Impact of Edge Computing on Cloud Infrastructure-A Review Paper

Ankur Mahida

Java Developer, Barclays

Abstract: ***Edge computing is an emerging trend that replaces cloud data centers with the provision of computational and data storage resources to the location where they are required, which is in complete contrast to solely relying on the Cloud. Due to the surge in connected devices and data from IoT, Augmented/virtual reality, and mobile applications, cloud computing needs help to keep up with the demands for huge bandwidth and low latency necessary for processing and analysis in real - time. Edge computing overcomes the challenges by decentralizing micro data centers and cloud services to the edge of the network, thus allowing data to be processed instantly and locally by the data generator. Such closeness allows for overcoming difficulties associated with overflow of capacity and latencies. The ecosystem of the edge, encompassing data centers at the edge, cloudlets, and end devices, works together with the centralized Cloud. The 5G deployment, content delivery, self - driving cars, industrial automation, and smart city infrastructure are some of the principal target domains for edge computing. Through the widespread adoption of edge computing, cloud infrastructure will be more diversified, as the edge will have specialized products. Cloud providers might have to tune their infrastructure for edge edge new needs. The collaboration of the edge, cloudlets, and Cloud will facilitate this, which is only possible with the Cloud. Edge computing is a novelty in the age of cloud computing, setting a new course for the distributed real - time computing paradigm.***

Keywords: Edge computing, cloud computing, cloudlets, Internet of Things, 5G, real - time applications

1. Introduction

The incredible spreading connectivity of billions of devices produces exponential data that forms new demands on traditional centralized cloud computing models. There will be a need to process the rapidly increasing amount of data produced by devices like smartphones, autonomous vehicles, and smart sensors located at the network periphery. This will present issues such as network congestion, bandwidth constraints, and latency delays if data centers are the only ones involved and they are located far from the devices. To tackle these challenges, the emerging computing paradigm of edge systems takes computation and data features to the origin of the data by distributing localized micro data centers and cloud services at the edge. Edge computing supports simultaneous analysis and processing of data at the source before sending only the most important information to the Cloud. This way, bandwidth usage and latency are significantly reduced. Although edge computing is anticipated to overturn the infrastructure deployment in the Cloud through supporting distributed and decentralized architectures, it brings about new challenges in managing this heterogeneous infrastructure extending from the Cloud to the edges and into the devices. The next research area this review focuses on is the maturing relationship between cloud computing and edge computing, specifically analyzing how this growth will be reflected in the supply chain of services and cloud infrastructure, as well as how the adoption and capability of edge computing will play out on the future architectures.

2. Problem Statement

Cloud computing architectures based on the centralized model no longer suffice the demands of the current world, which features more interconnectedness with billions of devices online [1]. A gap in network bandwidth is the consequence of placing cloud data centers thousands of miles away from most data sources and users. This large volume of data must be moved in the opposite direction from devices to the Cloud. Under this centralized system, shot bandwidth and high latency issues are due to the physical distance. With emerging devices such as wearables, smart homes, autonomous cars, and industrial devices worldwide that are connected to generating an upper limit of data in almost real - time, legacy cloud interfaces become overwhelmed, struggling to handle the mammoth data ingestion for quick analysis to enable real - time intelligent actions. Also, the single point of failure centralized clouds bring in case of downtime is caused by their fragileness. On the security front, it is the case that having large amounts of data in one place exposes the cloud data centers to great risks. Because distributed computing reduces risk, spaces become the center target in cyber - attacks. Compliance and data sovereignty are additional concerns under a centralized cloud model, as opposed to localized edge data centers that better fit local data residency guidelines [2]. The data distribution model against the Cloud's centralized nature is bringing challenges. Legacy cloud computing models have yet to be modified to address the new demands of an increasingly decentralized network where data is generated, processed, and utilized at the edge.

The boundaries of centralized clusters have led to the emergence of distributed and decentralized computing models, such as edge computing, that place the computation near the sources within the system [3]. The bottlenecks resulting from bandwidth, latency, outages, and centralization issues are tackled more efficiently when there is edge computing domination. The dissipative ecosystem collaborates with central clouds, moving the workload to be better handled on edge outside the core cloud and preserving the advantages of the major Cloud in data storage and historical data aggregation. In a few words, the expansion of connected devices and data volume nowadays is pushing for distributed edge solutions to offload the centralized Cloud and offer an effective solution in real - time and decentralized.

3. Solution

Edge computing offers a fix to the restraints of centralized clouds by making computation, analytics, and data storage closer to where data gets produced [4]. The model has a distributed architecture, where micro data centers and cloud services get "pushed" from colossal mega data centers centralized to numerous edge locations that are localized and closer to data - producing and consuming end users and devices. The network edge is so close that data processing, real - time analysis, and main storage can happen locally before selectively aggregating the key information and transmitting it to the Cloud. Edge computing eliminates the need for big datasets to be carried back and forth to the Cloud. Instead, it keeps data local by concentrating processing power at the edge, resulting in bandwidth loss and increased latency.

Some of the main features of edge computing enable its adoption as broadly as possible.5G networks offer the ultra - fast and low latency connection needed to transmit large volumes of data in real time between smart endpoints and edge data centers [5]. MEC is a novel architecture that brings the cloud services at the network edge closer to users. The MEC facility allows the telecom providers to serve as their edge infrastructure for third - party services. The advancement of high - performance flash storage media opens new possibilities to process and store data on the edge alongside the ability to do it locally. GPU and accelerator chips are also utilized at edge sites to offload edge inferencing and analysis for faster application performance. These facilitators allow executing workloads in the clouds and at the edges in a distributed manner for the best effects. Finally, one solution is to bring cloud resources closer to their data sources through edge computing that is distributed geographically to overcome the bandwidth and latency bottlenecks imposed by centralized clouds. Key enablers – 5G, advanced edge hardware, and MEC – push the idea of setting up micro data centers at the edge to process and store local data flows locally in real time. The symmetric architecture helps with network load lessening, application performance upgrades, and better user experience.

4. Uses

Edge computing is the right match for several emerging technologies and use cases that call for real - time data processing and low - latency interactions. Key examples include:

a) Internet of Things (IoT)

A huge number of edge devices and sensors, together with edge - native workloads, generate a substantial volume of edge data, which drives edge computing [6]. Cloud alone can not ensure real - time analytics and instant response to industrial and connective IoT. Edge computing deployed in the locality allows data filtering and reduction before analytics are sent to the Cloud.

b) *Content Delivery Networks (CDN) –*

Caching content at the edge makes it possible to distribute traffic and lower latency by keeping it closer to the end user [7]. Next comes the thing called edge computing, which takes this further by doing the computation at the edge nodes.

c) Augmented Reality/Virtual Reality

These technologies are based on big data transfers and intensive on - the - spot computing that requires very low latency. The enhanced computing power near where the data is generated also moves towards AR/VR.

d) Autonomous vehicles

Autonomous cars generate intensive data from sensors and cameras, and the analytics should be performed almost in real - time. This computation brings speed into the priority process without involving the Cloud, hence road safety and good driving.

e) Edge Computing Literally

Edge computing's ability to perform real - time analysis on live video streams is one of its greatest benefits. It makes edge nodes perform real - time analysis on locally downloaded streams before the condensed data is sent to the Cloud.

Activities that require immediate response times, real - time interactions, and processing of local data would execute at the edge because of the performance of those time - sensitive tasks, enabling the Cloud to apply to real - time needs. The fact that the network is split up cuts bandwidth usage and latency significantly, making the system more efficient.

5. Impact

Edge computing could transform infrastructure design, physical deployment, and service delivery positively. Edge computing fosters instances of distributed clouds since it offers the opportunity to deploy micro data centers and cloudlets at the edge of space and in place of centralized mega data centers [8]. Thus, what was a citizen of Asgard's physique becomes the exact stature that matches the bodily features of humans.

Cloud providers must rebuild their infrastructure to accommodate this distributed infrastructure for edging computing. Hence, newer edge data centers will augment the big centralized data facilities built on reducing size and optimization. Edge sites might require a configuration of dedicated hardware – e. g., GPUs, FPGAs, and accelerator chips – to support real - time analytics and inference with low latency [9]. Core interconnection to edges will become more important in networking topology. In the future, security architectures must encompass the edge because risks and their distribution may continue [10].

In this scenario, many of the core services currently on the Cloud will be re - architected for the continuum of Cloud and edge, leading to new hybrid and localized services. The example we can give is cloud storage, which can be tiered across the edge, cloudlets, and Cloud. Hyper - scale container orchestration will cover paradigms of edge and cloud servers. New edge - native offerings in the form of remote real - time machine learning pipelines, serverless functions, and streaming analytics will unfold.

In essence, computing breaks the cat mold of Cloud coming, which means meaning providers have to redesign their infrastructure system, provision hardware, and set up the service delivery models and software architectures for this distributed ecosystem. It transforms cloud computing, synchronizing in - network edges and functions with core clouds. The Cloud and edge together form a viable model, allowing the services that were otherwise unattainable. This transformation is ultimately expressed in industry - specific business models and end - user experiences.

6. Scope

Edge computing, however, deals with crucial challenges such as latency, bandwidth, and real - time processing demands that have recently emerged in the 5G and IoT spheres; its capabilities are, nevertheless, comparable to those of cloud computing in some respects.

a) Capacity per storage
Edge locations' storage capacity is limited compared to hyperscale cloud data centers and centralized [11]. Although sufficient for processing real - time data and temporary storage, edge computing cannot replace deep cloud storage for backup, disaster recovery, and other workloads that need full - scale archiving with high - grade backup.

b) Security
The security is improved by distributing threats around the network's edges rather than concentrating them in a central cloud. Nevertheless, the edge is mostly unsecured physically with no depth security controls, which commonly find a place in sophisticated cloud environments [12]. However, the Cloud will have to provide ample security for critical data.

c) Maintenance
Edge infrastructure is geographically distributed at times when it is remote in some cases, thereby making its upkeep more challenging compared to a few large centralized data centers. It can result in difficult operations such as monitoring, troubleshooting, and upgrading edge locations.

d) Usage
Edge computing can be found in distributed, localized, real - time applications, not the last ones being backend workflows like payroll, ERP, data warehousing, etc [13]. They still function well remotely. Applications for many gain a driving force of data aggregation in cloud computing.

Edge computing helps tackle a specific range of issues connected to distributed and localized environment - sensitive and real - time applications but still needs to improve in security, capacity, maintenance, and scope compared to the Cloud [14]. It works as an extension or a supplement instead of a cloud - based services provider.

The Cloud will continue to run the backend and core backend processes and serve as a powerful repository of authoritative data. It is a central facilitator for aggregating and correlating edge data distributed across different nodes. AI training, deep learning, backup, and disaster recovery, for which the Cloud is more appropriate, include high - computation tasks.

7. Conclusion

Edge computing will be the key enabler to transform the current cloud infrastructure when billions of interconnected devices need low - latency interactions on the edge. The centralized cloud computing paradigm, which aggregates all resources in a few mega data centers, does not suffice the performance needs of the upcoming technologies like 5G, IoT, connected vehicles, augmented reality/virtual reality that require near real - time data processing, and low millisecond latency. Edge computing is the answer to the problems of delay in data transfer and the challenges of limited bandwidth as it distributes local data centers, processing power, and cloud services near the data sources and the users at the network's edge. Although edge computing suffers from safety and maintenance drawbacks vis - a - vis the Cloud, its capability of immediate analysis and response for distributed data makes it practically irreplaceable. The future cloud infrastructure will probably be an organized system combining big cloud data centers with smaller, distributed edge data centers and smart end devices. Cloud providers will need to refactor their infrastructure, hardware, software stacks, and service delivery to take full advantage of edge computing, which covers the Cloud and the edge ports in a distributed environment. Edge computing enables the creation of breakthrough new services that were unthinkable to be implemented using centralized clouds before. But its full potential will be realized only through close cooperation between the Cloud, edge networks, telecom operators, and actual applications that will benefit from it. In essence, edge computing will help develop cloud computing to have network and localized/decentralized architectures to meet the growing demands of networks and application architectures. It comes with the added advantage of exploiting the potential of cloud infrastructure and is the gateway to innovation, which is a new era in cloud computing.

References

[1] A. Botta, W. de Donato, V. Persico, and A. Pescapé, “Integration of Cloud computing and Internet of Things: A survey, ” *Future Generation Computer Systems*, vol.56, pp.684–700, Mar.2016, doi: https: //doi. org/10.1016/j. future.2015.09.021.

[2] D. C. Klonoff, “Fog Computing and Edge Computing Architectures for Processing Data From Diabetes Devices Connected to the Medical Internet of Things, ” *Journal of Diabetes Science and Technology*, vol.11, no.4, pp.647–652, Jul.2017, doi: https: //doi. org/10.1177/1932296817717007.

[3] A. Aral and T. Ovatman, “A Decentralized Replica Placement Algorithm for Edge Computing, ” *IEEE Transactions on Network and Service Management*, vol.15, no.2, pp.516–529, Jun.2018, doi: https: //doi. org/10.1109/tnsm.2017.2788945.

[4] S. Nastic *et al.,* "A Serverless Real - Time Data Analytics Platform for Edge Computing, " *IEEE Internet Computing*, vol.21, no.4, pp.64–71, 2017, doi: https: //doi. org/10.1109/mic.2017.2911430.

[5] M. Ruffini, "Multidimensional Convergence in Future 5G Networks, " *Journal of Lightwave Technology*, vol.35, no.3, pp.535–549, Feb.2017, doi: https: //doi. org/10.1109/jlt.2016.2617896.

[6] W. Yu *et al.,* "A Survey on the Edge Computing for the Internet of Things, " *IEEE Access*, vol.6, pp.6900–6919, 2018, doi: https: //doi. org/10.1109/access.2017.2778504.

[7] K. Zhang, S. Leng, Y. He, S. Maharjan, and Y. Zhang, "Cooperative Content Caching in 5G Networks with Mobile Edge Computing, " *IEEE Wireless Communications*, vol.25, no.3, pp.80–87, Jun.2018, doi: https: //doi. org/10.1109/mwc.2018.1700303.

[8] K. Hwang, *Cloud computing for machine learning and cognitive applications*. Cambridge, Ma: The Mit Press, 2017.

[9] Y. Umuroglu *et al.,* "FINN, " *Proceedings of the 2017 ACM/SIGDA International Symposium on Field - Programmable Gate Arrays*, Feb.2017, doi: https: //doi. org/10.1145/3020078.3021744.

[10] S. Khan, S. Parkinson, and Y. Qin, "Fog computing security: a review of current applications and security solutions, " *Journal of Cloud Computing*, vol.6, no.1, Aug.2017, doi: https: //doi. org/10.1186/s13677 - 017 - 0090 - 3.

[11] K. Bilal, O. Khalid, A. Erbad, and S. U. Khan, "Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers, " *Computer Networks*, vol.130, pp.94–120, Jan.2018, doi: https: //doi. org/10.1016/j. comnet.2017.10.002.

[12] D. A. B. Fernandes, L. F. B. Soares, J. V. Gomes, M. M. Freire, and P. R. M. Inácio, "Security issues in cloud environments: a survey, " *International Journal of Information Security*, vol.13, no.2, pp.113–170, Sep.2013, doi: https: //doi. org/10.1007/s10207 - 013 - 0208 - 7.

[13] M. Frank, P. Roehrig, and B. Pring, *What to do when machines do everything: how to get ahead in a world of AI, algorithms, bots, and big data*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2017.

[14] F. Jameel, Z. Hamid, F. Jabeen, S. Zeadally, and M. A. Javed, "A Survey of Device - to - Device Communications: Research Issues and Challenges, " *IEEE Communications Surveys & Tutorials*, vol.20, no.3, pp.2133–2168, 2018, doi: https: //doi. org/10.1109/comst.2018.2828120