ARISTOTLE
CLOUD FEDERATION

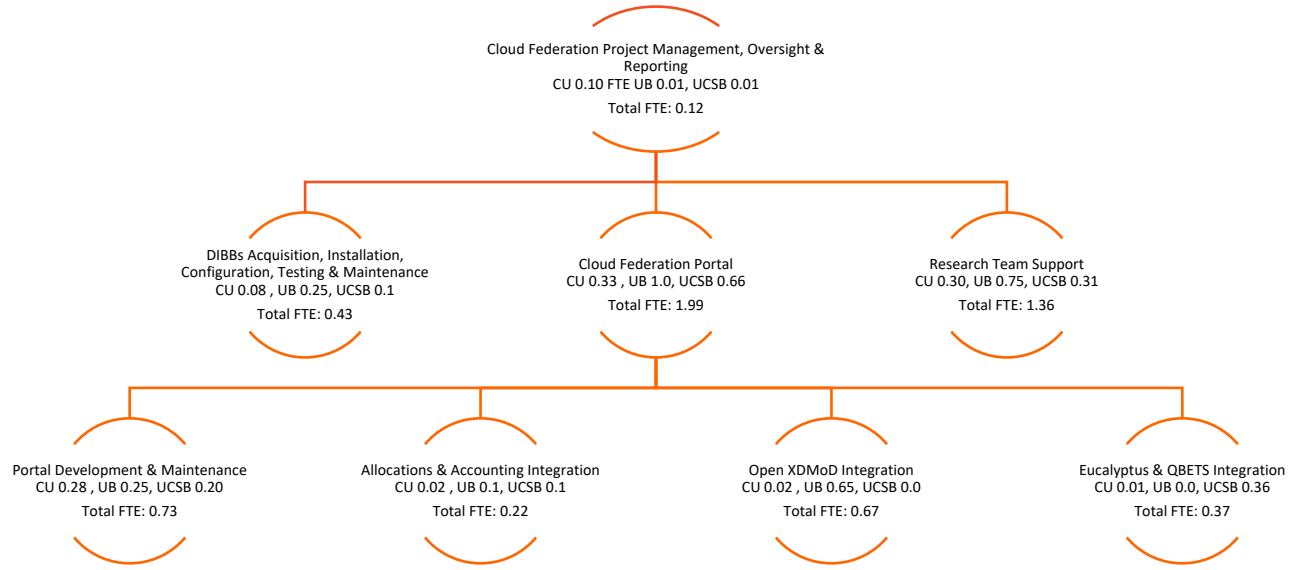
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 5: Quarterly Report 2

3/25/2020

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

This is the Program Year 5: 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

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/Partner Meetings
- Met with Red Hat and secured support for our switch to the OpenStack Platform Director model.

1.5 Project Planning Meetings/Status Calls

1/14/2020 status call:
- Apache Kafka distributed streaming platform is being tested on Aristotle by the SCIMMA project.
- A new use case in California will be using camera traps, solar panels, and an edge cloud to track ranging cattle in real time with Aristotle providing the back end.
- Another use case is focused on automating a UCSB student farm being created to produce food for food insecure students with the goal of nutrition equity. We are installing instrumentation, weather stations, and soil moisture sensors to automate farm management and enhance productivity. Again, Aristotle will provide the back end.
- In addition, the Citrus Under Screening project will be installing infrastructure next and integrating it with irrigation control systems, weather stations, etc. The goal is to protect citrus trees from the deadly citrus greening bacterium and Asian citrus psyllid spread.
- We are pursuing addition cloud credits from Amazon to run large-scale benchmarks of our DrAFTS 2.0 spot pricing instrument.

1/16/2020 progress call:
- We will be submitting an REU request to provide 6 undergraduates with the opportunity to work on cloud computing infrastructure research and a data visualization framework; high fidelity modeling and analytics for improved understanding of climate-relevant properties; transient detection in radio astronomy search data; and, cloud computing for science research, including assisting in the development of containers and container system monitoring using Docker, Terraform, and Ansible.
- We are planning to improve JSON file security and developing plans to share the Aristotle portal template via GitHub to the greater CI community.
- We are documenting how to effectively move an image from Aristotle to AWS.
- The Cornell Aristotle team is helping Dartmouth to investigate Magnum, an OpenStack API service that makes container orchestration engines available as resources in OpenStack.

#1541215

1/22/2020 progress call:

- Cornell is completing scripts for the Lake Problem and WRF codes; work on FRB search code continues to progress.
- Our investigations have demonstrated that AWS can use RightScale to tag instances, Terraform is ideal for scripting because it can tag everything, and Google will tag VMs.
- A request for additional AWS credits was submitted.

1/28/2020 status call:

- Aristotle team member Brandon Barker is writing a paper with Angela Douglas on results from their research on *Drosophila* gut microbiota interactions with the host. Barker built a simulation pipeline that takes organism models as inputs and gives metabolic fluxes/growth rates as outputs.
- We are working on running WRF on not only Aristotle resources but NSF's Jetstream.
- The SCIMMA project continues to focus on getting Apache Kafka working on OpenStack; testing is progressing on Aristotle resources. Note: Aristotle's science use case lead Adam Brazier is also SCIMMA's cyberinfrastructure lead.
- UB use case scientist Dominik Roesch plans to present an educational seminar on the Aristotle financial database framework to Cornell researchers and PhD students in March.
- Open XDMoD is not getting storage data in GBs.
- Dartmouth will be rebuilding their cloud with Globus Auth.
- UCSB plans to purchase additional storage to meet user demand.
- We had initial discussions on future Red Hat support.

2/1//2020 status call:

- The student run Edible Campus project is now pulling data down to Aristotle. The current goal is to understand growing conditions. We plan to work on analytics at some point.
- The cattle grazing camera project will provide answers to research questions such as whether cows eating grasses in a free-range is better or worse for native grasses. We are installing a giant radio network, cameras, and an edge cloud that will be sending data to Aristotle for processing.
- All Aristotle technical docs on the portal are now pointing to the original, single source GitHub docs in order to give more timely access to researchers.
- We are planning to benchmark select use case codes against established benchmarks to get an idea of speeds and costs using different container approaches (i.e., X-Containers, etc.).
- UCSB is working on a reinstall and is planning to onboard new projects.
- Red Hat has recommended going to 3:1 replication in Ceph.
- UB now has GPUs working and is testing the creation of different instance types with virtual GPUs. They will be sharing their findings and documentation and hope to roll out to production soon.
- Cornell is working on their test Ceph cluster domain upgrades before they go into production. They are also looking at upgrading some networking gear to keep up with the bandwidth.
- We discussed the concept of providing federated consulting to the national research community, delivering our expertise in cloud, application containerization, and other skill sets that will grow in researcher demand over the next five years.

2/25/2020 progress call:

- UB is getting their research software interface up to snuff so it can be used with their Lake Effect cloud.
- Aristotle PI Lifka attended the Dartmouth CIO workshop. Dartmouth, who is working with Aristotle to deploy their first cloud, has key alumni at AWS and Google.

#1541215

- We submitted a paper abstract to PEARC '20 entitled "Reproducible and Portable Workflows for Scientific Computing and HPC in the Cloud." The paper details how we packaged representative scientific applications in Docker and how we used Terraform and Ansible to automate the provisioning of compute resources and deployment of each application in a Multi-VM cluster. Variations of data management constraints, Multi-VM MPI communication, and embarrassingly parallel instance deployments and lessons learned are reported on.

3/10/2020 progress call:

- UCSB hired a new programmer who is focused on making DrAFTS 2.0 (the second degeneration Aristotle AWS pricing tool) production ready. DrAFTS 2.0 is now containerized and it will be set up to be supported as a long running service and be integrated into the Aristotle portal.
- UCSB researchers "love" Aristotle and the UCSB team has received another hardware donation.
- The Aristotle project summited our REU proposal to NSF OAC.
- UB added their GPUs and is currently trying to determine whether T4s are a viable use case instead of the more costly V100s. Testing is underway with researchers. T4s will likely be targeted for UB's cloud.

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

2.1 Hardware Acquisition

- Cornell worked with Red Hat to secure Red Hat support exceptions for all federation sites. These exceptions will enable Aristotle sites to obtain Red Hat assistance in implementing Red Hat's recommended OpenStack cloud installation path which previously did not exist.

2.2 Installation, Configuration, and Testing

- Cornell updated their test Ceph cluster to Nautilus. This was a two-step process: first to Luminous, then to Nautilus. Upgrade notes from the UB Aristotle team were invaluable to Cornell's successful upgrade. Separately, Cornell shared their vCompute Server (NVIDIA Virtual Compute Server) licensing information with UB. Learning from each other is a benefit of membership in a federation.
- UCSB worked with Red Hat on updating OpenStack to v15. They are also working on Globus transfer. For some unknown reason, their OpenStack TripleO deployment stopped working; Red Hat support is engaged in this problem and working on a solution.
- UB successfully deployed Globus transfer and demonstrated moving data between UB and Cornell. They officially have a Globus endpoint up and running with multiple DTNs running on OpenStack; upgraded all the original Ceph OSDs with Write Intensive Drives for the meta data operations; upgraded the 10gb NICs to Intel versions due to instability issues with Broadcom cards; added 2 V100 GPU servers to the compute pool; configured vGPUs on each Hypervisor to support 2 different types of GPU; and, added 1 T4-based GPU server to the compute pool. In addition, UB is testing performance and applications on various vGPU types to determine which types they will add going forward. They've requested a loaner Quadro RTX 8000 graphics card from Dell/NVIDIA so that they can test to see if this is a viable solution to use in the cloud. This particular GPU offers a larger memory option which would allow them to run more vGPUs per hypervisor. This option would give them more vGPU instances per dollar over the V100 GPU card. They just need to ensure that it provides adequate performance for their users. UB is also enabling support

#1541215

for Windows VMs (users will need to provide their own licensing). They have this working and will add the Horizon configuration changes to Puppet to decrypt the default password.

- Dartmouth's redeployment of OpenStack stalled this quarter but their staff is committed to working on this next quarter. The Aristotle/XSEDE CRI partnership team has offered to help Dartmouth on-site with their redeployment.

2.3 Federated Identity Management

Researchers use single sign-on at any member site.

2.4 Cloud Status by Site

The chart below shows each site's production cloud status. Dartmouth's cloud is in test mode.

	Cornell	Buffalo	UCSB
Cloud URL	https://redcloud.cac.cornell.edu	https://lakeeffect.ccr.buffalo.edu/ (access only to federation)	https://openstack.aristotle.ucsb.edu/
Status	Production	Production	Production
Software Stack	OpenStack	OpenStack	OpenStack
Hardware Vendors	Dell	Dell, Ace	Dell, HPE, DXC
DIBBs Purchased Cores	*616	**256	356
RAM/Core	8GB	up to 8GB	9GB Dell, 10GB HPE
Storage	Ceph (1244TB)	Ceph (768TB)	Ceph (528TB)
10gb Interconnect	Yes	Yes	Yes
Largest instance type	28core/240GB RAM	24core/192GB RAM	48core/119GB RAM
Globus File Transfer	Yes	In Progress	In Progress
Globus OAuth 2.0	Yes	Yes	Yes
Total Cores (DIBBs purchased cores + existing cores) = 2424	* 616 additional cores augmenting the existing Red Cloud (1252 total cores).	** 256 additional cores augmenting the existing Lake Effect Cloud (600 total cores).	***356 cores in UCSB Aristotle cloud (572 total cores, Aristotle is separate from UCSB campus cloud)

2.5 Tools

- Red Hat OpenStack – we successfully secured a support exception status for all three sites. We will start conversations with Red Hat about future license costs next.

3.0 Cloud Federation Portal Report

Content updates to the project portal are ongoing (https://federatedcloud.org).

#1541215

Open XDMoD continues to monitor data ingestion from all sites, as well as provide the utilization data (https://federatedcloud.org/using/federationstatus.php).

The portal planning table was not updated this quarter:

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	Transition to OpenStack console with 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 –5/2018**	**6/2017 – 10/2018**	**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. Note: due to OpenStack move, account creation across sites is delayed.	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 software changes were made to the portal platform this quarter.

3.2 Integrating DrAFTS into the Portal

UCSB has completed the process of hiring a programmer to implement the DrAFTS 2.0 prototype (developed by a graduate student working on the project). The student prototype is being handed off to the programmer to make it ready for general purpose usage.

3.3 Integrating Open XDMoD into the Portal

3.3.1 Application Kernels Containerization in the Cloud

The application kernel (AK) module of XDMoD monitors the performance of open cloud services by periodic execution of containerized application and analysis of their run time. In the part quarter, the AK module reported recurrent failure in the execution of tests on Red Cloud. An examination revealed that it was caused by an incomplete release of storage volume upon the deletion of compute instance. The Cornell team fixed this problem. Similar behavior is exhibited on the current UCSB OpenStack service, but it is not clear whether it is the same problem or not.

#1541215

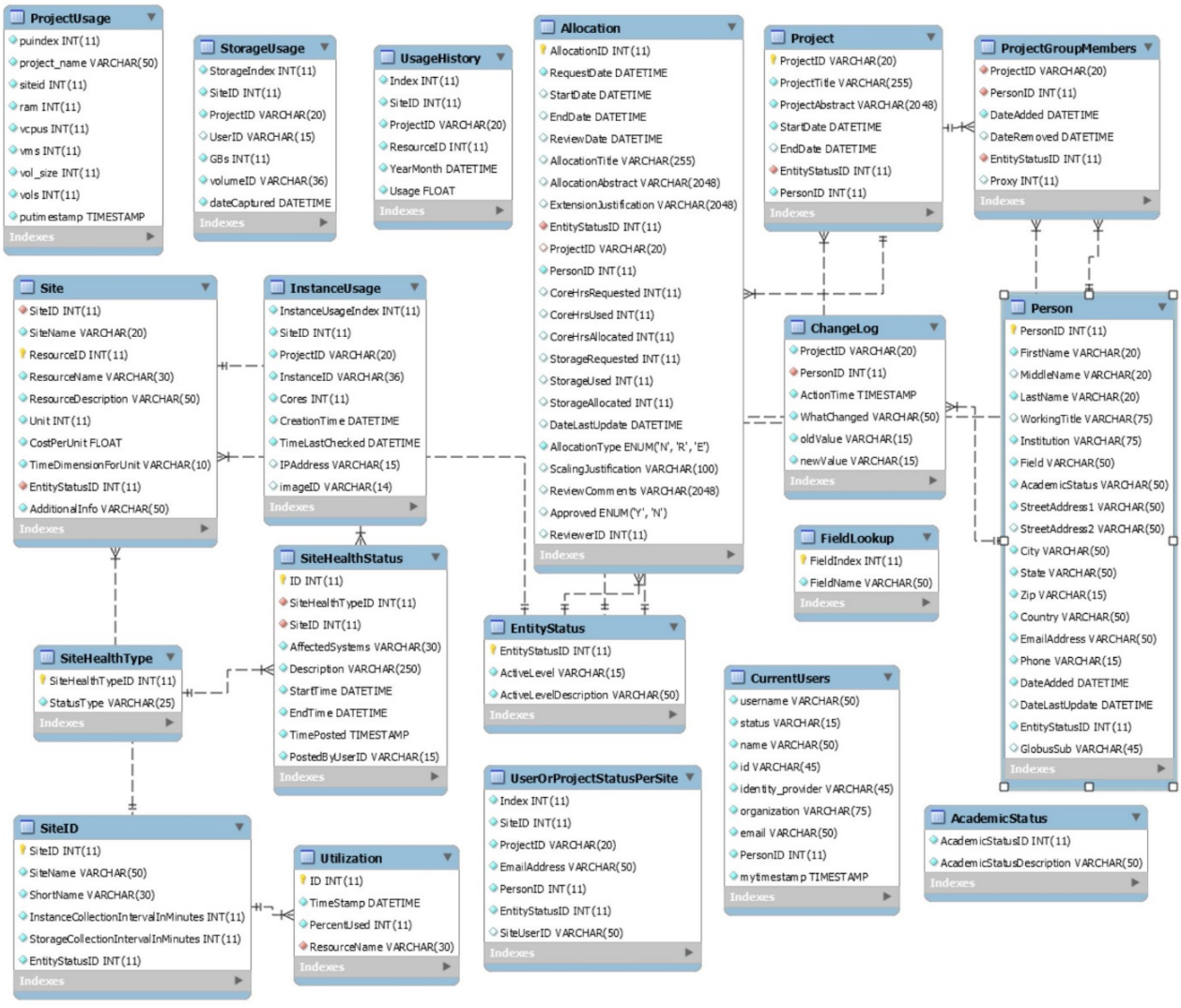
3.3.2 XDMoD Cloud Integration

RightScale XDMoD integration was evaluated. Currently, we are not able to retrieve the information that we need to integrate Rightscale into XDMoD. We're in the process of setting up a meeting with Rich Knepper, Bob DeLeon, Steve Guercio, Greg Dean, and Andrew Bruno to discuss whether it's possible to utilize the management side of RightScale which is where we can get the instance level events/data.

3.4 Allocations & Accounting

A usage report was created that show core usage, per site, per month. The draft report is currently out for review by the project directors.

No changes were made to the database schema this quarter:



#1541215

4.0 Research Team Support

The Aristotle Science Use Case team submitted a paper to PEARC '20 that shares lessons learned. The paper is entitled "Reproducible and Portable Workflows for Scientific Computing and HPC in the Cloud."

Profiling and testing of multi-instance MPI in the cloud continued with an eye toward an additional paper submission later this year.

An Aristotle partnership with the Scalable Cyberinfrastructure for Multi-Messenger Astronomy (SCIMMA) project is underway. We are currently prototyping a publish-subscribe service for distribution of transient astronomy data based on Apache Kafka.

4.1 Science Use Case Team Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data

UB Professor Varun Chandola and collaborators continued to develop a cloud enabled infrastructure to support the "Lower Great Lakes Resiliency Research Group." This involves building a framework on the Aristotle cloud that enables participants to interact and share data with each other. We have hired an undergraduate student (funded from a different grant) to develop this framework. The first version of the framework is already in use with a core group and will be extended to the larger group in April 2020.

Use Case 2: Global Market Efficiency Impact

UB Professor Dominik Roesch gave a guest lecture at Cornell, introducing the framework used to analyze financial data hosted on the Aristotle cloud. The talk—"Financial Market Frictions and Learning from the Stock Price: Let's 'Pay Attention to the Plumbing'"—was given as a guest lecture in a Market Microstructure PhD seminar hosted by Finance Professor Gideon Saar at the Cornell Johnson School.

Additionally, Professor Roesch is working with UB's Center for Computational Research to make this framework easily accessible to researchers within UB and, at a later stage, to researchers at other universities.

Use Case 3: Application of the Weather Research and Forecasting (WRF) Model for Climate-Relevant Simulations on the Cloud

Précis objectives of Cornell Professor Sara C. Pryor's and Post Doctorate Associate Tristan J. Shepherd's 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 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 (WRF v.3.8.1) code.

The research themes we have focused on are analyses of WRF simulations that were previously performed on Aristotle. These include:

Objective 1: Analyses of our convection-permitting high-resolution simulations with the WRF model conducted at 2 and 4 km grid-spacing (resolution) over the eastern USA to examine the presence/absence of low-level jets and examine the frequency of non-ideal wind speed profiles and the impact on those phenomena from varying resolution (Aird et al in review, Barthelmie at al. in review).

Objective 5: Analyses of our WRF output to characterize wind farm wakes (i.e., disruption of downstream near-surface properties) as simulated using the Fitch and EWP wind farm parameterizations applied at 2 and 4 km resolution, and developing a rigorous framework to use those simulations to provide guidance for a planned field experiment (Pryor et al. in review).

We have also continued new simulations for a domain centered over the Southern Great Plains in order to explore whether the inferences we have drawn regarding simulation sensitivity to resolution that have largely been based on our simulations of the US Midwest are fully generalizable irrespective of the base climate. These simulations are particularly interesting in terms of exploring what is possible on a single VM. The inner-most domain comprises 247 by 247 grid cells with 41 vertical levels (note: our attempts to run with a higher number of vertical levels failed in part due to numerical instability). There are three inner domains in these simulations; the first has no wind turbines operating, the second uses one wind farm parameterization (Fitch), and the third uses a second wind farm parameterization (EWP). Thus, it is very computationally expensive and RAM demanding, and it would make an exceptional candidate for a trial of simulations across multiple VMs. Due to the slow compute speeds on the single node VM, we are running one winter month, one spring month, one summer month, and one fall month. The status of the simulation is: January and April runs have completed; the simulation of July has commenced.

A third area of activity is focused on transitioning from conducting WRF simulations on Aristotle to XSEDE resources in anticipation of the conclusion of the Aristotle project at the end of summer 2020.

Finally, some of the Pryor group personnel time during the past three months has been directed towards supporting Cornell Aristotle staff efforts to develop a paper that was submitted to PEARC '20.

Activities planned for next quarter:

- Our activities will focus on completing the WRF simulations over the southern Great Plains and continuing the analysis of WRF output generated to date in support of Objectives 1, 4, and 5.

#1541215

Journal manuscripts published this quarter:

- Shepherd T.J., Barthelmie R.J. and Pryor S.C. (2020). Sensitivity of wind turbine array downstream effects to the parameterization used in WRF. *Journal of Applied Meteorology and Climatology,* 59, 333-336. https://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-19-0135.1
- Letson F.W., Barthelmie R.J. and Pryor S.C. (2020). RADAR-derived precipitation climatology for wind turbine blade leading edge erosion. *Wind Energy Science,* 5, 331-347. https://doi.org/10.5194/wes-5-331-2020
- Pryor S.C., Barthelmie R.J. and Shepherd T.J. (2020). 20% of US electricity from wind will have limited impacts on system efficiency and regional climate. *Nature: Scientific Reports*, 10, 541. https://www.nature.com/articles/s41598-019-57371-1
- Pryor S.C., Shepherd T.J., Volker P., Hahmann A.N. and Barthelmie R.J. (2020). 'Wind theft' from onshore wind turbine arrays: Sensitivity to warm farm parameterization and resolution. *Journal of Applied Meteorology and Climatology,* 59, 153-174. https://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-19-0235.1?mobileUi=0&journalCode=apme
- Pryor S.C., Shepherd T.J., Bukovsky M. and Barthelmie R.J. (2020). Assessing the stability of wind resource and operating conditions. *Journal of Physics: Conference Series*, 1452, 012084. https://iopscience.iop.org/article/10.1088/1742-6596/1452/1/012084
- Letson F., Shepherd T.J., Barthelmie R.J. and Pryor S.C. (2020). Modelling hail and convective storms with WRF for wind energy applications. *Journal of Physics: Conference Series*, 1452, 012051. https://iopscience.iop.org/article/10.1088/1742-6596/1452/1/012051

Journal manuscripts summited this quarter:

- Pryor S.C., Shepherd T.J., Volker P., Hahmann A.N. and Barthelmie R.J. (2020). Diagnosing systematic difference in predicted wind turbine array-array interactions. *Journal of Physics Conference Series: Special issue from The Science of Making Torque from Wind*. University of Delft, Delft, The Netherlands. (in review).
- Aird J., Barthelmie R.J., Shepherd T.J. and Pryor S.C. (2020). WRF-simulated springtime low-level jets over Iowa: Implications for wind energy. *Journal of Physics Conference Series: Special issue from The Science of Making Torque from Wind*. University of Delft, Delft, The Netherlands. (in review).
- Barthelmie R.J., Shepherd T.J. and Pryor S.C. (2020). Increasing turbine dimensions: Impact on shear and power. *Journal of Physics Conference Series: Special issue from The Science of Making Torque from Wind*. University of Delft, Delft, The Netherlands. (in review).

Use Case 4: Transient Detection in Radio Astronomy Search Data

Led by Cornell Professor Jim Cordes and CAC Computational Scientist Adam Brazier, the radio transient detection use case focused on running containerized deployments of the FRB_pipeline in the Aristotle federation and AWS this quarter. Results were included in a paper submission to the PEARC '20 conference.

We successfully ran multiple small-scale runs in both clouds, and attempted a full-scale run of the pipeline on data that was known to contain FRBs for verification. While we were able to set up a full-scale test running the Friends-Of-Friends (FOF) algorithm, we encountered several issues due to the large data size.

#1541215

Small data sizes required much less time to move data around, smaller amounts of disk and memory requirements during processing steps, and overall, significantly less time to complete runs. Due to these issues, the virtual machine size required for even the master node of a cluster becomes prohibitively expensive if attempting to run large amounts of data for a long time in AWS, and likely other public cloud providers. Ultimately, these runs demonstrate that this particular use case works well in the cloud for smaller data sizes, but not for large data chunks. Additional exploration into how to do more processing in parallel is needed. We are also planning future runs on XSEDE resources as a point of comparison. The containers we have curated for this use case will be converted from Docker to Singularity and optimized for community use on HPC resources.

Use Case 5: Water Resource Management Using OpenMORDM

A paper written by B.C. Trindade, D.F. Gold, P.M. Reed, H.B Zeffb, and G.W. Characklisb on WaterPaths is in review. The paper contains benchmarks for hybrid MPI and OpenMP workflows showing high levels of performance for Aristotle relative to TACC's Stampede and a local group HPC cluster hosted at CAC.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota

This quarter the Angela Douglas Lab focused on collating and integrating our data from all prior analyses for publication. Our manuscript addresses four main areas of *Drosophila* gut microbiota interactions with the host: (1) growth of *Drosophila* gut bacterial communities under different nutrient regimes, (2) patterns of metabolite consumption and release, (3) metabolic roles of individual bacteria, (4) net outputs from the bacterial communities.

From these analyses, we have obtained a comprehensive description of how interactions between gut-associated microorganisms are influenced by host diet and community composition. We have identified specific microbial-derived metabolites including acetate, succinate and alanine as candidate metabolites that modulate host health. We are currently working on a final draft of the results section of the manuscript and we are projected to complete the manuscript within the next few weeks with an anticipated submission date to a journal by the end of April.

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

Edible Campus (UCSB):
The Aristotle science use case team is now engaging students at the Edible Campus to develop instrumentation to aid in the growing of crops for food insecure students at UCSB. The team (including an undergraduate student who the project hopes to support with our current REU proposal) is developing image recognition software both to track farm operations and also to create long-term "time lapse" imaging for crop growth. During Q1, the team install cameras and weather sensors at the farm site and developed the data acquisition system (hosted in Aristotle). The camera installation has proved especially timely since the UCSB policies on social distancing and student interaction make personal supervision of farm operations infeasible.

Citrus Frost Prevention (Lindcove Research and Extension Center, Exeter, CA):
Inclement weather in the California Central Valley has delayed the construction of the CUP structure until April at the earliest. The science team is planning a trip by the end of March to repair a disabled FLUX Tower and also to install new weather sensors, but UCSB COVID-19 isolation protocols may require that

#1541215

trip be postponed. The science team is currently consulting with the Vice Chancellor for Research at UCSB to determine the appropriate actions to take.

Sedgwick Reserve (Santa Ynez, CA)
The science team has been working with Sedgwick IT personnel to install a new radio network to facilitate the remote camera trap operation. The radio network is operational but camera installation is on-hold due to recent rain events making the trap locations inaccessible. In addition, the Sedgwick staff is requesting a camera location that requires the installation of conduit and cabling (due to radio attenuation by the foliage). This installation is planned for early Q2, subject to UCSB COVID-19 restrictions on researcher visitations to the site.

5.0 Community Outreach and Education

5.1 Community Outreach

- Aristotle science use cases were presented at the Sustainability Leadership Summit held at Cornell: https://federatedcloud.org/papers/2019_SustainabilityPoster_CACServices.pdf
- Aristotle use case scientist Sara Pryor's paper that analyzes the impact of current and future wind turbine deployments was the focus of a recent Cornell news release: https://news.cornell.edu/stories/2020/02/quadrupling-turbines-us-can-meet-2030-wind-energy-goals. The publication in *Nature* acknowledges the Aristotle project.
- SCIMMA will be doing some testing on Aristotle. Cornell wrote the SCIMMA award news release featured in news outlets such as *InsideHPC*: https://insidehpc.com/2019/10/nsf-invests-in-multi-messenger-astrophysics/

5.2 Education

- Aristotle Cloud Federation co-PI Wolski shared poster/lessons learned at the 2020 NSF CSSI Meeting: https://federatedcloud.org/papers/WolskiPoster2020.pdf
- Nevena Golubovic, an Aristotle science use case student, was awarded her PhD by UCSB. Her dissertation on scalable analytics systems for multi-tier IOT deployments acknowledges the Aristotle project: https://cs.ucsb.edu/sites/default/files/docs/reports/golubovicdec2019.pdf
- An analysis of changes to Amazon's spot pricing was featured in *Data Centre Dynamics*: https://www.datacenterdynamics.com/en/analysis/amazons-spotty-pricing/
- Cornell presented a seminar topic entitled "Need computer time? XSEDE, Frontera and Cloud."
- A cloud seminar series is planned for this fall. Topics will include "Moving Researchers to the Cloud," "Containerization Techniques for Cloud and HPC," and "GPUs in the Cloud."
- The Aristotle team will be presenting a tutorial at PEARC '20 entitled "Deep Dive into Constructing Containers for Scientific Computing and Gateways."

#1541215