Jetstream: A Cloud System Enabling Learning in Higher Education Communities

Jeremy Fischer
Indiana University
Pervasive Technology Institute
Bloomington, IN
jeremy@iu.edu

David Y. Hancock
Indiana University
Pervasive Technology Institute
Bloomington, IN
dyhancoc@iu.edu

John Michael Lowe
Indiana University
Pervasive Technology Institute
Bloomington, IN
jomlowe@iu.edu

George Turner
Indiana University
Pervasive Technology Institute
Bloomington, IN
turnerg@iu.edu

Winona Snapp-Childs
Indiana University
Pervasive Technology Institute
Bloomington, IN
wsnappch@iu.edu

Craig A. Stewart
Indiana University
Pervasive Technology Institute
Bloomington, IN
stewart@iu.edu

ABSTRACT

Jetstream is the first production cloud funded by the NSF for conducting general-purpose science and engineering research as well as an easy-to-use platform for education activities. Unlike many high-performance computing systems, Jetstream uses the interactive Atmosphere graphical user interface developed as part of the iPlant (now CyVerse) project and focuses on interactive use on uniprocessor or multiprocessor. This interface provides for a lower barrier of entry for use by educators, students, practicing scientists, and engineers. A key part of Jetstream's mission is to extend the reach of the NSF's eXtreme Digital (XD) program to a community of users who have not previously utilized NSF XD program resources, including those communities and institutions that traditionally lack significant cyberinfrastructure resources. One manner in which Jetstream eases this access is via virtual desktops facilitating use in education and research at small colleges and universities, including Historically Black Colleges and Universities (HBCUs), Minority Serving Institutions (MSIs), Tribal colleges, and higher education institutions in states designated by the NSF as eligible for funding via the Experimental Program to Stimulate Competitive Research (EPSCoR). Jetstream entered into full production in September 2016 and during the first six months it has supported more than a dozen educational efforts across the United States. Here, we discuss how educators at institutions of higher education have been using Jetstream in the classroom and at student-focused workshops. Specifically, we explore success stories, difficulties encountered, and everything in between. We also discuss plans for increasing the use of cloud-based systems in higher education. A primary goal in this paper is to spark discussions between educators and information technologists on how to improve using cloud resources in education.

SIGUCCS '17, October 1–4, 2017, Seattle, WA, USA

ACM ISBN 978-1-4503-4919-2/17/10…$15.00
https://doi.org/10.1145/3123458.3123466

CCS CONCEPTS

• **Computer systems organization** → **Cloud computing**; • **Applied computing** → *Computer-assisted instruction*; *Interactive learning environments*;

KEYWORDS

cloud; openstack; digital; XSEDE; education; outreach; training; EOT; research; Jetstream; XD; Globus; Atmosphere; cyberinfrastructure

ACM Reference format:
Jeremy Fischer, David Y. Hancock, John Michael Lowe, George Turner, Winona Snapp-Childs, and Craig A. Stewart. 2017. Jetstream: A Cloud System Enabling Learning in Higher Education Communities. In *Proceedings of SIGUCCS '17, Seattle, WA, USA, October 1–4, 2017,* 6 pages.
https://doi.org/10.1145/3123458.3123466

1 INTRODUCTION

Jetstream is a recent addition to the XD (eXtreme Digital) national cyberinfrastructure (CI) funded by the National Science Foundation [20]. Previous awards have been aimed at traditional large-scale high-performance computing (HPC) systems such as Blue Waters, Kraken, and Stampede [15–17] These HPC systems have been utilized by thousands of researchers at almost 400 academic and research institutions in the last five years. [10]

Over time, the focus of the awards has changed somewhat. While the need for HPC resources has not decreased, the desire for different types of computing resources has also been seen as a priority for the national CI. This started with the NSF 13-528 (Track IIf) [18] awards for Comet and Wrangler, showing a diversification into heterogeneous HPC and virtual cluster systems and big data processing respectively.

The next step in these awards was looking at novel approaches to research computing. The solicitation for the NSF 14-536 (Track IIg) proposal asked applicants to "include capabilities suitable for addressing emerging computationally intense scientific and engineering research topics, workflows and communities that are not

Table 1: Survey Instrument Questions

Item	Answer Type
In what context did you use Jetstream?	Course, Workshop, Other
In what context did you use Jetstream? - Other Text	Open-ended Response
What was the course name or workshop/event title?	Open-ended Response
Institution	Open-ended Response
Location	Open-ended Response
Primary target audience of your class/workshop?	K-12 students, undergraduate students, graduate students, researchers, technical staff, faculty, other
How many students/participants did you have?	Open-ended Response
Do you consider the use of Jetstream in your class/workshop a success?	Yes/No
Considering the level of success (or failure) you experienced in using Jetstream in your class/workshop, please elaborate on your experience.	Open-ended Response
In your view, what did Jetstream do especially well in the context of your class/workshop?	Open-ended Response
Considering its educational uses, what aspects of Jetstream could be improved?	Open-ended Response

optimally served by current XD or Blue Waters resources." [19] While not specifically discussing cloud technologies, this was seen as a logical progression for NSF-funded research and education systems. The two winning awards, Jetstream and Bridges, incorporate cloud solutions as part or, in the case of Jetstream, all of their offerings to the US research and education communities.

Though exploring new methods and systems for research was a primary focus, the underlying mandate was to "less traditional computational science communities" [19] where ease of use would be a crucial aspect. In this regard, ease of use could mean both providing a graphical user interface (GUI) as well as creating an application library that would be useful and usable for researchers and educators. Jetstream outlined both of these as primary goals from the start [23] as well as other features and services designed to facilitate the use of Jetstream by educators and researchers who typically did not use the eXtreme Science and Engineering Discovery Environment (XSEDE) resources for their work [9]. Expanding cloud services and creating research-as-a-service will further the goal of creating a programmable cyberinfrastructure that will be flexible in meeting the needs of educators and researchers throughout the United States. Through the course of this paper, we examine what the Jetstream team has attempted to do to enhance educational use of Jetstream and look at use cases where varying degrees of success were achieved.

2 JETSTREAM DESIGN AND FEATURES

Since the mandate was to create a different type of resource for research and education, the Jetstream team took a different approach from the traditional HPC models. Providing on-demand cloud services wasn't enough to differentiate Jetstream from commercial offerings. Jetstream was designed from the start to focus on science, supported and run by staff that are experienced in supporting researchers already. One very important design goal was to make a more user-friendly research and education cloud. Jetstream partnered from the start with iPlant (now CyVerse) to adapt the Atmosphere web-based GUI for creating, managing, and accessing virtual machines (VM) [14]. This removed the need for researchers and educators to be experts in the methodologies to create and manage VMs. In addition, creating VMs with GUI desktop environments to let people have an environment similar to the desktop environments they are used to already was key for creating a more inviting user experience. Another key feature for making Jetstream use easier was creating a set of curated VMs. The Jetstream team manages a set of "Featured Images" that range from basic Linux operating systems that are routinely updated and patched for security items to images with research software such as Matlab, Galaxy, and R/RStudio. Jetstream users can use these stable and managed VMs to build their own customized workflows and preserve them for future use, creating reproducible, consistent environments for their research or classes and for the Jetstream user community if they choose.

3 USING JETSTREAM FOR EDUCATION

Jetstream's goal is to "provide self-serve academic cloud services, enabling researchers or students to select a virtual machine (VM) image from a published library" or to "customize their own virtual environment." This was the first step to creating a system not only suited for research but also instructional use [23]. Beyond that, the goals for providing for the best electronic learning environment are designing for interactivity, quality content presentation, and appropriate GUI [7]. A key feature of Jetstream is that it allows educators and researchers to create the environment tailored for their needs using command line interface (CLI) or GUI.

One of the best ways to expand the use of new resources beyond the typical researchers and research areas is to use the system to educate other researchers, educators, and students. The Jetstream education, outreach, and training (EOT) team has worked with members of various institutions, including both of Jetstream's physical hosts, Indiana University and University of Texas at Austin, as well as a number of others. Multiple workshops on using Jetstream have taken place at conferences and afternoon seminars to attract researchers and educators to Jetstream by showing its ease of use and discussing how Jetstream would benefit them.

One of the desired side effects of these workshops and seminars is for researchers and educators to create their own XSEDE startup allocations for Jetstream. Though there is no reported metric for defining the number of allocations applied for immediately after a workshop or seminar, anecdotal evidence shows that every workshop does yield a minimum of 2-3 new allocations shortly afterward from attendees. The EOT team, though, is constantly looking for opportunities to expand the reach of Jetstream to new audiences by conducting workshops or partnering with those who teach courses or workshops and helping them move their projects to Jetstream.

3.1 Overview of Education Allocations

Jetstream utilizes the XSEDE Resource Allocation System (XRAS) for awarding allocations. The process for obtaining an Education Allocation is fairly simple, requiring a current curriculum vitae (CV), the course description/syllabus, and a justification of resources [24]. From March 2016 to March 2017, there were twenty-three education allocations awarded on Jetstream totaling over 2.5 million cpu hours [10]. These allocations were awarded for educational program development and limited engagement workshops, as well as for semester-long courses. We will examine some of these varying educational uses as a basis for shaping Jetstream's role as an education resource.

4 MATERIALS AND METHODS

4.1 Participants

Participants for this survey were selected from the list of active Jetstream education allocations; as of March 2017 there were 23 active education allocations on Jetstream. Potential participants were principal investigators (PIs), co-principal investigators (Co-PIs), or key personnel who (as of March 31, 2017) had used or were actively using Jetstream resources to conduct workshops and courses using Jetstream cloud resources for instruction and course assignments. There were fourteen potential participants identified as having used at least a portion of their allocation (more than 0 Service Units, the "currency" of XSEDE, utilized); nine individuals agreed to participate and did so with informed consent. The courses and workshops covered in the survey results that gave permission to disclose are shown in Table 2.

4.2 Survey Instrument

The goal of the survey was to find how Jetstream was used, what went well, what could be improved, and if there were ways that educators felt cloud services in general could be made more attractive for classroom use. This survey was approved by the Human Subjects Office at Indiana University. Survey instrument items and types of responses are in Table 1.

4.3 Survey Administration

After potential participants were identified, they were sent an invitation to participate over email. A follow-up e-mail was sent one week after the initial invitation. A final e-mail was sent out two weeks after the initial request. This survey was unusual in that participants were able to share their identities and all but one did so.

5 RESULTS

Of the 14 active and used or in-use allocations identified, nine individuals (64.29%) agreed to participate. 100% of respondents reported that their use of Jetstream was a success. Several of the allocations had multiple PIs. Of the allocations, two had two respondents each, so results represent six of the allocations. Of these six, one specified that they did not want their results disclosed. One completed the survey but did not include enough information to discuss their workshop in any meaningful fashion. Thus, only four responses are presented below by individual course/workshop.

5.1 Course Review 1

The largest educational use of Jetstream to date was for the course from the Indiana University School of Informatics – "I535 Management, Access, and Use of Big and Complex Data". This course covered topics including "data integration, workflows and pipelines, and distributed noSQL stores" [21]. This course had almost 200 undergraduate and graduate students enrolled and utilized multiple customized images for two projects spanning six weeks of the semester. The instructors installed Hadoop and a modest data lake for doing data retrieval (MapReduce) exercises. They also utilized MongoDB for student exercises using noSQL databases.

This particular use-case started as a discretionary allocation for research awarded by the PI of Jetstream to a faculty partner, Dr. Plale, at Indiana University. The PI expected that the allocation would be used for prototyping Jetstream use for various projects that Dr. Plale was involved with, and not in a classroom setting. The first indication that it was being used for instructional use came when a couple of support requests came in from students in that course; however, it was not until there was a network outage on the IU-based Jetstream cloud that we become aware that there were so many students enrolled in that class.

This course tested Jetstream's limits early in its production. Having several hundred new users pushed on to the system, launching virtual machines (VMs), working with them, and suspending them helped us make Atmosphere more robust for both the users and support. There were several incidents where we had to work with the instructors to quickly solve unexpected issues related to the sudden growth and use of Jetstream. The instructors indicated that they were very pleased with the support provided by Jetstream for their students. They also expressed satisfaction overall with the use of Jetstream in their course. One felt that Jetstream could make tutorials available for students to help make Jetstream easier to learn to use. While the end results were very positive, we learned some valuable lessons about support, expectation management, handling outages, and generally trying to help users with the system.

5.2 Workshop Review 1

By contrast to the longer course by Dr. Plale described above, Dr. C. Titus Brown, Dr. Harriet Alexander, and Dr. Phillip Brooks led a recent two-day course on metagenome analysis at University of California at Santa Cruz [5]. This workshop has been held for

Table 2: Courses and Workshops Covered in Survey

Event Title	Type	Number of Students	Location	Audience
Analysis of large, complex biobehavioral, bioinformatic, and genomic data sets	Workshop	36	Brandeis University	Graduate students
Bioinformatics: Tools for Genome Analysis	Course	9	Online (via Johns Hopkins University)	Graduate students
'Digital Pedagogy' / 'Introduction to Text Analysis'	Course	30	University of Pittsburgh / Carnegie Mellon University	Graduate students
Management, Access, and Uses of Big and Complex Data	Course	200	Indiana University	Undergraduate and graduate students
Data-Driven Neuroimaging	Workshop	30	University of California San Francisco	Researchers and graduate students
2017 Metagenomics Workshop	Workshop	30	University of California Santa Cruz	Researchers, graduate students, and undergraduates

multiple years prior using Amazon Web Services (AWS) for the VM component. For the workshop, they used a base Ubuntu image and added several software packages like trimmomatic and fastqc for working with sequencer data, MEGAHIT assembler for assembling the metagenomes, Prokka for annotating them, and Jupyter Notebooks for doing the workshop exercises [6].

A two-day workshop doesn't allow much room for error on the part of the technology provider. While Dr. Brown was prepared for his part of the workshop, it was up to the Jetstream staff to be ready and able to handle any support issues. Due to staffing and budget constraints, Jetstream currently does not have a 24/7/365 support staff. This means that unless prior arrangements are made, there is a potential for support breakdown for courses or workshops held on the weekend. Fortunately, there were no issues of consequence and Dr. Brown reported that using "Jetstream went more smoothly than AWS in almost every way and seems to perfectly meet our needs for training!" [5]. He reported minor issues with the Jetstream web desktop and a few stalled instance startups, but no major issues.

Overall, this workshop was considered a success. A longer data-intensive workshop is scheduled for this summer using Jetstream. This will have up to 100 users for 30 days using Jetstream for sequencing and data analysis [4]. While this will not be the largest course or workshop to date, it will most likely be the most intense in terms of resources required and simultaneous load. This will be another milestone for Jetstream brought about by an educator well versed in using cloud technologies.

5.3 Workshop Review 2

The Berkeley Institute for Data Science (BIDS) and the University of Washington's eScience Institute teamed up with UCSF researchers to conduct a workshop on data-driven analysis and machine learning for neuroscience imaging data [8]. The demand for this workshop was high, with 60 applicants vying for 30 spots [3]. This workshop required a considerable amount of coordination amongst the three institutions involved, including the technical work before the actual workshop.

In order to utilize as much of the workshop time for actual learning and research techniques, IT staff created the workshop virtual machines in advance in a pre-loaded, customized Docker container that contained Jupyter Notebook, a selection of Python environments, tools, and libraries, as well as a selection of R utilities, based on the Jupyter Notebook Data Science Stack [22]. Additional neuroscience-specific packages such as Dipy [11]. The workshop coordinators then created a VM for each participant and provisioned the customized container on each VM so participants entered the workshop ready to go almost immediately.

This ensured that each person had the most up-to-date environment for the workshop. No installation time was required and no downloading of data or other tools was necessary. The workshop organizers wanted to remove this from their workshop as the process is often error prone [8]. While this could have been done using VM images, the process of using Docker containers worked well for the IT staff supporting the workshop creation. This allowed the the staff to have a consistent working environment for each participant and also saved time by having all data sets needed for the project preloaded into the container rather than being downloaded by each participant.

The workshop was considered a success and plans are being made for future versions. One respondent reported that better resource monitoring and management would help in the overall control of the workshop. With IT staff building and managing the workshop, it was a workable solution, but a goal for their organization is to make instructors more self-sufficient. Better usage and monitoring tools might make that a more achievable goal. Aaron Culich of Berkeley stated that one of Jetstream's strengths is providing "fast and flexible access to computational resources at no cost" and that it was a benefit to researchers to use in workshops because they could apply for their own allocations afterward.

5.4 Workshop Review 3

The last workshop examined was organized by Dr. Francesco Pontiggia, Research Computing specialist at Brandeis University, and taught by Dr. Kene Piasta. This workshop was a week of classroom

instruction followed by three months of independent project work. There were thirty-five graduate students enrolled with the purpose of learning methods for conducting statistical analysis of large, complex biobehavioral, bioinformatic, and proteomic data sets.

The VMs for the classroom portion were set up by Dr. Pontiggia in advance. The six shared VMs used R for the statistical analysis and then various other software packages depending on the particular track. The instructors also held a training session on using Jetstream as part of this classroom time to prepare the students for their independent project work. For the project work itself, they created a custom image with all the bioinformatics packages they expected students to need. However, as Dr. Pontiggia noted in the survey, "giving students and instructors the ability to modify, install software on their own and maintain access on demand later on to those images is something of great value, which would not have been easily achieved on our cluster." This flexibility for creating customized, savable, reproducible workflows is part of the design goal of Jetstream.

One issue during this workshop was in the shared VMs used during the classroom portion. In order to control the environment, students used shared VMs; however, individual accounts had to be added (because there was no easy way to use login credentials against the launched VMs inside another account). The workshop organizers reported that having some form of service to facilitate this, whether it's a lightweight directory access protocol (LDAP) server or something other service, it would allow for a shared service inside a specific user account. Nevertheless, overall, the organizers reported that Jetstream provided a comfortable, usable environment for the students. The participants, in general, were not extremely experienced computationally. Dr. Pontiggia in particular felt that having resources available on demand that had a look and feel like a normal workstation made it easier for the students to adapt to using analysis tools.

6 SUGGESTIONS FOR IMPROVEMENT

While the survey respondents all reported that their use of Jetstream was a success, it is important to the evolution and betterment of the system to find out how they felt we could improve the user experience. There were several requests for technical features like allowing persistent internet protocol (IP) addresses in the Atmosphere web interface (this is on the features roadmap but was not implemented at the time of writing). Other suggestions were for things like improved base images for instructional use with specific software. All technical requests are considered, though with budget and staffing being limited, the goal remains to choose those which provide maximum benefit to the most users.

One common thread in regards to improvement was how to better monitor users and usage. In Jetstream's Atmosphere interface, there is no easy way for non-admin users to see what users on their allocation are doing. We have had one successful experiment with using a community account for a two-day workshop. This is something we will explore further in the future. This solution, however, may not be advisable for longer courses where grading is based on coursework done on Jetstream as the potential for academic dishonesty is high. Other solutions may be found for this in the long run.

The last common suggestion was to improve video tutorials for using Jetstream. While we do provide fairly extensive documentation on our wiki [13], it is common practice now to provide more video tutorials for instruction. We plan to continue adding to the videos that we have created [12], looking at a range of areas from using the Atmosphere interface to utilizing more advanced application programming interface (API) techniques for working with Jetstream.

7 ENGAGING EDUCATORS

As with any new technology, the task of getting educators to examine and integrate it into their instructional environment required a concerted and multifaceted effort. Our survey respondents had several things to say about how to best engage educators. While this is a very small group and won't represent all educators, their contributions are valuable. Several of the educators noted that Jetstream simply needed "to get the word out more!" They didn't necessarily offer ideas on how best to do that, though. The Jetstream team has made efforts over the last year to appear at major conferences such as the Plant and Animal Genome (PAG) conference [2] and SuperComputing [1]. We are now targeting domain-specific conferences as well as doing direct outreach and communication, including tutorials, with universities and colleges. However, engaging a larger audience in a survey to find out what broader needs are for technology in education might be prudent. Given the very different needs from classroom to classroom and discipline to discipline, getting the input directly from educators at every level and a variety of institutions might help Jetstream understand how we can best serve the larger audience.

Engaging educators directly is crucial. At the same time, it is also important to engage those who support and train educators to use information technology (IT) components and infrastructure. For example, within the XSEDE framework, Campus Champions act as the liaisons between faculty and staff and XSEDE resources. They provide guidance and training on how to utilize XSEDE resources. This same approach might be worth exploring for Jetstream. If we engage and train the IT staff, they may suggest it as a means of enhancing classroom instruction as well as help their faculty use it. The difficulty again comes in how to best get in contact and train this particular group of people. Looking at EDUCAUSE and SIGUCCS amongst other conferences seems to be a potentially good way to identify and communicate with this group of people.

8 CONCLUSIONS

Initial educational uses of Jetstream have been successful. While there have been minor issues along the way, for the most part, instructors reported that Jetstream met their needs and did so in an easy-to-use, easy-to-implement way for their courses and workshops. In the end, though, the most user friendly system ever created will not be a success without actually having users. We must continue to find ways to increase adoption and usage. Finding new audiences must be a continuing effort. Afterwards, we must follow up to find how we can help individuals use Jetstream effectively in their instruction.

With resources and IT staff being in demand and also constrained at many institutions, utilizing cloud solutions may be of strategic and practical importance. Cloud services like Amazon and Azure are well known and offer introductory usage but often have other strings attached. Jetstream provides a smaller, more limited service but one that is free for US-based researchers and educators. Jetstream also endeavors to be easy to use and responsive to the needs of diverse research and education communities. These qualities give Jetstream the potential to be an extremely valuable tool for the US academic world.

9 ACKNOWLEDGMENTS

Implementation of Jetstream is supported by NSF award #1445604. The Indiana University Pervasive Technology Institute was established with support of funding from the Lilly Endowment, Inc., and partially supported this research. Any opinions expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation, the Lilly Endowment, or XSEDE leadership as a whole.

REFERENCES

[1] [n. d.]. International Conference for High Performance Computing, Networking, Storage and Analysis. http://dl.acm.org/citation.cfm?id=3014904. ([n. d.]).
[2] [n. d.]. International Plant & Animal Genome Conference. http://www.intlpag.org/. ([n. d.]).
[3] Berkeley Institute for Data Science. 2017. Berkeley and UW Data Scientists Team Up with UCSF Researchers to Deliver Data-Driven Analysis and Machine Learning for Neurosciences. https://bids.berkeley.edu/news/berkeley-and-uw-data-scientists-team-ucsf-researchers-deliver-data-driven-analysis-and-machine. (February 2017).
[4] C. Titus Brown. 2017. Request for Compute Infrastructure to Support the Data Intensive Biology Summer Institute for Sequence Analysis at UC Davis. http://ivory.idyll.org/blog/2017-dibsi-xsede-request.html. (February 2017).
[5] C. Titus Brown. 2017. A (revised and updated) shotgun metagenome workshop at UC Santa Cruz. http://ivory.idyll.org/blog/2017-metagenomics-at-ucsc.html. (2017).
[6] C. Titus Brown. 2017. Short read quality and trimming. https://2017-ucsc-metagenomics.readthedocs.io/en/latest/quality.html. (February 2017).
[7] Alexandru Butoi, Nicolae Tomai, and Loredana Mocean. 2013. Cloud-based mobile learning. *Informatica Economica* 17, 2 (2013), 27.
[8] Aaron Culich. 2017. Jetstream cloud support for multi-institutional data science workshops and research. http://research-it.berkeley.edu/blog/17/02/17/jetstream-cloud-support-multi-institutional-data-science-workshops-and-research. (February 2017).
[9] Jeremy Fischer, Steven Tuecke, Ian Foster, and Craig A. Stewart. 2015. Jetstream: A Distributed Cloud Infrastructure for Underresourced Higher Education Communities. http://doi.acm.org/10.1145/2753524.2753530. 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. https://doi.org/10.1145/2753524.2753530
[10] Thomas R. Furlani, Barry L. Schneider, Matthew D. Jones, John Towns, David L. Hart, Steven M. Gallo, Robert L. DeLeon, Charng-Da Lu, Amin Ghadersohi, Ryan J. Gentner, Abani K. Patra, Gregor von Laszewski, Fugang Wang, Jeffrey T. Palmer, and Nikolay Simakov. 2013. Using XDMoD to Facilitate XSEDE Operations, Planning and Analysis. http://doi.acm.org/10.1145/2484762.2484763. In *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery (XSEDE '13)*. ACM, New York, NY, USA, Article 46, 8 pages. https://doi.org/10.1145/2484762.2484763
[11] Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, and Ian Nimmo-Smith. 2014. Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics* 8 (2014), 8. https://doi.org/10.3389/fninf.2014.00008
[12] Indiana University. 2017. Jetstream - cloud-based computation videos. https://www.youtube.com/playlist?list=PLqi-7yMgvZy-0Sy94uyOm67MhvN6xgxZ3. (2017).
[13] Jetstream. 2017. Jetstream Help Wiki. https://wiki.jetstream-cloud.org. (2017).
[14] Nirav Merchant, Eric Lyons, Stephen Goff, Matthew Vaughn, Doreen Ware, David Micklos, and Parker Antin. 2016. The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLOS Biology* 14, 1 (01 2016), 1–9. https://doi.org/10.1371/journal.pbio.1002342
[15] National Science Foundation. 2007. Leadership Class Scientific and Engineering Computing: Breaking Through the Limits. https://www.nsf.gov/awardsearch/showAward?AWD_ID=0725070. (2007).
[16] National Science Foundation. 2007. A National Institute for Computational Sciences to Provide Leading-Edge Computational Support for Breakthrough Science and Engineering Research. https://nsf.gov/awardsearch/showAward?AWD_ID=0711134. (2007).
[17] National Science Foundation. 2011. Enabling, Enhancing, and Extending Petascale Computing for Science and Engineering. https://nsf.gov/awardsearch/showAward?AWD_ID=1134872. (2011).
[18] National Science Foundation. 2013. High Performance Computing System Acquisition: Building a More Inclusive Computing Environment for Science and Engineering. https://www.nsf.gov/pubs/2013/nsf13528/nsf13528.htm. (2013).
[19] National Science Foundation. 2014. High Performance Computing System Acquisition: Continuing the Building of a More Inclusive Computing Environment for Science and Engineering. https://www.nsf.gov/pubs/2014/nsf14536/nsf14536.htm. (2014).
[20] National Science Foundation. 2014. High Performance Computing System Acquisition: Jetstream - A Self-Provisioned, Scalable Science and Engineering Cloud Environment. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1445604. (2014).
[21] Beth Plale. 2016. Course Description: I535 Management, Access, and Use of Big and Complex Data. https://www.soic.indiana.edu/graduate/courses/index.html?number=i535&department=INFO. (2016).
[22] Project Jupyter. 2017. Jupyter Notebook Data Science Stack. https://github.com/jupyter/docker-stacks/tree/master/datascience-notebook. (February 2017).
[23] 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. http://doi.acm.org/10.1145/2792745.2792774. In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure (XSEDE '15)*. ACM, New York, NY, USA, Article 29, 8 pages. https://doi.org/10.1145/2792745.2792774
[24] XSEDE. 2012. Startup and Education Allocation Proposals. https://portal.xsede.org/allocations-overview#writing-startupeducation. (2012).