Future visions and research directions 2025-27 in the area of Cloud-to-Edge-to-IoT for European Data

Motivation

Data generation has undergone an explosion during the last decade: the world is becoming increasingly connected. In fact, 29 billion of connected devices on the global network was estimated by 2022, according to the Telecommunications Industry Association, being more than half (18 billion) IoT devices[1]. Europe is also supporting this movement through several policies and regulations (i.e., Digital Decade policy programme[2]). In this context, all the new digitization movements, paradigms, and trends revolve around two pillars: data and computational models/infrastructure. In recent years, organizations have become aware of how data is a strategic asset. Processed in the appropriate way, they can generate a competitive advantage for companies. The deployment of sensors or IoT devices in industrial, biomedical or Smart City sectors along with the massive use of mobile phones has fostered not only this unprecedented data capture, but also the growth of more and more sophisticated algorithmic schemes capable of extracting knowledge and giving value. As a consequence, computing infrastructure to capture and store all this data and execute complex algorithms needs to be put in place through innovative computing models in which to implement the different technological architectures of the organizations. The secure and smart exploitation of the data is the final goal of developing new and sophisticated computing paradigms, such Edge Computing, as well as tools, such as operationalization frameworks or methodologies, e.g., MLOps, that assist and complement them. In this document we focus on different lines of research in this field, describing them and identifying the challenges of interest.

Current status and research challenges

Next, we propose four large domains in the area of Cloud-to-Edge-to-IoT that we have identified as the main topics that should be studied in depth, and whose development will power the data economy. For each category we provide a rationale of its importance and the different challenges for which research and relevant solutions need to be provided.

AIOps

AIOps — artificial intelligence for IT operations — is a term coined by Gartner[3] that refers to the application of artificial intelligence (AI) paradigm to automate and streamline operational workflows. There are already mature research lines in this field addressing issues such as performance monitoring, detection of availability problems or DevOps root cause analysis (RCA[4]). In our opinion, it is necessary to extend the paradigm to include new technologies such as the Cloud Continuum, providing intelligence to the tools and solutions that are generated to extend the computing layers in a more automated and efficient way. Following this idea, we describe three new challenges that we believe can be solved with advanced research in this domain:

- The next version of the ***Cognitive Cloud***, that can be considered a self-adapting Cloud/Edge-based ecosystem that is capable of sensing its environment, learning from it, and opportunistically and dynamically adapt it using advanced AI techniques.
- ***Intelligent assistant for the operation in Cloud Continuum***, referring to a technology that combines all the technical knowledge about OT (operational technology), with powerful artificial

[1] Telecommunications Industry Association: https://enterprisersproject.com/article/2020/4/edge-computing-9-compelling-stats (2021), online; accessed 01 December 2021
[2] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en
[3] https://www.gartner.com/en/information-technology/glossary/aiops-artificial-intelligence-operations
[4] Sabharwal, N., & Bhardwaj, G. (2022). AIOps Supporting SRE and DevOps. In Hands-on AIOps: Best Practices Guide to Implementing AIOps (pp. 73-88). Berkeley, CA: Apress.

intelligence algorithms used to assist a specialized human technical team to automate/speed up the productivity of the different processes to be implemented in the Cloud Continuum.

- ***NLP as a technology to accelerate operations in the cloud continuum***. The process of operationalizing applications and solutions on the Cloud Continuum requires command input, leading to increased mean time to resolution (MTTR) and slower root cause analysis (RCA). Often these are repetitive operations over time where only certain parameters change. Therefore, being able to automate the process and understand the instructions in human language, will result in a greater stability of the environment and even achieve a reduction in the demand for requests to the human team.

Data at the edge

The conjunction of edge computing with the large amount of data that these edge devices under the IoT paradigm can generate, means that we are facing a deluge of data. The problem is that this data is discarded, either due to low storage capacity or the cost of transporting it to central nodes in the cloud. For this reason, it is important to be able to discern which data is relevant and where it should be stored so that its processing is optimal in terms of performance, security, or response to the user. These challenges can be handled through two new lines of research that, are important to consider: ***edge data quality*** and the inclusion of the ***edge layer in the data virtualization*** paradigm. The first refers to determining when and how important a piece of data is and if it meets adequate quality metrics. The second refers to data virtualization, which is a logical data layer that integrates all the heterogeneous data sources stored in different silos and on disparate systems, manages the unified data for centralized security and governance, and delivers it to users uniquely. The challenge is to add to this virtualization concept the data that is collected at the edge.

Hybrid computing

Hybrid computing is nothing more than a computing infrastructure architecture in which different paradigms are combined, from different types of clouds (public and private) to more complex solutions in which disruptive technologies are introduced, such as ***quantum computing***.

In addition, the challenge is not only in the architectural paradigm to mix such different architectures with complex infrastructure elements such as a container, an HPC or a quantum server, but we also find ourselves with the need to devise new patterns or models to implement and deploy the applications/services/processes in said architectures, with additional security or performance requirements. One of the most interesting forms of secure distributed processing that are being studied in recent times are the techniques of secure multi-party computation[5] (***MPC).***

Efficient & Green computing

Green Computing refers to the efficient use of computing resources with a focus on sustainability. Thus, by considering the environmental impact, Green Computing is moving towards the intelligent use of resources that minimizes the energy consumption and the waste of electronic devices. In this context, it is well known that centralized datacenters are one of the biggest carbon emitters in the world. For this reason, it is important to transfer computing to devices that have hitherto been underused, such as devices at the edge. The problem is current immature technological base mainly in the following aspects: the transfer of AI environments to said devices to enable analytical processing, the field known as Embedded AI[6], the improvement of communication networks in response to the user and privacy, investigating 6G networks and of course the intelligent management of the energy of the devices so that the solution itself does not become a problem in the future.

[5] Lindell, Y. ((2020)). *Secure multiparty computation.* . Communications of the ACM, 64(1), 86-96.

[6] Ray, P. P. ((2022)). A review on TinyML: State-of-the-art and prospects. *Journal of King Saud University-Computer and Information Sciences, 34(4),*, 1595-1623.