Cloud Computing with Nimbus

November 2009
Cloud Computing for Systems and Computational Biology
Kate Keahey
keahey@mcs.anl.gov
Nimbus project lead
University of Chicago
Argonne National Laboratory
Nimbus: Cloud Computing Software

- Allow providers to build clouds
  - Workspace Service: a service providing EC2-like functionality
  - WSRF and WS (EC2) interfaces
- Allow users to use cloud computing
  - Do whatever it takes to enable scientists to use IaaS
  - Context Broker: turnkey virtual clusters,
  - Also: protocol adapters, account managers and 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)
- Nimbus: http://workspace.globus.org
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
Science Clouds

● Goals
  ◆ Enable experimentation with IaaS
  ◆ Evolve software in response to user needs
  ◆ Exploration of cloud interoperability issues

● Participants
  ◆ University of Chicago (since 03/08), University of Florida (05/08, access via VPN), Wispy @ Purdue (09/08)
  ◆ International collaborators
  ◆ Using EC2 for large runs

● Science Clouds Marketplace: OSG cluster, Hadoop, etc.

● 100s of users, many diverse projects ranging across science, CS research, build&test, education, etc.

● Come and run: [http://scienceclouds.org](http://scienceclouds.org)
STAR experiment

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

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

Infrastructure-as-a-Service
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
A Large Ion Collider Experiment (ALICE)

Work by Artem Harutyunyan and Predrag Buncic, CERN

- Heavy ion simulations at CERN
- Problem: integrate elastic computing into current infrastructure
- Collaboration with CernVM project
- Elastically extend the ALICE testbed to accommodate more computing
Elastic Provisioning for ALICE HEP

- Queue Sensor
- ALICE queue
- AliEn
- Context Broker

Amazon and other commercial providers
Nimbus and other Science Clouds

Infrastructure-as-a-Service
Elastically Provisioned Resources

- **CHEP09 paper, Harutyunyan et al.**
- **Elastic resource base: OOI, ATLAS, ElasticSite, and others**
Sky Computing Environment

Work by A. Matsunaga, M. Tsugawa, University of Florida

Creating a seamless environment in a distributed domain
Hadoop in the Science Clouds

Papers:
Parting Thoughts

- **IaaS cloud computing is science-driven**
  - Scientific applications are successfully using the existing infrastructure for production runs
  - Promising new model for the future
- **We are just at the very beginning of the “cloud revolution”**
  - Cloud computing is not “done”
  - Significant challenges in building ecosystem, security, usage, price-performance, etc.
- **Lots of work to do!**
Nimbus: Friends and Family

- **Nimbus core team:**
  - UC/ANL: Kate Keahey, Tim Freeman, David LaBissoniere
  - UVIC: Ian Gable & team
  - UCSD: Alex Clemesha

- **Related technologies**
  - EBS: Marlon Pierce, Xiaoming Gao, Mike Lowe (IU)
  - ViNe: Mauricio Tsugawa, Jose Fortes (UFL)
  - Others:
    - Descher et al (Technical U of Vienna): privacy extensions