in a Yellow circle means that accomplishment is at least 60% of target and a metric should be a focus of attention and effort.**

| Test | Success criteria | Achieved values | Outcome |
|---|---|---|---|
| **Single-Node Performance Tests** | | | |
| High-Performance Linpack (HPL): Single node Linpack performance (for a problem size using at least half of the memory in a node) | Achieved performance in VM 80% or more of that achieved in Linux OS | Achieved in Linux OS: Jetstream-IU 697 GFLOPS; Jetstream-TACC 701 GFLOPS<br>•Achieved inside VM: 678 GFLOPS<br>•Performance in VM 97% of that in OS | G |
| STREAM: Single node OpenMP threaded performance (aggregate across the node). | 65 GB/s | Jetstream-IU: 100 GB/s; Jetstream-TACC 113 GB/s | G |
| 10 Gigabit Ethernet Bandwidth link performance: at least 1 GB/s for large-message point-to-point | 1 GB/s | 1.1 GB/s on the Indiana Cluster and 1.2 GB/s on the TACC cluster | G |
| **File System and Storage Benchmarks** | | | |
| 200 MB/s data transfer rate for data reads, 100 MB/s writes from within a virtual machine to the block storage (system totals). Initial tests done with dd. | 200 MB/s read<br>100 MB/s write | •Jetstream-IU 273 MB/s read, 170 MB/s write<br>•Jetstream-TACC 247 MB/s read, 251 MB/s write | G |
| Use IOR to analyze network bandwidth and I/O to the storage systems. (This test was not part of the initial acceptance test set; it was suggested by the review panel appointed by the NSF to review Jetstream). | N/A | •Jetstream-Indiana 50.8 MB/s read<br>•Jetstream-TACC 59.0 MB/s read<br>•Jetstream-Indiana 17.0 MB/s write<br>•Jetstream-TACC 19.7 MB/s write | **Not part of acceptance test** |
| **System Reliability & capacity** | | | |
| Uptime for a period of 14 days. | 95% uptime for 14 days | 100% uptime as an integrated resource during all of April | G |
| VMs simultaneously operating | 320 per location minimum | •Jetstream-IU: 998<br>•Jetstream-TACC: 832 | G |

### 2.2.2. *Specific Objectives:*

The specific objectives related to Jetstream are arranged here in terms of execution of terms in the Jetstream project plan; and achievement of operational and service delivery metrics established by mutual agreement between the National Science Foundation and IU and its partners and subcontractors; fulfillment of NSF review panel recommendations related to Jetstream.

The following are major specific objectives from the overall Jetstream project plan:

- March: The OpenStack APIs were accessible from programmatic access to Jetstream prior to acceptance, ahead of schedule and allowing gateways and users to interact in an automated fashion and leveraging Jetstream for persistent services.
- May: Acceptance process completed.
- August: We completed processing subcontracts for the full set of funded subcontractors for the Jetstream management and operations phase, including the subcontractors added as funded participants only in the Management and Operations phase: University of Texas San Antonio, Cornell University, Johns Hopkins University.
    - We also we engaged in considerable publicity calling attention to the "public launch" of Jetstream on 1 September
- September: We held the public launch of Jetstream. We also began reporting Jetstream to XCDB and charging usage against allocations.
- October. The Jetstream team continues to plan enhancements to available services and software features as the project has transitioned into the operations and management phase. Nineteen (19) technical representatives from Jetstream partners gathered at the

University of Arizona in October 2016 to plan and revise the project roadmap as it relates to new software development and feature enhancements. The team finished the prioritization process on sets of issues or major features, referred to as epics. The bulk of time was spent in the detailed specifications of a given epic or list of issues to include within an epic to ensure work was appropriately scoped and could be delivered in the intended release.

- Continuous improvement and ongoing operational management of the Jetstream system
    - Performed ongoing hardware and software maintenance and troubleshooting, and tuned system for performance, reliability, and stability.
    - Deployed accounting systems, interfacing with XSEDE and TACC.
    - Implemented image replication process between IU and TACC.
    - Responded to user issues both indirectly via staff contact and directly through XSEDE ticketing system.
    - Working with XSEDE, made adjustments to XSEDE ticketing system to provide more timely service to Jetstream users – many of whom are asking questions in "real time" as they do interactive work via Jetstream.
    - The Jetstream team has leveraged the pilot and production nature of the project to lead the way in the research cloud community with transparency by publicly sharing information such as technical reports (including benchmarks), architecture decisions, configuration files, and scripts. This stance has lead to presentations, often conducted jointly with unfunded collaborators, at multiple OpenStack Summits, the annual US Supercomputing conference, and a DOE workshop with international partners.
- The Jetstream user portal, Atmosphere, has undergone multiple significant upgrades adding functionality and addressing bugs with each release. The OpenStack services have also kept pace with the community release cycle (major updates every 6 months) ahead of what we committed to. We upgraded OpenStack and Ceph software. The most recent updates were as follows:
    - Upgraded to OpenStack Mitaka on July 11, 2016 (originally released April 7, 2016) & Ceph Jewel on May 24, 2016 (originally released April 21, 2016), within 6 months of software availability.
- Support for Science Gateways.
    - Jetstream delivers front ends for the following science gateways:
        - At least 10 science gateways are being supported.
        - SEAGrid. One of the most popular science gateways, SEAGrid has been available as a production service on Jetstream since September 7, 2016. SEAGrid was first used as part of an XSEDE outreach even April 21, 2016.
- Jetstream has provided back-end services for the Galaxy user portal at usegalaxy.org. Jetstream contributes significantly to providing computational services in support of Jetstream, and the movement of jobs from usegalaxy.org to Jetstream is a significant contributor to the load on Jetstream, as on average 210 VMs are running constantly on Jetstream doing computations routed to Jetstream via usegalaxy.org. (Galaxy is also available directly from Jetstream via standalone VM that is always kept up to date for the latest security patches).

The following are specific objectives specific to the project plan for Jetstream and work plans of individual funded subcontractors during PY2:

Indiana University

- Operated Jetstream at a high degree of availability. Jetstream was without interruptions in availability due to electrical power issues, in spite of several electrical outages created by regional construction events. There was one service outage of 12 hours: A significant fiber cut along the construction of new Interstate 69 did impact the Jetstream portal and availability of U for a period of 12 hours, TACC resources remained available via the API and all instances at both sites continued running. IU is discussing a new backup path for research traffic (current backup path resides in the same conduit).
- Ordered components and began installation of new Intrusion Detection System to increase security of Jetstream-Indiana
- Conducted significant benchmarks on both clouds to compare Ceph performance. These activities were done during production with no significant impact to user workloads.
- Provided major updates to online documentation and increased responsiveness for approving new imaging requests.

TACC

- Hardware: Replaced failed components under warranty via Dell support system.
- Software: Updated software to OpenStack Mitaka release. Updated Ceph storage system to Jewel release. Tuned RabbitMQ, Linux network stack, Ceph storage, and update firmware for performance and reliability. Ran IO and network benchmarks. Increased use of High-Availability features in OpenStack. Write scripts to automatically sync OpenStack images between IU and TACC. Implement TACC LDAP-based authentication for API users.
- Support: Answered user requests from email, ticket system, and in person. Implement feedback to improve documentation for future users. Fix technical issues and bugs that users encounter.
- Accounting: Configured Ceilometer metric-gathering system (an OpenStack component) for accounting purposes.

University of Arizona

- Accepted delivery of test and development system, began physical installation process.
- Developed a new allocation subsystem within Atmosphere to handle more sophisticated utilization and reporting model required by XSEDE.
- Integration with the TACC Accounting System (TAS), providing access to XSEDE allocation information and the ability to report utilization back to XSEDE.
- Enhanced integration of Atmosphere with Globus Auth, based on Globus requirements.
- Jetstream-AZ is now operational with operating with Ceph for storage management. We also added support for additional site-specific configurations, including: DNS servers; Timezones; Integration with NoVNC for web desktop; Migration to nginx, away from apache2, to improve the performance and stability of web shell and web desktop connections; Added support for distributing users across multiple OpenStack network routers for load balancing of network traffic; Numerous bug fixes, performance

enhancements, and features to improve the user experience of interacting with Atmosphere.

Johns Hopkins University

- Created a new version of the Galaxy standalone machine image with Galaxy application version 16.07 to provide the newest version and capabilities of Galaxy in a standalone VM image.
- Integrated this new machine image with a reference genome data repository used by the Galaxy Main server (usegalaxy.org) that made 3TB of reference genome data available to each launched instance (vs. ~50GB for the previous version of the image).
- Continuing to tune the ability to burst jobs from the Galaxy Main server onto Jetstream instances.

University of Chicago

- Removed the requirement that XSEDE users must have an active allocation to login to Jetstream. Now, anyone who has registered with XSEDE can login to Jetstream, assuming Jetstream itself authorizes their use.
- Debugged and resolved a critical integration issue between Globus Auth and Jetstream's Atmosphere instance that prevented Jetstream logins.
- Issued press release on Jetstream move to production, contributing to awareness.

### 2.2.3. *Significant results*

Accomplishment of system availability and functionality targets, and productivity targets, per cooperative agreement between NSF and IU and subcontractors:

Since the beginning of production operations IU and its partners have been achieving the operational, functionality, and availability goals set for Jetstream, in production operation, by agreement between the NSF and IU and its partners. However, in spite of outreach efforts, good user response to the capabilities of the system, and despite achieving or exceeding goals for operation of Jetstream, we are not yet achieving the goals set for usage levels of the system in terms of CPU utilization and number of VMs in use on a daily basis (levels of achievement outlined in Table 2). Performance on relevant metrics are shown in Table 3 and in Figures 1 through 4.

**Table 2. Status icons representing outcomes of performance tests.**

| Definition | Icon |
|---|---|
| Outcome is successful and complete | G |
| Outcome is in progress but not yet fulfilled or achieved | Y |
| Outcome is unsuccessful | R |
| Outcome is incomplete or metrics aren't available | W |

**Table 3. Operational availability, capacity, and usage metrics for Jetstream since NSF declaration that Jetstream was accepted and formally in production - beginning of June 201.6**

| | Q3<br>Jun - Aug | Q4<br>Sep - Nov | Annual running total | Status | Targets for metrics and Notes |
|---|---|---|---|---|---|
| System availability (uptime of an element of the production hardware, as % of wall clock time) | 99.29% | 100% | 99.65% | G | 95% |
| Capacity availability (% of the total capacity of Jetstream available for NSF use over time) | 99.65% | 99.14% | 99.34% | G | 95% |
| Job completion success - % of jobs submitted should complete without having to be resubmitted as a result of a failure in the hardware or system software. | 99.01% | 99.53% | 99.27% | G | 96% |
| Core cloud environment software will be upgraded to match current versions components such as operations systems and cloud software environments | Mitaka | Mitaka | -- | G | OpenStack updated prudently, generally within ≤12 months of major releases |
| Capacity of system allocated via NSF-specified allocation process | 96.5% | >100% | 98.25% | G | 90% |
| Total number of distinct users | 593 | 1,463 | 1,463 | G | 1,000; annual goal already exceeded |
| Total number of students having used Jetstream in an educational or training setting | 36 | 281 | 281 | G | 100; annual goal already exceeded |
| Total number of science gateways using Jetstream (running total) | 5 | 5 | 5 | G | 2; annual goal already exceeded |
| Use - average number of VMs active 24 hour average | Daily mean=253<br>Peak= 540 | Daily mean=342<br>Peak =551 | Daily mean=298<br>Peak=551 | Y | 320 |
| SUs available to user community per month | 1.47M | 1.47M | 2.94M | G | 1.47M (per month) |
| % of SUs available to user community that were used per month | -- | 79.81% | 79.81% | W | -- |
| CPU % utilization | Mean=0.60%<br>Peak=2.34% | Mean=0.88%<br>Peak=4.39% | Mean=0.74<br>Peak=4.39 | R | 6% |

| | Q3<br>Jun - Aug | Q4<br>Sep - Nov | Annual running total | Status | Targets for metrics and Notes |
|---|---|---|---|---|---|
| Total number of publications facilitated by use of Jetstream | 1 | 1 | 2 | W | 5 |
| Total number of VM images and/or data sets published with a DOI via IUScholarWorks | 9 | 1 | 10 | G | 10; annual goal already met |

As shown in Figure 1, since our publicity events and (unfunded) increased in outreach efforts began in earnest on September 1, 2016, the number of users has grown to more than 1,400. Similarly, we show in Figure 2 that SU utilization has been increasing over time. To better illustrate Jetstream is being used, we plotted the number of VMs launched per day in Figure 3 and the daytime (8:00am to 8:00pm) versus nighttime (8:00pm to 8:00am) running of VMs in Figure 4.

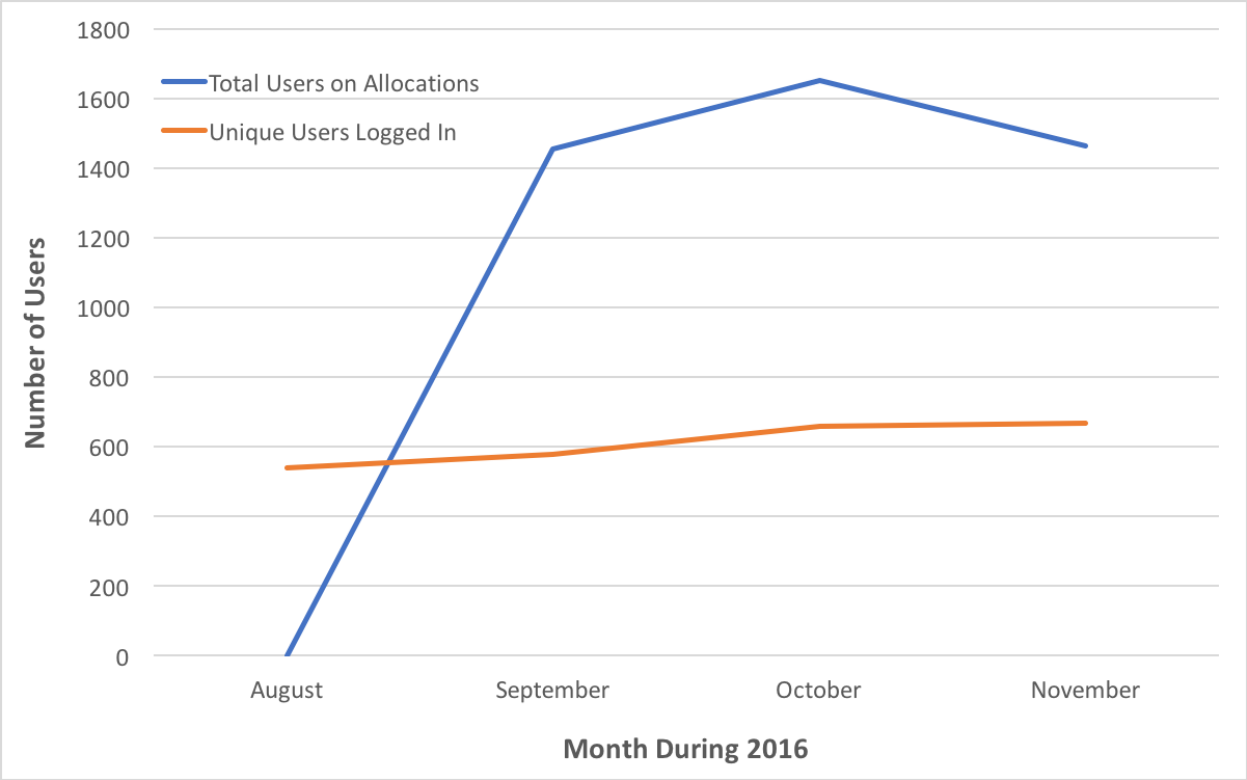


**Figure 1. Growth in distinct users of Jetstream, showing total number users overall and distinct users accessing the system per month.**

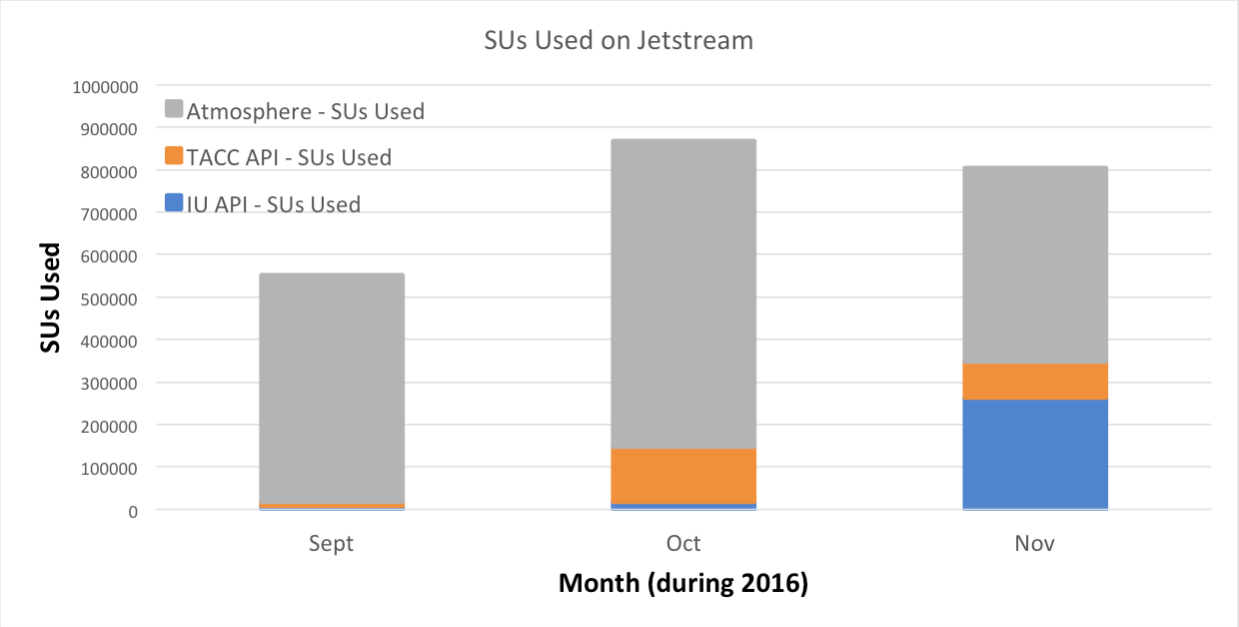


**Figure 2. SU consumption per month, starting with September 2016. (SU usage was charged against allocations starting effective 1 September 2016). In this graph, grey shows interactive use mediated by the Atmosphere interface; blue shows use on Jetstream-IU by applications that interact with the OpenStack API programmatically and represent a mix of jobs orchestrated by Mesos or a similar tool plus interactive and scripted access via Science Gateways; and orange shows use on Jetstream-TACC via the API.**

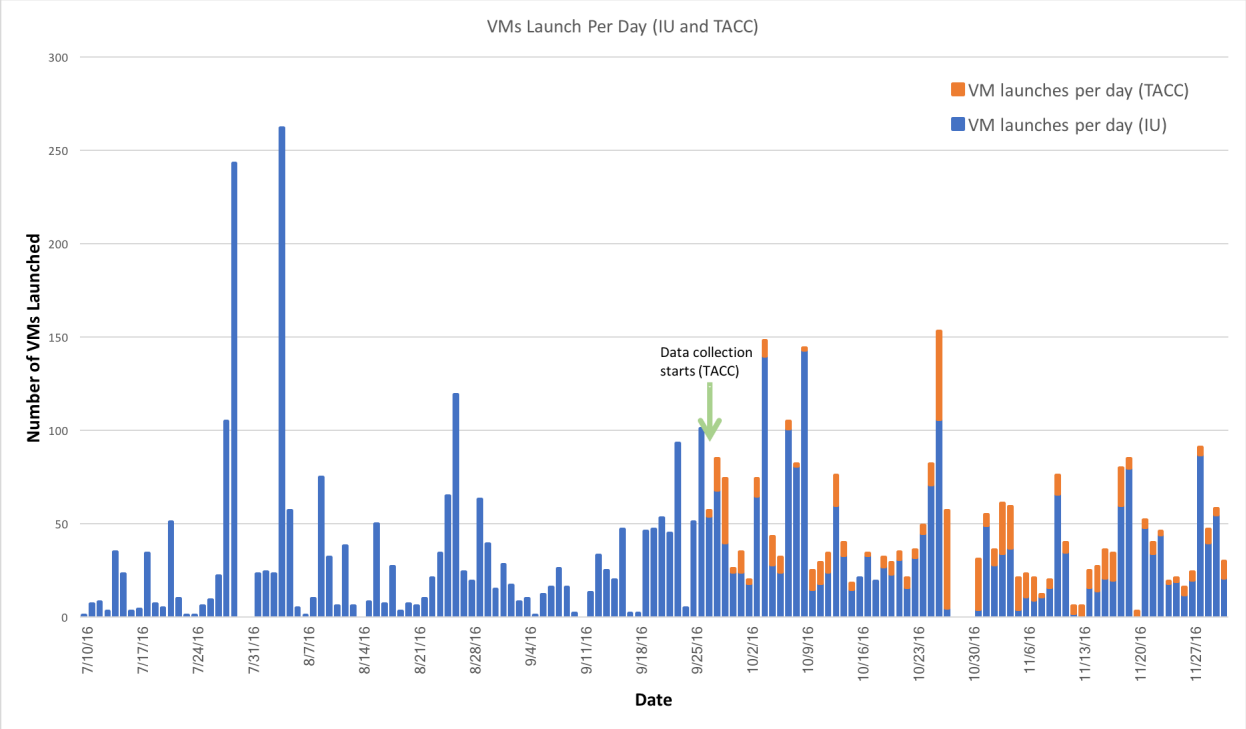


**Figure 3. Total numbers of VMs launched per day by cloud (VMs launched on Jetstream-IU in blue, and VMs launched on Jetstream-TACC in orange). Data for Jetstream-TACC are not available prior to Sep 25 2016.**

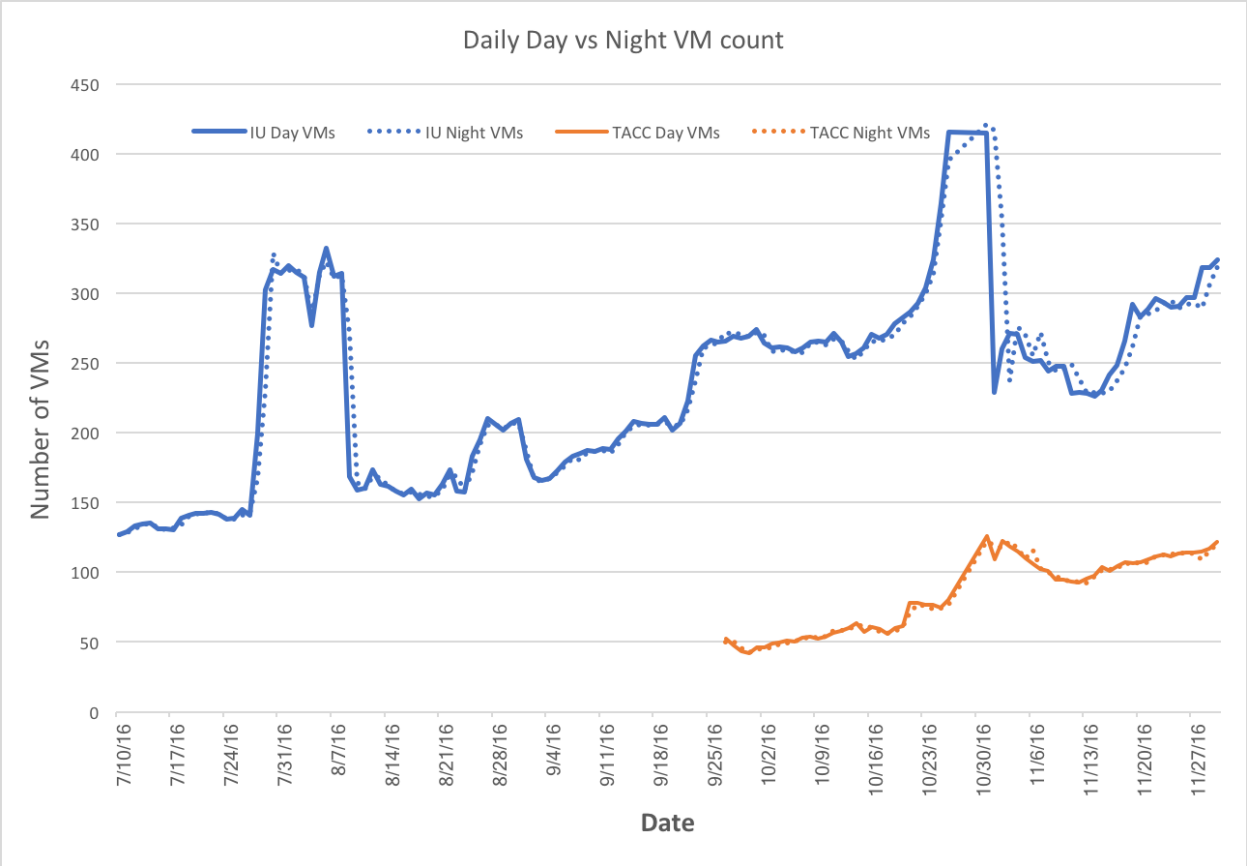


**Figure 4. Daytime (8am – 8pm) versus nighttime (8pm – 8am) VMs running during July - November 2016 by cloud (Jetstream-IU in blue, Jetstream-TACC in orange). There was little difference between daytime and nighttime running of VMs.**

Taking the data as a whole, what we see is success in meeting goals set in the NSF – awardee continuing agreement, good success in adding users, and levels of usage in terms of number of VMs in use, SUs consumed, and % CPU utilization that are below target levels and below system capability. In other words: growth in general in terms of number of users is acceptable to good; growth of intensive usage of the system lags behind the capability of the system to deliver computational resources.

Part of this lag is a result of a pattern in requests for application usage, and we have already observed and reacted to this trend. Data on resource requests for use of Jetstream are shown in Table 6. During the first round of research-level allocation requests after Jetstream went into production, during summer of 2016, there were no research-level requests for Jetstream.

### 2.2.4. *Key outcomes or Other achievements*

Fulfillment of NSF review panel recommendations related to Jetstream

The NSF-mandated and supervised acceptance review was notable for its amicable tenor and for the several very useful recommendations made by the review panel to the Jetstream project. The recommendations from the review panel are listed below, along with indications on the status of Jetstream project progress toward fulfilling these suggestions. (A sign of the quality and value of

the review process is the fact that Jetstream leadership has agreed that each suggestion is of merit and will be implemented if possible). Recommendations and status of responses are shown in Table 4.

**Table 4. Recommendations made regarding Jetstream project by NSF-appointed review panel**

| Recommendation (grouped by topic area) | Status |
|---|---|
| Share perspectives and experiences gained in Jetstream application of scientific impact metrics towards design of IU proposed acceptance tests benchmarks applied to Jetstream as contributions towards potential future NSF consideration of NSF system acquisition evaluative criteria. | In progress |
| Add an ISSO to the Jetstream advisory committee | Requires follow-up from PI Stewart |
| Develop a Secure Software MarketPlace, including perhaps in collaboration with ongoing synergistic developments @ the IU Center for Applied Cybersecurity Research, including with Von Welch, to further accelerate advances in "trusted virtual machines" | Initial plans have been made to certified as SWAMP-reviewed without critical security problems |
| Accelerate exploration of deploying trusted containers as a contributions further advancing system level security controls | Initial plans have been made; this activity will be included in a request for supplemental funding submitted to the NSF before the end of calendar 2016 |
| Expand the composition of the advisory board to include representatives of engineering communities | Completed |
| Further accelerate engagements with science and engineering research communities contributing to prioritize science and engineering applications, potentially through science gateways, including as contribution towards maintaining project focus and scope | In progress |
| Accelerate scalable outreach through students, including potentially those who have successfully used Jetstream including to perhaps produce a YouTube video as a potential contribution towards significantly broadening awareness of Jetstream, including perhaps if the resulting video were to "go viral" | Planned for PY3 |

The members of the Jetstream Stakeholder Advisory board are shown below in Table 5. Table The SAB has been consulted for advice and recommendations regarding the Jetstream project. Since the acceptance of the Jetstream system, new members have been added to the board.

**Table 5. Current members of the Jetstream Stakeholder Advisory Board**

| Person | Field | Institution | Home page |
|---|---|---|---|
| Gwen Jacobs (Chair) | Cyberinfrastructure, Biology | University of Hawaii | |
| Dennis Gannon | Computer science, cyberinfrastructure | Indiana (supposedly retired) | |
| Robert Ricci (Cloudlab PI) | Computer science | University of Utah | http://www.flux.utah.edu/users/ricci/ |
| Kate Keahey (Chameleon PI) | Computer science | University of Chicago, Argonne National Labs | www.mcs.anl.gov/~**keahey**/ |
| David Powers | HPC, cloud computing | Dow Chemical | https://www.linkedin.com/in/dave-powers-85229b5 |

| Person | Field | Institution | Home page |
|---|---|---|---|
| Joe Mambretti | Computer Science | Northwestern University | http://www.it.northwestern.edu/about /bios/mambretti.html |
| Ramamurthy Mohan | Atmospheric science | UCAR | http://www.unidata.ucar.edu/publications/directorspage/ |
| Julia Lane | Public policy, social science | NYU | http://wagner.nyu.edu/community/faculty/julia-ingrid-lane |
| Carl Boettiger | Environmental science | UC Berkely | http://www.carlboettiger.info |
| Robert Guralnick | Bioinformatics, field biology | Department of Natural History and the Florida Museum of Natural History | https://sites.google.com/site/robgur/ |
| Thomas Siegmund | Engineering (mechanical) | Purdue University | http://mymech.org |
| Scott Brandenberg | Engineering (civil and environmental) | UCLA | http://www.cee.ucla.edu/profile-Brandenberg/ |
| TBD | CISO (Chief Information Security Officer) | | |

## Encouragement of uptake and adoption of Jetstream as indicated in resource requests to the XSEDE Resource Allocation System

This report was delayed until early December in order to include the outcomes of XRAC meeting that was to decide the outcome of resource requests made in October was scheduled to be completed on 5 December 2016. As shown in Table 6, requests for Research Allocations has dramatically increased since the publicity campaign started in early fall of 2016.

**Table 6. New resource requests submitted via XRAS During Management and Operations Program Year 1 for use of Jetstream as of the full month of November (requests and the results of the December 4 and 5 allocation review).**

| | June | July | August | PY Q1 Total | Sept | October | November | PY Total Year to Date |
|---|---|---|---|---|---|---|---|---|
| *Startup* | | | | | | | | |
| Total requests | 3 | 7 | 3 | *13* | 12 | 19 | 7 | 44 |
| SUs requested | 110,062 | 550,000 | 150,000 | *810,062* | 1,130,000 | 1,719,000 | 500,000 | 3,659,062 |
| SUs awarded | 110,062 | 550,000 | 150,000 | *810,062* | 1,130,000 | 1,719,000 | 500,000 | 3,659,062 |
| *Educational* | | | | | | | | |
| Total requests | 1 | 0 | 5 | *6* | 1 | 0 | 0 | 7 |
| SUs requested | 50,000 | 0 | 342,000 | *392,000* | 50,000 | 0 | 0 | 442,000 |
| SUs awarded | 50,000 | 0 | 342,000 | *392,000* | 50,000 | 0 | 0 | 442,000 |
| *Campus Champion* | | | | | | | | |
| Total requests | 20 | 12 | 13 | 45 | 13 | 11 | 4 | 69 |
| SUs | 1,000,000 | 600,000 | 650,000 | 2,250,000 | 650,000 | 550,000 | 200,000 | 3,450,000 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| requested | | | | | | | | |
| SUs awarded | 1,000,000 | 600,000 | 650,000 | 2,250,000 | 650,000 | 550,000 | 200,000 | 3,450,000 |
| *Supplemental* | | | | | | | | |
| Total requests | 0 | 1 | 0 | 1 | 5 | 1 | 0 | 7 |
| SUs requested | 0 | 50,000 | 0 | 50,000 | 450,000 | 1,500,000 | 0 | 2,000,000 |
| SUs awarded | 0 | 50,000 | 0 | 50,000 | 450,000 | 1,500,000 | 0 | 2,000,000 |
| *Research allocations* | | | | | | | | |
| Total requests | N/A | 2 | N/A | 2 | N/A | 14 | N/A | 16 |
| SUs requested | N/A | 302,560 | N/A | 302,560 | N/A | 14,686,736 | N/A | 14,989,296 |
| SUs awarded | N/A | 302,560 | N/A | 302,560 | N/A | 13,186,736 | N/A | 13,489,296 |
| ***Total request and allocations*** | | | | | | | | |
| **Total requests** | **24** | **22** | **21** | **78** | **31** | **45** | **11** | **154** |
| **Total SUs requested** | **1,160,062** | **1,502,560** | **1,142,000** | **3,804,622** | **2,280,000** | **18,455,736** | **700,000** | **25,240,358** |
| **Total SUs awarded** | **1,160,062** | **1,502,560** | **1,142,000** | **3,804,622** | **2,280,000** | **16,955,736** | **700,000** | **23,740,286** |

Other achievements include:
Agave API Platform project (PI: Rion Dooley)

- Team has automated deployment of entire Agave sandboxes using Jetstream and their deployer/ansible project.
- Team has developed elastic scaling out to Jetstream for their JupyterHub deployments.

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

- Two doctoral candidates were able to complete their dissertations using the Jetstream system.
- Jetstream staff members have attended several conferences allowing them to keep current with the latest updates and developments.
- Jetstream has provided the following training opportunities:
  - Hands-on tutorial and resource diversity panel at XSEDE16 (July 2016)
  - Hands-on tutorial for XSEDE and ECSS staff at XSEDE16 (July 2016)
  - Hands-on tutorial at Internet2 Technology Exchange (September 2016)
  - TACC Staff Maria Esteva led Digital Preservation Workshop at the Institute for Research of Preservation of Cultural Patrimony at the Universidad Nacional de San Martin in Buenos Aires, Argentina. She used Jetstream instances to allow the students to "practice with different repository and data curation workflows".

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

The Jetstream team began a publicity campaign after the Jetstream system was accepted for production in May 2016. The Jetstream team has given dozens of presentations including

symposia, posters, conference talks, and tutorials in order to disseminate information to the relevant scientific and computing communities of interest. A full listing of these events is in Table 7 along with attendance information. In addition, several of the partner institutions had event space with Jetstream displays, demos, and talks at the SuperComputing16 conference -- we estimate that several hundred of the thousands individuals that were in attendance at SC16 visited the booths.

Sample of presentations:

- Presentation for The Federation of Earth Science Information Partners (ESIP) at their winter conference (January 2016)
- Presentation for Southern Partnership in Advanced Networking (SPAN) Spring Workshop (March 2016)
- Presentation for XSEDE Gateways Community monthly call (March 2016)
- Presentation to the Indiana University Security Office (March 2016)
- Presentation to multiple Indiana University School of Informatics classes (2016)
- Presentation for OpenStack Summit Austin (April 2016)
- Presentation and tutorial at Rocky Mountain Advanced Computing Symposium at HPC User Forum, Tucson, AZ (April 2016)
- Presentation for i-Light Conference (June 2016)
- Presentations at Galaxy Community Conference (June 2016)
- Presentation at RMACC, Ft. Collins, CO (August 2016)
- Presentation for Internet2 Technology Exchange (September 2016)
- Presentation for OpenStack Summit Barcelona (October 2016)
- Presentation at Johns Hopkins University (November 2017)
- Presentations at SC16 (November 2016)

Meetings with research groups interested in being future users of Jetstream:

- Galaxy Community Conference demo (June 2016)
- Hands-on tutorial and resource diversity panel at XSEDE16 (July 2016)
- Purdue Engineering demonstration (August 2016)
- Tapia Conference Booth and Poster Session (September 2016) - At the Jetstream Booth for this vendor / sponsor / recruiter fair, hundreds of attendees were made aware of Jetstream as a potential resource for education, professional development, and a platform for learning about cloud technologies. Information for the Tapia conference (which self-describes as "a premier venue to acknowledge, to promote and to celebrate diversity in computing" is online at http://tapiaconference.org).
- Collaboratory Research and Education Resources panel at SIGUCCS16 (November 2016)
- TACC Staff Maria Esteva led Digital Preservation Workshop at the Institute for Research of Preservation of Cultural Patrimony at the Universidad Nacional de San Martin in Buenos Aires, Argentina. She used Jetstream instances to allow the students to "practice with different repository and data curation workflows".

**Table 7. Presentations and (approximate) number of attendees since Jetstream went into production at the end of May 2016.**

| Presenters | Title | Event | Date | URL if applicable to source materials | Attendees |
|---|---|---|---|---|---|
| Tuecke, S | Building the Modern Research Data Portal using the Globus Platform | Tutorial @ 2016 NSF Cybersecurity Summit for Large Facilities and Cyberinfrastructure | 8/16/2016 | http://trustedci.org/s/BuildingTheModernResearchDataPortalUsingTheGlobusPlatform.pdf | 20 |
| Tuecke, S | Using Globus Auth to Streamline the Creation, Integration, and Use of Research Services | Computing Techniques Seminar @ Fermi Nat'l Laboratory, Batavia, IL | 11/09/2016 | -- | -- |
| Foster, I. | Streamlined data sharing and analysis to accelerate cancer research | National Cancer Institute "Big Data" seminar | 9/20/2016 | http://www.slideshare.net/ianfoster/streamlined-data-sharing-and-analysis-to-accelerate-cancer-research | 60 |
| Foster, I. | The Discovery Cloud: Accelerating and democratizing research on a global scale | International Conference on Cloud Engineering, Berlin Germany | 5/15/2016 | https://www.computer.org/csdl/proceedings/ic2e/2016/1961/00/1961a068-abs.html | 100 |
| Foster, I., and R. Madduri. | Big Data Technologies for Biomedical Knowledge Discovery | Tutorial at American Medical Informatics Institute annual meeting, Chicago, Illinois | 11/12/2016 | -- | 80 |
| Foster, I. | Globus Auth: A Research Identity and Access Management Platform | 12th IEEE International Conference on eScience | 10/25/2016 | http://www.slideshare.net/ianfoster/globus-auth-a-research-identity-and-access-management-platform | 40 |
| Tuecke, S., Vasiliadis, V., Ananthakrishnan, R., Martin, S. | Building the Modern Research Data Portal using the Globus Platform | SC16 Developer Tutorial (SC16, Salt Lake City, UT, November 2016). | 11/13/2016 | https://www.globusworld.org/files/2016/161113_Globus_SC16_Dev_Tutorial.pdf<br><br>https://www.globusworld.org/files/2016/161113_Globus_SC16_Admin_Tutorial.pdf | 26 |
| Vasiliadis, V. et al. | Building the Modern Research Data Portal (2 Day | GlobusWorld Tour – San Fransicso (ESNet) | 9/14/2016 | https://www.globusworld.org/files/2016/160913_GWTour_Berkeley.pdf | 42 |

| Presenters | Title | Event | Date | URL if applicable to source materials | Attendees |
|---|---|---|---|---|---|
| | Developer Tutorial) | | | | |
| Vasiliadis, V. et al. | Building the Modern Research Data Portal (2 Day Developer Tutorial) | Yale University | 10/13/2016 | https://www.globusworld.org/files/2016/161012_GWTour_Yale.pdf | 20 |
| Vasiliadis, V. et al. | Building the Modern Research Data Portal (2 Day Developer Tutorial) | NCAR, Boulder, CO | 10/26/2016 | https://www.globusworld.org/files/2016/161026_GWTour_NCAR.pdf | 35 |
| Fischer, J | Jetstream Overview | Presentation for The Federation of Earth Science Information Partners (ESIP) at their winter conference | 1/6/2016 | http://hdl.handle.net/2022/20577 | 20 |
| Fischer, J | Jetstream: A new national research and education cloud | Presentation for Southern Partnership in Advanced Networking (SPAN) Spring Workshop | 3/9/2016 | http://hdl.handle.net/2022/20724 | 42 |
| Fischer, J | Jetstream Overview | Presentation for XSEDE Gateways Community monthly call | 3/25/2016 | -- | 12 |
| Fischer, J. | ECSS Jetstream API Talk/Tutorial | Staff tutorial presented at XSEDE16 Conference, St. Louis, MO | 7/21/2016 | -- | 35 |
| Lowe, J.M. & Turner, G | Jetstream: a national research and education cloud | Presentation to the Indiana University Security Office | 3/1/2016 | http://jetstream-cloud.org/publications.php | 20 |
| Lowe, J.M. & Turner, G | Jetstream Overview | Presentation to IU School of Informatics and Computing Classes | 11/1/2016 | http://hdl.handle.net/2022/20744 | 30 |
| Lowe, J.M. | Deploying OpenStack for the National Science Foundation's Newest Supercomputers | Presentation for OpenStack Summit, Austin, TX | 4/28/2016 | http://hdl.handle.net/2022/20824 | 120 |
| Lowe, | Image is | Presentation for | 10/27/2016 | http://hdl.handle.net/2022/21069 | 30 |

| Presenters | Title | Event | Date | URL if applicable to source materials | Attendees |
|---|---|---|---|---|---|
| J.M. | Everything: Dynamic HPC VM Repositories using Murano | OpenStack Summit Barcelona | | | |
| Fischer, J | Jetstream Overview - Prepared for 2016 i-Light Conference | I-Light Conference | 6/19/2016 | http://hdl.handle.net/2022/20878 | 45 |
| Fischer, J. | Jetstream Overview | Seminar, Purdue University, West Lafayette, IN | 8/17/2016 | www.jetstream-cloud.org/files/Jetstream%20-%20Purdue%20Engineering%20Visit.pptx | 14 |
| Fischer, J. | Hands on with Jetstream. | Tutorial presented at the XSEDE 16 Conference, St. Louis, MO | 7/18/2016 | -- | 16 |
| Fischer, J. | Hands on with Jetstream. | Tutorial presented at the I2 Tech Exchange, Miami, FL | 9/27/2016 | http://www.jetstream-cloud.org/files/I2%20TechEx%20Tutorial%20-%20Hands%20on%20with%20Jetstream%20.pptx | 28 |
| Fischer, J. Afgan, E., Doak, T., Ganote, C., Hancock, D.Y., & Vaughn, M. | Using Galaxy with Jetstream | Poster presented at the Galaxy Community Conference 2016. Bloomington, IN. | 6/28/2016 | https://f1000research.com/posters/5-1544 | 25 |
| Goonasekera, N., Lonie, A., Taylor, J., Afgan, E. | CloudBridge – a Simple Cross-Cloud Python Library | XSEDE16 | 7/19/2016 | https://docs.google.com/presentation/d/109wUJAVA1yAf5t2U1qgBRJ8PbDbn-MsIx7_-IwAdmWc/edit?usp=sharing | 50 |
| Hancock, D.Y., Stewart, C.A., Fischer, J., Lowe, J.M., Rad, P. & Vaughn, M. | Resource Management from HPC to the Cloud. Proceedings of Accelerated Data and Computing (ADAC) Workshop: Resource Management from HPC to the Cloud. | Accelerated Data and Computing (ADAC) Workshop, Lugano, Switzerland | 6/14/2016 | http://computing.ornl.gov/workshops/adac/docs/S3_hancock_ADAC2016.pdf | 59 |
| Lowe, J.M. | OpenStack for HPC: Best | Symposium conducted at SuperComputing16, Salt Lake City, UT. | 11/15/2016 | -- | 150 |

| Presenters | Title | Event | Date | URL if applicable to source materials | Attendees |
|---|---|---|---|---|---|
| | practices for optimizing software-defined infrastructure. | (organized by J. Mills) | | | |
| Lowe, J.M. | Jetstream: A science and engineering cloud | Ceph in HPC environments. Symposium conducted at SuperComputing16, Salt Lake City, UT. (D. Fuller (Chair),) | 11/16/2016 | https://www.msi.umn.edu/sites/default/files/4-Jetstream.pdf | 75 |
| Lowe, J.M | HPC/Research Computing : Leveraging the Architectures, Flexibilities, and Tools Emerging from the Members of the OpenStack Scientific Community | Symposium conducted at SuperComputing16, Salt Lake City, UT. | 11/17/2016 | -- | 200 |
| Lowe, J.M. & Turner, G | Research Computing on Cloudy Platforms. | Talk presented at the Operating Innovative Networks Workshop. Bloomington, IN. | 7/12/2016 | www.jetstream-cloud.org/files/I2-OIN-IUB-160712.pptx. | 60 |
| Merchant, N | Cloud Computing for Education and Research | Plant and animal genome conference (PAG) | 1/10/2016 | -- | -- |
| Merchant. N. | Data Sharing and Infrastructure Working Group | South Big Data Hub Data Sharing and Infrastructure working group (virtual meeting) | 10/14/2016 | Info about workshop at https://southbdhub.wordpress.com/2016/10/13/working-group-to-host-series-of-demos/) | 25 |
| Merchant, N. | High Throughput Bioinformatics with CyVerse and Jetstream | CCL Workshop on Scalable Scientific Computing 2016, University of Notre Dame (Info about workshop at: http://ccl.cse.nd.edu/workshop/2016/) | 10/19/2016 | http://ccl.cse.nd.edu/workshop/2016/slides/CCtools-CyVerse-JS-2016.pptx | 12 |
| Mock, S. | A | Gateways 2016 | 11/3/2016 | https://drive.google.com/open?id=0By3fQ4sX63 | n/a |

| Presenters | Title | Event | Date | URL if applicable to source materials | Attendees |
|---|---|---|---|---|---|
| | Cyberinfrastructure for the Natural Hazards Community | Conference, San Diego, CA. | | N2azUyMzJtdm9JOWc | |
| Vaughn, M. | Jetstream: Accessible cloud computing for the national science and engineering communities | Rocky Mountain Advanced Computing Center; Fort Collins, CO | 8/1/2016 | -- | -- |
| Vaughn, M. | Jetstream: Cloud-first Computing Extending Our National Cyberinfrastructure | Texas Advanced Computing Center; Austin, TX | 4/1/2016 | -- | -- |
| Stewart, C.A. | Coevolution of biological research and cyberinfrastructure from now till the end of Moore's law. | Galaxy Community Conference, Bloomington, IN | 6/28/2016 | http://hdl.handle.net/2022/20903. | 250 |
| Stewart, C.A. | Cyberinfrastructure from scratch | Presented at University of Wyoming Laramie WY | 5/19/2016 | http://hdl.handle.net/ 2022/21103 | 20 |
| Stewart, C.A. | IU cyberinfrastructure – supercomputers and clouds in the context of initiatives toward exascale | IU Bloomington Astronomy Department symposium. Apr 26 2016. Presented at Swain Hall, Indiana University, Bloomington, IN. | 5/26/2016 | http://hdl.handle.net/ 2022/21101 | 40 |
| Stewart, C.A. | Jetstream - An NSF-funded cloud resource for researchers working in the long tail of | Presented at University of Wyoming, Laramie, WY | 5/19/2016 | http://hdl.handle.net/ 2022/21104 | 30 |

| Presenters | Title | Event | Date | URL if applicable to source materials | Attendees |
|---|---|---|---|---|---|
| | science. | | | | |
| Stewart, C.A. | Watson, Medicine, Health, and IU Grand Challenges | Presented at Cook Group, Bloomington, IN | 9/14/2016 | http://hdl.handle.net/ 2022/21102 | 40 |
| Stewart, C.A. | Jetstream – Performance, Early Experience, and early results | Presented at XSEDE16, Miami, FL | 7/19/2016 | http://hdl.handle.net/ 2022/21106 | 65 |
| Stewart, C.A. | Jetstream update for CASC (Coalition for Academic Scientific Computation) | Presented at CASC meeting, Arlington, VA | 9/15/2016 | http://hdl.handle.net/ 2022/21105 | 75 |
| Stewart, C.A., Hancock, D.A. | Jetstream Production Kickoff Event | Presentations in CIB, Bloomington IN | 9/1/2016 | -- | 120 |
| Stewart, C.A. | Jetstream - In production & supporting science | | | http://hdl.handle.net/ 2022/21090 | -- |
| Stewart, C.A., D. Y. Hancock, Jennifer Laherty | Jetstream - In production & supporting science | Presented in IU display in Exhibition Hall at SC'17, Salt Lake City, Utah. | 11/14/2016 | -- | 24 |
| Stewart, C.A., D. Y. Hancock, Jennifer Laherty | Jetstream - In production & supporting science | Presented in IU display in Exhibition Hall at SC'17, Salt Lake City, Utah. | 11/15/2016 | -- | 16 |
| **Total attendees at all presentations and tutorials, summed:** | | | | | **2,171** |

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

- The Jetstream team will continue to make a concerted effort to increase the number of research requests submitted to the XRAC committee.
- The Jetstream team will work with XSEDE and the NSF to plan other actions that can accelerate uptake of Jetstream.
- The Jetstream team will continue to work with XSEDE in the hope that future reports from XRAS do a significantly more complete job of reporting actual system resource

requests than the August XRAC did. It is obvious from this report that the August XRAC report created an incorrect overall perception of user interest in Jetstream.

- We will continue to investigate what we believe is the optimal amount of oversubscription of the system. We are in this area working in uncharted waters. XSEDE has years of experience in estimating what the relationship is between SUs allocated and SUs actually used. We will be working to adjust allocation amounts so as to arrive at a rational level of allocation and use as perceived by the user and by funding agencies and taxpayers: good user responsiveness on the part of the system (user view) and overall reasonable levels of use of the system (the view of the funders).
- The outcome of the technical roadmap meeting in October is that the following high priority items will receive development effort in the coming quarters: make it easier for users to onboard with Jetstream at a pre-startup level (using an approved community allocation), performing quality assurance work on end to end allocation data for SUs billed into XSEDE, integrating VM publication information info into the Jetstream portal, creating a monitoring/status reporting page for internal and external Jetstream services, and programmatically deploying globus personal connect services (already available in featured base images).
- Additionally our technical schedule currently states that we'll have a fully supported S3 block storage service at IU and TACC by the end of December 2016. The Globus S3 storage connector will be fully supported by the end of March 2017. Work is currently underway on both activities in parallel, and we'll begin offering these services to Jetstream users at the earliest opportunity.
- We will continue to pursue a federated keystone environment with partners involved in the XSEDE project, making it easier for users to leverage multiple OpenStack environments offered by XSEDE Service Providers.
- Jetstream-AZ will be leveraged to test a number of OpenStack service levels and provide a more comprehensive environment for quality assurance testing as it relates to future updates of the Jetstream user portal.
- Cornell virtual workshop is being developed. The full module is projected to be completed Q2 2017.
- We are seeking funding for an additional staff member to aid in education, outreach, and training efforts. Specific events include training sessions at partner institutions such as:
    - Jetstream Training: Half-day Seminar (Proposal: Virginia Trueheart)
        - The half-day training will cover the details of the machine architecture and the basics of running on the system.
            - Link to training proposal in Google Docs:https://docs.google.com/document/d/1Mwi-xTZx75Hu2ZdIUMNlvQCT7tuh3MKajBVLopOF4mM/edit

- Engineering outreach efforts: UTSA has been working with local college of engineering faculty. Based on their feedback, we have been building the following engineering research platforms for Machine Learning/Vision and Robotics similar to what Galaxy does for bioinformatics.

  The Machine Learning/Vision platform will include:
    - Environment:Jupyter Notebook Python

- Machine Learning Package Library and Environment : Tensorflow, Caffe, Torch, Theano
- Datasets:ImageNet, CIFAR100
- Algorithms: Recurrent Neural Network, Convolutional Neural Network, Support Vector Machine, K-nearest neighbor, K-Means, Random Forest, Decision Tree, Naive Bayes, Dimensionality Reduction Algorithms, Gradient Boost & Adaboost

The Robotic Operating System (ROS) platform will include:

- Environment: Jupyter Notebook Python, Indigo Igloo
- Robotic Operating System (ROS) Package library and environment: Gazibo, common_msgs, geometry, hector_localization, image_common
- Algorithms: Kalman Filtering, Multi Agent Control optimization, PID Control, Effort_controllers, Position_controllers, SWARM , Object Recognition systems

## 3. Products

Products resulting from this project during the specified reporting period are listed below (based on NSF classification of products).

Journal articles or Juried Conference Papers

*Journal articles*

1. Memarzadeh, M. and C. Boettiger. 2017. Measurement uncertainty matters: ecological management using POMDPs. Accepted pending revision; revisions in process. Preprint retrieved from http://biorxiv.org/content/early/2016/05/30/055319

*Juried Conference Papers*

2. Stewart, C.A., Hancock, D.Y., Vaughn, M., Fischer, J., Liming, L., Merchant, N., Miller, T., Lowe, J.M., Stanzione, D., Taylor, J., & Skidmore, E. (2016). Jetstream - Performance, Early Experiences, and Early Results. Proceedings of the XSEDE16 Conference. St. Louis, Mo. (This paper won the "Phil Andrews Memorial Award for Best Technical Paper at XSEDE'16)
3. Eric Coulter, Jeremy Fisher, Barbara Hallock, Richard Knepper, Dave Lifka, JP Navarro, Marlon Pierce, Craig Stewart (2016). Cyberinfrastructure as a Platform to Facilitate Effective Collaboration between Institutions and Support Collaboratories. *Proceedings of the 2016 ACM on SIGUCCS Annual Conference*.
4. Steven Tuecke, Rachana Ananthakrishnan, Kyle Chard, Mattias Lidman, Brendan McCollam, Stephen Rosen, Ian Foster (2016). Globus Auth: A Research Identity and Access Management Platform. *IEEE 12th International Conference on eScience*.

5. Stewart, C.A. (2016) Preserving Scientific Software . . . in a Usable Form? EDUCAUSE Review 51, no. 6 (November/December 2016). Retrieved from http://er.educause.edu/articles/2016/10/preserving-scientific-software-in-a-usable-form

Other Conference Presentations/Papers

1. Afgan, E. & Goonasekera, N. (2016). *CloudLaunch as a multi-cloud, multi-application launch platform*. Galaxy Community Conference. Bloomington, IN.
2. Anthonysamy WJB, Douglas MR, Bangs MR, Chafin TK, Martin BT, Mussmann SM, Douglas ME. 2016. *Phylogeography of the western rattlesnake complex (*Crotalus viridis spp.*): A genomic perspective.* Joint Meeting of Ichthyologists and Herpetologists, New Orleans, LA. Poster
3. Bangs MR, Douglas MR, Douglas ME, Chafin TK. 2016. *Comparative phytogeography and introgression of catostomids in the Little Colorado River using ddRAD*. Joint Meeting of Ichthyologists and Herpetologists, New Orleans, LA. Oral
4. Chafin TK**,** Douglas MR, Douglas ME. 2015. *Reduced-representation genomics reveals species level phylogeny of* Gila spp. *From the Colorado River basin*. Desert Fishes Council Meeting, Death Valley, CA. Oral.
5. Chafin TK, Douglas MR, Douglas ME. 2016. *Equivocal species boundaries in desert fishes: Gila (Cyprinidae) in the Upper Colorado River Basin*. Desert Fishes Council Meeting, Albuquerque, NM. Oral.
6. Chafin TK, Bangs MR, Douglas MR, Douglas ME. 2016. *Reduced-representation genomics reveals reticulate evolutionary history and species boundaries in* Gila *spp. of the Colorado River.* Joint Meeting of Ichthyologists and Herpetologists, New Orleans, LA. Oral.
7. Fischer, J. (2016). *{Hands on with Jetstream}*. XSEDE '16. St. Louis, MO.
8. Fischer, J. (2016). *{Jetstream Overview - Prepared for 2016 i-Light Conference}*. I-Light Conference.
9. Fischer, J. and Afgan, E. and Doak, T. and Ganote, C. and Hancock, D.Y. and Vaughn, M. (2016). *{Using Galaxy with Jetstream.}*. Galaxy Community Conference. Bloomington, IN.
10. Hancock, D.Y., Stewart, C.A., Fischer, J., Lowe, J.M., Rad, P. & Vaughn, M. (2016). Resource Management from HPC to the Cloud. Proceedings of Accelerated Data and Computing (ADAC) Workshop: Resource Management from HPC to the Cloud. (Lugano, Switzerland, June 2016). Available from: http://computing.ornl.gov/workshops/adac/docs/S3_hancock_ADAC2016.pdf
11. Lowe, John Michael & Budden, Robert (2016). *Deploying OpenStack for The National Science Foundation's Newest Supercomputers*. OpenStack Summit Austin. Austin, TX.
12. Lowe, John Michael, Budden, Robert, & Fischer, Jeremy (2016). *Image is Everything: Dynamic HPC VM Repositories using Murano*. OpenStack Summit Barcelona. Barcelona, Spain.
13. Lowe, John Michael (2016). *Jetstream: A science and engineering cloud*. SuperComputing16. Salt Lake City, UT.
14. Lowe, John Michael (2016). *OpenStack deployment for Jetstream*. SuperComputing16. Salt Lake City, UT.

15. Lowe, John Michael and Turner, George (2016). *{Research Computing on Cloudy Platforms}*. Operating Innovative Networks Workshop.
16. Lowe, J.M. (2016, November). Jetstream: A science and engineering cloud. In D. Fuller (Chair), Ceph in HPC environments. Symposium conducted at SuperComputing16, Salt Lake City, UT.
17. Martin BT, Douglas MR, PLacyk JS, Birkhead RD, Douglas ME. 2016. Phylogenomics and assessment of introgression in the North American Box Turtles (*Terrapene spp.*). Joint Meeting of Ichthyologists and Herpetologists, New Orleans, LA. Oral.
18. Stewart, Craig A. (2016). *{Coevolution of biological research and cyberinfrastructure from now till the end of Moore's law}*. Galaxy Community Conference. .
19. Stewart, C.A. (2014). *"Jetstream: A national science and engineering cloud"*. IEEE/ACM Supercomputing Conference. New Orleans, LA.
20. Stewart, C.A., Hancock, D. Y., & Laherty, J. (2016). Jetstream - In production & supporting science. Presented in IU display in Exhibition Hall at SC'17, Nov 15, Salt Lake City, Utah. Retrieved from http://hdl.handle.net/2022/21090

Other Products

"Other products" as defined by the NSF include: Databases; Physical collections; Audio or video products; Software or netware; Models; Educational aids or curricula; Instruments or equipment; Data and research materials (cell lines, DNA probes, animal models); Evaluation instruments; Survey instruments; Protocols. The depositing and curation of software, particularly VM images available for download and use from IU's persistent digital repository, are very important forms of software products. Those produced to date are listed below:

*Software or Netware*:

1. Boettiger, C. Bayesian Learning Algorithms for Markov Decision Processes. Planned for formal release in 2017; available as beta version from https://github.com/cboettig/mdplearning,
2. Boettiger, C. Ropensci. https://github.com/cboettig/appl
3. Boettiger, C. Rocker-org (Run earlier versions of R using docker) https://github.com/rocker-org
4. Boettiger, C. Rocker. Apps of interest to R users in docker containers. https://hub.docker.com/u/rocker/ (notable as this has > 100,000 pulls to date)
5. CentOS 7 Generic Cloud Raw VM Image Virtual Machine shared on IUScholarWorks.
6. Ubuntu 14.04 Trusty Tahr TLS Sample Image Virtual Machine shared on IUScholarWorks.
7. Ubuntu 14.04.3 Development GUI Virtual Machine shared on IUScholarWorks.
8. Ubuntu 14.04.3 Trusty Tahr Virtual Machine shared on IUScholarWorks.
9. CentOS 6 (6.7) Development Virtual Machine shared on IUScholarWorks.
10. CentOS 7 (7.2) Development Virtual Machine shared on IUScholarWorks.
11. CentOS 6 (6.7) Development GUI Virtual Machine shared on IUScholarWorks.

12. Decoding brain connectomes by encoding anatomical structure in multidimensional tensor-space Virtual Machine shared on IUScholarWorks.
13. Ubuntu 16.04 LTS Cloud Image (Test) Virtual Machine shared on IUScholarWorks.

Other Publications

*Technical papers and reports (not peer reviewed):*

1. Hancock, D.Y., Packard, M., Turner, G. & Stewart, C.A. (2016). Updated acceptance test results for the Jetstream production environment (PTI Technical Report PTI-TR16-005). Bloomington, IN: Indiana University. Retrieved from http://hdl.handle.net/2022/20958

*Presentations made available online (without paywall):*

2. Enis Afgan, Mo Heydarian (2016). *Galaxy CloudMan Performance on AWS*. Presentation given at Johns Hopkins University, Baltimore, MD. Retrieved from http://www.slideshare.net/afgane/galaxy-cloudman-performance-on-aws
3. Fischer, J. (2016). Jetstream Overview - Purdue University. Talk presented at Purdue University, West Lafayette, IN. Retrieved from www.jetstream-cloud.org/files/Jetstream%20-%20Purdue%20Engineering%20Visit.pptx
4. Fischer, J. (2016). Hands on with Jetstream. Tutorial presented at the I2 Technology Exchange, Miami, FL. Retrieved from http://www.jetstream-cloud.org/files/I2%20TechEx%20Tutorial%20-%20Hands%20on%20with%20Jetstream%20.pptx
5. Merchant, N. 2016. Cloud Computing for Education and Research. Talk presented at the Plant and Animal Genome XXIV Conference January 08 - 13, 2016, San Diego, CA. Retrieved from http://hdl.handle.net/2022/21131
6. 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
7. Stewart, C.A. (2016). Cyberinfrastructure from scratch. Talk presented at University of Wyoming, Laramie, WY. Retrieved from http://hdl.handle.net/2022/21103.
8. Stewart, C.A. (2016). IU cyberinfrastructure – supercomputers and clouds in the context of initiatives toward exascale. Talk presented at IU Bloomington Astronomy Department symposium, Bloomington, IN. Retrieved from http://hdl.handle.net/2022/21101.
9. Stewart, C.A. (2016). Jetstream – Performance, Early Experience, and early results. Talk presented at XSEDE16 conference, St. Louis, MO. Retrieved from http://hdl.handle.net/2022/21106.
10. Stewart, C.A. (2016). Watson, Medicine, Health, and IU Grand Challenges. Talk presented at Cook Group, Bloomington, IN. Retrieved from http://hdl.handle.net/2022/21102.

11. Stewart, C.A. (2016). Jetstream - An NSF-funded cloud resource for researchers working in the long tail of science. Talk presented at the University of Wyoming, Laramie, WY. Retrieved from http://hdl.handle.net/2022/21104
12. Stewart, C.A. (2016). Jetstream update for CASC (Coalition for Academic Scientific Computation). Talk presented at meeting of the Coalition for Academic Scientific Computation, Arlington, VA. Retrieved from http://hdl.handle.net/2022/21105
13. Stewart, C.A. and J. Fischer. (2016). Jetstream - In production & ready to go. Talk presented at XSEDE Resource Allocation Committee meeting, Orlando, FL. Retrieved from http://hdl.handle.net/2022/21132.

Thesis/Dissertations

1. Bangs MR. 2016. Fishes as a Template for Reticulate Evolution: A Case Study Involving *Catostomus* in the Colorado River Basin of Western North America. Doctoral Dissertation, University of Arkansas.
2. Steele ML. 2016. Where in the World are Pallas's Fish Eagles? The Migration and Ecology of *Haliaeetus leucoryphus* in Asia. Doctoral Dissertation, University of Arkansas.

Websites

1. Galaxy on Jetstream. https://wiki.galaxyproject.org/Cloud/Jetstream
2. On XSEDE, what is Jetstream? https://kb.iu.edu/d/bfde
3. How should I cite Jetstream in my published work? https://kb.iu.edu/d/addu
4. Accessing user guides for XSEDE digital services. https://kb.iu.edu/d/bazj

## 4. Participants/Organizations

### *4.1.* Individuals

| First Name | Last Name | Most Senior Project Role |
|---|---|---|
| Craig | Stewart | PD/PI |
| Ian | Foster | Co-PD/PI |
| Matthew | Vaughn | Co-PD/PI |
| Nirav | Merchant | Co-PD/PI |
| James | Taylor | Co-PD/PI |
| Enis | Afgan | Other |
| William (Joe) | Allen | Other |
| Brian | Beck | Other |
| Joseph | Bester | Other |
| Ashley | Bucholz | Other |
| Tharon | Carlson | Other |

| | | |
|---|---|---|
| Tim | Cockerill | Other |
| Jeremy | Fischer | Other |
| Steve | Gregory | Other |
| Bret | Hammond | Other |
| David | Hancock | Other |
| Matt | Hanlon* | Other |
| Amit | Juneja | Other |
| Jack | Kordas | Other |
| Andy | Lenards | Other |
| Mattias | Lidman | Other |
| Lee | Liming | Other |
| Peg | Lindenlaub | Other |
| John | Lowe | Other |
| Chris | Martin | Other |
| Nathaniel | Mendoza | Other |
| Andre | Mercer | Other |
| Therese | Miller | Other |
| Mike | Packard | Other |
| Karl | Pickett | Other |
| Julian | Pistorius | Other |
| Akhil | Seth | Other |
| Edwin | Skidmore | Other |
| Patrick | Storm | Other |
| Anusua | Trivedi | Other |
| Steve | Tuecke | Other |
| Nitesh | Turaga | Other |
| George | Turner | Other |

## *4.2.* Partner organizations

*Type of Partner organization:* Academic Institution
*Name*: University of Colorado/CIRES
*Location:* Boulder, CO
*Partner's contribution to the project*: Collaborative research
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.

*Type of Partner organization*: Academic Institution
Name: University of Texas at San Antonio
Location: San Antonio, TX
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.

*Type of partner organization*: Academic Institution
*Name*: University of Arizona
*Location*: Tucson, AZ
*Partner's contribution to the project: Facilities; In-kind support*

The University of Arizona develops the software product called Atmosphere, which is the primary, initial user interface to Jetstream and seamlessly orchestrates cloud resources across the two separate OpenStack sites, located at the Indiana University and TACC. Atmosphere was initially conceived and successfully used by CyVerse (formerly, named iPlant Collaborative) for their users, and has been adapted to Jetstream. During July and September, the Atmosphere development team primarily focused on integrating with the TACC Accounting System (TAS), modifying Atmosphere to the XSEDE allocations model, improving the user experience of the Atmosphere user interface, and numerous enhancements and bug fixes related to Jetstream's specific use cases and requirements. Another significant activity related to end-user and operational support tasks, when issues arose within the Jetstream system.

*Type of Partner organization:* Academic Institution
*Name:* Johns Hopkins University
*Location:* Baltimore, MD
*Partner's contribution to the project: Collaborative research; In-kind support*
*More detail on partner and contribution:* JHU has been working on tightening the integration between the Galaxy application and the Jetstream cloud. A new virtual machine image has been built with the latest Galaxy application release (v16.07). The image provides a toolset that closely matches that of the public Galaxy server available at usefgalaxy.org as well as a complete set of reference genomes available on that same server. Availability of the reference data was realized by enabling a connection to a (CVMFS-based) replica server that offers access to approximately 3TB of reference data (vs. ~50GB in the previous version of the image). An additional benefit of such integration is that the reference data is no longer stored on the machine image, reducing the size of that image to about 10GB, which leaves more space for user's analysis data. Beyond the standalone Galaxy image, JHU is continuing to improve the ability of the Galaxy Main server to offload user jobs onto VM instances running on Jetstream. The process of explicitly acquiring additional worker instances has been streamlined, with the intent to add support for dynamic instance acquisition and release. This feature is currently being experimented with.

*Type of Partner organization:* Academic Institution
*Name:* University of Chicago
*Location:* Chicago, IL
*Partner's contribution to the project: In-kind support; Collaborative research*
*More detail on partner and contribution:* During July through August, UChicago made certain that Jetstream's authentication/login function--which relies on Globus-based XSEDE user authentication--works continuously and satisfies user needs. The UChicago-based Globus user support team actively monitored XSEDE tickets related to Jetstream logins and was involved in debugging and resolving a handful of specific user requests related to this function. UChicago's secondary activity during this period was preliminary work toward future Jetstream features, specifically: (1) Jetstream users' ability to use Globus Transfer to move entire storage volumes to/from Jetstream, and (2) Jetstream users' ability to authenticate to Jetstream's OpenStack APIs using XSEDE identities (i.e., the same credentials they use with Jetstream's Web interface).

*Type of Partner organization:* Academic Institution
*Name:* University of Texas at Austin, Texas Advanced Computing Center (TACC)
*Location:* Austin, TX
*Partner's contribution to the project: In kind support; Facilities; Collaborative research*
*More detail on partner and contribution: TACC supports the project by hosting a portion of the production hardware as well as offering personnel to operate and function as a tier-2 user support.*

*Type of Partner organization:* Academic Institution
*Name:* Cornell University
*Location: Ithaca, New York*

*Partner's contribution to the project: In-kind support*
*More detail on partner and contribution: Creating Jetstream training video for outreach and education purposes.*

*Type of Partner organization:* Academic Institution
*Name:* Penn State University
*Location:* University Station, Pennsylvania
*Partner's contribution to the project: In kind support*
*More detail on partner and contribution:* Anton Nekrutenko, Penn State University, *development of the Galaxy software suite.*

*Type of Partner organization:* Academic Institution
*Name:* University of Hawaii
*Location:* Honolulu, Hawaii
*Partner's contribution to the project: Collaborative research*
More detail on partner and contribution: Gwen Jacobs, University of Hawaii, *serves as an exemplar for colleges and universities across the United States in utilizing Jetstream's unique features for faculty and students as well as developing VM images supporting the research needs in ocean science.*

*Type of Partner organization:* Academic Institution
*Name:* University of North Carolina at Chapel Hill
*Location:* Chapel Hill, North Carolina
*Partner's contribution to the project: Collaborative research*
More detail on partner and contribution: Thomas M. Carsey, University of North Carolina at Chapel Hill, The Odum Institute, *advises on software tools of interest to the social science research community; work to increase awareness of Jetstream to our user community, and serve as a pilot site for the implementation of Jetstream's distributed cloud services.*

*Type of Partner organization:* Academic Institution
*Name:* Jackson State University
*Location: Jackson, Mississippi*
*Partner's contribution to the project: Collaborative research*
More detail on partner and contribution: Jessie Walker, Jackson State University, *is an unfunded partner, leveraging the Jetstream project in support of academic and research endeavors.*

*Type of Partner organization:* Commercial vendor
*Name:* Dell, Inc.
*Location: Round Rock, Texas*
*Partner's contribution to the project: In-kind support*
*More detail on partner and contribution: Configure, manufacture, and support the Jetstream hardware utilized in the test and development as well as the production systems.*

*Type of Partner organization:* Commercial vendor
*Name:* The MathWorks, Inc
*Location: Natick, Massachusetts*
*Partner's contribution to the project: In kind support*
*More detail on partner and contribution: Enable end users of the Jetstream system to run MATLAB™ (and any other MathWorks products the end user is licensed to use). This is accomplished through the "bring your own license" arrangement which MathWorks has piloted with a small number of facilities worldwide. In addition, MathWorks, Inc. is allowing Indiana University to purchase a set of academic*

*licenses of MathWork products which can be made available to any academic user of the Jetstream system.*

### *4.3.* Have other collaborators or contacts been involved?

No

## 5. Impacts

### *5.1.* What is the impact on the development of the principal discipline(s) of the project?

Jetstream has had significant impact on the areas of computational science. In the area of computational science, as a first-of-a-kind production cloud funded by the NSF, Jetstream has been a pilot, a pathbreaker, and a tremendous learning experience. The Jetstream project has published lessons learned in peer-reviewed conference proceedings, which are important ways to have impact on the computational science community. In the early going of this project Jetstream may have had more impact on the community through extensive participation in many working groups, workshops, BOFs, and small meetings and discussions that are shaping the community approach to implementation, provision, and support of publically-funded cloud services for open science.

Three particular examples of significant impact on the field of computational science are as follows:

- The Jetstream team has initiated a multi-organization discussion about authentication to publically-funded clouds, with an aim of establishing a mechanism for multi-system authentication and credential management at the Keystone level of OpenStack
- The Jetstream team has initiated and is leading a multi-organization discussion about mechanisms for management of VM libraries and repositories. The one thing we have learned so far is that at least two mechanisms are likely needed in the future: one is the repository of static, unchanging VM images that we have already established through use of IU's persistent digital repository (Scholarworks.iu.edu). However, for operational work, some sort of scripted, "create on the fly" mechanism is essential so that you can create a current version of a particular VM functionality but incorporate whatever is the current version of relevant components at any given time. Some process involving some sort of automated on-the-fly or periodic rebuilding may be appropriate – e.g. something like use of Ansible scripts to build VMs.
- During PY2 hundreds of students took classes in computational sciences and did independent research in computational sciences based on use of Jetstream.
- Stewart et al. [2015] – the main paper describing plans for the Jetstream system - has already been cited by 5 other published technical papers, and read statistics from Researchgate.net indicate that it has been read more than 100 times just through that venue.
- Stewart et al. [2016] – the paper describing the implementation of, acceptance of, and early experiences with Jetstream – was given the best technical paper award at the XSEDE'16 conference, this indicating both its quality and impact as a contribution to the field of computational science.

## *5.2.* What is the impact on other disciplines?

Jetstream has had impact on research particularly in the biological sciences, where use of Jetstream has enabled researchers to enable analyses of evolution of fish, plants, and rattlesnakes; additionally, genomic analyses are being performed (via Galaxy).

Examples of the impact of Jetstream include:

- Jetstream is being recognized as a functional and reliable basis for computational science and application science research. Fourteen letters of commitment have been written that commit Jetstream resources in support of grant proposals in a number of disciplines. So far one award in the geological sciences and one award in bioinformatics have been made by the NSF in support of research that will be done with Jetstream as a primary computational and data management resource.
- One paper in the area of bioinformatics has been accepted, and two papers in evolutionary biology are in review.
- During PY2 we estimate that at least 100 students in domain sciences used Jetstream in class or independent research

## *5.3.* What is the impact on the development of human resources?

Jetstream is already having significant impact as a resource in the development of human resources, and Jetstream dissemination efforts are playing an important role in providing opportunities for self-directed learning. Key points include:

- Jetstream has been used directly and indirectly by more than 1,000 people to date.
- More than 2,000 people have attended some sort of seminar, tutorial, or talk about Jetstream.
- Hundreds of people have now read peer-reviewed publications thus adding significantly to the development of these individuals as computationally-oriented scientists.
- Every XSEDE staff member was given a short term startup account on Jetstream, providing specific professional development activities for cyberinfrastructure professionals already working in this field with at least some part of their salary funded by the NSF award to operate the eXtreme Science and Engineering Discovery Environment (XSEDE)

One example of the use of Jetstream in computational science is provided by this summary of a class taught during the fall of 2016: "A Fall 2016 data science course at Indiana University Bloomington with close to 200 students used as many Jetstream VM instances over a period of 6 weeks in analysis of datasets. The course, I535-I435-B669 Management, Access, and Uses of Big and Complex Data and taught by Prof. Beth Plale, includes graduate students of computer science and data science as well as undergraduate informatics majors. The course had all 195 students register accounts in the single. allocation TG-CIE160033. The students implemented two projects on Jetstream using images I535- 435-B669 Project A and I435-I535-B669 Project B. Students used their VMs over a period of about 6 weeks. The learning objectives of the project included: Set up and use a cloud hosted virtual machine on Jetstream; use a MongoDB database; Manipulate software tools that are installed at the cloud hosted sites to do analyses and produce a specific visualized result "

### *5.4.* What is the impact on physical resources that form infrastructure?

Nothing to report.

### *5.5.* What is the impact on institutional resources that form infrastructure?

Nothing to report.

### *5.6.* What is the impact on information resources that form infrastructure?

The implementation of Jetstream has set in motion what will eventually be a centralization of cyberinfrastructure resources supporting the geological sciences, simultaneously improving services and decreasing costs to the government and taxpayer. IRIS and UNAVCO are of the largest disseminators of geological information of geological information in the US; both operate with budgets in the several millions of dollars per year. IRIS is a collector and disseminator of seismic data. UNAVCO deals with geomorphology. Both are important sources of data for researchers and the public sector. For example, seismic data and geomorphology data play important roles in civic planning – particularly creation of new buildings and new areas of growth for cities and towns. Data from UNAVCO and IRIS help planners determine if a proposed subdivision will likely be flooded or if a large building in a city is proposed for an area at high risk of an earthquake. Both organizations deliver data used in emergency management planning. For decades, UNAVCO and IRIS have each run their own independent cyberinfrastructure systems. IRIS and UNAVCO have both determined that their existing resources for distributing information are insufficient to meet current and projected demand, and those resources are not available at present on a continuous, ongoing basis. IRIS and UNAVCO are now both planning to adopt use of Jetstream as their mechanism for disseminating data to the research community, and work to put this into practice has already begun. Both UNAVCO and IRIS have requested and received large awards of time on Jetstream. The plan of record is for both organizations to improve their services by abandoning their local, modest-scale data delivery systems and make use of Jetstream instead.

### *5.7.* What is the impact on technology transfer?

There has been important transfer of information from the Jetstream project to major open source technology projects, particularly OpenStack. There have been no formal invention disclosures yet, but Jetstream staff and research faculty are playing an important role in the development of OpenStack. OpenStack itself is continually transferred from its open source repository to anyone in the US that wishes to download, install, and use it.

### *5.8.* What is the impact on society beyond science and technology?

Through our impact on the OpenStack project, we are playing an important role in the ongoing improvement of the open source cloud technology that is the most widely used open source cloud environment for small business and industry. This aids the development of the cloud-based economy in the US generally.

# 6. Changes/Problems

## 6.1. Changes in approach and reasons for change

There have been no significant changes in approach. One notable addition to the operational processes used by the Jetstream team to manage the system is round the clock use of a Slack channel to enable ongoing communication and resolution of problems by all technical staff involved in system operations. Slack is a new technology and it represents a good way for anyone involved in running the system at any time to say "does anyone have an answer to the following question" and get an answer in real time from any other staff member online anywhere. Slack has proved to be particularly helpful in enhancing operational communications in the highly distributed Jetstream project.

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

We do not anticipate any delays in the future. During PY2 we experienced two planned delays and one unplanned delay, all of which are not resolved.

| Entity | Start | End | Duration | Planned/ Un-planned | Service, System, or Components | User Impact |
|---|---|---|---|---|---|---|
| Public announcement of NSF acceptance of Jetstream | 6/1/2016 | 9/1/2016 | 3 months | Planned | Given that the largest part of the user community we proposed as the main users of Jetstream would be distracted by the summer research season, we decided in consultation with the NSF to delay a major public launch and information campaign about Jetstream going into production mode until 1 September 2016 – timed to coincide with the beginning of fall semester at many colleges and universities in the US | This delay has had a net benefit in increasing user awareness and uptake of Jetstream. While this decision delayed a major public information campaign, it has made that public information campaign much more effective than it would have been had we proceeded with it over the summer |
| Processing of usage against allocations in the XSEDE user database for Jetstream generally | 6/1/2016 | 9/1/2016 | 3 months | Planned | Normally a declaration by the NSF that a CI resource had entered production would trigger a start of usage of that system being charged against allocations on that system within the XSEDE central database. Since we did not make a major announcement about Jetstream moving into production mode at the time this happened, we decided to delay charging of use | The impact to users was either nonexistent or positive. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | against allocations until the beginning of September, because to do otherwise would have been confusing to users and potentially have a negative impact on user perceptions of the system's value. | |
| Processing of usage against allocations in the XSEDE user database for Jetstream-TACC | 9/1/2016 | 9/24/2016 | 24 days | | Data within the OpenStack data management system were lost for the period between 9/1 and 9/24 2016. This was an inadvertent event related to the general difficulty in use and complexity of the OpenStack user data collection system. The root cause was the complexity and lack of clarity of the OpenStack documentation; extensive staff investigation of the system behavior that led to the loss of data was immediate correction and increased transfer of information and experience between staff is the ongoing preventive measure that we believe will avoid any recurrence. | Given the relatively limited duration of the data loss, there was no significant impact on users. The only impact on users is what amounts to an effective slight increase in the amount of resource allocated to users, since some usage was not charged against their allocations for the 24 days for which data was lost. |

## 6.3. Changes that have significant impact on expenditures

No significant changes in expenditures; execution of activities and expenditure of funds are proceeding according to plan.

## 6.4. Significant changes in use or care of human subjects

Nothing to report

## 6.5. Significant changes in the use or care of vertebrate animals

Nothing to report

## 6.6. Significant changes in the use or care of biohazards

Nothing to report

## 7. References


[1] V. Venkatesh, M. G. Morris, F. D. Davis, and G. . Davis, "User Acceptance of Information Technology: Toward a Unified View," *MIS Q.*, vol. 27, no. 3, pp. 425–478, 2003.

[2] D. Blankenberg, G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko, and J. Taylor, "Galaxy: a web-based genome analysis tool for experimentalists," *Curr Protoc Mol Biol*, p. Unit 19 10 1-21, 2010.

[3] "CyVerse." [Online]. Available: http://www.cyverse.org/. [Accessed: 15-Aug-2016].

[4] Indiana University, "IUScholarWorks." [Online]. Available: https://scholarworks.iu.edu. [Accessed: 29-Apr-2014].

[5] C. A. Stewart, D. Y. Hancock, M. Vaughn, J. Fischer, L. Liming, N. Merchant, T. Miller, J. M. Lowe, D. Stanzione, J. Taylor, and E. Skidmore, "Jetstream - Performance, Early Experiences, and Early Results," in *Proceedings of the XSEDE16 Conference*, 2016.