My Desktop
Prepare & Submit Proposals
Prepare and Submit Proposals
(Limited proposal types)
Demo Site: Prepare Proposals
Prepare Proposals in FastLane
Proposal Status
Awards & Reporting
Notifications & Requests
Project Reports
Award Functions
Manage Financials
Program Income Reporting
Grantee Cash Management Section Contacts
Administration
Look Up NSF ID

Preview of Award 1541215 - Final Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1541215
Project Title:	CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation
PD/PI Name:	David A Lifka, Principal Investigator Thomas R Furlani, Co-Principal Investigator Richard Wolski, Co-Principal Investigator
Recipient Organization:	Cornell University
Project/Grant Period:	10/01/2015 - 09/30/2021
Reporting Period:	10/01/2020 - 09/30/2021
Submitting Official (if other than PD\PI):	David A Lifka Principal Investigator
Submission Date:	09/29/2021
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	David A Lifka

Accomplishments

*** What are the major goals of the project?**

The major goals established for the Aristotle project were:

- Implement a scalable and sustainable multi-institutional cyberinfrastructure cloud federation model that provides data

analysis building blocks in support of multiple research disciplines requiring flexible workflows and analysis tools for large-scale data sets. Federation sites are Cornell University, University at Buffalo, and University of California, Santa Barbara.

- Support seven strategic science use cases from intentionally diverse disciplines (earth and atmospheric science, finance, chemistry, astronomy, civil engineering, microbiome, and agriculture) to demonstrate the potential of a federated cloud as a campus bridging paradigm. Explore data analysis techniques and their applicability to different disciplines. Document tools, workflows, challenges, and best practices for each use case.
- Encourage and reward data analysis resource sharing with a new allocations and accounting model that provides a fair exchange mechanism for resource access between and across multiple institutions.
- Develop and build a new tool for cloud metrics into Open XDMoD.
- Develop DrAFTS (Durability Agreement From Time Series) statistics to make online forecasts of cost and performance for Amazon Web Services and Aristotle Clouds.
- Investigate three container technologies (Docker, Singularity, X-Containers) and identify pain points experienced by users when selecting and implementing these technologies and Kubernetes orchestration for scientific software. Share lessons learned and best practices in a technical report and present the findings at a tutorial that works through the building and running of non-trivial containers.
- Containerize application kernels and test their performance on Google Cloud versus Aristotle Clouds, Comet, Bridges, Stampede 2, Bridges-2, and Expanse. Publish the performance comparisons for the cyberinfrastructure and research community.

*** What was accomplished under these goals and objectives (you must provide information for at least one of the 4 categories below)?**

Major Activities:

Investigated HPC Container Implementation and Orchestration to Identify CI Community Challenges and Opportunities

- Wrote a 48-page *Container Runtimes Technical Report* that details the challenges of implementing Docker, Singularity, and X-Containers and Kubernetes orchestration for HPC. Benchmark tests and WRF CONSUS experiences are included. Lessons learned in this study will be shared with the broader community via this report and an SC21 Conference tutorial. https://federatedcloud.org/papers/ContainerRuntimesTechReport.pdf

Compared Containerized Application Kernel Performance on Google Cloud, Campus Clouds and Supercomputers

- Wrote a *Containerized Application Kernels on Google Cloud* report that compares Google Cloud benchmark and application performance to Aristotle Clouds, Comet, Bridges, Stampede2, Bridges-2, and Expanse.

https://federatedcloud.org/papers/ContainerizedAppsOnGoogleCloud.pdf

Developed a Marketplace Pricing Tool based on DrAFTS 2.0 Market Prediction Technology

- Created the “Aristotle AWS Pricing Tool” that compares Aristotle resources to various Amazon Web Services alternatives based on performance, cost, and price-performance.

https://federatedcloud.org/using/drafts.php

Updated the Aristotle Portal

- Updated the Aristotle portal with how-to documentation and videos, publications, news and events.
- Added 10 new technologies and techniques developed by the Aristotle team. https://federatedcloud.org/using/technologies.php

Expanded and Improved Aristotle Infrastructure

- Added cores, GPUs, and storage to federation sites, improved OpenStack

operations, and shared cloud implementation insights with the broader community.
- Maintained a quality production environment at each site for the Aristotle use cases and new science projects.

Provided Uninterrupted Access to Federation Resources

- Provided users with access to federation resources during the no cost extension period so that scientists could continue their research uninterrupted.

Enabled Science Use Cases

1. **A Cloud-Based Framework for Visualization and Analysis of Big Geo Data** (Varun Chandola, UB). Aristotle hosted the OUTSTEPS integrated community platform and the webGlobe analytical ecosystem. OUTSTEPS is a multi-institution regional research network focused on sustainability of the Lower Great Lakes. A paper describing the digital platform and its capabilities is underway and a research proposal was submitted to the NSF.
2. **Global Market Efficiency Impact** (Dominik Roesch, UB). Roesch and his collaborators used Aristotle and the financial framework to investigate whether human traders still matter at a time in which trading is dominated by computers. The results were presented at the *Conference for Finance Market Regulation* and the *European Finance Annual Meeting*. Progress continued on a LASSO regression predicting short-term stock returns using the whole cross-section of international stocks and an investigation of Cum-Ex trading (labeled as the “Biggest Tax Heist Ever” by the *NYTimes*). Each of these projects require large datasets and computational power. In April, “Investor short-termism and real investment” was published in the *Journal of Finance*.
3. **Application of the Weather Research and Forecasting (WRF) Model for Climate-Relevant Simulations in the Cloud** (Sara C. Pryor, Cornell). The Pryor group investigated derechos, fast-moving, damaging systems associated with tornadoes, wind gusts, heavy precipitation, and hail; they are difficult to simulate. The researchers ran an 11-member ensemble of a derecho that impacted Washington, DC on Aristotle with varying lateral boundary conditions and microphysics schemes to identify the simulation configuration that yielded the greatest fidelity. Other work with the Aristotle science team focused on implementing a multi-node WRF simulation capability and analyzing existing simulations. ML approaches were used for wind gust detection and quantification, and wind farm wake simulations from the east-coastal offshore lease areas were analyzed using ultra-high resolutions with WRF. The goal is to define the optimal density of wind farms to optimize system-wide power production and minimize energy costs. Pryor’s research team benefitted from Aristotle’s cloud architecture because of the availability of large RAM, multi-processors for analysis speed using parallel processing, and large disk volumes that could host all WRF output for analysis. Four journal papers were published during this reporting period.
4. **Transient Detection in Radio Astronomy Search Data** (James Cordes, Cornell). The Cordes team used Aristotle to prototype Fast Radio Burst (FRB) search codes and developed Docker and Singularity containers for seamless application deployment on Bridges and public clouds. Of particular interest, is a data set for the direction to the Galactic Center, which harbors a supermassive black hole and likely has many neutron stars orbiting it. Clustering methods in the FRB pipeline were also improved.
5. **Water Resource Management Using Python Rhodium Framework** (Patrick Reed, Cornell). David Gold continued the work of Bernardo Trindade on OpenMORDM and WaterPaths. The Reed team wants to build on their initial scaling analysis for WaterPaths and use Aristotle to support water supply

infrastructure investment and portfolio management analyses. A full evaluation of the core steps of their workflow is needed, including generating alternatives with multi-objective optimization and using global sensitivity analysis to discover what factors and scenarios cause system failures. This workflow could then be tested on a range of use cases such as the Sedento Valley benchmarking system, the North Carolina Research Triangle, and the Tampa Bay Water Authority.

6. **Mapping Transcriptome Data to Metabolic Models of Gut Microbiota** (Angela Douglas, Cornell). The Douglas research team and Aristotle support staff reconstructed and analyzed 31 metabolic models for every combination of the 5 principal bacterial taxa in the gut microbiome of *Drosophila*. Their findings were published in the *mSystems Journal*. In addition, they rebuilt their VM to test the containerization facilities.
7. **Multi-Sourced Data Analytics to Improve Food Production and Security** (Kate McCurdy, Sedgwick Reserve; Elizabeth Grafton-Cardwell, UC Lindcove; Chandra Krintz, UCSB SmartFarm). FL and CA citrus are under threat from the Asian Citrus Psyllid and citrus greening disease. IoT and data visualization technologies developed by the Aristotle team and the Aristotle Cloud will be used by the Citrus Under Protective Screening (CUPS) project to instrument and assess the effectiveness of growing citrus under screens. A real-time frost prevention and prediction prototype was successful and will be deployed throughout the Lindcove growing facility. Aristotle will continue to model conditions at the Sedgwick Reserve and UCSB's Edible Campus student farm. New collaborations that will leverage the Aristotle cloud and the project's software artifacts include gathering and analyzing remote sensing data from the seabed and atmosphere off the coast of CA to study kelp forest dynamics (Woods Hole Oceanographic Institute/UCSB Marine Science), and studying atmospheric cloud formation and lifecycle dynamics using ground-based cloud observation networks that are being built at Sedgwick and Long Island, NY (DOE). New technologies were documented for the broader community (e.g., "Edge-adaptable serverless acceleration for machine learning Internet of Things applications," *Journal of Software: Practice and Experience*) and a new NSF grant, "Detroit – A New End-to-End System for Practical and Accessible IoT," will build off Aristotle successes to date.

See quarterly NSF Reports for more details.
https://federatedcloud.org/reports/

Specific Objectives:

Significant Results:

Key outcomes or
Other achievements:

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

Cross-Training & Knowledge Sharing

Expertise was shared between sites every two weeks on Aristotle team conference calls to ensure timely cross-training and knowledge sharing. In-depth follow-up calls occurred to solve specific technology implementation issues. Multiple sites sharing technology insights and rollout experiences saved systems staff time and costs, a clear benefit from working within a federation. Cross-site discussions and collaborations between use case scientists and Aristotle support teams were also beneficial as they facilitated the sharing of research tools.

The "New York Region" (1,024 computer cores and 896TB storage) of the Jetstream 2 cloud will be installed at Cornell by November. Eventually, it will be integrated with the Aristotle Red Cloud system. Cornell will use this system to further its exploration of federated clouds and to make OpenStack enhancements. Lessons learned will be disseminated to the Jetstream 2 partners and the broader CI community.

Knowledge gained from events such as the *Cloud Forum 2020* were beneficial to Aristotle staff and the wider community. The *Cloud Forum 2021* will be held at Cornell on November 15-18 and will be virtual. It will include research presentations showing the innovative use of cloud tools by faculty researchers.

Undergraduate & Graduate Student Development

Undergraduate students benefitted from active participation in the Aristotle project.

For example, as an UCSB undergrad, Kerem Celik developed a telemetry data visualizer which will be hosted on Aristotle as a service and used by Citrus Under Protective Screening (CUPS) researchers in Spring 2022. As part of the container runtime investigation, Cornell undergrad Jeffrey Lantz ran benchmarks on Google Cloud, Aristotle clouds, and an XSEDE system, and worked with Kubernetes.

In total, 19 NSF Research Experiences for Undergraduates (REU) students benefitted from participation in the Aristotle project since its inception. An *HPCwire* story highlighted the research experiences of the latest REU cohort.

https://www.hpcwire.com/off-the-wire/cornell-students-immersed-in-latest-cloud-technologies-thanks-to-nsf-research-experiences-for-undergraduates-program/

Faculty course development and delivery benefited from Aristotle infrastructure and expertise as well.

For example, during Spring 2021, UB professor Mohammed Zia used the Aristotle cloud service to host JupyterHub with the nbgrader module enabled on an Ubuntu image for his *Programming and Database Fundamentals for Data Scientists (EAS 503)* class. The nbgrader module allows him to create and grade programming assignments in JupyterHub, a streamlined method that enables more frequent distribution of assignments. Aristotle is cost-effective because Zia can choose the instance type based on the expected class size. There are usually 50 students in the spring semester and 120 students in the fall semester. Zia selects the instance type he needs based on the class size, and after the semester is over, deletes the instance. Ease of deployment and the ability to handle all of the student's interactive Python needs in one place are very beneficial.

Currently, the UCSB Aristotle team is engaged in a new collaboration with Math and Statistics faculty at California State University Long Beach who wish to automate and scale their calculus curriculum. Aristotle staff expertise in using, maintaining, and scaling campus clouds for academic use is the basis for this education-focused collaboration.

Aristotle science use case projects provided students with unique opportunities to develop their skills and apply those skills to their own research. For example, use case scientist Dominik Roesch taught three University at Buffalo PhD students to use the OneTick Time-Series financial framework with data hosted on Aristotle; the students are now using those skills to conduct their PhD research.

Students at the Aristotle sites learned about and, in some cases, created new cloud and edge computing technologies. Opportunities arose for students when use case projects required special capabilities such as automated deployment or IoT tools. For example, at UCSB, PhD student Wei-Tsung Lin built a serverless computing platform for multi-scale and heterogeneous environments using geo-replicated logs. Lin was awarded his PhD in June.

How-To Documentation & Training

Aristotle how-to user guides were developed on GitHub and are publicly available on the Aristotle portal.
https://federatedcloud.org/using/gettingstarted.php

Container investigation products (methods, tools, images, etc.) are included in the list of 28 technologies and techniques developed by the Aristotle team.
https://federatedcloud.org/using/technologies.php

Cornell launched a *YouTube* site in August 2021 to augment their wiki-based training with how-to videos. The first videos in the ongoing series include how to connect to Linux instances, launching an instance, how to create a security group, how to access the OpenStack web interface, and how to create key pairs in Red Cloud.
https://www.youtube.com/channel/UCVPGMVWhp3sqWZFU5NntjTA

In addition, Aristotle documentation on how to build your own federation is available and includes our database schema, Globus user credential lookup sample code, Single Sign-On with Globus Auth, and hardware and software recommendations.
https://federatedcloud.org/using/buildyourown.php

*** Have the results been disseminated to communities of interest? If so, please provide details.**

Scientific Meetings, Publications, and Conferences

Researchers and the Aristotle technical team presented results at scientific meetings where they referenced the Aristotle project and its contribution to their success. They also published papers acknowledging the project.

Twenty-seven new publications, conference presentations, and videos are featured in this reporting period and are readily available to the CI and research communities at the Aristotle portal. https://federatedcloud.org/about/publications.php

Examples include the *IEEE Conference on High Performance Extreme Computing (HPEC)* publication titled “Self-scaling clusters and reproducible containers to enable scientific computing” that shares our experience using containers and describes their utility with cluster benchmark runs in a self-scaling virtual cluster using the Slurm scheduler deployed in Jetstream and Aristotle Red Cloud. https://ieeexplore.ieee.org/document/9286208

A *Journal of Applied Meteorology and Climatology* publication titled “WRF modeling of deep convection and hail for wind power applications” is one of the many science use case impacts enabled by the federation.
https://journals.ametsoc.org/view/journals/apme/59/10/jamcD200033.xml

Aristotle PIs continued to be engaged in professional societies and events during this reporting period such as the Practice & Experience in Advanced Research Computing (PEARC) conference, Coalition for Academic Scientific Computing (CASC) meetings, and SC conference, each of which afforded opportunities to share project progress and findings.

At *PEARC’21,* the Aristotle team participated in a virtual tutorial with Indiana University colleagues titled “A Deep Dive into Constructing Containers for Scientific Computing and Gateways.” https://github.com/XSEDE/Container_Tutorial/blob/main/Gateways2020/Day2_3_AdvancedTopics.pdf
This tutorial was also delivered at the *Science Gateways Community Institute Coding Institute 2021, eScience 2021* and *Gateways 2000.*

At *SC20,* Aristotle project information and publications were featured such as “Reproducible and portable workflows for scientific computing and HPC in the cloud” by Peter Vaillancourt et al.
https://dl.acm.org/doi/abs/10.1145/3311790.3396659and “Red Cloud and Aristotle: campus clouds and federations” by Richard Knepper et al.
https://dl.acm.org/doi/10.1145/3355738.3355755
A poster highlighted the project’s goals, federated cloud metrics, and advances in science https://www.cac.cornell.edu/About/Pubs/WolskiPoster2020.pdf
At *SC21*, we will be presenting a tutorial on constructing containers and will share lessons learned and some performance data from our container runtimes investigation and report.
https://sc21.supercomputing.org/presentation/?id=tut135&sess=sess184

Through the year, PI David Lifka kept XSEDE management abreast of developments in cloud computing via his leadership role in XSEDE. In addition, Aristotle portal lead Susan Mehringer served as the XSEDE training lead and Aristotle container investigation lead Richard Knepper served as the XSEDE manager for Community Resource Integration (CRI). Each of these roles provided opportunities to share cloud computing insights and experiences with CI and research colleagues.

Aristotle Portal

The Aristotle portal provides scientists and the CI community with extensive information on the project, including detailed use case accomplishments and products, publications, user guides, and new technologies and techniques developed by the Aristotle team. We will continue to operate the portal and will respond to inquiries from organizations who wish to use or deploy Aristotle technologies.

https://federatedcloud.org/

Communicating to a General Audience

Aristotle communications broadened the public's understanding of the value of research computing in solving societal problems and contributed to science education. Here are a few examples of how the Aristotle scientists and educators communicated to the public:

- July 2021 – A *NYTimes* story explained what scientists know about the beta variant. It includes insights from Laith Abu-Raddad, an infectious disease researcher at Weill Cornell Medicine, whose research was made possible by Red Cloud. https://www.nytimes.com/2021/07/19/health/beta-variant-covid-19.html?searchResultPosition=1
- May 2021 – *The AI Guide: Making AI Human* interviewed Aristotle's multi-sourced data analytics to improve food production use case scientist Chandra Krintz to share with a general audience how artificial intelligence is transforming agriculture. https://www.youtube.com/watch?v=nm7q4_dLRIs
- April 2021 – *TED-Ed* launched the "How do wind turbines work" video produced by Aristotle use case scientist Sara C. Pryor and Rebecca Barthelmie. Using animation, the video explains to the public how wind turbines convert wind in electricity and the challenges of powering the world with wind energy. The video has 300,000+ views to date. https://www.youtube.com/watch?v=xy9nj94xvKA
- November 2020 – CASC published their *2021 Science at Full Speed* brochure to educate Congress, Research VPs, and the public on the value of scientific computation. The "Visionary Ideas for Virtual Education" section featured Aristotle REU projects. Nineteen REU students have participated in the Aristotle project since its inception. https://casc.org/wp-content/uploads/2021/01/CASC-2021-Brochure-FINAL-JW.pdf
- September 2020 – An *HPCwire* story featured students who were immersed in the latest cloud technologies thanks to the NSF REU program. Students described their experiences and the value of the program in their own words. https://www.hpcwire.com/off-the-wire/cornell-students-immersed-in-latest-cloud-technologies-thanks-to-nsf-research-experiences-for-undergraduates-program/

Products

Books

Book Chapters

Inventions

Journals or Juried Conference Papers

Dominik M. Rosch, Avanidhar Subrahmanyam & Mathijs A. van Dijk (2021). Investor short-termism and real investment. *Journal*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1016/j.finmar.2021.100645

Frederick Letson, Tristan J. Shepherd, Rebecca J. Barthelmie & Sara C. Pryor (2020). WRF modeling of deep convection and hail for wind power applications. *Journal*. 59 (10), 1717. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1175/jamc-d-20-0033.1

Gareth George, Fatih Bakir, Rich Wolski & Chandra Krintz (2020). NanoLambda: Implementing Functions as a Service at all resource scales for the Internet of Things. *Juried Conference Paper*. 220. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1109/SEC50012.2020.00035

Jeanie A. Aird, Rebecca J. Barthelmie, Tristan J. Shepherd & Sara C. Pryor (2020). WRF-simulated springtime low-level jets over Iowa: implications for wind energy. *Journal*. 1618 10.1088/1742-6596/16. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI:

Jeanie A. Aird, Rebecca J. Barthelmie, Tristan J. Shepherd & Sara C. Pryor (2021). WRF-simulated low-level jets over Iowa: characterization and sensitivity studies. *Journal*. 1015. Status = PUBLISHED; Acknowledgment of Federal

Support = Yes ; Peer Reviewed = Yes ; DOI: 10.5194/wes-6-1015-2021

Michael Zhang, Chandra Krintz & Rich Wolski (2020). Edge-adaptable serverless acceleration for machine learning Internet of Things applications. *Journal*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1002/spe.2944

Michael Zhang, Chandra Krintz & Rich Wolski (2021). Sparta: A Heat-Budget-based Scheduling Framework on IoT Edge Systems. *Juried Conference Paper*. . Status = ACCEPTED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Nana Y.D. Ankrah, Brandon E. Barker, Joan Song, Cindy Wu, John G. McMullen II & Angela E. Douglas (2021). Predicted metabolic function of the gut microbiota of Drosophila melanogaster. *Journal*. 6 (3), . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1128/mSystems.01369-20

Peter Z. Vaillancourt, J. Eric Coulter, Richard Knepper & Brandon Barker (2020). Self-scaling clusters and reproducible containers to enable scientific computing. *Juried Conference Paper*. 1. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1109/HPEC43674.2020.9286208

Rebecca J. Barthelmie, Tristan J. Shepherd & Sara C. Pryor (2020). Increasing turbine dimensions: impact on shear and power. *Journal*. 1618 . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1088/1742-6596/1618/6/062024

Sara C. Pryor, Tristan J. Shepherd, Patrick Volker, Andrea Hahmann & Rebecca J. Barthelmie (2020). Diagnosing systemic difference in predicted wind turbine array-array interactions. *Journal*. 1618 10.1088/1742-6596/16. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI:

Licenses

Other Conference Presentations / Papers

Eric Coulter, Peter Vaillancourt, Jeremy Fischer, Stephen Bird, Sanjana Sudarshan & Suresh Marru (2021). *A deep dive into constructing containers for scientific computing and gateways*. Practice & Experience in Advanced Research Computing (PEARC'21).. Virtual. Status = OTHER; Acknowledgement of Federal Support = Yes

Eric Coulter, Peter Vaillancourt, Jeremy Fischer, Stephen Bird, Sanjana Sudarshan & Suresh Marru (2020). *A deep dive into constructing containers for scientific computing and gateways*. Gateways 2020. Virtual. Status = OTHER; Acknowledgement of Federal Support = Yes

Sayed Mohammed Arshad Zaidi, Varun Chandola, Melissa R. Allen & Bhudhendra L. Bhaduri (2018). *Anomaly detection in Energy-Water Nexus: webGlobe – A cloud-based geospatial analysis framework for interacting with climate data*. American Geophysical Union Fall 2018 Meeting. Washington, DC. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Rebecca J. Barthelmie, Sara C. Pryor & Charlotte B. Hasager (2021). *Array-array interactions and downstream wake effects*. Wind Energy Science Conference 2021. Hannover, Germany. Status = OTHER; Acknowledgement of Federal Support = Yes

Tristan J. Shepherd, Rebecca J. Barthelmie & Sara C. Pryor (2019). *Assessment of wind turbine impact on future climate in GCM-driven WRF simulation*. North American Wind Energy Academy WindTech Conference. Amherst, MA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Rich Wolski, David Lifka & Tom Furlani (2020). *Data analysis and management for multi-campus cyberinfrastructure through cloud federation*. 2020 NSF Cyberinfrastructure for Sustained Scientific Innovation (CSSI) PI Meeting. Seattle, WA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Eric Coulter, Jeremy Fisher, Stephen Bird, Sanjana Sudarshan, Peter Vaillancourt & Suresh Marru (2021). *Deep dive into constructing containers for scientific computing and gateways*. Practice & Experience in Advanced Research Computing (PEARC'20). Virtual. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Rich Wolski (2019). *Devices-as-Services and the Internet as a Platform of Things (Keynote)*. Arm Research Summit 2019. Austin, TX. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Matt Ringgenberg, Jonathan Brogaard & Dominik Rosch (2021). *Does the trading floor matter?*. The Microstructure Exchange. Virtual. Status = OTHER; Acknowledgement of Federal Support = Yes

Rich Wolski (2021). *Experiences with IoT and the new internet as a platform of things*. 53rd IEEE/ACM International Symposium on Microarchitecture (MICRO-53). Virtual. Status = OTHER; Acknowledgement of Federal Support = Yes

Susan Mehringer (2018). *Federations: Lessons learned*. Coalition for Academic Scientific Computation Fall Meeting. Alexandria, VA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Dominik Rosch (2020). *Financial market frictions*. Cornell Johnson School. Ithaca, NY. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Richard Knepper, Resa Reynolds, Eric Coulter & Steve Bird (2019). *Gathering requirements for a campus cloud toolkit*. Practice and Experience in Advanced Research Computing (PEARC19). Chicago, IL. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Shami Chatterjee (2018). *New results in radio astronomy: Fast radio bursts and transients*. International Astronomical Union General Assembly. Vienna. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Tristan J. Shepherd, Rebecca J. Barthelmie & Sara C. Pryor (2019). *Quantifying array-array wind turbine impact on future climate in GCM driven WRF simulations*. North American Wind Energy Academy WindTech Conference. Amherst, MA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Tristan Shepherd, Bennett Wineholt, Rebecca J. Barthelmie & Sara C. Pryor (2019). *Quantifying weather and climate simulation reproducibility in the cloud*. 99th American Meteorological Society Annual Meeting: 5th Symposium on High Performance Computing for Weather, Water, and Climate. Phoenix, AZ. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Resa Reynolds (2019). *Red Cloud and the Aristotle Cloud Federation*. Dell XL HPC Consortium Fall Meeting. Scottsdale, AZ. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Rich Wolski, Chandra Krintz, Fatih Bakir, Wei-tsung Lin & Gareth George (2019). *Rethinking scalable services for the Internet of Things*. Big Data and Extreme-Scale Computing Workshop (BDEC2). San Diego, CA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Chandra Krintz (2020). *SmartFarm: computing research for the next-generation of precision agriculture*. American Association for the Advancement of Science Annual Meeting. Seattle, WA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Richard Knepper (2019). *Sustainability Research Impacts*. Cornell University Sustainability Leadership Summit. Ithaca, NY. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Shami Chatterjee (2018). *The dwarf galaxy host of a repeating fast radio burst*. International Astronomical Union General Assembly. Vienna. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Rebecca J. Barthelmie, Sara C. Pryor & Tristan J. Shepherd (2018). *Wakes from wind turbine array*. National Renewable Energy Laboratory. Golden, CO. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Sara C. Pryor, Tristan J. Shepherd, Melissa Bukovsky & Rebecca J. Barthelmie (2019). *Wind energy scenarios for climate change mitigation*. The Scenarios Forum 2019. Denver, CO. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Other Products

Audio or Video Products.

Ray Wang (2021). *How to connect to Linux instances*. YouTube.

https://www.youtube.com/watch?v=xqJuXhLRFek

Ray Wang (2021). *Launching an instance.* YouTube.
https://www.youtube.com/watch?v=jOx8jMqS_hQ&t=9s

Ray Wang (2021). *How to create a security group to launch a Red Cloud instance.* YouTube.
https://www.youtube.com/watch?v=8G6NiLoeqZw&t=4s

Ray Wang (2021). *How to access the OpenStack web interface for Red Cloud instances.* YouTube.
https://www.youtube.com/watch?v=sOl4intZ3tM

Ray Wang (2021). *How to create key pairs in Red Cloud.* YouTube.
https://www.youtube.com/watch?v=eQExMk_k6d8

Chandra Krintz. (2021). *Using artificial intelligence to transform agriculture*. The AI Guide: Making AI Human.
https://www.youtube.com/watch?v=nm7q4_dLRIs

Rebecca J. Barthelmie & Sara C. Pryor (2021). *How do wind turbines work?* TED-Ed animated video.
https://www.youtube.com/watch?v=xy9nj94xvKA&t=7s

Rich Wolski (2021). *Experiences with IoT and the new Internet as a Platform of Things*. Invited speaker, University of Chicago CERE Center for Unstoppable Computing. https://www.youtube.com/watch?v=p4CVwaGNUW0

Jeffrey Lantz (2021). *WRF in the Public Cloud*. Aristotle project student presentation, Cornell University.
https://www.youtube.com/watch?v=32ISd8LUQRw

Other Publications

Peter Z. Vaillancourt, Bennett Wineholt, Tristan J. Shepherd, Sara C. Pryor, Jeffrey Lantz, Richard Knepper, Rich Wolski, Christopher R. Myers, Ben Trumbore, Resa Reynolds, Jodie Sprouse, and David Lifka (2021). *Aristotle Cloud Federation: Container Runtimes Technical Report*. https://federatedcloud.org/papers/ContainerRuntimesTechReport.pdf. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Nikolay A. Simakov, Andrew E. Bruno, Salvatore J. Guercio, and Matthew D. Jones (2021). *Containerized Application Kernels on Google Cloud*. https://federatedcloud.org/papers/ContainerizedAppsOnGoogleCloud.pdf. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Patent Applications

Technologies or Techniques

The Aristotle team and its collaborators developed the following technologies and techniques:

Aristotle AWS Pricing Tool – developed a price performance analysis tool based on the original DrAFTS market prediction technology that helps users compare Aristotle resources to various AWS alternatives based on performance, cost, and price-performance. To do so, the tool runs the TOP500 LINPACK Benchmarks on all Aristotle instance types and all AWS instance types, and generates a report ranking them. Users can ask questions such as: Which AWS instance type is most equivalent to an instance type in Aristotle? Which is less costly? Or, if I want to spend 20% more and go 30% faster, which instance type should I use?
https://federatedcloud.org/using/drafts.php
http://169.231.235.92:5000/
Automated Deployment Methods – implemented a Slurm HPC cluster in a cloud with OpenHPC 2 series based on CentOS/Rocky Linux 8.
https://github.com/federatedcloud/wrf-cluster-openstack
Containerized Application Kernels – developed containerized Application Kernels (AK) and then compared AK performance on Google Cloud to performance on Aristotle Clouds, Stampede2, Bridges, Bridges-2, Comet, and Expanse.
https://federatedcloud.org/papers/ContainerizedAppsOnGoogleCloud.pdf
Kubernetes Implementation Code for MPI Clusters – this code reflects the current state of experimental support for MPI applications managed by Terraform Kubernetes constructs that allow for automatic node count scaling and cloud

portability. The software-based resource provisioning can best be attempted on AWS at the present time; it also works with Google Cloud Platform with MPI applications. Conversion to other cloud platforms should be possible with extensive changes to a platform-specific Terraform Kubernetes provider or other resource configurations. Familiarity with Kubernetes software concepts, resources provisioning on the desired cloud, and debugging parallel computing applications are recommended. The repository includes a “Getting Started with Kubernetes” tutorial as well.
https://github.com/federatedcloud/kubernetes-mpi-clusters
Metabolic Model and Container – developed code to predict metabolic function of the gut microbiota of Drosophila melanogaster using v.3.0.4 of the OpenCOBRA Toolbox and v7.51.1 of the Gurobi Optimizer. An optional, containerized environment for running the code is available as well as a tutorial for performing the simulations.
https://github.com/federatedcloud/DouglasMetabolicModels/releases/tag/v1.0.2 https://github.com/federatedcloud/COBRAContainers
NanoLambda – developed a portable platform that brings Functions as a Service (FaaS), high-level language programming, and familiar cloud service APIs to non-Linux and microcontroller-based IoT devices. NanoLambda couples a new, minimal Python runtime system that is designed for the least capable end of the IoT device spectrum, with API compatibility for AWS Lambda and S3. It transfers functions between IoT devices (sensors, edge, cloud), providing power and latency savings while retaining the programmer productivity benefits of a high-level language and FaaS. A key feature of NanoLambda is a scheduler that intelligently places function executions across multi-scale IoT deployments according to resource availability and power constraints https://sites.cs.ucsb.edu/~ckrintz/papers/nanolambda-sec20.pdf
Radio Astronomy Container – developed a single container of radio astronomy software that combines the pipeline components developed for pulsar and other transient detections that can be deployed either on the cloud with Docker or on an XSEDE HPC resource with Singularity.
https://github.com/federatedcloud/pulsar-pipeline-container
https://hub.docker.com/r/cornellcac/pulsar-pipeline
https://datasets.datalad.org/?dir=/shub/federatedcloud/pulsar-pipeline-container
Sparta – in order to protect edge clouds from overheating, we developed a heat-budget-based scheduling system called Sparta which leverage dynamic voltage and frequency scaling to adaptively control CPU temperature. Sparta takes machine learning applications, datasets, and a temperature threshold as input. It sets the initial frequency of the CPU based on historical data and then dynamically updates it, according to the applications’ execution profile and ambient temperature, to safeguard edge devices.
https://sites.cs.ucsb.edu/~ckrintz/papers/edge21.pdf
STOIC (Serverless Teleoperable Hybrid Cloud) – developed an IoT application and offloading system that extends the serverless model in 3 ways: (1) STOIC adopts a dynamic feedback control mechanism to precisely predict latency and dispatch workloads uniformly across edge and cloud systems using a distributed serverless framework, (2) STOIC leverages hardware acceleration (e.g., GPU resources) for serverless function execution when available from the underlying cloud system, (3) STOIC can be configured in multiple ways to overcome deployment variability associated with public cloud use.
https://sites.cs.ucsb.edu/~ckrintz/papers/stoic-smartedge20.pdf
WRF CONUS Benchmark Containers – implemented WRF 4.2.2 to run CONUS benchmarks on bare metal HPC in a Docker and a Singularity container.
https://github.com/federatedcloud/WRFv4-Benchmarking
WRF Docker Container – implemented a Docker container for WRF 3.8.1 with a Fitch patch. https://github.com/federatedcloud/Docker-WRF-3.8.1-Fitch

Thesis/Dissertations

Websites or Other Internet Sites
Aristotle Cloud Federation
https://federatedcloud.org

The Aristotle Cloud Federation portal is updated regularly to feature new web content and user guide documentation as well as Aristotle science use case accomplishments, plans, and products. Quarterly and annual National Science Foundation project reports are also available on the portal; they provide detailed descriptions of the project’s activities, challenges, and accomplishments.

Participants/Organizations

Research Experience for Undergraduates (REU) funding

Form of REU funding support:	REU supplement
How many REU applications were received during this reporting period?	Nothing to Report
How many REU applicants were selected and agreed to participate during this reporting period?	Nothing to Report
REU Comments:	No REUs this reporting period

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Lifka, David	PD/PI	1
Furlani, Thomas	Co PD/PI	1
Wolski, Richard	Co PD/PI	3

Full details of individuals who have worked on the project:

David A Lifka
Email: lifka@cac.cornell.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Programmatic oversight of the Aristotle Cloud Federation project ensuring deliverables outlined in the program execution plan are met on schedule.

Funding Support: No funding support from other projects used for this award.

Change in active other support: No

International Collaboration: No
International Travel: No

Thomas R Furlani
Email: furlani@buffalo.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Programmatic oversight of the UB subaward, and continuous interaction with the entire Aristotle technical team.

Funding Support: No funding support from other projects used for this award.

Change in active other support: No

International Collaboration: No
International Travel: No

Richard Wolski
Email: rich@cs.ucsb.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 3

Contribution to the Project: Architected deployments at UCSB for production and Aristotle services. On-boarded new science use cases.

Funding Support: No funding support from other projects used for this award.

Change in active other support: No

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?
Nothing to report.

Were other collaborators or contacts involved? If so, please provide details.
Nothing to report

Impacts

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

The Aristotle project advanced the knowledge of federated cloud computing and its potential role as a campus bridging paradigm. By building and deploying a federated cloud model with the necessary allocations, accounting, and cloud metrics, Cornell University, University at Buffalo, and UC Santa Barbara explored how cloud resources can be effectively shared between campuses and their impact on researchers who require flexible workflows and analysis tools for large-scale data sets. The project served as an important model for campus cyberinfrastructure that others may follow and provided elasticity by sharing resources, data, software, tools, and expertise that may not be available locally.

What is the impact on other disciplines?

Aristotle use case scientists strategically explored problems of increasing complexity and corresponding increases in data and, as a result, advanced scientific knowledge. Data challenges from a diversity of disciplines (earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, microbiome, agriculture) were addressed with collaborators from other academic institutions, public agencies, and research labs, as well as citizen scientists. The sharing of data infrastructure building blocks capacity and the movement of instances across institutional boundaries demonstrated the potential of creating wider science collaborations and increased data sharing. The creation and performance testing of containerized applications demonstrated how portability to from local cloud resources to other campus clouds, NSF clouds, public clouds, HPC clusters, and supercomputers may produce a positive impact on researcher productivity, flexibility, and time to science.

Cornell’s Aristotle Cloud (Red Cloud) will support future research in areas such as COVID-19 epidemiology, HIV modeling, cryptography, robot perception, and the understanding of the impurity of atoms in semiconductors and the ease with which ions can move in and out of batteries and other energy devices.

UB’s NSF award, “Collaborative Research: CCRI: New: Medium: A Development and Experimental Environment for Privacy-preserving and Secure (DEEPSECURE) Machine Learning,” will leverage the federation capabilities of UB’s cloud to branch out to Old Dominion's OpenStack cloud, developing a Software Development Environment and Multi-User Experimental Chamber for securing ML/AI Research.

UCSB’s use cases cover a broad range of research and instruction such as the application of IoT, edge, and cloud for

digital agriculture, the use of remote data to study kelp forest and atmospheric lifecycle dynamics; research support for Media Neuroscience and R-Shief; enablement of the GauchoSpace Learning Management System, and sandbox environments for capstone class projects in Computer Science, Data Science, and Statistics and Applied Probability. UCSB will apply the techniques and technologies developed for Aristotle use cases to support future data science and ML projects; build upon current resource capacity, computing cores, and storage; and, offer new data, message queuing, and Kubernetes services to the wider community.

What is the impact on the development of human resources?

Aristotle helped to pioneer the concept of federated cloud computing for research which may ultimately increase the availability of on demand resources, data, and analysis tools that engage underrepresented groups. In addition, it demonstrated how virtual laboratories in the cloud can enhance classroom learning. For example, Aristotle cloud resources were used to solve video analysis application problems in a *Cloud Computing, Edge Computing, and IoT* class (CS293B) taught at UCSB and were used to host JupyterHub with the nbgrader module in a *Programming and Database Fundamental for Data Scientists* class (EAS503) at the University at Buffalo. Aristotle is cost-effective because a professor can choose the instance type based on the expected class size and, after the semester is over, delete the instance.

The project also demonstrated how the availability of campus-to-campus and campus-to-public cloud computing capabilities can spurn the development and dissemination of ready-to-launch VMs and containers with training software and tools preloaded. This may reduce the redundant development and preparation of educational materials and reduce onsite computer lab administration, resulting in an increased focus on individual student learning needs.

Nineteen REU students participated in the Aristotle project since its inception. The latest round of REU student experiences were highlighted in an *HPCwire* story. Two former Aristotle REUs have gone on to complete their PhDs, and a third recently entered a PhD program.

What was the impact on teaching and educational experiences?

Undergraduate and graduate students derived value from the Aristotle project as did professors.

For example, as an UCSB undergrad, Kerem Celik developed a telemetry data visualizer which will be hosted on Aristotle as a service and used by Citrus Under Protective Screening (CUPS) researchers in Spring 2022. As part of the container runtime investigation, Cornell undergrad Jeffrey Lantz ran benchmarks on Google Cloud, Aristotle clouds, and an XSEDE system, and worked with Kubernetes. In total, 19 NSF Research Experiences for Undergraduates (REU) students benefitted from participation in the Aristotle project since its inception.

Faculty course development and delivery benefited from Aristotle infrastructure and expertise as well.

For example, during Spring 2021, UB professor Mohammed Zia used the Aristotle cloud service to host JupyterHub with the nbgrader module enabled on an Ubuntu image for his *Programming and Database Fundamentals for Data Scientists (EAS 503)* class. The nbgrader module allows him to create and grade programming assignments in JupyterHub, a streamlined method that enables more frequent distribution of assignments. Aristotle is cost-effective because Zia can choose the instance type based on the expected class size. There are usually 50 students in the spring semester and 120 students in the fall semester. Zia selects the instance type he needs based on the class size, and after the semester is over, deletes the instance. Ease of deployment and the ability to handle all of the student's interactive Python needs in one place are very beneficial.

Currently, the UCSB Aristotle team is engaged in a new collaboration with Math and Statistics faculty at California State University Long Beach who wish to automate and scale their calculus curriculum. Aristotle staff expertise in using, maintaining, and scaling campus clouds for academic use is the basis for this education-focused collaboration.

Aristotle science use case projects provided students with unique opportunities to develop their skills and apply those skills to their own research. For example, use case scientist Dominik Roesch taught three University at Buffalo PhD students to use the OneTick Time-Series financial framework with data hosted on Aristotle; the students are now using those skills to conduct their PhD research. Students at the Aristotle sites learned about and, in some cases, created new cloud and edge computing technologies. Opportunities arose for students when use case projects required special capabilities such as automated deployment or IoT tools. For example, at UCSB, PhD student Wei-Tsung Lin built a

serverless computing platform for multi-scale and heterogeneous environments using geo-replicated logs. Lin was awarded his PhD in June.

What is the impact on physical resources that form infrastructure?

The federated cloud model may impact the physical resources that form infrastructure by reducing the number of computer labs required for learning. Campus clouds may be installed so that researchers can cost effectively use local cloud resources and, when more capacity is needed, move their application container to the most suitable campus, public, or NSF cloud resource. Ultimately, federated clouds will likely become complementary resources to high-end supercomputers, e.g., performing on demand iterative tasks, streaming IoT data, etc. Integrating federated clouds (campus, labs, etc.) into the national cyberinfrastructure ecosystem may increase the sharing of resources and, subsequently, reduce physical resource expenditures at the local level.

The “New York Region” of the Jetstream 2 cloud will be integrated with Aristotle Red Cloud. Cornell will use this capability to further its exploration of federated clouds and to make OpenStack enhancements. Lessons learned will be disseminated to the Jetstream 2 partners and the broader community. Cornell was included in the Jetstream 2 project as a direct result of its Aristotle Cloud Federation work.

The University at Buffalo’s OpenStack cloud will continue to provide essential research support for UB faculty and their national collaborators. Support for containers and Kubernetes orchestration were added to prepare for new research workloads in the cloud. Multiple NSF grant proposal submissions were submitted that propose to utilize the work and infrastructure that has been put in place with the Aristotle project, thus requiring less hardware infrastructure funds and no start up time for building out the CI platform.

The Aristotle project germinated UCSB’s cloud infrastructure and spurred cloud computing forward in computer science, engineering, and the arts and sciences. UCSB will build upon current resource capacity, computing cores, and storage, and offer new data, message queuing, and Kubernetes services to the wider community.

What is the impact on institutional resources that form infrastructure?

Aristotle demonstrated the ability to maximize institutional resources through federation with other institutions by: (1) offloading variable computational and data analysis demands from local infrastructure, (2) starting coarsely parallel computations on demand, (3) providing heterogeneous instance types (CPUs, GPUs, etc.) and sizes to allow for unpredictable computational demands, (4) sharing knowledge and best practices between sysadmins and the broader community.

What is the impact on information resources that form infrastructure?

The Aristotle project demonstrated how a federated model can encourage the sharing of high-value processed datasets of general interest and separate data resources, and the creation of reproducible pipelines in the form of VMs or VM configurations, and containers that can run on multiple platforms (private cloud, campus clouds, HPC cluster systems, and national supercomputers and other CI resources).

We published the “Aristotle Cloud Federation: Container Runtimes Technical Report” to inform the CI community of the challenges and opportunities of using containers and orchestration for scientific HPC applications. Products are available from that investigation on GitHub and results will be shared next at the SC21 Conference. We also published the “Containerized Application Kernels on Google Cloud” report to share the performance of application kernels on Google Cloud versus Aristotle Clouds, Comet, Bridges, Stampede2, Bridges-2, and Expanse with the infrastructure and research community.

What is the impact on technology transfer?

Technologies generated by this project are open source, therefore technology transfer licensing, patent applications, etc. are not applicable. The project’s collaborative relationships for R&D included Amazon Web Services, Google Cloud, Microsoft Azure, Red Hat, Globus, Dell, HPE, Flexera, and VEXXHOST.

During this reporting period, we provided software testing services on Aristotle for X-Containers, a new container architecture that supports strong security isolation, high performance, and low cost running of application containers in the

cloud. It started as a research project at Cornell University and is now maintained by Exotanium, Inc., a start-up that completed the NSF I-Corps program and, today, offers solutions for cloud resource optimization. X-Containers can be deployed seamlessly in existing container orchestration platforms and automatically optimize the performance of an application without the need to modify source code or recompile.

What is the impact on society beyond science and technology?

Aristotle use case findings have the potential to impact wind turbine companies interested in harnessing the energy of the atmosphere and converting it into carbon-free electricity; policymakers regulating high-frequency trading; policymakers making water resources management decisions; health practitioners treating immune and metabolic diseases; oceanographers restoring nearshore coastal ecosystems and halting the decline of kelp forests; farmers increasing yields and protecting the environment by accessing on demand soil, water, and crop sensor data to aid decision-making; and, orange growers trying to protect citrus trees from the Huanglongbing bacteria (citrus greening disease) that has devasted FL and is now threatening CA.

What percentage of the award's budget was spent in a foreign country?
Nothing to report.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.

Change in primary performance site location
Nothing to report.