ARISTOTLE
CLOUD FEDERATION

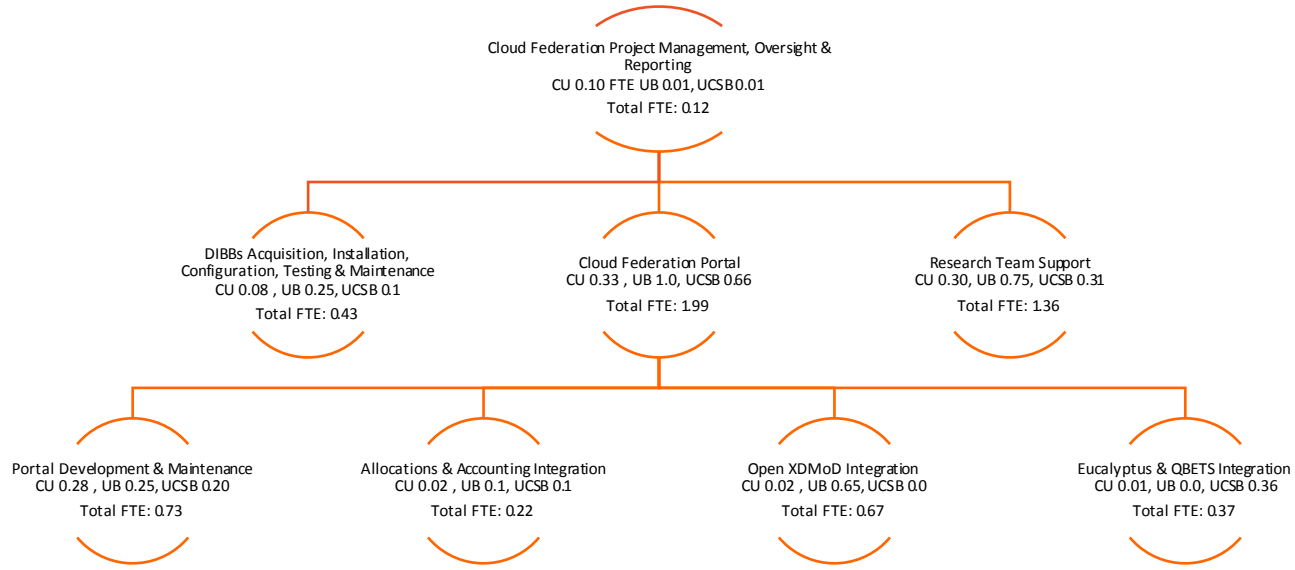
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 4: Quarterly Report 1

12/20/2018

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

This is the Program Year 4: 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

#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 and Knepper continued strategic discussions with senior management and technologists from AWS, Microsoft Azure, Google Cloud Platform, and IBM Cloud.

1.5 Project Status Calls

10/9/2018 status call:

- UB is finishing up testing of Open XDMoD 8.0. This version will enable the inclusion of initial cloud metrics into Open XDMoD. Once 8.0 is released and each site completes their installation, we will begin to collect OpenStack data and work on federating cloud data as well.
- Progress continues on implementing MPI and Docker for the OpenMORDM water resource management use case.
- Cornell is engaged in OpenStack deployment staff testing, UB has moved most of its instances from Eucalyptus to OpenStack, and UCSB is rebuilding their OpenStack system.
- Dartmouth is investigating how large a cloud to deploy and who to purchase it from. They joined the federation to explore the federated cloud computing model. See the *HPCwire* story: https://www.hpcwire.com/off-the-wire/dartmouth-joins-aristotle-cloud-federation-to-explore-federated-cloud-computing-model/
- Knepper is working on collecting documentation from all sites.
- UCSB plans to actuate the frost control fans automatically in the orange orchard at Visalia and the engineering team from Cal Poly are working on diagnosing problems at the Almond farm where they are measuring moisture on the sunny and shady side of almond trees for precision farming.
- Papers were submitted to Eurosys on CSPOT, our IOT OS that runs on Aristotle as well as everywhere else (https://www.cs.ucsb.edu/sites/cs.ucsb.edu/files/docs/reports/master_2.pdf). Papers were also submitted to an Ag science journal and *IEEE IOT – Design and Implementation.*
- UCSB PhD student Nevena Golubovic is giving a talk at the *IoT 2018 Conference* on Centaurus (the cloud service for K-means clustering running on Jetstream and Aristotle) and Aristotle use case scientist Chandra Krintz is giving keynotes at *IoT 2018* and the *2018 International Conference on Computer Aided Design*.
- The UCSB science team submitted to NSF Mediums and will be sending a third proposal to the Robert Wood Johnson Foundation based on Aristotle work.
- DrAFTS continues to collect data but we need a clear picture from AWS on what the pricing outlook will be.
- Microsoft's clouds for research director will be attending a meeting at UCSB.

#1541215

10/23/2018

- The Aristotle annual report was very well received by NSF. The Aristotle project review will be in Alexandria, VA in March.
- Dartmouth hopes to receive some hardware, go through Aristotle documentation, and do some initial testing by April 2019. Adding Dartmouth to the federation as the initial test site will be a good test for the quality of Aristotle's documentation.
- Lots of workloads make sense for XSEDE and some workloads make sense for the cloud; we see them as complementary resources. Cloud is increasingly being recognized as the right place for long tail science.
- We discussed the mechanics of integrating allocation requests into the portal dashboard: allocation details, proposal justifications, CVs, length of allocations, extension request process, etc.
- We decided accounting should be done through the identity management system.
- Lesson learned: edge computing must be able to operate at extreme temperatures. Equipment in the orange grove reached 150 degrees in the outbuilding and, as a result, we are shifting to military grade equipment.
- Our new, more watertight deployment in the almond grove is now collecting soil moisture data off a couple of hundred sensor streams (phase 1 was only 6 streams). We are using Aristotle exclusively for soil moisture clustering work; if Jetstream is available, we will add it to the resource pool.
- A new use case project was added that's focused on analyzing the positive and negative effects of cattle grazing on indigenous grasses. It will leverage lessons learned from our successful *Where's the Bear* camera trap data project which used edge computing, IoT, Aristotle, and public cloud.
- After porting, Centaurus is now spurning work in AWS using their Lambda stateless, event-driven program model. If we can do computation in Lambda, it will be a lot cheaper than using the spot market. We plan to do a cost study on this.
- UCSB is teaching a class in cloud computing. It starts in December and by the first week of January thirty students will want as many VMs as they can get from Aristotle.
- Cornell ordered more cores to add to the OpenStack side so they have enough cores to do their Eucalyptus to OpenStack migration. This is necessary because our cloud usage is over 85%.
- UB is refactoring their networking; they just need to add a couple extra new 10 gig switches. UB also created OpenStack scripts for creating and removing a private network that are now available in GitHub: https://github.com/smgallo/dev-tools/tree/master/openstack

11/7/2018

- Next quarter, we will evaluate whether to do an evictable class of service on Aristotle with DrAFTS.
- Cornell is deploying their new OpenStack nodes, staff testing is underway, and we are creating documentation and migrating our internal CAC infrastructure services into OpenStack. UCSB donated the head OpenStack node to the project. Separately, they received a gift of hardware from HPE/DXC which they are running Eucalyptus on for some statistics and data science users.
- WRF is being tested on OpenStack at Cornell and performance testing of MATLAB is also planned. We put 10TBs of data on Jetstream's Lustre and mounted it as VMs.
- Sansum Diabetes Research Institute is interested in developing a food as medicine use case on Aristotle.
- We plan to submit an REU request (10 REU students have made contributions to Aristotle use case research and performance technologies to date).

#1541215

12/4/2018

- Dartmouth has enough drives ordered to set up their Ceph cluster; after that's operational, the Dartmouth team will work on OpenStack.
- We discussed how to integrate Open XDMoD with the portal.
- The Aristotle database schema has been adjusted to incorporate allocations capabilities.
- UB has proper API calls that use either the project or PI name in order to pull the appropriate usage data out of XDMoD.
- The UCSB cloud team is currently focused on their OpenStack rebuild.

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

2.1 Hardware Acquisition

- Cornell purchased 16 nodes and a network switch, UB purchased network switches, and UCSB requested quotes.

2.2 Installation, Configuration, and Testing

- Cornell added 448 cores to the Red Cloud OpenStack cloud to allow the transition to OpenStack. The heavily used Red Cloud Eucalyptus cloud (85% utilization) was preventing us from migrating Eucalyptus nodes to OpenStack. After adding the cores, we were able to successfully move several science use cases to OpenStack. We're currently migrating the remaining use cases and plan to complete the migrations by February.

- Buffalo received the networking gear they needed to refactor their OpenStack network. The switches are now configured and installed in the racks. Next steps are to methodically juggle each OpenStack component, one at a time, in order to merge 5 separate physical networks into 2 bonded pairs on each OpenStack server. Tagged VLANs will be implemented to differentiate the provider networks. This work will be done "live" without disrupting users. Running instances will be migrated to other nodes while the controllers are converted. This work should be done by the end of December.

- UCSB successfully configured a separate Ceph cluster for OpenStack testing. They restructured their network environment to provide fault tolerance for critical services and to expand their network IP space in order to facilitate Eucalyptus usage during their OpenStack transition. They are continuing to develop their OpenStack deployment strategy using TripleO and have expanded their small OpenStack installation for beta tester use.

2.3 Federated Identity Management

Cornell, Buffalo, and UCSB are actively implementing federated logins. The federatedcloud.org site is publishing the Aristotle account list where Cornell, Buffalo, and UCSB will pull from to create accounts. Once all accounts are created at the three sites, Aristotle users will use Globus Auth to login at each site, using one set of credentials.

#1541215

2.4 Cloud Status by Site

The chart below shows each site's current production cloud status. Cornell and Buffalo have production OpenStack clouds. UCSB and Cornell have both OpenStack and Eucalyptus clouds.

	Cornell	Buffalo	UCSB
Cloud URL	https://redcloud.cac.cornell.edu	https://lakeeffect.ccr.buffalo.edu/	https://console.aristotle.ucsb.edu/
Status	Production	Production	Production
Software Stack	OpenStack	OpenStack	Eucalyptus 4.2.2
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 (1152TB)	Ceph (720TB)	Ceph (528TB)
10gb Interconnect	Yes	Yes	Yes
Largest instance type	28core/192GB RAM	24core/192GB RAM	48core/119GB RAM
Globus File Transfer	Yes	Yes	Planned
Globus OAuth 2.0	Yes	Yes	Planned
Total Cores (DIBBs purchased cores + existing cores) = 2060	* 616 additional cores augmenting the existing Red Cloud (1064 total cores).	** 256 additional cores augmenting the existing Lake Effect Cloud (424 total cores).	***356 cores in UCSB Aristotle cloud (572 total cores, Aristotle is separate from UCSB campus cloud)

2.5 Potential Tools

- Supercloud: nothing new to report
- Red Hat OpenStack: Cornell and UCSB are transitioning from Eucalyptus to Red Hat OpenStack. Buffalo has completely transitioned.

3.0 Cloud Federation Portal Report

Content updates to the project portal are ongoing (https://federatedcloud.org). Updates were made to many portal branches this quarter, including science use cases, publications, news, and events. Development of the portal user dashboard continues. Currently, we are adding functionality to manage project allocations and research team members; this is expected to be completed next quarter, and will include select cloud metrics from Open XDMoD.

A paper describing federating Open XDMoD for resources such as clouds was published in IEEE: Sperhac, J., Plessinger, B.D., Palmer, J.T., Chakraborty, R., Dean, G., Innus, M., Rathsam, R., Simakov, N. White, J., Furlani, T., Gallo, S.M., DeLeon, R.L., Jones, M.D., Cornelius, C. & Patra, A. (2018). Federated XDMoD to monitor affiliated computing resources. *2018 IEEE International Conference on Cluster Computing*. doi:ieeecomputersociety.org/10.1109/CLUSTER.2018.00074. https://federatedcloud.org/papers/hpcmaspa-ieee-approved-PID5492135.pdf

#1541215

We continue to monitor the Aristotle usage graph (https://federatedcloud.org/using/federationstatus.php) to ensure data is being collected consistently from all sites.

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.

#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	**3/2017–5/2018**	**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 Open XDMoD and DrAFTS into the Portal

DrAFTS continues to function as a pricing tool for AWS but without the reliability guarantees. UCSB personnel met with AWS representatives to discuss the changes AWS made with respect to spot instance pricing and reliability. The AWS team claimed that there was no real change but Aristotle data seemed to indicate otherwise. After an extensive analysis, the team determined both that Amazon has raised the prices of spot instances and also reduced their reliability. Neither of these changes is documented nor were they part of the AWS announcement of "enhancements" to spot instance pricing. Thus DrAFTS has proved a useful tool for auditing and understanding the cost and reliability of cloud resources (even when the vendor

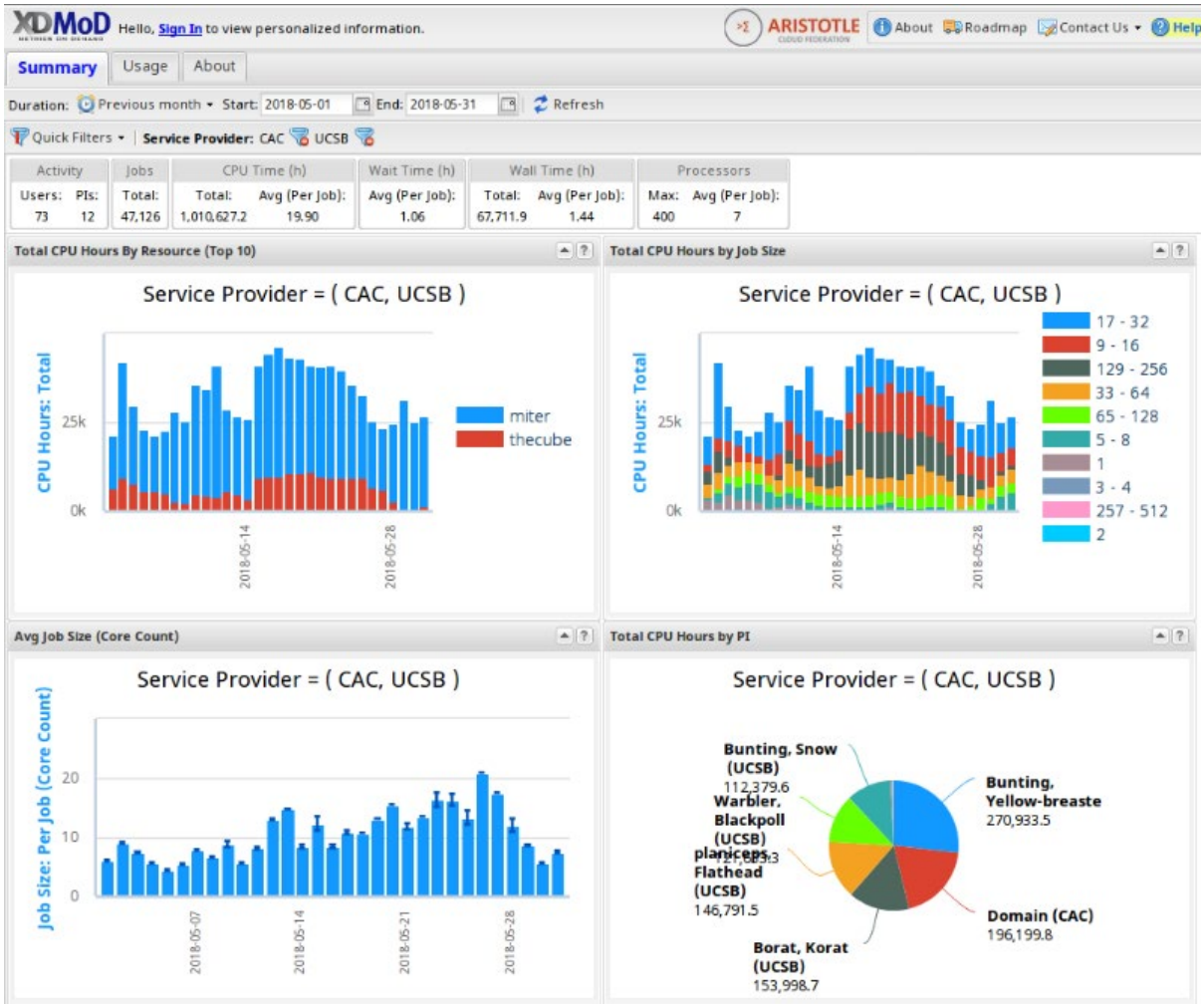
#1541215

fails to adequately announce service or pricing changes). We are exploring this new use of DrAFTS in the coming months as a way of determining reliability and price changes for volatile public cloud resources.

The UB team released Open XDMoD 8.0 which supports the ingestion of OpenStack log data for the generation of cloud metrics. It is available at https://github.com/ubccr/xdmod/releases. Work has begun on the federation of cloud data as well as the incorporation of additional cloud metrics.

Open XDMoD 8.1 will include REST endpoints that will allow the Aristotle portal to query cloud accounting data directly from XDMoD.

This is a screenshot of the federated Open XDMoD page:



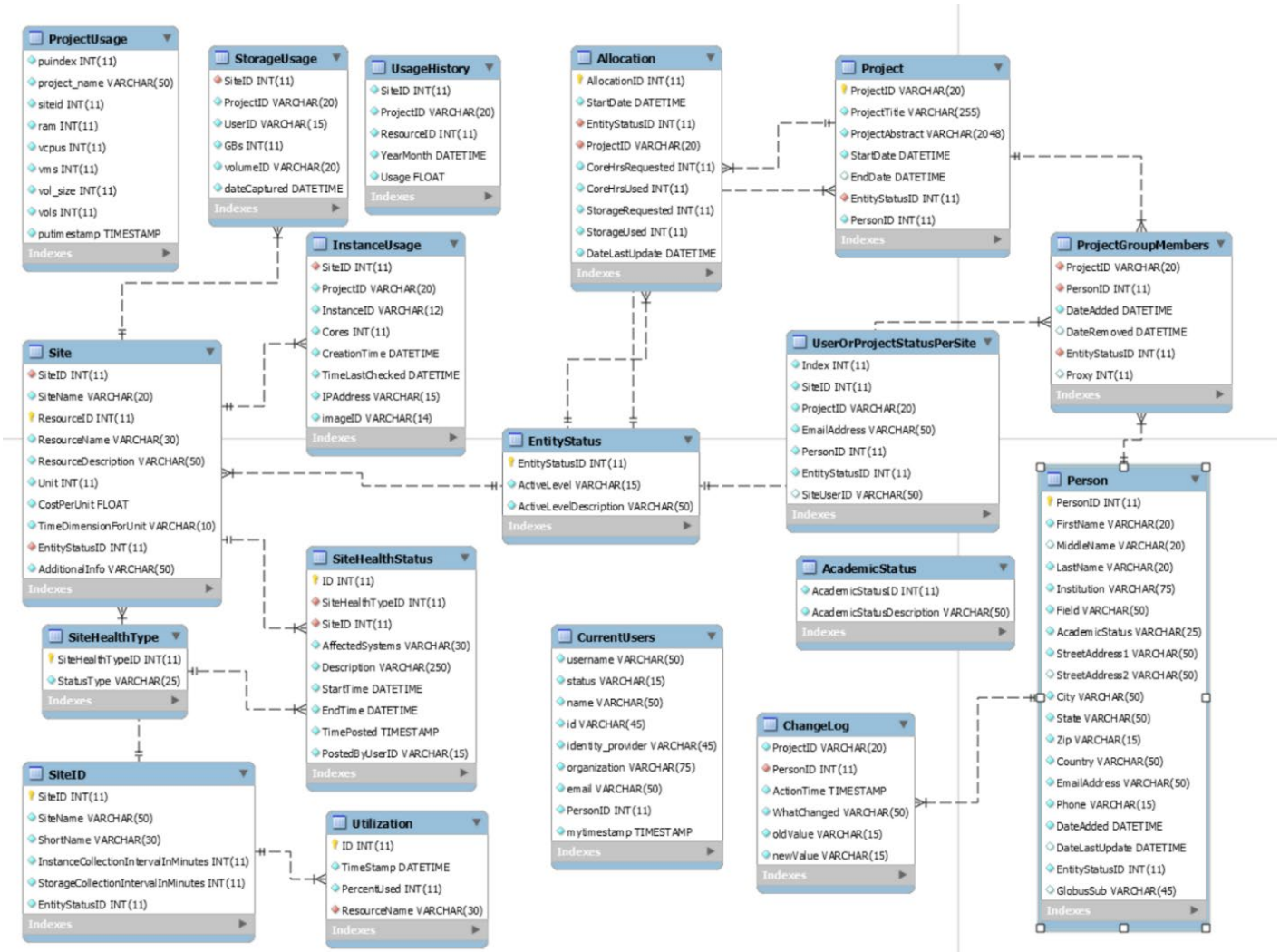
The Open XDMoD development timeline is available at:
https://docs.google.com/spreadsheets/d/1KlBIWY8ntCC35_5v7o19rro_oOM0Cre8WER-pIlSxMI/edit?usp=sharing

#1541215

3.3 Allocations & Accounting

- Created stored procedures for attaining allocation information.
- Created stored procedures for creating projects based on allocation information.
- Migrated the federatedcloud.org computer from Eucalyptus to OpenStack.
- Working towards getting information from the federated cloud database to CAC database for project and user creation in Ithaca.
- Adding calculations for CPU core hours and storage usage for new data format from OpenStack from Buffalo.

This is the database schema:



#1541215

4.0 Research Team Support

4.1 Science Use Case Team Updates

Aristotle science use case progress is highlighted below. Next quarter our use case support team will focus on completing cloud reproducibility work for Use Case 3, entering production runs for Use Case 4, completing initial performance characterization of the multi-instance MPI work for Use Case 5, and supporting the work of undergraduate student Jo Song for Use Case 6.

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

Varun Chandola co-organized a panel at the *American Geophysical Union Fall 2018 Meeting* on the "Application of Information and Data Science Methods and Technologies to Climate Research and Energy-Water Knowledge Discovery" (https://agu.confex.com/agu/fm18/meetingapp.cgi/Session/58681) and displayed a poster during the panel on work that is currently being deployed on the Aristotle cloud (see the AGU poster: https://federatedcloud.org/papers/AGUenergyWaterNexusPoster.pdf). Chandola's research team also added Redis for caching to webGlobe; this addition is expected to significantly improve server performance. The team is currently developing new machine learning components that will further enhance webGlobe (https://www.tandfonline.com/doi/full/10.1080/20964471.2018.1526057).

Use Case 2: Global Market Efficiency Impact

Progress on the analysis of financial markets efficiency continued this quarter: The Aristotle use case team:

- Investigated how price deviations (market inefficiencies) affect liquidity (the ease at which you can buy or sell). A paper by project lead Dominik Roesch received a revise and resubmit from one of the top three finance journals; that revision is underway.
- Investigated how the recent increase in the tick-size (the minimum price movement, of all U.S. stocks affects liquidity). This joint work with UB PhD candidate Albert Lee and Kee Chung also received a revise and resubmit request.
- Investigated the difference between asset prices during day and night. This joint work with Terrence Hendershott and Dmitry Livdan was accepted at two top finance conferences: the *American Finance Association Annual Meeting* and the *Jackson Hole Finance Conference*, both of which will be held in January 2019. Roesch also presented this paper at a Desmarais Global Finance Research Centre seminar at McGill University in November. The paper will be submitted to a journal next quarter.

In addition, Roesch is working with two other PhD students using the finance data framework and the underlying data hosted on Aristotle.

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

Progress continues on the application of the Weather Research and Forecasting (WRF) model for climate-relevant simulations on the cloud led by professor Sara C. Pryor and postdoc associate Tristan Shepherd.

Précis objectives of our 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)

#1541215

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).

The focus of this quarter's activities was:

- *Activity 1:* Writing a new proposal to seek continued access to XSEDE Jetstream resources to enable analyses of the simulations we are conducting on Aristotle.
- *Activity 2:* Completing the analysis of output from our long-term simulation with the WRF model at 12km over eastern North America for 2001-2016 for the assessment of year-to-year variability in the wind resource (paper published in *Wind Energy Science*). Additional simulations to examine stability in the future climate are planned pending computing resources.
- *Activity 3:* Continuing to address the impact of (a) changes to the computational node on which our simulations with the WRF model are running and (b) the compiler used in terms of the net effect on the simulated wind climate. These matters are ongoing, since we still do not have a compilation of the WRF model using the Intel compiler, but initial results regarding simulation sensitivity to computing platform are published in a forthcoming book chapter and are also the subject of a conference presentation to be given in January 2019.
- *Activity 4:* Optimization of wind turbine (WT) arrays to maximize system-wide power production (i.e., the system-wide capacity factor) requires high-fidelity simulations of array-array interactions at the regional scale. We are systematically comparing two parameterizations of wind farm impacts on atmospheric flow in the WRF model using real WT characteristics and locations for a domain centered on Iowa (state with highest WT density) (Fig. 1a). Our simulations employ three inner nests; noWT (no WT parameterization), and then two nests where we employ two different parameterizations of rotor aerodynamics (and thus the disturbance of the flow downstream of wind turbines); Fitch and EWP. Our first simulations are at a horizontal resolution of 4 km and with 41 vertical levels. Analyses of simulation output both Fitch and EWP indicate similar seasonality in system-wide capacity factors (CF, Fig. 1b). However, CF are systematically higher in simulations using EWP (Fig. 1d) because of the more rapid recovery of wakes. Output from the EWP nest also indicates much smaller near-surface climate impacts from WT (Fig. 1c). We are conducting simulations at enhanced vertical and horizontal resolution to examine the resolution-dependence.

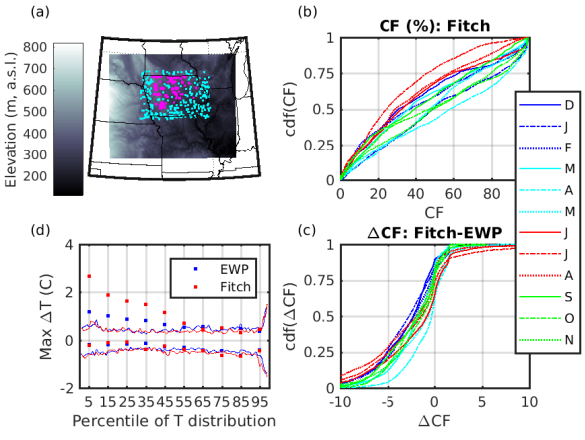


Fig 1: (a) WT grid cells (magenta) and background grid cells (cyan) in the inner domain. (b) System-wide 10-minute capacity factors (CF) from Fitch parameterization. (d) Difference in 10-minute CF (Fitch minus EWP) by calendar month. (c) Lines; maximum and minimum difference in T2m in background grid cells (EWP or Fitch minus noWT). Dots; maximum and minimum difference in any WT grid cell. Shown for 5:5:95th percentile T2m values during summer (JJA).

#1541215

Activities planned for next quarter:

- *Activity 1*: As described in previous reports difficulties remain with use of our XSEDE-Wrangler allocation. There has been sporadic email traffic about this over the last 3 months but our hopes for access to this resource for analyses of simulations we have conducted on Aristotle remains elusive. We will continue to work with CAC and the XSEDE staff.
- *Activity 2*: Complete enhanced high-resolution simulations of wind farm wakes from the two parameterizations (Fitch and EWP). These simulations are:
  1) Enhanced in the vertical (doubling of the number of vertical levels).
  2) Enhanced in the horizontal (halving of our horizontal resolution to 2 km).

  Our Aristotle VMs are too small to host all of the WRF output from a year simulation together with all the required input files, thus we pause the simulations periodically to transfer output to the NSERC HPSS, but the Globus transfer is currently very slow (ticket logged with CAC).
- *Activity 3*: I have acquired funding to support a 3-week visit for Tristan Shepherd to Pacific Northwest National Laboratory in Richland, Washington. The purpose of his visit is three-fold: (a) Assess the feasibility of porting a new global variable resolution atmospheric model (*Model for Prediction Across Scales* (*MPAS*), (https://mpas-dev.github.io/) to a Docker container on Aristotle as proof-of-principle for cloud-enabled simulations with this model. (b) Gain ability to configure/run MPAS. (c) Gain experience with analyzing wind output from MPAS. Naturally, it may not be possible for MPAS to be installed and run on Aristotle (due for example to the lack of an Intel compiler and limited RAM) and assessing whether it is feasible will require substantial assistance from the Aristotle science use case support team.

Forthcoming conferences:

- Pryor S.C., Shepherd T., Barthelmie R.J., Hahmann A., and Volker P. (2019): Wind farm wakes simulated using WRF. *Wakes Conference 2019*, Visby, Sweden, May 2019 *(abstract submitted)*.
- Pryor S.C., Shepherd T.J., Bukovsky M., and Barthelmie R.J. (2019): Wind energy scenarios for climate change mitigation. *The Scenarios Forum 2019*, Denver CO, March 2019. *(oral presentation)*.
- Shepherd T., Volker P., Barthelmie R.J., Hahmann A. and Pryor S.C. (2019): Sensitivity of wind turbine array downstream effects to the parameterization used in WRF. *99th American Meteorological Society Annual Meeting (10th Conference on Weather, Climate, and the New Energy Economy). (oral presentation)*.
- Shepherd T., Brazier A., Wineholt B., Barthelmie R.J. and Pryor S.C. (2019): Quantifying weather and climate simulation reproducibility in the cloud. *99th American Meteorological Society Annual Meeting (5th Symposium on High Performance Computing for Weather, Water, and Climate). (oral presentation)*.

Use Case 4: Transient Detection in Radio Astronomy Search Data

During this quarter, the use case team:

- Added a decimation method to the pipeline.
- Began implementing the Spitler and base PRESTO methods within the pipeline.
- Began work on a npz2fits_method to convert a NumPy Array to a PSRFITS file with correct headers. This will allow a seamless transition between the preferred storage format, PSRFITS, and the NumPy Array format which is simplest for manipulations in Python and will be particularly useful for creating simulated or injected data files.

Use Case 5: Water Resource Management Using OpenMORDM
This quarter we began working on regenerating all published results for the multi-node MPI Lake Problem simulation. This will be a valuable resource for others using the Lake Problem code, as well as serving as a general validation of our containerization methodology and as a useful benchmark between different container types (and vanilla VMs). We are simultaneously using the prior work on containerizing the Lake Problem code as a starting point for containerizing the Waterpaths code base, which will involve larger simulation runs on Jetstream once tested on Aristotle.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota
This quarter we focused on core model validation with SteadyCom and collecting and analyzing experimental data. In conjunction with prior work on constructing models and simulations using the models, we now have high confidence in our existing methods and our ability to simulate a wide variety of conditions moving forward.

Accomplishments this quarter included:

- Confirmed SteadyCom would successfully run in rich and minimal media on a multi-species model consisting of 5 bacterial species.
- Validated SteadyCom for feasibility when simulating multiple species that vary in population biomass by at least 6 orders of magnitude.
- Focused on metabolomic experiments that will serve to inform and validate future simulations.
- Added more unit tests to our simulation code repository.
- Began migrating work on our Windows VM in Aristotle/Eucalyptus to a Linux VM in Aristotle/OpenStack, which will aid in simplifying the workflow when dealing with the Linux-only Jetstream/OpenStack cloud.

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

Two publications were accepted and will be published in the *Journal of Big Data Intelligence* and *Proceedings of the International Conference on the Internet of Things*.

- Golubovic, N., Krintz, C., Wolski, R., Sethuramasamyraja, B. & Liu, B. A scalable system for executing and scoring K-means clustering techniques and its impact on applications in agriculture. Accepted in *International Journal of Big Data Science*. http://www.cs.ucsb.edu/~ckrintz/papers/centaurus-journal18.pdf
- Krintz, C., Wolski, R., Golubovic, N. & Bakir, F. Estimating outdoor temperature from CPU temperature for IoT applications in agriculture. Accepted in *International Conference on the Internet of Things (IoT)*. http://www.cs.ucsb.edu/~ckrintz/papers/iot18.pdf

We also delivered two invited keynotes:

- Krintz, C. SmartFarm: IoT systems that simplify and automate agriculture analytics. Keynote delivered at the *8th International Conference on Internet of Things (IoT 2018)* Santa Barbara, CA. http://lampz.tugraz.at/~iotc/iot-conf/iot2018/index.php/program/keynote-speakers/
- Krintz, C. Adventures and opportunities in cyber-physical systems and research. Keynote delivered at the *2018 International Conference on Computer Aided Design*, San Diego, CA https://iccad.com/adventures_and_opportunities_in_cyber_physical_systems_research

#1541215

Citrus frost prevention (Lindcove Research and Extension Center, Exeter, CA):

In this quarter, we investigated a number of different edge and sensing device configurations for deployment. We studied different combinations of single board computers, microcontrollers, edge clouds, and radio technologies (Wi-Fi, Zigbee) with various solar panel deployments. Our goal is to identify low-power and low-cost IoT deployments that produce accurate microclimate temperature prediction. This multi-tier IoT technology is part of a distributed application for frost detection and prevention for navel oranges. This problem is challenging due to the different requirements: fruit maturity, canopy size, temperature inversion, and mitigation strategies (irrigation, fans). We also installed an automatic fan that is triggered by temperature. Our goal is to link the two systems to reduce the cost of frost prevention and to study the impact of the fans on air movement on a per-block and per-tree basis going forward.



Low-power and low-cost IoT deployment such as solar panels and sensors will help researchers know when to turn on windfans to protect citrus from frost.

Precision irrigation for almond trees (Fresno State):

During this quarter, we developed the next version of our SmartFarm microcontroller sensor board. We evaluated a small (2 tree) deployment in the almond test farm and used it to improve the board and deployment. We then deployed a 9-board system across multiple rows of the field at two different locations. We are using the system to study the differences in soil moisture loss on different sides of the row (one which receives more direct sunlight than the other). Our goal is to develop row-specific irrigation strategies that account for these differences in an effort to reduce water use (without impacting yields). The deployment is a complete end-to-end, multi-tier IoT system that links Aristotle, edge systems, and sensors that target precision irrigation for almonds. We have published a paper related to this effort during this quarter and sent it to a journal on big data intelligence that provides decision support to farmers interested in site-specific farm management. We plan to extend this on-line tool to include the Fresno State precision irrigation deployment data and to develop new analyses that provide decision support for row-based irrigation during the next quarter.



UCSB Aristotle researchers and grad students and faculty from Cal Poly San Luis Obispo and Fresno State are working with farm consultants to instrument an almond test tree for differential irrigation. The goal is to see how much water can be saved by irrigating different sides (sunny and shade) of the root stock in proportion to its dryness

#1541215

5.0 Community Outreach and Education

5.1 Community Outreach

- Aristotle portal lead Susan Mehringer was a panelist on "Federations: Lessons Learned" at the Coalition for Academic Scientific Computation (CASC) Fall 2018 meeting in Alexandria, VA: https://federatedcloud.org/papers/CASC%20Aristotle%20Federation%20Lessons%20Learned.pdf
- An October *HPCwire* story—"Dartmouth joins Aristotle Cloud Federation to explore federated cloud computing model"—conveyed the motivations for and benefits of campus-to-campus cloud resource sharing: https://www.hpcwire.com/off-the-wire/dartmouth-joins-aristotle-cloud-federation-to-explore-federated-cloud-computing-model/.
- The Aristotle team exhibited at SC18 in Dallas and provided science use case updates: https://federatedcloud.org/science/SC18Briefing.pdf.

5.2 Education

- Aristotle hosted an "Openshift Container Application Training" all day workshop with a hands-on lab at Cornell on 12/5/2018.
- UCSB will be teaching a cloud computing class starting in December and will use Aristotle and other cloud resources.
- The 11/2/2018 Aristotle news release—"NSF REU students gain hands-on science experience working on Aristotle Cloud Federation project"—highlighted how REU students contributed to Aristotle science use case projects and participated in the development of emerging cloud computing technologies: https://www.cac.cornell.edu/about/news/181102.aspx.
- On 10/24/2018, we posted a "Centaurus K-Means Clustering as a Service" video to demonstrate the balancing of a scalable clustering analysis workload between two Aristotle clouds: https://federatedcloud.org/images/Balancing%20a%20Centaurus%20Workload%20on%202%20Aristotle%20Clouds%20Demo.mov.
- We also posted an "Aristotle WRF Container" video to demonstrate the pulling or caching of Docker images to get Weather Research and Forecasting Model apps running quickly: https://federatedcloud.org/images/Aristotle%20WRF%20Container%20Demo.mp4.
- UCSB PhD candidate Nevena Golubovic under the supervision of Aristotle use case scientists Chandra Krintz and Aristotle co-PI Rich Wolski, wrote a dissertation on tailoring cloud computing and IoT to agricultural processes. Her work has been recognized by the Institute for Energy Efficiency who on 11/20/2018 announced that she has been awarded the 2018-2019 J. Frenkel Foundation Fellowship for outstanding research contributions to the field of energy efficiency.

#1541215