

Experiments with Complex Scientific Applications on Hybrid Cloud Infrastructures

Maciej Malawski^{1,2}, Piotr Nowakowski¹, Tomasz Gubała¹, Marek Kasztelnik¹, Marian Bubak^{1,2}, <u>Rafael Ferreira da Silva³</u>, Ewa Deelman³, Jarek Nabrzyski⁴

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

AGH University of Science and Technology:

- ¹ACC 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



Information Sciences Institute



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



School of Engineering Information Sciences Institute

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

School of Engineering

Information Sciences Institute

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





M. Malawski, G. Juve, E. Deelman, J. Nabrzyski: Cost- and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds. SC 2012: 22



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, <u>http://dx.doi.org/10.1016/j.future.2013.01.004</u>

USCViterbi

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 <u>http://dx.doi.org/10.1007/978-3-642-55224-3_24</u>



5



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

Instance prices

h11.4.Marge co2.8.Marge co2.8.Marge m2.4.Marge m3.2.Marge m3.2.Mar

0.0 0.5 1.0 1.5 2.0 2.5 3.0

School of Engineering

A grant from Amazon has been obtained •

0

50



3.5 hi1.4xlarge 3.0 cc2.8xlarge 2.5 cg1.4xlarge 2.0 Price in \$ m2.4xlarge rs-30GB 1.5 m3 2x arge rs-15GB m2.2xlarae 1.0 1 x arge n1 x arg rs-8GB 0.5 rs-4GB m1.large rs-2GB rs-1GB ec1 medium m1.small n1.medium 0.0 rs-0.5GE 100 150 200

Computing time in seconds

M. Bubak, M. Kasztelnik, M. Malawski, J. Meizner, P. Nowakowski, S. Varma, Evaluation of Cloud Providers for VPH Applications, poster at CCGrid2013, Delft, the Netherlands, pp.13-16, 2013

6

200

150

Computing time

100

Computing time in seconds





Price in \$

Single-core price vs. time

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:
 - <u>Complex event processing</u>
 - <u>Time series database</u>
 - 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 <u>multi-site NSFCloud facilities</u>
 - Assess the impact of <u>network latency</u> and <u>bandwidth limitations</u>



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

School of Engineering

Information Sciences Institute

- A large-scale experimental testbed will allow investigating the <u>influence of the uncertainties</u>
- Development of new models to <u>mitigate</u> <u>uncertainties negative effects</u>





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



