Vrije Universiteit Amsterdam
https://www.overleaf.com/read/wjwrqfvtnjfx

Honours Programme, Problem Statement

Enhancing Datacenter Operations through Digital Twinning as a Service

---

Author: Daniel Ștefan Halasz (2804542)

1st supervisor: Prof. Dr. Ir. Alexandru Iosup

A report submitted in fulfillment of the requirements for the Honours Programme, which is an excellence annotation to the VU Bachelor of Science degree in Computer Science/Artificial Intelligence/Information Sciences

January 21, 2024

Abstract

Datacenter infrastructure has become a fundamental building block of our modern society. The rate of cloud adoption surpasses 90% within economic institutions, while government and public educational entities exhibit a 65% adoption rate [1]. Furthermore, in the Dutch digital economy, every €1 invested in the ICT infrastructure generates a notably higher added value of €15, which is projected to reach €22 by 2025.

Simulation constitutes an indispensable element in the process of operating a datacenter. It facilitates the conduct of scalable experiments involving millions of tasks and hundreds of thousands of computations, as well as the execution of what-if analyses within a time frame spanning minutes to hours [2], forever changing the way we design ICT infrastructure. However, integrating simulation tools into a natural workflow may pose a considerable challenge for any entity willing to manage a datacenter without a preexisting technical proficiency.

This paper aims to create a solution to this outstanding concern, by evaluating an ICT digital twin as a service. The objective of this work is to provide readily accessible means for any individual or organization to conduct analysis, automation, and optimization of the datacenter design process via simulation software, with the intent of minimizing the required technical expertise.

1 Introduction

```
‘‘‘
// TO-DO

1.1. Explain why cloud adoption is required, useful.

1.2. Elaborate the topic of economical impact of the ICT industry.

1.3. Explain why simulation was required in the first
place as a solution to the datacenter optimization problem.

1.4. (Essential) Introduce and elaborate the experience of a novice
data analyst trying to use a simulation tool such as OpenDC.
‘‘‘
```

2 PROBLEM STATEMENT

Currently, operating simulation software demands a high degree of technical expertise. An entity willing to use such software at an advanced level would require (some . . . describe) technical knowledge.

2.1 Describing in very technical, specific terms

1. Basic understanding of *Maven* artifacts and the ability to use a *Gradle* wrapper to compile and publish the OpenDC code base as a local repository, ready to be used as an API.
2. Ability to read Kotlin source code and a good understanding of the OpenDC code base, in order to create a meaningful simulation.
3. Ability to write Kotlin or Java code, compile it, and understand possible error messages.
4. Basic understanding of the Java Virtual Machine (JVM) and how it runs locally.

2.2 Describing in more general terms

1. Familiarity with software development principles and practices.
2. Understanding of how to use and interact with APIs.
3. Knowledge of how to troubleshoot and debug code.

3 Research Questions

```
// TO-DO

Possible research questions discussed

How to offer such a service as SaaS (users use this as complete software)
but also PaaS (users program). For the latter case, how to offer a simpler
programming interface?

How to design the backend for such a service? What architecture and which
components? What policies?

How to evaluate experimentally, through a prototype, that the design of this
system actually meets the goals?
```

References

[1] A. Iosup, F. Kuipers, A. L. Varbanescu, P. Grosso, A. Trivedi, J. Rellermeyer, L. Wang, A. Uta, and F. Regazzoni. Future computer systems and networking research in the netherlands: A manifesto, 2022.

[2] F. Mastenbroek, G. Andreadis, S. Jounaid, W. Lai, J. Burley, J. Bosch, E. van Eyk, L. Versluis, V. van Beek, and A. Iosup. OpenDC 2.0: Convenient modeling and simulation of emerging technologies in cloud datacenters. In *CCGRID*, 2021.