Advanced Strategies for Big Data Resource and Storage Optimization: An AI Perspective

Ayoub Sghir, Ayoub Allali, Najat Rafalia, Jaafar Abouchabaka
Department of Computer Science, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco

***Abstract*—The increasing use of advanced technologies with artificial intelligence in our daily lives has become an urgent necessity to facilitate tasks in a fluid and simple way, which leads to the generation of huge amounts of data. This data comes from various sources: media, social networks, connected objects., online transactions, and smart devices, among other sources. These data are generally organized into three categories: structured, unstructured, and semi-structured. This data is therefore called Big Data. This data is characterized by its enormous size and fast flow, as well as by the diversity of its sources. The importance of data lies in its ability to provide future perspectives and improve the decision-making process. To get the most out of this data, it must be stored and processed, but current technologies face many challenges and are often insufficient to cope with the huge amounts of data generated. It is necessary to look for advanced and highly efficient technologies, capable of storing the entirety of the data and processing it faster. We can also rely on artificial intelligence to help improve the use of storage and processing resources by compressing data or deleting excess data, thus saving storage space. This study discusses various approaches for optimizing Big Data processing, such as the use of AI compression techniques, the PSNR-SSIM method, and many others. The compression ratio for these algorithms is around 90%. With these technologies, it is possible to optimize the use of storage space, ensuring efficient and optimized management.**

***Keywords*—Artificial intelligence; big data; optimization; resource; storage**

I. Introduction

Big data has become the most widely used term in our digital world, particularly with the advent of the fifth generation and the use and reliance on machine learning and the Internet of Things. The Big Data industry is growing so fast that, according to a report by IDC (International Data Corporation), by 2025, around 30% of all data will be generated in real-time. According to the same entity, with around 150 billion connected devices worldwide in 2025, most of them will generate data in real-time, according to IDC. The global data domain is expected to grow from 23 zettabytes (ZB) in 2017 to 175 zettabytes (ZB) in 2025.

Today, numerous studies highlight the need to develop advanced and robust storage systems to effectively manage Big Data on a large scale [1]. Scalability considerations include reliability and availability as key technology design objectives for Big Data storage [2]. For this reason, these technologies have been designed to handle huge amounts of data, with their diverse sources and diversities, at high processing speeds.

Storage systems include Hadoop, NOSQL, Cloud Computing, NEWSQL and Data Lake. What sets these technologies apart is their ability to scale by adding more nodes to the cluster, as well as their ability to use a distributed file system across multiple machines for Hadoop. NoSQL databases also offer great flexibility for various data models (key-value, document, column, graph), enabling non-relational data to be managed with high performance. Cloud computing is also distinguished by its ability to adapt to demand, increasing storage capacity as needed, while offering rapid access to this massive data. Each technology has the capacity to store and process huge quantities of data.

Despite the power of these technologies, they also pose significant challenges, particularly with the growing sophistication of data flows. These challenges include the rapid growth of Big Data, the high cost of storage, real-time processing, data retention policies, and data duplication.

These challenges include the rapid growth of Big Data, the high storage cost, real-time processing, data retention policies and data duplication. All these challenges have led to the need to look for innovative solutions that can store and process all this data in real-time without consuming significant resources. Among these solutions are emerging technologies such as quantum storage, DNA data storage and holographic storage, a major revolution in the data storage industry, but at present, they are not the ideal solution, especially as most of them are still in the development phase. experimental framework and will not be accessible to everyone, as they will have high costs and complex management.

The best solution remains to work on developing and improving current technologies by enhancing the storage capacity of certain technologies to accelerate the pace of fast and immediate storage and processing, as well as integrating artificial intelligence into storage management frameworks for resource-constrained systems. This enables efficient storage utilization and improves system performance. AI-based solutions can be applied to improve resource utilization and system performance, in particular, their ability to learn and make decisions. The main objective of this study is to present techniques for optimizing Big Data.

This study proposes four techniques for optimizing storage, two of which are based on artificial intelligence (multi-agent systems and K-means clustering), batch processing, and other approaches based on statistical indicators. The objective of this project is to examine the effectiveness of these techniques on real data stored in MongoDB databases, focusing on four key aspects of the storage process: compression time, final size, access time, and CPU and RAM usage. Our findings highlight the strengths and weaknesses of each method and offer specific guidance for optimizing the efficiency, performance, and sustainability of storage in Big Data systems.

This study is organized as follows: In Section II, we provide an overview of the concept of Big Data along with the available technologies for its management and storage. Section III introduces some of the problems faced by data, especially at the storage level. Section IV describes our simulation methodology and the environment in which it operates. Section V details the role of AI for optimization. In Section VI, we present and discuss our simulation results. Finally, Section VII concludes this work and outlines future work possibilities.

II. Related Works

To improve dynamic storage management [7] this paper outlines different approaches for improving storage management to facilitate communication between AI agents in decentralized systems. It provides a multilayer communication environment that allows AI agents to dynamically identify communication needs to improve communication speed and efficiency. Decentralized storage systems are also compared in the paper and their performance is evaluated with regard to the efficiency of data storage, retrieval, and transfer between AI agents. This study aims to improve the scalability and interoperability of AI systems in decentralized environments by focusing on reducing communication delays.

In this paper [11], the authors present a model for improving data storage in smart grids, using an improved simulated annealing algorithm. This model aims to solve the problems of longitudinal and horizontal penetration between multi-level data centers in the information transmission networks of smart grids. A comparison of mathematical algorithms is also presented in the paper, based on the context of optimizing data storage in the Smart Grid. The aim is to evaluate the effectiveness of these algorithms in managing resources in a distributed computing system and to determine which algorithm offers superior performance in terms of task processing speed and efficient resource allocation.

In article [17], a method is examined for improving distributed storage using multi-agent systems. Its aim is to tackle the challenges of distributed storage, such as resource optimization and data management in a fragmented environment. To solve these problems, this method relies on multi-agent systems in which each agent manages a part of the overall task. The aim is to optimize storage efficiency while reducing costs and access times, minimizing obstacles, and improving overall system performance.

This article [21] presents a secure and efficient method for storing and managing data, using an intelligent optimization algorithm (IOA-DSAM). Images are divided into several parts, then encrypted using homomorphic encryption (HE). The improved fruit fly optimization algorithm (MFFO) improves encryption key generation. Once the shares have been received, they are decrypted and reconstructed. In this article, the author presents the various stages of the suggested model, such as share generation, the image encryption model and the more efficient key generation process. It also presents experimental results that compare the performance of the IOA-DSAM algorithm with other models. The results of the article offering better performance, thus proving its effectiveness for securing and managing images in a digital environment. The article concludes that this technique improves data security while offering optimized storage management.

In [25], a deep learning-based framework for storage management was proposed. This solution offers the ability to dynamically adjust key system parameters, such as cache size and queue depth. This method leads to performance improvements in terms of transfer rate, latency, and CPU utilization. However, this method focuses on optimizing system configurations without considering other crucial aspects such as compression, size, reads, and CPU/RAM. Furthermore, it does not address comparisons with other technologies and solutions. In the same context, in this article [13], particularly with regard to storage technologies, the author proposes a multi-agent approach to classify data as "hot" and "cold" in an HDFS cluster. He uses the classification to adapt redundancy and compression, with the aim of optimizing resource utilization. However, this method is confined to a specific Hadoop framework and does not systematically evaluate performance. Most research focuses on a single criterion, such as access time, system configuration, or a specific method, without adopting a comprehensive multi-criteria approach, leaving a gap in the literature on storage optimization. Furthermore, there is little research that systematically compares AI methods with conventional techniques in a practical application setting. Furthermore, validation often focuses on specific environments or technologies (such as HD file systems or system caches), which limits the ability to generalize. Our contribution offers a fair comparative study of four techniques, judged on four main axes, contributing to a more balanced and useful analysis of storage optimization.

III. Challenges in Big Data Storage

Data storage issues - The high cost of data storage and retention: Many companies have invested billions of dollars in data center development to build out the infrastructure elements of Big Data, due to the high cost of data storage and retention. In recent years, Amazon has invested $35 billion in AWS data centers in one region alone [14]. In the context of the Internet of Things, or with the fifth generation, the accumulation of information emitted by sensors generated by devices can quickly lead to data processing problems, especially as the available storage space is limited. This can lead to the loss of data or the deletion of readings, restricting data storage and monitoring capacity, and making data analysis extremely complex.

Data management and processing: The set of operations involved in metadata processing includes the collection, storage, cleansing, and management of huge quantities of data, whether in different forms or from different sources. Data ingestion into a data lake or streaming processing engine is one of the steps involved in Big Data processing. Although advanced data collection and processing mechanisms are available, there are certain obstacles, such as the lack of sufficiently advanced techniques or tools to extract and process data efficiently in real-time or close to real-time. There is still a lack of fast, efficient processing in current solutions. there is also a concern about data integration, as it is difficult to merge data from different sources and formats. Particularly given the complexity of processing algorithms, which require constant adjustments and improvements, especially when it comes to variations in formats and structures, resulting in prolonged integration times. It is also worth mentioning the importance of data consistency. A major problem during processing. Errors or inconsistencies can occur when updating data.

Data privacy and security: Numerous studies have explored the existence of numerous personal information breaches and data security vulnerabilities in Big Data environments. Hence the urgent need for real-time security measures. The study [16] highlights the complex nature of data security in the Big Data landscape while recognizing the need for innovative solutions that go beyond traditional security models. Delving deeper into data protection, it becomes clear that data security in Big Data is a multifaceted challenge that requires a deep understanding of the unique attributes of large datasets. As organizations continue to harness the power of Big Data, data control is a major challenge in addressing the security issues inherent in this transformative era [15]. Data protection is essential in the broader context of information management because information is sensitive. Rigorous and vigilant security measures are essential to prevent unauthorized access, data breaches, and misuse. This involves the application of specific information laws, which encourage the judicious use of data without touching or divulging personal or confidential information. Strong security systems and protocols are essential.

IV. Optimization Technique

Optimization is the process of making something, such as a system or process, more effective and efficient. By implementing strategies and techniques to optimize storage, we can reduce storage costs and improve data processing capabilities. Big Data storage management refers to the process of storing, structuring, and managing large quantities of data to facilitate retrieval and analysis. There are several techniques for managing and improving Big Data storage [9]. As illustrated in Fig. 1 below, here are the key methods for optimizing data storage in the context of Big Data:

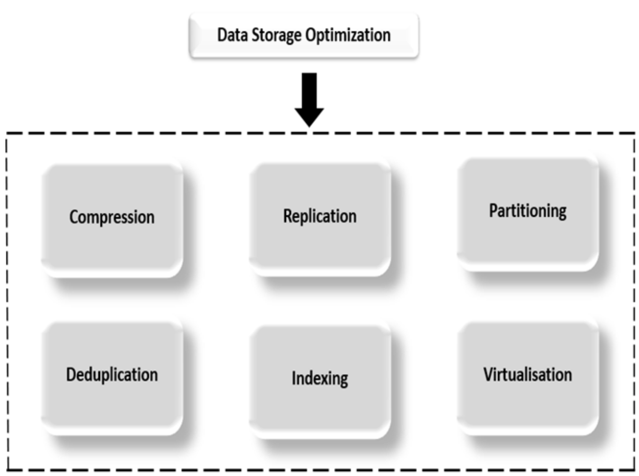

Fig. 1. Data storage optimization.

1) Compression: The compression process, among the most effective methods of optimizing data storage, is the use of coding to avoid data storage and transmission, which facilitates data transmission [18], restructuring, and other modifications to reduce the size of digital data files without altering their essential characteristics. The main aim of compression is to reduce the number of bits needed to code the data while reducing the memory required to store the data [19]. Compression can be divided into two categories: lossless and lossy. Lossless compression preserves the original data. While lossy compression reduces the original data size by compromising certain details. It permanently eliminates unnecessary bits. Although this only removes irrelevant information, it may result in the loss of some information due to compression. Lossy compression is often used for multimedia files such as audio, images, graphics, and video.

2) Replication: In a distributed environment, replication is frequently used to store multiple copies of the same data in different locations on multiple servers. When data is needed, it is retrieved from the nearest copy, thus avoiding delays and improving system performance [20]. Data replication offers the advantage of accelerating data access, reducing access times, and improving data availability. Replication is frequently used in distributed databases, cloud storage systems, and networks to improve consumer response speed and ensure service continuity.

3) Partitioning: A partitioning model is one of the most common methods used to optimize storage, and consists mainly of dividing a large dataset into smaller parts for easier handling, making it easier to manage large quantities of data and control their arrangement on disks. Each partition contains a part of the data, and these parts are usually spread over several servers or systems. These partitions are generally scattered over different database tables and contain their subset of data. It is possible to consider each partition as an independent database, but they are always part of the same logical database. Data partitioning can be achieved in several different ways. These include horizontal and vertical partitioning. The former is often used for the table mechanism. All partitions contain a separate part of the rows of a set of information. In a database, for example, records can be divided into partitions according to a specific range of values. In terms of vertical division, these partitions group together various subsets of columns. This is the opposite of horizontal division. All other types of partitioning are used to improve performance, whether for operations, aggregations or fast filters. In addition, they enable data to be distributed across several servers, enabling parallel processing and improving performance and utilization of computing resources.

4) Deduplication: Deduplication eliminates redundant data [22], to reduce the amount of data stored in a Big Data environment. Storage and storage optimization remain major challenges, especially with the exponential growth of data. After checking the data on the storage system, deduplication is used to increase and optimize storage space. As redundant data is written to storage, the deduplication process begins to be identified and eliminated. Duplicate data is identified, deleted, and replaced by a pointer to the first iteration of the data block. Deduplication also used the block verification technique, after dividing the data into blocks. A comparison is made between each block and other blocks already stored in the storage technologies to detect duplicates. With this method, users can separate specific workloads and quickly retrieve the updated backup, thus improving data storage efficiency.

5) Virtualization: Optimization through virtualization involves the creation of virtual structures for Big Data systems, such as servers, storage, or networks. Data virtualization can offer platforms that exploit all the information they collect to achieve different objectives. Data virtualization is a data management method that enables an application to retrieve and manipulate information without the need for technical details about the data, such as its source format or physical location [23]. About storage, it offers the possibility of adding more physical resources to a more flexible storage space, facilitating the dynamic localization of data. It improves efficiency by optimizing the use of resources and space. Virtualization makes storage and retrieval tasks simpler, further optimizing the storage infrastructure.

V. Role of AI for Optimization

While confronting the challenges of storage management methods in resource-constrained systems, there is an urgent need to look for more dynamic and fluid solutions. One such solution is artificial intelligence, which can store data intelligently, classify and label it, and retrieve it quickly and easily. The study establishes the fundamental concepts of Big Data and machine learning, highlighting their inter-relationships. It then examines the various ways in which Big Data improves machine learning algorithms, such as by improving data diversity, which promotes better generalization and reduces overlearning. In addition, Big Data uses artificial intelligence to obtain more precise information.

Artificial intelligence (AI) is a technology based on large databases and multiple algorithms, enabling the rapid discovery of new cultural systems and the creation of cognitive maps through data collection, data mining, and resource integration [6]. Simply put, AI aims to simulate human cognitive functions. AI represents a broad branch of computer science that seeks to create intelligent machines capable of performing tasks that usually require human intelligence [5]. Artificial intelligence plays a crucial role in storage and resource optimization. It enables machine-learning algorithms to reduce data duplication, remove redundancy, compress data, and load-balance to ensure optimal use of storage resources and dynamically compress information [7]. Analyzing historical data makes it possible to predict resource allocation and forecast future requirements. Artificial intelligence manages distributed storage environments, ensuring optimum use of resources. Finally, artificial intelligence simplifies data collection and resource allocation according to priorities and access patterns.

There are four technologies, illustrated in Fig. 2, which we will examine in this section.

1) Machine learning: is one of those technologies that has had a lot of success. To keep up with the expansion of the machine learning sector, some of the things that influence it, such as infrastructure and technical capabilities, also need to be increased [4]. Machine Learning is a branch of AI that enables machines to learn without the need for programming. Machine Learning is associated with Big Data, as it deals with data. Machine learning algorithms can be trained on datasets to spot patterns, relationships, and trends. In the field of storage management, machine learning algorithms can be trained on historical access patterns to predict future data requirements, identify the most efficient compression techniques for particular data types, and adjust real-time caching and retrieval strategies to suit system conditions [8].

2) Deep Learning (DL): is a branch of artificial intelligence that draws on the workings of the nervous system of living beings to solve more complex problems [24]. Deep learning algorithms process different types of information in a similar way to the way our neural networks react to the neural signals assigned to them. Deep learning is one of the recently recognized methods for storing and intelligently managing huge amounts of data. Thanks to deep learning, it operates on sophisticated models. Information can be reduced by identifying and eliminating repetitions, while preserving its reliability and quality.

In addition to optimizing the use of storage resources using deep learning algorithms, it is possible to optimize input/output (I/O) processing, thereby reducing access and disk access times. In addition, the use of neural networks to identify patterns and repetitions in data enables high-performance compression without compromising quality, thus optimizing the storage space used [25]. Deep learning reduces costs and optimizes the performance of Big Data management systems. This is particularly beneficial in a situation where the volume of data produced and collected continues to grow, requiring sophisticated, intelligent storage solutions.

3) Reinforcement learning: is an artificial intelligence technique that reproduces the human learning process to achieve its objectives. The learning paradigm used in RL is based on the ability of agents to make decisions by interacting with a context. The rewards or penalties the agent receives enable it to adapt its actions over time. This information-based method can be used to make dynamic storage choices in response to the changing requirements of Big Data systems. Several aspects of RL focus on optimizing Big Data management, including adaptive data positioning and real-time traffic management [3]. RL offers the possibility of dynamically adapting the choice of storage nodes based on available capacities and network conditions. This method is able to balance the load between servers and increase network throughput by adopting adaptive choices for each stage. RL is also used to improve the dynamic positioning of information in various storage strata, based on access patterns and the specific characteristics of the data, such as its temperature and input frequency, among other criteria [10]. The aim of this method is to optimize system performance by reducing processing time and resource expenditure.

4) A Multi-Agent System (MAS): is a distributed structure made up of various autonomous, intelligent software entities, referred to as agents. Agents interact with each other by transmitting information, and then collaborate with other agents to complete complex tasks. The decentralized approach of multi-agent systems (MAS) optimizes Big Data storage. Each agent, possessing learning skills, is able to make independent decisions regarding the management and processing of information. Coordination enables actors to optimize the use of storage resources by harmonizing workloads and dynamically modifying management policies in response to disparities between demand and resource availability.

As a key result, multi-agents facilitate the optimal distribution of data across multiple storage points, increase access speed and reduce throughput thanks to information compression and deduplication methods. Artificial intelligence enables these players to set up a decentralized structure and distributed storage [7-25] to exchange information efficiently with each other. This ensures immediate access to information. Agents are also able to adjust their procedures based on client and system performance, with the aim of selecting the most sophisticated technology for their underlying storage devices and ensure astute information management with the SMA monitor, which enables Data files to be accessed and classified as hot or cold [17]. Agents also have the ability to optimize storage system operation by optimizing replication and reacting to changes in data usage. This will reduce floor space and optimize resource utilization, while maintaining the right balance between efficiency and protection.

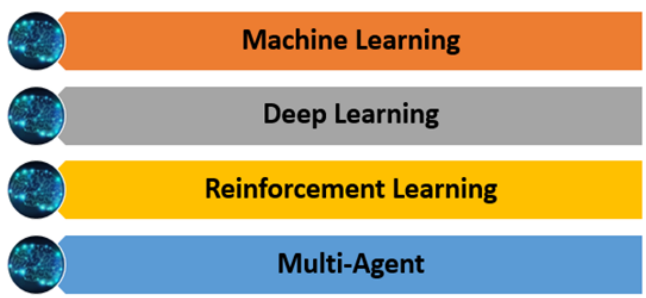

Fig. 2. AI-Driven techniques for big data storage optimization.

VI. Study of Methods and Results

There are four main stages in the Big Data management process, one of which is storage. One of the main challenges in Big Data is storage. They perform efficient, fluid storage of various forms of data while ensuring that fast access and real-time processing performance are maintained. It is essential to have an infrastructure and systems capable of storing and retrieving data rapidly without requiring considerable storage space. In this section, we look at the different optimization models used to improve Big Data storage, which achieved the best results based on four criteria (compressed size, compression time, read time, and average resources).

The data was chosen from the group of plant images, with a size of 900 MB. This is a specific unstructured data, which is considered as a type of Big Data. NoSQL databases were selected to store the data. MongoDB, a technology that manages data in the form of documents, offers a perfect solution for storing image-type data. It provides a distributed file system specially designed for MongoDB, called GridFS, specifically designed to store and retrieve image files. The storage optimization process is divided into three stages, as shown in Fig. 3:

When selecting the optimization algorithm, four algorithms are available. Two of these are based on the intelligent approach, namely multi-agent and K-means. The other algorithm is based on compression using batch processing and the PSNR-SSIM method. Phases 2 and 3 involve implementing the principles of the algorithms to improve storage in MongoDB technology.

The first method, entitled Smart Optimization and Batch Processing, is an algorithm that uses hybrid JPEG compression first, followed by gzip and batch processing. Its aim is to

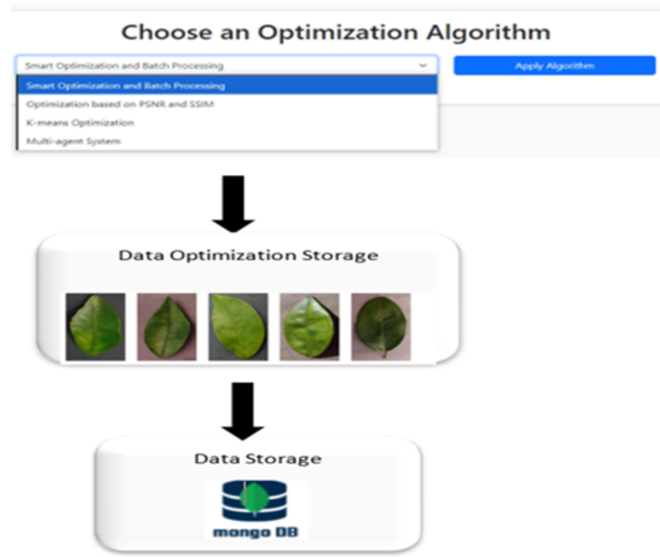

Fig. 3. Process architecture storage optimization.

divide images into several groups rather than processing them individually or all at once, thus effectively reducing image size and managing large numbers of images without overloading memory.

The second algorithms also use compression, but this time the PSNR-SSIM metric is used. As we know, pixels are small squares that represent the image. Digital image analysis is the specialty of uniform-density image processing. The image pixel matrix is two-dimensional. The numerical values of these pixels vary between 0 and 255. Color images are encoded on 3 RGB channels (red, green, and blue), with an integer ranging from 0 to 255 for each channel. To assess the quality of an image after compression, the PSNR-SSIM principle combines two metrics: the first PSNR measure evaluates the signal-to-noise ratio. It is measured in decibels (dB). The quality of the compressed image is improved as the PSNR value increases, while SSIM evaluates the structural conformity between the original image and the compressed image. Often, an SSIM greater than 0.85 is perceived as satisfactory quality for visual images.

Each time images are compressed; the process starts by calculating the PSNR and SSIM values between the original and compressed images, as soon as PSNR exceeds 30 dB and SSIM exceeds 0.85. The process is complete when all 70,000 data images have been processed. There are also two intelligence-based methods: unsupervised K-means and multi-agent.

The third K-means method is based on minimizing the number of colors present in the image and allows datasets to be grouped into K distinct clusters, based on a color grouping based on the specified number of clusters. Each group corresponds to a group of similar images. The K-means algorithm converts images into a single dimension and then groups them into K clusters based on the specified number of clusters. Each cluster corresponds to a group of similar images, and the distance between each image and the center of its cluster is minimized. Clustering aims to group similar images together for optimal processing and compression. Once the images have been grouped, each group is processed separately.

The multi-agent approach is the fourth method. This method implements autonomous agents that work together to perform tasks related to data compression and optimization management in a big data context. We employed three agents. Each agent has his or her own responsibilities. The first agent is responsible for applying random compression levels to images that are run on a regular basis. When activated, this agent compresses an image and sends the compressed file to the second agent. In this way, compressed files are received and saved in MongoDB GridFS. The role of the third agent is to supervise and monitor system performance. All these autonomous agents are defined to perform specific tasks in a coordinated way at different compression levels to reduce image size.

The four suggested algorithms were grouped and presented as diagrams. During the optimization process, the performance of the algorithms was evaluated in terms of size, compression time, read time, RAM usage, and CPU consumption.

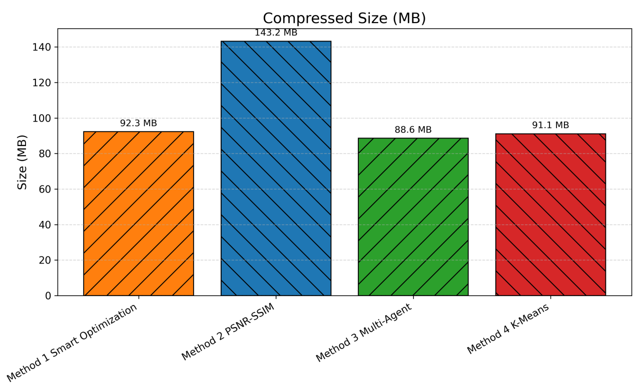

Fig. 4. Compressed size for all 4 methods.

Following the optimization of the four techniques, the results are illustrated in Fig. 4. This figure summarizes the graphs corresponding to the four techniques: Each instance produced a file size significantly smaller than that of the initial file. Technique 2 produced a 140 MB document, with a compression ratio of approximately 84.44%. This indicates that compression reduced the data size by 84.44%. The other methods, SMA and K-means, performed best (88.6 MB and 91.6 MB), achieving a compression rate of approximately 91%. In contrast, PSNR-SSIM produced a larger compression size (143.2 MB), thus reducing its effectiveness in contexts where storage space is critical. The results of the four methods indicate highly effective compression, with a higher compression index. Thus, our multi-criteria study indicates that examining compressed size alone is not sufficient to judge a method; it must be combined with other parameters such as the time required for compression, the time required for reading, and resource utilization in order to make a rigorous choice in the context of Big Data.

The time required to compress the data was considered as a second criterion, and the image data was correctly compressed and stored by the storage system. However, according to Fig. 5, the two artificial intelligence algorithms achieved a

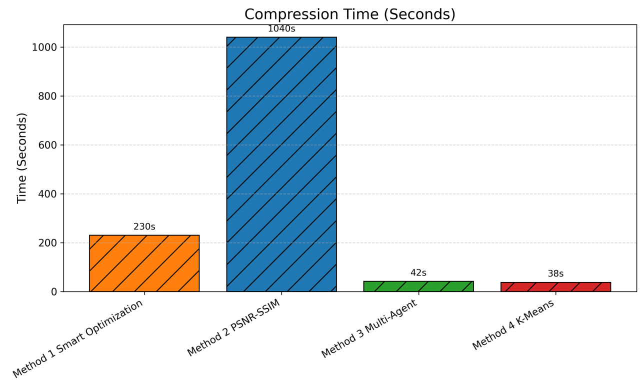

Fig. 5. Compression time for all 4 methods.

compression time of 31 and 26 seconds, respectively, for multi-agent method 3 and K-means method 4. It is observed that the AI-based methods require little time to compress the data, while the other methods require more time. Methods 3 and 4 demonstrate considerably higher processing efficiency.

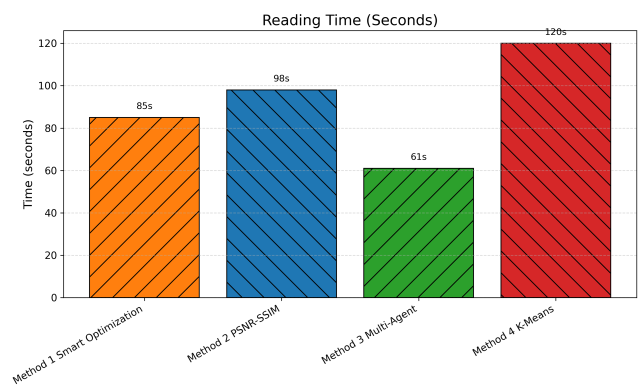

Fig. 6. Reading time for all methods.

The third criterion evaluated in this study was data reading time (see Fig. 6). For all four methods, reading time ranked well against traditional methods. However, there is one method that scored well for data reading, namely the multi-agent method. This method required just 60 seconds to read compressed data from MongoDB (GridFS file management system). This compares favorably with other techniques, which are more time-consuming to varying degrees.

The execution time analysis reveals that the SMA technique is fast, making it a suitable option for contexts, where regular access to data is essential. The Smart-Optimization technique, which takes 85 seconds, offers a satisfactory balance between performance and flexibility. However, although the K-means method offers remarkable compression performance, it suffers from a major drawback in terms of read time, which is 120 seconds. The PSNR-SSIM technique, with an average time of 98 seconds, is average but less efficient than the SMA method.

In Fig. 7, resource performance was evaluated and two comparisons were made for the four methods. The illustration compares CPU and RAM usage for four optimization techniques. Intelligent optimization and PSNR-SSIM require

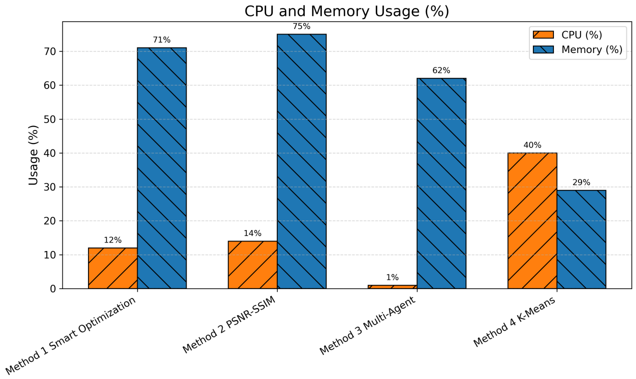

Fig. 7. CPU and RAM usage for all methods.

a lot of memory ($\tilde{7}$0%), but little CPU ($\tilde{1}$4%). Multi-Agent uses only 0.4% of the CPU while utilizing 62% of the RAM, demonstrating optimal task allocation. On the other hand, K-means uses a powerful CPU ($\tilde{4}$0%) while consuming little memory (29%). According to these results, multi-agent is efficient for handling low CPU loads, while K-Means is better suited to computationally-intensive systems.The analytical findings indicate that the selection of the compression method is based not only on performance criteria such as time, size, and readability, but also on the available hardware configuration. This is essential for a multi-criteria evaluation that takes resource constraints into account.

This study aims to highlight methods that can improve storage efficiency in the context of Big Data, which were the subject of this study because of the results obtained. All four algorithms showed high performance in comparison with the identified criteria (compressed size - compression time - read time - average CPU, RAM), although AI-based algorithms have an advantage. Other compression algorithms such as RLE (Run-Length Encoding) and Huffman were used to test these techniques, as were the mathematical algorithms such as PSO and Simulated Annealing. However, the results obtained did not match the algorithms presented in this report, particularly about compressed size and compression ratio, which were high, and temporal compression, which was higher.

Our multi-criteria evaluation offers added value by demonstrating that examining a single indicator (such as compression time) is not sufficient to judge the suitability of a storage method. Thus, this research helps to inform practical decisions regarding the storage and management of Big Data.

VII. Conclusion

One of the biggest challenges on a global scale is how to handle the constant flow of big data. This study discusses some of the key methods used to improve data storage, such as compression, duplication, deduplication, segmentation, and virtualization, with particular consideration given to the challenges associated with contemporary storage technologies. These methods form the basis for contemporary solutions, but they remain insufficient without the incorporation of artificial intelligence and its performance and resource requirements. To overcome these constraints, we have suggested a comparative analysis of four approaches designed to address the challenges associated with data storage and management. All of these methods are based on the principle of storage optimization, which consists of maximizing capacity through algorithms designed to improve the use of storage resources. We propose two techniques based on artificial intelligence (multi-agent systems and K-means clustering), as well as two other approaches, PSNR-SSIM and smart storage optimization. These techniques were implemented on a real dataset consisting of images. The performance of the four algorithms compared to established benchmarks was high, although the algorithms based on artificial intelligence had a slight advantage. These results have several implications. Scientifically speaking, these algorithms are all based on the principle of storage optimization, which aims to improve capacity through algorithms designed to maximize the use of storage resources. The application of storage optimization methods in image databases has proven effective in terms of storage size and processing speed. Once decompressed, the data retains its quality without any loss of information, highlighting the power and efficiency of the algorithms. This research could reveal promising new prospects. Future advances include cutting-edge intelligent algorithms and the incorporation of other innovative methods such as deep learning, reinforcement learning, and quantum optimization. These algorithms can handle various types of data of varying sizes, thanks to a better understanding of information and how it is used. They also facilitate informed decision-making regarding data management and storage, compression, and rapid insertion in an optimization context. This improves the efficiency of existing systems while reducing storage costs, particularly when integrating an intelligent structure into an intelligent environment to develop and optimize storage in the era of big data.

Acknowledgment

This work is supported by National Center for Scientific and Technical Research (CNRST), Rabat, Morocco

References

[1] L. Theodorakopoulos, A. Theodoropoulou, Y. Stamatiou, “A State-of-the-Art Review in Big Data Management Engineering: Real-Life Case Studies, Challenges, and Future Research Directions,” eng, Vol. 3, Issue. 3, doi: 10.3390/eng5030068, 2024.

[2] A. Siddiqa, A. Gani, and A. Karim, “Big data storage technologies: a survey,” Frontiers of Information Technology & Electronic Engineering, vol. 18, pp. 1040–1070, 2017. doi: 10.1631/fitee.1500441

[3] W. Yang, Y. Qin, and Z. Yang, “A reinforcement learning based data storage and traffic management in information-centric data center networks,” Mobile Networks and Applications, vol. 27, 2022. doi: 10.1007/s11036-020-01629-w

[4] J. P. Bharadiya, “A comparative study of business intelligence and artificial intelligence with big data analytics,” American Journal of Artificial Intelligence, vol. 7, no. 1, pp. 24–30, 2023. doi: 10.11648/j.ajai.20230701.14

[5] A. Charalambous and N. Dodlek, “Big data, machine learning, and artificial intelligence to advance cancer care: opportunities and challenges,” Seminars in Oncology Nursing, vol. 39, 2023. doi: 10.1016/j.soncn.2023.151429

[6] J. Zhang, “Multiple roles of big data platform and artificial intelligence architecture,” Applied Mathematics, Modeling and Computer Simulation, pp. 985–963, 2023. doi: 10.3233/atde231039

[7] A. Tara, H. K. Turesson, and N. Natea, “Dynamic storage optimization for communication between ai agents,” Future Internet, vol. 16, 2024. doi: 10.3390/fi16080274

[8] S. B. Mohammed, B. Krothapalli, and C. Althati, "Advanced techniques for storage optimization in resource-constrained systems using ai and machine learning," Journal of Science & Technology, vol. 4, no. 1, pp. 89–124, 2023.

[9] P. H. Ahmad and M. Rai, "Analysis of optimization strategies for big data storage management: a study," in Proc. 2023 4th Int. Conf. on Electronics and Sustainable Communication Systems, 2023. doi: 10.1109/icesc57686.2023.10193738

[10] T. Zhang *et al.*, "Data management of scientific applications in a reinforcement learning-based hierarchical storage system," Expert Systems With Applications, vol. 237, pt. B, Mar. 2024. doi: 10.1016/j.eswa.2023.121443

[11] Q. Wang, D. Yu, J. Zhou, and C. Jin, "Data storage optimization model based on improved simulated annealing algorithm," Sustainability, vol. 15, 2023. doi: 10.3390/su15097388

[12] D. Sawitri, "Big data challenges and opportunities in the development of digital technology," Jurnal Info Sains: Informatika dan Sains, vol. 14, no. 2, pp. 215–224, 2024. doi: 10.54209/infosains.v14i02.4619

[13] M. Sais, N. Rafalia, R. Mahdaoui, J. Abouchabaka, "Distributed storage optimization using multi-agent systems in Hadoop", E3S Web of Conferences, doi: 10.1051/e3sconf/202341201091, 2023.

[14] A. S. Pothukuchi, L. V. Kota, and V. Mallikarjunaradhya, "A critical analysis of the challenges and opportunities to optimize storage costs for big data in the cloud," Asian Journal of Multidisciplinary Research & Review, vol. 3, no. 1, pp. 132–144, 2022.

[15] A. O. Adenubi, A. P. Oduroye, and A. Akanni, "Data security in big data: challenges, strategies, and future trends," Int. J. of Research in Education Humanities and Commerce, 2024. doi: 10.37602/ijrehc.2024.5201

[16] L. Wang and Y. Zhang, "Access controls in big data environments," Journal of Information Security, vol. 16, no. 3, pp. 112–128, 2019.

[17] M. Sais, N. Rafalia, R. Mahdaoui, and J. Abouchabaka, "Distributed storage optimization using multi-agent systems in Hadoop," E3S Web of Conferences, vol. 412, 2023. doi: 10.1051/e3sconf/202341201091

[18] L. A. Fitriya, T. W. Purboyo, and A. L. Prasasti, "A review of data compression techniques," International Journal of Applied Engineering Research, vol. 12, no. 19, pp. 8956–89633, 2017.

[19] R. V. Bidwe, S. Mishra, S. Patil, K. Shaw, D. R. Vora, K. Kotecha, and B. Zope, "Deep learning approaches for video compression: A bibliometric analysis," Big Data Cognitive Computing, vol. 6, 2022. doi: 10.3390/bdcc6020044

[20] A. Govardhan and R. Dugyani, "Survey on data replication in cloud systems," Web Intelligence, vol. 22, pp. 83–109, 2024. doi: 10.3233/web-230087

[21] S. Ayoub *et al.*, "Intelligent optimization algorithm based data storage and management: an empirical study," in Proc. 2023 2nd Int. Conf. on Electronics and Renewable Systems (ICEARS), 2023. doi: 10.1109/icears56392.2023.10085038

[22] N. Sharma, A. V. K. Prasad, and V. Kakulapati, "Data deduplication techniques for big data storage systems," Int. J. of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 10, pp. 1145–1150, Aug. 2019. doi: 10.35940/ijitee.j9129.0881019

[23] A. Gulamraza, "Big data virtualization," Int. Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 4, pp. 5104–5107, 2023.

[24] B. Zohuri and F. M. Rahmani, "Artificial intelligence driven resiliency with machine learning and deep learning components," Journal of Communication and Computer, 2019. doi: 10.17265/1548-7709/2019.01.001

[25] Ch. Cheng, Ch. Zhou, Y. Zhao, J. Cao, "Dynamic Optimization of Storage Systems Using Reinforcement Learning Techniques, "arXiv, 2024.