ARISTOTLE
CLOUD FEDERATION

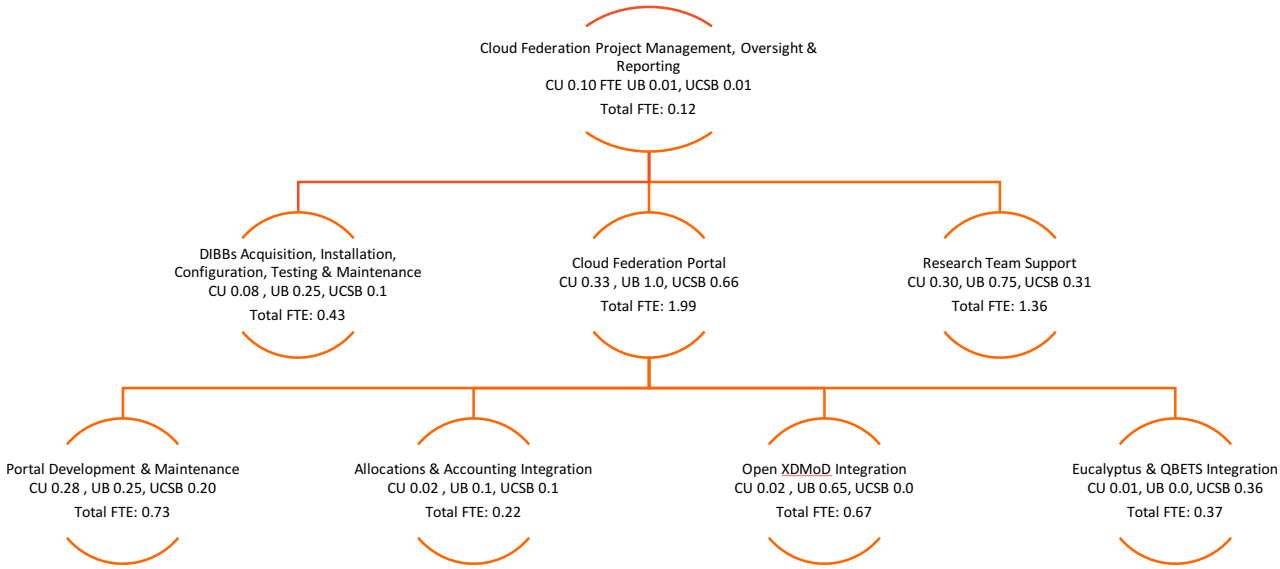
CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation

Program Year 1: Quarterly Report 1

12/22/2015

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

This is the first required quarterly report for Program Year 1 of the Aristotle Cloud Federation team. We report on plans and activities for each area of the project Work Breakdown Structure (WBS).



#1541215

#1541215

1.0 Cloud Federation Project Management, Oversight & Reporting Report

1.1 Work Breakdown Structure (WBS) Leadership

We established the WBS Leadership for the Aristotle Team. For the time being, Rich Wolski will serve as UCSB Lead on all Aristotle Cloud Federation activities until he can hire additional support staff. Recruiting efforts are underway. Wolski will remain the UCSB lead for the Project Management WBS group.

- Cloud Federation Project Management, Oversight & Reporting
  - Team Lead Lifka, UB Lead Tom Fulani, UCSB Lead Wolski
- DIBBs Acquisition, Installation, Configuration, Testing & Maintenance
  - Team Lead Resa Reynolds, UB Lead Guercio Salvatore & Andrew Bruno, UCSB Wolski/TBD
- Cloud Federation Portal
  - Team Lead Susan Mehringer
    - Portal Development & Maintenance
      - Lead Mehringer
    - Allocations & Accounting Integration
      - Lead Lucia Walle
    - Open XDMoD Integration
      - UB Lead Steven Gallo
    - Eucalyptus and QBETS Integration
      - UCSB Lead Wolski/TBD
- Research Team Support
  - Team Lead Adam Brazier, UB Lead Varun Chandola, UCSB Lead Wolski/TBD

1.2 External Advisory Committee

We established an External Advisory Committee:

Name	Affiliation	Project	Email
John Towns	NCSA	XSEDE	jtowns@illinois.edu
Craig Stewart	IU	Jetstream	stewart@indiana.edu
Jamie Kinney	AWS	SciCo	jkinney@amazon.com
Rick Wagner	SDSC	Comet	rpwagner@sdsc.edu
Ian Foster	UC	Globus	foster@cs.uchicago.edu
Dan Nurmi	HP	Eucalyptus	nurmi@hp.com
Ben Rosen	Dell	Big Data & Cloud	ben_rosen@dell.com
Steve Johnson	WCM	NIH CTSC	johnsos@med.cornell.edu

1.3 Subcontracts

The subcontract with UCSB was completed on 12/02/2015.

1.4 Project Change Request

A Project Change Request by Co-PI Wolski to use funds originally proposed for a postdoctoral researcher to fund UCSB technical staff for Aristotle support/maintenance functions with research functions carried out by Wolski and a Graduate Research Assistant was approved by the NSF Division of Grants and Agreements on 12/09/2015.

1.5 Project Execution Plan

The Project Execution Plan (PEP) was approved by NSF on 12/18/2015.

1.6 Team Meetings

We established every other Tuesday from 12-1pm Eastern Time as our all-hands team call starting 10/27/2015. We have had three team meetings thus far. Topics of discussion included:

- Review of Project Plan expectations and initial action items for all WBS areas and associated leaders.
- Lifka will circulate monthly reports to the team for review before submitting to NSF Program Director Amy Walton.
- Reynolds setup a team mailing list on Google (completed 10/27/2015).
- PEP & Cooperative Agreement completion.
- Lifka requested status updates for monthly reports from all WBS areas.
- NSF PEP approved; order hardware whenever ready.
- UCSB expects quotes early January. At that time, the Aristotle team will review two different UCSB configuration options in terms of density, scale out, and other issues.
- CU had a positive exchange with the Euca team at HP. The Euca team is very interested in the idea of providing InCommon support through the Euca portal.
- Lifka will be setting up the first External Advisory Committee call in January.
- UB will be talking with UCSB about QBETS in early January.

1.7 PI Meetings

PI meetings included:

- Lifka and the CU team had followup discussions with Cycle Computing about joint efforts to harden CloudLaunch and make it available as open source to the community. In addition, Cycle Computing will discuss a partnership with AWS to have CloudLaunch queues available to the community on AWS resources. We will also discuss the possibility of having AWS offer XSEDE allocations on this resource.
- Lifka met with Professor Hakim Weatherspoon (http://www.cs.cornell.edu/~hweather) and Robbert van Renesse (http://www.cs.cornell.edu/Info/People/rvr/) about partnering with them on their Supercloud project which is jointly funded by NSF-CISE (program director Amy Apon) and NIST. This technology has the potential of making it very easy to burst between Aristotle and other clouds including Jetstream, Chameleon, CloudLab, and all the major public cloud providers. It removes the burden of researchers having to have different virtual machines for each cloud software stack. Lifka will be following up with Craig Stewart to pilot the tool between Aristotle and Jetstream early in the New Year.
- Lifka, Furlani and their team members met at SC15 to discuss the phased implementation of Open XDMoD.

1.8 Project Planning and Preparation

A Google shared folder with all project planning documents has been established. Team members are tasked with updating those documents as part of the monthly reporting process. Mehringer and Reynolds have been working on a detailed multi-phase plan for all Aristotle components and capabilities required for a targeted January 2016 go-live. In addition:

- Discussions were held on User Portal requirements (see initial requirements outlined in section 3.0).
- UB and UCSB have begun initial conversations around QBETS integration into Open XDMoD.
- Requirements analysis for the allocations and accounting system is underway (see section 3.6).
- Requirements analysis for the science use cases, i.e., hardware, programming software, and infrastructure software stack requirements, is underway (see section 4.2).
- Hardware options and quotes were gathered in anticipation of the project PEP being approved.

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

2.1 Federation Resource Status Updates

- **CU**

  Red Cloud has been in production at Cornell (and available to the national community under a cost recovery model) for close to 5 years. Recently CAC established a second geographical location at Weill Cornell Medicine (WCM) in Manhattan, NY for fail over and additional capacity. Red Cloud details and user documentation are available at www.cac.cornell.edu/redcloud/. Much of this documentation, including "how to burst to AWS," will be useful to the Aristotle Cloud Federation. On 10/27/2015 CAC upgraded Red Cloud at WCM to Eucalyptus 4.2. The upgrade went very smoothly and CAC subsequently upgraded Red Cloud resources in Ithaca.

  The CU team obtained and finalized Dell quotes for the first hardware purchase. CU upgraded Red Cloud to Eucalyptus 4.2 and worked with HP to obtain a hotfix for 4.2 that will be included in 4.2.1.

  The CU team built the Aristotle User Portal web site infrastructure using a Red Cloud instance. User Portal content will be developed next. This portal will eventually move to Amazon, but for now it lives at Cornell (see specific information in portal section 3.0). The team also finalized Dell quotes to augment Red Cloud. Portal, allocations, and accounting discussions were held to ensure that the infrastructure will meet all requirements for these components.

- **UB**

  Procurement (5/14-7/1) (UB funds not DIBBs):
  The PO for UB cloud hardware and software went out on 5/14/2015. Various components were delivered throughout the month of June. Everything from the original PO was received by July 1st. During pre-installation, discussions with Eucalyptus engineers determined that UB/CCR needed to order an additional server to be used as a dedicated Cloud Controller and the servers that were ordered for the Cluster Controllers needed more memory. CCR also had to make additional network connections in order to support the configuration that was required. This additional equipment was received mid-July.

Hardware Installation (7/1-7-15):
After all the additional equipment arrived, everything was racked and cabled up. The servers were provisioned utilizing Foreman and Puppet. UB/CCR IT staff created Puppet configurations for each server type based on the specifications provided by Eucalyptus engineers and tested the installations several times. New servers can be provisioned in a matter of minutes with minimal effort. At this point, everything was ready for the installation of the Eucalyptus software stack.

Cloud Installation (8/18-8/20):
Two engineers from HP were on site at CCR for 3 days to perform the cloud software installation and configuration. After day two, the cloud was mostly functional minus a few features that needed to be further investigated by Eucalyptus. Working with the Eucalyptus engineers, CCR staff, led by Guercio and Bruno, were able to solve all the problems that came up during the installation. Most of these issues were due the complexities of UB's installation, which features a split-cloud setup consisting of a private and public cloud that are independent of each other but share the same cloud controller. This is a nonstandard setup and required significant customizations on the network site to keep the two separate.

Training (9/22-9/24):
One of the HP engineers that performed the installation came back to CCR to perform administrative training. Each of CCR's 5 systems administrators completed the course material. After this point, the UB cloud was fully functional.

Current Cloud Status (10/16):
At the current moment, the cloud is up and running and CCR can create instances and this has been tested. CCR admin staff are using a cloud instance to perform the system administration/programming test for potential candidates. The UB cloud is not fully integrated with the rest of CCR's services infrastructure yet and therefore users are not yet able to self-provision their own instances. It is anticipated that this capability will be available soon.

Next Steps:
CCR staff, led by Guercio and Bruno, are currently working on integrating the cloud infrastructure with their existing core services. Specifically, they are trying to get the cloud to pull in the user account information from the existing central authentication system so that they can manage the Eucalyptus accounts in the same place that the rest of the CCR accounts are managed. Also in progress is creating the usage policies/terms and governance for the cloud so that CCR users understand proper usage. CCR is exploring the use of charge back model similar to what Cornell has in place for Red Cloud. This will help alleviate many of the concerns with possible system abuse and misuse. CCR is also working through the branding element of this service offering. Similarly to how Cornell markets Red Cloud, CCR would like to present this service to the University community to fill in some of the gaps that currently exist in the University's IT Service Catalogue.

The Buffalo team has been working with their Central IT department and NYSERNet to secure a /22 block of IP addresses. The campus currently does not have anything greater than /24 blocks. They are working on rolling out their existing Eucalyptus cloud into a production state and have integrated it into their FreeIPA and Foreman/Puppet infrastructure. Buffalo has also worked on refreshing the previous hardware quotes with some slight changes based on feedback from Eucalyptus during the installation of their first cloud.

NSF
#1541215

New Cloud Equipment:
UB submitted the P.O.s for their cloud equipment. Their cloud nodes will be similar to Red Cloud nodes, each with 24 cores and 192GB RAM. They will also use a Dell EqualLogic SAN (168TB) and a Dell Force10 S4820T 10Gb switch. UB upgraded their cloud to Eucalyptus 4.2.1 as well.

- **UCSB**
  The Engineering Computing Infrastructure (ECI) cloud has been operational and in production for approximately 2 years at UCSB. It consists of:
  - 22 nodes with 32GB memory and 4 Intel E3-1230 3.3 GHz cores each.
  - 8 nodes with 64GB memory and 6 Intel E5-2620 2.1 GHz cores each.
  - 100GB Seagate 600 Pro SSD per node and 200GB STA disk per node.
  - Dell EqualLogic PS4100E SAN with 13TB of storage.

  The system is maintained by the UCSB College of Engineering's technical staff. The current system is running Eucalyptus 4.1.2, but because it is used for both educational and research purposes, it will not be upgraded to Eucalyptus 4.2 until the quarter break in December.

  The UCSB team apprised the campus NOC of the scope and scale of Aristotle's bandwidth and addressing needs and researched current 10G switch offerings. They also worked with vendors on updated hardware quotes based on the Dell quotes submitted with the proposal.

  Most recently, the UCSB team spent time researching InCommon setup requirements with the UCSB Identity Group. Cornell will be sharing our InCommon setup information to assist. They continued to work with HP on their hardware proposal.

Reynolds created the following planning doc in the Aristotle Google shared folder. She is actively working with the team to identify target dates for all capability roll outs and resource specifications.

	CU	UB	UCSB
Cloud URL	euca4.cac.cornell.edu	ccr-cbls-2.ccr.buffalo.edu	TBD
EUCA Version	4.2 with hotfixes	4.2.1	4.1.2
Migrate to 4.2.1	As soon as it's available	12/2015	3/1/2016
Globus	Yes	Not currently, but is planned	?
InCommon	Yes	Not currently, but is planned	Planned
Hardware quotes	Dell quotes posted, waiting to process	Quotes sent to processing	Waiting for quotes
Hardware vendor	Dell	TBD	TBD
# cores	144	112-140 (target)	TBD
ram/core	4GB/8GB	6GB	TBD
10gb interconnect	Yes	Yes	Yes

2.2 Eucalyptus

Eucalyptus 4.2.1 is now available. Cornell has installed it on Red Cloud NYC. 4.2.1 resolved an issue with security groups and fixed several issues with upgrading to 4.2.x from earlier versions. The Aristotle team tested and provided feedback to HP Helion Eucalyptus team on the usability of the federation

NSF
#1541215

features in 4.2 in the context of the InCommon federation. Aristotle partners will complete upgrading to 4.2.1 after the New Year and go live with this version.

2.3 Positive Industry Influence: Eucalyptus

Aristotle is having a positive influence with industry as they recognize the value/importance of federation through our partnership. For example, we had a very productive meeting with the HP Helion Eucalyptus team of Dan Nurmi, Chris Grzegorczyk, and Dmitrii Zagorodnov about using InCommon in Eucalyptus. They are very supportive of implementing this feature. It turns out they've been thinking about federation and have been looking for a real world application to guide them how it should be implemented. The HP Helion Eucalyptus team proposed to add support for InCommon logins for the Eucalyptus web console. Each cloud would be a service provider that accepts security tokens from InCommon identity providers. Users would be able to log into the web console using their InCommon credentials. This would be a purely web console feature and should be achievable.

This feature should take care of the authentication part, but the Aristotle team will have to take care of authorization part (i.e., account creation/deletion/suspension, etc.) as part of account management.

2.4 CloudLaunch

CloudLaunch is a new capability being developed by Cornell. CloudLaunch consists of extensions to the SLURM scheduler which allow researchers to use cloud resources with traditional HPC methods. Each site will offer a "login node" where researchers can log in, compile codes, check output, and submit, cancel and check job status. When a job is submitted to a SLURM queue on this login node, the requested number of virtual "node instances" will be spun up and the job run on them as if it were a dedicated HPC cluster. When the jobs complete, SLURM will shut the virtual node instances down making the cloud resources available for other types of work. Note: job queues will have different node instances associated with them (e.g., number of cores, RAM/core). Recent testing is progressing nicely and we expect several of our proposed science use cases will want to take advantage of the CloudLaunch capability.

Once testing is complete, we will investigate how to have queues that point to federation resources as well as AWS and, hopefully, NSF cloud resources eventually for users that have allocations.

Our goal is to make this code available to the community and, hopefully, gain community support. Lifka has already discussed this with John Towns and Rick Wagner (Aristotle External Advisors). We believe creating a group of folks interested in contributing cloud-related tools, virtual machines, and cloudy-ready, pre-packaged software will be of great value to the NSF community and likely something that makes sense for XSEDE to host. CloudLaunch continues to be tested and hardened at CU.

Lifka had discussions with Cycle Computing at SC15 around Cycle Computing hardening and supporting CloudLaunch while continuing to make it available to the community as open source. Their goal would be to have AWS login nodes running CloudLaunch that would allow anyone to login and submit jobs to AWS. This would be ideal for the Aristotle Federation. Users could simply login to a different login node to submit jobs that require more hardware resources than are available in the Federation. Lifka and the CU team followed up with Cycle Computing to discuss next steps. Cornell agreed to share the existing CloudLaunch code with Cycle Computing to help them understand its value to the HPC community as they explore running it in the cloud. More details to come next month.

2.5 RT (Request Tracker)

Aristotle incidents will be tracked using CU's RT ticketing system. Initially, researchers will send email to Aristotle-help@cornell.edu to report problems, request support, or contact the team. Once the Aristotle portal is "live," we plan to add an alias of help@federatedcloud.org. Researchers will then be able to use either email to submit requests. Brazier, the science lead, will assign science-related tickets and Reynolds, the infrastructure lead, will assign infrastructure-related tickets as necessary.

Latest status: Aristotle-help@cornell.edu is now available to report problems, request support, or contact the team. The alias help@federatedcloud.org will be working soon.

Discussion items can go to the Aristotle-team list for discussion, but problems/questions should go to the ticket system for a number of reasons:

- For project reporting purposes, we need it to keep project problem tracking in one place.
- It is the best way to track open problems.
- Researchers will get a response even when their personal contact is out of touch.
- Consultants can find project history.

If a researcher persists in writing to the consultant directly, the consultant should make that contact into a ticket (retraining).

3.0 Cloud Federation Portal Report

As a team we agreed on what each phase of "Aristotle Production" would include. Phase 1 was the primary and obvious focus. We expect to refine Phase 2 and beyond as we progress. Each step is focused on providing full federation capability while making each of the components stand-along ready and redistributable. Phase 1 largely leverages existing components in use at each site. Our current component plan is as follows:

	Phase 1	**Phase 2**	**Phase 3**
	Now - Jan 2016	**Jan 2016 - 18 mo. mark**	**18 mo. mark - ??**
Allocations & Accounting	Allocations: Fair division of resources across three sites and projects based on project readiness. Accounting: Implement the accounting and tracking systems currently used on Red Cloud. UCSB & UB report the same data back to CAC, i.e., poll data and send reports to CAC.	Move portal & database to AWS.	Make available as download from GitHub.
Documentation & Training	Create basic user docs, focused on materials that will get users started. Draw from existing Red Cloud docs and the user project requirements.	Move the docs into a repository for the federation to draw from.	Make available as download from GitHub.

#1541215

Usage & Status	Show % utilization graphs. Show available resources. Show user balance.	Incorporate Open XDMoD.	Incorporate QBETs (via Open XDMoD). Make available as download from GitHub.
User Authorization & Keys	Login to the portal using InCommon.	Get 4.2.1 federated key after InCommon login.	
"Euca Tools"	Identify common Euca portal tasks to be embedded in the portal via a button to a script. Identify which images should be created.	Create a repository to give back to Euca.	
Systems	Get Globus running on all sites. Order and install hardware at all sites. Determine software requirements for portal and accounting elements (see details in Aristotle spreadsheet).		

3.1 InCommon Access

We have successfully demonstrated that we can use InCommon to password protect user access to sensitive project data and to their Eucalyptus keys.

3.2 Software Requirements

Software requirements for the portal were defined. We will use open source software that meets the requirements for the key functionality elements of authentication and Open XDMoD. Timeline planning for incorporating Open XDMoD was defined in more detail, first for incorporating Open XDMoD at each site, then for all sites on a federated portal, and later incorporating QBETs. We determined that the first phase of the project portal template will include project and user info, basic system status, documentation, and utilize InCommon authentication. Development of the portal template has begun.

Portal software requirements are as follows:

- Ubuntu 12.04 or CentOS 6
- Apache
- MySQL 5.1 or 5.5
- PHP 5.3+
- Java
- PhantomJS
- Cron
- Logrotate
- MTA with sendmail compatibility (e.g. postfix, exim or sendmail)
- User docs, team contribution: MediaWiki -or- draw user docs directly from GitHub
- Log in: InCommon
- Styles: Bootstrap (optional; implementation sites can style as they wish)

#1541215

The software requirements are largely driven by Open XDMoD. See details here: http://xdmod.sourceforge.net/software-requirements.html

3.3 Portal Platform

The initial 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.

3.4 Usage Data Collection

Usage data collection plans for Phase 1 were further refined. The current plan is to have all three sites collect core hour and storage (EBS) usage every ten minutes, using a perl script which polls Eucalyptus, and another script or program that computes usage by account and/or project, and logs the data in a local database. The data will be pushed from each site database into the federation database, at least every two hours. Next month, the scripts and database tables will be shared to ensure data collection is cohesive across sites, and the plan to put the data into the federation database will be refined. The database schema design is further discussed in section 3.6.

3.5 Open XDMoD

A release of Open XDMoD that collects cloud level metrics is expected in 2016. The UB team has begun a specification document that will be shared with the team for discussion. Until then, CU and UCSB can begin gaining experience with Open XDMoD on batch-based systems, or with the CloudLaunch system in development at CU.

3.6 Allocations & Accounting

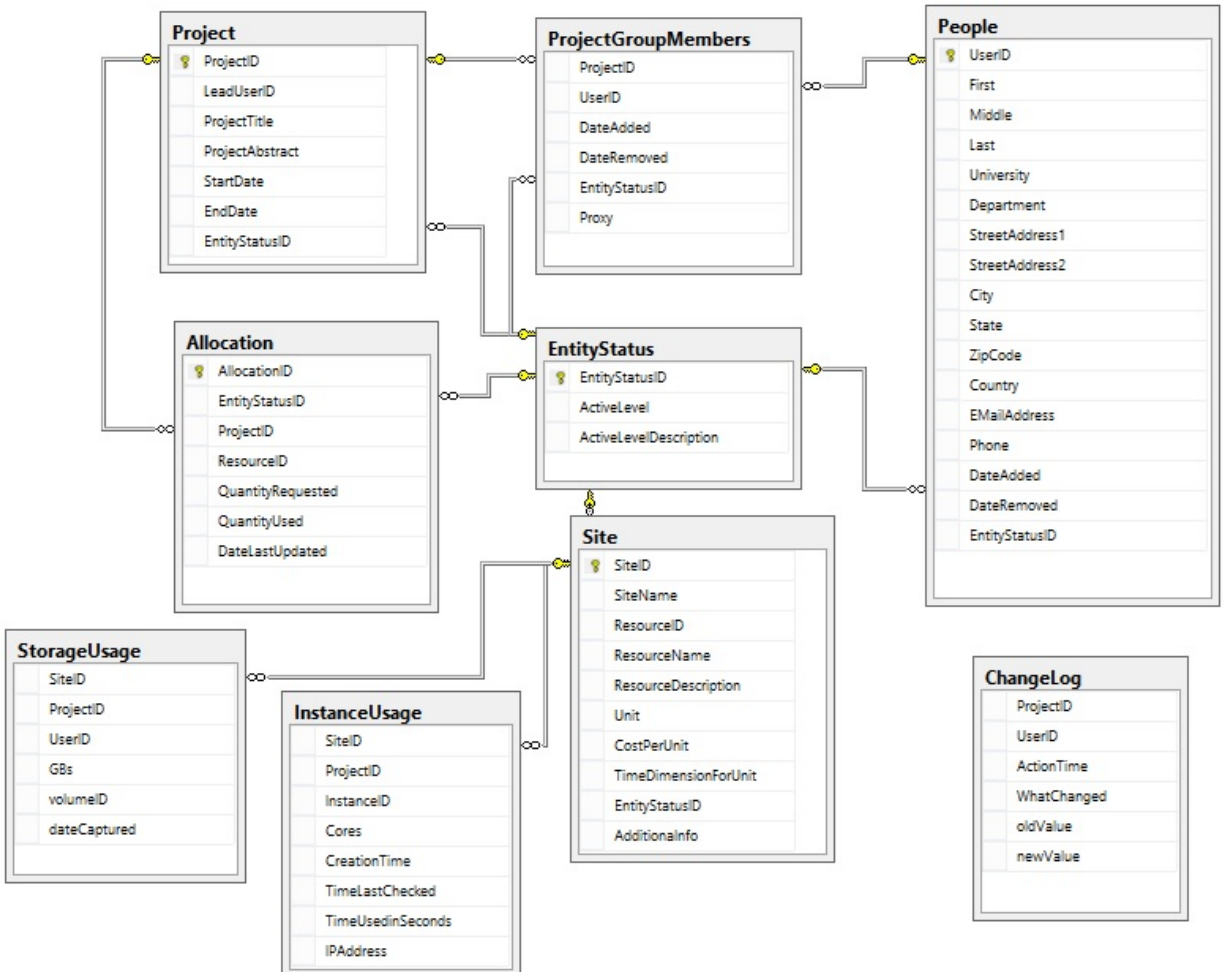
Allocations and Accounting components are also being planned in a phased approach. Initially, we will leverage the CAC Red Cloud Accounting and Allocation software we have been running in production for close to five years. We will help the partner sites implement usage accounting that reports back to CU and measures usage of all federation resources. At the same time, we are actively developing a longer term plan for breaking out the allocations, accounting and all critical federation components into redistributable building blocks that others can use to deploy their own cloud federation or to join the Aristotle Cloud Federation. Our goal (part of the project plan) is to have this plan completed, agreed upon and reviewed by a subset of the External Advisory Committee.

A proposed schema for a project, user, accounting and usage database was reviewed by the team. The goal is to keep this simple yet have all the information we will need for adding and removing projects and users while being able to report usage and allocations. This essentially provides the essential components of the CAC accounting system for cloud allocations and accounting.

Work on the schema for the project, user, accounting, and usage database is progressing. Specific use cases were reviewed, discussed, and a requirements analysis document was written. The allocations and use cases team worked together to revise and update the schema to incorporate this input (see below). In addition to making entity and table name changes, it was decided that the Allocation table will be updated with the quantity used for each resource (core hours and storage) so when viewing the information on the web pages it will show how much of a resource was requested and how much was used. A ChangeLog table was also added to reflect changes made and by whom.

NSF
#1541215

The next step will be to create the database using MariaDB and begin building Stored Procedures and API calls to update and record information. Then, Stored Procedures will be created to interact with other accounting components, user creation, project creation, web portal and Eucalyptus scenarios.



4.0 Research Team Support

4.1 Help Ticket Queues/Access

Reynolds, Steven Lee, and Mehringer will setup RT help ticket queues and provide appropriate partner access and ticket redirection in order to provide timely, collaborative support to the federation users.

4.2 Science Use Cases

In order for the Aristotle Cloud Federation to succeed, it must support the 7 proposed science use cases effectively and, ultimately, provide options/paths for faster “time to science." In this section, we provide a progress report for each of the 7 science teams. Initially this involves getting project plans in place with each science team and understanding how they currently do their science and their requirements, and to help them understand what their initial allocations are likely to be. We have limited staff effort in this award so our focus is to provide initial hand holding as necessary and use those experiences to create and

#1541215

improve user documentation and training materials so that future users can be as self-sufficient as possible.

Our next focus will be to get project plans in place with each science team and help them understand what their initial allocations are likely to be. Initial discussions with the 7 science teams are well underway. It is clear that all science teams are anxious to get started. Patrick Reed, Sarah Pryor, Angela Douglas, and Varun Chandola have all expressed a desire to get their research teams trained as soon as possible. We are considering providing limited exploratory allocations to Red Cloud to get them started while we await the first installment of the DIBBs infrastructure that will be allocated to the science teams.

We have created a requirements template and documentation for each team's requirements below.

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

Hardware

16 node cluster. Each node with 8 cores and 28GB memory.	1 16 core node with 1 high-performance NVIDIA GPU (1,536 CUDA cores and 4GB video memory).
25K CPU core hours.	1TB storage.

Programming Software

Java	Python

Software Infrastructure Stack

Apache Spark on cluster.	CUDA on single node.

Use Case 2: Global Market Efficiency Impact

Varun Chandola and Cristian Tiu (UB) had a kickoff meeting on October 30th to initiate plans for porting of UB science cases on the Aristotle Cloud Federation. Brazier also had a call with Chandola (UB) to help get the UB science teams up and running.

Roesch: Hardware

1 node cluster with 16 cores and 128GB memory.	40K CPU core hours (depends on how fast it is to start/stop VM).
10TB (but depends on what data will be available).	

Roesch: Programming Software

OneTick (Proprietary software)	Perl
R	MySQL
GNU parallel	Git

Roesch: Software Infrastructure Stack

Maybe none. OS, assuming Linux for now.

Tiu: Hardware

1 node/>=8 cores, 64GB memory.	CPU cycles required: unclear on CPU cycles required.
>= 2TB storage for now	

Tiu: Programming Software

MATLAB (w/ certain toolboxes)	R
MySQL	Julia

Tiu: Software Infrastructure Stack

Assuming Linux

#1541215

Wolfe: Hardware

1 node, 8 cores, ? GB memory (typically 32 or 64GB).	CPU cycles - not sure.
Storage - 5TB.	

Wolfe: Programming Software

Python	Perl
7zip	OCR software TBD.
SAS	

Wolfe: Software Infrastructure Stack

Linux	Bash

Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-Relevant Aerosol Properties

Adam Brazier installed WRF-Chem on a Red Cloud virtual machine in single-threaded mode to evaluate the required software stack; installation instructions were recorded on the project's GitHub wiki. Issue: we are still investigating whether ifort is necessary or if gfortran is sufficient.

Hardware

About 10TB base data and 25TB of data output for each year processed.	At least 4GB per core, with distributed memory; scaling not tested yet with new data and models, requirements may be higher.
Processing estimated to take >60 000 core-hours for each year of data processed.	

Programming Software

WRF	WRF-Chem
Fortran	Perl 5
MATLAB	Netcdf
Flex	Curl-devel
Hdf5-devel	Byacc
Libfl-devel	Csh
Time	JASPAR

Software Infrastructure Stack

Linux	

Use Case 4: Transient Detection in Radio Astronomy Search Data

Jim Cordes is interested in NANOGrav research as well as searching the PALFA data set for transient events. We will focus initially on the PALFA work, but happily support NANOGrav and other projects once it is working successfully and assuming we have adequate resources to allocate to him for the additional projects.

Hardware

About 30TB of data, growing at 5TB/year.	4GB per core.
Processing takes 1.5-2 core-hours per beam, about 210 000 beams. About 500 000 core-hours per complete reprocessing.	

Programming Software

PRESTO pulsar package and dependencies	

Software Infrastructure Stack

Linux	Windows Server
SQL Server	ASP.NET MVC
Python 2.7	

#1541215

Use Case 5: Water Resource Management Using OpenMORDM

The Reed team's biggest usage has been 526,000 cores. They are a likely candidate to work with AWS. We will work to make the connections with Reed's team and the AWS SciCo Team. We believe this would be an exciting capability, demonstrating value on an important science problem: drought management. We think AWS may consider supporting this use case because their research product is in part for municipalities that could in principle pay for the necessary computing for the product in operations mode. Reed's collaborators at Penn State expressed interest in joining the federation with a new computer they are acquiring. There are several technical and logistical issues to discuss but, as we had hoped, we are already seeing interest from other institutions in participating in the Aristotle Cloud Federation!

Hardware

32 compute nodes.	512 core, Dual 8-core E5-2680 CPUs @ 2.7 Ghz.
128GB of RAM (8GB/core).	10GB Ethernet and InfiniBand QDR full interconnect.
8TB /home (NFS mounted across 10GB Ethernet to all compute from head node).	900GB local /tmp on each node.
95TB Lustre /scratch space.	

Programming Software

Intel Compilers (including MKL)	Openmpi 1.6.5
Mathematica	MATLAB
SAS	Boost
cmake	eclipse
hdf5	netcdf
valgrind	visit
zlib	acml
R	BLAS, LAPACK libraries

Software Infrastructure Stack

CloudLaunch or similar, to launch and manage cluster for MPI runs then take it down again.	Linux

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

Brandon Barker and Angela Douglas have established requirements.

Hardware

About 200GB storage initially (years 1-3), ramping up to 8TB storage in years 4-5 to allow for analysis of metagenome data.	64GB per node, 4+ cores/node, 0-2 nodes active during years 1-3.
Allocation of a minimum of 8,500 core-hours over the course of the project, possibly scaling up within an order of magnitude.	

Programming Software

Python	MATLAB
libSBML	Gurobi
SBML Toolbox	COBRApy
COBRA Toolbox	Docker or Nix
Git	

Software Infrastructure Stack

Linux	

#1541215

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production
UCSB has identified Andreas Boshke as their science team lead supporting Science Use Case 7. They are using an open source sensor network framework called GSN, but we have had to fork it for them to make it build in the cloud. Not sure in which category GSN goes -- probably software stack. Also -- there isn't now, but there may need to be, Windows images.

Hardware

8 to 32 nodes, at least 4 cores per node (8 better).	At least 4GB memory per node (8 is better).
At least 50GB storage (100GB better) for every 4 cores in ephemeral disk.	5TB in object storage and/or volumes (split yet to be determined).
Core Hours: Probably 96 hours/per week/per node (e.g., we run an ensemble for 96 hours/week). This is really a guess at this point.	

Programming Software

R	MPI
PostgreSQL	MySQL
Hadoop/Pig/Hive	Spark
Storm	Zookeeper
Puppet	MATLAB

Software Infrastructure Stack

Linux (Centos and Ubuntu)	Python 2.7
Apache	Tomcat
AppScale	

4.3 Publications
Co-PI Rich Wolski, UCSB and John Brevik, California State University, Long Beach submitted a paper entitled "Providing Statistical Reliability Guarantees in the AWS Spot Tier" to the 24th High Performance Computing Symposium (HPC 2016), April 3-6, 2016, Pasadena, CA.

Co-PI Furlani, UB, submitted a paper entitled "Providing Statistical Reliability Guarantee for Cloud Clusters," to the 2016 USENIX Workshop on Cool Topics in Sustainable Data Centers (CoolDC '16), March 2016, Santa Clara, CA.

5.0 Outreach Activities

5.1 CASC Presentation
Lifka presented information about the Aristotle Cloud Federation at the Fall Coalition for Academic Scientific Computation (CASC) meeting on October 14th in Arlington, Virginia. Several institutions expressed immediate interest in joining the federation. Lifka explained that is certainly a goal and we will report back to CASC when we are ready.

5.2 Media Outlets/Social Media
We generated national, international, and local media coverage across 21 media outlets including major IT magazines (CIO, Campus Technology, Government Technology News, NetworkWorld) as well as HPC verticals (ACM TechNews, HPCwire, InsideBIGDATA, Scientific Computing Magazine). A strong social media presence (Twitter, LinkedIn, etc.) was generated as well in partnership with NSF, AWS, Cornell, and our other partners.

5.2.1 National Media Coverage

- *ACM TechNews* - Cornell Leads New NSF Federated Cloud Project
- *Campus Technology* - Cornell to Lead NSF-Funded Cloud Federation for Big Data Analysis
- *Cloudwards* - Aristotle: Academic Focused Cloud Funded
- *CIO* - University researchers get $5M grant to build 'Aristotle Cloud'
- *Cloud Strategy Magazine* - Cornell Leads New NSF Federated Cloud Project
- *Data Center Talk* - Cornell to Head the $5M Federal Cloud Computing Program
- *Global Wireless Research* - University researchers get $5M grant to build 'Aristotle Cloud'
- *Government Technology News* - Wisdom of the Clouds: Aristotle Cloud Federation
- *HPCwire* - Cornell Leads New NSF Federated Cloud Project
- *InsideBIGDATA* - Cornell to Lead Aristotle Cloud Federation for Research
- *NetworkWorld* - University researchers get $5M grant to build 'Aristotle Cloud'
- *Next Generation Communications* - Cornell Leads New NSF Aristotle Cloud Federation Project
- *Scientific Computing* - Cornell to lead $5M NSF Federated Cloud Project
- *Web Host Industry Review* – National Science Foundation Sponsors $5 Million Cloud Project

5.2.2 International Media Coverage

- *CIOL (India)* - Cornell University to develop federated cloud
- *ComputerWorld (Australia)* - University researchers get $5M grant to build 'Aristotle Cloud'
- *Primeur Magazine (UK)*- Cornell leads new National Science Foundation federated cloud project
- *TechWorld (Australia)*- University researchers get $5M grant to build 'Aristotle Cloud'

5.2.3 Local Media Coverage

- *Cornell Chronicle* - Aristotle: A federated cloud for academic research
- *Cornell Daily Sun*- Cornell to Lead $5M Federal Cloud Computing Program
- *Ithaca Journal* - Cornell to head effort to streamline data flows

5.3 CloudLaunch Backgrounder

Cornell produced a backgrounder on CloudLaunch that describes its features and benefits: https://www.cac.cornell.edu/technologies/CloudLaunch.pdf.

5.4 SC15 Conference

The federation was also featured at the Cornell SC15 exhibit and Aristotle news flyers were distributed at our exhibit as well as the University at Buffalo's exhibit. The CloudLaunch backgrounder was also distributed.