M. Marazakis: “Challenges towards effective support of Computing Continuum use cases”
Expression-of-Interest on future visions and research directions 2025-27 in the area of Cloud-to-Edge-to-IoT for European Data – submitted: April 30, 2023

Challenges towards effective support of Computing Continuum use cases

Author: Dr. Manolis Marazakis

Principal Staff Research Scientist, Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH), Greece. Email: maraz@ics.forth.gr

Motivation

Edge applications, aimed at effectively processing data in a timely manner, require a fluid integration of resources at the edge, the core, and along the data path to support dynamic and data-driven application workflows [1]. This vision of combining real-time data with complex models and data analytics to monitor and manage systems of interest motivates a *Computing Continuum (CC)* view of computing.

The Computing Continuum raises challenges at multiple levels [2]:

- Application level, having to bridge simulations, machine learning and data-driven analytics;
- Middleware level, necessitating tools to enable efficient deployment, scheduling and orchestration of the workflow components across a widely distributed, dynamically changing, and heterogeneous infrastructure;
- System level, requiring a resource management approach that can allocate a suitable set of components of the infrastructure to run the application workflow in a dynamic and adaptive manner.

The degree of dynamic adaptation required will vary over a broad range, from adjustments due to availability and component-level load changes to resource elasticity enabling users to build applications and integrate data for thematic specialisation and decision support, within ever shortening response time windows [3].

This position paper identifies challenges, for a near-/mid-term time horizon (2025-27), towards realizing an effective platform for CC use cases, and offers suggestions for initiating relevant research and innovation (R&I) lines.

Current Status

Today’s software approaches available to address the needs of CC use cases consist of several separate software stacks optimised for different goals [2], specific to the target infrastructure (eg. physical simulation, ML/AI-driven modelling and inference, data processing and analysis). There is no simple solution making it possible to deploy and orchestrate a combination of consistent interoperable components across the entire CC scope.

Therefore, there is an increasing need for integrated software ecosystems which combine current "island" solutions and bridge the gaps between them. These ecosystems must support the entire lifecycle of CC use cases, including initial modelling, programming, deployment, execution, optimisation, as well as monitoring and control [2]. Establishing and managing trust over time when sharing systems, software and data is a further major concern, together with support for reproducability of workflow results.

Research Challenges

To address the challenges identified, the following two broad areas of R&I actions need to be considered:

- Software interoperability and composability enhancements, to enable smoother integration in the context of CC use cases. In particular, we need to facilitate integration of HPC, AI/ML and data analytics processing, including hybrid applications such as AI-enabled simulations. Given the diversity of the existing software stacks and execution platforms, we expect that such applications cannot run efficiently on a single homogenous platform, but rather will require usage of distributed and dynamically allocated resources.

- Federated usage of compute, storage and communication resources, of different calibers, that operate autonomously under the purvue of independent authorities. Challenges, with both technical and organizational aspects, include support for interoperable resource allocation and accounting, as well as identity and access management (IAM). To address the heterogeneity of available platforms at large scale, scheduling/orchestration and monitoring middleware is a major challenge. Moreover, security assurances regarding platforms, interconnects, and data are essential enablers of CC use cases.

The authors’ view is that lower-TRL project concepts should be encouraged, to be followed by pilot deployments at sufficient scale to allow for realistic evaluation. To define a clear research framework, it is recommended to engage several technical communities that already have their own technical agendas (eg. ETP4HPC, BDVA/AIRO) and possibly invest in short-term feasibility and planning studies, starting from from systematic reviews of what is currently available and comparing with what is to be achieved in the given time frame. There is great value in expanding coverage of cross-over technical areas (eg. HPC-Cloud-Edge-IoT convergence), both for demonstration/validation purposes of concepts but also as proving ground for results from basic research areas (eg. algorithms and distributed system protocols).

Conclusion

The emergence of complex workflows combining simulations, data analytics and learning, running on hybrid infrastructures necessitates coverage of cross-over technical areas, such convergence of HPC-Cloud-Edge technologies, with an emphasis on open interfaces and resource management concerns (esp. efficiency metrics).

References

[1] Balouek-Thomert, Daniel , Gibert Renart, Eduard , Reza Zamani, Ali , Simonet, Anthony & Parashar, Manish: "Towards a computing continuum: Enabling edge-to-cloud integration for data-driven workflows", The International Journal of High Performance Computing Applications 2019 33:6, 1159-1174. https://doi.org/10.1177/1094342019877383

[2] Antoniu Gabriel, Valduriez, Patrick, Hoppe, Hans-Christian, & Krüger, Jens. (2021). Towards Integrated Hardware/Software Ecosystems for the Edge-Cloud-HPC Continuum. Zenodo. https://doi.org/10.5281/zenodo.5534464

[3] Marazakis, Manolis, Duranton, Marc, Pleiter, Dirk, Taffoni, Giuliano, & Hoppe, Hans-Christian. (2022). HPC for Urgent Decision-Making. Zenodo. https://doi.org/10.5281/zenodo.6107362