



Red Cloud and the Aristotle Cloud Federation

Resa Reynolds

Cornell University Center for Advanced Computing (CAC)

resa.reynolds@cornell.edu

Cornell University Center for Advanced Computing (CAC)

- Cost Recovery Center (core facility)
- Mission: To accelerate discovery and broaden impact
- Services: Computing, Consulting, Training and Education
 - Systems: HPC clusters, storage leases, Red Cloud (IaaS)
 - Consulting: data management, parallel programming, visualization
 - Training and Education: Cornell Virtual Workshops
- Research Impact
 - Support Cornell faculty with over \$100 million in research funding from NSF, NIH, USDA, DOE, NASA ...
 - Management roles in national cyberinfrastructure program
 - NSF Computer, Information Science & Engineering Advisory Committee

Red Cloud

- Launched 2011 as a Pilot
- Motivation
 - Workloads not suitable for HPC batch queues
 - Development platform (sandbox)
- Infrastructure as a Service
 - 2 Dell C6000 chassis, 8 nodes/96 cores
 - Enable use of a self-service science infrastructure
 - No CPU/RAM over-subscription
 - Sold as a subscription model to limit accidental research budget



Red Cloud infrastructure

- Cloud stack: OpenStack
- Compute: 39 nodes (Dell C6xxx), 1092 CPU cores
 - Instance types up to 28 cores/240GB RAM
 - Just added GPU option, PE940 with 56core/240GB RAM and 4 V100 GPUs
- Storage
 - Ceph cluster with ~1.3 PB capacity (Dell PE740xd)
- Networking
 - 10 Gbit Ethernet interconnect between cloud components and a 10Gbit connection to the world
- Provisioning tool: Puppet



On-demand scalable infrastructure, deployed in minutes

- Red Cloud gives users a fully customizable computing resource
 - Many instance sizes to choose from
 - Root access to instances
 - Block storage allocated in increments of GB
 - Define network access policies via security groups
 - 3 modes of accessing the cloud: web console, command line client, API
- Application examples
 - Develop/test code and upload production workload to public clouds
 - Web portals/Research Wikis
 - Interactive workloads
 - CAC has virtualized the center's internal infrastructure: web portals, file servers, Nagios monitor, etc.

Red Cloud Security

- Red Cloud installations are on isolated subnets
- Leverages campus IT security staff for edge firewall and network monitoring
- Accounts are integrated with CAC's Active Directory and a fully automated account management system
- Users must define network access policies for their instances
 - **Default policy denies all inbound access**
 - Users must explicitly grant access by protocol, IP address and range
 - User education and easy-to-use management GUI are critical
- Base Linux and Windows images are maintained by CAC staff
- Users are reminded to patch their running instances

Lessons learned

- User education and communications are critical
- Capacity planning
 - Recovery model supports adding more cores when cloud has 60% sustained usage. Usage is sometimes inflated when users don't shutdown instances.
- Advocating to enable research
 - Analyzing Twitter data during Super Bowl overwhelms Internet link



Fast Forward to 2015



- Red Cloud is successful (from 96 to 492 cores in 4 years)
- Cornell users begin looking for additional resources and find public clouds too expensive
- Cornell, Buffalo, and UC Santa Barbara, who each have on-prem private clouds, write and win a NSF CC*DNI DIBBs award to federate the three clouds – dubbed **Aristotle Cloud Federation**

PI David Lifka, Cornell University, lifka@cornell.edu

Co-PI Tom Furlani, U. at Buffalo, furlani@buffalo.edu

Co-PI Rich Wolski, UC Santa Barbara, rich@cs.ucsb.edu

federatedcloud.org

Aristotle Cloud Federation

Project goals

- Optimize “time to science”
- Encourage use of self-service science infrastructure
- Demonstrate the value of sharing resources and data across institutional boundaries
- Sustainable: Only capitalize infrastructure your site can keep busy
- Cross-institution allocations
- Burst to remote federated cloud sites or public clouds during peak usage
- Use Open XDMoD to monitor clouds and provide metrics



Aristotle Cloud Federation

Where are we today?



Aristotle Cloud Federation

- Infrastructure
 - Successfully maintaining cloud and storage assets at each site
 - Globus Auth implemented for single sign-on across federated sites
 - Globus transfer used for moving images and data sets between sites
 - Federated OpenXDMOD collecting usage to a central location



Aristotle Cloud Federation

- Portal – federatedcloud.org
 - Project accounting (add/remove projects/users)
 - Site Status
 - Allocations
 - Usage as provided by OpenXDMoD
 - Science Use Case information



Aristotle Cloud Federation

Research: 7 teams across the 3 sites (more than 40 global collaborators)

- Representing diversity of analysis requirements and cloud usage modalities
 - Earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, food science
- All teams are running successfully across the federation



Aristotle Cloud Federation

Final year goals:

- Migrate research images to containers for portability
- Add a fourth institution to the federation (Dartmouth).
- Investigate ways to provide users better estimates for running in the public cloud.
- Package web portal code and documentation for public consumption.
- Assist those who interested in building their own federations.
- Present results at conferences



Aristotle Cloud Federation

Post NSF Award?

- Aristotle Federation will run beyond the award.
- Opportunities for other institutions to join Aristotle.

Questions?

