ARISTOTLE
CLOUD FEDERATION

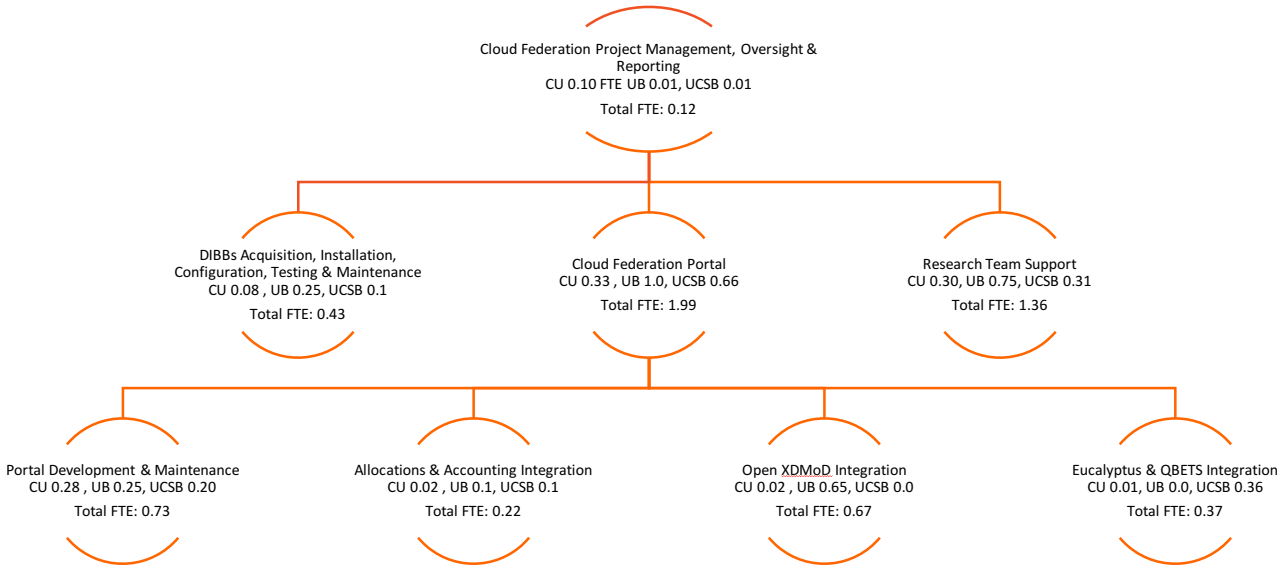
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Monthly Report 12/21/2016

Report 15 of 18

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

This is the fifteenth required monthly report of the Aristotle Cloud Federation team. We report on plans and activities for each area of the project Work Breakdown Structure (WBS).



Contents

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 month.

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 is preparing to Chair the 1st NSF Data Infrastructure Building Blocks PI Workshop (DIBBs17) scheduled for January 11-12, 2017 in Arlington, VA.

1.5 Status Calls
12/20/2016 project status call topics:

- Discussion of how Cornell is testing and positively impacting the rollout of Eucalyptus 4.4 with OAuth by suggesting improvements to the user interface and identifying bugs.
- Identification of next steps in identifying where the network bottleneck is between New York and California (possibly Tulsa area).
- Plans to have a Globus endpoint at UB.

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

2.1 Hardware Acquisition

- UB's 10G network equipment arrived for their public Ethernet connection. This should be installed, tested and in production by the end of December.
- UB received quotes for node controllers from Dell, HPE, and Ace Computer. Ace Computer has the lowest bid by quite a margin (2K less per server). These are Supermicro boxes and are identically configured including 5-year NBD on-site warranty. UB has no experience with the vendor, but talked with them at SC16.
- Cornell is working with Dell, HPE, and Red Barn (local Supermicro reseller) revising year 2 hardware quotes. Quotes should be finalized by the end of December.
- UCSB is reviewing quotes for year 2 hardware from Dell and IRON Systems; a quote from HPE is expected soon. UCSB is also waiting on facilities improvements for deployment of an HPE equipment donation into their cloud infrastructure (indirectly affects Aristotle).

2.2 Software Installation, Configuration, and Testing

- UCSB spent time troubleshooting a network delay which happens intermittently when some new instances are started. The delays cause instances not to be able get user data from CLC and the SSH key injection fails.

- Cornell upgraded a test cloud to HPE Helion Eucalyptus version 4.4 nightly builds and worked with Eucalyptus engineers to successfully test login via Globus Auth 2.0 on the Eucalyptus Console. Regression testing was performed to make sure Eucalyptus version 4.4 nightly builds work without major issues. Cornell raised two issues with the Eucalyptus engineers:
    - Two login buttons (one for Eucalyptus local cloud login and one for Globus Auth 2.0 login) on the same screen are confusing as presented,
    - An outstanding bug in version 4.3 (EUCA-12799) was found that needs to be fixed in version 4.4 before that version can be used in production.
- The Cornell infrastructure and web portal teams met for a quick demo of Eucalyptus 4.4 using Globus Auth 2.0. Discussions followed regarding how to provision and remove users via role policies and how we plan to use Globus Auth APIs to implement cloud federation.
- Based on feedback from UCSB, the Cornell team is working closely with the Cornell Network staff to test and investigate the bandwidth between Cornell and UCSB. Upon completion, a full report with findings will be distributed to all three sites.

There were no updates to the infrastructure planning table this month:

	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	4.3	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)	

2.4 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**

  Cornell, working closely with the HPE Helion Eucalyptus engineers, has successfully logged into the Eucalyptus Console using Globus Auth 2.0. HPE's support for OAuth 2 will be critical for implementing the Aristotle Cloud Federation.

- **Supercloud**

  Nothing new to report this month.
  .

3.0 Cloud Federation Portal Report

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

The usage graph https://federatedcloud.org/using/federationstatus.php was completed last month; it shows basic early usage data from all 3 sites. For ease of conformity between federated sites, we have requested that the REST API code provided by UB via GitHub be modified to report time in UTC/GMT. The UB team investigated why the timestamp field in the cloud data is reporting local time and found that the script was using the UNIX date command without the -u option for UTC. This has been updated in GitHub and some of the web service code has been refactored, redesigned, and improved. An SQL script has been written to update the times correctly in the database at the 3 sites. Currently, this is in test at UB, and is expected to be implemented early in January 2017.

There were no changes to the portal planning table this month:

Portal Framework			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 12/2016**	**1/2017 - End**	**1/2017 - End**
Gather portal requirements, including software requirements, metrics, allocations, and accounting. Install web site software.	Implement content/functionality as shown in following sections. Add page hit tracking with Google Analytics, as well as writing any site downloads to the database.	Implement content/functionality as shown in following sections. Add additional information/tools as needed, such as selecting where to run based on software/hardware needs and availability.	Release portal template via GitHub. Update periodically.

Documentation			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**4/2016 – 10/2016**	**11/2016 – End**	**1/2017 - End**
Basic user docs, focused on getting started. Draw from existing materials. Available through CU doc pages.	Update materials to be federation-specific and move to portal access.	Add more advanced topics as needed, 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 1-2 day visits.	Hold 1 day training for local researchers. Offer Webinar for remote researchers. Use recording/materials to provide asynchronous training on the portal.	Add more advanced topics as needed. Check and update materials periodically, based on training feedback and new functionality.	Release training materials via GitHub. Update periodically.
User Authorization and Keys			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 1/2016	**2/2016 – 5/2016**	**6/2016 – 3/2017**	**1/2017 – 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.	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	N/A	N/A
Allocations and Accounting			
Phase 1	**Phase 2**	**Phase 3**	**Phase 4**
10/2015 – 3/2016	**3/2016 –3/2017**	**3/2017 – 6/2017**	**6/2017 – End**
Plan requirements and use cases for allocations and account data collection across the federation. Design database schema for Users, Projects and collections of CPU usage and Storage Usage of the federated cloud.	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 – 3/2017**	**10/2017 - End**
Buffalo team utilize Cornell scripts to design a REST API for basic cloud data and deploy at 3 sites and publish usage data to project portal (completed). Buffalo currently standardizing the API by using UTC across the sites and refactoring the code efficiency. Buffalo also completed a redesign of the XDMoD data warehouse to support cloud metrics and is moving this into the testing phase.	Provide documentation for installing XDMoD and SUPReMM at individual sites. Install Open XDMoD/SUPReMM at individual sites and begin data collection. This includes installation of SUPReMM and the data collection piece at the federation sites. Begin integration with federated authentication providers. Currently waiting for latest release of Open XDMoD (v.6.5.1) which will be available at year end at http://open.xdmod.org/. Note: this version does not support cloud metrics but will give sites an opportunity to get infrastructure in place for a future version that does.	Federated data collection will ship data from XDMoD instances at the individual sites to a master XDMoD instance at UB where overall cloud data will be displayed. This is in alpha testing at UB with completion planned for 3/2017.	A prototype cloud realm using Euca data is planned for 10/2017. When completed, federated data from all 3 sites will be available at the master XDMoD instance. Release materials via GitHub. Update periodically.

3.1 Software Requirements & Portal Platform

Continued work on implementing Globus OAuth2 authentication is covered in the previous infrastructure section. The next release of Eucalyptus is expected near the end of 2016 which will allow us to incorporate OAuth to facilitate seamless support from the portal to the Eucalyptus console. The Cornell team has been working with the HPE team on testing Globus Auth support for the Euca Console in the upcoming Eucalyptus 4.4 release. Testing and debugging on nightly builds are ongoing on a test cloud at Cornell.

3.2 Integrating Open XDMoD and QBETs into the Portal

UB continues work on benchmarking proposed changes.

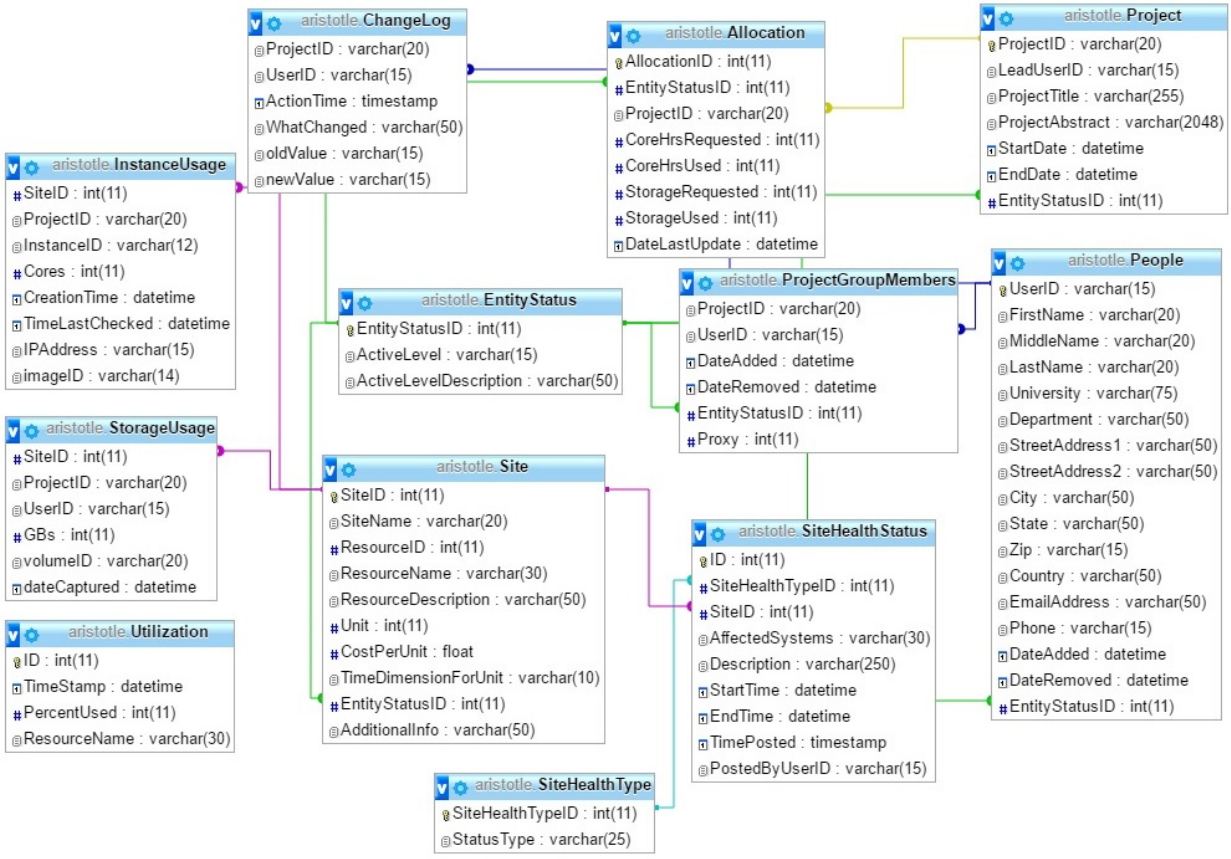
UB expects to implement Open XDMoD and QBETS in the first quarter 2017.

3.3 Allocations & Accounting

Development on accounting and allocations is proceeding. The database and tables with test data are complete, and interface implementation is starting. Currently, we are working on php scripts to import

project usage data from UB and UCSB using the REST API and adding the data to the database. We are also working on developing Stored Procedures for collating and reporting usage data.

There were no changes to the database schema this month:



4.0 Research Team Support

4.1. General Update

- An allocation request to Jetstream for 2.4 million SUs (1 SU = 1 vCPU-hour) was approved for work extending Aristotle Use Cases with increased scale.
- MPI/Docker work continues with substantial progress. Bennett Wineholt (Cornell systems team), working with Brandon Barker, has successfully executed basic multi host MPI commands using sample code and is working on making the process more robust.

4.2 Science Use Case Team Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data
The UB science team continues to develop the interface for running sustainability-related codes on climate simulation data using the Lake Effect cloud. The key issue is engineering an interface which can account for resource usage for the users. The web-based interface is nearly ready and should be available to Aristotle Cloud Federation researchers in two weeks. Eventually, we will be releasing it to the general scientific community.

Use Case 2: Global Market Efficiency Impact
UB researchers setup the analysis framework on the VM and started using the VM to analyze the finance data. So far, we have experienced no roadblocks.

Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-Relevant Aerosol Properties
Tristan Shepherd (Pryor group postdoc) starting working on the basic image and is discussing use case tests with Brandon Barker. MPI work continues as described above.

Use Case 4: Transient Detection in Radio Astronomy Search Data
No new updates this month.

Use Case 5: Water Resource Management Using OpenMORDM
Brandon Barker has handed off an initial VM, container, and container source (Docker file) for OpenMORDM for Dave Hadka to test. MPI work continues with a target of 8 8-core nodes.

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota
A manuscript is still being prepared by Cornell's Nana Ankrah, including reference to simulation data generated on Aristotle.

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production & Security
Agricultural Food Security Project: For the irrigation and water study projects (oaks and grapes, respectively) the team is using the irrigation hiatus to rework the micro-controller platform. The Xbee radios appear to be attenuated by the surrounding foliage. To compensate, the radios are being fit with one range antennas that can be raised on an aerial. The first test of the system looks promising. A full deployment is scheduled for this week.

The team has also begun looking at new clustering algorithms for the EC (electroconductivity) data gathered by via the Veris device. They have developed a time-series based clustering methodology (that uses QBETS for change point identification). They have also obtained a copy of the MZA software. MZA (Management Zone Analysis) uses fuzzy clustering to determine soil management zones, with a plan to compare the time series clustering, k-means, and MZA in terms of clustering results. The first comparison will be based on information-theoretic scoring (e.g. AIC). However, the team has ordered a SoilCares scanner device for taking soil samples at scale and plans to use it to determine ground truth at Sedgwick. The SoilCares device will be shipping in the early spring, hopefully, in time to modify the grape irrigation schedule.

5.0 Outreach Activities

5.1 Community Outreach

- 61 PIs and co-PIs have registered for the 1st NSF Data Infrastructure Building Blocks PI Workshop (DIBBs17) led and chaired by Cornell. See https://dibbs17.org for details on the Jan. 11-12 event being held at the Westin Arlington in Arlington, VA. A news release is also available at https://www.cac.cornell.edu/about/news/161209.aspx.