

Towards Bridging the Gap Between Scalability and Elasticity

Nicolas Ferry, Gunnar Brataas, Alessandro Rossini, Franck Chauvel, Arnor Solberg

SINTEF

CLOSER 2014 – Special session: Multi-Cloud

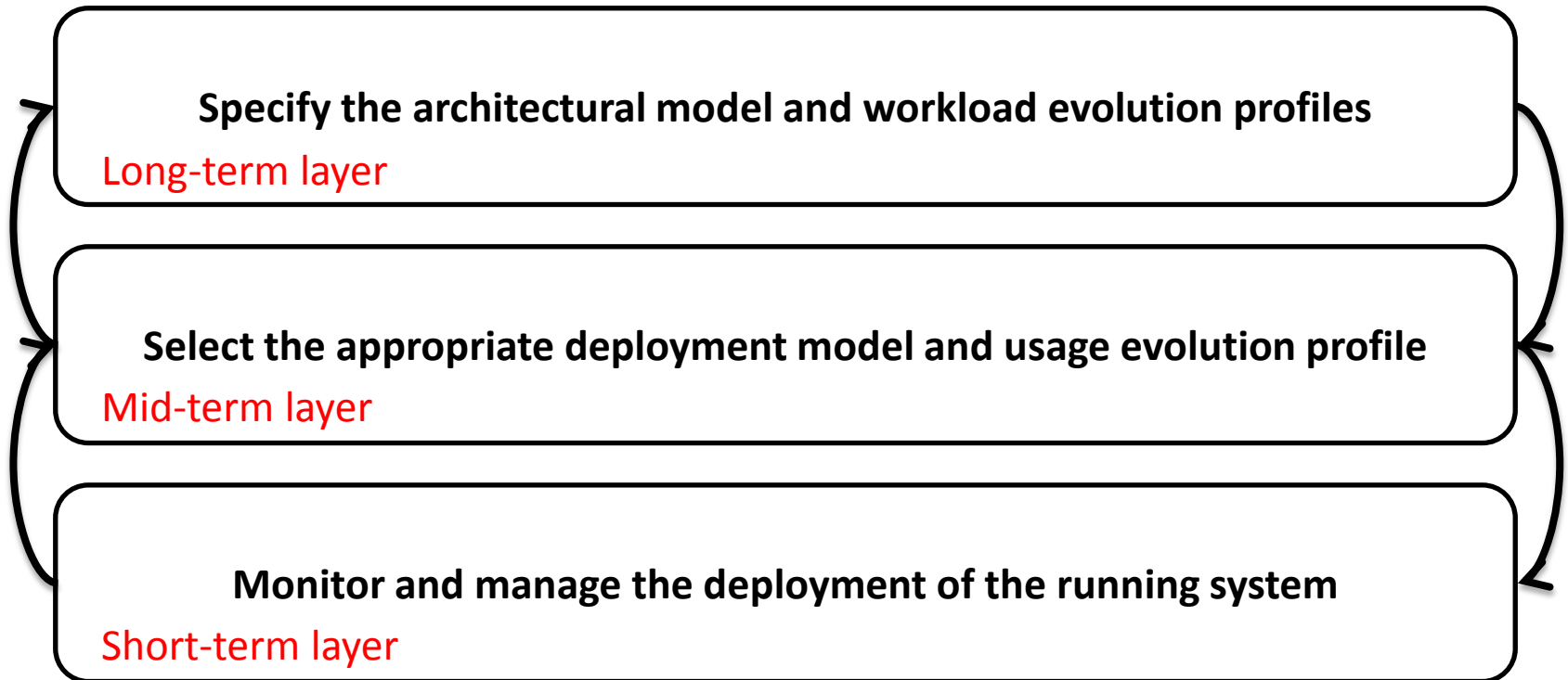
Introduction

Scalability as the ability to fulfill the SLAs against the changing workload

- Typically two independent activities:
 - Reasoning about long-term workload evolution (ScaleDL)
 - Handling short-term workload evolutions (CloudML)
- Model-driven self-management of multi-cloud systems scalability combining these two activities

Architecture

- Inspired by 3-layer architecture from self-adaptive systems



ScaleDL

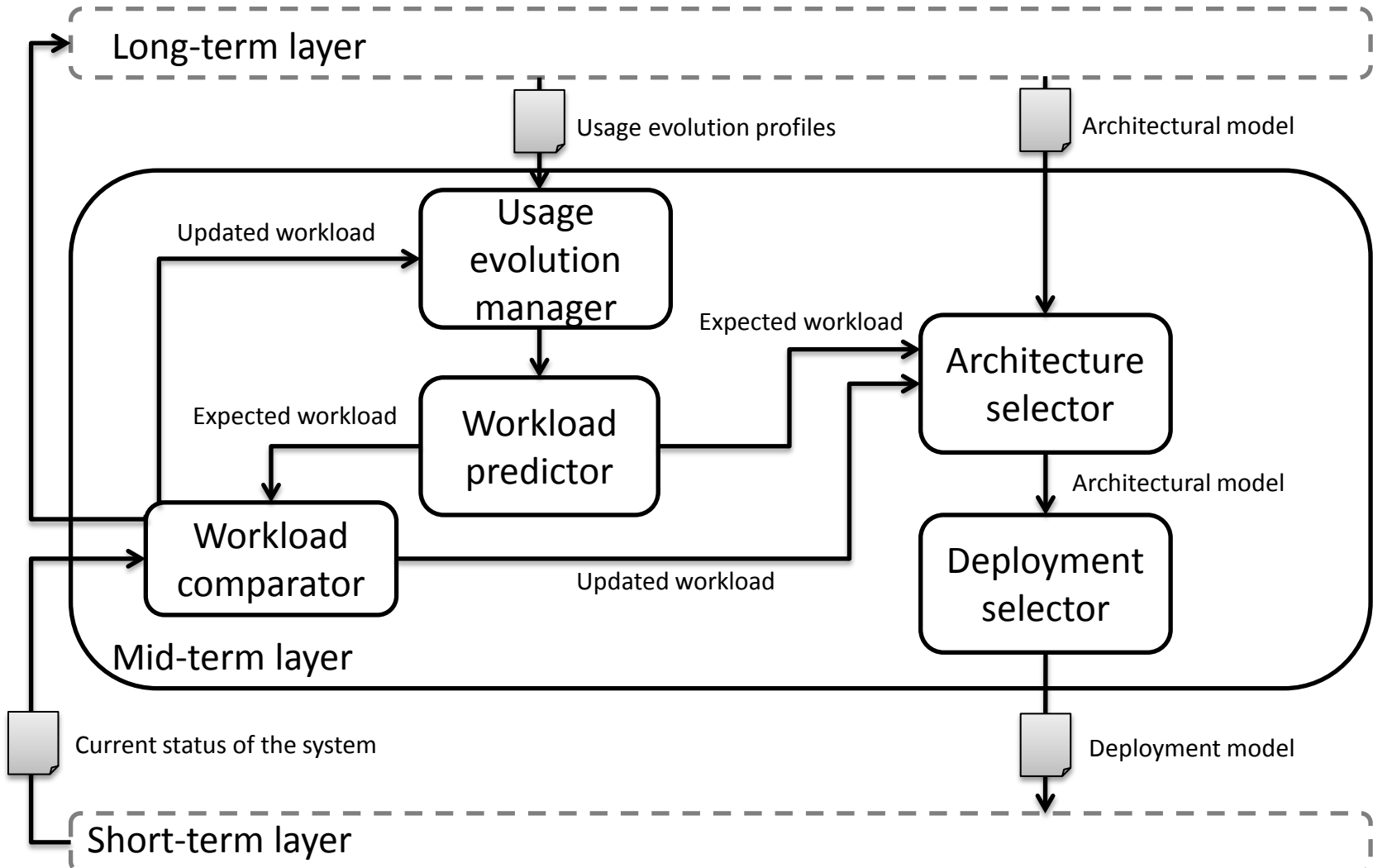
(Long term layer)

- Main interaction point for designers
- ScaleDL leverages upon four sub-languages
 - Usage evolution profiles:
 - Define profile for the evolution of the workload of each service offered by the application
 - Work
 - Load
 - Quality metrics
 - Three forms: **stable, gradual changes and spikes**

Mid-term layer

- Bridge the gap between short and long-term layers
- Responsible for adapting the system on the basis of the workload predictions from long term layer proactively

Mid-term layer



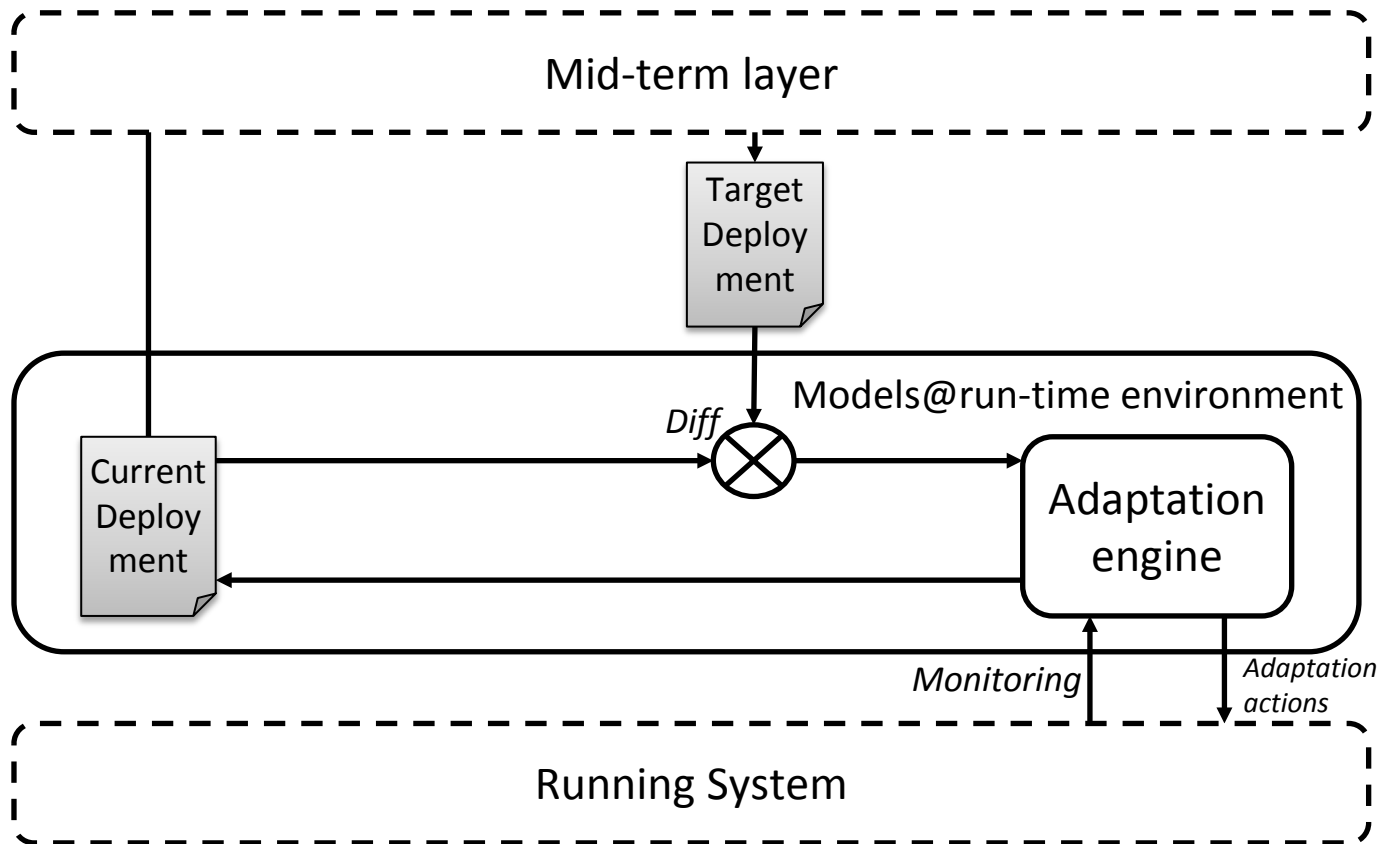
CloudMF

(Short-term layer)

- Two main components:
 - A **modelling environment** with a tool-supported domain-specific modelling language (DSML) to model the provisioning and deployment of multi-cloud systems
 - A **models@run-time environment** for enacting the provisioning, deployment and adaptation of these systems
- Manage workload evolution within the free space defined by the upper layers

Models@runtime

- Causally connected abstract representation of the system



Conclusion

- Approach for the self-management of the scalability concern of multi-cloud systems
- Based on a 3-layer architecture
- Future Work:
 - Finalize implementation
 - Collect past workload variations to improve accuracy of load predictions

Thank you !



Contact: nicolas.ferry@sintef.no