Nimbus Update

March 2010

OSG All Hands Meeting

Kate Keahey

keahey@mcs.anl.gov

Nimbus Project
University of Chicago
Argonne National Laboratory
Nimbus: Cloud Computing for Science

- Allow providers to build clouds
  - Workspace Service: a service providing EC2-like functionality
  - WSRF-style and EC2-style (both WS and REST) interfaces
  - Support for Xen and KVM
- Allow users to use cloud computing
  - Do whatever it takes to enable scientists to use IaaS
  - Context Broker: turnkey virtual clusters
  - Currently investigating scaling tools
- Allow developers to experiment with Nimbus
  - For research or usability/performance improvements
  - Open source, extensible software
  - Community extensions and contributions: UVIC (monitoring), IU (EBS, research), Technical University of Vienna (privacy, research)
The Workspace Service
The workspace service publishes information about each workspace.

Users can find out information about their workspace (e.g. what IP the workspace was bound to).

Users can interact directly with their workspaces the same way they would with a physical machine.
Turnkey Virtual Clusters

- Turnkey, tightly-coupled cluster
  - Shared trust/security context
  - Shared configuration/context information
- Context Broker goals
  - Every appliance
  - Every cloud provider
  - Multiple distributed cloud providers
- Used to contextualize 100s of virtual nodes for EC2 HEP STAR runs, Hadoop nodes, HEP Alice nodes...
- Working with rPath on developing appliances, standardization

www.nimbusproject.org
Nimbus CloudKit

- Nimbus: www.nimbusproject.org
- Science Clouds: www.scienceclouds.org
- GitHub for code management
- Latest release is 2.4 (02/02/10)
Science in the Clouds: Update on Applications Working with Nimbus
STAR experiment

- STAR: a nuclear physics experiment at Brookhaven National Laboratory
- Studies fundamental properties of nuclear matter
- Problems:
  - Complexity
  - Consistency
  - Availability

Work by Jerome Lauret, Leve Hajdu, Lidia Didenko (BNL), Doug Olson (LBNL)

www.nimbusproject.org
STAR Virtual Clusters

- **Virtual resources**
  - A virtual OSG STAR cluster: OSG headnode (gridmapfiles, host certificates, NFS, Torque), worker nodes: SL4 + STAR
  - One-click virtual cluster deployment via Nimbus Context Broker
- **From Science Clouds to EC2 runs**
- **Running production codes since 2007**
- **The Quark Matter run**: producing just-in-time results for a conference: [http://www.isgtw.org/?pid=1001735](http://www.isgtw.org/?pid=1001735)
Priceless?

- **Compute costs: $ 5,630.30**
  - 300+ nodes over ~10 days,
  - Instances, 32-bit, 1.7 GB memory:
    - EC2 default: 1 EC2 CPU unit
    - High-CPU Medium Instances: 5 EC2 CPU units (2 cores)
  - ~36,000 compute hours total
- **Data transfer costs: $ 136.38**
  - Small I/O needs: moved <1TB of data over duration
- **Storage costs: $ 4.69**
  - Images only, all data transferred at run-time
- **Producing the result before the deadline...**
  - $ 5,771.37
Modeling the Progression of Epidemics

Work by Ron Price and others, Public Health Informatics, University of Utah

- Can we use clouds to acquire on-demand resources for modeling the progression of epidemics?
- What is the efficiency of simulations in the cloud?
  - Compare execution on:
    - a physical machine
    - 10 VMs on the cloud
    - The Nimbus cloud only
  - 2.5 hrs versus 17 minutes
  - Speedup = 8.81
  - 9 times faster
Ocean Observatory Initiative

- Highly Available Services
- Rapidly provision resources
- Scale to demand
OOI Architecture

- VM
- Application Software
- EPU Worker
- EPU Worker
- EPU Worker
- EPU
- HA Service (OOI Application)
- Availability and Scaling Strategy

www.nimbusproject.org
Elastically Provisioned Resources

- CHEP 2009 paper, Harutyunyan et al., Collab. with CernVM

www.nimbusproject.org
Hadoop in the Science Clouds

Papers:


www.nimbusproject.org
Keeping the Blood Flowing

Work by Karnidakis & Sherwin and Dong & Karonis (Brown, Imperial College, NIU)

- **Arterial Tree Models**
  - The NECTAR project
  - Simulating the blood flow in ALL major arteries of the human.
  - Running across TeraGrid

- **Investigating cloud computing**
  - Environments
  - Network
  - Efficiency

www.nimbusproject.org
Nimbus for FutureGrid

- **FutureGrid**: experimental testbed
- **Infrastructure-as-a-Service for FutureGrid**
  - Dynamic cloud deployment
  - Improved accounting, administration, user experience
- **Continue to provide IaaS cycles for science**
  - Encourage new and experimental uses of the infrastructure
  - Continue reaching out to scientists in the US and collaborators abroad
- **Develop new capabilities**
  - Automate building of “sky computing” scenarios
  - Provide additional tools on the cloud federation level

www.nimbusproject.org
Nimbus: Friends and Family

- **Nimbus core team:**
  - UC/ANL: Kate Keahey, Tim Freeman, David LaBissoniere, John Bresnahan
  - UVIC: Ian Gable & team:
    - Patrick Armstrong, Adam Bishop, Mike Paterson, Duncan Penfold-Brown
  - UCSD: Alex Clemesha

- **Contributors:**
  - http://www.nimbusproject.org/about/people/

- **Other efforts:**
  - ViNe: Mauricio Tsugawa, Jose Fortes (UFL)

www.nimbusproject.org