

First International Workshop on Software Architecture for Data-Intensive Systems (SADIS@ECSA 2025)

Ilias Gerostathopoulos¹[0000–0001–9333–7101], Marcela Tuler De
Oliveira²[0000–0001–5739–5590], and Tomas Bures³[0000–0003–3622–9918]

¹ Vrije Universiteit Amsterdam, Amsterdam, The Netherlands,
`i.g.gerostathopoulos@vu.nl`

² Delft University of Technology, Delft, The Netherlands,
`M.TulerdeOliveira@tudelft.nl`

³ Charles University in Prague, Prague, Czech Republic, `bures@d3s.mff.cuni.cz`

Abstract. Modern software systems increasingly rely on data-centric processes to fuel intelligent features, scalable services, and real-time insights. As these systems evolve, designing effective software architectures for data-intensive environments becomes critical yet challenging. These architectures must address complex concerns such as handling high-velocity data streams, ensuring security and privacy, integrating heterogeneous data sources and heterogeneous computation fabrics (including GPUs/TPUs and edge devices), and adapting to changing workloads. The workshop aims to bring together academic researchers, industry practitioners, and thought leaders to explore cutting-edge approaches for building, deploying, and maintaining robust data-centric solutions. We invite discussions on a range of topics at the intersection of data-intensive systems and software architecture, including machine learning (ML) pipelines, ML Operations (MLOps) frameworks and patterns, domain-specific languages, reference architectures, and visualization techniques. Emphasis is placed on practical experiences, innovative tools and platforms, and theoretical advancements that tackle the ever-growing demands placed on data-intensive systems. By fostering interdisciplinary collaboration and knowledge exchange, the workshop seeks to shape the future of data-driven architecture design and empower professionals and researchers to address emerging challenges in this rapidly evolving domain.

Keywords: Data-centric, data-intensive systems, software architectures

1 Motivation and Objectives

The notion “data is the new oil”, declared by the mathematician Clive Humby in 2006, emphasizes that data, like oil, is in principle valuable but can only be really used if properly processed. In 2025, data-driven methods (including ML) and the underlying data-intensive systems are the main force of innovation, since

they make data valuable by extracting insights from it. Entire business sectors are either already transformed or under disruption by such innovation, including predictive maintenance, cyber-security, mobility, public health, emergency coordination, to mention just a few.

Data-intensive systems include ML-based systems, but also systems built with workflow automation engines, simulation-heavy and visualization-heavy systems, and data analytics applications. This broad definition covers several types, phases and maturity levels of data-driven methods, which may start from simple automation of data pipelines to complex MLOps processes where ML models are coupled with online monitoring and re-trained on the fly in the presence of new data.

This is often not restricted to simple centralized solutions, but addressing the needs of Cyber-Physical Systems or IoT, it often calls for decentralized and heterogeneous approaches that involve end-user devices.

We need better methods to design and architect data-intensive systems. Currently, similar to the design and development of ML-systems, such endeavors are mostly based on the skills and intuition of experienced engineers. Given the extreme need for data-intensive systems now and in the near future, there is a dire need for scalable solutions that can be provided by the software engineering and software architecture communities.

In response, SADIS aims to provide a platform for researchers, practitioners, and industry professionals to exchange ideas, share experiences, and discuss the latest developments in software architecture approaches for data-intensive systems. The goal is to address the challenges of designing, developing, and maintaining scalable, secure, and efficient systems in the era of big data and artificial intelligence. The workshop will focus on innovative methods, frameworks, and tools to support data-intensive workflows, improve integration, and enable adaptability to real-world demands.

In terms of topics of interest, the following is a non-exhaustive list.

- ML training pipelines
- Extreme flexibility
- User in the loop aspects
- Online optimization
- Architectural approaches for MLOps
- Decentralized architectures for data-intensive systems
- Data-intensive CPS/IoT and edge-cloud systems
- Patterns
- Model-driven approaches for data-intensive systems and data workflows
- Domain-specific languages for data-intensive systems and data workflows
- Reference architectures for data-intensive systems and data workflows
- Security and privacy concerns for data pipelines
- Distributed ledgers and NFTs
- Integration with data sources
- Digital Twins
- Micro-frontends and architectures for data visualization and presentation

- Approaches for integration of data pipelines
- Adaptive visualization approaches
- Explainable AI
- Building knowledge across experiments
- Continuous data-driven optimization via experiments
- Architecture approaches, challenges, implications of AutoML

2 Format

We expect to organise a full-day workshop. We plan to invite a keynote speaker to open the workshop. We would like to have two to three sessions of full and short papers. We plan to end the workshop with a panel to summarize and discuss the emerged ideas, and to initiate collaborations among the participants to the workshop, including the non-speakers attendees.

We do not need any special services, and we do not have any logistic and/or equipment constraints.

3 Publicity and Participation

We plan to publish the Call for Papers through the regular mailing lists, project networks, social networks such as Twitter and LinkedIn, and through reaching out via direct emails to researchers working on software architectures and intensive systems. We also plan to attract participants from industry⁴. We will publish our workshop on the ECSA 2025 website and on the workshop website. We expect around 20-30 participants to our open workshop.

4 Submissions

The workshop invites three types of submissions:

- Research Contributions (long papers, up to 12 pages LNCS): Full research papers presenting novel contributions and theoretical advancements in software architectures for data-intensive systems. These submissions should provide an in-depth analysis, rigorous evaluation, and significant contributions to the field.
- Case Studies/Experience Reports from Industry (short papers, up to 8 pages LNCS): Contributions that share practical insights, challenges, and lessons learned from real-world implementations of software architectures. While these papers may not focus on novelty, they offer valuable perspectives on addressing industrial challenges.

⁴ Including our partners from the currently running ExtremeXP EU project <https://extremexp.eu>

- Artifacts/Benchmarks/Tools/Platforms (short papers, up to 8 pages LNCS): Papers showcasing innovative tools, platforms, or benchmarks that support the design, development, or evaluation of data-intensive systems. These contributions should provide practical solutions or demonstrate the potential for real-world application.

Submissions must be written in English and follow the LNCS format guidelines. All submissions will be reviewed by at least three members of the Program Committee (or their sub-reviewers) and will be evaluated based on relevance, quality, and originality.

At least one author of each accepted paper is required to register, attend, and present the work at the workshop. The workshop especially encourages PhD students and early-career researchers to contribute visionary ideas and preliminary work, fostering innovation and creativity in the field.

Accepted papers will be included in the workshop proceedings, which we expect to be published along with other workshop papers at ECSA in the ECSA companion proceedings.

5 Organizers

Ilias Gerostathopoulos is an Assistant Professor of computer science with Vrije Universiteit Amsterdam, The Netherlands. His research interests include software engineering, software architecture, and self-adaptive systems. He received the Ph.D. degree in computer science in 2015 from the Department of Distributed and Dependable Systems, Faculty of Mathematics and Physics, Charles University, Prague. He was also as a Postdoctoral Researcher with the Department of Informatics, Technical University of Munich. He serves in program committees of several international conferences in the fields of software architecture and self-adaptive systems. Ilias has also co-chaired several tracks of such conferences (including the technical tracks of ECSA 2022 and ACSOS 2022) and co-organized five international workshops 9 (at ECSA, ICSE, SASO, and ACSOS).

Marcela Tuler de Oliveira is an Assistant Professor in Trustworthy Data Systems at TU Delft's Department of Engineering Systems & Services as part of the ICT section. She is a Telecommunications Engineer and a thought leader in digital solutions for cross-organizational data sharing. She has participated in the program committees of international conferences IEEE CloudNet 2024 and IEEE BCCA 2025.

Tomáš Bureš is an Full Professor and a vice-chair of the Department of Distributed and Dependable Systems at Charles University, in Prague, Czech Republic. His research focus is in component-based development, cyber-physical systems, and self-adaptive systems. He served as a general chair of ECSA 2022, program chair of ECSA 2019, SEAA 2018, QoSA 2014, and as a track chair at numerous occasions. Tomas was also a co-organizer of SEsCPS 2015 – 2019 (workshop on Software Engineering for Smart Cyber-Physical Systems) at ICSE.

6 Other Initiatives (instead of Previous Editions)

In our community, we see emerging initiatives that examine the interplay between ML and software architecture (SA) (e.g. SAML⁵), focus on the timely topic of generative AI for SA (e.g. SAGAI⁶), and take a broader scope in examining the synergies between software engineering and AI systems (e.g. CAIN⁷). We aim to complement these initiatives by focusing on SA for data-intensive systems.

⁵ <https://sa-ml.github.io/saml2025/>

⁶ <https://www.iese.fraunhofer.de/en/events/sagai.html>

⁷ <https://conf.researchr.org/home/cain-2025>

1st International Workshop on Software Architecture for Data-Intensive Systems (SADIS@ECSA 2025)

In conjunction with ECSA'25, September 2025, Limassol, Cyprus

IMPORTANT DATES

Abstracts: May 12, 2025

Papers: May 19, 2025

Acceptance notification: June 20, 2025

Camera-ready copies: June 27, 2025

Workshop: September 2025

ORGANISERS

Ilias Gerostathopoulos

Vrije Universiteit Amsterdam, Netherlands

Marcela Tuler De Oliveira

Delft University of Technology, Netherlands

Tomas Bures

Charles University, Czech Republic

PROGRAM COMMITTEE (TENTATIVE)

Paolo Arcaini

National Institute of Informatics, Japan

Steffen Becker

University of Stuttgart, Germany

Justus Bogner

Vrije Universiteit Amsterdam, Netherlands

Martina De Sanctis

Gran Sasso Science Institute, Italy

Giorgos Giannopoulos

Athena Research Center, Greece

Rihan Hai

Delft University of Technology, Netherlands

Petr Hnětynka

Charles University, Czech Republic

Amela Karahasanovic

SINTEF, Norway

Henry Muccini

Università degli Studi dell'Aquila, Italy

Erik Gøsta Nilsson

SINTEF, Norway

George Papastefanatos

Athena Research Center, Greece

Patrizio Pellicione

Gran Sasso Science Institute, Italy

Marie Platenius-Mohr

ABB AG Corporate Research Center, Germany

Romina Spalazzese

Malmö University, Sweden

Vlado Stankovski

University of Ljubljana, Slovenia

Catia Trubiani

Gran Sasso Science Institute, Italy

Karthik Vaidhyanathan

IIIT Hyderabad, India

Kristina Veljković

University of Ljubljana, Slovenia

Yiannis Verginadis

Athens University of Economics and Business, Greece

Eleni Zarogianni

Intracom Telecom, Greece

MOTIVATION AND CONTEXT

Modern software systems increasingly rely on data-centric processes to fuel intelligent features, scalable services, and real-time insights. As these systems evolve, designing effective software architectures for data-intensive environments becomes critical yet challenging. These architectures must address complex concerns such as handling high-velocity data streams, ensuring security and privacy, integrating heterogeneous data sources and heterogeneous computation fabrics (including GPUs/TPUs and edge devices), and adapting to changing workloads.

The workshop aims to bring together academic researchers, industry practitioners, and thought leaders to explore cutting-edge approaches for building, deploying, and maintaining robust data-centric solutions. We invite discussions on a range of topics at the intersection of data-intensive systems and software architecture, including machine learning (ML) pipelines, ML Operations (MLOps) frameworks and patterns, domain-specific languages, reference architectures, and visualization techniques. Emphasis is placed on practical experiences, innovative tools and platforms, and theoretical advancements that tackle the ever-growing demands placed on data-intensive systems. By fostering interdisciplinary collaboration and knowledge exchange, the workshop seeks to shape the future of data-driven architecture design and empower professionals and researchers to address emerging challenges in this rapidly evolving domain.

WORKSHOP STRUCTURE AND PLANNED OUTCOMES

The full-day workshop will open with a keynote talk, followed by presentations of the papers and Q&A. We plan to end the workshop with a panel to summarize and discuss the emerged ideas, and to initiate collaborations among the participants to the workshop, including the non-speakers attendees.

TOPICS (in relation to SW architecture of data-intensive systems)

- ML training pipelines
- Extreme flexibility
- User in the loop aspects
- Online optimization
- Architectural approaches for MLOps
- Decentralized architectures for data-intensive systems
- Data-intensive CPS/IoT and edge-cloud systems
- Patterns
- Model-driven approaches for data-intensive systems and data workflows
- Domain-specific languages for data-intensive systems and data workflows
- Reference architectures for data-intensive systems and data workflows
- Security and privacy concerns for data pipelines
- Distributed ledgers and NFTs
- Integration with data sources
- Digital Twins
- Micro-frontends and architectures for data visualization and presentation
- Approaches for integration of data pipelines
- Adaptive visualization approaches
- Explainable AI
- Building knowledge across experiments
- Continuous data-driven optimization via experiments
- Architecture approaches, challenges, implications of AutoML

SUBMISSIONS

The workshop invites three types of submissions:

- **Research Contributions (long papers, up to 12 pages LNCS):** Full research papers presenting novel contributions and theoretical advancements in software architectures for data-intensive systems. These submissions should provide an in-depth analysis, rigorous evaluation, and significant contributions to the field.
- **Case Studies/Experience Reports from Industry (short papers, up to 8 pages LNCS):** Contributions that share practical insights, challenges, and lessons learned from real-world implementations of software architectures. While these papers may not focus on novelty, they offer valuable perspectives on addressing industrial challenges.
- **Artifacts/Benchmarks/Tools/Platforms (short papers, up to 8 pages LNCS):** Papers showcasing innovative tools, platforms, or benchmarks that support the design, development, or evaluation of data-intensive systems. These contributions should provide practical solutions or demonstrate the potential for real-world application.

Submissions must be written in English and follow the LNCS format guidelines. All submissions will be reviewed by at least three members of the Program Committee (or their sub-reviewers) and will be evaluated based on relevance, quality, and originality.

At least one author of each accepted paper is required to register, attend, and present the work at the workshop. The workshop especially encourages PhD students and early-career researchers to contribute visionary ideas and preliminary work, fostering innovation and creativity in the field.

Accepted papers will be included in the workshop proceedings, which we expect to be published along with other workshop papers at ECSA in the ECSA companion proceedings.