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

PI: David Lifka, Cornell; Co-PIs: Rich Wolski, UC Santa Barbara; Tom Furlani, University at Buffalo

Award #: 1541215

Aristotle Goals

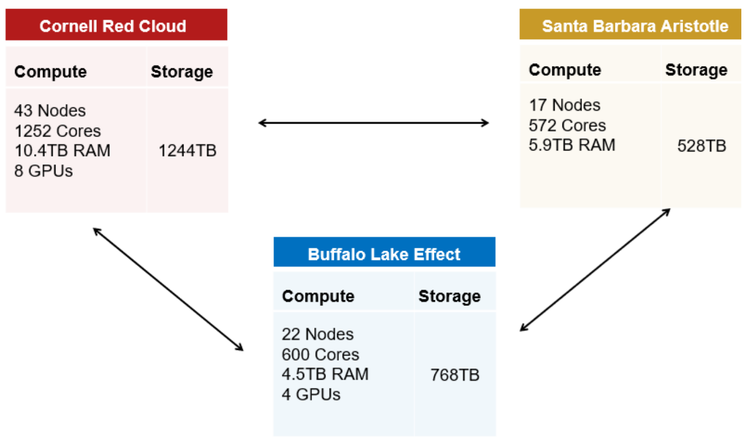
ARISTOTLE CLOUD FEDERATION

- Share cloud resources between institutions.
- Enable cross-institution allocations.
- Develop federated accounting and cloud metrics tools; build into Open XDMoD.
- Containerize intentionally diverse science use case applications.
- Burst to public or NSF clouds during peak usage.
- Forecast Spot market prices with a new prediction tool: DrAFTS 2.0
- Deliver federated cloud tools and services via the Aristotle portal.

federatedcloud.org

Federated Model

- Federation located at Cornell (CU), UC Santa Barbara, and University at Buffalo (UB).
- OpenStack software, Ceph storage.
- Total model deployment = 2424 cores (including campus resources) and 2540TB storage.



New Federation Members

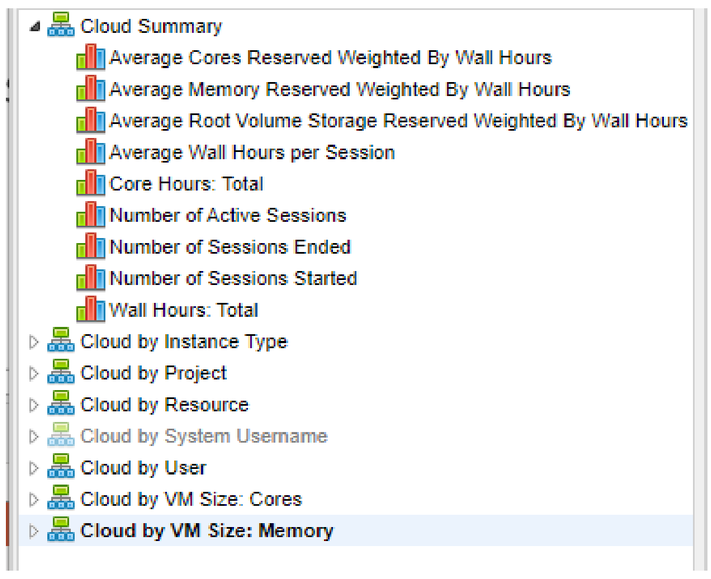
- Members quickly learn from one another rather than going it alone when deploying campus clouds.
- For example: Dartmouth joined the Aristotle federation and is deploying their first OpenStack cloud. New use cases will include digital humanities and cloud-based Jupyter notebooks.
- Aristotle partnered with XSEDE CRI to provide cloud implementation documentation and services.

Industry Participants

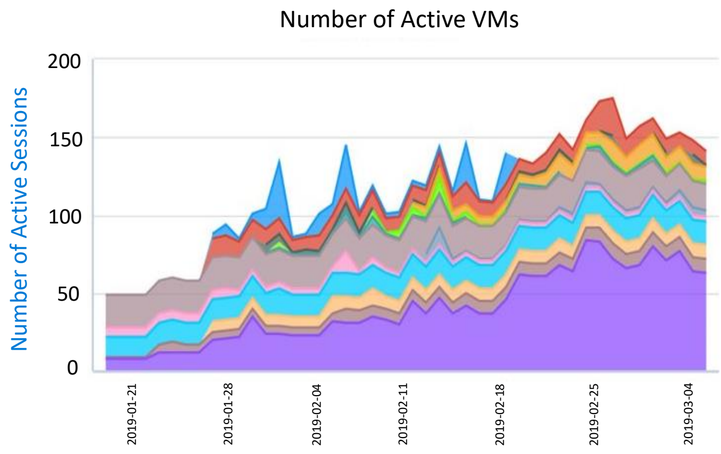
- Red Hat
- AWS
- Azure
- Google Cloud
- Globus
- Dell
- HPE
- RightScale

Federated Cloud Metrics

- Developed federated version of Open XDMoD with Aristotle cloud metrics.



- Data from all 3 sites available at Open XDMoD Federation Master instance.
- Stakeholders can see usage data displayed through an intuitive graphical interface.



Number of active virtual machines per project across all 3 clouds (2019 Q1)

- Built federated accounting database to encompass cloud resources, allocations, and usage by site.
- Users manage projects via the portal dashboard where they access all federation sites with single sign-on, see cloud metrics, and see availability across sites.

New Technologies

- Examples: *CSPOT* – a portable, multi-scale, serverless platform for IoT apps; *Centaurus* – cloud service for K-means clustering; *Multi-Cloud Run* – uses Python, Celery, and other tools to run apps across multiple cloud sites.
- See 85 publications and conference presentations at federatedcloud.org.

Cloud Marketplace Investigation

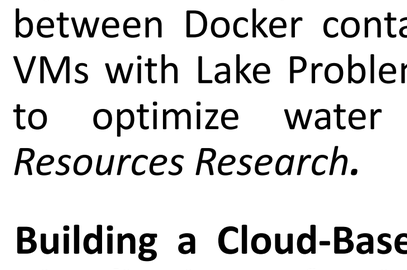
- Analyzing cost differences of running Linear Algebra benchmarks, WRF, OpenMORDM, and Terrraform and Ansible automated cluster creation on Aristotle, AWS, Azure, and Google.

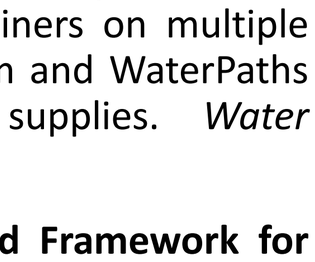
Workforce Development

- 16 REUs plus grad students. Nevena Golubovic awarded PhD by UCSB based on her Aristotle work.

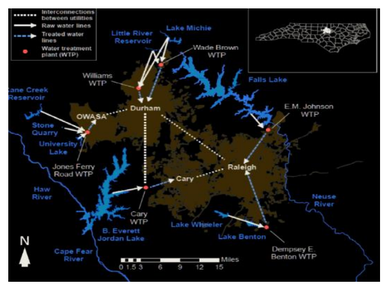
Advances in Science

Application Containerization: All Cornell science use cases containerized. UB containerized app kernels in Docker (HPCC, HPCG, NAMD, NWChem, GAMESS) and ran on OpenStack.





Application of WRF Model in the Cloud (S. Pryor et al.) – 150TB runs to assess year-to-year wind variability. Enhanced wind farm wake simulations to maximize wind turbine power production. *Journal of Physics.*



Water Resource Management Using OpenMORDM (P. Reed et al.) – MPI runs between Docker containers on multiple VMs with Lake Problem and WaterPaths to optimize water supplies. *Water Resources Research.*

Building a Cloud-Based Framework for Visualization of Big Geo Data (V. Chandola et al.) – UB developed and released the *webGlobe* framework that allows researchers to visualize/analyze NetCDF data sets on Aristotle. Currently enhancing with ML. *BigSpatial Journal*.

Transient Detection in Radio Astronomy Data (J. Cordes et al.) – CU created a container and pipeline for FRB detection.

Global Market Efficiency (D. Roesch et al.) – UB set up VMs with a 25TB DB of all US stock transactions and tick by tick data to analyze day vs. night prices and liquidity. *Journal of Financial Economics*.

Multi-Sourced Data Analytics to Improve Food Production & Security (K. McCurdy, C. Krintz, et al.) – California citrus is threatened by HLB citrus greening caused by the Asian citrus psyllid. Aristotle will provide the computational infrastructure necessary to analyze effects of growing Citrus Under Protective Screening (CUPS).

Mapping Transcriptome Data to Metabolic Models of Gut Microbiota (A. Douglas et al.) – CU migrated a Windows VM to Linux and using SteadyCom, investigated the scope of metabolic interactions that occur in the Drosophila gut microbial community. Identified the exchange of 159 unique metabolites. *mBio Journal.*