ARISTOTLE

CLOUD FEDERATION

Award No. 1541215

Building a Federated Cloud Model: Aristotle Cloud Federation

David Lifka, PI
Director, Cornell University Center for Advanced Computing (CAC)
Interim CIO & Vice President, Cornell Information Technologies
lifka@cornell.edu

Thomas Furlani, Co-PI
Director, Center for Computational Research (CCR), University at Buffalo
furlani@buffalo.edu

Rich Wolski, Co-PI
Professor, Computer Science, University of California, Santa Barbara
Co-Founder, Eucalyptus Systems
rich@cs.ucsb.edu

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

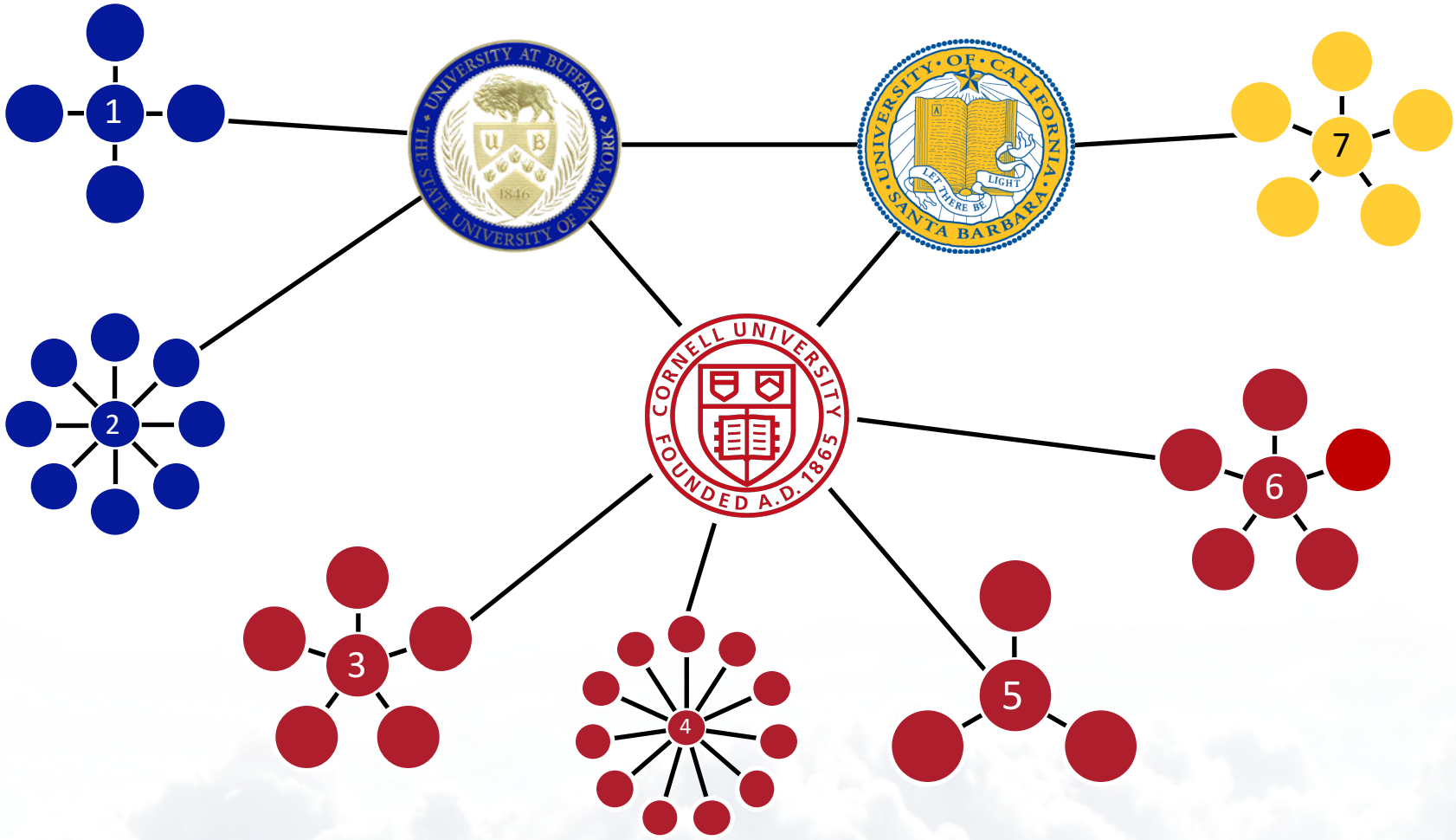
Aristotle Overview & Goals

- $5 million, 5-year NSF award to deploy a federated DIBBs cloud at Cornell, U. Buffalo, UC Santa Barbara
  - 7 science teams, over 40 global collaborators
    - Requiring flexible workflows/analysis tools for large-scale data
    - Representing diversity of analysis requirements and cloud usage modalities
      - Earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, food science
- Project goals
  - Optimize "time to science"
  - Demonstrate the value of sharing resources and data across institutional boundaries

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

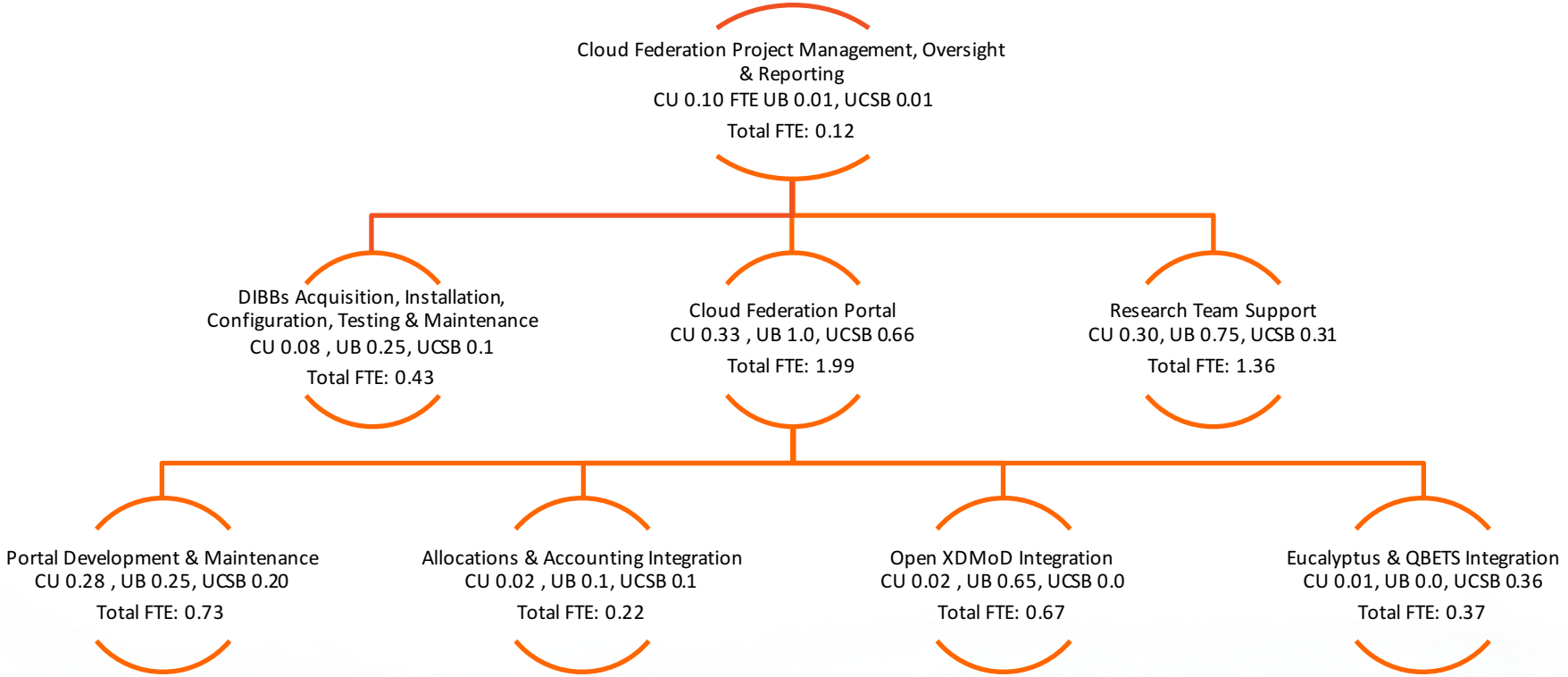
Multi-Disciplinary Collaborations in a Multi-Campus Cloud Federation



Award No. 1541215

ARISTOTLE
CLOUD FEDERATION

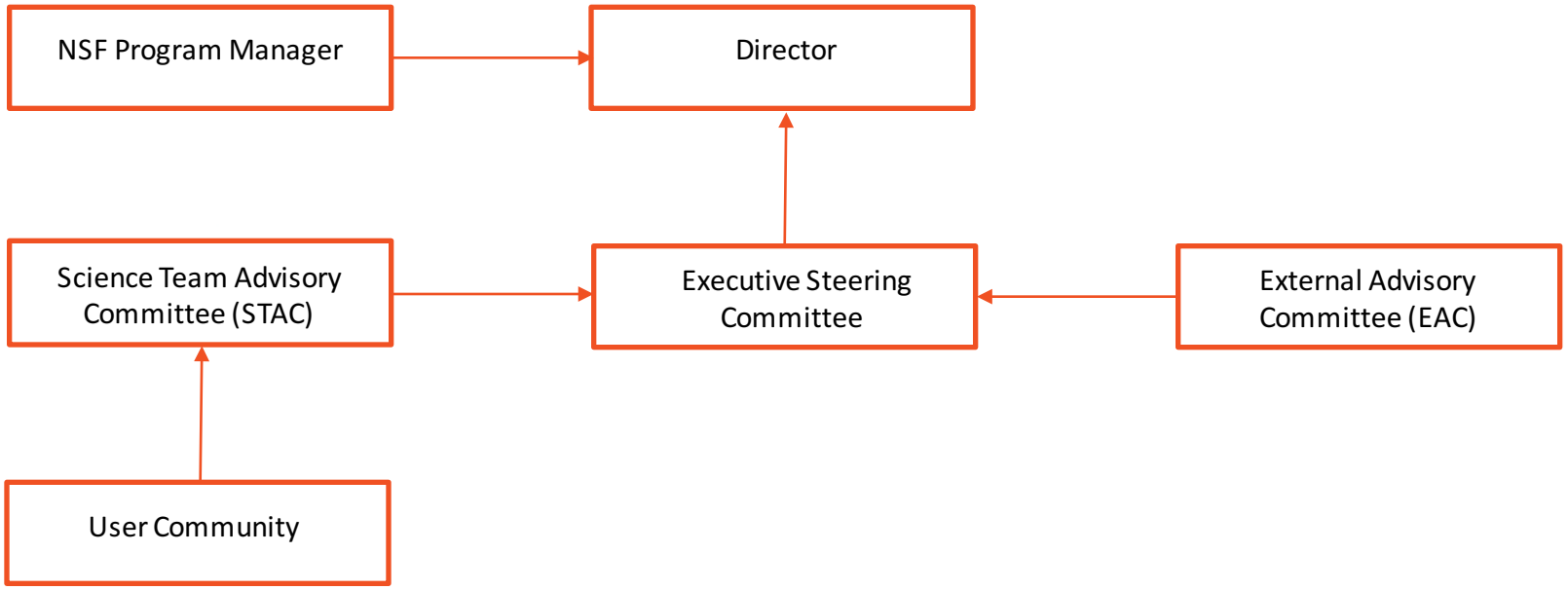
Project Work Breakdown Structure



Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Project Governance



External Advisory Committee Introductions

Name	Affiliation	Project	Email
Amy Walton	NSF	Aristotle	awalton@nsf.gov
Adam Brazier	CAC	Aristotle	brazier@cornell.edu
Susan Mehringer	CAC	Aristotle	susan@cac.cornell.edu
Paul Redfern	CAC	Aristotle	red@cac.cornell.edu
Resa Reynolds	CAC	Aristotle	rda1@cornell.edu
Jamie Kinney	AWS	SciCo	jkinney@amazon.com
Dmitrii Zagorodnov	HPE	Eucalyptus	cedzc@hpe.com
Ben Rosen	Dell	Big Data & Cloud	ben_rosen@dell.com
Craig Stewart	IU	Jetstream	stewart@indiana.edu
John Towns	NCSA	XSEDE	jtowns@illinois.edu
Rick Wagner	SDSC	Comet	rpwagner@sdsc.edu
Ian Foster	UC	Globus	foster@cs.uchicago.edu
Steve Johnson	WCM	NIH CTSC	johnsos@med.cornell.edu

Project Office Update (Q1/Q2)

- Completed Cooperative Agreement & PEP (12/18/15)
  - Established External Advisory Committee
- Completed UB and UCSB subcontracts
- Established biweekly team calls
- Established monthly/quarterly reporting process
- Discussions around enabling technologies
  - Open XDMoD data collection for Aristotle federation @SC15
  - Cycle Computing leverage @SC15
  - Supercloud with Cornell Computer Science and Jetstream
  - QBETS integration into Open XDMoD and Globus Genomics
  - Key portal components and capabilities
  - AWS SciCo partnership
- Agreement on allocations and accounting principles

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Infrastructure Update (Q1/Q2)

- Cornell Red Cloud in production for 5 years
  - Red Cloud Eucalyptus upgraded to 4.2
  - Federation capabilities tested between Ithaca and WCM NYC
  - Current Aristotle portal site running in Red Cloud
  - January 2016 first installment of DIBB infrastructure ordered/installation complete
- UB Lake Effect cloud installed/operational 10/16/15
  - Lake Effect Eucalyptus upgraded to 4.2
  - Integrating into existing UB core services
  - Leveraging Euca Availability Zones
  - January 2016 first installment of DIBBs infrastructure ordered/installation nearly complete
- UCSB Engineering Computing Infrastructure cloud in production for 2 years
  - UCSB Eucalyptus cloud upgraded to 4.2
  - Working on InCommon setup
  - First installment of DIBBs infrastructure ordered/arriving now

Award No. 1541215

Portal Update (Q1/Q2)

- **Goal:** Informed Bursting
  - Across the Aristotle federation
  - NSF Clouds
  - AWS
- Defined portal components and developer responsibilities
- Completed InCommon integration proof of concept
  - Support from HPE Helion Eucalyptus team for InCommon access to Euca Portal in development
- Completed setup of Aristotle RT system at Cornell
  - help@federatedcloud.org
  - Established ticket routing roles
- Completed Allocations & Accounting database schema and database
  - Based on streamlined version of Cornell CAC Allocations & Accounting system
- Will draw basic user documentation from Red Cloud
- Utilization and allocation balance reporting
  - Ability to show DIBBs usage vs. local institutional usage
- Integration of XDMoD and QBETs
- Globus Online up at Cornell; UB and UCSB in progress

Award No. 1541215

Research Teams Update (Q1/Q2)

- 7 initial science teams
  - Use Case 1: A Cloud-Based Framework for Visualization & Analysis of Big Geospatial Data (UB)
  - Use Case 2: Global Market Efficiency Impact (UB)
  - Use Case 3: High Fidelity Modeling and Analytics for Improved Understanding of Climate-Relevant Aerosol Properties (CU)
  - Use Case 4: Transient Detection in Radio Astronomy Search Data (CU)
  - Use Case 5: Water Resource Management Using OpenMORDM (CU)
  - Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota (CU)
  - Use Case 7: Multi-Sourced Data Analytics to Improve Food Production (UCSB)
- Science team meetings started from day 1
  - Understand current workflows and processes
  - Document software and hardware requirements
  - Develop cloud adoption strategy
  - Start creating VMs (e.g., geospatial analytics, finance analytics, Windows instance for MATLAB VMs already completed)
  - Demo proof of concepts to science teams

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Use Case 1: A Cloud-Based Framework for Visualization and Analysis of Big Geospatial Data (UB - Chandola)

- *What?*
  - Integrate, visualize and analyze heterogeneous geospatial data from around the world
- *Why Aristotle?*
  - Offload variable computational demands from local infrastructure
  - Create an expert service hub for widely distributed data
  - Scalable and elastic computing

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Use Case 2: Global Market Efficiency Impact (UB- Tin, Roesch, Wolfe)

- *What?*
  - Measure market efficiencies with high volumes of historical and current intraday market data
- *Why Aristotle?*
  - Burst to integrate new data
  - Tens of TBs moved to available processing resources
  - Share data and pipelines around the world for collaboration and reproducibility

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

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

- *What?*
  - Model the behavior of atmospheric aerosol particles to predict and ameliorate climate and human health impacts
- *Why Aristotle?*
  - Burst to process new data
  - Share a high-value processed dataset of general interest
  - Build an extensible toolkit for researchers, transferrable to other groups

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Use Case 4: Transient Detection in Radio Astronomy Search Data (CU - Cordes)

- *What?*
  - Low-latency/on-demand search for time-varying, including one-time, signals in radio astronomy data
  - Build on existing Cornell/CAC collaboration*
- *Why Aristotle?*
  - Burst on-demand to analyze new data
  - High-speed transport of very large data sets into and around the federation

** "A repeating Fast Radio Burst," Nature 531, 202–205 (10 March 2016)*

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Use Case 5: Water Resource Management Using OpenMORDM (CU - Reed)

- *What?*
  - Simulate complex water resource management outcomes on-demand to achieve substantial beneficial environmental impact on municipalities located throughout the US
- *Why Aristotle?*
  - Start coarsely parallel computations on-demand
  - Very large computations, ideal for transfer to AWS or other clouds
  - Make available to many municipalities

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Use Case 6: Mapping Transcriptome Data to Metabolic Models of Gut Microbiota (CU - Douglas)

- *What?*
  - Use host and gut-microbial expression data to infer which metabolic reactions occur in each organism as a result of dietary inputs
- *Why Aristotle?*
  - Heterogeneous instance types and sizes with high availability to allow for unpredictable computational demands
  - Allows generation of reproducible computational biology pipelines in the form of VMs or VM configurations

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Use Case 7: Multi-Sourced Data Analytics to Improve Food Production (UCSB/Sedgwick Reserve - McCurdy)

- *What?*
  - Interdisciplinary analysis of data from a wide variety of sources (including the general public) to study and improve ecological outcomes
- *Why Aristotle?*
  - Scalable and portable infrastructure
  - Access to multiple data sources, many of which are already in public and private clouds

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

Outreach Activities

- **CASC Presentation**
  Lifka presented info on Aristotle at the fall Coalition for Academic Scientific Computation (CASC) meeting in Arlington. 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.

- **Media Outlets/Social Media**
  Generated media coverage across 23 global media outlets including major IT outlets (CIO, Campus Technology, Government Technology News, NetworkWorld) as well as HPC verticals (ACM TechNews, HPCwire, InsideBIGDATA, Scientific Computing World). A strong social media presence (Twitter, LinkedIn, etc.) was generated in collaboration with NSF, AWS, and other project partners.

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

- **SC15 Conference**
  The Aristotle federation was featured at Cornell's and UB's SC15 exhibits and news and backgrounders were distributed to attendees.

Outreach Activities (cont.)

- **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
  Scientific Computing World - Will the cloud change scientific computing?
  Web Host Industry Review - National Science Foundation Sponsors $5 Million Cloud Project

- **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'

- **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

Sharing and Working with the Community

- We're developing and identifying cloud-related tools
  - QBETS
    http://www.cs.ucsb.edu/~nurmi/nurmi_qbets.pdf
  - Open XDMoD with cloud support
    http://xdmod.sourceforge.net/
  - CloudLaunch (potential work with Cycle Computing and AWS)
    https://www.cac.cornell.edu/technologies/CloudLaunch.pdf
  - Supercloud (testing on Jetstream)
    http://fireless.cs.cornell.edu/publications/supercloud.pdf
- We want to create a "Cloudy" community focus group
  - Identify and share cloud tools, services, and solutions
  - Leverage new computing and scaling paradigms (e.g., containerization)
  - Help each other and the greater research community achieve faster time to science

Award No. 1541215

ARISTOTLE CLOUD FEDERATION

ARISTOTLE

CLOUD FEDERATION

Award No. 1541215