

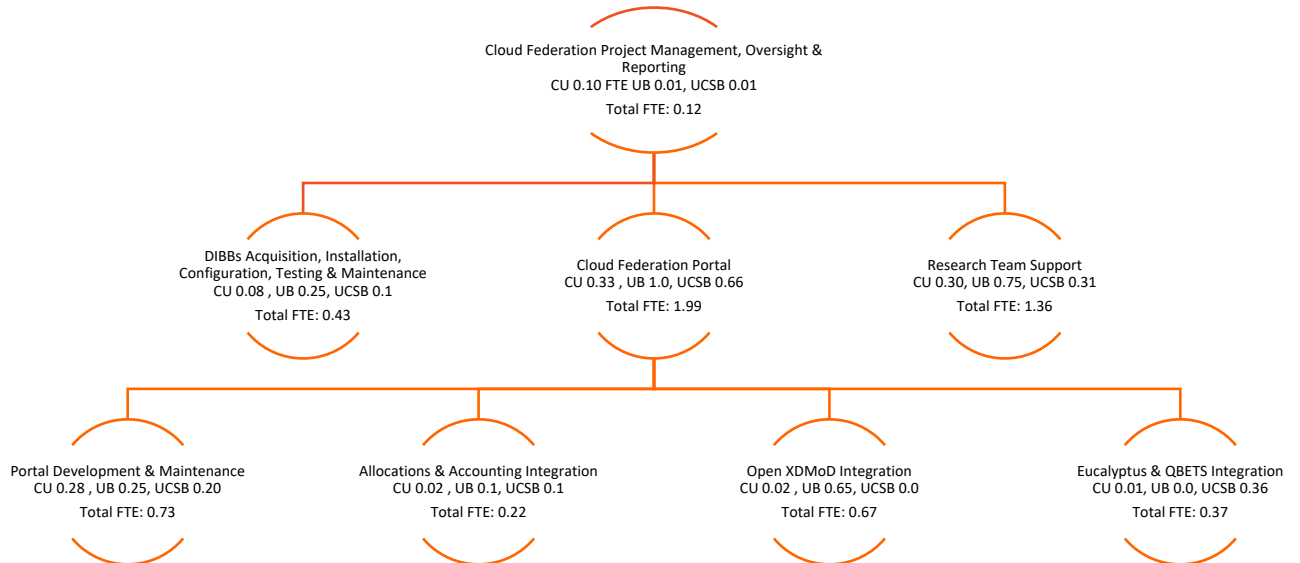
# CC\*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

## Program Year 4: Quarterly Report 3

6/28/2019

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

This is the Program Year 4: Quarterly Report 3 of the Aristotle Cloud Federation team. We report on plans and activities for each area of the project Work Breakdown Structure (WBS).



## Contents

<b>1.0 Cloud Federation Project Management, Oversight &amp; Reporting Report .....</b>	<b>3</b>
1.1 Subcontracts .....	3
1.2 Project Change Request.....	3
1.3 Project Execution Plan.....	3
1.4 PI/Partner Meetings.....	3
1.5 Project Planning/Status Calls .....	3
<b>2.0 DIBBs Acquisition, Installation, Configuration, Testing &amp; Maintenance Report.....</b>	<b>7</b>
2.1 Hardware Acquisition.....	7
2.2 Installation, Configuration, and Testing.....	7
2.3 Federated Identity Management.....	7
2.4 Cloud Status by Site.....	7
2.5 Tools.....	8
<b>3.0 Cloud Federation Portal Report.....</b>	<b>8</b>
3.1 Software Requirements & Portal Platform .....	10
3.2 Integrating Open XDMoD and DrAFTS into the Portal .....	10
3.3 Allocations & Accounting .....	11
<b>4.0 Research Team Support .....</b>	<b>12</b>
4.1 Science Use Case Team Updates .....	12
Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data	12
Use Case 2: Global Market Efficiency Impact.....	13
Use Case 3: Application of the Weather Research and Forecasting (WRF) Model for Climate-Relevant Simulations on the Cloud.....	13
Use Case 4: Transient Detection in Radio Astronomy Search Data .....	14
Use Case 5: Water Resource Management Using OpenMORDM .....	14
Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota.....	15
Use Case 7: Multi-Sourced Data Analytics to Improve Food Production & Security.....	15
<b>5.0 Community Outreach and Education .....</b>	<b>16</b>
5.1 Community Outreach .....	16
5.2 Education .....	16

## **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/Partner Meetings**

The results of partner meetings are included in Section 1.5.

### **1.5 Project Planning/Status Calls**

4/9/2019 status call:

- The project team discussed comments and suggestions made by the NSF review panel and set the following project goals for the next 18 months: (1) complete the portal with all sites at same functionality level and fully interoperable, (2) package the capabilities so others can build their own cloud federations (or join ours), (3) investigate pricing capabilities with RightScale, focusing on two or more science use cases, and benchmarking on Aristotle and a public cloud(s), (4) Work with the XSEDE CRI to get other campuses to implement OpenStack clouds for research; Dartmouth is helping us harden our OpenStack documentation and training, (5) demo cloud bursting as an application level service (note: the main point of the project is to enable campus cloud sharing and to investigate how one might trade resources, not to burst to public clouds), (6) add new use cases if they are ready to use the federation and do not require in-depth consulting support; new use cases will push us in new directions (e.g., “science as a service” use cases have VMs that have been running 2+ years on Aristotle).

4/23/2019 status call:

- Cornell is in the process of completing a script to automate user setup. We discussed how often the automated scripts should run and decided that a 4-hour cycle was too long for cloud users. We will try to reduce that time to 30 minutes (users ready to run in 30 minutes or less after making an account request). UB and UCSB will need to poll Cornell data more frequently than every 30 minutes for this to work.
- UB XDMoD 8.1 is being released this week. UB will update each site to 8.1 and make sure the cloud is working as far as data collection goes. A separate side bar is planned on how to pull accounting data out of XDMoD and feed it into the database at Cornell.
- A paper on DrAFTS will appear in a July conference in Prague and is expected to generate valuable community feedback. We are currently prioritizing DrAFTS’ features and looking at new features to add to the DrAFTS interface.

- Cornell officially turned off all Eucalyptus cloud systems on April 23; all nodes have been successfully moved over to the OpenStack cloud platform. UCSB is gearing up to update their OpenStack software stack. UCSB and UB will be ordering PY4 hardware.
- Patrick Reed's grad student is working with Cornell cloud engineer Bennett Wineholt on getting MPI running on Reed's water resource management application. The radio astronomy REU student is working with Pete Vaillancourt.
- We have 3 new potential science projects. The most exciting is a USDA-funded project with UCSB to investigate growing citrus trees in green houses. This has never been done before. The Asian Citrus Psyllid bug has wiped out in just 5 years most of the citrus trees in central Florida and is now in California and moving up the freeway into the Central Valley. The citrus industry is panicked and wants to build massive screened enclosures and grow oranges in them. All sorts of instrumentation will be required to measure the effects on the trees and fruit, whether the bug made it through the barriers due to a breach, etc., and the processing and storage of this data would occur in the Aristotle cloud. A second potential project will use lessons learned from the Aristotle image project and our IoT and edge computing experience to do shoreline marine ecology research on the last pieces of beachfront property in the Central Valley which has been bought by ESRI (GIS software) founder and environmental scientist Jack Dangermond and donated to UCSB as a reserve. A third potential project is another camera trap study like Aristotle's "Where's the Bear?" project.
- We are currently setting up the RightScale Optima platform to investigate cost reporting and monitoring. UB and UCSB are interested in getting a better understanding of the API capabilities of the platform.

5/7/2019 status call:

- We are working with the Cornell Cloudification team to separate Cornell AWS bills from AWS Aristotle credits which we will use for our open cloud marketplace investigation. Technical meetings (Knepper, Wineholt, Brazier, and Vaillancourt) are being scheduled with the RightScale technical team in order to define next steps. Cornell's Cloudification team helps units move apps to the cloud—all AWS, Azure, and Google accounts go through this team so that users can securely log into public cloud platforms. Our focus is on trying to define some notion of benchmark performance per dollar for selected use cases.
- UB plans to do some work this summer on containerizing some of their application kernels and benchmarking AWS vs. their local Aristotle cloud to see the cost differences. Open XDMoD app kernels do various benchmarks for Amber, WRF, etc. UB has a student working on containerizing the app kernels in Docker so they can fire them up and get performance information back.
- UB is also working on some XDMoD 8.1 release issues and on getting 8.1 up on UCSB and Cornell cloud sites.
- Cornell added GPUs to their development cloud and is testing them before turning them over for general production use. UB is investigating the addition of GPUs to their Aristotle cloud as well. Cornell will help UB make the GPU transition using their lessons learned. A lot of users are requesting GPUs.
- UCSB will write up how to get an application to run multicloud (across multiple sites); they've used Python, Celery distributed task queue, and other tools to achieve this. It was successfully demonstrated at our 36-month NSF review.
- Cornell is leveraging AWS AppStream to replace student labs on campus. This will reduce student lab costs (hardware, maintenance, etc.) and provide the students with more mobility.

5/21/2019 status call:

- Cornell is continuing to test its first GPUs as part of OpenStack.
- UB is purchasing Dell PowerEdge servers to double the size of their Ceph storage. UB is also beginning to test their NVIDIA Tesla T4 GPUs on OpenStack. They're hoping virtual GPU support will be included in the next release of Red Hat (they plan to run Queens until the next long-term support version comes out).
- Knepper's open cloud marketplace investigation team had two meetings with RightScale on how their billing centers work, how policies and allocations might work, and the platform's API capabilities. Reynolds and Wineholt are analyzing the API capabilities and looking at how to do our own orchestration to start/stop instances. The federation portal team may pull a subset of the RightScale information into the portal.
- Dartmouth is providing feedback on where there are holes in our OpenStack cloud implementation documentation.
- UB is making a change to the Open XDMoD reporting API so it's easier to provide a project name and then pull out that project's usage from XDMoD. Once that change is implemented, we will figure out how to get that into the portal. All sites are now up to date and on version 8.1.2.
- Globus is using the new version of the DrAFTS tool for AWS pricing analysis.
- UCSB met with the citrus greening project PI. The project is going forward and will soon be breaking ground (building screen houses for citrus trees to investigate whether this approach is feasible to protect citrus from the Asian Citrus Psyllid bug). They discussed power, sensor placements, etc. The Aristotle cloud will be the back end for this project.

5/22/2019 planning call:

- The open cloud marketplace investigation team will eventually produce a white paper on cost monitoring capabilities, explain how we tested them, and provide suggestions for future investigations. They will take select science use cases (e.g., WRF, OpenMORDM), run them on public cloud resources using the credits we've received, process billing where possible through RightScale, and see what it costs to run on a public cloud platform (AWS, Google, and/or Azure). Instance types will be made as close as possible. The plan is to visualize the results in the portal and show job characteristics, time to completion, and costs. UB plans to run Open XDMoD app kernels on the instances.

6/4/2019 status call:

- Cornell is helping Dartmouth design a bigger, more production-ready OpenStack cloud, including how to lay out the network, etc. This is now a top priority at Dartmouth. Steven Lee from Cornell may travel to Dartmouth to assist their systems team onsite if funding is available.
- The Aristotle portal accounting system was adjusted so that users may be added every 30 minutes rather than every 4 hours. Four hours was deemed too long a wait time for users eager to run on the cloud.
- Changes were made to Open XDMoD in order to get accounting data out of XDMoD. The portal is now querying XDMoD for the accounting data. The setup is done; you can now query individual users.
- We will be pursuing additional AWS credits for UCSB to use on their DrAFTS 2.0 project.
- UCSB has Globus working and is checking with Cornell and UB to ensure it's working there.

- UB's servers will double the size of the Ceph storage; they're expected to arrive in a couple of weeks.

6/4/2019 planning call:

- We received cloud credits from Microsoft's Elizabeth Bruce for the open cloud marketplace investigation and hooked them into the CAC RightScale outlook.com address. We are currently not seeing our usage because these are credits (essentially a bill showing zero usage). We hope to see an improvement in this situation with the Azure team by the end of the summer. Right now, we're just seeing a big zero. Google credits are also coming up zero. We are talking with RightScale and Google to try to figure out how to get better information (how to see the credits and how the credits work). The RightScale platform is currently showing the end of this process rather than credits as they are used. Cornell's Cloudification team has the same issue with their use of Azure credits—there's very little they can actually see in terms of usage except here's what we're billing you. Ideally, researchers would like to see how credits are being used, as they're being used.
- A ticket was submitted to RightScale to figure out why we can't see the AWS billing dashboard (preferred since AWS does per second billing). If we could see the dashboard, we could provide users who desire a more detailed view that information.
- We are working to access the most appropriate technical experts at RightScale to assist us.

6/18/2019 status call:

- Cornell is helping Dartmouth deploy their first Aristotle cloud. Dartmouth is currently getting quotes for ~600TB of Ceph and has requirements into networking for permanent DLANS.
- Cornell is testing their first GPU nodes in preparation for Aristotle deployment.
- UB is (1) purchasing GPUs which will enter a test phase soon, (2) working on having a node only accessible by a particular project, (3) working on having all software containerized for production cloud installations, (4) planning to add compute and storage nodes.
- A call is planned to address Cornell questions regarding collecting instance and volume data from XDMoD's central location.
- UCSB's REU student will be working on DrAFTS 2.0.
- User- and group-creating scripts are now running every 30 minutes so new users can get up and running faster.
- "Where's the Bear" continues to run with a graduate student responsible for operations.
- The "save the oranges project" (US citrus trees are being decimated by Asian insect) is scheduled to install screen growing areas in July.
- UB's finance project (Dominik Roesch) has a paper coming out in a financial economics journal. Two new PhD students have joined his research team and are using the Aristotle framework.
- Cornell cloud engineer Wineholt built a new machine for Sarah Pryor's research group so they can do WRF runs with a 150TB volume.
- Cornell is onboarding a new REU student who will be working on the water resources management project and on the containerization of use case applications.
- Dartmouth is investigating moving some current use cases to federation resources so their users can get started early prior to Dartmouth deploying their own cloud. This is one of the benefits of being part of a federation: taking advantage of lessons learned plus accessing cloud resources early at other sites.

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

### 2.1 Hardware Acquisition

- Cornell, USCB, and Dartmouth made no purchases this quarter.
- UB purchased storage assets to add to their Ceph cluster; they will be installed next quarter.

### 2.2 Installation, Configuration, and Testing

- Cornell has been investigating Red Hat's TripleO, an OpenStack deployment and management tool, using a few of the oldest Red Cloud nodes. TripleO is the deployment method Red Hat will be supporting. Cornell also installed a server with 4 GPUs into Red Cloud as a proof of concept; researchers are being asked to test and provide feedback.
- UB prepared a second rack with power and networking to expand their OpenStack. They are updating to the latest OpenStack Queens packages and will schedule a Ceph upgrade v2 to v3.2 in the near future. Their development cloud Ceph was upgraded successfully a few weeks ago. They are working with Red Hat on what they think is a small bug in the latest dashboard packages in which horizon ignores the API settings in the config file and uses API v1 by default for Cinder. They are also investigating what it would take for their current cloud to be managed by Red Hat Director Platform because it seems that this is the only supported way going forward.
- UCSB setup a rack of servers to create an OpenStack and Ceph testing environment. They continued restructuring their network environment to provide added fault tolerance for critical services. They too are preparing for OpenStack and Ceph upgrades. They successfully investigated and corrected their Globus authentication problem within the HA/containerized environment.
- Dartmouth has Globus single sign-on working so that once their OpenStack is running, they will easily become a member of the Aristotle Federation. They are collecting quotes for a production Ceph cluster to replace their small test cluster. Internal and external VLANs have been created for production deployment of the Ceph cluster and OpenStack. They are tearing out the test PackStack deployment and will redeploy using TripleO on CentOS using 100-200 cores pulled from their on-premise HPC cluster. They are actively seeking science use cases; something that could run soon on another site and/or something to run at Dartmouth as they get their OpenStack cloud deployed.

### 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 current production cloud status. Dartmouth's will be coming soon.

	Cornell	Buffalo	UCSB
<b>Cloud URL</b>	<a href="https://redcloud.cac.cornell.edu">https://redcloud.cac.cornell.edu</a>	<a href="https://lakeeffect.ccr.buffalo.edu/">https://lakeeffect.ccr.buffalo.edu/</a>	<a href="https://openstack.aristotle.ucsb.edu/">https://openstack.aristotle.ucsb.edu/</a>
<b>Status</b>	Production	Production	Production
<b>Software Stack</b>	OpenStack	OpenStack	OpenStack
<b>Hardware Vendors</b>	Dell	Dell, Ace	Dell, HPE, DXC
<b>DIBBs Purchased Cores</b>	*616	**256	356
<b>RAM/Core</b>	8GB	up to 8GB	9GB Dell, 10GB HPE
<b>Storage</b>	Ceph (1392TB)	Ceph (720TB)	Ceph (528TB)
<b>10gb Interconnect</b>	Yes	Yes	Yes
<b>Largest instance type</b>	28core/192GB RAM	24core/192GB RAM	48core/119GB RAM
<b>Globus File Transfer</b>	Yes	Yes	Planned
<b>Globus OAuth 2.0</b>	Yes	Yes	Yes
<b>Total Cores (DIBBs purchased cores + existing cores) = 2060</b>	* 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 Tools

- Red Hat OpenStack – Cornell, UB, and UCSB all have production OpenStack clouds. Dartmouth is building theirs.

## 3.0 Cloud Federation Portal Report

Content updates to the project portal are ongoing (<https://federatedcloud.org>). The portal user dashboard added cloud metrics from Open XDMoD to the project member section.

Open XDMoD is now monitoring data ingestion from all sites, as well as providing 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	Implement content/functionality as shown in following sections. Add additional information/tools as needed, such as selecting where to run based on	Release portal template via GitHub. Update periodically.



	downloads to the database.	software/hardware needs and availability.	
<b>Documentation</b>			
<b>Phase 1</b>	<b>Phase 2</b>	<b>Phase 3</b>	<b>Phase 4</b>
<b>10/2015 – 3/2016</b>	<b>4/2016 – 10/2016</b>	<b>11/2016 – End</b>	<b>1/2017 - End</b>
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.
<b>Training</b>			
<b>Phase 1</b>	<b>Phase 2</b>	<b>Phase 3</b>	<b>Phase 4</b>
<b>10/2015 – 3/2016</b>	<b>4/2016 – 12/2017</b>	<b>4/2017 – 12/2017</b>	<b>1/2018 - End</b>
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.
<b>User Authorization and Keys</b>			
<b>Phase 1</b>	<b>Phase 2</b>	<b>Phase 3</b>	<b>Phase 4</b>
<b>10/2015 – 1/2016</b>	<b>2/2016 – 5/2016</b>	<b>6/2016 – 3/2017</b>	<b>4/2017 – End</b>
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.
<b>Euca Tools</b>			
<b>Phase 1</b>	<b>Phase 2</b>	<b>Phase 3</b>	<b>Phase 4</b>
<b>10/2015 – 3/2016</b>	<b>4/2016 – 12/2016</b>	<b>1/2017 – End</b>	<b>1/2017 – End</b>
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

We now have a replacement for the DrAFTS system, the Aristotle AWS SpotPrice Visualizer. We also have a plan for making it into a price comparison tool that we will be working on during the summer with our REU student. This work will take place while we work on the price prediction portion of DrAFTS.

## Aristotle AWS SpotPrice Visualizer (Beta Version 1.0)

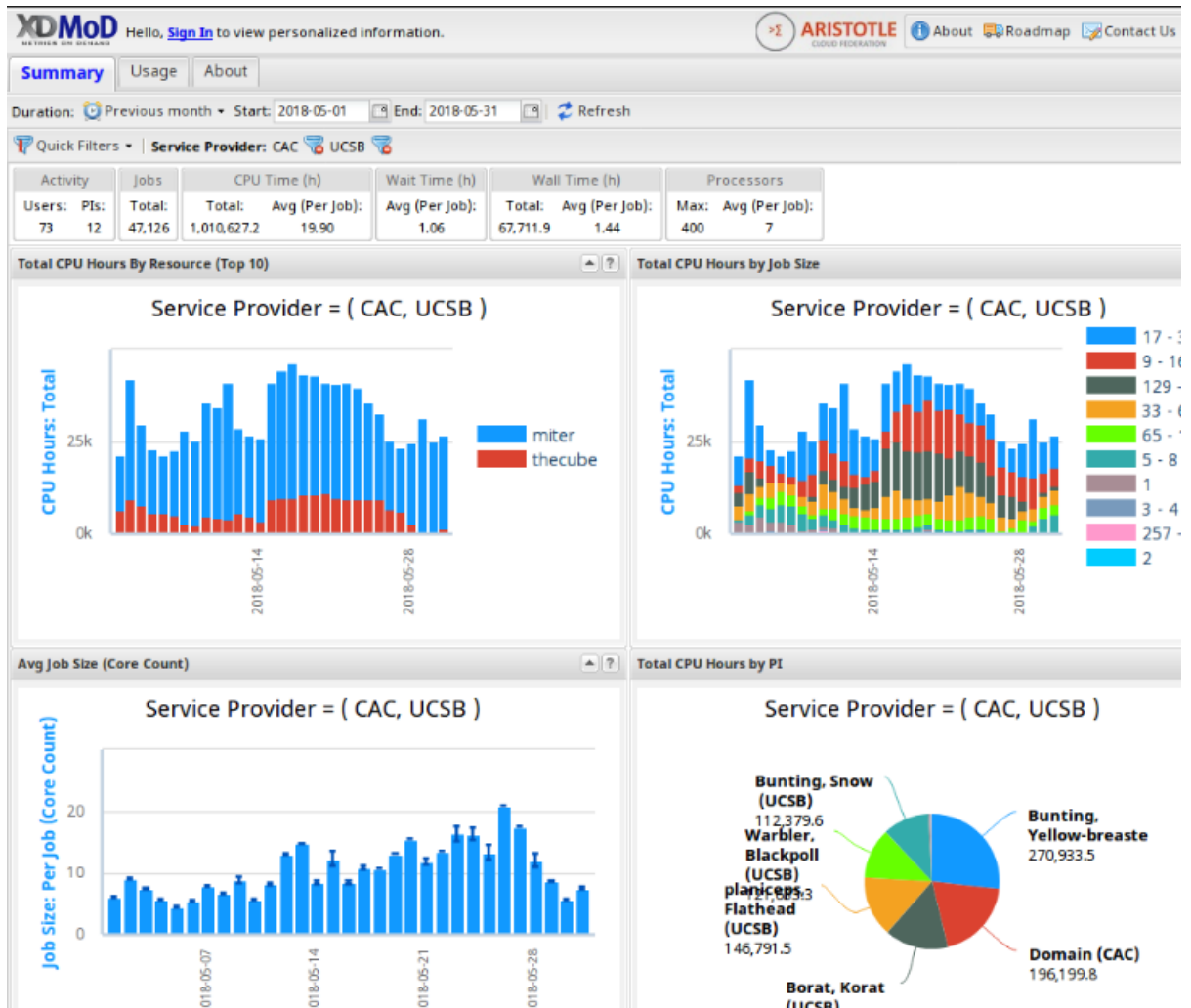
For usage instructions see the [/about](#) page

Start Date: 
 Stop Date: 
 Threshold 
 Show % Change ☒

	use1-az1	use1-az2	use1-az3	use1-az4	use1-az5	use1-az6	use2-az1	use2-az2	use2-az3	usw1-az1	usw1-az3	usw2-az1	usw2-az2	usw2-az3
a1.2xlarge		\$0.07 %0		\$0.07 %0		\$0.07 %0	\$0.04 %0	\$0.04 %0					\$0.31 %0	\$0.31 %0
a1.4xlarge		\$0.13 %0		\$0.26 %0		\$0.13 %0	\$0.08 %0	\$0.08 %0					\$0.62 %0	\$0.62 %0
a1.large		\$0.02 %0		\$0.02 %0		\$0.02 %0	\$0.01 %0	\$0.01 %0					\$0.08 %0	\$0.08 %0
a1.medium		\$0.01 %0		\$0.01 %0		\$0.01 %0	\$0 %0	\$0 %0					\$0.04 %0	\$0.04 %0
a1.xlarge		\$0.03 %0		\$0.03 %0		\$0.03 %0	\$0.02 %0	\$0.02 %0					\$0.15 %0	\$0.15 %0
c1.medium	\$0.01 %0	\$0.01 %0		\$0.01 %0		\$0.01 %0				\$0.01 %0	\$0.01 %0	\$0.01 %0	\$0.01 %0	\$0.1 %28
c1.xlarge	\$0.12 %0	\$0.11 %2		\$0.1 %8		\$0.11 %3				\$0.08 %3	\$0.06 %0	\$0.11 %8	\$0.11 %12	\$0.07 %2
c3.2xlarge		\$0.12 %0	\$0.12 %0	\$0.12 %0		\$0.12 %0				\$0.11 %0	\$0.11 %0	\$0.12 %0	\$0.12 %0	\$0.12 %0
c3.4xlarge		\$0.23 %0	\$0.23 %0	\$0.23 %0		\$0.23 %0				\$0.22 %0	\$0.22 %0	\$0.23 %0	\$0.23 %0	\$0.23 %0
c3.8xlarge		\$0.47 %0	\$0.47 %0	\$0.47 %0		\$0.47 %0				\$0.45 %0	\$0.45 %0	\$0.47 %0	\$0.47 %0	\$0.47 %0
c3.large		\$0.03 %0	\$0.03 %0	\$0.03 %0		\$0.03 %0				\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0
c3.xlarge		\$0.06 %0	\$0.06 %0	\$0.06 %0		\$0.06 %0				\$0.06 %0	\$0.06 %3	\$0.06 %0	\$0.06 %0	\$0.06 %0
c4.2xlarge	\$0.19 %0	\$0.13 %0	\$0.12 %0	\$0.12 %0	\$0.15 %0	\$0.12 %0	\$0.08 %3	\$0.07 %0	\$0.07 %0	\$0.12 %0	\$0.12 %0	\$0.12 %0	\$0.13 %0	\$0.12 %1
c4.4xlarge	\$0.32 %3	\$0.31 %2	\$0.29 %3	\$0.29 %3	\$0.41 %10	\$0.29 %3	\$0.15 %0	\$0.15 %2	\$0.14 %0	\$0.24 %1	\$0.24 %1	\$0.22 %5	\$0.25 %0	\$0.25 %0
c4.8xlarge	\$0.55 %0	\$0.52 %3	\$0.5 %0	\$0.55 %3	\$0.5 %0	\$0.53 %4	\$0.3 %2	\$0.3 %5	\$0.3 %10	\$0.47 %0	\$0.47 %1	\$0.49 %0	\$0.49 %0	\$0.51 %1
c4.xlarge	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.02 %0	\$0.02 %0	\$0.02 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0	\$0.03 %0
c4.xlarge	\$0.06 %0	\$0.06 %0	\$0.06 %0	\$0.06 %0	\$0.06 %0	\$0.06 %0	\$0.04 %0	\$0.04 %0	\$0.04 %1	\$0.06 %0	\$0.06 %0	\$0.06 %0	\$0.06 %0	\$0.06 %0
c5.18xlarge	\$1.17 %0	\$1.17 %0		\$1.17 %0	\$1.17 %0	\$1.17 %0	\$0.72 %2	\$0.7 %1	\$0.7 %1	\$1.11 %0	\$1.11 %0	\$1.16 %0	\$1.16 %0	\$1.16 %0
c5.2xlarge	\$0.13 %0	\$0.15 %2		\$0.14 %1	\$0.13 %0	\$0.15 %3	\$0.08 %0	\$0.08 %0	\$0.08 %0	\$0.12 %0	\$0.13 %0	\$0.14 %1	\$0.13 %1	\$0.14 %0
c5.4xlarge	\$0.29 %1	\$0.37 %3		\$0.36 %0	\$0.35 %2	\$0.36 %2	\$0.15 %0	\$0.15 %0	\$0.15 %0	\$0.25 %8	\$0.27 %8	\$0.27 %1	\$0.28 %4	\$0.26 %0

The XDMoD team has release Open XDMoD 8.1.2 and updated all Aristotle sites to this version. OpenStack log data is now flowing into the federation as shown in the chart below. The XDMoD team will be working with Cornell to begin the process of extracting usage accounting data from XDMoD for inclusion in the Aristotle portal. Work will begin shortly on examining the RightScale API to determine what public cloud information is available for inclusion into XDMoD.

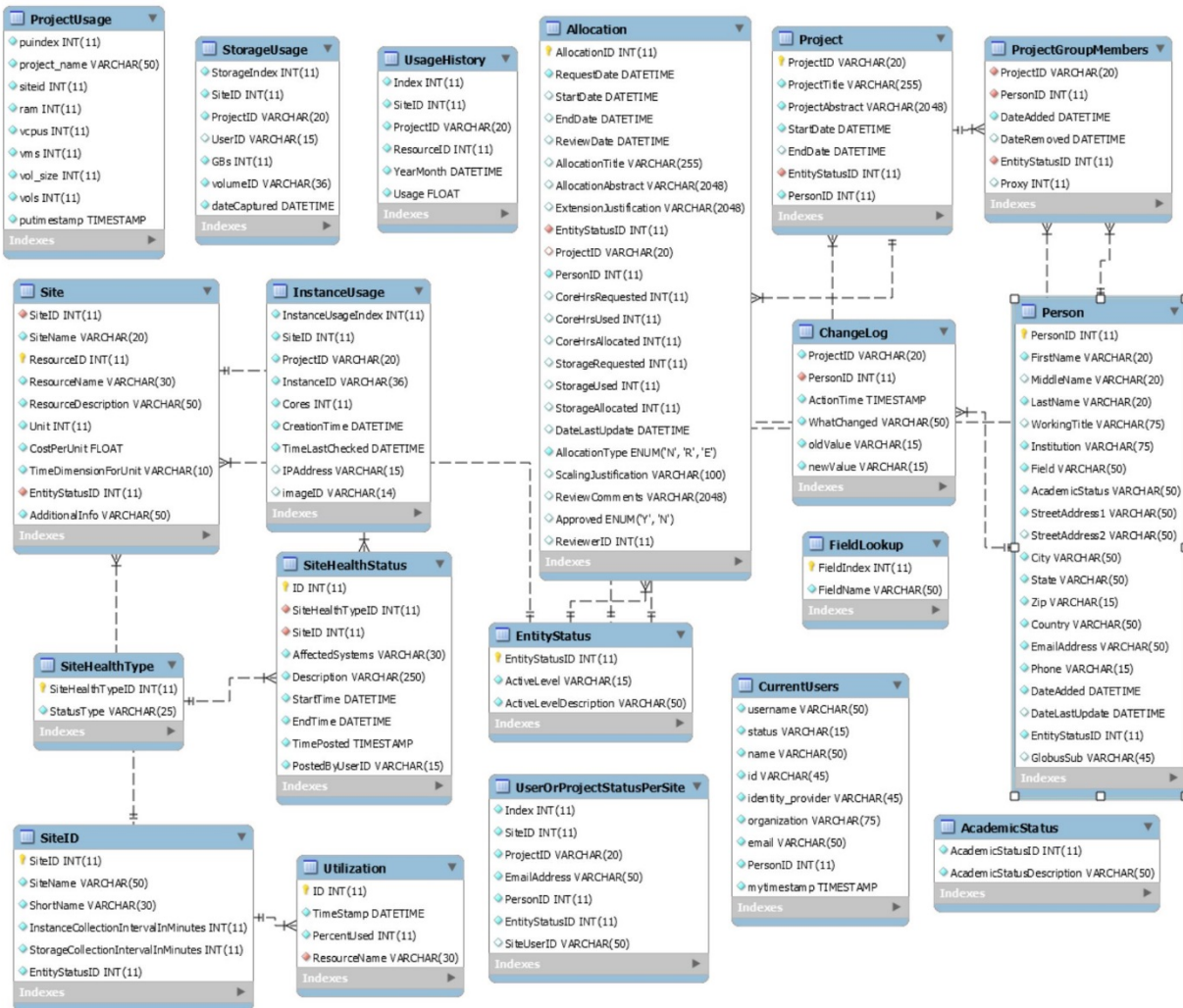
This is a screenshot of a federated Open XDMoD page:



### 3.3 Allocations & Accounting

- Created stored procedures and scripts for automating project and account creation.
- Automated scripts now run every 30 minutes.

No changes were made to the database schema this quarter:



## 4.0 Research Team Support

### 4.1 Science Use Case Team Updates

Aristotle science use case progress is highlighted below.

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

Varun Chandola (UB) and his collaborators are developing a Jupyter Notebook extension of the webGlobe framework that will allow users to analyze the underlying data in a Notebook.

### Use Case 2: Global Market Efficiency Impact

We prepared a paper— “Tick Size, Liquidity for Small and Large Orders, and Price Informativeness: Evidence from the Tick Size Pilot” (Kee H. Chung, Albert J. Lee, Dominik Roesch) for forthcoming publication in the *Journal of Financial Economics*. The paper acknowledges the Aristotle project (NSF 1541215): [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3220470](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3220470). We are also resubmitting a paper on how price deviations affect liquidity to the same journal, and wrapped up our financial data analyses for a third paper whose first draft will be completed in the coming weeks. In addition, work continues with two new PhD students who are using our framework and underlying financial data hosted on Aristotle.

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

Progress continues on the application of the WRF model led by Cornell 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 90<sup>th</sup> percentile value)
3. Quantify the platform dependence of wind simulations (i.e., quantify the differences in near-surface wind regimes from simulations conducted on conventional HPC and the cloud)
4. Examine inter-annual variability in near-surface wind speeds (can we simulate it, what is the source?)
5. Evaluate impact of large wind turbine (WT) developments on downstream climate (local to mesoscale)

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

Activities on Aristotle were temporarily stopped due to the urgent need to complete the transition of work from Eucalyptus to the OpenStack cloud platform. Thus, the simulations with WRF that we were conducting to assess resolution sensitivity had to be terminated after 9-months of simulation instead of the 12-months as originally planned. Nevertheless, these data are a treasure trove of information regarding the performance and sensitivity of key parameterizations within the WRF model to resolution. Analyses of this, and other output from our simulations on Aristotle, is being conducted on our Jetstream allocation and is the focus of a journal manuscript that was submitted by Tristan Sheperd in May and was also the subject of conference presentations during May and June at the Wakes conference and the Wind Energy Science conference by Sara C. Pryor (see details below).

Our primary activities on the Aristotle resources during this quarter focused on rebuilding instances in OpenStack, and selection of and integration of an REU student within the project. Recruitment of the REU student was greatly accelerated in time because we only received notification of the NSF resource allocation on April 30. We immediately advertised the REU positions with a due date of May 11 and began the process of generating an RCO Aristotle instance on which the student could access and analyze a 60TB subset of output from WRF simulations. Our project received 25 REU applications. Pryor interviewed three candidates on May 14 and selected one. Unfortunately, due to personal circumstances, the top candidate had to decline our offer of employment. We rapidly contacted the alternative candidate and Ryan joined the team and started employment on June 3. The analysis focus for this REU is extreme winds. I supplied

three key references on this topic and a MATLAB code for initial calculation of the extreme winds for him to modify for computational efficiency. There were a few teething problems with initiating the new instance but I believe it is now running and Ryan should now be able to commence the data analyses.

*Activities planned for next quarter:*

- Our activities will focus on WRF output and conducting (if feasible) additional simulations to test numerical stability to model formulation.

*Journal manuscripts:*

- Pryor, S.C., Shepherd, T.J., Barthelmie, R.J., Hahmann, A. Volker, P. (2019). Wind farm wakes simulated using WRF. *Journal of Physics: Conference Series (in press)*.

*Conference presentations:*

- Pryor, S.C., Shepherd, T., Voker, P., Hahmann, A., and Barthelmie, R.J. (2019). ‘Wind theft’ from onshore arrays: Sensitivity to wind farm parameterization and resolution. *Wind Energy Science Conference 2019*, Cork, Ireland, June 2019 (oral presentation).
- 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 (oral presentation).
- Pryor, S.C., Shepherd, T., Bukovsky, M., and Barthelmie, R.J. (2019). Wind energy scenarios for climate change mitigation. *The Scenarios Forum 2019*, Denver, CO, March 2019 (oral presentation)
- Barthelmie, R.J., Pryor, S.C., and Shepherd, T.J. (2019). Wakes from wind turbine arrays. IEA Wakes Task 23, NREL, CO, March 2019 (oral presentation).

#### **Use Case 4: Transient Detection in Radio Astronomy Search Data**

We focused on making improvement to the FRB Pipeline this quarter. These include some preparation for potentially parallelizing certain methods, updating to the Python 3 version of PRESTO (which is still in active development), and generalizing the pipeline to include some functionality for general radio transient detection beyond just FRBs. Updates to the user documentation for the design and usage of the pipeline are in progress.

We are also interested in adding checkpoint/restart functionality to the pipeline to support running tests on AWS spot instances as part of the Aristotle RightScale investigation. This includes adding features to improve robustness, such as logging functionality, prior to testing on AWS. The addition of these features is currently in progress.

#### **Use Case 5: Water Resource Management Using OpenMORDM**

Progress running WaterPaths on the cloud was slow this quarter. The initial reason was issues with WaterPaths' GitHub repository that prevented Cornell from having a stable test case. Once a stable WaterPaths+BorgMS test code and required data were established, we tested it with MPI across two nodes, and will be testing it on eight nodes once our multi-node setup is completed. Ideally, we will use these test results to show that WaterPaths can be used in the cloud with a performance comparable to that under HPC; that would encourage WaterPaths' potential users to adopt it in the absence of HPC resources.

Our REU student, Jackie Zheng, assisted with some cleanup of the OpenMORDM Lake Problem project to get his feet wet with container technology. He has finished that work and is now moving on to deployment automation looking at Ansible and Terraform to assist with multi-node deployments, in particular for the



MPI Aristotle projects.

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

Using SteadyCom, our team (Angela Douglas, Cornell) investigated the scope of feasible metabolic interactions that occur among a *Drosophila* gut microbial community (composed of *Acetobacter fabarum*, *Acetobacter pomorum*, *Acetobacter tropicalis*, *Lactobacillus brevis*, and *Lactobacillus plantarum*). From our simulations we identified a total of 159 unique metabolites that are exchanged between the *Drosophila* gut microbial community. The metabolites exchanged between community members are derived from central carbon (48%), amino acid (19%), vitamin (11%), nucleotide (11%), sulfur (8%) and nitrogen (3%) metabolism pathways. Of the 159 metabolites, our simulations revealed a subset of 20 metabolites that are most frequently exchanged (produced or consumed) between the gut microbial community.

Our simulations show that the gut microbial community is an important sink for EAAs (arginine, methionine, phenylalanine, tryptophan), nEAAs (asparagine, proline, tyrosine, serine, glutamine, glycine), TCA cycle intermediates (malate), glycerolipid intermediates (glycerol) and B vitamins (pantothenate, nicotinamide ribonucleotide.). Our simulations also show that the gut microbial community is an important source of TCA cycle intermediates 2-oxoglutarate and succinate. With regards to short chain fatty acids and other bioenergetic substrates, our simulations show the gut microbial community serves both as important source and sink for acetate, ethanol, lactate and formate.

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

#### *Where's the Bear:*

Graduate student Michael Zhang is taking over the “Where’s the Bear” project at the Sedgwick Reserve. His plan is to use the Pacific Research Platform (PRP: <https://ucsd-prp.gitlab.io>) to train image recognition models in conjunction with Aristotle. The PRP provides a shared GPU environment based on Kubernetes and containers. The goal is to schedule model training on the PRP and “burst” to Aristotle when GPUs (which cannot be shared on the PRP) are not available. In addition, he’s automating image acquisition and classification at Sedgwick. The eventual goal will be to use this system to provide real-time tracking of cattle that now graze on the property.

#### *Citrus Under Protective Screening (CUPS):*

CUPS installation is scheduled for July. The science team has been acting as technology consultants to the project, meeting with automation vendors and designing an instrumentation plan with the goal of having instrumentation available when the first trees are planted. In addition, the science team has been working with local growers who are interested in the project, either as a remedy for citrus greening or as a model for technological deployment and management. Aristotle will serve as the data hosting service for the CUPS project once the installation is complete. The UCSB team has scheduled a maintenance period during the month of August to allow for new capacity and software upgrades ahead of the CUPS launch. In addition, University of California Agriculture and Natural Resources (UC ANR) is making available a high-end weather station for the project. The science team has been training on the new equipment and performing initial integration experiments with Aristotle serving as the data repository.

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

The new instrumentation hardware was installed a month ago and upgraded last week. Both the initial installation and the upgrade have been fully operational since inception, and the solar and battery power infrastructure is now stable. In addition to greater scalability, the initial analysis of this hardware indicates

that it provides additional temperature and humidity accuracy. The analysis and data gathering are using Aristotle as the compute and storage platform. Also, the team has procured actuation controllers to be able to power on and off the local wind machines. Integration of these controllers will take place between now and the next frost season.

## 5.0 Community Outreach and Education

### 5.1 Community Outreach

- “Cornell investigating multi-cloud cost management with RightScale.” See *HPCwire*: <https://www.hpcwire.com/off-the-wire/cornell-investigating-multi-cloud-cost-management-with-rightscale/>

### 5.2 Education

- Aristotle use case scientist Varun Chandola is part of a team who received a Responsible Computer Science award to integrate ethics into computer science and engineering classes. See *UBNow*: <http://www.buffalo.edu/ubnow/stories/2019/06/rudra-computer-ethics.html>
- Cornell is refining OpenStack cloud implementation documentation in conjunction with Dartmouth which will be used to educate users and share lessons learned on deploying OpenStack clouds on U.S. campuses.