QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

A New Efficient Method for Information Security in Hadoop

Shadan Muhammed Jihad Abdalwahid
Department of Information System Engineering, Erbil Technical Engineering College, Erbil Polytechnic University, Erbil, Iraq.
shadan.abdalwahid@epu.edu.iq

Banar Fareed Ibrahim
Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Erbil, Kurdistan Region, Iraq
Bnar.fareed@lfu.edu.krd

Sami Hassan Ismael
Technical Institute of Bardarash, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
sami.hussein@dpu.edu.krd.

Shahab Wahhab Kareem
Department of Information System Engineering, Erbil Technical Engineering College, Erbil Polytechnic University, Erbil, Kurdistan Region, Iraq.
shahab.kareem@epu.edu.iq

ARTICLE INFO

Article History:
Received: 22/2/2022
Accepted: 12/5/2022
Published: Summer 2022

Keywords: *Big Data, Hadoop, RSA, Rabin, Cryptography.*

Doi:
10.25212/lfu.qzj.7.2.42

ABSTRACT

A significant difficulty in developing a specialized Hadoop for cloud computing, and to tackle this problem, an anticipated safe cloud computing system has been built. Hadoop was utilized in this field to develop and improve the security of handling and gathering data from users. Other Apache Big Data technologies include Hadoop, which uses the Map-Reduce architecture to prepare enormous amounts of data. Hadoop is one of the most significant tools for dealing with Big Data issues. Data storage security is one of the most challenging tasks, and the Hadoop distributed file system (HDFS) lacks a well-defined security policy. The proposed technique encrypts all of the files stored in HDFS with public-key cryptography to safeguard them all. The acquired data is encrypted in HDFS during the data collecting process using the proposed data encryption method (Rabin RZ). The

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

proposed method is compared to the Paillier method and the default Rivest–Shamir–Adleman (RSA) cryptosystem, respectively. When compared to other cryptosystems, the proposed method has a more powerful computational complexity and a smaller latency than the alternatives.

1. Introduction

In recent years, cloud technology has attracted increasing attention from both businesses and the wider community. User access to a diverse range of resources is provided by cloud computing, including computing platforms, storage facilities, computing power, and internet-based software applications. Amazon, Google, IBM, Microsoft, and other major cloud computing providers are currently available on the market. It is becoming increasingly essential to protect data from different types of users as the number of businesses utilizing cloud resources grows. Cloud computing is currently being utilized in a massive amount across various industries (Shahab Wahhab Kareem, Raghad Zuhair Yousif, Shadan Mohammed Jihad Abdalwahid, 2020). Massive amounts of data are generated daily in our lives. To store these enormous amounts of data, consumers are turning to cloud computing services. The security, protection, and processing of data that is the property of the user are some of the most significant challenges that cloud computing must overcome. The term "big data" refers to the massive collection of data that is assigned to processing and retrieval. Moreover, big data must also be associated with the compilation of incredibly sensitive and important data from social networking sites and also matters related to the government, resulting in security concerns for those involved. It is necessary to encrypt the data collected using effective features to keep it safe from illegal disclosure. It can categorize the qualities of Big Data into four categories, which are denoted by the letters V for volume (Shadan Mohammed Jihad ABDALWAHID, Raghad Zuhair YOUSIF, Shahab Wahhab KAREEM, 2019) (shadan MohammedJihad abdalwahid Shahab Wahhab Kareem, Raghad Zuhair Yousif, 2020) , V for veracity (ascertain ability), V for variety (variety), and V for velocity (speed). Every V has a real activity that stems from producing as a result of Big Knowledge (Cao, et al., 2022). As a result, volume refers to the amount of information that has been delivered and that

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

has the potential to be stored, which could be in the hundreds of terabytes. The term "data quality" is defined by the term "veracity," which in this context, in addition, the terms "data confidentiality," "data integrity," "data privacy," and "data availability" are used to describe data. The term "variety" refers to the types of information and their structures, which include structured, semi-structured, and unstructured information. The input and output rates of data stream generation are represented by the term "velocity." A level of abstraction is provided in this context so that the enormous information frameworks can store information freely and at a high rate of activity. According to this perspective, the proposed scheme can be considered an opportunity to improve on what was previously conveyed in the paper [4] and encryption and decryption of Big Data files using Hadoop-integrated AES and OTP algorithms at both the decipherment and decryption stages (Raghad Z.Yousif, Shahab.Wahhab Kareem, Ammar. O.Hasan, 2016).

Rivest, Shamir, and Adleman devised the RSA cryptosystem in 1977. It is a widely used system for assuring the privacy and validity of digital data in all applications where data protection is a concern. RSA remains the most used cryptosystem, particularly for high-end devices such as E-commerce and VPN servers. Modular exponentiation is the foundation of RSA (M, 2021). Its security depends on the intractability of the factorization problem, and it may be used to ensure either secrecy or digital signatures [5].

The difficulty of finding a square root modulo a composite number inspired Rabin to create a public-key cryptosystem. Rabin's work is theoretically significant since it gave the first proved security for public-key cryptosystems: The Rabin cryptosystem's security is exactly the integer factorization problem's intractability (IFP). (Recall our explanation of the RSA: whether the RSA issue is analogous to the IFP is unknown.) The Rabin cryptosystem's encryption method is also incredibly efficient, making it ideal for specific applications such as encryption conducted by hand-held devices [4]. Under the Decisional Composite Residuosity Assumption, the Paillier cryptosystem is a probabilistic, additively homomorphism encryption method that is known to be semantically safe. Such schemes have an additive homomorphic property: any

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

encryption of m1 and m2 may be achieved as E(m1,r1) E(m2,r2)E(m1+m2, r1r2). In the context of electronic voting, verifiable encryption, and threshold methods, this trait is quite valuable (Poongodi, et al., 2022). Surprisingly, with the right assumptions, such techniques may be simply shown as semantically safe. The Paillier scheme's security is based on the difficulty of computing discrete logarithms [2].

The following describes the structure of this paper: Section II provides an overview of the security framework. Enterprises are gradually believing in Hadoop for storing their important data and processing it. Hadoop is an open-source context designed for distributed storage and concurrent processing on big databases. Big Data in Hadoop is introduced in Section III, which is based on HDFS and MapReduce and is a part of the Hadoop ecosystem. Afterward, Section IV describes the proposed optimized hybrid decipherment algorithm and compares it to the classical public-key cryptosystems, before demonstrating how it can be used to secure Big Data at the Hadoop data processing cluster. In the following section, you will find a discussion of the simulation results, and in Section VI, you will find a list of conclusions.

2. Literature Review

The data and the analyses conducted on the data must be precise, and the establishments must ensure this. Large-scale data processing has become almost indispensable for many governments and business applications due to the incredible rate at which data is being generated, accumulated, and investigated through computer systems. Many factors have played a role in the massive increase in data, including the emergence and widespread use of IoT technology, object localization, and tracking, as well as the increasing adoption of medical applications that receive the data. There are some drawbacks to the widespread use of big data. The information gathered usually includes some personal information about individuals, including, for example, information that would be problematic if discovered by adversaries (Sharma, Miran, & Ahmed, 2022). To facilitate the possession and purchase of stolen personal information, criminal organizations establish underground markets. Intelligence provided by the government agencies relies on specific, eavesdropping systems used by corporations and the government's

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

opposition, as well as competitive advantage systems, to gather information. Recent high-profile cyber-attacks on commercial enterprises have demonstrated the potential for harm, and several governments have paid millions of dollars to these organizations, resulting in significant harm to the individuals and organizations who have been the targets of these attacks. Furthermore, security within each cloud collaboration is still in the early stages of development, because there are numerous security flaws in the cloud, and users' information is at risk. Administrators of cloud computing services have no idea where or in what format the data is being stored, and they have no way of knowing where the data is stored. Customers must be assured in this situation that adequate safety measures will be implemented to protect their data from information spillage and to maintain control over information flow and control of information. A significant feature of Hadoop is the division of data and computation among a considerable number of hosts and the implementation of application computations corresponding close to their data. Furthermore, the handling and examination of massive amounts of data on a cloud server farm is a fundamental issue. A few recently made accessible distributed structures, for example, HADOOP, which was created to store and process Big Data, and the Google File System, are examples of distributed structures that are currently available. However, the distributed HADOOP system is well-known among industries as well as experimentation networks, and for good reason. HADOOP is comprised of several different components, including I HDFS, which is used for storing large and unstructured data sets, and (ii) the Map-Reduce framework, which is used for processing large amounts of data. Hadoop is typically used in forms that have massive data nodes, up to and including petabytes in size (KAREEM, 2020).

The Hadoop software does not include any security mechanisms. Several studies have examined the use of ciphering algorithms in Hadoop data encryption and storage, followed by encrypted data storage in HDFS. Cryptanalysis schemes perform a variety of replacements and manipulations on the clear message to modify it within the cipher text, which must be composed entirely of arbitrary and unintelligible data. For the sake of information security, a variety of ciphering algorithms have been developed and put into practice. As a result, there are two significant categories of

cryptosystems: symmetric-key and private-key) cryptosystems (Shahab Wahhab Kareem, Yahya Tareq Hussein, 2017) including the Advanced Encryption Standard (AES), Data Encryption Standard (DES), and Triple-DES; (ii) symmetric-key encryption, such as the AES, DES, and Triple DES; and (public-key) cryptosystems such as the Elliptic Curve Diffie-Hellman (ECDH) and RSA. In a paper, the security architecture for Hadoop was examined. As a result, AES encryption and decryption classes have been added to the library for data encryption and decryption. It is used to demonstrate two integrations, one HDFS-RSA integration, and the other HDFS pairing integration. These integrations are amazing different kinds of extensions for HDFS. The results of the analyses revealed that there was adequate insurance for understanding tasks and critical insurance for composing activities. Several cryptographic schemes were integrated with the cloud data storage system on Hadoop for encrypting data files in HDFS based on DES and RSA, then referred to IDEA for securing the private key in RSA for each user, the encryption of the HDFS files is implemented when they are put away as support in the wake of transferring information, In this paper, a modified asymmetric key cryptosystem is proposed to secure large amounts of data.

The proposed algorithm is based on a previous NSCT algorithm. This advancement is based on the addition of a key expansion mechanism to create keys using the F-function, as well as a reduction in the number of cycles to save time. The proposed algorithm has been constructed in three phases; the first stage uses FELICS simulation to determine the algorithm's efficiency. In the second stage, the proposed technique is implemented in Java to encrypt the text and determine the encryption's execution time. Finally, the proposed method is implemented in the MATLAB language to encrypt photos and determine the technique's effectiveness in terms of defending against hacker assaults (Galal A. Al-Rummana, Abdulrazzaq H. A. Al-Ahdal, G. N. Shinde, 2021).

To encrypt the data gathered in HDFS, this study presents a hybrid solution combining two prominent asymmetric key cryptosystems (RSA and ElGamal). Thus, data is encrypted before being stored in HDFS using the suggested cryptosystem. The user of the cloud has two options for uploading data: non-secure or secure. In comparison

to the RSA cryptosystem alone, the hybrid technique has a higher computational complexity and lower latency (Shahab Wahhab Kareem, Raghad Zuhair Yousif, Shadan Mohammed Jihad Abdalwahid, 2020).

This study presents a hybrid approach for encrypting files stored in HDFS that combines two well-known asymmetric key cryptosystems (RSA and Paillier). The suggested cryptosystem is used to encrypt data before it is saved in HDFS. Each cloud user has two options for uploading files: non-secure or secure. In comparison to the RSA cryptosystem alone, the hybrid system has a higher computational complexity and shorter delay (Shadan Mohammed Jihad Abdalwahid, Raghad Zuhair Yousif, Shahab Wahhab Kareem, 2019).

The proposed method was used in this research to improve the completion of the encryption/decryption file by combining (OU) methods on Hadoop.
Wherever the data is encrypted in HDFS and then decrypted in the Map Task. Using encryption, the suggested cloud computing protection system ensured the user's confidentiality and privacy. The proposed approach for converting between the system and the cloud user.

The Okamoto-Uchiyama cryptosystem was implemented and the results were compared to the RSA cryptosystem (KAREEM, 2020).

To encrypt files saved in HDFS, this work presents a hybrid technique combining two well-known asymmetric key cryptosystems (RSA and Rabin). As a result, the suggested cryptosystem is used to encrypt data before it is stored in HDFS. In the proposed system, cloud users can upload data in one of two ways: secure or non-secure. In compared to the RSA cryptosystem alone, the hybrid technique has a higher computational complexity and lower latency (Raghad Z. Yousif1, Shahab W. Kareem, Shadan M. Abdalwahid, 2020).

The encryption technique of an autonomous single-chip computer and embedded system based on big data is investigated in this paper using research methods such as literature data, mathematical statistics, and logical analysis.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

It primarily enhances and improves the present data encryption algorithm. The data encryption system can encrypt and store data, which can be said to realize transparent encryption and decryption in the reading and writing process. According to studies, although the system's built-in function takes 5-6 seconds to open a file, the function takes 45-48 seconds, which is somewhat longer (Chen, 2021).

The purpose of this work is to give a comprehensive analysis of privacy preservation strategies in big data, as well as the issues that these mechanisms face. In addition, this article discusses contemporary privacy-preserving strategies in big data, such as concealing a needle in a haystack, identity-based anonymization, differential privacy, privacy-preserving big data publication, and quick anonymization of huge data streams. This study discusses the privacy and security implications of big data in healthcare. A comparative assessment of different contemporary large data privacy approaches is also carried out (Priyank Jain, Manasi Gyanchandani and Nilay Khare, 2016).

The time complexity of the some of the several methods are presented in the table 1.

Table 1: Computational complexity of some cryptography algorithms

Methods	**Encryption**	**Decryption**
RSA	$T(c) = O (\log n)^3$	$T(M) = O (\log n)^3$
Paillier	$T (c) = 2O (\log n)^2$	$T(M) = O (\log n)^3$
Hybrid system in (Shadan Mohammed Jihad Abdalwahid, Raghad Zuhair Yousif, Shahab Wahhab Kareem, 2019)	$T(c) = 2O (\log n)^3 + O (\log n)$	$T(M) = 2 O (\log n)^3 + 3 + O (\log n)$
Rabin	$T (c) = O (\log n)^2$	$T(M) = O (\log n)^3$
Hybrid system in (Raghad Z. Yousif1, Shahab W. Kareem, Shadan M. Abdalwahid, 2020)	$T(c) = O (\log n)^3$	$T(M) = 2 O (\log n)^3$
Okamoto-Uchiyama (KAREEM, 2020)	$T(C) = O (\log n)^3 + O (\log n)$	$T(M) = 2 O (\log n)^3 + O (\log n)$
ElGamal	$T(C) = 2 O(\log n)^3 + O(\log n)$	$T(M) = 2 O(\log n)^3 + O(\log n)$
Proposed system in (Shahab Wahhab Kareem, Raghad Zuhair Yousif, Shadan Mohammed Jihad Abdalwahid, 2020)	$T(C) = O(\log n)^3 + O(\log n)$	$T(M) = 2 O(\log n)^3 + O(\log n)$

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

3. Issues Relating to Security

Big data is concerned with the storage, processing, and recovery of information. To accomplish this, a variety of technologies, including memory management, transaction management, visualization, and network networking, are implemented. As a result, the security issues associated with these technologies are also applicable to big data. The four major Big Data security issues are authentication, data-level security, network-level security, and generic issues. A study conducted by (M. Elsayed and M. Zulkernine,, 2018), and a study conducted by (Shadan Mohammed Jihad Abdalwahid, Raghad Zuhair Yousif, Shahab Wahhab Kareem, 2019) (Shadan Mohammed Jihad ABDALWAHID, Raghad Zuhair YOUSIF, Shahab Wahhab KAREEM, 2019) (Shahab Wahhab Kareem, Raghad Zuhair Yousif, Shadan Mohammed Jihad Abdalwahid, 2020) and (K. Sekar and M. Padmavathamma,, 2016) found that

Authentication Level Issues: a collection of requirements for functional and non-functional system models, a timeline for the development of the system, and a compendium of expressions used throughout the scheme. There are a large number of clusters and nodes that are present. Each node possesses a different set of priorities or rights than the others (Ghiasvand, et al., 2022). Administrative nodes have complete access to all data. However, if a malicious node has organizational priority over another, it may steal or manipulate critical user data. A large number of nodes are joining clusters to benefit from faster execution through parallel processing. In the absence of authentication, any malicious node can cause havoc on the cluster. In the world of big data, logging is significant. If no logging is provided, no activity involving the modification or deletion of data will be recorded. If a new node is added to the cluster and there is no logging, the new node will not be recognized (Tang, 2020).

Issues at the Data Level: Data is the most important component of big data, as well as the most difficult to work with. Even though data is nothing, some government or social networking websites may contain sensitive and personal information about individuals. The integrity and availability of data, as well as the protection and distribution of data, are the primary issues that the Data level could address. To

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

increase efficiency, big data environments such as Hadoop store data without encrypting it. If a hacker gains access to the machines, he or she will be unable to be stopped. To provide quick access to information across a large number of nodes, a distributed data store is used. While it is possible to recover data from a node if the hacker deletes or manipulates a replica, it is much more difficult to recover data from another node (KAREEM, 2020).

Network-Level Issues: Clusters consist of a large number of nodes, each of which is responsible for computing or processing data on the cluster's behalf. It is possible to perform this data processing anywhere in the cluster, regardless of which node is being used. The identification of which node data is being processed becomes difficult as a result of these limitations. As a result of this difficulty, it will be more difficult to determine how to ensure node safety (Amin Salih Mohammed, Shahab Wahhab Kareem, Ahmed khazal al azzawi and M. Sivaram, 2018) (Sami H. Ismael, Shahab Wahhab Kareem, Firas H. Almukhtar, 2020) (Sardar M. R. K Al-Jumur, Shahab Wahhab Kareem, Raghad z.yousif, 2021) (Shahab Kareem, Mehmet C Okur, 2018). Communication and sharing of data and resources between two or more nodes are possible through the network. Network communication is frequently facilitated through the use of RPC (Remote Procedure Call) calls. However, RPC is secure, until and unless it is encrypted (Roojwan S. Ismael, Rami S. Youail, Shahab Wahhab Kareem, 2014) (Hadeer Mahmoud, Abdelfatah Hegazy, and Mohamed H. Khafagy, 2018).

The use of many technologies in the big data environment to process data for some traditional security tools for security purposes is also a concern at the general level. Standard tools have been developed over many years. As a result, these tools may not perform well when dealing with new distributed forms of big data. To store and process large amounts of data, various data processing, data recovery, and data storage technologies are employed (Raghad Z.Yousif, Shahab.Wahhab Kareem, Ammar. O.Hasan, 2016). As a result of the many different technologies used to store and process big data, there may be some complexities as a result of these other technologies (Shahab Wahhab Kareem, Yahya Tareq Hussein, 2017) (Shahab Kareem, Mehmet C Okur, 2018) (K. Sekar and M. Padmavathamma,, 2016) (P. Merla and Y.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Liang, 2017) (Shahab Wahhab Kareem and Mehmet Cudi Okur, 2020) (Zahari Mahad and Muhammad Rezal Kamel Ariffin, 2015).

4. Bigdata And Hadoop

When storing data, the Hadoop distributed file system (HDFS) is utilized, and when analyzing data, MapReduce is utilized. These are the two most important components of the Hadoop architecture (Abdiaziz Omar Hassan, Abdulkadir Abdulahi Hasan, 2021). Hadoop was formed in 2005 by Doug Cutting, a Yahoo employee for Nutch Search Engine Project. In the Hadoop distributed file system, which has a default block size of 64 megabytes, HDFS (Hadoop Distributed File System) can be used for distributed storage of large datasets on a Hadoop cluster. HDFS (Hadoop Distributed File System) is an open-source file management system (Bhardwaj, A., Singh, V. K., Vanraj, & Narayan, Y., December 17-20, 2015 (pp. 1-5)). The term Big data is widely known across the world. To manage such kinds of data, Big data mining methods are designed. These mining methods are undoubtedly efficient in dealing with big data; however, a lack of confidentiality can be found in such approaches. As such a massive quantity of data can have certain sensitive information that requires privacy. To preserve privacy while mining big data, the 'Hadoop framework' is used. HDFS operates the file management system for utilizing as the distributed storage of large datasets on a Hadoop cluster (After storing the input files in HDFS, it calculates the results using a program identified by MapReduce, and the results are eventually moved to the output folder of HDFS after being processed (Dubey, A. K., Jain, V., & Mittal, A. P, 2015). Hadoop MapReduce is the software form created to process massive volumes of data sets over a large number of computing resources (Seema Maitrey and C.K. Jha, 2015). MapReduce is the fundamental distribution scheme used by the Hadoop engine to distribute work across a cluster of computers. When processing enormous amounts of data in parallel on large hardware clusters in a suitable and nearly error-free manner, the input data, which resides in the cluster on a distributed file system, is divided into groups of equal size. This is designed to enable and optimize the enormous amounts of data collection in parallel on large clusters of hardware. A Hadoop cluster balances computation volume, storage volume, and I/O bandwidth by merely adding commodity servers. Hadoop groups at Yahoo! span

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

40,000 servers, and stock 40 petabytes of application data, with the biggest cluster being 4000 servers. Globally, several organizations confirmed using Hadoop. This recommends that the Hadoop framework is the most efficient to process and store big data. A large volume of informational collections is transformed into organized key-esteem matches and provided as data sources, as specified by the name of the MapReduce framework in Hadoop; the first stage is the map, and the second stage is the reduce. For example, a large volume of informational collections is transformed into organized key-esteem matches and provided as data sources (Masoumeh RezaeiJam, Leili Mohammad Khanli, Mohammad Kazem Akbari, and Morteza Sargolzaei Javan, 2014). The MapReduce calculation information stream is depicted in Figure 1. show though the mapper does not legitimately compose to the memory block, it takes advantage of the compositions being buffered. Every mapper has a round memory cushion with a default size of 100 MB that can be changed by changing the property of the round memory cushion (io. sort. MB). Hadoop comprises various elements, out of which the most fundamental is the Hadoop Distributed File System (HDFS) which is located at the bottom and saves the files among the storage nodes in the Hadoop cluster. The other one is the Map-Reduce engine built on top of the HDFS and scrutinizes the big data. It's an exceptionally astute flushing device. When the cushion has been filled to a specific level, it begins to leak the support substance into the circle around it. Before the spill occurs in the process, the string allotments the information to the reducers. It must go through the foundation string to perform a kind of in-memory operation inside the key-based parcel before the spill occurs in the block. The yield in the memory sort is increased if a combiner is available, which is the case in most cases (Seema Maitrey and C.K. Jha, 2015).

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

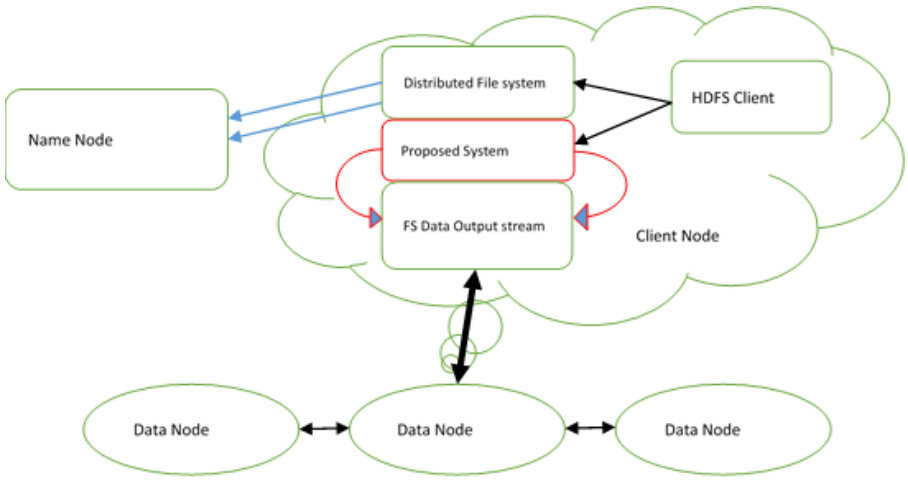

Figure (1): Data Flow of Mapreduce computation [26]

5. Proposed Algorithms

Hadoop is the leading provider of large-scale cloud processing and warehouse services, so it must employ cutting-edge encryption techniques to ensure protection. The strategy is based on cascading two public-key cryptosystems on top of each other (Rabin RZ). Hybridization is a concept that was developed to overcome the limitations of practicing a specific cryptosystem alone while also extending protection. It is anticipated that each of the records corresponding to HDFS will be ciphered before being stored in the database. The HDFS customer qualifies for key generational upgrades (public and personal keys). The recommended hybrid method is used to cipher the file while simultaneously buffering it to HDFS to simulate an unstructured file. The HDFS begins to transmit the cipher files to the information nodes, the first step in the encryption process. Figure 2 depicts the steps that must be taken. High-Density File System (HDFS) is composed of a reputation Node that stores Metadata and controls the namespace of the filing system and manages clients' access to specific cyphering files. The cyphering data is created from one or more segments that have been collected in a very large collection of knowledge nodes. The hybrid method that is recommended is illustrated in figure 3. According to Figure 3, the model for creating keys (both public and private) is based on the mechanism used by the RSA. The ability to factor large numbers could impact the level of protection

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

provided by the RSA algorithm. Algorithm1 is a representation of the cryptographic system.

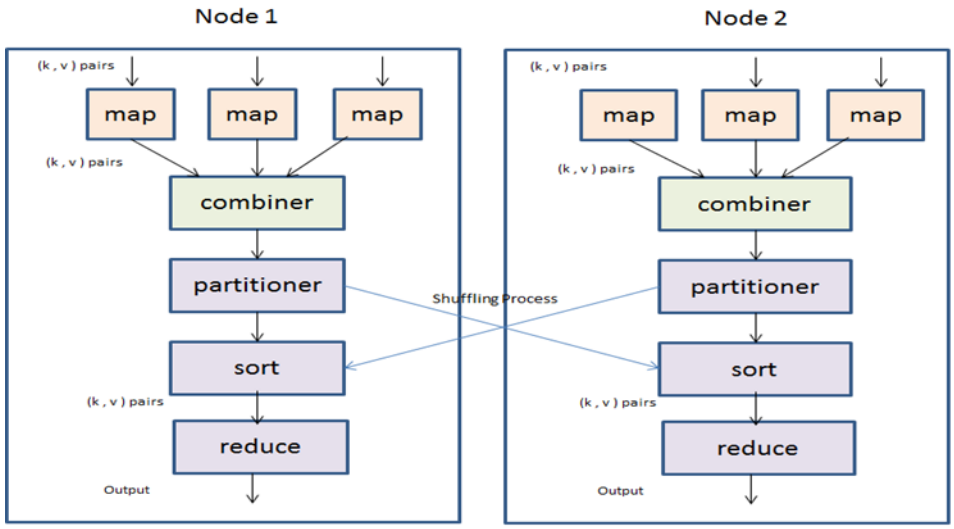

Figure (2): Encryption procedure in HDFS (Shadan Mohammed Jihad ABDALWAHID, Raghad Zuhair YOUSIF, Shahab Wahhab KAREEM, 2019)

Algorithm 1: - The proposed algorithm's key generation procedure
The input must be a prime number of size n bits.
OUTPUT: The N=p.q public key is returned.
The message was sent between user A and user B.

To solve this problem, generate two primes, each approximately the same size and satisfying (p and q are 3 mod 4; 2np2n+1). Calculate N, which is equal to the product of p and q multiplied together (N=p*q).
The encryption process is also based primarily on the Rabin encryption procedure, but the clear text or message (m) is raised to power 2e rather than e(where N is the public key), as is the case in the Rabin cryptosystem, and the ciphertext c is calculated as follows:

$$C = M^2 \ mod \ N \qquad (1)$$

The following sections provide a detailed explanation of the encryption procedures:

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Algorithm2: -The proposed algorithm's encryption process is described in detail.
INPUT: Encoding the plaintext and obtaining the user's public key (N=p*q) are both performed.

OUTPUT: The cipher-text C has been encoded.
User A sends a message to user B, who then receives it.
User B is required to complete the following steps to encode the data:-

(a) Find out what A's true public key is (N=p q).

(b) Explain how information such as the number m in the period [M 22n,2n-1 Z pq] should be interpreted.

(c) Determine the value of C = (M2) module N.

(d) Copy the encoding text to the variable A.

According to (2), the decoding process emulates the Rabin decryption scenario in which the cipher text is raised to the power of the private key (d). However, the output is not the direct message (M), but the message raised to power 2 as described in (1).

VC (mod d) is an abbreviation for Visual Communication (2)

To retrieve the message M first solves the V using the Chinese Remainder Theorem (CRT) and the pair (p,q) and then returns four messages, which are designated as M1, M2, M3, and M4 respectively. The following step is to loop through all of the messages and compute ($W_i = \frac{C-M_i^2}{N}$), and then return Mi, which is W_i^Z. In (Algorithm3), the following procedure is described in greater detail:

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

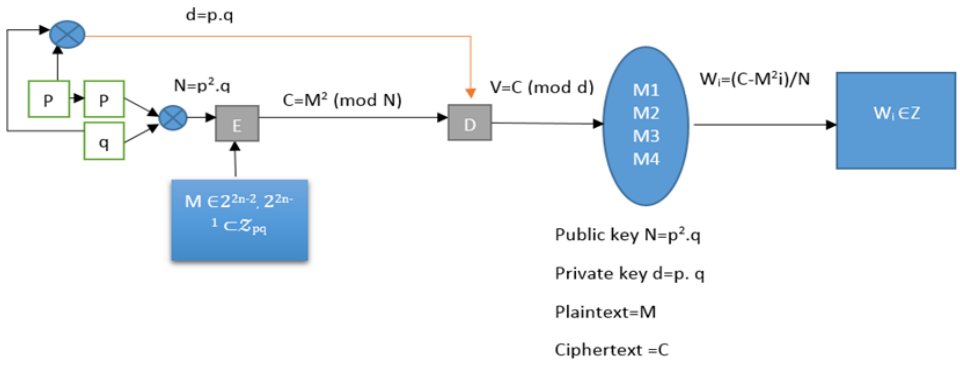

Figure (3): Process of the proposed public-key algorithm.

Decryption process of the proposed algorithm (Algorithm3)

NPUT: The encoded text was accepted, and the private key was also accepted (d, p, q).

OUTPUT: Convert the encoded text to plain text using the M character encoding.

The following steps are required by B to recover the original text M from the decoding text:

(a) Determine V=C mod d by using the formula V=C mod d

Calculate the square root of V through the CRT by employing (p, q)

(c) Pay attention to the four possible messages M1, M2, M3, and M4.

(d) Calculate ($W_i = \frac{C-M_i^2}{N}$), in a loop from (1 to 4).

(e) Return the plain text Mi that produces the result Wi Z. (Zahari Mahad and Muhammad Rezal Kamel Ariffin, 2015).

The mathematical example of the Rabin RZ is below: The scenario is A (Bob) will send his public key to B (Bob) and Bob will encrypt to Along. Along will choose the primes p=100669, q=69859 and compute N=707968400363899 and d=7032635671. Let says Bob want to sends a message M=1439948310 to Alice. Bob will compute, and compute and. Let says Bob want to sends a message to Alice. Bob will compute

519659206359828≡14399483102(mod 70796840036899)

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

And send to Alice. To decrypt, Along computes

Then, Alice uses the CRT and his private keys to compute the four square roots of 3691358296 modulo d, which are

M1=3890433108 , M2=1439948310 , M3=5592687361 , M4=3142202563
Then, to determine the correct message Along computes for i=1 to 4:
Wi=C-Mi2/N

In this example only M2 produce W∈Z.

Following the proposed approach coding scheme, the data nodes are effective in segmenting production, replication, and deletion based on instructions from the name node. In addition, information about the client is obtained from a variety of sources and scrambled on the server by utilizing asymmetric mechanism cryptography instruments. Soon after encoding, information is stored in the cloud, specifically via the Hadoop File System (HDFS), where it will be accessible by a cluster of computers. Whenever a user requests information, the server will provide the encrypted information so that it can be decrypted. The user then uses the corresponding keys to retrieve the decrypted data, which is accomplished through the use of a hybrid approach, as proposed in this paper. Figure 3 depicts a flowchart of the data encryption process, which includes several steps.

6. Experimental Results and Analysis

When evaluating the performance of encrypted HDFS, the MapReduce and HDFS frameworks are utilized. Each node is equipped with an i5 processor, 16 GB of memory, and a 1 TB hard drive. When processing enormous amounts of data in parallel on large clusters of hardware in a suitable and nearly error-free manner, the input file, which resides in the cluster on a distributed file system, is divided into groups of equal size to facilitate and simplify the enormous amounts of data processing in parallel on large clusters of hardware. Duration of encryption: When using the RAS alone or the proposed method, the time required to decrypt the Hadoop separation dataset files and convert them to plaintext using the private key

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

is measured in minutes and seconds. It is measured in terms of seconds. In this case, essentially, the Encryption Time is equal to the system's current time before encryption lets the system's modern time after encryption, which results in the encryption time being given. The Decryption Time, on the other hand, is analogous to the system current time before decryption subtracted from it by the system current time after decryption, as shown in the example below. Using different file sizes, the ramifications of the association between encryption schemes, such as those used by RSA and Paillier, as well as the proposed system, were depicted in Figure 4. When comparing the proposed method to the RSA and Paillier methods for file sizes ranging from 100 MB to 1 GB with a step size of 100 MB or greater, it is clear that the proposed method demonstrated efficient time consumption when compared to the RSA and Paillier methods. Furthermore, the proposed method for the encryption stage is significantly faster than the default RSA and Paillier methods. RSA, Paillier, and the proposed method all have a period during the decryption stage, as shown in Figure 5. The encrypted files that have been applied to the current stage are of a variety of different sizes. The proposed ciphered method requires less decryption time than the other methods when all of the previously mentioned systems (including the proposed method) are used. Table 2 compares the computational complexity of the proposed method with the RSA and Paillier cryptosystems, and it is clear that the proposed method has a lower computational complexity when compared to the individual cryptosystems (RSA or Rabin).

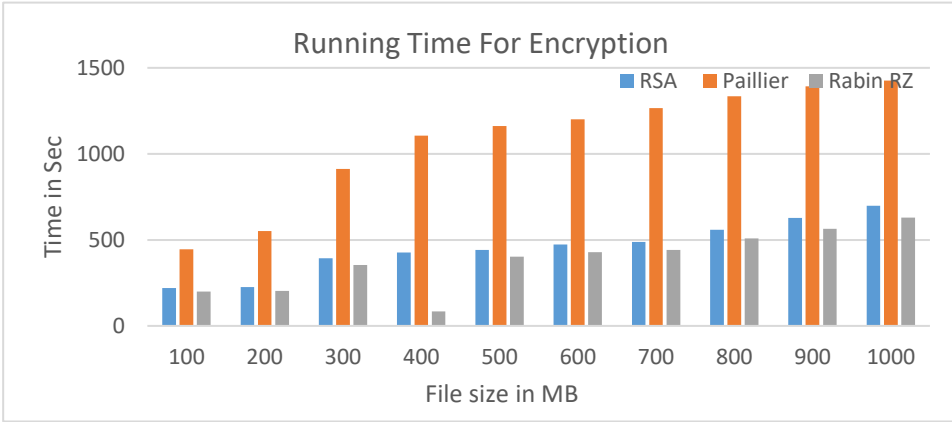

Figure (4): Running time in second for Encryption Process.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Table 2. Computational Complexity of the Proposed Method, RSA, and Rabin.

Method	Encryption	Decryption
RSA	$T(c) = O(\log n)^3$	$T(M) = O(\log n)^3$
Rabin	$T(c) = O(5n^2+2n)$	$T(M) = O(2n^3+12n^2+2n)^3$
Rabin-RZ	$T(c) = O(4n^2++3n)$	$T(M) = O(2n^3+12n^2+4n)^3$

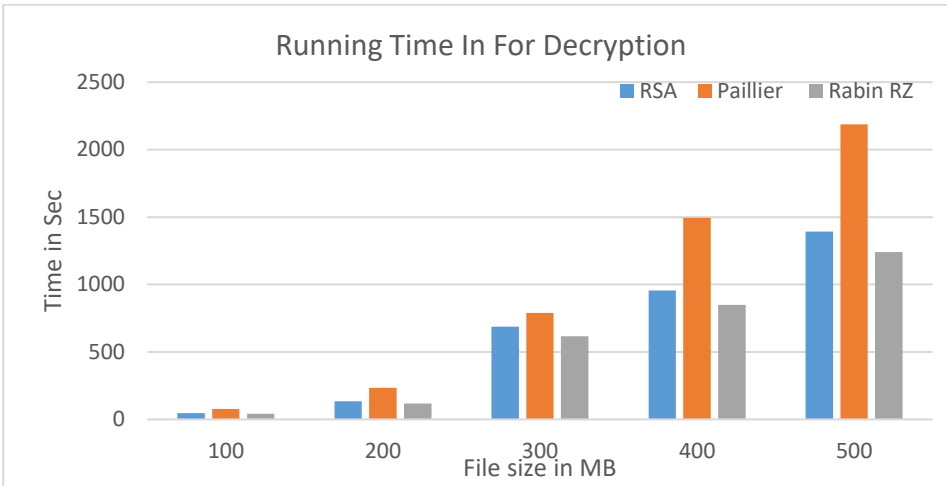

Figure (5): Running time in second for Decryption Process.

7. Conclusions

While Hadoop helps us overcome the challenges presented by big data in businesses and organizations, it is not a data protection tool. Information collected in Hadoop may be compromised by an intruder or snooper. The veracity of information is constantly in question. Before storing the information in HDFS, the proposed Hybrid cipher asymmetric key algorithm encrypts the content of the file by protecting it from various network intrusions, as demonstrated in the following figure. It is now possible to save files or data in Hadoop without having to be concerned about security issues because the encryption method has been applied to the files before they are collected in Hadoop. In addition to the most popular cloud computing service models such as Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service, the proposed system includes the following additional features: (PaaS). It also

helps with information management and safety concerns (Verification, Credibility, Ease of access, and Secrecy) in protection, as well as an authentication scheme for data transfer. The proposed encryption method is much quicker than the usual RSA and Paillier methods. During the decryption step, RSA, Paillier, and the proposed technique all have a period. When all of the previously listed systems (including the proposed approach) are applied, the proposed ciphered method takes less time to decrypt than the other ways. Despite its greater complexity, the proposed method showed outstanding performance when applied to a wide range of file sizes during the encryption and decryption stages (double the computational complexity in the decryption stages). The integration of the RSA asymmetric key cryptosystem would be a longer-term task. Continuing the effects of the complex system study, it was discovered that the encryption and decryption rates for encryption and decryption were significantly higher than the average. The proposed method has the potential to produce an excellent principle for the purpose, which relies on fast encryption and decryption blocks to accomplish its goals. The decryption breakdown of the Rabin cryptosystem has been accomplished more efficiently than previously achieved by other systems.

References

Abdiaziz Omar Hassan, Abdulkadir Abdulahi Hasan. (2021). Simplified Data Processing for Large Cluster: A MapReduce and Hadoop Based Study. *Advances in Applied Sciences, 6*(3), 43-48.

Amin Salih Mohammed, Shahab Wahhab Kareem, Ahmed khazal al azzawi and M. Sivaram. (2018, 10 12). Time Series Prediction Using SRE- NAR and SRE- ADALINE. *Jour of Adv Research in Dynamical & Control Systems*.

Bhardwaj, A., Singh, V. K., Vanraj, & Narayan, Y. (December 17-20, 2015 (pp. 1-5)). Bhardwaj, A., Singh, V. K., Vanraj, & Analyzing BigData with Hadoop cluster in HDInsight azure Cloud. *Annual IEEE India Conference (INDICON).* New Delhi, India.

Cao, Y., Dhahad, H. A., Alsharif, S., Sharma, K., El-Shafay, A. S., & Kh, T. I. (2022). Development of a MSW-fueled sustainable co-generation of hydrogen and electricity plant for a better environment comparing PEM and alkaline electrolyzers. *Sustainable Cities and Society*, *81*, 103801.

Chen, J. (2021). Encryption Algorithm of Automatic Single Chip Computer and Embedded System Based on Big Data. *Journal of Physics: Conference Series , 2037*.

Dubey, A. K., Jain, V., & Mittal, A. P. (2015). Stock Market Prediction using Hadoop Map-Reduce Ecosystem. *2nd International Conference on Computing for Sustainable Global Development* (hal. 616-621). IEEE.

Galal A. Al-Rummana, Abdulrazzaq H. A. Al-Ahdal, G. N. Shinde. (2021). Faster Big Data Encryption Technique Using Key Generation. *International Journal of Advanced Trends in Computer Science and Engineering , 3*(10), 1776 - 1783.

Ghiasvand, A., Noori, S. M., Suksatan, W., Tomków, J., Memon, S., & Derazkola, H. A. (2022). Effect of Tool Positioning Factors on the Strength of Dissimilar Friction Stir Welded Joints of AA7075-T6 and AA6061-T6. *Materials*, *15*(7), 2463.

Hadeer Mahmoud, Abdelfatah Hegazy, and Mohamed H. Khafagy. (2018). An approach for Big Data Security based on Hadoop Distributed File system. *International Conference on Innovative Trends in Computer Engineering (ITCE 2018).* Aswan : Aswan University.

Hsiao, F.-H. (2018). Chaotic synchronization cryptosystems combined with RSA encryption algorithm. *Fuzzy Sets and Systems, 342*, 109-137.

J. Wu, J. Shi, and T. Li. (2020). A novel image encryption approach based on a hyperchaotic system, pixel-level filtering with variable kernels, and DNA-level diffusion. *Entropy,, 22*(5).

K. Sekar and M. Padmavathamma,. (2016). Comparative study of encryption algorithm over big data in cloud systems. *3rd International Conference on Computing for Sustainable Global Development (INDIACom),* (hal. 525 - 528). IEEE .

KAREEM, S. W. (2020). Secure Cloud Approach Based on Okamoto-Uchiyama Cryptosystem. *Journal of Applied Computer Science & Mathematics, 14*(1), 9-13.

M. Elsayed and M. Zulkernine,. (2018). Towards Security Monitoring for Cloud Analytic Applications. *IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity),*, *6*, hal. 45-56. Omaha, NE, USA.

Masoumeh RezaeiJam, Leili Mohammad Khanli, Mohammad Kazem Akbari, and Morteza Sargolzaei Javan. (2014). A Survey on Security of Hadoop. *4th International Conference on Computer and Knowledge Engineering (ICCKE)* (hal. 716-721). IEEE.

Mohammed, A. G. (2021). A Study of Scheduling Algorithms in LTE-Advanced HetNet. *QALAAI ZANIST JOURNAL*, *6*(3), 945-968.

P. Merla and Y. Liang. (2017). Data analysis using hadoop MapReduce environment. *IEEE International Conference on Big Data (Big Data)* (hal. 4783-4785). Boston: IEEE.

Priyank Jain, Manasi Gyanchandani and Nilay Khare. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data, 25*(3).

Poongodi, M., Bourouis, S., Ahmed, A. N., Vijayaragavan, M., Venkatesan, K. G. S., Alhakami, W., & Hamdi, M. (2022). A Novel Secured Multi-Access Edge Computing based VANET with Neuro fuzzy systems based Blockchain Framework. *Computer Communications*.

Raghad Z. Yousif1, Shahab W. Kareem, Shadan M. Abdalwahid. (2020). Enhancing Approach for Information Security in Hadoop. *Polytechnic Journal. 2020., 10*(1), 808.

Raghad Z.Yousif, Shahab.Wahhab Kareem, Ammar. O.Hasan. (2016). Design Security System Based on AES and MD5 for Smart Card. *charmo university.* Sulaimanyia: Charmo university.

Riguang Lin and Sheng Li . (2021). An Image Encryption Scheme Based on Lorenz Hyperchaotic System and RSA Algorithm. *Security and Communication Networks , 2021*.

Roojwan S. Ismael, Rami S. Youail, Shahab Wahhab Kareem. (2014). Image Encryption by Using RC4 Algorithm. *EUROPEAN ACADEMIC RESEARCH, Vol. II*(Issue 4), 5833-5839.

Saliha, R. K. (2022). Optimizing RSA cryptosystem using Hermite polynomials. *Int. J. Nonlinear Anal. Appl., 13*(1), 955-961.

Sami H. Ismael, Shahab Wahhab Kareem, Firas H. Almukhtar. (2020, 14 1). Medical Image Classification Using Different Machine Learning Algorithms. *AL-Rafidain Journal of Computer Sciences and Mathematics*, hal. 135-147.

Sardar M. R. K Al-Jumur, Shahab Wahhab Kareem, Raghad z.yousif. (2021, 22 2). Predicting temperature of erbil city applying deep learning and neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, hal. 944-952.

Seema Maitrey and C.K. Jha. (2015). MapReduce: Simplified Data Analysis of Big Data. *Procedia Computer Science, 57*(1), 563-571.

Shadan Mohammed Jihad Abdalwahid, Raghad Zuhair Yousif, Shahab Wahhab Kareem. (2019, 15). Enhancing approach using hybrid pailler and RSA for information security in bigdata. *Applied Computer Science*.

Shadan Mohammed Jihad ABDALWAHID, Raghad Zuhair YOUSIF, Shahab Wahhab KAREEM. (2019). ENHANCING APPROACH USING HYBRID PAILLER AND RSA FOR INFORMATION SECURITY IN BIGDATA. *Applied Computer Science, 15*(4), 63-74.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

shadan MohammedJihad abdalwahid Shahab Wahhab Kareem, Raghad Zuhair Yousif. (2020, 20 3). An approach for enhancing data confidentiality in Hadoop. *Indonesian Journal of Electrical Engineering and Computer Science*, hal. 1547-1555.

Shahab Kareem, Mehmet C Okur. (2018). Bayesian Network Structure Learning Using Hybrid Bee Optimization and Greedy Search. *Çukurova University.* Adana, Turkey.

Shahab Wahhab Kareem and Mehmet Cudi Okur. (2020). Evaluation of Bayesian Network Structure Learning Using Elephant Swarm Water Search Algorithm. Dalam S. C. Shi, *Handbook of Research on Advancements of Swarm Intelligence Algorithms for Solving Real-World Problems* (hal. 139-159). Chapter 8: IGI Global.

Shahab Wahhab Kareem, Raghad Zuhair Yousif, Shadan Mohammed Jihad Abdalwahid. (2020, 20 3). An approach for enhancing data confidentiality in hadoop. *Indonesian Journal of Electrical Engineering and Computer Science*, hal. pp. 1547~1555.

Shahab Wahhab Kareem, Yahya Tareq Hussein. (2017). Survey and New Security methodology of Routing Protocol in AD-Hoc Network. *QALAAI ZANIST JOURNAL.* Erbil.

Sharma, A., Miran, A., & Ahmed, Z. R. (2022). The 3D Facemask Recognition: Minimization for Spreading COVID-19 and Enhance Security. In *ICT Analysis and Applications* (pp. 619-637). Springer, Singapore.

Tang, Z. (2020). A Preliminary Study on Data Security Technology in Big Data Cloud Computing Environment. *International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE* (hal. 437–442). Bangkok, Thailand: EIDWT .

Zahari Mahad and Muhammad Rezal Kamel Ariffin. (2015). Rabin-RZ: A New Efficient Method to Overcome Rabin Cryptosystem Decryption Failure Problem. *International Journal of Cryptology Research, 5*(1), 11-20.

شێوازێکی کارای نوێ بۆ پاراستنی زانیاری لە هادوپ.

پەرەسەندنی هەدۆپێکی دیاریکراو بۆ کۆمپیوتەری هەوری تەحەدایەکی سەرەکیە، بۆ ئەوەی ئەم تەحەدایە بەدی بهێنیت، سیستەمێکی ژمێریاری هەوری پارێزراوی پێشبینیکراو پەرەی پێدرا. کۆمەڵێک هاوکاری هەور هەیە کە دەتوانرێت بەکار بهێنرێت بۆ جێبەجێکردنی سیستەمی پاراستن، یان خزمەتگوزاری جۆراوجۆر، لەوانە سۆفتوێر وەک خزمەتگوزاری (SaaS)، ژێرخان وەک خزمەتگوزاری و پلاتفۆرم وەک خزمەتگوزاری. لەم ناوچەیەدا بۆ دروستکردن و باشترکردنی ئاسایشی بەڕێوەبردن و کۆکردنەوەی داتا بەکارهێنەران بەکارهاتووە و پێی دەوترێت هادۆپ ئامرازەکانی تر کە لەلایەن

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (7), No (2), Summer 2022
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Apacheبەرهەم دەهێنرێت بۆ زاڵبوون بەسەر تەحەداکانی داتای گەورەدا، هادوپ دەگرێتەوە، کە بەکارهێنانی نژیاروانیی Map-Reduce دەکات بۆ ئامادەکردنی بڕێکی زۆر لە داتاکان. هادوپ یەکێکە لە گرنگترین ئامرازەکان بۆ زاڵبوون بەسەر تەحەداکانی داتای گەورە. ئەمە یەکێکە لە قورسترین بەرەنگارییەکان بۆ دڵنیابوون لە ئاسایشی کۆگای داتا، وە سیستەمی فایلی دابەشکراوی هادوپ (HDFS) ستراتیژیەکی ئەمنی بە ڕوونی پێناسە نەکراوە. ستراتیجی پێشنیارکراو بەکارهێنانی کلیلی گشتی نهێنی دەکات بۆ ئەوەی هەموو ئەو فایلانەی لە HDFS خەزن کراون بۆ مەبەستی پاراستنی هەموویان. لە کۆکراوەی داتادا، داتا کۆکراوەکان لە HDFS کۆکراوەتەوە بە بەکارهێنانی شێوازی پێشنیارکراو (Rabin RZ) لە ڕەمزاندنی داتا. بەکارهێنەرانی هەور لەوانەیە دوو بژاردەیان هەبێت کاتێک دێت بۆ بارکردنی داتا بۆ هەور: بارکردنی پارێزراو و بارکردنی نائەمن. بارکردنی پارێزراو بژاردەی نموونەییە. شێوازی پێشنیارکراو بە پێی شێوازی پایلیەر و سیستەمی گریمانەیی کریپتۆسی ئار ئێس ئەی، بە هەندێک بەراورد دەکرێت. کاتێک بە بەراورد لەگەڵ سیستەمی کریپتۆسیستەمەکانی تر، شێوازی پێشنیارکراو ئاڵۆزییەکی ژمێریاریی بەهێزتری هەیە و لە جێگرەوەکان پێچانەوەیەکی بچووکتری هەیە.

طريقة جديدة فعالة لأمن المعلومات فيHadoop

يعد تطوير Hadoop محددًا للحوسبة السحابية تحديًا أساسيًا ، ومن أجل مواجهة هذا التحدي ، تم تطوير نظام حوسبة سحابية آمن متصور. هناك مجموعة متنوعة من عمليات التعاون السحابية التي يمكن استخدامها لتنفيذ نظام الأمان ، أو مجموعة متنوعة من الخدمات ، بما في ذلك البرامج كخدمة (SaaS) والبنية التحتية كخدمة ومنصة كخدمة. تم استخدامه في هذا المجال لبناء وتحسين أمان إدارة وجمع البيانات من المستخدمين ، وكان يسمى .Hadoop تشمل الأدوات الأخرى التي أنتجتها Apache للتغلب على تحديات البيانات الضخمة Hadoop، والتي تستخدم بنية Map-Reduce لإعداد كمية كبيرة من البيانات. يعد Hadoop أحد أهم الأدوات للتغلب على تحديات البيانات الضخمة. يعد ضمان أمان تخزين البيانات أحد أصعب التحديات ، ولا يحتوي نظام الملفات الموزعة Hadoop (HDFS) على استراتيجية أمان محددة بوضوح. تستخدم الإستراتيجية الموصى بها تشفير المفتاح العام لتشفير جميع الملفات المخزنة في HDFS بغرض حمايتها جميعًا. في سياق جمع البيانات ، يتم تشفير البيانات المجمعة في HDFS باستخدام الطريقة المقترحة (رابين RZ) لتشفير البيانات. قد يكون لدى مستخدمي السحابة خياران عندما يتعلق الأمر بتحميل البيانات إلى السحابة: التحميل الآمن والتحميل غير الآمن. التحميل الآمن هو الخيار الافتراضي. تتم مقارنة الطريقة المقترحة مع طريقة Paillierونظام تشفير RSA الافتراضي ، على التوالي. عند مقارنتها بأنظمة التشفير الأخرى ، فإن الطريقة المقترحة لها تعقيد حسابي أقوى وزمن انتقال أصغر من البدائل.