Synthesizing of Process-Aware Digital Twin Cockpits from Event Logs

Dorina Bano[1] Judith Michael[2] Bernhard Rumpe[2] Simon Varga[2] Mathias Weske[1]

Abstract: In this work, we summarize our article "Process-Aware Digital Twin Cockpit Synthesis from Event Logs" published in the Journal of Computer Languages (COLA). The engineering of digital twins and their user interfaces with explicated processes, namely process-aware digital twin cockpits (PADTCs), is challenging due to the complexity of the systems and the need for information from different disciplines within the engineering process. Therefore, we have investigated how to facilitate their engineering by using already existing data, namely event logs. We present a low-code development approach that reduces the amount of hand-written code needed to derive PADTCs using process mining techniques. We describe what models could be derived from event log data, which generative steps are needed for the engineering of PADTCs, and how process mining could be incorporated into the resulting application. A PADTC prototype is created based on the MIMIC III dataset, which simulates an automated hospital transportation system. Initially, our approach requires no hand-written code and empowers the domain expert to iteratively create PADTC prototypes.

Keywords: Process-Aware Digital Twin Cockpit; Low-Code Development Approaches; Sensor Data; Event Log; Process Mining; Process-Awareness

1 Summary

Improving the engineering process of digital twins (DTs) with a higher degree of automation enables domain experts to take an active role in this process and allows for an iterative evolvement of the DT. A DT of an actual object can include process-aware DT cockpits (PADTCs) which allow for user interaction and provide means to handle processes of the actual object and its context. Full automation of the DT engineering process and its services might be a big vision. However, the automated engineering of PADTCs is the first reachable goal. Within [Ba22], we have suggested a low-code development approach for generating PADTCs from event logs. Our approach includes *four phases*.

In the *preparation phase*, we analyze the real world and extract models from data about the actual object. We extract an event log from the sensor data of an actual object. Using this event log, we perform data-to-model transformations and discover domain information in a class diagram, process models in BPMN, and roles in a tagging model.

In the *generation phase*, we apply model-to-model and model-to-code transformations to generate a PADTC. We are using the models from the preparation phases as input for

[1] Hasso Plattner Institute, University of Potsdam, Potsdam, Germany {dorina.bano,mathias.weske}@hpi.de
[2] Software Engineering, RWTH Aachen University, Aachen, Germany {michael,rumpe,varga}@se-rwth.de

model-to-model transformations and generate data models (as views) and GUI models. These generated models and the domain model are the generator input to synthesize the PADTC source code.

In the *adaption phase*, we can extend the generated parts by handwritten additions on models and code levels. We can add handwritten models which are used as additional input to synthesize the code of the PADTC with a code generator. Additionally, we can add handwritten code in addition to the generated code.

For *runtime of the PADTC*, we connected the PADTC to DT services to get a fully functional digital twin. For a process-aware digital twin, especially the connection to services for process discovery, process conformance checking and process prediction services are relevant. Additionally, the PADTC visualizes live data from the actual object or third-party applications and enables domain users to interact with the DT. The DT can influence the actual object via user commands from the PADTC or automatic commands from self-adaptive services.

Using this low-code development approach, we have created a PADTC prototype that simulates a hospital transportation system. For its creation, we have used event logs extracted from the MIMIC III dataset [Jo16], a dataset with health-related data of patients who stayed in critical care units. However, our approach is not limited to this domain and could be applied to engineer PADTCs for areas where process data is available, e.g., tools and machines in production, products, experiments, software systems, or organizations.

The application has shown, that we can achieve a high degree of automation in the PADTC engineering process and reduce the amount of hand-written code needed: Developers need no hand-written code in the first place to get a deployable prototype. This can be extended by hand-written models and code to create a digital twin.

Data Availability. The original publication is accessible under `http://doi.org/10.1016/j.cola.2022.101121`, a preprint is available at `https://www.se-rwth.de/publications`. Unfortunately, we could not provide more detailed artifacts for reasons of confidentiality.

Acknowledgments. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2023 Internet of Production - 390621612. Website: https://www.iop.rwth-aachen.de

Bibliography

[Ba22] Bano, Dorina; Michael, Judith; Rumpe, Bernhard; Varga, Simon; Weske, Matthias: Process-Aware Digital Twin Cockpit Synthesis from Event Logs. Journal of Computer Languages (COLA), 70, June 2022.

[Jo16] Johnson, Alistair EW; Pollard, Tom J; Shen, Lu; Lehman, Li-wei H; Feng, Mengling; Ghassemi, Mohammad; Moody, Benjamin; Szolovits, Peter; Celi, Leo Anthony; Mark, Roger G: MIMIC-III, a freely accessible critical care database. Scientific data, 3, 2016.