Experiments with Complex Scientific Applications on Hybrid Cloud Infrastructures

Maciej Malawski$^{1,2}$, Piotr Nowakowski$^1$, Tomasz Gubała$^1$, Marek Kasztelnik$^1$, Marian Bubak$^{1,2}$, Rafael Ferreira da Silva$^3$, Ewa Deelman$^3$, Jarek Nabrzyski$^4$

NSFCloud Workshop on Experimental Support for Cloud Computing
December 11-12, 2014, Arlington, VA

AGH University of Science and Technology:
$^1$ ACC Cyfronet AGH, ul. Nawojki 11, 30-950 Kraków, Poland
$^2$ Department of Computer Science, al. Mickiewicza 30, 30-095 Kraków, Poland
$^3$ University of Southern California, Information Sciences Institute, Marina Del Rey, CA, USA
$^4$ Center for Research Computing, University of Notre Dame, IN, USA
Research Challenges

• Execution of complex scientific applications on clouds: workflows and their ensembles
  • Pegasus Workflow Management System (OCI SI2-SSI #1148515)
  • HyperFlow Workflow Engine

• Platform for deployment and sharing of scientific applications on hybrid clouds
  • Atmosphere Framework

• Algorithms for scheduling, provisioning and cost optimization:
  • Dynamic and Static Algorithms
  • Mathematical Programming
  • Cloud Workflow Simulator
Research: The Atmosphere Framework
Hybrid cloud as a means of provisioning computing power for virtual experiments

GUI host (provisions end-user features and access options)

Cloud Management Portlets
Provide GUI elements which enable service developers and end users to interact with the Atmosphere platform and create/deploy services on the available cloud resources

Secure RESTful API (Cloud Facade)

Atmosphere Core Services Host

Atmosphere Core
- Authentication and authorization logic
- Communication with underlying computational clouds
- Launching and monitoring service instances
- Creating new service templates
- Billing and accounting
- Logging and administrative services

Atmosphere Registry (AIR)
- User accounts
- Available cloud sites
- Services and templates

OpenStack cloud site at ACC CYFRONET AGH
- 96 CPU cores
- 184 GB RAM
- 4 TB storage
- Private IP space

VPH-Share cloud site at UNIVIE
- 128 CPU cores
- 256 GB RAM
- 4 TB storage
- Private IP space

Amazon Elastic Compute Cloud (EC2) – European availability zone
- Massive (functionally limitless) hardware resource pool
- Public IP space
Research: Simulation and Scheduling of Large-Scale Scientific Workflows on IaaS Clouds

• Large-scale scientific workflows from Pegasus WMS
  • Workflows of 100,000 tasks

• Workflow Ensembles
  • Schedule as many workflows as possible within a budget and deadline
  • Uses a Cloud Workflow Simulator

Research: Cost Optimization of Applications on Clouds

- Infrastructure model
  - Multiple compute and storage clouds
  - Heterogeneous instance types

- Application model
  - Bag of tasks
  - Multi-level workflows

- Modeling with AMPL and CMPL
  - Modeling Language for Mathematical Programming

- Cost optimization
  - Under deadline constraints

- Mixed integer programming
  - Bonmin, Cplex solvers

---

M. Malawski, K. Figiela, J. Nabrzyski, Cost minimization for computational applications on hybrid cloud infrastructures, Future Generation Computer Systems, 29(7), 2013, pp.1786-1794, [http://dx.doi.org/10.1016/j.future.2013.01.004](http://dx.doi.org/10.1016/j.future.2013.01.004)

M. Malawski, K. Figiela, M. Bubak, E. Deelman, J. Nabrzyski, Cost Optimization of Execution of Multi-level Deadline-Constrained Scientific Workflows on Clouds. PPAM, 2013, 251-260 [http://dx.doi.org/10.1007/978-3-642-55224-3_24](http://dx.doi.org/10.1007/978-3-642-55224-3_24)
Research: Cloud Performance Evaluation

- Performance of VM deployment times
  - Virtualization overhead
- Evaluation of open source cloud stacks
  - Eucalyptus, OpenNebula, OpenStack
- Survey of European public cloud providers
- Performance evaluation of top cloud providers
  - EC2, RackSpace, SoftLayer
  - A grant from Amazon has been obtained
Experiment: Evaluation of autoscaling techniques for Atmosphere cloud platform

• Challenges
  • Requires repeated tests under varying workloads
  • Experiments in an isolated environment

• Goals
  • Perform autoscaling based on:
    • Complex event processing
    • Time series database
  • Build an isolated environment on NSFCloud
Experiment: Scalability of Scientific Workflows in HyperFlow Model

• Challenges
  • Issues on data transfers and data locality
  • Calibrate the performance models of applications

• Goals
  • Execute large-scale deployments on multi-site NSFCloud facilities
  • Assess the impact of network latency and bandwidth limitations
Experiment: Influence of Variability of Clouds on the Quality of Algorithms

• Challenges
  • Static scheduling methods assume that the estimates of task runtimes are available
  • The runtime variations and various uncertainties influence the actual execution

• Goals
  • A large-scale experimental testbed will allow investigating the influence of the uncertainties
  • Development of new models to mitigate uncertainties negative effects
Experiment: Interoperation of Cloud Testbed of PL-Grid Infrastructure with NSFCloud

- **PL-Grid**
  - One of the largest national grid infrastructures in Europe (2500+ users, 500+ teams)
  - Cloud testbed based on OpenNebula and OpenStack

- **Goals**
  - Possibility to run transatlantic and global-scale experiments
  - Evaluation of impact of wide-area and high-latency networks
Experiments with Complex Scientific Applications on Hybrid Cloud Infrastructures

Thank you.

DICE Team at AGH: http://dice.cyfronet.pl
Center for Research Computing at Notre Dame: https://crc.nd.edu
Pegasus Team at USC: http://pegasus.isi.edu