ARISTOTLE
CLOUD FEDERATION

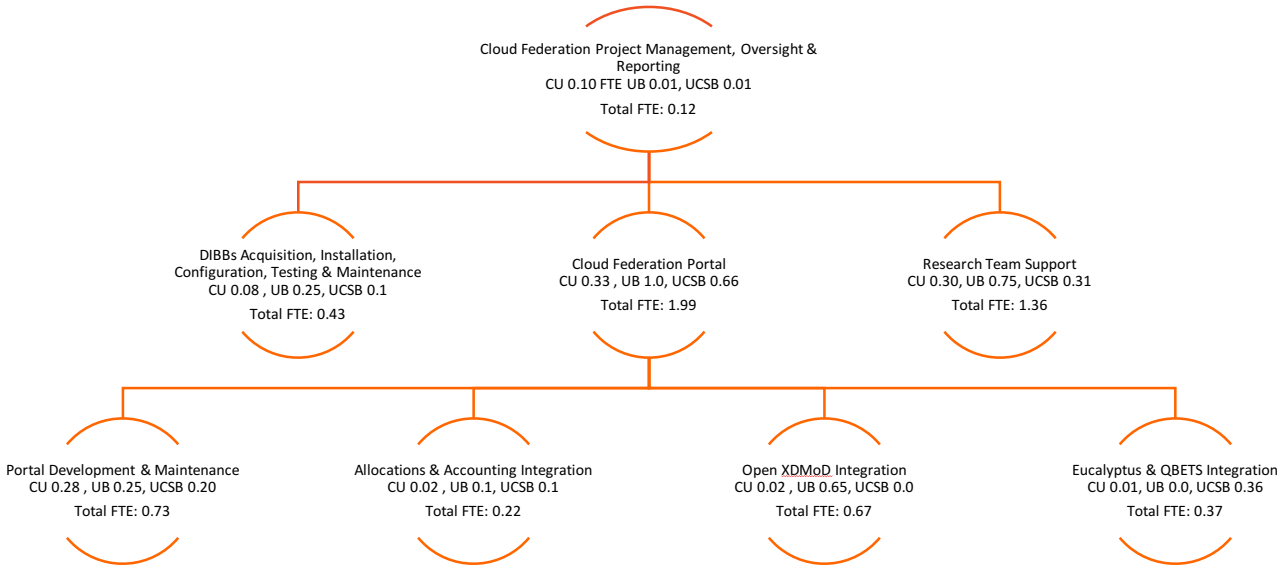
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 2: Quarterly Report 4

9/19/2017

Submitted by David Lifka (PI)
lifka@cornell.edu

This is the Program Year 2: Quarterly Report 4 of the Aristotle Cloud Federation team. We report on plans and activities for each area of the project Work Breakdown Structure (WBS).



#1541215

Contents

#1541215

1.0 Cloud Federation Project Management, Oversight & Reporting Report

1.1 Subcontracts
All subcontracts are in place. Nothing new to report.

1.2 Project Change Request
No new project change requests were made this quarter.

1.3 Project Execution Plan
The Project Execution Plan (PEP) was approved by NSF on 12/18/2015. We are operating as planned and continuously updating our PEP on a monthly basis.

1.4 PI Meetings
There were no Aristotle PIs meetings this quarter.

1.5 Project Status Calls

7/17/2017 call:

- Discussed Singularity, a container gear for HPC systems and how it may solve some of the issues Docker has with MPI. UB may consider Singularity for security purposes.
- The big geo data case team plans to expand access to their application that allows visualization and analysis of a wide variety of geospatial data sets. An access link will be available at the Aristotle portal.
- The Open XDMoD team plans to parse some Euca data to see how it works.
- The science use case teams at Cornell and UB are very happy with the progress and contributions their REU students have made thus far. Besides work on the Aristotle platform, the students are making progress using Jetstream, in particular for the Sara Pryor modeling project which is taking advantage of Jetstream's MATLAB and FBToolkit capabilities.
- UCSB is looking at AWS Lambda for Function as a Service (FaaS) for either HPC or Internet of Things event-driven applications. FaaS allows users to run application functionalities (i.e., microservices) without the users needing to directly manage the resources.
- Cornell has their new Euca 4.4 cluster with Ceph backend ready for testing and is working out the logistics of how to migrate it to a production cloud.
- UB has begun working on an OpenStack development cloud and the integration of Ceph with Puppet modules for easy build and tear down. They are maintaining, as are Cornell and UCSAB, their Euca production cloud in the meantime so that no science use case research is impeded.
- UCSB is retrofitting their network for future expansion and integrating it with a donated cloud infrastructure provided by HPE.

8/15/2017 call:

- After multiple discussions with the cloud team at Red Hat, they appear amenable to consider Aristotle as one site (rather than three). We are pursuing a packaged academic support subscription for both Red Hat OpenStack and Red Hat Ceph. Red Hat focus appears to be more on who has access to the bits rather than who is consuming the bits. They have received approval for the Aristotle project from senior management and we are in final stages of securing an agreement.

NSF
#1541215

- UCSB is creating a new website for DrAFTS. It will include demo code with a new front-end and programmatic API. The demo code will be beta tested by the Globus team in the fall.
- UCSB is also starting to work on OpenStack deployment efforts and a way to maintain cloud and Ceph sharing. There was also a discussion regarding whether UCSB should upgrade to Eucalyptus 4.4. or go directly to OpenStack.
- UB's strategy is to move from an OpenStack test cloud to a mini production cloud. They will migrate instances from the Euca cloud and leverage Red Hat support where needed once the academic support subscription is in place.
- The Open XDMoD for cloud team at UB brought in Euca logs into the data warehouse by capturing log events. They are examining the data and seeing what is there and not there in order to figure out what metrics to tackle first. They also set up Nagios monitoring on the UCSB cloud.
- The Patrick Reed water management resource decision-making application appears to be a strong candidate for containerization.

8/29/2017 call:

- Negotiations concluded with Red Hat for a one site, packaged academic support subscription for OpenStack and Ceph. Subscription details were clarified. UB plans to use Red Hat for OpenStack and Ceph. Cornell plans to use Red Hat for OpenStack and possibly Ceph.
- Cornell (Brazier) will be working with the Jetstream team to facilitate access to Jetstream for UCSB.
- Besides using Aristotle, Pryor is now using Jetstream for processing outcomes with MATLAB.
- Angela Douglas's use case benefitted from an REU student who conducted gut microbial classification.
- Aristotle analysis at UCSB led to new, more accurate irrigation scheduling at Sedgwick that uses much less water. Plan to automate camera trap image delivery in the fall when students return and a network upgrade is completed. UCSB is also investigating AWS image analysis services. UCSB has determined that not upgrading to Eucalyptus 4.4 will not negatively impact their science use cases and will therefore be moving directly to OpenStack.

1.6 Supplemental Proposal

In June, Cornell PI Lifka and the Aristotle co-PIs submitted a proposal to NSF titled "Adapting the Federated Cloud to New Software Environments: Supplement to Aristotle" to support 3 principle activities:

- the transition of Aristotle from Eucalyptus to OpenStack,
- an exploration of containerization for scientifically important benchmarks,
- an exploration of the use of containers by the Aristotle science teams to achieve cross-cloud deployment portability.

The $986,580 supplement proposal was approved.

Deliverables include:

- Conversion from Eucalyptus to OpenStack at each of the partner sites and the successful transition of currently running science VMs to the new OpenStack architecture.
- Delivery of a set of containerized applications that can be used to demonstrate functionality of the application in a container as well as performance of the application on multiple systems. Lessons

NSF
#1541215

learned from the process will be documented and disseminated to the cyberinfrastructure community.

2.0 DIBBs Acquisition, Installation, Configuration, Testing & Maintenance Report

2.1 Hardware Acquisition

- The Cornell team completed negotiations with Red Hat for OpenStack support.
- Cornell purchased 4 new servers using non-DIBBs funds to add 116 cores to Red Cloud as the system was running at 80% steadily.
- UB ordered the 10G network switches and servers required for a basic OpenStack/Ceph foundation that they will migrate Lake Effect cloud instances to.
- UCSB had no new hardware additions this quarter.

2.2 Software Installation, Configuration, and Testing

- Cornell installed, tested, and moved into production a Eucalyptus 4.4 cloud and a 1PB Ceph storage pool. The Ceph storage will be used by both the production Eucalyptus cloud as well as a new OpenStack cloud. As nodes are freed up from the Eucalyptus 4.2 cloud, they will be split between the 4.4 cloud and an OpenStack test cloud.
- UB is moving full steam ahead on building OpenStack. They have built a development OpenStack cloud that they are testing/developing with. They are using Foreman with Puppet to automate the installations of OpenStack and are well on their way to having it all working. The next step is to begin to install the production cloud. They are still waiting for the servers that were ordered and the Red Hat support contract to kick in so that they can get support when needed. As they free up node controllers from the Eucalyptus cloud, they will migrate them to the OpenStack Cloud. They will then move on to migrating the Aristotle cloud instances/hardware.
- UCSB spent time this quarter troubleshooting hardware problems with a Node Controller CPU resetting/failing; it was brought offline until repaired. Like Cornell and Buffalo, UCSB is prepping for OpenStack deployment using Red Hat's OpenStack Standard Release (currently version 11). They are continuing their work on improving the redundancy of Ceph connectivity — using bonded NICs to stacked switches. They also worked on stabilizing the Buffalo log-ingestor API client, along with monitoring from both UCSB and UB via NAGIOS. They attended Red Hat's OpenStack CL210 training which was held onsite.

The infrastructure status table was updated this quarter:

	Cornell (CU)	Buffalo (UB)	Santa Barbara (UCSB)
Cloud URL	https://euca4.cac.cornell.edu	https://console.ccr-cbls-2.ccr.buffalo.edu/	https://console.aristotle.ucsb.edu
Cloud Status	Production	Production	Production
Euca Version	4.4	4.4	4.2.2

#1541215

Hardware Year 1	Dell	Dell	Dell
Hardware Year 2	Dell	Dell, Ace	Dell, HPE
DIBBs Purchased Cores	*168	**256	356
RAM/Core	4GB/6GB	up to 8GB	9GB Dell, 10GB HPE
Storage	SAN (226TB), Ceph (1152TB)	SAN (336TB) Ceph (144TB)	Ceph (528TB)
10Gb Interconnect	Yes	Yes	Yes
Largest Instance Type	28 core/192GB RAM	24 core/192GB RAM	16 core/16GB RAM
Global File Transfer	Yes	Planned	Planned
Globus OAuth 2.0	Planned	Planned	Planned
	* 168 additional cores augmenting the existing Red Cloud (488 total cores)	** 256 additional cores augmenting the existing Lake Effect Cloud (424 total cores)	

2.3 Potential Tools

- DXC Technology Eucalyptus – waiting on a bug fix. All reams have a work-around plan provide by the Buffalo team.
- Supercloud - nothing new to report.

3.0 Cloud Federation Portal Report

Content updates to the project are ongoing: https://federatedcloud.org. These include adding a link to the webGlobe analysis and visualization tool for geospatial data which runs on top of Aristotle.

We continue to monitor the usage graph (https://federatedcloud.org/using/federationstatus.php) to ensure data is being collected consistently from all sites. We began implementing software to verify that the data ingestion API is running. Nagios is configured at the University of Buffalo to verify that the API is currently working, and is planned for the other federation sites. The software has successfully identified issues during the past two Eucalyptus upgrades. The checks being performed are:

- First result is the most recent record from the database. Ensure this record is not more than 1-hour old, otherwise the API is likely down.
- Check all records to ensure Free >= 0 (should always be 0 or positive).

#1541215

- Check all records to ensure Capacity > 0 (should never be 0).

No changes were made to the portal planning table this month.

Portal Framework			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 12/2016**	**1/2017 - End**	**1/2017 - End**
Gather portal requirements, including software requirements, metrics, allocations, and accounting. Install web site software.	Implement content/functionality as shown in following sections. Add page hit tracking with Google Analytics, as well as writing any site downloads to the database.	Implement content/functionality as shown in following sections. Add additional information/tools as needed, such as selecting where to run based on software/hardware needs and availability.	Release portal template via GitHub. Update periodically.
Documentation			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 10/2016**	**11/2016 – End**	**1/2017 - End**
Basic user docs, focused on getting started. Draw from existing materials. Available through CU doc pages.	Update materials to be federation-specific and move to portal access.	Add more advanced topics as needed and after implementation in Science Use Cases, including documents on “Best Practices” and “Lessons Learned.” Check and update docs periodically, based on ongoing collection of user feedback	Release documents via GitHub. Update periodically.
Training			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 12/2017**	**4/2017 – 12/2017**	**1/2018 - End**
Cross-training expertise across the Aristotle team via calls and science group visits.	Hold training for local researchers. Offer Webinar for remote researchers. Use recording/materials to provide asynchronous training on the portal.	Add more advanced topics as needed. Check and update materials periodically, based on training feedback and new functionality.	Release training materials via GitHub. Update periodically.
User Authorization and Keys			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 1/2016	**2/2016 – 5/2016**	**6/2016 – 3/2017**	**4/2017 – End**
Plan how to achieve seamless login and key	Login to the portal using InCommon.	Beta testing Euca 4.4 with Euca console supporting Globus Auth. Will deploy	Move seamlessly to Euca console after portal Globus Auth login.

#1541215

transfer from portal to Euca dashboard.		and transition to Euca 4.4 on new Ceph-based cloud.	
Euca Tools			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 12/2016**	**1/2017 – End**	**1/2017 – End**
Establish requirements, plan implementation.	No longer relevant since Globus Auth will let us interface with Euca web console	N/A	N/A
Allocations and Accounting			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2017	**3/2017 –5/2017**	**6/2017 – 10/2017**	**6/2017 – End**
Plan requirements and use cases for allocations and account data collection across the federation. Design database schema for Users, Projects and collections of CPU usage and Storage Usage of the federated cloud.	Display usage and CPU hours by account or project on the portal. Integration hooks for user and project creation/deletion and synchronization across sites.	Automate project (account) creation by researcher, via the portal.	Report on usage by account, if the researcher has multiple funding sources. Release database schema via GitHub.

3.1 Software Requirements & Portal Platform

No activity this quarter.

3.2 Integrating Open XDMoD and DrAFTS into the Portal

The Aristotle Buffalo team augmented the Euca log scraper to track data on volume creation and destruction so that better reporting on storage usage can be produced.

Data is being ingested periodically (and working on a daily ingestion process) and work has begun on developing the following initial metrics (more are planned; these are the low-hanging fruit):

- Resources: tracking state and settings of whole resources.
- Requests: tracking successful and failed requests for start, stop, suspend, resume, create, attach, etc.
- Events: tracking activity of instances in clouds (e.g., volumes attached).
- States: tracking time instances spent in various states (e.g., running, migrating).

We have also re-engaged with the Jetstream team to look at bringing in data from OpenStack.

The timeline table is below; there were no changes this month.

#1541215

Task	Order	Days	%	Feb 1	Feb 2	March 3	March 4	March 5	March 6	April 7	April 8	April 9	April 10	May 11	May 12	May 13	May 14	June 15	June 16	June 17	June 18	July 19	July 20	July 21	July 22	August 23	August 24	August 25	August 26	September 27	September 28	September 29	September 30
Update cloud schema docs	1	5	100																														
Implement static dimension tables	2	3	100																														
Document JSON import schema	3	2	100																														
Documentation Updated	3M	0			X																												
Define file-based API	4	2	100																														
ETL config file references	7	6	100																														
Subquery support	8	3																															
Improve start/end/number handling	9	3	100																														
Action date chunk handling	11	3	100																														
ActionState	12	3	100																														
Pre-Cloud Refactoring	13	6	100																														
ETL Required Additions	13M	0											X																				
XDMoD cloud event mapper	14	3	25																														
Updates to StructuredFile ingestor	15	6	100																														
Build single record ingestor	16	4	100																														
Build directory crawler	17	6	100																														
Aggregator	18	4																															
Testing	19	4																															
Eucalypus Data Ingested	19M	0																X															
Migrate from modw.jobfact	20	5	50																														
Implement realm & metrics	21	12																															
PEARC17	22	5																															
Implement drilldowns	23	10																															
Testing	24	10																															
Cloud Beta (Accounting)	24M	0																											X				
Summarization Infrastructure	25	20	75																														
Develop summarization tools	26	40	75																														
Summary docs in mongodb	27	15	25																														
Ingest summaries into XDMoD	28	15																															
Display perf data via existing Job Viewer	29	20	25																														
Cloud Beta (Performance)	29M	0																							X								

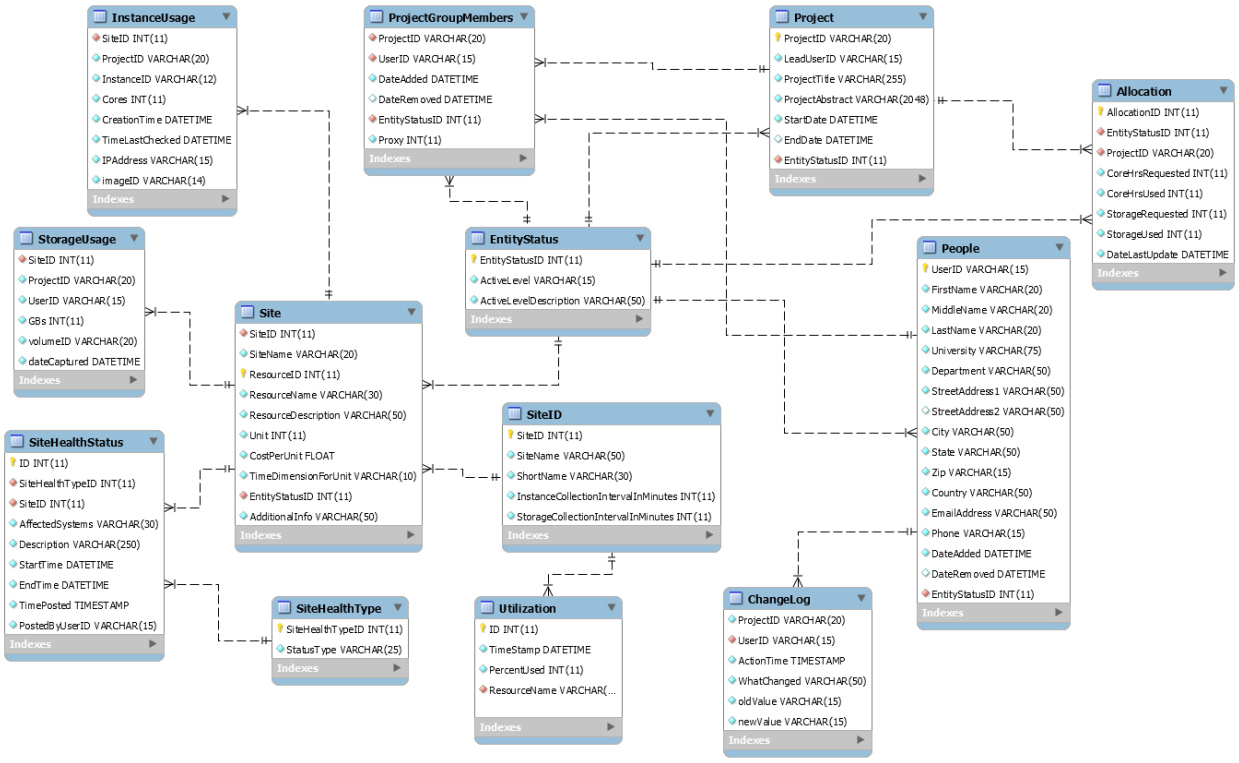
ETLv2 Housekeeping
Cloud Data
Cloud Realm
SUPReMM
Milestone

3.3 Allocations & Accounting

The login procedure was updated to write any missing member data to the accounting database.

There were no changes made to the database schema this quarter:

#1541215



4.0 Research Team Support

4.1. General Update

- Aristotle REU students completed their funded work. For example, Cornell's Thomas Biondi processed over 10TB of complex datasets produced by weather models on multiple computing platforms; used Aristotle to store the large amounts of data while simultaneously running the data through code on a Linux virtual machine; applied statistical metrics to compute weather model accuracy; used machine learning algorithms to predict days with high wind speeds, and created visualization to present the data in a way that a lay audience could understand. Biondi is an author on a submission to the 98th American Meteorological Society Annual Meeting (31st Conference on Climate Variability and Change) in January 2018. In another example, Cornell REU students Elizabeth Holzknecht and Shiva Laskhamann worked on bringing radio astronomy search data compression to a stable state.
- Containerization work has commenced with a focus on training the scientists in the research groups so that they can train the other members of their groups. Cross cloud-stack testing of the containerized science Use Case solutions will initially take place on Jetstream (which is based on OpenStack) while the Aristotle transition to OpenStack is underway.
- Cornell Professor Sara Pryor (Use Case 3) and UB Professor Varun Chandola (Use Case 1) have begun cross-site, cross-use case collaboration discussions. Pryor is exploring the possibility of using webGlobe visualization technologies developed and deployed on Aristotle by Chandola for the output of her Aristotle science runs. Pryor is also is planning a submission for an additional Jetstream allocation to continue the research that they have been conducting with the Aristotle-originated Jetstream allocation.

4.2 Science Use Case Team Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data
The big geo team is currently developing an application based on webGlobe to understand the Energy Water Nexus in future climate scenarios. This effort is part of a separately funded grant (DOE) whose focus is to understand how the relationship between energy and water will evolve under various future climate scenarios. The team is currently developing a version of webGlobe to answer those questions by integrating climate simulations along with other energy and water usage data sets (GIS format) and developing analytical tools for the combined analysis

Use Case 2: Global Market Efficiency Impact
Dominik Roesch (Aristotle UB science team lead) will present the project's approach to analyzing financial data and ideas about market efficiency to the Federal Reserve Bank of Kansas City. During the summer, Professor Roesch introduced the framework to analyze financial data to two UB PhDs. Both are about to finish writing their first queries (similar to a stored procedure in SQL) that will allow the finance research group to start analyzing data. One of the projects will be related to the recent tick-size pilot study by the SEC. Roesch also completed replicating a previous paper related to another project and shared first results with the co-authors.

Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate
Précis purpose of current suite of simulations:

1. Quantify impact of resolution (to convective permitting scales) on near-surface flow (i.e. wind speed) regime fidelity.
2. Examine scales of coherence in wind fields. Specifically spatial scales of calms (i.e. wind speeds < 4 m/s), and spatial scales of intense wind speeds (i.e. wind speeds > the local 90th percentile value).
3. Quantify the platform dependence of wind simulations (i.e. quantify the differences in nearsurface wind regimes from simulations conducted on conventional HPC and the Aristotle cloud).
4. Examine inter-annual variability in near-surface wind speeds (Can we simulate it? What is the source?)
5. Evaluate impact of large wind turbine (WT) developments on downstream climate (local to mesoscale).

Activities this quarter have focused on two key aspects:

1. Continuation of the numerical simulations with the Weather Research and Forecasting (WRF) model – see status below.
2. Commence evaluation of 10-minute wind speeds from simulations on the two computing platforms and relative to in situ measurements (from the National Weather Service Automated Surface Observing System (ASOS). This work was partly undertaken by REU undergraduate student Thomas Biondi and is largely being conducted on the NSF Jetstream cloud to enable Aristotle activities to continue to focus on the simulations.

#1541215

Status of WRF simulations *(note: Cori is a DOE Cray)*:

Simulation (Platform)	Domains	Wind turbine installed capacity	Status	Purpose
2008 eastern USA no WT (Cori)	12 km, 4 km	None	Completed	1,2
2008 eastern USA WT (Cori)	12 km, 4 km	Current	Completed	2
2008 eastern USA (Cori)	12 km	None	Completed	4,5
2008 eastern USA (Aristotle)	12 km	None	Completed	5
2008 Iowa WT (Aristotle)	12 km, 4 km	Current	Completed	1
2008 Iowa no_WT (Aristotle)	12 km, 4 km	None	Completed	1
2008 eastern USA double WT capacity [2WT] (Cori)	12 km, 4 km	Doubled	Running	1
2008 eastern USA quadruple WT capacity [4WT] (Cori)	12 km, 4 km	Quadrupled	Running	1
Contemporary climate decade eastern USA (Aristotle)	12 km	None	Running	3
Future climate decade	12 km	None	Pending	3

Example results based on analyses of output from simulations of 2008:

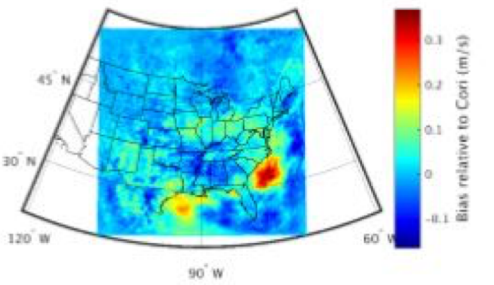


Figure 1: Mean difference (bias) in wind speeds (ms^{-1}) at 10-m a.g.l. computed from all 10-minute values from 12 km simulations on Cori and Aristotle. Implication: Compiler differences matter to simulated wind fields particularly in areas with complex discontinuities in land surface conditions.

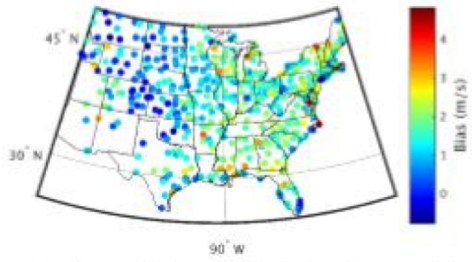
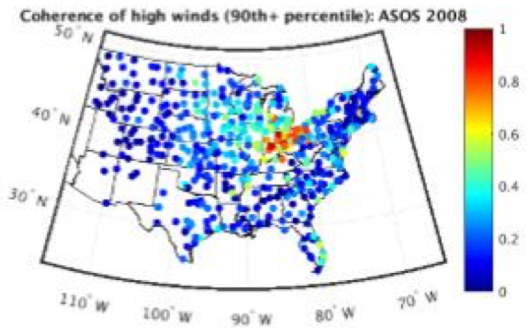


Figure 2: Mean difference (bias) in wind speeds (ms^{-1}) at 10-m a.g.l. as simulated with WRF on Cori at 12 km versus observations from approx. 600 NWS Automated Surface Observing Stations. Implication: Substantial spatial variability in model fidelity.



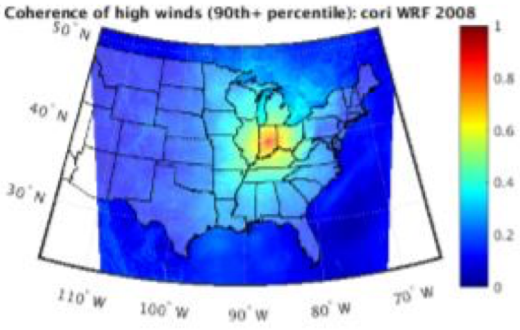


Figure 3: Diagnostic analysis of model skill. Left: Coherence of the occurrence of wind speeds above the locally determined 90th percentile value. Right: Same analysis but for output at 12 km from WRF simulations conducted on Cori. Interpretation: A value of 0.2 indicates that 1 in 5 events at the centroid station/grid cell (s. Indiana) is associated with a local exceedence at given station/grid cell. Implication: WRF captures broad scales of coherence.

Conference abstracts submitted:

- Pryor S.C., Barthelmie R.J., Biondi T. and Shepherd T. (2018): Improved characterization of the magnitude and causes of spatio-temporal variability in wind resources. *98th American Meteorological Society Annual Meeting (31st Conference on Climate Variability and Change),* Austin TX, January 2018 *(abstract submitted).*
- Shepherd T., Barthelmie R.J. and Pryor S.C. (2018): Assessing the fidelity of the North American wind climate and impacts of wind farms using high resolution modeling. *98th American Meteorological Society Annual Meeting (21st Conference on Planned and Inadvertent Weather Modification),* Austin TX, January 2018 *(abstract submitted).*
- Pryor S.C., Barthelmie R.J. and Shepherd T. (2017): High-fidelity simulations of the downstream impacts of high density wind turbine deployments. *4th Workshop on Systems Engineering for Wind Energy,* Roskilde, Denmark, September 2017 *(abstract submitted).*

Other presentations:
Pryor S.C. (2017): High resolution WRF simulations for resource assessment and quantification of downstream impacts of high density wind turbine deployments. *DTU Wind Energy Dept.* Roskilde, Denmark, August 23, 2017 (*invited presentation*).

Use Case 4: Transient Detection in Radio Astronomy Search Data
Working with recent PhD recipient Robert Wharton and Aristotle Science Team Lead Adam Brazier, REU students Holzknecht and Lakshmanan completed code to down-resolve the data and are working on testing it with Laura Spitler's code to evaluate the effects on sensitivity to single pulse/transient sources of the down resolution process. This process has been performed on PALFA data, which is the initial target for the project, but should be generalizable to other data in the PSRFITS (a standard for data storage) data format.

Use Case 5: Water Resource Management Using OpenMORDM
Brandon Barker (Cornell CAC) successfully trained two lab members in Docker basics: users are now familiar with building container images, stopping and starting images, and other container management tasks. One user was able to update the Git commit of three MORDM-related software packages in the Docker file and rebuild the image without issue. Users are currently working to test out a variety of software in their single node container environment, and have expressed a strong interest in moving forward to a multi-node environment. Unlike other MPI container use cases studied so far on Aristotle, this use case is expected to have a relatively low amount of communication, making it more likely to be directly amenable to a multi-node environment in the cloud.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota
We (the Angela Douglas Aristotle research team) have now completed the reconstruction of the metabolic model for the fifth bacterium *Acetobacter fabarum* DLS54. Over the coming months, we will be able to conduct a detailed manual curation of the five bacterial models, prior to the analysis of among-partner transport fluxes to generate a single multi-compartment model. Our gut transcriptome data are also completely analyzed (Bost et al. submitted) and in a suitable format for generating the sixth metabolic sub-model for the host.

Working with Brandon Barker, we have initiated a computational strategy, 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. We have installed the latest version of the COBRA

NSF
#1541215

Toolbox that supports SteadyCom on a Windows system running on Aristotle Red Cloud, including some detailed debugging. The next step is to verify that SteadyCom can run on three-compartment models already reconstructed in the laboratory. We expect that to be done by the next quarter.

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production & Security

The science team has been on something of a hiatus this summer. All of the graduate students are on internships except for two, who graduated. It is expected that activity will pick back up at the end of September when the fall quarter starts. Additionally, August is a lull for the growers as they get ready for harvest in the fall, so that this is their down time. There are results, however, to report:

- The Sedgwick Team is now using soil moisture monitoring to implement their irrigation schedule for the grapes, in some sense adopting the research project as part of their production farming practices. That's good, but it also presents some support problems that Rich Wolski's Aristotle science team are working through.
- In addition, the soil moisture monitoring has exposed a problem with weeds that Sedgwick wasn't aware of. The monitoring infrastructure measures moisture at three depths: 1 foot, 2 feet, and 3 feet. Typically, in a well-maintained field, the moisture forms a gradient (either drier on top and wetter on the bottom or vice versa). However, we are seeing a "moisture inversion" in which the middle layer is driest. We thought it was a bug in the devices that the students built, however it now appears that it is an actual effect. A couple of farm consultants have suggested that it is due to the type of weeds that are present in the field. The Sedgwick farm management team are investigating.
- The team is also conducting a long-term battery-scheduling experiment with the Oaks monitoring project. We have discovered that gypsum block moisture sensors (the cheap ones we are using) are ineffective for this type of monitoring. Gypsum block (the team have used the brand "Watermark") cannot operate when the soil is very dry and in Santa Ynez in the summer, the soil is extremely dry. However, tests are being run to see how long we can run one of our solar powered devices using a scheduling algorithm that turns the device on and off according to how much sunlight it predicts it will receive in the following day. That experiment has been running since June 1st and data from it will be analyzed in the coming months.

The Wolski team is also waiting on a new network switch for the Sedgwick administration building before they can hook up the new batch of camera traps. That problem seems to be with campus IT and not anything related to the science.

A paper on cloud power management was published in IEEE Cloud 2017: Wolski, R. and Brevik, J, QPRED: Using Quantile Predictions to Improve Power Usage for Private Clouds, Proceedings of IEEE Cloud 2017, June 2017. Aristotle is acknowledged in the paper.

5.0 Community Outreach and Education

5.1 Community Outreach

- Co-PI Wolski was a keynote panelist on "Cloud Computing Status and Future" at the 10th IEEE International Conference on Cloud Computing, June 25-30, 2017, Honolulu.

#1541215

- Steven Lee (Cornell) was a speaker on "Campus-based Hybrid Clouds" at the 2017 NSF Cybersecurity Summit for Large Facilities and Cyberinfrastructure, August 15-17, 2017, Arlington, VA.
- Aristotle will be featured at the SC17 Conference, November 12-17, 2017, Denver in the Cornell exhibit. In addition, the Aristotle team at UCSB has had a paper accepted by SC17 on DrAFTs.

5.2 Education

- Six REU students (4 at Cornell and 2 at UCSB) participated in the project over the summer and increased their understanding and hands-on experience with cloud technologies, edge computing the Internet-of-Things, and machine learning, as well as their respective domain sciences. The use case scientists feel that the REU students clearly added value to their research projects.