ARISTOTLE
CLOUD FEDERATION

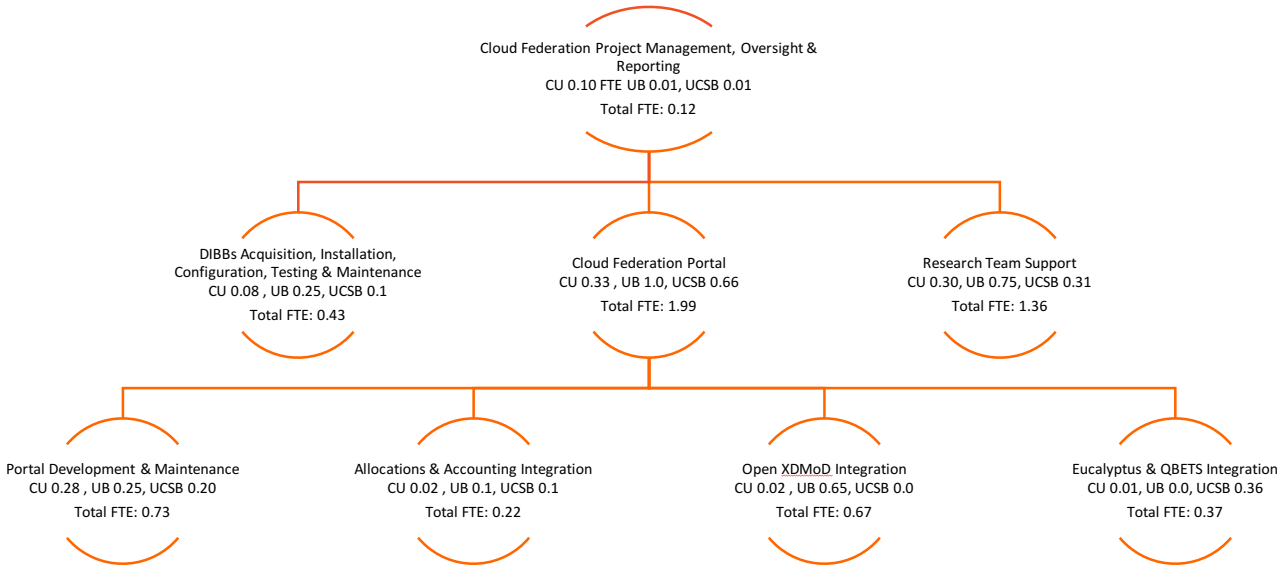
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 1: Quarterly Report 4

9/30/2016

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

This is the "Program Year 1: Quarterly Report 4" 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
A budget change request was approved by Amy Walton on 9/15/2016 for Hewlett Packard Enterprise (HPE) support. The agreement provides the Aristotle partners with an HPE professional maintenance offering, including support for SAN drivers. As indicated by the deep discounts HPE is offering, HPE is strongly interested in learning more about the NSF cyberinfrastructure community, and how company capabilities can better support this community.

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 had discussions with Amy Walton and Bob Chadduck about leading a DIBBs PI workshop. Cornell submitted a proposal on 9/27/2016 to lead a January 2017 DIBBs PI Workshop in Arlington, VA.

An Aristotle Executive Advisory Committee meeting occurred 9/20/2016:

- Attendees included Dmitrii Calzago (HPE), Ian Foster (ANL/U. Chicago), Steve Johnson (Weill Cornell Medicine), Sanjay Padhi (Amazon Web Services), Ben Rosen (Dell), Craig Stewart (Indiana U.), John Towns (XSEDE), Aristotle PIs and Co-PIs, and the Aristotle Portal, Infrastructure, and Science Use Case leads.
- Cornell CS (Zhiming Shen, Hakim Weatherspoon, and Robbert van Renesse) demonstrated Supercloud, a cloud architecture that enables application migration as a service across different cloud providers. Shen started a Supercloud instance and demonstrated how the user has full control over the location where the virtual machine runs with latency minimized by software defined networking. An instance was migrated from AWS to Google Compute Engine and between Red Cloud and AWS. Previous demos showed migration from Red Cloud to Jetstream and back. Discussions will continue on how this work might be further developed to bring value to the national community (John Towns expressed interest in this technology for XSEDE).
- Updates on the portal were provided, including plans for Aristotle to move to OAuth2 (the federated identity management used by XSEDE) which will be built into Eucalyptus 4.4 (Dmitrii Calzago, HPE, asked Cornell to alpha/beta OAuth2). The portal itself will eventually move as a package to AWS where it will be readily available as a template for any organization to use.
- The Infrastructure team explained that all 3 campus clouds are up with users running locally and one use case ran cross-site (at UB and CU) via the manual sharing of accounts. This will eventually be automated when all core functionality, i.e., allocations and accounting, are in place.
- The Science Use Case lead provided updates on the 7 use cases which are documented in monthly reports, and results to date and lessons learned are shared at quarterly Science Team Advisory Committee meetings.
- Wolski described how UCSB built a system for predicting the "bid price" that an AWS user should bid in the spot market to ensure a minimum duration of execution before AWS terminates the instance. Tests have shown substantial savings. This technology, called DrAFTS (Durability

#1541215

Agreements from Time Series), was developed as part of the Aristotle project. It uses QBETs (Quantile Bound Estimation Time Series) internally.

- Furlani's team is developing a federated Open XDMoD capability which requires re-engineering the Open XDMoD data warehouse. Cloud metrics will be added to Open XDMoD; QBETS services will integrate its predictive capabilities with other metrics generated by Open XDMoD (bandwidth, VM duration, storage load, etc.) in an attempt to generate statistical bounds on guaranteed delivered performance levels.

1.5 Status Calls

7/5/2016 call:

- Discussions regarding how long it should take for an instance store to fire up.
- Specific plans for portal additions are now available in a shared doc.
- Site-specific documentation should be placed in GitHub. Adam Brazier is the POC to provide your GitHub ID for access.
- Usage data graphs will be developed. CU-written scripts will collect the data and write to the database. Eventually, we will use the REST API.
- Discussions regarding how to manage instances across sites occurred. PCP will allow UB to collect information; sites will need to add it to their controller. Users won't have to do anything. We will inform them that this is running on their instances.

8/2/2016 call:

- Science Team Advisory Committee (STAC) meeting minutes were posted at https://federatedcloud.org/science/advisorycommittee.php. Amy Walton kicked off the meeting. Participation was high.
- Discussions regarding support implications if we decide to use Ceph as our storage platform.

8/16/2016 call:

- Discussions regarding how detailed usage graphs should be.
- Preparation for the 4.3 HPE Helion Eucalyptus upgrade. Cornell will take the lead.
- UCSB experiencing slow downs in speed between East and West coast. Cornell is checking on internal speeds and potentially NYSERNet speeds or other possible factors.
- UB plans to scale their geo use case on CU's Red Cloud using 32 and 64 cores.

9/13/2016 call:

- UB is working on getting usage data onto the interim Aristotle usage graph.
- CU is testing HPE Helion Eucalyptus 4.3 and will provide lessons learned to other sites. Upgrading to 4.3 is necessary to facilitate a smooth update to 4.4 which will feature OAuth2.
- CU continues to progress on using Docker to a create a cluster to run MPI on. Testing continues with WRF-Chem.
- Initial feedback from Sedgwick Reserve personnel on the precision agriculture project is they were shocked to learn they may have been over watering by a factor of 3; Sedgwick is currently working on getting equipment to automatically control the water flow.

9/26/2016 call:

- Infrastructure network testing continues. Plans to move to Eucalyptus 4.3 are underway.
- Back-end of the portal database is working (major achievement); the next step will be to integrate the other sites (OAuth2 will be required; anticipated late 2016 when Eucalyptus 4.4 is released).

#1541215

- The Aristotle annual report was completed and submitted on 9/15/2016.

1.6 Project Planning and Preparation
The project web site went live at https://federatedcloud.org on 6/30/2016.

All of these efforts are described in more detail in this quarter's report.

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

2.1 Federation Resource Status Updates
We're happy to report that all Aristotle clouds—Cornell Red Cloud, Buffalo Lake Effect, and UC Santa Barbara Aristotle—are now in production with researchers running at each of the three sites.

UCSB completed development of initial health/status tests, successfully stress tested the new installation, and Sedgwick researchers started to use the cloud. UCSB also configured Nagios (OMD) for monitoring; developed a provisioning process using LDAP Syncs "accounting groups" method and their Campus director; and, researched a backup solution for Ceph.

- **Log History**
  UB noticed that instance history is deleted from log files if the instance is started, stopped, and then deleted. We decided that this is not a problem because it will be rare for a researcher to delete their instance.

- **OAuth2 Support for Single Sign-In Credentials**
  HPE Helion Eucalyptus has announced support for OAuth2 in version 4.4 which is slated for release in later 2016. Eucalyptus 4.3 is out now. The biggest 4.3 release change is it requires CentOS 7. Cornell has a test cluster and will use it to test the 4.3 upgrade procedures and then share results with UB and UCSB. UB is setting up a small test cluster as well. Once at 4.3, UB and Cornell will alpha and beta test the OAuth2 code due to be released in v. 4.4.

- **Network Bandwidth**
  Further network testing between sites this month has provided more information but no conclusions. The sites will continue to analyze where the bottlenecks are (if any) and provide a report on the results at the conclusion of testing.

- **Ceph Storage**
  Cornell and UCSB discussed recent progress investigating Ceph. Cornell shared iozone results that compared reads and writes to Ceph vs. reads and write to the SAN. The Ceph numbers look great. UB's development cluster will deploy Ceph storage as well in order to test its features and get some Ceph experience.

- **Storage Performance**
  UCSB brought a reserved node controller (NC) online and reconfigured to a separate NC to use SSD as ephemeral storage for the Sedgwick users with the goal of improving performance.

#1541215

The infrastructure planning table was updated this quarter:

	Cornell (CU)	Buffalo (UB)	Santa Barbara (UCSB)
Cloud URL	https://euca4.cac.cornell.edu	https://console.ccr-cbls-2.ccr.buffalo.edu/	https://console.aristotle.ucsb.edu
Cloud Status	Production	Production	Production
Euca Version	4.2.2	4.2.2	4.2.2
Globus	Yes	Planned	Planned
InCommon	Yes	Yes	Yes
Hardware Vendor	Dell	Dell	Dell
# Cores	*168	**144	140
RAM/Core	4GB/6GB	up to 8GB	up to 9GB
Storage	SAN (226TB)	SAN (336TB)	Ceph (288TB)
10Gb Interconnect	Yes	10Gb inter-cluster; 1Gb external, 10Gb external planned	Yes
Largest Instance Type	28 core/192GB RAM	24 core/192GB RAM	16 core/16GB RAM
	* 168 additional cores augmenting the existing Red Cloud (376 total cores)	** 144 additional cores augmenting the existing Lake Effect Cloud (312 total cores)	

2.2 Potential Tools

- **CloudLaunch**
  The Cornell team continues to work on deploying a virtual cluster in Red Cloud with a generic compute node image for functional testing, including running sample jobs.

- **HPE Helion Eucalyptus**
  The HPE Eucalyptus team announced that they are close to having code in support of OAuth2 authentication ready for alpha testers. Cornell volunteered to do early testing on their test cluster.

- **Supercloud**
  A demo migrating a streaming video application between clouds (Aristotle's Red Cloud, AWS, Google Compute Engine, and Microsoft Azure) was provided to the Aristotle Executive Advisory Committee.

#1541215

2.3 Industry Influence

Cornell and HPE had discussions regarding HPE support for Globus Auth. The HPE development team has confirmed that OAuth 2 support will be in Eucalyptus 4.4.

3.0 Cloud Federation Portal Report

The new portal design went live on 6/30/2016 at https://federatedcloud.org.
There were content updates and additions this quarter to the portal. We added a basic usage graph (currently only Cornell's early usage is displayed) and links to user documentation from the developer's pages located at GitHub. To facilitate document links from the portal, a public repository was created to house the production documents. Documents under development or with sensitive project data are in a private repository. We are currently implementing a process to request and approve access to online reports by individual.

The portal planning table below was unchanged this quarter.

Portal Framework			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 10/2016**	**11/2016 - 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, 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 – 10/2016**	**11/2016 – 3/2017**	**4/2017 - End**
Cross-training expertise across the Aristotle team via calls and 1-2 day visits.	Hold 1 day training for local researchers. Offer Webinar for remote researchers. Use recording and materials to provide training	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

	asynchronously on the portal.		
User Authorization and Keys			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 1/2016	**2/2016 – 5/2016**	**6/2016 – 9/2016**	**10/2016 – End**
Plan how to achieve seamless login and key transfer from portal to Euca dashboard.	Login to the portal using InCommon.	Switch to Globus Auth in order to better interface with the Euca web console Get 4.2.1 federated key.	Move seamlessly to Euca console after portal 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		Test access to Euca console.
Allocations and Accounting			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**3/2016 – 8/2016**	**9/2016 – 12/2016**	**1/2017 – 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.	Implement project (account) creation in the database and display on the portal. Integration hooks for user and project creation/deletion and synchronization across sites.	Automate project (account) creation by researcher, via the portal.	Report on usage by account, if the researcher has multiple funding sources. Release database schema via GitHub.
Metrics and Usage			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 7/2016	**7/2016 – 9/2016**	**10/2016 – 12/2016**	**1/2017 - End**
Implement graphs of basic usage data, including % utilization, available resources, and user balance, using scripts from Cornell and U Buffalo for basic data collection.	Provide documentation for installing XDMoD and SUPReMM at individual sites. Install Open XDMoD/SUPReMM at individual sites and begin data collection. This includes the installation of SUPReMM and the data collection piece at the federation sites. Begin integration with federated authentication providers.	Federated data collection across sites. Ship data from the individual sites to UB. We can summarize data remotely and send the summarized data or collect all raw data and summarize locally. Other job information will be federated as well using the prototype model under development with OSG. Display federated metrics in Open XDMoD at UB.	Release materials via GitHub. Update periodically.

#1541215

3.1 Software Requirements & Portal Platform

Work on implementing Globus authentication was delayed early in the quarter due to a version problem; the widely-used league/oath2-client requires php 5.5 or higher, while 5.4.16 is the version provided with the currently available software stack.

A newer version of php (5.6) was installed in September so we can continue our work to implement Globus OAuth2 authentication. The instance for the main portal site was moved to a single-function project for security reasons. This will allow us to move forward with allocation and project management on the portal.

The next release of Eucalyptus, expected near the end of 2016, will allow us to incorporate OAuth support to facilitate seamless support from the portal to the Eucalyptus console.

3.2 Integrating Open XDMoD and QBETs into the Portal

Basic usage graphs will be made available on the portal to show basic data until Open XDMoD and QBETS is implemented late this year. The code provided by UB via GitHub has been implemented at Cornell; Buffalo and UC Santa Barbara will provide usage data using the same mechanism.

UB began data warehouse refactoring to support metrics for cloud and innovative resources. The existing XDMoD data warehouse was developed primarily to report on data generated by individual HPC jobs run on traditional HPC resources. With the advent of open source cloud solutions such as Eucalyptus and OpenStack, as well as non-traditional HPC resources such as Hadoop running alongside traditional HPC clusters at many centers, we re-examined the infrastructure used to store and report on center utilization as well as the definition of an HPC job within XDMoD. The capabilities of the XDMoD data warehouse will be updated to support these new resource types and become more flexible to better manage new types developed in the future.

A review of the existing data warehouse was completed to analyze the impact of any changes. We then completed benchmarking the existing data warehouse and we're currently working on benchmarking the proposed changes.

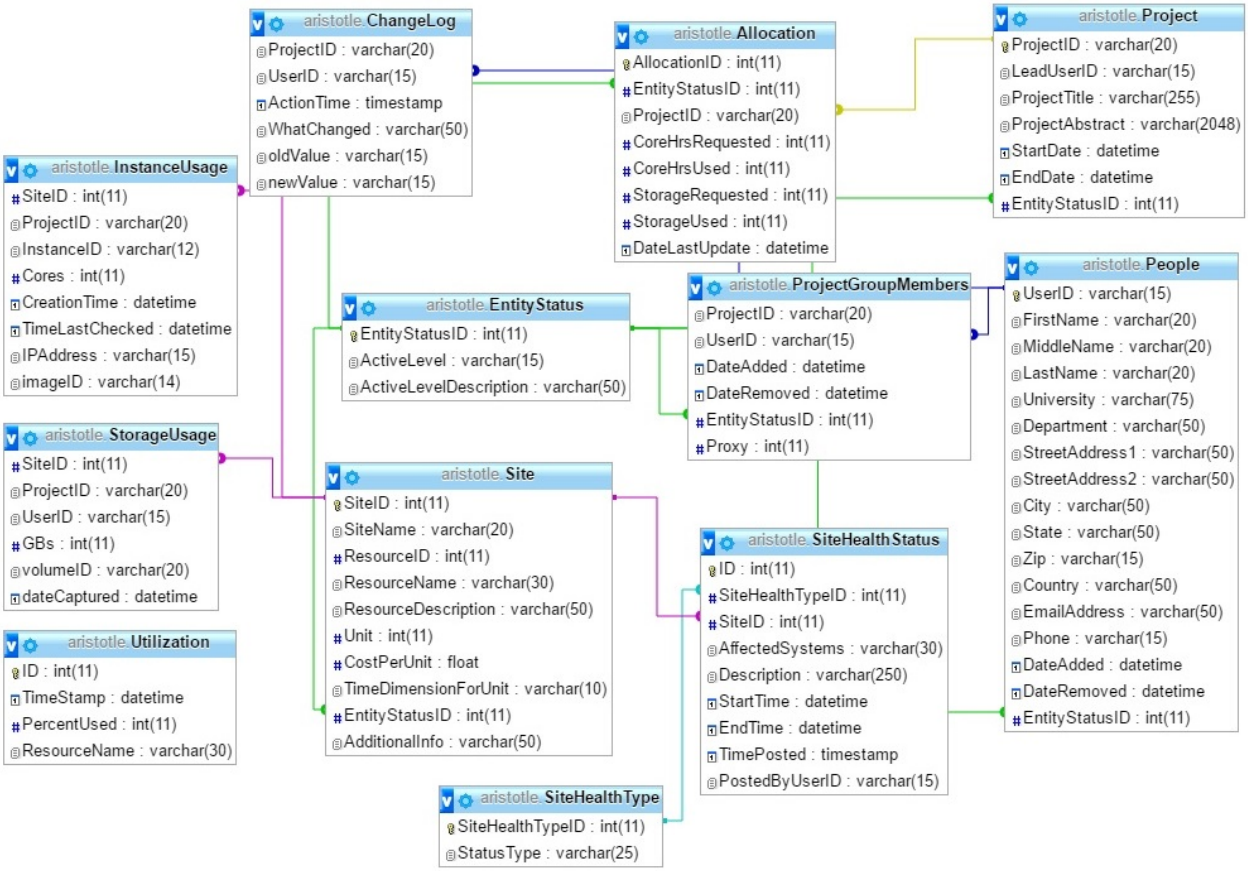
3.3 Allocations & Accounting

The usage graph mentioned in Section 3.0 is intended to show basic data until Open XDMoD and QBETS are implemented. As part of the graph creation, GetUtilization, a new stored procedure, was added to pull in the current usage data when the page is loaded.

The REST API code provided by UB via GitHub will be used by each site to share this data with the portal. All three sites are working on installing the API. UCSB continues work on the installation. The log ingestion glitch was resolved by better understanding of the UB code. Cornell continues to work on implementing the Aristotle Usage REST API for the Ithaca site (http://aristotle-usage.cac.cornell.edu). UB plans to provide this data after migrating existing HPC jobs.

There were no changes to the database schema (see page 10) this quarter.

#1541215



4.0 Research Team Support

4.1. General Update

- Brazier and Barker (CU) have been producing user documentation for Aristotle and some initial documentation has been made available via the portal. Additional Use Case level documentation is being produced and will be made public at a later date when the work is completed and vetted by the science teams as suitable for public release; this will, in particular, include advice for others about installation of domain-specific software stacks in a cloud environment and will also cover some difficult installations (such as the parallelized version of WRF-Chem) in a thorough way useful to the science community. This will be valuable to scientists who are planning to migrate their computational operations and/or data analyses to the cloud.
- The Aristotle Science Team Advisory Committee (STAC) quarterly meeting took place on 7/27/2016. Participation was high. STAC minutes were distributed and posted on the Aristotle portal. Brazier prepared Science Team goals for the next quarter.
- Work continues on MPI using Docker containers.
- All projects have agreed upon goals for the next project year.

#1541215

4.2 Science Use Case Team Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data
We created general purpose scripts to dynamically generate Spark clusters on the Aristotle cloud (UB "Lake Effect" cloud) and migrated the virtual machines. The Gaussian process-based change detection algorithm allows for identifying changes from observational or simulation data in a distributed fashion. We are currently studying the scalability of the algorithm on the cloud. The results will be presented at *BigSpatial 2016 – the 5th International Workshop on Analytics for Big Geospatial Data* workshop in early November. The HTML5-based interface to the underlying Spark cluster on the Aristotle cloud is ready to be used by the general public. We are finalizing access control mechanisms to audit the cloud usage by users. We developed a revamped browser-based interface (webglobe) and PI Chandola presented its capabilities at the *Advances in Virtual Globe Technology Using NASA World Wind, an Open-Source Geobrowser* meeting at MITRE, McLean, VA, 8/10-11. Chandola's UB student, Dinh Tran, created documents on how to create a Spark cluster in a cloud and how to port OpenNebula VM's to the Aristotle cloud. Scalability tests were run Red Cloud at CU before migrating the application to Lake Effect at UB.

Use Case 2: Global Market Efficiency Impact
The software required was installed and the researchers plan to begin initial test runs on the UB Lake Effect cloud using the TRTH database in the October/November time frame.

Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-Relevant Aerosol Properties
Paola Crippa tested the WRF-Chem installation running on a single instance and it passed tests. Brandon Barker is approaching this use case as the first target for the MPI/Docker project.

Use Case 4: Transient Detection in Radio Astronomy Search Data
Work continues on building a pipeline architecture and encoding, and how best to reduce the data by de-resolution in time and frequency. PALFA data was temporarily unavailable due to a machine failure (of a non-Aristotle file server); tens of terabytes of PALFA data are available though, and more are coming in on another server. Brazier met with Robert Wharton and Shami Chatterjee to discuss the project and characterize the data. Data reduction code is being prepared.

Use Case 5: Water Resource Management Using OpenMORDM
We are making plans for building the software stack and testing it. This project will be a candidate for the work Barker is conducting on implanting MPI in Aristotle.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota
Barker met with one of the project postdocs to discuss and decide on possible algorithms to use for simulating symbiont metabolism. The research team created results using Windows instances developed by the Science Use Case team at Cornell and running on Red Cloud. Nana Y.D. Ankrah (Cornell) reported that with metabolic modeling, we have been able to show that (1) the metabolism of whitefly symbionts are co-evolved to minimize the overlap of inputs (competition) and outputs (efficiency), (2) differences in the metabolic inputs from the whitefly bacteriocyte drive the difference in EAA released from symbionts although the relationship between substrate and product concentration is not always linear for all EAAs, and (3) the whitefly bacteriocyte acts as a sink for NH3 and facilitates symbiont N recycling for EAA production through shared biosynthetic pathways.

#1541215

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

- *Drought Ecology of the California Live Oaks*: Last quarter, Sedgwick initiated an experiment to investigate the use of slow drip "water boxes" to help seedling development for California Live Oaks. There is little data on how drought conditions affect slow-growing trees in California, like the Live Oak, especially during the seedling stage. The project installed WATERMARK sensors in each of three test seedlings to measure "slow drip" effects from water boxes; however, the sensor control platform (new for this project) is not yet functional. Initial, short-term tests show good results, but the robustness necessary for a long-term study is not yet possible. The team is actively developing the reliability engineering required. The movement of data from the sensor to the UCSB Aristotle cloud for back-end processing is functional for collecting moisture sensor data to improve seedling development of CA Live Oaks; testing continues.

- *Precision Agriculture:* The goal of this work is to develop an Aristotle-based notification system that alerts Sedgwick grape vineyard personnel when irrigation is necessary and the amount of water to use. Moisture sensor analysis in the vineyard indicates that the grapes are receiving too much water during a wetting event. We are working to understand how much less water we can use and with what frequency. Early analysis both of the soil moisture transfer rate and the irrigation system has resulted in two unsuccessful watering tests (that delivered too little water).

  Using the sensing technology that is in place, vineyard personnel have started doing irrigation scheduling in August. Part of the current issue is that they do not know what the "drain rate" is for the vineyard under study. That is, the sensors are generating an alert for when to put water on the grapes, but the current soil content map generated by the agronomists seems to be making inaccurate predictions of when to turn the water off. The project (a collaboration with Fresno State) is looking at directly correlating Electrical Conductivity (EC) with the different soil types in the vineyard. With this correlation, it should be possible to create alerts for both water start and shutoff. Looking ahead, the goal is to "close the loop," i.e., to make the process completely automated.

- *Wildlife Survey:* A grad student and undergraduate student developed the animal-identification system that runs Google's TensorFlow in parallel on Aristotle to auto-identify animals from camera-trap images (200,000 photos/month). The project tested Caffe in July and then switched to Google's TensorFlow machine learning system in August with better results. Using an open database of animal images to train a neural network, the students developed the software framework necessary to process the camera-trap data for Sedgwick so that specific animal species can be identified automatically. This framework is running on the new Aristotle cloud system. Currently, the camera traps take approximately 200,000 photos per month that must be manually classified. Aristotle has already enabled the development of software to identify images that were triggered by non-animal movement (e.g., wind blowing leaves in front of the motion detector). This new system will be able to provide additional classification indicating the likely species of animal

  As part of this effort, the student developed an OCR (optical character recognition) system to read the camera meta-data from the images. The Sedgwick camera traps print the meta-data on the images themselves (they do not have an alternative metadata store). Again, using machine-learning, the student's system (migrating to the new Aristotle hardware) recovers the meta data so it can be used to index the images.

NSF
#1541215

While TensorFlow provides a better success rate for camera trap pictures using an existing animal imagery database, the rate is not high. Part of the problem is that there are relatively few pictures of some species of animals. For example, in almost 200,000 images from July, there are only 6 pictures of bears (although many pictures of deer). The project is now looking at "mashing up" images from Google's image search with backgrounds taken by the camera traps. That is, by superimposing images of different animals from Google on the fixed and empty background taken from the camera traps, the hope is that the training data will become more effective.

Finally, the transfer of Sedgwick to eMammal (https://emammal.si.edu) continues. However, the estimated time for one-month's worth of image transfer is approximately 14 days because of the repository write speed. The repository is currently hosted in Box. The project attempted to use AWS S3 as an alternative and the speed increase was insignificant. The current plan is to use Aristotle as the alternative data repository (eMammal has an S3 interface and can interact with Aristotle without modification). The speed increase may be as high as two orders of magnitude.

5.0 Outreach Activities

5.1 Community Outreach

- Tom Furlani presented federated XDMoD for Aristotle plans at the *XSEDE16 BOF—Coin of the Realm: Current Practices and Future Opportunities in Processing XSEDE Allocation Awards & Usage Data.*

- Rich Wolski et al. wrote a technical report on the *Probabilistic Guarantees of Execution Duration for Amazon Spot Instances* that is available at the Aristotle portal: https://federatedcloud.org/about/publications.php.

- A Securities and Exchange Commission Branch Chief FCW presentation on *Challenges and Opportunities in the Cloud* highlighted the NSF Aristotle Cloud Federation as "what's next" in cloud computing: http://www.digitalgovernment.com/media/Downloads/asset_upload_file34_5802.pdf.

- The Mapping Transcriptome Data to Metabolic Models of Gut Microbiota use case researchers presented findings at an American Society for Microbiology (ASM) conference (http://conferences.asm.org/images/1905%20asm%20beneficial%20microbes%20web2.pdf) that were based on results from using a Windows instance in Aristotle: Ankrah, N.Y.D., Luan, J. & Douglas, A. Evolution of Metabolite Exchange in a Three-Partner Symbiosis. *6th ASM Conference on Beneficial Microbes*, September 10, 2016, Seattle, WA. The Aristotle team configured a Windows instance with MATLAB and associated software (COBRA Toolbox, Gurobi, etc.) to support this research.