

УДК 004.056.53:004.021

DOI <https://doi.org/10.32689/maup.it.2025.4.22>

**Дмитро ПЕРСУНОВ**

аспірант спеціальності 126 Інформаційні системи та технології,  
Державне некомерційне підприємство «Державний університет «Київський авіаційний інститут»,  
5812850@stud.kai.edu.ua  
ORCID: 0009-0000-5877-3493

## ТЕХНОЛОГІЯ КОНТРОЛЮ ТА ВІДНОВЛЕННЯ ЦІЛІСНОСТІ ДАНИХ У РОЗПОДІЛЕНИХ СХОВИЩАХ

**Анотація.** У статті запропоновано інформаційну технологію контролю та відновлення цілісності даних у розподілених багатоджерельних сховищах, яка поєднує формальні критерії структурної, логічної та семантичної відповідності з адаптивними методами обчислення узгодженості та реконструкції даних. Розроблена математична модель забезпечує кількісну оцінку рівня порушень цілісності та визначає найбільш достовірні значення в умовах конфліктів або відсутності фрагментів інформації. Запропонований підхід апробовано на моделюванні багатоджерельного набору даних, що демонструє ефективність технології у виявленні аномалій та формуванні узгодженого інформаційного простору.

**Мета.** Підвищення рівня цілісності даних у розподілених багатоджерельних сховищах шляхом розроблення інформаційної технології контролю та відновлення даних на основі формалізованої математичної моделі.

**Методологія.** Методологічною основою дослідження є математичне моделювання процесів забезпечення цілісності даних у розподіленому середовищі, аналіз структурних і логічних порушень, формалізація функцій узгодженості та визначення оптимальних значень параметрів на основі вагових коефіцієнтів довіри до джерел. Практична частина реалізована у середовищі Python із використанням прототипу інформаційної технології, що виконує виявлення аномалій, оцінювання узгодженості та реконструкцію даних для формування узгодженого набору.

**Наукова новизна.** Запропоновано формалізовану модель узгодженості багатоджерельних даних, що поєднує структурні, логічні та семантичні критерії з адаптивним відновленням значень.

**Висновки.** Запропонована інформаційна технологія забезпечує автоматичне виявлення порушень цілісності та формування узгоджених даних у розподіленому багатоджерельному середовищі. Результати моделювання підтверджують практичну ефективність моделі у виявленні аномалій та відновленні параметрів за критерієм максимізації узгодженості.

**Ключові слова:** цілісність даних, багатоджерельні сховища, інформаційна технологія, узгодженість, аномалії, відновлення даних, математична модель.

## Dmitry PERSUNOV. TECHNOLOGY FOR CONTROLLING AND RESTORING DATA INTEGRITY IN DISTRIBUTED STORAGE SYSTEMS

**Abstract.** The article proposes an information technology for controlling and restoring data integrity in distributed multi-source storage systems, which combines formal criteria of structural, logical and semantic consistency with adaptive methods of calculating consistency and reconstructing data. The developed mathematical model provides a quantitative assessment of the level of integrity violations and determines the most reliable values in conditions of conflict or absence of information fragments. The proposed approach has been tested on a multi-source data set, demonstrating the effectiveness of the technology in detecting anomalies and forming a consistent information space.

**Objective.** Improving data integrity in distributed multi-source storage systems by developing information technology for data control and recovery based on a formalised mathematical model.

**Methodology.** The methodological basis of the study is mathematical modelling of data integrity assurance processes in a distributed environment, analysis of structural and logical violations, formalisation of consistency functions, and determination of optimal parameter values based on source trust weighting coefficients. The practical part is implemented in the Python environment using a prototype information technology that performs anomaly detection, consistency assessment, and data reconstruction to form a consistent set.

**Key words:** data integrity, multi-source storage, information technology, consistency, anomalies, data recovery, mathematical model.

**Вступ.** Розподілені інформаційні системи, що працюють із різнотипними та незалежними джерелами даних, стикаються з проблемою втрати цілісності через пропуски, аномалії, конфлікти та неоднорідність форматів. У таких умовах неможливо гарантувати достовірність і узгодженість інформації без спеціалізованих механізмів перевірки та відновлення даних. Наявні підходи або орієнтовані на централізовані системи, або не враховують багатоджерельність та асинхронність потоків, що створює потребу у розробленні інформаційної технології, здатної забезпечити повний цикл контролю та реконструкції даних у розподіленому середовищі.

© Д. Персунів, 2025

Стаття поширюється на умовах ліцензії CC BY 4.0

**Виклад основного матеріалу.** Цілісність даних у розподілених сховищах є однією з ключових характеристик інформаційної системи, що визначає її здатність забезпечувати достовірність, узгодженість та немодифікованість даних під час їх зберігання, оброблення та передавання [1–3]. У розподілених IT-інфраструктурах порушення цілісності виникають через асинхронність взаємодії між вузлами, неоднорідність джерел, затримки у мережі, збої апаратного рівня, некоректне об'єднання даних, а також людський фактор. Враховуючи зростання обсягів інформації та зміщення до багатоджерельних моделей, домінують сценарії, у яких дані надходять із різних сенсорів, модулів, підсистем і потребують узгодження в момент інтеграції [4–6].

Розподілені сховища характеризуються відсутністю єдиного центру контролю, що ускладнює гарантування узгодженості. Традиційні підходи, такі як транзакційні моделі ACID чи жорсткі схеми реплікації, стають менш ефективними при великих масштабах або гетерогенних потоках. Тому сучасні інформаційні технології переходять до комбінованих стратегій: часткова реплікація, адаптивні механізми перевірки, інтелектуальний аналіз аномалій, локально-орієнтоване відновлення даних та інкрементальна валідація об'єктів.

Цілісність у розподілених сховищах розглядають у трьох аспектах: структурна цілісність, логічна цілісність, семантична цілісність [7].

Для кожного з цих аспектів механізми перевірки застосовуються окремо, але інтегровано в межах єдиної інформаційної технології. Це дає змогу знизити ймовірність накопичення помилок у довготривалих обчислювальних процесах.

Існуючі засоби забезпечення цілісності включають контрольні-сумаційні методи, алгоритми хешування, транзакційні логіки, механізми цифрового підпису, протоколи узгодження (Raft, Paxos), методи виявлення аномалій та оцінювання достовірності даних [8]. Проте основним викликом залишається те, що більшість цих підходів орієнтовані або на цілісність окремих фрагментів, або на централізовані моделі, які не враховують багатоджерельність та асинхронність потоків.

Таким чином, виникає необхідність у цілісній інформаційній технології, яка поєднує механізми контролю, виявлення порушень, оцінювання достовірності та відновлення даних у межах розподіленого середовища. Запропонований підхід передбачає інтеграцію математичних методів оцінювання узгодженості, алгоритмів локальної реконструкції фрагментів, а також інтелектуальних засобів аналізу аномалій у потоках даних.

З урахуванням теоретичних положень виникає необхідність у формальному описі процесів контролю та відновлення цілісності у розподілених сховищах. Теоретичні аспекти дозволяють окреслити основні джерела порушень, типи невідповідностей і характер взаємодії між даними з різних вузлів, однак для побудови інформаційної технології потрібна математична модель, яка забезпечує можливість кількісного оцінювання цілісності, виявлення критичних відхилень і вибору оптимального варіанта відновлення. Саме тому було розроблено формалізовану модель, що описує структуру багатоджерельного набору даних, функції перевірки узгодженості та правила реконструкції пошкоджених фрагментів.

*Крок 1. Формалізація багатоджерельних даних.*

Розподілене сховище розглядається як множина незалежних джерел:

$$S = \{S_1, S_2, \dots, S_m\}.$$

Кожне джерело генерує набір фрагментів:

$$D = \{d_1, d_2, \dots, d_n\}, d_i \in S_j.$$

*Крок 2. Формалізація показників цілісності.*

Для кожного фрагмента даних вводимо три складові цілісності:

$\alpha_s$  – структурна відповідність;

$\alpha_l$  – логічна відповідність;

$\alpha_c$  – семантична відповідність.

Тоді інтегральна функція цілісності:

$$I(d_i) = f(\alpha_s(d_i), \alpha_l(d_i), \alpha_c(d_i)).$$

Таким чином можна формально оцінити, наскільки фрагмент відповідає вимогам цілісності у трьох вимірах.

*Крок 3. Виявлення порушення цілісності.*

Порушення визначається як відхилення від нормованої інтегральної оцінки:

$$\Delta(d_i) = 1 - I(d_i),$$

де  $\Delta(d_i) = 0$  – повна відповідність,  $\Delta(d_i) \rightarrow 1$  – критичне порушення. Це ключова метрика, на основі якої запускається процес відновлення.

**Крок 4. Оцінювання узгодженості між даними з різних джерел.**

Для даних, що описують один об'єкт або подію, вводиться метрика розбіжності:  $dist(d_i, d_k) = |d_i - d_k|$ , також вводиться ваговий коефіцієнт довіри до джерела:  $w_j = g(S_j)$ . Тоді узагальнена оцінка узгодженості:  $I_{total} = \sum_{j=0}^m w_j \cdot I(d_j)$ .

На цьому кроці визначається, наскільки дані з різних джерел «погоджуються» між собою та які з них є найбільш достовірними.

**Крок 5. Модель відновлення даних.**

Пошкоджений або суперечливий фрагмент  $d_i$  замінюється на значення, яке має максимальну узгодженість:  $d^* = \underset{d_i \in D}{argmax} I(d_i)$ .

Це правило дозволяє обрати найбільш достовірний варіант з наявних даних або реконструйованих значень.

Запропонована математична модель описує формальний механізм контролю та відновлення цілісності даних у розподіленому багатоджерельному середовищі. Вона узагальнює структуру даних, способи їх класифікації, критерії відповідності та правила визначення розбіжностей між фрагментами, що надходять із різних джерел. Модель визначає, яким чином можна кількісно оцінити рівень структурної, логічної та семантичної цілісності, і встановлює інтегральні показники, що відображають ступінь відхилення від нормативного стану. Завдяки цьому процес контролю стає не описовим, а формалізованим, що дозволяє задавати конкретні порогові значення, критерії аномалій та правила ухвалення рішень.

Крім того, математична модель визначає процедуру відновлення даних через пошук найбільш узгодженого та достовірного значення серед альтернативних фрагментів. Це робиться за допомогою оцінки довіри до джерел, функцій подібності та оптимізаційних критеріїв. У результаті ця модель виступає центральною теоретичною основою інформаційної технології, забезпечуючи формальний опис усіх ключових процесів: аналізу, виявлення, оцінювання та відновлення даних.

Оскільки математична модель формалізує логіку контролю та відновлення цілісності, наступним етапом є її практична реалізація. Для перевірки працездатності запропонованої математичної моделі було розроблено прототип інформаційної технології контролю та відновлення цілісності даних у середовищі Python з використанням бібліотеки pandas. Як тестовий приклад розглядалося багатоджерельне сховище з трьома джерелами даних, що описують один і той самий набір об'єктів, але містять пропуски, суперечливі та аномальні значення. Такий підхід дає змогу змодельовати типову ситуацію в розподіленій інформаційній системі, де дані формуються незалежними підсистемами та потребують узгодження.

Кожне джерело подано у вигляді таблиці (DataFrame), що містить ідентифікатор об'єкта та значення контрольованого параметра. Під час моделювання навмисно формуються ситуації, коли для одного й того самого об'єкта значення параметра в різних джерелах відрізняються, відсутні або виходять за допустимі межі. На основі запропонованої математичної моделі виконується послідовний контроль структурної, логічної та семантичної цілісності, обчислюються показники узгодженості та обирається найбільш достовірне значення параметра для запису в узгоджене сховище.

Результат моделювання (рис. 1) коду на Python, який реалізує об'єднання даних з трьох джерел, виявлення аномальних значень та відновлення параметра за критерієм зваженого середнього з урахуванням коефіцієнтів довіри до джерел.

	id	value_s1	value_s2	value_s3	restored_value
0	1	10.0	10.5	9.5	10.050000
1	2	20.0	NaN	21.0	20.384615
2	3	30.0	29.0	100.0	29.533333
3	4	NaN	40.0	39.0	39.583333

In [1]: |

**Рис. 1. Результати роботи прототипу інформаційної технології**

Три незалежні джерела містили різні типи порушень: пропущені значення, випадкові відхилення та аномальні дані, що виходять за межі допустимого діапазону. Застосування запропонованої математичної моделі дозволило автоматично виявити такі порушення та обчислити інтегральний показник узгодженості для кожного фрагмента.

На основі коефіцієнтів довіри до джерел виконувалось відновлення значень за правилом максимізації узгодженості. У випадку наявності декількох коректних варіантів використовувалося зважене середнє значення, що враховує якість кожного джерела. Таким чином, для кожного об'єкта було отримано відновлене значення параметра, яке є найбільш достовірним серед доступних альтернатив.

Результати моделювання показують, що аномальні значення (наприклад, 100.0 для об'єкта з  $id = 3$ ) були коректно виключені з обчислення, пропуски успішно компенсовані за рахунок інших джерел, а фінальні значення в усіх випадках збігаються з очікуваним поведінковим діапазоном. Це підтверджує ефективність інформаційної технології при виявленні та усуненні порушень цілісності в умовах неоднорідності даних.

**Висновки.** У статті вирішено актуальну задачу забезпечення цілісності даних у розподілених багатоджерельних сховищах, що характеризуються неоднорідністю форматів, різною якістю інформації та відсутністю централізованого контролю. Розроблено цілісну інформаційну технологію, яка реалізує повний цикл: від формалізованого аналізу структурних і семантичних порушень та оцінювання узгодженості до відновлення значень на основі вагових коефіцієнтів довіри до джерел та оптимізаційного критерію. Запропонована математична модель дозволяє кількісно оцінювати ступінь порушення цілісності та вибирати найбільш достовірні значення серед доступних альтернатив, що забезпечує підвищення точності та надійності інтегрованої інформації. Реалізація прототипу в середовищі Python підтвердила ефективність розробленої технології, оскільки всі види порушень – пропуски, аномалії, відхилення та конфлікти – були коректно виявлені та виправлені. Таким чином, запропонована інформаційна технологія може бути використана як основа для побудови масштабованих систем моніторингу, інтеграції та аналізу даних у складних інформаційних інфраструктурах, де особливо важливими є достовірність і повнота даних.

#### Список використаних джерел:

1. Alkhafaji M. J. A., Kashkevich S., Shyshatskyi A., Sova O., Nalapko O., Buyalo O., Yula O., Shaposhnikova O., Matsyi O., Dvorskyi M. Development of a method for managing a group of unmanned aerial vehicles using a population algorithm. *Eastern-European Journal of Enterprise Technologies*, 2024. 6 (9 (132)), 108–116. <https://doi.org/10.15587/1729-4061.2024.318600>.
2. G. R. Annapureddy. The Impact of Blockchain on Data Integrity and Security in Distributed Systems. *The Eastasouth Journal of Information System and Computer Science*, Vol. 2 No.2, 2024.
3. P. Goswami et al. Investigation on Storage Level Data Integrity Strategies in Cloud Computing. *Journal of Cloud Computing*, 2024.
4. B. A. Jalil et al. A Survey on Data Integrity Verification Schemes in the Cloud. *AIP Conference Proceedings*, 2023.
5. Kashkevich S. (Ed.) (2025). Decision support systems: mathematical support. Kharkiv: TECHNOLOGY CENTER PC, 202. <https://doi.org/10.15587/978-617-8360-13-9>.
6. Owaid S. R., Kashkevich S., Shyshatskyi A., Radzivilov H., Sova O., Zarubenko A., Veretnov A., Lazuta R., Noskov O., Voznytsia A. Development of heterogeneous data processing method in organizational and technical systems. *Eastern-European Journal of Enterprise Technologies*, 2025. 1 (4 (133)), 64–71. <https://doi.org/10.15587/1729-4061.2025.322629>.
7. Owaid S. R., Miahkykh H., Odarushchenko E., Kashkevich S., Shyshatskyi A., Plekhova G., Hrymud A., Petruk S., Shaposhnikova O., Stryhun V. Development of a method for detecting cyber attacks on information systems based on artificial intelligence technologies. *Eastern-European Journal of Enterprise Technologies*, 2025. 3 (9 (135)), 33–39. <https://doi.org/10.15587/1729-4061.2025.329258>.
8. C. Shi A. Manschula T. Mahmud et al. Revisiting Computational Storage for Data Integrity and Security. arXiv preprint, 2025.

Дата надходження статті: 21.11.2025

Дата прийняття статті: 10.12.2025

Опубліковано: 30.12.2025