ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/sriramaraju-sagi/53

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 1 & Iss: 1

Research Article

Bridging the Data Divide: A Study on Seamless Data Transfer across Edge, Core, and Cloud

Sriramaraju Sagi*

Sriramaraju Sagi, NetApp, USA

Citation: Sriramaraju Sagi. Bridging the Data Divide: A Study on Seamless Data Transfer Across Edge, Core, and Cloud. *J Artif Intell Mach Learn & Data Sci 2022*, 1(1), 116-119. DOI: doi.org/10.51219/JAIMLD/sriramaraju-sagi/53

Received: October 10, 2022; **Accepted:** October 18, 2022; **Published:** October 20, 2022

***Corresponding author:** Sriramaraju Sagi, NetApp, USA

ABSTRACT

This research delves into enhancing the efficiency of transferring data across environments, like edge, core and cloud in data center setups. With the rise of transformation, the increasing volume of data generated at the edge by devices calls for effective strategies to move data to central hubs and the cloud. Our study focuses on developing an approach to managing data flow from edge locations to the cloud. We leveraged NetApps SnapMirror technology for data migration from the edge to core systems, where data lakes were used to prepare information for core applications like Oracle and SAP HANA. A hybrid cloud model then facilitated transfer of data to the cloud for analysis, backup, and disaster recovery purposes. The results highlight how a unified infrastructure platform can boost efficiency, minimize delays, enhance data security and enable real time processing across computing environments. This research serves as a guide for optimizing data movement in businesses driven by data showcasing how integrated infrastructure solutions play a role in meeting the evolving demands of today's world.

Keywords: Data Movement Optimization, Edge Computing, Core Data Centers, Data Lakes

1. Introduction

The rapid advancement of transformation has led to shifts in how data is created and processed owing to the growing presence of Internet of Things (IoT) devices and the rise of edge computing. This transformation has sparked an increase in the volume of data generated at the edges of networks necessitating techniques for moving this data from peripheral locations to central data centers and eventually to the cloud. The significance and relevance of data movement lies in its ability to enhance efficiency, improve decision making processes and create opportunities for innovation across various industries.

Edge computing plays a role by reducing latency optimizing usage and enabling real time processing through the relocation of computation and data storage closer to where they are needed. However, integrating edge computing into data center operations and cloud services presents challenges. These challenges include managing volumes of data ensuring the accuracy and security of data during transit and optimizing infrastructure for data transfer. Addressing these obstacles is vital as it directly impacts business's ability to leverage their data effectively thereby influencing their competitiveness, operational efficiency, and innovative capabilities.

The transfer of data, from sites to the central operations data center and to the cloud is crucial for a variety of reasons. Initially it supports the consolidation, analysis and long-term storage of data which're essential for gaining insights, enhancing decision making and driving business intelligence. Additionally, it allows for the execution of analytics, Artificial intelligence (AI) and machine learning (ML) models that depend on vast information collected from diverse sources. Efficient data movement ensures compliance with data governance and privacy regulations by transferring and storing data in locations.

This research seeks to explore the obstacles and possibilities in managing data movement from edge locations to core operations

data centers and the cloud. It will pinpoint the challenges and opportunities in handling data transfer across computational environments. The research will assess existing data center infrastructures in terms of scalability, reliability, security, and performance. It will suggest infrastructure solutions by proposing enhancements in architecture, technology integration and best practices for managing data, across edge, core and cloud environments. The research will also show how operational efficiency can be influenced, illustrating how the right data center setup can improve efficiency, lower delays, boost data protection and support real time data processing abilities.

2. Enhancing Data Flow in Modern Data Centers

Data centers play a role in today's landscape serving as central hubs for storing organizing and exchanging data in global IT activities. The rapid growth in data volume driven by advancements in technologies, cloud computing and AI has created a need for enhanced data transfer capabilities within these facilities. The demand for data management and transmission systems is not a response to the increasing data flow but also a proactive measure to ensure the scalability, efficiency, and reliability of data center operations to meet future digital needs.

The flow of data within data centers involves a network of processes that encompass the intake, storage, manipulation, and movement of data across various parts of the IT infrastructure. The efficiency of these processes is critical to the performance of the data center and influences factors like system speed and energy usage. However traditional data center setups often struggle to adapt to the nature of data streams marked by intermittent spikes in traffic and the need for real time processing capabilities.

Enhancing data flow, in data centers comes with a range of obstacles. These hurdles include dealing with the growing amount of data addressing delays caused by data being spread out geographically managing security risks from connecting networks and systems and coping with the increased energy consumption needed to handle datasets effectively. These challenges make it tough to ensure access to data for analysis and processing, posing obstacles to organizations sustainability objectives.

To tackle these issues the industry is adopting strategies in data center management. This includes utilizing Software Defined Networking (SDN) to adjust bandwidth and routes based on data flow needs. Edge computing is another approach that processes data closer to its source reducing reliance on data centers and cutting down on latency and bandwidth usage. Innovative storage solutions like all flash arrays and object storage enhance data retrieval speeds and system performance. Additionally leveraging AI and machine learning algorithms can forecast patterns in data flow optimizing operations while minimizing delays. Implementing energy infrastructure designs such as cooling technologies and renewable energy sources can also help reduce the impact.

Improving data transmission in data centers is essential to meet the growing demands of technologies and to prepare for the future of global IT operations. By implementing technologies and strategies data centers can effectively manage volumes of data while maintaining their performance, efficiency, and scalability. The evolution of data center operations towards intelligent systems showcases not only technological advancements but also underscores the crucial role that data centers play in shaping today's digital society.

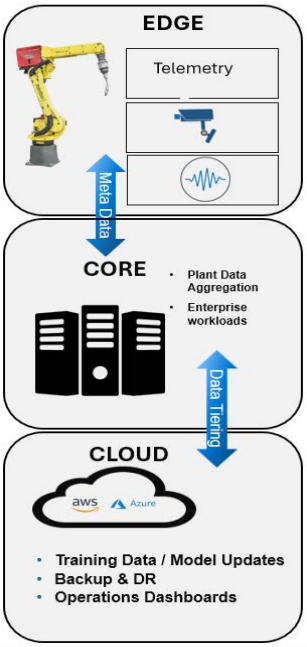

Figure 1: Enhancing Data Flow Edge to Core to Cloud

3. Literature Review

Various studies have proposed methods to optimize data flow, from the edge to the core and eventually to the Cloud[1,2] both explore the use of software defined networks (SDN) to improve data transfer efficiency. Aujla focused on workload slicing while Muñoz integrates Internet of Things (IoT) transport SDN and edge/cloud computing. The research highlights the effectiveness of their approach in managing data applications in an edge cloud setting. Evaluation of this approach considers factors like energy consumption, delay, SLA violations, migration rate and cost reduction. Notably there is a decrease in energy usage across the environment and networks because of this scheme. The study introduces a novel multilayer transport architecture that considers IoT systems. This architecture aims to distribute analytics between the cloud and network edges while also regulating IoT traffic flow. Experimental validation confirms the efficiency of this architecture.

Both[3,4] highlight the benefits of combining edge and cloud computing. Uddin talks, about the shift from data centers to ones while Balouek Thomert focuses on merging resources for data driven processes. Moving towards data centers near network edges allows for operations in real time. Although edge to cloud computing offers advantages it also comes with challenges. Hence exploring edge architecture is crucial to leverage the potential of this computing approach. The research delves into ways of using real time data with models underscoring the need for a seamless computing environment and streamlining the integration of edge, to cloud systems to support data centric operations

Each[5,6] offer solutions for placing operators in a distributed manner and migrating data in a time-efficient manner. Renart's solution is specifically designed for IoT data analytics, while Marium's solution is focused on cloud migration. The strategy effectively reduces end-to-end latency, edge-to-cloud data transfers, and message costs in IoT applications, surpassing typical cloud configurations, Random, and LB placement approaches. The results demonstrate the efficacy of the suggested operator placement method in optimizing performance indicators. The researchers investigated contextual elements that

contribute to shorter transmission time. They employed Entity Framework Core technology and Object Related Mapping to enhance time efficiency. Therefore, they successfully improved data transmission between Azure and Gear Host cloud.

Both[7,8] tackle the issue of optimizing data transfer within the cloud. Janet specifically emphasizes the use of effective compression methods, while Teli focuses on minimizing costs. The primary discoveries involve addressing the difficulty of scheduling large-scale data transfers at various levels of urgency to optimize the utilization of bandwidth between data centers. This includes proposing the enhancement of data transfer through the implementation of effective compression and decompression techniques on the client's end, as well as recommending the utilization of techniques such as "snappy" and "BDT" to enhance the efficiency of data movement. The study addresses the issue of transferring large amounts of data across cloud data centers. It presents an effective algorithm for optimizing the cost of data transfer and demonstrates that using a graph model as the chosen mechanism offers a superior option for minimizing expenses.

4. Results

Our study employed a unified infrastructure system, developed in partnership with Cisco and NetApp to investigate how data is transferred efficiently from the edge to the core and eventually to the cloud within data center setups. In this section we present an overview of the outcomes derived from our setup emphasizing the seamless capture of edge data its transmission to a central operational data center and subsequent integration with cloud services for analysis, backup, and disaster recovery purposes.

The initiative centered around a unified infrastructure system that computing, storage and networking components into a unit for easy management. This system facilitated a rapid setup process enabling efficient data gathering at the edge. Through this framework adoption we were able to create an environment closely mirroring real world data center operations. This setting offers insights into implementing data transfers from edge locations to central hubs and onward to cloud platforms.

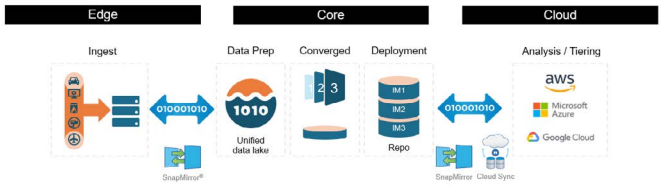

Figure 2: Experimental Lab setup from Edge to Core to Cloud

Data was smoothly brought in at the edge capturing real time information, from devices and other edge technologies. In the data transfer phase, we utilized NetApps Snap Mirror technology to efficiently replicate data from the edge to our primary operations data center. This technology facilitated an efficient and secure data transfer process ensuring the integrity and availability of data throughout. This phase demonstrated the practicality of employing technology for data replication needs within a remote data center setup.

Upon arrival at the central operations data center the data underwent a preparation process involving utilization of data lakes. Here the collected data was optimized for use in core applications like Oracle and SAP HANA. This step was crucial for ensuring that the data was in a format for analysis and processing, by these applications highlighting the effectiveness of using data lakes to handle amounts of diverse data sources.

In the phase of our study, we moved the data we had prepared to the cloud to fulfill needs, for analysis, backup and disaster recovery. We adopted a mix of infrastructure and cloud services to ensure data exchange. This method not simplified large scale data analysis at a cost but also showcased the flexibility and resilience of hybrid cloud structures for implementing data organization strategies. Leveraging cloud services allowed us to carry out data analytics by tapping into the clouds computing power. Moreover, we used cloud services for creating backup and disaster recovery mechanisms enhancing data management and security.

The outcomes of our study emphasize the effectiveness of utilizing an integrated infrastructure platform in collaboration with Cisco and NetApp to optimize data flow across edge devices, central systems, and cloud platforms. By leveraging NetApps Snap Mirror technology for transferring data integrating data lakes for preprocessing information and adopting a hybrid cloud approach for analyzing and organizing data tiers we effectively demonstrated a streamlined method, for overseeing data center operations. This study offers a roadmap for improving data transfer, between computing setups. Demonstrates how these structures can adapt to the changing requirements of modern data focused businesses.

5. Conclusion

The outcomes of our test confirm the effectiveness of using a converged infrastructure approach in collaboration with Cisco and NetApp to facilitate the transfer of data from edge devices to data centers and then to cloud based services. By utilizing NetApps SnapMirror for replicating data integrating data lakes for preparing data at the core and adopting a cloud strategy for data processing we have demonstrated a practical and efficient method for managing data across edge, central and cloud environments. This strategy improved the scalability and adaptability of operations in data centers. Also ensures the security and compliance of data handling procedures addressing key challenges in contemporary practices of moving data. Our research adds to the existing knowledge base on optimizing data flow within data centers by providing insights and strategies that organizations can implement to enhance their IT infrastructure amid increasing demands for handling data.

6. Acknowledgement

This article was supported by the technical team from Cisco and NetApp, we sincerely thank them for their contribution and valuable input.

7. References

1. Aujla GS, Kumar N, Zomaya AY, Ranjan R. Optimal decision making for big data processing at edge-cloud environment: An SDN perspective. IEEE Transactions on Industrial Informatics, 2018;14: 778-789.

2. Muñoz R, Vilalta R, Yoshikane N, et al. Integration of IoT, transport SDN, and Edge/Cloud computing for dynamic distribution of iot analytics and efficient use of network resources. Journal of Lightwave Technology, 2018;36: 1420-1428.

3. Uddin MF, Ahmad S. A review on edge to cloud: paradigm shift from large data centers to small centers of data everywhere. 2020 International Conference on Inventive Computation Technologies, 2020; 318-322.

4. Balouek-Thomert D, Renart EG, Parashar M, Simonet A, Zamani AR, Parashar M. Towards a computing continuum: Enabling edge-to-cloud integration for data-driven workflows. The International Journal of High Performance Computing Applications 2019;33: 1159 - 1174.

5. Renart EG, Veith ADS, Balouek-Thomert D, Assuncao MDD, Lefevre L, Parashar M. Distributed Operator Placement for IoT Data Analytics Across Edge and Cloud Resources. 2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2019; 459-468.

6. Marium SM, Thebo LA, Jaffari SNA, Memon MH. Time Efficient Data Migration among Clouds. arXiv, 2018.

7. Janet J, Balakrishnan S, Prasad ER. Optimizing data movement within cloud environment using efficient compression techniques. 2016 International Conference on Information Communication and Embedded Systems, 2016; 1-5.

8. Teli P, Thomas MV, Chandrasekaran K. An efficient approach for cost optimization of the movement of big data. Open J. Big Data, 2015;1: 4-15.

9. Bangui H, Rakrak S, Raghay S, Buhnova B. Moving to the edge-cloud-of-things: recent advances and future research directions. Electronics, 2018;7: 309.