ARISTOTLE
CLOUD FEDERATION

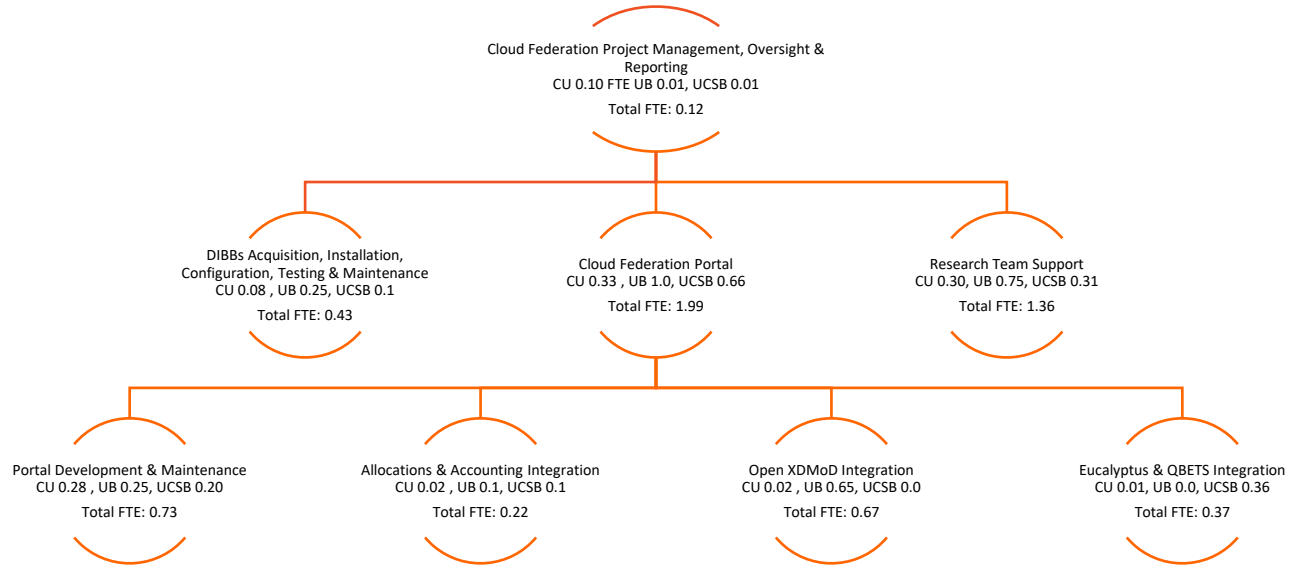
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 3: Quarterly Report 2

3/28/2018

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

This is the Program Year 3: Quarterly Report 2 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
- Lifka continued strategic discussions with major public cloud providers.

1.5 Project Status Calls

1/16/2018 status call:

- UB is working federated cloud features into Open XDMoD, including some beta cloud metrics such as no VM started, VM wall clock time, etc. More complex cloud metrics will follow. Open XDMoD version 7.5 is scheduled for a July release.
- UCSB continues development of the new DrAFTS website which is now in beta and gathering user feedback. Early beta testers include Globus Genomics at Argonne and New Zealand who are providing feedback on how the API is working. Globus Genomics is interested in doing a collaborative pricing study on the AWS spot market.
- The CENTAURUS clustering system works great on Aristotle but is not scaling well on Jetstream. We are analyzing whether the issue is Jetstream stability and whether we need to re-engineer CENTAURUS to be Jetstream-proofed.
- We are making progress on new food production use case projects. Our soil moisture monitoring of California almond trees project features a multi-scale IoT infrastructure that uses solar panel powered sensors to measure soil moisture and temperatures at different depths. Our citrus orchard project is deploying radio-based networking throughout the orchards to take weather measurements and predict temperatures. Power is the challenge rather than the Aristotle cloud where the analytics run; a big solar array is needed. The analytics result in spatial cluster in soil and differential meteorology for frost prediction.
- UCSB has transitioned its original Aristotle infrastructure to OpenStack; it is currently in beta.
- Once Cornell's PY3 hardware arrives, we will get our OpenStack beta up and running for user testing.

1/30/2018 status call:

- We will continue work on the Aristotle dashboard once it gets moved over to the new database.
- Testing of the OpenStack version of Open XDMoD v. 7.5 is well underway and the ability to federate feature will be completed in several weeks and then be distributed to federation sites for

#1541215

beta testing. Cloud metric features are under development; the goal is to integrate the initial cloud metrics features by July, and work on the development of more complex metrics.

- Cornell's PY3 servers arrived in late March. We will rack and stack, build our OpenStack infrastructure, and start staff testing in the next 4 weeks. Our staff will travel to UB or use concalls for help as needed since UB was the first federation site to deploy OpenStack and is now working with friendly users. The ability for federation members to quickly learn from one another rather than going it alone is a major benefit of participating in a federation.
- UCSB is currently testing the free version of Tesora's Percona database as a way to handle all databases within OpenStack. Percona is a cluster environment with a SQL-like model for replication and failover. UCSB is also completing their order for PY3 hardware. They are also following up with Jetstream regarding their Ceph outage and other issues.
- UB big geo data use case scientist Varun Chandola is preparing a paper on machine learning techniques.
- Cornell's Sara Pryor will complete a long simulation in April and is also leveraging Jetstream for MATLAB work. Jetstream appears to be an ideal resource for MATLAB work.
- The radio astronomy project is working to containerize their use case; stability of individual software components is a challenge.
- Patrick Reed's water resources management student is investigating Jetstream performance vs. a dedicated cluster and is looking at cross instance MPI performance. CAC consultant Bennett believes that if he is saturating a very large number of cores, the physical resources may be constrained. Netflix, for example, starts 15 AWS instances, then cancels 5 to be sure they get what they need.
- Cornell and UCSB use case scientists were very pleased with the productivity of REU students last summer and appreciated the support provided by the NSF. They would like us to submit a REU supplemental for summer 2018.

2/13/2018 status call:

- Cornell is building their OpenStack platform, but has had to order 5 more servers to add capacity because their cloud is running at 90% utilization. They will deploy an OpenStack development environment so they do not interfere with production use of Aristotle. UB has also been running at about 90% and had to ask some users to stop instances and turned on hyperthreading for more space. A few Dell CPUs are intermittently going in and out; this may be a legacy bios problem.
- Cornell is making database improvements to handle Aristotle federation requirements. Emails of each user will be kept in a federated cloud database. Project, allocation information, etc. will be added to the dashboard interface and implemented according to plan.
- Open XDMoD will be installed at all 3 sites so that we can test out the federated capability in advance to adding all the cloud metrics. The beta version should be completed by the end of March with just HPC data so that all 3 sites can get data flowing to the federated master at UB. It can be installed on any machine that federation sites have available. Open XDMoD will run on a VM and the database can be on the same VM. We will need to decide what we want to show in the federated data, e.g., fields of science, etc.
- We defined Aristotle's 36-month review critical path stacks. Goals include federation and migration (workflows of containers or VMs) between sites with initial dashboard and Open XDMoD functionality.
- Cornell's radio astronomy use case (Jim Cordes) is updating their pipeline prior to initial testing of containers.

#1541215

- Use case scientist Sara Pryor's 17-year simulations ran successfully in 4-5 day increments.

2/27/2018 status call:

- Networking is operational for the IoT/edge cloud/Aristotle cloud citrus orchard project. It is now taking temperature data at 2 elevations at 1 location. This will be expanded to 12 orchard locations. The goal is automated frost prevention.
- The "Where's the Bear?" automatic wildlife identification project is working on adding a wildlife accounting capability to the system.
- UCSB submitted 2 papers: one on how clustering works on soil moisture measurements (solved on the Aristotle and Jetstream clouds) and a second paper on how Aristotle is used for sensor and cloud applications (submitted to IEEE Cloud Conference).
- Dashboard tools need to be defined to reflect member status, e.g., proxy tools, project member tools such a view my usage, etc.). We need to complete the OpenStack installations at all 3 sites before we can implement these.
- UCSB is working on stabilizing the new DrAFTS site. It is written in Python and uses Docker.
- Discussions occurred regarding using a lease concept rather than rates to track usage between sites. If you want capacity at another site for a large-scale experiment, you should be able to find a lease for a long-fixed period at another site. If Cornell is leasing so much per year from UCSB, then UCSB should be entitled to that much back. We believe that the economic model we want to pursue is leases. If we do go that way, we would start monitoring who is using what where, analyze the potential of use of leases, and then develop a programmatic interface for that. We would track usage across the federation and even things up at the end of the month. We would learn how to create equity and, from that, how to properly word a lease from a business perspective.

3/13/2018 status call:

- Cornell will be updating the Aristotle database with the latest functionality.
- Cornell is also standing up OpenStack and will be moving nodes so that by mid-April we should have our test environment completely set up.
- UB is also standing up an OpenStack development cloud so that they can perform their most intrusive testing without interrupting their production OpenStack cloud. They are troubleshooting performance issues that appear to be network latency issues rather than an issue with Ceph. And, their finance use case researchers continue to crunch data. UB's CSE611 students produced a backend for the data so that researchers have a dashboard to use and can keep the finance data up-to-date more easily.
- The Aristotle project will be requesting a 6-month extension to our Jetstream allocation.

3/27/2018 status call:

- Cornell continues to work on the portal dashboard.
- UB is adding initial cloud metrics to Open XDMoD. The beta RPM for Open XDMoD 8.0 will be ready for availability in early April. Sites will download it and then start to get data flowing in a federated mode. It will run at individual sites while the master runs at UB. Documentation will be available. For purpose of the portal, we plan to show overall usage and also usage by PI science teams. On our next status call, we will discuss what needs to happen next. Installation of federated Open XDMoD should be pretty simple and we don't anticipate too much trouble.

#1541215

- Cornell and UCSB are busy standing up their OpenStack clouds.
- Cornell and UB had some good collaboration conversations over the past couple of weeks regarding Ceph performance. UB is all set; they will be putting in new SSD in late March and add them into Ceph.
- The Aristotle Jetstream allocation has been extended for another 6 months.
- Steven Lee and Bennett Wineholt met with Tristan Shepherd of Sara Pryor's use case team and fixed a NERSC Cori Cray XC40 to Red Cloud data performance problem. The data transfer speed is very good now.

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

2.1 Hardware Acquisition

- All 3 sites installed their PY3 hardware. Cornell installed infrastructure servers and a network switch for their Red Cloud OpenStack cloud; UB installed a Ceph storage cluster for their Lake Effect OpenStack cloud; and, UCSB installed OpenStack node controllers and network switches.

2.2 Installation, Configuration, and Testing

- Cornell installed and tested a Foreman/Puppet system to automate all future Aristotle server deployments (including OpenStack). We also began installing and configuring Red Cloud OpenStack controllers.
- UB relocated hardware from the Lake Effect Eucalyptus cloud to the Lake Effect OpenStack cloud and installed and configured OpenStack. We implemented auto provisioning for OpenStack and Ceph using Foreman and Puppet allowing us to re-provision our cloud in less than 20 minutes. During testing, we noticed the performance of the Ceph cluster was not meeting expectations. After discussions with the Cornell and UCSB teams, we decided that SSD drives are necessary journal writes. The SSD drives were ordered and will be added to the Ceph cluster, and the OSDs will be reconfigured to maximize performance. Finally, we are standing up a development OpenStack cloud to use for test and development once the Lake Effect OpenStack cloud officially goes into production. We on-boarded a few friendly users to help us further test the OpenStack cloud. These users include XDMoD developers who will be working on the monitoring piece. A web service was created to handle the reporting of OpenStack usage.
- UCSB prepared for OpenStack deployment using Red Hat's OpenStack standard release (version 11). We continue to work on improving redundancy of Ceph cluster connectivity and configured the networks. Finally, we tested OpenStack-Ansible automated installations for OpenStack deployment.

The infrastructure table below was updated to reflect the current status of each site's production Eucalyptus cloud. All three sites are in the process of standing up OpenStack environments. All of the hardware will be transitioned from Eucalyptus to OpenStack.

This table will be updated next quarter to reflect OpenStack status at each site.

#1541215

	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	4.3.1	4.2.2
Hardware Vendors	Dell	Dell , Ace	Dell, HPE
DIBBs Purchased Cores	*168	**256	356***
RAM/Core	8GB	up to 8GB	9GB Dell, 10GB HPE
Storage	Ceph (1152TB)	SAN (336TB) Ceph (384TB)	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	Yes	Planned	Planned
	*168 cores added to Cornell Red Cloud (478 total cores)	**256 cores added to Buffalo Lake Effect Cloud (424 total cores)	***356 cores in UCSB Aristotle cloud (572 total cores, Aristotle is separate from UCSB campus cloud)

2.3 Potential Tools

- Supercloud - nothing new to report.
- Red Hat OpenStack – we are currently transitioning from Eucalyptus to Red Hat OpenStack.

3.0 Cloud Federation Portal Report

Content updates to the project are ongoing (https://federatedcloud.org). Updates were made to many portal branches this quarter, including publications, partners, and news and events,

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 continue to implement software to verify that the data ingestion API is running. Nagios server at Cornell is now monitoring the usage report API at all 3 sites in the Aristotle federation. When the usage report API at a site becomes unreachable by the Aristotle portal, Nagios will alert the infrastructure team at Cornell to take appropriate corrective action.

#1541215

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 were no changes 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.

#1541215

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

The MariaDB database was upgrade from version 5.5.6 to 10.2.13. This version provides more functionality for generating reports.

3.2 Integrating Open XDMoD and DrAFTS into the Portal

The DrAFTS team is testing their methodology against Amazon's new spot pricing policy which "smooths" price changes. In late January, Amazon introduce a new method for computing spot prices. While the exact algorithm remains hidden, the effect is to cause spot prices to remain stable over long periods of time. We are currently testing DrAFTS to determine the effect on DrAFTS bids. Under the old regime, DrAFTS could provide guarantees for up to 48 hours of durability. We suspect that under this new regime, the durability guarantees will be significantly longer, however we must allow long periods of time to elapse before we can validate this conjecture. This effort is taking place now; we are running jobs in the AWS spot market using DrAFTS bids to determine durability.

The UB team released Open XDMoD 7.5 which includes Single Sign-On (SSO) integrated with Globus to allow any user who has a Globus identity to use that identity to log into XDMoD. Work is nearing completion to provide a beta version of federated XDMoD which will support the collection and aggregation of data from individually managed centers into a single federated instance of XDMoD in order

#1541215

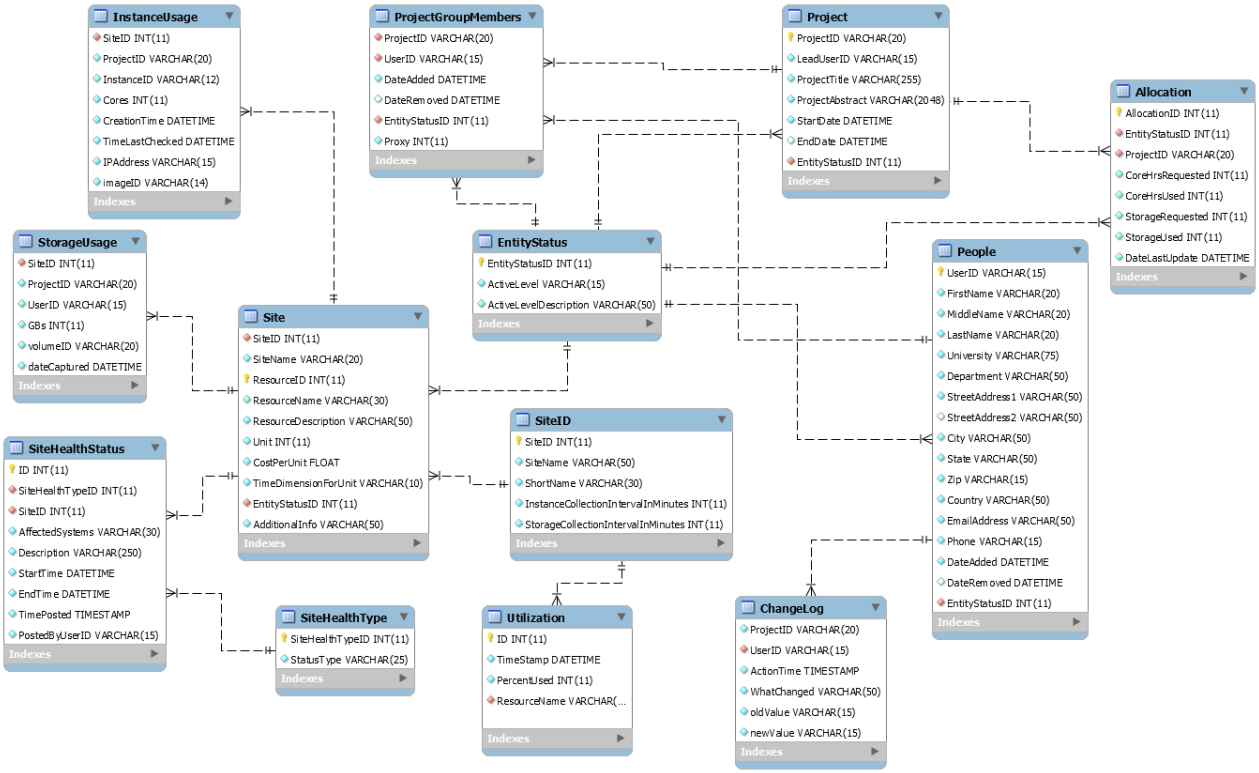
to display federation-wide metrics. We aggregated nearly 2 years of Eucalyptus cloud data to use in creating the Cloud Metrics realm. In addition, we began to process OpenStack log data from both our own test OpenStack installation and data from the Jetstream Atmosphere team. This data will augment the existing Eucalyptus data and be used as a proof-of-concept for simultaneously reporting on data from 3 different cloud providers and data formats. We plan to include the following initial cloud metrics:

- Number of Instances: Running
- Number of Instances: Started
- Number of Instances: Ended
- Number of Cores: Utilized
- CPU Hours: Per Instance
- CPU Hours: Total
- Wallhours: Per Instance
- Wallhours: Total

The XDMoD timeline has been updated and is available online: https://docs.google.com/spreadsheets/d/1KlBIWY8ntCC35_5v7o19rro_oOM0Cre8WER-pIlSxMI/edit?usp=sharing

3.3 Allocations & Accounting

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



#1541215

4.0 Research Team Support

4.1 Science Use Case Team Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data
We continued development of an Energy-Water Knowledge Discovery Framework portal using the webGlobe technology currently running on the Aristotle cloud. An extended abstract on that work was accepted by the University Consortium for Geographic Information Science 2018 Symposium (May presentation):

Berres, A., Karthik, R., Sorokine, A., Nugent, P., Allen, M., McManamay, R., Chandola, V., Zaidi, A. & Sanyal, J. (2018). *EWN-KDF: A Knowledge Discovery Framework to Understand the Energy Water Nexus*, University Consortium for Geographic Science 2018 Symposium and CaGIS AutoCaro, Madison, WI.

Use Case 2: Global Market Efficiency Impact
Several months ago, we started working with a PhD student from Germany and in April we will start crunching data hosted in Aristotle. Also, in February, we started a project as part of a course (CSE611) taught by Alan Hunt. A goal in that course is to program a backend that keeps the financial research data in Aristotle up-to-date and to create a website to serve as a dashboard. The dashboard will give researchers convenient access to the latest financial data and enable us to showcase the research we can do with the data and software available on Aristotle.

Papers status:

1. An investigation into how price deviations (market inefficiencies) affect liquidity (the ease at which you can buy or sell) > we are revising this paper based on comments.
2. An investigation of how the recent increase in tick-size, the minimum price movement, of US stocks affect liquidity (joint work with UB PhD student Albert Lee and Kee Chung) > variable constructions are completed and we are currently writing the first draft; the first version will be completed in April.
3. An investigation of how market inefficiencies affect the allocation of capital requires the construction of several different stock-day efficiency measures for all US stocks (joint work with Avanidhar Subrahmanyam and Mathijs Van Dijk) > we finished constructing our variables and started looking at results.
4. An investigation of how much information is incorporated into prices, separately for each US trading venue, and how this affects trading volume at a particular exchange (joint work with UB PhD student Seungoh Han and Brian Wolfe) > this research is currently on hold.

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 C. 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).

#1541215

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.

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	Completed, additional runs started to test sensitivity to LBC	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	Completed	5
Contemporary climate decade eastern USA (Aristotle)	12 km	None	Completed	2,3,4
2015 eastern USA no WT (Cori) GCM driven	12 km, 4 km	None	Running	1,2
2015 eastern USA WT (Cori) GCM driven	12 km, 4 km	Current	Running	5
2015 eastern USA quadruple WT capacity [4WT] (Cori) GCM driven	12 km, 4 km	Quadrupled	Running	5
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)

#1541215

We have completed simulations to test the sensitivity of the climate impacts to the precise description of the wind turbine (WT) aerodynamics (the extraction of momentum and introduction of turbulence (TKE) behind the turbine rotor). We conducted two sets of paired simulations for the nested domain shown in Fig. 1. The simulations are:

(a) A simulation without any wind turbines ('noWT1'), followed by an identical simulation using the 'Fitch' wind farm parameterization (defined as 'WT1').
(b) A simulation without any wind turbines ('noWT2'), followed by an identical simulation using the Explicit Wake Parameterization, EWP (defined as 'WT2').

The key difference between these two parameterizations is that while the Fitch scheme applies a (local) drag force and additional TKE to all model grid cells that intersect the turbine rotor, the EWP scheme parameterizes the unresolved wake expansion within the grid-cell and applies a grid-cell averaged drag force. Additional TKE results solely from enhanced vertical shear due to WT wake(s). In each pair of WRF simulations, all other settings are unchanged.

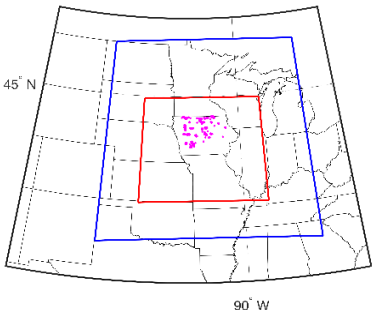


Fig. 1: Map of the central/eastern USA, with the simulation domains (d01 outlined in blue has a resolution of 12 km and d02 in red has 4 km resolution) and the locations of WT in Iowa as of the end of 2014 (magenta dots). The grey lines denote state boundaries. d01 comprises 320 by 320 grid cells, while d02 comprises 246 by 204 grid cells, 299 of which have WT within them. There are 41-layers in the vertical.

An example of our results is shown in Figure 2, and the results are discussed in detail in Pryor et al. (2018, in review). On average spatial fields of the pairwise differences (WTX minus noWTX) indicate the perturbation to the wind speed field due to the presence of WT is generally more marked in the simulations using the Fitch parameterization than when EWP is applied (Figure 2). Accordingly, the WT impacts on near-surface air temperature and specific humidity (along with the other parameters evaluated) are also typically more modest when the EWP parameterization is applied. Thus our previous inference (described in Pryor et al, 2017, in review) of only very modest climate impacts from WT is strongly supported by our most recent results (presented in Pryor et al. 2018).

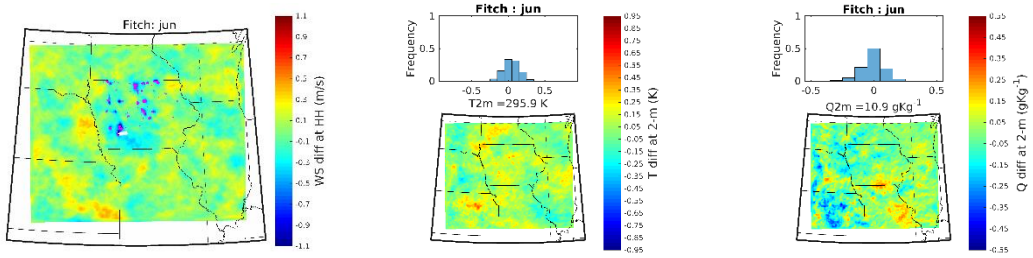
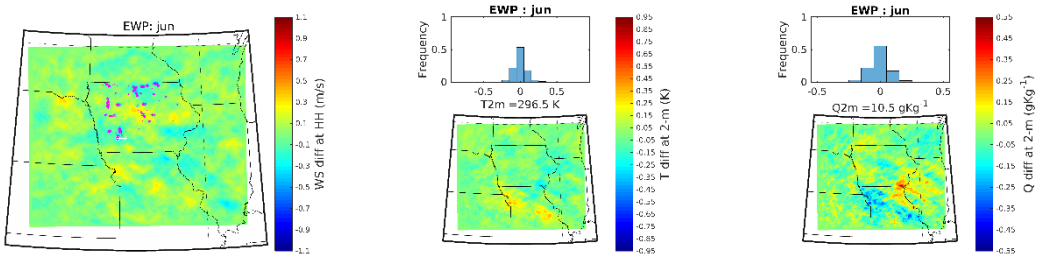


Fig. 2: Above: Mean impact of WT on wind speed in the third model layer (WTHH, left), near-surface (2-m) air temperature (center) and near-surface (2-m) specific humidity (right) from WT1 minus noWT1 simulations for; June. Values below the histograms indicate mean d02 values from the noWT1 simulation. Below: For WT2 minus noWT2.

#1541215



Activities planned for next 3 months (PY3Q3):
It is our intention that work over the coming 3 months will focus on analyses of our long-term (nearly 2 decades long) simulation to address Question 4 but we are experiencing extreme difficulties with data transfer speeds to the large (100TB) disk mounted to a new instance created for this purpose. Thus, while we wait for that transfer to complete we will likely undertake an additional simulation for the domain shown in Figure 2 to test the sensitivity of the regional/local climate impacts from operating WT to exact way we initialize the WRF model (referred to in the table above as LBC). We also plan to make major progress on the "upscaling" of our results regarding the impact of WT on regional climate using our simulations for the entire eastern USA.

Journal manuscripts:
Pryor, S.C., Barthelmie, R.J., Hahmann, A., Shepherd, T. & Volker, P. (2018). Contemporary wind turbine deployments have a minor impact on regional climate. *Journal of Physics: Conference Series*, in review.

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

Conference presentations abstracts submitted:
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*? European Geosciences Union General Assembly 2018, Vienna, Austria. *(Invited presentation, April 2018).*

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

Use Case 4: Transient Detection in Radio Astronomy Search Data
Following an all-hands meeting, it was decided to move away from using the PALFA pipeline, which primarily runs the PRESTO package of Scott Ransom, as used by Laura Spitler for several reasons: (1) the pipeline, although under active development, does a lot more than we need, so that keeping up with changes would require fairly substantial effort, (2) the functionality used by Laura Spitler is now substantially changed in the new pipeline, including revisions to the underlying PRESTO code. PRESTO retains and continues to improve its functionality for computational efficient operations on the radio astronomy data, so the project will utilize that, via its Python bindings, on a container which has the newest version of PRESTO installed. A new pipeline design has been created and responsibilities for various code areas given to students Elizabeth Holzknecht, Shiva Lakshamanan, and Shen Wang and to CAC consultant Peter Vaillencourt. The new pipeline is extensible and will allow for modular changes via a configuration file parsed by the master pipeline script. Delivery of an early prototype is expected in April.

#1541215

Use Case 5: Water Resource Management Using OpenMORDM
Work continues on Jetstream deployment. Bernado Trinvale has finished his MPI/OpenMP water utility planning and management optimization model and will now run it to try to optimize a test case in North Carolina on different computing platforms and compare their performance. This will help promote the use of cloud computing among researchers and practitioners in the field of water resources systems. Julianne Quinn ran benchmarks on Jetstream for her model; initial performance has been slower than expected and we are investigating the cause (contention or an artifact of hyperthreading are under consideration). Reported results for comparison of Jetstream and a local cluster, TheCUBE, are below, with comparisons from runs on XSEDE resources Stampede2 and Comet. We ran 1 trial run for a 1-island parallel configuration of Borg multiobjective search algorithm for 1000 function evaluations of the Red River, Vietnam food-energy-water case study. Here, are the timings on Jetstream vs. TheCUBE:

Number of Cores	Time on Jetstream	Time on TheCUBE
2 cores	40:15:24	18:05:44
4 cores	13:49:48	06:17:57
6 cores	08:41:21	03:54:17

For perspective, on Stampede2, we ran 1 trial of a 16-island configuration with 400,000 function evaluations per island. That ran on 136 nodes using all 68 cores/node and it took 46:43:16. On Comet, we ran 1 seed of an 8-island configuration with 1M function evaluations per island. That ran on 72 nodes using all 24 cores/node and it took 43:44:23.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota
Our main focus this quarter has been to optimize the use of the computational framework SteadyCom to analyze the metabolic models of the gut microbial communities in Drosophila. In particular, we have debugged a few issues in our model and configuration, and we can now run SteadyCom on replicated Lactobacillus plantarum models using the latest code from the COBRA Toolbox for MATLAB. We now have the COBRA Toolbox running on both Red Cloud and Jetstream systems. With these key capabilities in place, we have started investigating parameters to SteadyCom that may be of use in fixing community composition, which is normally allowed to vary in SteadyCom.

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production & Security
Where's the Bear (WTB): We have a preliminary version of WTB that does image counting. The goal is to use it to try and make estimates of population sizes (deer, bear, bobcats, etc.). It uses some new algorithms for determining bounding boxes within images that one of the students on the project brought back from an internship. Unfortunately, it isn't working well enough yet. We are also attempting to recruit human image taggers (or to contract with a tagging service) to get "ground truth" for a large classification and counting study.

Food Security: The science team has two preliminary deployments "in the field." The first is monitoring sunny-side/shady side moisture content in the soil in an almond orchard located in Fresno, California. The data is somewhat inconclusive as of yet due to on-going hardware reliability issues. One of the team collaborators has designed a low-power controller board that has a bug which causes data corruption periodically. We are working on both a software work around (filer the data) and a hardware fix (respin the board). Once we get these issues worked out, the next step will be a full deployment for 30 trees consisting of 180 moisture sensors.

#1541215

The other deployment is at a citrus orchard in Exeter, California. The purposes of this deployment are to detect an air temperature differential at night during periods of time where frost could form. Warm air at 10-meter altitude can be mixed with cooler air at 1-meter altitude to prevent frost from forming. The science team is working on a deployment of 24 low cost temperature sensors. To keep the cost low, the installation labor burden low, and the resulting installation safe for normal farm operations (e.g., tractor proof), the team is developing an innovative set of analytics to get highly accurate temperature readings with a minimally invasive instrumentation footprint. The first experimental system was installed in February and this setup failed. Revision 2 was installed in mid-March and, so far, it is holding up. The next step is to instrument in the orchard itself (currently we are measuring temperature on the ground in a farm-safe location). We are planning an installation trip in April to try and get the next phase under way. We are doing much of the analytics and all of the website hosting on Aristotle. The websites can be seen here: http://169.231.235.221

5.0 Community Outreach and Education

5.1 Community Outreach

- Aristotle use case scientist T. J. Shepherd described how North American wind climate impacts wind farms at the January American Meteorological Society Meeting in Austin, TX. See the video: https://ams.confex.com/ams/98Annual/videogateway.cgi/id/44205?recordingid=44205&uniqueid=Paper331226&entry_password=508882
- UCSB Aristotle use case scientists described the "Where's the Bear?" project in a February news article and video: http://www.news.ucsb.edu/2018/018702/wheres-bear
- Aristotle co-PI Tom Furlani participated in a "Metrics Panel" at the Coalition for Academic Scientific Computation Spring 2018 meeting in Alexandria, VA; PI David Lifka presented "Findings of the External Advisory Committee."
- Aristotle infrastructure lead Resa Reynolds attended the March Dell EMC Community Meeting in Austin, TX; Resa is President of Dell XL and a member of the Board of Directors for the Dell EMC HPC Community.

5.2 Education

- Cornell hosted an all-day AWS Research Immersion Workshop in March that featured a review of AWS basics, a deep dive on foundational services (computer, storage, containers, etc.), research computing patterns, and a hands-on lab.