



EXTENSIBLE STORAGE MANAGEMENT AND DATA AVAILABILITY TECHNIQUES

Anozie, E. L.¹, Okoronkwo, M. C.², Emesiani, J. U.³, Eze, R. C.⁴

Department of Computer Science, State University of Medical and Applied Sciences, Igbo-Eno, Enugu State, Nigeria.¹

Department of Computer Science, University of Nigeria, Nsukka, Enugu State, Nigeria²

Department of Computer Science, University of Nigeria, Nsukka, Enugu State, Nigeria³

Department of Electrical and Electronic Engineering, State University of Medical and Applied Sciences, Igbo-Eno, Enugu State, Nigeria.⁴

Abstract: Storage management is a critical component of modern information technology, encompassing a wide range of techniques and strategies to efficiently store, organize and retrieve data. With the high growth of digital information in recent years, effective storage management has become paramount for businesses and individuals. Effective storage management improves data availability by ensuring the data is properly structured, stored, and accessible, even during periods of high load or during system maintenance. Extensible storage management and data availability techniques encompass a broad range of strategies and technologies designed to manage data storage systems dynamically and ensure data is available whenever needed. In this work, we looked at the techniques, current trends and applications of storage management, benefits and challenges of storage management. We also highlighted the components and challenges of data availability, best practices and tools to ensure data availability and the impact of data availability on business operations.

Keywords: Storage Management, Data Availability, Database Management System

I. INTRODUCTION

With the ever-increasing reliance on data-driven technologies, including artificial intelligence, machine learning, and data analytics, the need for uninterrupted access to data has become paramount. Organizations rely heavily on data to make informed decisions, drive innovation, identify trends and patterns, personalize user experiences, and optimize business processes and gain a competitive edge [1]. This state of guaranteed access to data is known as data availability. Without data availability, businesses would face significant challenges in accessing the necessary information at crucial times. Therefore, organizations strive to establish robust systems and processes that ensure data availability at all times [2]. One of the key factors that contribute to data availability is data storage [1].

Data storage management has become increasingly central to the success of any enterprise as remote work, cloud services, and cyber security concerns have come to dominate the technological environment. Storage management is responsible for the efficient and reliable storage of data. Storage management and data availability go hand in hand. They are crucial for maintaining optimal performance and reliability. Effective storage management improves data availability by ensuring the data is properly structured, stored, and accessible, even during periods of high load or during system maintenance.

II. EXTENSIBLE STORAGE MANAGEMENT

Extensible storage management refers to the ability to dynamically extend and manage storage resources to meet evolving data demands. It involves scaling storage capacities, integrating diverse storage systems, and automating management tasks. Storage management encompasses all the processes and technologies used to manage the resources, software, and hardware that make up a data storage system [3]. It enables users to optimize the use of storage devices and to protect the integrity of data for any media on which it resides [4]. Storage management is the way organizations deliver fast access to data while ensuring data integrity, policy and regulatory compliance, and efficient storage resource use. It involves developing a plan to provision, configure, back up, and monitor data storage infrastructure to maintain performance, prevent data loss, and quickly address problems [5]. It entails the processes and technologies organizations use to maintain a data storage system. It enables them to expand their storage capacity without major overhauls, ensuring that data remains accessible and manageable. In database management system (DBMS), storage management refers to the techniques used to store and retrieve data efficiently. The purpose of storage management is to ensure that the storage



resources are used efficiently, and to provide reliable and secure access to data and also to help organizations find a balance between costs, performance and storage capacity [6]. It tries to improve and maximize the efficiency of data storage resources [7].

Techniques of Storage Management

Some common storage management techniques in a DBMS include:

- **Indexing:** This is creating indexes on database tables to improve the speed of data retrieval. These indexes are specialized data structures that store a subset of the data and provide quick access to specific records
- **Clustering:** This is organizing data on disk according to the relationships between data records to minimize disk I/O operations.
- **Data Partitioning:** It means dividing large data into smaller and more manageable chunks to improve performance.
- **Transaction logging:** It is the recording of all database transactions in a log file to ensure data integrity.
- **Provisioning:** This is assigning storage capacity by analyzing current capabilities, such as storage on physical drives or the cloud, and deciding the proper information to store in each location.
- **Virtualization:** This is leveraging virtual machines (VMs) to decouple software from hardware operating environments.
- **Data compression:** This is the act of reducing the size of data sets without compromising them. Compressing data allows users to free up storage space, improve file transfer speeds. Data compression works by either removing unnecessary bits of information or redundancies within data.
- **Data migration:** This method entails moving data from one location to another. This can include the physical location, such as from one hard drive to another or the application that uses the data. Data migration is often necessary when introducing new hardware or software components into an organization. Important factors to consider while implementing data migration include ensuring network bandwidth, effective transfer speeds, data integrity and ample storage space for the new location throughout the transfer.
- **Data replication:** It involves storing the same data in multiple storage locations for redundancy, resilience, and reliability. There are two types of data replication: synchronous and asynchronous. Synchronous data replication is when companies copy any changes to an original data set in the replicated data set. Asynchronous replication only occurs when a professional enters a command into the database, so it is not an automatic process. With this type, a company has more control over the resources used to replicate data but may not possess real-time data backups.
- **Automation:** It is the process of having tools automatically manage your data. Rather than updating your data manually, you can use software tools to accomplish the task. This would ensure that the database contains updated information for all users and prevents users from viewing outdated information if a user forgets to submit changes [6][3].
- **Backup and recovery:** Regularly backing up data to protect against data loss due to hardware failure, user error, or natural disasters.
- **Data deduplication:** It involves removing duplicate data to reduce storage requirements and improve performance.
- **Data archiving:** This is storing older and less-frequently accessed data on slower, less expensive storage devices, to free up faster storage for more critical data.

Current Trends in Storage Management

- **DNA data storage:** DNA storage is a technology that uses DNA molecules to store and retrieve large amounts of data, using the unique structure and properties of DNA to encode and decode digital information, and providing a highly durable, compact and secure solution for long term data storage, with potential applications in fields such as genomics, archiving and data preservation. The benefits of DNA data storage are increased capacity, life span and sustainability [9].
- **AI storage management:** Artificial intelligence for IT operations (AIOPs) enable monitoring, diagnostics, predictive analysis and prescriptive functions for storage infrastructure and applications. AIOPs can tell an organization what is happening with storage and why, what could happen and what to do about it. AIOPs for storage uses machine learning to collect and analyze telemetry data. This information is turned into predictive analytics. AIOPs include automation, performance management and service management and automate many of the decisions involved in scaling and securing storage systems. AI-driven storage management removes much complexities and manual tasks of traditional storage resource management. AI in storage is a potential tool for detecting ransomware or other attacks early enough to take action that mitigates or eliminates damages [10].
- **Object-based storage:** This is a storage architecture that treats data as objects rather than files. It is used to archive, backup or retain large amount of unstructured data in the cloud or on-premises. It is scalable, flexible and cost-effective [11].

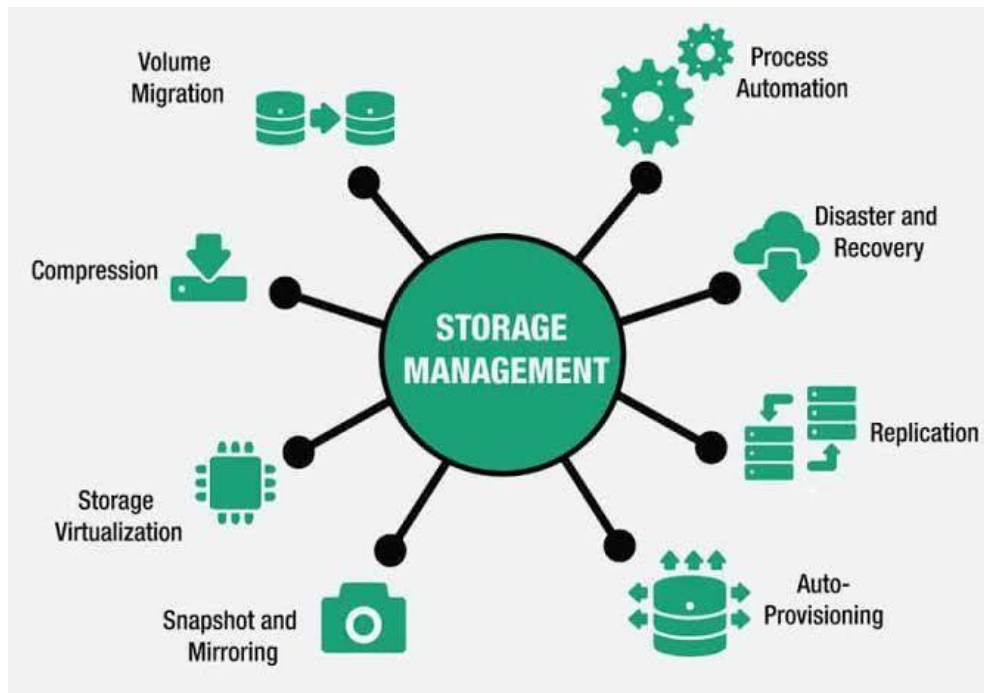


Fig. 1: Techniques of Storage Management [7]

- **Software-Defined Storage (SDS):** This enables organizations to manage storage on commodity x86 servers and it is more cost-effective and easier to manage than monolithic storage area network (SAN) and network-attached storage (NAS) hardware. SDS disaggregates storage from on-premises hardware, such as SSDs on a cluster of x86 servers, and combines them into a unified pool of storage that can be centrally managed and provisioned through a single management tool. SDS enables better storage utilization, scalability and flexibility. One of the approaches to SDS is the use of hyper converged infrastructure (HCI) software. HCI is an integrated software stack that allows IT administrators to manage server, storage, networking and virtualization through a single interface [12].
- **Edge Computing:** This involves adopting edge data center, which are small or micro data centers located at the edge of the network and closer to users and devices. Having computing power and storage at the edge enables organizations to process, analyze and store data where it is generated, thereby improving efficiency. Edge computing improves performance, reduces latency and lowers bandwidth costs, as organizations do not have to move the data off-site to central data centers or the public cloud to process it [12].
- **Multicloud and Hybrid cloud storage:** This refers to the practice of using services and resources from multiple heterogeneous cloud service providers, including Amazon web services, Google cloud platform or Microsoft Azure, as well as specialized platform-as-a-service, infrastructure-as-a-service or software-as-a-service providers in a way that integrates them in a unified ecosystem. For example, in a multicloud architecture, an organization could harness the computational agility of AWS for its data analytics, leverage Microsoft Azure for its suite of productivity tools and tap into the Google Cloud Platform to power its machine learning capabilities. Hybrid cloud blends the use of private cloud environment with public cloud environments. Businesses leverage multi-platform, multicloud and hybrid cloud architectures to get the best [13].

Applications of Storage Management

- **Financial services:** Banks and financial institutions use storage management techniques to store and manage financial data such as transaction records and customer information. It also helps them to comply with regulatory requirements, protect sensitive data and improve operational efficiency.
- **Healthcare:** Hospitals and clinics use storage management to store and manage electronic health records and medical images like X-rays, CT scans and MRIs and helps healthcare providers to access patient data quickly and securely.
- **E-commerce and Retail:** Retailers use storage management to store and manage inventory data, customer data and sales data. It helps retailers to optimize their inventory level, personalize customer experiences and improve supply chain and marketing strategies.
- **Entertainment:** In media and entertainment industry, storage management is key in managing and distributing large volumes of video and audio content, ensuring seamless delivery to customers.



- **Education:** Educational institutions like universities or colleges rely on storage management to manage and store students' data, course registration, results, course materials and research data. It helps institutions to improve students' outcome and facilitate collaboration.
- **Scientific Research:** Scientific research generates massive amounts of data, such as genomic data, simulations and sensor data which require effective storage management.
- **Human resources:** Records about workers, wages, payroll, deductions, salaries and other information are managed using storage management techniques.
- **Industry:** Manufacturing companies use storage management to store and manage production data, quality control and supply chain data. It helps manufacturers to optimize their production processes, improve product quality and reduce costs.
- **Internet of Things (IoT):** IoT devices generate vast amounts of data and storage management is essential for collecting, analyzing and storing this data for various applications.
- **Agriculture:** Storage management techniques are used to manage the records of agricultural products like crops, machinery, fertilizers and livestock.
- **Telecommunications:** Telecommunication companies use storage management techniques to store customers' information, call details and monthly post-paid bills.
- **Content Management Systems (CMS):** CMSs like Wordpress rely on storage management to store and retrieve text, images, videos and other contents for websites and web applications.
- **Data Warehousing:** Storage management is essential in data warehousing systems, where large amounts of data are collected, stored and organized for business intelligence.
- **Airlines:** Airlines use storage management techniques to keep records of flight bookings and scheduling information.

Benefits of Storage Management

- **Enhances performance:** Storage management improves the performance of the existing storage resources. For example, compressing data reduces the amount of storage and improves file transfer speeds.
- **Improves security:** With a storage management plan, organizations can ensure they store sensitive data in proper locations and protect it against external parties. In addition, storage management usually includes developing a data recovery plan, which protects data in the event of disaster, such as a computer failing or a fire within an office.
- **Promotes adaptability:** Storage management practices include regularly analyzing storage needs and ensuring you have the proper equipment and tools to effectively store data. This makes organizations more adaptable to changes, as their storage capabilities grow in accordance with their needs.
- **Access consistency:** As storage is better managed, user access can remain consistent despite fluctuations in demand. That is because with a good storage management program, as data volumes rise, the system in place knows where and how to store the incoming data, sending it out of the traffic flow just as demand surges.
- **Cost-savings:** By centralizing storage management, storage administrators can use resources more efficiently, responding with capacity on as needed basis rather than planning for and purchasing capacity in advance of a spike in demand.
- **Easier administration:** With a storage management program in place, the centralization helps administrators keep track of multiple storage systems, leading to better resource utilization.
- **Higher availability:** Storage techniques help to backup and restore data fast leading to higher availability and reliability on data.
- **Speed and Flexibility:** Storage management solution work in real-time and adapt to sudden changes in the storage resources. It also helps improve flexibility and reduces wasted storage [7][6][5].

Challenges of Storage Management

- **Complexity of storage management:** Storage management can be complex, especially as the size of the storage environment grows.
- **Limited physical storage capacity:** Operating systems can only manage the physical storage space that is available, and as such, there is a limit to how much data can be stored.
- **Aligning storage with overall IT infrastructure:** If storage deployments are not carefully planned and allocated, applications and storage types may not line up for optimal performance.
- **Performance degradation with increased storage utilization:** As more data is stored, the system's performance can decrease due to increased disk access time, fragmentation.
- Backup and recovery of data can also be challenging, especially if the data is stored on multiple systems or devices [5][4].



III. DATA AVAILABILITY

Data availability is the process of ensuring that data is available to end users and applications, when and where they need it. It defines the degree or extent to which data is readily usable along with the necessary IT and management procedures, tools and technologies required to enable, manage and continue to make data available [14]. Data availability can also be defined as the ability to access and utilize data promptly and efficiently when and where it is required [15]. It enables organization to ensure that all of its business-related data are available to the organization, partners, or end-users at any time of the day, whenever and wherever required [16]. Data availability calls for implementing products, services, policies and procedures that ensure that data is available in normal and even in disaster recovery operations [14]. Ensuring data availability is very important to ensure successful day-to-day operations [16].

Components of Data Availability

To ensure efficient data availability, several key components must be considered in designing data systems.

- **Data storage:** It encompasses the physical or virtual means used to store data, such as databases, cloud storage, or distributed file systems.
- **Data replication:** It involves creating duplicate copies of data across different nodes or locations, ensuring redundancy and mitigating the risks of data loss or unavailability.
- **Data synchronization:** It is the process of maintaining consistency between multiple copies of data. It ensures that changes made to one copy are propagated to all other copies, guaranteeing that users accessing different replicas see the same version of the data.
- **Data integration:** It refers to the process of combining data from different sources and making it available in a unified format. It plays a crucial role in enabling organizations to access and analyze diverse datasets seamlessly.
- **Data recovery:** It involves having robust backup and disaster recovery mechanisms in place. This ensures that even in the event of hardware failures, natural disasters, or cyber-attacks, data can be recovered promptly to minimize downtime and maintain data availability.
- **Data security:** It involves implementing measures to protect data from unauthorized access, modification, or destruction. This includes implementing strong authentication and access control mechanisms, encrypting sensitive data, and regularly monitoring and auditing data systems for potential vulnerabilities or breaches [15].

Challenges of Data Availability

- **Host server failure:** If the server that stores your data fails, your data will become unavailable.
- **Storage failures:** If your physical storage device fails, you can no longer access the data it stores.
- **Network crashes:** Failure on the network means that any data accessed through it will become unavailable.
- **Data compatibility issues:** Data that is usable and working on a specific platform or environment might not be compatible on another platform.
- **Poor data quality:** Low quality datasets may contain incomplete, inconsistent, or redundant data, which could be useless for your IT operations.
- **Slow data transfers:** Data transfers could be slow depending on where the data is stored and where it is used.
- **Security and data breaches:** Malicious actors could obtain and block access to an organization's data, such as in a ransomware [17][18].

IV. BEST PRACTICES AND TOOLS FOR EXTENSIBLE STORAGE MANAGEMENT AND DATA AVAILABILITY

- **Redundancy and backups:** Backing up data is an essential aspect of data availability. Data backups should be stored in separate locations or in a distributed network. This way, if data is lost or corrupted, it can be restored quickly. You should frequently update your backups so you can restore the most recent versions of your data.
- **The use of data loss prevention tools (DLP):** Data loss prevention tools help to mitigate data breaches and damage to data centers. Most data loss prevention tools use cloud-based or third-party storage to prevent data loss. Some of the common features of these tools include monitoring, forensic analysis, and threat blocking.
- **Erasur coding:** This data protection method breaks data into fragments, expands it and then encodes it with redundant data pieces. The data is then stored across a set of different locations or storage devices. If a drive fails or data becomes corrupt, the data can be reconstructed from the segments stored on the other drives.
- **Avoid single points of failure:** Businesses should have multiple access routes. This way, in case of failure of one access route or network component, other networks continue to provide data.
- **Use of Software-Based Infrastructure:** Organizations should use software defined infrastructure and storage wherever possible, as it helps enhance data availability. Hosting data on software instead of hardware makes data movement easier and faster.



- **Removal of unnecessary data:** If certain data are outdated or not needed, they should be archived in different storage systems or securely dispose them. This will reduce clutter in the primary storage centre and ensure that only the necessary data reach the users [18][17][16].
- **Data life cycle management:** Implementing data life cycle management policies to manage data from creation to deletion. This includes archiving old data, ensuring that active data is stored on high-performance systems, and securely disposing of obsolete data.
- **Regular audits and monitoring:** Conducting regular audits and continuous monitoring to ensure storage systems and data availability techniques are functioning optimally.
- **Security integration:** Integrating robust security measures with data availability techniques. This includes encryption, access controls, and regular security assessments to protect against data breaches and cyber-attacks.

V. IMPACT OF DATA AVAILABILITY ON BUSINESS OPERATIONS

Data availability has profound impacts on business operations across multiple domains.

- It enables organizations to make data-driven decisions promptly by providing the necessary information at the right time. This empowers businesses to respond rapidly to changing market conditions, identify emerging opportunities, and mitigate potential risks.
- Data availability plays a vital role in optimizing internal processes and workflows. With data availability, organizations can ensure that employees have the information they need to perform their tasks effectively, enhancing productivity and driving business growth.
- Data availability contributes to enhanced customer experiences. By leveraging data to understand customer preferences, behavior patterns, and needs, businesses can personalize interactions, provide tailored products and services, and establish long-lasting relationships with their customers.
- Data availability also plays a crucial role in ensuring data security and compliance. Organizations must have access to their data to implement robust security measures, such as encryption and access controls, to protect sensitive information from unauthorized access or breaches.
- Data availability enables businesses to comply with regulatory requirements and industry standards by providing the necessary data for audits, reporting, and compliance assessments.
- By incorporating data availability into the decision-making process, organizations can minimize the risk of errors, optimize resource allocation, and increase the likelihood of successful outcomes [15].

VI. CONCLUSION

Extensible storage management and data availability techniques are indispensable in this data-driven age. They offer the scalability, flexibility, and reliability required to effectively manage enormous volumes of data. They are crucial for organizations to maintain operational efficiency, ensure data integrity and meet business objectives. As technology continues to evolve, these techniques will adapt to ensure that data remains accessible and secure, supporting business operations and enabling growth.

REFERENCES

- [1] C. S. Mullins, "The importance of storage management for database administration," Jun. 2013. [Online]. Available: <https://www.dbta.com/Columns/DBA-Corner/The-Importance-of-Storage-Management-for-Database-Administration-90178.aspx>
- [2] R. Castagna, "12 ways to manage your data storage strategy," May 2021. [Online]. Available: <https://www.techtarget.com/searchstorage/tip/12-ways-to-manage-your-data-storage-strategy>
- [3] "What is storage management?" [Online]. Available: <https://www.purestorage.com/knowledge/what-is-storage-management.html>
- [4] "Storage management," Feb. 2023. [Online]. Available: <https://www.geeksforgeeks.org/storage-management/>
- [5] "What is storage management?" [Online]. Available: <https://www.hpe.com/us/en/what-is/storage-management.html>
- [6] "What is storage management and how does it work?" Jun. 2022. [Online]. Available: <https://www.indeed.com/career-advice/career-development/what-is-storage-management>
- [7] M. Wilson, "Storage management—what is it and how to implement," Jul. 2023. [Online]. Available: <https://www.pcwdld.com/storage-management>
- [8] J. Howarth, "5 important data storage trends 2023–2025," Jan. 2023. [Online]. Available: <https://explodingtopics.com/blog/data-storage-trends>
- [9] D. Raffo, "How to use AI in storage management," Apr. 2022. [Online]. Available:



- <https://www.techtarget.com/searchstorage/tip/How-to-use-AI-in-storage-management>
- [10] D. Robb, "5 top cloud object storage trends in 2022," May 2022. [Online]. Available: <https://www.enterprisestorageforum.com/cloud/cloud-object-storage-trends/>
 - [11] W. Wong, "5 storage trends that can transform the public sector," Jul. 2023. [Online]. Available: <https://insights.samsung.com/2023/07/28/5-storage-trends-that-can-transform-the-public-sector-2/>
 - [12] F. Okeke, "Multicloud cheat sheet: Definition, benefits and use cases," Aug. 2023. [Online]. Available: <https://www.techrepublic.com/article/multicloud-the-smart-persons-guide/>
 - [13] M. Rouse, "What does data availability mean?" Jun. 2017. [Online]. Available: <https://www.techopedia.com/definition/14678/data-availability>
 - [14] M. Matson, "Data availability definition: Tech terms explained." [Online]. Available: <https://www.playerzero.ai/advanced/tech-terms-explained/data-availability-definition-tech-terms-explained>
 - [15] "What is data availability?" [Online]. Available: <https://www.tibco.com/reference-center/what-is-data-availability>
 - [16] "Data availability: Ensuring the continued functioning of business operations." [Online]. Available: <https://cloudian.com/guides/data-protection/data-availability/> [Accessed: Aug. 7, 2023].
 - [17] A. S. Gillis, "Data availability," Sep. 2021. [Online]. Available: <https://www.techtarget.com/searchstorage/definition/data-availability> [Accessed: Aug. 6, 2023].