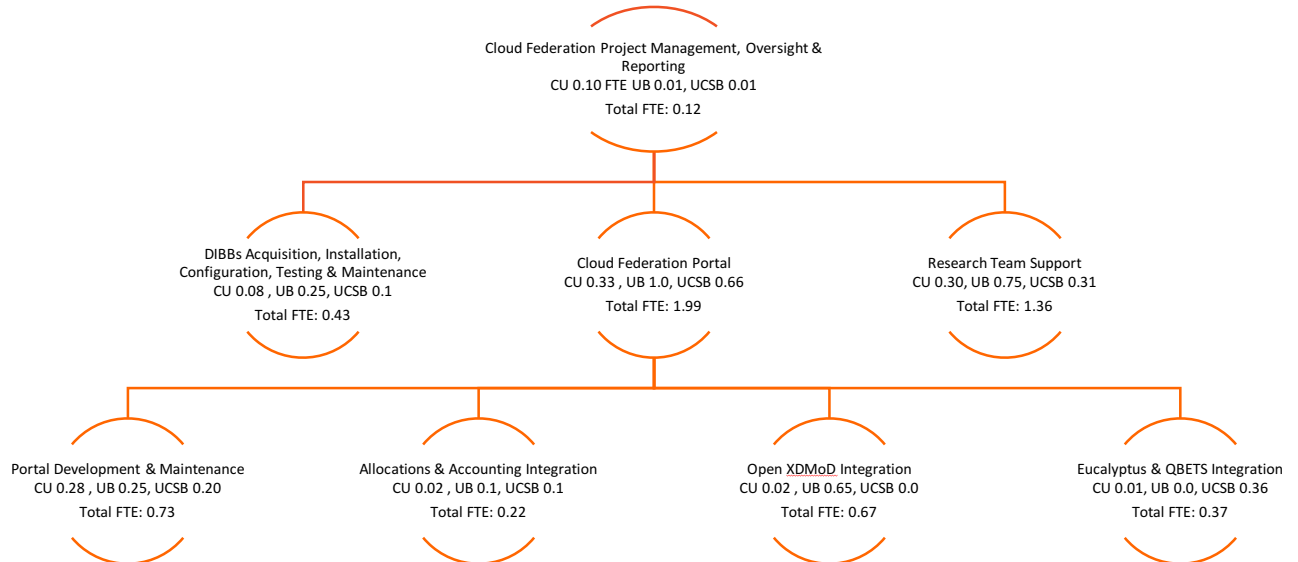# CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Monthly Report 1/29/2016

Report 4 of 18

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

This is the fourth 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).

Cloud Federation Project Management, Oversight & Reporting
CU 0.10 FTE UB 0.01, UCSB 0.01
Total FTE: 0.12

DIBBs Acquisition, Installation, Configuration, Testing & Maintenance
CU 0.08 , UB 0.25, UCSB 0.1
Total FTE: 0.43

Cloud Federation Portal
CU 0.33 , UB 1.0, UCSB 0.66
Total FTE: 1.99

Research Team Support
CU 0.30, UB 0.75, UCSB 0.31
Total FTE: 1.36

Portal Development & Maintenance
CU 0.28 , UB 0.25, UCSB 0.20
Total FTE: 0.73

Allocations & Accounting Integration
CU 0.02 , UB 0.1, UCSB 0.1
Total FTE: 0.22

Open XDMoD Integration
CU 0.02 , UB 0.65, UCSB 0.0
Total FTE: 0.67

Eucalyptus & QBETS Integration
CU 0.01, UB 0.0, UCSB 0.36
Total FTE: 0.37

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
As follow-up to Lifka's Cycle Computing meetings, CU sent CloudLaunch to Cycle Computing so that they can review it and consider a partnership with AWS to have CloudLaunch queues available to the community on AWS resources. We expect to continue these conversations in February 2016.

Lifka had a call with the AWS Scientific Computing (SciCo) team supporting Cornell and Aristotle. There continues to be great interest in our project and in broader interactions with the NSF HPC community. A project update was given and we discussed how we might work together in the coming months. We also discussed the potential for allocating AWS time to XSEDE and leveraging CloudLaunch with Cycle Computing and Supercloud from Cornell. These potential activities may facilitate making the transition to cloud easier for the science community.

Lifka arranged a call with George Turner, an IU Jetstream team lead, and Cornell Professor Hakim Weatherspoon (http://www.cs.cornell.edu/~hweather), Research Scientist Robbert van Renesse, (http://www.cs.cornell.edu/Info/People/rvr/) and their team about partnering with Cornell CS on their Supercloud project which is jointly funded by NSF CISE (program director Amy Apon) and NIST. This technology has the potential to make it easy to burst between Aristotle and other clouds including Jetstream, Chameleon, CloudLab, and the major public cloud providers. It removes the burden of requiring researchers to have different virtual machines for each cloud software stack. The meeting went very well and there is definite interest on the part of the Jetstream team. IU will provide Weatherspoon and van Renesse's team with Jetstream accounts for testing. Once testing is complete, the Cornell CS team will do a live demonstration of Supercloud bursting to and from Aristotle resources for the Jetstream team. Assuming a successful demonstration, Cornell CS will then give a demonstration to the Aristotle PIs and project leads. Rich Wolski's depth of cloud technology knowledge will really help us understand the benefits and limitations of Supercloud. Lifka also discussed this technology with XSEDE PI John Towns who is also extremely interested. After our in depth evaluation of Supercloud, we hope to provide a demonstration to our External Advisory Committee to better understand how we might move this technology from the lab into the hands of the national community. We will continue to report on this effort.

1.5 Status Call with NSF Program Manager
Our monthly status call with NSF program director Amy Walton was held on 1/11/2016. Topics of discussion included:
- Aristotle project may have a pathfinder role in helping NSF map out what roles clouds (NSF, public, etc.) might play in future NSF cyberinfrastructure.

- Possibility of a NSF DIBBs workshop in conjunction with the fall 2016 CASC meeting.
- Potential NSF reviewers for next round of DIBBs proposals: Dave Lifka or Susan Mehringer.
- Hewlett-Packard Enterprise (HPE) Helion Eucalyptus team considering building InCommon support into the Eucalyptus portal.
- Key project deliverables to keep in mind: Can others pick up Aristotle components and use/customize them? Are lessons learned documented?
- Collection of cloud metrics will be added to Open XDMoD in CY2016.
- A new version of allocations and accounting database schema was developed last month.
- We are starting to get science teams prepared, e.g., Brazier got WRF-Chem atmospheric chemistry observation and modeling software running on a Red Cloud virtual machine, we started to identify queries that will stress astronomy data in new ways, etc. Some use case scientists know how to make their codes run in the cloud; others don't know what their workflows are. We will take an agile approach. Understanding how to easily containerize codes will be an important lesson learned.
- Amy Walton pleased with quality of Aristotle project reports.

1.6 Project Planning and Preparation

Project planning and preparation by the Aristotle team continued in January 2016. Requirements analysis for the allocations and accounting system included how using Eucalyptus availability zones will allow us to partition our local cloud installations between local university users and Aristotle-funded users. We discussed and gained a better understanding of Eucalyptus provided federation capabilities and how we might leverage them. CU and UB placed orders for their first year hardware installment from Dell and are preparing for installation and testing. UCSB is still negotiating their hardware purchase with Dell, HPE, and other vendors. There was also continued User Portal planning and development, and ongoing work with the science teams. All of these efforts are described in more detail in this monthly report.

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

2.1 Federation Resource Status Updates

- **CU**
  The CU team is awaiting the arrival of our first year's Aristotle hardware. The result will add 168 cores (6GB/core) and 120TB SAN storage to the Red Cloud infrastructure.

- **UB**
  The UB team is busy cabling their new Aristotle servers, storage, and network switch. The design and layout for the UB Aristotle cloud is underway; this planning will aid UB in the installation and configuration of the cloud software. In addition, UB will soon upgrade the Center for Computational Research's (CCR) existing cloud infrastructure to HPE Helion Eucalyptus version 4.2.1. This upgrade experience will be beneficial in the installation of the UB Aristotle cloud.

- **UCSB**
  The UCSB team continues to negotiate with HPE and sent RFPs to other vendors too. UCSB is investigating a variety of storage and compute configurations.

The three sites briefly discussed using HPE Helion Eucalyptus availability zones during our monthly meeting. We plan to deep dive into this topic in February 2016.

The CU/UB/UCSB infrastructure planning table has been updated below:

	CU	UB	UCSB
Cloud URL	euca4.cac.cornell.edu	ccr-cbls-2.ccr.buffalo.edu	TBD
EUCA Version	4.2.1	4.2.1	4.1.2
Migrate to 4.2.1	X	X	3/1/2016
Globus	Yes	Not currently, but is planned.	?
InCommon	Yes	Not currently, but is planned.	Planned
Hardware Quotes	Quotes processed.	Processed. New equipment delivered.	Under negotiations.
Hardware Vendor	Dell	Dell	TBD
# Cores	168	112-140 (target)	TBD
Ram/Core	4GB/6GB/8GB	6GB	TBD
10Gb Interconnect	Yes	Yes	Yes

2.2 Industry Influence: Eucalyptus

This month, the CU team had a couple of meetings with HPE's Chris Grzegorczyk to discuss using InCommon with HPE Helion Eucalyptus. Specifically, we wanted to share Aristotle requirements for InCommon integration with Chris so that he could in turn let us know how HPE might be able to assist.

Chris offered two possibilities:
1) Add InCommon support to HPE Helion Eucalyptus' current cloud federation so that a user could log in via InCommon credentials and run on any cloud in the federation using the same account ID, the same keys, and subject to the same IAM policies;
Or,
2) Allow users to log into the web console using their InCommon credentials. The user's account and permissions will need to be administered separately in each cloud.

Given the decentralization of infrastructure in academe, we chose the second approach. HPE Helion Eucalyptus will add InCommon login support on the web console and create tools for cloud administrators to associate InCommon DNs to HPE Helion Eucalyptus users. Chris will send the Aristotle team a description of this feature for review and further discussion.

2.3 Potential Tools: CloudLaunch & Supercloud

Steven Lee of the CU infrastructure team packaged and sent the CU CloudLaunch (https://www.cac.cornell.edu/technologies/CloudLaunch.pdf) software stack to Cycle Computing for evaluation by their engineers.

The CU infrastructure team also met with Cornell's Supercloud developers and Indiana's Jetstream team to facilitate future testing of Supercloud, a Cornell CS tool that enables users to easily run images across cloud platforms. Supercloud's Xen-Blanket wrappers allow images created under HPE Helion Eucalyptus to run on an OpenStack cloud, AWS, Azure, etc.

Indiana will provide Cornell access to Jetstream for a Supercloud proof of concept.

3.0 Cloud Federation Portal Report

The format of the portal planning table (below) was modified this month to allow much more detail to be added. Dates were also updated to reflect gating factors and additional process steps.

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.	Update materials to be federation-specific.	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 asynchronously 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 – 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.	Get 4.2.1 federated key after InCommon login.	Move seamlessly to Euca console after portal InCommon login.

6

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.	Implement minimal set of Euca Tools to bridge portal to Euca console.	Add/refine/update, based on ongoing collection of user feedback.	Release via GitHub. Update periodically.

Allocations and Accounting			
Phase 1	Phase 2	Phase 3	Phase 4
10/2015 – 3/2016	3/2016 – 5/2016	6/2016 – 9/2016	10/2016 – 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 – 3/2016	2/2016 – 5/2016	6/2016 – 10/2016	11/2016 - End
Implement graphs of basic usage data, including % utilization, available resources, and user balance, using scripts from Cornell CAC 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.

3.1 Software Requirements & Portal Platform

Last month the basic platform was built on a cloud instance, running Centos7 with the LAMP stack installed, and populated with a "Coming Soon" page at http://www.federatedcloud.org. This month an SSL certificate was requested for the site, but it hasn't been received yet and applied.

### 3.2 Integrating QBETS into Open XDMoD

On 1/11/2016 Steven Gallo, Joseph White, and Robert DeLeon (UB) had a conference call with Rich Wolski (USCB) on integrating QBETS into XDMoD and Aristotle. We discussed how QBETS works,

how it could be combined with data from XDMoD, and potential uses for it in modeling Aristotle usage, e.g., predicting the value of sharing between different clouds within the federation and bursting to AWS. Wolski subsequently prepared a stand-alone version of QBETS with documentation and test data and put it in a GitHub repository accessible to UB personnel. UB is currently preparing to run QBETS test cases with HPC usage data to simulate cloud usage. Further discussion on the integration of QBETS into XDMoD will happen at a later date.

3.3 Allocations & Accounting
This month the team began biweekly meetings to flesh out use cases for allocations across both the federation and different funded projects. A researcher may fall into a very simple case of using resources from one allocation at one site, while another may draw from several funding sources and work on multiple nodes in the federation. Our goals are to define all use cases, then define requirements for users, data collection and reporting, and finally, design a phased implementation plan.

No changes were made this month to the allocations and accounting database schema below:
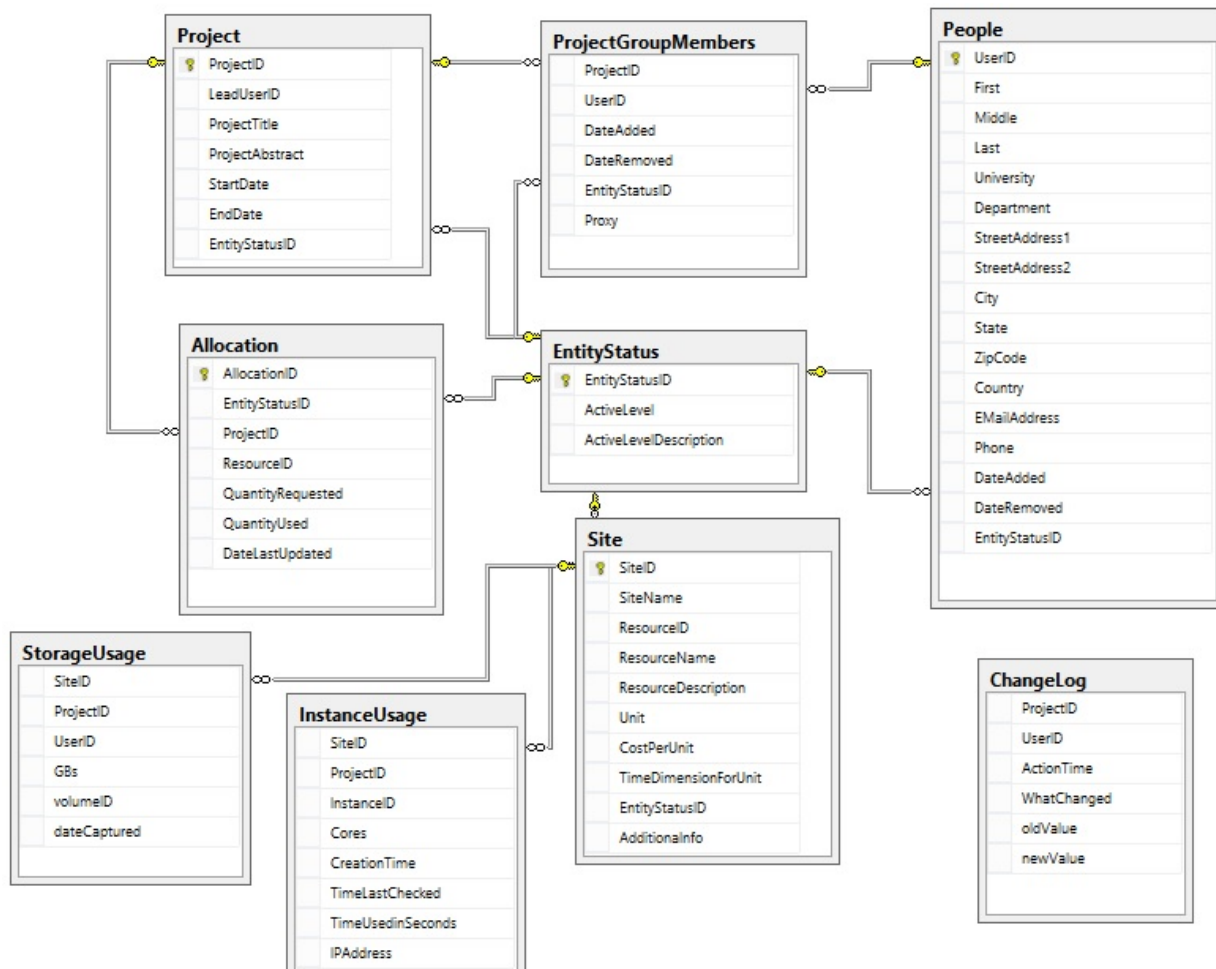
4.0 Research Team Support

4.1 General Support Packages
We are compiling a list of packages of general utility for a standard base image, package list, or scripted installation. This work is in progress and subject to review; so far, the list includes:

wget	mlocate
gcc-fortran	csh
tcsh	perl
time	redhat-lsb
redhat-lsb.i686	compat-libstdc++-33.x86_64
compat-libstdc++-33.i686	

Additionally, a brief list of optional items, such as X11, will be produced to help users who may be more familiar with desktop Linux installations which are more software-rich.

4.2 Science Use Case Updates

Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data and Use Case 2: Global Market Efficiency Impact
UB science team lead, Varun Chandola, is currently installing, testing and running his VM environment on the existing UB CCR cloud and will migrate to the UB Aristotle cloud as it becomes available. This will allow UB use case scientists to begin leveraging the Aristotle cloud almost immediately.

Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-Relevant Aerosol Properties
Sara Pryor's graduate student received training on WRF-Chem and will assist in the installation of the distributed-memory version of WRF-Chem.

Use Case 4: Transient Detection in Radio Astronomy Search Data
Work has begun on creating an instance, with necessary software installed, for the astronomy search data use case.

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production
This month's work on the UCSB-led use case included:
- Installed the IRROMETER soil moisture monitoring system and currently integrating data acquisition with Aristotle.
- Recruiting a student researcher for the "Where's the Bear?" photo image analysis project.
- Presentation and recruiting opportunity presented at freshmen seminar on ecological science and sustainability (Jan. 12, 2016).
- IRROMETER data acquisition not functional yet. January 15, 2016 debugging field trip unsuccessful; scheduled vendor to provide on-site service on February 3, 2016.
- On-site data acquisition "weigh station" hardware is on order from Sedgwick.

5.0 Outreach Activities

5.1 Media Outreach
A January 9, 2016 *Cloudhostingzine* article on "Cloud Computing for Academic Research" featured the Aristotle Cloud Federation.

Cornell contacted the Editor of *Scientific Computing World* and pitched including federated clouds as part of a February 2016 forthcoming feature on "Barriers to the Cloud in HPC." Subsequently, Lifka was interviewed by Editor Tom Wilkie and it is anticipated that the Aristotle Cloud Federation will be included in the article.

5.2 HPE Helion Eucalyptus Outreach
Cornell had several calls with Colby Dyess, HPE cloud marketing director, and will be working with the HPE cloud referencing team (Sean Garcia and Tracy Roberts) to capture the Aristotle Cloud Federation story and lay the groundwork for future outreach activities, e.g., case studies, blog posts, etc. Cornell also introduced the HPE Eucalyptus manager to the AWS Scientific Computing manager to help build a federated cloud ecosystem.