

Experiments with Complex Scientific Applications on Hybrid Cloud Infrastructures

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

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



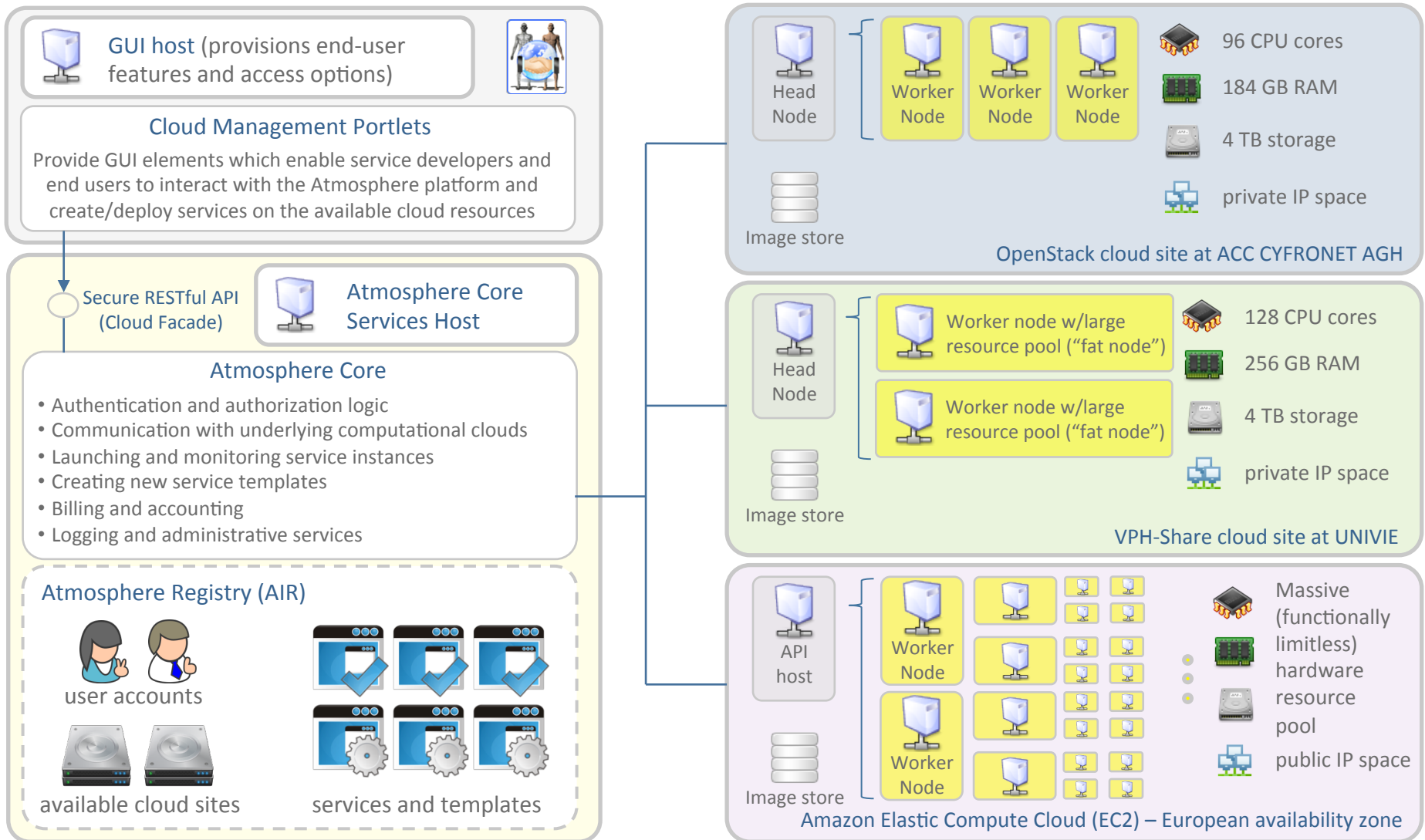
Research Challenges

- 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



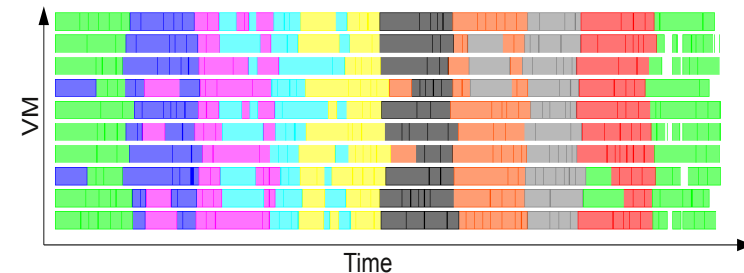
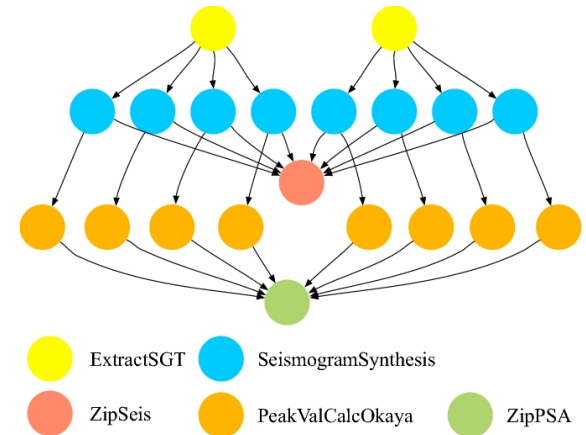
Research: The Atmosphere Framework

Hybrid cloud as a means of provisioning computing power for virtual experiments



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

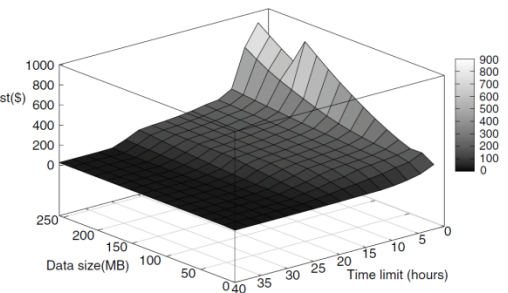
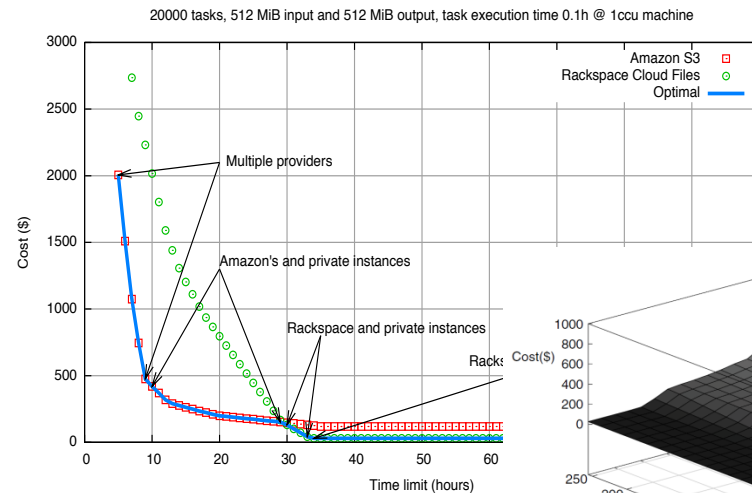
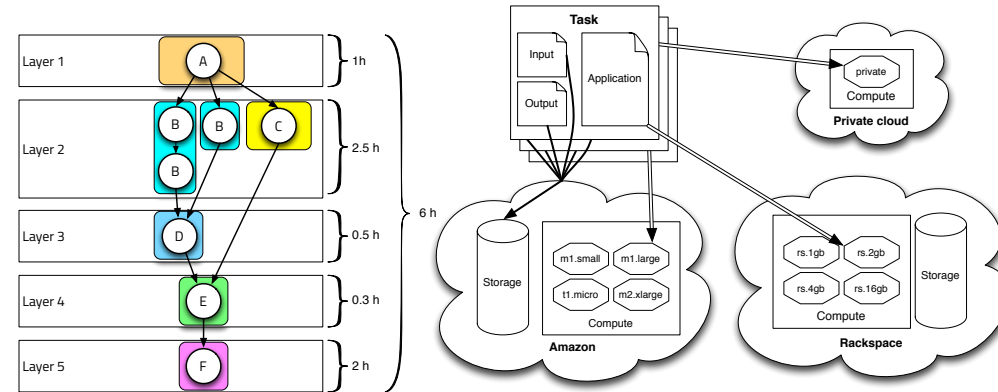


M. Malawski, G. Juve, E. Deelman, J. Nabrzyski: Cost- and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds. SC 2012: 22



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

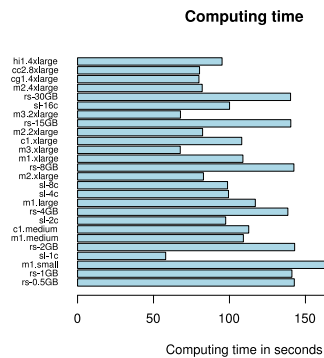
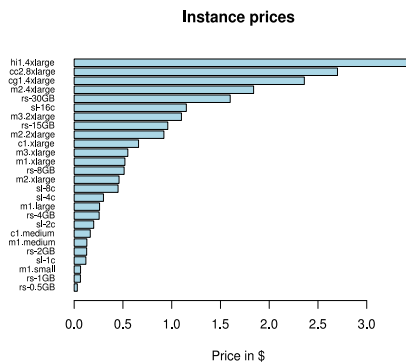
M. Malawski, K. Figiela, M. Bubak, E. Deelman, J. Nabrzyski, *Cost Optimization of Execution of Multi-level Deadline-Constrained Scientific Workflows on Clouds*. PPAM, 2013, 251-260 http://dx.doi.org/10.1007/978-3-642-55224-3_24



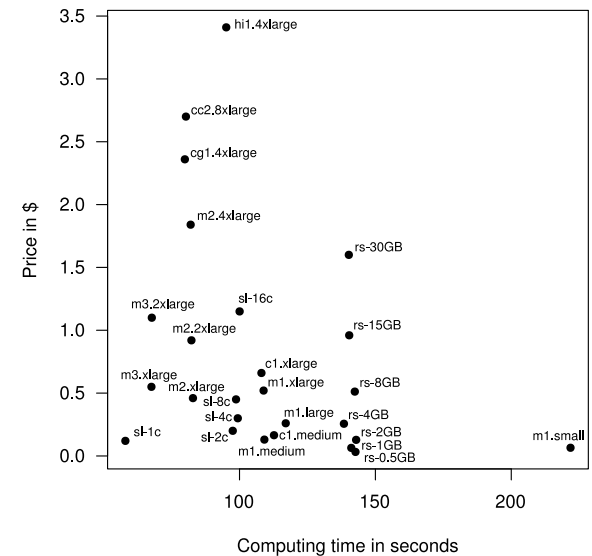
Research: Cloud Performance Evaluation

- 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

IaaS Provider	EEA Zoning	jClouds API Support	BLOB storage support	Per-hour instance billing	API Access	Published price	VM Image Import / Export	Relational DB support	Score
Weight	20	20	10	5	5	5	3	2	
1 Amazon AWS	1	1	1	1	1	1	0	1	27
2 Rackspace	1	1	1	1	1	1	0	1	27
3 SoftLayer	1	1	1	1	1	1	0	0	25
4 CloudSigma	1	1	0	1	1	1	1	0	18
5 ElasticHosts	1	1	0	1	1	1	1	0	18
6 Serverlove	1	1	0	1	1	1	1	0	18
7 GoGrid	1	1	0	1	1	1	0	0	15
8 Terremark ecloud	1	1	0	1	1	0	1	0	13
9 RimuHosting	1	1	0	0	1	1	0	1	12
10 Stratogen	1	1	0	0	1	0	1	0	8
11 Bluelock	1	1	0	0	1	0	0	0	5
12 Fujitsu GCP	1	1	0	0	1	0	0	0	5



Single-core price vs. time



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



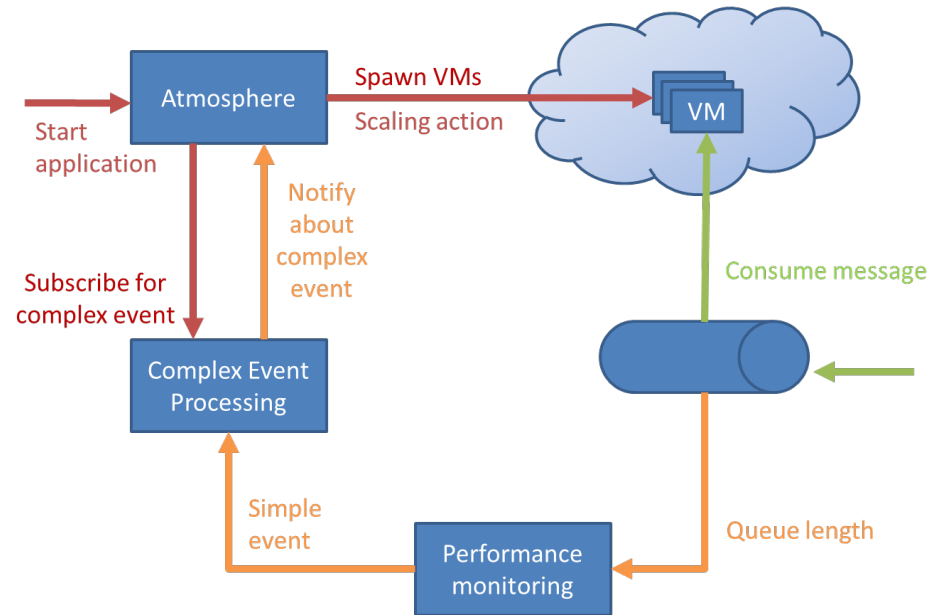
Experiment: Evaluation of autoscaling techniques for Atmosphere cloud platform

- 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



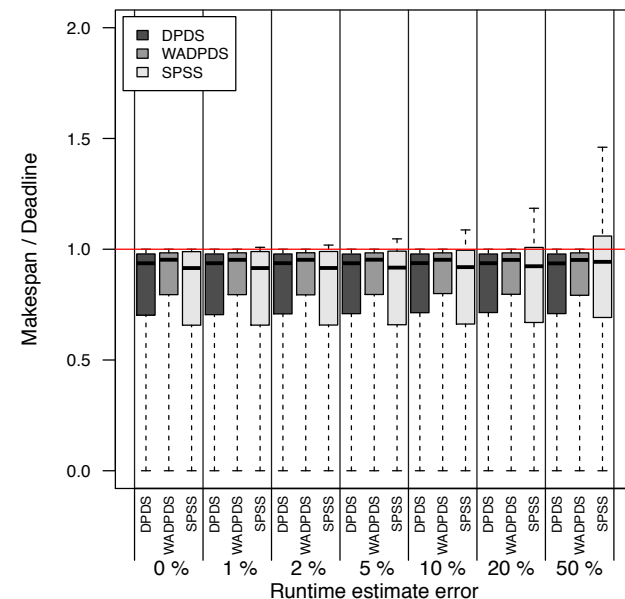
Experiment: Influence of Variability of Clouds on the Quality of Algorithms

- Challenges

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

- Goals

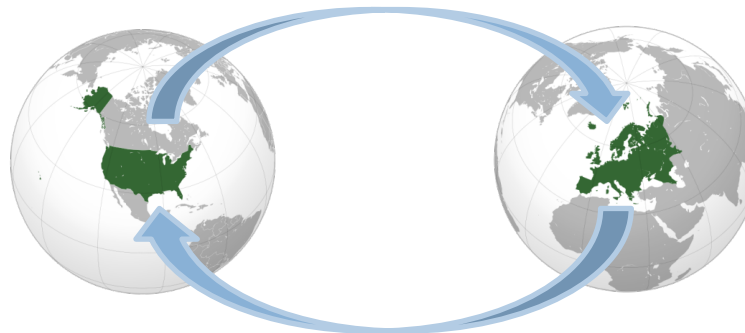
- A large-scale experimental testbed will allow investigating the influence of the uncertainties
- Development of new models to mitigate uncertainties negative effects



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>

