

Workshop Proposal

Title

9th International Workshop on Serverless Computing (WoSC9) part of ACM/IFIP Middleware 2023 in DAMSLab, Department of Arts, University of Bologna, Italy, Dec 11-15, 2023

The names, affiliations and e-mail addresses of the organizing committee

Paul Castro, IBM Research

Paul Castro, Ph.D. is a Research Staff Member at the IBM Watson Research Center. He has been active in research on mobile and pervasive computing, cloud infrastructure, wireless location systems, location databases, stream processing, and enterprise web applications and has been awarded several patents in these areas. He has worked on cloud services for supporting mobile applications running on various smartphone platforms. Work from his research in the area of multi-device application support was recently released as part of the IBM Bluemix Mobile Backend as a Service. He has earned two IBM Technical Achievement Awards for the IBM SmartCloud Web Meetings for mobile clients and the Intelligent Notification System. Most recently, he worked on IBM OpenWhisk for Bluemix, with a focus on mobile solutions.

Pedro García López, University Rovira i Virgili

Pedro Garcia Lopez is full professor in Computer Science and Artificial Intelligence at Universitat Rovira i Virgili (URV). He leads CLOUDLAB: Cloud and Distributed Systems research group and he currently coordinates three large European research projects related to Serverless technologies (cloudstars.eu, neardata.eu, cloudskin.eu). His research topics are Serverless Computing, Big Data analytics, Cloud Computing, distributed systems, Cloud Storage, software architectures and middleware.

Vatche Isahagian, IBM Research

Vatche Ishakian is a Research Staff Member in the AI Engineering group at the IBM Thomas J. Watson Research Center. He earned his PhD in Computer Science from Boston University under the supervision of Professor Azer Bestavros. His research interests include developing platforms and tools in support of AI cloud applications.

Vinod Muthusamy, IBM Research

Vinod Muthusamy is a Research Staff Member in the AI Lifecycle Acceleration group at the IBM Thomas J. Watson Research Center. His current research interests include cloud platforms that support a variety of workloads, programming models and technologies to compose services, and tools to support the development and maintenance of AI applications.

Aleksander Slominski, IBM Research

Aleksander Slominski is Research Staff Member in the Serverless Group at the IBM T.J. Watson Research Center. He is interested in development of applications for next generation Internet, Web Services, Orchestration, Components, AI, Workflows, and Clouds.

Brief Technical Description

Over the last nine years, Serverless Computing (Serverless) has gained an enthusiastic following in industry as a compelling paradigm for the deployment of cloud applications, and is enabled by the recent shift of enterprise application architectures to containers and microservices. Many of the major cloud vendors have released serverless

platforms, including Amazon Lambda, Google Cloud Functions, Microsoft Azure Functions, IBM Cloud Functions. Open source projects are gaining popularity in providing serverless computing as a service.

Recently, Kubernetes gained in popularity in enterprise and in academia. Several open source projects such as OpenFaaS and Knative aim to provide developers with serverless experience on top of Kubernetes by hiding low-level details. Auto-scalable Multi-tenant Kubernetes deployments like Google Cloud Run or IBM Code Engine also overcome previous limitations of Serverless Functions like duration, networking, and higher granularity (more vCPUs).

Serverless on the cloud is a somewhat mature research area with many conferences accepting papers on this topic. In the spirit of having this workshop serve as a venue for future and exploratory research directions, we will be evolving the workshop to include hybrid cloud environments, as well as edge and IoT devices. These next-gen computing architectures are becoming more common but have little support from serverless platforms and bring new challenges to old concerns such as resource optimization, scaling, cost, monitoring, and ease of use. The serverless experience becomes an important topic for emerging topics such as DevOps and Platform Engineering in industry and will be critical to the success of next-gen computing.

Building on the recent advances in generative AI, including Large Language Models (LLMs) and other types of Foundations Models (FMs), the workshop also plans to explore the use of hybrid serverless platforms to fine-tune, serve, and manage the lifecycle of LLMs with a focus on aspects such as use cases, resource allocations, optimizations, and using AI to improve serverless experience.

This workshop brings together researchers and practitioners to discuss their experiences and thoughts on future directions of serverless research.

Interest to the Research Community

Serverless architectures offer different tradeoffs in terms of control, cost, and flexibility compared to distributed applications built on an Infrastructure as a Service (IaaS) substrate. For example, a serverless architecture requires developers to more carefully consider the resources used by their code (time to execute, memory used, etc.) when modularizing their applications. This is in contrast to concerns around latency, scalability, and elasticity, which is where significant development effort has traditionally been spent when building cloud services. In addition, tools and techniques to monitor and debug applications aren't applicable in serverless architectures, and new approaches are needed. As well, test and development pipelines may need to be adapted. Another decision that developers face is the appropriateness of the serverless ecosystem to their application requirements. A rich ecosystem of services built into the platform is typically easier to compose and would offer better performance. However, composing external services may be unavoidable, and in such cases, many of the benefits of serverless disappear, including performance and availability guarantees. This presents an important research challenge, and it is not clear how existing results and best practices, such as workflow composition research, can be applied to composition in a serverless environment. Finally, an intriguing area of research lies in the emergence of innovative FaaS runtimes tailored to various application domains. The investigation of serverless FaaS is not limited to low-latency HPC scenarios; it is also being explored in edge computing environments. Moreover, FaaS is serving as a source of inspiration for groundbreaking interactive burst-parallel extreme data analytics environments in the Cloud. Additionally, serverless AI services are expected to present formidable research objectives in the coming years.

The history of the workshop

This would be the fifth time of the workshop at Middleware and we want to build on our previous experience and continue to draw the attention of distributed systems researchers to this young research field and further drive the adoption and development of available technology.

There have been seven previous editions of this workshop:

- WoSC8: 22nd ACM/IFIP International Conference Middleware, Nov 7, 2021 in hybrid location; virtual and in Québec, Canada
- WoSC7: 21th ACM/IFIP International Conference Middleware, Dec 7, 2021 in virtual location

- WoSC6: 20th ACM/IFIP International Conference Middleware, Dec 8, 2020 in virtual location
- WoSC5: 20th ACM/IFIP International Conference Middleware, Dec 9, 2019 in UC Davis, CA, USA
- WoSC4: 11th IEEE/ACM UCC / 5th IEEE/ACM BDCAT 2018, Zurich, Switzerland
- WoSC3: IEEE CLOUD 2018, San Francisco, CA, USA
- WoSC2: Middleware 2017, Las Vegas, NV, USA
- WoSC1: ICDCS 2017, Atlanta, GA, USA

These are linked on the series' permanent web presence at: <https://www.serverlesscomputing.org/workshops/>

The organizers of this proposed workshop are the founding members of this workshop and have co-chaired previous iterations.

Similar Workshops

Since WoSC has been the first academic workshop dedicated to Serverless Computing that has more than one edition. As far as we know in 2023 there is only one other serverless workshop: SESAME The 1st Workshop on SErverless Systems, Applications and MEthodologies co-located with EuroSys 2023. In previous years there were few serverless workshops co-located with the conferences such as HiPS 2022 Workshops on High Performance Serverless Computing (in conjunction with ACM HPDC) in 2021 and 2022, then in 2019 The First IEEE SERVICES Workshop on Serverless Computing SWoSC and a regional workshop - European Symposium on Serverless Computing and Applications ESSCA 2018 that was co-organized with WoSC4 in 2018. Checking with WikiCFP those are the only similar workshops.

A short description of the intended format of the workshop

Hybrid workshop (in-person and remote participants) with invited keynote speakers, paper presentations (pre-recorded videos as backup, live Q&A), possible poster & demo session, and panel Q&A.

Review process

Submitted papers must have at most 6 pages of technical content, including text, figures, and appendices, but excluding any number of additional pages for bibliographic references. Note that submissions must be double-blind: authors' names must not appear, and authors must make a good faith attempt to anonymize their submissions. Submitted papers must adhere to the formatting instructions of the ACM SIGPLAN style, which can be found on the ACM template page. The font size has to be set to 10pt.

We will use the HotCRP system in a way consistent with how Middleware workshops are using it.

Tentative workshop paper submission and notification deadlines

- Submission: September 25, 2023
- Notification: September 26, 2022
- Camera Ready: October 20, 2023
- Author registration deadline: TBD
- Conference: October 27, 2023

Primary contacts of the organizing committee

Aleksander Slominski, IBM Research

Estimated number of participants, length, and timetable of the workshop

We expect the workshop length and agenda to be similar to the previous iterations of the workshop, which included

invited talks, paper presentations, and panel discussions. We will provide a final agenda once we know the number of invited talks and accepted papers.

Approximate attendance: 30-50 participants

A preliminary list of PC members

Workshop co-chairs

- Paul Castro, IBM Research
- Pedro García López, Universitat Rovira i Virgili
- Vatche Isahagian, IBM Research
- Aleksander Slominski, IBM Research
- Vinod Muthusamy, IBM Research

Technical Program Committee (tentative)

- Gul Agha, University of Illinois at Urbana-Champaign
- Azer Bestavros, Boston University
- Flavio Esposito, Saint Louis University
- Rodrigo Fonseca, Brown University
- Ian Foster, University of Chicago and Argonne National Laboratory
- Geoffrey Fox, Indiana University
- Dennis Gannon, Indiana University & Formerly Microsoft Research
- Arno Jacobsen, MSR (Middleware Systems Research Group)
- Wes Lloyd, University of Washington Tacoma
- Vi-nja Kri-anovi-, Josip Juraj Strossmayer University of Osijek
- Tyler Harter, GSL, Microsoft
- Maciej Malawski, AGH University of Science and Technology
- Pietro Michiardi, Eurecom
- Lucas Nussbaum, LORIA, France
- Maciej Pawlik, Academic Computer Centre CYFRONET of the University of Science and Technology in Cracow
- Per Persson, Ericsson Research
- Peter Pietzuch, Imperial College
- Rodric Rabbah, Apache OpenWhisk
- Josef Spillner, Zurich University of Applied Sciences
- Rich Wolski, University of California, Santa Barbara

Steering Committee (tentative)

- Geoffrey Fox, Indiana University
- Dennis Gannon, Indiana University & Formerly Microsoft Research
- Arno Jacobsen, MSR (Middleware Systems Research Group)