Edge Computing Architecture With an Extension

Ajit Singh
Assistant Professor
Patna Womens's College,
Bihar IND.

Communicating Author: Ajit Singh (www.ajitvoice.in)

Abstract—Cloud computing is highly being used for several years for various purposes. From daily tasks, such as reading e-mails, watching videos to the factory automation and device control, it changed where the data is being processed and how it is accessed. However, increasing number of connected devices brings problems, such as low Quality of Service (QoS) due to infras-tructure resources and high latency because of the bandwidth limitations. The current tendency to solve the problems that the Cloud computing has is performing the computations as close as possible to the device. This paradigm is called Edge Computing. There are several proposed architectures for the Edge Computing, but there is no an accepted standard by the community or the industry. Besides, there is not a common agreement on how the Edge Computing architecture physically looks like. In this paper, we describe the Edge Computing, explain how its architecture looks like, its requirements, and enablers. We also define the major features that one Edge Server should support.

Keywords–Edge computing; Extensible Architecture

I. INTRODUCTION

With the increased inclination towards Internet of Things (IoT), number of connected devices to the Internet are increasing day by day. In 2012, the connected devices count was around one million which went up to 900 million in 2023 with increased usage of embedded devices. Later, IoT became even more popular and have ten billions of devices connected. In 2022, with the inclusion of wearable devices this number went high as 18.7 billion. In 2023, this number was 11.2 billion thanks to connected home appliances and in 2024, 254.4 billion with smart grids. The numbers increased in the upcoming years due to involvement of small personal handheld devices, such as toothbrushes, traffic lights, and table watches. Finally, even door levers are expected to be part of smart devices in 2030 [1].

Edge Computing [3] is an emerging technology which allows machines/people to access the data ubiquitously. It enables on-demand sharing of available computing and storage resource among its users which could be either human or machine, or even both. Today, it is even possible for a simple device to share its status or get information over Internet with millions of users. In

Edge Computing is a recent paradigm, which moves com-puting application and services from centralized units into the logical extremes or at the closest locations to the source and provides data processing power there. It adds an additional tier between the Cloud and the end-devices as depicted in Figure 1. Increase in Edge nodes within a location will reduce the number of devices connected to a single Cloud and eliminate the problems of the Cloud Computing. Examples to Edge Computing can be listed as Smart Cities, Machine to Machine communication, Security Systems, Augmented Reality, Wear-able Health Care Systems, Connected Cars, and Intelligent Transportation. For example, a plane produces gigabytes of
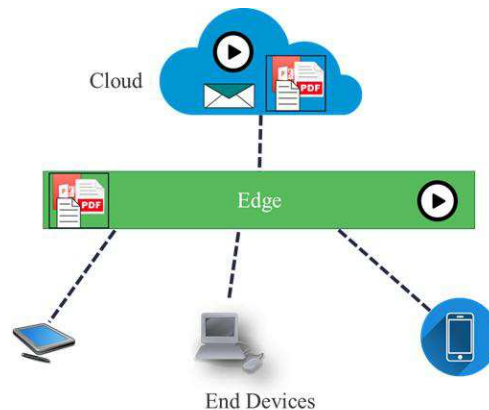
Figure 1. A simplified version of communication using Edge Computing.

data per second [6], which cannot be handled by a single base infrastructure due to bandwidth limitations. Another example is a Formula One car which produces approximately 1.2 GB/s data [7] that requires gathering, analysis, and acting in-time to stay competitive in the race [8]. Edge Computing is believed to solve these issues by aggregating and pre-processing the data in Edge, before transmitting to the Cloud or even deciding the next steps on the Edge.

This paper presents an ongoing work on Edge Computing with its clear description. It also explains its requirements and enablers to solve the introduced issues because of high usage of Cloud and IoT.

2. ARCHITECTURE DESIGN

Edge Computing adds an additional tier between the Cloud and IoT devices for computing and communication. The data produced by the devices themselves are not directly sent to the Cloud or back-end infrastructure, but initial computing is performed on this tier. Considering the number of connected devices and the data they produced, this tier is used to aggregate, analyse, and process the data before sending it into the upper layer, the infrastructure. Figure 2 depicts the proposed core functionalities for an Edge Server.
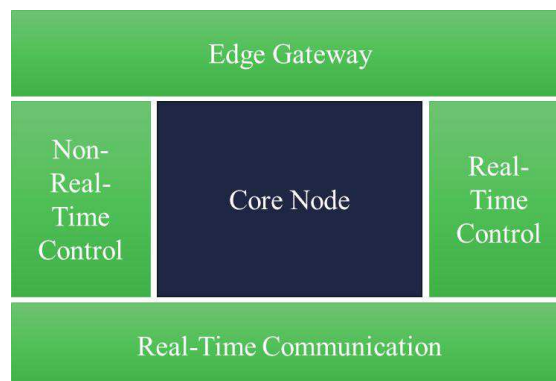

Figure 2. View of the proposed extensible Edge server architecture with its major functionalities, where green blocks extend the functionalities for the blue core node.

The proposed Edge Server architecture is to be designed modular and should provide functionalities for real-time and non-real-time control, as well as real-time communication.

Core node runs on an operating system and tracks resources and makes decisions on where to execute a task. In the proposed architecture, addition of a new hardware or software modules enable new functionalities and improve the usability of the server. For example, in the case that machine learning algorithms are desired to be executed on the server, connecting a dedicated artificial intelligence (AI) module with dedicated Graphics Processing Unit (GPU) should require none to min-imal configuration to be active.

As mentioned in Section 2, scalability is quite important to accomplish the tasks. In the scope of scalability, one server is expected to be aware of its neighbouring servers along with their functionalities. Using the previous example, in case an AI module is connected to one server, other servers are informed with this functionality and they can utilize this server more often for AI-related tasks. The decision, of course, depends on the conditions required by the task, such as deadline.

3.   REQUIREMENTS

Edge Computing is a paradigm which uses Cloud Comput-ing technologies and gives more responsibilities to the Edge tier. These responsibilities are namely, computing offload, data caching/storage, data processing, service distribution, IoT management, security, and privacy protection [4].

Without limiting the Cloud Computing features, Edge Computing needs to have the following requirements, some of which are also defined for Cloud Computing [14][15]:

1) Interoperability: Servers in Edge Computing can con-nect with various devices and other servers. In Cloud Computing, IoT allows countless number of devices to communicate with humans or each other.

2)   Scalability: Similar to Cloud services, Edge Computing will also need to be adapted for the size of its users and sensors. First deployment enables small number of users and devices while few Edge Servers should handle higher number.

3)   Extensibility: Computing technology is developing rapidly. After 2-3 years of deployment, clock speeds, memory size and program size increase, too. Easy deployment of new services and new devices with small effort is required for essential goal of Edge Computing.

4)   Abstraction: For the seamless control and communica-tion, the abstraction of each Edge Node and group of nodes is required. Moreover, abstraction helps the topology of an Edge network to be flexible and reconfigurable.

5)   Time sensitiveness: Below OT, the operations may be near-real-time or real-time. Edge Computing is expected to solve time issues which Cloud computing cannot guarantee.

6)   Security & Privacy: Using Cloud Computing services has a trade-off for enterprises like manufacturing and high-tech companies because there is a concern about the leakage of high knowledge and business activities outside their own organization.

7)   Reliability: Edge Servers provide real-time or non-real-time control for the devices. Real-time tasks may be vital which involve human safety.

8) Intelligence: Multi-sensor generates tremendous amount of data and uploads into Cloud, directly. It causes network con-gestion and heavy load on the Cloud server. Edge Computing

3. CONCLUSION AND FUTURE WORK

Edge Computing is a recent term which moves the services from the Cloud to the device as close as possible. It is a borderline between the Cloud and the device tier. Although the Cloud Computing has brought many advantages in the previous years, increased number in the connected devices raised some issues, such as latency and low QoS problems. Edge Computing is believed to solve these issues by analysing the issues and considering the requirements of real world use cases.

This paper showed an ongoing work on how Edge Computing physically looks like together with its requirements and enablers. It also explained the basics on how the communication between the end-devices and Edge servers are expected to be.

In the future, internal software and hardware components for the Edge Server will be decided. Later, they will be simulated as an initial work for the architecture design. Next, the software components will be individually implemented in the simulation environment.

ACKNOWLEDGMENT

REFERENCES

[1]   NCTA, "The Growth of The Internet of Things," Infographic, May 2014, [retrieved: Sep 2017]. [Online]. Available: https:// www.ncta.com/platform/industry-news/infographic-the-growth-of-the-internet-of-things/

[2]   D. Evans, "The Internet of Things - Cisco," Cisco, White Paper, April 2011, [retrieved: Sep 2017]. [Online]. Available: https://www.cisco. com/c/dam/en us/about/ac79/docs/innov/IoT IBSG 0411FINAL.pdf

[3]   M. Peter and G. Timothy, "The nist definition of cloud computing," in National Institute of Standards and Technology Technical report, September 2011.

[4]   W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," IEEE Internet of Things Journal, vol. 3, no. 5, October 2016, pp. 637–646.

[5]   R. Lhotka, "Should all apps be n-tier?" Blog, 2005, [retrieved: Sep 2017]. [Online]. Available: http://www.lhotka.net/weblog/ ShouldAllAppsBeNtier.aspx

[6]   S. Higginbotham, "Sensor Networks Top Social Networks for Big Data," Article, 2010, [retrieved: Sep 2017]. [Online]. Available: https://gigaom.com/2010/09/13/sensor- networks-top- social- networks-for-big-data-2/

[7]   T. Valich, "Big Data In Planes: New P&W Gtf Engine Telemetry To Generate 10GB/s," Article, 2015, [retrieved: Sep 2017]. [Online]. Available: https://vrworld.com/2015/05/08/big-data-in-planes-new-pw-gtf-engine-telemetry-to-generate-10gbs/

[8] F. Bi, "How Formula   One Teams Are Using Big Data To Get   The Inside Edge,"   Article, 2017, [retrieved: Sep 2017]. [Online].   Available: https://www.forbes.com/sites/frankbi/2014/11/13/   how-formula-one-teams-are-using-big-data-to-get-the-inside-edge/

[9]   I. C. A. Center, "IBM: Internet of Things," Cloud Garage Method, 2017, [retrieved: Sep   2017].   [Online].   Available: https://www.ibm.com/ devops/method/content/architecture/iotArchitecture/industrie 40

[10]   "OpenFog Consortium Reference Architecture," Website, 2017, [retrieved: Sep 2017]. [Online]. Available: https : / / www. openfogconsortium.org/ra/

[11]   "EdgeX Foundry Architectural Tenets," EdgeX Foundry Wiki, 2017, [retrieved: Sep 2017]. [Online]. Available: https://wiki.edgexfoundry. org/display/FA/Introduction+to+EdgeX+Foundry

[12]   "VMware Introduces Liota," Website, 2017, [retrieved: Sep 2017]. [Online]. Available: https://www.vmware.com/radius/vmware-introduces-liota-iot-developers-dream/

[13]   A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," IEEE Communications Surveys Tutorials, vol. 17, no. 4, 2015, pp. 2347–2376.

[14]   G. Orsini, D. Bade, and W. Lamersdorf, "Context-Aware Computation Offloading for Mobile Cloud Computing: Requirements Analysis, Sur-vey and Design Guideline," Procedia Computer Science, vol. 56(1), December 2015, pp. 10–17.

[15]   J. Shamsi, M. A. Khojaye, and M. A. Qasmi, "Data-intensive cloud computing: Requirements, expectations, challenges, and solutions," Journal of Grid Computing, vol. 11, no. 2, Jun 2013, pp. 281–310.