Information Modelling and Knowledge Bases XXXV
M. Tropmann-Frick et al. (Eds.)

doi:10.3233/FAIA231165

Opportunities and Pitfalls of IT Architectures for Edge Computing

Luka ČETINA[a,1] and Luka PAVLIČ[a]
[a] *University of Maribor, Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia*
ORCiD ID: Luka Četina https://orcid.org/0009-0000-0267-1092,
Luka Pavlič https://orcid.org/0000-0001-5477-4747

Abstract. Networks and data centres are under a lot of stress due to the rapid growth of connected devices and the significant amount of data that they produce. While data centres are struggling with heavy workloads, the cloud-native IT solutions underutilise the increasingly powerful clients. By bringing computational capabilities closer to the edge, the Edge Computing direction offers a way to alleviate the pressure on data centres, reduce network traffic, and opens the way for novel applications that would take full advantage of the benefits it brings. Edge Computing can also supplement existing applications by giving them the resources to run new, more complex operations, or improve existing ones partially. Edge applications are less reliant on the cloud, and provide more stability and customisation to globally distributed systems.

Keywords. Edge Computing, Cloud Computing, IT architectures

1. Introduction

The number of devices with Internet connectivity is growing rapidly, as well as the volume of data these devices produce, causing increasing strain for the network [1]. Currently, the prevalent paradigm is Cloud Computing, wherein all the data from devices are sent to the cloud for storage and processing [2]. With applications being deployed worldwide and users creating large quantities of data, sending and processing everything in a central location is becoming unsustainable. Networks are struggling to handle the data being sent, and energy usage of data centres is expected to triple in the next decade [3]. Reducing the load for data centres is not only beneficial for the environment, but also for businesses, as less energy means more savings, especially at times when energy costs are high. Edge Computing is sometimes heralded as the solution for these problems, and, while it is no silver bullet, it presents an opportunity to rethink the classic cloud paradigm in domains where the centralisation of storage and processing power is becoming a burden and hindering innovation.

Edge Computing is not a new technology, but rather a topology that brings processing power closer to the places where data are produced and consumed, also called the edge [4]. The edge is where people and things interact with the digital world, and devices such as mobile phones, smart devices, cars and game consoles, which facilitate

[1] Corresponding Author: Luka Četina, Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 46, 2000 Maribor, Slovenia; E-mail: luka.cetina@um.si

this interaction are called edge devices. In Edge Computing, data produced on the edge are not sent to the cloud, but are processed on nearby edge nodes, which are devices deployed close to the edge, capable of running computations [3]. The nodes process the data and produce an answer in real time, which is then sent to edge devices. Due to the many different types of devices, the line between an edge device and an edge node is not always clear. A mobile phone, for example, is an edge device, as it enables the interaction between people and the digital world, yet it also has computing power which can be used to process the produced data [5]. In this paradigm, edge nodes can process some of the data edge devices produce, thereby reducing the amount of data that must be transported to the cloud and processed there. This alleviates the pressure on data centres, and enables edge devices to shift the load onto edge nodes in real time [1].

However, the potential benefits of this are irrelevant if they are not exploited through innovative IT solutions taking advantage of them. As pointed out by [6], while the infrastructure enables edge computing, IT solutions help us realise the benefits. This raises the question of how to design IT solutions not only to work in an edge environment, but also to take full advantage of the benefits it offers. It is also important to be aware of the pitfalls of Edge Computing, as it may not be beneficial in every domain.

The objectives of this chapter are presenting the benefits of utilising Edge Computing on the one hand, and avoiding possible obstacles when designing suitable IT architectures. This is why we present several aspects of Edge Computing specifics related to IT architectures:

- The nature of Edge Computing (advantages, differences between Cloud and Edge Computing),
- The current state and trends while developing edge computing-enabled systems,
- The advantages and limitations of IT architectures for Edge Computing,
- The characteristics of edge computing-enabled system development (typical architecture, different system types, best development practices).

The rest of the chapter is organised as follows. In the next Section, Edge Computing concepts are presented in detail. The current state-of-the-art and future trends are presented in Section three. Possible edge computing opportunities and pitfalls from the IT architecture point of view are presented in Sections four and five. The Chapter is concluded with Section six.

2. Edge computing IT architectures

Edge-computing-based IT solutions can be classified as solutions, designed specifically to take advantage of unique features of edge systems, such as: bandwidth scalability, low-latency offload, privacy-preserving denaturing and WAN-failure resiliency [7].

A typical edge IT solution architecture is three-tier, and is composed of:

- **Edge devices,** which typically run the part of the solution that produces and consumes data. These devices typically do not have enough resources for virtualization, which makes development more challenging, as multiple versions of the software have to be produced to work on different devices. The software is typically installed using application stores, which is a common practice in smartphones, and increasingly in desktop operating systems.

- **Edge nodes** typically run microservices that handle requests from edge devices and communicate with the cloud. Since these devices have more resources available compared to the edge devices, virtualization such as containers is used typically.
- **Cloud** typically contains the services that handle more complex operations, data storage and provide an overview of the entire system.

The described architecture is a basic blueprint, where additional tiers can be added or existing ones removed, such as the cloud. Depending on how well the IT solutions take advantage of the edge architecture, we can classify them into three types, as shown in Figure 1.

Edge-enhanced, device-native IT solutions typically run on devices, and do not require the cloud or edge nodes to function satisfactorily. They can, however, take advantage of edge nodes if they are available, and use their resources to offer new functionalities or improve existing ones.

Edge-enhanced, cloud-native IT solutions are currently the most common type of edge-based IT solutions. Cloud and edge solution development share some similarities, which means that adapting a cloud-native solution to work on the edge is easier than developing a new one, designed specifically for the edge .

Edge-native IT solutions are designed to work in an edge system and utilise its resources fully. They require edge nodes to be available at all times, and lose functionality when otherwise. Despite edge-enhanced, cloud-native solutions being the most common, edge-native solutions are expected to be the most revolutionary and popularise Edge Computing further. While they are in a large part reliant on the underlying infrastructure, history shows that novel applications drive hardware acceptance. A nice example are spreadsheets, which popularised personal computers in the 80's [7].

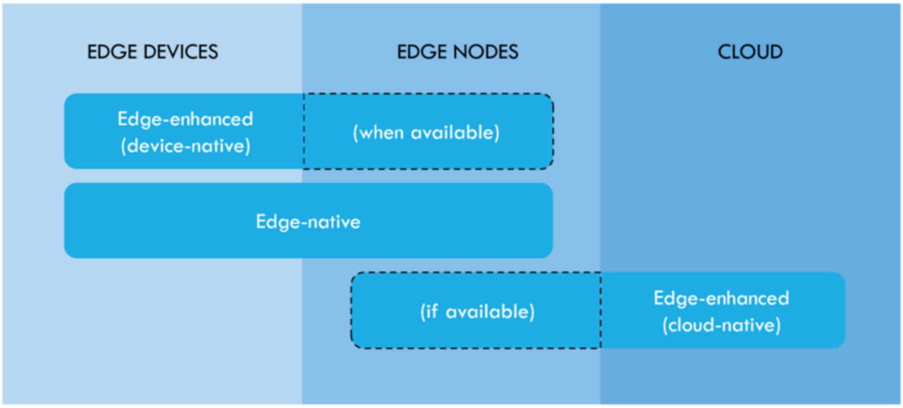

Figure 1: Types of edge-computing-based IT solutions

3. Current trends

The area of Edge Computing-based IT solutions is still developing, therefore the cloud-native solutions being extended to the edge are currently the most common. Even the platforms and tools used most were often intended for cloud-native development [7]. Even so, tends show that edge-native solutions are becoming more common, and the framework and tool ecosystem is evolving as well [8] with tools like OpenYurt [9], KubeEdge [10] and EdgeX [11]. Because the area of Edge Computing is still evolving standardisation is becoming an issue, and should be addressed as noted by [12], while [13] found service migration, security and privacy preservation and deployment as the most pressing issues.

Currently the most researched topic in Edge Computing is Machine and Deep Learning on the edge, followed closely by computational offloading. Less commonly researched topics are security and privacy in Edge Computing, its use in smart cities and homes and the industrial Internet. In 2021 COVID-19 was a hot topic, as a number of articles were published on how Edge Computing could be used to help during the pandemic [8]. Edge-based IT solutions are currently used typically for video surveillance and video processing [1], [7], [8], while another common use is running Deep Learning on the edge [1]. In recent years a number of authors have addressed the topic of combining edge-based IT solutions with technologies like Blockchain and 5G networks [8]. There has also been an increase in the number of articles dealing with the application of Edge Computing in Smart Healthcare and intelligent analysis in recent years [14], as well as combining Edge Computing and technologies such as digital twins, that can be used to provide support for remote working, retail industries and digital advertising [13].

Future research directions of Edge Computing research involve accelerating AI services and improving the efficiency of edge AI, integration with the upcoming 6G and using low-delay edge networks to create digital twins. While Edge Computing is currently focused on supplementing the cloud, it is expected to decouple from the cloud in the future, and instead become edgeless computing, where devices communicate with each other instead of via the cloud. There might be no need for performing computations on the cloud, as devices will be able to share processing power as needed. The cloud could still function as a data repository, but the processing would happen on devices closer to the data source [13].

4. Opportunities

In an edge-based IT architecture the resources for performing basic computations are deployed much closer to end devices than in a cloud-based architecture, therefore edge devices can perform complex operations that would not be feasible without the help of edge nodes. Using those resources, edge devices can offer advanced functionalities that would otherwise be too demanding. That is crucial, as many edge devices are portable (eg. mobile phones, smartwatches, wearables), meaning they have very limited resources compared to the cloud. Due to being highly portable the chances of such devices being in an area where Internet connectivity may be limited or non-existent is high, which presents a challenge for IT solutions that are not designed to work offline. With strategically placed edge nodes, portable edge devices could offer functionalities with high added value in real time, without being dependent on the cloud [1].

Edge-based IT solutions that complement human cognition are often touted as a killer, or revolutionary application of Edge Computing, and the authors in [7] are convinced that they will drive wider adoption of Edge Computing. Such solutions help improve tasks such as face-recognition, long-term memory or navigation, making them of immense value to people with cognitive impairment. The aforementioned use cases can be quite demanding computationally, which means that most devices portable enough to be carried around easily, will struggle to perform them in real time. Current solutions designed to help people with cognitive impairment often rely on humans to perform desired tasks, as is the case for the app "Be My Eyes", where volunteers help people with impaired vision read labels or navigate their environment via video call [15]. A similar solution based on Edge Computing could make use of nearby computational resources, and perform at least some operations without involving volunteers, thereby empowering people who would otherwise be reliant on others.

Due to the large amount of data gathered by web giants like Facebook and Google and many instances of questionable uses of those data, many countries are passing a stricter regulative, detailing which data can be stored and how they can be processed, to respect users' privacy. This can mean, for example, that data gathered from EU citizens can only be kept in data centres located in the EU, or additional safety measures must be taken [16]. If a solution uses a single data centre located in the US and all data are sent there, violations of the data regulative can occur quickly. In an edge-based IT architecture data are kept and processed as close to the source as possible, and only necessary data are sent to the cloud. Since less data are transferred over the network, they are less likely to be intercepted by a malicious third party, which is beneficial to users and companies alike. A key enabler of cloud-based solutions is the assumption that all data will eventually be uploaded to the cloud, where they will be processed and stored. The authors in [5] pointed out that companies often shy away from sharing data due to safety concerns and data transfer costs, which limits possible collaboration and innovation. An edge-network connecting all parties could be established, where edge-nodes would intercept the produced data and send them to all relevant systems, ensuring that data are shared as they are created, reducing the need for intensive batch transfers of data between different systems. The authors saw the most potential in Healthcare, where sensitive data are handled and are rarely shared, due to security risks and outdated systems. Edge-based IT solutions could enable wider collaboration and data sharing, bringing value for organisations and better services for individuals.

Because edge-based IT solutions are deployed close to the edge they can adapt better to the needs of an environment than traditional cloud-based ones. Different edge-nodes can run different versions of the solutions, making sure that, while on the move, users always have access to the solution that is most relevant to their current environment. Companies with many geographically distributed stores or warehouses can also benefit from using Edge Computing-based IT solutions, as it allows them to adapt their system easily to the needs and specifics of each location. A great example of this is the American fast food franchise Chick-fil-A, which abandoned its cloud-based IT solution in favour of an edge-based one [17]. The cloud-based solutions were used to help them coordinate work in the restaurant and provide better service to the customers, but was entirely reliant on the cloud, meaning that it ceased to function in cases of Internet outage. The company also discovered that a uniform solution might not be the best option, as different restaurants were struggling with very different issues and needed a system more adapted to their needs. They decided to develop an edge-based solution and deploy it on an edge node installed in each restaurant. Along with the edge node they installed hundreds of sensors, tracking everything from the queue length to the temperature of the oil in the fryer, and used all those data and the edge node to predict which items would be in demand and must be prepared in advance. Kubernetes [18] was used, to make redeployment and updates as simple as possible, as one of their goals was to create a system that could be customised easily for each individual restaurant. The edge-based solution depends only on the sensors and edge nodes, which means that Internet outages are no longer a major concern.

Edge Computing-based IT solutions can take advantage of edge nodes to take the pressure off the cloud in times of increased traffic. This can be beneficial to web stores that see a high increase in traffic a couple times a year when their servers struggle to process all requests. The most common requests, like adding or removing an article from the basket, are trivial, but can cause problems due to how often they are performed. Such operations could be performed by edge nodes, to give users a responsive experience and synchronised with the cloud in the background, thereby improving the user experience and reducing the number of requests the cloud has to handle [5].

5. Pitfalls

Edge-based IT architectures have great potential, but are in large part dependent on the underlying infrastructure [7], which can be expensive to set up and maintain. Before we decide to develop an edge-based IT solution we must know exactly what we want to achieve and if an edge-based architecture will help or hinder us. Not only is setting up the infrastructure demanding, but deploying our solution to all edge devices and nodes can also be a challenge, due to the large number of devices that can vary greatly in resources and availability [19]. If a device is unavailable during redeployment, the older and the newer versions of the solution must be able to coexist until all devices are updated. Shipping even a small fix can require a redeploy, which takes quite a bit of time and resources. If we are not aware of the unique properties of Edge Computing-based IT solutions, we might find that the cost of maintaining and managing the infrastructure nullifies the benefits of using Edge Computing. Most edge devices are not capable of virtualization, which means that many different versions of the solution must be developed, one for each type of device [5]. Many edge nodes also means that there are more potential points of attack, making it harder to prevent data breaches and maintain data security.

Despite offering more independence from the cloud, most edge-based IT solutions are still dependent on the cloud in some way. When the calculations the solutions need to run are too complex for edge nodes, the solution loses some functionality if the cloud is not available to handle the excess load. The IT solution that offers cognitive assistance mentioned above, would only be capable of running basic operations without edge nodes, making it useless for users who need a response in real-time. Scalability is another issue, as the underlying edge infrastructure is difficult to upgrade due to devices being numerous and distributed geographically. If the number of requests increases suddenly our solution might run out of resources quickly, and be unable to provide responses to all requests. Cloud-native solutions, on the other hand, can scale up quickly to receive more resources when needed. When developing Edge Computing-based IT solutions we must anticipate that the load might have to be reduced, shifted or stopped entirely at any moment. Therefore, one of the main benefits of edge-based IT architectures – independence from the cloud, can turn quickly into a weakness if used improperly [20].

6. Conclusion

Edge-based IT solutions are gaining traction, and many organisations are wondering whether they are the right fit for their use case. Edge based IT solutions have many advantages, such as offering complex functionalities on devices with very limited resources, adapting to the environment and transferring less data through the network. However, they are limited by their dependency on the underlying infrastructure, which can be expensive and hard to manage due to the large number of geographically distributed edge nodes. The most common type of edge-based IT solution is currently the edge-enhanced, cloud-native IT solution, but edge-native IT solutions are evolving quickly, and are expected to become the most common and provide the killer application of Edge Computing. Because the area of Edge Computing is still evolving, it lacks tools, frameworks and standardisation. Current trends for Edge Computing research include improving edge AI and integration with promising technologies such as digital twins, 6G and finding further use cases with Blockchain.

This chapter introduces the concept of Edge Computing and IT architectures that are used commonly. We present current trends, different types of edge-based IT solutions and the opportunities as well as pitfalls they bring.

Acknowledgements

The authors acknowledge financial support from the Slovenian Research Agency (Research Core Funding No. P2-0057)

References

[1] K. Cao, Y. Liu, G. Meng, and Q. Sun, "An Overview on Edge Computing Research," *IEEE Access*, vol. 8, pp. 85714–85728, 2020, doi: 10.1109/ACCESS.2020.2991734.

[2] L. Mei, W. K. Chan, and T. H. Tse, "A Tale of Clouds: Paradigm Comparisons and Some Thoughts on Research Issues," in *2008 IEEE Asia-Pacific Services Computing Conference*, Dec. 2008, pp. 464–469. doi: 10.1109/APSCC.2008.168.

[3] B. Varghese, N. Wang, S. Barbhuiya, P. Kilpatrick, and D. S. Nikolopoulos, "Challenges and Opportunities in Edge Computing," in *2016 IEEE International Conference on Smart Cloud (SmartCloud)*, Nov. 2016, pp. 20–26. doi: 10.1109/SmartCloud.2016.18.

[4] B. Gill and D. Smith, "The Edge Completes the Cloud: A Gartner Trend Insight Report".

[5] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

[6] I. Verma, "Developing at the edge: Best practices for edge computing," Red Hat Developer. Accessed: Feb. 13, 2023. [Online]. Available: https://developers.redhat.com/blog/2020/07/16/developing-at-the-edge-best-practices-for-edge-computing

[7] M. Satyanarayanan, G. Klas, M. Silva, and S. Mangiante, "The Seminal Role of Edge-Native Applications," in *2019 IEEE International Conference on Edge Computing (EDGE)*, Jul. 2019, pp. 33–40. doi: 10.1109/EDGE.2019.00022.

[8] B. Liu, Z. Luo, H. Chen, and C. Li, "A Survey of State-of-the-art on Edge Computing: Theoretical Models, Technologies, Directions, and Development Paths," *IEEE Access*, vol. 10, pp. 54038–54063, 2022, doi: 10.1109/ACCESS.2022.3176106.

[9] "An open platform that extends upstream Kubernetes to Edge | OpenYurt." Accessed: Oct. 03, 2023. [Online]. Available: https://openyurt.io/

[10] "KubeEdge." Accessed: Oct. 03, 2023. [Online]. Available: https://kubeedge.io/

[11] T. L. Foundation, "EdgeX Foundry | The Open Source Edge Platform," https://www.edgexfoundry.org. Accessed: Oct. 03, 2023. [Online]. Available: https://www.edgexfoundry.org

[12] M. Mahbub and R. M. Shubair, “Contemporary advances in multi-access edge computing: A survey of fundamentals, architecture, technologies, deployment cases, security, challenges, and directions,” *J. Netw. Comput. Appl.*, vol. 219, p. 103726, Oct. 2023, doi: 10.1016/j.jnca.2023.103726.

[13] X. Kong, Y. Wu, H. Wang, and F. Xia, “Edge Computing for Internet of Everything: A Survey,” *IEEE Internet Things J.*, vol. 9, no. 23, pp. 23472–23485, Dec. 2022, doi: 10.1109/JIOT.2022.3200431.

[14] J.-H. Syu, J. C.-W. Lin, G. Srivastava, and K. Yu, “A Comprehensive Survey on Artificial Intelligence Empowered Edge Computing on Consumer Electronics,” *IEEE Trans. Consum. Electron.*, pp. 1–1, 2023, doi: 10.1109/TCE.2023.3318150.

[15] “Be My Eyes - See the world together.” Accessed: Mar. 13, 2023. [Online]. Available: https://www.bemyeyes.com/

[16] “Data protection under GDPR,” Your Europe. Accessed: Apr. 14, 2023. [Online]. Available: https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm

[17] B. Chambers, “Enterprise Restaurant Compute,” chick-fil-atech. Accessed: Mar. 14, 2023. [Online]. Available: https://medium.com/chick-fil-atech/enterprise-restaurant-compute-f5e2fd63d20f

[18] “Production-Grade Container Orchestration,” Kubernetes. Accessed: Apr. 17, 2023. [Online]. Available: https://kubernetes.io/

[19] “IBM Edge Computing Field Guide”.

[20] J. Wang, Z. Feng, S. George, R. Iyengar, P. Pillai, and M. Satyanarayanan, “Towards scalable edge-native applications,” in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, in SEC ’19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 152–165. doi: 10.1145/3318216.3363308.