Modeling of Data Farming Application for Multi-Cloud Execution

Michał Orzechowski, Dariusz Król,
Renata Słota, Jacek Kitowski
AGH

Agenda

1. Data Farming in Scalarm Platform
2. Modeling in Multi-Cloud Systems
3. Scalarm Model in CAMEL
4. Results and Conclusions

PAASAGE

SEVENTH FRAMEWORK PROGRAMME

What is data farming?

*Data farming is a **methodology** based on the idea that by **repeatedly** running a simulation model on a **vast parameter space**, enough output data can be gathered to provide an **meaningful insight** into relation of model's properties and behaviours, with respect to simulation's input parameters.*

PAASAGE

SEVENTH FRAMEWORK PROGRAMME

Data Farming

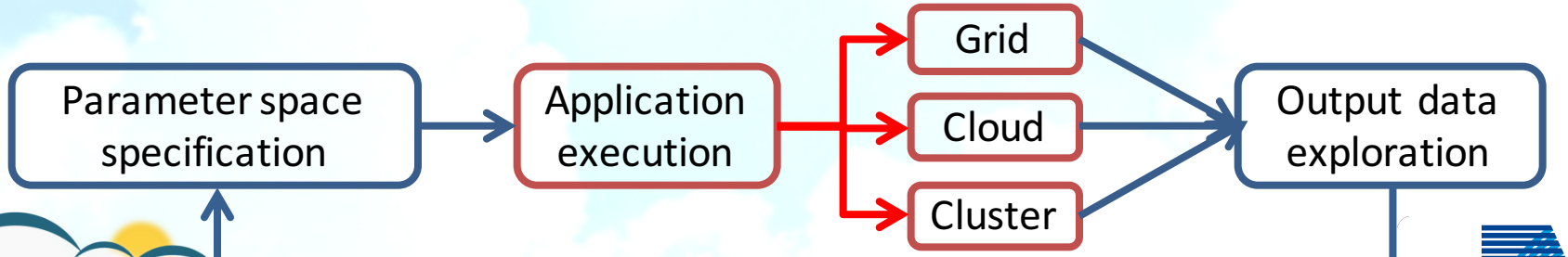
- Provides insights (not answers)
- Decision support
- Study rarely occurring phenomena
  - Military operations
  - Social and cultural behavior modeling
  - Natural disasters
- Heavily dependent on model quality

Scalarm

Massively Self-Scalable Platform for Data Farming

Scalarm Data Farming

- Complete platform for conducting data farming experiments
- Scalable and adaptable to particular problem size and different simulation types
- Supporting online analysis of experiment partial results



PAASAGE

SEVENTH FRAMEWORK PROGRAMME

Modelling Scalarm for Multi-Cloud

Multi-Cloud System

- type of multiple cloud system (usually a Hybrid Cloud: Public + Private)
- acts as intermediary
- follow constrains; like new locations laws
- replicate the applications or services to ensure high availability
- optimize cost or improve quality of service
- a third party provides (builds) a unique entry point to multiple clouds
- not intrusive for Cloud Providers; as compared to Federated Clouds

PAASAGE

SEVENTH FRAMEWORK PROGRAMME

The PaaSage Project

- We are a member of a PaaSage EU FP7 Project
- PaaSage Project:
  - is a multi-cloud system
  - aims to create an open source integrated platform
  - together with an accompanying methodology that will
  - enable model-based development, deployment and management of applications independently of the existing underlying Cloud infrastructures

Most crucially: PaaSage provides a modelling language CAMEL, that serves as a interface for application developer

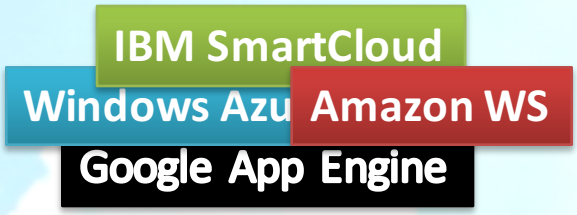
In order to be able to leverage PaaSage one has to model application in CAMEL

PAASAGE

SEVENTH FRAMEWORK PROGRAMME

Model Driven Execution in PaaSage

- PaaSage uses model driven development, for deployment management and execution of software; independently from Cloud providers
- To run an application in PaaSage one have to model it
- The modelling is done using Cloud Application Modelling and Execution Language (CAMEL)



Provider Model

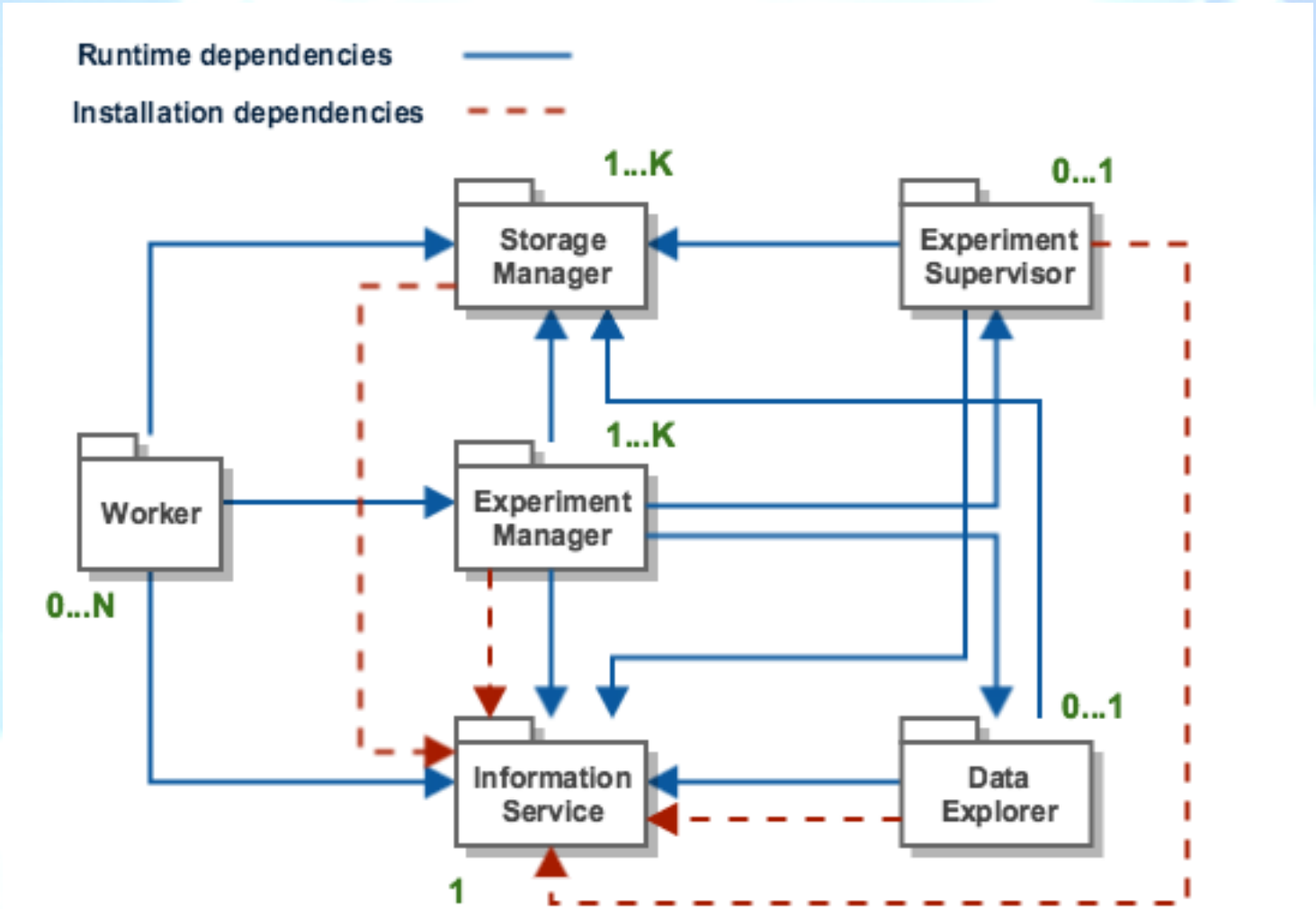
Deployment Model

Scalability Model

Requirements Model

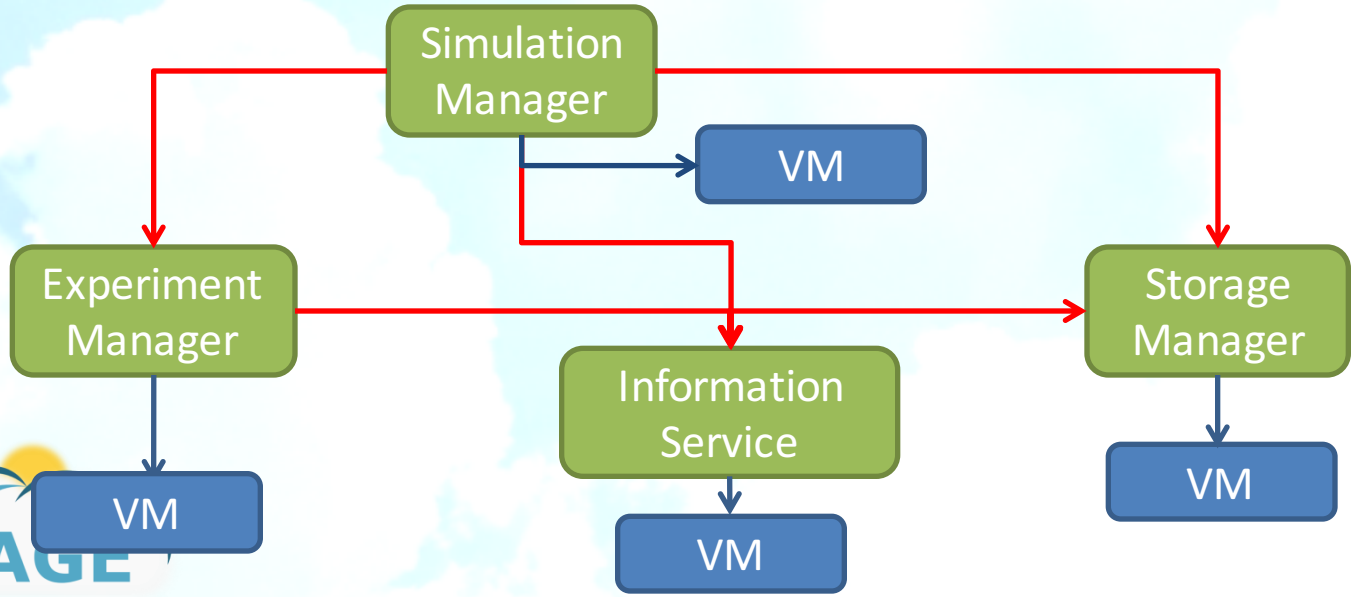
Execution Model

Scalarm Dependencies



Deployment Model

- The Deployment Model corresponds to component based approach, describes how Scalarm services interact with each other; in which order they should be started and on what types of virtual machines.
- Each of Scalam services was modelled so it would run on different virtual machine of specified minimum parameters
- The dependency model is inferred from declared communication requirements between components



Requirements Model

- The Requirements Model defines:
  - platform-level optimization metrics
  - priorities for each metric
  - components which relate to optimization metrics
- We defined optimization metrics for Scalarm:
  - minimize performance degradation of Experiment Manager
  - minimize data faring experiment makespan
  - minimize maximum cost ($) of running data farming experiment

PaaSage Platform is responsible for upholding those optimisation metrics.

Scalability Model

- The Scalability Model uses information from deployment model and defines:
  - scalability rules,
  - actions that should be taken when scalability event is triggered
  - monitored parameters and metrics that aggregate them
- We defined metrics for each of Scalarm services:
  - experiment execution progress metric
  - I/O and disk space metrics
  - response time and CPU utilization metrics
- Listed metrics are responsible for triggering the scalability rules
- Scalability rules describe what should happen when the scalability is triggered - in case of Scalarm extra nodes for each of described services are deployed.

Conclusions

- Modeling languages in multi-cloud systems successfully hide complexity of underlying infrastructures
- The usage of modeling languages for multiple cloud deployments makes sense for large scale deployments
- Modeling languages pose a high learning curve for programmers

PAASAGE

SEVENTH FRAMEWORK PROGRAMME

Thank you!

Michał Orzechowski

Phd Student, Department of Computer Science

AGH University of Science and Technology

PAASAGE

SEVENTH FRAMEWORK PROGRAMME