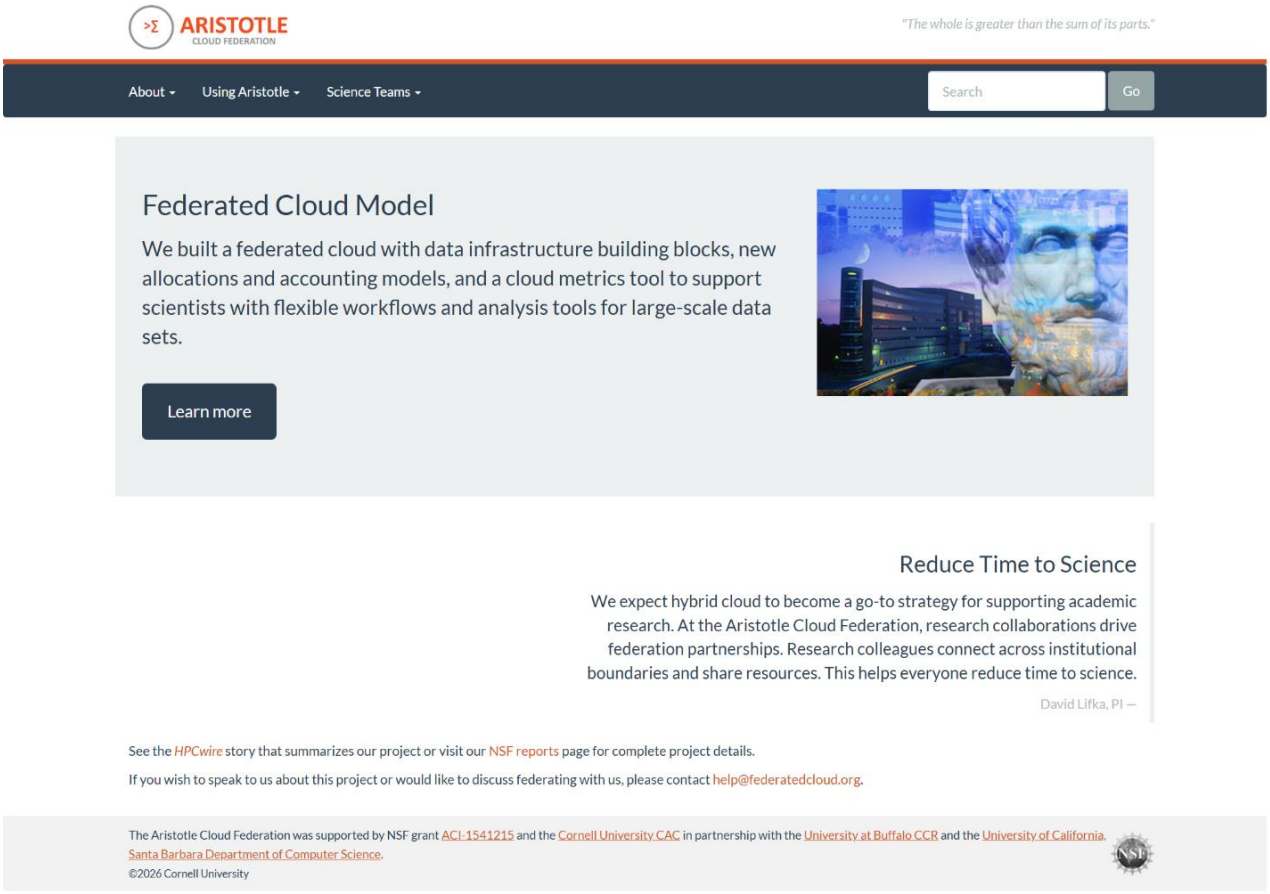
ARISTOTLE CLOUD FEDERATION

2015-2021 David Lifka, PI, Cornell. Partners: UC, Santa Barbara and U. at Buffalo

Visit the NSF Project Website for Information on the 2015-2021 Aristotle Cloud Federation DIBBs Project including the Project Abstract, 48 Publications Produced as the Result of the Research, and the Project Outcomes Report.

The project website (federatedcloud.org) was retired August 2026. Below are assorted screenshots from that site.



Project Summary

Project Overview

The ability to aggregate, share, and analyze important large data sets while optimizing time to science has become essential to support multi-disciplinary and multi-institutional data-driven discovery. Cornell University and its partners, the University at Buffalo and the University of California, Santa Barbara, created a federated cloud in New York State and California.

Each site operated standard cloud infrastructure components augmented with DIBBs storage assets including data analysis servers, scalable storage, and a Globus file transfer and sharing endpoint.

This project served as an important model for campus cyberinfrastructure that other institutions can follow providing flexibility in support of multi-discipline scientific workflows, elasticity by sharing resources, and an allocation model that provides a fair exchange mechanism for resource access between and across multiple institutions.

See the *HPCwire* story that summarizes our project or visit our NSF reports page for complete project details.

If you wish to speak to us about this project or would like to discuss federating with us, please contact help@federatedcloud.org.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Project Summary

Intellectual Merit

The goal of this partnership was to 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.

The project supported a rich set of open source software and cloud usage modalities (e.g., Apache Hadoop for large data sets, Virtual Machine (VM) snapshots of complex software systems, dynamically-sized SLURM clusters, etc.) and implemented and optimized frameworks for query based exploration, modeling, and analysis.

Research scientists and users strategically explored problems of increasing complexity and corresponding increases in data. Data challenges from a diversity of fields (earth & atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, etc.) were addressed with collaborators from over 40 academic institutions, public agencies, and research labs as well as citizen scientists. In addition to advancing scientific knowledge, this project advanced the knowledge of federated and hybrid cloud computing and their potential roles as campus bridging paradigms.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Project Summary

Broader Impact

Societal benefits of the scientific research enabled by this partnership included a better understanding of how US and international market inefficiencies can lead to financial instability; how wind speeds and atmospheric properties impact weather and wind turbine performance; how US water supply systems and river basins can be better managed and protected; how microbial therapies can support health and resolve clinical conditions; and how environmental stewardship can increase food production and security.

The federated cloud with DIBBs storage assets benefited the broader NSF community by serving as a model for campus-to-campus cyberinfrastructure that is scalable and sustainable. It demonstrated how sharing data infrastructure building block capacity across institutional boundaries can create wider science and engineering collaborations and increase data sharing. A community portal provided how-to guides to institutions that wish to replicate this model. Technical documentation and webinars featured best practices and included how to create VM instances, run at federated sites, predict price performance of Aristotle cloud and AWS instances, use containers effectively, and access, move, and store large-scale data.

A new tool for cloud metrics was developed and built into Open XDMoD and features QBETs statistics to make online forecasts of future performance and allocations levels.

A new model for allocations and accounting was also implemented.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Project Summary

Federated Cloud Model Goals

- 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 issues of building and running non-trivial containers.
- Containerize application kernels and test their performance on Google Cloud versus Aristotle Clouds, Comet, Bridges, Stampede2, Bridges-2, and Expanse. Publish the performance comparisons for the cyberinfrastructure and research community.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Aristotle Leadership Team

David Lifka, PhD
Principal Investigator and Project Director
Vice President for Information Technologies & CIO
Cornell University

Tom Furlani, PhD
Co-Principal Investigator
CIO
Roswell Park Comprehensive Cancer Center

Rich Wolski, PhD
Co-Principal Investigator
Professor, Department of Computer Science
University of California, Santa Barbara

Resa Reynolds
Infrastructure Team Lead
Assistant Director, Systems
Cornell University Center for Advanced Computing

Susan Mehringer
Portal Team Lead
Assistant Director, Consulting
Cornell University Center for Advanced Computing

Adam Brazier, PhD
Science Team Lead
Computational Scientist
Cornell University Center for Advanced Computing

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Executive Advisory Committee

Dmitrii Calzago
Chief Technical Officer
AppScale Systems

Ian Foster, PhD
Director of the Computation Institute, ANL/University of Chicago

Steve Johnson, PhD
Professor, Department of Population Health
NYU Langone Health

Sanjay Padhi, PhD
Research & Technical Computing, Amazon Web Services

Craig Stewart, PhD
Consultant

John Towns
PI and Project Director, XSEDE
Executive Director, Science & Technology, NCSA
Deputy CIO, Research IT, University of Illinois

Rick Wagner
Professional Services Manager
Globus

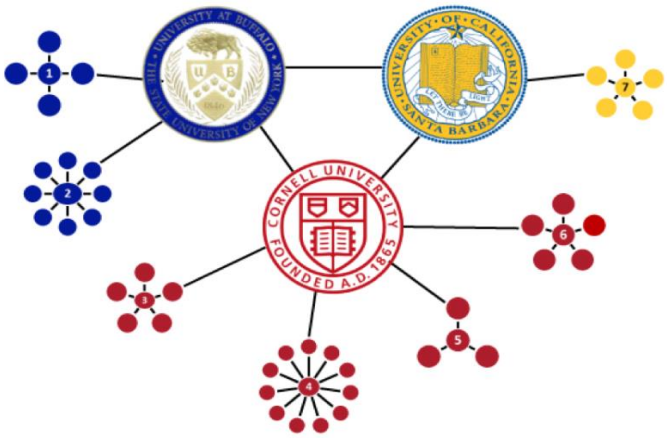
The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Federation Resources

Introduction

The Cornell University Center for Advanced Computing, with partners University at Buffalo and University of California, Santa Barbara, deployed data infrastructure building blocks (DIBBs) for multi-institutional cyberinfrastructure (CI). Our approach combined data analytics and flexible workflow management in the form of a federated cloud model capable of supporting large-scale, shared, and collaborative data analysis. This project was metrics driven, leveraging system analytics provided by Open XDMoD and DrAFTS, to ensure effective resource utilization and optimal time to science.

This project was multi-institutional in two important ways. First, it supported 7 science use cases with researchers and their collaborators at each partner site and extended to collaborators located at other institutions. This provided access to extended research groups with common data interests and requirements without each group having to replicate critical data assets.



Multi-disciplinary collaborations in a multi-campus cloud federation

Second, our cloud federation model enabled sharing the project's storage assets among the three partner institutions, providing their researchers and collaborators elasticity that might not have been financially or logistically possible at each individual site, particularly as the number of researchers requiring large scale data analysis grew.

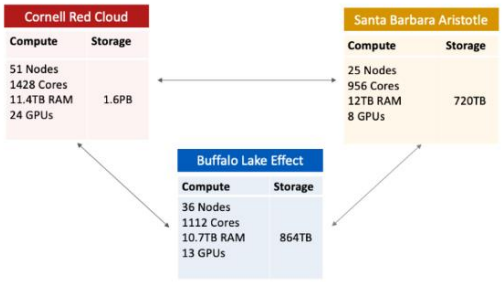
An important goal of this project was to demonstrate a new model of cloud federation that is a sustainable and effective way for institutions to augment campus CI for collaborative data analysis. This model included investigating a common allocation mechanism and explored accounting mechanisms that would provide a transparent resource exchange between partner institutions.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Federation Resources

Resources by Site

Aristotle Cloud Federation model resources were increased each year of the 5-year grant and grew by 2020 to the configuration detailed below. Since then, the clouds have continued to grow to meet research computing and data demands at Cornell, the University at Buffalo, and UC Santa Barbara.



2020 Site Details

	Cornell	Santa Barbara	Buffalo
Cloud URL	Red Cloud	Aristotle	Lake Effect
Status	Production	Production	Production
Vendor	Dell	Dell, HPE, DXC	Dell, ACE
Storage	1.6PB	720TB	864TB
DIBBs Purchased Cores	616*	740**	792***
RAM/Core	8GB	Dell: 9GB HPE: 10GB	Up to 8GB
Software Stack	OpenStack	OpenStack	OpenStack
10Gb Interconnect	Yes	Yes	Yes
Globus File Transfer	Yes	Yes	Yes
Globus OAuth 2.0	Yes	Yes	Yes

* 616 cores added to Cornell Red Cloud (1428 total)
** 740 cores in Santa Barbara Aristotle cloud (956 total, Aristotle is separate from UCSB campus cloud)
*** 1112 total cores

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Federation Resources

Cornell Red Cloud OpenStack Instances

Instance Name	RAM (MB)	VCPUs
c1.m8	8192	1
c2.m16	16384	2
c4.m32	32768	4
c8.m64	65536	8
c14.m112	114688	14
c20.m160	163840	20
c28.m224	229376	28

Buffalo Lake Effect Instances

Instance Name	VCPUs	RAM (GB)	Total Disk (GB)
c1.m1	1	1	20
c1.m4	1	4	20
c2.m4	2	4	20
c2.m8	2	8	20
c4.m8	4	8	20
c4.m16	4	16	20
c8.m16	8	16	20
c8.m32	8	32	20
c16.m32	16	32	20
c16.m64	16	64	20

Santa Barbara Aristotle Instances

Instance Name	VCPUs	RAM (MB)	Total Disk (GB)
c1.m8	1	8192	10
c4.16	4	16384	50
c8.m96	8	98304	100
c12.m64	12	65536	200
m1.tiny	1	512	0
m1.small	1	2048	10
m1.large	4	8192	10
m1.xlarge	8	8192	50

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

ARISTOTLE
CLOUD FEDERATION

"The whole is greater than the sum of its parts."

About ▾ Using Aristotle ▾ Science Teams ▾ Search Go

Use Cases

Seven strategic science use cases from intentionally diverse disciplines–earth and atmospheric science, finance, chemistry, astronomy, civil engineering, genomics, and food science–were selected to demonstrate the potential of a federated cloud in broadening cyberinfrastructure impact.

These use cases allowed us to explore data analysis techniques such as MapReduce, data stream processing, and machine learning and their applicability to different disciplines. We documented the tools, workflows, challenges, and best practices for each science case.

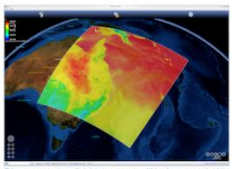
Learn more about the use case teams, rationale for using Aristotle, accomplishments, plans, and products.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Use Case

A Cloud-Based Framework for Visualization and Analysis of Big Geospatial Data

V. Chandola (University at Buffalo) with R. Vatsavai (North Carolina State University), P. Hogan (NASA Ames Research Center), B. B. Bhaduri (Oak Ridge National Laboratory)



Disparate geospatial data sources will be integrated with webGlobe and Aristotle

What?

- Integration, visualization, and analysis of geospatial data from around the world

Why Aristotle?

- Offloading of variable computational demands from local infrastructure
- Expert service hub for widely distributed data
- Scalable and elastic computing

Accomplishments

- Developed the capabilities to create Spark analytical clusters on-demand.
- Deployed a new distributed machine learning method for change detection in sustainability data.
- Ran scalability tests on Red Cloud at Cornell before migrating the application to Lake Effect cloud at the University at Buffalo.
- Developed webGlobe, a browser-based, cloud-driven interactive 3D user interface that allows scientists to upload, visualize, and analyze Network Common Data Form (NetCDF) data sets and is, at present, the only browser-based system available with this functionality.
- Completed initial runs of Gaussian Process-based change detection algorithm on 200 years of climate simulation data.
- Refined webGlobe's integrated visualization and analysis capabilities that support a variety of data formats prevalent in the climate and earth science communities.
- Continued development of an Energy-Water Knowledge Discovery Framework portal using webGlobe technology currently running on the Aristotle cloud.
- Performed comparative evaluations of various machine learning methods on Aristotle resources to better understand the Energy-Water nexus.
- Shipped a version of webGlobe to collaborators at Oak Ridge National Laboratory to support their research activities in the area of climate data analytics.
- Co-organized panel and gave poster presentation at the AGU Meeting on "Application of Information and Data Science Methods and Technologies to Climate Research and Energy-Water Knowledge Discovery."
- Developing new machine learning components to enhance webGlobe and a Jupyter Notebook extension of the framework to allow users to analyze underlying data.
- Created an interaction platform and tailored webGlobe for the Lower Great Lakes Resiliency Effort. Both are hosted on Aristotle.

Plans

- Continue development of iGlobe-based application to support sustainability research. Plan to create a data and computing infrastructure, powered by Aristotle, to support a large collection of researchers, community stakeholders, and planners, working on the sustainable development of the Lower Great Lakes area.
- Currently working on a paper that describes the OUTSTEP Community Platform and its digital capabilities.

Products

- Technology demonstration at the NSF-funded Organizing Urban Transects for a Sustainable Transformation of Economic Partnerships across the Lower Great Lakes Workshop (OUTSTEP 2019).
- Zaidi, S.M.A., Chandola, V., Allen, M.R. & Bhaduri, B.L. (2018). WebGlobe - A cloud-based geospatial analysis framework for interacting with climate data. *Proceedings of the 7th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*.
- Zaidi, S.M.A., Chandola, V., Allen, M.R. & Bahduri, B.L. (2018). *Anomaly detection in energy-water nexus.* Poster at the American Geophysical Union Fall Meeting, Washington, DC and co-organizer of a panel on the *Application of information and data science methods and technologies to climate research and energy-water knowledge discovery*.
- Mohammed, S.Z.A., Allen, M.R., Chandola, V. & Bhaduri, B.L. (2018). Machine learning for the energy-water nexus: challenges and opportunities. *Big Earth Data Journal*.
- Sharma, A., Mohammed, S.Z.A. & Chandola, V. (2018). A cloud-based geospatial analysis framework for interacting with climate data. *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.
- Allen, M., Mohammed, S.Z.A., Chandola, V., Morton, A., Brelsford, C.M., McManamay, R.A., Binita, K.C., Sanyal, J., Stewart, R.N. & Bhaduri, B.L. (2018). A survey of analytical methods for inclusion in a new energy-water nexus knowledge discovery framework. *Big Earth Data Journal*.
- Morton, A., Piburn, J., Stewart, R. & Chandola, V. (2017). *Development of a suite of analytical tools for energy and water infrastructure knowledge discovery.* Presentation at the American Geophysical Union Fall Meeting, New Orleans, LA.
- Sanyal, J., Chandola, V., Sorokine, A., Allen, M.R., Berres, A., Pang, H., Karthik, R., Nugent, P., McManamay, R. & Bhaduri, B.L. (2017). *Towards a web-enabled geovisualization and analytics platform for the energy and water nexus.* Presentation at the American Geophysical Union Fall Meeting, New Orleans, LA.
- Morton, A., Piburn, J., Stewart, R. & Chandola, V. (2017). *Leveraging advances in population modeling to support energy and water nexus knowledge discovery.* Poster at the American Geophysical Union Fall Meeting, New Orleans, LA.
- Co-Chaired 6th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (BigSpatial-2017).
- *WebGlobe Capabilities video* (2017) demonstrates webGlobe, a new tool developed by University at Buffalo use case scientist Varun Chandola. Powered by Aristotle, webGlobe provides access to and visualization of remotely available NetCDF geo data.
- Tran, D., Khoo, E.R. & Chandola, V. (2016). *WebGlobe: Big geospatial analytics on the cloud.* Presentation at MITRE Corporation, McLean, VA.

Learn more Next

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Use Case

Global Market Efficiency Impact

C. Tiu, D. Rösch, B. Wolfe (University at Buffalo) with S. Titman, K. Brown (University at Texas Austin), U. Yoeli (University of Texas), M. Carlson (University of British Columbia), A. Reed (TIAA), A. Subrahmanyam, R. Roll (UCLA), M. van Dijk, D. Bongaerts (Erasmus University)

What?

- Measure market efficiencies with high volumes of historical and current intraday market data

Why Aristotle?

- Bursting to integrate new data
- Tens of TBs moved to available processing resources
- Sharing data and pipelines around the world for collaboration and reproducibility

Accomplishments

- Developed the capability to migrate OpenNebula images to the cloud.
- Setup the framework on a VM and began to import and successfully analyze 2TBs of international tick-by-tick financial data from the Thomson Reuters Tick History (TRTH) database.
- Imported 25TBs Trade and Quote (TAQ) data up until 2017 for intraday trades and quote transactions for all US exchange-listed stocks, and computed 9 measures of efficiency.
- Created VMs in collaboration with Varun Chandola and Jialiang Jiang (UB) that share the same image and hold all the data; they are executing well.
- Demonstrated the framework to PhD students so they can analyze the data independently on Aristotle.
- Launched new investigations, i.e., how price deviations (market inefficiencies) affect liquidity (the ease at which you can buy or sell), how the recent increase in tick-size (the minimum price movement of US stocks) affects liquidity, etc.
- Gave a guest lecture at Cornell introducing the financial framework and data hosted on Aristotle.

Plans

- Use Aristotle and the financial framework to investigate whether human trades still matter at a time when trading is dominated by computers. Also, employ a LASSO regression to predict short-term stock returns using the whole cross-section of international stocks and investigate CumEx trading (labeled as the "Biggest Tax Heist Ever" by the *NYTimes*). All projects require large datasets and computational power. Collaborators include UCLA and the University of Utah.
- Continue to mentor PhD student researchers.

Products

- D. M. Rosch. A. Subrahmanyam & M.A. van Dijk (2021). *Investor short-termism and real investment.*. Journal of Financial Markets.
- J. Brogarrd, M.C. Ringgenberg & D. Rosch (2021). *Does the Trading Floor Matter?*. 8th Annual Conference on Financial Market Regulation.
- Presented "Does the Trading Floor Matter" at the 49th European Finance Annual Meeting in Milan (2021).
- Rosch, D. (2020). *Financial market frictions*. Presentation at the Cornell University Johnson School, Ithaca, NY.
- Hendershott, T., Livdan, D. & Roesch, D. (2020). Asset pricing: A tale of night and day. *Journal of Financial Economics*.
- Hendershott, T., Livdan, D. & Roesch, D. (2019). *Asset pricing: A tale of night and day*. Presentations at American Finance Association 2019 Annual Meeting, Atlanta, GA; Jackson Hole Finance Conference, Jackson Hole, WY; and, Desmarais Global Finance Research Centre, Montreal, Canada.
- Chung, K.H., Lee, A.J. & Rösch, D. (2019). Tick size, liquidity for small and large orders, and price informativeness: evidence from the tick size pilot program. *Journal of Financial Economics (accepted)*.
- Hendershott, T., Livdan, D. & Rösch, D. (2018). *Asset pricing: a tale of night and day*. Presentation at the Research in Behavioral Finance Conference, Amsterdam, NL.
- Rösch, D. (2017). *Investigating the efficiency of financial stock markets with high-frequency data*. Presentation at the 2nd Federal Reserve Bank Economic Research in High Performance Computing Environments Workshop, Kansas City, KS.

Time-variation in global and local market efficiency studies need large data storage

Learn more Next

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Use Case

Application of the Weather Research and Forecasting (WRF) Model for Climate-Relevant Simulations on the Cloud

S.C. Pryor (Cornell) with R. Levy (NASA Goddard), F. Yu (SUNY Albany), A. Hodzic (NCAR), P. Crippa (University of Newcastle), H. Matsui (Japan Agency for Marine-Earth Science and Technology), and B. Barker, P. Vaillancourt, B. Wineholt (Cornell CAC)

What?

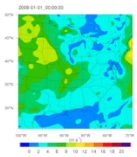
- Model the variability of wind speeds and atmospheric properties

Why Aristotle?

- Bursting to process new data
- Sharing of a high-value processed dataset of general interest

Accomplishments

- Built a physics-only version of the Weather Research and Forecasting (WRF) model using Docker and compiled it with parallel NetCDF to evaluate cloud-based performance.
- Ran high-resolution simulations to quantify wind climate and analyze the impact of large wind turbine developments on downstream climate (local to mesoscale).
- Evaluated 10-minute wind speeds from simulations relative to in-situ measurements from the National Weather Service Automated Surface Observing System (ASOS) on Jetstream to allow Aristotle to continue to focus on the numerical simulations.
- Analyzed high-resolution numerical simulations of the effects of wind turbines (WT) on regional climate.
- Completed simulations to test the sensitivity of the climate impacts to the precise description of the WT aerodynamics (the extraction of momentum and introduction of turbulence behind the turbine rotor).
- Analyzed long-term simulations with the WRF model to examine inter-annual variability of annual mean wind speeds at/near typical wind turbine hub-heights, and applied the power curve of the most commonly deployed WT to post-process the 10-minute wind speed output into estimated annual energy production (goal is to create a more robust prediction of the value of wind energy projects: analyses rendered possible by mounting a 100TB hard drive on an Aristotle instance).
- Analyzed output from simulation of WRF model at 12Km over eastern N. America for 2001-2016 for the assessment of year-to-year variability in the wind resource (150TB WRF runs).
- Enhanced high-resolution simulations of wind farm wakes from 2 parameterizations to advance methods to optimized wind turbine arrays and maximize system-wide power production. Analysis of output from simulation on Aristotle were conducted on Jetstream.
- Began new simulations for a domain centered over the Southern Great Plains on a single VM; this case would make an exceptional candidate for a trial of simulations across multiple VMs.
- Used machine learning and Red Cloud's large RAM and multi-processors to detect and quantify wind gusts at Newark, Boston, and Chicago O'Hare Airports. Found Artificial Neural Networks exhibit a higher skill than logistic and linear regression models for wind gust occurrence and magnitude.
- Simulated wind farm wakes from the east-coastal offshore lease areas using ultra-high resolutions with WRF on NERSC-Cori. Performed the data analyses on Aristotle's Red Cloud.
- 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.
- Used machine learning approaches for wind gust detection and quantification.
- Analyzed wind farm wake simulations from the east-coastal offshore lease areas 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 the levelized cost-of-energy.
- Benefitted from Aristotle's large RAM, multi-processors for analysis speed using parallel processors, and large disk volumes that enabled all WRF output to be hosted for analysis.



Policymakers and the wind energy industry need better climate and wind speed variability models. (click image for video)

Plans

- Complete simulations of derechos, fast-moving damaging deeply convective systems, associated with tornadoes, wind gusts, very heavy precipitation and hail.
- Under a new grant, examine the wind resources and optimize wind farm layouts for the offshore wind turbines along the U.S. east coast. This work strongly leverages our previous simulations on Aristotle and will entail additional WRF simulations with wind farm parameterizations enabled. The goal is to define the optimal density of wind farms to optimize system-wide power production and minimize the levelized cost-of-energy.

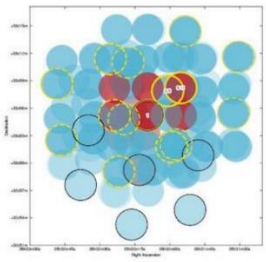
Products

- Aird, J.A., Barthelmie, R.J., Shepherd, T.J. & Pryor, S. C. (2022). Occurrence of low-level jets over the Eastern U.S. coastal zone at heights relevant to wind energy.. *Energies*.
- Aird, J.A., Barthelmie, R.J, Shepherd, T.J. & Pryor, S.C. (2021). WRF-simulated low-level jets over Iowa: characterization and sensitivity studies. *Wind Energy Science*.
- Barthelmie, R.J., Shepherd, T.J. & Pryor, S.C. (2021). *Offshore wakes in the U.S. east coast lease areas.* Presentation at the Wind Energy Science Conference (WESC), Hanover, Germany.
- Letson, F., Shepherd, T.J., Barthelmie, R.J. & Pryor, S.C. (2020). WRF modeling of deep convection and hail for wind power applications. *Journal of Applied Meteorology and Climatology*.
- Pryor, S.C., Shepherd, T., Volker, P., Hahmann, A. & Barthelmie, R.J. (2020). Diagnosing systematic differences in predicted wind turbine array-array interactions. *Journal of Physics: Conference Series*.
- Aird, J., Barthelmie, R.J., Shepherd, T.J. & Pryor, S.C. (2020). WRF-simulated springtime low-level jets over Iowa: Implications for wind energy. *Journal of Physics: Conference Series*.
- Barthelmie, R.J., Shepherd, T.J. & Pryor, S.C. (2020). Increasing turbine dimensions: Impact on shear and power. *Journal of Physics: Conference Series*.
- Shepherd, T.J., Barthelmie, R.J. & Pryor, S.C. (2020). Sensitivity of wind turbine array downstream effects to the parameterization used in WRF. *Journal of Applied Meteorology and Climatology*.
- Letson, F.W., Barthelmie, R.J. & Pryor, S.C. (2020). RADAR-derived precipitation climatology for wind turbine blade leading edge erosion.. *Wind Energy Science*.
- Pryor, S.C., Barthelmie, R.J. & Shepherd, T.J. (2020). 20% of US electricity from wind will have limited impacts on system efficiency and regional climate. *Nature: Scientific Reports*.
- Pryor, S.C., Shepherd, T.J., Volker, P., Hahmann, A.N. & Barthelmie, R.J. (2020). 'Wind theft' from onshore wind turbine arrays: Sensitivity to wind farm parameterization and resolution. *Journal of Applied Meteorology and Climatology*.
- Pryor, S.C., Shepherd, T.J., Bukovsky, M. & Barthelmie, R.J. (2020). Assessing the stability of wind resource and operating conditions. *Journal of Physics: Conference Series*.
- Letson, F., Shepherd, T.J., Barthelmie, R.J. & Pryor, S.C. (2020). Modelling hail and convective storms with WRF for wind energy applications. *Journal of Physics: Conference Series*.
- Pryor, S.C., Shepherd, T.J., Bukovsky, M. & Barthelmie, R.J. (2019) *Assessing the stability of wind resource and operating conditions*. Presentation at the North American Wind Energy Academy WindTech Conference, Amherst, MA.
- Leston, F., Shepherd, T.J., Barthelmie, R.J. & Pryor, S.C. (2019). *Modeling hail and convective storms with WRF for wind energy applications*. Presentation at the North American Wind Energy Academy WindTech Conference, Amherst, MA.
- Shepherd, T.J., Barthelmie, R.J. & Pryor, S.C. (2019). *Assessment of wind turbine impact on future climate in GCM-driven WRF simulations*. Presentation at the North American Wind Energy Academy WindTech Conference, Amherst, MA.
- Shepherd, T.J., Barthelmie, R.J. & Pryor, S.C. (2019). *Quantifying array-array effects using WRF model simulations: A sensitivity analysis*. Presentation at the North American Wind Energy Academy WindTech Conference, Amherst, MA.
- Shepherd, T., Wineholt, B., Barthelmie, R. & Pryor, S.C. (2019). *Quantifying weather and climate simulation reproducibility in the cloud*. Presentation at 99th American Meteorological Society Annual Meeting: 5th Symposium on High Performance Computing for Weather, Water, and Climate, Phoenix, AZ.
- Pryor, S.C., Shepherd, T.J., Barthelmie, R.J., Hahmann, A.N. & Volker, P. (2019) Wind farm wakes simulated using WRF. *Journal of Physics*.
- Barthelmie, R., Pryor, S.C. & Shepherd, T. J. (2019). *Wakes from wind turbine arrays*. Presentation at National Renewable Energy Laboratory, Golden, CO.
- Pryor, S.C., Shepherd, T.J., Bukovsky, M. & Barthelmie, R.J. (2019). *Wind energy scenarios for climate change mitigation*. Presentation at The Scenarios Forum 2019, Denver, CO.
- Letson, F., Barthelmie, R.J. & Pryor, S.C. (2019). RADAR-derived precipitation climatology for wind turbine blade leading edge erosion. *Wind Energy Science*.
- Pryor, S.C. & Hahmann, A.N. (2018). Downscaling wind: Oxford Research Encyclopedias: Climate Science. Oxford University Press, Von Storch, H., editor.
- *Aristotle WRF Container video* (2018) demonstrates the pulling or caching of Docker images to get the WRF app running quickly.
- Pryor, S.C., Shepherd, T.J. & Barthelmie, R.J. (2018). Inter-annual variability of wind climates and wind turbine annual energy production. *Wind Energy Science*.
- Pryor, S.C., Barthelmie, R.J. & Shepherd, T.J. (2018). The influence of real-world wind turbine deployments on regional climate. *Journal of Geophysical Research: Atmospheres* 123(11).
- Pryor, S.C., Barthelmie, R.J., Hahmann, A., Shepherd, T. & Volker, P. (2018): Downstream effects from contemporary wind turbine deployments. *Journal of Physics: Conf. Series 1037*.
- Pryor, S.C., Barthelmie, R.J., Hahmann, A., Shepherd, T. & Volker, P. (2018). *Contemporary wind turbine deployments have a minor impact on regional climate*. Presentation at the Science of Making Torque from Wind, Milan Italy.
- Shepherd, T.J., Volker, P., Barthelmie, R.J., Hahmann, A. & Pryor, S.C. (2018). *Sensitivity of wind turbine array downstream effects to the parameterization used in WRF*. Presentation at WRF/MPAS User's Workshop, Boulder, CO.
- Pryor, S.C., Barthelmie, R.J. & Shepherd, T. (2018). *Do current and near-term future wind turbine deployments have a substantial impact on regional climate?* Presentation at European Geosciences Union General Assembly 2018, Vienna, Austria.
- Pryor, S.C., Barthelmie, R.J. & Shepherd, T. (2018). *Assessing the fidelity of the North American wind climate and impacts of wind farms using high resolution modeling*. Presentation (recorded) at the 98th American Meteorological Society Annual Meeting (21st Conference on Planned and Inadvertent Weather Modification), Austin, TX.
- Pryor, S.C., Barthelmie, R.J. & Shepherd, T. (2018). *Improved characterization of the magnitude and causes of spatio-temporal variability in wind resources*. Presentation at the 98th American Meteorological Society Annual Meeting (31st Conference on Climate Variability and Change), Austin, TX.
- Pryor, S.C. (2017). *High resolution WRF simulations for resource assessment and quantification of downstream impacts of high density wind turbine deployments*. Presentation at DTU Wind Energy, Roskilde, Denmark.

Use Case

Transient Detection in Radio Astronomy Search Data

J. Cordes (Cornell) with M. McLaughlin (West Virginia University), V. Kaspi (McGill University), B. Allen (Albert Einstein Institute), A. Brazier, P. Vaillancourt (Cornell CAC), and PALFA collaborators located at the National Radio Observatory, Max Planck Institute for Radio Astronomy, University of British Columbia-Vancouver, University of Texas at Brownville, Naval Research Laboratory, Franklin and Marshall College, University of Wisconsin-Milwaukee, and Columbia University



On demand access to new survey data will provide astronomers and the public with more opportunities to make discoveries

What?

- Low-latency/on-demand searching of time-varying, including one-time signals in radio astronomy data
- Builds on existing Cornell/CAC collaboration

Why Aristotle?

- Bursting on-demand to analyze new data
- High-speed transport of very large data sets into and around the Federation

Accomplishments

- Developed a new containerized solution that features a full suite of all relevant pulsar-processing software, Python Conda distributions, the core PRESTO pulsar search package to perform transient search using the Spitler modulation index method, and the decimation code.
- Implemented an improved reproduction of the PALFA2 pipeline for detecting single pulse candidates that may be a Fast Radio Burst (FRB) source; this pipeline includes PRESO functionality, modulation index calculation, parameter customization, and the production of graphic data output.
- Built a new flexible framework for running radio astronomy searches; data are read from their native format into NumPy arrays, and the pipeline's routines are selected from a configuration file and includes a friends-of-friends search, and also allows the running of the pipeline of Laura Spitler who discovered FRB 1211102.
- Improved the Fast Radio Burst (FRB) Pipeline, a customizable scientific software pipeline for detecting single pulse candidates that may be FRB sources in radio astronomy data, by preparing to parallelize certain methods, updating the Python 3 version of PRESTO, and adding functionality for general radio transient detection beyond just FRBs.
- Added checkpoint/restart functionality to the pipeline to support tests on AWS spot instances as part of the RightScale investigation.
- Updated container and pipeline documentation.
- Implemented the workflow for the "Friends-Of-Friends" algorithm in parallel and prepared to deploy it in containerized form to a cluster of Spot instances in an automated cloud deployment.
- Demonstrated Friends of Friends workflows work well for smaller data sizes, but are prohibitively expensive if attempting to run large amounts of data for a long time. Exploration of how to do more processing in parallel is needed.
- Completed and tested the single container of radio astronomy software combining the pipeline components developed for Pulsar and other transient detections that can be deployed either on the cloud (Docker) or on an XSEDE HPC resource (Singularity), and released the container build files and images publicly via GitHub and DockerHub.
- Created a Singularity container packaging our pulsar and single pulse search pipeline, and secured trial XSEDE allocations for deploying the container on Bridges and for running pulsar/FRB searches on large data volumes.

Plans

- Continue to make improvements to the FOF algorithm in the FRB_pipeline and complete the comparison with Single Pulse Search method in PRESTO with the modulation index calculation. A new dataset that has not been processed yet will then be explored with the FOF algorithm in search of FRB detections.
- Continue to use Aristotle for the prototyping of FRB search codes and the development of deployable Docker and Singularity containers for deployment on large-scale XSEDE resources and public clouds.

Products

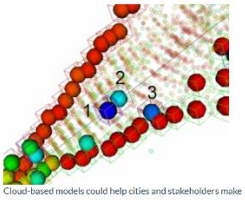
- Publicly released the container packaging of our pulsar and single pulse search pipeline. See the GitHub repo containing both the Dockerfile and Singularity definition file; the Docker Hub image; and, the Singularity Hub image.
- Chatterjee, S. (2018). *New results in radio astronomy: Fast radio burst and transients.* Presentation at the International Astronomical Union (IAU) General Assembly, Vienna.
- Chatterjee, S. (2018). *The dwarf galaxy host of a fast radio burst*. Presentation at the International Astronomical Union (IAU) General Assembly, Vienna.

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Use Case

Water Resource Management Using OPENMORDM Data

P. Reed (Cornell) with G. Characklis (UNC Chapel Hill), K. Keller (Penn State), and B. Barker, P. Vaillancourt, B. Wineholt (Cornell CAC). Testing and evaluation of the MORDM framework will leverage a collective partnership between Reed and the water utilities from the North Carolina cities of Cary, Durham, Raleigh, and Chapel Hill



Cloud-based models could help cities and stakeholders make complex water resource management decisions faster

What?

- On-demand simulation of complex water resource management outcomes to achieve substantial beneficial environmental impact on municipalities across the country

Why Aristotle?

- Coarsely parallel computations started on demand
- Very large computations, ideal for transfer to AWS or other cloud, to make available to many municipalities

Accomplishments

- Built and tested Parallel Platypus software (the Python version of Open MORDM) with a VM, container, and container source (Docker file) developed by the Aristotle team for use as a virtual MPI cluster on Aristotle.
- Added Docker documentation to the Aristotle wiki.
- Containerized and published build scripts for the Lake Problem code and successfully ran it in MPI across multiple virtual machines using Docker; this will allow distributed scientific software to be executed faster at cloud scale, both in existing institutional cloud, XSEDE resources, and public provider clouds.
- Regenerated all published results for the multi-node MPI Lake Problem simulation which will serve as validation of our containerization methodology.
- Ran MPI between Docker containers on multiple VMs with the Lake Problem code on Aristotle and Jetstream.
- Tested the Open MORDM group MPI WaterPaths software in a Docker container on single and multiple VMs. The goal is to show that WaterPaths can be used in the cloud with a performance comparable to that under HPC.
- Successfully performed OpenMORDM WaterPaths production runs at scale using up to 224 CPU cores on 8 worker VMs organized into an on-demand cloud MPI cluster with Docker container-based software deployment.
- Developed Terraform and Ansible scripts for on-demand creation of OpenStack cluster VMs with Docker installed and provisioned a cluster-shared NFS storage server.
- Modified an existing Docker container to create a VM for initial tests with WaterPaths, which automatically clones and compiles WaterPaths and Borg and downloads all necessary data files.
- Ran standalone WaterPaths simulations and small-scale optimize runs with four MPI processes (one master and three slaves).
- Defined guidelines for our next steps, which will include timings and MPI profiling. The Aristotle team is currently working on (1) using the same size VM with 28 cores, setting up 3 Docker containers to use 8 cores each, (2) setting up a small cluster of 3 VMs, each with 8 cores, with a single Docker container on each that uses all 8 cores, (3) setting up a larger cluster of 8 VMs, each with 28 cores, with a single Docker container on each VM. The timings and MPI profiling will allow for at least two publications, one focused on WaterPaths and the other on the cloud itself.

Plans

- David Gold is continuing the work of Bernardo Trindade on MORDM and WaterPaths.
- Consider building on our initial scaling analysis for WaterPaths and using Aristotle to support our water supply infrastructure investment and portfolio management analysis.
- Use Aristotle to more fully evaluate the core steps of our workflow: (1) generate alternatives with multi-objective optimization, (2) ensemble Monte Carlo evaluation of candidate alternatives in a broader sampling of state-of-the-world, (3) elicit acceptable limits and requirements on the range of performance attained on key measures of performance, (4) global sensitivity analysis to discover what factor and scenarios cause system failures. The workflow can be tested on a range of broader case studies such as the Sedento Valley benchmarking system, North Carolina Research Triangle, and the Tampa Bay Water Authority.
- Developed 3 new collaborations that wish to leverage Aristotle infrastructure, tools and techniques: (1) Woods Hole Oceanographic Institute and UCSB Marine Sciences plan to study forest dynamics by fusing and analyzing remote sensing data from the seabed and atmosphere off the coast of CA, (2) DOE researchers are building ground-based cloud observation networks to study cloud formation and lifecycle dynamics and will use Aristotle artifacts to deploy two new observation networks at Sedgwick and Long Island, (3) the Math Dept. at Cal State Long Beach will use Aristotle science team expertise in using, maintaining, and scaling clouds to automate and scale its calculus curriculum.

Products

- Trindade, B.C., Gold, D.F., Reed, P.M., Zeff, H.B & Characklis, G.W. (2020). Water pathways: An open source stochastic simulation system for integrated water supply portfolio management and infrastructure investment planning. *Environmental Modelling & Software.*

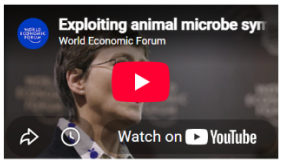
Learn more Next

The Aristotle Cloud Federation was supported by NSF grant ACI-1541215 and the Cornell University CAC in partnership with the University at Buffalo CCR and the University of California, Santa Barbara Department of Computer Science.

Use Case

Mapping Transcriptome Data to Metabolic Models of Gut Microbiota

A. Douglas (Cornell) with J. Chaston (Brigham Young University), A. Moya (University of Valencia), G. Thomas (University of York), A. Heddi (INSA Lyon), and B. Barker (Cornell CAC)



Angela Douglas explains how exploiting animal-microbe symbiosis may lead to pest-resistant crops

What?

- Use host and gut-microbial express data to infer which metabolic reactions occur in each organism as a result of dietary inputs

Why Aristotle?

- Heterogeneous instance types and sizes with high availability to allow for unpredictable computational demands
- Allows generation of reproducible computational biology pipelines in the form of VMs or VM configurations

Accomplishments

- Created a Windows instance for MATLAB and a Samba server for file sharing across Windows and Linux.
- Developed a transcriptomic pipeline for Docker.
- Documented scalable file access for cross-cloud and many-instance file/app access.
- Investigated Cornell Supercloud as possible means to scale out to other clouds.
- Developed and published a computational model for the whiteflies system.
- Analyzed and compared three independently-evolved communities in xylem-feeding insects. Generated advanced draft metabolic reconstructions for 5 bacteria needed to simulate Drosophila-microbial community metabolic interactions. The draft models of the 2 Lactobacilli have an average of 586 genes, 1193 metabolites and 750 reactions. The draft models of the 3 Acetobacter have an average of 686 genes, 1362 metabolites and 1129 reactions.
- Initiated a computational strategy on Aristotle, based on a new optimization framework SteadyCom, to obtain the metabolic fluxes for optimized relative abundances of the partners under equilibrium conditions of constant growth rates.
- Integrated the SteadyCom computational framework to model multi-species metabolic interaction in the insect gut.
- Performed SteadyCom on test models in Aristotle using a newly confirmed Windows VM with MATLAB and the COBRA Toolbox.
- Built a Linux VM with Docker, MATLAB, and Gurobi to containerize the simulation environment.
- Perform a Flux Variability Analysis on the four microbes composing the Drosophila gut microbiome.
- Constructed multi-species networks in silico of 2, 3, 4 and 5 bacterial species, quantified predicted metabolite fluxes, and wrote a MATLAB function to restore constraints to a multi-species model created from a single species.
- Verified multi-species models behave as expected with standard flux algorithms.
- Migrated a Windows VM to Linux VM in OpenStack, and wrote scripts and library functions to expedite running model simulations and choosing ideal model parameters.
- Using SteadyCom, investigated the scope of metabolic interactions that occur among a Drosophila gut microbial community.
- Identified 159 unique metabolites that are exchanged and showed that the gut microbial community is an important source of TCA cycle intermediates 2-oxoglutarate and succinate.
- Aristotle VM enabled metabolic interaction research was published in mBio Journal.
- Developed a new method, Semi Dynamic SteadyCom, to investigate priority effects among Drosophila gut microbes. Semi Dynamic SteadyCom employs SteadyCom with discrete time steps.
- Developed a modeling framework with strong static typing, named hCOBRA, in the Haskell programming language. hCOBRA extends the COBRA Toolbox for MATLAB with the aim of making large scale simulations more reliable and more scalable.
- Integrated functionality from the Funflow library into hCOBRA to allow caching of workflow results in a distributed setting.
- Added R-based analyses to our hCOBRA workflows through the HaskellR library.
- Expanded our analysis to investigate priority effects in a 5-member Drosophila gut microbe community comprising three Acetobacters and two Lactobacilli, which will require continued Aristotle cloud resources. Analyzed data from these simulations and prepared a manuscript.

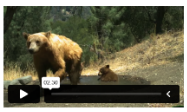
Products

- Ankrah, N.Y.D., Barker, B.E., Song, J., Wu, C., McMullen II, J.G. & Douglas, A.E. (2021). The predictive metabolic function of the gut microbiota of Drosophila melangogaster. *mSystems Journal.*
- Ankrah, N.Y.D., Wilkes, R.A., Zhang, F.Q., Zhu, D., Kaweesi, T., Aristilde, L. & Douglas, A.E. (2020). Syntrophic splitting of central carbon metabolism in host cells bearing functionally different symbiotic bacteria. *ISME Journal.*
- Ankrah, N.Y.D., Chouaia, B. & Douglas, A.E. (2018). The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. mBio.
- Douglas, A.E. (2018). Fundamentals of Microbiome Science: How Microbes Shape Animal Biology, Princeton University Press.
- Song, J., Barker, B., Ankrah, N. & Douglas, A. (2018). *Genome-scale metabolic modeling of gut microbiota in the fly gut.* NSF REU student poster at American Society for Engineering Education 2018 St. Lawrence Section Meeting, Ithaca, NY.
- Ankrah, N.Y.D., Luan, J. & Douglas, A.E. (2017). Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. Journal of Bacteriology 199(15).
- *Exploiting Animal-Microbe Symbiosis to Produce Pest-Resistant Crops* video (2017) features Aristotle use case scientist Angela Douglas describing at the World Economic Forum how understanding the gut microbiome may lead to the development of microbial therapies for everything from making honey bees healthier to restoring coral reefs.
- Douglas, A.E. (2017). *The Drosophila model for gut microbiome research.* Presentation at the NIH Common Fund: The Human Microbiome: Emerging Themes at the Horizon of the 21st Century, Bethesda, MD.
- Douglas, A.E. (2017). *Metabolic conversations in insect microbiomes.* Keynote lecture at the 14th Symposium on Bacterial Genetics and Ecology, Aberdeen, Scotland.
- Ankrah, N.Y.D. (2017). *Nutritional roles of beneficial bacteria associated with insects revealed by metabolic modeling.* Winning poster at the 2017 Copenhagen Bioscience Conference: Data-Driven Biotechnology – Bench, Bioreactor and Bedside, Copenhagen, Denmark.
- Douglas, A.E. (2017). *Metabolism and microbiomes: Metabolic models meet experimental data in insect-microbial symbiosis.* Presentation at Metabolism and Metabolomics: Disease Models and Model Organisms – 36th Annual Summer Symposium in Molecular Biology at The Pennsylvania State University, State College, PA.
- *Host-Microbe Interactions Affect Metabolism and Nutrition video* (2017) interview by Kiel University Professor Thomas Bosch of Aristotle use case scientist Angela Douglas.

Use Case

Multi-Sourced Data Analytics to Improve Food Production & Security

K. McCurdy (Sedgwick Reserve, University of California, Santa Barbara Natural Reserve System) with B. Roberts, B. Sethuramasamyraja (California State University, Fresno), B. Liu (California State Polytechnic University, San Luis Obispo), C. Krintz, R. Wolski (UC Santa Barbara)



UCSB use case scientists are identifying wildlife images with machine learning

What?

- Interdisciplinary analysis of data from a wide variety of sources, including the general public, to study and improve ecological outcomes

Why Aristotle?

- Scalable and portable infrastructure
- Access to multiple data sources, many of which are already in public and private clouds

Accomplishments

- Completed large-scale Google TensorFlow training and parallel image classification runs on Aristotle for the *"Where's the Bear?"* camera trap application analysis project (20 training runs using 2000 cores hours per run and a classification run using 1800 core hours to classify a test sample of 10,000 images prior to classifying 240,000 images from a single camera).
- Used soil moisture sensing devices, edge computing, and Aristotle to schedule vineyard irrigation (saved 66% of the water used previously).
- Installed hardware and software to instrument almond trees in a Fresno, CA test orchard to see how much water can be saved by irrigating the different sides of the root stock in proportion to its dryness.
- Developed innovative analytics with a minimally invasive instrumentation footprint for an Exeter, CA citrus test orchard to get highly accurate temperature readings at night when frost could form.
- Progress was made developing a low-power, low-cost, multi-tier IoT deployment for citrus frost prevention and the differential irrigation of almond trees.
- The Sedgwick Reserve "Where's the Bear" project is using the Pacific Research Platform's Kubernetes and containers environment to train image recognition models in conjunction with Aristotle.
- A new use case called Citrus Under Protective Screening (CUPS) was initiated. CUPS is a potential remedy for citrus greening which has devastated citrus in FL and is now threatening CA. Aristotle will serve the data hosting service for CUPS.
- 8 new publications (including a Best Student Paper award) and 2 keynote presentations were produced in PY4.
- Installed cameras and weather sensors at the UCSB Edible Campus farm site and developed the data acquisition system hosted on Aristotle.
- Installed sensors inside and outside the Citrus Under Protective Screening facility and deployed an IoT system that provides both real-time measurements via a visualization tool developed by Aristotle REU student Kareme Celik and a prototype frost alerting system.
- Developed the Aristotle AWS Pricing Tool to help users compare Aristotle resources to various AWS alternatives based on performance, cost, and price-performance.
- Developed 3 new collaborations that wish to leverage Aristotle infrastructure, tools and techniques: (1) Woods Hole Oceanographic Institute and UCSB Marine Sciences plan to study forest dynamics by fusing and analyzing remote sensing data from the seabed and atmosphere off the coast of CA, (2) DOE researchers are building ground-based cloud observation networks to study cloud formation and lifecycle dynamics and will use Aristotle artifacts to deploy two new observation networks at Sedgwick and Long Island, (3) the Math Dept. at Cal State Long Beach will use Aristotle science team expertise in using, maintaining, and scaling clouds to automate and scale its calculus curriculum.
- A new UCSB grant from the NSF titled "Detroit - A New End-to-End System for Practical and Accessible IoT" will leverage the Aristotle legacy.

Plans

- Develop a hybrid machine-learning and CFD model to support Citrus Under Protective Screening (CUPS) spraying operations and frost prevention.
- Enhance REU student Kerem Celik's telemetry data visualizer which works either as a local tool or as a service hosted in Aristotle.
- Develop sustainable land use practices at Sedgwick that employ livestock as part of the management lifecycle and install new monitoring infrastructure.
- Provide the instrumentation and analytics necessary to evaluate the first CUPS installation (at scale) in California. Aristotle will be providing the computational infrastructure to the team that is necessary to analyze the effects of CUPS on citrus production.

Products

.

- Zhang, M., Krintz, C. & Wolski, R. (2021). Sparta: A Heat-Budget-based Scheduling Framework on IoT Edge Systems. *International Conference on Edge Computing*.
- Zhang, M, Krintz, C. & Wolski, R. (2020). Edge-adaptable serverless acceleration for machine learning Internet of Things applications. *Journal of Software: Practice and Experience*..
- George, G., Bakir, F., Wolski, R. & Krintz, C. (2020). NanoLambda: Implementing Functions as a Service at all resource scales for the Internet of Things. *P2020 IEEE/ACM Symposium on Edge Computing (SEC)*..
- Dimopoulos, S., Krintz, C. & Wolski, R. (2020). Fair scheduling for deadline-driven, resource-constrained, multi-analytics workloads. *2020 International Conference on Computing, Networking and Communications (ICNC)*.
- Vaillancourt, P., Wineholt, B., Barker, B., Deliyannis, P., Zheng, J., Suresh, A., Brazier, A., Knepper, R. & Wolski, R. (2020). Reproducible and portable workflows for scientific computing and HPC in the cloud. *Practice and Experience in Advanced Research Computing (PEARC'20)*.
- Wolski, R., Lifka, D. & Furlani, T. (2020). *Data analysis and management for multi-campus cyberinfrastructure through cloud federation*. Poster presentation at the 2020 NSF Cyberinfrastructure for Sustained Scientific Innovation (CSSI) Principal Investigator Meeting, Seattle, WA.
- Wolski, R., Krintz, C., Bakir, F., Lin, W.T. & George, G. (2019). *Rethinking scalable services for the Internet of Things*. Presentation at the Big Data and Extreme-Scale Computing Workshop (BDEC2), San Diego, CA.
- Wolski, R., Krintz, C., Bakir, F., George, G. & Lin, W.T. (2019). CSPOT: portable, multi-scale functions-as-a-service for IoT. *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing (SEC'19)*.
- Dimopoulos, S., Krintz, C. & Wolski, R. (2019). Towards distributed, fair, deadline-driven resource allocation for cloudlets. *Proceedings of the 4th Workshop for Edge Clouds and Cloudlets*.
- Golubovic, N., Krintz, C. & Wolski, R. (2019). A scalable system for executing and scoring K-means clustering techniques and its impact on applications in agriculture. *International Journal of Big Data Intelligence*.
- Bakir, F., Woksi, R., Krintz, C. & Ramachandran, G. (2019). Devices-as-services: Rethinking scalable service architectures for the internet of things. *Proceedings of the 2019 USENIX Conference*.
- Golubovic, N., Wolski, R., Krintz, C. & Mock, M. (2019). Improving the accuracy of outdoor temperature prediction by IoT devices. *IEEE Conference on IoT*.
- Zhang, M., Krintz, C. Wolski, R. & Mock, M. (2019). Seneca: Fast and low cost hyperparameter search for machine learning methods. *IEEE Cloud 2019*.
- Carson, K., Thomason, J., Wolski, R., Krintz, C. & Mock, M. (2019). Mandrake: Implementing durability for edge clouds. *IEEE International Conference on Edge computing*.
- Lin, W-t., Bakir, F., Krintz, C. & Mock, M. (2019). Data repair for distributed event-based IoT applications. *Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems*.
- George, G., Wolski, R., Krintz, C. & Brevik, J. (2019). Analyzing AWS spot instance pricing (2019). IEEE International Conference on Cloud Engineering (IC2E).
- Krintz, C. *SmartFarm: IoT systems that simplify and automate agriculture analytics*. Keynote at 8th International Conference on Internet of Things (IoT 2018) Santa Barbara, CA.
- Krintz, C. *Adventures and opportunities in cyber-physical systems and research*. Keynote at 2018 International Conference on Computer Aided Design, San Diego, CA.
- *Centaurus K-Means Clustering as a Service video* (2018) demonstrates the balancing of a scalable clustering analysis workload between two Aristotle clouds.
- Lin, W.T., Krintz, C. & Wolski, R. (2018). Tracing function dependencies across clouds. *Proceedings of the IEEE 11th International Conference on Cloud Computing*.
- Jayathilaka, H., Krintz, C. & Wolski, R. (2018). Detecting performance anomalies in cloud platform applications. *IEEE Transactions on Cloud Computing*.
- Golubovic, N., Gill, A., Krintz, C. & Wolski, R. (2018). CENTAURUS: A cloud service for K-means clustering. *2017 IEEE 15 Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Cyber Science and Technology*
- *Where's the Bear* article (2018) highlights how UCSB use case scientists Chandra Krintz & Rich Wolski are identifying wildlife with machine learning and their IoTDI publication details how this automated wildlife image processing uses IoT, edge cloud systems, and Aristotle.
- Wolski, R., Brevik, J., Chard, R. & Chard, K. (2017). Probabilistic guarantees of execution duration for Amazon spot instances. *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis - SC17*.
- Dimopoulos, S., Krintz, C. & Wolski, R. (2017). PYTHIA: Admission control for multi-framework, deadline-driven, big data workloads. *Proceedings of 2017 IEEE 10th International Conference on Cloud Computing – CLOUD*.
- Wolski, R. & Brevik. J. (2017). QPRED: Using quantile predictions to improve power usage for private clouds. *Proceedings of 2017 IEEE 10th International Conference on Cloud Computing – CLOUD*.
- Wolski, R. (2017). *Keynote 2: Cloud computing status and future*. Panelist at 10th IEEE International Conference on Cloud Computing, Honolulu, HI.
- Elias, A.R., Golubovic, N., Krintz, C. & Wolski, R. (2017). Where's the Bear? - Automating wildlife image processing using IoT and edge cloud systems. *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation - IoTDI*.
- Jayathilaka, H., Krintz, C. & Wolski, R. (2017). Performance monitoring and root cause analysis for cloud-hosted web applications. *Proceedings of the 26th International Conference on World Wide Web – WWW*.
- Brevik, J. & Wolski, R. (2016). Providing statistical reliability guarantees in the AWS spot tier. *Proceedings of the 24th High Performance Computing Symposium*.