ARISTOTLE
CLOUD FEDERATION

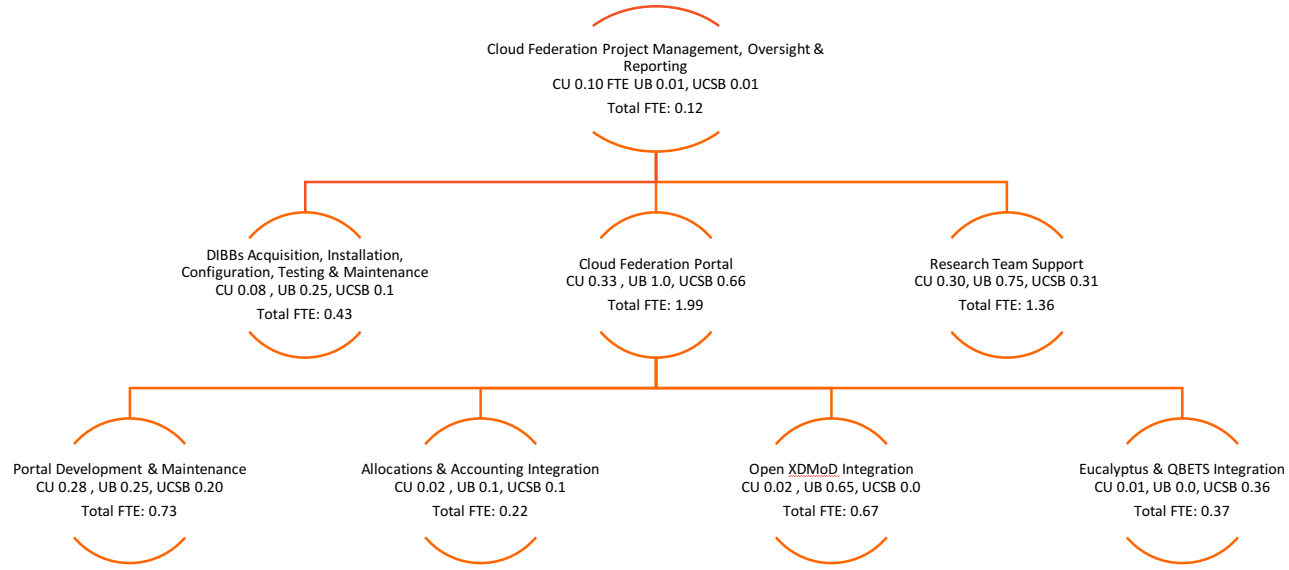
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 3: Quarterly Report 1

12/20/2017

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

This is the Program Year 3: Quarterly Report 1 of the Aristotle Cloud Federation team. We report on plans and activities for each area of the project Work Breakdown Structure (WBS).



#1541215

Contents

NSF
#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
- Google heard about the Aristotle project and invited David Lifka to a Google meeting at EDUCAUSE in Philadelphia. Google is very interested in becoming part of the federation and making their resources available through the Aristotle portal like Amazon. Lifka will explore this interest which could potentially expand the future computational landscape for NSF.
- Lifka had a separate meeting with Amazon at EDUCAUSE which included discussions regarding how to get university research computing directors to understand the value of cloud. The fact that Aristotle is taking a federated approach to cloud is appealing to public cloud providers.
- Lifka also met with Dell executives regarding competitive bidding and bid responsiveness.

1.5 Project Status Calls

10/10/2017 call:

- Cornell held a meeting to discuss their containerization strategy: they are planning to containerize "everything" and are assessing which use cases to containerize first.
- Sara Pryor's (Cornell) use case team submitted a Jetstream allocation request (it has been approved). They plan to run coarsely parallel jobs as soon as the containers are built. UCSB plans to get their K-Means Clustering Service running on Jetstream.
- UCSB is planning a trip to Fresno State to install soil moisture analysis systems at a science preserve. A farm consultant has a lot of data that UCSB would like to crank through Jetstream next quarter.
- The Aristotle infrastructure team had a separate call to ensure that everyone is on the same page regarding their network setup so that workloads will work between sites. Cornell is ordering equipment. UB has gone to Red Hat Enterprise License on the OpenStack side and converted Centos Ceph to Red Hat. They are using Puppet for builds. UB is also combining all their cloud resources into a single cloud (called "Lake Effect") instead of their current 2 cloud setup. Networking is a lot more flexible with OpenStack.
- UCSB is ordering additional equipment. They have 2 classes running in the cloud and will need approximately 200 IP addresses and a lot more network space.
- The Open XDMoD team plans on getting version 7.1 out by the end of October. It will include the Globus authentication piece and some cloud metrics. OpenStack has much better accounting for what the federation needs. UB has done some alpha testing with DrAFTS and knows what needs to be done. UCSB is going to update their DrAFTS website.

#1541215

10/24/2017 call:

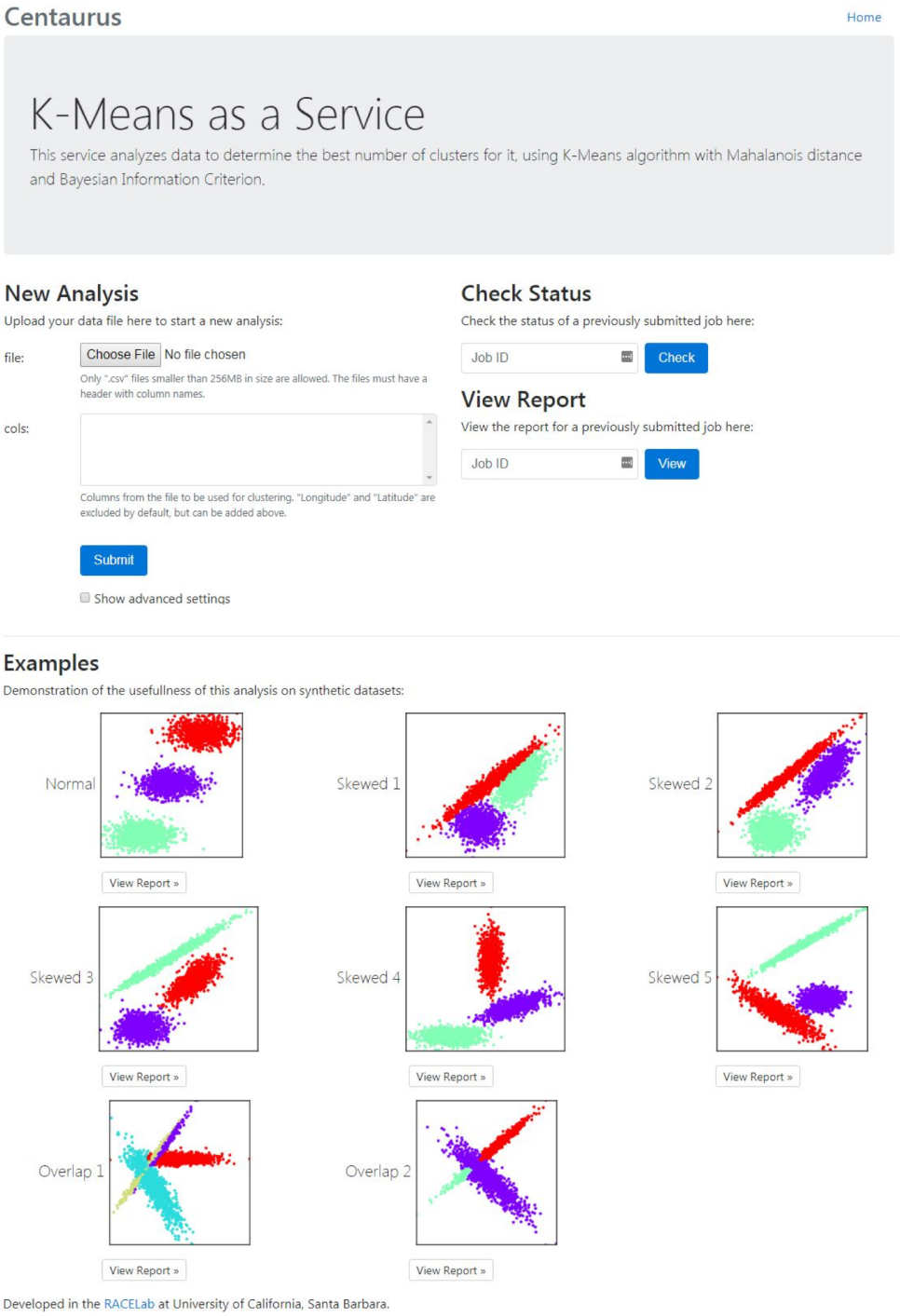
- A separate CU/UB/UCSB infrastructure call discussed OpenStack plans for each site and the best path forward. Steven Lee (Cornell) plans to take a Puppet class. UB is leading the Aristotle OpenStack approach and continuing to test their OpenStack installation. Year 3 purchase plans include UB adding resources to their Ceph cluster and CU adding storage. UCSB plans to expend some funds on management servers and the remainder on compute. UCSB will likely support Eucalyptus a bit longer than the other sites.
- The portal team is waiting for UB to provide the link for webGlobe (the interface UB developed for the visualization and analysis of geo data). In addition, there's plans to add dashboard buttons to the portal that will take users directly to Cornell, UB, and UCSB consoles.
- UB is engaged in getting cloud metrics working and will send a list the metrics that they plan on implementing. UCSB's new DrAFTS site is up, but user interface items need to be completed as well as possible refactoring behind the scenes with containers so that DrAFTS will be easier to deploy and be more scalable. One challenge is if a user wants to run a VM for a month, we can't tell them what to bid because Amazon limits the history you can see. We will try to work through this by using containers. This will require refactoring the guts of DrAFTS so that we can get predictions in seconds rather than minutes.
- UCSB continues to prepare for the deployment of sensors in Fresno. Thirty soil sensors are planned. A student completed end testing of a sensor to the cloud; it was moderately successful. The plan is to analyze the sensor data to determine the optimum water required for almond trees on both the shady and sunny side of the trees.
- Lifka emphasized the need to be in production with OpenStack in 11 months or earlier.
- Wolski said a UCSB startup is taking over Eucalyptus.

11/7/2017 call:

- Current Red Cloud hardware has been upgraded to Eucalyptus 4.4; some hardware will be dedicated to preparing for the transition to OpenStack. Steven Lee completed Puppet training. Puppet is used by UB and will be used by Cornell.
- NSF's NanoGRAV Physics Frontiers Center project (NanoGRAV cyberinfrastructure lead is Aristotle Science Use Case lead Adam Brazier) is sold on containers, but not all open source packages are available (recommend putting them on GitHub for accessibility) and a Docker file broke one piece of software.
- Cornell's new Deputy Director Rich Knepper looked into extending Jetstream cycles for the Aristotle project and an extension request has been submitted by Brazier.
- UCSB did some DrAFTS refactoring planning: a student will be taking a step-by-step approach to refactoring and, once that is completed, the new DrAFTS site will go public (likely next quarter).
- UCSB's almond tree experiment is progressing. A test run was completed successfully at Cal Poly with the edge cloud and the Aristotle cloud, and the field deployment at Fresno is planned for December 1.
- UCSB's Centaurus K-Means as a Service is running on Jetstream (subject to Jetstream outages) and when working is able to process hundreds of thousands of soil electrical conductivity (EC) data sets. To help cover the outage periods for the multi-sourced analytics to improve food production science team, we run a backup service on Aristotle. The team has access to both services and can use either one when an analysis is needed. The Centaurus service itself is still under development, but it is now being used to generate data provided by one of the farm consultants located in the

NSF
#1541215

California Central Valley. When the Centaurus service is more widely available (after authentication is built in), it will enable scientists analyzing data to determine the best number of clusters for it, using K-Means algorithm with Mahalanois distance and Bayesian Information Criterion.



Centaurus K-Means as a Service portal

#1541215

11/21/2017 call:

- PY2 annual report (https://federatedcloud.org/reports/reportPY2AnnualWithAddendum.pdf) was updated with science use case accomplishments. In addition, use case descriptions on the Aristotle portal were expanded to include the latest information on each use case's accomplishments, plans, and products to date (https://federatedcloud.org/science/usecasegeo.php).
- Cornell is planning their OpenStack deployment.
- UB is integrating OAuth with OpenStack for federated login.
- UB is expanding their PY3 hardware vendor list to explore other vendor options.

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

2.1 Hardware Acquisition

- Cornell's PY3 hardware procurement will focus on the servers and networking required for our OpenStack cluster; compute cores will be added as well.
- UB ordered 4 Ceph storage OSDs from Dell. The servers they are currently using for Ceph are re-purposed machines and will be migrated to their Development OpenStack Cloud once the new equipment is installed.
- UCSB is researching PY3 hardware options and preparing for OpenStack deployment using Red Hat's OpenStack standard release (currently version 11).

2.2 Software Installation, Configuration, and Testing

- The Aristotle infrastructure teams held a conference call to discuss hardware deployment and configuration tools. UB uses Puppet and Foreman. UCSB uses Ansible. Both are valid tools. Cornell found the call helpful in gaining insight from UB and UCSB which will help us make an informed decision on tool selection.
- Cornell successfully migrated all users to Eucalyptus 4.4.2 with support for OAuth2 authentication. This migration allowed us to decommission old cluster infrastructure hardware for use as Puppet and Foreman servers. Based on information from UB, Cornell chose Puppet and Foreman for deployment and configuration management. Cornell staff attended Puppet training in early November and can now more easily deploy new hardware and ensure consistent server configurations.
- UB continues to make progress on their OpenStack installation with a Ceph storage backend. They implemented federated logins using OAuth (users can login with OAuth through the web and command line) and users were mapped to the Aristotle project. A problem with VPC networking was discovered; UB will need to engage Red Hat Support to help resolve.

NSF
#1541215

The infrastructure status table was updated this quarter:

	Cornell (CU)	Buffalo (UB)	Santa Barbara (UCSB)
Cloud URL	https://euca44.cac.cornell.edu	https://console.ccr-cbls-2.ccr.buffalo.edu/	https://console.aristotle.ucsb.edu
Cloud Status	Production	Production	Production
Euca Version	4.4.2	4.4.0	4.2.2
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	48 core/119GB 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

- Supercloud - nothing new to report.
- Red Hat is now an Aristotle support partner for OpenStack and Ceph.

#1541215

3.0 Cloud Federation Portal Report

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

We continue to monitor the Aristotle 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 UB 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).
- Check all records to ensure Capacity > 0 (should never be 0).

The portal planning table was updated 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.

#1541215

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 transfer from portal to Euca dashboard.	Login to the portal using InCommon.	Beta testing Euca 4.4 with Euca console supporting Globus Auth. Will deploy and transition to Euca 4.4 on new Ceph-based cloud.	Transition to OpenStack console with Globus Auth login.
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.

#1541215

3.2 Integrating Open XDMoD and DrAFTS into the Portal

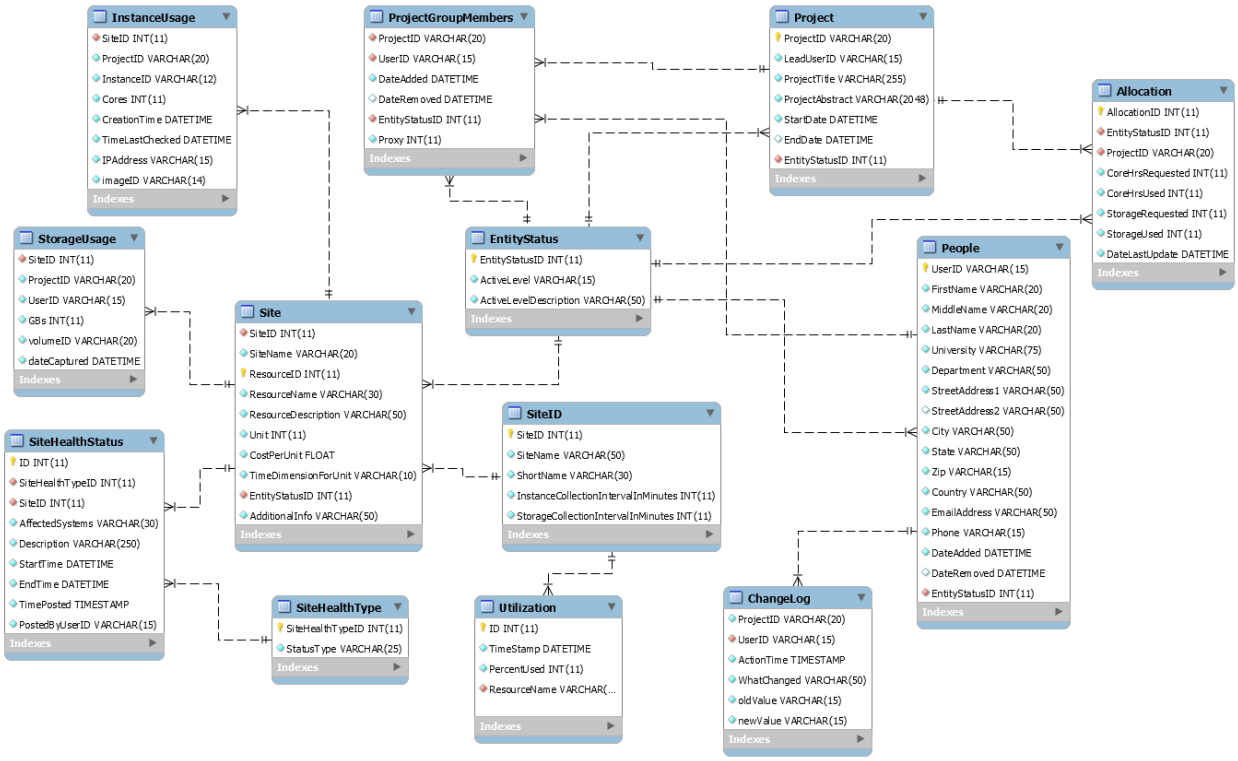
The new DRaFTS service and website (http://predictspotprice-beta.cs.ucsb.edu) is in beta testing and has been released to the Globus Genomics group for testing.

The Aristotle Buffalo team ingested Eucalyptus log data and is continuing to move forward with the creation of a cloud metrics realm based on Eucalyptus data. At the same time, we are working towards obtaining examples of OpenStack log data from Jetstream and our own test OpenStack installation. We integrated Globus Auth into the upcoming XDMoD 7.1 release which will be used to provide Single Sign-On (SSO) access to users across multiple sites and will facilitate access to federated Aristotle data. Changes to the federated XDMoD version are being tested after making the changes necessary to store cloud data in the federation. These changes are currently working for new installations and tests are being run on installations with existing data.

The timeline table for integrating Open XDMoD and DrAFTS into the Aristotle portal has been updated and is now available online: https://docs.google.com/spreadsheets/d/1KlBIWY8ntCC35_5v7o19rro_oOM0Cre8WER-pIlSxMI/edit#gid=76989353.

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

- Cornell professor Sara Pryor's Jetstream allocation request to continue her Aristotle Use Case 3 research was approved.
- Containerization work continues in Use Cases 3, 4, 5, and 6.
- An Aristotle Containerization Working Group (CWG) was formed to (1) evaluate requirements, (2) design an implementation matching those requirements, (3) ensure a consistent environment across the cloud federation. The CWG will be led by Adam Brazier (Cornell). UCSB has joined; UB is identifying their representative(s).
- Bennett Wineholt (Cornell) created a Docker Registry to allow and enable the distribution of Container images across the federation and to other clouds (e.g, Jetstream). Wineholt successfully tested pulling from the Docker Registry on Jetstream and running a Docker capable Ubuntu VM on Jetstream using the OpenStack API. We have the IP ranges for UCSB to allow access to them; IP ranges for UB will be established next.

4.2 Science Use Case Team Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data
The UB research team further refined the integrated visualization and analysis framework called webGlobe. The system is capable of supporting a variety of data formats prevalent in the climate and earth science communities. It is now available to the cloud federation through Globus authentication (https://federatedcloud.org/using/webglobe.php). webGlobe was developed to better understand the interaction between Energy and Water (the energy-water nexus) as part of a DOE grant (https://udi.ornl.gov/research/energy). A joint publication on using machine learning methods for understanding the energy water nexus is under preparation. The experimental results are supported by webGlobe running on the Aristotle cloud. webGlobe details were presented at the American Geophysical Union Fall Meeting (https://agu.confex.com/agu/fm17/meetingapp.cgi/Paper/205848).

Conference presentation given:

April Morton, Jesse Piburn, Robert Stewart, and Varun Chandola (2017). Development of a suite of analytical tools for energy and water infrastructure knowledge discovery. *American Geophysical Union Fall Meeting*, New Orleans, LA.

Use Case 2: Global Market Efficiency Impact
Good progress was made this quarter. Working with Varun Chandola and Jialiang Jiang (UB), this use case is now executing on two VMs; both VMs share the same images that hold all the data and are working well. While the initial project is currently on hold, financial tick data is being used to extend a sample in a different research project; that project plans to submit a paper on their results next quarter. Initial analysis for another paper has been completed, but it was concluded that the sample was not big enough; the VMs are being used to extend that sample and that work will be completed by February. Use case PI Roesch demonstrated the framework to two UB PhD students with whom he launched new projects. Both students will soon be able to analyze the data independently. For the project focused on a policy change in tick size and minimum stock price movement (https://www.sec.gov/oiea/investor-alerts-bulletins/ia_ticksize.html),

NSF
#1541215

all data are collected and the initial analysis is completed, The team is now looking deeper into the effects of this policy change. It is hoped that the first version of a paper will be ready by March. Finally, one other project used the existing data and that team is now computing the variable of interest. By the end of next quarter this computation should completed.

In summary, 4 papers are in preparation:

1. an investigation into how price deviations (market inefficiencies) affect liquidity (the ease at which you can buy or sell)
2. an investigation of how market inefficiencies affect the allocation of capital. We are currently constructing several different stock-day efficiency measures for all US stocks. (joint work with Avanidhar Subrahmanyam and Mathijs Van Dijk)
3. an investigation of how the recent increase in the tick-size, the minimum price movement, of US stocks affects liquidity (joint work with UB PhD student Albert Lee and Kee Chung)
4. an investigation of how much information is incorporated into prices, separately for each US trading venue, and how this affects trading volume at the particular exchange (joint work with UB PhD student Seungoh Han and Brian Wolfe).

Conference presentation given:

Dominik Roesch (2017). Investigating the efficiency of financial stock markets with high frequency data. *2nd Federal Reserve Bank Economic Research in High Performance Computing Environments Workshop*. Kansas City, KS.

Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate

Post Doc Associate Tristan Sheperd and Cornell professor/Aristotle use case lead Sara Pryor report that the précis objectives of their current suite of simulations are:

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 near-surface wind regimes from simulations conducted on conventional HPC and the 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).

We are addressing these objectives by conducting and analyzing the output from high-resolution numerical simulations with the Weather and Research Forecasting model (WRF, v3.8.1).

Our activities this quarter have focused on two key aspects:

1. Continuation of the numerical simulations with the Weather Research and Forecasting model (WRF) – see status below.
2. Analysis of WRF model output in the context of our 5 research objectives (listed above). This analysis component 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. Highlighted in orange are simulations that have been completed and/or begun this quarter.

Simulation (Platform)	Domains	Wind turbine installed capacity	Status	Purpose
2008 eastern USA no WT (Cori)	12 km, 4 km	None	Completed	1,2,5
2008 eastern USA WT (Cori)	12 km, 4 km	Current	Completed	5
2008 eastern USA (Cori)	12 km	None	Completed	1,2,4,5
2008 eastern USA (Aristotle)	12 km	None	Completed	1,2,5
2008 Iowa WT (Aristotle)	12 km, 4 km	Current	Completed	5
2008 Iowa no_WT (Aristotle)	12 km, 4 km	None	Completed	1,2,5
2008 eastern USA double WT capacity [2WT] (Cori)	12 km, 4 km	Doubled	Completed	1,5
2008 eastern USA quadruple WT capacity [4WT] (Cori)	12 km, 4 km	Quadrupled	Completed	1,5
2008 Iowa WT (Aristotle)	12 km, 4 km	Current. Different parameterization of the action of WT on the flow field	Pending	5
2015 eastern USA no WT (Cori)	12 km, 4 km	None	Completed	1,2
2015 eastern USA WT (Cori)	12 km, 4 km	Current	Completed	5
2015 eastern USA quadruple WT capacity [4WT] (Cori)	12 km, 4 km	Quadrupled	Running	5
Contemporary climate decade eastern USA (Aristotle)	12 km	None	2001-2010 Completed,2011-2016 Running	2,3,4
Future climate decade	12 km	None	Pending	2,3,4

Example results based on analyses of output from simulations with and without the influence of wind turbines (objective 5 above)

We have completed analyses of our high-resolution numerical simulations of the effects of wind turbines (WT) on regional climate for a domain centered on the highest density of current wind turbine deployments in the contiguous US (Iowa) and submitted a journal manuscript describing our research. Our analyses indicate that impacts from WT on near-surface climate variables such as; temperature, specific humidity, precipitation, sensible and latent heat fluxes are statistically significant only in summer, are of very small magnitude and are highly localized (see examples in Fig. 1 below). Our finding that the WT impact is maximized in summer is due to the lower wind speeds during this season and is consistent with WT aerodynamics. This finding may indicate that previous studies that have considered shorter duration summertime simulations from very high-density WT arrays in the US Central Plains have overstated the climate impacts of realistic WT deployments. Our initial research thus implies that further expansion of wind turbine deployments can likely be realized without leading to substantial downstream impacts on weather and climate.

In the coming quarter (Jan-Mar 2018) we hope to examine the robustness of our inferences using a different parameterization of the WT rotor aerodynamics (pending establishment of a new Docker on Aristotle and successful compilation of the modified WRF code) and by evaluating output of our simulations with WT deployments over the entire eastern USA (simulations complete, analysis commencing largely on Jetstream).

#1541215

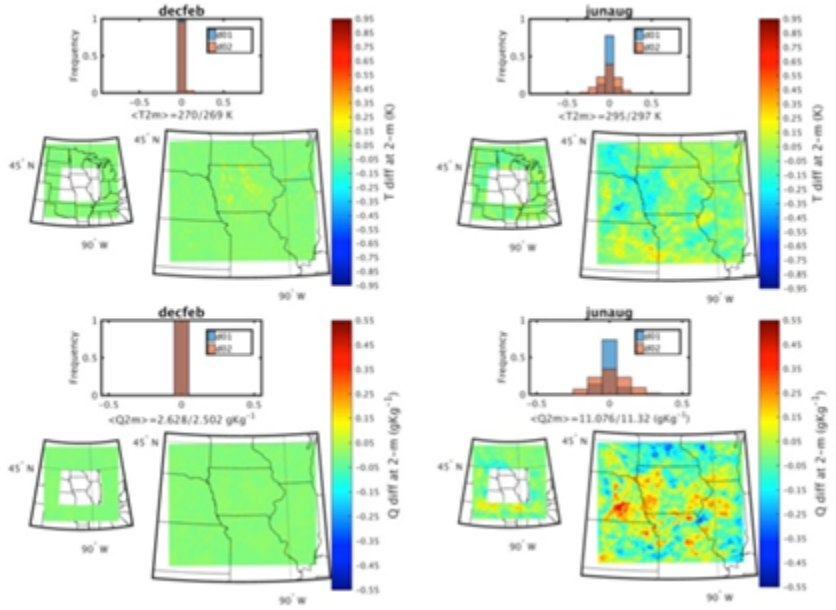


Fig. 1. Maps and histograms of the mean difference in each grid cell of (above) air temperature at 2-m (T2M) and (below) specific humidity at 2-m (Q2M) in simulations with WT minus without WT (i.e. WT1 – noWT) during winter (left) and summer (right). The values below the histograms denote the seasonal mean T2M and Q2M in d01 and d02 in the noWT simulation. Figure adapted from those in Pryor et al. (2017), in review.

Journal manuscripts:

Pryor S.C., Barthelmie R.J. and Shepherd T. (2017): The influence of real-world wind turbine deployments on regional climate. *Journal of Geophysical Research: Atmospheres* (2017JD028114), in review.

Conference presentations given/abstracts submitted:

Pryor S.C., Barthelmie R.J., Hahmann A., Shepherd T., and Volker P. (2018): Contemporary wind turbine deployments have a minor impact on regional climate. *Science of Making Torque from Wind*, Milan Italy, May 2018 *(abstract 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 accepted).*

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 accepted).*

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 *(presented).*

#1541215

Use Case 4: Transient Detection in Radio Astronomy Search Data
We moved software for this use case to a new containerized solution. It can now be deployed on a variety of workstations, i.e., it is a reproducible environment that can be spun up anywhere. The container includes a full suite of all relevant pulsar-processing software, and Python 2 and Python 3 Conda distributions. It executes modification and recompilation of the core PRESTO pulsar search package to perform the transient search using the Spitler modulation index method. We are currently adding the decimation code. Use case PI James Cordes, Aristotle science lead Adam Brazier, visiting PhD student Shen Wang, and former REU students Elizabeth Holznecht and Shiva Lackshmanan are all at work developing on this platform. Our near-term aim is to (1) deploy this container to Jetstream, (2) test for OpenStack compliance, (3) initiate large-scale decimation runs.

Use Case 5: Water Resource Management Using OpenMORDM
Brandon Barker and Bennett Wineholt (Cornell), working with Julianne Quinn and Bernado Trinivale, ran an OpenMORDM borg example `./borg.exe -n 100 -v 11 -o 2 dtlz2_embedded.exe` on a single medium size Jetstream VM, and an OpenMORDM MPI example `mpirun dtlz2_ms.exe` on Jetstream. While both of these were single VM runs, past work demonstrated related runs on Docker and between multiple VMs (not on Jetstream), as well as MPI running between multiple Docker containers on Red Cloud (Aristotle). Based on these initial successes, our short-term goals are to (1) run Julianne's OpenMORDM code across many cores on many Jetstream VMs utilizing Docker containers and a scripted cluster setup, (2) run Bernardo's MPI simulation code that depends on significant input data in a similar scenario utilizing IU Jetstream Wrangler storage.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota
The Angela Douglas Aristotle research team has completed manual curation of the first bacterial model, Lactobacillus plantarum DmCS001. In addition to completing this model, we generated two additional bacterial models that are arginine and histidine auxotrophic mutants of the wild type L. plantarum model to use as a primer to develop and extend the necessary computational framework (SteadyCom) to model multi-species metabolic interaction in the insect gut. We are also at the terminal stages of manual curation of a second gut bacterium Acetobacter pomorum DmCS004.

Manual curation of this model will be completed by February 2018. We have set up a simulation environment and successfully performed SteadyCom on test models in Red Cloud (Aristotle) using a newly configured Windows VM with MATLAB and the COBRA Toolbox. In an effort to containerize the simulation environment and also make use of the larger Jetstream resource at Indiana, we built a Linux VM on Jetstream with Docker and MATLAB, and worked with Gurobi to obtain the necessary Docker-friendly licenses for their commercial solvers.

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production & Security
Where's the Bear: UCSB Aristotle researchers visited the Sedgwick Reserve in December to collect the latest set of camera trap images (~150GB this quarter). The IT staff installed long-distance radio networks to various sites around the reserve. The next step is to attempt to automate the image capture. It is still a manual process to download the images from the cameras, but with the networking (which has not yet come out of testing), it will be possible to go from camera to the Aristotle cloud automatically. The hope is to begin testing early next year.

NSF
#1541215

Food Security: UCSB Aristotle researchers visited a test orchard at Fresno State in December (see photos: https://photos.app.goo.gl/RZGikmiJsMId1A4q1). The goal of this SmartFarm deployment is to study differential irrigation (both sunny side and shady side) for almonds trees. In the photos, Nevena Golubovic (UCSB PhD student studying IoT for Agriculture) works with Agricultural Science graduate students from Cal Poly San Luis Obispo and Fresno State, as well as faculty researchers and farm consultants to instrument a test tree. If successful, the plan is to instrument 30 trees within the test orchard over a full growing season to determine how much water can be saved by irrigating the different sides of the root stock in proportion to its dryness.

5.0 Community Outreach and Education

5.1 Community Outreach

- Aristotle PI Lifka accepted an invitation from NSF to join the Directorate for Computer & Information Science & Engineering (CISE) Advisory Committee.
- Cornell CAC Deputy Director Rich Knepper served on a "Cloud Day: Current Use Cases at Academic Institutions" panel moderated by Ruth Marinshaw of Stanford at the CASC Fall 2017 Meeting.
- Cornell featured the Aristotle Cloud Federation at the November SC17 Conference in Denver in presentations, graphics, and literature. Co-PI UB featured Aristotle as well and our new partnership with Red Hat (https://federatedcloud.org/papers/SC17flyer.pdf).
- Aristotle Infrastructure lead Resa Reynolds (Cornell) led the Dell XL meeting at SC17. Reynolds is President of Dell XL (https://www.dellhpc.org/dellxl.html). She is also a member of the Board of Directors for the Dell EMC HPC Community.
- Aristotle co-PI Rich Wolski gave a presentation on DrAFTS (which was developed under Aristotle funding) at SC17 and a corresponding publication was accepted by the Conference Committee (http://www.cs.ucsb.edu/~rich/publications/drafts-sc17.pdf).

  Publication:

  Rich Wolski, John Brevik, Ryan Chard, and Kyle Chard. 2017. Probabilistic Guarantees of Execution Duration for Amazon Spot Instances. In *Proceedings of SC17, Denver, CO, USA, November 12–17, 2017*, 11 pages. DOI: 10.1145/3126908.3126953

  Abstract:

  We propose DrAFTS - a methodology for implementing probabilistic guarantees of instance reliability in the Amazon Spot tier. Amazon offers "unreliable" virtual machine instances (ones that may be terminated at any time) at a potentially large discount relative to "reliable" On-demand and Reserved instances. Our method predicts the "bid values" that users can specify to provision Spot instances which ensure at least a fixed duration of execution with a given probability. We illustrate the method and test its validity using Spot pricing data post facto, both randomly and using real-world workload traces. We also test the efficacy of the method experimentally by using it to launch Spot instances and then observing the instance termination rate. Our results indicate that it is possible to obtain the same level of reliability from unreliable instances that the Amazon service level agreement guarantees for reliable instances with a greatly reduced cost.

5.2 Education

- Steven Lantz (Cornell) led a hands-on workshop on parallel computing with MATLAB and scaling to Red Cloud (see description at https://www.cac.cornell.edu/about/news/170927.aspx). In addition, the workshop—in collaboration with MathWorks—featured a seminar on parallelizing MATLAB without CUDA or MPI programming and a seminar on machine learning. ~30 graduate student and faculty researchers participated.