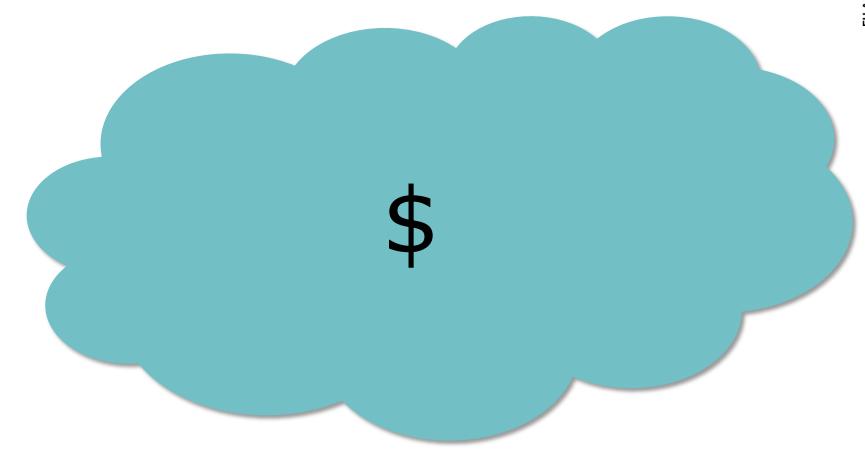
# Adaptive Scheduling of Workflows on Clouds with Deadlines

Michael Gerhards<sup>1,2</sup>, Volker Sander<sup>1</sup>, Adam Belloum<sup>2</sup>, Marian Bubak<sup>2</sup> Cracow Grid Workshop (CWG)- 27/10/2014

# Wanna save Money using the Cloud?



## Temporary Cloud Computing Infrastructure

### **Properties**

- Horizontal scalability
- Vertical scalability
- Typically hourly billing
- On demand provisioning

## Questions

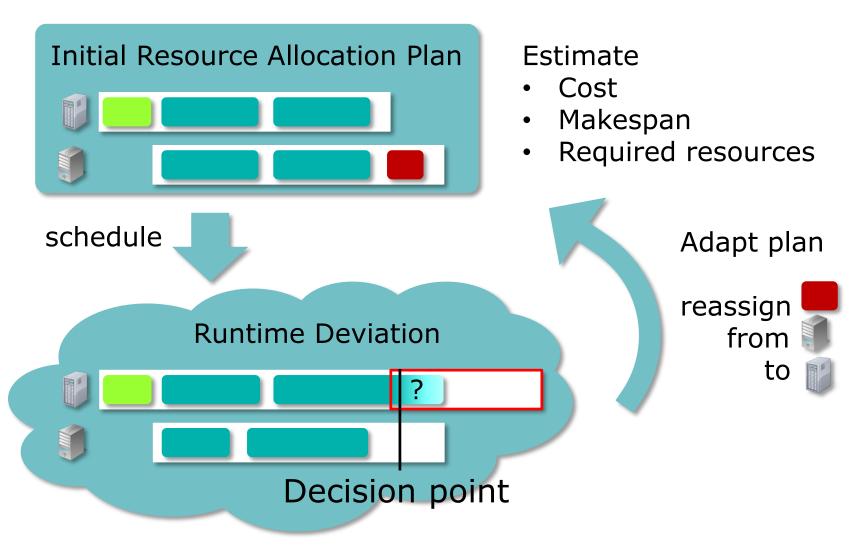
- How many resources?
- Type of resources?
- How long?
- When?

DEADLINE

VS.

Trade-off between deadline and cost constraints

## Idea: Adaptive Scheduling



#### Solution: Visit us at our Poster



Michael



Volker

CGW Workshop 14, Krakow, Poland, October 27-29, 2014

#### Adaptive Scheduling of Workflows on Clouds with Deadlines

Michael Gerhards<sup>1,2</sup>, Volker Sander<sup>1</sup>, Adam Belloum<sup>2</sup>, Marian Bubak<sup>2</sup>

FH Aachen - University of Applied Sciences, Heinrich-Mussmann-Str. 1, 52428 Juelich, Germans - Per Jacom - Omero до и прилого экспеков, frant us reasonaire set, з достой повы http://www.fin-searchen.de/forsetung/projeck/- kidongrapy/veroffendischungen/ - Elmiversily of Amsterdam, Science Park 107, 1096 GH Amsterdam, The Metherlands http://www.mb.http://www.committen.de/people m.gerhards@ft-aachen.de, v.sander@ft-aachen.de, a.s.z.belloum@uva.nl, bubak@agh.edu.pl

#### Problem Statement

Cloud computing offers new scheduling apportunities. Any cloud-aware scheduling algorithm has to decide on the number of actually used resources Could computing offers new scheduling oper-strainties. Any cloud-evaner scheduling apportion has to decode on the number of actually used resources and for each resource about its instead scheduling apportion and direct scheduling control evaners. Healther static nor dynamic scheduling strategies allows address the occurrent. Neither static nor dynamic scheduling strategies allows address the district and inharacteristics of a cloud ecosystem. While dynamic scheduling strategies allows address the scheduling strategies allows address the scheduling strategies allows address the scheduling strategies and inharacteristics of a cloud ecosystem. While dynamic scheduling of scheduling or nect on unexpected behaviors and numtime varietions, an initial setup of resources is required. The proposed adaptive scheduling of scheduling or scheduling of scheduling schedu

#### Solution Application Model Cloud Ecosystem Runtime Profiles Cloud infrastructures provide horizontal and vertical scalable on-demand resources following typically an hourly The targeted application models are large scale computational workflows having thousands of tasks and following All thousands of tasks of a workflow template only follow few task types. A runtime profile for each task type is well-structured data dependencies such billed pay-as-you-go cost model. Any cloud-aware scheduling algorithm has to created based on gathered monitoring/provenance data of previous as the templates of the Pegasus workflows Epigenomics, Cybershake, Inspiral, Montage, and SIPHT. The maximal makespan is limited by a deadline. runs. The runtime profile of a task type also depends on its multithreading capabilities that are applicable on a particular cloud resource. decide on the number of actually used resources and for each resource about its instance size and occupation time to address the occurred monetary cost. Resource Allocation Plan Static Scheduling Adaptive Scheduling Adaptive Scheduling An adaptation of the resource allocation plan is performed whenever an event occurs, e.g. a runtime deviation exceeds a threshold. Pessible adaptations are scaling up/out to hold the deadline and scaling down/in to reduce cost but also reassigning a task to another resource, Here, the red task is reassigned from The initial resource allocation plan for a particular workflow run is created using a particular workflow run is created using a static scheduling strategy considering the workflow dependency graph, the runtime profiles of the occurring task types, and the cloud billing model as parameters. The workflow engine instantiates on Resource Allocation Plan demand all required cloud resources given by the resource allocation plan. Whenever a resource the second to the first resource to substitute the unused capacity of plan. Whenever a resource finished executing its currently assigned task, the workflow engine schedules the next task for execution selected from the resource allocation substitute the unused capacity of the there started new billing hour. Strategic decision points derived from workflow template's flow structure enable a proactive adaptation for future tasks

Remapper Actual Cloud Resource Usage billing hour started

schedule task on

resource

#### **Runtime Deviation** A task's runtime behavior might deviate

plan assignments.

A task's runtime behavior might deviate from its prediction derived from its runtime profile. Hence, the static resource allocation plan might become uneconomic or might even fail in holding the deadline. Here, the third task hosted by the first resource starts an expensed additional billion beautiquities. unexpected additional billing hour during its execution.

COMMIT/

Deadline conforma ost saving executi if workflow instance

#### Research Objectives

much earlier.

from 🐌

to 👊

- Find deviation thresholds/events to identify adaptation demand for a workflow instance according to its actual execution in a particular cloud economy. Derive strategic decision points from the flow structure of a workflow





Adam



Marian

Michael GERHARDS, M.Sc.

FH Aachen – University of Applied Sciences Campus Jülich

Germany Heinrich-Mußmannstraße 1 52428 Jülich T +49. 241. 6009 53791

M.Gerhards@fh-aachen.de, V.Sander@fh-aachen.de, A.S.Z.Belloum@uva.nl, Bubak@agh.edu.pl

http://www.fh-aachen.de/en/research/projekt-hixforagws/veroeffentlichungen/http://commit-nl.nl/