

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

Monthly Report 12/22/2015

Report 3 of 18

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

This is the third 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).







1.0 Cloud Federation Project Management, Oversight & Reporting Report	3
1.1 Subcontracts	3
1.2 Project Change Request	3
1.3 Project Execution Plan	3
1.4 PI Meetings	3
1.5 Team Meeting	3
1.6 Project Planning and Preparation	3
2.0 DIBBs Acquisition, Installation, Configuration, Testing & Maintenance Report	4
2.1 Federation Resource Status Updates	4
2.2 Positive Industry Influence: Eucalyptus	4
2.3 CloudLaunch	5
2.4 RT (Request Tracker)	5
3.0 Cloud Federation Portal Report	5
3.1 Software Requirements & Portal Platform	6
3.2 Usage Data Collection	6
3.3 Open XDMoD	6
3.4 Allocations & Accounting	6
4.0 Research Team Support	7
4.1 Science Use Cases	7
Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-	
Relevant Aerosol Properties	7
Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota	8
4.2 Publications	8
5.0 Outreach Activities	9
5.1 CASC Presentation	9
5.2 CloudLaunch Backgrounder	9
5.3 SC15 Conference	9





1.0 Cloud Federation Project Management, Oversight & Reporting Report

1.1 Subcontracts

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

1.2 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.

1.3 Project Execution Plan

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

1.4 PI Meetings

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 (<u>http://www.cs.cornell.edu/~hweather</u>) and Robbert van Renesse (<u>http://www.cs.cornell.edu/Info/People/rvr/</u>) 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.

1.5 Team Meeting

We held an all-hands team call on 12/22/2015. Topics of discussion included:

- 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.

1.6 Project Planning and Preparation

A great deal of other planning and preparation was done by the Aristotle team in December including requirements analysis for the allocations and accounting system, hardware options and quotes were gathered in anticipation of the project PEP being approved, User Portal planning and development, and ongoing work with the science teams.





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

2.1 Federation Resource Status Updates

• **CU**

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

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 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.

Progress on software, hardware quotes, and other configuration items are reflected in this updated planning table:

	CU	UB	UCSB
		ccr-cbls-	
Cloud URL	euca4.cac.cornell.edu	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
		Not currently, but is	
Globus	Yes	planned.	?
		Not currently, but is	
InCommon	Yes	planned.	Planned
	Dell quotes posted, waiting to	Quotes sent to	
Hardware quotes	process.	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 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 on 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 the authorization part (i.e., account creation/deletion/suspension, etc.) as part of account management.

2.3 CloudLaunch

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.

2.4 RT (Request Tracker)

Aristotle incidents will be tracked using CU's RT ticketing system.

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

3.0 Cloud Federation Portal Report

We use the following table for Portal planning. There were no updates to the table this month.

	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.
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 Software Requirements & Portal Platform

Last month the software requirements for the portal were defined, focusing on using open source software that meets requirements for the key functionality elements of authentication and Open XDMoD. This month the initial platform was built on a cloud instance, running Centos7 with the LAMP stack installed, and populated with a "Coming Soon" page at <u>http://www.federatedcloud.org</u>.

3.2 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 3.4.

3.3 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.

3.4 Allocations & Accounting

Work continued on the schema for a project, user, accounting, and usage database proposed last month. This month, 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.

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 Science Use Cases

We have completed the initial requirements analysis for the remaining two Science Use Cases—Use Cases 3 and 6—and investigated the installation of the key software stack for Use Case 3

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.





Use Case 3 (Cont.)

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 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	64GB per node, 4+ cores/node, 0-2 nodes active during years 1-3.
data.	
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

MATLAB		
Gurobi		
COBRApy		
Docker or Nix		
Software Infrastructure Stack		

4.2 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 CloudLaunch Backgrounder

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

5.3 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.

