Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 8, No. 6, December, 2024

BSHPC: Improve Big Data Privacy Based on Blockchain and High-Performance Computing (HPC)

Albandari Alsumayt [1*]

[1] *Computer Science Department, Applied College, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia.*

Abstract

The vast expansion and sharp rise in data across many facets of society have made it increasingly difficult to manage big data effectively. Using traditional methods to ensure the security and privacy of users' data is no longer sufficient. In keeping with this worry, massive data storage is still crucial. High-Performance Computing (HPC) is examined to determine the need for handling blockchain issues and protecting large data in a decentralized manner that strives for resilience. This study proposes the Big Data Storage High-Performance Computing (BSHPC) approach, which addresses big data considerations in storage management to maintain accuracy and enables the usage of blockchain. The best storage management is the primary benefit of BSHPC, as only critical data is kept on the blockchain, and other data may be kept in an off-chain database using the interplanetary file system (IPFS). Furthermore, the network's node authentication in this strategy depends on trustworthy nodes. On HPC computers, data authenticity and provenance tracking would be guaranteed, and managing large data across blockchains would be more secure. The proposed method is simulated using the Python-MPI version, and the results confirm the effectiveness of the proposed method based on performance and transactions. Moreover, the proposed method is evaluated with another study in the literature on MEC-based sharing, and it proves its effectiveness.

Keywords:

HPC; IOT;
Blockchain;
Big Data;
BSHPC;
Security; Privacy.

Article History:

Received:	14	February	2024
Revised:	20	October	2024
Accepted:	07	November	2024
Published:	01	December	2024

1- Introduction

Massive data produced from various sources on the Internet and by social media is called 'big data'. Big data refers to a complex and massive block of data that requires the use of a standard database management system for its proper manipulation. Moreover, the handling and maintenance of this data—which can be described as structured, semi-structured, or unstructured—can be carried out on various platforms. The rise in the amount of big data in recent years has led to a significant conversion of data levels, from petabytes to zettabytes. Additionally, as decision-making strategies are reliant on big data, the storage of huge data needs to be organized in a way that makes data retrieval both easy and secure. In this regard, many aspects of privacy and security—such as privacy, confidentiality, integrity, key management, and auditing—must be considered in the management of big data [1, 2].

Due to the complexity of big data, its storage and retrieval cannot be carried out using traditional methods. Many organizations have responded to this situation by choosing cloud storage servers to store the load of this massive data. As a result, the use of cloud servers has spread, owing to their low cost, reliability, quick accessibility, and capability to replicate data in pressing or urgent situations such as workload balance and disaster response [3].

Cloud servers provide external host solutions for accessing data anywhere and anytime. However, the level of trustworthiness and security of storing big data in the cloud is questionable, particularly in terms of the locus of proper ownership. To combat the main challenges of leakage and compromise in big data cloud storage, security methods such

* **CONTACT**: afaalsumayt@iau.edu.sa

DOI: http://dx.doi.org/10.28991/ESJ-2024-08-06-011

as encryption, hashing and asymmetric and symmetric techniques can be used. Unfortunately, existing methods such as these cannot guarantee the privacy and security of big data; thus, these considerations remain a hot topic of research. There are several stages through which big data passes in its management, and at each of these stages, the data is exposed to risk [4, 5]. The stages progress from data collection and integration to storage and management to the final process stage, storage. For example, data is collected from the web or social media; next, the collected data is classified as semi-structured, structured, or unstructured; the cleansing data are then integrated, and an integrated data model is produced; lastly, the data are stored and prepared for analysis and processing. Thus, the data would finally be ready for business decisions such as healthcare, banking, and education [6].

In recent years, a new technology called 'blockchain' – a chain of information-containing blocks – has been used in aspects of data maintenance such as preserving the confidentiality of data. Practically speaking, while most believe Bitcoin to be blockchain and vice versa Bitcoin is a cryptocurrency that utilizes blockchain technology. Created by Satoshi Nakamoto, blockchain technology consists of blocks; each block, as the name implies, is made up of several transactions, and each transaction is recorded as a hash. To wit, each block is given a unique address after its creation called a hash, and any additional changes to the block will change its hash. Many features are included in the blockchain: distributed, secure, immutable, consensus, decentralized, fast settlement, non-animus, transparency, smart contracts, and spread [7].

All executed transactions are included as a distributed public ledger. As constructed, blockchain can present an accurate record that can be trusted by anyone in the network. By exchanging information and putting together a shared, public, and trusted universal ledger, any node within the network can reach a similar consensus. Importantly, there is no centralized authority to manage the blockchain database: blockchain occurs on peer-to-peer (P2P) networks, as each node contains a copy of the blockchain ledger [8]. This leads to the building of decentralized trust since the database can only be accessed by equal and actual users. Nodes must agree regarding the specific issues that are required to validate and authorize transactions. The decentralized trust in blockchain allows low transaction costs and ensures non-reversible operations; in fact, the only trust required in the blockchain setup is trust in the code and in the system itself. As mentioned previously, the name 'blockchain' stems from the database storage of transactions classified in blocks. If a new transaction occurs, then it would be broadcasted by the sender in the P2P network to other nodes within the network [9]. When other nodes receive the transaction details, then nodes validate the process and keep that in the transaction pool. 'Validate' here refers to applying predefined checks for all structures and actions related to the transaction. A node called 'miners' is responsible for creating a new block and adding it to the transaction pool, and in mining blocks, the proof of the job is found by using variable data from the header of the new block.

Overall, continuous mathematical calculations and block hashing are the primary concerns, and this process requires substantial power and special mining hardware. When the miner node succeeds in finding a solution for the block, it is then called a 'winner'. Thus, the candidate block is the new block that enters the chain, as the block is added based on its time of arrival; thus, the last block in the blockchain is the last transaction added. Once a new block is initialized or mined, then a timestamp will be attached to it and propagated to the network. After receiving the block, each node will validate the transaction and add the new block to its local copy. The block now is a non-reversible and authorized piece of the blockchain, as the block is accepted by the greatest number of nodes. Some blocks could be inspected as a method of transactional clearing. At specific periods, the approval of valid transactions is carried out in groups to solve the double spending issue and bypass conflicts. As such, each block is a pointer to its previous block, so each block saves the hash value and some metadata of the previous block. The blockchain designation is derivative of this process, specifically in how the blocks are linked to perform a chain of many blocks (i.e., blockchain) [10, 11]. There are three types of blockchains: public, private, and consortium. A public blockchain enables any user to participate in the network, as it is considered a decentralized sector and not controlled by any organization. As its name suggests, a private blockchain is not open publicly and is only used within an organization. Lastly, a consortium blockchain occurs between both public and private blockchains and is usually employed when there are many user roles to perform. Figure 1 shows the architecture of the blockchain [12].

HPC (High-Performance Computing) refers to a technology that uses high-quality processors in parallel ways to maintain huge, big data and solve complicated issues at high speed. HPC usually performs tasks at a rate of more than one million times faster than other systems. HPC, an innovation that is widely used in industries such as energy exploration, aerospace, automotive financial computing, weather forecasting, and nuclear simulation, among others, is primarily concentrated on increasing computing performance through computer architecture, software development, and parallel algorithms. It is becoming a crucial method for conducting studies in science. Today, clusters and grids are the two most popular and distinctive ways to use HPC analogy in both industry and academics.

With the rising adoption of HPC technology, there has been a concomitant—and significant—rise in concerns regarding security issues. Attacks occur in different organizations and in a complex manner. HPC environments such as grids, computer clusters, and supercomputers supply the computational resources, while data comes from different devices and resources such as sensors, satellites, telescopes, transaction processing systems, Enterprise Software (CRM, SCM, and ERP), surveys, photos, images, conversation data, MIS, E-commerce and external sources [6]. HPC security represents one of the most worrisome issues facing users and HPC system developers, whereas in the past, the

majority of HPC systems were initially created by a small group of devoted users for private usage and thus the level of interest was not as intense. Because HPC systems are now frequently shared by several users, particularly following the advent of the 'HPC-as-a-service' concept, it has become one of the key concerns and barriers to their expansion.

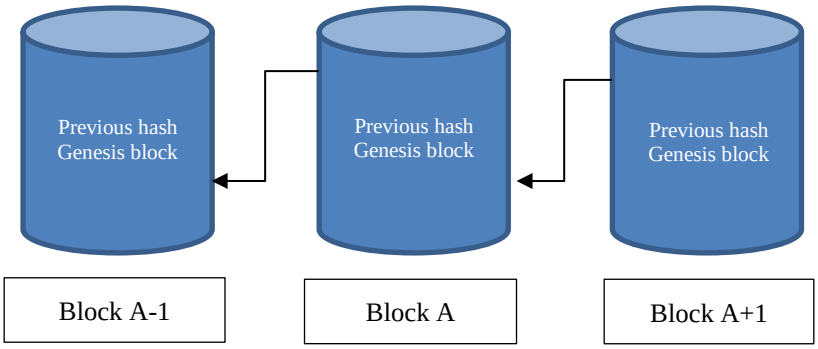

Figure 1. Blockchain architecture

To create trustworthy and responsible systems, the security of HPC systems continues to be an area of focus, particularly due to the introduction of log files to identify hostile intrusions and behaviors. To prevent malicious activities, it is necessary to limit who may utilize computing power as a tool, or perhaps better [13]. Therefore, the security concerns associated with big data and blockchain resources have risen dramatically in ordinary corporations [14].

Current blockchain architectures copy the full chain of data on the internal storage of all contributing nodes, which is one of the compelling hurdles for integrating blockchains with HPC. If present, the (compute) node-local discs (also known as burst buffers) are not intended for archiving huge data, but rather for buffering hot, typically small-sized data as opposed to data stored on distant storage environments such as *Lustre* [15]. As a result, this design is scarcely relevant to HPC systems. In other words, reality creates two major obstacles to the adoption of blockchains on HPC platforms. Regarding storage design, most scientific software is installed on HPC clusters with shared-nothing architecture, which is significantly different from the architecture of currently available blockchain systems.

Sensitive data can be extremely big, depending on where it comes from. Since they need enough computational power, the associated real-time decryption and encryption processes, during their transit and storage, also pose difficulties. These issues can be effectively solved with HPC, for example, for any real-time Big Data analyses and/or visualizations. Furthermore, the security and privacy of Big Data can be significantly enforced through the usage of blockchain in a variety of ways. The integration of HPC with Blockchain is anticipated to address several issues with big data protection and security monitoring, as well as Big Data analytics. In such situations, more efficient methods are needed to accomplish the desired goals [16-18].

In addition, a study by Esang et al. [19] examines several access control methods and guidelines intended to protect private information on global HPC clusters. In addition, this study addresses the difficulties, developing tools, and control policies related to data access control for high-performance computing (HPC), offering recommendations for administrators of systems, academics, and organizations using HPC systems. Another study by Akhtar et al. [20] provides an automata-theoretic method for creating an integrated access control policy that is economical. The proposed method reduces the conventional balanced set cover issue to this compound policy-generating problem. It demonstrates how a combined policy lowers the cost of policy review over the blockchain and accurately reflects all relevant access control policies by using Ethereum as the base of the blockchain. A study by Zhang et al. [21] created a new algorithm for cross-domain shared identity authentication that utilizes electronic signatures and certificates. Subsequently, it uses the blockchain to exchange data across domains and lower the cost of computing to increase efficiency. It utilizes an elliptic-curve cryptography approach to create an efficient agreement on the key algorithm that protects session keys to achieve fast key agreement. The performance of BCAE was validated by thorough testing on a real smart healthcare problem, and the efficiency of the proposed work was proven through rigorous security assessments. Moreover, another work by Irawan & Trihatmojo [22] poses a new algorithm for cross-domain shared identity authentication that utilizes electronic signatures and certificates. It uses the blockchain to exchange data across domains and lower the cost of computing to increase efficiency. It utilizes an elliptic-curve cryptography approach to create an efficient agreement on the key algorithm that protects session keys to achieve fast key agreement. The performance of BCAE was validated by thorough testing on a real smart healthcare problem, and the efficiency of the proposed work was proven through rigorous security assessments.

Due to the high demand for blockchain technology, data is getting bigger and bigger, and the volume is increasing drastically, which invites high computing. Given this demand, blockchain has garnered significant research interest in many sectors, such as cryptocurrency. While HPC has dealt with massive amounts of big data analytics for many years, the shared-storage system infrastructure of HPC systems and the MPI programming model for scientific programmers are the two obstacles that have hindered the HPC and scientific technology communities from implementing blockchains

into their ecosystems, regardless of resilience being one of their top system design objectives. The majority of blockchain systems assume that instead of MPI and rankings, the underlying systems are shared-nothing clusters with the TCP/IP network stack. The consensus protocols for the special I/O subsystem pose one of the most compelling difficulties for integrating blockchain technology into HPC systems. Recent initiatives have attempted to address some of the issues with implementing blockchain or blockchain-like data caching in the HPC industry. However, these systems either adhere to standard methods [23], making them inappropriate for permissioned environments such as HPC, or they do not yet have the most crucial functionality that users want.

In this study, a new method is proposed based on secure big data and blockchain architecture for HPC systems; it aims to preserve security, track provenance, and ensure data fidelity on HPC systems. The main idea of the proposed blockchain architecture is to use trust nodes as a key to authenticate nodes on HPC systems instead of using multiple keys to authenticate devices. The proposed method avoids the shortcomings of existing methods, reduces overhead, improves scalable performance, and is lightweight and distributed in memory.

Provides an overview of big data, blockchain, and HPC definitions, characteristics, and challenges. Thus, the contributions of this study are as follows:

- Analyze various big data, blockchain, and HPC security issues; additionally, the analysis of existing studies in the literature would assist in taking advantage of the beneficial aspects of existing methods while avoiding their limitations. The gap from the limitations of the existing studies would help to design a developed method.
- Develop reliable, flexible, and fault monitoring methods to manage and store big data as blockchain on HPC supercomputers. Thus, the proposed novel method in the paper is named Big Data Storage-based High-performance Computing (BSHPC). The proposed method is based on using trust nodes to complete the authentication process and ensure the application of HPC is secured.

The remainder of this paper is organized as follows: Section 2 discusses some related work. Section 3 illustrates the security analysis of big data over IoT. Section 4 explains the proposed method in detail. Section 5 illustrates the implementation of the proposed method and the evaluation process. Section 6 outlines the discussion of the proposed method. Section 7 concludes the paper and suggests future work.

2- Literature Review

Many research interests focus on blockchains in many sectors, such as smart government [24] and cryptocurrency [25, 26], the HPC [27], and computing scientific communities. Shared storage system infrastructure is not practical with the massive growth of big data and the need to retrieve and deal with this huge data. Blockchain infrastructure systems help to deal with big data and overcome the shortage with shared storage systems. Banks are now able to analyze data in real-time in search of trends according to the blockchain, which stores records of each transaction. Thus, it is possible to maximize the security of banking transactions combined with blockchain and big data [28, 29]. As there are many advantages to using blockchain technology for data storage, there are also some concerns to be aware of, such as storage restrictions: the space and scalability of blockchain storage may be restricted, making it difficult to store significant volumes of data [30].

A study by Al-Mamun et al. [31] proposed HPChain, which is a blockchain framework for HPC systems. It uses two keys compared with blockchain systems. This framework is based on the optimization of HPC systems, which is compatible with a consensus protocol. A new consensus protocol designed specifically for HPC system infrastructure is used by HPChain. It may seem simple to convert a shared-nothing protocol to a shared-storage one, but doing so while maintaining good security proves to be a difficult task. In particular, the initial consensus of the blockchain presumes that no single party might manage more than 50% of the nodes (each of which has its own unique storing device), but in a shared storage structure, a single compromised storage node could be the same as several nodes in terms of the blockchain. To achieve this, we build the protocol so that quorum voting takes into consideration both the storage and computation of nodes.

HPC and big data are distinct systems and are different technically in terms of their ecosystems. Both have separate hardware, resource management, a file system, and a programming paradigm. Although applications that use data are now a significant portion of the burden in an HPC environment, computationally demanding tasks still make up the majority of HPC workloads. Numerous software structures have been created for distributed systems, cluster administration, parallel programming models, and frameworks for machine learning because of recent developments in data-intensive applications. A widely recognized standard programming framework for high-speed computing is Open MP/MPI. With a distinct population of developers who utilize Java and other higher-level languages and a focus on usability, analytics for big data have been developed from a new perspective. This allows for the solution of issue domains without the need for complex programming [32]. In addition, a study focuses on applications that deal with intensive data used in HPC infrastructure and use map-reduce programming models. The throughput increases because of the rise of parallelism, which leads to high energy efficiency. Thus, the mission is completed in less time [33]. Hadoop works on HPC clusters with various cores, and every node is compatible with performing multiple maps/decrease missions by using cores. Therefore, the latter decreases the cost and data movement and raises the throughput regarding the access of map-reduce tasks, as it cannot predict throughput and energy consumption.

A study regarding data provenance on HPC methods is proposed to enable enhancement in many scenarios as a permissioned blockchain idea. To improve the latter method, another study by Nanayakkara et al. [34], which uses BigchainDB, considers the Practical Byzantine Fault Tolerant (PBFT). In addition, it used blockchain features such as owner-controlled assets, immutability, and decentralization. Moreover, it also used database features such as low latency, indexed query of data, and performance rate. Another study by Aniello et al. [35] proposes a method to provide reliable data integrity in distributed database systems regarding proof of work and the leader rotation approach. In addition, a study by Al-Mamun et al. [36] poses the SciChain method, which ensures reliable and secure data provenance on HPC systems. It enables dealing with HPC systems and assigning the blockchain in memory locally with low overhead memory. The limitation of the aforementioned approaches is not identifying the underlying platform framework than shared zero clusters using the blockchain systems, which can bridge gaps between blockchain and HPC technology. Rahman et al. [37] posed a MEC-based sharing economy system that influences the blockchain and off-chain framework to save ledgers that support the AI model, which is considered as the future framework of smart cities via IoT data.

Another research by Stergiou et al. [38] is the Integrated Federated Model, or InFeMo, which is the name of the model. InFeMo offers an innovative integrated model by combining all of the current Cloud models with a federated learning environment and other relevant technologies that may have integrated use with each other. Furthermore, the suggested model attempts to expand the range of managing data by providing users with a more cost-effective system design and environment. Additionally, the user could use less time queuing in every step queue by implementing the InFeMo. Moreover, a study by Zhou [39] to stop human trafficking can use data already collected from web advertising. Finding signs of human trafficking is much easier with the use of online marketing. It provides the enhanced SVM method for additional analysis of the crawled data. On unobserved data, the trained model is used to recognize labels. The approach can successfully identify adverts for concealed human trafficking, according to the results. A study by Saba et al. [40] proposes a safe blockchain model for HPC in a big data atmosphere to secure operations for financial communication with the software-defined network (SDN) architecture's smart services. The SDN controller first establishes a connection between IoT gadgets and keeps worldwide as well as local records to retain security privileges. Second, the artificial intelligence strategy is investigated to promote network flexibility and extract the most recent routing data for sending financial data, employing trustworthy and fault-tolerant approaches. The suggested protocol deals with the financial security of large data by looking into cryptographic methodologies while also ensuring data integrity with an excellent degree of network availability. Table 1 illustrates the summary of these studies.

Table 1. Summary of the existing studies

Study	Method name	Advantages	Limitations
Peng et al. (2023) [27]	A Prototype Evaluation of a Tamper-resistant HPC	The technology establishes a peer-to-peer file system for saving files and utilises a collective blockchain for cross-organizational verification and data accessibility.	Need experiments to ensure works perfect regards latency
Meiryani et al. (2023) [28]	to analyse how using blockchain technology will affect digitising accounting records.	Integrity and Reliability	Need more experiments in other sectors
Ullah & Al-Turjman (2023) [29]	a conceptual framework used in smart cities	Explains a multilayer, blockchain smart contract adoption framework.	Limited and need to deploy in real life
Saba et al. (2023) [40]	a safe blockchain approach for big data environments and HPC	High integrity and availability	Slow due to the cryptographic methods
Al-Mamun et al. (2022) [31]	HPChain	For the quorum voting, each of the storage and computing nodes are considered. Also, security and scalability	Off-chain data becomes insecure while HPChain is down
Zhou (2022) [39]	SVM (supported vector machine) is used to investigates internet textual data used to categorise advertisements	Detect malicious advertisement	Scalability
Wang et al. (2022) [33]	Applications used in HPC infrastructure, and use map-reduce programming models	High efficiency energy	Requires high security and reliability
Aniello et al. 2017 [35]	Big data security and privacy challenges	Consider the HPC and blockchain perspective in big data	Not consider authentication and trust ability
Nanayakkara et al. (2021) [34]	BigchainDB	Fault tolerance owner-controlled assets, immutability, and decentralised	Security issues
Stergiou et al. (2021) [38]	InFeMo: Flexible Big Data Management Through a Federated Cloud System	Low waiting time	Need to be more efficient
Shah et al. (2020) [30]	The IPFS is used to encode and distribute the user's file among several network peers.	Less waiting time in procedure queues	Security issues
Rahman et al. (2019) [37]	MEC-based sharing economy system	Use off-chain storage	Security issues
AlMamun et al. (2018) [23]	permissioned blockchain idea	A lightweight system	Scalability
Usman et al. (2018) [32]	Integration between HPC and big data	Improve the usage of big data	Security issues and different architecture
Usman et al. (2017) [14]	track the HPC data provenance by utilising a distributed in memory ledger	Takes advantages of the existing methods	Resilience, Power constraints, and Performance ratio

3- Security Analysis of Big Data Over IoT

Big data can be located anywhere in our lives. For example, smart cities demonstrate how big data can be employed over IoT. Specifically, big data may help in the expansion of resources and services, and in this sense, cutting-edge methods and technologies enable very effective and efficient data analysis. Along with increasing customer happiness and business potential, these tools and channels can encourage interaction and collaboration between organizations providing services to diverse parts of a smart city [41, 42]. Many challenges can affect the performance of this technology in the IoT:

- Complexity in collecting and dealing with data: The massive growth of data leads to the consideration of myriad aspects, such as the rising amount of data, numerous data divisions, bad data quality, lack of systems and procedures, difficulty integrating data, security and privacy issues, inadequate governance of information, difficulty with data computerization, and difficult data analysis.
- High cost: The costs associated with large data include computing load, demanding data administration, as well as the process and difficult statistical examinations.
- Integration of data diversity: The number of data mining tasks has substantially expanded with the continuous growth of datasets. Furthermore, it is crucial to perform data reduction, data selection, and feature selection, particularly if working with huge datasets.
- Analytical challenges: Big data presents several significant analytical hurdles, including the procedures that need to be followed if the data volume increases too much.
- Scalability: The platform should be scalable to accommodate the expanding demands of a business. Variations in the total amount of devices, services, and users should be addressed, and the processing, storing, and administration of data should all be continuously available.
- Privacy and security issues such as Denial of Service attacks (DoS) in data storage and acquisition: Enormous quantities of data are related, analyzed, and mined for useful patterns in big data analysis, and different organizations have different rules in place to protect their private data. The preservation of sensitive information is a significant problem in big data mining. As big data comes with significant security issues, privacy is evolving into a big data analytics issue. Authentication, authorization, and encryption approaches may enhance the privacy of massive data. Big data applications must contend with several security measures, including the size of the network, the variety of gadgets, real-time security monitoring, and the absence of an intrusion detection system. Thus, information security has focused on the particular security issue brought on by large data.

4- Material and Methods

A new blockchain architecture for HPC systems is proposed to deal with big data, and it is named BSHPC. It is a trustworthy method to be compatible with HPC supercomputers that have shared storage architecture. Many aspects need to be considered in the BGHPC, such as real-time fault monitoring and parallel resiliency. The proposed method can be deployed with big data and blockchain to HPC systems.

4-1- BSHPC (Big Data Storage Based High-Performance Computing) Architecture

A significant efficiency and growing need for new techniques to allow big data and HPC systems to use millions-fold parallelism, upgrade data locality, synchronization, and unified storage applications to avoid the complexity and cost of moving HPC convergence and big data.

4-2- Design

The proposed method is based on using internal filtering in storage in the blockchain. The reason for the latter is to improve both the performance and applicability of original services that are built upon blockchain. The proposed method is designed to use various storage layers to make provenance services lightweight and fast in an HPC system. Figure 2 shows the proposed method architecture.

The suggested architecture can be deployed in different resource-limited systems with constrained memory size and no local disks. The only need is remote storage servers to handle the blockchain. The main contribution of this paper is as follows:

- Design an HPC system to deal with big data and blockchain to facilitate parallel block processing to obtain a lightweight memory as deleted unwanted data regularly.
- Consider many aspects to ensure that the system works. Many aspects need to be considered in the proposed method, such as low latency, scalability, time series friendliness, and flexibility.

In low latency to hold interactive analysis, which needs quick and real-time response for any updates. Moreover, scalability, the suggested architecture requires storing historical log information and even future events for systems. Big data is the result of an increasing volume of data over time. The main challenge of big data is heterogeneity, or variety.

Data is collected from different IoT sources with different forms and would be semi-structured, structured, or unstructured. According to Weitzenboeck et al. [43], the majority of collected data, around 80%, is unstructured. The conversion process of unstructured data is costly compared with other forms. In addition, it is impractical to transform all data, as the majority of data is unstructured. Various data need many steps and procedures methods, such as data cleaning, integration, and data normalization. Another challenge is security. Data storage locations and plain data can be vulnerable to intruders who compromise data and cause data breaches. Security should be integrated between all stages of big data, starting from collection, storage, transfer, analysis, sharing, and extraction of any knowledge. The integration between big data and blockchain can preserve the confidentiality of data. Real-time cryptography of data when dealing with big data in transfer and storage in blockchain leads to challenges. HPC is an ideal and effective solution to ease the usage of blockchain integrated with big data. Therefore, this combination would enforce the privacy and security of big data. Both HPC and blockchain are the keys to solving security challenges related to big data. To achieve the expected outcomes, the appropriate algorithm is needed. In addition, key management, integrity, availability, monitoring, and auditing are the other fundamental security divisions in big data with HPC [16]. Big Data is driving the expansion of public clouds faster than any other application, and HPC on demand is a new force poised to take on this challenge. The more data are received, the more processing power would be required to analyze it. Simply said, the expansion of cloud-based HPC and big data go hand in hand. Only by deploying HPC-class assets to boost computing speed and density can there be sufficient scale to meet demand.

Therefore, the suggested architecture should manage that perfectly. In addition, time series friendliness is another aspect that needs to be thought about. For example, time series data gives an overview vision of the system behavior for a specific time and predicts future behaviors. Also, fixability is essential to add new types of events and edit any existing scheme to conciliate changes in software updates and system configuration. Moreover, this is because the time series and job scheduling-based clustering technique for blockchain systems is given. Data from time series are analyzed to identify trends and related patterns. Smart meters, surveillance systems, and weather forecasts are a few examples of applications for this type of study.

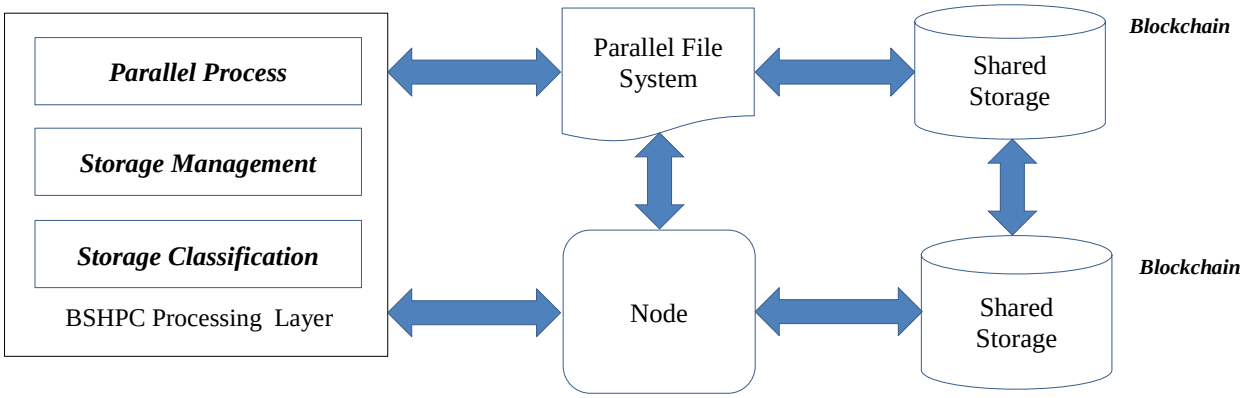

Figure 2. The architecture of the BSHPC method

BSHPC serves as a distributed, real-time practical compatible with the shared infrastructure of HPC supercomputers. The proposed method avoids the shortcomings of the state-of-the-art blockchain systems by completely classifying unimportant resources and deleting them to decrease the load of the blockchain and improve the level of HPC systems.

Based on Figure 2, many components need to be illustrated as follows:

- Parallel process: It enables parallel block processing and validation by injecting the parallel distribution method for distributing individual blocks with transactions into the conceptually separated cluster of compute nodes. Block validation is unaffected by the ledgers' persistence process in the sharing blockchain's remote storage, which also allows for independent parallel synchronization with global nodes. A transaction is a block that contains all of its attributes, including the parent block hash, date, transaction list, block ID, and transaction list. Additionally, it has all attributes, like Node-ID, that are at runtime contained in a Block item and a Transaction item.
- Storage management: To increase HPC, storage management organizes the large data processing, whether it is on a blockchain or off. preserving the validated block in the nodes' memory and enduring in the shared storage to control the resilient distributed ledger. The proposed method uses the interplanetary file system (IPFS). Designed by Joan Bennett in 2015, IPFS is a decentralized file system protocol that is run by Protocol Labs [44]. Devices operating the IPFS client software make up the IPFS network. Everyone may access the IPFS network as a local user to save and retrieve files or as an IPFS node executing the IPFS client. Textual content, video, audio, and pictures may all be stored, which is particularly helpful for non-fungible tokens (NFTs). IPFS data is identifiable by content instead of location, unlike HTTP. A data hash is created each time we send a file to IPFS. It is possible

to retrieve the file by using this hash, which distinguishes the content. In addition, Proof-of-Stake (PoW) technology is used. It is an incredibly reliable and safe consensus method. The two main parts of PoW are power and miners. Miners are the people or organizations that operate and oversee nodes, or machines, in order to sustain the network. Miners instruct nodes to use cognitive capacity from power to solve ever harder difficulties in mathematics. A proof-of-work blockchain maintains a single, valid record of data, with the ability to add new blocks of operations to the chain as long as the miner resolves the challenges [45].

- Storage classification: the classification of the data based on the flag that is given to the block. If data is used for a long period, it will be saved on the blockchain. Other non-important data will be saved on the off-chain database. The option of flag blocks as important or not is based on the data itself.
- Parallel file system: to control the equally distributed block distribution throughout the clusters. To manage the clever load-balancing of the workload that supports parallel block processing, the proposed method first establishes a set of administrators and logically divides the computational nodes into a collection of clusters. An administrator oversees each cluster. Subsequently, each cluster is further separated into several sub-clusters, each of which processes a block.
- Node: keeping tabs on each transaction, both live and pending. By combining the date and the address of the entity, such as a node that issues the transactions, a distinctive hash is created for each active transaction to avoid the execution of repeated transactions.
- Shared storage: A block is verified using two successive processes. The block is first verified using a parallel technique supported by each node's in-memory blockchain technology. It is better to bypass any classic serialized clock computation to reduce the block validation time. Second, the distant storage collaborates in the block validation process with comparatively little overhead if the majority of the compute nodes, such as at least 55%, cannot manage to agree on the validity of a block.

Consensus protocol refers to the guarantees of correct and stable operations that are handled in blockchain systems. Therefore, nodes in the network can agree on specific transactions using the consensus protocol. The proposed BSHPC is deployed in HPC systems. Many aspects are considered, such as memory scalability, fault control, resilient distributed ledger, and shared storage consensus protocol. The latter develops parallelly the attaining of consensus as it does not need extra computational processes nor overhead. Thus, the system achieves the best possible throughput level via an in-memory blockchain, as only needed information will be kept; otherwise, it will be on an off-chain database. In BSHPC, a validation process of the block is done with two consecutive steps. First, use a parallel mechanism that supports in-memory blockchain in every node. Thus, there will be a significant decrease in the block validation time, and there is a bypass for traditional sterilized processing methods based on blockchain systems. Second, in case many compute nodes fail to join the consensus regarding the block validity, then the remote storage would participate in the process of validation, which can decrease the overhead and increase the throughput. Trusted nodes are the key to authenticating other nodes. The definition of trusted nodes refers to the nodes that do not perform malicious attacks or drop blocks from them. Trusted nodes help to evaluate networks to determine if they comply with security policy standards. These prerequisites include setting up an agent administrator, running updated antivirus software, and connecting to a VPN. By doing this, only reputable devices can join the network. This method raises the threshold for security and improves the reliability of the network infrastructure.

More details regarding the aforementioned aspects related to the proposed method will be discussed in detail below:

- ***Resilient ledger:*** A resilient distributed ledger system among computing nodes is called the first section. There must be at least one irreversible, persistent storage that cannot be compromised to keep back the ledger copies in case of disasters for example, more than 50% of computational nodes smash and destroy their ledgers. This is because all computing nodes fetch and update the ledger with a few latest blocks only in temporary memory or, in ongoing local storage with a short lifetime—purged once the task is complete. Although storage-only compute nodes are not strictly less dependable than those with permanent storage, the data stored there would be lost if the process that launches the memory were to fail.
- ***Fault control:*** many procedures are considered to monitor the occurrence of faults in the proposed method. For instance, communication status is checked regularly to ensure that no fault exists. In addition, to capture any unforeseen exceptions, place checkpoints throughout the node validity phase.
- ***Memory scalability:*** The block is verified using three steps. First, the block is verified using a parallel technique based on every node's in-memory blockchain technology. Second, the conventional serialized block processing mechanism is used by cutting-edge blockchain technologies to reduce block validity time. Third, trust nodes are considered to decide if at least 55% of compute nodes could not have a decision about the block validation status. Therefore, remote storage will participate in the block validation stage to minimize overhead. Trust nodes can be nominated if the node does not have any fault problems three successive times in a specific period.

5- Results

The prototype system of the suggested BSHPC architecture in consensus protocols and user space using the Python-MPI version (3.1.4). MPI for Python enables bindings for the Message Passing Interface (MPI) standard and provides Python applications to use many processors on clusters, supercomputers, and workstations. To develop the proposed method, MPI, which is a message-passing application programming interface that eases the implementation process. To reach the highest performance, every CPU is assigned a singular process. The core modules of the prototype are only released at this stage. Other complementary contents and plugins are tested using extra edge cases, and after the tests are completed, they will be released. Figure 3 shows the implementation architecture of the BSHPC method. 800 cores on an HPC cluster are deployed in the suggested prototype. Figure 3 shows the BSHPC implementation with MPI and shared storage. The implementation architecture is generally of the proposed method on MPI. 800 cores have been deployed in the prototype system on an HPC cluster. The steps below show the performance of the proposed method as follows:

- BSHPC service API passes a new transaction from other sources, such as the HPC cluster, and it is hashed using SHA-512 to ensure confidentiality, and this refers to step 1.
- In step 2, the transactions are moved in a queue and encapsulated in a block.
- In step 3, the latter is added to the block queue.
- The storage processing module checks both the clusters of nodes and the data category in step 4.
- After that, in step 5, the block is flagged to either the blockchain or off-chain database.
- In step 6, the blocks are validated, and the consensus agent gathers the consensus of the block from the cluster.
- In step 7, once the block is successfully validated, the BSHPC method will decide to append the block to remote storage or an off-chain database based on the importance level of the data if it is flagged by the trusted nodes as important or not important. demonstrates the throughput at various node scales. Thus, Figure 4 shows the proposed method's performance in terms of the number of transactions with a high number of nodes. Classification of data level, whereas high priority is to be saved on the blockchain or low priority to be saved on the off-chain database based on the nature of the data, whereas permanent data or temporary. The flag is added based on the admin's decision when uploading data.
- Fault monitoring works all the time during the block validation process to track every node and collect initiatives to retrieve any failure and return the BSHPC method, and that is in step 8.
- In the final step, step 9, validation of source data is done and appended to the distributed ledger. Applications access the data via the BSHPC method as a validated source. The data structure form is a linked list, and it is stored in a hash table, and every tuple is connected to a block. For every hash table, a connected hash is considered a primary key to identify a transaction or a block. Features of both a block and a transaction, such as a parent block hash, transaction list, timestamp, block, transaction ID, transaction date, node ID, and application ID, are encrypted and encapsulated in a block object within the transaction object.

In the proposed method, all transactions are appended to the transaction queue. If the queue reaches the total limit, then the nodes are encrypted in the block. It is the responsibility to validate the blocks and send consensus to the related coordinators, such as MPI communicators, via the parallel module. The main role of the coordinator is to store block nodes locally in the memory after the validation process. The data is stored in a compute node format, as every node can communicate with the corresponding coordinator via the parallel method.

The experiments are conducted using 100 physical nodes. Every node is built with an Intel Core-i9 GHz core CPU with 512 GB memory. Every node could be emulated using user-level threads. In a normal configuration on HPC systems and within the compute nodes, there is no local disk used. Some software, such as Ubuntu 20.04, Python 3.11.1, NumPy 1.24.2, and mpi4py 3.1.4, are installed on each node. 800 cores are deployed in the proposed prototype. To evaluate the blockchain, YCSB [46] is used as a benchmark for evaluating blockchains. Data is generated in standard format in the YCSB benchmark. Thus, data is flexible to classify scientific data and divide it from the arbitrary applications. In this paper, it is assumed that the data is in the format that enables them to deal with blockchains. In the proposed system, the default block size is set to 6, and it deploys more than four million transactions in the experiments. Figure 4 shows the performance of the proposed method and the relationship between the number of nodes and transaction seconds. The increasing number of transactions per second raises subsequently with the number of nodes. Block size in blockchain technology relates to the volume of transactional data that each block in the chain could hold. As mentioned earlier, the blocks are saved in the nodes. It shows the number of nodes, for example, at 6 MB, and then the number of transactions increases per second.

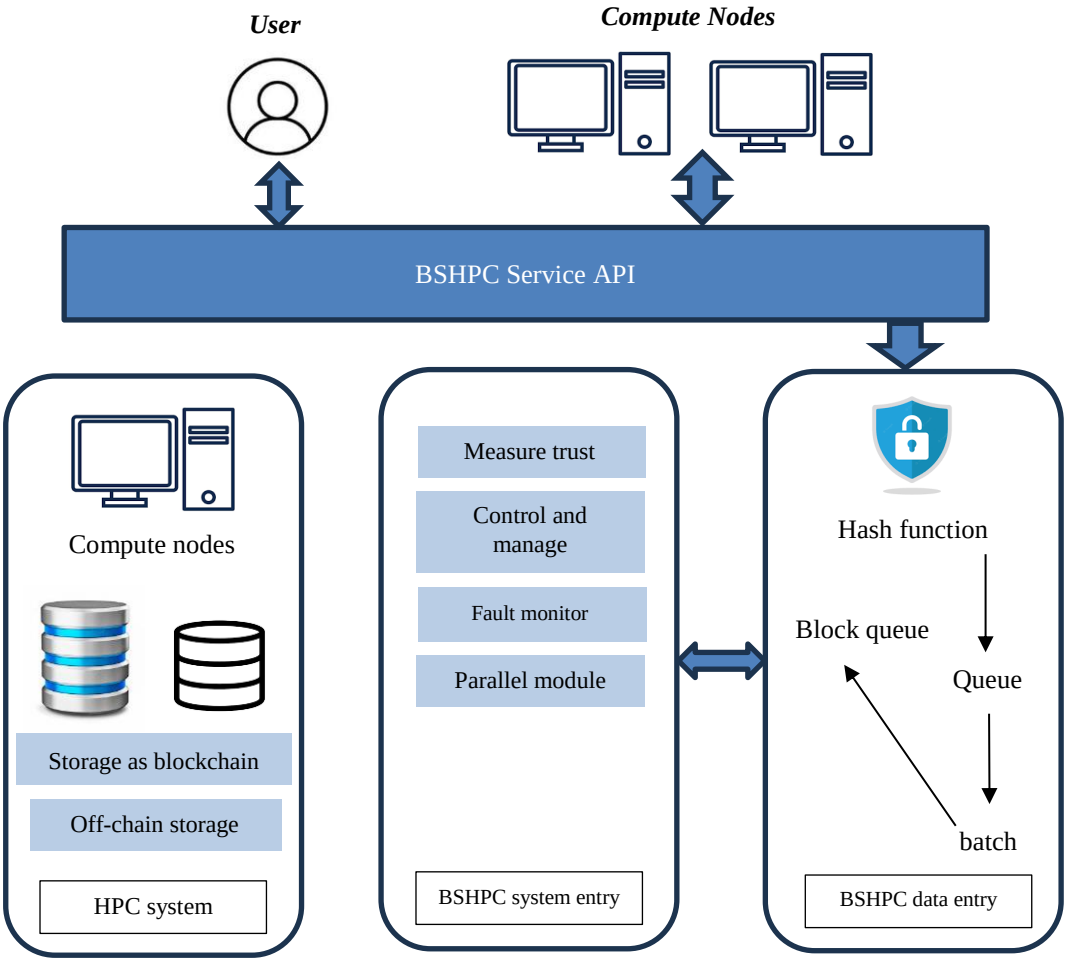

Figure 3. Proposed method implementation

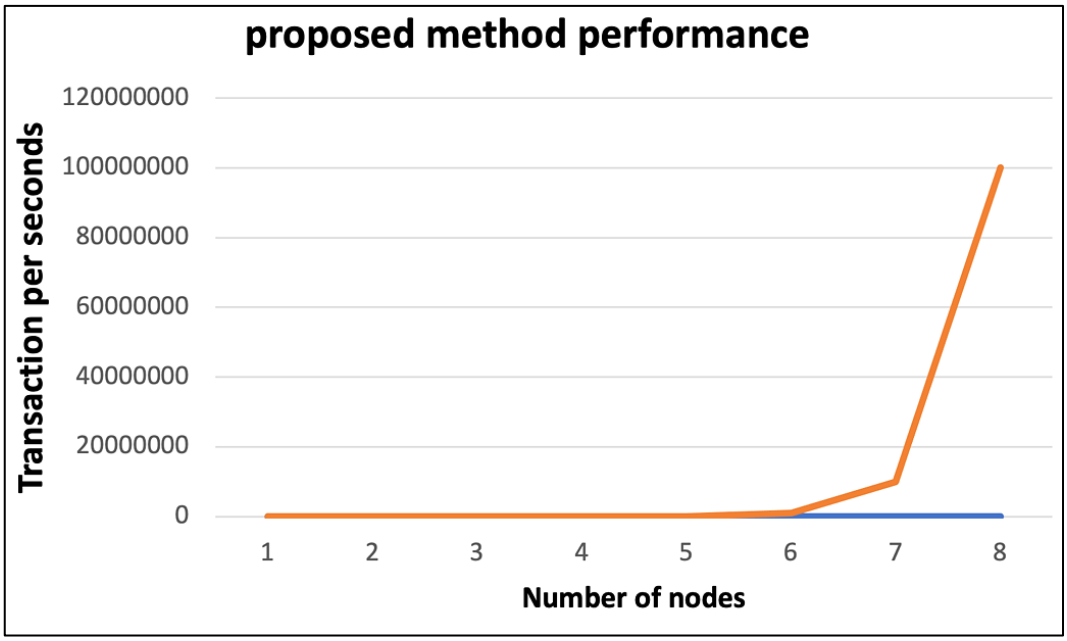

Figure 4. The proposed method performance

For comparison, the proposed method BSHPC is compared to the MEC-based sharing economy system [37]. Figure 5 shows this comparison between the proposed method and another system in services such as blocks reading and writing time, and ledger validity and confirms the effectiveness of the proposed method BSHPC. The proposed method scores lower delay time in services than the MEC-based sharing method.

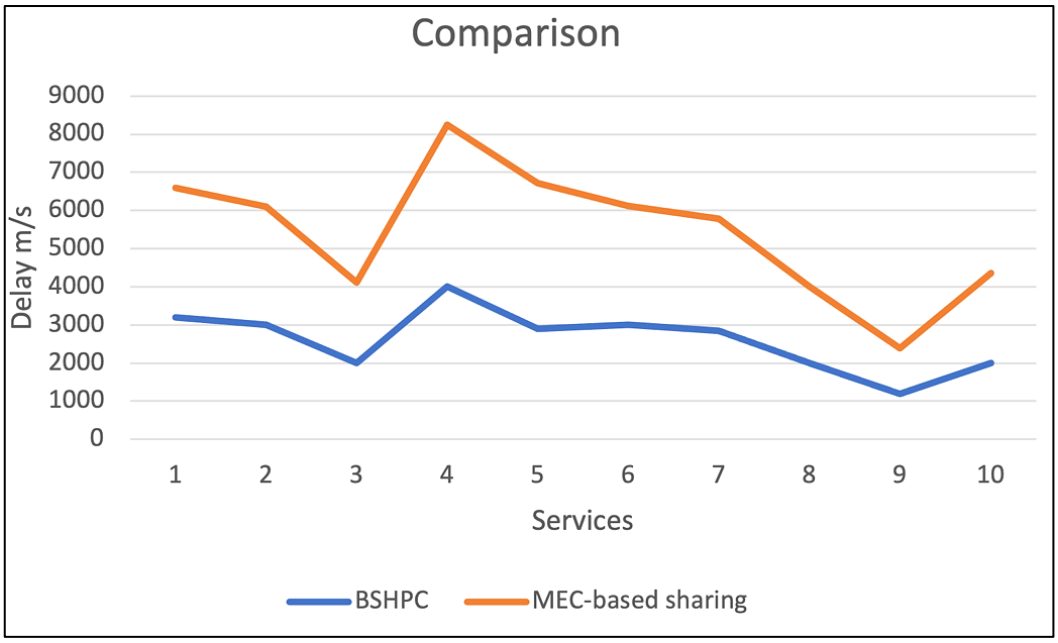

Figure 5. Evaluation between the proposed method and another system (MEC)

6- Discussion

The massive growth of data in many organizations has made the operations related to this big data a complicated proposition. Different research, industrial, and academic sectors in different fields share HPC systems and store big data in a blockchain manner. In addition, as the area of confidential and sensitive data is of rising concern, there is an essential need for HPC systems that apply security requirements and policies to process and host data. Many services offer HPC for confidential data. For instance, the University of Oslo has developed TSD services to deal with sensitive data [47], and based on a dedicated infrastructure that has a limited scale, the services probably work. In this study, a secure infrastructure is used to deal with big data imperceptive of blockchain and HPC.

The usage of massively parallel computation enables trust in the analysis. However, many issues need to be considered. First, there is the integrity of computational operations, as any node performs computations and is influenced to edit the outcome. Second, hardware errors, or the MTBF (Mean Time Between Failures) that come from commodity hardware, can give incorrect results from computations. Third, there are errors with the initial data, especially with raw data that is analyzed, as the operation includes new raw data until the outcome can be identified.

The main idea of the proposed method is to improve the HPC when dealing with blockchain and big data. The storage of big data will be classified based on the importance of the data. Thus, essential data will be kept in a blockchain manner and, otherwise, on an off-chain database, in consideration of improving efficiency and storage management. In the design of the proposed architecture, many considerations are taken into account. The first is low latency, as the proposed model needs to work in near-real-time query response and serve interactive analysis to manage timely visual updates. The suggested model operates with low latency, as nodes are passed subsequently at each specific time. The second, scalability, relates to how the suggested model requires the storage of all log data and future events, as scalability enables the accommodation of the rising volume of data. The third, being time series friendly, serves as the base of the HPC interest practitioners and provides an overview of the system's behaviors over time. The fourth, flexibility, concerns how the schema of different types of events from multiple systems' elements is not feasible and posits the essential role of flexibility in adding new events and editing existing schema to manage changes in software updates and system configurations.

BSHPC provides many services to ensure the efficiency and privacy of data. For example, block validation is important to check the validity of the block transaction, such as authenticated provenance operations. A node within the cluster receives a block, then it checks the transaction in the block regarding entities. The node gives a vote after the validation of the block. The validator node provides a hash using a private key combined with a timestamp after processing all the transactions in the block to ensure the integrity of the data. If the block is invalid, the validator node puts the hash value as a null value. BSHPC stores both transactions and blocks in the memory ledger and even in remote storage. Thus, if a node tries to save a block in the storage node, it checks if the block is stored or not. Essential data is considered important and will be stored in the blockchain. Otherwise, as noted previously, if the node is not essential, it will be stored off-chain in the database to manage memory and decrease the overhead. The evaluation process in Figure 5 supports the effectiveness of the proposed method over the other method; this effectiveness stems from the fact that the delay is lesser than that of the system in Rahman et al. [37].

Regards the reliability and trustworthiness of the proposed system, the system is designed with randomized transactions for five minutes and continues the execution ten times to demonstrate the system's dependability and trustworthiness. In each of the ten executions, the proper blockchains are held by over 50% of nodes: Four tests show lower ratios since stopping the operation once over 50% of nodes contain the correct blocks, but nine of the 10 executions provide more than 95% accuracy. In the end, we must ensure that the data of a minimum of 51% of nodes has not been altered, which is what we have done.

All in all, the integration of big data, blockchain, and HPC can improve data security drastically. Organizations and business sectors can use the integration of these technologies to ensure security and privacy in many ways—for example, identity and access management, secure communications, distributed resiliency, product tracking, carrying out complex calculations, and having strong encryption and authentication. HPC can enhance safety on both a digital and physical level. For instance, it enables the quick analysis and correlation of enormous volumes of diverse data for airport surveillance. A facial recognition mechanism, more video input, 'watch list' data, flight information, and risk intelligence databases can all be compared by an HPC-assisted airport security system to identify identity problems as they materialize. Combining all these data sources manually would be difficult, and using normal technology would not even be an option. These HPC-driven physical security systems are still in their early stages of development, but some of their key enablers are more powerful, more affordable HPC clusters as well as improvements in APIs and interoperability [48, 13]. Authentication is applied in the proposed method, used to grant storage nodes identification access. The system accepts the nodes that have completed the authentication of identities. It has an authentication tool and a credential issuer installed. These nodes and users can receive identity certificates from the seller's identity, and the authenticator can confirm that the identity issuer provided the certificates correctly.

Many organizations can receive the advantages of the proposed system when dealing with big data and ensuring HPC; this would include banks, education systems, health monitoring record keepers, and the like, as well as any in possession of big data that needs quality and precise queries. Big data analysis necessitates a significant investment in storage and computational power. Additionally, in some instances, such as those involving life or death, powerful electronic processing is required. It is generally acknowledged that HPC is an exciting opportunity for facilitating the processing of large amounts of data. However, in recent times, it has faced many significant research difficulties, including scalability of computing speed for large amounts of data involving high velocity, high diversity, and high amounts; machine learning with extremely large datasets; and informal data processing with HPC. Regarding the growth of big data, the proposed method can still obtain good results, as the data storage time is not long for unimportant data that will be saved on the off-chain database.

To meet the challenges posed by the future, there appear to be two key views. It is important to evaluate networks to examine if they comply with security policy standards. These prerequisites include setting up an agent administrator, running updated antivirus software, and connecting to a VPN. By doing this, only reputable devices may join the network. This method raises the threshold for security and improves the reliability of the network infrastructure.

7- Conclusion

Blockchain is considered a disruptive ledger technology that supports big data systems and provides a high level of security and network management. Network management refers to the applications, devices, and procedures required to set up, utilize, preserve, control, and protect the network infrastructure that collectively makes up network management. Network management's main responsibility is to make sure users have productive, successful, and rapid access to network services. The integration of these two technologies enables many services such as big data storage, analytics, acquisition, and privacy preservation. Blockchain plays an important role in big data applications such as smart healthcare, smart cities, smart grids, and smart transportation. While the use of a blockchain framework for big data encompasses many challenges, both in terms of its characteristics and in its deployment, these challenges could be resolved via big data services. Employing the immutable ledger with these technologies provides integrity, accompanied by big data analytics, and enables the best decision-making solutions.

The proposed BSHPC method poses a technique that enables the lightweight use of big data over the blockchain, which enhances the HPC. With big data, many data details are not essential, and it is not necessary to exhaust the network and blockchain of these details. Therefore, the data from the whole big data can be categorized as essential and not essential, and only essential data would be kept in the blockchain. A system prototype of the suggested BSHPC method was implemented and evaluated using more than four million transactions over 800 cores. High performance and overhead reliability were compared to the state-of-the-art systems, and the performance of the proposed method was examined to test its efficiency, such as data integration methods for blockchain data sources. The results of this evaluation affirm the power and strengths of the proposed method compared with another similar method. For future studies, other aspects such as fault tolerance will be considered, and the number of nodes will be increased to test the effects of this change.

8- Declarations

8-1- Data Availability Statement

Data sharing is not applicable to this article.

8-2- Funding and Acknowledgments

The author received no financial support for the research, authorship, and/or publication of this article.

8-3- Institutional Review Board Statement

Not applicable.

8-4- Informed Consent Statement

Not applicable.

8-5- Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the author.

9- References

[1] Payton, T., & Claypoole, T. (2023). Privacy in the age of Big data: Recognizing threats, defending your rights, and protecting your family. Rowman & Littlefield, Maryland, United States.

[2] Vasa, J., & Thakkar, A. (2023). Deep Learning: Differential Privacy Preservation in the Era of Big Data. Journal of Computer Information Systems, 63(3), 608–631. doi:10.1080/08874417.2022.2089775.

[3] Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., ... & Jarić, I. (2022). Big-data approaches lead to an increased understanding of the ecology of animal movement. Science, 375(6582), eabg1780. doi:10.1126/science.abg1780.

[4] Lin, G., Zhang, H., Song, X., & Shibasaki, R. (2023). Blockchain for location-based big data-driven services. Handbook of Mobility Data Mining, Volume 3: Mobility Data-Driven Applications, 153. doi:10.1016/B978-0-323-95892-9.00009-7.

[5] Wang, R., Xu, C., Dong, R., Luo, Z., Zheng, R., & Zhang, X. (2023). A secured big-data sharing platform for materials genome engineering: State-of-the-art, challenges and architecture. Future Generation Computer Systems, 142, 59–74. doi:10.1016/j.future.2022.12.026.

[6] Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., De-la-Hoz-Franco, E., & De-La-Hoz-Valdiris, E. (2022). Trends and Future Perspective Challenges in Big Data. Smart Innovation, Systems and Technologies, 253, 309–325. doi:10.1007/978-981-16-5036-9_30.

[7] Oliveira, T. A., Oliver, M., & Ramalhinho, H. (2020). Challenges for connecting citizens and smart cities: ICT, e-governance and blockchain. Sustainability (Switzerland), 12(7), 2926. doi:10.3390/su12072926.

[8] Krichen, M., Ammi, M., Mihoub, A., & Almutiq, M. (2022). Blockchain for Modern Applications: A Survey. Sensors, 22(14), 5274. doi:10.3390/s22145274.

[9] Dharma Putra, G., Kang, C., Kanhere, S. S., & Won-Ki Hong, J. (2022). DeTRM: Decentralized Trust and Reputation Management for Blockchain-based Supply Chains. IEEE International Conference on Blockchain and Cryptocurrency, ICBC 2022, 1–5. doi:10.1109/ICBC54727.2022.9805565.

[10] Shrimali, B., & Patel, H. B. (2022). Blockchain state-of-the-art: architecture, use cases, consensus, challenges and opportunities. Journal of King Saud University-Computer and Information Sciences, 34(9), 6793-6807. doi:10.1016/j.jksuci.2021.08.005.

[11] Li, X., Jiang, P., Chen, T., Luo, X., & Wen, Q. (2020). A survey on the security of blockchain systems. Future Generation Computer Systems, 107, 841–853. doi:10.1016/j.future.2017.08.020.

[12] Zhang, S., & Lee, J. H. (2020). Analysis of the main consensus protocols of blockchain. ICT Express, 6(2), 93–97. doi:10.1016/j.icte.2019.08.001.

[13] Luo, Z., Qu, Z., Nguyen, T., Zeng, H., & Lu, Z. (2019). Security of HPC Systems: From a Log-analyzing Perspective. ICST Transactions on Security and Safety, 6(21), 163134. doi:10.4108/eai.19-8-2019.163134.

[14] Usman, S., Mehmood, R., & Katib, I. (2018). Big data and HPC convergence: The cutting edge and outlook. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 224, 11–26. doi:10.1007/978-3-319-94180-6_4.

[15] Schwan, P. (2003). Lustre: Building a File System for 1,000-node Clusters. Proceedings of the Linux Symposium, 401–409.

[16] Khalil Alsulbi, Maher Khemakhem, Abdullah Basuhail, Fathy Eassa, Kamal Mansur Jambi, & Khalid Almarhabi. (2021). Big Data Security and Privacy: A Taxonomy with Some HPC and Blockchain Perspectives. IJCSNS International Journal of Computer Science and Network Security, 21(7), 43–55,.

[17] Georgiou, Y., Zhou, N., Zhong, L., Hoppe, D., Pospieszny, M., Papadopoulou, N., Nikas, K., Nikolos, O. L., Kranas, P., Karagiorgou, S., Pascolo, E., Mercier, M., & Velho, P. (2020). Converging HPC, Big Data and Cloud Technologies for Precision Agriculture Data Analytics on Supercomputers. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12321 LNCS, 368–379. doi:10.1007/978-3-030-59851-8_25.

[18] Nystrom, N. A., Buitrago, P. A., & Blood, P. D. (2019). Bridges: Converging HPC, AI, and Big Data for Enabling Discovery. Contemporary High Performance Computing, 355–383. doi:10.1201/9781351036863-14.

[19] Esang, M. O., Akpan, E. I. O., Jimoh, T. G., Ajibola, H. R., & Dan, E. E. (2024). Data Access Control In High-Performance Computing: Preventing Unauthorized Access To Sensitive Data In Shared Clusters. International Journal of Agribusiness and Sustainable Development Research, 1(1), 38–45.

[20] Akhtar, A., Barati, M., Shafiq, B., Rana, O., Afzal, A., Vaidya, J., & Shamail, S. (2024). Blockchain Based Auditable Access Control For Business Processes With Event Driven Policies. IEEE Transactions on Dependable and Secure Computing, 4699 - 4716. doi:10.1109/TDSC.2024.3356811.

[21] Zhang, S., Yan, Z., Liang, W., Li, K. C., & Di Martino, B. (2024). BCAE: A Blockchain-Based Cross Domain Authentication Scheme for Edge Computing. IEEE Internet of Things Journal, 11(13), 24035–24048. doi:10.1109/JIOT.2024.3387934.

[22] Irawan, B., & Trihatmojo, D. (2024). Decentralized Trusted Storage of Audio-Video Log Data Based on Blockchain Technology and IPFS. International Journal of Science, Technology & Management, 5(2), 473–484. doi:10.46729/ijstm.v5i2.1084.

[23] Al-Mamun, A., Li, T., Sadoghi, M., & Zhao, D. (2018). In-memory Blockchain: Toward Efficient and Trustworthy Data Provenance for HPC Systems. Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018, 3808–3813. doi:10.1109/BigData.2018.8621897.

[24] Ølnes, S. (2016). Beyond Bitcoin enabling smart government using blockchain technology. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9820 LNCS, 253–264. doi:10.1007/978-3-319-44421-5_20.

[25] Eyal, I., Gencer, A. E., Sirer, E. G., & Van Renesse, R. (2016). Bitcoin-NG: A scalable blockchain protocol. Proceedings of the 13th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2016, 16, 45–59.

[26] Gilad, Y., Hemo, R., Micali, S., Vlachos, G., & Zeldovich, N. (2017). Algorand: Scaling Byzantine Agreements for Cryptocurrencies. SOSP 2017 - Proceedings of the 26th ACM Symposium on Operating Systems Principles, 51–68. doi:10.1145/3132747.3132757.

[27] Peng, S., Bao, W., Liu, H., Xiao, X., Shang, J., Han, L., Wang, S., Xie, X., & Xu, Y. (2023). A peer-to-peer file storage and sharing system based on consortium blockchain. Future Generation Computer Systems, 141, 197–204. doi:10.1016/j.future.2022.11.010.

[28] Meiryani, Marcelino, Rusmanto, T., Lesmana, T., Modjo, M. I., & Budiarto, A. Y. (2023). Blockchain Technology in Digitalization of Recording Accounting Transactions. Journal of Theoretical and Applied Information Technology, 101(9), 3351–3361.

[29] Ullah, F., & Al-Turjman, F. (2023). A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities. Neural Computing and Applications, 35(7), 5033–5054. doi:10.1007/s00521-021-05800-6.

[30] Shah, M., Shaikh, M., Mishra, V., & Tuscano, G. (2020). Decentralized Cloud Storage Using Blockchain. Proceedings of the 4th International Conference on Trends in Electronics and Informatics, ICOEI 2020, 384–389. doi:10.1109/ICOEI48184.2020.9143004.

[31] Al-Mamun, A., Li, T., Sadoghi, M., Jiang, L., Shen, H., Zhao, D., & Shen, H.-I. (2019). HPChain: An MPI-Based Blockchain Framework for Data Fidelity in High-Performance Computing Systems. ACM Reference Format, 17–19.

[32] Usman, S., Mehmood, R., & Katib, I. (2020). Big data and HPC convergence for smart infrastructures: A review and proposed architecture. Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies, 561-586. doi:10.1007/978-3-030-13705-2_23.

[33] Wang, J., Xu, C., Zhang, J., & Zhong, R. (2022). Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems, 62, 738–752. doi:10.1016/j.jmsy.2021.03.005.

[34] Nanayakkara, S., Rodrigo, M. N. N., Perera, S., Weerasuriya, G. T., & Hijazi, A. A. (2021). A methodology for selection of a Blockchain platform to develop an enterprise system. Journal of Industrial Information Integration, 23, 100215. doi:10.1016/j.jii.2021.100215.

[35] Aniello, L., Baldoni, R., Gaetani, E., Lombardi, F., Margheri, A., & Sassone, V. (2017). A Prototype Evaluation of a Tamper-Resistant High Performance Blockchain-Based Transaction Log for a Distributed Database. Proceedings - 2017 13th European Dependable Computing Conference, EDCC 2017, 151–154. doi:10.1109/EDCC.2017.31.

[36] Al-Mamun, A., Yan, F., & Zhao, D. (2021). SciChain: Blockchain-enabled lightweight and efficient data provenance for reproducible scientific computing. Proceedings - International Conference on Data Engineering, 2021-April, 1853–1858. doi:10.1109/ICDE51399.2021.00166.

[37] Rahman, M. A., Rashid, M. M., Shamim Hossain, M., Hassanain, E., Alhamid, M. F., & Guizani, M. (2019). Blockchain and IoT-Based Cognitive Edge Framework for Sharing Economy Services in a Smart City. IEEE Access, 7, 18611–18621. doi:10.1109/ACCESS.2019.2896065.

[38] Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2022). InFeMo: Flexible Big Data Management Through a Federated Cloud System. ACM Transactions on Internet Technology, 22(2), 1–22. doi:10.1145/3426972.

[39] Zhou, Q. (2022). A Study on Human Transiting Based on Big Data and Web Semantics: Distinguishment and Detection. International Journal on Semantic Web and Information Systems, 18(1), 1–18,. doi:10.4018/IJSWIS.310055.

[40] Saba, T., Haseeb, K., Rehman, A., & Jeon, G. (2024). Blockchain-Enabled Intelligent IoT Protocol for High-Performance and Secured Big Financial Data Transaction. IEEE Transactions on Computational Social Systems, 11(2), 1667–1674. doi:10.1109/TCSS.2023.3268592.

[41] Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I. A. T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. IEEE Access, 5, 5247–5261. doi:10.1109/ACCESS.2017.2689040.

[42] Talebkhah, M., Sali, A., Marjani, M., Gordan, M., Hashim, S. J., & Rokhani, F. Z. (2021). IoT and Big Data Applications in Smart Cities: Recent Advances, Challenges, and Critical Issues. IEEE Access, 9, 55465–55484. doi:10.1109/ACCESS.2021.3070905.

[43] Weitzenboeck, E. M., Lison, P., Cyndecka, M., & Langford, M. (2022). The GDPR and unstructured data: is anonymization possible? International Data Privacy Law, 12(3), 184–206. doi:10.1093/idpl/ipac008.

[44] N. Sangeeta, & Nam, S. Y. (2023). Blockchain and Interplanetary File System (IPFS)-Based Data Storage System for Vehicular Networks with Keyword Search Capability. Electronics (Switzerland), 12(7), 1545. doi:10.3390/electronics12071545.

[45] Wendl, M., Doan, M. H., & Sassen, R. (2023). The environmental impact of cryptocurrencies using proof of work and proof of stake consensus algorithms: A systematic review. Journal of Environmental Management, 326, 116530. doi:10.1016/j.jenvman.2022.116530.

[46] Sidhanta, S., Mukhopadhyay, S., & Golab, W. (2019). Dyn-YCSB: Benchmarking adaptive frameworks. In Proceedings - 2019 IEEE World Congress on Services, SERVICES 2019, 2642, 392–393. doi:10.1109/SERVICES.2019.00119.

[47] Øvrelid, E., Bygstad, B., & Thomassen, G. (2021). TSD: A research platform for sensitive data. Procedia Computer Science, 181, 127–134. doi:10.1016/j.procs.2021.01.112.

[48] Singh, A. K., & Sharma, S. D. (2019). High performance computing (HPC) data center for information as a service (IaaS) security checklist: Cloud data governance. Webology, 16(2), 83–96. doi:10.14704/web/v16i2/a192.