Sunfish: Enabling Predictive Analytics for Datacenters Through Digital Twinning

M. Kwiatkowski Drs. D. Niewenhuis[1] Prof. A. Iosup[2]

[1]Daily Supervisor

[2]Main Supervisor

Vrije Universiteit Amsterdam

June 11, 2026

Motivation

VU VRIJE UNIVERSITEIT AMSTERDAM

Context

Heterogeneous datacenter architectures are common [4] and modern computational needs of AI drive managers to diversify datacenters even more [1]. In result datacenters become extremely complex and hard to operate with millions of CPU's, GPU's etc.

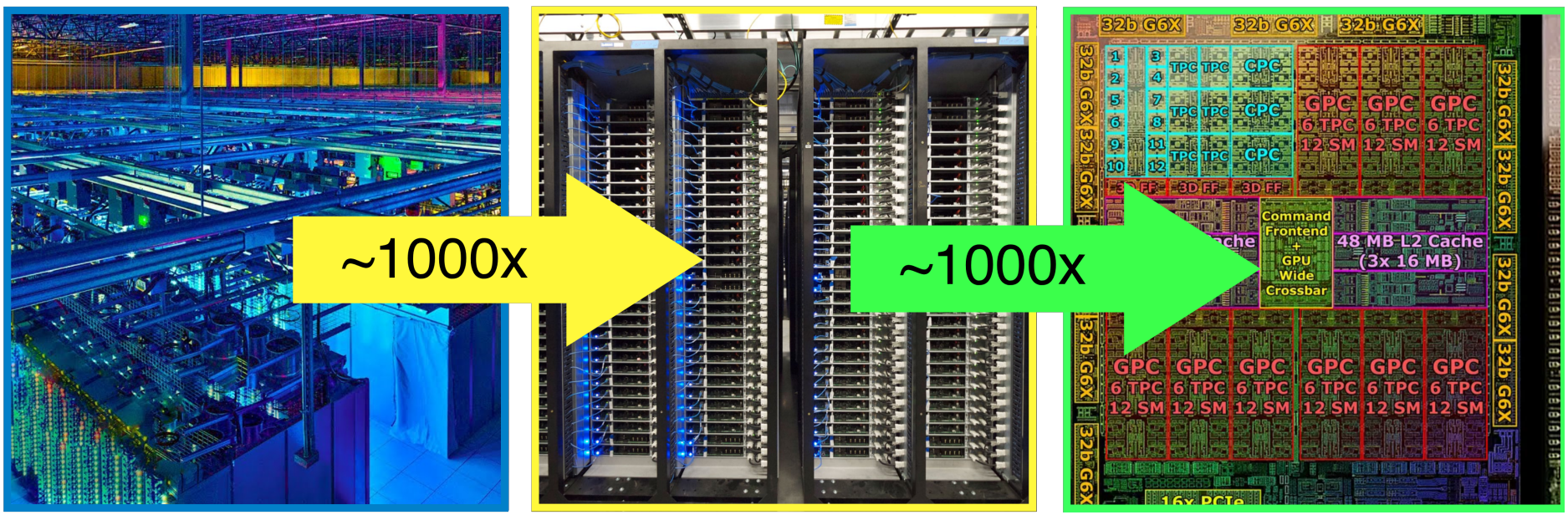

Figure 1.1: Society depends on datacenters to keep running, and therefore we cannot afford to let these systems break down or experience significant performance-related issues. With millions of servers in the largest datacenters, real-time management becomes very difficult. Left to right: a Google datacenter, server racks, Ada Lovelace AD102 GPU architecture.

Problem Statement

VU VRIJE UNIVERSITEIT AMSTERDAM

We need tools to tackle datacenter complexity

We need Datacenter Digital Twins (DCDT) to be better able to detect and solve issues in this critical infrastructure [1]. However, DCDT's are still actively developed, and lack crucial features *e.g.,* predictive analytics [5] to prevent unexpected job failures.

Research Questions

VU VRIJE UNIVERSITEIT AMSTERDAM

Main Research Question

How to enable predictive analytics for datacenters through digital twinning?

Research Question 1

How to asses the current state-of-the-art of digital twinning for datacenters?

Research Question 2

How to design a datacenter digital twin reference architecture using discrete-event simulation and predictive data analytics?

Research Question 3

How to evaluate and validate a datacenter digital twin architecture in relation to system requirements?

RQ1: System Model

VU VRIJE UNIVERSITEIT AMSTERDAM

RQ2: Reference Architecture

RQ3: Experimental Results I

RQ3: Experimental Results II

Discussion

Extra Slides: References I

Jyotika Athavale, Cullen E. Bash, Wesley Brewer, Matthias Maiterth, Dejan S. Milojicic, Harry Petty, and Soumyendu Sarkar.
Digital twins for data centers.
Computer, 57(10):151–158, 2024.
URL https://doi.org/10.1109/MC.2024.3436945.

Douglas Donnellan, Andy Lawrence, and Rose Weinshenk.
Executive summary: Annual outage analysis 2025, May 2025.
URL https://uptimeinstitute.com/resources/research-and-reports/annual-outage-analysis-2025.

Alexandru Iosup, Fernando Kuipers, Ana Lucia Varbanescu, Paola Grosso, Animesh Trivedi, Jan S. Rellermeyer, Lin Wang, Alexandru Uta, and Francesco Regazzoni.
Future computer systems and networking research in the netherlands: A manifesto.
CoRR, abs/2206.03259, 2022.
URL https://doi.org/10.48550/arXiv.2206.03259.

Dejan S. Milojicic, Paolo Faraboschi, Nicolas Dubé, and Duncan Roweth.
Future of HPC: diversifying heterogeneity.
In *Design, Automation & Test in Europe Conference & Exhibition, DATE 2021, Grenoble, France, February 1-5, 2021*, pages 276–281. IEEE, 2021.
URL https://doi.org/10.23919/DATE51398.2021.9474063.

National Academy of Engineering, National Academies of Sciences Engineering, and Medicine.
Foundational research gaps and future directions for digital twins.
The National Academies Press, Washington, DC, 2024.
ISBN 978-0-309-70042-9.
URL https://nap.nationalacademies.org/catalog/26894/foundational-research-gaps-and-future-directions-for-digital-twins.

Extra Slides: References II

Sacheendra Talluri, Leon Overweel, Laurens Versluis, Animesh Trivedi, and Alexandru Iosup.

Empirical characterization of user reports about cloud failures.
In Esam El-Araby, Vana Kalogeraki, Danilo Pianini, Fr édéric Lassabe, Barry Porter, Sona Ghahremani, Ingrid Nunes, Mohamed Bakhouya, and Sven Tomforde, editors, *IEEE International Conference on Autonomic Computing and Self-Organizing Systems, ACSOS 2021, Washington, DC, USA, September 27 - Oct. 1, 2021*, pages 158–163. IEEE, 2021.
URL https://doi.org/10.1109/ACSOS52086.2021.00039.

Fei Tao, Meng Zhang, Yushan Liu, and A.Y.C. Nee.

Digital twin driven prognostics and health management for complex equipment.
CIRP Annals, 67(1):169–172, 2018.
ISSN 0007-8506.
URL https://www.sciencedirect.com/science/article/pii/S0007850618300799.

Extra Slides: Societal Impact

Why is this research important today?

Over 3 million jobs in the Netherlands directly depend on cloud services, which are hosted in datacenters [3]. Already the rapid expansion of datacenters has increased the presence of service failures across all cloud services [6]. We need to act now.

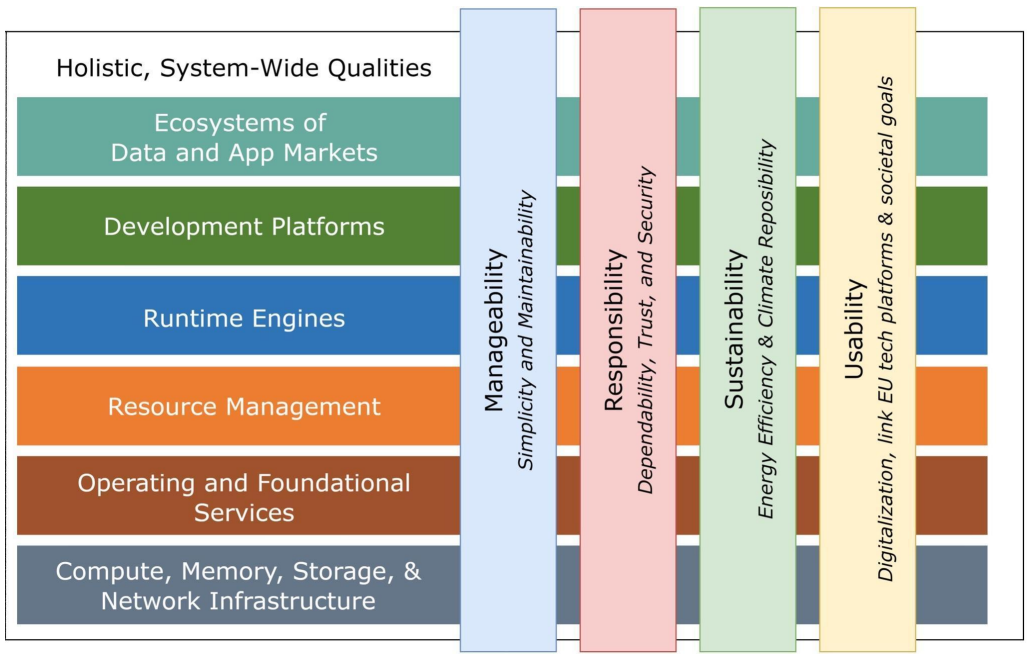

Figure E.1: Horizontally: the most important research areas in computer science in Netherlands. Vertically: qualities we should ensure across all research areas with the most outstanding impact on society. Datacenter manageability is a top-priority [3].

Extra Slides: Why Digital Twinning?

Definition

A DCDT mirrors the structure, context and behaviour of a datacenter [1]. The prerequisite to any digital twin is good monitoring and sensing capabilities in the physical entity. Datacenters meet this requirement easily because they already connect hundreds of monitoring sensors.

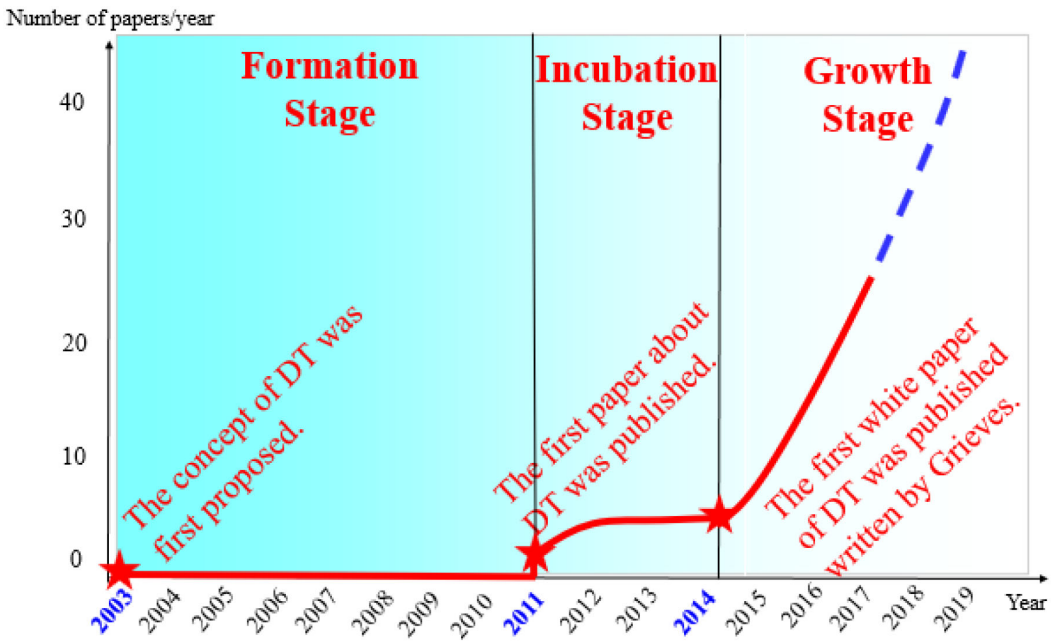

Figure E.2: Due to insufficient technological foundations, little work is available on DTs between 2003 and 2018, and it is only with the rapid growth of cloud computing, Internet-of-Things and Big Data analytics that DTs have reemerged [7]. That is why nobody used digital twins to mirror datacenters earlier.

Extra Slides: Why not pure simulation?

Predicting job failures

Preventing failure-caused outages in advance can reduce huge operational costs, as over 20% of all reported outages amount to more than 1 million US$ [2]. Only a constant bi-directional interaction (digital twin $\Longleftrightarrow$ physical entity) can achieve this.

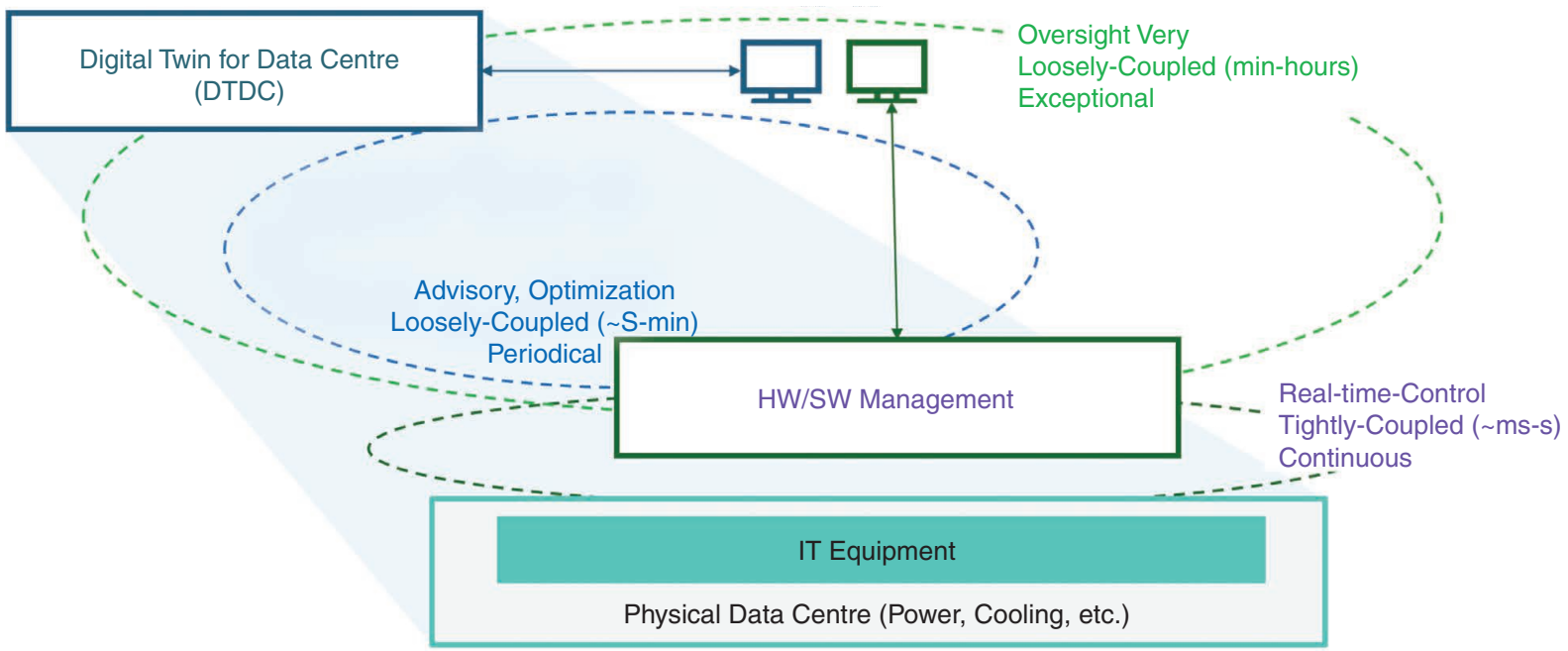

Figure E.3: Real-time control that is tightly-coupled with the IT equipment is a prerequisite for timely predictions within seconds/minutes [1].