ДОДАТОК А

Перелік джерел посилання за науковими напрямами керівника та науковців кафедри програмної інженерії

1. Cherednichenko O., Kyrychenko I., Tereshchenko G., Myand D., Pylypenko S., Comparison of Blockchain–Based Data Storage Systems, COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine, vol. 3688, p. 10, https://ceur-ws.org/Vol-3688/paper10.pdf.

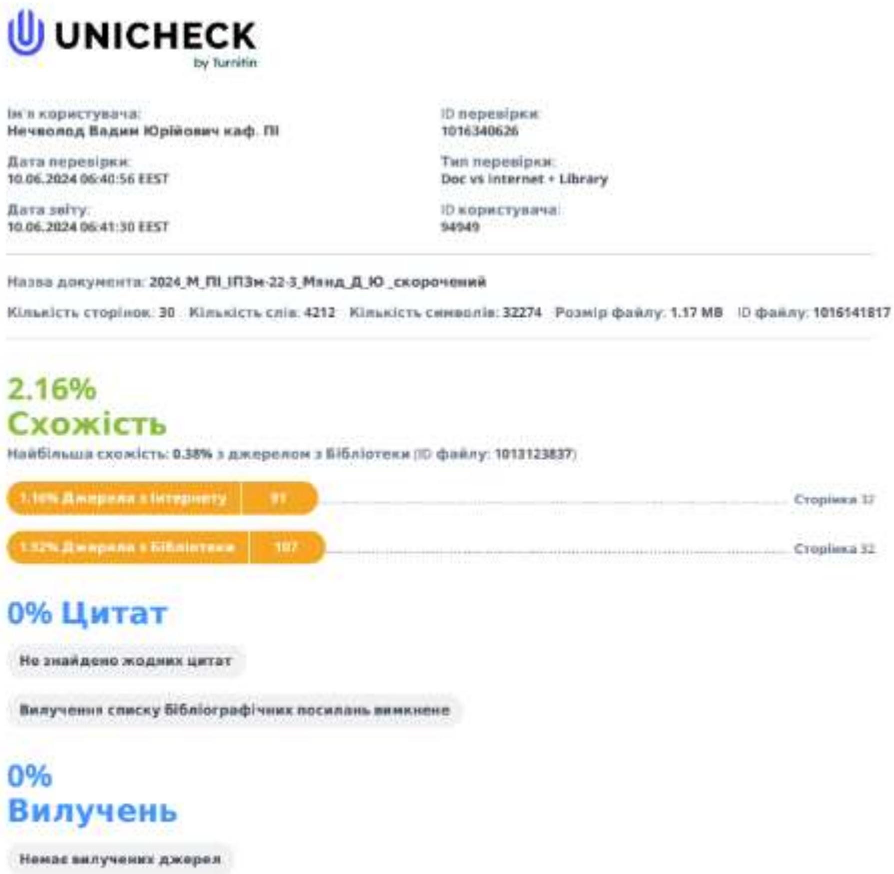
2. Кириченко І. В. Застосування симетричних алгоритмів в блокчейні / І. В. Кириченко, Г. Ю. Терещенко, І. В. Груздо // Бионика интеллекта. – 2020. – № 1(94). – С. 71–75.

3. Kyrychenko I., Shyshlo O., Shanidze N., Minimizing Security Risks and Improving System Reliability in Blockchain Applications: a Testing Method Analysis, CEUR Workshop Proceedings, 2023, 3403, - C423–433, https://ceur-ws.org/Vol-3403/paper33.pdf.

17. Tereshchenko Glib, Chetverykov Grigori, Overview of image storage models in Big Data conditions, Комп’ютерна математика в науці, інженерії та освіті (CMSEE-2020). – Матеріали V Всеукраїнської науково-технічної конференції – Полтава, 27 листопада 2020 р. – С. 18–21. УДК 00.89:004.043, http://kist.ntu.edu.ua/konferencii/27_konf_2020.pdf.

ДОДАТОК Б

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



UNICHECK by Turnitin

Ім'я користувача:
Нечволод Вадим Юрійович каф. ПІ

ID перевірки:
1016340626

Дата перевірки:
10.06.2024 06:40:56 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
10.06.2024 06:41:30 EEST

ID користувача:
94949

Назва документа: 2024_М_ПІ_ІПЗм-22-3_Мянд_Д_Ю _скорочений

Кількість сторінок: 30 Кількість слів: 4212 Кількість символів: 32274 Розмір файлу: 1.17 MB ID файлу: 1016141817

2.16%
Схожість

Найбільша схожість: 0.38% з джерелом з Бібліотеки (ID файлу: 1013123837)

1.16% Джерела з Інтернету 91 Сторінка 32

1.92% Джерела з Бібліотеки 107 Сторінка 32

0% Цитат

Не знайдено жодних цитат

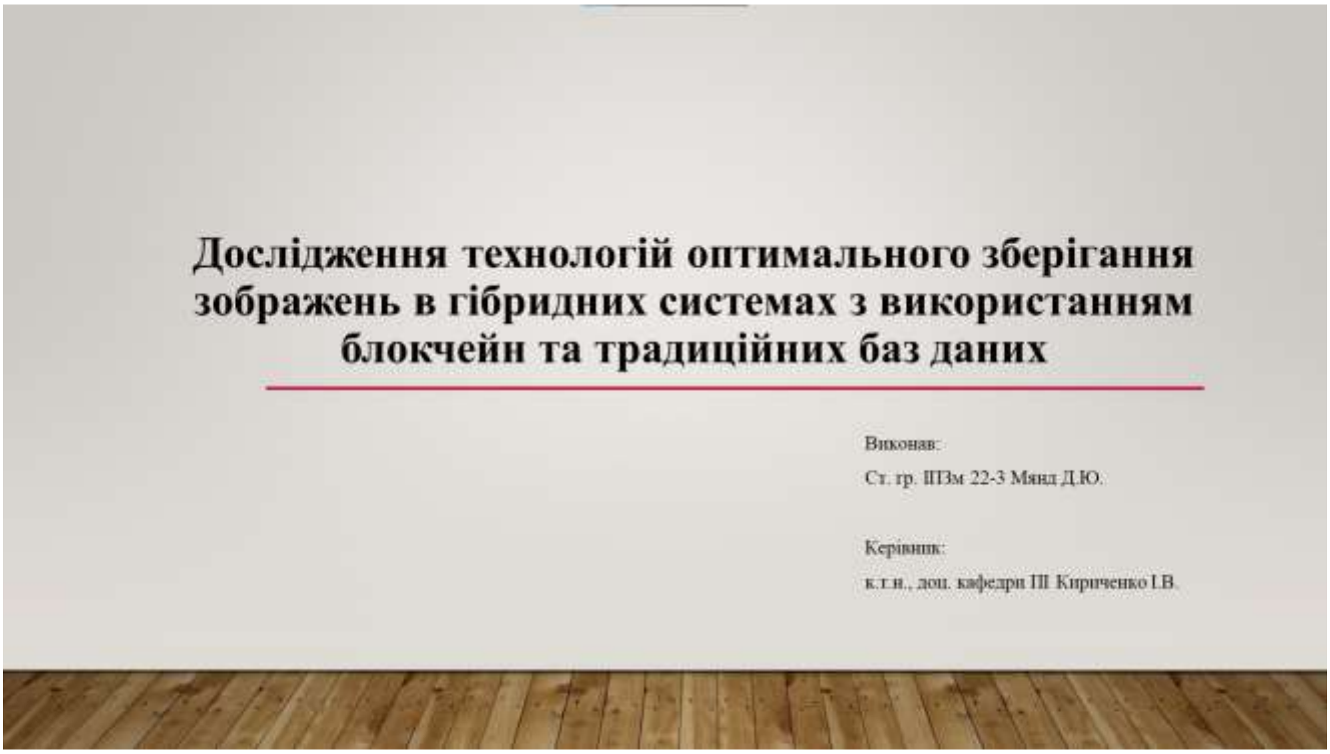
Вилучення списку бібліографічних посилань вимкнене

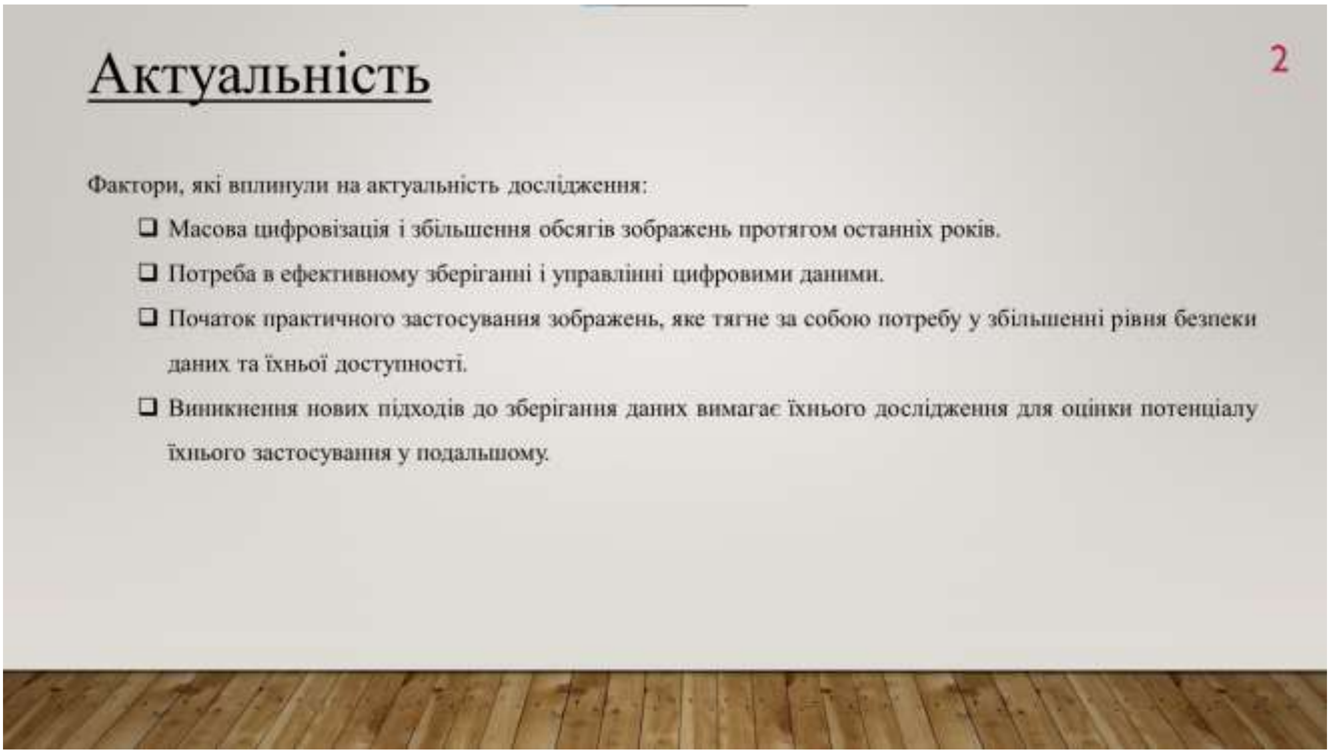
0%
Вилучень

Немає вилучених джерел

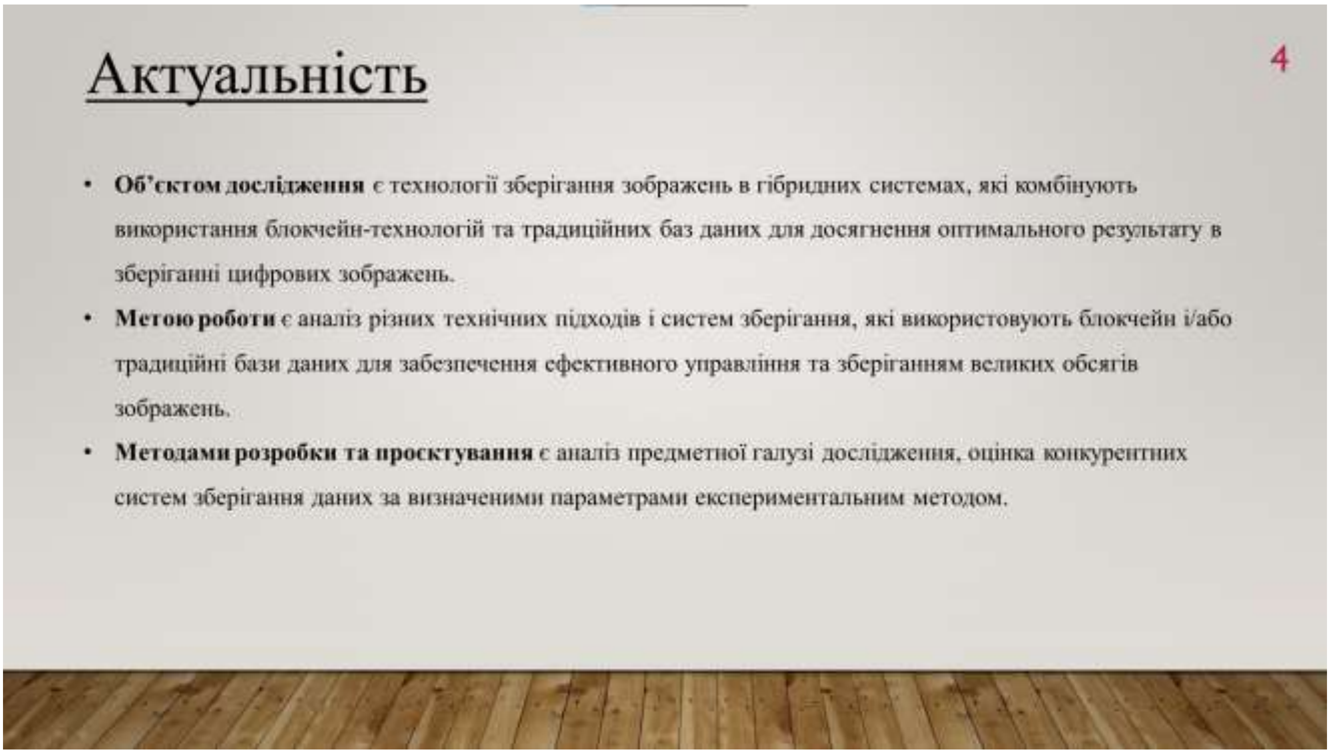
ДОДАТОК В

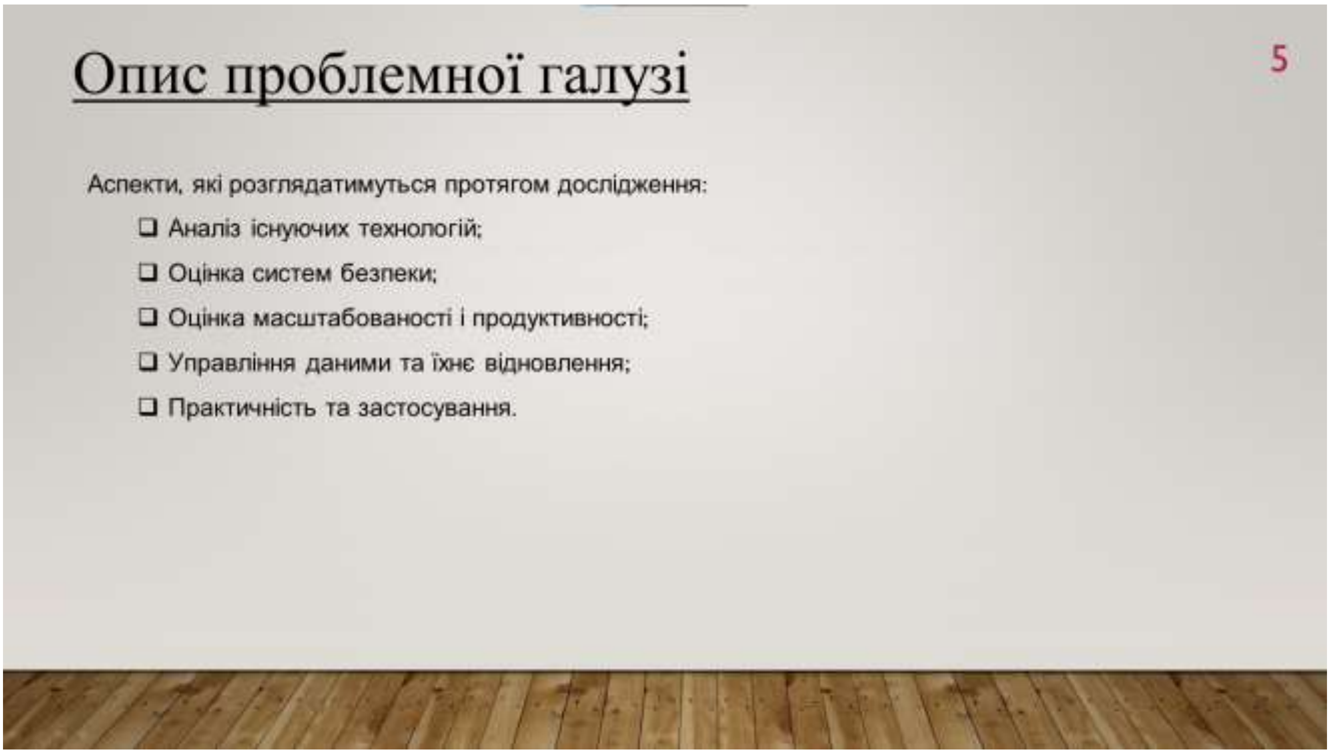
Слайди презентації

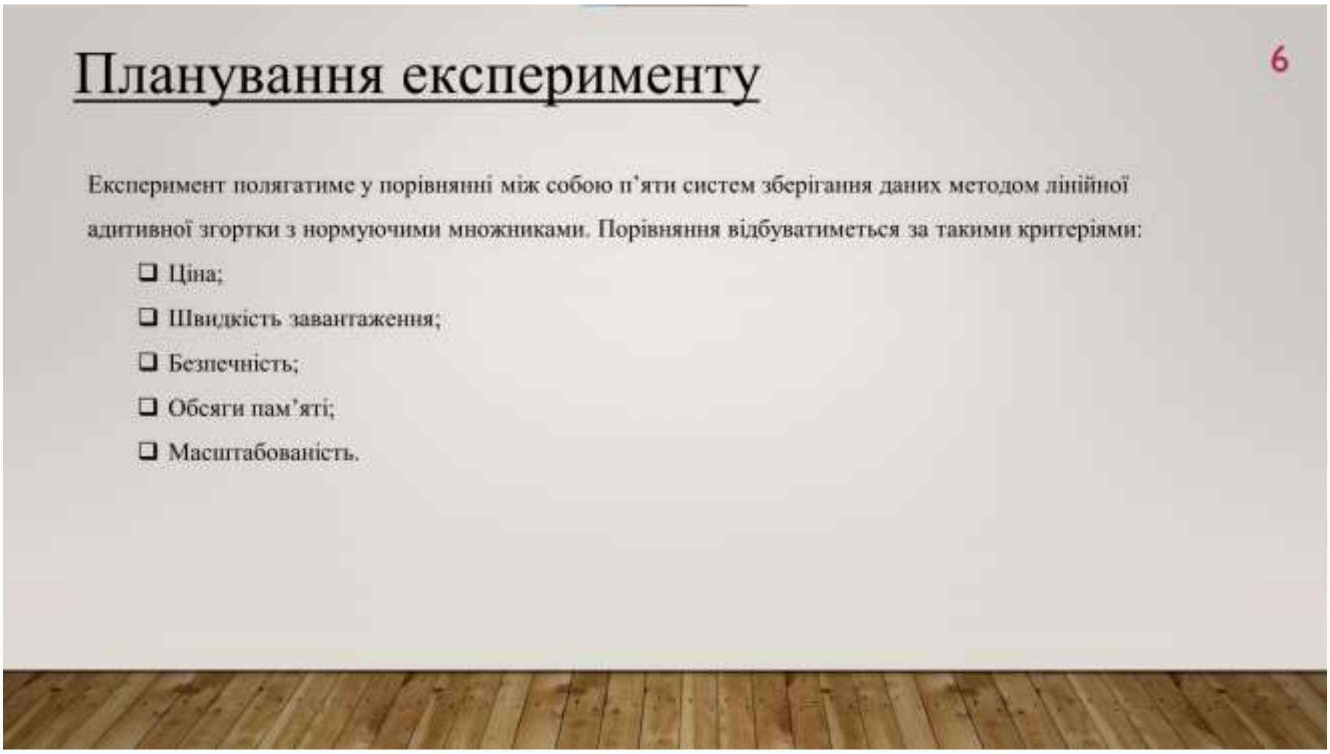


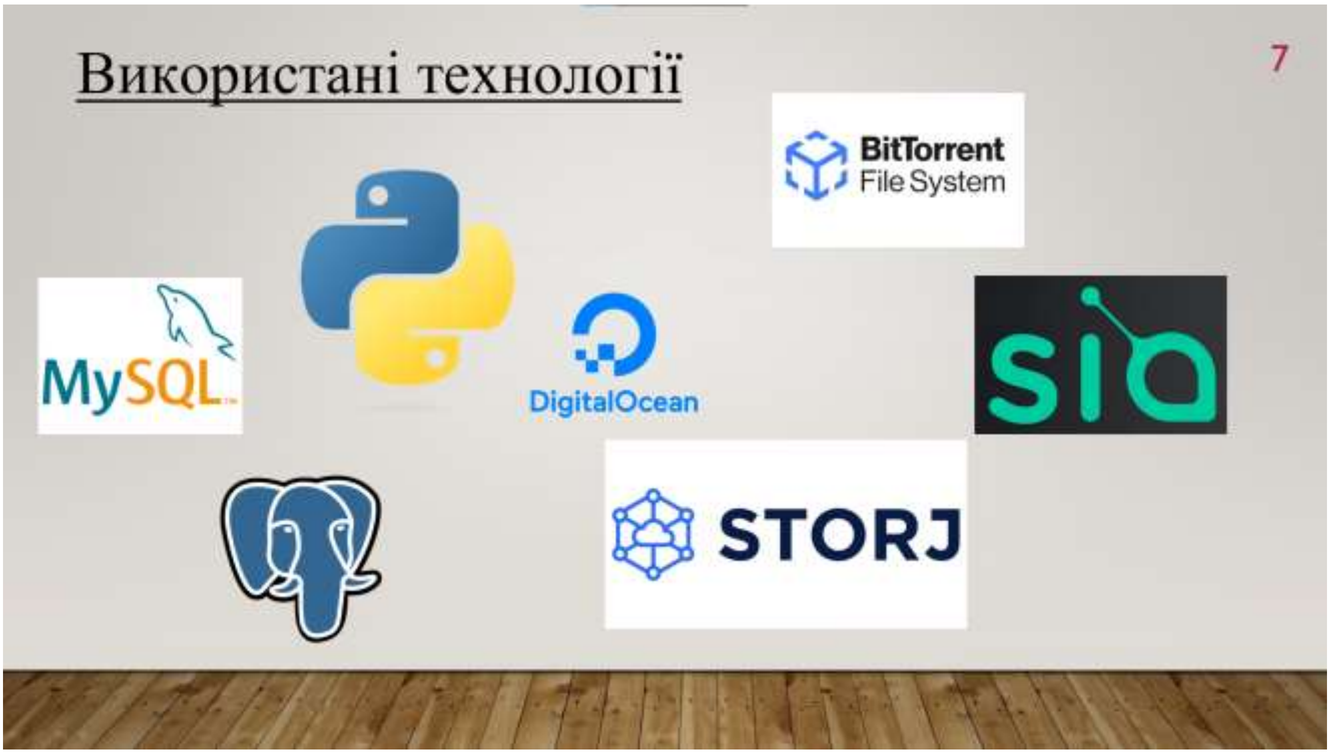



Актуальність
3
Основні проблеми, які роблять дослідження актуальним:
Безпека;
Масштабованість;
Інтеграція;
Цілісність;
Ефективне управління.

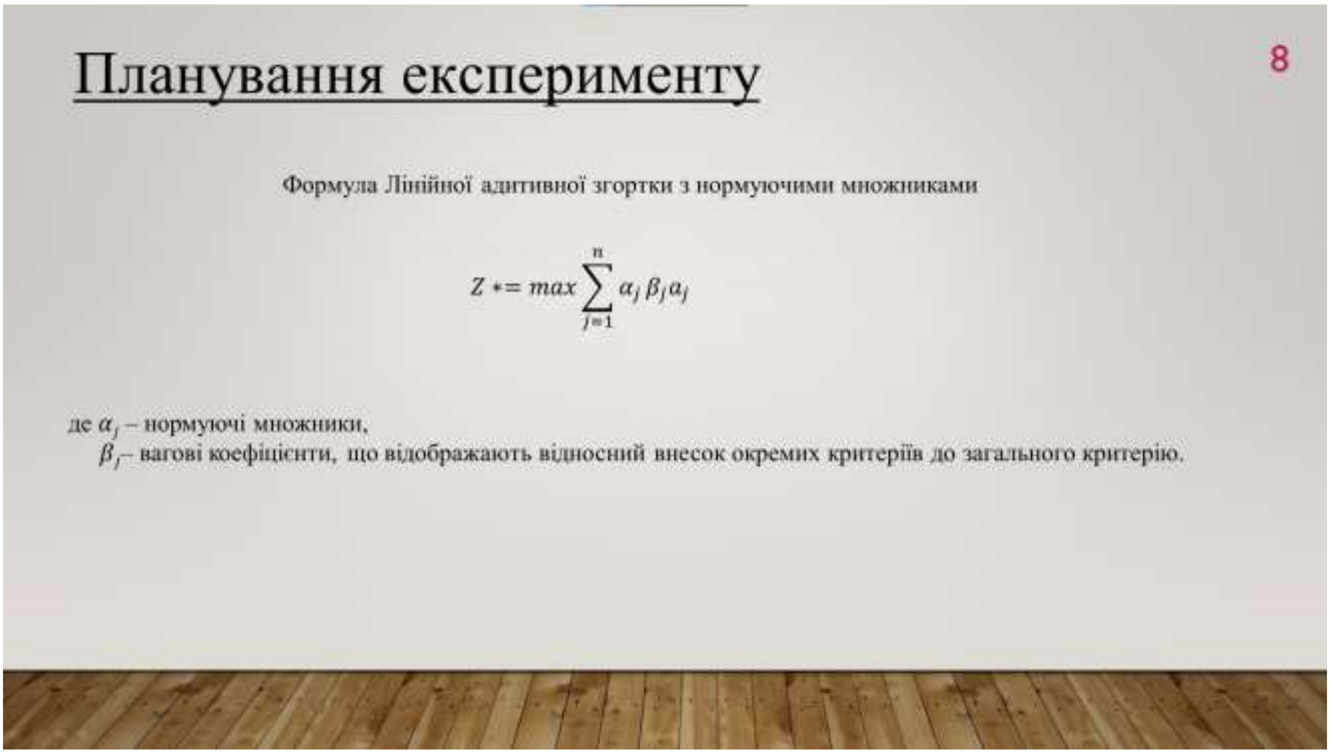

Актуальність
4
Об'єктом дослідження є технології зберігання зображень в гібридних системах, які комбінують використання блокчейн-технологій та традиційних баз даних для досягнення оптимального результату в зберіганні цифрових зображень.
Метою роботи є аналіз різних технічних підходів і систем зберігання, які використовують блокчейн і/або традиційні бази даних для забезпечення ефективного управління та зберіганням великих обсягів зображень.
Методами розробки та проєктування є аналіз предметної галузі дослідження, оцінка конкурентних систем зберігання даних за визначеними параметрами експериментальним методом.


Опис проблемної галузі
5
Аспекти, які розглядатимуться протягом дослідження:
Аналіз існуючих технологій;
Оцінка систем безпеки;
Оцінка масштабованості і продуктивності;
Управління даними та їхнє відновлення;
Практичність та застосування.


Планування експерименту
6
Експеримент полягатиме у порівнянні між собою п'яти систем зберігання даних методом лінійної адитивної згортки з нормуючими множниками. Порівняння відбуватиметься за такими критеріями:
Ціна;
Швидкість завантаження;
Безпечність;
Обсяги пам'яті;
Масштабованість.



Використані технології
7
BitTorrent
File System
MySQL
DigitalOcean
sia
STORJ



Планування експерименту
8
Формула Лінійної адитивної згортки з нормуючими множниками
$Z *= max \sum_{j=1}^{n} \alpha_j \beta_j a_j$
де α_j – нормуючі множники,
β_j– вагові коефіцієнти, що відображають відносний внесок окремих критеріїв до загального критерію.

Планування експерименту

9

Проміжні дані до проведення програмної частини експерименту

	Ціна	Швидкість	Безпечність	Обсяг пам'яті	Масштабованість
BitTorrent (BTFS)	10	TBD	9	2	TBD
Sia	5	TBD	9	2	TBD
Storj	5	TBD	10	2	TBD
MySQL	1	TBD	7	1	TBD
PostgreSQL	1	TBD	8	1	TBD

Програмна реалізація експерименту

10

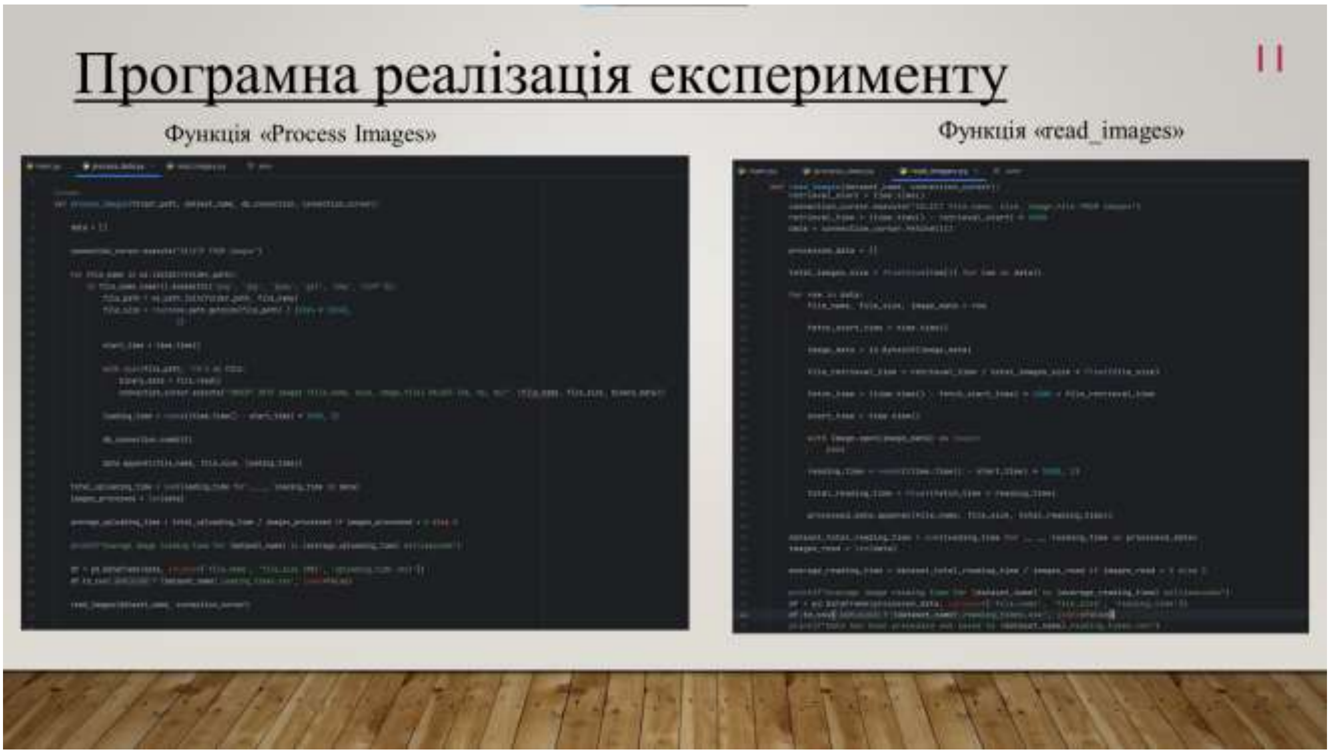
Підготовка до роботи зі сховищем (MySQL)

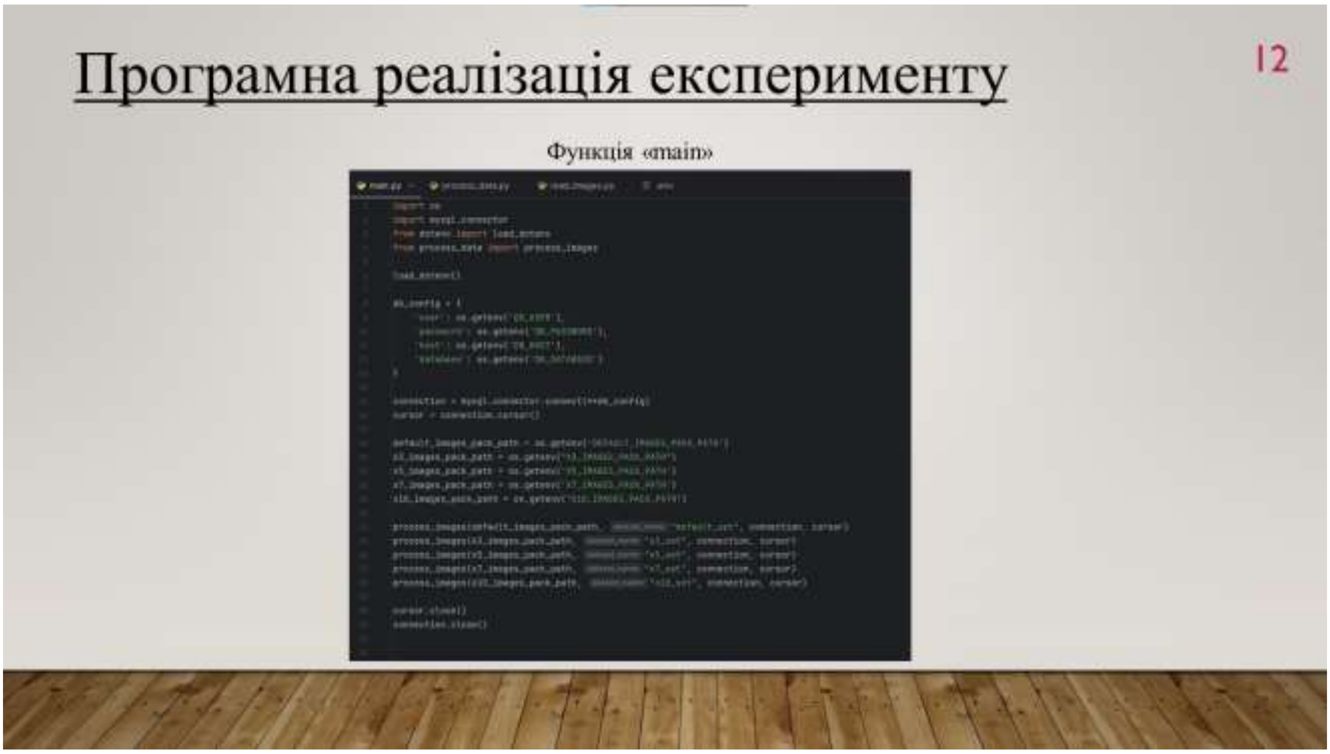
Запити, які використовуються в програмній реалізації:

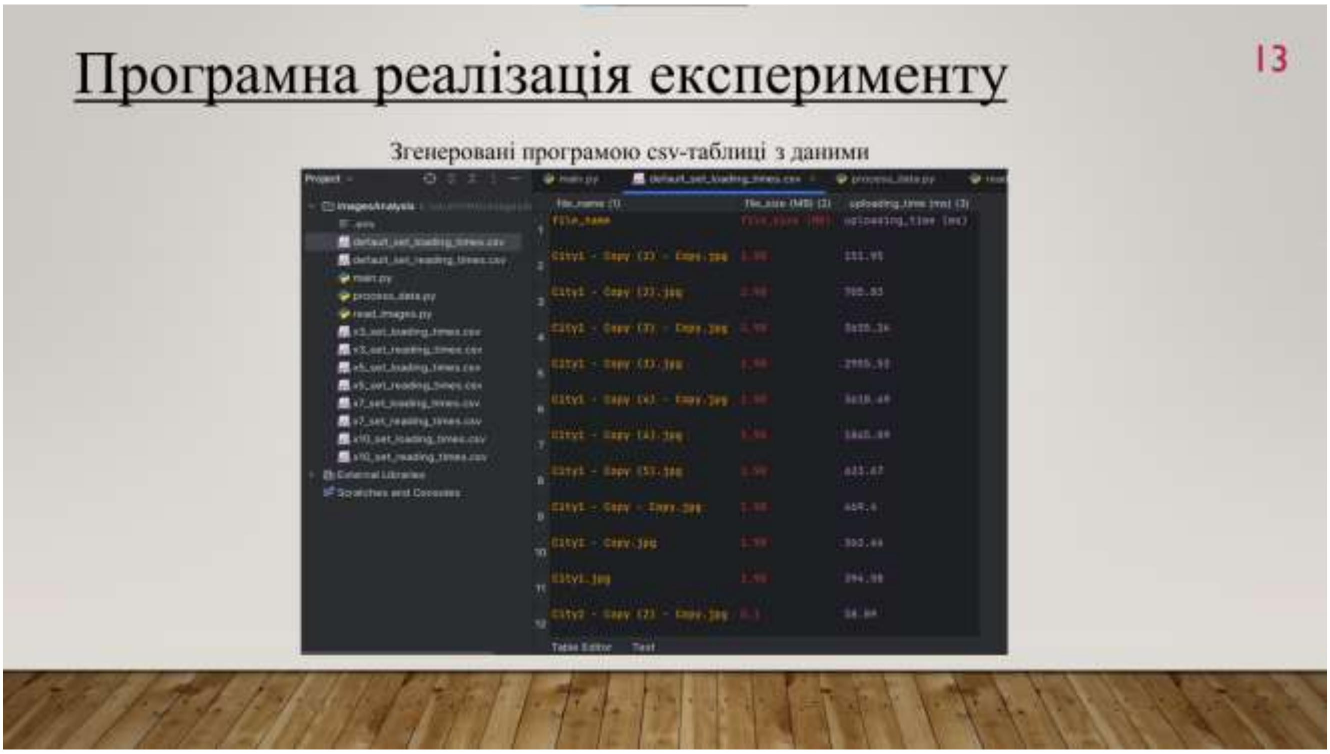
- «CREATE TABLE» – створення таблиці для зображень;
- «DELETE FROM» – очистка таблиці від даних;
- «INSERT INTO» – завантаження даних в таблицю;
- «SELECT» – отримання даних з таблиці.

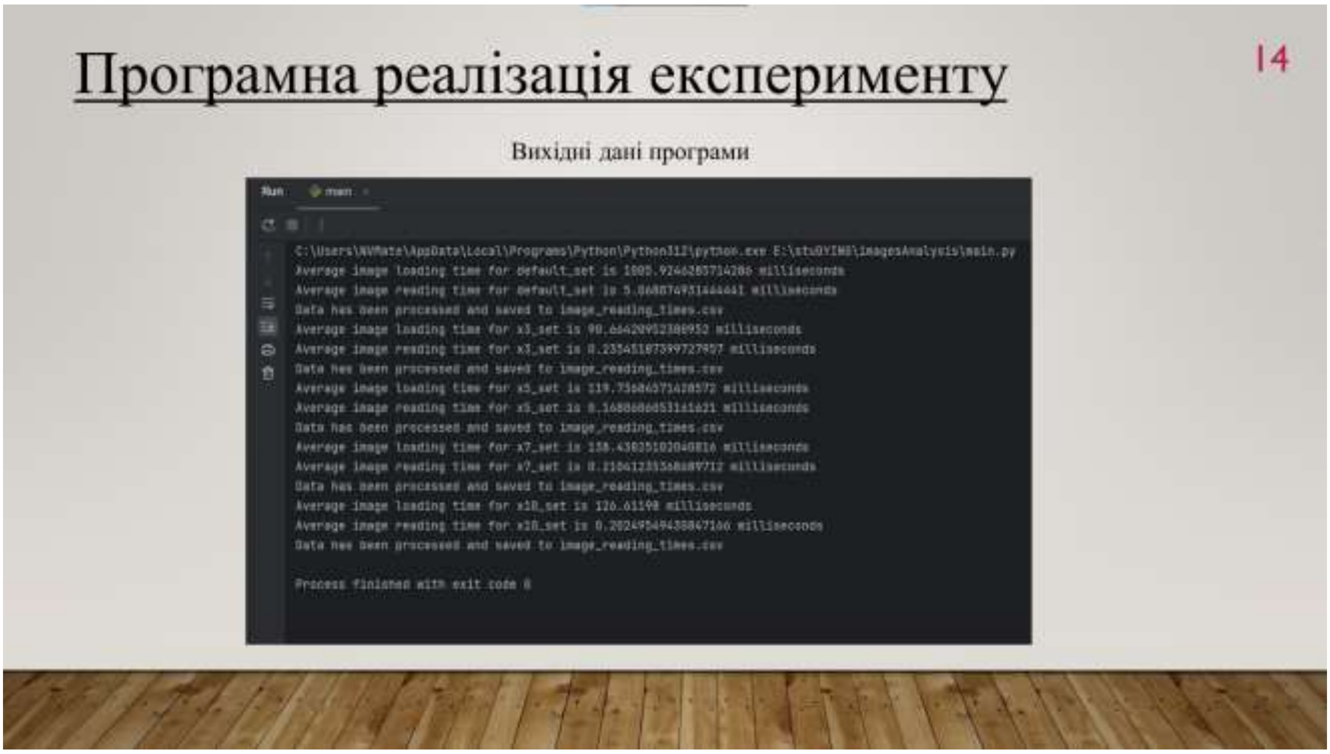
Схема бази даних

images	
id	integer
file_name	varchar(255)
size	decimal(6,3)
image_file	blob


Програмна реалізація експерименту
11
Функція «Process Images»
Функція «read_images»


Програмна реалізація експерименту
12
Функція «main»

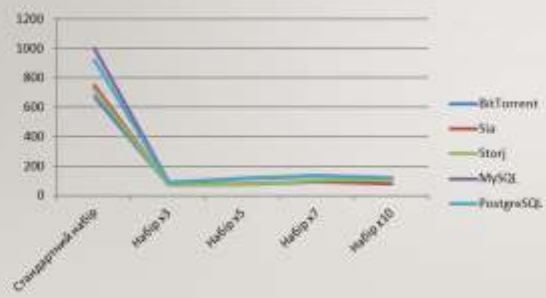

Програмна реалізація експерименту
13
Згенеровані програмою csv-таблиці з даними


Програмна реалізація експерименту
14
Вихідні дані програми
Run
main
Data has been processed and saved to image_reading_times.csv
Process finished with exit code 0

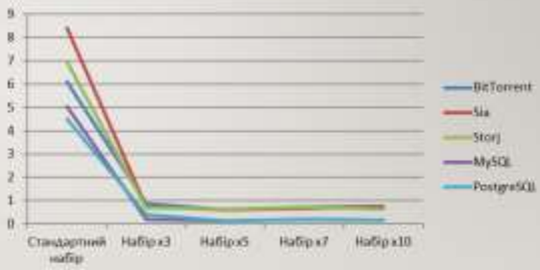
Програмна реалізація експерименту

15

Порівняльний графік часу загрузки зображень



Порівняльний графік часу зчитування зображень



Підсумки експерименту

16

Фінальні дані експерименту

	Ціна	Швидкість	Безпечність	Обсяг пам'яті	Масштабованість
BitTorrent (BTFS)	10	10	9	2	3.16
Sia	5	7.35	9	2	1
Storj	5	7.32	10	2	2.29
MySQL	1	1	7	1	8.45
PostgreSQL	1	1.84	8	1	10

Підсумки експерименту

17

Результати експерименту

Система	Результати згортки
BitTorrent (BTFS)	1.393
Sia	0.987
Storj	1.060
MySQL	0.705
PostgreSQL	0.820

Висновки

18

- Проаналізовано предметну галузь, розглянуто декілька варіантів систем, підходящих для зберігання зображень – дві традиційних і дві, які використовують блокчейн технології для зберігання даних. Для усіх систем було визначено основні критерії для порівняння їх одне з одним. Для кожної системи було проведено детальний розгляд кожного параметра кожної окремо взятої системи зберігання даних;
- Проведено експерименти зі збором параметрів, значення яких мали бути з'ясовані практичним методом. Для цього було створено Python-проєкт, який завантажував в системи зберігання даних зображення вимірюючи час завантаження та створюючи таблиці-звіти із виміряними показниками. Також проєкт вимірював і генерував звіти з часом зчитування зображень;
- Було з'ясовано, що для використання в малобюджетних проєктах варто звертатися до блокчейн-сховищ, бо за свої гроші вони дають гарні показники, але для великих проєктів краще підійдуть традиційні системи зберігання даних. Оптимальним варіантом при рівній важливості кожного з визначених критеріїв є використання системи зберігання BitTorrent (BTFS).
- Кваліфікаційна робота пройшла апробацію на науковій конференції (індексується SCOPUS) - 8th International Conference on Computational Linguistics and Intelligent Systems» (April 12-13 2024, Lviv, Ukraine)


дякую за увагу

ДОДАТОК Г

Апробація результатів роботи

Comparison of Blockchain–Based Data Storage Systems

Olga Cherednichenko[1], Iryna Kyrychenko[2], Glib Tereshchenko[2], Denys Miand[2] and Serhii Pylypenko[2]

[1] Univ Lyon, Univ Lyon 2, UR ERIC – 5 avenue Mendès France, 69676 Bron Cedex, France
[2] Kharkiv National University of Radio Electronics, Nauky Ave, 14, Kharkiv, 61166, Ukraine

Abstract

In the present paper blockchain-based data storage systems are getting reviewed. A comparative description is given to different data storage systems that use blockchain technologies in order to store data. The authors review pros and cons of the selected blockchain-based data storage systems. As a result, the optimal data storage system is chosen and suggestions on how to improve the storage systems are given.

Keywords

Data storage system, blockchain, costs, United States, research, linear adaptive convolution, decentralization, capacity, safety, node, anonymity, privacy, speed of transactions, standardization, regulation

1. Introduction

Our world changes constantly and lots of things get discovered every day. One of such discoveries happened about 30 years ago and gained a huge popularity in the last decade. This discovery is a blockchain. It started as a solution for timestamping of the digital documents, so no documents get misdated.

However, the most recent decade has proven that there are more use cases in which blockchain technologies can come in handy. Blockchain-based data storage systems are among these technologies as well. Based on blockchain, these technologies makes data storage cheaper and more reliable compared to centralized data storages. Also, blockchain-based data storages don't require advanced hardware to maintain their data centers since they simply don't have them. This allows profiting for businesses and individuals who are free to rent out their drives' free space.

Data storage in blockchain-based systems is implemented in following steps:

1. Data breaks into pieces (shards) during sharding
2. Shards of data get encrypted
3. For each shards is generated a hash which is sent to a blockchain ledge
4. Shards get replicated
5. Copies of shards are sent to nodes with synced ledger
6. User receives a key to decrypt their data

Currently there are four big blockchain-based storage networks:

- BitTorrent
- Filecoin
- Sia
- Storj

COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine
olga.cherednichenko@univ-lyon2.fr (O. Cherednichenko); iryna.kyrychenko@nure.ua (I. Kyrychenko); glib.tereshchenko@nure.ua (G. Tereshchenko); denys.miand @nure.ua (D. Miand); serhii.pylypenko @nure.ua (S. Pylypenko)
0000-0002-9391-5220 (O. Cherednichenko); 0000-0002-7686-6439 (I. Kyrychenko); 0000-0001-8731-2135 (G. Tereshchenko); 0009-0006-9691-1782 (D. Miand); 0009-0009-9049-7172 (S. Pylypenko)

This work's goal is to give a comparative characteristic of the data storage systems mentioned before to show their differences with pros and cons to give further suggestions on how these data storage systems could be improved.

2. Related Works

In [1] it was mentioned that scalability, security and decentralization belong to blockchain trilemma pictured on Figure 1.

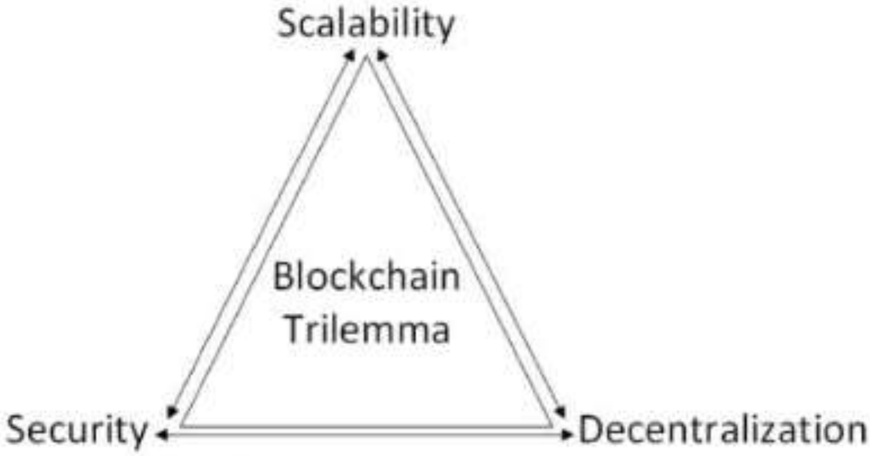


Figure 1: The blockchain trilemma

The blockchain trilemma offered by Vitalik Buterin states that it is inevitable that these parameters will be traded between each other. Author of the trilemma states that blockchain itself is decentralized and security is an essential part of it. However, it affects scalability and Bitcoin network is a good example of it since reducing latency to improve transaction throughput may result in weakening of security due to a higher probability of creating forks in the blockchain.

It totally makes sense, and it affects blockchain-based data storage systems as well. Main purpose of this research is to find out which of the most popular blockchain-based data storage systems is the best for use. These three parameters from the blockchain trilemma will be reviewed as well during this research.

Authors of [2] make a good work proving that modern challenges require better solutions than centralized cloud data storages due to high risks of data being accessed by dangerous individuals thus possessing threat to security and privacy of clients. This may lead to bad consequences for both individual users and organizations. There step in IPFS (InterPlanetary File Systems) that require stable connection between nodes which may lead to some performance issues instead rising the security level of the data.

However, using blockchain-based data storage systems gives more privacy and security it doesn't ensure that nothing bad happens to data. There still could be malicious nodes, so data tampering and other malicious activites, such as stealing of a content under the copyright, spreading of illegal content, etc may take place. Ways of protecting from it were researched in detail in [3]. Authors of this article suggest the way of protection for users' data. Although the object of their research is a medical data storage, it is still mostly applicable to general blockchain-based data storage systems.

For example, these measures offered by authors may help in preventing data losses or unwanted data access if the following measures are taken:

- System immutability: once data gets into storage it cannot be neither removed nor changed
- Decentralization: data must be stored in multiple different locations with an independent access to data to prevent a single point failure
- Digital signatures: ensure non-repudation by using signatures based on asymmetric key encryption
- Blockchain hashing algorithm: ensures integrity of data
- Data audibility: each node must be able to perform and audit
- Scalability
- Easy tracking: user should be able to see his data at any moment

These methods seem completely reasonable and having them applied would mean having the security grade of a given data storage system close to the highest grade possible.

3. Methods

In this part the problems of local storages and general cloud services are being stated to explain why it makes sense to switch from older technologies everyone is used to the newer ones. Also, there will be explained how the problem of selection of the optimal data storage system is going to be solved. In the last subsection there will be given more information on selected data-storage systems with their pros and cons.

3.1. Problems to solve

First thing first, we should investigate feasibility of switching from local data storages and usual cloud storages to blockchain-based storages. There are some problems that cause switching from "non-modern" data storage systems [4-6]:

- Growth of data volumes: modern world goes through the phase where data volumes grow exponentially each day due to increase of its quality (higher resolutions for images, bigger data sets for data mining, etc.) and capabilities of data storing hardware.
- Security and privacy problems: with technologies improvements come improvements of hackers' skills. However, people strive to keep their data as secure and private as possible. There comes the need in newer ways to protect user's data.
- Importance of effective data management: searching, indexing, metadata control are important features that allow users save a lot of time to get to needed data units.
- Scalability: bigger data volumes require more powerful data storage systems which will be able to store users' data and process it quickly.
- Trendiness: as blockchain is trendy nowadays, people try hard to learn it and test these futuristic and not yet properly explored things.

If you give it a thought, there are many problems which are if not solved, then are greatly relieved in blockchain-based data storage systems [7, 8].

3.2. Way to solve

Main purpose of this work is to determined differences between the most popular storage systems [9] that use blockchain to operate.

For this research has been decided to use linear additive convolution with normalizing factors as it gives precise results while still being simple.

Next step after deciding on using this method is to decide on parameters which values will play its role in selecting some blockchain-based data storage as an optimal one.

There are the parameters which are crucial when deciding if the blockchain-based data storage is worth using it (grades will be varying from 0 to 10 where 0 – worst result of a modern worl, and 10 – best result so far):

1. Cost – represents the cost of using and deployment of a specific data storage system. Mark is based on usage cost compared to other systems [10, 11].
2. Speed – represents the speed of the data processing and transmission. Speed estimates are based on total bandwidth flow and transaction processing speed for each system.
3. Security and privacy – represent level of users' data protection and how unlikely the data could be stolen or accessed by hackers. Evaluated relative to other blockchain-based data storage system in comparison.
4. Memory capacity – represents the amount of memory available or space for data storage in each system. The score is calculated relative to other systems.
5. Centralization – represents the centralization level of the particular data storage system. Even though it might count as a part of security parameter, we believe it should be a separate parameter as it is one of the main features provided by blockchain-based systems.

This method should let us compare all the chosen blockchain-based data storage systems to get a better understanding which one of them is optimal for migration there.

3.3. Familiarization with competing data-storage systems

This part will serve as a guide to what each of the competing blockchain-based data storage systems is [12-14] and what is its pros and cons. It will reveal more details about these systems which should give a better understanding of the current state of each system.

1. BitTorrent – a protocol for sharing and distributing files which uses a decentralized network to upload and transfer parts of files between users. The only cons of this data storage system are connected with it being secure so nobody has access to uploaded files, therefore the data could contain a harmful software or illegal content (either protected by copyrights or just forbidden by law).

Main benefits of BitTorrent:

- Decentralization: users can upload and share files simultaneously without overloading centralized servers;
- Upload speed: sharding allows faster receiving of big amounts of data
- Bandwidth: each user uploads and receives files at the same time which lowers the pressure on the network
- TRON-based: affiliation with a famous and long-maintained network is a plus since this is more stable, so data should be accessible all the time and won't get lost in the future due to the network's shutdown

2. Filecoin – decentralized network which allows storing and sharing of data using cryptocurrencies and blockchain technologies.

Main pros of Filecoin:

- Decentralization: the network uses nodes all around the world to operate so no main server is necessary for it to work
- High level of customization: storage providers are free to set the conditions of their data providing, be it the cost of the space they provide, who can use their service or under which jurisdiction they want to work with their potential client
- Proof-of-storage protocol: the system uses a unique protocol which checks if the data is really stored on chain before the storage providers receive their fee
- Easy provider switching: users are free to change their storage provider since there is an implemented content addressing for files so user doesn't have to download their files to reupload them to another provider's storage if they decide to do so

Main cons of Filecoin:

- Entry barrier: Filecoin uses its own token for transactions inside the system what forces users to spend additional time to trade the tokens
- Prices instability: as all the economy is built on the tokens which are not some stablecoins, the prices can vary from very high to very low

3. Sia – decentralized data storage platform which uses blockchain technologies to create a distributed storage network.

Main pros of Sia:

- High privacy: the network's decentralization makes any third parties unable to access your files or prevent you from accessing them
- Easily interactable: Sia exposes endpoints which are easily understandable since they are using JSON. The endpoints give access to the renter-host protocol which allows massive renting operations
- Open source: all software is completely open source which allows users to contribute to the code and even build applications and businesses on top of Sia
- High throughput: Sia stores redundant file segments on nodes across the globe providing high throughput

Main cons of Sia:

- Potentially dangerous files: due to absence of control over transferred files (which is usually one of pros), may happen a scenario where illegal or harmful files may be transferred over the network
- Dynamic pricing: due to instability of tokens which are used to purchase services on the network, a provider may earn money which wouldn't cost anything on the next day

4. Storj – decentralized data storage platform which uses blockchain technologies and cryptography protocols to maintain data security and privacy.

Main pros of Storj:

- Quick data transfer: due to operations being performed simultaneously data transfer appears to happen quicker
- Security: every file is encrypted which prevents data from being either censored, or hacked
- Sustainability: as promoted by the company, when using this system, due to absence of data centers, "user's carbon footprint is dramatically reduced"
- Stable cost: unlike usual pattern where the cost of services may vary due to token price volatility, Storj provides costs in real money what makes it clear how much user has to pay

Main cons of Storj:

- Data recovery: if a storage of one of the nodes gets broken, then part of data may get lost
- Hosting complexity: process of becoming a storage provider is difficult due to users reviews

Summing up all the information given in this section, these systems look alike with most of their features being similar, yet with little details added to some of them.

4. Experiment

In order to decide on optimal blockchain-based data storage system the earlier considered methods will be put to use.

First of all, we have to grade every system by selected criteria:

BitTorrent:

1. Cost (7/10): offers good prices in comparison to cloud storages, but not cheapest in blockchain-based data storages industry
2. Speed (8/10): known for its pretty fast data transfer capability due to its peer-to-peer technology
3. Security and privacy (8/10): provides good level of security and privacy, but not the highest in the industry [15, 16]
4. Memory capacity (8/10): offers a large amount of storage, but not unlimited
5. Decentralization (7/10): is decentralized, however its level of centralization is lower than level of its competitors [17]

Filecoin:

1. Cost (8/10): offers a competitive pricing which makes the model affordable for many users
2. Speed (7/10): its speed is decent, however not the transfers are not the fastest
3. Security and privacy (9/10): known for its good security measures [18]
4. Memory capacity (9/10): has a large decentralized storage network
5. Decentralization (8/10): is decentralized, but there are still some aspects of centralization

Sia:

1. Cost (9/10): known for its high cost-effectiveness
2. Speed (8/10): offer a good speed of transfer
3. Security and privacy (9/10): provides a high level of security and privacy due to its decentralized nature
4. Memory capacity (8/10): offers large data storage, but not unlimited
5. Decentralization (9/10): the system is highly decentralized

Storj:

1. Cost (8/10): offers competitive pricing model
2. Speed (7/10): good transfer speed, but not the best
3. Security and privacy (9/10): offers a high level of security and privacy due to its decentralized nature
4. Memory capacity (8/10): offers large data storage, but not unlimited
5. Decentralization (8/10): the system is decentralized, but there are still some aspects of decentralization

Below you can find Table 1 with all this data:

Table 1
Parameters of the blockchain-based data storage systems

System	Cost	Speed	Security and privacy	Memory capacity	Decentralization
BitTorrent	7	8	8	8	7
Filecoin	8	7	9	9	8
Sia	9	8	9	8	9
Storj	8	7	9	8	8

The formula (1) of the convolution with normalizing factors is as follows:

$$Z^* = \max_{i=1,m} \sum_{j=1}^{n} \alpha_j a_{ij}, \quad (1)$$

where $\alpha_j = \frac{1}{\sum_{i=1}^{m} a_{ij}}$ are normalizing factors.

Here are the normalizing factors for the convolution:

- Cost: 1/ (7+8+9+8) = 1/32 = 0.031
- Speed: 1/ (8+7+8+7) = 1/30 = 0.33
- Security and privacy = 1/ (8+9+9+9) = 1/35 = 0.29
- Memory capacity = 1/ (8+9+8+8) = 1/33 = 0.3
- Decentralization = 1/ (7+8+9+8) = 1/32 = 0.31

Next step is counting of the convolution index.

Results will be as follows:

- BitTorrent: 7/32 + 8/30 + 8/35 + 8/33 + 7/32= 1.175
- Filecoin: 8/32 + 7/30 + 9/35 + 9/33 + 8/32 = 1.263
- Sia: 9/32 + 8/30 + 9/35 + 8/33 + 9/32 = 1.329

- Storj: 8/32 + 7/30 + 9/35 + 8/33 + 8/32 = 1.233

Now when we've got every data storage system's convolution index the experiment is over and it's time to sum up all our results.

5. Results

In the Table 2 below there is data from Table 1 with respective convolution index in front of every present data storage system.

Table 2

Parameters of the blockchain-based data storage systems with convolution index

System	Cost	Speed	Security and privacy	Memory capacity	Decentraliza-tion	Convolution index
BitTorret	7	8	8	8	7	1.175
Filecoin	8	7	9	9	8	1.263
Sia	9	8	9	8	9	1.329
Storj	8	7	9	8	8	1.233

So, even though the results are very close, the Sia has the highest convolution index. The lowest one, on the other hand, was acquired by BitTorrent. The final results are with each blockchain-based data storage system's placement is shown in Table 3 below:

Table 3

Final result of the experiment

System	Convolution index	Placement
BitTorrent	1.175	4
Filecoin	1.263	2
Sia	1.329	1
Storj	1.233	3

So, according to the results, Sia received the highest grade, Filecoin has the second highest grade, Storj has the third one and BitTorrent has the lowest. The results do not differ very much. However, they allow us to tell which of the data storage systems is slightly better.

The method we used allowed us to choose optimal blockchain-based data storage system having in mind all the parameters are equally important to us [19, 20]. It allows us to spot even minor differences while deciding on which of the data storage systems is better.

6. Discussions

First of all, it is important to discuss which other ways to discover the most suitable blockchain-based data storage system we had. There are some more ways that come to mind if we decide not to use the chosen method:

- Feature by feature comparison – rough method which doesn't give any precise results and is strictly subjective. The point is to choose main features and compare data storage systems to one another. After finding differences, choose which one has better ones. Hasn't been selected due to being inaccurate.
- Linear additive convolution with weighted coefficients – another type of convolution which is more suitable if we have exact preferences on features. Difference with the chosen method is that multiplicators are selected based on our preferences instead of sum of grades on the parameter. Hasn't been chosen because it's been decided to study generally better blockchain-based data storage, not raiting one parameter over another.

Summing up the analytics above, it seemed more logical for us to use linear additive convolution with normalizing factors.

Since we've got the results, we can now also form suggestions on improvement for every system on the list:

- BitTorrent: Since it has the lowest grades for almost every parameter, then it would have a lot of aspects to work on. The cost problem could be fixed by adding liquidity to the chain's token. It would add some stability to the token price and storage costs should be reduced by the market. Regarding speed, memory capacity and decentralization, their situation could be improved by attracting new users into system, which, due to the blockchain trilemma, would lower the level of security [21]. The security level's loss can also be compensated by adding the features from the research [22].
- Filecoin: It has almost the highest marks of security and memory capacity. The security is again to be improved by adding measures from [22-24]. Speed, on the other hand, has the lowest value among all the participants. It can be improved by enlarging the quantity of storage providers or getting new storage providers with high storage capacity; however, it would lower the security level. Costs could be normalized by, again, adding liquidity or getting more users. New users would also raise level of decentralization and speed level in exchange for privacy [23, 24].
- Sia: It has got the biggest grade for almost every parameter. However, it can still be improved. It is mostly about getting new customers and storage providers with bigger storages and in more places around the world.
- Storj: Since it has its cost in dollars, then liquidity suggestion would not help there. However, all other tips given to another data storage systems are applicable there.

As many can notice, the ways to improve blockchain-based data storage systems are not too diverse. Therefore, all the systems have received pretty much alike suggestions. In [25] authors propose a model for secured data storage in decentralized cloud by using Etherium. The proposed system solves the challenges of cloud computing by integrating MetaMask for user registration, Ethereum smart contracts for secure data storage, ECC for efficient encryption, and IPFS for decentralized file sharing. It provides enhanced security, user privacy and decentralization by leveraging the strengths of each technology in a unified approach. Alike methods are also mentioned in [26, 27].

7. Conclusions

The purposes of the research were to find optimal blockchain-based data storage system, to see how they differ and what can be done to improve them.

During the completion of this paper, we've researched differences between most popular blockchain-based data storage systems in order to decide which of them was better. It appeared that Sia was the better one, Filecoin was the second one, third one was Storj and Bittorent was the last. However, it doesn't mean any of them is either bad or superior – they are all decent and have its own pros and cons. Then we have added suggestions for each one of them on how they can be improved.

The choice of method to decide which one of the systems was better has been highlighted as well as it was explained why other methods would be less effective during this research.

However, as stated before, this work doesn't give a definite answer to the question "What blockchain-based data storage system is better?" or "Which blockchain-based data storage system should I be using?" as there are still a lot of subjective factors like looks of interface, companies-partners, etc. which have not been taken into account since it has been decided to avoid subjective factors during decision-making to make the decision more clear and objective.

However, we think this topic deserves more researches from other perspectives or using other methods. In example, it is worth to make some research with statistics gathering on the subjective aspects of the given blockchain-based data storage systems. It would give even more precise results than this research and could even probably answer the questions stated before.

Acknowledgments

The research study depicted in this paper is funded by the French National Research Agency (ANR), project ANR-19-CE23-0005 BI4people (Business intelligence for the people).

References

[1] Nana Kwadwo Akarasi-Mensah, Eric Tutu Tchao, Axel Sikora, Andrew Selasi Agbemenu, Henry Nunoo-Mensah, Abdul-Rahman Ahmed, Dominik Welte, Eliel Keelson, An Overview of Technologies for Improving Storage Efficiency in Blockchain-Based IIoT Applications, Advanced Internet of Things (IoT): Sensing Techniques, Hardware, and Software Architectures for the Next Generation Internet of Things (2022) 11(16), 2513; https://doi.org/10.3390/electronics11162513.

[2] Luon-Chang Lin, Yi-Hsuan Kuo, Ching-Chung Chang, Jui-Chuan Liu, Chin-Chen Chang, Symmetry in Blockchain-Powered Secure Decentralized Data Storage: Mitigating Risks and Ensuring Confidentiality, Symmetry 2024, 16(2), 147; https://doi.org/10.3390/sym16020147.

[3] H. Javed, Z. Abaid, S. Akbar, K. Ullah, A. Ahmad, H. Ali, Y. Y. Ghadi, Blockchain-Based Logging to Defeat Malicious Insiders: The Case of Remote Health Monitoring Systems, IEEE Access (volume 12) (2023). doi:10.1109/ACCESS.2023.3346432.

[4] Kyrychenko, I., Shyshlo, O., Shanidze, N. "Minimizing Security Risks and Improving System Reliability in Blockchain Applications: a Testing Method Analysis", 2023 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2023), 2023. – CEUR-WS, 2023, ISSN 16130073. - Volume 3403, PP. 423-433.

[5] Biryukova, M., Kyrychenko, I., Shanidze, N., Shanidze, O. "Social Computing of the Social Well-being of Refugees and Internally Displaced Persons in Ukraine Using Data Mining Methods", 2023 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2023), 2023. – CEUR-WS, 2023, ISSN 16130073. - Volume 3403, PP. 410-422.

[6] K. Smelyakov, A. Chupryna, D. Sandrkin and M. Kolisnyk; Search by Image Engine for Big Data Warehouse; 2020 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 2020, pp. 1-4, doi: 10.1109/eStream50540.2020.9108782.

[7] M. Hayat, H. Winkler; From Traditional Product Lifecycle Management Systems to Blockchain-Based Platforms, Logistics 2022, 6(3), 40; https://doi.org/10.3390/logistics6030040.

[8] Iuon-Chang Lin, Chin-Chen Chang, Yu-Sung Chang; Data Security and Preservation Mechanisms for Industrial Control Network Using IOTA, Symmetry 2022, 14(2), 237; https://doi.org/10.3390/sym14020237.

[9] A. Chiarini, L. Compagnucci; Blockchain, Data Protection and P2P Energy Trading: A Review on Legal and Economic Challenges, Sustainability 2022, 14(23), 16305; https://doi.org/10.3390/su142316305.

[10] K. Azbeg, O. Ouchetto, Said Jai Andaloussi; BlockMedCare: A healthcare system based on IoT, Blockchain and IPFS for data management security, Egyptian Informatics Journal Volume 23, Issue 2, July 2022, Pages 329-343, https://doi.org/10.1016/j.eij.2022.02.004.

[11] S. Zaman, M. R. A. Khandaker, R. T. Khan, F. Tariq, K.-K. Wong; Thinking Out of the Blocks: Holochain for Distributed Security in IoT Healthcare, IEEE Access (Volume: 10), 30 March 2022, pp. 37064 – 37081, DOI: 10.1109/ACCESS.2022.3163580.

[12] R. P. Pinto, B. M. C. Silva, P. R. M. Inácio; A System for the Promotion of Traceability and Ownership of Health Data Using Blockchain, IEEE Access (Volume: 10), 31 August 2022, pp. 92760 – 92773, DOI: 10.1109/ACCESS.2022.3203193.

[13] T. L. Tan, I. Salam, M. Singh; Blockchain-based healthcare management system with two-side verifiability, PLOS ONE, April 2022, DOI : 10.1371/journal.pone.0266916.

[14] T. M. Ghazal, M. K. Hasan, S. N. H. S. Abdullah, K. A. A. Bakar, H. A. Hamadi; Private blockchain-based encryption framework using computational intelligence approach, Egyptian Informatics Journal Volume 23, Issue 4, December 2022, Pages 69-75, https://doi.org/10.1016/j.eij.2022.06.007.

[15] M. Alizadeh, K. Andersson, O. Schelén; Comparative Analysis of Decentralized Identity Approaches, IEEE Access (Volume: 10), 2022, pp. 92273 – 92283, DOI: 10.1109/ACCESS.2022.3202553.

[16] H. R. Hasan, K. Salah, I. Yaqoob, R. Jayaraman, S. Pesic, M. Omar; Trustworthy IoT Data Streaming Using Blockchain and IPFS, IEEE Access (Volume: 10), 07 February 2022, pp. 17707 – 17721, DOI: 10.1109/ACCESS.2022.3149312.

[17] S. Tang, Z. Wang, J. Dong, Y. Ma; Blockchain-Enabled Social Security Services Using Smart Contracts, IEEE Access (Volume: 10), 14 July 2022, pp. 73857 – 73870, DOI: 10.1109/ACCESS.2022.3190963.

[18] Z. Ullah, B. Raza, H. Shah, S. Khan, A. Waheed; Towards Blockchain-Based Secure Storage and Trusted Data Sharing Scheme for IoT Environment, IEEE Access (Volume: 10), 01 April 2022, pp. 36978 – 36994, DOI: 10.1109/ACCESS.2022.3164081.

[19] L. Wang, Y. He, Z. Wu; Design of a Blockchain-Enabled Traceability System Framework for Food Supply Chains, Foods 2022, 11(5), 744; https://doi.org/10.3390/foods11050744.

[20] T. Manoj, M. Krishnamoorthi, V. G. Narendra, D. T. Pham; A Blockchain Based Decentralized Identifiers for Entity Authentication in Electronic Health Records, Cogent Engineering Volume 9, 2022 - Issue 1, 08 Mar 2022, https://doi.org/10.1080/23311916.2022.2035134.

[21] N. Alrebdi, A. Alabdulatif, C. Iwendi, Z. Lian; SVBE: searchable and verifiable blockchain-based electronic medical records system, Scientific Reports volume 12, Article number: 266 (2022), https://doi.org/10.1038/s41598-021-04124-8.

[22] M. U. Javed, A. Jamal, E. H. Alkhammash, M. Hadjouni, S. A. Bahaj, N. Javaid; Secure Message Handling in Vehicular Energy Networks Using Blockchain and Artificially Intelligent IPFS, IEEE Access (Volume: 10), 28 July 2022, pp. 82063 – 82075, DOI: 10.1109/ACCESS.2022.3194513.

[23] A. Ali, M. A. Almaiah, F. Hajjej, M. F. Pasha, O. H. Fang, R. Khan, J. Teo, M. Zakarya; An Industrial IoT-Based Blockchain-Enabled Secure Searchable Encryption Approach for Healthcare Systems Using Neural Network, Sensors 2022, 22(2), 572; https://doi.org/10.3390/s22020572.

[24] H. Li, M. Li; Patent data access control and protection using blockchain technology, Scientific Reports volume 12, Article number: 2772 (2022), https://doi.org/10.1038/s41598-022-05215-w.

[25] N. Khan, H. Aljoaey, M. Tabassum, A. Farzamnia, T. Sharma, Y. H. Tung; Proposed Model for Secured Data Storage in Decentralized Cloud by Blockchain Ethereum, Electronics 2022, 11(22), 3686; https://doi.org/10.3390/electronics11223686.

[26] P. Kang, W. Yang, J. Zheng; Blockchain Private File Storage-Sharing Method Based on IPFS, Sensors 2022, 22(14), 5100; https://doi.org/10.3390/s22145100.

[27] C. Zakaret, N. Peladarinos, V. Cheimaras, E. Tserepas, P. Papageorgas, M. Aillerie, D. Piromalis, K. Agavanakis; Blockchain and Secure Element, a Hybrid Approach for Secure Energy Smart Meter Gateways, Sensors 2022, 22(24), 9664; https://doi.org/10.3390/s22249664.

ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент (посада) | програмної інженерії (кафедра) | ІПЗм 22-3 (група)

Мянд Д. Ю.
[illegible]

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
7.10.6	Пояснення познак, які входять до формули чи рівняння, треба подавати безпосередньо під формулою або рівнянням у тій послідовності, у якій їх наведено у формулі або рівнянні. Пояснення познак треба подавати без абзацного відступу з нового рядка, починаючи зі слова «де» без двокрапки. Позначки, яким встановлюють визначення чи пояснення, рекомендовано ви-рівнювати у вертикальному напрямку.	17
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

Експерт ________ (підпис) Вадим НЕЧВОЛОД (прізвище, ініціали)

10.06.2024