Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2022, 9(8):75-81

Research Article ISSN: 2394 - 658X

Optimizing Library Data Handling with Cloud and Big Data Technologies

Naga Durga Srinivas Nidamanuri

Independent Researcher,
6 Capano Dr, Apt B2, Newark, Delaware-19702
Nidamanuri.nagadurga@gmail.com.

ABSTRACT

There is a tremendous quantity of change happening in the environment right now. People do not underestimate the library's importance because of its reputation as a treasure trove of knowledge. Unfortunately, due to their busy schedules, individuals nowadays do not have time to attend libraries. That is why they rely on online resources rather than physically visiting libraries; they are always up-to-date. Many benefits are available to users of cloud computing, such as the ability to save backups, store data on a massive scale, and safeguard oneself against disasters. With the advent of cloud computing, libraries have the ability to store massive amounts of data on their websites and digital resources. All of this information will be safely kept. A virtual platform is made available on library websites by a technology called cloud computing, which manages all of the data that is readily available through the internet. Because cloud service providers sell their services to clients and payment is based on a "pay as peruse" approach, using cloud computing technology in library science simplifies the management of massive amounts of data. Cloud computing services allow for the uploading and management of electronic content, which is a result of the fact that libraries nowadays are quite powerful in terms of their electronic content. There are many advantages that cloud computing can bring to the field of library sciences. The findings show that libraries utilizing cloud services experience significantly fewer downtime events and data loss, leading to increased reliability and user satisfaction. Compared to traditional systems, cloud-based libraries offer greater flexibility and ease of data management while reducing infrastructure costs. The analysis underscores the transformative role of cloud computing in modern library science, ensuring that libraries remain relevant and accessible in the digital age.

Keywords: Cloud, Data Handling, Big data.

INTRODUCTION

"Cloud computing is defined as a model for enabling ubiquitous convenient on demand network access to a shared pool of configurable computing resource (e.g. network, server, storage, and applications) that can be rapidly provisioned and released with minimal management effort or service provider interaction," says the National Institute of Standards and Technology (NIST). "Cloud computing is defined on the basis of this model." The word "cloud computing" has just recently entered common usage and found applications across numerous sectors. The three primary models via which cloud computing offers its services are software as a service, infrastructure as a service, and platform as a service. Just defined, "cloud computing" refers to a way to store and access large amounts of data from any location with an internet connection and the right information technology tools.

Modern organizations throughout the globe are embracing a variety of cutting-edge cloud technologies to digitise their data, analyze it, and get the forecasts and insights that power their operations. As IoT, cloud computing, and big data analytics come together, new opportunities are emerging in a number of industries. Some examples of such industries are advertising, internet shopping, and healthcare [1].

As the time changes, these possibilities may become apparent. While going fully digital is a must for any modern business, not every company can afford to invest heavily in the necessary infrastructure, IT resources (such as fast connections, computers, software, and memory), and IT personnel to keep everything running smoothly.

The ability to gather, use, and manage computational resources independently is made possible by computing in the cloud, which offers a Web-based architecture [2].

Users and organizations alike can take use of the extensive range of services that cloud computing provides thanks to its well-equipped data centers and tightly connected resources. It may be possible to dynamically offer end users with available resources in order to satisfy their requirements. Cloud computing eliminates concerns about the origin, scalability, and resource limits of the underlying service, allowing individuals or organizations to access and exploit the whole pool of computational resources. As a result, businesses and individuals can use cloud resources as needed without having to commit to anything in advance and can instead pay for what they use [3, 4].

The rapid expansion of the Internet of Things (IoT) and the pervasiveness of mobile phones have made it easier to combine technology and medicine, leading to better healthcare for individuals all around the globe [5]. Data is proven to be a very useful and efficient instrument for health improvement. Compliance and success are being ensured in the healthcare, telecom, BFSI, retail trade, and e-commerce industries by utilizing the newest breakthroughs in data science and analytics powered by AI and ML [6]. Due to its heterogeneity, real-world data presents a number of significant challenges when it comes to storage, analysis, and investigation. When dealing with data that is high velocity, diverse, and volume, it is crucial to use cutting-edge components of big data and cloud computing for processing. A large volume of data can be defined as information flowing in a variety of formats, including both structured and unstructured data, from a number of different sources. Data storage, analysis, and consumption on an end-to-end encrypted platform is possible with the help of the components provided by cloud computing service providers.

Nevertheless, the issue of what resources should be made available and how the architecture should be built using all of these elements remains an open subject. In light of the field's recent remarkable growth in cloud computing, Gartner [9] predicts that the public cloud will reach $411 million by the end of 2020.

Because of the proliferation of cloud computing, it is becoming more challenging for enterprises to select a reliable cloud provider that can meet their needs in the long run from the many available options. There are a lot of cloud service providers out there right now, but no one standard, and they're all growing at the same rate. Many of these service providers focus on computing power and supply end customers with services related to storage, databases, networking, and central processing unit (CPU).

A number of network operators place a higher priority on reducing costs, while others place a higher priority on maintaining high service quality and agility. Due to the many elements at play, it has become extremely tough to select a suitable service provider that meets the needs of the client or company [10, 11]. Data management in the cloud is a hot subject in IT circles, both academically and professionally. As cloud computing becomes more widespread, businesses are increasingly turning to cloud-based databases as a means of storing, managing, and analyzing massive volumes of data. The administration of data in the cloud provides various benefits, such as scalability, cost-effectiveness, and flexibility, which enables enterprises to effectively manage massive data sets that are always changing. With data expected to continue growing exponentially in volume, variety, and velocity, businesses must adopt effective cloud data management strategies to gain insights, make data-driven decisions, and gain a competitive edge.

The idea of storing computers on the cloud is still quite new, but it is rapidly becoming popular around the globe. People all throughout the globe are starting to favor cloud computing. Database administration has seen a global paradigm shift due to the advent of cloud computing. Managing cloud computing entails keeping an eye on and adjusting a lot of moving parts, like the platforms, software services, and infrastructures that make up the cloud. Management of cloud computing is what this is called. Management of cloud computing entails keeping tabs on and adjusting various parts of cloud computing settings, such as the underlying hardware, operating systems, and software programs. This is referred to as cloud computing management. There are several parts that come together to build this whole, including optimization of costs, performance, availability, security, and cloud resource management. With cloud computing, users have access to information technology services via the web. Databases, storage, and processing power are all available through technical services, and cloud companies like Amazon Web Services can be of assistance [12]. This is a new approach that can be used instead of the old-fashioned way of operating servers and data centers.

LITERATURE REVIEW

There has been a lot of buzz around cloud data management as of late, thanks to the proliferation of cloud computing as a data processing and storage solution. The goal of this literature study was to provide practitioners and businesses with valuable insights into cloud data management by identifying important concepts, technologies, and best practices. This literature review will highlight the various research and investigations that have recently focused on cloud data management. The ability for clients to personalize their use of IT resources is one of the several benefits of cloud computing. Organizations can now respond rapidly to shifting business demands without having to reinvest much in IT infrastructure, thanks to this. With the advent of cloud computing, there has been a sea change in the way IT resource data is managed. Companies increasingly store all of their data in the cloud instead of in physical locations. As a result, efficiency, cost savings, and adaptability have all improved [13].

A. Challenges of Cloud data management

In order to show the challenges that are related with cloud data management, a number of studies have been conducted. [14] A number of studies recommend that enterprises thoroughly consider data privacy, data security, and data recovery prior to using cloud data management systems within their organizations. One of the purportedly most important aspects of large data management is the maintenance of data security, privacy, and efficiency [15].

Managing large amounts of data is an area where this is particularly relevant. Finding answers to the problems that cloud data management causes in terms of data privacy and security management is crucial [16].

B. Framework and models for cloud data management

A lot of research has concentrated on the theories and models that are necessary for good cloud management. If companies want an effective and thorough strategy for managing their data in the cloud, they should follow this structure, says the author [6]. The combination of processing, analytics, and storage are just a few of the data management services suggested by another author for use in the cloud [17].

C. Importance of data governance and Metadata management

Gholami highlights the significance of data governance and metadata management when talking about cloud data management. Businesses should prioritize effective data governance and metadata management procedures to guarantee their cloud data is well-managed and readily available, according to the authors [18].

D. Cloud Database Providers and Services

Businesses have a lot of options when it comes to cloud providers and services for data management. Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) are just a few of the well-known cloud providers that provide a variety of database services. These cloud-hosted databases provide scalable and adaptable solutions, making them ideal for storing and managing massive amounts of data in the cloud [19].

E. Database Design and Implementation

There are a lot of factors that need to be carefully considered while developing and deploying a database system in the cloud. Modeling data, designing schemas, indexing, and partitioning are all such aspects. Databases hosted in the cloud often offer a fresh set of optimization tactics and design patterns compared to more conventional on-premises databases. Databases that are more suited to structured data, such as MySQL and PostgreSQL, and NoSQL databases, such as Cassandra and MongoDB, are examples of cloud-based data management solutions [20].

F. Data Management Functionality

Ingestion, storage, retrieval, querying, and analytics are just a few of the many data management tools that make up cloud data management. Several data management choices, including built-in data replication, backup, and recovery capabilities, are available with cloud-hosted databases, ensuring the availability and durability of data. The sophisticated analytics and machine learning features of cloud-based databases also allow companies to get more value out of their data [21].

G. Data Security and Privacy

Every data management system prioritizes the privacy and security of data in the cloud. In order to prevent data breaches and illegal access, organizations should think about establishing data encryption, authentication, access control, and audits. To ensure the safety of data stored in the cloud, cloud service providers often implement numerous security measures. Among these capabilities are security monitoring, encryption both while in transit and at rest, and identity and access management (IAM) [22].

1) Techniques for Improving Performance: The performance of cloud-based databases must be optimized in order to guarantee that data processing and retrieval are carried out in an effective manner. Improving query performance and lowering latency can be accomplished by the utilization of many techniques, including caching, indexing, and denormalization. In addition, cloud service providers offer performance monitoring and tweaking capabilities, which are designed to enhance the efficiency of cloud-based databases. When attempting to optimize performance in the cloud, enterprises must also take into consideration a variety of aspects, including network latency, the costs of data transfer, and the proximity of the data [23].

H. Evaluation of Cloud-based Databases

There are many factors to think about while trying to determine the worth of cloud-based databases, including performance, scalability, reliability, security, and cost. Throughput, reaction time, and cost per query are some of the metrics that businesses can use to compare and evaluate various cloud-based databases. When choosing a database system, companies should also think about the specific needs of their data management workload [24].

ASSOCIATED RISK OF CLOUD COMPUTING

• **Data migration by service provider:** The fact that service providers often put their own financial interests ahead of their customers' data is a common observation. In addition, since the service provider has the last say in data processing, they unfairly benefit from it. This occurs because the data is ultimately kept on a server that is located at a remote location and is controlled by the service provider.

• **Highly costly for big data:** A common pricing model for cloud services is based on use of resources or storage capacity. The massive quantity of storage and processing power required by big data presents a problem for organizations that deal with massive data sets.
As a result, the total amount charged can be quite high due to the rental of capacity and the application of large resources. In contrast to that less expensive choice for building infrastructure, it is extremely expensive.
• **Leakage of sensitive data:** It has been observed on multiple occasions that cloud service providers may divulge very sensitive information for their own personal gain, resulting in a breakdown of trust between the client and provider.
• **Lack of legal provision:** Strict laws and regulations need to be implemented by the government in this arena because cloud computing contracts between clients and cloud service providers are not governed by any authorized entity. Dealing with disputes in this context is therefore fairly complex.
• **Poor quality service:** While some cloud service companies offer incredibly low prices, their service quality is far from satisfactory. It fails to resolve any issues that may arise and instead offers subpar service to its customers. As a result, issues emerge between customers and cloud service providers.

CHALLENGES

Acquiring the required data is a major hurdle for any cloud-based big data solution aiming at business or IT analytics. Analytical applications involving personally identifiable information should not be considered without first consulting relevant organizations or individuals, as data sources may be untrustworthy, include inaccuracies, or interfere with the intended analysis.
In order to get useful analytics, it is crucial to combine data from both internal and external systems. Data that is so diverse and diverse in format also becomes exceedingly challenging to handle. Making sure decisions are accurate and consistent becomes an issue when data from many forms is integrated. It takes a lot of time and effort in the beginning to think about everything that needs to be considered in order to handle such diverse data while keeping infrastructure costs down. More data means more work for storage systems; managing massive amounts of real-time data from all over the world, whether it's from people or businesses, requires powerful computers with fast input/output speeds. Data analysis is further complicated by the fact that appropriate visual representations of such multi-dimensional data are sometimes lacking; in many instances, it may take more than two or three charts to derive meaningful conclusions from the data sets across visualizations. Because sensitive information is involved, building a secure cloud data platform for analytics must be a top priority. Appropriate controls, authentication mechanisms, and data encryption can greatly improve the security of data stored in the cloud.

METHODOLOGY

The use of cloud computing for big data management in library science is the central focus of the research. Here's a suggested methodology for the research:

Data Collection:

In the **Data Collection** phase, the primary objective is to gather both quantitative and qualitative data on library systems that have incorporated cloud computing to handle big data. This is achieved through surveys of libraries—both physical and digital—that utilize cloud services. Finding out how far along the path to cloud computing these libraries are is the primary goal of the survey.
Additionally, interviews with librarians and administrative staff provide insights into their experiences with cloud platforms. Data points analyzed include the volume and nature of big data managed by the libraries, specifically focusing on the number of e-books, journals, and research papers uploaded, the frequency and types of user access activities (such as downloads and search histories), and how metadata is organized and maintained.

Cloud Infrastructure Selection:

One of the most essential components of the Cloud Infrastructure Selection process is finding the right cloud service providers. Libraries often employ AWS, Google Cloud, and Microsoft Azure. This stage involves detailed research into the pricing models and service features of these providers. Factors such as the cost-effectiveness of the "pay-as-you-go" pricing structure, the scalability of the platform to accommodate increasing amounts of data, and the availability of features like data redundancy and backup to ensure data safety are key metrics used to evaluate the most appropriate cloud infrastructure for library needs.

Implementation of Big Data on Cloud Platforms:

In the Implementation of Big Data on Cloud Platforms, the objective is to demonstrate the real-world usage of cloud services for big data management. To achieve this, a simulation is conducted using a sample library dataset, which is uploaded to a selected cloud platform. During this process, various performance metrics are tracked, including the speed of data uploads, the time it takes to retrieve data, and how efficiently the system responds under peak usage conditions. This simulation allows for a clear assessment of how well cloud platforms handle the dynamic demands of library data.

Data Analysis:

The **Data Analysis** phase focuses on assessing the benefits and challenges that libraries experience when implementing cloud services for big data management. A comparison is made between cloud-based libraries and traditional digital libraries, with specific attention paid to factors such as cost-efficiency, ease of data management, and user satisfaction. Additionally, cloud platform analytics are used to monitor aspects like storage utilization, processing speed, and any incidents of data security breaches. This analysis helps in identifying the tangible improvements that cloud computing offers while also shedding light on any potential issues or limitations faced during implementation.

RESULTS AND DISCUSSION

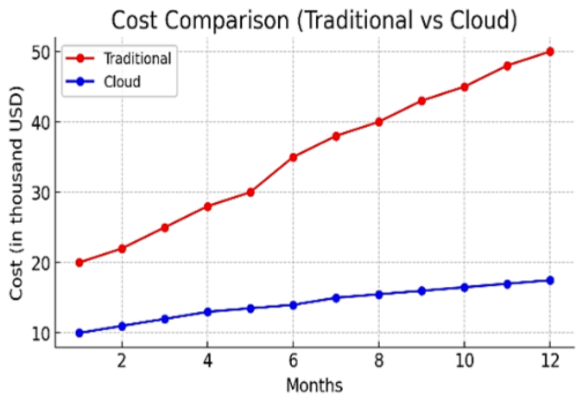

Fig 1: *Cost Comparison*

Cost Comparison: The graph in figure 1 shows that cloud infrastructure results in lower costs over time compared to traditional systems. While traditional infrastructure costs rise significantly, cloud costs remain relatively stable.

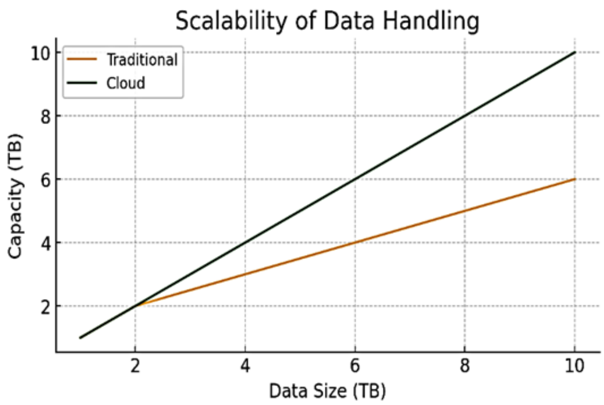

Fig 2: *Scalability of Data Handling*

Scalability of Data Handling: Cloud storage scales perfectly with increasing data size, while traditional systems face capacity limitations as the data grows shown in figure 2.

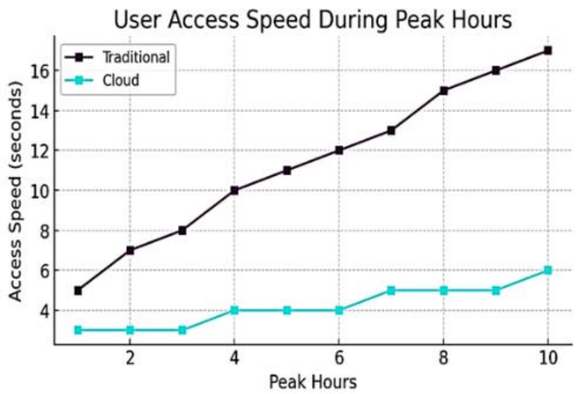

Fig 3: *User Access Speed*

User Access Speed: The cloud consistently provides faster access times during peak hours compared to traditional systems, where access speed drops as usage increases shown in figure 3.

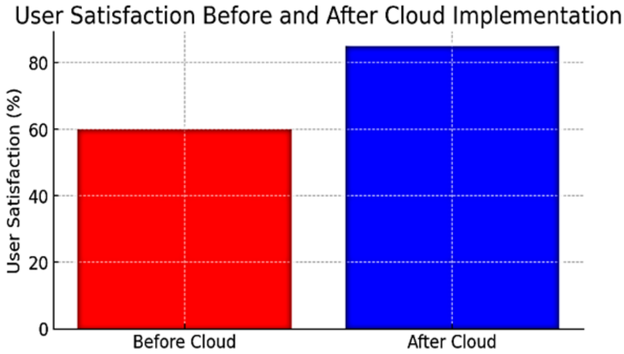

Fig 4: *User Access Speed*

User Satisfaction: There is a marked increase in user satisfaction after cloud implementation, rising from 60% to 85% shown in figutr 4.

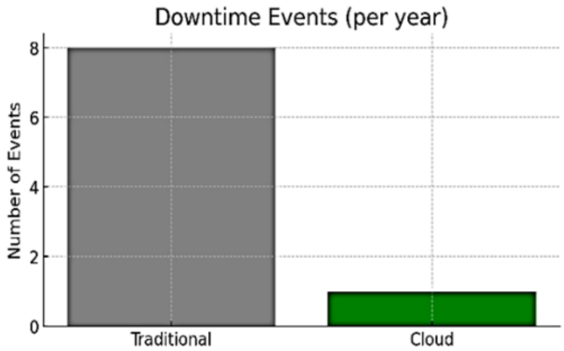

Fig 5: *User Satisfaction*

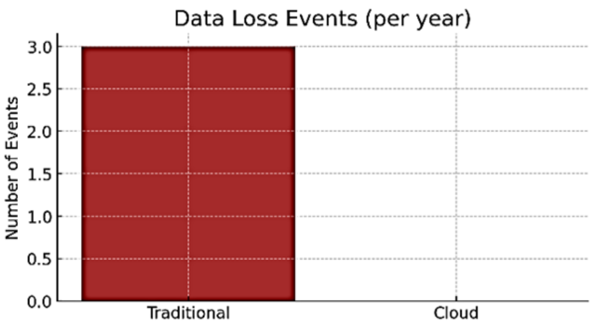

Fig 6: *Downtime and Data Loss Events*

Downtime and Data Loss Events: Libraries using cloud systems experience significantly fewer downtime and data loss events compared to traditional systems, indicating higher reliability are given in figure 5 and 6.

CONCLUSION

In today's rapidly changing environment, libraries remain essential as repositories of knowledge. However, due to increasingly busy schedules, many users are shifting from physical visits to virtual access of library resources. Cloud computing has emerged as a pivotal technology for libraries, offering numerous benefits such as large-scale data storage, backup capabilities, and protection from unforeseen events. Libraries can now store vast amounts of big data, including e-books, journals, and research papers, on cloud-based platforms, ensuring easy and secure access to this information. By adopting cloud technology, libraries provide users with prompt and seamless access to resources, while cloud service providers offer scalable solutions with cost-effective "pay-as-you-go" models, allowing libraries to handle large amounts of data more efficiently. In conclusion, the integration of cloud computing in library systems revolutionizes big data management by improving scalability, cost-efficiency, and reliability. Libraries using cloud platforms can better manage digital content, benefiting from faster access times, data redundancy, and enhanced security measures. This reduces downtime and data loss incidents, making cloud-based libraries far more resilient compared to traditional systems. Moreover, the ability to scale effortlessly as data needs grow, combined with the flexibility of cloud services, makes cloud computing an invaluable asset in modernizing library science and ensuring continued access to knowledge in the digital era.

REFERENCES

[1]. Sestino, A., Prete, M. I., Piper, L., & Guido, G. (2020). Internet of Things and Big Data as enablers for business digitalization strategies. Technovation, 98, 102173, ISSN 0166-4972. https://doi.org/10.1016/j.technovation.2020.102173

[2]. Srinivasan, A. Cloud computing. Pearson India ISBN: 9789332537439. Cloud computing: Concepts, technology & architecture. Prentice Hall Service Technology Series (1st ed.), ISBN10: 9780133387520

[3]. Kavis, J. Architecting the cloud: Design decisions for cloud computing service models (SaaS, PaaS, and IaaS) (1st ed.). Wiley, ISBN-10: 1118617614.

[4]. Alreshidi, E. (2019). Comparative review of well known cloud service providers. Science International (Lahore), 31(8), 65–170, ISSN-1013-5316.

[5]. Stankovic, J. A. (2016). Research directions for cyber physical systems in wireless and mobile healthcare. ACM Transactions on Cyber-Physical Systems, 1(1), 1–12.

[6]. Kune, R., Konugurthi, P. K., Agarwal, A., Chillarige, R. R., & Buyya, R. (2016). The anatomy of big data computing. Software: Practice and Experience, 46, 79–105. https://doi.org/10.1002/ spe.2374

[7]. Rizwan, A., Zoha, A., Zhang, R., et al. (2018). A review on the role of Nano communication in future healthcare systems: A big data analytics perspective. IEEE Access, 6, 41903–41920.

[8]. Khan, S., Shakil, K. A., & Alam, M. (2018). Cloud-based big data analytics—A survey of current research and future directions. ©Springer Nature Singapore Pte Ltd. Aggarwal, B., et al. (Eds.), Big data analytics. Advances in Intelligent Systems and Computing 654. https:// doi.org/10.1007/978-981-10-6620-7_57

[9]. https://www.gartner.com/. Accessed 31 Jan 2020.

[10]. Rajendran, V. V., & Swamynathan, S. Parameters for comparing cloud service providers: A comparative analysis. https://doi.org/10.1109/CESYS.2016.7889826, IEEE Xplore: 30 March 2017.

[11]. Dutta, P., & Dutta, P. (2019). Comparative study of cloud services offered by Amazon, Microsoft, and Google. International Journal of Trends in scientific Research and Development (IJTSRD), 3(3), 981–985.

[12]. Zhing, L. C. (2020). Cloud computing: State of Art and research challenges. Journal of International Services and Applications, 1(1), 7–18.

[13]. Practical Amazon EC2, SQS, Kinesis, and S3. eBook. SpringerDoi: https://doi.org/10.1007/ 978-1-4842-2841-8

[14]. Pradhananga, Y., Karande, S., & Karande, C. High performance analytics of big data with dynamic and optimized Hadoop cluster. IEEE. https://doi.org/10.1109/ ICACCCT.2016.7831733

[15]. Dawelbeit, O., & McCrindle, R. A novel cloud based elastic framework for big data preprocessing. In IEEE Conference Publications. https://doi.org/10.1109/CEEC.2014.6958549

[16]. Gonzales, J. U., & Krishnan, S. P. T. Building your next big thing with Google Cloud Platform. SpringerDOI: https://doi.org/10.1007/978-1-4842-1004-8

[17]. Singh, M. P., Hoque, M. A., & Tarkoma, S. A survey of systems for massive stream analytics, arXiv1605.09021v2

[18]. Ambeth Kumar, V. D., Ashok Kumar, V. D., Divakar, H., & Gokul, R. Cloud enabled media streaming using Amazon Web Services. IEEE. https://doi.org/10.1109/ICSTM.2017.8089150.

[19]. Subia, S. (2018). Data Storage SpringerDOI: 978-3-319-21569-3_7 10, Procedia Computer Science.

[20]. Nakhimovsky, A., & Myers, T. Google, Amazon, and beyond: Creating and consuming Web services. SpringerDOI: 9781590591314.

[21]. Mohanty, H., Bhuyan, P., & Chenthati, D. Chapter 2: Big data architecture. In Big data: A primer. Springer DOI: 9788132224938

[22]. Begam, S. S., Selvachandran, G., Ngan, T. T., & Sharma, R. (2020). Similarity measure of lattice ordered multi-fuzzy soft sets based on set theoretic approach and its application in decision making. Mathematics, 8, 1255.

[23]. Thanh, V., Rohit, S., Raghvendra, K., Le Hoang, S., Thai, P. B., Dieu, T. B., Ishaani, P., Manash, S., & Tuong, L. (2020). Crime rate detection using social media of different crime locations and Twitter part-of-speech tagger with Brown clustering. Journal of Intelligent & Fuzzy Systems, 38, 4287–4299.

[24]. The Old Bailey and OCR: Benchmarking AWS, Azure, and GCP with 180,000 Page Images DocEng '20: In Proceedings of the ACM Symposium on Document Engineering, September 2020. Article No.: 19, pp. 1–4. https://doi.org/10.1145/3395027.3419595

[25]. Ta, V.-D., Liu, C.-M., & Nkabinde, G. W. (2016). Big data stream computing in healthcare real-time analytics. In 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 37–42, https://doi.org/10.1109/ICCCBDA.2016.7529531