

PERMISSIONED VS PUBLIC БЛОКЧЕЙН У BIG DATA СЕРЕДОВИЩІ: ЗАБЕЗПЕЧЕННЯ ПРОЗОРОСТІ ТА НЕЗМІННОСТІ ДАНИХ

Вінницький національний технічний університет

Анотація

У роботі розглянуто особливості використання блокчейн-технологій у середовищі великих даних (Big Data) для забезпечення прозорості, цілісності та довіри до інформації. Проведено порівняльний аналіз двох основних типів блокчейн-архітектур – публічної (public) та дозволеної (permissioned) з точки зору їх застосування у розподілених аналітичних системах. Визначено ключові відмінності між ними за такими критеріями, як рівень децентралізації, продуктивність, керованість доступом, масштабованість і безпека. Особливу увагу приділено можливостям інтеграції блокчейну з Big Data платформами (Apache Hadoop, Spark, Kafka) для фіксації метаданих, аудиту та контролю достовірності даних. На основі аналізу запропоновано рекомендації щодо вибору оптимального типу блокчейну залежно від вимог до прозорості та швидкодії аналітичних процесів.

Ключові слова: Big Data, blockchain, permissioned blockchain, public blockchain, прозорість даних, незмінність даних, безпека, децентралізація, аналітика даних, Hyperledger Fabric.

Abstract

The article discusses the features of using blockchain technologies in the Big Data environment to ensure transparency integrity, and trust in information. A comparative analysis of two main types of blockchain architectures – public and permissioned – is conducted in terms of their application in distributed analytical systems. Key differences between them are identified based on criteria such as the level of decentralization, performance, access control, scalability and security. Particular attention is paid to the possibilities of integrating blockchain with Big Data platforms (Apache Hadoop, Spark, Kafka) for metadata recording, auditing, and data authenticity control. Based on the analysis, recommendations are made on choosing the optimal type of blockchain depending on the requirements for transparency and speed of analytical processes.

Keywords: Big Data, blockchain, permissioned blockchain, public blockchain, data transparency, data immutability, security, decentralisation, data analytics, Hyperledger Fabric.

Вступ

Сучасні системи Big Data обробляють величезні обсяги інформації, що надходить із різних джерел – від корпоративних баз даних до сенсорних мереж та IoT-пристроїв. Однією з головних проблем таких систем залишається забезпечення достовірності та незмінності даних протягом усього життєвого циклу від збору до аналітики. Традиційні централізовані методи зберігання й контролю даних не завжди гарантують прозорість і захист від внутрішніх маніпуляцій, особливо у багатокористувацьких або міжорганізаційних середовищах.

Блокчейн-технологія, що базується на принципах децентралізації, криптографічного захисту та консенсусу, пропонує ефективне рішення для підвищення надійності та довіри до даних у Big Data інфраструктурі. Проте існують різні підходи до її реалізації – публічні (public) та дозволені (permissioned) блокчейни, які суттєво відрізняються за архітектурою, швидкістю та рівнем доступу.

Результати дослідження

Блокчейн як технологія розподіленого реєстру (Distributed Ledger Technology, DLT) став основою сучасних систем довіреної обробки даних. Його головна перевага полягає в незмінності й прозорості записів, які підтверджуються всіма учасниками мережі. Кожен блок містить криптографічний хеш попереднього, формуючи ланцюг, що неможливо підробити без порушення цілісності всієї структури [1].

Публічні (public) блокчейни, такі як Bitcoin, Ethereum чи Avalanche, характеризуються відкритою архітектурою – будь-який користувач може брати участь у перевірці транзакцій, створенні блоків і читанні даних. Це забезпечує максимальну децентралізацію, але водночас створює проблеми масштабованості та продуктивності, особливо коли обсяги транзакцій сягають мільйонів записів на хвилину, як у Big Data-аналітиці [2].

Консенсус у таких мережах досягається через алгоритми Proof of Work (PoW) або Proof of Stake (PoS), що потребують значних обчислювальних ресурсів. Наприклад, енергоспоживання одного блоку Ethereum у період PoW досягало 1000 кВт·год [9].

Дозвільні (permissioned) блокчейни – Hyperledger Fabric, Quorum, Corda – функціонують за іншим принципом: учасники мережі проходять автентифікацію, а права доступу контролюються централізовано або федеративно. У таких системах використовуються ефективніші алгоритми консенсусу – Practical Byzantine Fault Tolerance (PBFT), Raft, IBFT – які забезпечують високу швидкість (до тисяч транзакцій за секунду) і низьку затримку при обробці даних [5].

Саме ці характеристики роблять permissioned-блокчейни перспективними для інтеграції у Big Data середовища, де критичними є оперативність обчислень і керованість доступом до інформації.

Для кращого розуміння відмінностей між типами блокчейнів доцільно розглянути особливості кількох найпоширеніших платформ – Ethereum, Hyperledger Fabric, Quorum та Avalanche. У таблиці 1 наведено порівняльний аналіз платформ.

Таблиця 1 – Порівняння найпоширеніших платформ

Платформа	Тип	Алгоритм консенсусу	Пропускна здатність (TPS)	Середня затримка	Конфіденційність
Ethereum	Public	PoS (раніше PoW)	~30 TPS	Висока	Низька
Hyperledger Fabric	Permissioned	PBFT / Raft	1000–3500 TPS	Низька	Висока (приватні канали)
Quorum	Permissioned	IBFT / Raft	400–1000 TPS	Низька	Висока
Avalanche	Public	Avalanche Consensus	4500–7000 TPS	Низька	Середня

Публічні платформи, як Ethereum, орієнтовані на максимальну прозорість і універсальність, але їх продуктивність обмежена через необхідність підтвердження транзакцій усією мережею. Це робить їх менш придатними для потокових аналітичних задач. Avalanche, хоч і публічна система, демонструє значне зростання швидкодії завдяки новому типу консенсусу, заснованому на стохастичному опитуванні вузлів. Проте навіть у цьому випадку відсутність контрольованого доступу може бути ризиком для конфіденційних даних [4].

У протилежність їм, Hyperledger Fabric створений спеціально для корпоративних сценаріїв. Його архітектура базується на модульності, де ролі вузлів (peer, orderer, client) чітко розділені. Fabric підтримує приватні канали зв'язку, дозволяючи зберігати комерційно важливі дані в окремих секторах мережі [3]. Quorum, розроблений JPMorgan, поєднує переваги Ethereum (смарт-контракти, сумісність) із permissioned архітектурою, що дозволяє інтегруватися з фінансовими Big Data системами для аудиту транзакцій [7].

Отже, загалом permissioned-блокчейни демонструють кращу продуктивність, масштабованість та відповідність корпоративним вимогам, тоді як публічні – переважають у прозорості та відкритості екосистеми [2].

Вибір технології блокчейну для Big Data аналітики має ґрунтуватися на системі показників ефективності. Серед ключових – пропускна здатність, latency, енергоспоживання, масштабованість і конфіденційність.

Для Big Data критичним є обсяг даних, які можна обробити за одиницю часу. У системах потокової аналітики, як-от Apache Kafka або Spark Streaming, події надходять безперервно, тому блокчейн повинен здатись реєструвати транзакції з мінімальною затримкою. Публічні системи, як Ethereum, обмежуються десятками транзакцій за секунду, тоді як Hyperledger Fabric забезпечує понад 3000 TPS [9]. Це пояснюється різницею в механізмах консенсусу: PoS потребує глобального узгодження, а PBFT – лише між довіреними вузлами.

Питання енергоефективності є особливо важливим при масштабному розгортанні аналітичних платформ. За даними [6], публічні блокчейни з PoW можуть споживати в сотні разів більше енергії, ніж permissioned рішення з Raft. Таким чином, для корпоративних середовищ, орієнтованих на сталий розвиток, більш доцільним є використання permissioned архітектур.

Big Data інфраструктура зазвичай потребує масштабування горизонтального типу. Permissioned системи, такі як Hyperledger Fabric, дозволяють додавати нові вузли без зниження продуктивності завдяки шардінгу та розподілу ролей між peer-вузлами. Публічні ж мережі мають обмеження, пов'язані з необхідністю підтвердження транзакцій великою кількістю вузлів [4].

Конфіденційність даних є одним із ключових викликів Big Data аналітики. Permissioned блокчейни дозволяють створювати приватні канали, де транзакції бачать лише уповноважені учасники. У

публічних мережах навіть зашифровані дані можуть бути піддані кореляційному аналізу, що загрожує витоком [6].

Інтеграція блокчейну у Big Data середовище спрямована на створення трастової інфраструктури даних, де кожна операція має підтвердження походження (data provenance). Зв'язок між цими технологіями реалізується через middleware-рівень, який з'єднує системи збору, зберігання та верифікації даних.

Одним із найефективніших підходів є поєднання Apache Kafka з permissioned блокчейном, наприклад, Quorum. Kafka забезпечує безперервну обробку подій у реальному часі, а блокчейн гарантує їх незмінність і перевірюваність. Така архітектура підходить для аудиту транзакцій у фінансових або IoT-системах [5].

Apache Spark також може бути інтегрований із блокчейном як джерело достовірних даних. У дослідженні [4] наведено приклад системи, де результати машинного навчання, виконані в Spark, автоматично зберігаються в Hyperledger Fabric для забезпечення відтворюваності аналітичних результатів.

Надоор, у свою чергу, може виступати як сховище великих наборів даних, а блокчейн – як «мета-рівень довіри», що контролює цілісність і походження файлів у HDFS.

Таким чином, блокчейн не замінює Big Data технології, а доповнює їх, створюючи рівень довіри й захисту, що критично важливо при обміні даними між різними організаціями.

Аналізуючи наведені аспекти, можна сформулювати такі рекомендації. Для аналітики поточкових даних найкраще підходять permissioned рішення (Hyperledger Fabric, Quorum), які забезпечують низький latency і контроль доступу. У фінансових системах перевагу мають Quorum або Corda, які підтримують приватні транзакції. Для аудиту дій користувачів і зберігання логів доцільне використання гібридних моделей, що поєднують permissioned ядро з публічними каналами для верифікації результатів. Публічні блокчейни – Ethereum, Avalanche – залишаються оптимальними для відкритих аналітичних платформ, де ключовою є довіра та прозорість [4].

Висновки

У ході дослідження проаналізовано принципи функціонування двох основних типів блокчейн-архітектур – публічних (public) та дозвільних (permissioned) – у контексті їх інтеграції з інфраструктурою Big Data. Визначено, що блокчейн може виконувати роль механізму забезпечення довіри, який гарантує незмінність, автентичність і контрольовану прозорість даних під час їх зберігання й аналітичної обробки [1].

Проведене порівняння показало, що public блокчейни (зокрема Ethereum, Avalanche) забезпечують високий рівень децентралізації та відкритості, що робить їх доцільними для відкритих аналітичних систем, публічних моніторингових проєктів і платформ колективного доступу до даних. Проте обмежена пропускна здатність і підвищене енергоспоживання таких систем зменшують їхню ефективність у задачах потокової аналітики великих обсягів даних у реальному часі [8].

Натомість permissioned блокчейни (Hyperledger Fabric, Quorum, Corda) демонструють вищу швидкість, масштабованість і контроль доступу, що робить їх більш придатними для корпоративних або міжорганізаційних Big Data середовищ. У таких сценаріях критично важливими є конфіденційність, аудит, відновлення довіри між партнерами та інтеграція з існуючими аналітичними платформами

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Bitcoin - Open source P2P money. URL: <https://bitcoin.org/bitcoin.pdf> (date of access: 20.10.2025).
2. Ethereum.org: The complete guide to Ethereum. URL: https://ethereum.org/content/whitepaper/whitepaper-pdf/Ethereum_Whitepaper_-_Buterin_2014.pdf (date of access: 20.10.2025).
3. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. arXiv.org. URL: <https://arxiv.org/abs/1801.10228> (date of access: 22.10.2025).
4. Amir Masoud Rahmani V. B. Blockchain integration in big data: Review, vision, and opportunities. *WILEY Online Library*. URL: <https://onlinelibrary.wiley.com/doi/10.1002/spy2.392> (date of access: 22.10.2025).
5. Artem Barger E. A. Hyperledger fabric: a distributed operating system for permissioned blockchains. *ACM*. URL: <https://dl.acm.org/doi/10.1145/3190508.3190538> (date of access: 22.10.2025).
6. Reynold S. Xin M. Z. Apache Spark: a unified engine for big data processing. *ACM*. URL: <https://dl.acm.org/doi/10.1145/2934664> (date of access: 23.10.2025).

7. Unleashing the Potential of Permissioned Blockchain: Addressing Privacy, Security, and Interoperability Concerns in Healthcare Data Management. *MDPI*. URL: <https://www.mdpi.com/2079-9292/13/24/5050> (date of access: 23.10.2025).

8. Peter Mell D. Y., Nik Roby. *NIST Technical Series Publications*. URL: <https://nvlpubs.nist.gov/nistpubs/ir/2018/nist.ir.8202.pdf> (date of access: 25.10.2025).

9. A Comparative Study of Consensus Mechanisms in Blockchain for IoT Networks. *ResearchGate*. URL: https://www.researchgate.net/publication/363059813_A_Comparative_Study_of_Consensus_Mechanisms_in_Blockchain_for_IoT_Networks (date of access: 25.10.2025).

Пінчук Дар'я Олександрівна – студентка групи 1KITC-226, Факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: dashapinchukschool@gmail.com

Науковий керівник: **Грицак Анатолій Васильович** – кандидат технічних наук, доцент, доцент кафедри менеджменту та безпеки інформаційних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: grytsak.a.v@gmail.com

Pinchuk Daria O. – student of group 1KITS-22b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: dashapinchukschool@gmail.com

Supervisor: **Hrytsak Anatoly V.** – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Management and Security of Information Systems Vinnytsia National Technical University, Vinnytsia, e-mail: grytsak.a.v@gmail.com